

Multivariable Calculus

Anderson Trimm

Gwinnett School of Mathematics, Science and Technology

These are the notes for the Spring Semester 2020 course in Multivariable Calculus at GSMST. They will continually be updated throughout the course. The latest PDF can always be accessed at https://github.com/atrimm/mvc/blob/master/Course%20Notes/multivariable_calculus_2020.pdf.

Contents

1	Curves	4
1.1	Vector-valued functions	4
1.1.1	Definitions	4
1.1.2	Review: limits of single-variable functions	6
1.1.3	Limits of vector-valued functions	12
1.1.4	Derivatives of vector-valued functions	16
1.2	Parametrized Curves	18
1.3	Reparametrizations	21
1.4	Arc Length	24
1.5	Curvature	29
1.6	Curvature of a unit speed curve	29
1.7	Osculating Plane	31
1.8	Arbitrary parametrizations	32
1.9	Plane curves	35
1.10	Local-global connection	38
1.11	Space Curves	41
1.12	The Fundamental Theorems of Plane and Space Curves	51
1.13	Summary of formulas	52
1.13.1	Plane Curves	52
1.13.2	Space Curves	52
2	Multivariable Functions	53
2.1	Basic Definitions	53
2.2	Visualizing Multivariable Functions	55
2.3	Limits	60
2.4	Partial Derivatives	66
2.4.1	Motivation	66
2.4.2	Definition of Partial Derivatives	68
2.4.3	Interpretation of Partial Derivatives	69
2.4.4	Functions of More than Two Variables	71
2.5	The Differential	72
2.5.1	Differential of a Single-variable Function	72
2.5.2	Differential of a Multivariable Function	73
2.5.3	Matrix Representation of the Differential	76
2.6	The Composite Mapping Theorem (Chain Rule)	78
2.7	Higher Derivatives	80

A	Justification of the Arc Length Formula	80
B	Topology of \mathbb{R}^n	85
B.1	Normed Linear Spaces	85
B.2	Metric Spaces	86
B.3	Continuous maps	88
C	Affine Space	89

Quick reference

1.1	Definition (Vector-valued function)	4
1.3	Definition (Limit of a single-variable function)	7
1.4	Theorem (Uniqueness of limits)	7
1.6	Lemma (Triangle Inequality)	8
1.7	Theorem (Limit laws for single-variable functions)	9
1.8	Definition (Continuity)	11
1.9	Theorem (New continuous functions from old)	11
1.10	Theorem (A composition of continuous functions is continuous)	11
1.12	Definition (Limit of a vector-valued function)	12
1.14	Theorem (Limit of a vector-valued function)	13
1.15	Corollary (Uniqueness of the limit of a vector-valued function)	13
1.17	Theorem (Limit laws for vector-valued functions)	14
1.18	Definition (Continuity)	15
1.19	Theorem (Continuity of vector-valued functions)	15
1.21	Theorem (New continuous vector-valued functions from old)	16
1.22	Theorem (Continuity of composite function)	16
1.23	Definition (Derivative of a vector-valued function)	16
1.24	Theorem (Derivative of a vector-valued function)	17
1.25	Theorem (Differentiation formulas)	17
1.26	Definition (Higher derivatives)	17
1.27	Definition (Parametrized curve)	18
1.31	Definition (Regular curve)	21
1.32	Definition (Reparametrization)	22
1.33	Proposition (Reparametrized curves are equivalent)	23
1.34	Definition (Curve)	23
1.35	Definition (Oriented curves)	24
1.39	Proposition (Invariance of arc length)	26
1.40	Proposition (Any regular curve can be parametrized by arc length)	27
1.45	Definition (Curvature)	30
1.47	Definition (Unit normal vector)	31
1.48	Definition (Osculating circle)	32
1.50	Definition (Curvature in an arbitrary parametrization)	34
1.51	Definition (Unit tangent and Unit normal in an arbitrary parametrization)	34
1.59	Theorem (Angle-Sum Theorem)	38
1.60	Definition (Simple closed curve)	38
1.62	Theorem (Total Curvature Theorem)	38

1.64	Definition (Rotation index)	41
1.68	Definition (Unit binormal vector)	43
1.70	Definition (Frenet frame)	43
1.71	Definition (Torsion for unit speed curve)	44
1.78	Theorem (Fundamental Theorem of the Local Theory of Curves)	51
2.5	Example (Elevation maps)	56
2.7	Definition (Level surfaces)	57
2.9	Definition (Limit)	60
2.13	Theorem (Limit laws for multivariable functions)	64
2.15	Definition (Continuous function)	64
2.17	Example (Heat index)	66
2.18	Definition (Partial derivative)	68
2.20	Definition (Differentiable function)	72
2.22	Definition (Differential)	73
2.23	Definition (Differentiable multivariable function)	73
2.24	Definition (Differential of a multivariable function)	74
2.25	Proposition (Uniqueness of the differential)	74
2.26	Theorem (A differentiable function is continuous)	75
2.27	Definition (Continuously differentiable function)	76
2.34	Theorem (Chain rule)	79
B.1	Definition (Normed linear space)	85
B.3	Definition (Continuous linear maps)	85
B.4	Proposition (Equivalent definition of a bounded linear maps)	85
B.5	Definition (Metric space)	86
B.8	Definition (Open ball)	86
B.10	Definition (Interior and boundary points)	86
B.11	Definition (Open and closed sets)	86
B.13	Definition (Neighborhood)	87
B.15	Definition (Limit points and isolated points)	87
B.21	Definition (Bounded sets)	88
B.23	Definition (Lipschitz continuity)	88
C.1	Definition (Affine space)	89
C.3	Definition (Affine map)	90
C.4	Proposition (Uniqueness of the differential)	90
C.5	Definition (Affine isomorphism)	91

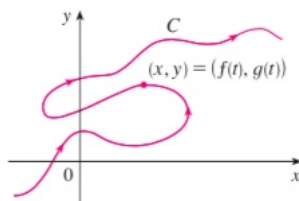
1 Curves

In this section we study functions with one input and multiple outputs.

1.1 Vector-valued functions

1.1.1 Definitions

Suppose a particle moves in the plane along the following curve C :



Since the curve fails the vertical line test, C cannot be described as the graph of a function $y = f(x)$. Note however that the x - and y -coords of the particle are functions of time

$$x = f(t), \quad y = g(t)$$

so the curve C can be described as the image of function $\mathbf{r} : I \rightarrow \mathbb{R}^2$ defined by

$$\mathbf{r}(t) = (f(t), g(t)),$$

where $I = [a, b]$ is an interval in \mathbb{R} . [\[Add mapping diagram.\]](#)

Definition 1.1 (Vector-valued function). Let $U \subseteq \mathbb{R}$. A mapping $\mathbf{r} : U \rightarrow \mathbb{R}^n$ called a *vector-valued function*. The value of \mathbf{r} at $t \in U$ can be written as

$$\mathbf{r}(t) = (r_1(t), r_2(t), \dots, r_n(t))$$

where the n functions $r_i : U \rightarrow \mathbb{R}, i = 1, \dots, n$ are called the *component functions* of \mathbf{r} .

Unless specified otherwise, we will take the domain U of a vector-valued function to be the largest domain on which all of the component functions are defined.

Example 1.2. Consider the vector-valued function $\mathbf{r} : U \rightarrow \mathbb{R}^3$ defined by

$$\mathbf{r}(t) = (t^3, \ln(3 - t), \sqrt{t}).$$

The component functions of $\mathbf{r}(t)$ are

$$r_1(t) = t^3, \quad r_2(t) = \ln(3 - t), \quad r_3(t) = \sqrt{t}.$$

The domains of each of these functions, respectively, are

$$U_1 = \mathbb{R}, \quad U_2 = (-\infty, 3), \quad U_3 = [0, \infty),$$

so the domain U of $\mathbf{r}(t)$ is

$$U = U_1 \cap U_2 \cap U_3 = [0, 3).$$

Exercise 1.1. Consider the vector-valued function $\mathbf{r} : U \rightarrow \mathbb{R}^3$ defined by

$$\mathbf{r}(t) = \left(\frac{t-2}{t+2}, \sin t, \ln(9-t^2) \right).$$

What is the domain U of the function?

Solution. The component functions of $\mathbf{r}(t)$ are

$$r_1 = \frac{t-2}{t+2}, \quad r_2 = \sin t, \quad r_3 = \ln(9-t^2).$$

The domains of $r_1(t)$ and $r_2(t)$ are given, respectively, by

$$U_1 = (-\infty, 2) \cup (2, \infty), \quad U_2 = \mathbb{R}.$$

To find the domain of $r_3(t)$, we need to solve the inequality

$$9 - t^2 > 0.$$

The graph of the function $y = 9 - x^2$ is a concave-down parabola with y -intercept 9 and x -intercepts ± 3 .

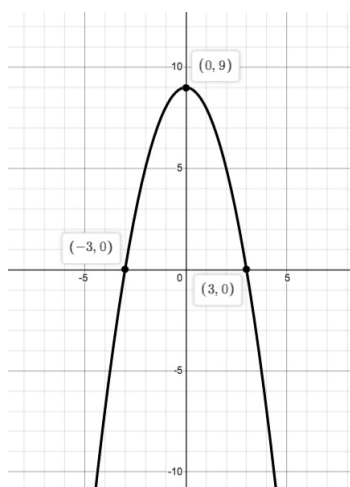


Figure 1: Graph of $y = 9 - x^2$.

We have $y > 0$ where the graph is above the x -axis, so $y > 0$ when $-3 < x < 3$. The domain of $r_3(t)$ is therefore

$$U_3 = (-3, 3).$$

The domain of $\mathbf{r}(t)$ is then

$$U = U_1 \cap U_2 \cap U_3 = (-3, -2) \cup (-2, 3).$$

□

1.1.2 Review: limits of single-variable functions

Before considering limits of vector-valued functions, let's review the definition for a real-valued function $y = f(x)$ of a single real variable x .

To motivate the definition, consider the function

$$f(x) = \begin{cases} 2x - 1, & \text{if } x \neq 3 \\ 6, & \text{if } x = 3 \end{cases}$$

whose graph is shown in the figure below.

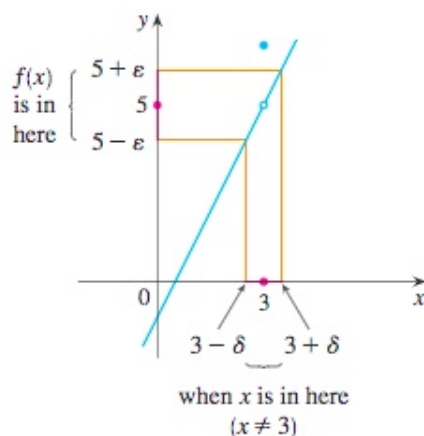


Figure 2: Graph of the function $y = f(x)$ in the example above.

From the graph, we see that when x is close to 3 but not equal to 3, then $f(x)$ is close to 5, and so $\lim_{x \rightarrow 3} f(x) = 5$.

To obtain more detailed information about how $f(x)$ varies when x is close to 3, we ask the following question:

How close to 3 does x have to be so that $f(x)$ differs from 5 by less than 0.1?

The distance from x to 3 is $|x - 3|$ and the distance from $f(x)$ to 5 is $|f(x) - 5|$, so our problem is to find a number δ such that

$$|f(x) - 5| < 0.1 \quad \text{if} \quad 0 < |x - 3| < \delta.$$

If $x \neq 3$, then

$$|f(x) - 5| = |(2x - 1) - 5| = |2x - 6| = 2|x - 3|$$

so we see that by taking $\delta = \frac{1}{2}(0.1) = 0.05$, we have $|f(x) - 5| < 2(0.05) = 0.1$. Thus, an answer to the problem is given by $\delta = 0.05$; that is, if x is within a distance of 0.05 from 3, then $f(x)$ will be within a distance of 0.1 from 5.

If we change the number 0.1 in our problem to the smaller number 0.01, then by using the same method we find that $f(x)$ will differ from 5 by less than 0.01 provided that x differs from 3 by less than $\frac{1}{2}(0.01) = 0.005$; that is,

$$|f(x) - 5| < 0.01 \quad \text{if} \quad 0 < |x - 3| < 0.005.$$

Similarly,

$$|f(x) - 5| < 0.001 \quad \text{if} \quad 0 < |x - 3| < 0.0005.$$

Think of the numbers 0.1, 0.01, 0.001 above as *error tolerances* that we might allow. That is, when challenged with an error tolerance, it is our task to find a corresponding δ so that whenever x is within a distance of δ from 3, $f(x) \approx 5$, within the given error tolerance.

Now for 5 to be the precise limit of $f(x)$ as x approaches 3, we must not only be able to bring the difference between $f(x)$ and 5 below each of these numbers; we must be able to bring it below *any* positive number. And, by exactly the same reasoning, we can. That is, if ϵ is any positive number, then by choosing $\delta = \frac{\epsilon}{2}$, we find

$$|f(x) - 5| < \epsilon \quad \text{if} \quad 0 < |x - 3| < \delta = \frac{\epsilon}{2}. \quad (1.1)$$

This is a precise way of saying that $f(x)$ is close to 5 when x is close to 3, because Equation (1.1) says that we can make the values of $f(x)$ within an arbitrary distance ϵ from 5 by taking the values of x within a distance $\frac{\epsilon}{2}$ from 3 (but $x \neq 3$).

Note that Equation (1.1) can be rewritten as follows:

$$\text{if} \quad 3 - \delta < x < 3 + \delta \quad (x \neq 3) \quad \text{then} \quad 5 - \epsilon < f(x) < 5 + \epsilon$$

as illustrated in the figure above. This says that by taking the values of x ($x \neq 3$) to lie in the interval $(3 - \delta, 3 + \delta)$ we can make the values of $f(x)$ lie in the interval $(5 - \epsilon, 5 + \epsilon)$.

Following the reasoning in this example, the precise definition of a limit is the following.

Definition 1.3 (Limit of a single-variable function). Let (a, b) be an open interval containing the point x_0 and let $f(x)$ be a real-valued function defined on this interval, except possibly at x_0 itself. A number L is called the *limit of $f(x)$ as x approaches x_0* if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $0 < |x - x_0| < \delta$. If such an L exists, we write

$$\lim_{x \rightarrow x_0} f(x) = L.$$

Theorem 1.4 (Uniqueness of limits). If $f(x)$ has a limit L at x_0 , then the limit is unique.

Proof. Suppose that $\lim_{x \rightarrow x_0} f(x) = L$ and $\lim_{x \rightarrow x_0} f(x) = L'$. Then, given any $\epsilon > 0$ there exist positive numbers δ_1 and δ_2 such that

$$|f(x) - L| < \frac{\epsilon}{2} \quad \text{if} \quad |x - x_0| < \delta_1$$

and

$$|f(x) - L'| < \frac{\epsilon}{2} \quad \text{if} \quad |x - x_0| < \delta_2.$$

Then by taking $|x - x_0| < \delta = \min\{\delta_1, \delta_2\}$, we have

$$|L - L'| = |L - L' + f(x) - f(x)| = |L - f(x) + f(x) - L'| \leq |f(x) - L| + |f(x) - L'| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

For this to be true for all $\epsilon > 0$, we must have $L - L' = 0$, or $L = L'$. \square

Exercise 1.2. Use Definition 1.3 to prove that $\lim_{x \rightarrow 3} (4x - 5) = 7$.

Solution. Let $\epsilon > 0$. For all $x \neq 3$,

$$|f(x) - 7| = |(4x - 5) - 7| = |4x - 12| = 4|x - 3|.$$

By taking $\delta = \frac{\epsilon}{4}$, we have $0 < |x - 3| < \frac{\epsilon}{4}$ and therefore

$$|f(x) - 7| = 4|x - 3| < 4 \cdot \frac{\epsilon}{4} = \epsilon,$$

which proves that $\lim_{x \rightarrow 3} (4x - 5) = 7$. □

Example 1.5. We now use Definition 1.3 to prove that $\lim_{x \rightarrow 3} x^2 = 9$.

Let $\epsilon > 0$. For all $x \neq 3$, we have

$$|f(x) - 9| = |x^2 - 9| = |(x + 3)(x - 3)| = |x + 3||x - 3|.$$

Notice that if we can find a positive number C such that $|x + 3| < C$, then

$$|x + 3||x - 3| < C|x - 3|$$

and we can make $C|x - 3| < \epsilon$ by taking $|x - 3| < \frac{\epsilon}{C} = \delta$. We can find such a number C if we restrict x to lie in some interval centered at 3. Since we are only interested in values of x that are close to 3, this is exactly what we want. Let's assume that $|x - 3| < \alpha$ for some positive number α , say $\alpha = 1$ (it does not matter what number we take here). Then we have

$$x - 3 < 1 \quad \text{or} \quad -x + 3 < 1.$$

The first inequality says $x < 4$ and the second says $2 < x$, so $|x - 3| < 1$ implies that

$$2 < x < 4.$$

Adding 3 to both sides of this inequality gives

$$5 < x + 3 < 7,$$

and therefore $|x + 3| < |7| = 7 = C$. But now there are two restrictions on $|x - 3|$, namely

$$|x - 3| < 1 \quad \text{and} \quad |x - 3| < \frac{\epsilon}{C} = \frac{\epsilon}{7}.$$

To make sure that both of these inequalities are satisfied, we take $\delta = \min\{1, \frac{\epsilon}{7}\}$. Since $0 < |x - 3| < \delta$ implies $|x^2 - 9| < \epsilon$, this proves that $\lim_{x \rightarrow 3} x^2 = 9$.

The previous example shows that it is not always easy to prove that a function has a particular limit using Definition 1.3. In fact, if we had considered a more complicated function such as

$$f(x) = \frac{6x^2 - 8x + 9}{2x^2 - 1}$$

then proving that $\lim_{x \rightarrow 1} f(x) = 7$ using Definition 1.3 would require a great deal of ingenuity. Instead, we prove the following theorems, which makes evaluating limits much easier.

Lemma 1.6 (Triangle Inequality). For all $x, y \in \mathbb{R}$,

$$|x + y| \leq |x| + |y|.$$

Proof. We have

$$\begin{aligned} |x + y|^2 &= (x + y)^2 = x^2 + y^2 + 2xy \\ &= |x|^2 + |y|^2 + 2xy \\ &\leq |x|^2 + |y|^2 + 2|x||y| \\ &= (|x| + |y|)^2. \end{aligned}$$

Since both sides are nonnegative, this implies that

$$|x + y| \leq |x| + |y|.$$

□

Theorem 1.7 (Limit laws for single-variable functions). Suppose $f(x)$ and $g(x)$ are defined on the same open set containing x_0 , and that

$$\lim_{x \rightarrow x_0} f(x) = L \quad \text{and} \quad \lim_{x \rightarrow x_0} g(x) = M.$$

Then

- (i) $\lim_{x \rightarrow x_0} c = c$ for any constant $c \in \mathbb{R}$.
- (ii) $\lim_{x \rightarrow x_0} x = x_0$.
- (iii) $\lim_{x \rightarrow x_0} cf(x) = cL$ for any $c \in \mathbb{R}$;
- (iv) $\lim_{x \rightarrow x_0} (f(x) + g(x)) = L + M$;
- (v) $\lim_{x \rightarrow x_0} (f(x)g(x)) = LM$;
- (vi) $\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \frac{L}{M}$ whenever $M \neq 0$.

Proof. (i) Let $\epsilon > 0$. Since $|c - c| = 0$, $|c - c| < \epsilon$ whenever $|x - x_0| < \delta$ for any positive number δ .

(ii) Given $\epsilon > 0$, by taking $\delta = \epsilon$ we have $|x - x_0| < \epsilon$ whenever $|x - x_0| < \delta = \epsilon$.

(iii) Since $\lim_{x \rightarrow x_0} f(x) = L$, given $\epsilon > 0$ there exists a corresponding $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $0 < |x - x_0| < \delta$. Then $|cf(x) - cL| = |c||f(x) - L| < \epsilon$ whenever $0 < |x - x_0| < \frac{\epsilon}{|c|}$.

(iv) We have

$$|f(x) + g(x) - (L + M)| = |(f(x) - L) + (g(x) - M)| \leq |f(x) - L| + |g(x) - M|$$

by the Triangle Inequality (Lemma 1.6). Since $\lim_{x \rightarrow x_0} f(x) = L$ and $\lim_{x \rightarrow x_0} g(x) = M$, given $\epsilon > 0$ there exist positive numbers δ_1 and δ_2 such that

$$|f(x) - L| < \frac{\epsilon}{2} \quad \text{if} \quad |x - x_0| < \delta_1$$

and

$$|g(x) - M| < \frac{\epsilon}{2} \quad \text{if} \quad |x - x_0| < \delta_2.$$

By taking $|x - x_0| < \delta = \min\{\delta_1, \delta_2\}$, we have

$$|f(x) + g(x) - (L + M)| \leq |f(x) - L| + |g(x) - M| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which proves that $\lim_{x \rightarrow x_0} (f(x) + g(x)) = L + M$.

(v) First, note that

$$f(x)g(x) - LM = (f(x) - L)(g(x) - M) + L(g(x) - M) + M(f(x) - L).$$

Let $\epsilon > 0$. Since $\lim_{x \rightarrow x_0} f(x) = L$ there exists $\delta_1 > 0$ such that $|f(x) - L| < \sqrt{\epsilon}$ whenever $|x - x_0| < \delta_1$. Since $\lim_{x \rightarrow x_0} g(x) = M$ there exists $\delta_2 > 0$ such that $|g(x) - M| < \sqrt{\epsilon}$ whenever $|x - x_0| < \delta_2$. Then, whenever $|x - x_0| < \delta = \min\{\delta_1, \delta_2\}$, we have

$$|(f(x) - L)(g(x) - M)| = |f(x) - L||g(x) - M| < (\sqrt{\epsilon})^2 = \epsilon$$

which shows that $\lim_{x \rightarrow x_0} (f(x) - L)(g(x) - M) = 0$. By (iii),

$$\lim_{x \rightarrow x_0} L(g(x) - M) = L \lim_{x \rightarrow x_0} (g(x) - M) = L \cdot 0 = 0,$$

and

$$\lim_{x \rightarrow x_0} M(f(x) - L) = M \lim_{x \rightarrow x_0} (f(x) - L) = M \cdot 0 = 0.$$

Applying (iv),

$$\begin{aligned} \lim_{x \rightarrow x_0} (f(x)g(x) - LM) &= \lim_{x \rightarrow x_0} (f(x) - L)(g(x) - M) + \lim_{x \rightarrow x_0} L(g(x) - M) + \lim_{x \rightarrow x_0} M(f(x) - L) \\ &= 0 + 0 + 0 \\ &= 0, \end{aligned}$$

and therefore

$$\lim_{x \rightarrow x_0} f(x)g(x) = LM.$$

(vi) First, note that since $|M| > 0$ and $\lim_{x \rightarrow x_0} g(x) = M$, there exists $\delta_1 > 0$ such that $|g(x)| > \frac{1}{2}|M|$ whenever $|x - x_0| < \delta_1$ [Draw a picture.]. Let $\epsilon > 0$. Choose $\delta_2 > 0$ such that $|x - x_0| < \delta_2$ implies that $|g(x) - M| < \frac{1}{2}|M|^2\epsilon$. Then, for $|x - x_0| < \delta = \min\{\delta_1, \delta_2\}$, we have

$$\begin{aligned} \left| \frac{1}{g(x)} - \frac{1}{M} \right| &= \left| \frac{M - g(x)}{Mg(x)} \right| \\ &= \frac{|g(x) - M|}{|Mg(x)|} \\ &< \frac{\frac{1}{2}|M|^2\epsilon}{\frac{1}{2}|M|^2} \\ &= \epsilon, \end{aligned}$$

and therefore $\lim_{x \rightarrow x_0} \frac{1}{g(x)} = \frac{1}{M}$. It then follows from (v) that

$$\begin{aligned} \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} &= \lim_{x \rightarrow x_0} f(x) \lim_{x \rightarrow x_0} \frac{1}{g(x)} \\ &= \frac{L}{M}. \end{aligned}$$

□

Using Theorem 1.7, it is much easier to prove the limits in the examples above. For instance

$$\begin{aligned}\lim_{x \rightarrow 3}(4x - 5) &= (\lim_{x \rightarrow 3} 4)(\lim_{x \rightarrow 3} x) + (\lim_{x \rightarrow 3} (-5)) \\ &= 4(3) + (-5) \\ &= 12 - 5 \\ &= 7,\end{aligned}$$

and

$$\lim_{x \rightarrow 3} x^2 = (\lim_{x \rightarrow 3} x)(\lim_{x \rightarrow 3} x) = (3)(3) = 9.$$

Note that, in both of these examples, the function $f(x)$ is actually defined at x_0 and $\lim_{x \rightarrow x_0} f(x) = f(x_0)$; that is, the limit of $f(x)$ as x approaches x_0 is equal to the value of $f(x)$ at x_0 .

Definition 1.8 (Continuity). Let $f(x)$ be defined on an open interval (a, b) containing a point x_0 . We say that $f(x)$ is *continuous at x_0* if $\lim_{x \rightarrow x_0} f(x) = f(x_0)$. We then say that $f(x)$ is *continuous on (a, b)* if $f(x)$ is continuous at every point in (a, b) .

The limit laws in Theorem 1.7 imply that

- Polynomials are continuous on \mathbb{R} ;
- Rational functions are continuous wherever they are defined;
- The absolute value function $f(x) = |x|$ is continuous;

Trig functions, and exponential and logarithmic functions are all also continuous wherever they are defined.

Exercise 1.3. Prove that $f(x) = |x|$ is continuous on \mathbb{R} .

Solution. If $x > 0$, then $f(x) = x$ which is continuous since it is a polynomial. The same is true for $x < 0$ since then $f(x) = -x$. By taking $\delta = \epsilon$, $|f(x) - 0| = ||x|| = |x| < \epsilon$ whenever $|x| < \delta = \epsilon$, so $\lim_{x \rightarrow 0} f(x) = 0 = f(0)$, which shows that $f(x)$ is also continuous at $x = 0$. Thus, $f(x)$ is continuous on \mathbb{R} . \square

Theorem 1.9 (New continuous functions from old). Let $f(x)$ and $g(x)$ be defined on the same open interval containing x_0 . If $f(x)$ and $g(x)$ are continuous at x_0 , then so are

- (i) $cf(x)$
- (ii) [Finish.]

Theorem 1.10 (A composition of continuous functions is continuous). Suppose $f(x)$ is defined on an open interval containing x_0 and $g(x)$ is defined on an open interval containing $f(x_0)$. If f is continuous at x_0 and $g(x)$ is continuous at $f(x_0)$, then $(g \circ f)(x)$ is continuous at x_0 .

Proof. Let $\epsilon > 0$. Since g is continuous at $f(x_0)$, corresponding to ϵ there exists $\eta > 0$ such that $|g(f(x)) - g(f(x_0))| < \epsilon$ whenever $|f(x) - f(x_0)| < \eta$. Since f is continuous at x_0 , corresponding to η there exists $\delta > 0$ such that $|f(x) - f(x_0)| < \eta$ whenever $|x - x_0| < \delta$. This shows that $|g(f(x)) - g(f(x_0))| < \epsilon$ whenever $|x - x_0| < \delta$, proving that $(g \circ f)(x)$ is continuous at x_0 . \square

Example 1.11. Consider the function $f(x) = e^{x^2}$. We can view $f(x)$ as the composition $(h \circ g)(x)$, where $h(x) = e^x$ and $g(x) = x^2$. Since $h(x)$ and $g(x)$ are continuous on \mathbb{R} , by Theorem 1.10 so is $f(x)$.

1.1.3 Limits of vector-valued functions

Throughout this section, let $I = (a, b)$ denote an open interval in \mathbb{R} containing a point t_0 , and let $\mathbf{r}(t)$ be a vector-valued function defined on I , except perhaps at t_0 itself.

Definition 1.12 (Limit of a vector-valued function). A fixed vector $\mathbf{L} \in \mathbb{R}^n$ is said to be the *limit* as $\mathbf{r}(t)$ approaches t_0 if for every $\epsilon > 0$ there exists a corresponding $\delta > 0$ such that

$$0 < |t - t_0| < \delta \implies \|\mathbf{r}(t) - \mathbf{L}\| < \epsilon.$$

If \mathbf{L} exists, we write $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{L}$.

We will now show that the limit of a vector-valued function can be computed in terms of the limits of its component functions. We will need the following lemma.

Lemma 1.13. Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$ be vectors in \mathbb{R}^n . Then

$$|x_i - y_i| \leq \|\mathbf{x} - \mathbf{y}\| \leq \sum_{i=1}^n |x_i - y_i|$$

for all $i = 1, 2, \dots, n$.

Proof. For any fixed index i , we have

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\|^2 &= \sum_{i=1}^n (x_i - y_i)^2 \\ &= (x_i - y_i)^2 + \underbrace{\sum_{j \neq i} (x_j - y_j)^2}_{\geq 0} \\ &\geq (x_i - y_i)^2. \end{aligned}$$

Since both sides are nonnegative, this implies that

$$\|\mathbf{x} - \mathbf{y}\| \geq \sqrt{(x_i - y_i)^2} = |x_i - y_i|,$$

so the first inequality holds.

To see that the second inequality holds, note that

$$\begin{aligned} \left(\sum_{i=1}^n |x_i - y_i| \right)^2 &= \sum_{i=1}^n |x_i - y_i|^2 + 2 \underbrace{\sum_{1 \leq i < j \leq n} |x_i - y_i| |x_j - y_j|}_{\geq 0} \\ &\geq \sum_{i=1}^n |x_i - y_i|^2 \\ &= \sum_{i=1}^n (x_i - y_i)^2 \\ &= \|\mathbf{x} - \mathbf{y}\|^2. \end{aligned}$$

Since both sides are nonnegative, this implies that

$$\sum_{i=1}^n |x_i - y_i| \geq \|\mathbf{x} - \mathbf{y}\|,$$

so the second inequality holds. □

Exercise 1.4. Verify that

$$\left(\sum_{i=1}^n |x_i - y_i| \right)^2 = \sum_{i=1}^n |x_i - y_i|^2 + 2 \sum_{1 \leq i < j \leq n} |x_i - y_i| |x_j - y_j|$$

for $n = 3$ by explicitly writing out both sides.

Theorem 1.14 (Limit of a vector-valued function). Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a vector-valued function. Then

$$\lim_{t \rightarrow t_0} \mathbf{r}(t) = (\lim_{t \rightarrow t_0} r_1(t), \lim_{t \rightarrow t_0} r_2(t), \lim_{t \rightarrow t_0} r_3(t)). \quad (1.2)$$

Proof. Let $\mathbf{L} = (L_1, L_2, \dots, L_n)$ be a fixed vector in \mathbb{R}^n . We will prove that $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{L}$ if and only if $\lim_{t \rightarrow t_0} r_i(t) = L_i$ for all $i = 1, \dots, n$; that is, both sides of Equation (1.2) are either undefined, or they are both equal to \mathbf{L} and hence to each other.

(\Rightarrow) First, suppose that $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{L}$. Then, given $\epsilon > 0$, there exists $\delta > 0$ such that $\|\mathbf{r}(t) - \mathbf{L}\| < \epsilon$ whenever $0 < |t - t_0| < \delta$. By Lemma 1.13, for each $i = 1, \dots, n$

$$|r_i(t) - L_i| < \|\mathbf{r}(t) - \mathbf{L}\|$$

so we have $|r_i(t) - L_i| < \epsilon$ for each $i = 1, \dots, n$ whenever $0 < |t - t_0| < \delta$. Thus, $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{L}$ implies that $\lim_{t \rightarrow t_0} r_i(t) = L_i$ for all $i = 1, \dots, n$.

(\Leftarrow) Now suppose that $\lim_{t \rightarrow t_0} r_i(t) = L_i$ for all $i = 1, \dots, n$. Given $\epsilon > 0$, there exist positive numbers $\delta_1, \delta_2, \dots, \delta_n$ such that $|r_i(t) - L_i| < \frac{\epsilon}{n}$ whenever $0 < |t - t_0| < \delta_i$. By Lemma 1.13,

$$\|\mathbf{r}(t) - \mathbf{L}\| < \sum_{i=1}^n |r_i(t) - L_i|,$$

so by taking $\delta = \min\{\delta_1, \delta_2, \dots, \delta_n\}$, we have

$$\|\mathbf{r}(t) - \mathbf{L}\| < \sum_{i=1}^n |r_i(t) - L_i| < \epsilon$$

whenever $|t - t_0| < \delta$. Thus, $\lim_{t \rightarrow t_0} r_i(t) = L_i$ for all $i = 1, \dots, n$ implies that $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{L}$. \square

Corollary 1.15 (Uniqueness of the limit of a vector-valued function). If $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{L}$, then the limit is unique.

Proof. Since the limits $\lim_{t \rightarrow t_0} r_i(t) = L_i$ are unique (if they exist) by Theorem 1.4, it follows immediately from Theorem 1.14 that $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{L}$ is unique if it exists. \square

Example 1.16. Let $\mathbf{r}(t) = (1 + t^3, te^{-t}, \frac{\sin t}{t})$. Since

$$\lim_{t \rightarrow 0} (1 + t^3) = 1,$$

$$\lim_{t \rightarrow 0} te^{-t} = \lim_{t \rightarrow 0} t \lim_{t \rightarrow 0} e^{-t} = 0 \cdot 1 = 0,$$

$$\lim_{t \rightarrow 0} \frac{\sin t}{t} = \lim_{t \rightarrow 0} \cos t = 1 \quad (\text{by L'Hospital's rule})$$

by Theorem 1.14

$$\begin{aligned} \lim_{t \rightarrow 0} \mathbf{r}(t) &= \left(\lim_{t \rightarrow 0} (1 + t^3), \lim_{t \rightarrow 0} te^{-t}, \lim_{t \rightarrow 0} \frac{\sin t}{t} \right) \\ &= (1, 0, 1). \end{aligned}$$

Exercise 1.5. Find $\lim_{t \rightarrow 1} \mathbf{r}(t)$, where $\mathbf{r}(t) = \left(\frac{t^2-t}{t-1}, \sqrt{t+8}, \frac{\sin(\pi t)}{\ln(t)} \right)$, if it exists.

Solution. Since

$$\begin{aligned} \lim_{t \rightarrow 1} \frac{t^2-t}{t-1} &= \lim_{t \rightarrow 1} \frac{t(t-1)}{t-1} = \lim_{t \rightarrow 1} t = 1, \\ \lim_{t \rightarrow 1} \sqrt{t+8} &= \sqrt{1+8} = \sqrt{9} = 3, \\ \lim_{t \rightarrow 1} \frac{\sin(\pi t)}{\ln(t)} &= \lim_{t \rightarrow 1} \frac{\pi \cos(\pi t)}{\frac{1}{t}} = \lim_{t \rightarrow 1} \pi t \cos(\pi t) = \pi(1) \cos(\pi) = -\pi, \end{aligned}$$

by Theorem 1.14

$$\begin{aligned} \lim_{t \rightarrow 1} \mathbf{r}(t) &= \left(\lim_{t \rightarrow 1} \frac{t^2-t}{t-1}, \lim_{t \rightarrow 1} \sqrt{t+8}, \lim_{t \rightarrow 1} \frac{\sin(\pi t)}{\ln(t)} \right) \\ &= (1, 3, -\pi). \end{aligned}$$

□

Theorem 1.17 (Limit laws for vector-valued functions). Let \mathbf{u}, \mathbf{v} be vector valued functions into \mathbb{R}^n defined on the same open interval containing t_0 and let $c \in \mathbb{R}$ be a constant. Then

- (i) $\lim_{t \rightarrow t_0} (c_1, c_2, \dots, c_n) = (c_1, c_2, \dots, c_n)$ if (c_1, c_2, \dots, c_n) is a constant vector in \mathbb{R}^n .
- (ii) $\lim_{t \rightarrow t_0} c\mathbf{u}(t) = c \lim_{t \rightarrow t_0} \mathbf{u}(t)$
- (iii) $\lim_{t \rightarrow t_0} [\mathbf{u}(t) + \mathbf{v}(t)] = \lim_{t \rightarrow t_0} \mathbf{u}(t) + \lim_{t \rightarrow t_0} \mathbf{v}(t)$
- (iv) $\lim_{t \rightarrow t_0} [\mathbf{u}(t) \cdot \mathbf{v}(t)] = \lim_{t \rightarrow t_0} \mathbf{u}(t) \cdot \lim_{t \rightarrow t_0} \mathbf{v}(t)$
- (v) $\lim_{t \rightarrow t_0} [\mathbf{u}(t) \times \mathbf{v}(t)] = \lim_{t \rightarrow t_0} \mathbf{u}(t) \times \lim_{t \rightarrow t_0} \mathbf{v}(t)$ (for $n = 3$)

Proof. The proof of each of these follows by applying Theorems 1.7 and 1.14.

- (i) If (c_1, c_2, \dots, c_n) is a constant vector in \mathbb{R}^n , then

$$\lim_{t \rightarrow t_0} (c_1, c_2, \dots, c_n) = \left(\lim_{t \rightarrow t_0} c_1, \lim_{t \rightarrow t_0} c_2, \dots, \lim_{t \rightarrow t_0} c_n \right) = (c_1, c_2, \dots, c_n).$$

- (ii) If $c \in \mathbb{R}$ is a constant, then

$$\begin{aligned} \lim_{t \rightarrow t_0} c\mathbf{u}(t) &= \lim_{t \rightarrow t_0} c(u_1(t), u_2(t), \dots, u_n(t)) \\ &= \lim_{t \rightarrow t_0} (cu_1(t), cu_2(t), \dots, cu_n(t)) \\ &= \left(\lim_{t \rightarrow t_0} cu_1(t), \lim_{t \rightarrow t_0} cu_2(t), \dots, \lim_{t \rightarrow t_0} cu_n(t) \right) \\ &= \left(c \lim_{t \rightarrow t_0} u_1(t), c \lim_{t \rightarrow t_0} u_2(t), \dots, c \lim_{t \rightarrow t_0} u_n(t) \right) \\ &= c \left(\lim_{t \rightarrow t_0} u_1(t), \lim_{t \rightarrow t_0} u_2(t), \dots, \lim_{t \rightarrow t_0} u_n(t) \right) \\ &= c \lim_{t \rightarrow t_0} (u_1(t), u_2(t), \dots, u_n(t)) \\ &= c \lim_{t \rightarrow t_0} \mathbf{u}(t). \end{aligned}$$

(iii) If $\mathbf{u}(t), \mathbf{v}(t)$ are vector-valued functions into \mathbb{R}^n , then

$$\begin{aligned}
 \lim_{t \rightarrow t_0} [\mathbf{u}(t) + \mathbf{v}(t)] &= \lim_{t \rightarrow t_0} [(u_1(t), \dots, u_n(t)) + (v_1(t), \dots, v_n(t))] \\
 &= \lim_{t \rightarrow t_0} (u_1(t) + v_1(t), \dots, u_n(t) + v_n(t)) \\
 &= (\lim_{t \rightarrow t_0} (u_1(t) + v_1(t)), \dots, \lim_{t \rightarrow t_0} (u_n(t) + v_n(t))) \\
 &= (\lim_{t \rightarrow t_0} u_1(t) + \lim_{t \rightarrow t_0} v_1(t), \dots, \lim_{t \rightarrow t_0} u_n(t) + \lim_{t \rightarrow t_0} v_n(t)) \\
 &= (\lim_{t \rightarrow t_0} u_1(t), \lim_{t \rightarrow t_0} u_2(t), \lim_{t \rightarrow t_0} u_3(t)) + (\lim_{t \rightarrow t_0} v_1(t), \lim_{t \rightarrow t_0} v_2(t), \lim_{t \rightarrow t_0} v_3(t)) \\
 &= \lim_{t \rightarrow t_0} (u_1(t), u_2(t), u_3(t)) + \lim_{t \rightarrow t_0} (v_1(t), v_2(t), v_3(t)) \\
 &= \lim_{t \rightarrow t_0} \mathbf{u}(t) + \lim_{t \rightarrow t_0} \mathbf{v}(t).
 \end{aligned}$$

(iv) If $\mathbf{u}(t), \mathbf{v}(t)$ are vector-valued functions into \mathbb{R}^n , then

$$\begin{aligned}
 \lim_{t \rightarrow t_0} [\mathbf{u}(t) \cdot \mathbf{v}(t)] &= \lim_{t \rightarrow t_0} \sum_{i=1}^n u_i(t) v_i(t) \\
 &= \sum_{i=1}^n \lim_{t \rightarrow t_0} u_i(t) v_i(t) \\
 &= \sum_{i=1}^n \lim_{t \rightarrow t_0} u_i(t) \lim_{t \rightarrow t_0} v_i(t) \\
 &= \lim_{t \rightarrow t_0} \mathbf{u}(t) \cdot \lim_{t \rightarrow t_0} \mathbf{v}(t).
 \end{aligned}$$

(v) If \mathbf{u}, \mathbf{v} are vector-valued functions into \mathbb{R}^3 , then

$$\begin{aligned}
 \lim_{t \rightarrow t_0} [\mathbf{u}(t) \times \mathbf{v}(t)] &= \lim_{t \rightarrow t_0} (u_2(t)v_3(t) - u_3(t)v_2(t), -u_1(t)v_3(t) + u_3(t)v_1(t), u_1(t)v_2(t) - u_2(t)v_1(t)) \\
 &= (\lim_{t \rightarrow t_0} (u_2(t)v_3(t) - u_3(t)v_2(t)), \lim_{t \rightarrow t_0} (-u_1(t)v_3(t) + u_3(t)v_1(t)), \lim_{t \rightarrow t_0} (u_1(t)v_2(t) - u_2(t)v_1(t))) \\
 &= (\lim_{t \rightarrow t_0} u_2(t) \lim_{t \rightarrow t_0} v_3(t) - \lim_{t \rightarrow t_0} u_3(t) \lim_{t \rightarrow t_0} v_2(t), -\lim_{t \rightarrow t_0} u_1(t) \lim_{t \rightarrow t_0} v_3(t) + \lim_{t \rightarrow t_0} u_3(t) \lim_{t \rightarrow t_0} v_1(t), \\
 &\quad \lim_{t \rightarrow t_0} u_1(t) \lim_{t \rightarrow t_0} v_2(t) - \lim_{t \rightarrow t_0} u_2(t) \lim_{t \rightarrow t_0} v_1(t)) \\
 &= \lim_{t \rightarrow t_0} \mathbf{u}(t) \times \lim_{t \rightarrow t_0} \mathbf{v}(t).
 \end{aligned}$$

□

[Add examples.]

Definition 1.18 (Continuity). Let $\mathbf{r}(t)$ be defined on an open interval (a, b) containing a point t_0 . We say that $\mathbf{r}(t)$ is *continuous at* t_0 if $\lim_{t \rightarrow t_0} \mathbf{r}(t) = \mathbf{r}(t_0)$. We then say that $\mathbf{r}(t)$ is *continuous on* (a, b) if $\mathbf{r}(t)$ is continuous at every point in (a, b) .

Theorem 1.19 (Continuity of vector-valued functions). A vector-valued function $\mathbf{r}(t) = (r_1(t), \dots, r_n(t))$ is continuous at t_0 if and only if its component functions are all continuous at t_0 .

Proof. This follows immediately from Theorem 1.14.

□

Example 1.20. The vector-valued function $\mathbf{r}(t) = (\cos t, \sin t, t)$ is continuous on \mathbb{R} since its component functions are each continuous on \mathbb{R} .

Theorem 1.21 (New continuous vector-valued functions from old). Let \mathbf{u} and \mathbf{v} be two vector-valued functions on \mathbb{R}^n which are continuous at t_0 and let c be a constant. Then the following functions are also continuous at t_0 :

- (i) $c\mathbf{u}(t)$
- (ii) $\mathbf{u}(t) + \mathbf{v}(t)$
- (iii) $\mathbf{u}(t) \cdot \mathbf{v}(t)$
- (iv) $\mathbf{u}(t) \times \mathbf{v}(t)$ (for $n = 3$)

Proof. The proof follows immediately from Theorem 1.17. □

Theorem 1.22 (Continuity of composite function). Let $\mathbf{r} : \mathbb{R} \rightarrow \mathbb{R}^3$ be a continuous vector-valued function and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ a continuous real-valued function. Then the composite function $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi : \mathbb{R} \rightarrow \mathbb{R}^3$ is continuous.

Proof. Since $\mathbf{r}(t)$ is continuous, given $\epsilon > 0$ there exists $\eta > 0$ such that $\|\mathbf{r}(\varphi(t)) - \mathbf{r}(\varphi(t_0))\| < \epsilon$ whenever $|\varphi(t) - \varphi(t_0)| < \eta$. Since $\varphi(t)$ is continuous, corresponding to η there exists $\delta > 0$ such that $|\varphi(t) - \varphi(t_0)| < \eta$ whenever $|t - t_0| < \delta$. Thus, given $\epsilon > 0$, there exists $\delta > 0$ such that $\|\mathbf{r}(\varphi(t)) - \mathbf{r}(\varphi(t_0))\| < \epsilon$ whenever $|t - t_0| < \delta$. □

[Add example.]

1.1.4 Derivatives of vector-valued functions

Definition 1.23 (Derivative of a vector-valued function). The *derivative* of a vector-valued function $\mathbf{r}(t)$ is the limit

$$\mathbf{r}'(t) = \lim_{h \rightarrow 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}.$$

The function \mathbf{r} is said to be *differentiable* at t_0 if $\mathbf{r}'(t_0)$ exists, and \mathbf{r} is said to be *differentiable* on (a, b) if $\mathbf{r}'(t)$ exists for all $t \in (a, b)$.

Geometrically, $\mathbf{r}'(t_0)$ is the *tangent vector* to the curve \mathbf{r} at $\mathbf{r}(t)$.

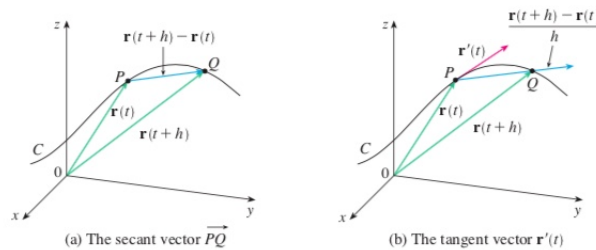


Figure 3: The derivative $\mathbf{r}'(t)$ is the tangent vector to the curve \mathbf{r} at $\mathbf{r}(t)$.

The next theorem shows that we can compute the derivative of a vector-valued function in terms of the derivatives of its component functions.

Theorem 1.24 (Derivative of a vector-valued function). The derivative of a vector-valued function $\mathbf{r}(t) = (r_1(t), r_2(t), r_3(t))$ is given by

$$\mathbf{r}'(t) = (r_1'(t), r_2'(t), r_3'(t)).$$

Proof. By Theorem 1.14

$$\begin{aligned} \mathbf{r}'(t) &= \lim_{h \rightarrow 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(r_1(t+h), r_2(t+h), r_3(t+h)) - (r_1(t), r_2(t), r_3(t))}{h} \\ &= \lim_{h \rightarrow 0} \frac{(r_1(t+h) - r_1(t), r_2(t+h) - r_2(t), r_3(t+h) - r_3(t))}{h} \\ &= \lim_{h \rightarrow 0} \left(\frac{r_1(t+h) - r_1(t)}{h}, \frac{r_2(t+h) - r_2(t)}{h}, \frac{r_3(t+h) - r_3(t)}{h} \right) \\ &= \left(\lim_{h \rightarrow 0} \frac{r_1(t+h) - r_1(t)}{h}, \lim_{h \rightarrow 0} \frac{r_2(t+h) - r_2(t)}{h}, \lim_{h \rightarrow 0} \frac{r_3(t+h) - r_3(t)}{h} \right) \\ &= (r_1'(t), r_2'(t), r_3'(t)). \end{aligned}$$

□

[Add examples.]

Theorem 1.25 (Differentiation formulas). Let \mathbf{u}, \mathbf{v} be differentiable vector-valued functions into \mathbb{R}^n , c a constant, and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ a real-valued function. Then

- (1) $[\mathbf{u}(t) + \mathbf{v}(t)]' = \mathbf{u}'(t) + \mathbf{v}'(t)$
- (2) $[c\mathbf{u}(t)]' = c\mathbf{u}'(t)$
- (3) $[\varphi(t)\mathbf{u}(t)]' = \varphi'(t)\mathbf{u}(t) + \varphi(t)\mathbf{u}'(t)$
- (4) $[\mathbf{u}(t) \cdot \mathbf{v}(t)]' = \mathbf{u}'(t) \cdot \mathbf{v}(t) + \mathbf{u}(t) \cdot \mathbf{v}'(t)$
- (5) $[\mathbf{u}(t) \times \mathbf{v}(t)]' = \mathbf{u}'(t) \times \mathbf{v}(t) + \mathbf{u}(t) \times \mathbf{v}'(t)$ (for $n = 3$)
- (6) $[\mathbf{u}(\varphi(t))]' = \varphi'(t)\mathbf{u}'(\varphi(t))$

Proof. [Add proof.]

□

Definition 1.26 (Higher derivatives). (a) The *second derivative*, \mathbf{r}'' , of a vector-valued function \mathbf{r} is the derivative of \mathbf{r}' : $\mathbf{r}'' = (\mathbf{r}')'$. Thus,

$$\mathbf{r}''(t) = (f''(t), g''(t), h''(t))$$

Similarly, for k a positive integer, the k th derivative of \mathbf{r} is given by the formula

$$\mathbf{r}^{(k)}(t) = (f^{(k)}(t), g^{(k)}(t), h^{(k)}(t)).$$

- (b) A vector-valued function is said to be of class \mathcal{C}^k if its first k derivatives exist and are continuous and of class \mathcal{C}^∞ if all of its derivatives exist. Functions in the class \mathcal{C}^∞ are also called *smooth*.

1.2 Parametrized Curves

We will now focus on a special class of vector-valued functions, which model the motion of a particle through space. Our interest will primarily be in the geometry of the trajectory of the particle.

Definition 1.27 (Parametrized curve). Let I be an interval. A *parametrized curve* is a smooth mapping $\mathbf{r} : I \rightarrow \mathbb{R}^n$. For $n = 2$, a curve is also called a *plane curve* while for $n = 3$ it is also called a *space curve*. The variable t is called the *parameter*. The image $\mathbf{r}(I)$ is called the *trace* of the curve \mathbf{r} .

Remark. If I is an interval containing boundary points, such as $[a, b]$, then we define $f'(a)$ as the right-hand limit

$$f'(a) = \lim_{h \rightarrow 0^+} \frac{f(a+h) - f(a)}{h}$$

and, $f'(b)$ as the left-hand limit

$$f'(b) = \lim_{h \rightarrow 0^-} \frac{f(b+h) - f(b)}{h}.$$

□

Example 1.28. The graph of any smooth function $y = f(x)$ can be written as a parametrized curve by defining

$$\begin{aligned}\mathbf{r} : \mathbb{R} &\rightarrow \mathbb{R}^2 \\ \mathbf{r}(t) &= (t, f(t)).\end{aligned}$$

For example, the function $y = x^2$ can be written the parametrized curve $\mathbf{r}(t) = (t, t^2)$ for all $-\infty < t < \infty$. The trace of a plane curve can be plotted in *Mathematica* as shown below:

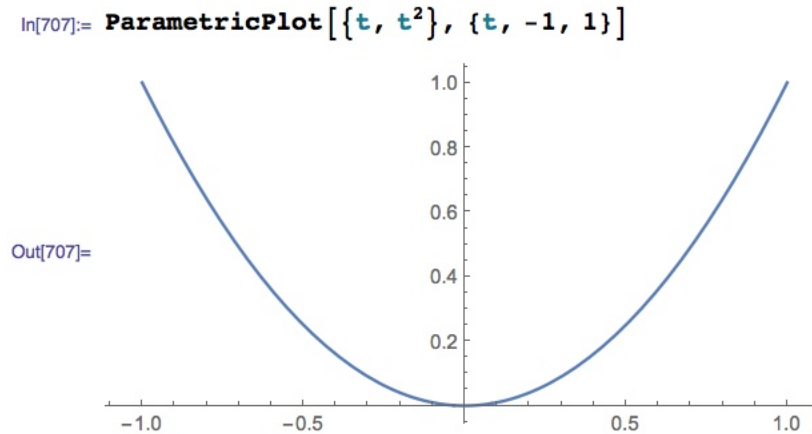


Figure 4: The trace of the plane curve $\mathbf{r}(t) = (t, t^2)$.

The particle moves down the parabola from the upper left starting at $t = -\infty$, reaches the origin at $t = 0$, and then continues up the parabola to the upper right for all $t > 0$.

Exercise 1.6. Sketch the trace of the plane curve $\mathbf{r}(t) = (t^2 - 2t, t + 1)$, $-\infty < t < \infty$ in by making a table of the coordinates $(x(t), y(t))$ for integer values of t from $-2 \leq t \leq 4$. Compare your sketch with the curve produced in *Mathematica* using the “ParametricPlot” command.

Solution. The curve is a parabola, which we can confirm by eliminating t as follows. First, solve for t in terms of y to obtain

$$y = t + 1 \implies y - 1 = t.$$

Substituting into $x(t)$ then gives x in terms of y :

$$x = t^2 - 2t = (y - 1)^2 - 2(y - 1) = y^2 - 4y + 3 = (y - 2)^2 - 1.$$

which is the equation of a parabola in vertex form, with vertex at $(x, y) = (-1, 2)$.

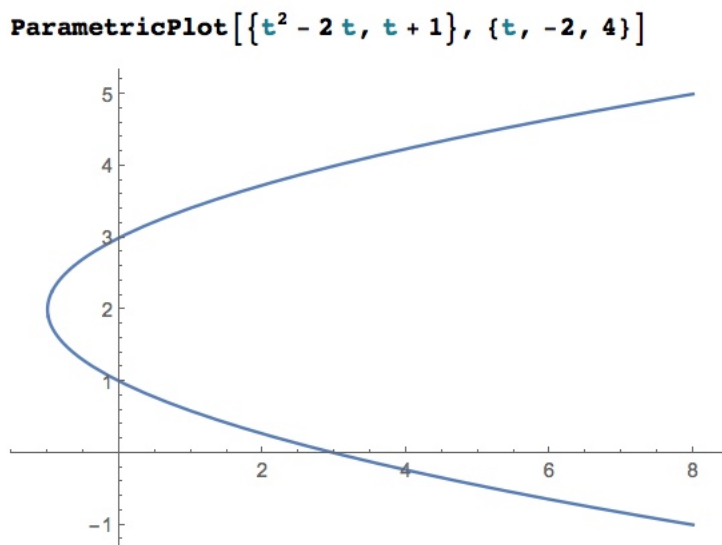


Figure 5: The trace of the plane curve $\mathbf{r}(t) = (t^2 - 2t, t + 1)$ from $-2 \leq t \leq 4$.

The particle travels upwards along the parabola from $t = -\infty$, reaches the vertex at $t = 1$, and then continues upward to the right for $t > 0$. □

Example 1.29. Consider the following three plane curves [Add definition of derivative for at boundary points of a closed interval.]

- (1) $\mathbf{r}_1(t) = (\cos t, \sin t), 0 \leq t \leq 2\pi$,
- (2) $\mathbf{r}_2(t) = (-\sin 2t, \cos 2t), 0 \leq t \leq 2\pi$,
- (3) $\mathbf{r}_3(t) = (\cos(-t), \sin(-t)), 0 \leq t \leq 2\pi$.

For all three curves, $x(t)^2 + y(t)^2 = 1$, so the trace of each curve is the unit circle. However, the curves are different. The reader can easily verify the following by making a table for each curve and sketching the trace:

- (1) In \mathbf{r}_1 , the particle begins at $(x, y) = (1, 0)$ at $t = 0$ and completes one full circle, moving in the counterclockwise direction.
- (2) In \mathbf{r}_2 , the particle begins at $(x, y) = (0, 1)$ at $t = 0$ and completes *two* full circles, moving in the counterclockwise direction.
- (3) In \mathbf{r}_3 , the particle begins at $(x, y) = (1, 0)$ at $t = 0$ and completes one full circle, moving in the *clockwise* direction.

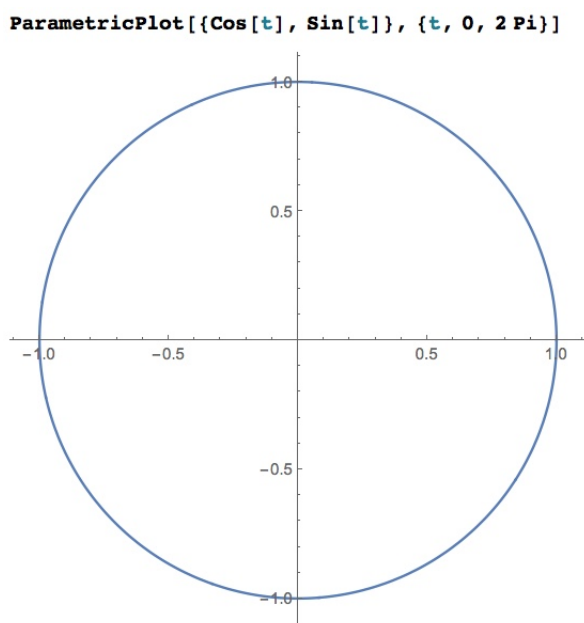
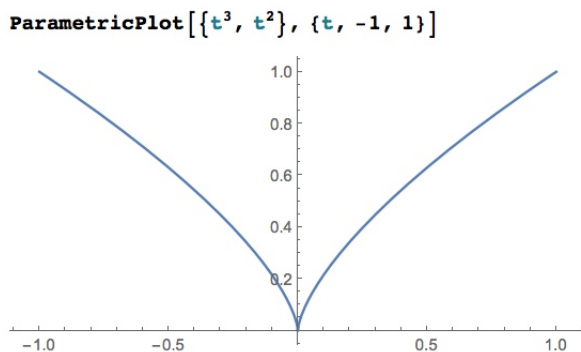


Figure 6: The trace of the curves $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$.

The previous example shows that a parametrized curve is more than just the trace: it is the trace together with a choice of how to traverse the trace of curve.

Example 1.30. Consider now the trace of the plane curve $\mathbf{r}(t) = (t^3, t^2)$, $-\infty < t < \infty$.



Since the component functions are smooth, the reader may be surprised by the cusp at the origin. Solving for y in terms of x by eliminating t , we find that $y = x^{2/3}$, which is indeed not

differentiable at the origin. Note that $\mathbf{r}'(t) = (3t^2, 2t) \neq (0, 0)$ everywhere except the origin, where the tangent vector vanishes.

Definition 1.31 (Regular curve). Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a parametrized curve.

- (a) A point where $\mathbf{r}'(t) = \mathbf{0}$ is called a *singular point*, while a point where $\mathbf{r}'(t) \neq \mathbf{0}$ is called a *regular point*.
- (b) A *regular curve* is a curve with no singular points; that is, it is a curve for which $\mathbf{r}'(t) \neq \mathbf{0}$ for all $t \in I$.

In our study of the differential geometry of curves, we will need the curve to have a tangent vector at every point. Thus, from now on we will assume all curves are regular.

1.3 Reparametrizations

Recall the three regular curves from Example 1.29:

- (1) $\mathbf{r}_1(t) = (\cos t, \sin t), 0 \leq t \leq 2\pi,$
- (2) $\mathbf{r}_2(t) = (-\sin 2t, \cos 2t), 0 \leq t \leq 2\pi,$
- (3) $\mathbf{r}_3(t) = (\cos(-t), \sin(-t)), 0 \leq t \leq 2\pi.$

We have noticed that all three curves have exactly the same trace; namely, the unit circle. If we are only interested in the geometry of the trace of the curve, then we should view all parametrized curves with the same trace as equivalent. We now formalize this idea.

First, notice that if we define the function

$$\varphi : [0, 2\pi] \rightarrow [0, 2\pi]$$

by

$$\varphi(t) = 2t + \frac{\pi}{2}$$

then we see that \mathbf{r}_2 is equal to the composition $\mathbf{r}_2 = \mathbf{r}_1 \circ \varphi$, since for all $t \in [0, 2\pi]$

$$\begin{aligned} \mathbf{r}_1(\varphi(t)) &= (\cos(2t + \frac{\pi}{2}), \sin(2t + \frac{\pi}{2})) \\ &= (\underbrace{\cos(2t) \cos(\frac{\pi}{2})}_{=0} - \underbrace{\sin(2t) \sin(\frac{\pi}{2})}_{=1}, \underbrace{\sin(2t) \cos(\frac{\pi}{2})}_{=0} + \underbrace{\cos(2t) \sin(\frac{\pi}{2})}_{=1}) \\ &= (-\sin(2t), \cos(2t)) \\ &= \mathbf{r}_2(t). \end{aligned}$$

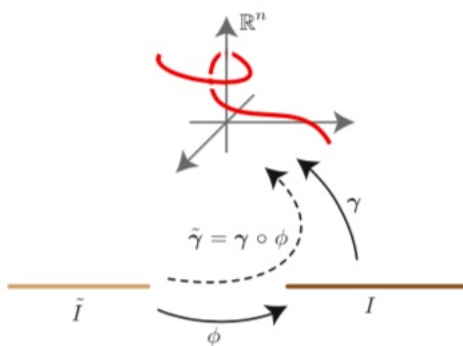
To understand this relationship, recall that a parametrized curve $\mathbf{r}(t)$ is a function which gives the position of the particle at time t , with respect to a given clock. If φ denotes the time with respect to a different clock, then $\mathbf{r}(\varphi)$ gives the position of the particle at time φ . The particles follow the same trajectory, but are at different positions at different times if the two clocks are not synchronized. In this example, the clocks are related by $\varphi = 2t + \frac{\pi}{2}$; that is, the φ -clock is $\frac{\pi}{2}$ seconds ahead of the t -clock (since $\varphi = \frac{\pi}{2}$ when $t = 0$), and runs half as fast (since if the time interval between two successive ticks of the t -clock is $\Delta t = 1$ s, then the time interval between two

successive ticks of the φ -clock is $\Delta\varphi = 2\Delta t = 2$ s; that is, if the t -clock ticks once every second, then the φ -clock ticks once every two seconds). Thus, if $\mathbf{r}_1(t)$ describes a single trip around the unit circle in the counterclockwise sense at unit speed, then, with respect to the t -clock, $\mathbf{r}_1(\varphi) = \mathbf{r}_2(t)$ describes a particle that is already at $(0, 1)$ when $t = 0$ and completes two trips around the circle at double speed in 2π seconds.

Exercise 1.7. Verify that by defining $\varphi : [0, 2\pi] \rightarrow [0, 2\pi]$ by $\varphi(t) = 2\pi - t$, then $\mathbf{r}_3 = \mathbf{r}_1 \circ \varphi$. Interpret this as above.

Solution. We see that $\mathbf{r}_1(\varphi(t)) = (\cos(2\pi - t), \sin(2\pi - t)) = (\cos(-t), \sin(-t)) = \mathbf{r}_3(t)$ for all t . The φ -clock is related to the t -clock by “time-reversal”; that is, the clocks run at the same speed, but the φ clock runs backwards with respect to the t -clock, since $\varphi(t)$ is a decreasing function of t with $\varphi(0) = 2\pi$ and $\varphi(2\pi) = 0$. \square

Definition 1.32 (Reparametrization). Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a regular curve. A *reparametrization* of \mathbf{r} is a function of the form $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi : \tilde{I} \rightarrow \mathbb{R}^n$, where \tilde{I} is an interval and $\varphi : \tilde{I} \rightarrow I$ is a smooth bijection with nowhere vanishing derivative ($\varphi'(t) \neq 0$ for all $t \in \tilde{I}$).¹



Remark. Note that the hypotheses on φ are all required to ensure that $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$ has the same trace as \mathbf{r} :

- The requirement that φ is smooth ensures that $\tilde{\mathbf{r}}$ is smooth, since the composition of smooth functions is smooth.
- The requirement that φ is a bijection is also essential. If φ is not surjective, then the trace of $\tilde{\mathbf{r}}$ will only be a proper subset of the trace of \mathbf{r} . If φ is not injective, then $\tilde{\mathbf{r}}$ will have self-intersections at points where \mathbf{r} does not (since if $\varphi(t_1) = \varphi(t_2)$ for $t_1 \neq t_2$, then even if $\mathbf{r}(t_1) \neq \mathbf{r}(t_2)$, we will have $\tilde{\mathbf{r}}(\varphi(t_1)) = \tilde{\mathbf{r}}(\varphi(t_2))$). [Maybe say this better or give an example.]
- The requirement that $\varphi'(t) \neq 0$ for all t ensures that $\tilde{\mathbf{r}}(t)$ is regular, since if $\mathbf{r}(t)$ is regular and $\varphi'(t) \neq 0$, then by the chain rule

$$\tilde{\mathbf{r}}'(t) = \varphi'(t)\mathbf{r}'(t) \neq 0.$$

¹While in each of the examples above we had $\varphi : \tilde{I} \rightarrow I$ with $\tilde{I} = I$, these intervals need not be the same. For instance, in Example 1.7, we could have instead taken $\varphi : [0, 2\pi] \rightarrow [-2\pi, 0]$ where $\varphi(t) = -t$.

[Give examples or an exercise in which each of these separately fail.] □

Proposition 1.33 (Reparametrized curves are equivalent). The relation $\tilde{\mathbf{r}} \sim \mathbf{r}$ if $\tilde{\mathbf{r}}$ is a reparametrization of \mathbf{r} is an equivalence relation.

Proof. (1) (Reflexivity) Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a parametrized curve. Take $\varphi = id_I : I \rightarrow I$ to be the identity map on I . This is a smooth bijection, and since $id_I(t) = t$, $id'_I(t) = 1 \neq 0$ for all $t \in I$. Since $\mathbf{r} = \mathbf{r} \circ id_I$, $\mathbf{r} \sim \mathbf{r}$. Thus, \sim is reflexive.

(2) (Symmetry) Suppose $\mathbf{r} : I \rightarrow \mathbb{R}^n$ and $\tilde{\mathbf{r}} : \tilde{I} \rightarrow \mathbb{R}^n$ are parametrized curves with $\tilde{\mathbf{r}} \sim \mathbf{r}$. Then there exists a smooth bijection $\varphi : \tilde{I} \rightarrow I$ with $\varphi'(t) \neq 0$ for all $t \in \tilde{I}$. It follows from the Inverse Function Theorem that φ has a smooth inverse, $\varphi^{-1} : I \rightarrow \tilde{I}$, where $(\varphi^{-1})'(t) = 1/\varphi'(\varphi^{-1}(t))$. Since $\varphi'(t)$ is never zero, neither is $(\varphi^{-1})'(t)$.² This proves that $\tilde{\mathbf{r}} \sim \mathbf{r}$ implies that $\mathbf{r} \sim \tilde{\mathbf{r}}$. Thus, \sim is symmetric.

(3) (Transitivity) Let

$$\begin{aligned}\mathbf{r}_1 &: I_1 \rightarrow \mathbb{R}^n, \\ \mathbf{r}_2 &: I_2 \rightarrow \mathbb{R}^n, \\ \mathbf{r}_3 &: I_3 \rightarrow \mathbb{R}^n,\end{aligned}$$

be parametrized curves and suppose $\mathbf{r}_1 \sim \mathbf{r}_2$ and $\mathbf{r}_2 \sim \mathbf{r}_3$. Then there exist smooth bijections

$$\begin{aligned}\varphi &: I_2 \rightarrow I_1, \\ \psi &: I_3 \rightarrow I_2\end{aligned}$$

with nowhere vanishing first derivatives such that

$$\begin{aligned}\mathbf{r}_2 &= \mathbf{r}_1 \circ \varphi, \\ \mathbf{r}_3 &= \mathbf{r}_2 \circ \psi\end{aligned}$$

Then $\varphi \circ \psi : I_3 \rightarrow I_1$ is a smooth bijection with nowhere vanishing derivative such that

$$\mathbf{r}_3 = \mathbf{r}_1 \circ \varphi \circ \psi.$$

Thus, $\mathbf{r}_1 \sim \mathbf{r}_3$.

This shows that \sim is reflexive, symmetric, and transitive, and is therefore an equivalence relation. □

Definition 1.34 (Curve). We call the equivalence class

$$[\mathbf{r}] \equiv \{\tilde{\mathbf{r}} : \tilde{\mathbf{r}} = \mathbf{r} \circ \varphi \text{ for some } \varphi\}$$

a *curve* in \mathbb{R}^n . Any parametrized curve $\tilde{\mathbf{r}}(t)$ in $[\mathbf{r}]$ is said to be a *representative* of the equivalence class $[\mathbf{r}]$. [Show that if two curves have the same trace, then they are equivalent.]

²The *Inverse Function Theorem* states that if $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuously differentiable with $f'(a) \neq 0$ then f is invertible in a neighborhood of a , the inverse is continuously differentiable, and the derivative of the inverse function at $b = f(a)$ is the reciprocal of the derivative of f at a . Applying this to the k -th derivative, it follows as a corollary that we can replace “continuously differentiable” everywhere in the theorem above by “smooth”. The *Gluing Lemma* then allows us to patch together the inverse on each neighborhood into an inverse defined on the entire interval I .

Let $\mathbf{r}(t)$ be a parametrized curve and let $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$ be a reparametrization of $\mathbf{r}(t)$. Since $\varphi'(t)$ is nowhere 0, φ must be either monotonically increasing ($\varphi' > 0$) or monotonically decreasing ($\varphi' < 0$) on all of \tilde{I} .

Definition 1.35 (Oriented curves). A reparametrization $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$ is said to be *orientation-preserving* if $\varphi' > 0$ and *orientation-reversing* if $\varphi' < 0$. Thus, each curve $[\mathbf{r}]$ is partitioned into two subsets

$$\begin{aligned} [\mathbf{r}]_+ &= \{\tilde{\mathbf{r}} : \tilde{\mathbf{r}} = \mathbf{r} \circ \varphi \text{ with } \varphi' > 0\}, \\ [\mathbf{r}]_- &= \{\tilde{\mathbf{r}} : \tilde{\mathbf{r}} = \mathbf{r} \circ \varphi \text{ with } \varphi' < 0\}. \end{aligned}$$

Each subset is called an *oriented curve*. A choice of orientation of a curve $[\mathbf{r}]$ is a choice of one of these two subsets.

Example 1.36. Consider again three parametrized curves from Example 1.29 and :

- (1) $\mathbf{r}_1(t) = (\cos t, \sin t), 0 \leq t \leq 2\pi,$
- (2) $\mathbf{r}_2(t) = (-\sin 2t, \cos 2t), 0 \leq t \leq 2\pi,$
- (3) $\mathbf{r}_3(t) = (\cos(-t), \sin(-t)), 0 \leq t \leq 2\pi.$

We have seen that $\mathbf{r}_2 = \mathbf{r}_1 \circ \varphi$ where $\varphi(t) = 2t + \frac{\pi}{2}$. Since $\varphi'(t) = 2 > 0$, these two curves have the same orientation (i.e., they are representatives of the same oriented curve). We also saw that $\mathbf{r}_3 = \mathbf{r}_1 \circ \psi$ where $\psi(t) = 2\pi - t$. Since $\psi'(t) = -1 < 0$, these two curves have opposite orientations.

[Add exercises.]

1.4 Arc Length

We have just seen that a given curve can be parametrized in many ways. Out of these many options, there is a natural choice for the parameter t , namely, the *arc length* measured from any point \mathbf{r}_0 on the curve. This is because the arc length is an *invariant* of the curve, which is independent of the parametrization. For a curve given by the graph of a function $y = f(x)$, the arc length from a to b is given by

$$s = \int_a^b \sqrt{1 + (f'(x))^2} dx. \quad (1.3)$$

Writing the function as a parametrized curve $\mathbf{r}(t) = (t, f(t))$, we have $\mathbf{r}'(t) = (1, f'(t))$ and $\|\mathbf{r}'(t)\| = \sqrt{1 + (f'(t))^2}$, so we can write Equation (1.3) as

$$s = \int_a^b \|\mathbf{r}'(t)\| dt. \quad (1.4)$$

[Add example.] How do we compute the arc length if $\mathbf{r}(t)$ is not the graph of a function? If $\mathbf{r}(t)$ is not the graph of a function, it turns out that Equation (1.4) is still valid. In this case, we obtain the formula as the limit of a sequence of polygonal approximations of the curve (the proof is given in Appendix A).

Definition 1.37. Arc length Let $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$ be a regular curve. We define the *arc length* between the points $\mathbf{r}(a)$ and $\mathbf{r}(b)$ on the curve by

$$s = \int_a^b \|\mathbf{r}'(t)\| dt.$$

Example 1.38. We now compute the arc length along the portion of the unit circle from $(1, 0)$ to $(0, 1)$. If we use the parametrization

$$\mathbf{r}_1(t) = (\cos t, \sin t), 0 \leq t \leq \frac{\pi}{2},$$

then we have

$$\begin{aligned} \mathbf{r}'_1(t) &= (-\sin t, \cos t), \\ \|\mathbf{r}'_1(t)\| &= \sqrt{(-\sin t)^2 + (\cos t)^2} = \sqrt{1} = 1. \end{aligned}$$

Equation (1.4) then gives

$$s = \int_0^{\frac{\pi}{2}} \|\mathbf{r}'_1(t)\| dt = \int_0^{\frac{\pi}{2}} dt = \frac{\pi}{2}.$$

Exercise 1.8. Use the parametrization $\mathbf{r}_2(t) = (\cos(2t), \sin(2t)), 0 \leq t \leq \frac{\pi}{4}$ to compute the arc length of the same portion of the unit circle as in Example 1.38. Verify that you get the same answer.

Solution. Using the parametrization $\mathbf{r}_2(t) = (\cos(2t), \sin(2t)), 0 \leq t \leq \frac{\pi}{4}$, we have

$$\begin{aligned} \mathbf{r}'_2(t) &= (-2 \sin(2t), 2 \cos(2t)), \\ \|\mathbf{r}'_2(t)\| &= \sqrt{4 \sin^2(2t) + 4 \cos^2(2t)} = \sqrt{4(\sin^2(2t) + \cos^2(2t))} = \sqrt{4} = 2 \end{aligned}$$

and therefore

$$s = \int_0^{\frac{\pi}{4}} \|\mathbf{r}'_2(t)\| dt = 2 \int_0^{\frac{\pi}{4}} dt = 2 \cdot \frac{\pi}{4} = \frac{\pi}{2}.$$

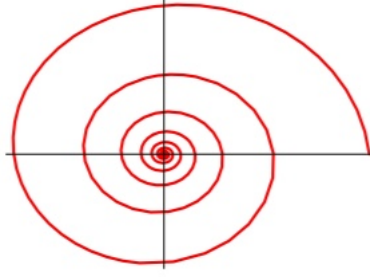
□

Exercise 1.9. A *logarithmic spiral* is a plane curve of the form $\mathbf{r}(t) = c(e^{\lambda t} \cos t, e^{\lambda t} \sin t)$, $t \in \mathbb{R}$ where $c, \lambda \in \mathbb{R}$ and $c \neq 0$. Below is the restriction of $\mathbf{r}(t)$ to $[0, \infty)$ with $\lambda < 0$.

Use an improper integral to prove that such a restriction has finite arc length even though it makes infinitely many loops around the origin.

Solution. We have

$$\begin{aligned} \mathbf{r}'(t) &= c(\lambda e^{\lambda t} \cos t - e^{\lambda t} \sin t, \lambda e^{\lambda t} \sin t + e^{\lambda t} \cos t) \\ &= ce^{\lambda t}(\lambda \cos t - \sin t, \lambda \sin t + \cos t) \end{aligned}$$



and therefore

$$\begin{aligned} ||\mathbf{r}'(t)|| &= |c|e^{\lambda t} \sqrt{(\lambda \cos t - \sin t)^2 + (\lambda \sin t + \cos t)^2} \\ &= |c|e^{\lambda t} \sqrt{\lambda^2 \cos^2 t + \sin^2 t - 2\lambda \sin t \cos t + \lambda^2 \sin^2 t + \cos^2 t + 2\lambda \sin t \cos t} \\ &= |c|e^{\lambda t} \sqrt{\lambda^2 + 1}. \end{aligned}$$

The arc length is then given by the integral

$$s = \int_0^\infty ||\mathbf{r}'(t)|| dt = |c| \sqrt{\lambda^2 + 1} \int_0^\infty e^{-|\lambda|t} dt.$$

Substituting

$$u = -|\lambda|t, \quad du = -|\lambda|dt,$$

this becomes

$$\begin{aligned} s &= -\frac{|c|\sqrt{\lambda^2 + 1}}{|\lambda|} \int_0^{-\infty} e^u du \\ &= -\frac{|c|\sqrt{\lambda^2 + 1}}{|\lambda|} \lim_{k \rightarrow -\infty} (e^k - 1) \\ &= -\frac{|c|\sqrt{\lambda^2 + 1}}{|\lambda|} (0 - 1) \\ &= \frac{|c|\sqrt{\lambda^2 + 1}}{|\lambda|}. \end{aligned}$$

□

Proposition 1.39 (Invariance of arc length). The arc length is independent of parametrization.

Proof. Let $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$ be a reparametrization of \mathbf{r} . Then since

$$\tilde{\mathbf{r}}'(t) = [\mathbf{r}(\varphi(t))]' = \varphi'(t)\mathbf{r}'(\varphi(t)),$$

we have

$$\begin{aligned} s &= \int_{t_0}^{t_1} \|\mathbf{r}'(t)\| dt \\ &= \int_{t_0}^{t_1} \|\varphi'(t) \mathbf{r}'(\varphi(t))\| dt \\ &= \int_{t_0}^{t_1} \|\mathbf{r}'(\varphi(t))\| |\varphi'(t)| dt \end{aligned}$$

If $\varphi'(t) > 0$, then $|\varphi'(t)| = \varphi'(t)$ and by substituting

$$u = \varphi(t), \quad du = \varphi'(t) dt$$

we obtain

$$s = \int_{\varphi(t_0)}^{\varphi(t_1)} \|\mathbf{r}'(u)\| du.$$

If $\varphi'(t) < 0$, then $|\varphi'(t)| = -\varphi'(t)$ and the same substitution gives

$$s = - \int_{\varphi(t_0)}^{\varphi(t_1)} \|\mathbf{r}'(u)\| du = \int_{\varphi(t_1)}^{\varphi(t_0)} \|\mathbf{r}'(u)\| du.$$

[Include figure?]

□

Proposition 1.40 (Any regular curve can be parametrized by arc length). Any regular curve $\mathbf{r} : I \rightarrow \mathbb{R}^n$ can be parametrized by arc length.

Proof. Choose $t_0 \in I$ and consider the arc length function $s : I \rightarrow \mathbb{R}$ defined by

$$s(t) = \int_{t_0}^t \|\mathbf{r}'(u)\| du.$$

Let $\tilde{I} = s(I)$ denote the image of I under s . By the Fundamental Theorem of Calculus, $s'(t) = \|\mathbf{r}'(t)\| \neq 0$ (since the curve is regular), so by the Inverse Function Theorem s has an inverse $\varphi : \tilde{I} \rightarrow I$, which is also a smooth bijection with a nowhere-vanishing derivative. Then $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$ is parametrized by arc length, since $\mathbf{r}(\varphi(s))$ is the position of a point along the curve at the time when it achieves arc length s measured from t_0 . □

Example 1.41. Consider the helix $\mathbf{r}(t) = (\cos t, \sin t, t)$, $t \in \mathbb{R}$.

Let us reparametrize the helix with respect to the arc length measured from $(1, 0, 0)$ in the direction of increasing t . Since $\mathbf{r}(0) = (1, 0, 0)$, following the proof above we define

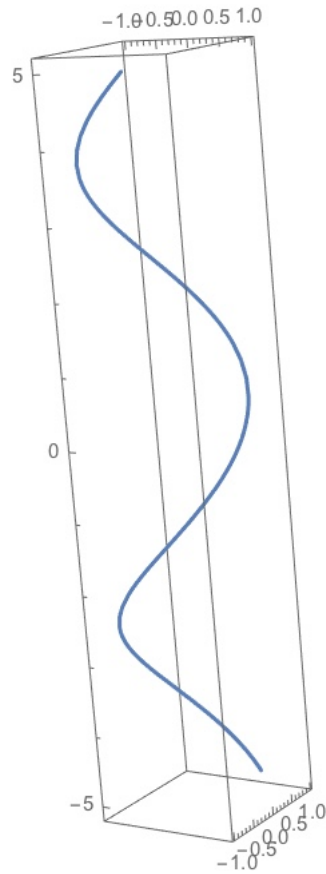
$$s(t) = \int_0^t \|\mathbf{r}'(u)\| du.$$

Since $\mathbf{r}'(u) = (-\sin u, \cos u, 1)$, we have $\|\mathbf{r}'(u)\| = \sqrt{2}$ and therefore

$$\begin{aligned} s(t) &= \int_0^t \|\mathbf{r}'(u)\| du \\ &= \sqrt{2} \int_0^t du \\ &= \sqrt{2}t. \end{aligned}$$

In[728]:= ParametricPlot3D[{Cos[t], Sin[t], t}, {t, -5, 5}]

Out[728]=



The inverse function is then easily obtained:

$$\varphi(s) = \frac{s}{\sqrt{2}}$$

and

$$\mathbf{r}(\varphi(s)) = \left(\cos\left(\frac{s}{\sqrt{2}}\right), \sin\left(\frac{s}{\sqrt{2}}\right), \frac{s}{\sqrt{2}} \right).$$

Example 1.42. Let's reparametrize the plane curve $\mathbf{r}(t) = (\cos(3t), \sin(2t))$ with respect to arc length measured from $(1, 0)$ in the direction of increasing t . Since

$$\begin{aligned} \mathbf{r}'(t) &= (-3\sin(3t), 2\cos(2t)) \\ \|\mathbf{r}'(t)\| &= \sqrt{9\sin^2(3t) + 4\cos^2(2t)}, \end{aligned}$$

the arc length function is given by

$$s(t) = \int_0^t \sqrt{9\sin^2(3t) + 4\cos^2(2t)} dt$$

This integral cannot be evaluated in terms of elementary functions. [Explain what this means.] This example illustrates that, while our proof gave an explicit method to parametrize any curve by arc length, in practice it is usually not computationally reasonable to implement. However, we will use the fact that every curve can be parametrized by arc length for theoretical purposes.

Proposition 1.43. A curve parametrized by arc length is a *unit speed curve*; that is, $\|\mathbf{r}'(t)\| = 1$ for all $t \in I$.

Proof. If \mathbf{r} is parametrized by arc length, then by the Fundamental Theorem of Calculus

$$s'(t) = \frac{d}{dt} \int_{t_0}^t \|\mathbf{r}'(u)\| du = \|\mathbf{r}'(t)\| = 1.$$

Conversely, suppose $\|\mathbf{r}'(t)\| = 1$ for all t . Then

$$s(t) = \int_{t_0}^t \|\mathbf{r}'(u)\| du = \int_{t_0}^t 1 du = t - t_0$$

so t is equal to the arc length accumulated from the point t_0 . □

Exercise 1.10. Verify that the helix parametrized by arc length in Example 1.41 is a unit speed curve.

Solution. We have $\mathbf{r}'(s) = \frac{1}{\sqrt{2}}(-\sin(\frac{s}{\sqrt{2}}), \cos(\frac{s}{\sqrt{2}}), 1)$ and therefore

$$\begin{aligned} \|\mathbf{r}'(s)\| &= \frac{1}{\sqrt{2}} \sqrt{\sin^2(\frac{s}{\sqrt{2}}) + \cos^2(\frac{s}{\sqrt{2}}) + 1} \\ &= \frac{1}{\sqrt{2}} \sqrt{2} \\ &= 1. \end{aligned}$$

□

1.5 Curvature

In this section, we will quantify how sharply a curve “bends” at each point. This will lead us to define a function called the *curvature function* for the curve.

1.6 Curvature of a unit speed curve

We begin by considering curves parametrized by arc length. To define the curvature function, we will need the following lemma.

Lemma 1.44. Let $\mathbf{u}, \mathbf{v} : I \rightarrow \mathbb{R}^n$ be a pair of curves.

- (a) If \mathbf{u} has constant nonzero length (that is, if $\|\mathbf{u}(t)\| = c > 0$ for all $t \in I$), then $\mathbf{u}'(t)$ is orthogonal to $\mathbf{u}(t)$ for all $t \in I$.

(b) If $\mathbf{u}(t)$ is orthogonal to $\mathbf{v}(t)$ for all $t \in I$, then

$$\mathbf{u}'(t) \cdot \mathbf{v}(t) = -\mathbf{u}(t) \cdot \mathbf{v}'(t) = 0$$

for all $t \in I$.

Proof. (1) Suppose $\|\mathbf{u}(t)\| = c$ for all t . Differentiating both sides, gives $\|\mathbf{u}(t)\|' = 0$. That is,

$$\begin{aligned} 0 &= \|\mathbf{u}(t)\|' \\ &= \left(\sqrt{\mathbf{u}(t) \cdot \mathbf{u}(t)} \right)' \\ &= \frac{2\mathbf{u}'(t) \cdot \mathbf{u}(t)}{2\sqrt{\mathbf{u}(t) \cdot \mathbf{u}(t)}} \\ &= \frac{\mathbf{u}'(t) \cdot \mathbf{u}(t)}{\|\mathbf{u}(t)\|} \\ &= \frac{\mathbf{u}'(t) \cdot \mathbf{u}(t)}{c} \end{aligned}$$

and therefore $\mathbf{u}'(t) \cdot \mathbf{u}(t) = 0$ for all t .

(2) Suppose $\mathbf{u}(t) \cdot \mathbf{v}(t) = 0$ for all t . Differentiating both sides, we obtain

$$0 = \mathbf{u}'(t) \cdot \mathbf{v}(t) + \mathbf{u}(t) \cdot \mathbf{v}'(t)$$

and therefore

$$\mathbf{u}'(t) \cdot \mathbf{v}(t) = -\mathbf{u}(t) \cdot \mathbf{v}'(t)$$

for all t .

□

Note that both of these hypotheses of Lemma 1.44 are true for an orthonormal set of vectors $\{\mathbf{u}, \mathbf{v}\}$.

Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a curve parametrized by arc length. Since $\mathbf{r}'(s)$ has unit length, by part (a) of Lemma 1.44, $\mathbf{r}''(s)$ is orthogonal to $\mathbf{r}'(s)$. Thus, $\|\mathbf{r}''(s)\|$ measures the rate at which the curve is pulling away from the tangent line at $\mathbf{r}(s)$.

Definition 1.45 (Curvature). Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a curve parametrized by arc length $s \in I$. Define the *curvature function* $\kappa : I \rightarrow [0, \infty)$ by $\kappa(s) = \|\mathbf{r}''(s)\|$. The number $\kappa(s)$ is called the *curvature* of the curve \mathbf{r} at s .

Example 1.46. (a) If \mathbf{r} is a straight line, then $\mathbf{r}(s) = \mathbf{r}'(s_0)s + \mathbf{r}(s_0)$, where $\mathbf{r}(s_0), \mathbf{r}'(s_0)$ are constant vectors and we have $\kappa(s) = \|\mathbf{r}''(s)\| \equiv 0$. Conversely, if $\kappa = \|\mathbf{r}''(s)\| \equiv 0$, then integrating twice gives $\mathbf{r}(s) = \mathbf{r}'(s_0)s + \mathbf{r}(s_0)$, so the curve is a straight line.

(b) If $\mathbf{r}(s) = (R \cos(\frac{s}{R}), R \sin(\frac{s}{R}))$ is a circle of radius R , then

$$\mathbf{r}''(s) = \left(-\frac{1}{R} \cos\left(\frac{s}{R}\right), -\frac{1}{R} \sin\left(\frac{s}{R}\right) \right)$$

and therefore

$$\kappa(s) = \|\mathbf{r}''(s)\| = \frac{1}{R}.$$

Definition 1.47 (Unit normal vector). At points where $\kappa(s) \neq 0$, the vector

$$\hat{\mathbf{n}}(s) = \frac{\mathbf{r}''(s)}{\|\mathbf{r}''(s)\|}$$

is a unit vector orthogonal to $\mathbf{r}'(s)$. This is called the *unit normal vector* to \mathbf{r} at s .

1.7 Osculating Plane

Let $\mathbf{r}(s)$ be a unit-speed curve. Consider a point s_0 where $\kappa(s_0) \neq 0$. Then, for sufficiently small $h = s - s_0$, we can approximate $\mathbf{r}(s)$ by its second-order Taylor polynomial:

$$\mathbf{r}(s_0 + h) = \mathbf{r}(s_0) + h\mathbf{r}'(s_0) + \frac{h^2}{2}\mathbf{r}''(s_0) + \mathbf{E}(h),$$

$$\mathbf{r}(s_0 + h) = \mathbf{r}(s_0) + h\hat{\mathbf{t}} + \frac{h^2}{2}\kappa(s_0)\hat{\mathbf{n}}(s_0) + \mathbf{E}(h),$$

where $\lim_{h \rightarrow 0} \frac{\|\mathbf{E}(h)\|}{h^2} = 0$. Thus, sufficiently near $\mathbf{r}(s_0)$, the trace of a curve lies in the plane spanned by $\{\hat{\mathbf{t}}(s_0), \hat{\mathbf{n}}(s_0)\}$, called the *osculating plane at $\mathbf{r}(s_0)$* . With respect to the coordinate $\mathbf{r}(s_0 + h) - \mathbf{r}(s_0)$ centered at $\mathbf{r}(s_0)$, the trace of the curve in the osculating plane is given by

$$\left\{ \left(h, \frac{\kappa(s_0)}{2}h^2 \right) : -\epsilon < h < \epsilon \right\} \subseteq \mathbb{R}^2,$$

which is a parabola $y = \frac{\kappa(s_0)}{2}x^2$ whose vertex is at the origin and whose concavity is $y'' = \kappa(s_0)$.

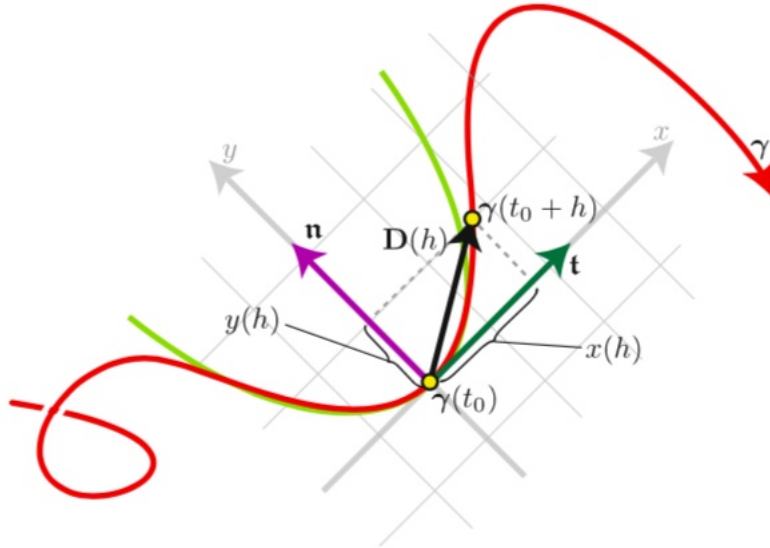


Figure 7: The trace of \mathbf{r} is approximated near $\mathbf{r}(s_0)$ by the parabola with concavity $\kappa(s_0)$ in the osculating plane.

The trace of the curve is also well approximated by a certain circle in the osculated plane.

Definition 1.48 (Osculating circle). The *osculating circle* to γ is the circle of radius $\kappa(s_0)$ in the osculating plane centered at $\mathbf{r}(s_0) + \frac{1}{\kappa(s_0)}\hat{\mathbf{n}}$.

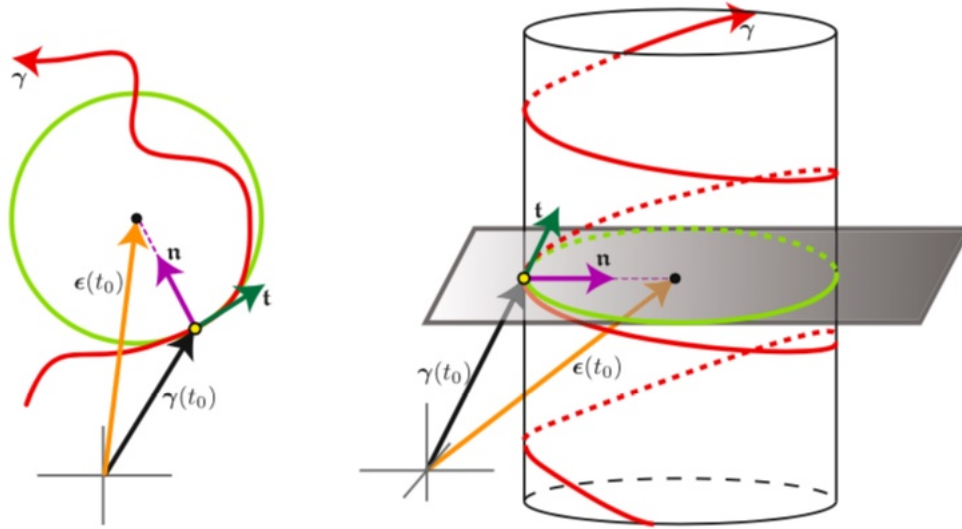


Figure 8: The osculating circle for a plane curve (left) and a space curve (right).

Thus, we may also interpret the curvature as the radius of the osculating circle. [\[Add Exercise 1.42?\]](#)

At points where $\kappa(s) = 0$, the normal vector (and therefore the osculating plane) is not defined. To proceed with our analysis, we will need the osculating plane, so from now on we will only consider curves for which $\kappa(s) \neq 0$ for all $s \in I$.

1.8 Arbitrary parametrizations

As noted previously, in practice it is often too difficult to parametrize a curve by arc length, and we are forced to work with other parametrizations.

Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a curve with an arbitrary parametrization $t \in I$. One might guess that the curvature. Since the speed $\|\mathbf{r}'(t)\|$ is no longer constant, the acceleration vector $\mathbf{r}''(t)$ will have a nonzero component in the direction of the velocity vector $\mathbf{r}'(t)$. We might therefore be tempted to define the curvature function $\kappa(t)$ as the magnitude of the component of the acceleration perpendicular to $\mathbf{r}'(t)$; that is, to define $\kappa(t) := \|\mathbf{r}''_{\perp}(t)\|$, where

$$\begin{aligned}\mathbf{r}''_{\perp}(t) &= \mathbf{r}''(t) - \text{proj}_{\mathbf{r}'(t)} \mathbf{r}''(t) \\ &= \mathbf{r}''(t) - \frac{\mathbf{r}''(t) \cdot \mathbf{r}'(t)}{\mathbf{r}'(t) \cdot \mathbf{r}'(t)} \mathbf{r}'(t).\end{aligned}$$

However, this definition is problematic, as the next example shows.

Example 1.49. Consider the parabola which is the graph of $y = x^2$. Describe the parabola as a

parametrized curve in the following two parametrizations,

$$\begin{aligned}\mathbf{r}_1(t) &= (t, t^2), & t \in \mathbb{R}, \\ \mathbf{r}_2(t) &= (2t - 2, (2t - 2)^2), & t \in \mathbb{R}.\end{aligned}$$

In the first parametrization, the origin is the point $\mathbf{r}(0) = (0, 0)$. Since

$$\begin{aligned}\mathbf{r}'_1(t) &= (1, 2t), \\ \mathbf{r}''_1(t) &= (0, 2),\end{aligned}$$

$$\begin{aligned}\mathbf{r}''_{1\perp}(0) &= (0, 2) - \frac{(0, 2) \cdot (1, 0)}{(1, 0) \cdot (1, 0)}(1, 0) \\ &= (0, 2),\end{aligned}$$

and therefore $\kappa(0) = \|\mathbf{r}''_{1\perp}(0)\| = 2$.

In the second parametrization, the origin is $\mathbf{r}_2(1) = (0, 0)$. Since

$$\begin{aligned}\mathbf{r}'_2(t) &= (2, 4(2t - 2)) = (2, 8t - 8) \\ \mathbf{r}''_2(t) &= (0, 8)\end{aligned}$$

we have

$$\begin{aligned}\mathbf{r}''_{2\perp}(1) &= (0, 8) - \frac{(0, 8) \cdot (2, 0)}{(2, 0) \cdot (2, 0)}(2, 0) \\ &= (0, 8)\end{aligned}$$

and therefore

$$\kappa(1) = \|\mathbf{r}''_{2\perp}(1)\| = 8.$$

Using our definition of $\kappa(t)$, we have just obtained *different* values of the curvature at the origin of the parabola by using different parametrizations. The curvature should be a geometric property of the curve that does not depend on parametrization, so this definition turns out to be bad.

To find a satisfactory definition of curvature, we need to study more carefully the dependence of the derivatives of $\mathbf{r}(t)$ on parametrization. Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a regular curve with an arbitrary parametrization, and let $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$ be a reparametrization of \mathbf{r} . Then, by the chain rule,

$$\begin{aligned}\tilde{\mathbf{r}}'(t) &= \mathbf{r}'(\varphi(t))\varphi'(t), \\ \tilde{\mathbf{r}}''(t) &= \mathbf{r}''(\varphi(t))(\varphi'(t))^2 + \mathbf{r}'(\varphi(t))\varphi''(t), \\ \tilde{\mathbf{r}}''_{\perp}(t) &= \mathbf{r}''_{\perp}(\varphi(t))(\varphi'(t))^2.\end{aligned}$$

From these expressions, we see that the quantity

$$\frac{\|\mathbf{r}''_{\perp}(t)\|}{\|\mathbf{r}'(t)\|^2} \tag{1.5}$$

is invariant under reparametrizations. Moreover, when $t = s$ is the arc length, $\|\mathbf{r}'(s)\| = 1$ and $\mathbf{r}''(s) = \mathbf{r}''_{\perp}(s)$, so the expression in (1.5) agrees with our definition of the curvature function of a unit speed curve in Definition 1.45.

Thus, we arrive at our definition of the curvature function in an arbitrary parametrization.

Definition 1.50 (Curvature in an arbitrary parametrization). Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a regular curve with an arbitrary parametrization $t \in I$. We define the *curvature function* $\kappa : I \rightarrow [0, \infty)$ for all $t \in I$ by

$$\kappa(t) = \frac{\|\mathbf{r}''_{\perp}(t)\|}{\|\mathbf{r}'(t)\|^2}. \quad (1.6)$$

Rewriting (1.6) as

$$\|\mathbf{r}''_{\perp}(t)\| = \kappa(t)\|\mathbf{r}'(t)\|^2,$$

this expression says, in physical terms, that the magnitude of the centripetal acceleration at t depends not only on the curvature, but also on the speed at t . This explains the results of Example 1.49.

Definition 1.51 (Unit tangent and Unit normal in an arbitrary parametrization). Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a regular curve with an arbitrary parametrization and let $t \in I$ be a point where $\kappa(t) \neq 0$. We define the *unit tangent* and *unit normal vectors* at t to be

$$\hat{\mathbf{t}}(t) = \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}, \quad \hat{\mathbf{n}}(t) = \frac{\mathbf{r}''_{\perp}(t)}{\|\mathbf{r}''_{\perp}(t)\|}.$$

The next proposition shows that we can compute the unit normal vector in terms of the unit tangent vector.

Proposition 1.52. Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a regular curve. For any point where $\kappa(t) \neq 0$, we have

$$\hat{\mathbf{n}}(t) = \frac{\hat{\mathbf{t}}'(t)}{\|\hat{\mathbf{t}}'(t)\|}.$$

Proof. By the quotient rule,

$$\hat{\mathbf{t}}'(t) = \frac{\|\mathbf{r}'(t)\|\mathbf{r}''(t) - \mathbf{r}'(t)\|\mathbf{r}'(t)\|'}{\|\mathbf{r}'(t)\|^2}.$$

Now

$$\begin{aligned} \|\mathbf{r}'(t)\|' &= \frac{d}{dt}(\mathbf{r}'(t) \cdot \mathbf{r}'(t))^{1/2} \\ &= \frac{1}{2}(\mathbf{r}'(t) \cdot \mathbf{r}'(t))^{-1/2}(2\mathbf{r}''(t) \cdot \mathbf{r}'(t)) \\ &= \frac{\mathbf{r}''(t) \cdot \mathbf{r}'(t)}{\|\mathbf{r}'(t)\|} \end{aligned}$$

so we have

$$\begin{aligned} \hat{\mathbf{t}}'(t) &= \frac{\|\mathbf{r}'(t)\|\mathbf{r}''(t) - \mathbf{r}'(t)\|\mathbf{r}'(t)\|'}{\|\mathbf{r}'(t)\|^2} \\ &= \frac{\mathbf{r}''(t)}{\|\mathbf{r}'(t)\|} - \frac{\mathbf{r}''(t) \cdot \mathbf{r}'(t)}{\|\mathbf{r}'(t)\|^3} \mathbf{r}'(t) \\ &= \frac{1}{\|\mathbf{r}'(t)\|} \left(\mathbf{r}''(t) - \frac{\mathbf{r}''(t) \cdot \mathbf{r}'(t)}{\|\mathbf{r}'(t)\|^2} \mathbf{r}'(t) \right) \\ &= \frac{\mathbf{r}''_{\perp}(t)}{\|\mathbf{r}'(t)\|}. \end{aligned}$$

and therefore

$$\begin{aligned} \frac{\hat{\mathbf{t}}'(t)}{\|\hat{\mathbf{t}}'(t)\|} &= \frac{\frac{\mathbf{r}'_{\perp}(t)}{\|\mathbf{r}'(t)\|}}{\frac{\|\mathbf{r}'_{\perp}(t)\|}{\|\mathbf{r}'(t)\|}} \\ &= \frac{\mathbf{r}''_{\perp}(t)}{\|\mathbf{r}''_{\perp}(t)\|} \\ &\equiv \hat{\mathbf{n}}(t). \end{aligned}$$

□

1.9 Plane curves

Our results so far about curves in \mathbb{R}^n have been valid for all n . In this section, we consider the special case $n = 2$. This is the only dimension in which the terms “clockwise” and “counterclockwise” make sense for describing how a regular curve is turning.

Proposition 1.53. The map $R : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $R(x, y) = (-y, x)$ is an isomorphism.

Proof. The map R is the same as multiplication by the matrix

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

which is linear. The kernel of this map is $\{(0, 0)\}$, so R is injective. Given any $(x, y) \in \mathbb{R}^2$, we see that $(x, y) = R(y, -x)$, hence R is also surjective. Thus, R is an isomorphism. □

Geometrically, R rotates each vector $(x, y) \in \mathbb{R}^2$ counterclockwise by 90° . One can immediately check that $R^2(\mathbf{v}) \equiv R(R(\mathbf{v})) = -\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^2$.

Suppose now that $\mathbf{r} : I \rightarrow \mathbb{R}^2$ is a unit-speed curve. Then, for all $s \in I$, by Lemma 1.44 $\mathbf{r}''(t)$ and $R(\mathbf{r}'(t))$ are both orthogonal to $\mathbf{r}'(t)$, so they must be parallel. We can therefore write

$$\mathbf{r}''(t) = \kappa_s(t)R(\mathbf{r}'(t)) \tag{1.7}$$

for some scalar $\kappa_s(t) \in \mathbb{R}$. We call $\kappa_s : I \rightarrow \mathbb{R}$ the *signed curvature function*. It is negative if the curve is turning clockwise at t , and positive if it is turning counterclockwise, as shown below: This measurement is called “signed curvature” because its sign is interpreted as above, while its absolute value equals the curvature:

Proposition 1.54. Let $\mathbf{r} : I \rightarrow \mathbb{R}^2$ be a regular unit-speed plane curve. Then $|\kappa_s(t)| = \kappa(t)$.

Proof. Since $\|R(\mathbf{r}'(t))\| = \|\mathbf{r}'(t)\| = 1$,

$$\kappa(t) = \|\mathbf{r}''(t)\| = \|\kappa_s(t)R(\mathbf{r}'(t))\| = |\kappa_s(t)| \|R(\mathbf{r}'(t))\| = |\kappa_s(t)|.$$

□

Taking the dot product of (1.7) with $R(\mathbf{r}'(t))$, we obtain

$$\begin{aligned} \mathbf{r}''(t) \cdot R(\mathbf{r}'(t)) &= \kappa_s(t)R(\mathbf{r}'(t)) \cdot R(\mathbf{r}'(t)) \\ &= \kappa_s(t)\mathbf{r}'(t) \cdot \mathbf{r}'(t) \\ &= \kappa_s(t)\|\mathbf{r}'(t)\|^2 \\ &= \kappa_s(t), \end{aligned}$$

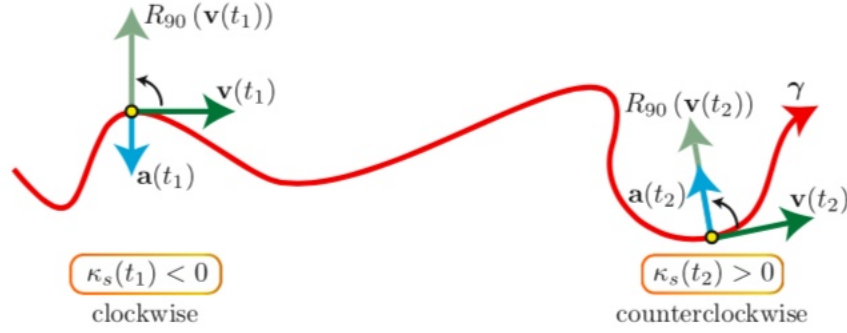


Figure 9: Signed curvature is positive when the path is turning counterclockwise, and negative when turning clockwise.

so we can write $\kappa_s(t)$ as

$$\kappa_s(t) = \mathbf{r}''(t) \cdot R(\mathbf{r}'(t)). \quad (1.8)$$

Definition 1.55. Let $\mathbf{r} : I \rightarrow \mathbb{R}^2$ be a regular plane curve with an arbitrary parametrization. Then we define, for all $t \in I$,

$$\kappa_s(t) = \frac{\mathbf{r}''(t) \cdot R\left(\frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}\right)}{\|\mathbf{r}'(t)\|^2} = \frac{\mathbf{r}''(t) \cdot R(\mathbf{r}'(t))}{\|\mathbf{r}'(t)\|^3}. \quad (1.9)$$

Note that this definition of $\kappa_s(t)$ agrees with our previous definition when $t = s$ is the arc length. This could also be verified by comparing the definition $\kappa(t) = \frac{\|\mathbf{r}''_{\perp}(t)\|}{\|\mathbf{r}'(t)\|^2}$ to the above formula for $\kappa_s(t)$, noticing that

$$\mathbf{r}''(t) \cdot R\left(\frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}\right) = \pm \|\mathbf{r}''_{\perp}(t)\|.$$

Proposition 1.56. The formula in (1.9) is invariant under orientation-preserving reparametrizations.

Proof. Let $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$ be an orientation-preserving reparametrization of \mathbf{r} (i.e., $\varphi'(t) > 0$ for all $t \in I$). Then

$$\begin{aligned} \tilde{\mathbf{r}}'(t) &= \mathbf{r}'(\varphi(t))\varphi'(t), \\ R(\tilde{\mathbf{r}}'(t)) &= \varphi'(t)R(\mathbf{r}'(\varphi(t))), \\ \tilde{\mathbf{r}}''(t) &= \mathbf{r}''(\varphi(t))(\varphi'(t))^2 + \mathbf{r}'(\varphi(t))\varphi''(t) \end{aligned}$$

and

$$\begin{aligned}
 \frac{R(\tilde{\mathbf{r}}'(t)) \cdot \tilde{\mathbf{r}}''(t)}{\|\tilde{\mathbf{r}}'(t)\|^3} &= \frac{\varphi'(t)R(\tilde{\mathbf{r}}'(t)) \cdot \tilde{\mathbf{r}}''(t)(\varphi'(t))^2 + \varphi'(t)R(\tilde{\mathbf{r}}'(t)) \cdot \mathbf{r}'(\varphi(t))\varphi''(t)}{(\varphi'(t))^3\|\tilde{\mathbf{r}}'(t)\|^3} \\
 &= \frac{\varphi'(t)R(\tilde{\mathbf{r}}'(t)) \cdot \tilde{\mathbf{r}}''(t)(\varphi'(t))^2 + 0}{(\varphi'(t))^3\|\tilde{\mathbf{r}}'(t)\|^3} \\
 &= \frac{\varphi'(t)R(\tilde{\mathbf{r}}'(t)) \cdot \tilde{\mathbf{r}}''(t)(\varphi'(t))^2}{(\varphi'(t))^3\|\tilde{\mathbf{r}}'(t)\|^3} \\
 &= \frac{(\varphi'(t))^3 R(\tilde{\mathbf{r}}'(t)) \cdot \tilde{\mathbf{r}}''(t)}{(\varphi'(t))^3\|\tilde{\mathbf{r}}'(t)\|^3} \\
 &= \frac{R(\tilde{\mathbf{r}}'(t)) \cdot \tilde{\mathbf{r}}''(t)}{\|\tilde{\mathbf{r}}'(t)\|^3}.
 \end{aligned}$$

[Where did we use $\varphi'(t) > 0$?]

□

It follows from Proposition 1.56 that $|\kappa_s(t)| = \kappa(t)$ even for non-unit-speed curves.

Proposition 1.57. Let $\mathbf{r}(t) = (x(t), y(t))$ be a regular plane curve. Then the signed curvature function is given by

$$\kappa_s(t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{(x'(t)^2 + y'(t)^2)^{\frac{3}{2}}}.$$

Proof. [Finish...]

□

Example 1.58. Consider the plane curve $\mathbf{r}(t) = (t, \ln(\cos t))$, $-\frac{\pi}{2} < t < \frac{\pi}{2}$. Then

$$\begin{aligned}
 \mathbf{r}'(t) &= (1, -\tan t), \\
 \mathbf{r}''(t) &= (0, -\sec^2 t).
 \end{aligned}$$

The unit tangent vector is

$$\begin{aligned}
 \hat{\mathbf{t}}(t) &= \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|} \\
 &= \frac{1}{\sec t}(1, -\tan t) \\
 &= (\cos t, -\sin t).
 \end{aligned}$$

The unit normal vector is then

$$\begin{aligned}
 \hat{\mathbf{n}}(t) &= \frac{\hat{\mathbf{t}}'(t)}{\|\hat{\mathbf{t}}'(t)\|} \\
 &= (-\sin t, \cos t).
 \end{aligned}$$

The signed curvature function is

$$\begin{aligned}
 \kappa_s(t) &= \frac{1 \cdot (-\sec^2 t) - 0}{1 + \tan^2 t} \\
 &= \frac{-\sec^2 t}{\sec^2 t} \\
 &= -\sec t
 \end{aligned}$$

and the curvature function is then

$$\kappa(t) = |\kappa_s(t)| = |\sec t| = \sec t$$

since $\sec t > 0$ for all $-\frac{\pi}{2} < t < \frac{\pi}{2}$.

1.10 Local-global connection

Recall the following theorem about triangles:

Theorem 1.59 (Angle-Sum Theorem). The sum of the interior angles of a triangle is 180° .

This theorem shows how local data (the angle measures) influences the global shape by giving the condition for the three angles to fit together to form a triangle. In this section, we will prove a theorem for plane curves of a similar flavor.

Definition 1.60 (Simple closed curve). A *closed curve* is a regular curve of the form $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$ such that $\mathbf{r}(a) = \mathbf{r}(b)$ and all derivatives match:

$$\mathbf{r}'(a) = \mathbf{r}'(b), \quad \mathbf{r}''(a) = \mathbf{r}''(b), \text{ etc.}$$

If additionally \mathbf{r} is 1-1 on the domain $[a, b)$, then it is called a *simple closed curve*. (See the Figure below.)

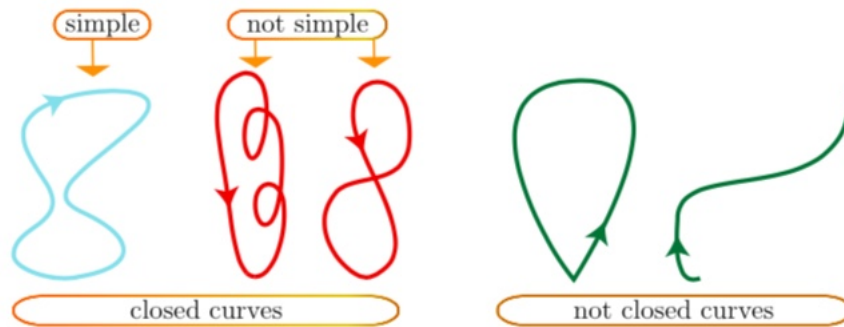


Figure 10: A simple closed curve closes up smoothly and does not have any self-intersections.

Example 1.61. The circle $\mathbf{r}(t) = (\cos t, \sin t), t \in [0, 2\pi]$ is a simple closed curve.

In this section we will prove the following theorem:

Theorem 1.62 (Total Curvature Theorem). If $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^2$ is a unit speed simple closed curve. Then

$$\int_a^b \kappa_s(t) dt = \pm 2\pi.$$

[Explain the sign.]

Proposition 1.63. Let $\mathbf{r} : I \rightarrow \mathbb{R}^2$ be a unit-speed plane curve. Then there exists a smooth *angle function*, $\theta : I \rightarrow \mathbb{R}$, such that for all $t \in I$, we have

$$\mathbf{v}(t) = (\cos \theta(t), \sin \theta(t)). \quad (1.10)$$

This function is unique up to adding an integer multiple of 2π .

Proof.

1. First, note that for each $t_0 \in I$, there exists a neighborhood $U \subseteq I$ containing t_0 and a smooth *local angle function* $\theta : U \rightarrow \mathbb{R}$ such that

$$\mathbf{v}(t) = (\cos \theta(t), \sin \theta(t)).$$

To see this, write \mathbf{v} in terms of its components as $\mathbf{v}(t) = (v_x(t), v_y(t))$. Since the curve is regular, $\mathbf{v}(t) \neq 0$ for all $t \in I$. It follows that, for all $t \in I$, at least one of the component functions is non-zero. Since the component functions v_i ($i = x, y$) are continuous, if $v_i(t_0) \neq 0$, then $v_i(t)$ must be non-zero throughout a neighborhood containing t_0 .³ Thus, each $t_0 \in I$ has a neighborhood where at least one of the component functions is non-zero. Explicitly, the following local angle functions are well-defined on each type of interval: [\[Add explanation of how to get \(or at least verify\) these. Maybe add a figure?\]](#)

- If $v_x > 0$ on U , for each $t \in U$ define $\theta : U \rightarrow \mathbb{R}$ by⁴

$$\theta(t) = \tan^{-1}\left(\frac{v_y}{v_x}\right)$$

- If $v_y > 0$ on U , for each $t \in U$ define $\theta : U \rightarrow \mathbb{R}$ by

$$\theta(t) = \cos^{-1}(v_x)$$

- If $v_x < 0$ on U , for each $t \in U$ define $\theta : U \rightarrow \mathbb{R}$ by

$$\theta(t) = \pi - \tan^{-1}\left(\frac{v_y}{v_x}\right)$$

- If $v_y < 0$ on U , for each $t \in U$ define $\theta : U \rightarrow \mathbb{R}$ by

$$\theta(t) = 2\pi - \cos^{-1}(v_x)$$

These are, of course, far from unique. While this shows each point has a neighborhood on which we can define a local angle function, to prove the theorem we need to define a *global angle function*, which is defined on *all* of I .

³Proof: Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be continuous and non-zero at some point x_0 . Let $\epsilon = \frac{|f(x_0)|}{2}$. Then $\epsilon > 0$ so by continuity at x_0 there exists $\delta > 0$ such that $|f(x) - f(x_0)| < \frac{|f(x_0)|}{2}$ whenever $|x - x_0| < \delta$. If there is a point x in the neighborhood $|x - x_0| < \delta$ for which $f(x) = 0$, then we have $|0 - f(x_0)| = |f(x_0)| < \frac{|f(x_0)|}{2}$, which is a contradiction. Thus, $f(x)$ must be nonzero throughout a neighborhood of x_0 .

⁴Since $\|\mathbf{v}(t)\| = 1$ for all $t \in I$, $\mathbf{v}(t)$ lies on the unit circle. If $v_x > 0$, then we must have $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ up to an integer multiple of 2π . Basic trigonometry then gives the stated formula for θ . The rest of the stated functions are obtained similarly. The actual neighborhood U on which each local angle function is defined will depend on the functions $v_x(t)$ and $v_y(t)$.

2. To construct such a function, we will need the following result. Let θ be a local angle function defined on a neighborhood U of a point t_0 . Then for all $t \in U$, we have

$$\begin{aligned}\mathbf{v}(t) &= (\cos \theta(t), \sin \theta(t)) \\ \mathbf{a}(t) &= \theta'(t)(-\sin \theta(t), \cos \theta(t)) \\ R(\mathbf{v}(t)) &= (-\sin \theta(t), \cos \theta(t))\end{aligned}$$

and therefore

$$\kappa_s(t) \equiv R(\mathbf{v}(t)) \cdot \mathbf{a}(t) = \theta'(t)(\sin^2 \theta(t) + \cos^2 \theta(t)) = \theta'(t)$$

for all $t \in U$. Suppose we have another function $\psi(t)$ defined on U such that

- (a) $(\cos \psi(t_0), \sin \psi(t_0)) = (\cos \theta(t_0), \sin \theta(t_0)) = \mathbf{v}(t_0)$, and
- (b) $\psi'(t) = \kappa_s(t)$ for all $t \in U$.

Then by (b) we have for all $t \in U$

$$\psi'(t) - \theta'(t) = (\psi(t) - \theta(t))' = 0$$

and therefore $\psi(t) = \theta(t) + c$, where c is a constant. It then follows from (a) that $c = 2\pi n$ for some $n \in \mathbb{Z}$. Thus, $\psi(t) = \theta(t) + 2\pi n$ ($n \in \mathbb{Z}$) for all $t \in U$.

3. We now construct a global angle function. Let $t_0 \in I$. Choose a neighborhood V of t_0 such that a local angle function $\theta(t)$ is defined on V , and let $\theta_0 = \theta(t_0)$. Define $\Theta : I \rightarrow \mathbb{R}$ for all $t \in I$ by

$$\Theta(t) = \int_{t_0}^t \kappa_s(u) du + \theta_0.$$

By the Fundamental Theorem of Calculus, $\Theta'(t) = \kappa_s(t)$ for all $t \in I$. Let

$$\hat{I} = \{t \in I : \mathbf{v}(t) = (\cos \Theta(t), \sin \Theta(t))\}.$$

In particular, $t_0 \in \hat{I}$ since $\Theta(t_0) = \theta_0$, and $\mathbf{v}(t_0) = (\cos \theta_0, \sin \theta_0)$, so $\hat{I} \neq \emptyset$. Suppose then that t_1 is any element of \hat{I} . Then $\mathbf{v}(t_1) = (\cos \Theta(t_1), \sin \Theta(t_1))$. By the previous discussion, there exists a neighborhood U of t_1 and a local angle function θ defined on U such that $\mathbf{v}(t) = (\cos \theta(t), \sin \theta(t))$ for all $t \in U$. Since $\Theta'(t) = \theta'(t)$ for all $t \in U$, and $(\cos \Theta(t_1), \sin \Theta(t_1)) = (\cos \theta(t_1), \sin \theta(t_1)) = \mathbf{v}(t_1)$, by the previous discussion $\Theta(t) = \theta(t) + 2\pi n$ on U , so $(\cos \Theta(t), \sin \Theta(t)) = (\cos(\theta(t) + 2\pi n), \sin(\theta(t) + 2\pi n)) = (\cos \theta(t), \sin \theta(t)) = \mathbf{v}(t)$ for all $t \in U$. This shows that each $t \in \hat{I}$ has a neighborhood U contained in \hat{I} , and therefore \hat{I} is open in I .

Now suppose $t_1 \in I - \hat{I}$. Then $\mathbf{v}(t_1) \neq (\cos \Theta(t_1), \sin \Theta(t_1))$. It follows that there is a neighborhood U_1 of t_1 on which $\mathbf{v}(t) \neq (\cos \Theta(t), \sin \Theta(t))$ for all $t \in U_1$. To see this, suppose otherwise. Then, given any neighborhood U of t_1 , there exists $t_2 \in U - \{t_1\}$ such that $\mathbf{v}(t_2) = (\cos \Theta(t_2), \sin \Theta(t_2))$. Let U be a neighborhood of t_1 on which the local angle function $\theta(t)$ is defined. By the same argument as above, we then have $\mathbf{v}(t) = (\cos \theta(t), \sin \theta(t)) = (\cos \Theta(t), \sin \Theta(t))$, for all $t \in U$, contradicting the fact that $\mathbf{v}(t_1) \neq (\cos \Theta(t_1), \sin \Theta(t_1))$. This shows that every point of $I - \hat{I}$ has a neighborhood contained in $I - \hat{I}$, so $I - \hat{I}$ is open in I and therefore \hat{I} is closed on I . Since I is an interval in \mathbb{R} , I is connected and therefore

has no nonempty subset that is both open and closed. We conclude that $\hat{I} = I$, proving that Θ is a global angle function on I . [\[Need an appendix on basic topological ideas used in the text.\]](#)

If ψ is another global angle function, then $\mathbf{v}(t) = (\cos \Theta(t), \sin \Theta(t)) = (\cos \psi(t), \sin \psi(t))$ for all $t \in I$, and therefore $\psi(t) = \Theta(t) + 2\pi n$, $n \in \mathbb{Z}$ for all $t \in I$, which shows that Θ is unique up to adding an integer multiple of 2π .

□

Definition 1.64 (Rotation index). The *rotation index* of a unit-speed closed plane curve $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^2$ is the integer

$$\text{Ind}(\mathbf{r}) = \frac{1}{2\pi}(\Theta(b) - \Theta(a))$$

where Θ is the global angle function of Proposition 1.63.

Exercise 1.11. Let $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^2$ be a regular (not necessarily unit-speed) closed plane curve.

- (1) Prove there exists a smooth function $\Theta : [a, b] \rightarrow \mathbb{R}$ such that for all $t \in [a, b]$, the unit tangent vector satisfies

$$\hat{\mathbf{t}}(t) = (\cos \Theta(t), \sin \Theta(t))$$

and that the rotation index of \mathbf{r} equals $\frac{1}{2\pi}(\Theta(b) - \Theta(a))$.

- (2) Prove that the rotation index of \mathbf{r} equals

$$\text{Ind}(\mathbf{r}) = \frac{1}{2\pi} \int_a^b \kappa_s(t) \|\mathbf{r}'(t)\| dt.$$

Hint: For both parts, verify that the formula is valid for unit-speed curves and is unchanged by orientation-preserving reparametrizations.

Example 1.65.

1.11 Space Curves

In this section, we will consider regular curves $\mathbf{r} : I \rightarrow \mathbb{R}^3$ where $\kappa(t) \neq 0$ for all $t \in I$. In the previous section, we found that the unit tangent vector, unit normal vector, and curvature function are given by

$$\begin{aligned} \hat{\mathbf{t}}(t) &= \frac{\mathbf{r}'(t)}{\|\mathbf{r}'(t)\|}, & \hat{\mathbf{n}}(t) &= \frac{\hat{\mathbf{t}}'(t)}{\|\hat{\mathbf{t}}'(t)\|}, \\ \kappa(t) &= \frac{\|\mathbf{r}''_{\perp}(t)\|}{\|\mathbf{r}'(t)\|^2}. \end{aligned}$$

In \mathbb{R}^3 , we can use the cross-product to simplify the computation of the curvature function.

Proposition 1.66. Let $\mathbf{r} : I \rightarrow \mathbb{R}^3$ be a regular space curve. Then for all $t \in I$,

$$\kappa(t) = \frac{\|\mathbf{r}'(t) \times \mathbf{r}''(t)\|}{\|\mathbf{r}'(t)\|^3}.$$

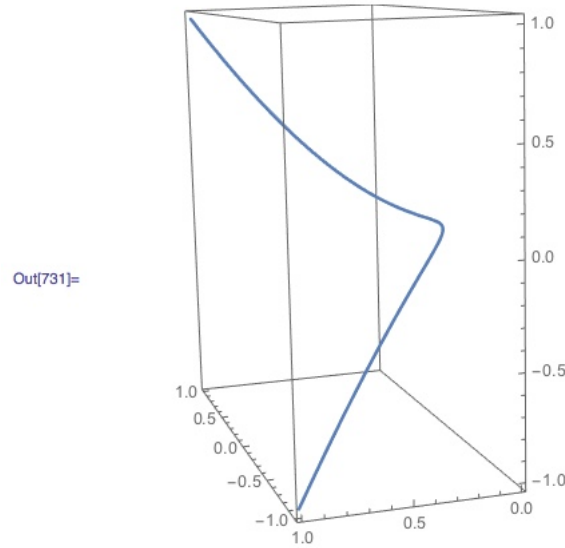
Proof. Since $\|\mathbf{r}_{\perp}''(t)\| = \|\mathbf{r}''(t)\| \sin \theta$ (where θ is the angle between $\mathbf{r}'(t)$ and $\mathbf{r}''(t)$), we have

$$\kappa(t) = \frac{\|\mathbf{r}_{\perp}''(t)\|}{\|\mathbf{r}'(t)\|^2} = \frac{\|\mathbf{r}''(t)\| \sin \theta}{\|\mathbf{r}'(t)\|^2} = \frac{\|\mathbf{r}'(t)\| \|\mathbf{r}''(t)\|}{\|\mathbf{r}'(t)\|^3} = \frac{\|\mathbf{r}'(t) \times \mathbf{r}''(t)\|}{\|\mathbf{r}'(t)\|^3}.$$

□

Example 1.67. The space curve $\mathbf{r}(t) = (t, t^2, t^3)$, $t \in \mathbb{R}$ is called the *twisted cubic*. The unit tangent

In[731]:= ParametricPlot3D[{t, t², t³}, {t, -1, 1}]



and unit normal vectors are rather tedious to compute by hand. These can easily be computed in *Mathematica* as follows:

One can easily verify that these vectors form an orthonormal set: The curvature function is then given by

$$\begin{aligned} \kappa(t) &= \frac{1}{(1 + 4t^2 + 9t^4)^{3/2}} \left\| \begin{pmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ 1 & 2t & 3t^2 \\ 0 & 2 & 6t \end{pmatrix} \right\| \\ &= \frac{2\|(1, -3t, 3t^2)\|}{(1 + 4t^2 + 9t^4)^{3/2}} \\ &= \frac{2\sqrt{1 + 9t^2 + 9t^4}}{(1 + 4t^2 + 9t^4)^{3/2}} \\ &= 2\sqrt{\frac{1 + 9t^2 + 9t^4}{(1 + 4t^2 + 9t^4)^3}}, \end{aligned}$$

or using *Mathematica*:

```

In[752]:= r[t_] := {t, t^2, t^3};

In[753]:= T[t_] := Simplify[ $\frac{\mathbf{r}'[t]}{\text{Norm}[\mathbf{r}'[t]]}$ , Assumptions -> t ∈ Reals]

In[754]:= n[t_] := Simplify[ $\frac{\mathbf{T}'[t]}{\text{Norm}[\mathbf{T}'[t]]}$ , Assumptions -> t ∈ Reals]

In[762]:= T[t]
Out[762]:=  $\left\{ \frac{1}{\sqrt{1 + 4 t^2 + 9 t^4}}, \frac{2 t}{\sqrt{1 + 4 t^2 + 9 t^4}}, \frac{3 t^2}{\sqrt{1 + 4 t^2 + 9 t^4}} \right\}$ 

In[764]:= n[t]
Out[764]:=  $\left\{ -\frac{t (2 + 9 t^2)}{\sqrt{(1 + 4 t^2 + 9 t^4) (1 + 9 t^2 + 9 t^4)}}, \frac{1 - 9 t^4}{\sqrt{1 + 13 t^2 + 54 t^4 + 117 t^6 + 81 t^8}}, \frac{3 t (1 + 2 t^2)}{\sqrt{(1 + 4 t^2 + 9 t^4) (1 + 9 t^2 + 9 t^4)}} \right\}$ 

In[767]:= Simplify[Norm[T[t]], Assumptions -> t ∈ Reals]
Out[767]:= 1

In[768]:= Simplify[Norm[n[t]], Assumptions -> t ∈ Reals]
Out[768]:= 1

In[769]:= Simplify[Dot[T[t], n[t]], Assumptions -> t ∈ Reals]
Out[769]:= 0
    
```

We can also use the cross product to add one more vector to the set $\{\hat{\mathbf{t}}(t), \hat{\mathbf{n}}(t)\}$.

Definition 1.68 (Unit binormal vector). The vector

$$\hat{\mathbf{b}}(t) = \hat{\mathbf{t}}(t) \times \hat{\mathbf{n}}(t)$$

is called the *unit binormal vector*.

Proposition 1.69. The set $\{\hat{\mathbf{t}}(t), \hat{\mathbf{n}}(t), \hat{\mathbf{b}}(t)\}$ is an orthonormal basis for \mathbb{R}^3 .

Proof. We have already seen that $\{\hat{\mathbf{t}}(t), \hat{\mathbf{n}}(t)\}$ is an orthonormal set. It follows immediately from the properties of the cross product [\[Reference this in my linear algebra notes.\]](#) that $\hat{\mathbf{b}}(t)$ is orthogonal to both $\hat{\mathbf{t}}(t)$ and $\hat{\mathbf{n}}(t)$. Also

$$\|\hat{\mathbf{b}}(t)\| = \|\hat{\mathbf{t}}(t)\| \|\hat{\mathbf{n}}(t)\| \left| \sin\left(\frac{\pi}{2}\right) \right| = 1 \cdot 1 \cdot 1 = 1.$$

Thus, $\{\hat{\mathbf{t}}(t), \hat{\mathbf{n}}(t), \hat{\mathbf{b}}(t)\}$ is an orthonormal basis for \mathbb{R}^3 . □

Definition 1.70 (Frenet frame). A basis for \mathbb{R}^n together with a choice of origin is called a *frame*. The orthonormal frame $\{\hat{\mathbf{t}}(t), \hat{\mathbf{n}}(t), \hat{\mathbf{b}}(t)\}$ for \mathbb{R}^3 (whose origin is understood to be $\mathbf{r}(t)$) is called the *Frenet frame*.

$$\begin{aligned} \text{In[760]} &:= \kappa = \text{Simplify}\left[\frac{\text{Norm}[\text{Cross}[\mathbf{r}'[t], \mathbf{r}''[t]]]}{(\text{Norm}[\mathbf{r}'[t]])^3}, \text{Assumptions} \rightarrow t \in \text{Reals}\right] \\ \text{Out[760]} &= 2 \sqrt{\frac{1 + 9t^2 + 9t^4}{(1 + 4t^2 + 9t^4)^3}} \end{aligned}$$

The curvature function of the curve \mathbf{r} gave us a quantitative measure of to what extent the trace of the curve fails to be a straight line in a neighborhood of the point $\mathbf{r}(t)$. We now want a quantitative measure of the extent to which the curve fails to lie in the osculating plane in a neighborhood of $\mathbf{r}(t)$. Similar to the definition of curvature, a good candidate for such a function is $\|\mathbf{b}'(t)\|$, which gives the rate at which the osculating plane is tilting at $\mathbf{r}(t)$. However, we will instead find it more useful to define a *signed* measurement of the rate at which the osculating plane is tilting.

First, note that $\hat{\mathbf{b}}'(t)$ is parallel to $\hat{\mathbf{n}}(t)$. To see this, note that since $\{\hat{\mathbf{t}}(t), \hat{\mathbf{n}}(t), \hat{\mathbf{b}}(t)\}$ is an orthonormal basis, we can expand $\hat{\mathbf{b}}'$ in this basis as [\[Reference this in my linear algebra notes.\]](#)

$$\hat{\mathbf{b}}'(t) = (\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{t}}(t))\hat{\mathbf{t}}(t) + (\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{n}}(t))\hat{\mathbf{n}}(t) + (\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{b}}(t))\hat{\mathbf{b}}(t).$$

Since $\|\hat{\mathbf{b}}(t)\| = 1$ for all $t \in I$, by part (a) of Lemma 1.44 we have $\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{b}}(t) = 0$. By part (b) of the same lemma we also have

$$\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{t}}(t) = -\hat{\mathbf{b}}(t) \cdot \hat{\mathbf{t}}'(t) = -\hat{\mathbf{b}}(t) \cdot (\|\mathbf{t}'(t)\|\hat{\mathbf{n}}(t)) = -\|\mathbf{t}'(t)\|(\hat{\mathbf{b}}(t) \cdot \hat{\mathbf{n}}(t)) = 0,$$

since $\hat{\mathbf{b}}(t)$ and $\hat{\mathbf{n}}(t)$ are orthogonal. Thus, $\hat{\mathbf{b}}'(t) = (\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{n}}(t))\hat{\mathbf{n}}(t)$.

Definition 1.71 (Torsion for unit speed curve). Let $\mathbf{r} : I \rightarrow \mathbb{R}^3$ be a regular curve parametrized by arc length s , such that $\kappa(s) \neq 0$ for all $s \in I$. The *torsion function* of \mathbf{r} is defined by

$$\tau(s) = -\hat{\mathbf{b}}'(s) \cdot \hat{\mathbf{n}}(s).$$

The minus sign in the definition above is conventional, and many textbooks define torsion without this sign. Before we move on to interpreting this formula, we need a formula valid for a regular curve with arbitrary parametrization.

Proposition 1.72. Let $\mathbf{r} : I \rightarrow \mathbb{R}^3$ be a regular space curve with an arbitrary parametrization. Then for every $t \in I$ with $\kappa(t) \neq 0$, the expression

$$\tau(t) = \frac{-\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{n}}(t)}{\|\mathbf{r}'(t)\|} \quad (1.11)$$

is invariant under reparametrizations and agrees with the torsion function $\tau(s)$ of Definition 1.71.

Proof. If the curve is parametrized by arc length, then $\|\mathbf{r}'(t)\| = 1$ for all t and thus the formula becomes that of Definition 1.71.

To see that this formula is independent of parametrization, let us first consider the transformation of the vectors in the Frenet frame change under a reparametrization:

$$\begin{aligned}\tilde{\mathbf{t}}(t) &= \frac{\tilde{\mathbf{r}}'(t)}{\|\tilde{\mathbf{r}}'(t)\|} = \frac{\varphi'(t)\mathbf{r}'(\varphi(t))}{\|\varphi'(t)\mathbf{r}'(\varphi(t))\|} = \frac{\varphi'(t)\mathbf{r}'(\varphi(t))}{|\varphi'(t)|\|\mathbf{r}'(\varphi(t))\|} = \text{sgn}(\varphi'(t))\hat{\mathbf{t}}(\varphi(t)) \\ \tilde{\mathbf{n}}(t) &= \frac{\tilde{\mathbf{r}}''_{\perp}(t)}{\|\tilde{\mathbf{r}}''_{\perp}(t)\|} = \frac{(\varphi'(t))^2\mathbf{r}''_{\perp}(\varphi(t))}{\|(\varphi'(t))^2\mathbf{r}''_{\perp}(\varphi(t))\|} = \frac{(\varphi'(t))^2\mathbf{r}''_{\perp}(\varphi(t))}{(\varphi'(t))^2\|\mathbf{r}''_{\perp}(\varphi(t))\|} = \hat{\mathbf{n}}(\varphi(t)) \\ \tilde{\mathbf{b}}(t) &= \tilde{\mathbf{t}}(t) \times \tilde{\mathbf{n}}(t) = \text{sgn}(\varphi'(t))\hat{\mathbf{b}}(\varphi(t))\end{aligned}$$

If φ is an orientation-preserving reparametrization ($\varphi'(t) > 0$), then $\tilde{\mathbf{t}} = \hat{\mathbf{t}} \circ \varphi$, $\tilde{\mathbf{n}} = \hat{\mathbf{n}} \circ \varphi$, $\tilde{\mathbf{b}} = \hat{\mathbf{b}} \circ \varphi$, and

$$\tilde{\tau}(t) = -\frac{\tilde{\mathbf{b}}'(t) \cdot \tilde{\mathbf{n}}(t)}{\|\tilde{\mathbf{r}}'(t)\|} = \frac{-\varphi'(t)\hat{\mathbf{b}}(\varphi(t)) \cdot \hat{\mathbf{n}}(\varphi(t))}{\|\varphi'(t)\mathbf{r}'(\varphi(t))\|} = \frac{-\hat{\mathbf{b}}(\varphi(t)) \cdot \hat{\mathbf{n}}(\varphi(t))}{\|\mathbf{r}'(\varphi(t))\|} = \tau(\varphi(t)).$$

If φ is an orientation-reversion reparametrization ($\varphi'(t) < 0$), then $\tilde{\mathbf{t}} = -\hat{\mathbf{t}} \circ \varphi$, $\tilde{\mathbf{n}} = \hat{\mathbf{n}} \circ \varphi$, $\tilde{\mathbf{b}} = -\hat{\mathbf{b}} \circ \varphi$, and

$$\tilde{\tau}(t) = -\frac{\tilde{\mathbf{b}}'(t) \cdot \tilde{\mathbf{n}}(t)}{\|\tilde{\mathbf{r}}'(t)\|} = \frac{\varphi'(t)\hat{\mathbf{b}}(\varphi(t)) \cdot \hat{\mathbf{n}}(\varphi(t))}{-\varphi'(t)\|\mathbf{r}'(\varphi(t))\|} = \frac{-\hat{\mathbf{b}}(\varphi(t)) \cdot \hat{\mathbf{n}}(\varphi(t))}{\|\mathbf{r}'(\varphi(t))\|} = \tau(\varphi(t)).$$

so the sign changes in the formula cancel, yielding the same conclusion: $\tilde{\tau} = \tau \circ \varphi$. Thus, the expression in (1.12) is invariant under reparametrizations. \square

Definition 1.73. Let $\mathbf{r} : I \rightarrow \mathbb{R}^3$ be a regular space curve with an arbitrary parametrization. Then for every $t \in I$ with $\kappa(t) \neq 0$, we define the *torsion function* of the curve by the expression

$$\tau(t) = \frac{-\hat{\mathbf{b}}'(t) \cdot \hat{\mathbf{n}}(t)}{\|\mathbf{r}'(t)\|}. \quad (1.12)$$

Proposition 1.74. Let $\mathbf{r} : I \rightarrow \mathbb{R}^3$ be a regular space curve with $\kappa(t) \neq 0$ for all $t \in I$. Then the trace of \mathbf{r} is constrained to a plane if and only if $\tau(t) = 0$ for all $t \in I$.

Proof. Let $\mathbf{w} = (a, b, c)$ be a fixed vector in \mathbb{R}^3 to serve as the normal vector of our plane. First, suppose that the trace of \mathbf{r} is constrained to the plane

$$P = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \cdot \mathbf{w} = d\}.$$

Since the trace of \mathbf{r} lies in the plane P , $\mathbf{r}(t) \cdot \mathbf{w} = d$ is a constant function of t , so its derivatives vanish:

$$\begin{aligned}0 &= (\mathbf{r}(t) \cdot \mathbf{w})' = \mathbf{r}'(t) \cdot \mathbf{w} \\ 0 &= (\mathbf{r}(t) \cdot \mathbf{w})'' = \mathbf{r}''(t) \cdot \mathbf{w}.\end{aligned}$$

From this it follows that $\hat{\mathbf{t}}(t)$ and $\hat{\mathbf{n}}(t)$ are both orthogonal to \mathbf{w} , so their cross product must be parallel to \mathbf{w} : $\hat{\mathbf{b}}(t) = \pm \frac{\mathbf{w}}{\|\mathbf{w}\|}$. Since $\hat{\mathbf{b}}(t)$ is a continuous function of t , this sign cannot change abruptly, so it must be constant on I . Thus, $\hat{\mathbf{b}}$ is constant, so $\tau(t) = 0$ for all $t \in I$.

Conversely, suppose that $\tau(t) = 0$ for all $t \in I$. This implies that $\hat{\mathbf{b}}'(t) = 0$ for all $t \in I$, so $\hat{\mathbf{b}}(t) = \mathbf{w}$ (a constant vector) for all $t \in I$. Notice that

$$(\mathbf{w} \cdot \mathbf{r}(t))' = \mathbf{w} \cdot \mathbf{r}'(t) = \|\mathbf{r}'(t)\|(\hat{\mathbf{b}}(t) \cdot \hat{\mathbf{t}}(t)) = 0,$$

since $\hat{\mathbf{b}}(t)$ and $\hat{\mathbf{t}}(t)$ are orthogonal. This $\mathbf{w} \cdot \mathbf{r}(t) = d$ is a constant function. In other words, the trace of \mathbf{r} lies in the plane

$$P = \{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \cdot \mathbf{w} = d\}.$$

□

Thus, roughly, torsion measures the failure of the trace of the curve to remain in a single plane. To formulate this idea more precisely, we must first compute the derivatives of the vectors in the Frenet frame. In the following analysis, we will assume that \mathbf{r} is parametrized by arc length.

Proposition 1.75. Let $\mathbf{r} : I \rightarrow \mathbb{R}^3$ be a regular curve parametrized by arc length. At every time $s \in I$ with $\kappa(s) \neq 0$, the derivatives of the unit tangent, normal, and binormal vectors are given by the *Frenet equations*:

$$\begin{aligned}\hat{\mathbf{t}}'(s) &= \kappa(s)\hat{\mathbf{n}}(s) \\ \hat{\mathbf{n}}'(s) &= -\kappa(s)\hat{\mathbf{t}}(s) + \tau(s)\hat{\mathbf{b}}(s) \\ \hat{\mathbf{b}}'(s) &= -\tau(s)\hat{\mathbf{n}}(s).\end{aligned}$$

Proof. The first and third of these follow immediately from the definitions of κ and τ , since

$$\begin{aligned}\kappa(s) &= \hat{\mathbf{t}}'(s) \cdot \hat{\mathbf{n}}(s) \\ \tau(s) &= -\hat{\mathbf{b}}'(s) \cdot \hat{\mathbf{n}}(s).\end{aligned}$$

To show the second, expand $\hat{\mathbf{n}}'(s)$ in the Frenet frame as

$$\begin{aligned}\hat{\mathbf{n}}'(s) &= \underbrace{(\hat{\mathbf{n}}'(s) \cdot \hat{\mathbf{t}}(s))}_{=-\hat{\mathbf{n}}(s) \cdot \hat{\mathbf{t}}'(s) = -\kappa(s)} \hat{\mathbf{t}}(s) + \underbrace{(\hat{\mathbf{n}}'(s) \cdot \hat{\mathbf{n}}(s))}_{=0} \hat{\mathbf{n}}(s) + \underbrace{(\hat{\mathbf{n}}'(s) \cdot \hat{\mathbf{b}}(s))}_{=-\hat{\mathbf{n}}(s) \cdot \hat{\mathbf{b}}'(s) = \tau(s)} \hat{\mathbf{b}}(s) \\ &= -\kappa(s)\hat{\mathbf{t}}(s) + \tau(s)\hat{\mathbf{b}}(s).\end{aligned}$$

□

We can now describe more precisely how torsion measures the failure of the trace of the curve to remain in a single plane. Assume now that $\mathbf{r} : I \rightarrow \mathbb{R}^3$ is a unit-speed curve, and that $s \in I$ with $\kappa(s) \neq 0$. We saw above that the trace of a second-order Taylor polynomial for \mathbf{r} at s is a parabola in the osculating plane. We will now show that if $\tau(s) \neq 0$, the third-order Taylor polynomial for \mathbf{r} at s leaves the osculating plane.

The third-order Taylor polynomial for the displacement vector $\mathbf{D}(h) \equiv \mathbf{r}(s+h) - \mathbf{r}(s)$ at $s \in I$ is given by

$$\mathbf{D}(h) \equiv \mathbf{r}(s+h) - \mathbf{r}(s) = h\mathbf{r}'(s) + \frac{h^2}{2}\mathbf{r}''(s) + \frac{h^3}{6}\mathbf{r}'''(s)$$

Using Proposition 1.75,

$$\begin{aligned}\mathbf{r}'(s) &= \hat{\mathbf{t}}(s) \\ \mathbf{r}''(s) &= \kappa(s)\hat{\mathbf{n}}(s) \\ \mathbf{r}'''(s) &= (\kappa(s)\hat{\mathbf{n}}(s))' = \kappa'(s)\hat{\mathbf{n}}(s) + \kappa(s)\hat{\mathbf{n}}'(s) \\ &= \kappa'(s)\hat{\mathbf{n}}(s) + \kappa(s)(-\kappa(s)\hat{\mathbf{t}}(s) + \tau(s)\hat{\mathbf{b}}(s)) \\ &= \kappa'(s)\hat{\mathbf{n}}(s) - (\kappa(s))^2\hat{\mathbf{t}}(s) + \kappa(s)\tau(s)\hat{\mathbf{b}}(s)\end{aligned}$$

so the third-order Taylor polynomial of $\mathbf{D}(h)$ is

$$\begin{aligned}\mathbf{D}(h) &= h\hat{\mathbf{t}}(s) + \kappa(s)\frac{h^2}{2}\hat{\mathbf{n}}(s) + \frac{h^3}{6}(\kappa'(s)\hat{\mathbf{n}}(s) - (\kappa(s))^2\hat{\mathbf{t}}(s) + \kappa(s)\tau(s)\hat{\mathbf{b}}(s)) \\ &= (h - (\kappa(s))^2\frac{h^3}{6})\hat{\mathbf{t}}(s) + (\kappa(s)\frac{h^2}{2} + \kappa'(s)\frac{h^3}{6})\hat{\mathbf{n}}(s) + \kappa(s)\tau(s)\frac{h^3}{6}\hat{\mathbf{b}}(s)\end{aligned}$$

If we choose a coordinate system centered at $\mathbf{r}(s)$ where $\hat{\mathbf{t}}(s) = (1, 0, 0)$, $\hat{\mathbf{n}}(s) = (0, 1, 0)$, and $\hat{\mathbf{b}}(s) = (0, 0, 1)$, then we have

$$[\mathbf{D}(h)]_B = (x(h), y(h), z(h)) = \left(h - (\kappa(s))^2\frac{h^3}{6}, \kappa(s)\frac{h^2}{2} + \kappa'(s)\frac{h^3}{6}, \kappa(s)\tau(s)\frac{h^3}{6} \right).$$

This representation is called the *local canonical form* of \mathbf{r} , in a neighborhood of s . Below, we plot the projections of the trace of \mathbf{r} , for small h , in the tn , tb , and nb planes. The tn plane is the osculating plane, while the tb and nb planes are called the *rectifying plane* and *normal plane*, respectively.

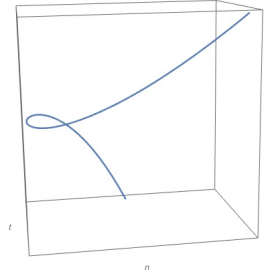


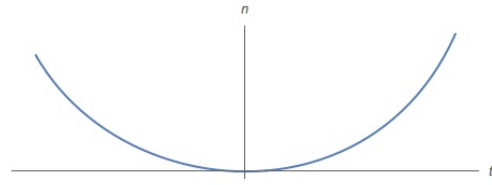
Figure 11: Trace of a regular space curve in the neighborhood of a point with non-zero curvature and torsion.

Since $\kappa(s) > 0$, the Taylor polynomial for $z(h)$ implies that if $\tau(s) > 0$, then $z(h) > 0$ for sufficiently small positive h . On the other hand, if $\tau(s) < 0$, then $z(h) < 0$ for sufficiently small positive h . Thus *positive torsion* at s implies that the curve \mathbf{r} is passing through the osculating plane at s from below. Negative torsion implies from above. Here “above” means the direction of $\hat{\mathbf{b}}(s)$.

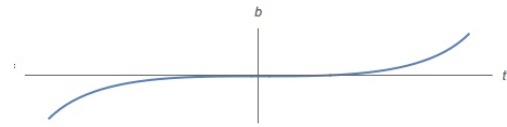
To avoid taking the derivative of the cross product, here is another formula for the torsion which is often more computationally useful.

Proposition 1.76. Let $\mathbf{r} : I \rightarrow \mathbb{R}^3$ be a regular space curve with an arbitrary parametrization. Then for every $t \in I$ with $\kappa(t) \neq 0$, the torsion is given by

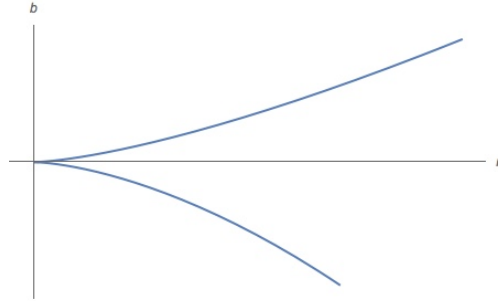
$$\tau(t) = \frac{\mathbf{r}'(t) \times \mathbf{r}''(t) \cdot \mathbf{r}'''(t)}{\|\mathbf{r}'(t) \times \mathbf{r}''(t)\|^2}. \quad (1.13)$$



(a) Projection onto the osculating plane.



(b) Projection onto the rectifying plane.



(c) Projection onto the normal plane.

Figure 12: Projections of a regular space curve in the neighborhood of a point with non-zero curvature and torsion onto the osculating, rectifying, and normal planes.

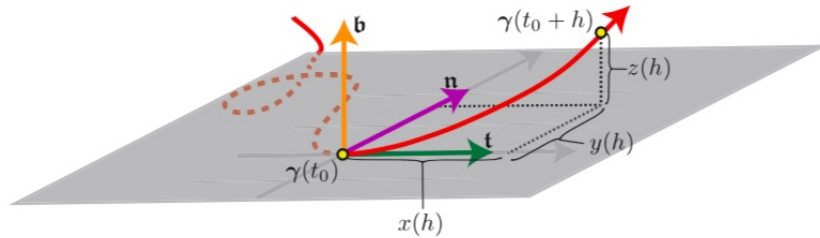


Figure 13: Positive torsion implies that the curve passes through the osculating plane from below.

In physical terms, the vector $\mathbf{r}'''(t)$ (the rate of change of acceleration) is called the *jerk* at time t .

Proof. We will prove that this formula holds for a curve parametrized by arc length, and that it is also independent of parametrization, which then implies it holds for a curve with any parametrization.

For a curve parametrized by arc length, $\mathbf{r}'(s) = \hat{\mathbf{t}}(s)$ and $\mathbf{r}''(s) = \kappa(s)\hat{\mathbf{n}}(s)$, and therefore $\mathbf{r}'(s) \times \mathbf{r}''(s) = \kappa(s)(\hat{\mathbf{t}}(s) \times \hat{\mathbf{n}}(s)) = \kappa(s)\hat{\mathbf{b}}(s)$ and $\|\mathbf{r}'(s) \times \mathbf{r}''(s)\| = \|\kappa(s)\hat{\mathbf{b}}(s)\| = \kappa(s)$. Plugging

into (1.13), we have

$$\begin{aligned}
 \tau(s) &= \frac{\hat{\mathbf{b}}(s) \cdot \mathbf{r}'''(s)}{\kappa(s)} \\
 &= \frac{\hat{\mathbf{b}}(s) \cdot (\kappa(s)\hat{\mathbf{n}}(s))'}{\kappa(s)} \\
 &= \frac{\kappa'(s)\hat{\mathbf{b}}(s) \cdot \hat{\mathbf{n}}(s) + \kappa(s)\hat{\mathbf{b}}(s) \cdot \hat{\mathbf{n}}'(s)}{\kappa(s)} \\
 &= \frac{\kappa'(s) \cdot 0 + \kappa(s)\hat{\mathbf{b}}(s) \cdot \hat{\mathbf{n}}'(s)}{\kappa(s)} \\
 &= \hat{\mathbf{b}}(s) \cdot \hat{\mathbf{n}}'(s) \\
 &= -\hat{\mathbf{b}}'(s) \cdot \hat{\mathbf{n}}(s)
 \end{aligned}$$

in agreement with Definition 1.71.

To check that this formula is independent of parametrization, as we computed previously, if $\tilde{\mathbf{r}} = \mathbf{r} \circ \varphi$, then

$$\begin{aligned}
 \tilde{\mathbf{r}}'(t) &= \varphi'(t)\mathbf{r}'(\varphi(t)) \\
 \tilde{\mathbf{r}}''(t) &= \varphi''(t)\mathbf{r}'(\varphi(t)) + (\varphi'(t))^2\mathbf{r}''(\varphi(t)) \\
 \tilde{\mathbf{r}}'''(t) &= \varphi'''(t)\mathbf{r}'(\varphi(t)) + 3\varphi'(t)\varphi''(t)\mathbf{r}''(\varphi(t)) + (\varphi'(t))^3\mathbf{r}'''(\varphi(t)).
 \end{aligned}$$

Then

$$\begin{aligned}
 \tilde{\mathbf{r}}'(t) \times \tilde{\mathbf{r}}''(t) &= \varphi'(t)\mathbf{r}'(\varphi(t)) \times [\varphi''(t)\mathbf{r}'(\varphi(t)) + (\varphi'(t))^2\mathbf{r}''(\varphi(t))] \\
 &= \varphi'(t)\varphi''(t) \underbrace{\mathbf{r}'(\varphi(t)) \times \mathbf{r}'(\varphi(t))}_{=0} + (\varphi'(t))^3(\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t))) \\
 &= (\varphi'(t))^3(\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t)))
 \end{aligned}$$

and therefore

$$\tilde{\tau}(t) = \frac{(\varphi'(t))^3(\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t))) \cdot (\varphi'''(t)\mathbf{r}'(\varphi(t)) + 3\varphi'(t)\varphi''(t)\mathbf{r}''(\varphi(t)) + (\varphi'(t))^3\mathbf{r}'''(\varphi(t)))}{(\varphi'(t))^6\|\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t))\|^2}$$

Using the cyclic property of the triple product [Refer to my Linear Algebra notes.], we have

$$\mathbf{r}'(t) \times \mathbf{r}''(t) \cdot \mathbf{r}'(t) = \underbrace{\mathbf{r}'(t) \times \mathbf{r}'(t)}_{=0} \cdot \mathbf{r}''(t) = 0$$

and

$$\mathbf{r}'(t) \times \mathbf{r}''(t) \cdot \mathbf{r}''(t) = \underbrace{\mathbf{r}''(t) \times \mathbf{r}''(t)}_{=0} \cdot \mathbf{r}'(t) = 0$$

so the formula becomes

$$\begin{aligned}
 \tilde{\tau}(t) &= \frac{(\varphi'(t))^6(\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t))) \cdot \mathbf{r}'''(\varphi(t))}{(\varphi'(t))^6\|\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t))\|^2} \\
 &= \frac{(\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t))) \cdot \mathbf{r}'''(\varphi(t))}{\|\mathbf{r}'(\varphi(t)) \times \mathbf{r}''(\varphi(t))\|^2} \\
 &= \tau(\varphi(t))
 \end{aligned}$$

which shows τ is invariant under reparametrizations.

□

Example 1.77. Consider the twisted cubic $\mathbf{r}(t) = (t, t^2, t^3)$ of Example 1.67. The unit binormal vector is given by

$$\begin{aligned} \text{In[863]} &:= \text{Simplify}[\text{Cross}[\mathbf{T}[t], \mathbf{n}[t]], \text{Assumptions} \rightarrow t \in \text{Reals}] \\ \text{Out[863]} &:= \left\{ \frac{3t^2}{\sqrt{1+9t^2+9t^4}}, -\frac{3t}{\sqrt{1+9t^2+9t^4}}, \frac{1}{\sqrt{1+9t^2+9t^4}} \right\} \end{aligned}$$

The torsion function is

$$\begin{aligned} \text{In[865]} &:= \tau = \text{Simplify}\left[\frac{\text{Dot}[\text{Cross}[\mathbf{r}'[t], \mathbf{r}''[t]], \mathbf{r}'''[t]]}{(\text{Norm}[\text{Cross}[\mathbf{r}'[t], \mathbf{r}''[t]])^2}, \text{Assumptions} \rightarrow t \in \text{Reals}\right] \\ \text{Out[865]} &:= \frac{3}{1+9t^2+9t^4} \end{aligned}$$

Notice that the torsion takes its minimal value of 3 at $t = 0$, and then strictly decreases as $|t| \rightarrow \infty$.

Physically, we can think of a curve in \mathbb{R}^3 as being obtained from a straight line by bending (curvature) and twisting (torsion).

1.12 The Fundamental Theorems of Plane and Space Curves

Theorem 1.78 (Fundamental Theorem of the Local Theory of Curves). (1) Given a smooth function $\kappa_s(s)$, $s \in I$, then there exists a regular parametrized plane curve $\mathbf{r} : I \rightarrow \mathbb{R}^2$ such that s is the arc length and $\kappa_s(s)$ is the signed curvature. This curve is unique up to rotations and translations.

(2) Given smooth functions $\kappa(s) > 0$ and $\tau(s)$, $s \in I$, there exists a regular parametrized curve $\mathbf{r} : I \rightarrow \mathbb{R}^3$ such that s is the arc length, $\kappa(s)$ is the curvature, and $\tau(s)$ is the torsion of \mathbf{r} . This curve is unique up to rotations and translations.

The complete proof of this theorem involves the theorem of existence and uniqueness of solutions of ordinary differential equations. However, we can prove the construction of the plane curve in part (1).

Proof. (1) Choose any $t_0 \in I$. For this, define the angle function $\Theta : I \rightarrow \mathbb{R}$ by

$$\Theta(t) = \int_{t_0}^t \kappa_s(u) du + \Theta(t_0)$$

and the velocity function as $\mathbf{v}(t) = (\cos \Theta(t), \sin \Theta(t))$. Then define $\mathbf{r} : I \rightarrow \mathbb{R}^2$ by

$$\mathbf{r}(t) = \int_{t_0}^t \mathbf{v}(u) du + \mathbf{r}(t_0).$$

As desired, \mathbf{r} is a unit-speed plane curve for which Θ is an angle function. Its signed curvature function is $\kappa_s = \Theta'$. The initial conditions $\mathbf{r}(t_0), \Theta(t_0)$ can be chosen arbitrarily.

For the second statement, let $\mathbf{r}, \hat{\mathbf{r}} : I \rightarrow \mathbb{R}^2$ be two unit-speed plane curves each of whose signed curvature functions equals κ_s . Each angle function (denoted Θ and $\hat{\Theta}$, respectively) is an antiderivative of κ_s ; since antiderivatives are unique up to an additive constant, we have $\hat{\Theta} = \Theta + \Theta_0$ for some $\Theta_0 \in \mathbb{R}$. Let $\mathbf{v}, \hat{\mathbf{v}}$ denote the velocity functions of the two curves. Then

$$\hat{\mathbf{v}}(t) = (\cos \hat{\Theta}, \sin \hat{\Theta}) = (\cos(\Theta(t) + \Theta_0), \sin(\Theta(t) + \Theta_0)) = L_A \mathbf{v}(t)$$

where $L_A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is the linear transformation which multiplies \mathbf{v} by the rotation matrix

$$A = \begin{bmatrix} \cos \Theta_0 & -\sin \Theta_0 \\ \sin \Theta_0 & \cos \Theta_0 \end{bmatrix}.$$

[Show this.] Let $T_{\mathbf{q}} : \mathbb{R} \rightarrow \mathbb{R}$ denote translation by the fixed vector \mathbf{q} :

$$T_{\mathbf{q}} \mathbf{v} = \mathbf{v} + \mathbf{q}$$

and let $f = T_{\mathbf{q}} \circ L_A$. Consider the curve $\tilde{\mathbf{r}} = f \circ \mathbf{r}$, where $\mathbf{q} \in \mathbb{R}^2$ has been chosen so that $\tilde{\mathbf{r}}(t_0) = \hat{\mathbf{r}}(t_0)$. Since $\tilde{\mathbf{r}}$ and $\hat{\mathbf{r}}$ have the same derivative function and the same initial value, the uniqueness of antiderivatives implies that $\tilde{\mathbf{r}} = \hat{\mathbf{r}}$. [Add some details.] \square

1.13 Summary of formulas

Let $\mathbf{v} \equiv \mathbf{r}'$, $\mathbf{a} \equiv \mathbf{r}''$, $\mathbf{j} \equiv \mathbf{r}'''$ and drop the dependence on the parameter t .

1.13.1 Plane Curves

1. Unit tangent and unit normal vectors

$$\hat{\mathbf{t}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}, \quad \hat{\mathbf{n}} = \frac{\hat{\mathbf{t}}'}{\|\hat{\mathbf{t}}'\|}$$

2. Signed curvature function ($\mathbf{r} = (x, y)$)

$$\kappa_s = \frac{x'y'' - x''y'}{((x')^2 + (y')^2)^{\frac{3}{2}}}.$$

3. Curvature function ($\mathbf{r} = (x, y)$)

$$\kappa = |\kappa_s| = \frac{|x'y'' - x''y'|}{((x')^2 + (y')^2)^{\frac{3}{2}}}.$$

1.13.2 Space Curves

1. Frenet frame

$$\hat{\mathbf{t}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}, \quad \hat{\mathbf{n}} = \frac{\hat{\mathbf{t}}'}{\|\hat{\mathbf{t}}'\|}, \quad \hat{\mathbf{b}} = \hat{\mathbf{t}} \times \hat{\mathbf{n}}.$$

2. Curvature function

$$\kappa = \frac{\|\mathbf{v} \times \mathbf{a}\|}{\|\mathbf{v}\|^3}.$$

3. Torsion function

$$\tau = \frac{\mathbf{v} \times \mathbf{a} \cdot \mathbf{j}}{\|\mathbf{v} \times \mathbf{a}\|^2}.$$

2 Multivariable Functions

2.1 Basic Definitions

In the last section we studied functions with one input and multiple outputs. In this section, we study functions with multiple input and one output. That is, we study functions of the form

$$f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}.$$

Examples abound.

Example 2.1. (a) The volume of a right circular cylinder of radius r and height h is a function $V : U \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $V(r, h) = \pi r^2 h$.

(b) The Euclidean norm on \mathbb{R}^n is a function $\rho : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $\rho(x_1, x_2, \dots, x_n) = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$.

(c) The ideal gas law is a function giving the pressure in terms of the temperature and volume:

$$p : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$p(T, V) = \frac{nRT}{V}.$$

In calculus, one generally defines a function by writing $y = f(x)$, where it is understood that the domain of the function is the largest set for which it is defined. This is common practice for multivariable functions as well, and one writes $z = f(x, y)$ meaning that the value of the function f at $(x, y) \in \mathbb{R}^2$ is z . Similarly, one writes $u = f(x, y, z)$, $w = f(x, y, z, u)$, etc., if f depends on more variables.

Example 2.2. Consider the function $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f(x, y) = \sqrt{y - x^2}$. For the square root to be defined, we must have $y - x^2 \geq 0$. Taking the domain D to be the largest set where the function is defined, we have then $D = \{(x, y) \in \mathbb{R}^2 : y \geq x^2\}$, which is the shaded region above (and including) the parabola $y = x^2$ shown below. Note that D is closed (since it contains its boundary $y = x^2$) and unbounded. (See Appendix B for definitions of these terms.)

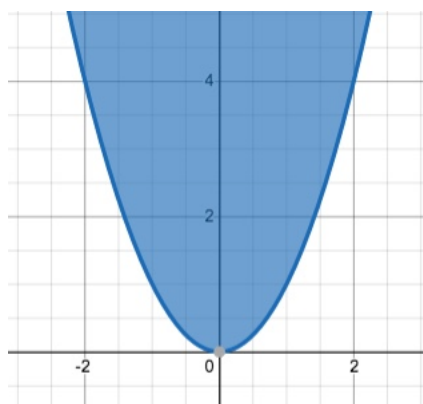


Figure 14: The domain of the function $z = \sqrt{y - x^2}$.

Exercise 2.1. Describe the domain D of the function $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$f(x, y) = \frac{1}{\sqrt{16 - x^2 - y^2}}.$$

Solution. For the square root to be defined, we must have $16 \geq x^2 + y^2$. However, we cannot take $16 = x^2 + y^2$ since then we divide by zero. Therefore

$$D = \{(x, y) : x^2 + y^2 < 16\} = B_4(0)$$

is the open ball of radius 4 centered at $(0, 0)$. Note that D is open and bounded. □

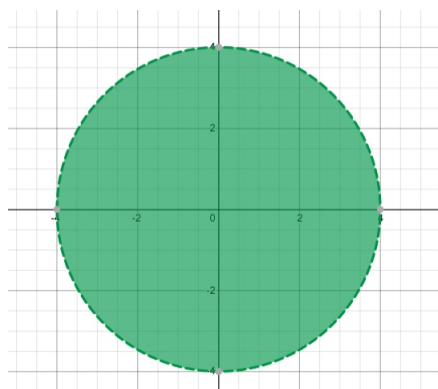


Figure 15: The domain of the function $z = \frac{1}{\sqrt{16 - x^2 - y^2}}$.

Exercise 2.2. For each of the following functions, evaluate $f(3, 2)$ and find and sketch the domain.

1. $f(x, y) = \frac{\sqrt{x+y+1}}{x-1}$
2. $f(x, y) = x \ln(y^2 - x)$

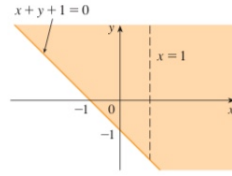
SOLUTION

(a)
$$f(3, 2) = \frac{\sqrt{3+2+1}}{3-1} = \frac{\sqrt{6}}{2}$$

The expression for f makes sense if the denominator is not 0 and the quantity under the square root sign is nonnegative. So the domain of f is

$$D = \{(x, y) \mid x + y + 1 \geq 0, x \neq 1\}$$

The inequality $x + y + 1 \geq 0$, or $y \geq -x - 1$, describes the points that lie on or above



the line $y = -x - 1$, while $x \neq 1$ means that the points on the line $x = 1$ must be excluded from the domain. (See Figure 2.)

$$(b) \quad f(3, 2) = 3 \ln(2^2 - 3) = 3 \ln 1 = 0$$

Since $\ln(y^2 - x)$ is defined only when $y^2 - x > 0$, that is, $x < y^2$, the domain of f is $D = \{(x, y) \mid x < y^2\}$. This is the set of points to the left of the parabola $x = y^2$. (See Figure 3.)

2.2 Visualizing Multivariable Functions

Definition 2.3. Let $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ be a real-valued function of two variables.

- (1) The *graph* of f is the set

$$\Gamma(f) = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\}$$

- (2) A *level curve* of f is the set

$$L(f) = \{(x, y) \in U \subset \mathbb{R}^2 : f(x, y) = \text{constant}\}$$

- (3) To each level curve in the domain we associate a *contour line*, which is the intersection of the plane $z = c$ with the surface $z = f(x, y)$.

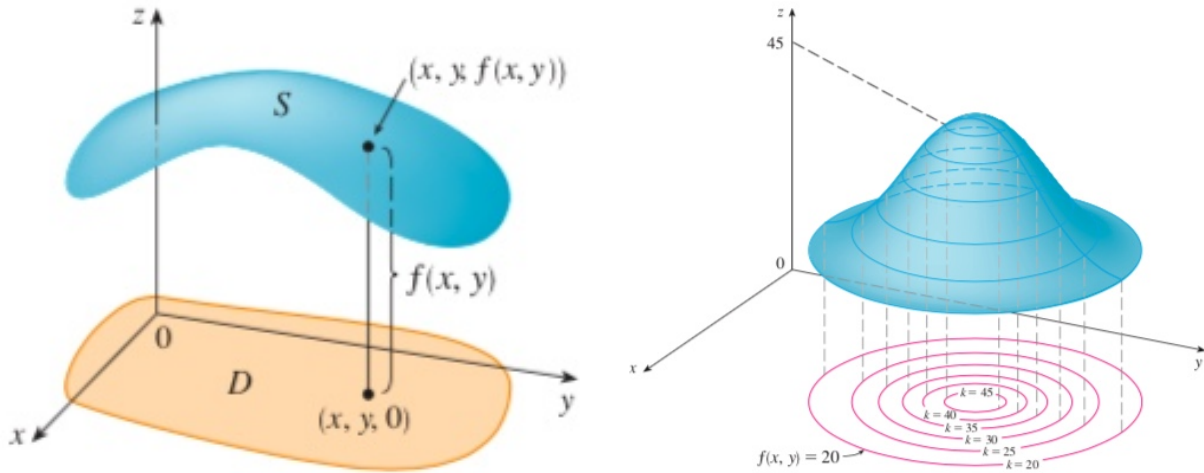


Figure 16: The graph of a function $z = f(x, y)$ is shown on the left. On the right, contour lines are shown in the domain which are the projections onto the xy -plane of the contour lines on the graph.

Example 2.6. Consider the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f(x, y) = y^2$. For any fixed value x_0 of x , a cross section of the graph is the parabola $z = y^2$ in the zy plane at x_0 . Since $f(x, y)$ is independent of x , the graph has a translational symmetry in the x direction. The graph and level curves can be plotted in *Mathematica* using the commands shown below.

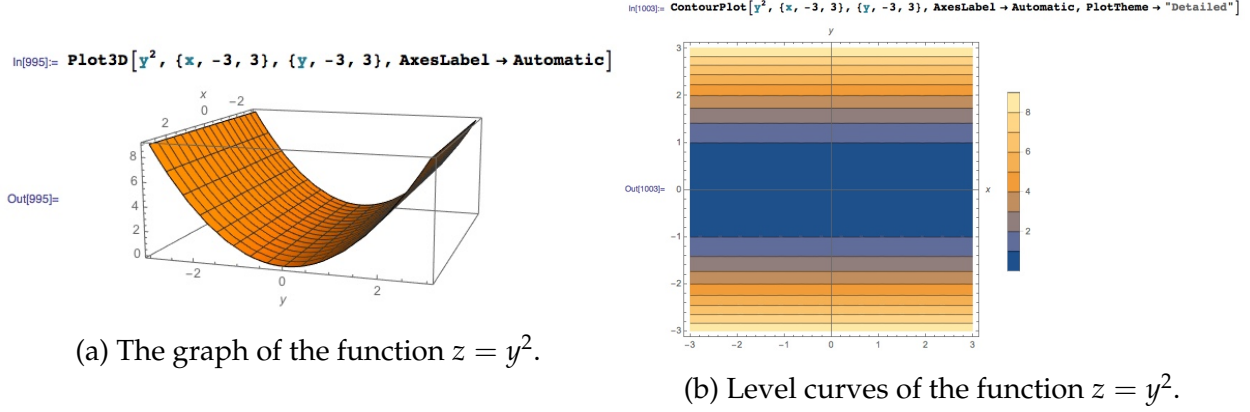


Figure 18: The graph and level curves of the function $z = y^2$.

Exercise 2.3. Sketch the graph and level curves of the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f(x, y) = x^2 + y^2$.

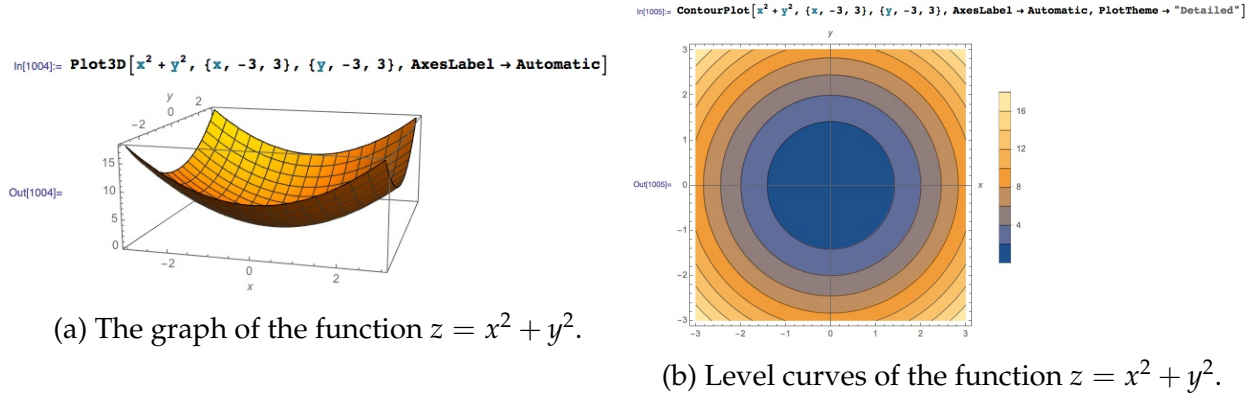


Figure 19: The graph and level curves of the function $z = x^2 + y^2$.

Solution.

If $f : U \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ is a function of three variables, its graph $\{(x, y, z, f(x, y, z))\}$ is a subset of \mathbb{R}^4 , and therefore is impossible to visualize. Instead, we visualize the function by its three-dimensional *level surfaces*.

Definition 2.7 (Level surfaces). A *level surface* of f is the set

$$L(f) = \{(x, y, z) \in U \subset \mathbb{R}^3 : f(x, y, z) = \text{constant}\}$$

Example 2.8. The level surfaces of the function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ defined by $f(x, y, z) = x^2 + y^2 + z^2$ are concentric spheres. These can be plotted in *Mathematica* using the command shown below:

```
In[1011]:= ContourPlot3D[x^2 + y^2 + z^2, {x, -4, 4}, {y, -4, 4}, {z, -4, 4}, AxesLabel -> Automatic,
PlotTheme -> "Detailed"]
```

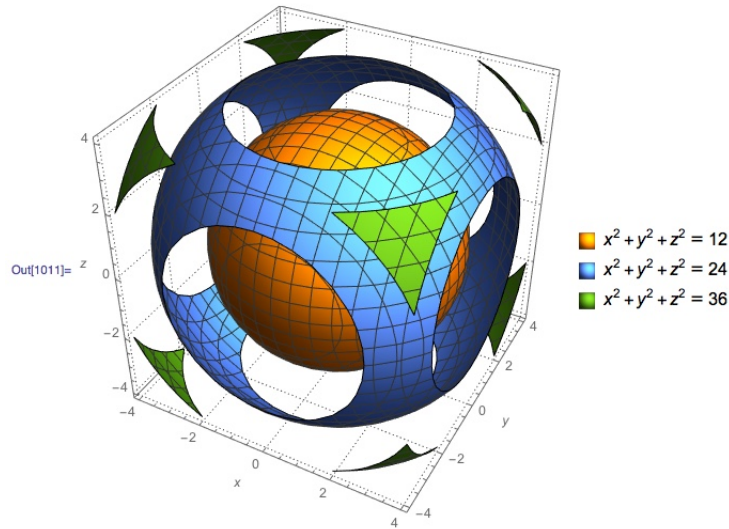
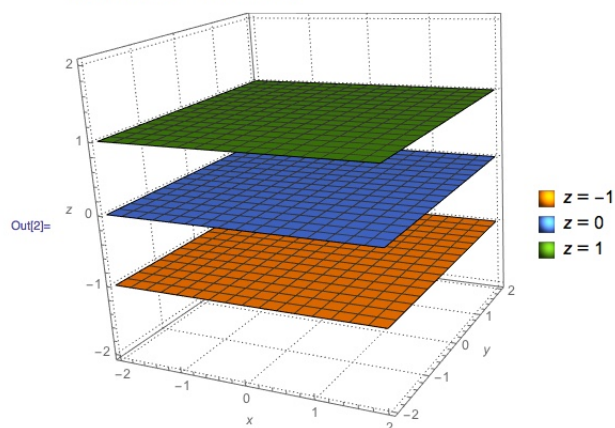


Figure 20: Level surfaces of $f(x, y, z) = x^2 + y^2 + z^2$.

Exercise 2.4. Plot the level surfaces of the function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ defined by $f(x, y, z) = z$.

```
In[2]:= ContourPlot3D[z, {x, -2, 2}, {y, -2, 2}, {z, -2, 2}, AxesLabel -> Automatic,
PlotTheme -> "Detailed"]
```



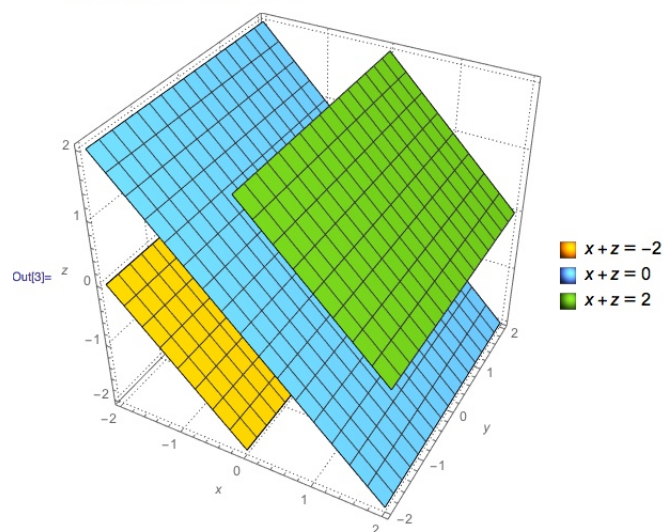
Solution.

Figure 21: Level surfaces of $f(x, y, z) = z$.

□

Exercise 2.5. Plot the level surfaces of the function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ defined by $f(x, y, z) = x + z$.

```
In[3]:= ContourPlot3D[x + z, {x, -2, 2}, {y, -2, 2}, {z, -2, 2}, AxesLabel -> Automatic,
PlotTheme -> "Detailed"]
```



Solution.

Figure 22: Level surfaces of $f(x, y, z) = x + z$.

□

2.3 Limits

Let's compare the behavior of the functions

$$f(x, y) = \frac{\sin(x^2 + y^2)}{x^2 + y^2} \quad \text{and} \quad g(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$$

as $(x, y) \rightarrow (0, 0)$. Note that neither function is defined at $(0, 0)$:

TABLE 1 Values of $f(x, y)$

$x \backslash y$	-1.0	-0.5	-0.2	0	0.2	0.5	1.0
-1.0	0.455	0.759	0.829	0.841	0.829	0.759	0.455
-0.5	0.759	0.959	0.986	0.990	0.986	0.959	0.759
-0.2	0.829	0.986	0.999	1.000	0.999	0.986	0.829
0	0.841	0.990	1.000		1.000	0.990	0.841
0.2	0.829	0.986	0.999	1.000	0.999	0.986	0.829
0.5	0.759	0.959	0.986	0.990	0.986	0.959	0.759
1.0	0.455	0.759	0.829	0.841	0.829	0.759	0.455

TABLE 2 Values of $g(x, y)$

$x \backslash y$	-1.0	-0.5	-0.2	0	0.2	0.5	1.0
-1.0	0.000	0.600	0.923	1.000	0.923	0.600	0.000
-0.5	-0.600	0.000	0.724	1.000	0.724	0.000	-0.600
-0.2	-0.923	-0.724	0.000	1.000	0.000	-0.724	-0.923
0	-1.000	-1.000	-1.000		-1.000	-1.000	-1.000
0.2	-0.923	-0.724	0.000	1.000	0.000	-0.724	-0.923
0.5	-0.600	0.000	0.724	1.000	0.724	0.000	-0.600
1.0	0.000	0.600	0.923	1.000	0.923	0.600	0.000

It looks like $\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 1$ (since the values of the function approach 1 from all directions) while $\lim_{(x,y) \rightarrow (0,0)} g(x, y)$ does not exist (since the values of the function near $(0, 0)$ do not approach any fixed value).

Definition 2.9 (Limit). Let $f : U - \{\mathbf{x}_0\} \subset \mathbb{R}^n \rightarrow \mathbb{R}$, and $L \in \mathbb{R}$. We say that $f(\mathbf{x}) \rightarrow L$ as $\mathbf{x} \rightarrow \mathbf{x}_0$ if for every $\epsilon > 0$ there exists a corresponding $\delta > 0$ such that

$$0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta \implies |f(\mathbf{x}) - L| < \epsilon.$$

In this case L is called the *limit* of f as \mathbf{x} approaches \mathbf{x}_0 , and we write $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = L$.

This definition says that the distance between $f(\mathbf{x})$ and L can be made arbitrarily small by making the distance between \mathbf{x} and \mathbf{x}_0 sufficiently small.⁵

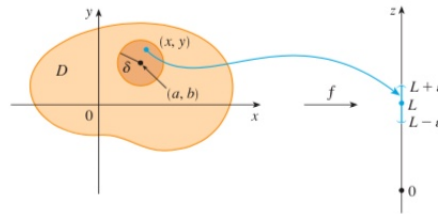


Figure 23: A function $f(\mathbf{x})$ has limit L as $\mathbf{x} \rightarrow \mathbf{x}_0$ if $f(\mathbf{x})$ can be made arbitrarily close to L by taking \mathbf{x} sufficiently close to \mathbf{x}_0 .

⁵Two points x_1, x_2 are said to be *arbitrarily close* if, for every $\epsilon > 0$, $d(x_1, x_2) < \epsilon$. A point x_1 is said to be *sufficiently close* to x_2 if there exists some (fixed) $\delta > 0$ such that $d(x_1, x_2) < \delta$.

As in single-variable calculus, another illustration of the definition of a limit can be given in terms of its graph: For any $\epsilon > 0$, we can find $\delta > 0$ such that if $\mathbf{x} \in B_\delta(\mathbf{x}_0)$ and $\mathbf{x} \neq \mathbf{x}_0$, then $f(\mathbf{x})$ lies between the horizontal planes $z = L - \epsilon$ and $z = L + \epsilon$.

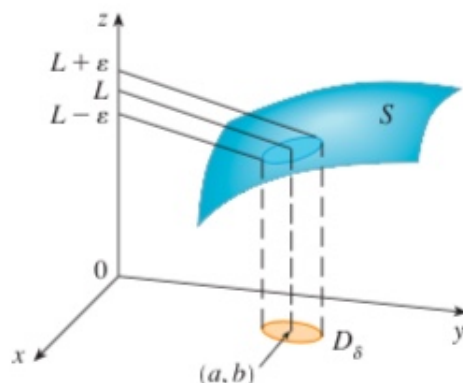


Figure 24: A function f has limit L as $\mathbf{x} \rightarrow \mathbf{x}_0$ if the graph can be bounded between two arbitrarily close planes whenever for all \mathbf{x} sufficiently close to \mathbf{x}_0 .

For functions of a single variable, when we let $x \rightarrow a$, there are only two directions of approach: from the left or from the right. Recall that if $\lim_{x \rightarrow a^-} f(x) \neq \lim_{x \rightarrow a^+} f(x)$, then $\lim_{x \rightarrow a} f(x)$ does not exist. For functions of two variables, we can let (x, y) approach (a, b) from an infinite number of directions, as long as (x, y) stays within the domain of f .

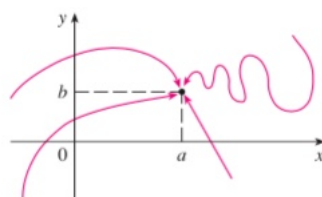


Figure 25: Various paths in the domain of a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ along which we can approach a point (a, b) .

Note that the definition of a limit only depends on the *distance* from (x, y) and (a, b) . It does *not* refer to the direction of the approach. Therefore, if the limit exists, then $f(x, y)$ must approach the same limit no matter how (x, y) approaches (a, b) . Thus, if we can find two different paths of approach along which the function $f(x, y)$ has different limits, then it follows that $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$ does not exist.

Example 2.10. Consider $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$. Consider approaching $(0, 0)$ along the x -axis, that is, along the curve $\mathbf{r}(t) = (t, 0)$. Then

$$\lim_{(x(t), y(t)) \rightarrow (0,0)} \frac{x(t)^2 - y(t)^2}{x(t)^2 + y(t)^2} = \lim_{t \rightarrow 0} \frac{t^2 - 0}{t^2 + 0} = \lim_{t \rightarrow 0} \frac{t^2}{t^2} = \lim_{t \rightarrow 0} 1 = 1.$$

Approaching $(0,0)$ along the y -axis, i.e., along the path $\mathbf{r}(t) = (0, t)$, we find

$$\lim_{(x(t), y(t)) \rightarrow (0,0)} \frac{x(t)^2 - y(t)^2}{x(t)^2 + y(t)^2} = \lim_{t \rightarrow 0} \frac{0 - t^2}{0 + t^2} = \lim_{t \rightarrow 0} \frac{-t^2}{t^2} = \lim_{t \rightarrow 0} (-1) = -1.$$

We have found two paths along which $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$ has different values, so $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$ does not exist.

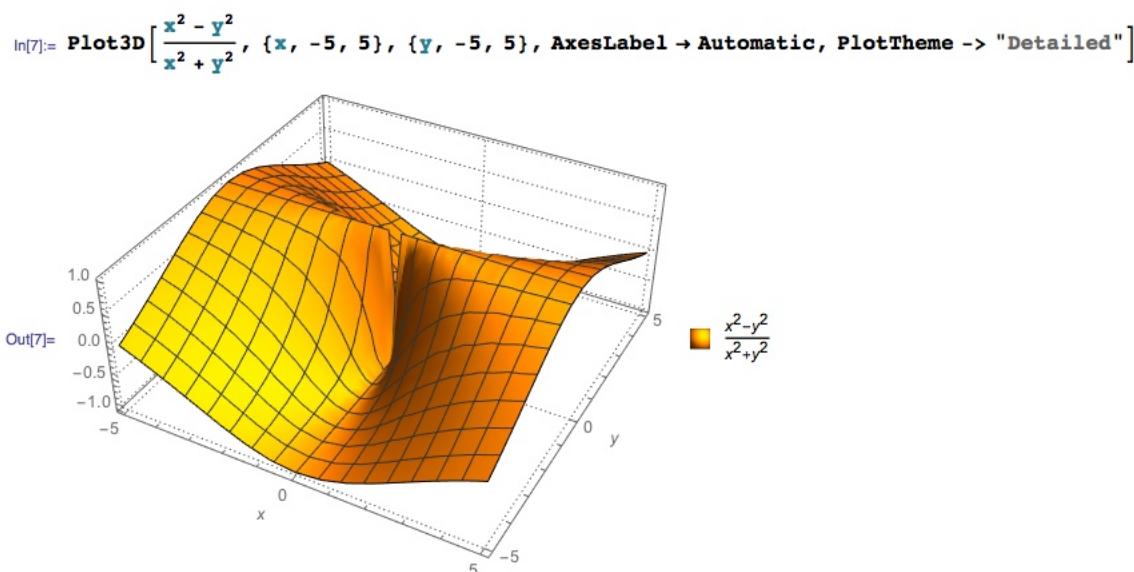


Figure 26: The function $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$ has no limit as $(x, y) \rightarrow (0, 0)$.

Example 2.11. Consider $\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2}$. Approaching $(0,0)$ along the x -axis gives

$$\lim_{x \rightarrow 0} \frac{x \cdot 0}{x^2 + 0} = \lim_{x \rightarrow 0} \frac{0}{x^2} = 0.$$

Approaching $(0,0)$ along the y -axis gives

$$\lim_{y \rightarrow 0} \frac{0 \cdot y}{0 + y^2} = \lim_{y \rightarrow 0} \frac{0}{y^2} = 0.$$

While these limits agree, this is *not* sufficient to conclude that $\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2} = 0$, since we must show that we obtain the same limit along *every* path approaching $(0,0)$. Indeed, if we approach $(0,0)$ along the line $y = x$, we obtain

$$\lim_{x \rightarrow 0} \frac{x^2}{x^2 + x^2} = \lim_{x \rightarrow 0} \frac{x^2}{2x^2} = \lim_{x \rightarrow 0} \frac{1}{2} = \frac{1}{2}.$$

Thus, $\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2}$ does not exist.

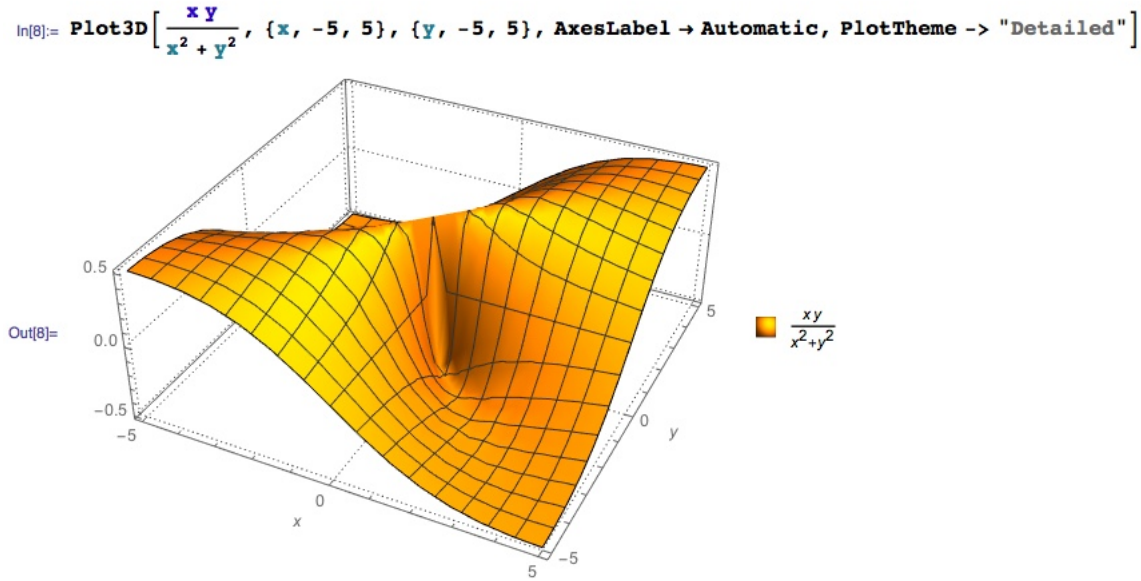


Figure 27: The function $f(x, y) = \frac{xy}{x^2+y^2}$ has no limit as $(x, y) \rightarrow (0, 0)$.

The previous example showed that it is not sufficient to compute the limit of a multivariable function by restricting the function to paths in the domain. We now show how to use Definition 2.9 to show that a limit *does* exist.

Theorem 2.12. Let a, b, c be fixed real numbers. Then

- (a) $\lim_{(x,y) \rightarrow (a,b)} c = c$.
- (b) $\lim_{(x,y) \rightarrow (a,b)} x = a$.
- (c) $\lim_{(x,y) \rightarrow (a,b)} y = b$.

Proof.

- (a) Let $\epsilon > 0$. Since $|c - c| = 0$, by taking δ to be any positive real number, then $0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta$ implies $|f(x, y) - f(a, b)| = |c - c| = 0 < \epsilon$. This proves that $\lim_{(x,y) \rightarrow (a,b)} c = c$.
- (b) Let $\epsilon > 0$. We have $|f(x, y) - f(a, b)| = |x - a|$. Since

$$|x - a| = \sqrt{(x - a)^2} \leq \sqrt{(x - a)^2 + (y - b)^2},$$

by taking $\delta = \epsilon$, $0 < \sqrt{(x - a)^2 + (y - b)^2} < \delta$ implies that $|f(x, y) - f(a, b)| = |x - a| < \epsilon$. This proves that $\lim_{(x,y) \rightarrow (a,b)} x = a$.

- (c) The proof is essentially the same as that of part (b) and is left as an exercise.

□

Exercise 2.6. Prove part (c) of Proposition ??.

Solution. Let $\epsilon > 0$. We have $|f(x, y) - f(a, b)| = |y - b|$. Since

$$|y - b| = \sqrt{(y - b)^2} \leq \sqrt{(x - a)^2 + (y - b)^2},$$

by taking $\delta = \epsilon$, $0 < \sqrt{(x - a)^2 + (y - b)^2} < \delta$ implies that $|f(x, y) - f(a, b)| = |y - b| < \epsilon$. This proves that $\lim_{(x,y) \rightarrow (a,b)} y = b$. \square

The next theorem shows that all of the limit properties of single-variable functions carry over to multivariable functions.

Theorem 2.13 (Limit laws for multivariable functions). Suppose $f(x, y)$ and $g(x, y)$ are defined on the same open set containing (x_0, y_0) , and that

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x, y) = L \quad \text{and} \quad \lim_{(x,y) \rightarrow (x_0,y_0)} g(x, y) = M.$$

Then

- (i) $\lim_{(x,y) \rightarrow (x_0,y_0)} cf(x, y) = cL$ for any $c \in \mathbb{R}$;
- (ii) $\lim_{(x,y) \rightarrow (x_0,y_0)} (f(x, y) + g(x, y)) = L + M$;
- (iii) $\lim_{(x,y) \rightarrow (x_0,y_0)} (f(x, y)g(x, y)) = LM$;
- (iv) $\lim_{(x,y) \rightarrow (x_0,y_0)} \frac{f(x, y)}{g(x, y)} = \frac{L}{M}$ whenever $M \neq 0$.

Proof. The proofs of these are identical to the proofs of the corresponding properties in Theorem 1.7, with the metric $d_{\mathbb{R}}(x, y) = |x - y|$ on \mathbb{R} replaced by the metric $d_{\mathbb{R}^2}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ on \mathbb{R}^2 . \square

Example 2.14. To evaluate $\lim_{(x,y) \rightarrow (1,2)} (x^2y^3 - x^3y^2 + 3x + 2y)$ using Theorems 2.13 and 2.12, we have

$$\begin{aligned} \lim_{(x,y) \rightarrow (1,2)} (x^2y^3 - x^3y^2 + 3x + 2y) &= \lim_{(x,y) \rightarrow (1,2)} (x^2y^3) + \lim_{(x,y) \rightarrow (1,2)} (-x^3y^2) + \lim_{(x,y) \rightarrow (1,2)} (3x + 2y) \\ &= \lim_{(x,y) \rightarrow (1,2)} x^2 \cdot \lim_{(x,y) \rightarrow (1,2)} y^3 - \lim_{(x,y) \rightarrow (1,2)} x^3 \cdot \lim_{(x,y) \rightarrow (1,2)} y^2 + 3 \lim_{(x,y) \rightarrow (1,2)} x + 2 \lim_{(x,y) \rightarrow (1,2)} y \\ &= \left(\lim_{(x,y) \rightarrow (1,2)} x \right)^2 \cdot \left(\lim_{(x,y) \rightarrow (1,2)} y \right)^3 - \left(\lim_{(x,y) \rightarrow (1,2)} x \right)^3 \cdot \left(\lim_{(x,y) \rightarrow (1,2)} y \right)^2 + 3 \lim_{(x,y) \rightarrow (1,2)} x + 2 \lim_{(x,y) \rightarrow (1,2)} y \\ &= (1)^2(2)^3 - (1)^3(2)^2 + 3(1) + 2(2) \\ &= 8 - 4 + 3 + 4 \\ &= 11. \end{aligned}$$

Definition 2.15 (Continuous function). A function f of two variables is *continuous at* (a, b) if

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b).$$

The function f is *continuous on* D if f is continuous at every point (a, b) in D .

⁶In fact, these properties hold for maps between any metric spaces. For a function $f : X \rightarrow Y$ between metric spaces, the definition of $\lim_{x \rightarrow x_0} f(x) = L$ becomes: for every $\epsilon > 0$ there exists $\delta > 0$ such that $d_Y(f(x), L) < \epsilon$ whenever $0 < d_X(x, x_0) < \delta$. Since the proofs of the limit properties only depend on properties shared by every metric (e.g., the triangle inequality), the proofs are exactly the same, with the Euclidean metrics on \mathbb{R}^2, \mathbb{R} replaced by the metrics on X and Y .

As in single-variable calculus, continuous functions map nearby points in D to nearby points in \mathbb{R} , so a surface that is the graph of a continuous function has no holes or breaks.

Using the limit properties, we find that sums, products, and quotients of continuous functions are continuous on their domains. Similarly, compositions of continuous functions are continuous.

It follows from Theorem 2.12 that $f(x, y) = x$, $g(x, y) = y$, and $h(x, y) = c$ are continuous. Since any polynomial, e.g.,

$$f(x, y) = x^4 + 5x^3y^2 + 6xy^4 - 7y + 6$$

can be built out of these by multiplication and addition, all polynomials are continuous on \mathbb{R}^2 . Likewise, any rational function is continuous on its domain because it is a quotient of continuous functions. E.g.,

$$g(x, y) = \frac{2xy + 1}{x^2 + y^2}.$$

Example 2.16. The function $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$ is continuous everywhere except at $(0, 0)$.

Exercise 2.7. Where is the function

$$g(x, y) = \begin{cases} \frac{x^4 - x^2y^2}{x + y}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

continuous?

Solution. Since a rational function is continuous everywhere it is defined, $g(x, y)$ is continuous for all $(x, y) \neq (0, 0)$. Since for $(x, y) \neq (0, 0)$

$$\frac{x^4 - x^2y^2}{x + y} = \frac{x^2(x^2 - y^2)}{x + y} = \frac{x^2(x + y)(x - y)}{x + y} = x^2(x - y),$$

we have

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^4 - x^2y^2}{x + y} = \lim_{(x,y) \rightarrow (0,0)} x^2(x - y) = 0.$$

Thus, $g(x, y)$ is continuous everywhere. □

2.4 Partial Derivatives

2.4.1 Motivation

Example 2.17 (Heat index).

On a hot day, extreme humidity makes us think the temperature is higher than it really is, whereas in very dry air we perceive the temperature to be lower than the thermometer indicates. The National Weather Service has devised the *heat index* (also called the temperature-humidity index, or humidex, in some countries) to describe the combined effects of temperature and humidity. The heat index I is the perceived air temperature when the actual temperature is T and the relative humidity is H . So I is a function of T and H and we can write $I = f(T, H)$. The following table of values of I is an excerpt from a table compiled by the National Weather Service.

		Relative humidity (%)								
Actual temperature (°F)	$T \backslash H$	50	55	60	65	70	75	80	85	90
	90	96	98	100	103	106	109	112	115	119
	92	100	103	105	108	112	115	119	123	128
	94	104	107	111	114	118	122	127	132	137
	96	109	113	116	121	125	130	135	141	146
	98	114	118	123	127	133	138	144	150	157
	100	119	124	129	135	141	147	154	161	168

If we concentrate on the highlighted column of the table, which corresponds to a relative humidity of $H = 70\%$, we are considering the heat index as a function of the single variable T for a fixed value of H . Let's write $g(T) = f(T, 70)$. Then $g(T)$ describes how the heat index I increases as the actual temperature T increases when the relative humidity is 70%. The derivative of g when $T = 96^\circ\text{F}$ is the rate of change of I with respect to T when $T = 96^\circ\text{F}$:

$$g'(96) = \lim_{h \rightarrow 0} \frac{g(96 + h) - g(96)}{h} = \lim_{h \rightarrow 0} \frac{f(96 + h, 70) - f(96, 70)}{h}$$

We can approximate $g'(96)$ using the values in Table 1 by taking $h = 2$ and -2 :

$$g'(96) \approx \frac{g(98) - g(96)}{2} = \frac{f(98, 70) - f(96, 70)}{2} = \frac{133 - 125}{2} = 4$$

$$g'(96) \approx \frac{g(94) - g(96)}{-2} = \frac{f(94, 70) - f(96, 70)}{-2} = \frac{118 - 125}{-2} = 3.5$$

Averaging these values, we can say that the derivative $g'(96)$ is approximately 3.75. This means that, when the actual temperature is 96°F and the relative humidity is 70%, the apparent temperature (heat index) rises by about 3.75°F for every degree that the actual temperature rises!

Now let's look at the highlighted row in Table 1, which corresponds to a fixed temperature of $T = 96^\circ\text{F}$. The numbers in this row are values of the function $G(H) = f(96, H)$, which describes how the heat index increases as the relative humidity H increases when the actual temperature is $T = 96^\circ\text{F}$. The derivative of this function when $H = 70\%$ is the rate of change of I with respect to H when $H = 70\%$:

$$G'(70) = \lim_{h \rightarrow 0} \frac{G(70 + h) - G(70)}{h} = \lim_{h \rightarrow 0} \frac{f(96, 70 + h) - f(96, 70)}{h}$$

By taking $h = 5$ and -5 , we approximate $G'(70)$ using the tabular values:

$$G'(70) \approx \frac{G(75) - G(70)}{5} = \frac{f(96, 75) - f(96, 70)}{5} = \frac{130 - 125}{5} = 1$$

$$G'(70) \approx \frac{G(65) - G(70)}{-5} = \frac{f(96, 65) - f(96, 70)}{-5} = \frac{121 - 125}{-5} = 0.8$$

By averaging these values we get the estimate $G'(70) \approx 0.9$. This says that, when the temperature is 96°F and the relative humidity is 70%, the heat index rises about 0.9°F for every percent the relative humidity rises.

In general, if $f(x, y)$ is a function of two variables x and y , and we consider the values of the function obtained by varying x (say) while keeping y fixed to a constant value $y = y_0$, then we are really considering a function of a *single* variable $g(x) = f(x, y_0)$. Then if $g(x)$ is differentiable at $x = x_0$, we call $g'(x)$ a *partial derivative* of $f(x, y)$. We now define this formally.

2.4.2 Definition of Partial Derivatives

Definition 2.18 (Partial derivative). Let $f : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined on an open subset of \mathbb{R}^2 containing the point (x_0, y_0) .⁷ Then the real number

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

is called the *partial derivative of f with respect to x at the point (x_0, y_0)* , provided this limit exists. Similarly, the real number

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

is called the *partial derivative of f with respect to y at the point (x_0, y_0)* , provided this limit exists. If $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are defined at all points of U , then they are functions on U .

Remark. Note that $\frac{\partial f}{\partial x}$ is not a quotient, but a *limit* of a quotient. Thus, the symbols “ ∂f ” or “ ∂x ” do not have independent meaning. Similar remarks apply to $\frac{\partial f}{\partial y}$. While this is the most common notation for the partial derivatives (and the one we will use almost exclusively), other notations are also frequently used. \square

Notations for Partial Derivatives If $z = f(x, y)$, we write

$$f_x(x, y) = f_x = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) = \frac{\partial z}{\partial x} = f_1 = D_1 f = D_x f$$

$$f_y(x, y) = f_y = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x, y) = \frac{\partial z}{\partial y} = f_2 = D_2 f = D_y f$$

Figure 28: Notations for partial derivatives.

To compute the partial derivative of a function $f(x, y)$ with respect to x , one simply treats y as a constant and differentiates $f(x, y)$ with respect to x . To compute the partial derivative with respect to y , treat x as a constant and differentiate $f(x, y)$ with respect to y .

Example 2.19. If $f(x, y) = x^3 + x^2y^3 - 2y^2$, then

$$\begin{aligned} \frac{\partial f}{\partial x} &= 3x^2 + 2xy^3 \\ \frac{\partial f}{\partial y} &= 3x^2y^2 - 4y. \end{aligned}$$

⁷Since U is open, there exists a neighborhood of (x_0, y_0) contained in U . Thus, if h is sufficiently close to zero, then all points of the form $(x_0 + h, y_0)$ will be in U , so this definition makes sense.

2.4.3 Interpretation of Partial Derivatives

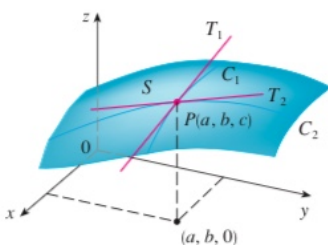


FIGURE 1

The partial derivatives of f at (a, b) are the slopes of the tangents to C_1 and C_2 .

To give a geometric interpretation of partial derivatives, we recall that the equation $z = f(x, y)$ represents a surface S (the graph of f). If $f(a, b) = c$, then the point $P(a, b, c)$ lies on S . By fixing $y = b$, we are restricting our attention to the curve C_1 in which the vertical plane $y = b$ intersects S . (In other words, C_1 is the trace of S in the plane $y = b$.) Likewise, the vertical plane $x = a$ intersects S in a curve C_2 . Both of the curves C_1 and C_2 pass through the point P . (See Figure 1.)

Notice that the curve C_1 is the graph of the function $g(x) = f(x, b)$, so the slope of its tangent T_1 at P is $g'(a) = f_x(a, b)$. The curve C_2 is the graph of the function $G(y) = f(a, y)$, so the slope of its tangent T_2 at P is $G'(b) = f_y(a, b)$.

Thus the partial derivatives $f_x(a, b)$ and $f_y(a, b)$ can be interpreted geometrically as the slopes of the tangent lines at $P(a, b, c)$ to the traces C_1 and C_2 of S in the planes $y = b$ and $x = a$.

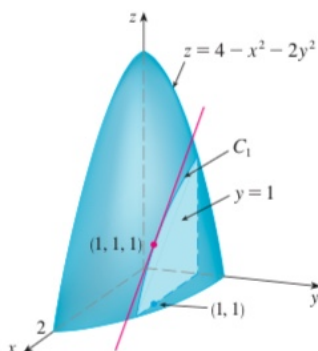
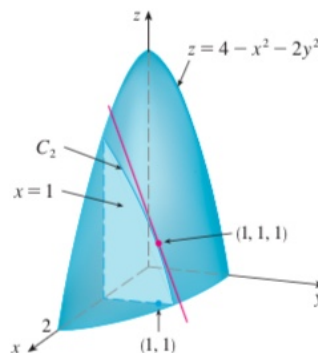


FIGURE 2



As we have seen in the case of the heat index function, partial derivatives can also be interpreted as *rates of change*. If $z = f(x, y)$, then $\partial z / \partial x$ represents the rate of change of z with respect to x when y is fixed. Similarly, $\partial z / \partial y$ represents the rate of change of z with respect to y when x is fixed.

EXAMPLE 2 If $f(x, y) = 4 - x^2 - 2y^2$, find $f_x(1, 1)$ and $f_y(1, 1)$ and interpret these numbers as slopes.

SOLUTION We have

$$f_x(x, y) = -2x \qquad f_y(x, y) = -4y$$

$$f_x(1, 1) = -2 \qquad f_y(1, 1) = -4$$

The graph of f is the paraboloid $z = 4 - x^2 - 2y^2$ and the vertical plane $y = 1$ intersects it in the parabola $z = 2 - x^2$, $y = 1$. (As in the preceding discussion, we label it C_1 in Figure 2.) The slope of the tangent line to this parabola at the point $(1, 1, 1)$ is $f_x(1, 1) = -2$. Similarly, the curve C_2 in which the plane $x = 1$ intersects the paraboloid is the parabola $z = 3 - 2y^2$, $x = 1$, and the slope of the tangent line at $(1, 1, 1)$ is $f_y(1, 1) = -4$. (See Figure 3.)

Figure 4 is a computer-drawn counterpart to Figure 2. Part (a) shows the plane $y = 1$ intersecting the surface to form the curve C_1 and part (b) shows C_1 and T_1 . [We have used the vector equations $\mathbf{r}(t) = \langle t, 1, 2 - t^2 \rangle$ for C_1 and $\mathbf{r}(t) = \langle 1 + t, 1, 1 - 2t \rangle$ for T_1 .] Similarly, Figure 5 corresponds to Figure 3.

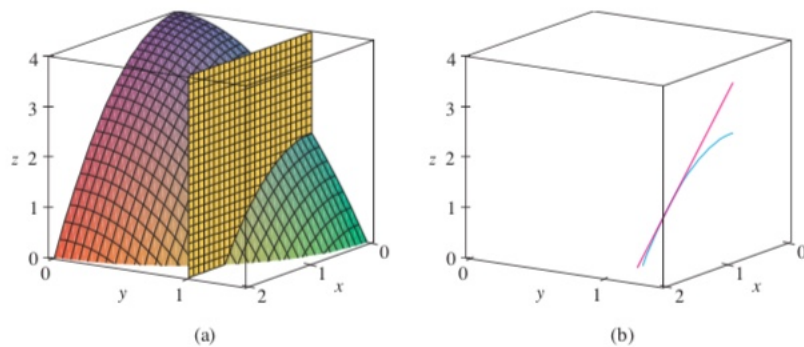


FIGURE 4

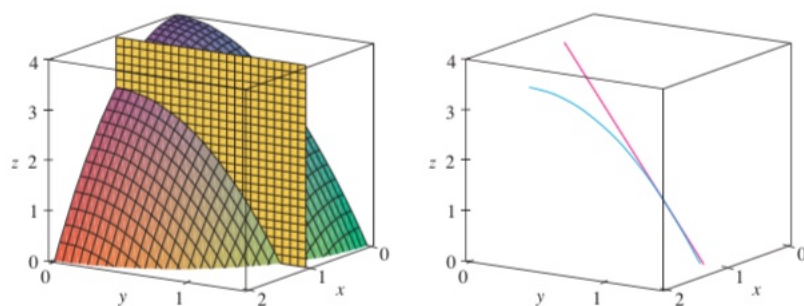


FIGURE 5

Exercise 2.8. If $f(x, y) = \sin\left(\frac{x}{1+y}\right)$, calculate $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

SOLUTION Using the Chain Rule for functions of one variable, we have

$$\frac{\partial f}{\partial x} = \cos\left(\frac{x}{1+y}\right) \cdot \frac{\partial}{\partial x}\left(\frac{x}{1+y}\right) = \cos\left(\frac{x}{1+y}\right) \cdot \frac{1}{1+y}$$

$$\frac{\partial f}{\partial y} = \cos\left(\frac{x}{1+y}\right) \cdot \frac{\partial}{\partial y}\left(\frac{x}{1+y}\right) = -\cos\left(\frac{x}{1+y}\right) \cdot \frac{x}{(1+y)^2}$$

[More Examples?]

[Implicit differentiation here?]

2.4.4 Functions of More than Two Variables

Partial derivatives can also be defined for functions of three or more variables. For example, if f is a function of three variables x , y , and z , then its partial derivative with respect to x is defined as

$$f_x(x, y, z) = \lim_{h \rightarrow 0} \frac{f(x + h, y, z) - f(x, y, z)}{h}$$

and it is found by regarding y and z as constants and differentiating $f(x, y, z)$ with respect to x . If $w = f(x, y, z)$, then $f_x = \partial w / \partial x$ can be interpreted as the rate of change of w with respect to x when y and z are held fixed. But we can't interpret it geometrically because the graph of f lies in four-dimensional space.

In general, if u is a function of n variables, $u = f(x_1, x_2, \dots, x_n)$, its partial derivative with respect to the i th variable x_i is

$$\frac{\partial u}{\partial x_i} = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}$$

That is, we hold all the variables other than x_i constant and differentiate f with respect to x_i .

Exercise 2.9. Compute $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ if $f(x, y, z) = e^{xy} \ln z$.

SOLUTION Holding y and z constant and differentiating with respect to x , we have

$$f_x = ye^{xy} \ln z$$

Similarly, $f_y = xe^{xy} \ln z$ and $f_z = \frac{e^{xy}}{z}$

2.5 The Differential

2.5.1 Differential of a Single-variable Function

Let $f(x)$ be defined on an open subset U of \mathbb{R} . If f is differentiable on U , then

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \quad (2.1)$$

exists for each $x \in U$. Rearrange (2.1) as follows:

$$\begin{aligned} \lim_{h \rightarrow 0} \left[\frac{f(x+h) - f(x)}{h} \right] - f'(x) &= 0 \\ \lim_{h \rightarrow 0} \left[\frac{f(x+h) - f(x)}{h} - f'(x) \right] &= 0 \\ \lim_{h \rightarrow 0} \left[\frac{f(x+h) - f(x)}{h} - \frac{f'(x)h}{h} \right] &= 0 \\ \lim_{h \rightarrow 0} \left[\frac{f(x+h) - f(x) - f'(x)h}{h} \right] &= 0 \end{aligned}$$

Let $r(h) \equiv f(x+h) - f(x) - f'(x)h$. Then existence of the limit in equation (2.1) implies that

$$f(x+h) = f(x) + f'(x)h + r(h) \quad (2.2)$$

where $\lim_{h \rightarrow 0} \frac{r(h)}{h} = 0$. This says that, up to a small error term, $f(x)$ is an affine map (see Appendix C) with differential

$$\begin{aligned} df_x : \mathbb{R} &\rightarrow \mathbb{R} \\ df_x(h) &= f'(x)h. \end{aligned}$$

That is, df_x is the linear function which multiplies $h \in \mathbb{R}$ by the derivative of f at x . [Refer to Linear Algebra notes. We have seen that every linear map from $\mathbb{R} \rightarrow \mathbb{R}$ is multiplication by some fixed real number.]

We now give an equivalent definition of differentiability which will readily generalize to multivariable functions. We will say that $f(x)$ is *old style differentiable* at x if

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

exists.

Definition 2.20 (Differentiable function). A function $f : U \subseteq \mathbb{R} \rightarrow \mathbb{R}$ defined on an open subset U of \mathbb{R} is *differentiable* at $x \in U$ if there exists a linear map $T : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$f(x+h) = f(x) + Th + r(h)$$

where $\lim_{h \rightarrow 0} \frac{r(h)}{h} = 0$.

Proposition 2.21. If $T, T' : \mathbb{R} \rightarrow \mathbb{R}$ are two linear maps which satisfy Definition 2.20, then $T' = T$.

Proof. Suppose

$$\begin{aligned} f(x+h) &= f(x) + Th + r(h), \text{ and} \\ f(x+h) &= f(x) + T'h + r'(h) \end{aligned}$$

where $\lim_{h \rightarrow 0} \frac{r(h)}{h} = 0$ and $\lim_{h \rightarrow 0} \frac{r'(h)}{h} = 0$. Then

$$f(x) + Th + r(h) = f(x) + T'h + r'(h)$$

implies

$$T - T' = \frac{r'(h)}{h} - \frac{r(h)}{h}$$

and therefore

$$T - T' = \lim_{h \rightarrow 0} (T - T') = \lim_{h \rightarrow 0} \left(\frac{r'(h)}{h} - \frac{r(h)}{h} \right) = \lim_{h \rightarrow 0} \frac{r'(h)}{h} - \lim_{h \rightarrow 0} \frac{r(h)}{h} = 0 + 0 = 0,$$

hence $T = T'$. □

Definition 2.22 (Differential). If f is differentiable at x , we call the unique linear map T in Definition 2.20 the *differential of f at x* and write $T = df_x$.

We have seen above that if f is old style differentiable, then it is differentiable in the sense of Definition 2.20. Conversely, if f is differentiable at x then there exists a unique linear map T such that

$$f(x+h) = f(x) + Th + r(h),$$

which implies that

$$T = \frac{f(x+h) - f(x) - r(h)}{h}$$

and therefore

$$T = \lim_{h \rightarrow 0} T = \lim_{h \rightarrow 0} \left(\frac{f(x+h) - f(x)}{h} - \frac{r(h)}{h} \right) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} - \lim_{h \rightarrow 0} \frac{r(h)}{h} = f'(x) + 0 = f'(x),$$

so f is old style differentiable at x .

2.5.2 Differential of a Multivariable Function

Following the previous discussion, we will define a multivariable function to be differentiable if it is well-approximated by an affine map, up to a small error term.

Definition 2.23 (Differentiable multivariable function). Let $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be defined on an open subset of \mathbb{R}^n . We say that f is *differentiable* at $x \in U$ if there exists a linear map $T : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + T\mathbf{h} + r(\mathbf{h})$$

where $\lim_{\mathbf{h} \rightarrow 0} \frac{|r(\mathbf{h})|}{\|\mathbf{h}\|} = 0$. Equivalently, f is differentiable at $x \in U$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $\|\mathbf{h}\| < \delta$ implies

$$|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - T\mathbf{h}| < \epsilon \|\mathbf{h}\|.$$

Definition 2.24 (Differential of a multivariable function). If f is differentiable at \mathbf{x} , we call the unique linear map T in Definition 2.23 the *differential of f at \mathbf{x}* and write $T = df_{\mathbf{x}}$.

Proposition 2.25 (Uniqueness of the differential). If $T, T' : \mathbb{R}^n \rightarrow \mathbb{R}$ are two linear maps which satisfy Definition 2.23, then $T' = T$.

Proof. Suppose $T, T' : \mathbb{R}^n \rightarrow \mathbb{R}$ are linear maps such that

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) &= f(\mathbf{x}) + T\mathbf{h} + r(\mathbf{h}), \\ f(\mathbf{x} + \mathbf{h}) &= f(\mathbf{x}) + T'\mathbf{h} + r'(\mathbf{h}) \end{aligned}$$

with

$$\lim_{\mathbf{h} \rightarrow 0} \frac{|r(\mathbf{h})|}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \rightarrow 0} \frac{|r'(\mathbf{h})|}{\|\mathbf{h}\|} = 0.$$

Then, since T, T' are linear,

$$\begin{aligned} |(T - T')\mathbf{h}| &= |T\mathbf{h} - T'\mathbf{h}| \\ &= |T\mathbf{h} - T'\mathbf{h} + f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - (f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}))| \\ &= |T - (f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})) + f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - T'| \\ &\leq |T - (f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}))| + |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - T'| \\ &= |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - T| + |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - T'| \\ &= |r(\mathbf{h})| + |r'(\mathbf{h})| \end{aligned}$$

and therefore

$$\frac{|(T - T')\mathbf{h}|}{\|\mathbf{h}\|} \leq \frac{|r(\mathbf{h})|}{\|\mathbf{h}\|} + \frac{|r'(\mathbf{h})|}{\|\mathbf{h}\|}.$$

Since

$$\lim_{\mathbf{h} \rightarrow 0} \frac{|r(\mathbf{h})|}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \rightarrow 0} \frac{|r'(\mathbf{h})|}{\|\mathbf{h}\|} = 0,$$

given $\epsilon > 0$ there exist positive real numbers δ_1, δ_2 such that

$$\frac{|r(\mathbf{h})|}{\|\mathbf{h}\|} < \frac{\epsilon}{2}$$

whenever $\|\mathbf{h}\| < \delta_1$ and

$$\frac{|r'(\mathbf{h})|}{\|\mathbf{h}\|} < \frac{\epsilon}{2}$$

whenever $\|\mathbf{h}\| < \delta_2$. By choosing $\delta = \min\{\delta_1, \delta_2\}$, we have

$$\frac{|(T - T')\mathbf{h}|}{\|\mathbf{h}\|} \leq \frac{|r(\mathbf{h})|}{\|\mathbf{h}\|} + \frac{|r'(\mathbf{h})|}{\|\mathbf{h}\|} < \frac{\epsilon}{2} + \frac{\epsilon}{2}$$

whenever $\|\mathbf{h}\| < \delta$, and therefore

$$\lim_{\mathbf{h} \rightarrow 0} \frac{|(T - T')\mathbf{h}|}{\|\mathbf{h}\|} = 0.$$

Since this limit holds as $\mathbf{h} \rightarrow \mathbf{0}$ along any path, for any fixed $\mathbf{h} \neq \mathbf{0}$ we must have

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{|(T - T')(\mathbf{th})|}{\|\mathbf{th}\|} = 0.$$

Since $T - T'$ is linear, this becomes

$$\begin{aligned} 0 &= \lim_{t \rightarrow 0} \frac{|(T - T')(\mathbf{th})|}{\|\mathbf{th}\|} \\ &= \lim_{t \rightarrow 0} \frac{|t(T - T')\mathbf{h}|}{\|t\mathbf{h}\|} \\ &= \lim_{t \rightarrow 0} \frac{|t| |(T - T')\mathbf{h}|}{|t| \|\mathbf{h}\|} \\ &= \lim_{t \rightarrow 0} \frac{|(T - T')\mathbf{h}|}{\|\mathbf{h}\|} \\ &= \frac{|(T - T')\mathbf{h}|}{\|\mathbf{h}\|}, \end{aligned}$$

which implies that $(T - T')\mathbf{h} = \mathbf{0}$ and therefore $\mathbf{h} \in \ker(T - T')$. Since \mathbf{h} was an arbitrary nonzero vector, this means that $\ker(T - T') = \mathbb{R}^2$, so $T - T'$ is the zero map, and therefore $T = T'$. \square

Remark. (a) If $f : A \rightarrow B$ is an affine map (Definition C.3), then f is differentiable for all $p \in A$ and df_p is independent of p . In other words, an affine map has a constant differential.

(b) If f is a linear map, then $d_p f = f$.

(c) If $f : U \rightarrow B$ is differentiable at $p \in U$, then the differential is a map

$$df : U \rightarrow \text{Hom}(V, W).$$

\square

As in the single variable case, a differentiable function $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous.

Theorem 2.26 (A differentiable function is continuous). If $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable at $\mathbf{x} \in U$, then it is continuous at $\mathbf{x} \in U$.

Proof. Let $df_{\mathbf{x}}$ be the differential of f at \mathbf{x} and let $\epsilon > 0$.

$$\begin{aligned} |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})| &= |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - df_{\mathbf{x}}(\mathbf{h}) + df_{\mathbf{x}}(\mathbf{h})| \\ &\leq |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - df_{\mathbf{x}}(\mathbf{h})| + |df_{\mathbf{x}}(\mathbf{h})|. \end{aligned}$$

Since f is differentiable, there exists $\delta_0 > 0$ such that $|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - df_{\mathbf{x}}(\mathbf{h})| < \frac{\epsilon}{2} \|\mathbf{h}\|$ whenever $\|\mathbf{h}\| < \delta_0$. By Theorem B.26, there exists $C > 0$ such that $|df_{\mathbf{x}}(\mathbf{h})| < \epsilon C \|\mathbf{h}\|$. Thus

$$\begin{aligned} |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})| &\leq |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - df_{\mathbf{x}}(\mathbf{h})| + |df_{\mathbf{x}}(\mathbf{h})| \\ &< \epsilon(1 + C) \|\mathbf{h}\|. \end{aligned}$$

Let $\delta = \min\{\delta_0, \frac{1}{1+C}\}$. Then for $\|\mathbf{h}\| < \delta$ implies

$$\begin{aligned} |f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})| &\leq \epsilon(1 + C) \|\mathbf{h}\| \\ &< \epsilon(1 + C) \frac{1}{(1 + C)} \\ &= \epsilon, \end{aligned}$$

hence f is continuous at \mathbf{x} . \square

Definition 2.27 (Continuously differentiable function). If f is differentiable on U and $df : U \rightarrow \text{Hom}(V, W)$ is continuous, then we say that f is *continuously differentiable*.

Note that the existence of both partial derivatives of a function f at (x, y) is *not* sufficient for f to be continuous at (x, y) .

Example 2.28. Let

$$f(x, y) = \begin{cases} 0 & \text{if } xy \neq 0, \\ 1 & \text{if } xy = 0 \end{cases}$$

Since $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ is 1 along the x -axis and 0 along the line $y = x$, the limit does not exist and therefore $f(x, y)$ is discontinuous at $(0, 0)$. However,

$$\begin{aligned} D_x f(0, 0) &= \lim_{h \rightarrow 0} \frac{f(h, 0) - f(0, 0)}{h} = \lim_{h \rightarrow 0} \frac{1 - 1}{h} = 0 \\ D_y f(0, 0) &= \lim_{h \rightarrow 0} \frac{f(0, h) - f(0, 0)}{h} = \lim_{h \rightarrow 0} \frac{1 - 1}{h} = 0 \end{aligned}$$

so the partial derivatives both exist at $(0, 0)$.

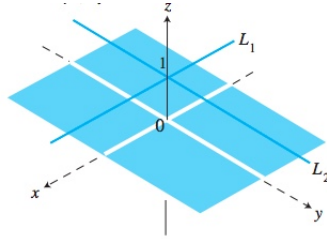


Figure 29: Graph of the function $f(x, y)$ in Example 2.28

Thus, the existence of both partial derivatives at (x, y) is not strong enough to guarantee continuity at (x, y) . This also makes intuitive sense: continuity at (a, b) requires that $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b)$ when the limit is taken along *any* path passing through (a, b) . The partial derivatives, on the other hand, only depend on what is happening along lines through (a, b) parallel to the x and y axes, and therefore they failed to detect the discontinuity in the previous example.

2.5.3 Matrix Representation of the Differential

In this section we clarify the relation between the differential of a function f and the partial derivatives of f .

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Then $df_{\mathbf{x}}$ is a linear map for each $\mathbf{x} \in \mathbb{R}^n$, so it can be represented by a matrix [Refer to my [Linear Algebra notes](#).]

Theorem 2.29. Let U be an open subset of \mathbb{R}^n and $f : U \rightarrow \mathbb{R}$. If f is differentiable at $\mathbf{x} \in U$, then the partial derivatives $\frac{\partial f}{\partial x_j}(\mathbf{x})$ all exist, and

$$df_{\mathbf{x}}(\mathbf{e}_j) = \sum_{i=1}^m \frac{\partial f}{\partial x_j}(\mathbf{x}), \quad (1 \leq j \leq n),$$

where $\{\mathbf{e}_j\}_{j=1}^n$ denotes the standard basis for \mathbb{R}^n .

Proof. The j th column of the standard matrix of $df_{\mathbf{x}}$ is $df_{\mathbf{x}}(\mathbf{e}_j)$. Since, f is differentiable at \mathbf{x} ,

$$f(\mathbf{x} + t\mathbf{e}_j) - f(\mathbf{x}) = df_{\mathbf{x}}(t\mathbf{e}_j) + r(t\mathbf{e}_j)$$

where $r(t\mathbf{e}_j)/t \rightarrow 0$ as $t \rightarrow 0$. Since $df_{\mathbf{x}}$ is linear, $df_{\mathbf{x}}(t\mathbf{e}_j) = tdf_{\mathbf{x}}(\mathbf{e}_j)$. Dividing both sides of the equation above by t and taking $t \rightarrow 0$, we have

$$\begin{aligned} \frac{\partial f}{\partial x_j}(\mathbf{x}) &\equiv \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{e}_j) - f(\mathbf{x})}{t} = \lim_{t \rightarrow 0} [df_{\mathbf{x}}(\mathbf{e}_j) + \frac{r(t\mathbf{e}_j)}{t}] \\ &= \lim_{t \rightarrow 0} df_{\mathbf{x}}(\mathbf{e}_j) + \lim_{t \rightarrow 0} \frac{r(t\mathbf{e}_j)}{t} \\ &= df_{\mathbf{x}}(\mathbf{e}_j) + 0 \\ &= df_{\mathbf{x}}(\mathbf{e}_j). \end{aligned}$$

□

Definition 2.30. The $1 \times n$ matrix representing $df_{\mathbf{x}}$ is called the *gradient of f at \mathbf{x}* , and is denoted by

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right).$$

Example 2.31. [Choose a function which is not linear.] Let $f(x, y) = 3x + 2y$. Then $\nabla f(x, y) = (3, 2)$ and therefore

$$\begin{aligned} df_{(x,y)}(h_1, h_2) &= \nabla f(x, y) \cdot (h_1, h_2) \\ &= 3h_1 + 2h_2 \\ &= f(h_1, h_2). \end{aligned}$$

As we saw in Example 2.28, the existence of all partial derivatives is not enough to guarantee differentiability (their existence was not even enough to guarantee continuity), that is, the existence of the differential df . However, as we will now show, if the partial derivatives all exist and are *continuous*, then the differential *does* exist. Thus, **any function whose partial derivatives exist and are continuous is differentiable in the sense of Definition 2.23.**

Theorem 2.32. Let U be an open subset of \mathbb{R}^n and let $f : U \rightarrow \mathbb{R}$. Then, if each of the partial derivatives $\frac{\partial f}{\partial x_i}$ exists and is continuous on U , then f is differentiable on U .

Proof. To cut down on writing, we will prove this for the case of $n = 2$. The proof for $n > 2$ is exactly the same. We need to show that, for any $(x, y) \in U$ and $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x + h_1, y + h_2) - f(x, y) - df_{(x,y)}\mathbf{h}| < \epsilon \|\mathbf{h}\|$$

whenever $\|\mathbf{h}\| \equiv \sqrt{h_1^2 + h_2^2} < \delta$.

$$\begin{aligned} &|f(x + h_1, y + h_2) - f(x, y) - df_{(x,y)}\mathbf{h}| \\ &= |f(x + h_1, y + h_2) - f(x, y) - \frac{\partial f}{\partial x}(x, y)h_1 - \frac{\partial f}{\partial y}(x, y)h_2| \\ &= |f(x + h_1, y + h_2) - f(x, y + h_2) + f(x, y + h_2) - f(x, y) - \frac{\partial f}{\partial x}(x, y)h_1 - \frac{\partial f}{\partial y}(x, y)h_2| \end{aligned}$$

By the Mean Value Theorem, there exists $u_1 \in (x, x + h_1)$ and $u_2 \in (y, y + h_2)$ such that

$$\begin{aligned} f(x + h_1, y + h_2) - f(x, y + h_2) &= \frac{\partial f}{\partial x}(u_1, y + h_2)h_1 \\ f(x, y + h_2) - f(x, y) &= \frac{\partial f}{\partial y}(x, u_2)h_2 \end{aligned}$$

so the above is equal to

$$\begin{aligned} &|(\frac{\partial f}{\partial x}(u_1, y + h_2) - \frac{\partial f}{\partial x}(x, y))h_1 + (\frac{\partial f}{\partial y}(x, u_2) - \frac{\partial f}{\partial y}(x, y))h_2| \\ &\leq |(\frac{\partial f}{\partial x}(u_1, y + h_2) - \frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, u_2) - \frac{\partial f}{\partial y}(x, y))| \|\mathbf{h}\| \\ &\leq (|\frac{\partial f}{\partial x}(u_1, y + h_2) - \frac{\partial f}{\partial x}(x, y)| + |\frac{\partial f}{\partial y}(x, u_2) - \frac{\partial f}{\partial y}(x, y)|) \|\mathbf{h}\| \end{aligned}$$

where in the second line we have used the Cauchy-Schwarz inequality⁸ and in the third line the fact that $\sqrt{x^2 + y^2} \leq |x| + |y|$. Since the partial derivatives are continuous, and since u_1 and u_2 lie between x and $x + h_1$ and y and $y + h_2$, respectively, there exists a δ_1 such that the first absolute value in the parenthesis is less than $\epsilon/2$ whenever $\|\mathbf{h}\| < \delta_1$ and a δ_2 such that the second absolute value in the parenthesis is less than $\epsilon/2$ whenever $\|\mathbf{h}\| < \delta_2$. Thus, by taking $\|\mathbf{h}\| < \delta = \min\{\delta_1, \delta_2\}$, the entire parenthesis is less than ϵ . \square

Example 2.33. Let $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ be defined by $f(x, y, z) = 3x^2y^3z$. Since the partial derivatives

$$\frac{\partial f}{\partial x}(x, y, z) = 6xy^3z, \quad \frac{\partial f}{\partial y}(x, y, z) = 9x^2y^2z, \quad \frac{\partial f}{\partial z}(x, y, z) = 3x^2y^3$$

are continuous, the function f is differentiable with gradient

$$\nabla f(x, y, z) = [6xy^3z \quad 9x^2y^2z \quad 3x^2y^3].$$

[Write equation of tangent plane at (x, y) for a differentiable function $f(x, y)$.]
[Directional derivatives here?]

2.6 The Composite Mapping Theorem (Chain Rule)

Suppose a particle moves along a path C along which the temperature varies. The path is the image of

$$\mathbf{r} : I \rightarrow \mathbb{R}^2$$

while the temperature at a point $(x, y) \in C$ is $z = T(x, y)$. Thus, $T : C \subset \mathbb{R}^2 \rightarrow \mathbb{R}$. The composite function

$$F \equiv T \circ \mathbf{r} : I \rightarrow \mathbb{R}$$

defined by $F(t) = T(\mathbf{r}(t))$ gives the temperature felt by the particle at time t . Then $F'(t)$ is the rate at which the temperature at the position of the particle is changing with time at $t \in I$. How do we compute $F'(t)$?

⁸The Cauchy-Schwarz inequality states that $|df_{(x,h)} \cdot \mathbf{h}| \leq |df_{(x,h)}| \|\mathbf{h}\|$.

Theorem 2.34 (Chain rule). Let $\mathbf{r} : I \rightarrow \mathbb{R}^2$ and $f : U \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$, where I is an interval in \mathbb{R} and U is an open subset of \mathbb{R}^2 containing the trace of the curve \mathbf{r} . If \mathbf{r} is differentiable at t and f is differentiable at $\mathbf{r}(t)$, then $f \circ \mathbf{r} : I \rightarrow \mathbb{R}$ is differentiable at t and

$$d(f \circ \mathbf{r})_t = df_{\mathbf{r}(t)} \circ d\mathbf{r}_t.$$

[Draw a picture.]

Proof. Since \mathbf{r} is differentiable at t and f is differentiable at $\mathbf{r}(t)$, given $\epsilon_1, \epsilon_2 > 0$ there exist $\delta_1, \delta_2 > 0$ such that if $|k| < \delta_1$ and $\|\mathbf{h}\| < \delta_2$, then $t + k \in I$, $\mathbf{r}(t) + \mathbf{h} \in U$ and

$$\|\mathbf{r}(t + k) - \mathbf{r}(t) - d\mathbf{r}_t(k)\| \leq \epsilon_1 |k|, \quad (2.3)$$

$$|f(\mathbf{r}(t) + \mathbf{h}) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(\mathbf{h})| \leq \epsilon_2 \|\mathbf{h}\|. \quad (2.4)$$

Set $\delta = \min\{\delta_1, \frac{\delta_2}{\epsilon_1 + \|d\mathbf{r}_t\|}\}$. Let $k \in \mathbb{R}$ with $|k| < \delta$ and set $\mathbf{h} = \mathbf{r}(t + k) - \mathbf{r}(t)$. Then (2.3) implies

$$\|\mathbf{h} - d\mathbf{r}_t(k)\| \leq \epsilon_1 |k|.$$

By the triangle inequality, we then have

$$\begin{aligned} \|\mathbf{h}\| &= \|\mathbf{h} - d\mathbf{r}_t(k) + d\mathbf{r}_t(k)\| \\ &\leq \|\mathbf{h} - d\mathbf{r}_t(k)\| + \|d\mathbf{r}_t(k)\| \\ &\leq \epsilon_1 |k| + \|d\mathbf{r}_t\| |k| \\ &= (\epsilon_1 + \|d\mathbf{r}_t\|) |k| \\ &< (\epsilon_1 + \|d\mathbf{r}_t\|) \frac{\delta_2}{(\epsilon_1 + \|d\mathbf{r}_t\|)} \\ &= \delta_2 \end{aligned}$$

and therefore by (2.4)

$$\begin{aligned} |f(\mathbf{r}(t + k)) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(d\mathbf{r}_t(k))| &= |f(\mathbf{r}(t + k) - \mathbf{r}(t) + \mathbf{r}(t)) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(d\mathbf{r}_t(k))| \\ &= |f(\mathbf{r}(t) + \mathbf{h}) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(d\mathbf{r}_t(k))| \\ &= |f(\mathbf{r}(t) + \mathbf{h}) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(\mathbf{h}) + df_{\mathbf{r}(t)}(\mathbf{h}) - df_{\mathbf{r}(t)}(d\mathbf{r}_t(k))| \\ &\leq |f(\mathbf{r}(t) + \mathbf{h}) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(\mathbf{h})| + |df_{\mathbf{r}(t)}(\mathbf{h}) - df_{\mathbf{r}(t)}(d\mathbf{r}_t(k))| \\ &= |f(\mathbf{r}(t) + \mathbf{h}) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(\mathbf{h})| + |df_{\mathbf{r}(t)}(\mathbf{h} - d\mathbf{r}_t(k))| \\ &\leq \epsilon_2 \|\mathbf{h}\| + \|df_{\mathbf{r}(t)}\| \epsilon_1 |k| \\ &\leq \epsilon_2 (\epsilon_1 + \|d\mathbf{r}_t\|) |k| + \epsilon_1 \|df_{\mathbf{r}(t)}\| |k| \\ &= (\epsilon_2 (\epsilon_1 + \|d\mathbf{r}_t\|) + \epsilon_1 \|df_{\mathbf{r}(t)}\|) |k| \end{aligned}$$

Thus, given any $\epsilon > 0$, we can find $\delta_1, \delta_2 > 0$ such that (2.3) and (2.4) hold for

$$\begin{aligned} \epsilon_1 &= \frac{1}{2} \frac{\epsilon}{\|df_{\mathbf{r}(t)}\|} \\ \epsilon_2 &= \frac{1}{2} \frac{\epsilon}{\epsilon_1 + \|d\mathbf{r}_t\|}. \end{aligned}$$

Defining δ as before, by taking $|k| < \delta$ we therefore have

$$|f(\mathbf{r}(t + k)) - f(\mathbf{r}(t)) - df_{\mathbf{r}(t)}(d\mathbf{r}_t(k))| \leq \epsilon |k|.$$

This proves that $f \circ \mathbf{r}$ is differentiable at t with differential $df_{\mathbf{r}(t)} \circ d\mathbf{r}_t$. □

Since the matrix representing a composition of linear maps is the product of the matrices representing each linear map [Refer to my Linear Algebra notes], the standard matrix of $d(f \circ \mathbf{r})_t$ is

$$\begin{aligned} [d(f \circ \mathbf{r})_t] &= \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} \\ &= \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}, \end{aligned}$$

evaluated at $(x(t), y(t))$.

Example 2.35. Let $f(x, y) = x^2y + 3xy^4$ and $\mathbf{r}(t) = (x(t), y(t)) = (\sin 2t, \cos t)$. If $F(t) = f(x(t), y(t))$, then

$$\begin{aligned} \frac{dF}{dt} &= \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} \\ &= (2xy + 3y^4)2\cos(2t) + (x^2 + 12xy^3)(-\sin t)|_{(x,y)=(\sin 2t, \cos t)} \\ &= (2\sin(2t)\cos(t) + 3\cos^4(t))2\cos(2t) - (\sin^2(2t) + 12\sin(2t)\cos^3(t))\sin(t) \\ &= 4\sin(2t)\cos(2t)\cos(t) + 6\cos^4(t)\cos(2t) - \sin^2(2t)\sin(t) - 12\sin(2t)\cos^3(t)\sin(t). \end{aligned}$$

Exercise 2.10. Find $(f \circ \mathbf{r})'(t)$ for each function.

(1) $f(x, y) = x^2 + y^2 + xy$, $\mathbf{r}(t) = (\sin t, e^t)$.

(2) $f(x, y) = \cos(x + 4y)$, $\mathbf{r}(t) = (5t^4, t^{-1})$.

2.7 Higher Derivatives

Exercise 2.11. Compute the second partial derivatives of

$$f(x, y) = x^3 + x^2y^3 - 2y^2.$$

Notice for the function in Exercise 2.11 that $\frac{\partial f}{\partial x \partial y} = \frac{\partial f}{\partial y \partial x}$. This is no coincidence.

A Justification of the Arc Length Formula

Proposition A.1. The arc length formula

$$s = \int_a^b \|\mathbf{r}'(t)\| dt. \quad (\text{A.1})$$

can be obtained as the limit of a polygonal approximation of the curve.

Proof. [This proof references an idea here and there which are most likely unfamiliar. Attempts to give a rough explanation are given, but may not be entirely satisfactory.] Let $\mathbf{r} : I \rightarrow \mathbb{R}^n$ be a parametrized curve and let $[a, b] \subseteq I$. Let $a \equiv t_0 < t_1 < t_2 < \cdots < t_n \equiv b$ be a partition of $[a, b]$ where, for simplicity, we take each interval $t_i - t_{i-1}$ to have equal width $\Delta t \equiv \frac{b-a}{n}$. Approximate

If f is a function of two variables, then its partial derivatives f_x and f_y are also functions of two variables, so we can consider their partial derivatives $(f_x)_x$, $(f_x)_y$, $(f_y)_x$, and $(f_y)_y$, which are called the **second partial derivatives** of f . If $z = f(x, y)$, we use the following notation:

$$(f_x)_x = f_{xx} = f_{11} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 z}{\partial x^2}$$

$$(f_x)_y = f_{xy} = f_{12} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 z}{\partial y \partial x}$$

$$(f_y)_x = f_{yx} = f_{21} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 z}{\partial x \partial y}$$

$$(f_y)_y = f_{yy} = f_{22} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial y} \right) = \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 z}{\partial y^2}$$

Thus the notation f_{xy} (or $\partial^2 f / \partial y \partial x$) means that we first differentiate with respect to x and then with respect to y , whereas in computing f_{yx} the order is reversed.

the length of curve between $\mathbf{r}(t_{i-1})$ and $\mathbf{r}(t_i)$ by a straight line. We then define the arc length s between $\mathbf{r}(a)$ and $\mathbf{r}(b)$ by

$$\begin{aligned} s &\equiv \lim_{n \rightarrow \infty} \sum_{i=1}^n \|\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})\| \\ &\equiv \lim_{n \rightarrow \infty} \sum_{i=1}^n \sqrt{(x(t_i) - x(t_{i-1}))^2 + (y(t_i) - y(t_{i-1}))^2}. \end{aligned}$$

Since the coordinate functions $x(t)$ and $y(t)$ are continuous on $[a, b]$ and differentiable on (a, b) , by the Mean Value Theorem there exist t_i^* and t_i^{**} in (a, b) such that

$$\begin{aligned} x(t_i) - x(t_{i-1}) &= x'(t_i^*)(t_i - t_{i-1}) \\ &= x'(t_i^*)\Delta t \\ y(t_i) - y(t_{i-1}) &= y'(t_i^{**})(t_i - t_{i-1}) \\ &= y'(t_i^{**})\Delta t \end{aligned}$$

and therefore

$$s = \lim_{n \rightarrow \infty} \sum_{i=1}^n \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2} \Delta t.$$

If t_i^* were equal to t_i^{**} for all $i = 1, \dots, n$, then this would be a Riemann sum for $\int_a^b \|\mathbf{r}'(t)\| dt$. However, in general they are not equal, so we will instead show that the difference between this

SOLUTION In Example 1 we found that

$$f_x(x, y) = 3x^2 + 2xy^3 \quad f_y(x, y) = 3x^2y^2 - 4y$$

Therefore

$$\begin{aligned} f_{xx} &= \frac{\partial}{\partial x} (3x^2 + 2xy^3) = 6x + 2y^3 & f_{xy} &= \frac{\partial}{\partial y} (3x^2 + 2xy^3) = 6xy^2 \\ f_{yx} &= \frac{\partial}{\partial x} (3x^2y^2 - 4y) = 6xy^2 & f_{yy} &= \frac{\partial}{\partial y} (3x^2y^2 - 4y) = 6x^2y - 4 \end{aligned}$$

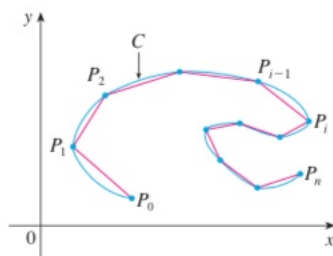


Figure 30: A polygonal approximation of a regular curve.

expression and the Riemann sum

$$\int_a^b \|\mathbf{r}'(t)\| dt \equiv \lim_{n \rightarrow \infty} \sum_{i=1}^n \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} \Delta t$$

goes to zero as $n \rightarrow \infty$.⁹ First, note that since \mathbf{r} is, in particular, continuously differentiable on each closed interval $[t_{i-1}, t_i]$, the functions x' and y' are uniformly continuous on $[t_{i-1}, t_i]$.¹⁰ Thus, given any $\epsilon > 0$ there exists a $\delta > 0$ such that $|t_i^* - t_i^{**}| < \delta$ implies $|y'(t_i^*) - y'(t_i^{**})| < \epsilon$. Since $|t_i^* - t_i^{**}| < t_i - t_{i-1} = \Delta t$, if we take $\Delta t < \delta$, then we will have $|y'(t_i^*) - y'(t_i^{**})| < \epsilon$. Note that this is possible since, given any $\delta > 0$, by the Archimedean property of the real numbers we can find a positive integer n such that $\Delta t = \frac{b-a}{n} < \delta$. Assuming now that we have chosen $\Delta t < \delta$, we

⁹Intuitively, this should make sense, since t_i^* and t_i^{**} both lie within $[t_{i-1}, t_i]$ and by shrinking Δt , each of these subintervals also shrinks and therefore we should be able to make the points t_i^* and t_i^{**} become arbitrarily close to each other by making Δt arbitrarily small.

¹⁰The proof of this involves the notion of “compactness”, the explanation of which would take us beyond the scope of this course. A function is said to be *uniformly continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that for any point $t \in [t_{i-1}, t_i]$, $|t' - t| < \delta$ implies $\|f(t) - f(t')\| < \epsilon$; that is, the *same* δ works for every point $t \in [t_{i-1}, t_i]$. This is not true when a function is continuous but not uniformly continuous, since in that case δ depends on both ϵ and t .

then have

$$\begin{aligned}
 & \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2} \right| \\
 &= \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2 + (y'(t_i^*))^2 - (y'(t_i^*))^2} \right| \\
 &= \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2 + (y'(t_i^{**}))^2 - (y'(t_i^*))^2} \right|.
 \end{aligned}$$

Now $(y'(t_i^{**}))^2 - (y'(t_i^*))^2 \leq |(y'(t_i^{**}))^2 - (y'(t_i^*))^2|$, so, by the triangle inequality,

$$\begin{aligned}
 \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2 + (y'(t_i^{**}))^2 - (y'(t_i^*))^2} &\leq \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2 + |(y'(t_i^{**}))^2 - (y'(t_i^*))^2|} \\
 &\leq \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} + \sqrt{|(y'(t_i^{**}))^2 - (y'(t_i^*))^2|}
 \end{aligned}$$

and therefore

$$\begin{aligned}
 & \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2} \right| \\
 &= \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2 + (y'(t_i^{**}))^2 - (y'(t_i^*))^2} \right| \\
 &\leq \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \left(\sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} + \sqrt{|(y'(t_i^{**}))^2 - (y'(t_i^*))^2|} \right) \right| \\
 &= \left| -\sqrt{|(y'(t_i^{**}))^2 - (y'(t_i^*))^2|} \right| \\
 &= \sqrt{|(y'(t_i^{**}))^2 - (y'(t_i^*))^2|} \\
 &= \sqrt{|(y'(t_i^{**}) + (y'(t_i^*))(y'(t_i^{**}) - (y'(t_i^*)))|} \\
 &= \sqrt{|(y'(t_i^{**}) + (y'(t_i^*)))|} \sqrt{|(y'(t_i^{**}) - (y'(t_i^*)))|}
 \end{aligned}$$

Since y' is continuous and $[a, b]$ a closed interval, $y'([a, b])$ is bounded by some $M > 0$, so ¹¹

$$\begin{aligned}
 &= \sqrt{|(y'(t_i^{**}) + (y'(t_i^*)))|} \sqrt{|(y'(t_i^{**}) - (y'(t_i^*)))|} \\
 &\leq \sqrt{2M} \sqrt{|(y'(t_i^{**}) - (y'(t_i^*)))|},
 \end{aligned}$$

so by choosing $\Delta t < \delta$ such that $|(y'(t_i^{**}) - (y'(t_i^*)))| < \frac{\epsilon^2}{4M^2(b-a)^2}$, then

$$\begin{aligned}
 & \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2} \right| \\
 &\leq \sqrt{2M} \sqrt{|(y'(t_i^{**}) - (y'(t_i^*)))|} \\
 &< \sqrt{2M} \sqrt{\frac{\epsilon^2}{4M^2(b-a)^2}} \\
 &= \frac{\epsilon}{b-a}
 \end{aligned}$$

¹¹Here we have used the facts that the image of a compact set under a continuous mapping is compact, and that a compact set is bounded.

and therefore

$$\begin{aligned}
 & \left| \sum_{i=1}^n \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} \Delta t - \sum_{i=1}^n \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2} \Delta t \right| \\
 &= \left| \sum_{i=1}^n \left(\sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2} \right) \right| \Delta t \\
 &\leq \sum_{i=1}^n \left| \sqrt{(x'(t_i^*))^2 + (y'(t_i^*))^2} - \sqrt{(x'(t_i^*))^2 + (y'(t_i^{**}))^2} \right| \Delta t \\
 &< n \frac{\epsilon}{b-a} \Delta t \\
 &= n \frac{\epsilon}{b-a} \frac{(b-a)}{n} \\
 &= \epsilon.
 \end{aligned}$$

□

B Topology of \mathbb{R}^n

A *topology* is a structure on a set which allows us to define the notion of continuity. In this appendix, we collect some basic definitions used throughout the rest of the notes.

B.1 Normed Linear Spaces

Definition B.1 (Normed linear space). Let V be a real vector space. A *norm* on V is a function

$$\rho : V \rightarrow \mathbb{R}^{\geq 0}$$

such that for all $\xi, \eta \in V$ and $c \in \mathbb{R}$ we have

1. $\rho(\xi) = 0$ if and only if $\xi = 0$,
2. $\rho(c\xi) = |c|\rho(\xi)$,
3. $\rho(\xi + \eta) \leq \rho(\xi) + \rho(\eta)$.

The pair (V, ρ) is called a *normed linear space*.

Property (3) is called the *triangle inequality*. We often use the notation $\|\xi\| = \rho(\xi)$ for the norm.

Example B.2. For $V = \mathbb{R}^n$, define $\rho : \mathbb{R}^n \rightarrow \mathbb{R}^{\geq 0}$ by $\rho(\mathbf{x}) = \sqrt{\sum_{i=1}^n x_i^2}$. This norm on \mathbb{R}^n is called the *Euclidean norm* and the normed linear space (\mathbb{R}^n, ρ) is called *Euclidean space*.

Exercise B.1. Prove that the Euclidean norm on \mathbb{R}^n is a norm.

[Def of equivalent norms.]

[All norms on a finite dimensional normed linear space are equivalent.]

Definition B.3 (Continuous linear maps). Let V, W be normed linear spaces. A linear map $T : V \rightarrow W$ is *bounded* if there exists $C > 0$ such that

$$\|T\mathbf{v}\|_W \leq C\|\mathbf{v}\|_V, \quad \text{for all } \mathbf{v} \in V.$$

[Define operator norm on $\text{Hom}(V, W)$.]

Proposition B.4 (Equivalent definition of a bounded linear maps). Let $T : V \rightarrow W$ be a bounded linear map, and let $S(V) \equiv \{\mathbf{v} \in V : \|\mathbf{v}\|_V = 1\} \subseteq V$ be the unit sphere in V . Then T is bounded if and only if $T(S(V)) \subseteq W$ is bounded.

Proof. Suppose T is bounded and suppose $\mathbf{v} \in S(V)$. Then $\|\mathbf{v}\|_V = 1$ and therefore

$$\|T\mathbf{v}\|_W \leq C\|\mathbf{v}\|_V = C.$$

Thus, $T(S(V))$ is bounded.

Conversely, suppose $T(S(V))$ is bounded. For any $\mathbf{v} \in V$, $\frac{\mathbf{v}}{\|\mathbf{v}\|_V} \in S(V)$ and by assumption there exists $C > 0$ such that

$$\|T\frac{\mathbf{v}}{\|\mathbf{v}\|_V}\|_W = \frac{1}{\|\mathbf{v}\|_V}\|T\mathbf{v}\|_V \leq C$$

and therefore

$$\|T\mathbf{v}\|_W \leq C\|\mathbf{v}\|_V.$$

Hence, T is bounded. □

B.2 Metric Spaces

Definition B.5 (Metric space). A metric space (X, d) is a set X together with a function

$$d : X \times X \rightarrow \mathbb{R}^{\geq 0}$$

such that for all $x, y, z \in X$ we have

1. $d(x, y) = 0$ if and only if $x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, y) \leq d(x, z) + d(z, y)$.

Again, property (3) is called the *triangle inequality*.

Proposition B.6. Let (V, ρ) be a normed linear space. Then V inherits a metric $d_\rho : V \times V \rightarrow \mathbb{R}^{\geq 0}$ defined by

$$d_\rho(\xi, \eta) = \rho(\xi - \eta), \quad \xi, \eta \in V.$$

Proof. 1. $d_\rho(\xi, \eta) = \rho(\eta - \xi) = 0$ if and only if $\eta - \xi = 0$ if and only if $\eta = \xi$.

$$2. \quad d_\rho(\eta, \xi) = \rho(\xi - \eta) = \rho(-1(\eta - \xi)) = |-1|\rho(\eta - \xi) = \rho(\eta - \xi) = d_\rho(\xi, \eta).$$

$$3. \quad d_\rho(\xi, \eta) = \rho(\eta - \xi) = \rho(\eta - \alpha + \alpha - \xi) \leq \rho(\eta - \alpha) + \rho(\alpha - \xi) = d_\rho(\alpha, \eta) + d_\rho(\xi, \alpha).$$

□

Example B.7. The metric $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^{\geq 0}$ defined by

$$||\mathbf{x} - \mathbf{y}|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

(called the *Euclidean metric*) gives (\mathbb{R}^n, d) the structure of a metric space.

Definition B.8 (Open ball). Let $r > 0$. The subset

$$B_r(\mathbf{x}_0) := \{\mathbf{x} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{x}_0) < r\}$$

is called the *open ball of radius r centered at \mathbf{x}_0* .

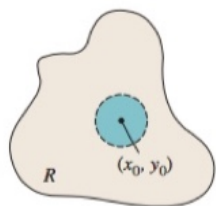
Example B.9. (1) If $n = 1$, then $B_r(x_0)$ is the open interval $(x_0 - r, x_0 + r)$.

(2) If $n = 2$, then $B_r(\mathbf{x}_0)$ is the open disc of radius r centered at \mathbf{x}_0 .

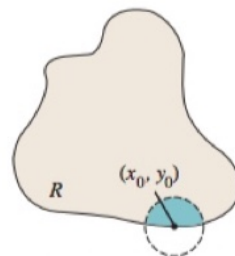
Definition B.10 (Interior and boundary points). Let U be a subset of \mathbb{R}^n and let $\mathbf{x}_0 \in U$.

- (1) The point \mathbf{x}_0 is an *interior point* of U if there exists $r > 0$ such that $B_r(\mathbf{x}_0) \subseteq U$. The *interior* of U is the set of interior points of U .
- (2) The point \mathbf{x}_0 is a *boundary point* of U if for every $r > 0$, $B_r(\mathbf{x}_0)$ contains a point of U as well as a point of $\mathbb{R}^n - U$. The boundary point itself need not belong to U . The *boundary* of U is the set of boundary points of U .

Definition B.11 (Open and closed sets). (1) A subset $U \subseteq \mathbb{R}^n$ is *open* if it consists entirely of interior points.



(a) Interior point.



(b) Boundary point.

(2) A subset $U \subseteq \mathbb{R}^n$ is *closed* if it contains all its boundary points. [Update: A subset $U \subseteq \mathbb{R}^n$ is *closed* if it contains all its limit points.]

Example B.12. Every open ball is an open set.

Proof. Let $\mathbf{x} \in B_r(\mathbf{x}_0)$. Then $h = r - d(\mathbf{x}, \mathbf{x}_0) > 0$. Let $\mathbf{y} \in B_h(\mathbf{x})$. Then $d(\mathbf{y}, \mathbf{x}) < h$. It then follows from the triangle inequality that

$$\begin{aligned} d(\mathbf{x}_0, \mathbf{y}) &\leq d(\mathbf{x}_0, \mathbf{x}) + d(\mathbf{x}, \mathbf{y}) \\ &< r - h + h = r, \end{aligned}$$

which shows that $\mathbf{y} \in B_r(\mathbf{x}_0)$, and therefore $B_h(\mathbf{x}) \subseteq B_r(\mathbf{x}_0)$. Thus, $B_r(\mathbf{x}_0)$ is open. \square

Definition B.13 (Neighborhood). A *neighborhood* of a point $\mathbf{x} \in \mathbb{R}^n$ is an open set containing \mathbf{x} .

Remark. Note that a set can be neither open nor closed, or can be both open and closed. \square

Example B.14. (a) \mathbb{R}^n is open since for any point $\mathbf{x} \in \mathbb{R}^n$ we can find an open ball centered at \mathbf{x} contained in \mathbb{R}^n . The boundary of \mathbb{R}^n is empty, so \mathbb{R}^n trivially contains all of its boundary points. Thus, \mathbb{R}^n is also closed.

(b) The half-open interval $(0, 1] \subseteq \mathbb{R}$ is neither open nor closed.

Definition B.15 (Limit points and isolated points). Let U be a subset of \mathbb{R}^n .

(a) A point \mathbf{x} is a *limit point* of the set U if every neighborhood of \mathbf{x} contains a point of U different from \mathbf{x} .

(b) If $\mathbf{x} \in U$ and \mathbf{x} is not a limit point of U , then \mathbf{x} is called an *isolated point* of U .

Theorem B.16. If \mathbf{x} is a limit point of a set U , then every neighborhood of \mathbf{x} contains infinitely many points of U .

Proof. (By contradiction) Suppose V is a neighborhood of \mathbf{x} which contains only finitely many points of U . Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be the points of $V \cap U$ which are distinct from \mathbf{x} , and let $r = \min\{d(\mathbf{x}, \mathbf{x}_i)\}_{i=1}^n$. Since the minimum of a finite set of positive numbers is positive, $r > 0$. Then the neighborhood $B_r(\mathbf{x})$ contains no point of U distinct from \mathbf{x} , which is a contradiction since \mathbf{x} is a limit point of U . \square

Corollary B.17. A finite point set has no limit points.

Example B.18. Consider the set $U = \{\frac{1}{n}\}_{n=1}^{\infty} \subseteq \mathbb{R}$. By the Archimedian property of the real numbers ¹², given any $\epsilon > 0$ there exists $n \in \mathcal{N}$ such that $\frac{1}{n} < \epsilon$. Thus, 0 is a limit point of U , since every open ball $B_{\epsilon}(0)$ contains a point of U . Note however that every element of U itself is an isolated point of U since, for any $\frac{1}{n} \in U$, the open ball $B_{d(\frac{1}{n}, \frac{1}{n+1})}(\frac{1}{n})$ contains only $\frac{1}{n}$. This illustrates the difference between a set having a limit point and containing a limit point.

Theorem B.19. A set U is open if and only if its complement is closed.

Proof. Suppose $U^c \equiv \mathbb{R}^n - U$ is closed. Let $\mathbf{x} \in U$. Then $\mathbf{x} \notin U^c$ and \mathbf{x} is not a limit point of U^c . Hence there exists a neighborhood N of \mathbf{x} such that $U^c \cap N$ is empty, that is, $N \subseteq U$. Thus \mathbf{x} is an interior point of U , so U is open.

Conversely, suppose that U is open. Let \mathbf{x} be a limit point of U^c . Then every neighborhood of \mathbf{x} contains a point of U^c , so that \mathbf{x} is not an interior point of U . Since U is open, this means that $\mathbf{x} \in U^c$. It follows that U^c is closed. \square

Corollary B.20. A set V is closed if and only if its complement is open.

[Need to define relative open sets in order to define a connected subset of \mathbb{R}^n .]

Definition B.21 (Bounded sets). A subset $U \subseteq \mathbb{R}^n$ is *bounded* if it is contained in some ball of finite radius. If U is not bounded, then we say it is *unbounded*.

Example B.22. 1. An open ball is bounded.

2. A line is unbounded.

B.3 Continuous maps

Definition B.23 (Lipschitz continuity). Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \rightarrow Y$ is *Lipschitz continuous* if there exists a positive real number K such that

$$d_Y(f(x_1) - f(x_2)) \leq K d_X(x_1 - x_2)$$

for all $x_1, x_2 \in X$. The real number K is said to be a *Lipschitz constant* for the function f .

Lemma B.24. If f is Lipschitz continuous, then it is uniformly continuous.

Proof. Given $\epsilon > 0$, choose $\delta = \frac{\epsilon}{K}$. Then

$$d_Y(f(x_1) - f(x_2)) \leq K \frac{\epsilon}{K} = \epsilon$$

whenever $d_X(x_1, x_2) < \delta$. \square

Theorem B.25. Let $T : V \rightarrow W$ be a linear map between normed linear spaces. The following are equivalent:

- (i) T is bounded
- (ii) T is uniformly continuous
- (iii) T is continuous at $0 \in V$

¹²[Explain this.]

Proof. To see that (i) implies (ii), if T is bounded then

$$\|T(\mathbf{v} - \mathbf{w})\|_W = \|T\mathbf{v} - T\mathbf{w}\|_W \leq C\|\mathbf{v} - \mathbf{w}\|_V,$$

which says that T is Lipschitz continuous, and therefore is uniformly continuous. To see that (ii) implies (iii), just take $\mathbf{w} = 0$ above. To prove that (iii) implies (i), if T is not bounded then choose $\mathbf{v}_n \in V$ such that $\|\mathbf{v}_n\|_V = 1$ and $\|T\mathbf{v}_n\|_W > n$. Then $\mathbf{w}_n \equiv \mathbf{v}_n/n$ satisfies $\lim_{n \rightarrow \infty} \mathbf{w}_n = 0$ but $\|T\mathbf{w}_n\|_W > 1$, so the sequence $(T\mathbf{w}_n)$ does not converge to $0 \in W$. Hence, T is not continuous at $0 \in V$. \square

Theorem B.26. Let V, W be normed linear spaces and assume V is finite dimensional. Then any linear map $T : V \rightarrow W$ is bounded.

Proof. [Finish.] \square

C Affine Space

In our discussion of differentiability, it will be helpful to view points and vectors as distinct. This leads to the notion of an affine space over a vector space.

Definition C.1 (Affine space). Let V be a vector space. An *affine space* A over V is a set A together with a map

$$\psi : V \times A \rightarrow A \tag{C.1}$$

$$\psi(\mathbf{v}, p) = p + \mathbf{v} \tag{C.2}$$

with the following properties: ¹³

1. $p + \mathbf{0} = p$ for all $p \in A$.
2. $p + (\mathbf{v} + \mathbf{w}) = (p + \mathbf{v}) + \mathbf{w}$ for all $p \in A, \mathbf{v}, \mathbf{w} \in V$.
3. For every $p_1, p_2 \in A$, there exists a unique $\mathbf{v} \in V$ such that $p_2 = p_1 + \mathbf{v}$. Since \mathbf{v} is uniquely determined by p_1 and p_2 , we write $\mathbf{v} = p_2 - p_1$.

Elements of A are called *points*; elements of V are called *vectors*. While we have not been very careful about the distinction between points and vectors so far, they play very different roles. For example, we might model time by an affine space A over a one-dimensional vector space V . Points of A represent instants of time, whereas elements of V represent intervals of time. Notice that it makes sense to add intervals of time, say 3 hours + 4 hours = 7 hours, whereas 3:00 PM + 4:00 PM does not make good sense. Then again, that the difference 4:00 PM – 3:00 PM (the same day) is the time interval 1 hour is part of our intuition about time. Similarly, in a flat model of the Earth we do not try to make sense of the sum of Chicago and New York, but their difference as a displacement vector does make sense.

Proposition C.2. Every vector space V is an affine space over itself.

¹³Recall that a vector space consists of the data $(V, 0, +, \cdot)$, that is, a set V , a distinguished vector in V (the zero vector), together with operations of vector addition and scalar multiplication. In particular, the vector space axioms tell us that $(V, 0, +)$ is an *abelian group*. A map (C.1) satisfying properties 1-3 is said to be a *simply transitive action* of the abelian group $(V, 0, +)$ on A .

Proof. Take $A = V$ and define

$$\begin{aligned}\psi : V \times V &\rightarrow V \\ \psi(\mathbf{v}, \mathbf{w}) &= \mathbf{v} + \mathbf{w}\end{aligned}$$

to be vector addition. Then

1. $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$ by definition of the zero vector.
2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ by associativity of vector addition.
3. For any $\mathbf{v}, \mathbf{w} \in V$, we have

$$\begin{aligned}\mathbf{w} &= \mathbf{0} + \mathbf{w} \\ &= (\mathbf{v} - \mathbf{v}) + \mathbf{w} \\ &= \mathbf{v} + (\mathbf{w} - \mathbf{v}).\end{aligned}$$

To see that $\mathbf{w} - \mathbf{v}$ is the unique vector with this property, suppose that

$$\mathbf{w} = \mathbf{v} + \mathbf{u}$$

for some vector $\mathbf{u} \in V$. Then

$$\mathbf{v} + (\mathbf{w} - \mathbf{v}) = \mathbf{v} + \mathbf{u}.$$

Adding $-\mathbf{v}$ to both sides gives

$$\begin{aligned}-\mathbf{v} + (\mathbf{v} + (\mathbf{w} - \mathbf{v})) &= -\mathbf{v} + (\mathbf{v} + \mathbf{u}) \\ (-\mathbf{v} + \mathbf{v}) + (\mathbf{w} - \mathbf{v}) &= (-\mathbf{v} + \mathbf{v}) + \mathbf{u} \\ \mathbf{0} + (\mathbf{w} - \mathbf{v}) &= \mathbf{0} + \mathbf{u} \\ \mathbf{w} - \mathbf{v} &= \mathbf{u}\end{aligned}$$

which shows that $\mathbf{w} - \mathbf{v}$ is unique. □

The affine space in Proposition C.2 is called the *canonical affine space over V* . Loosely, one says that one obtains this affine space from V by “forgetting the zero vector”.

Definition C.3 (Affine map). Let A be an affine space over a vector space V and B be an affine space over a vector space W . A function $f : A \rightarrow B$ is an *affine map* if there exists a linear map $T : V \rightarrow W$ such that

$$f(p + \mathbf{v}) = f(p) + T\mathbf{v}$$

for all $p \in A, \mathbf{v} \in V$. The linear map T is called the *differential* of f , and we write $T = df$.

Proposition C.4 (Uniqueness of the differential). Let $f : A \rightarrow B$ be an affine map. Then the differential of f is unique.

Proof. Suppose there exist linear maps $T, T' : V \rightarrow V$ such that

$$f(p + \mathbf{v}) = f(p) + T\mathbf{v} \quad \text{and} \quad f(p + \mathbf{v}) = f(p) + T'\mathbf{v}.$$

Then

$$f(p) + T\mathbf{v} = f(p) + T'\mathbf{v}.$$

Since $T\mathbf{v} \in W, -T\mathbf{v} \in W$, so we can act with $-T\mathbf{v}$ on both sides of this equation to obtain

$$(f(p) + T\mathbf{v}) - T\mathbf{v} = (f(p) + T'\mathbf{v}) - T\mathbf{v}.$$

By properties 1 and 2 of Definition C.1,

$$\begin{aligned} (f(p) + T\mathbf{v}) - T\mathbf{v} &= (f(p) + T'\mathbf{v}) - T\mathbf{v} \\ f(p) + (T\mathbf{v} - T\mathbf{v}) &= f(p) + (T'\mathbf{v} - T\mathbf{v}) \\ f(p) + \mathbf{0} &= f(p) + (T'\mathbf{v} - T\mathbf{v}) \\ f(p) &= f(p) + (T'\mathbf{v} - T\mathbf{v}). \end{aligned}$$

By property 3 of Definition C.1, $\mathbf{u} = \mathbf{0}$ is the unique vector such that $f(p) = f(p) + \mathbf{u}$, so we must have

$$T'\mathbf{v} - T\mathbf{v} = (T' - T)\mathbf{v} = \mathbf{0},$$

which says that $\mathbf{v} \in \ker(T' - T)$. Since \mathbf{v} was an arbitrary vector in V , we have $\ker(T' - T) = V$, so $T' - T$ is the zero map, and therefore $T' = T$. \square

Definition C.5 (Affine isomorphism). Let A and B be affine spaces over vector spaces V and W , respectively. A bijective affine map $f : A \rightarrow B$ is said to be an *affine isomorphism*.

Proposition C.6. Let $f : A \rightarrow B$ and $g : B \rightarrow C$ be affine maps. Then $g \circ f : A \rightarrow C$ is an affine map.

Proof. If

$$\begin{aligned} f(p + \mathbf{v}) &= f(p) + df(\mathbf{v}), \\ g(p + \mathbf{v}) &= g(p) + dg(\mathbf{v}), \end{aligned}$$

then

$$\begin{aligned} g(f(p + \mathbf{v})) &= g(f(p) + df(\mathbf{v})) \\ &= g(f(p)) + dg(df(\mathbf{v})) \\ &= (g \circ f)(p) + (dg \circ df)(\mathbf{v}), \end{aligned}$$

so $g \circ f$ is an affine map whose differential is $dg \circ df$. \square

[Add more if we end up using it.]