# Differential Vector Calculus

Joan E. Licata

March 20, 2012

**Preface**

This book is written for the differential vector calculus portion of Math 51 at Stanford, and I hope you will find it helpful, interesting, and accessible. Although I have made every effort to proofread for typographical and mathematical errors, I am sure that some have slipped through. As a favor to future students, I ask you to let me know about any mistakes you find. Please email me with any comments, corrections, or suggestions you come up with! There is also a list of known errata on my webpage, and this will be updated as new items come to light.

**Using this book**

Every section comes in three parts. The body of the section introduces important definitions, theorems, and examples. A Worked Examples section follows the main exposition. In the Worked Examples, you will find more detailed explanations and in some cases, more complicated problems. These examples are meant to help the transition from reading about mathematics to practicing it. Finally, every section ends with a collection of exercises. In most sections, there are more exercises than would reasonably be assigned in a single course, so the Exercises sections offer opportunities to get additional practice. Answers and hints for most of the odd-numbered problems appear in the back of the book, along with an Appendix which contains some of the more technical proofs. Finally, there is an index of terms and notation immediately before the Answers section.

**Acknowledgments**

The development of this text was made possible by an award from the Hoagland Award Fund for Innovations in Undergraduate Teaching, and I greatly appreciate their support. Many of the exercises were developed by Jeffrey Danciger, Robin Koytcheff, and Yu-jong Tzeng. I am also grateful to Steven Levandosky for sharing the style files from his Linear Algebra book. Finally, I am extremely grateful to the students and teaching staff of previous Math 51 courses who caught typos, sent suggestions, and recommended improvements.

The figures in this text were created using Mathematica 6.0.3 and Adobe Illustrator CS3. The topographical map in Section 1 is courtesy of the United States Geographical Survey.

I hope you will find this text to be a useful resource, and I would appreciate any suggestions for improvement. Good luck, and enjoy Math 51!

Best,

Joan Licata

# Contents

# 1 Functions of several variables

In this section we introduce some of the terminology and notation that will be useful for studying the calculus of functions of several variables.

## 1.1 Multivariable functions

We think of a function $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ as a rule which assigns a unique element in $\mathbf{R}^m$ to each element in $\mathbf{R}^n$. More generally, a function may only be defined on a subset of Euclidean space, and we write

$$\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$$

when $\mathbf{f}$ assigns a unique element of $\mathbf{R}^m$ to each vector in the subset $\mathcal{D}^n \subset \mathbf{R}^n$. In this case, we call the set $\mathcal{D}^n$ the *domain* of the function $\mathbf{f}$ and $\mathbf{R}^m$ the *codomain* of $\mathbf{f}$. The image of $\mathcal{D}^n$ under $\mathbf{f}$ is called the *range* of $\mathbf{f}$.

**Example 1.1.** Suppose that $M$ is an $m \times n$ matrix and $F : \mathbf{R}^n \to \mathbf{R}^m$ is the linear transformation defined by

$$F(\mathbf{x}) = M\mathbf{x}.$$

The domain of $F$ is $\mathbf{R}^n$ and the codomain of $F$ is $\mathbf{R}^m$. The range of $F$ is the column space of $M$.

When the codomain of a function is $\mathbf{R}$, we say that the function is *real-valued* or *scalar-valued*. When the codomain of a function is $\mathbf{R}^n$ for $n > 1$, we say that the function is *vector-valued*.

We will be particularly interested in real-valued multivariable functions. In fact, any vector-valued multivariable function can be described in terms of a sequence of scalar-valued functions.

**Definition 1.1.** Given a function $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$, we can write $\mathbf{f}$ as

$$\mathbf{f}(x_1, x_2, ...x_n) = (f_1(x_1, x_2, ...x_n), f_2(x_1, x_2, ...x_n), ...f_m(x_1, x_2, ...x_n)).$$

The function $f_i : \mathcal{D}^n \to \mathbf{R}$ is called the $i^{th}$ *coordinate function of* $\mathbf{f}$.

Many properties of vector-valued functions are easiest to study via the corresponding coordinate functions. For example, a function $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ is even if and only if all its coordinate functions are even, and $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ is odd if and only if all its coordinate functions are odd. (See Exercises 10 and 11.)

## 1.2 Graphs

It's often useful to view a function as a set of pairs of points:

**Definition 1.2.** Given a function $f : \mathcal{D}^n \to \mathbf{R}^m$, the *graph of* $f$ is the set of points

$$\Gamma_f = \{(\mathbf{a}, \mathbf{f}(\mathbf{a})) \in \mathbf{R}^{n+m} \mid \mathbf{a} \in \mathcal{D}^n\}.$$

**Example 1.2.** *Describe the graph of the function* $f(x, y) = 1 + \sqrt{1 - x^2 - y^2}$.

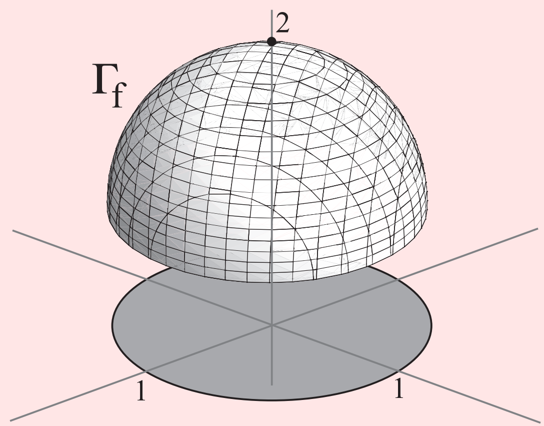
When the domain of a function isn't explicitly given, we interpret it as being the largest subset of Euclidean space that makes sense. Since the square root function is only defined for non-negative numbers, the domain of $f$ is the set

$$\mathcal{D}^2 = \{(x, y) \mid 1 - x^2 - y^2 \geq 0\}.$$

By definition, the graph of $f$ is the set

$$\{(x, y, 1 + \sqrt{1 - x^2 - y^2}) \mid 1 - x^2 - y^2 \geq 0\}.$$

The figure shows the graph $\Gamma_f$, together with the shaded unit disc indicating the $x$ and $y$ values of points in $\mathcal{D}^2$.



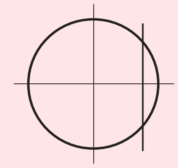

## 1.3 The Line Test

Given a function $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$, the graph of $\mathbf{f}$ is a subset of $\mathbf{R}^{n+m}$. However, not all subsets of $\mathbf{R}^{m+n}$ are graphs of functions.

The definition of a function implies that each point in the domain is paired with just one point in the range. Thus, if $(x, y_1)$ and $(x, y_2)$ are two points in the graph of a single variable function, then $y_1 = y_2$. Checking this condition is sometimes referred to as the *Vertical Line Test*.

**Example 1.3.** *Show that the unit circle in* $\mathbf{R}^2$ *is not the graph of a function of* $x$.

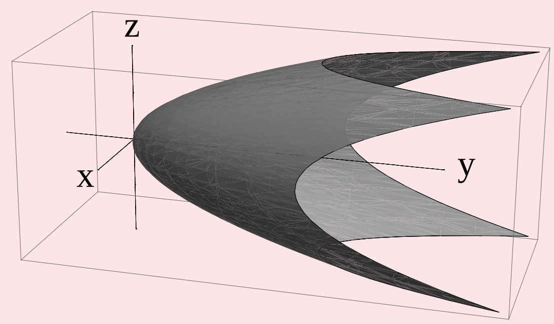
The unit circle is the set of points satisfying $x^2 + y^2 = 1$, but the line $x = \frac{2}{3}$ intersects the circle twice. This shows that the unit circle is not the graph of $y$ as a function of $x$.



We can generalize this idea to real-valued functions of several variables:

**Proposition 1.1** (The Line Test)**.** Let $\Gamma \subset \mathbf{R}^{n+1}$ be the graph of a function $f : \mathbf{R}^n \to \mathbf{R}$. If the codomain of $f$ is identified with the subspace spanned by $\mathbf{e_i}$, then any line in $\mathbf{R}^{n+1}$ with direction vector $\mathbf{e_i}$ intersects $\Gamma$ no more than once.

**Example 1.4.** *Let $S$ be the set of points in $\mathbf{R}^3$ which satisfy $x^2 + z^2 - y = 0$. Is $S$ the graph of a function of $x$ and $y$? Of $x$ and $z$? Of $y$ and $z$?*




There are vertical lines which intersect $S$ more than once, so the $\mathbf{e_3}$ line test shows that $S$ is not the graph of a function $f(x, y) = z$.
Similarly, $S$ is not the graph of a function of $y$ and $z$, because the line below intersects $S$ in two points:

$$\mathbf{L}(t) = \mathbf{e_2} + t\mathbf{e_1}.$$

However, $S$ is the graph of the function $h(x, z) = x^2 + z^2$. Every line parallel to $\mathbf{e_2}$ intersects $S$ exactly once in a point of the form $(x, x^2 + z^2, z)$.

## 1.4 Level sets

If $f : \mathcal{D}^n \to \mathbf{R}$ is a real-valued function of several variables, it is sometimes useful to consider all the points in $\mathcal{D}^n$ which map to the same point in the codomain.

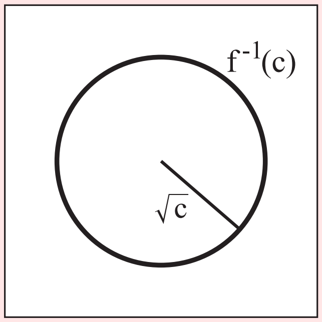
**Definition 1.3.** Given $f : \mathcal{D}^n \to \mathbf{R}$, the *level set of $f$ at height $c$* is the pre-image of $c \in \mathbf{R}$:

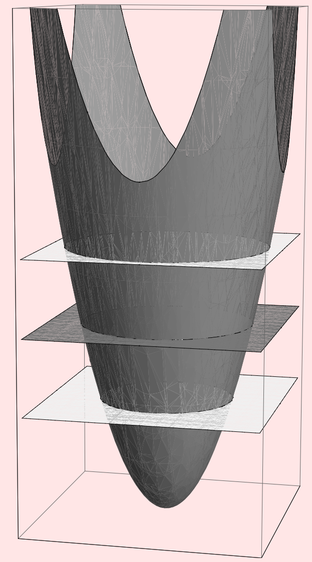
$$f^{-1}(c) = \{\mathbf{x} \in \mathcal{D}^n \mid f(\mathbf{x}) = c\}.$$

**Example 1.5.** *Let $f(x, y) = x^2 + y^2$. What do the level sets of $f$ look like?*
For any $c \in \mathbf{R}$, let

$$f^{-1}(c) = \{(x, y) \mid x^2 + y^2 = c\}.$$

If $c < 0$, then $f^{-1}(c)$ is empty.
If $c = 0$, then $f^{-1}(0)$ is the single point $(0, 0)$.
If $c > 0$, then $f^{-1}(c)$ is the circle in $\mathbf{R}^2$ centered at the origin with radius $\sqrt{c}$.



Each level set $f^{-1}(c)$ is a subset of the domain of $f$. In order to picture these level sets, it's convenient to think of the intersection of the graph $\Gamma_f$ with the plane $z = c$.

**Example 1.6.** *Show that the unit sphere in* $\mathbf{R}^3$ *is a level set of a function of three variables.* Recall that the unit sphere in $\mathbf{R}^3$ is the set of points

$$\{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 = 1\}.$$

Define the function $F : \mathbf{R}^3 \to \mathbf{R}$ by $F(x, y, z) = x^2 + y^2 + z^2$. Then we see that the unit sphere is exactly the level set of $F$ at height $1$:

$$\{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 = 1\} = F^{-1}(1).$$

Example 1.6 illustrates the general principle that many surfaces can be described as level sets of functions of three variables. In fact, this is true for the graph of any function of two variables:

**Proposition 1.2.** Given $f : \mathbf{R}^2 \to \mathbf{R}$, define $F : \mathbf{R}^3 \to \mathbf{R}$ by $F(x, y, z) = f(x, y) - z$. Then the graph of $f$ is the level set of $F$ at height $0$:

$$\Gamma_f = F^{-1}(0).$$

The next example illustrates the fact that some level sets of functions of three variables are *not* graphs of functions of two variables.

**Example 1.7.** Define the function $g : \mathbf{R}^2 \to \mathbf{R}$ by

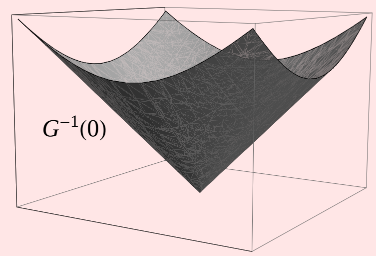
$$g(x, y) = \sqrt{x^2 + y^2}.$$

The graph of $g$ is the cone

$$\Gamma_g = \{(x, y, \sqrt{x^2 + y^2}) \mid (x, y) \in \mathbf{R}^2.\}$$

This cone is the level set at height $0$ for the function

$$G(x, y, z) = \sqrt{x^2 + y^2} - z.$$



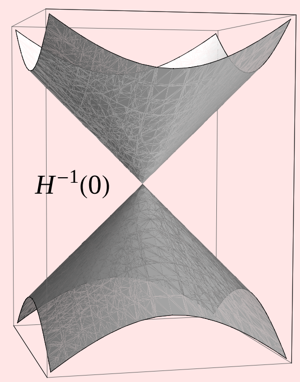

Now consider the function

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

The level set of $H$ at height $0$ is the double cone

$$H^{-1}(0) = \{(x, y, z) \mid x^2 + y^2 = z^2\}.$$

Although $\Gamma_g$ is a subset of $H^{-1}(0)$, the Vertical Line Test shows that $H^{-1}(0)$ is not the graph of any function of $x$ and $y$.



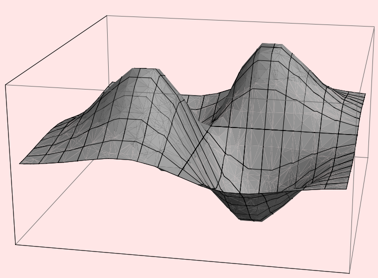

## 1.5 Contour maps

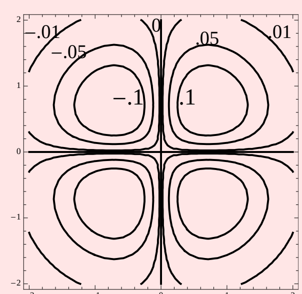
Given $f : \mathbf{R}^2 \to \mathbf{R}$, one may consider a collection of level sets drawn together in $\mathbf{R}^2$. We call such a picture a *contour map*, and it can be a helpful tool for understanding $\Gamma_f$.

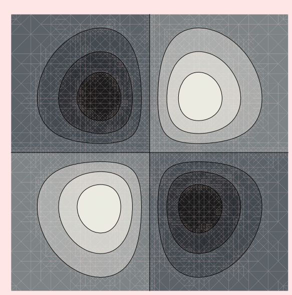
**Example 1.8.** Consider the function

$$f(x, y) = xye^{-x^2-y^2}.$$

There are several ways to visualize the graph of $f$:



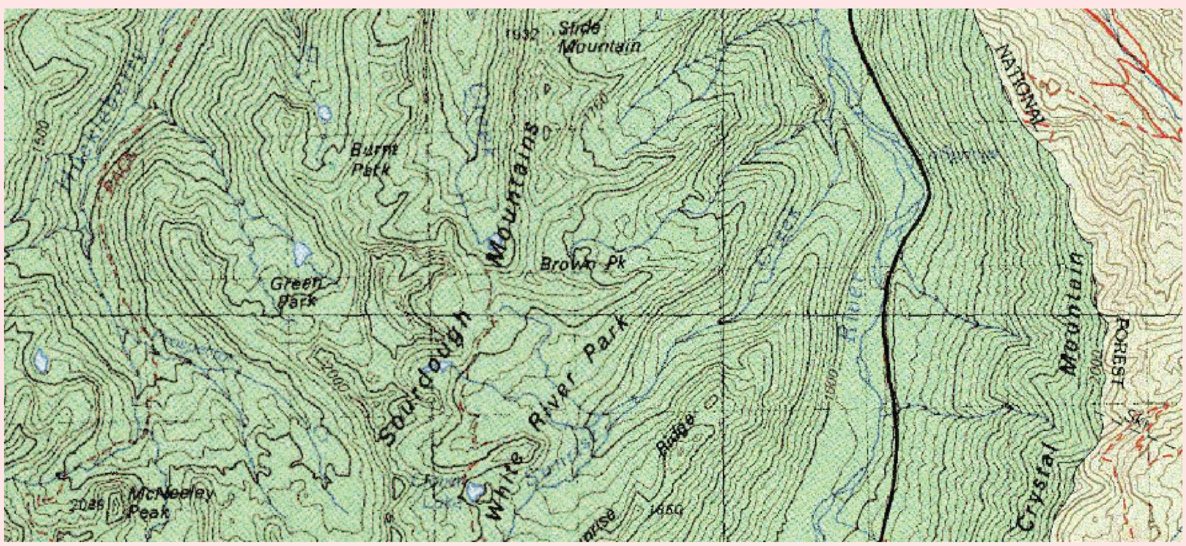


The first picture shows a computer-generated sketch of $\Gamma_f$. The center picture shows a contour map for $f$ where some of the contours are labeled by their heights. The third picture is also a contour map, but here the height is indicated by the shading, with darker tones indicating lower values of $c$.

In general, it's helpful to indicate the heights of the level sets on a contour map.

**Example 1.9.** A *topographical map* is a contour map for the function which associates height above sea level to each point on a map. The figure below shows part of a topographical map for Mount Ranier National Park. Each curve is a level set for the altitude function, so a hiker whose path follows one of these curves will stay at the same elevation.

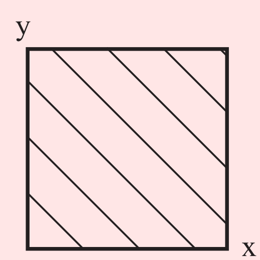



This map appears courtesy of the U. S. Geological Survey.

**Example 1.10.** *Draw a contour map for the function* $f(x, y) = \sin(x + y)$.

First, consider the level set of $f$ at height 0:

$$f^{-1}(0) = \{(x, y) \mid \sin(x + y) = 0\}.$$




This is the set of lines defined by

$$\{(x, y) \mid (x + y) = n\pi, n \in \mathbb{Z}\}.$$

In fact, the level set for any choice of $c \in [-1, 1]$ is a set of lines with slope $-1$.

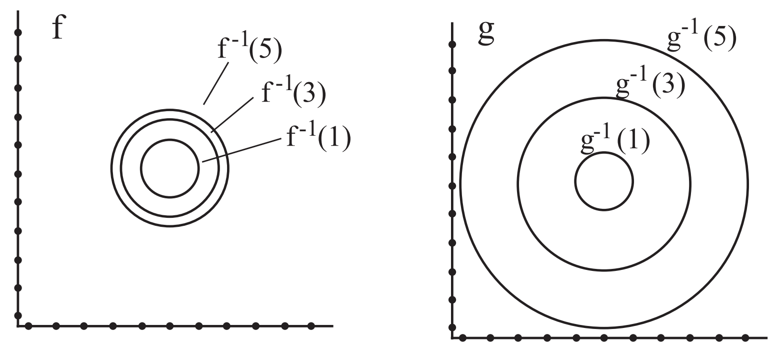
## 1.6 Worked Examples

**Example 1.11.** *Compare contour maps for the functions*

$$f(x, y) = x^2 + y^2 \quad \textit{and} \quad g(x, y) = \sqrt{x^2 + y^2}.$$

A contour map for $f$ or for $g$ consists of a collection of concentric circles in $\mathbf{R}^2$; two examples are shown below. Since every level set for $\Gamma_f$ is symmetric with respect to rotation around the origin, the surface $\Gamma_f$ is symmetric with respect to rotation around the $z$ axis. Similarly, $\Gamma_g$ is also symmetric with respect to rotation around the $z$ axis.

Although a contour map gives some information about the function, more information is required to completely describe the graph. For example, we can't distinguish $f$ and $g$ via their contour maps unless we add more information to the diagrams.




If we label each level sets on a contour map with its height, we get more information about the graph. In this example, labeling the level sets allows us to compare how quickly $f(x, y)$ and $g(x, y)$ increase as $||(x, y)||$ increases: $\Gamma_f$ is "steeper" than $\Gamma_g$.

**Example 1.12.** *Suppose that $T : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation. Show that the graph $\Gamma_T$ is a subspace of $\mathbf{R}^{m+n}$.*

By definition, the graph of $T$ is the set of points

$$\Gamma_T = \{\big((\mathbf{x}, \mathbf{f}(\mathbf{x})\big)\} \in \mathbf{R}^{m+n}.$$

In order to show that $\Gamma_T$ is a subspace, we must show that it satisfies the three subspace axioms:

1. $\mathbf{0} \in \Gamma_T$;
2. If $\mathbf{a}, \mathbf{b} \in \Gamma_T$, then $\mathbf{a} + \mathbf{b} \in \Gamma_T$;
3. If $\mathbf{a}$ and $c$ is a scalar, then $c\mathbf{a} \in \Gamma_T$.

We now show that each axiom holds:

1. Since $T$ is a linear transformation, $T(\mathbf{0}) = \mathbf{0}$. This implies that $\Gamma_T$ contains the point $(\mathbf{0}, \mathbf{T}(\mathbf{0})) = (\mathbf{0}, \mathbf{0})$, which is the zero vector in $\mathbf{R}^{m+n}$. Thus, the first axiom holds.
2. Suppose that $(\mathbf{a}, \mathbf{T}(\mathbf{a}))$ and $(\mathbf{b}, \mathbf{T}(\mathbf{b}))$ are two vectors in $\Gamma_T$.
$$\begin{aligned}(\mathbf{a}, \mathbf{T}(\mathbf{a})) + (\mathbf{b}, \mathbf{T}(\mathbf{b})) &= (\mathbf{a} + \mathbf{b}, \mathbf{T}(\mathbf{a}) + \mathbf{T}(\mathbf{b})) \\ &= (\mathbf{a} + \mathbf{b}, \mathbf{T}(\mathbf{a} + \mathbf{b}))\end{aligned}$$
Since the vector $(\mathbf{a} + \mathbf{b}, \mathbf{T}(\mathbf{a} + \mathbf{b}))$ is also in $\Gamma_T$, this proves that the second axiom holds.

3. Let $(\mathbf{a}, \mathbf{T}(\mathbf{a}))$ be a vector in $\Gamma_T$. Since $T$ is linear,

$$\mathbf{T}(c\mathbf{a}) = c\mathbf{T}(\mathbf{a}).$$

This implies that $c(\mathbf{a}, \mathbf{T}(\mathbf{a})) = (c\mathbf{a}, c\mathbf{T}(\mathbf{a})) = (c\mathbf{a}, \mathbf{T}(c\mathbf{a}))$. Since the vector $(c\mathbf{a}, \mathbf{T}(c\mathbf{a}))$ is in $\Gamma_T$, the third axiom holds as well.

This shows that the graph of a linear transformation is a subspace.

## 1.7 Exercises

1. For each of the following functions, describe the largest subset of Euclidean space which could be the domain of the function. Then describe the range of the function, assuming this domain.

   (a) $f(x, y) = \sqrt{xy}$

   (b) $g(x, y) = \sqrt{x^2 + y^2}$

   (c) $h(x, y) = \ln(x + 2y^2)$

   (d) $g(x, y, z) = e^{xyz}$

   (e) $f(t) = t \cos t$

   (f) $\mathbf{h}(x, y) = (x^2 + y^2, x + y)$

   (g) $g(x, y) = \frac{x}{y}$

2. Let $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$. Identify each of the following statements as always true or sometimes false.

   (a) $\Gamma_f$ is a subset of the codomain of $f$.

   (b) The range of $f$ is a subset of the codomain of $f$.

   (c) The graph of $f$ is a subset of the domain of $f$.

   (d) The domain of $f$ is a subset of the range of $f$.

   (e) $\mathbf{R}^{n+m}$ is the graph of $f$.

3. Let $A = \{a_1, a_2, a_3\}$ and let $B = \{b_1, b_2, b_3\}$, where $A$ and $B$ are subsets of some Euclidean spaces. Could the following be the graph of a function from $A$ to $B$? Why or why not?

$$\{(a_1, b_3), (a_2, b_3), (a_1, b_2), (a_3, b_1)\}$$

4. The line test helps identify subsets of $\mathbf{R}^{n+1}$ which cannot be graphs of real-valued functions. Generalize Proposition 1.1 to give a necessary condition for a subset of $\mathbf{R}^{n+2}$ to be the graph of a function $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^2$.

5. Show that the set

$$T = \{(w, x, y, z) \in \mathbf{R}^4 \mid y = w \text{ and } x^2 = z^4\}$$

is not the graph of any function of $w$ and $x$.

6. Let $\Gamma_f$ be the graph of $f : \mathcal{D} \to \mathbf{R}$. Show that the set

$$S = \{(w, x, y, z) \in \mathbf{R}^4 \mid y = f(w) \text{ and } x = z\}$$

is the graph of some function $\mathbf{g} : \mathcal{D}^2 \to \mathbf{R}^2$.

7. Define the *$n$-sphere* $S^n$ as

$$S^n = \{\mathbf{x} \in \mathbf{R}^{n+1} \mid ||\mathbf{x}|| = 1\}.$$

Show that $S^n$ is not the graph of any function from $\mathbf{R}^n$ to $\mathbf{R}$.

8. A *monomial in $n$ variables* is a product of the form

$$cx_1^{a_1} x_2^{a_2} \cdots x_n^{a_n},$$

where $c$ is a scalar and $a_i$ is a non-negative integer for all $i$. A *polynomial function of $n$ variables* is a function which can be expressed as a finite sum of monomials in $n$ variables.

   (a) When is a polynomial function of one variable a linear transformation?

   (b) Generalize your first answer to describe the multivariable polynomials which are also linear transformations.

9. Let $M$ be a $p \times q$ matrix, and let $T$ be the linear transformation defined by

$$T(\mathbf{x}) = M\mathbf{x}.$$

Identify each of the following statements as always true or sometimes false.

   (a) $T$ has $q$ coordinate functions.

   (b) All the coordinate functions have the same range.

   (c) All the coordinate functions have the same domain.

   (d) The column space of $M$ is the codomain of $T$.

   (e) The kernel of each coordinate function lies in the null space of $M$.

10. A function $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ is *even* if

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(-\mathbf{x}).$$

Show that a function is even if and only if all its coordinate functions are even.

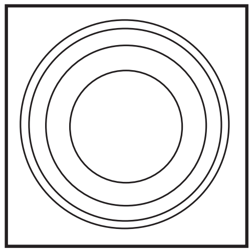
11. A function $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ is *odd* if

$$\mathbf{f}(-\mathbf{x}) = -\mathbf{f}(\mathbf{x}).$$

Show that a function is odd if and only if all its coordinate functions are odd.

12. Is every plane in $\mathbf{R}^3$ the graph of some function of $x$ and $y$? Explain your answer.

13. Which of the functions to the right could have the contour map shown below?



(a) $f(x, y) = (x + 1)^2 + (y - 1)^2$
(b) $f(x, y) = (x + y)^2 - (x - y)^2$
(c) $f(x, y) = (x + y)^2 + (x - y)^2$
(d) $f(x, y) = (x - y)^2 + (y - x)^2$
(e) $f(x, y) = (x + y)^2 + (y - x)^2$
(f) $f(x, y) = \cos(x^2 + y^2)$
(g) $f(x, y) = 7e^{2x^2} e^{2y^2}$

In Exercises 14-23, draw a contour map for $f$ and label the level sets in your map with their heights.

14. $f(x, y) = 4x^2 + y^2 + 5$

15. $f(x, y) = 9 - x^2 - y^2$

16. $f(x, y) = \sin(x - 3y)$

17. $f(x, y) = \frac{1}{xy}$

18. $f(x, y) = \frac{x}{y}$

19. $f(x, y) = (xy)^3$

20. $f(x, y) = e^{x-y}$

21. $f(x, y) = 4x^2 + 12xy + 9y^2$

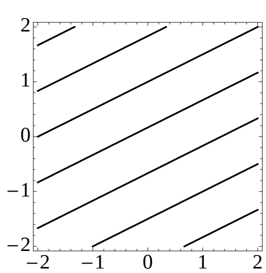
22. $f(x, y) = x - 3y^2 - 7$

23. $f(x, y) = 4(x - y)^2 + (x + y)^2$

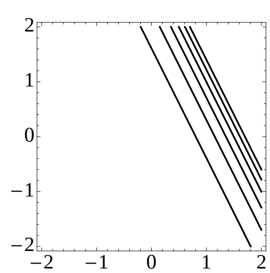
24. Match the following functions with their contour maps:

(a) $f(x, y) = e^{2x+y}$
(b) $f(x, y) = \sin(x^2 + 2y^2)$
(c) $f(x, y) = 3x - 6y - 4$
(d) $f(x, y) = x(x + y)$
(e) $f(x, y) = 3(x + y)^2 + (x - y)^2$
(f) $f(x, y) = 4x^2 - y^2$

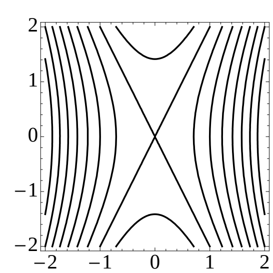
1.



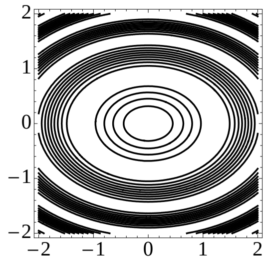

2.



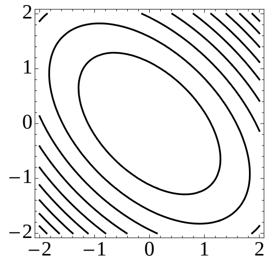

3.



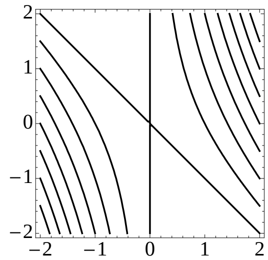
4.



5.



6.



25. Match the following functions with their contour maps:

(a) $f(x, y) = \sin x + \sin y$

(b) $f(x, y) = \sin x + y$

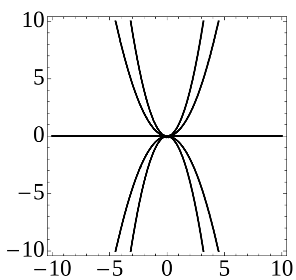
(c) $f(x, y) = \frac{y}{x^2}$

(d) $f(x, y) = x^3 + y^3$

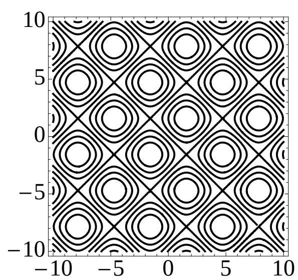
(e) $f(x, y) = y \sin x$

(f) $f(x, y) = \frac{y^2}{x}$

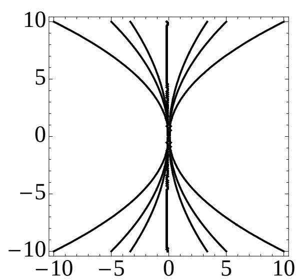
1.



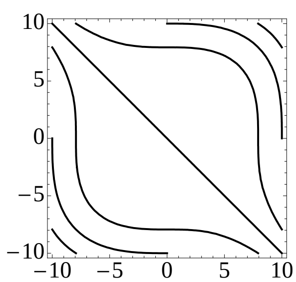
3.



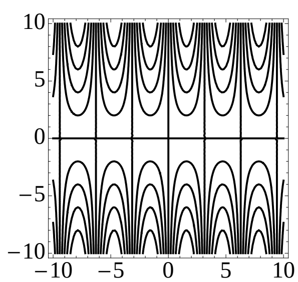
5.



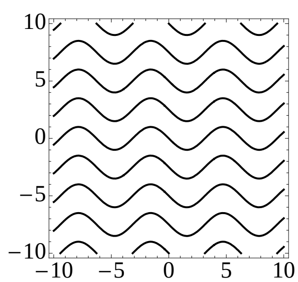
2.



4.



6.



26. Match the following functions with their contour maps:

(a) $f(x, y) = \cos(|x + y|)$

(b) $f(x, y) = \cos(|x|) + \cos(|y|)$

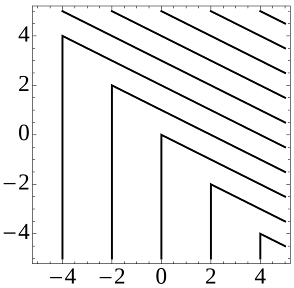
(c) $f(x, y) = |x + y| + y$

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

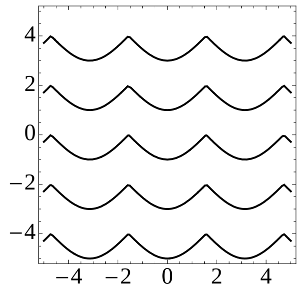
(e) $f(x, y) = |x - y| + |y - x|$

(f) $f(x, y) = |\cos x| + y$

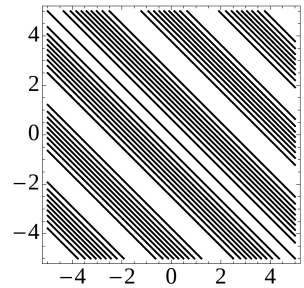
1.


4
2
0
−2
−4
−4 −2 0 2 4

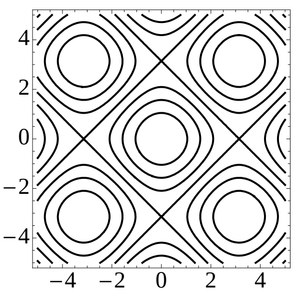
2.


4
2
0
−2
−4
−4 −2 0 2 4

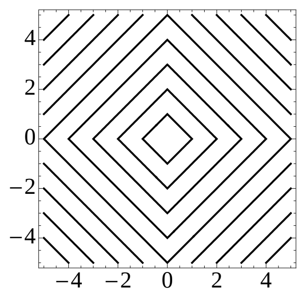
3.


4
2
0
−2
−4
−4 −2 0 2 4

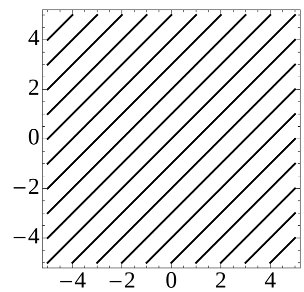
4.


4
2
0
−2
−4
−4 −2 0 2 4

5.


4
2
0
−2
−4
−4 −2 0 2 4

6.


4
2
0
−2
−4
−4 −2 0 2 4

# 2 Review: Limits and continuity for single variable functions

This section provides a review of some of the most important concepts from single variable calculus. The treatment of limits and continuity here is fairly formal, in order to emphasize the parallels when we explore the same topics for multivariable functions in Section 5.

## 2.1 Limits

**Definition 2.1.** Let $f : \mathcal{D}^1 \to \mathbf{R}$ have the property that every punctured neighborhood of the point $a$ contains some point of $\mathcal{D}^1$.
Then for $L \in \mathbf{R}$,

$$\lim_{x \to a} f(x) = L$$

if for every $\epsilon > 0$, there exists some $\delta > 0$ such that

$$\text{if } x \in \mathcal{D}^1 \text{ and } 0 < |x - a| < \delta, \text{ then } |f(x) - L| < \epsilon.$$

In this case, we say that $L$ *is the limit of* $f$ *as* $x$ *approaches* $a$.

Writing

$$\lim_{x \to a} f(x) = L$$

asserts both that the limit exists and that it equals $L$. In words, this means that $f(x)$ can be made *arbitrarily* close to $L$ be ensuring that $x$ is *sufficiently* close to $a$.

**Example 2.1.** *Find the limit*

$$\lim_{x \to 1} \frac{x^2 - x}{x - 1}.$$

Notice that away from the point $x = 1$, we have equality between the given function and the function $x$:

$$\frac{(x)(x-1)}{x-1} = x \text{ for } x \neq 1.$$

Thus,

$$\lim_{x \to 1} \frac{(x)(x-1)}{x-1} = \lim_{x \to 1} x = 1,$$

since the limit as $x$ approaches $1$ describes the value of the function near $x$, but not at $x$.
The last equality may look obvious, but you should check that it satisfies Definition 2.1; for any $\epsilon > 0$, set $\delta = \epsilon$.

**Example 2.2.** *Let $f(x) = x^2 + 1$. Prove that*

$$\lim_{x \to 1} f(x) = 2$$

*by showing that for every $\epsilon > 0$, if $\delta = \min(\frac{\epsilon}{4}, 2)$ then*

$$|x - 1| < \delta \;\; \textit{implies} \;\; |f(x) - 2| < \epsilon.$$

To see that is true, suppose $0 < |x - 1| < \delta$. Then

$$\begin{aligned} |f(x) - 2| &= |x^2 + 1 - 2| \\ &= |x^2 - 1| \\ &= |(x-1)^2 + 2(x-1)| \\ &\leq |(x-1)^2| + 2|x - 1| \\ &< \delta^2 + 2\delta \\ &\leq 4\delta. \end{aligned}$$

Since $4\delta \leq \epsilon$, this proves that

$$|x - 1| < \delta \;\; \text{implies} \;\; |f(x) - 2| < \epsilon.$$

Although this proof is correct, the choice of $\delta = \min(\frac{\epsilon}{4}, 2)$ may seem somewhat mysterious. The Worked Examples section discusses how one might come up with this value.

**Theorem 1** (Limit Laws)**.** Suppose that $f(x)$ and $g(x)$ are functions whose limits exist as $x$ approaches $a$:

$$\lim_{x \to a} f(x) = L_f \;\; \text{and} \;\; \lim_{x \to a} g(x) = L_g.$$

Then the following statements hold:

$$\lim_{x \to a} (f + g)(x) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x) = L_f + L_g \tag{2.1}$$

$$\lim_{x \to a} (fg)(x) = \lim_{x \to a} f(x) \lim_{x \to a} g(x) = L_f L_g \tag{2.2}$$

$$\lim_{x \to a} cf(x) = c \lim_{x \to a} f(x) = cL_f \text{ for all scalars } c. \tag{2.3}$$

Note that the limit laws provide an easier proof of the limit in Example 2.2, because we can decompose the function $x^2 + 1$ into $(x)(x) + (1)$. (See Exercise 39.)

*Proof of Theorem 1.* We prove the limit law in Equation 2.1, and the others are left as exercises.

Fix $\epsilon > 0$. By hypothesis, there is some $\delta_f > 0$ with the property that

$$|x - a| < \delta_f \quad \text{implies} \quad |f(x) - L_f| < \frac{\epsilon}{2}.$$

Similarly, there is some $\delta_g > 0$ with the property that

$$|x - a| < \delta_g \quad \text{implies} \quad |g(x) - L_g| < \frac{\epsilon}{2}.$$

Let $\delta = \min(\delta_f, \delta_g)$. If $|x - a| < \delta$, then we have the following:

$$\begin{aligned} |(f+g)(x) - (L_f + L_g)| &= |f(x) - L_f + g(x) - L_g| \\ &\leq |f(x) - L_f| + |g(x) - L_g| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon. \end{aligned}$$

This proves that $\lim_{x \to a}(f+g)(x) = L_f + L_g$. □

**Theorem 2.** Suppose there exists some $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x)| < M$, and suppose that

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

Then

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

The proof of Theorem 2 is Exercise 41.

## 2.2 Limits from above and below

Given a function $f : \mathbf{R} \to \mathbf{R}$, we can define a special kind of limit that depends on how $x$ approaches the point $a$:

**Definition 2.2.** If for every $\epsilon > 0$, there exists some $\delta > 0$ such that

$$0 < x - a < \delta \quad \text{implies} \quad |f(x) - L| < \epsilon,$$

then we write

$$\lim_{x \to a^+} f(x) = L.$$

In this case we say that $L$ is the *limit of $f$ as $x$ approaches $a$ from above*.

**Definition 2.3.** If for every $\epsilon > 0$, there exists some $\delta > 0$ such that

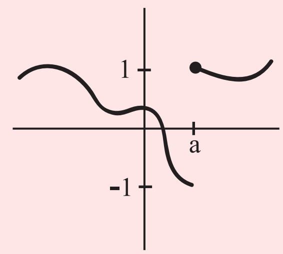
$$0 < a - x < \delta \quad \text{implies} \quad |f(x) - L| < \epsilon,$$

then we write

$$\lim_{x \to a^-} f(x) = L$$

In this case we say that $L$ is the *limit of $f$ as $x$ approaches $a$ from below*.

**Example 2.3.** The figure on the right shows the graph of a function $f$ with the property that the limits as $x$ approaches $a$ from above and below both exist, but are not equal:




$$\lim_{x \to a^-} f(x) = -1 \quad \text{and} \quad \lim_{x \to a^+} f(x) = 1.$$

**Theorem 3.**

$$\lim_{x \to a} f(x) = L \quad \text{if and only if} \quad \lim_{x \to a^-} f(x) = L = \lim_{x \to a^+} f(x).$$

The proof of this theorem is Exercise 40.

## 2.3 Limits and infinity

It's sometimes useful to identify a special way in which a limit can fail to exist.

**Definition 2.4.** Given the function $f : \mathcal{D}^1 \to \mathbf{R}$, we write

$$\lim_{x \to a} f(x) = \infty$$

if for every $N \in \mathbf{R}$, there exists a $\delta > 0$ such that $0 < |x - a| < \delta$ implies $f(x) > N$. Similarly, we write

$$\lim_{x \to a} f(x) = -\infty$$

if for every $N \in \mathbf{R}$, there exists a $\delta > 0$ such that $0 < |x - a| < \delta$ implies $f(x) < N$.

This notation is helpful because it provides more information about what happens to $f(x)$ near $a$. However, it should *not* be confused with saying that the limit exists in the sense of Definition 2.1.

**Proposition 2.1.**

$$\text{If} \quad \lim_{x\to a} f(x) = \infty, \quad \text{then} \quad \lim_{x\to a} \frac{1}{f(x)} = 0.$$

## 2.4 Continuity

**Definition 2.5.** The function $f : \mathcal{D}^1 \to \mathbf{R}$ is *continuous at $a$* if

$$\lim_{x\to a} f(x) = f(a).$$

A function is *continuous* if it is continuous at all points in its domain.

Note that this definition requires three things to be true:

- the point $a$ is in the domain of $f$;
- the limit of $f$ as $x$ approaches $a$ exists;
- the limit of $f$ as $x$ approaches $a$ is equal to $f(a)$.

**Example 2.4.** *Let $f : \mathbf{R} \to \mathbf{R}$ be the function defined piecewise by*

$$f(x) = \begin{cases} \frac{x^2-1}{x+1} & \text{if } x \neq -1 \\ c & \text{if } x = -1. \end{cases}$$

*Is there any value of $c$ for which $f$ is continuous?*
We first note that away from $x = -1$,

$$f(x) = \frac{(x-1)(x+1)}{x+1} = x - 1.$$

Setting $\delta = \epsilon$ shows that the function $g(x) = x - 1$ is continuous at every point. (Try this!)
Since

$$\lim_{x\to -1} f(x) = \lim_{x\to -1} g(x) = -2,$$

setting $c = -2$ makes $f$ a continuous function.

**Proposition 2.2.** Suppose that $f$ and $g$ are continuous at $a$, and let $c \in \mathbf{R}$. Then the following functions are continuous at $a$:

- $f + g$
- $fg$
- $cf$.

The proof of this proposition follows from the limit laws and is left as an exercise. Continuous functions also behave well with respect to composition.

**Theorem 4.** Suppose that $f : \mathcal{D}^1 \to \mathbf{R}$ is continuous at the point $a$, and suppose further that $g : \mathcal{E}^1 \to \mathbf{R}$ is continuous at $f(a)$. Then the composition $(g \circ f)(x)$ is continuous at $a$.

This theorem is proved in the Worked Examples section.

It is useful to have a library of functions which are known to be continuous. Except when explicitly asked to prove this, you may assume that each of the following functions is continuous on its entire domain:

- Polynomials
- Ratios of continuous functions
- Exponential functions: $a^x$ for $a \in \mathbf{R}$
- Trigonometric functions: $\sin(x), \cos(x)$.

## 2.5 The Squeeze Theorem

We end the section with a theorem that can be helpful for computing limits.

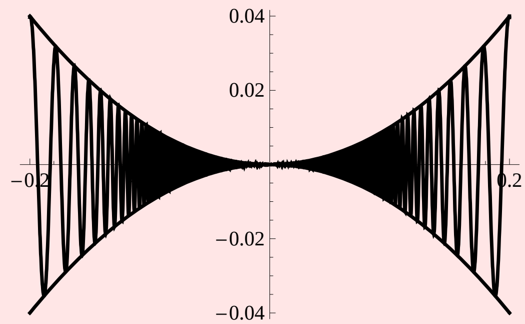
**Theorem 5** (The Squeeze Theorem)**.** Suppose that there is some open interval containing $a$ on which $f(x) \leq g(x) \leq h(x)$. Suppose further that

$$\lim_{x \to a} f(x) = \lim_{x \to a} h(x).$$

Then $\lim_{x \to a} g(x)$ exists and

$$\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = \lim_{x \to a} h(x).$$

**Example 2.5.** *Let* $g(x) = x^2 \cos(\frac{1}{x})$. *Show that* $\lim_{x\to 0} g(x) = 0$.




First, observe that $-1 \leq \cos(\frac{1}{x}) \leq 1$ for all values of $x$. Thus

$$-x^2 \leq x^2 \cos(\frac{1}{x}) \leq x^2.$$

Furthermore,

$$\lim_{x\to 0} x^2 = 0 = \lim_{x\to 0} -x^2.$$

(This follows from the fact that $x^2$ is a polynomial, and therefore continuous, or you can prove the limit directly without too much work.)

Applying the Squeeze Theorem proves the desired limit.

*Proof of the Squeeze Theorem.* Suppose that $f(x)$, $g(x)$, and $h(x)$ satisfy the hypotheses of the Squeeze Theorem, and let

$$L = \lim_{x\to a} f(x) = \lim_{x\to a} h(x).$$

Fix some $\epsilon > 0$.

By hypothesis, there exists some $\delta_f > 0$ such that $0 < |x-a| < \delta_f$ implies $|f(x)-L| < \epsilon$.
Similarly, there exists some $\delta_h > 0$ such that $0 < |x-a| < \delta_h$ implies $|h(x) - L| < \epsilon$.
Let $\delta = \min(\delta_f, \delta_h)$.
For $0 < |x-a| < \delta$, we have

$$L - \epsilon < f(x) \leq g(x) \leq h(x) < L + \epsilon.$$

This shows that $0 < |x-a| < \delta$ implies $|g(x) - L| < \epsilon$, as desired.

□

## 2.6 Worked Examples

**Example 2.6.** We begin by revisiting the function $f(x) = x^2 + 1$ from Example 2.2. We were asked to prove

$$\lim_{x\to 1} f(x) = 2.$$

The proof presented in Example 2.2 began by fixing $\epsilon > 0$ and then showing

$$|x-1| < \frac{\epsilon}{4} \text{ implies } |f(x) - 2| < \epsilon.$$

In order to motivate the equation $\delta = \frac{\epsilon}{4}$, recall that the proof relies on bounding $|f(x) - 2|$ from above.

Our only tool for controlling this quantity is to require that $|x-1|$ be very small, so we try to rewrite $|f(x)-2|$ in terms of $|x-1|$:

$$\begin{aligned}|f(x)-2| &= |x^2+1-2| \\ &= |x^2-1| \\ &= |(x-1)^2+2(x-1)| \\ &\leq |(x-1)|^2+2|x-1|.\end{aligned}$$

Thus, if we can find a $\delta$ such that

$$\delta^2+2\delta<\epsilon, \tag{2.4}$$

we will have the desired result that

$$|x-1|<\delta \text{ implies } |f(x)-2|<\epsilon.$$

Although $\delta=\frac{\epsilon}{4}$ satisfies Equation 2.6 when $\delta$ is small enough, we could equally well take the smaller value $\delta=\frac{\epsilon}{100}$, or use the quadratic formula to set $\delta<-1+\sqrt{1+\epsilon}$. The choice of $\delta$ is not unique, but it is often easier to follow a proof where the releationship between $\delta$ and $\epsilon$ is clear.

**Remark**: Equations involving absolute value can be split into cases and solved, but it's often helpful to think of "$0<|x-c|<\delta$" as a statement about distance in the real number line. That is, $x$ is a solution to this inequality if the distance between the points $x$ and $c$ is less than $\delta$ but greater than $0$. Writing $\lim_{x\to c} f(x)=L$ means that $f(x)$ can be made arbitrarily close to $L$ by making $x$ sufficiently close to $c$.

**Example 2.7.** *Show the limit does not exist:*

$$\lim_{x\to 2}\frac{x+2}{x-2}.$$

As $x\to 2$, the denominator approaches zero. Since the numerators is bounded away from $0$ as $x$ approaches $2$, we would expect the value of the ratio to tend towards infinity. Thus, we might be inclined to guess

$$\lim_{x\to 2}\frac{x+2}{x-2}=\infty.$$

However, Theorem 3 shows us otherwise.

If $x<2$, then $x-2<0$. Thus

$$\lim_{x\to 2^-}\frac{x+2}{x-2}=-\infty.$$

If $x>2$, then $x-2>0$. Thus

$$\lim_{x\to 2^+}\frac{x+2}{x-2}=\infty.$$

This proves that the limit does not exist.

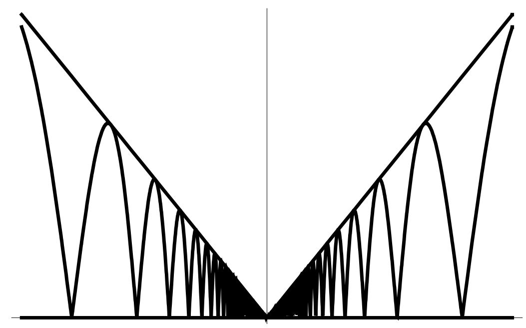
**Example 2.8.** *Evaluate the limit*
$$\lim_{x\to 0} x\sin(\frac{1}{x}).$$
If we plug in a sequence of points approaching $0$ or study the graph, we might guess that the limit equals $0$. We'll use the Squeeze Theorem to prove that this is correct.

Instead of applying the Squeeze Theorem directly, we first note that for any function $f(x)$,
$$\lim_{x\to 0} f(x) = 0 \text{ if and only if } \lim_{x\to 0} |f(x)| = 0.$$

The advantage of the this observation is that if we substitute $|x\sin(\frac{1}{x})|$ for $x\sin(\frac{1}{x})$, we get one of our squeezing functions "for free":
$$0 \le |x\sin(\frac{1}{x})|.$$

To get an upper bound for $|x\sin(\frac{1}{x})|$, we use the inequality
$$|\sin(\frac{1}{x})| \le 1 \quad \text{for all } x \neq 0.$$

Thus, in a punctured neighborhood of $0$, we have
$$0 \le |x\sin(\frac{1}{x})| \le |x|.$$



We will state, but not prove, that the limit of the constant function $c$, as $x$ approaches any value, is $c$. (Try this!) Finally, we must show that $\lim_{x\to 0} |x| = 0$.

Fix $\epsilon > 0$. We need to find $\delta > 0$ such that
$$0 < |x - 0| < \delta \quad \text{implies} \quad ||x| - 0| < \epsilon.$$
However, $||x| - 0| = |x|$, so setting $\delta = \epsilon$ proves that the limit of $|x|$ as $x \to 0$ is $0$.

We have shown that
$$0 \le |x\sin(\frac{1}{x})| \le |x|$$
and
$$\lim_{x\to 0} 0 = 0 = \lim_{x\to 0} |x|,$$
so the Squeeze Theorem proves the desired limit.

**Example 2.9** (Proof of Theorem 4)**.** As in the statement of Theorem 4, suppose that $f$ is continuous at $a \in \mathbf{R}$ and that $g$ is continuous at $f(a)$. In order to show that $g \circ f$ is continuous at $a$, we must show that
$$\lim_{x\to a} (g \circ f)(x) = (g \circ f)(a).$$

First, we show that $a$ is in the domain of the composition. By hypothesis, $a$ is in the domain of $f$ and $f(a)$ is in the domain of $g$. Since $(g \circ f)(a) = g(f(a))$, $a$ is in the domain of $g \circ f$.

Next, we show that

$$\lim_{x \to a} (g \circ f)(x)$$

exists. In order to apply Definition 2.1, we need a candidate for the limit, but this is supplied by the statement we are trying to prove: $g(f(a))$.

Fix $\epsilon > 0$.

We would like to find a $\delta > 0$ with the property that

$$0 < |x - a| < \delta \quad \text{implies} \quad |g(f(x)) - g(f(a))| < \epsilon.$$

Since $g$ is continuous, we know that making $|f(x) - f(a)|$ small enough ensures that $|g(f(x)) - g(f(a))| < \epsilon$. On the other hand, since $f$ is continuous, making $|x - a|$ small enough will ensure that $|f(x) - f(a)|$ is small. In order to complete the proof, we make this chain of reasoning more precise.

By hypothesis,

$$\lim_{y \to f(a)} g(y)$$

exists, so there is some $\delta_g > 0$ such that

$$0 < |y - f(a)| < \delta_g \quad \text{implies} \quad |g(y) - g(f(a)| < \epsilon.$$

Since $g$ is continuous at $f(a)$, we actually get the stronger statement

$$|y - f(a)| < \delta_g \quad \text{implies} \quad |g(y) - g(f(a))| < \epsilon. \tag{2.5}$$

Since $f$ is continuous at $a$, there exists some $\delta_f$ such that

$$0 < |x - a| < \delta_f \quad \text{implies} \quad |f(x) - f(a)| < \delta_g. \tag{2.6}$$

Putting Equations 2.5 and 2.6 together, we see that

$$0 < |x - a| < \delta_f \quad \text{implies} \quad |g(f(x)) - g(f(a))| < \epsilon.$$

This proves Theorem 4.

## 2.7 Exercises

In Exercises 1 through 12 , use a $\delta/\epsilon$ or $\delta/N$ argument to prove the given limit.

1. $$\lim_{x \to 0} 2x + 3 = 3$$

2. $$\lim_{x \to 0} 5 + 3x = 5$$

3. $$\lim_{x \to 0} x^2 + 2 = 2$$

4. $$\lim_{x \to 0} \sqrt{x} = 0$$

5. $$\lim_{x \to 2} 2x - 1 = 3$$

6. $$\lim_{x \to 0} 2x^3 + x^2 + 1 = 1$$

7. $$\lim_{x\to 3} 2x^2 - 3x = 9$$

8. $$\lim_{x\to 1} x^2 + x = 2$$

9. $$\lim_{x\to -1} 3x + x^2 = -2$$

10. $$\lim_{x\to 0} \frac{-1}{x^4 + 5x^2} = -\infty$$

11. $$\lim_{x\to 1} x^3 + 2x^2 = 3$$

12. $$\lim_{x\to 2} \frac{x}{x^3 - 4x^2 + 4x} = \infty$$

In Exercises 13 through 33, determine whether the limit exists. If it does, find it.

13. $$\lim_{x\to -1} \frac{x+1}{x^2+8x+7}$$

14. $$\lim_{x\to -2} \frac{x^2-4x-12}{x+2}$$

15. $$\lim_{x\to 3} \frac{x^2-4x+3}{x^2-5x+6}$$

16. $$\lim_{x\to 2} \frac{x^2-4}{\sqrt{x-2}}$$

17. $$\lim_{x\to -6} \frac{x^2+3x-18}{\sqrt{x+6}}$$

18. $$\lim_{x\to 1} \frac{x^2+x+1}{x^2-5x-14}$$

19. $$\lim_{x\to 0} \frac{|x|}{x}$$

20. $$\lim_{x\to 0} \frac{x^2}{|x|}$$

21. $$\lim_{x\to 0} \frac{(x^3-x^2-6x)\cos x}{x^2-2x}$$

22. $$\lim_{x\to 2\pi} \frac{(x-2\pi)\sin x}{x^2-4\pi^2}$$

23. $$\lim_{x\to 0} \left(\frac{x^2}{x-2} + \frac{1}{x}\right)$$

24. $$\lim_{x\to -2} \frac{x^2-3x+7}{x^2-2x-8}$$

25. $$\lim_{x\to 2} \frac{x^3-8}{x-2}$$

26. $$\lim_{x\to 2} \frac{-x^2-x+6}{x^4-13x^2+36}$$

27. $$\lim_{x\to 0} \frac{\cos x}{x}$$

28. $$\lim_{x\to 1} \frac{x^3-3x^2+3x-1}{x^2-2x+1}$$

29. $$\lim_{x\to 3} \frac{2x^2-7x+3}{x^2-6x+9}$$

30. $$\lim_{x\to 0} x^2 e^{\sin(1/x)}$$

In Exercises 31 and 32, determine if the given function is continuous.

31.
$$g(x) = \begin{cases} \frac{x^2+x-12}{x-3} & \text{if } x \neq 3 \\ 7 & \text{if } x = 3 \end{cases}$$

32.
$$f(x) = \begin{cases} \frac{x^2+x-2}{e^{-x}(x-1)} & \text{if } x \neq 1 \\ 3 & \text{if } x = 1 \end{cases}$$

In Exercises 33 and 34, determine if there is a value for $a$ which makes the given function continuous.

33.
$$g(x) = \begin{cases} \frac{x-3}{x^2+x-12} & \text{if } x \neq 3, -4 \\ \frac{1}{7} & \text{if } x = 3 \\ a & \text{if } x = -4 \end{cases}$$

34.
$$h(x) = \begin{cases} (x^2 - 4x)\cos(\ln|x|) & \text{if } x \neq 0 \\ a & \text{if } x = 0 \end{cases}$$

In Exercises 35 and 36, prove that the given function is continuous at the indicated point.

35. $f(x) = 17x - 4$, $a = 12$

36. $f(x) = x^2 + 4x - 5$, $a = -3$

37. A *polynomial of degree n* is a function of the form
$$p(x) = \sum_{i=0}^{n} c_i x^i,$$
where each $c_i$ is a scalar.
   (a) Show that $f(x) = x$ is a continuous function.
   (b) Show that every constant function is continuous.
   (c) Use the limit laws to prove that an arbitrary polynomial is continuous.

38. Define
$$f(x) = \begin{cases} x^2 & \text{if } x \text{ is a rational number} \\ 0 & \text{if } x \text{ is an irrational number.} \end{cases}$$
Show $f$ is continuous at $0$.

39. Suppose that the limits of $f$, $g$, and $fg$ all exist as $x$ approaches $a$. Show that the following statements are true:
   (a)
$$\lim_{x\to a}(fg)(x) = \lim_{x\to a} f(x) \lim_{x\to a} g(x)$$
   (b)
$$\lim_{x\to a} cf(x) = c \lim_{x\to a} f(x)$$

40. Prove Theorem 3.

41. Prove Theorem 2.

# 3 Review: Derivatives of single variable functions

The derivative is a fundamental object of study in calculus. Here we recall the basic definitions and applications of the derivative of a single variable function.

## 3.1 The derivative of a single variable function

**Definition 3.1.** If $f : \mathcal{D}^1 \to \mathbf{R}$ is defined on an interval containing the point $a$, then the *derivative* of $f$ at $a$ is

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

provided this limit exists.

The substitution $x = a + h$ yields the following equivalent definition:

**Definition 3.2.** If $f : \mathcal{D}^1 \to \mathbf{R}$ is defined on an interval containing the point $a$, then the *derivative* of $f$ at $a$ is

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},$$

provided this limit exists.

If the derivative of $f$ at $a$ exists, we say that $f$ is *differentiable* at $a$.

Given $f : \mathcal{D}^1 \to \mathbf{R}$, we can define a new function $f' : \mathcal{E}^1 \to \mathbf{R}$, which we call the *derivative of* $f$. The domain $\mathcal{E}^1$ is the set of points in $\mathcal{D}^1$ where $f$ is differentiable, and $f'(x)$ is defined pointwise using Definition 3.1. Note that the set $\mathcal{D}^1$ may be strictly larger than the set $\mathcal{E}^1$, since a function may be defined at a point where it fails to be differentiable.

**Example 3.1.** *Find the derivative of $f(x) = x^2$.*
We compute the limit given in Definition 3.1 for an arbitrary point $a$:

$$\begin{aligned} f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\ &= \lim_{h \to 0} \frac{(a+h)^2 - a^2}{h} \\ &= \lim_{h \to 0} \frac{a^2 + 2ha + h^2 - a^2}{h} \\ &= \lim_{h \to 0} 2a + h \\ &= 2a. \end{aligned}$$

Since this limit exists for all values of $a \in \mathbf{R}$, the derivative of $f$ is given by $f'(x) = 2x$.

A variety of notation is used to denote the derivative of the function $f$ at the point $a$:

$$f'(a) = \frac{df}{dx}(a) = Df(a).$$

Similarly, the derivative function of $f$ can be denoted by any of the following:

$$f', \ \frac{df}{dx}, \ \frac{d}{dx}f, \ \text{ or } \ Df.$$

### 3.1.1 Secants and tangents

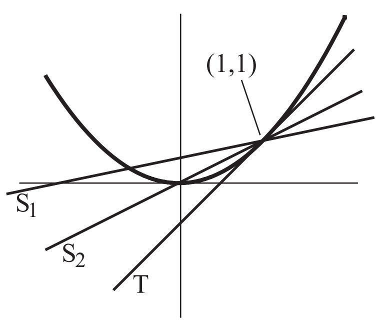
Suppose that $f : \mathcal{D}^1 \to \mathbf{R}$ is differentiable at the point $a$. For each fixed value of $x$, the ratio

$$\frac{f(x) - f(a)}{x - a}$$

is the slope of the line in $\mathbf{R}^2$ through the points $\big(x, f(x)\big)$ and $\big(a, f(a)\big)$, both of which lie on $\Gamma_f$. A line with this form is called a *secant line* for $\Gamma_f$. As the point $x$ approaches $a$, the slope of the secant line approaches $f'(a)$. The line

$$y = f'(a)(x - a) + f(a)$$

is called the *tangent line* to $\Gamma_f$ at $(a, f(a))$. The slope of the secant line is the *average* rate of change of $y$ over the interval between $x$ and $a$. Since the length of this interval shrinks to $0$ as $x$ approaches $a$, we sometimes say that $f'(a)$ is the *instantaneous rate of change* of $y$ with respect to $x$ at $a$.




The graph of $f(x) = x^2$ is shown on the left, together with the tangent line to $\Gamma_f$ at the point $(1, 1)$, which is labeled $T$. The figure also shows two secant lines, labeled $S_1$ and $S_2$, which correspond to the values $x = \frac{-1}{2}$ and $x = 0$, respectively. Note that as $x \to 1$, the secant lines approach the tangent line.

### 3.1.2 Higher derivatives

When $f'$ is itself a differentiable function, the derivative of $f'$ is called the *second derivative of $f$*. It is denoted by

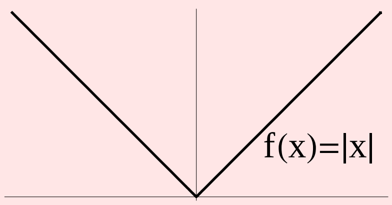
$$f'' = \frac{d}{dx}f' = \frac{d^2 f}{dx^2}.$$

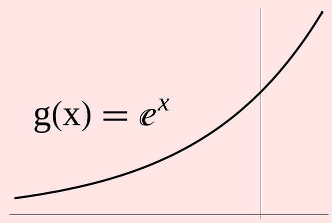
Similarly, the derivative of the second derivative is called the third derivative. Repeating the differentiation process $n$ times yields the $n^{th}$ derivative of the original function $f$, which we denote by $f^{(n)}$:

$$f^{(n)}(x) = \frac{d}{dx} f^{(n-1)}(x).$$

If $f$ is a function which can be differentiated infinitely-many times, we say that $f$ is *smooth*.

**Example 3.2.** *Consider the graphs of the functions $f(x) = |x|$ and $g(x) = e^x$.*




The figure shows that $\Gamma_f$ has a sharp point at the origin, and $f$ is not differentiable at $x = 0$. (See Exercise 7.) On the other hand, $g$ is differentiable at every point in $\mathbf{R}$, and $\Gamma_g$ has no corners or points. In fact, the function $g$ is smooth.

## 3.2 L'Hôpital's Rule

We end this section by recalling l'Hôpital's Rule, which is a technique for computing limits of real-valued functions of one variable.

**Theorem 6** (l'Hôpital's Rule)**.** Suppose that $f$ and $g$ are differentiable at $a$. If

$$\lim_{x \to a} f(x) = 0 = \lim_{x \to a} g(x),$$

then

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

See any text on single variable calculus for a proof.

## 3.3 Worked Examples

**Example 3.3.** *Compute the limit*

$$\lim_{x \to 0} e^{\frac{\sin x}{e^x - 1}} .$$

The function is defined by a complicated formula, so let's start by just studying the exponent. As $x$ approaches $0$, we see that

$$\lim_{x\to 0}(e^x - 1) = 0 \quad \text{and} \quad \lim_{x\to 0}\sin x = 0.$$

The exponent $\frac{\sin x}{e^x-1}$ is the kind of ratio that shows up in Theorem 6, so l'Hôpital's Rule tells us that

$$\lim_{x\to 0}\frac{\sin x}{e^x-1} = \lim_{x\to 0}\frac{\frac{d}{dx}\sin x}{\frac{d}{dx}(e^x-1)} = \lim_{x\to 0}\frac{\cos x}{e^x} = 1.$$

The last equality follows because the two functions in the ratio are continuous at $0$.

In order to use this result to solve the original problem, observe that

$$\lim_{x\to 0}\ln(e^{\frac{\sin x}{e^x-1}}) = \lim_{x\to 0}\frac{\sin x}{e^x-1} = 1.$$

Since the exponent approaches $1$ as $x$ approaches $0$, we might guess that the original limit is $e^1$. In order to confirm this, we make use of the following fact:

$$\text{if } \lim_{x\to a}\ln(f(x)) = \ln L, \quad \text{then} \quad \lim_{x\to a}f(x) = L.$$

(This is an application of Theorem 4.)

We've shown that

$$\lim_{x\to 0}\ln(e^{\frac{\sin x}{e^x-1}}) = 1 = \ln e,$$

so we see that

$$\lim_{x\to 0}e^{\frac{\sin x}{e^x-1}} = e.$$

## 3.4 Exercises

In Exercises 1 through 6, use Definition 3.1 to compute the derivative of the given function at the indicated point. Use your result to find an equation for the tangent line to the graph of $f$ at $a$. (You may use the fact that $\lim_{x\to 0}\frac{\sin x}{x} = 1$.)

1. $f(x) = x^3$, $a = 1$.
2. $g(x) = x^5$, $a = -1$.
3. $f(x) = \sin x$, $a = 0$.
4. $h(x) = \cos x$, $a = \frac{\pi}{2}$.
5. $g(x) = x^2 - 2x^3$, $a = 2$.
6. $h(x) = \tan x$, $a = 0$.

7. Show that $f(x) = |x|$ is not differentiable at $0$.

8. Use Definition 3.1 to show that

$$\frac{d}{dx}x^n = nx^{n-1}.$$

9. If every secant line to the graph of $f$ is the same, what does that tell you about the function $f$?

10. Use Definition 3.1 to prove the *product rule*:

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x)$$

11. Let $f : \mathbf{R} \to \mathbf{R}$ be defined piecewise as follows:

$$f(x) = \begin{cases} 2x+7 & \text{if } x \leq 2 \\ x^2 - 2x + 11 & \text{if } x > 2 \end{cases}$$

   (a) Show that $f$ is differentiable for all $x \in \mathbf{R}$ and compute the derivative.

   (b) Show that $f$ is not twice differentiable at $x = 2$.

In Exercises 12 through 21, use l'Hôpital's Rule to determine if the limit exists. If it does, compute it.

12. $$\lim_{x\to 0} \frac{1-\cos x}{x^2}$$

13. $$\lim_{x\to 0} \frac{\sin x}{x}$$

14. $$\lim_{x\to 1} \ln(x^2)\tan(\frac{\pi}{2}x)$$

15. $$\lim_{x\to 0} x\ln(x)$$

16. $$\lim_{x\to 0} x^x$$

17. $$\lim_{x\to 0} \frac{1}{x}e^{\frac{-1}{x}}$$

18. $$\lim_{x\to -1} \frac{3x^{4/3} - 10x^{2/5} + 7}{12x^{1/3} + 7x^{4/7} + 5}$$

19. $$\lim_{x\to 1} \frac{x^{4/3} + x^{3/7} - 2}{x^{1/3} - x^{4/7}}$$

20. $$\lim_{x\to -1} \frac{3x^{4/3} - 10x^{2/5} + 7}{4x^{1/3} + 2x^{4/7} + 3}$$

21. $$\lim_{x\to 3} (x-3)^2\, e^{1/(x-3)^2}$$

# 4 Parameterized curves

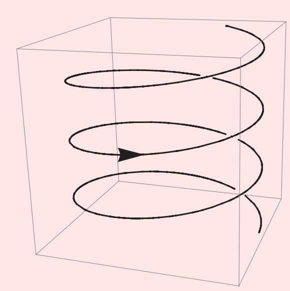
In this section we focus on a special class of multivariable functions called *parameterized curves*. Parameterized curves can be used to model the motion of a particle in two- or three-dimensional space, and they are also a helpful tool for extending notions from single-variable calculus to functions of several variables.

## 4.1 Parameterized curves

**Definition 4.1.** Let $\mathcal{I}$ be an interval in $\mathbf{R}$. A *parameterized curve* is a function $\mathbf{g} : \mathcal{I} \to \mathbf{R}^n$ with the property that each coordinate function $g_i : \mathcal{I} \to \mathbf{R}$ is continuous.

**Example 4.1.** *Consider the parameterized curve defined by* $\mathbf{g}(t) = \begin{bmatrix} \cos t \\ \sin t \\ t \end{bmatrix}$ *for* $t \in \mathbf{R}$.

Each coordinate function is a continuous function of $t$. The image of g is a *helix* in $\mathbf{R}^3$.
The helix is *oriented* by the increasing value of the parameter. When drawing the image of a parameterized curve, we indicate this orientation using an arrowhead.



**Example 4.2.** The parametric equation of a line provides a familiar example of a parameterized curve. The line in $\mathbf{R}^3$ which passes through the point $\mathbf{x} = (x_1, x_2, x_3)$ and has direction vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ is the image of the parameterized curve

$$\mathbf{g}(t) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + t \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} x_1 + tv_1 \\ x_2 + tv_2 \\ x_3 + tv_3 \end{bmatrix} \text{ for } t \in \mathbf{R}.$$

Replacing $t$ by $-t$ in this equation wouldn't change the image of g, but it would change the orientation.

**Definition 4.2.** Let $C \subset \mathbf{R}^n$, and suppose that $\mathbf{g} : \mathcal{I} \to \mathbf{R}^n$ is a parameterized curve such that $\mathbf{g}(\mathcal{I}) = C$. Then $\mathbf{g}$ is a *parameterization* of $C$.

Parameterizations are not unique; in fact, if $C \subset \mathbf{R}^n$ has one parameterization, then it has an infinite number of parameterizations. For this reason, we sometimes just say "curve" to refer to the image of g if we don't care about emphasizing a particular parameterization.

**Example 4.3.** *Parameterize the unit circle in* $\mathbf{R}^2$.
Since the unit circle is defined by the equation $x^2 + y^2 = 1$, the image of any curve $\mathbf{g}(t) = \big(x(t), y(t)\big)$ which satisfies $\big(x(t)\big)^2 + \big(y(t)\big)^2 = 1$ will lie on the unit circle. Here's one function that satisfies this condition:

$$\mathbf{f}(t) = \begin{bmatrix} \cos 2\pi t \\ \sin 2\pi t \end{bmatrix} \text{ for } t \in [0, 1]$$

Furthermore, the image of $\mathbf{f}$ contains every point on the unit circle, so $\mathbf{f}$ parameterizes the unit circle.
The following functions satisfy the same properties, so they provide alternative parameterizations:

$$\mathbf{g}(t) = \begin{bmatrix} \cos 2\pi t^2 \\ \sin 2\pi t^2 \end{bmatrix} \text{ for } t \in [0, 1]$$
$$\mathbf{h}(t) = \begin{bmatrix} \sin t \\ \cos t \end{bmatrix} \text{ for } t \in [0, 4\pi]$$

Note that $\mathbf{f}$ and $\mathbf{g}$ give the unit circle a positive (counter-clockwise) orientation, while $\mathbf{h}$ orients the circle negatively (clockwise).

**Example 4.4.** *Show that the graph of any continuous single variable function is a parameterized curve.*
Suppose that $f : \mathbf{R} \to \mathbf{R}$ is continuous. By definition, the graph of $f$ is the set of points

$$\Gamma_f = \{\big(x, f(x)\big) \in \mathbf{R}^2 \mid x \in \mathbf{R}\}.$$

We can parameterize $\Gamma_f \subset \mathbf{R}^2$ by $\mathbf{g}(t) = \begin{bmatrix} t \\ f(t) \end{bmatrix}$.

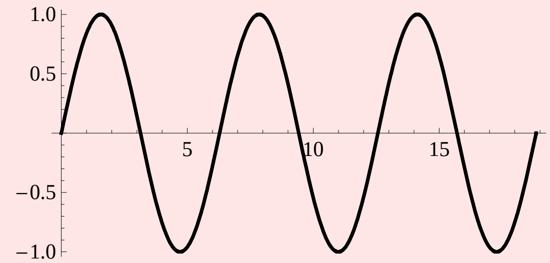
For example, suppose that $f : \mathbf{R} \to \mathbf{R}$ is defined by $f(x) = \sin(x)$.

The graph of $f$ is the set

$$\Gamma_f = \{(x, \sin(x)) \in \mathbf{R}^2 \mid x \in \mathbf{R}\}.$$

We can parameterize this curve by

$$\mathbf{g}(t) = \begin{bmatrix} t \\ \sin(t) \end{bmatrix}.$$




## 4.2 Calculus of space curves

Since each coordinate functions of a parameterized curve is a function of a single variable, we can use the tools of single-variable calculus to study parameterized curves.

**Definition 4.3.** Suppose that each coordinate function of the paramterized curve $\mathbf{g} : \mathcal{I} \to \mathbf{R}^n$ is differentiable at $a$. The *tangent vector* to $\mathbf{g}(\mathcal{I})$ at $\mathbf{g}(a)$ is the vector

$$\mathbf{g}'(a) = \begin{bmatrix} g_1'(a) \\ g_2'(a) \\ \vdots \\ g_n'(a) \end{bmatrix}.$$

If $\mathbf{g}'(a) \neq \mathbf{0}$, the *unit tangent vector to* $\mathbf{g}(\mathcal{I})$ *at* $\mathbf{g}(a)$ is the vector

$$T_{\mathbf{g}}(a) = \frac{\mathbf{g}'(a)}{||\mathbf{g}'(a)||}.$$

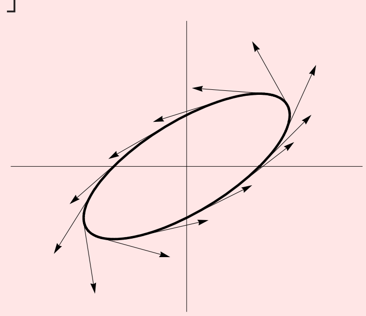
**Example 4.5.** *Consider the parameterized curve* $\mathbf{f} : [0, 2\pi] \to \mathbf{R}^2$ *defined by*

$$\mathbf{f}(t) = \begin{bmatrix} \cos t + \sin t \\ \sin t \end{bmatrix}.$$

The image of $\mathbf{f}$ is the ellipsoid $E$:

$$E = \{(x, y) | x^2 - 2xy + 2y^2 = 1\}.$$

You can check that the coordinate functions satisfy this equation. The figure to the right shows $E$ together with a collection of its unit tangent vectors.



Tangent vectors can be used to define the tangent line to curves which are not graphs.

**Definition 4.4.** Suppose that each coordinate function of the parameterized curve $\mathbf{g} : \mathcal{I} \to \mathbf{R}^n$ is differentiable at $a$, and suppose that $\mathbf{g}'(a) \neq \mathbf{0}$. Then the *tangent line to* $\mathbf{g}(\mathcal{I})$ *at* $\mathbf{g}(a)$ is the line

$$\{\mathbf{g}(a) + t\mathbf{g}'(a) \mid t \in \mathbf{R}\}.$$

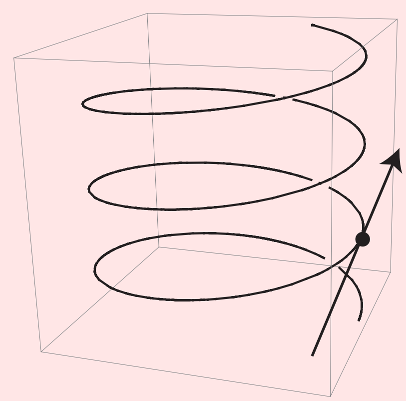
**Example 4.6.** *Find an equation for the tangent line to the helix at the point* $(1, 0, 0)$.
Recall from Example 4.1 that the helix is the image of the parameterized curve $\mathbf{g}(t) = \begin{bmatrix} \cos t \\ \sin t \\ t \end{bmatrix}$ for $t \in \mathbf{R}$. The point $(1, 0, 0)$ is $\mathbf{g}(0)$.

Differentiate $\mathbf{g}$ to find an expression for the tangent vector to the helix at $t = 0$:

$$\mathbf{g}'(0) = \begin{bmatrix} -\sin t \\ \cos t \\ 1 \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Then Definition 4.4 implies that the tangent line to the helix at $(1, 0, 0)$ is the line

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$



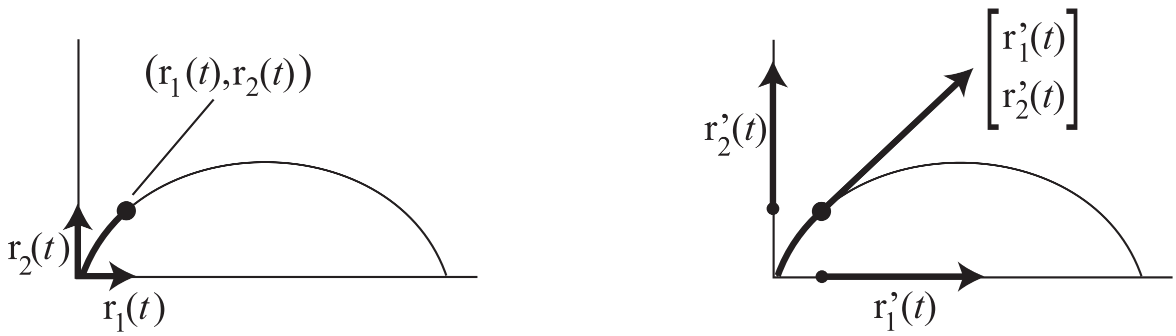
### 4.2.1 Position, velocity, and acceleration

In this section we describe how parameterized curves can be used to study the motion of a particle in $\mathbf{R}^2$ or $\mathbf{R}^3$. We first recall how these ideas appear in singe-variable calculus.

Given a continuous function $r : \mathbf{R} \to \mathbf{R}$, we can interpret $r(t)$ as describing the position of a particle on the real number line at time $t$. If the function $r$ is differentiable, then $r'(t) = v(t)$ describes the velocity of the particle, and $r''(t) = v'(t) = a(t)$ describes its acceleration. Both velocity and acceleration should be thought of as vectors in the one-dimensional space $\mathbf{R}$, whereas the speed of the particle, $|r'(t)|$, is a scalar.

An analogous physical interpretation applies to parameterized curves $\mathbf{r} : \mathcal{I} \to \mathbf{R}^n$. As before, interpret $\mathbf{r}(t)$ as the location of a particle at time $t$. In this case, the coordinate function $r_i(t)$ describes the position of the particle with respect to the $i^{th}$ coordinate.

Similarly, we can interpret the tangent vector $\mathbf{r}'(t)$ as the *velocity* of the particle. The component $r_i'(t)$ describes the velocity of the particle in the direction $\mathbf{e_i}$. If we differentiate each coordinate function again, we get the vector $\mathbf{r}''(t) = \mathbf{a}(t)$ describing the acceleration of the particle.




In the figures above, the dot is traveling along the curved arc. At the time $t$, the value of $r_1(t)$ is the $x$-coordinate of the dot, and $r_1'(t)$ is its velocity in the horizontal direction. The projection of the vector $\mathbf{r}'(t)$ to the $x$-axis is the vector $r_1'(t)$.

The magnitude of the vector has physical significance: the greater the magnitude, the faster the particle is moving. Different parameterizations of a curve may yield tangent vectors with different magnitudes. Physically, this just means that particles may travel along the same path in space at different speeds.

## 4.3 Worked Examples

**Example 4.7.** *Suppose that $f : \mathbf{R} \to \mathbf{R}$ is a differentiable single-variable function. Find the tangent line to $\Gamma_f$ at the point $(a, f(a))$.*

In Example 4.4 we parameterized the graph of a continuous single-variable function $f$ by

$$\mathbf{g}(t) = \begin{bmatrix} t \\ f(t) \end{bmatrix}.$$

Definition 4.4 implies that the tangent line to $\Gamma_f$ at $(a, f(a))$ is described by the equation

$$\mathbf{x}(t) = \mathbf{g}(a) + t\mathbf{g}'(a) \quad \text{for } t \in \mathbf{R}. \tag{4.1}$$

On the other hand, single-variable calculus tells us that the equation for the tangent line to the graph of $f$ at $(a, f(a))$ is

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

Let's see that this agrees with equation we just derived. First, observe that

$$\mathbf{g}'(a) = \begin{bmatrix} 1 \\ f'(t) \end{bmatrix}_a = \begin{bmatrix} 1 \\ f'(a) \end{bmatrix}.$$

If we rewrite Equation 4.1 in coordinates, we get the pair of equations

$$\begin{aligned} x &= a + t \\ y &= f(a) + tf'(a). \end{aligned}$$

When $f'(a) \neq 0$, we can solve both equations for the parameter $t$:

$$t = x - a$$
$$t = \frac{1}{f'(a)}\big(y - f(a)\big).$$

Setting these equal to each other yields the familiar point-slope form of the tangent line:

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

This shows that we can recover the formula for tangent lines to graphs that we had seen knew, but Definition 4.4 extends this idea to a much larger class of curves.

**Example 4.8.** *Suppose that a particle is traveling around the circle defined by*

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

*Show that its position and velocity vectors are always orthogonal.*

Describe the position of the particle at time $t$ by the parameterized curve

$$\mathbf{r}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}.$$

Then for all $t$,

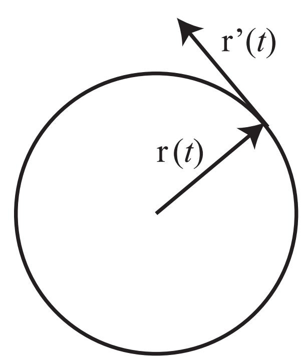
$$x^2(t) + y^2(t) = c^2.$$

Differentiate this equation with respect to $t$:

$$2x(t)x'(t) + 2y(t)y'(t) = 0.$$

Observe that the previous equation can be rewritten:

$$2\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \cdot \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = 2\mathbf{r}(t)\cdot\mathbf{r}'(t) = 2\mathbf{r}(t)\cdot\mathbf{v}(t) = 0.$$




This shows that the position and velocity vectors are orthogonal.

## 4.4 Exercises

In Exercises 1 through 4, find the tangent line to the image of the given parameterized curve at the indicated point.

1. $\mathbf{g}(t) = \big(t^2 - 2, t + 4\big)$, $(2, 6)$
2. $\mathbf{g}(t) = \big(2 \sin t \cos t, \cos(2t)\big)$, $(0, 1)$
3. $\mathbf{g}(t) = \big(e^t, \cos t^2, \ln(t + 1)\big)$, $(1, 1, 0)$
4. $\mathbf{g}(t) = \big(t^2 - 5t + 6, t^2 - 4, t^2 - 9\big)$, $(0, 0, -5)$

Each of the equations in Exercises 5-8 defines a surface in $\mathbf{R}^3$. For each one, find an example of a parameterized curve g whose image lies on the surface and which satisfies $\mathrm{g}'(t) \neq \mathbf{0}$.

5. $x^2 + y^2 + z^2 = 4$
6. $x^2 + y^2 - 2y + 1 = 1$
7. $y^2 + z^2 = 3 + \sin x$
8. $z = x^2 - y^2$

9. As in Example 4.1, define the *helix* as the image of the parameterized curve

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

   (a) Show that the image of each of the parameterized curves below lies on the helix.
      i. $\mathbf{g}(t) = (\cos e^t, \sin e^t, e^t)$
      ii. $\mathbf{h}(t) = (\cos t^3, \sin t^3, t^3)$
      iii. $\mathbf{r}(t) = (\cos(-t), \sin(-t), -t)$

   (b) Which of these parameterized curves can be used to find the tangent line to the helix at $(1, 0, 0)$? Why or why not?

10. Let $f : \mathbf{R} \to \mathbf{R}$ be continuous and let $h : \mathbf{R} \to \mathbf{R}$ be continuous and onto.

   (a) Show that the composition $\mathbf{r}(t) = \begin{bmatrix} h(t) \\ f(h(t)) \end{bmatrix}$ is a parameterization of $\Gamma_f$.

   (b) Construct an example which proves that the hypothesis that $h$ is onto is necessary.

   (c) Do the parameterized curves $\begin{bmatrix} t \\ f(t) \end{bmatrix}$ and $\begin{bmatrix} h(t) \\ f(h(t)) \end{bmatrix}$ induce the same orientation on $\Gamma_f$?

11. Consider the parametrized curve $\mathbf{r}(t) = \big(\cos(t), \sin(2t)\big)$ for $t \in [0, 2\pi)$.

   (a) Sketch the image of $\mathbf{r}$.

   (b) The curve intersects itself at the point $(0, 0)$. Find the equations of both tangent lines to the curve at this point.

12. Let $\mathbf{r} : \mathcal{I} \to \mathbf{R}^3$ be a parameterized curve, and let $g : \mathcal{D}^1 \to \mathcal{D}^1$ be a differentiable function which maps onto $\mathcal{D}^1$. Show that $\mathbf{r} \circ g$ is a parameterized curve with the same image as $\mathbf{r}$.

In Exercises 13-16, calculate the velocity, the unit tangent vector and the acceleration as a function of time for the given position function.

13. $\mathbf{r}(t) = (t^3 + t - 3, e^{2t^2})$ for $t \in \mathbf{R}$

14. $\mathbf{r}(t) = (\ln t + 3t, \sqrt{t^2+1})$ for $t > 0$

15. $\mathbf{r}(t) = (\cos 4t, \sin 2t, t^2)$ for $t \in \mathbf{R}$

16. $\mathbf{r}(t) = (3t+4, 4\sin^2 t + 3, 7t^3 - 4t^2 + 2)$ for $t \in \mathbf{R}$.

17. Let
$$\mathbf{r}(t) = (2t+2, 4t^2, \ln(t))$$
be defined for $t > 0$, and let
$$\tilde{\mathbf{r}}(s) = (e^{s+3} + 2, e^{2s+6}, s + 3 - \ln(2))$$
be defined for all $s$.
    (a) Evaluate $\mathbf{r}(t)$ at $t = \frac{1}{2}e$, and $\tilde{\mathbf{r}}(s)$ at $s = -3$.
    (b) Evaluate $\mathbf{r}'(t)$ at $t = \frac{1}{2}e$, and $\tilde{\mathbf{r}}'(s)$ at $s = -3$. Are these two tangent vectors parallel?
    (c) Show that $\mathbf{r}$ and $\tilde{\mathbf{r}}$ are parameterizations of the same curve by finding a function $g : \mathbf{R}^+ \to \mathbf{R}$ such that $\mathbf{r}(t) = \tilde{\mathbf{r}}(g(t))$.

18. For which of the following parameterized curves does the image lie on the unit sphere in $\mathbf{R}^3$?
    (a) $\mathbf{f}(t) = \big(\sin(3t), 0, \cos(3t)\big)$
    (b) $\mathbf{g}(t) = \big(t^2, \cos t, \sin t\big)$
    (c) $\mathbf{h}(t) = \big(\frac{1}{\sqrt{2}}, \frac{-\cos t}{\sqrt{2}}, \frac{\sin t}{\sqrt{2}}\big)$
    (d) $\mathbf{r}(t) = \big(\sec t, \frac{1}{\sqrt{2}}, \cos t\big)$

19. Suppose that the motion of a particle is described by the function
$$\mathbf{f}(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix},$$
and suppose that $\mathbf{f}(t) \cdot \mathbf{f}'(t) = 0$ for all $t$. Show that if $\mathbf{f}'(t) \neq \mathbf{0}$, then the particle is traveling in a circle.

    (Hint: You may find Example 4.8 helpful.)

20. Show that if $\mathbf{g}(t)$ is a parameterized curve whose image lies on a sphere in $\mathbf{R}^3$, then $\mathbf{g}(t) \cdot \mathbf{g}'(t) = 0$ for all $t$.

    (Hint: You may find Example 4.8 helpful.)

21. Let $\mathbf{r}, \mathbf{s} : \mathcal{D}^1 \to \mathbf{R}^3$ be two parameterized curves. Show that
$$\frac{d}{dt}\big(\mathbf{r}(t) \cdot \mathbf{s}(t)\big) = \mathbf{r}'(t) \cdot \mathbf{s}(t) + \mathbf{r}(t) \cdot \mathbf{s}'(t).$$

22. Show that
$$\big(\mathbf{r} \times \mathbf{s}\big)'(t) = \mathbf{r}(t)' \times \mathbf{s}(t) + \mathbf{r}(t) \times \mathbf{s}'(t).$$

23. *Tangential and Normal acceleration* Let $\mathbf{r} : \mathcal{D}^1 \to \mathbf{R}^2$ be a twice differentiable function. Suppose the velocity $\mathbf{v} = \mathbf{r}'$ is non-zero at time $t = t_0$. It is useful to describe the acceleration as the sum of two components, the *tangential* acceleration and the *normal* (also called centripetal) acceleration

$$\mathbf{a} = \mathbf{a}_T + \mathbf{a}_N$$

where $\mathbf{a}_T$ is parallel to $\mathbf{v}$ and $\mathbf{a}_N$ is perpendicular to $\mathbf{v}$.

(a) Let $\mathbf{r}(t) = (2t^3 - t^2 - 3t, t^2 + 8)$. Compute $\mathbf{a}_T(1)$ and $\mathbf{a}_N(1)$.

(b) Let $\mathbf{r}(t) = (\cos t, \sin t)$. Compute $\mathbf{a}_T$ and $\mathbf{a}_N$ for all $t$.

(c) Let $\mathbf{r}(t) = (t^5 + 5t^3 + 2t, 4t^5 + 20t^3 + 8t)$. Compute $\mathbf{a}_T$ and $\mathbf{a}_N$ for all $t$.

(d) Does $\mathbf{a}_T$ or $\mathbf{a}_N$ tell you more about the shape of the curve traced out by $\mathbf{r}$? Explain.

# 5 Limits and continuity for functions of several variables

The limit is the most important idea in calculus. In this section we define limits for multivariable functions and discuss techniques for computing them. Although we introduce the formal definition of the limit here, it's studied in more depth in Section 6; here, we'll focus on tools that don't require using $\epsilon$'s and $\delta$'s. This section has a lot of material, so there's a synopsis (Section 5.5) after the Worked Examples.

## 5.1 The definition of the limit

The definition of "limit" for multivariable functions is nearly identical to the definition for the single variable case.

The limit of $\mathbf{f}$ as $\mathbf{x}$ approaches $\mathbf{a}$ is $\mathbf{L}$ if the image of $\mathbf{f}$ can be made *arbitrarily* close to $\mathbf{L}$ by restricting to points in the domain which are *sufficiently* close to $\mathbf{a}$.

**Definition 5.1.** Let $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ be a function with the property that every punctured neighborhood of $\mathbf{a} \in \mathbf{R}^n$ contains some point of $\mathcal{D}^n$.
Then for $\mathbf{L} \in \mathbf{R}^m$,

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$$

if for every $\epsilon > 0$, there exists some $\delta > 0$ such that

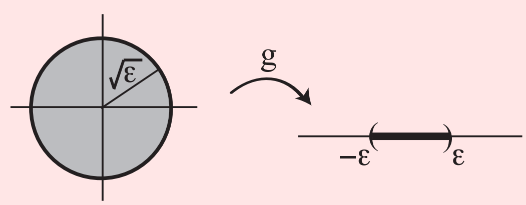
$$\text{if } \mathbf{x} \in \mathcal{D}^n \text{ and } 0 < ||\mathbf{x} - \mathbf{a}|| < \delta, \quad \text{then} \quad ||\mathbf{f}(\mathbf{x}) - \mathbf{L}|| < \epsilon.$$

In this case, we say that $\mathbf{L}$ is *the limit of* $\mathbf{f}$ *as* $\mathbf{x}$ *approaches* $\mathbf{a}$.

**Example 5.1.** *Define $g : \mathbf{R}^2 \to \mathbf{R}$ by $g(x, y) = x^2 + y^2$. For any $\epsilon > 0$, show that if*

$$0 < ||(x, y) - (0, 0)|| < \sqrt{\epsilon}, \quad \textit{then} \quad |g(x, y) - 0| < \epsilon.$$

The function $g$ takes points in $\mathbf{R}^2$ to points in $\mathbf{R}$, and we need to show that if a point in $\mathbf{R}^2$ has distance less than $\sqrt{\epsilon}$ from $(0, 0)$, then its image in $\mathbf{R}$ has distance less than $\epsilon$ from $0$.




First, notice that

$$||(x, y) - (0, 0)|| = \sqrt{(x - 0)^2 + (y - 0)^2} = \sqrt{x^2 + y^2} = \sqrt{g(x, y)}.$$

Thus, if $0 < ||(x, y) - (0, 0)|| < \sqrt{\epsilon}$, then $\sqrt{g(x, y)} < \sqrt{\epsilon}$. This implies $|g(x, y)| < \epsilon$, which is what we wanted.
This proves that the limit of $g$ as $(x, y)$ approaches $(0, 0)$ equals $0$.

The kind of argument used in Example 5.1 is developed further in Section 6, but in the remainder of this section we'll focus on other techniques for studying limits.

The next theorem implies that limits of vector-valued functions are no harder to find than limits of scalar-valued functions.

**Theorem 7.** Suppose that $f : \mathcal{D}^n \to \mathbf{R}^m$ can be written in terms of coordinate functions as $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), ...f_m(\mathbf{x}))$. Then

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) \text{ exists if and only if } \lim_{\mathbf{x}\to\mathbf{a}} f_i(\mathbf{x}) \text{ exists for all } i \in \{1, 2, ...m\}.$$

## 5.2 Continuity

There are special functions for which limits are particularly easy to find.

**Definition 5.2.** The function $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ is *continuous at* $\mathbf{a}$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}).$$

A function is *continuous* if it is continuous at all points in its domain.

As in the single variable case, observe that the definition requires three things: that $\mathbf{a} \in \mathcal{D}^n$; that the limit exists as $\mathbf{x}$ approaches $\mathbf{a}$; and that the value of the function and the limit agree.

When a function is continuous, it's possible to find its limit at any point simply by evaluation. Many of the most common classes of functions are continuous.

**Theorem 8.** The following classes of functions are continuous:

- polynomials;
- rational functions;
- the exponential function;
- the sine and cosine functions.

**Example 5.2.** *Evaluate the limit of* $f(x, y) = 3x^2 + xy - 2y$ *as* $(x, y)$ *approaches* $(3, 1)$.
The function $f$ is continuous, so it suffices to evaluate $f$ at $(3, 1)$:

$$\lim_{(x,y)\to(3,1)} f(x, y) = f(3, 1) = 3(3)^2 + (3)(1) - 2(1) = 28.$$

**Example 5.3.** *Evaluate the limit of* $\frac{xy-y}{x^2-1}$ *as* $(x, y)$ *approaches* $(1, 0)$.
The rational function $\frac{xy-y}{x^2-1}$ is not defined at $(1, 0)$, so we can't simply plug in $x = 1$ and $y = 0$.
Instead, factor the numerator and denominator:

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

The function $\frac{y}{x+1}$ is continuous at $(1, 0)$, so evaluation shows that

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

Since the existence of the limit depends on what happens *near* $(1, 0)$, but not *at* $(1, 0)$, we evaluate the original limit using a continuous function that agrees with it near the origin.

The next few theorems show how the library of continuous functions established in Theorem 8 can be used to show that other functions are continuous as well.

**Theorem 9.** Suppose that $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$and $\mathbf{g} : \mathcal{D}^n \to \mathbf{R}^m$ are continuous at $\mathbf{a}$. Then $\mathbf{f} + \mathbf{g}$ is continuous at $\mathbf{a}$. Furthermore, if $c \in \mathbf{R}$, then $c\mathbf{f}$ is continuous at $\mathbf{a}$.

**Theorem 10.** Suppose that $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ is continuous at some point $\mathbf{a}$. If $\mathbf{g} : \mathcal{E}^m \to \mathbf{R}^p$ is continuous at $\mathbf{f}(\mathbf{a})$, then the composition $\mathbf{g} \circ \mathbf{f}$ is continuous at $\mathbf{a}$.

The proofs of these theorems are similar to those of their single variable analogues. They follow from the Multivariable Limit Laws, which are discussed in Section 6.1.

## 5.3 Techniques for limit problems

Limit problems tend to come in two flavors: finding limits or proving they don't exist. In this section we discuss techniques for both types of problems.

### 5.3.1 The Squeeze Theorem

The Squeeze Theorem for multivariable functions is similar to the single variable version. In both cases, we show that the limit of a given function exists by bounding it between two functions whose limits are known.

**Theorem 11.** Suppose that $f(\mathbf{x})$, $g(\mathbf{x})$, and $h(\mathbf{x})$ are real-valued functions with the property that on some punctured neighborhood of $\mathbf{a}$,

$$f(\mathbf{x}) \leq g(\mathbf{x}) \leq h(\mathbf{x}).$$

$$\text{If } \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \lim_{\mathbf{x}\to\mathbf{a}} h(\mathbf{x}), \quad \text{then} \quad \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = \lim_{\mathbf{x}\to\mathbf{a}} h(\mathbf{x}).$$

In many cases, we will use the continuous functions from Theorem 8 to provide the bounding functions $f$ and $h$.

**Example 5.4.** *Use the Squeeze Theorem to show that*

$$\lim_{(x,y)\to(0,0)} \frac{2xy}{\sqrt{x^2+y^2}} = 0.$$

In order to apply the Squeeze Theorem, we need to find functions $h$ and $f$ which provide upper and lower bounds for $\frac{2xy}{\sqrt{x^2+y^2}}$ near $(0,0)$.
One way to make a ratio bigger is to divide by a smaller denominator. Thus,

$$\frac{2xy}{\sqrt{x^2+y^2}} \leq \frac{2|xy|}{\sqrt{x^2}} = 2|y| \text{ when } x \neq 0.$$

On the other hand, if $x = 0$ and $y \neq 0$, then

$$\frac{2(0)y}{\sqrt{0^2+y^2}} = 0 \leq 2|y|.$$

This shows that the function $h(x,y) = 2|y|$ bounds $\frac{2xy}{\sqrt{x^2+y^2}}$ from above for all points $(x,y) \neq (0,0)$. To get a lower bound $f(x,y)$, we use a similar argument to show

$$-2|y| \leq \frac{2xy}{\sqrt{x^2+y^2}}$$

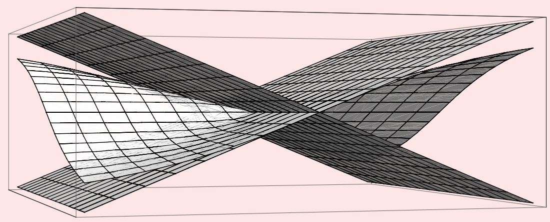
The absolute value of a continuous function is continuous, so Theorem 8 implies

$$\lim_{(x,y)\to(0,0)} f(x,y) = \lim_{(x,y)\to(0,0)} -2|y| = 0 = \lim_{(x,y)\to(0,0)} 2|y| = \lim_{(x,y)\to(0,0)} h(x,y).$$

Since the limits of the two bounding functions are $0$, the Squeeze Theorem proves

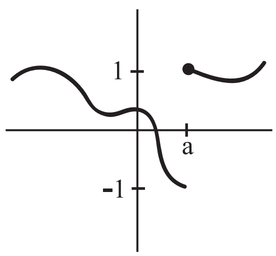
$$\lim_{(x,y)\to(0,0)} \frac{2xy}{\sqrt{x^2+y^2}} = 0.$$

The figure below illustrates how the Squeeze Theorem works in this example. The graphs of the bounding functions $2|y|$ and $-2|y|$ trap the graph of $\frac{2xy}{\sqrt{x^2+y^2}}$ between them.



### 5.3.2 Evaluating a limit along a curve

Recall from single variable calculus that a function $f$ can have a limit as $x$ approaches $a$ from above or as $x$ approaches $a$ from below.

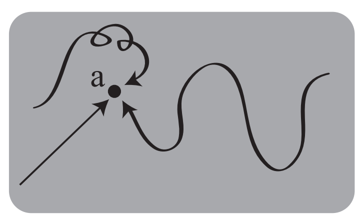



This graph shows that the limit as $x$ approaches $a$ from the left exists, and the limit as $x$ approaches $a$ from the right exists, but these values disagree. See Section 2 for more discussion of the single variable case.

The limit of $f$ as $x$ approaches $a$ exists only when the limit from above and the limit from below both exist and are the same. That is, we say that this limit of $f$ exists only when it's irrelevant *how* $x$ approaches $a$.

A similar idea appears in multivariable limits, except that now **x** can approach **a** along infinitely-many different paths, not just from the left and right.

Just as we can show a single variable limit fails to exist if the limits from above and below disagree, we can show a multivariable limit fails to exist if the limits along different paths in the domain disagree.




**Example 5.5.**

$$\textit{Let } f(x,y) = \frac{x^3}{y^3}. \quad \textit{Show that } \lim_{(x,y)\to(0,0)} f(x,y) \textit{ does not exist.}$$

We'll show that approaching $(0,0)$ along different paths in $\mathbf{R}^2$ leads to different values for the limit along the curve.

First, consider what happens if we approach $(0,0)$ along the line $y = x$.

Substituting $y = x$ into the formula for $f$, we get

$$\lim_{x \to 0} f(x, x) = \lim_{x \to 0} \frac{x^3}{x^3} = \lim_{x \to 0} 1 = 1.$$

This implies that if the limit of $f$ as $(x, y)$ approaches $(0, 0)$ exists, then it must equal $1$. However, suppose we approach $(0, 0)$ along the line $y = 2x$ instead. Then

$$\lim_{x \to 0} f(x, 2x) = \lim_{x \to 0} \frac{x^3}{(2x)^3} = \lim_{x \to 0} \frac{x^3}{8x^3} = \frac{1}{8}.$$

Since this disagrees with the previous value, $\lim_{(x,y) \to (0,0)} f(x, y)$ does *not* exist.

The next example is similar, but we use parameterized curves to emphasize that one can consider any path of approach, not just lines.

**Example 5.6.**

*For* $f(x, y) = \dfrac{x^3}{y^2}$, *show that* $\displaystyle\lim_{(x,y) \to (0,0)} f(x, y)$ *does not exist.*

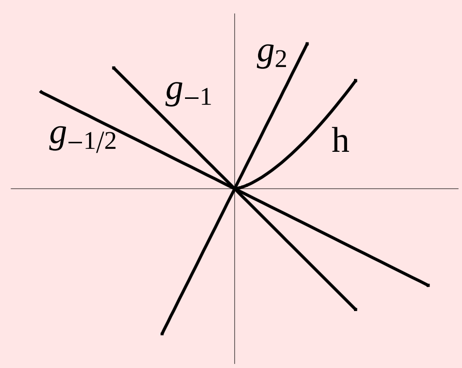
Let's start by approaching $(0, 0)$ along the line $y = mx$. We can parameterize this line by the function $\mathbf{g}_m(t) = (t, mt)$ for any value of $m \in \mathbf{R}$. The point $\mathbf{g}_m(t)$ approaches $(0, 0)$ as $t$ approaches $0$, so we'll study the limit of the composition $f \circ \mathbf{g}_m$:

$$\lim_{t \to 0} (f \circ \mathbf{g}_m)(t) = \lim_{t \to 0} \frac{(x(t))^3}{(y(t))^2} = \lim_{t \to 0} \frac{t^3}{(mt)^2} = \frac{1}{m^2} \lim_{t \to 0} t = 0.$$

This shows that if the limit of $f$ exists as $(x, y) \to (0, 0)$, it must equal $0$. However, even though the parameterized curves $\mathbf{g}_m$ describe (nearly) all the lines through the origin, this is not enough to show that the limit of $f$ exists.

This time, let's approach $(0, 0)$ along the curve defined by $y^2 = x^3$ in the first quadrant. We can parameterize this by $\mathbf{h}(t) = (t, t^{\frac{3}{2}})$ for $t > 0$. Then

$$\begin{aligned} \lim_{t \to 0} (f \circ \mathbf{h})(t) &= \lim_{t \to 0} \frac{(x(t))^3}{(y(t))^2} \\ &= \lim_{t \to 0} \frac{t^3}{(t^{\frac{3}{2}})^2} \\ &= \lim_{t \to 0} 1 \\ &= 1. \end{aligned}$$



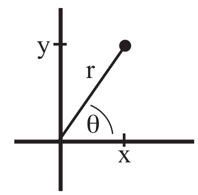
Since approaching along a line yields a different value than approaching along $y^2 = x^3$, the multivariable limit does not exist.

### 5.3.3 Polar coordinates

A point in $\mathbf{R}^2$ can be described by either Cartesian coordinates or polar coordinates.

The two coordinate systems are related by

$$x = r\cos\theta$$
$$y = r\sin\theta.$$




Letting $(x, y) \to (0, 0)$ is equivalent to letting $r \to 0$, and sometimes limits at the origin may be evaluated more easily by rewriting the function in polar coordinates. In particular, it may be easier to find bounding functions when applying the Squeeze Theorem. Examples 5.8 and 5.9 in the Worked Examples section illustrate this technique.

## 5.4 Worked Examples

**Example 5.7.** *Evaluate*

$$\lim_{(x,y)\to(0,0)} \frac{x}{1-\sqrt{1-y^2}} .$$

*or show the limit does not exist.*

With an open-ended problem like this, the first task is to guess whether or not the limit exists. One way to do this is to search for reasonable candidate for the limit. With a candidate in hand, we can then try to construct a delta/epsilon argument or apply the Squeeze Theorem. However, if we can't find a candidate value for the limit, then our efforts are better directed to trying to prove the limit doesn't exist.

In both cases, a natural first step is to try to evaluate the limit along a path in the domain.

To start with, parameterize the $y$-axis by $\mathbf{g}(t) = (0, t)$. Then $\mathbf{g}(0) = (0, 0)$, and

$$\lim_{(x,y)\to(0,0)} \frac{x}{1-\sqrt{1-y^2}} = \lim_{t\to 0} \frac{0}{1-\sqrt{1-t^2}}$$

if both limits exist.

The composition $\frac{0}{1-\sqrt{1-y^2}}$ is identically equal to $0$ where it's defined, so the right-hand limit equals $0$. We now have a candidate value for the multivariable limit.

However, it's worth looking a little harder before trying to prove that this is correct.

The original function isn't defined at any point on the $y$ axis. In fact, for any fixed, non-zero value of $x$, if we let $y \to 0$, the value of the function will blow up to infinity. Although we can't yet rule out the possibility that letting $x \to 0$ somehow cancels this effect, this observation should make us cautious.

Consider the path $\mathbf{h}(t) = (1 - \sqrt{1-t^2}, t)$. The image of $\mathbf{h}$ contains $(0,0)$, so evaluating the limit along this path shows that

$$\lim_{(x,y)\to(0,0)} \frac{x}{1-\sqrt{1-y^2}} = \lim_{t\to 0} \frac{1-\sqrt{1-t^2}}{1-\sqrt{1-t^2}}$$

if both limits exist.

The right-hand limit is clearly equal to $1$, which proves that the original multivariable limit does not exist.

**Example 5.8.** *Use polor coordinates to prove the limit from Example 5.1:*

$$\lim_{(x,y)\to(0,0)} \frac{2xy}{\sqrt{x^2+y^2}}$$

Rewriting the function in polar coordinates yields

$$\lim_{r\to 0} \frac{2(r\cos\theta)(r\sin\theta)}{\sqrt{r^2\cos^2\theta + r^2\sin^2\theta}} = \lim_{r\to 0} \frac{2r^2\cos\theta\sin\theta}{|r|} = \lim_{r\to 0} 2|r|\cos\theta\sin\theta \text{ for } r \neq 0.$$

This is an easier limit to study. First, observe that since $|\cos\theta\sin\theta| \leq 1$, the following bounds hold for all values of $\theta$ and $r$:

$$-2|r| \leq 2|r|\cos\theta\sin\theta \leq 2|r|.$$

We note that

$$\lim_{r\to 0} \pm 2|r| = 0,$$

so the Squeeze Theorem implies that

$$\lim_{r\to 0} \frac{(r\cos\theta)(r\sin\theta)}{\sqrt{r^2\cos^2\theta + r^2\sin^2\theta}} = \lim_{(x,y)\to(0,0)} \frac{2xy}{\sqrt{x^2+y^2}} = 0.$$

It's important that the limit didn't depend on $\theta$, as you can see in the next example. You can take the limit as $r$ approaches $0$, but this gives you no control over the variable $\theta$.

**Example 5.9.** In Example 5.5, we showed that $\lim_{(x,y)\to(0,0)} \frac{x^3}{y^3}$ does not exist. This is also clear if we change to polar coordinates:

$$\lim_{(x,y)\to(0,0)} \frac{x^3}{y^3} = \lim_{r\to 0} \frac{r^3\cos^3\theta}{r^3\sin^3\theta} = \lim_{r\to 0} \cot^3\theta.$$

The value of the function is independent of $r$ but changes with $\theta$, so the desired limit does not exist.

## 5.5 Overview of the section

This informal overview summarizes the main points in this section. It's not meant as a substitute for the more detailed discussion found earlier, but you may find it helpful in organizing the information.

- **Definitions**

  We say that $\mathbf{L}$ is the *limit of* $\mathbf{f}$ *as* $\mathbf{x} \to \mathbf{a}$ if $\mathbf{f}(\mathbf{x})$ can be made arbitrarily close to $\mathbf{L}$ by ensuring that $\mathbf{x}$ is sufficiently close to $\mathbf{a}$. The formal definition of a limit expresses this relationship in terms of balls around $\mathbf{a}$ and $\mathbf{L}$: for any $\epsilon$, there exists a $\delta$ with the property that if $\mathbf{x}$ lies in a ball of radius $\delta$ around $\mathbf{a}$, then $\mathbf{f}(\mathbf{x})$ lies in in a ball of radius $\epsilon$ around $\mathbf{L}$.

  A function $\mathbf{f}$ is *continuous at* $\mathbf{a}$ if $\mathbf{f}(\mathbf{a})$ is equal to the limit of $\mathbf{f}$ as $\mathbf{x}$ approaches $\mathbf{a}$.

- **Continuous functions** It's often useful to have a library of functions known to be continuous. For example, the following functions are continuous at all points in their domains:

  - rational functions (including polynomials);
  - some trigonometric functions;
  - exponential functions.

  Furthermore, you can build new continuous functions by scalar multiplication, addition, composition, and multiplication (for scalar-valued functions).

- **Showing a limit exists** To show a limit exists, the following techniques may be helpful:

  - Use algebraic manipulation to compare the function to a function known to be continuous.
  - Bound the function by functions known to be continuous and use the Squeeze Theorem.
  - For functions defined on $\mathbf{R}^2$ and limits at the origin, change to polar coordinates and evaluate the limit as $r \to 0$.

- **Showing a limit doesn't exist** The basic technique for showing a limit doesn't exist is to restrict to a curve in the domain and evaluate the limit along the curve. Sometimes it's helpful to try the following types of curves:

  - the coordinate axes;
  - lines through the point $\mathbf{a}$;
  - curves satisfying $x^m = y^n$ for integer values of $m$ and $n$.

## 5.6 Exercises

In Exercises 1 - 27, find the given limit or show it does not exist.

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

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

3.
$$\lim_{(x,y)\to(0,0)} \frac{x^2+y^2}{x^2-y^2}$$

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

5.
$$\lim_{(x,y)\to(0,0)} \frac{4x^2-9y^2}{xy}$$

6.
$$\lim_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2-3x+xy-3y}$$

7.
$$\lim_{(x,y)\to(0,0)} \frac{x^3-3x^2y+3xy^2-y^3}{x^2-2xy+y^2}$$

8.
$$\lim_{(x,y)\to(0,0)} \frac{x^2}{y^2}$$

9.
$$\lim_{(x,y)\to(0,0)} y\sin\left(\frac{x^2}{y^2}\right)$$

.

10.
$$\lim_{(x,y)\to(0,0)} \frac{x^2-6x+5}{y^2}$$

11.
$$\lim_{(x,y)\to(0,0)} \frac{x^2+3xy+y^2}{x^2+y^2}$$

12.
$$\lim_{(x,y)\to(0,0)} \frac{x^2y-xy^2}{x^3+y^3}$$

13.
$$\lim_{(x,y)\to(0,0)} \frac{x^3-y^3}{x^2+y^2}$$

14.
$$\lim_{(x,y)\to(1,1)} \frac{x^2-2xy+y^2}{x^2+5xy-6y^2}$$

15.
$$\lim_{(x,y)\to(2,1)} \frac{x^2-6xy+8y^2}{x^2+3xy-10y^2}$$

16.
$$\lim_{(x,y)\to(1,1)} \frac{xy-1}{xy-x-y+1}$$

17.
$$\lim_{(x,y)\to(-1,2)} \frac{x^2-1}{4x^2-y^2}$$

18.
$$\lim_{(x,y)\to(-1,2)} \frac{x+y}{x^2-y^2}$$

19.
$$\lim_{(x,y)\to(1,1)} \frac{x^4+y^4}{x^2+y^2}$$

20.
$$\lim_{(x,y)\to(0,0)} \frac{9x-4y}{3\sqrt{x}+2\sqrt{y}}$$

21.
$$\lim_{(x,y)\to(0,0)} \frac{x^2y-xy^2}{x^2+y^2}$$

22.
$$\lim_{(x,y)\to(0,0)} \frac{x^4+2x^2y^2+y^4}{e^{1/(x^2y^2)}}$$

23.
$$\lim_{(x,y)\to(0,0)} (x^2+y^2)\sin\big(\ln(x^2+y^2)\big)$$

24.
$$\lim_{(x,y,z)\to(0,0,0)} \frac{xy+xz+yz}{\sqrt{x^2+y^2+z^2}}$$

25.
$$\lim_{(x,y,z)\to(0,0,0)} \frac{\cos(xyz)}{\ln|x+y+z|}$$

26.
$$\lim_{(x,y,z)\to(0,0,0)} \frac{e^{x^2+y^2+z^2}}{x^2+y^2+z^2}$$

27.
$$\lim_{(x,y,z)\to(0,0,0)} \frac{e^{x+y+z}}{\cos(x+y+z)}$$

28. Find the limit or show it does not exist.

$$\lim_{(x_1,x_2,\dots,x_n)\to(0,0,\dots,0)} \frac{x_1+x_2+\dots+x_n}{x_1^2+x_2^2+\dots+x_n^2}$$

(Hint: You may find it helpful to try setting $n$ equal to $1$ or $2$ first.)

29. Find the limit or show it does not exist.

$$\lim_{(x_1,\dots,x_n)\to(0,0,\dots,0)} (x_1^2+\dots+x_n^2)\left[\cos(1/x_1)+2\cos(1/x_2)+\dots+n\cos(1/x_n)\right]=0.$$

(Hint: You may find it helpful to try setting $n$ equal to $1$ or $2$ first.)

30. Is there a value of $a$ such that the function

$$f(x,y)=\begin{cases} \frac{x^4+10x^2y^2+21y^4}{x^2+3y^2} & \text{if } (x,y)\neq(0,0) \\ a & \text{if } (x,y)=(0,0) \end{cases}$$

is continuous? If so, find it.

31. Is there a value of $a$ such that the function

$$f(x,y)=\begin{cases} \frac{x^2+2y^2}{2x^2+y^2} & \text{if } (x,y)\neq(0,0) \\ a & \text{if } (x,y)=(0,0) \end{cases}$$

is continuous? If so, find it.

32. Is the function

$$f(x,y)=\begin{cases} \frac{(x+y)^2}{\sqrt{x^2+y^2}} & \text{if } (x,y)\neq(0,0) \\ 1 & \text{if } (x,y)=(0,0) \end{cases}$$

continuous? Explain your answer.

33. Is the function

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

continuous? Explain your answer.

# 6 The definition of the limit*

In this section we revisit the definition of the limit. Although the definition may look off-puttingly formal, the goal of this section is making this topic seem more approachable. First, recall the definition from Section 5:

**Definition 6.1.** Let $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ be a function with the property that every punctured neighborhood of $\mathbf{a} \in \mathbf{R}^n$ contains some point of $\mathcal{D}^n$.
Then for $\mathbf{L} \in \mathbf{R}^m$,

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$$

if for every $\epsilon > 0$, there exists some $\delta > 0$ such that for every $\mathbf{x} \in \mathcal{D}^n$,

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

In this case, we say that $\mathbf{L}$ is *the limit of* $\mathbf{f}$ *as* $\mathbf{x}$ *approaches* $\mathbf{a}$.

To say that the limit of $f$ as $\mathbf{x}$ approaches $\mathbf{a}$ equals $\mathbf{L}$ means that $\mathbf{f}(\mathbf{x})$ can be guaranteed to be as close to $\mathbf{L}$ as we want, as long as $\mathbf{x}$ is chosen close enough to $\mathbf{a}$. In practice, we let the variable $\epsilon$ denote the tolerance in the codomain and then try to solve for $\delta$ as a function of $\epsilon$.

**Example 6.1.** *Prove the given limit:*

$$\lim_{(x,y)\to(0,0)} x^2 y = 0.$$

Fix $\epsilon > 0$. We must find some $\delta > 0$ with the property that

$$0 < ||(x, y) - (0, 0)|| < \delta \quad \text{implies} \quad |x^2 y - 0| < \epsilon.$$

Although we don't yet know how $\delta$ should depend on $\epsilon$, we can nevertheless study what consequences follow from picking an arbitrary $\delta$.
For example, setting $||(x, y) - (0, 0)|| < \delta$ implies that

$$|x| < \delta \quad \text{and} \quad |y| < \delta.$$

What we're really interested in is the effect this choice has on $|f(x, y) - 0|$, so we plug in these inequalities:

$$\begin{aligned} |f(x,y) - 0| &= |x^2 y - 0| \\ &= |x|^2 |y| \\ &< (\delta)^2 \delta \\ &= \delta^3. \end{aligned}$$

Thus, if $\delta^3 < \epsilon$, then we get the bound we want, and this proves that the limit equals zero. This argument has all the right information, but we can organize it more clearly. A clear solution might look like this:
We claim
$$\lim_{(x,y)\to(0,0)} x^2y = 0.$$
Proof: Fix $\epsilon > 0$, and set $\delta = \sqrt[3]{\epsilon}$.
We will show that if $0 < ||(x,y)-(0,0)|| < \delta$, then $|x^2y - 0| < \epsilon$.
Suppose $(x,y)$ satisfies $0 < ||(x,y)-(0,0)|| < \delta$.
Then
$$||(x,y)-(0,0)|| = \sqrt{(x-0)^2+(y-0)^2} = \sqrt{x^2+y^2} < \delta.$$
This implies
$$x^2+y^2 < \delta^2.$$
Therefore
$$|x| < \delta \quad \text{and} \quad |y| < \delta.$$
Therefore
$$|f(x,y)-0| = |x^2y| < \delta^2|\delta| = (\sqrt[3]{\epsilon})^3 = \epsilon,$$
as desired.

## 6.1 Multivariable Limit Laws

Another approach to Example 6.1 begins with the observation that $x^2y$ is the product of simpler functions. The Multivariable Limit Laws describe how limits behave under sums and products.

**Theorem 12.** [Multivariable Limit Laws] Suppose that $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ and $\mathbf{g} : \mathcal{D}^n \to \mathbf{R}^m$ are two functions with the same domain. Furthermore, suppose that
$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L_f} \text{ and } \lim_{\mathbf{x}\to\mathbf{a}} \mathbf{g}(\mathbf{x}) = \mathbf{L_g}.$$
Then
$$\lim_{\mathbf{x}\to\mathbf{a}} (\mathbf{f}+\mathbf{g})(\mathbf{x}) = \mathbf{L_f} + \mathbf{L_g} \tag{6.1}$$
$$\lim_{\mathbf{x}\to\mathbf{a}} (c\mathbf{g})(\mathbf{x}) = c\mathbf{L_g} \text{ for any scalar } c \tag{6.2}$$
$$\lim_{\mathbf{x}\to\mathbf{a}} (fg)(\mathbf{x}) = L_f L_g \quad (m = 1 \text{ only}). \tag{6.3}$$

When evaluating the limit of a complicated function, the Limit Laws sometimes allow you to decompose it into a collection of functions which are easier to understand.

**Example 6.2.** *Evaluate the limit*

$$\lim_{(x,y)\to(0,0)} x^2y.$$

This time we begin by writing $x^2y$ as a product of the functions $f(x,y) = x$ and $g(x,y) = y$. According to Theorem 12,

$$\lim_{(x,y)\to(0,0)} x^2y = \big(\lim_{(x,y)\to(0,0)} f(x,y)\big)^2\big(\lim_{(x,y)\to(0,0)} g(x,y)\big), \tag{6.4}$$

as long as the limits of $f$ and $g$ exist as $(x,y)$ approaches $(0,0)$.
We'll show that the limit of $f$ as $(x,y)$ approaches $(0,0)$ is $0$.
Given any $\epsilon > 0$, we claim that the condition of Definition 6.1 is satisfied if we set $\delta = \epsilon$.
To see this, rewrite $0 < ||(x,y) - (0,0)|| < \epsilon$ as $0 < \sqrt{x^2+y^2} < \epsilon$.
Since $\sqrt{x^2} \le \sqrt{x^2+y^2}$, this implies

$$\sqrt{x^2} = |x| = |x-0| = |f(x,y) - 0| < \epsilon.$$

This proves that the limit of $f$ as $(x,y)$ approaches $(0,0)$ equals $0$, and a similar argument shows that the limit of $g$ as $(x,y)$ approaches $(0,0)$ also equals $0$.
Putting this together with Equation 6.4, we see

$$\lim_{(x,y)\to(0,0)} x^2y = \big(\lim_{(x,y)\to(0,0)} f(x,y)\big)^2\big(\lim_{(x,y)\to(0,0)} g(x,y)\big) = (0)^2(0) = 0.$$

We'll prove the first statement in Theorem 12, and the others are left as exercises.

*Proof of Theorem 12, Part (1).* According to Theorem 7, the limit of a multivariable function exists if and only if the limits of each of its coordinate functions exist. It therefore suffices to prove Theorem 12 for the case when the codomain is $\mathbf{R}$.

As stated in the hypotheses, suppose that

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L_f \quad \text{and} \quad \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = L_g.$$

In order to show that $\lim_{\mathbf{x}\to\mathbf{a}}(\mathbf{f}+\mathbf{g})(\mathbf{x}) = L_f + L_g$, we begin by picking $\epsilon > 0$.

Since the limit of $f$ exists as $\mathbf{x}$ approaches $\mathbf{a}$, there exists some $\delta_f > 0$ with the property that

$$0 < ||\mathbf{x} - \mathbf{a}|| < \delta_f \quad \text{implies} \quad |f(\mathbf{x}) - L_f|| < \frac{\epsilon}{2}.$$

Similarly, the existence of the limit of $g$ as $\mathbf{x}$ approaches $\mathbf{a}$ implies that there exists some $\delta_g > 0$ such that

$$0 < ||\mathbf{x} - \mathbf{a}|| < \delta_g \quad \text{implies} \quad |g(\mathbf{x}) - L_g|| < \frac{\epsilon}{2}.$$

Define $\delta$ to be the smaller of $\delta_f$ and $\delta_g$. Then

$$0 < ||\mathbf{x} - \mathbf{a}|| < \delta$$

implies

$$|(f+g)(\mathbf{x})-(L_f+L_g)| \leq |f(\mathbf{x})-L_f|+|g(\mathbf{x})-L_g| < \frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

(We used the Triangle Inequality to get the first "$\leq$".)

This proves that the limit of $f+g$ as $\mathbf{x}$ approaches $\mathbf{a}$ is $L_f+L_g$.

$\square$

## 6.2 Exercises

For more practice with $\epsilon/\delta$ arguments, try the first 12 exercises in Section 2.

In Exercises 1 through 10, prove the given limit without invoking continuity.

1. $$\lim_{(x,y)\to(0,0)} x^2+y^2=0$$

2. $$\lim_{(x,y)\to(0,0)} 2x+2y=0$$

3. $$\lim_{(x,y)\to(1,2)} x^2+y^2=5$$

4. $$\lim_{(x,y,z)\to(2,1,2)} 4x-5y-7z=-4$$

5. $$\lim_{(x,y)\to(2,1)} x^2+2y=6$$

6. $$\lim_{\mathbf{x}\to\mathbf{0}} 5\mathbf{x}=\mathbf{0}$$

7. $$\lim_{(x,y)\to(0,0)} \frac{2x^2y}{\sqrt{x^4+y^4}}=0$$

8. $$\lim_{(x,y)\to(0,0)} 3x^2-2xy=0$$

9. $$\lim_{(x,y)\to(1,2)} xy+y=4$$

10. $$\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^2+y^2}=0$$

In Exercises 11 through 15, use Definition 6.1 to prove the given statement.

11. Show that if $T:\mathbf{R}^n\to\mathbf{R}^m$ is a linear transformation, then $T$ is continuous.

12. Prove Part (2) of Theorem 12.

13. Prove Part (3) of Theorem 12.

14. Let $\mathbf{x}=\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}$ and $\mathbf{a}=\begin{bmatrix} a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix}$.

    (a) Show that if $|x_i-a_i|<\delta$ for $i\in\{1,2,...n\}$, then $||\mathbf{x}-\mathbf{a}||<\sqrt{n\delta^2}$.

    (b) Show that if $||\mathbf{x}-\mathbf{a}||<\delta$, then $|x_i-a_i|<\delta$ for all $i\in\{1,2,...n\}$.

(c) Use Theorem 12 to show that multivariable polynomials are continuous. (Hint: Example 6.2 may also be helpful.)

(d) Prove Theorem 7.

15. Let $f, g : \mathbf{R}^n \to \mathbf{R}$ and suppose there exist $\delta > 0$ and $M \in \mathbf{R}$ such that $||\mathbf{x} - \mathbf{a}|| < \delta$ implies $|f(\mathbf{x})| < M$.

(a) Show that if $\lim_{\mathbf{x}\to\mathbf{a}} |g(\mathbf{x})|$ exists, then

$$\lim_{\mathbf{x}\to\mathbf{a}} |f(\mathbf{x})g(\mathbf{x})| < M \lim_{\mathbf{x}\to\mathbf{a}} |g(\mathbf{x})|.$$

(b) Show that

$$\lim_{(x,y,z)\to(0,0,0)} \frac{xyz}{|xyz|} \sin(xyz) = 0.$$

16. Definition 6.1 requires that every punctured neighborhood of $\mathbf{a}$ has non-empty intersection with the domain of $\mathbf{f}$. Show that if this condition is removed from the definition, then the limit of $\mathbf{f}$ as $\mathbf{x}$ approaches $\mathbf{a}$ is not always unique.

# 7 Partial Derivatives and the Total Derivative

In this section we begin to develop the idea of the derivative of a multivariable function. Where possible, we emphasize the parallels with single variable derivatives.

## 7.1 Partial derivatives

Recall that for a function $f : \mathbf{R} \to \mathbf{R}$, the derivative of $f$ at $a$ is the limit

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

provided this limit exists. (See Section 3 for a review of single variable differentiation.)

When $\mathbf{g} : \mathbf{R}^n \to \mathbf{R}^m$ is a multivariable function, the numerator and denominator of the analogous ratio are vectors rather than scalars. However, we can still try to construct the derivative as a tool for measuring a rate of change. In particular, we could ask how a change in one coordinate in the domain affects the value of a particular coordinate in the range. This suggests the following definition:

**Definition 7.1.** Given $f : \mathcal{D}^n \to \mathbf{R}$, let a be a vector in $\mathcal{D}^n$. Then the *partial derivative* of $f$ with respect to $x_j$ at $\mathbf{a}$ is defined by

$$\frac{\partial f}{\partial x_j}(\mathbf{a}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{e_j}) - f(\mathbf{a})}{h},$$

provided this limit exists.

You might also see the notation $D_{x_j} f(\mathbf{a})$, $D_j f(\mathbf{a})$, and $f_{x_j}(\mathbf{a})$ used to denote the $j^{th}$ partial derivative of $f$ at $\mathbf{a}$. When the coordinates of $\mathbf{R}^3$ are identified with the variables $x, y$, and $z$, we sometimes write $f_x(\mathbf{a})$, $f_y(\mathbf{a})$ and $f_z(\mathbf{a})$ instead of $f_{x_1}(\mathbf{a})$, $f_{x_2}(\mathbf{a})$, and $f_{x_3}(\mathbf{a})$.

**Example 7.1.** *Compute the partial derivatives of $f(x, y) = x^2 y$ at the point $(a, b)$.*
We evaluate the limit from Definition 7.1:

$$\begin{aligned}
\frac{\partial f}{\partial x}(a, b) &= \lim_{h \to 0} \frac{(a+h)^2 b - a^2 b}{h} \\
&= b \lim_{h \to 0} \frac{(a+h)^2 - a^2}{h} \\
&= b(\frac{d}{dx} x^2)|_a \\
&= b(2x)|_a \\
&= 2ba
\end{aligned}$$

Having fixed $y = b$ initially, $y$ behaves like a constant throughout the computation.

Similarly,

$$\frac{\partial f}{\partial y}(a,b) = \lim_{h\to 0} \frac{a^2(b+h) - a^2 b}{h} = a^2 \lim_{h\to 0} \frac{(b+h)-b}{h} = a^2.$$

This example shows that partial derivatives are no harder to compute than derivatives of single variable functions. To compute $\frac{\partial f_i}{\partial x_j}$, simply treat $x_k$ as a constant for $k \neq j$ and apply single variable differentiation techniques to $f_i(x_j)$.

Each coordinate function of $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ has $n$ different partial derivatives at $\mathbf{a}$, and it's often useful to organize all this information simultaneously.

**Definition 7.2.** Given a function $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$, the *matrix of partial derivatives of* $\mathbf{f}$ *at the point* $\mathbf{a}$ is the matrix $D\mathbf{f}(\mathbf{a})$ whose $(i,j)$ entry is $\frac{\partial f_i}{\partial x_j}(\mathbf{a})$, provided this is defined.

$$D\mathbf{f}(\mathbf{a}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \dots & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{bmatrix}$$

**Example 7.2.** *Compute the matrix of partial derivatives of* $\mathbf{f}(x,y) = (3xy^2, e^{2y+\sin x})$ *at the point* $(0,-1)$.

First, compute the partial derivatives of $\mathbf{f}$ at $(0,-1)$:

$$\frac{\partial f_1}{\partial x}(0,-1) = 3y^2|_{(0,-1)} = 3 \qquad \frac{\partial f_1}{\partial y}(0,-1) = 6xy|_{(0,-1)} = 0$$

$$\frac{\partial f_2}{\partial x}(0,-1) = (\cos x)e^{2y+\sin x}|_{(0,-1)} = e^{-2} \qquad \frac{\partial f_2}{\partial y}(0,-1) = 2e^{2y+\sin x}|_{(0,-1)} = 2e^{-2}.$$

Then the matrix of partial derivatives of $\mathbf{f}$ at $(0,-1)$ is

$$D\mathbf{f}(0,-1) = \begin{bmatrix} 3 & 0 \\ e^{-2} & 2e^{-2} \end{bmatrix}.$$

**Example 7.3.** *Compute the matrix of partial derivtives of the linear transformation*

$$\mathbf{g}(x,y,z) = (3x+2y, x-y-z, x+5z)$$

*at the point* $(7,2,6)$.

The partial derivatives of $\mathbf{g}$ are all constants. (Check this!) Thus, the partial derivatives are independent of the point $(7, 2, 6)$. For *any* point $(a, b, c)$, the matrix $D\mathbf{g}(a, b, c)$ is the same:

$$D\mathbf{g}(a, b, c) = \begin{bmatrix} 3 & 2 & 0 \\ 1 & -1 & -1 \\ 1 & 0 & 5 \end{bmatrix}$$

This example illustrates a general property of linear transformations.

**Proposition 7.1.** Suppose that $T : \mathbf{R}^n \to \mathbf{R}^m$ is the linear transformation defined by

$$T(\mathbf{x}) = M\mathbf{x}$$

for some $m \times n$ matrix $M$. Then

$$DT(\mathbf{x}) = M$$

for all points $\mathbf{x} \in \mathbf{R}^n$.

You will prove this proposition in Exercise 38.

## 7.2 Higher derivatives

One may also view the partial derivatives of a function $\mathbf{f}$ as functions which are defined on a subset of the domain of $\mathbf{f}$. This is analogous to the statement in single-variable calculus that the derivative of $x^2$ is $2x$; this statement makes sense for any value of $x$. When a partial derivative is continuous on some domain, one may try to differentiate it.

**Definition 7.3.** Given $f : \mathcal{D}^n \to \mathbf{R}$ and $\mathbf{a} \in \mathcal{D}^n$, suppose that $\frac{\partial f}{\partial x_j}(\mathbf{x})$ exists in a neighborhood of $\mathbf{a}$ for some $j \in \{1, 2, ...n\}$. Then

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_j}(\mathbf{a} + h\mathbf{e_i}) - \frac{\partial f}{\partial x_j}(\mathbf{a})}{h},$$

provided the limit exists.
The quantity $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$ is called the *second partial derivative of $f$ with respect to $x_i$ and $x_j$ at* $\mathbf{a}$.

If $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$ exists, it can be denoted in several ways:

$$\left(\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)\right)(\mathbf{a}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = f_{x_j x_i}(\mathbf{a}).$$

When $i = j$, the notation $\frac{\partial^2 f}{\partial x_i^2}(\mathbf{a})$ is also used to denote the second partial derivative of $f$ with respect to $x_i$ at the point $\mathbf{a}$. When $i \neq j$, the second partial derivative $f_{x_i x_j}(\mathbf{a})$ is called a *mixed partial derivative*.

**Example 7.4.** *Compute the second partial derivatives of $f(x, y) = 3xy^2$.*

$$f_{xx} = \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(3y^2) = 0 \qquad f_{xy} = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(3y^2) = 6y$$
$$f_{yx} = \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(6xy) = 6y \qquad f_{yy} = \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(6xy) = 6x$$

In Example 7.4, observe that, $f_{xy} = f_{yx}$. This is not a coincidence:

**Theorem 13** (Clairaut's Theorem)**.** Given $f : \mathcal{D}^n \to \mathbf{R}$, suppose that for all $i, j \in \{1, 2, ...n\}$, the mixed partial derivative $f_{x_i x_j}(\mathbf{x})$ exists in a neighborhood of $\mathbf{a}$. If $f_{x_i x_j}(\mathbf{x})$ is continuous at $\mathbf{a}$ for all $i, j \in \{1, 2, ...n\}$, then

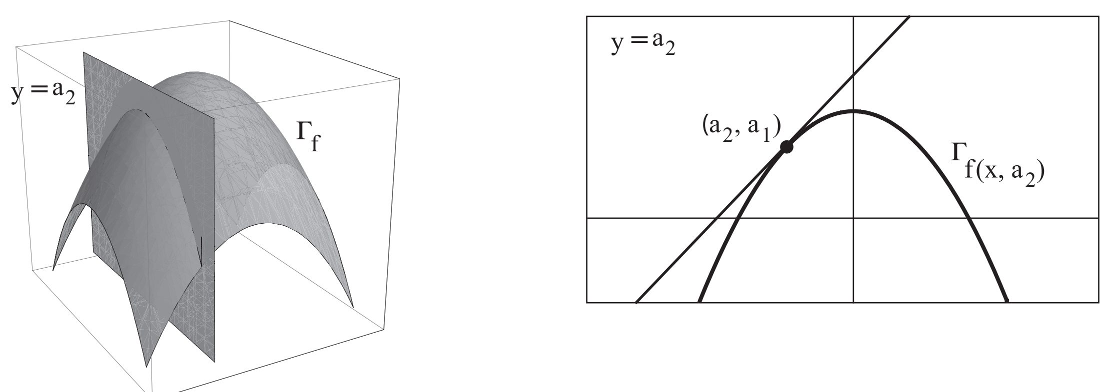
$$f_{x_i x_j}(\mathbf{a}) = f_{x_j x_i}(\mathbf{a}).$$

Clairaut's Theorem says that if a function is sufficiently "well-behaved", then the value of a mixed partial derivative is independent of the order in which the differentiation is performed. The proof of this theorem may be found in Section 15.1 in the Appendix.

## 7.3 Interpreting partial derivatives geometrically

If $f : \mathcal{D}^1 \to \mathbf{R}$ is differentiable at $a$, then the slope of the tangent line to $\Gamma_f$ at $(a, f(a))$ is $f'(a)$. Partial derivatives of multivariable functions may similarly be interpreted as slopes.

Suppose $f : \mathcal{D}^2 \to \mathbf{R}$ is a function whose partial derivatives exist everywhere, and let $(a_1, a_2) \in \mathcal{D}^2$. The graph of $f$ is a surface, and the intersection of $\Gamma_f$ with the plane $y = a_2$ is the graph of the single variable function $f(x, a_2)$.



The slope of the tangent line to this curve at the point $(a_1, a_2)$ is the derivative of the single variable function $f(x, a_2)$ evaluated at the point $x = a_1$.

On the other hand, if we fix $x = a_1$ and let $y$ vary, we see that the graph of the function $f(a_1, y)$ is the intersection of $\Gamma_f$ with the plane $x = a_1$. The tangent line to this curve at the point $\big(a_1, a_2, f(a_1, a_2)\big)$ has slope $\frac{\partial f}{\partial y}(a_1, a_2)$.

**Example 7.5.** *The graph of $f(x, y) = 1 - x^2 - y^2$ intersects the plane $y = \frac{1}{2}$ in a parabola. Find an equation for the tangent line to this parabola at the point $(\frac{-1}{2}, \frac{1}{2}, \frac{1}{2})$.*
The figure on the previous page illustrates this example. The intersection of $y = \frac{1}{2}$ with $\Gamma_f$ is the graph of the function

$$f(x, \frac{1}{2}) = 1 - x^2 - \left(\frac{1}{2}\right)^2 = \frac{3}{4} - x^2.$$

Equivalently, this intersection is the image of the parameterized curve

$$\mathbf{g}(t) = \begin{bmatrix} t \\ 1/2 \\ \frac{3}{4} - t^2 \end{bmatrix}.$$

If we compute the tangent line to this curve at the point $(\frac{-1}{2}, \frac{1}{2}, \frac{1}{2})$, we get

$$\begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ for } t \in \mathbf{R}.$$

On the other hand, the partial derivative of $f$ with respect to $x$ at $(\frac{-1}{2}, \frac{1}{2})$ is

$$f_x(\frac{-1}{2}, \frac{1}{2}) = -2x|_{\frac{-1}{2}} = 1.$$

This implies that when $y$ is kept constant, $z$ changes at the same rate as $x$; you should compare this to the direction vector of the tangent line above.
We could similarly study the intersection between $\Gamma_f$ and the plane $x = \frac{-1}{2}$. In this case, the tangent line to the intersection at $(\frac{-1}{2}, \frac{1}{2}, \frac{1}{2})$ is

$$\begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \end{bmatrix} + s \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \text{ for } s \in \mathbf{R}.$$

(Check this!)
The plane which contains both these lines is called the *tangent plane to $\Gamma_f$ at $(\frac{-1}{2}, \frac{1}{2}, \frac{1}{2})$*. We'll study tangent planes again in Section 10.

Just as we can interpret partial derivatives as slopes of certain lines, we can also interpret the second partial derivatives of $f : \mathcal{D}^2 \to \mathbf{R}$ geometrically. Since $\frac{\partial f}{\partial x}(a_1, a_2)$ describes the slope of a tangent line in the plane $y = a_2$, the second partial derivatives $f_{xx}$ and $f_{xy}$ describe how this slope varies as $x$ and $y$ change. This idea is explored further in the Worked Examples.

## 7.4 Differentiability

In single variable calculus, the derivative of a differentiable function can be used to estimate how a small change in $x$ affects the value of $f(x)$:

$$f(x) - f(a) \approx f'(a)\big(x - a\big).$$

In this case, the difference in the codomain is approximately equal to the difference in the domain times the derivative.

If $\mathbf{f}$ is a multivariable function, the matrix of partial derivatives $D\mathbf{f}(\mathbf{a})$ plays the same role as the the derivative $f'(a)$ in the single variable case:

$$\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) \approx D\mathbf{f}(\mathbf{a})\big(\mathbf{x} - \mathbf{a}\big).$$

Again, the difference in the codomain is approximately equal to the difference in the domain multiplied by the derivative.

In order to describe how good this approximation is, we introduce the notation $\mathbf{R}(\mathbf{x})$ to represent the discrepancy between the actual difference and the estimate coming from the matrix of partial derivatives:

$$\mathbf{R}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

The vector $\mathbf{R}(\mathbf{x})$ is the error associated to the approximation provided by the matrix of partial derivatives. We say that $\mathbf{f}$ is *differentiable at* $\mathbf{a}$ if this error shrinks to $\mathbf{0}$ quickly enough as $\mathbf{x}$ approaches $\mathbf{a}$. The next definition gives a precise formulation of what "quickly enough" means.

**Definition 7.4.** Given $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$, suppose that $\mathcal{D}^n$ contains a neighborhood of the point $\mathbf{a}$. If

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{\mathbf{R}(\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} = \mathbf{0},$$

then $\mathbf{f}$ is *differentiable at* $\mathbf{a}$
The function $\mathbf{f}$ is *differentiable* if it is differentiable at every point in its domain.

Observe that when $f : \mathcal{D}^1 \to \mathbf{R}$ is a single variable function, Definition 7.4 reduces to the statement

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = Df(a).$$

The existence of this limit is the familiar criterion for differentiability for a single variable function.

The limit in Definition 7.4 may be difficult to compute, and the next theorem provides conditions for establishing differentiability that are often easier to verify.

**Theorem 14.** Suppose that the following statements are true:

- there exists some $\epsilon > 0$ such that for all $i \in \{1, 2, ...m\}$ and $j \in \{1, 2, ...n\}$,

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}) \quad \text{exists for all } \mathbf{x} \text{ satisfying} \quad ||\mathbf{x} - \mathbf{a}|| < \epsilon;$$

- for all $i \in \{1, 2, ...m\}$ and $j \in \{1, 2, ...n\}$, the partial derivative $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ is continuous at $\mathbf{a}$.

Then $\mathbf{f}$ is differentiable at $\mathbf{a}$.

It is often easy to verify that the partial derivatives of a function are continuous at a point $\mathbf{a}$ and that they are defined on an open ball containing $\mathbf{a}$, so Theorem 14 is a useful way to establish that a function is differentiable at a given point. The proof of this theorem may be found in Section 15.2 in the Appendix.

We'll return to the idea of approximating a function using its matrix of partial derivatives in Section 11.

**Example 7.6.** *Let $f(x, y) = xy$ for $(x, y) \in \mathbf{R}^2$. Show that $f$ is differentiable at $(2, 1)$.*
We first compute the matrix of partial derivatives $Df(2, 1)$. Since $f$ is a polynomial, the partial derivatives exist and are continuous at every point in $\mathbf{R}^2$:

$$Df(2, 1) = [y \;\; x]_{(2,1)} = [1 \;\; 2].$$

Theorem 14 therefore implies that $\mathbf{f}$ is differentiable at $(2, 1)$. This completes the problem, but in order to illustrate Definition 7.4, we'll solve the same problem using the definition.
Consider the remainder function $R(x, y)$:

$$\begin{aligned} R(x, y) &= f(x, y) - f(2, 1) - Df(2, 1) \begin{bmatrix} x - 2 \\ y - 1 \end{bmatrix} \\ &= f(x, y) - f(2, 1) - [1 \;\; 2] \begin{bmatrix} x - 2 \\ y - 1 \end{bmatrix} \\ &= xy - 2 - (x - 2) - 2(y - 1) \\ &= xy - 2 - x + 2 - 2y + 2 \\ &= xy - x - 2y + 2 \end{aligned}$$

To prove that $f$ is differentiable at $(2, 1)$, we must show that

$$\lim_{(x,y)\to(2,1)} \frac{R(x, y)}{||(x, y) - (2, 1)||} = 0.$$

Therefore, we compute the limit

$$\begin{aligned}
\lim_{(x,y)\to(2,1)} \frac{|R(x,y)|}{||(x,y)-(2,1)||} &= \lim_{(x,y)\to(2,1)} \frac{|xy-x-2y+2|}{\sqrt{(x-2)^2+(y-1)^2}} \\
&\leq \lim_{(x,y)\to(2,1)} \frac{|y(x-2)+(2-x)|}{|x-2|} \\
&= \lim_{(x,y)\to(2,1)} |y-1| \\
&= 0
\end{aligned}$$

Since the limit of the ratio vanishes, this proves that $f$ is differentiable at $(2,1)$.

The matrix of partial derivatives of $\mathbf{f}$ at $\mathbf{a}$ defines a linear transformation, which is sometimes called the *(total) derivative* of $\mathbf{f}$ at $\mathbf{a}$.

**Example 7.7.** *Let $f(x,y)=(x^2y,\sin(xy),e^{x^2})$. Compute the total derivative of $\mathbf{f}$ at $(2,0)$.*
Denote the coordinate functions of $\mathbf{f}$ by $f_i$:

$$\begin{aligned}
f_1(x,y) &= x^2y \\
f_2(x,y) &= \sin(xy) \\
f_3(x,y) &= e^{x^2}.
\end{aligned}$$

You may check that all of the partial derivatives are continuous functions. (Try this!) It therefore follows from Theorem 14 that $\mathbf{f}$ is differentiable at every point in $\mathbf{R}^2$.
We find the matrix $D\mathbf{f}(2,0)$ by evaluating the partial derivatives at the point $(2,0)$:

$$D\mathbf{f}(2,0) = \left[\begin{array}{cc} 2xy & x^2 \\ y\cos(xy) & x\cos(xy) \\ 2xe^{x^2} & 0 \end{array}\right]_{(2,0)} = \left[\begin{array}{cc} 0 & 4 \\ 0 & 2 \\ 4e^4 & 0 \end{array}\right]. \tag{7.1}$$

Thus, the total derivative $D\mathbf{f}(2,0)$ is the linear map from $\mathbf{R}^2$ to $\mathbf{R}^3$ defined by

$$D\mathbf{f}(2,0)(x,y) = \left[\begin{array}{cc} 0 & 4 \\ 0 & 2 \\ 4e^4 & 0 \end{array}\right]\left[\begin{array}{c} x \\ y \end{array}\right].$$

As in the single variable case, a function $\mathbf{f}:\mathcal{D}^n\to\mathbf{R}^m$ gives rise to the *total derivative function*, $D\mathbf{f}$, whose domain is the set of points where $\mathbf{f}$ is differentiable. When $\mathbf{f}$ is a differentiable function, $D\mathbf{f}(\mathbf{a})$ is a linear transformation and $D\mathbf{f}$ is a function which assigns a linear map to each point in its domain.

## 7.5 Worked Examples

**Example 7.8.** If $g(x, y) = xy^2$, then $\frac{\partial^2 g}{\partial y \partial x}(0, -0.5) = -1$. In this example, we'll interpret the mixed partial derivative geometrically.

Begin by considering the partial derivative with respect to $x$ at an arbitrary point $(a, b)$:

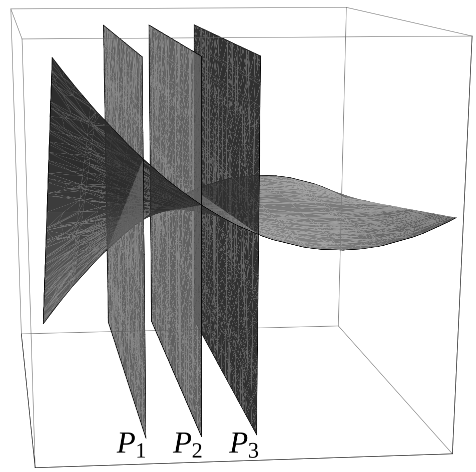
$$\frac{\partial g}{\partial x}(a, b) = y^2|_{(a,b)} = b^2.$$

This tells us that $\Gamma_g$ intersects the plane $y = b$ in a curve whose tangent line has slope $b^2$. Furthermore, since the slope doesn't depend on $x$, we see that this intersection is a line. In particular, $\Gamma_g$ intersects the plane $y = -0.5$ in a line with slope $0.25$.

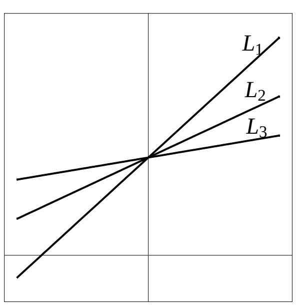
Now, let the value of $y$ vary and study what happens to $\frac{\partial g}{\partial x}$. As $b$ increases near $-0.5$, $b^2$ will decrease. For example, consider the three planes shown on the right:

$$P_1 \text{ is } y = -0.7 \text{ and } \frac{\partial g}{\partial x}(0, -.7) = 0.49$$
$$P_2 \text{ is } y = -0.5 \text{ and } \frac{\partial g}{\partial x}(0, -.5) = 0.25$$
$$P_3 \text{ is } y = -0.3 \text{ and } \frac{\partial g}{\partial x}(0, -.3) = 0.09.$$

The slope of the line of intersection decreases as $b$ increases.




This figure below uses one plane to show the lines of intersection between $\Gamma_g$ and the $P_i$.



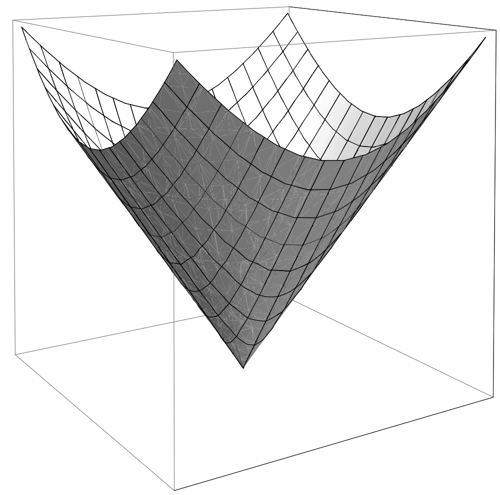

Since the slopes of these lines decrease as $b$ increases from $-0.7$ to $-0.3$, we would expect the mixed partial derivative $\frac{\partial^2 g}{\partial y \partial x}(a, b) = -1$ to be negative for

$$-0.7 \leq b \leq -0.3$$

and for all $a$.

This example began with the statement $\frac{\partial^2 g}{\partial y \partial x}(0, -0.5) = -1$, and we can interpret this as a description of the rate of change of the slope of the intersection between $\Gamma_g$ and planes $y = b$ as $b$ increases near $-0.5$.

**Example 7.9.** *Show that the function $f(\mathbf{x}) = ||\mathbf{x}||$ is not differentiable at $\mathbf{0}$.*



Suppose that $f$ were differentiable at $\mathbf{0}$. In this case, the matrix of partial derivatives of $f$ at $\mathbf{0}$ satisfies

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{||\mathbf{x}|| - Df(\mathbf{0})(\mathbf{x})}{||\mathbf{x}||} = 0.$$

Letting $\mathbf{x} = h\mathbf{e_i}$, this becomes

$$\lim_{h\to 0} \frac{h - Df(\mathbf{0})(h\mathbf{e_i})}{h} = 0.$$

Since multiplication by $Df(\mathbf{0})$ is a linear transformation, we can factor out the scalar $h$:

$$\lim_{h\to 0} \frac{h - hDf(\mathbf{0})(\mathbf{e_i})}{h} = \lim_{h\to 0}(1 - Df(\mathbf{0})(\mathbf{e_i})) = 0.$$

This implies that $Df(\mathbf{0})(\mathbf{e_i}) = 1$. On the other hand, repeating the same computation with $\mathbf{x} = -h\mathbf{e_i}$ implies that $Df(\mathbf{0})(\mathbf{e_i}) = -1$. Since we have arrived at a contradiction, the assumption that $f$ is differentiable at $\mathbf{0}$ must be false.

When $n = 1$, this proof recovers the familiar fact that the absolute value function is not differentiable at $0$.

## 7.6 Exercises

In Exercises 1 through 4, find the partial derivatives of the given function by explicitly evaluating the appropriate limits.

1. $f(x, y) = xy^2$
2. $f(x, y, z) = x\cos z + y$
3. $f(t, s) = \frac{e^{t+1}}{s}$
4. $g(s, t) = e^{st}$

In Exercises 5 through 13, compute the matrix of partial derivatives $D\mathbf{f}(\mathbf{a})$ of the given function at the indicated point.

5. $\mathbf{f}(x, y) = (3x + 2y, 7x - 11y)$, $\mathbf{a} = (2, 0)$
6. $\mathbf{f}(x, y) = (\sin(xy), \cos(x + y))$, $\mathbf{a} = (0, \pi)$
7. $\mathbf{f}(x, y) = (e^{\cos x \sin y}, 2x + \sin y^2)$, $\mathbf{a} = (0, 0)$
8. $\mathbf{f}(x, y, z) = (\sqrt{xy}, \ln(x + y), x^2z^3, x)$, $\mathbf{a} = (1, 4, -1)$
9. $\mathbf{f}(w, x, y, z) = (1, z + w, \frac{1}{x^2+y^2})$, $\mathbf{a} = (0, 3, 4, 0)$
10. $\mathbf{f}(x, y) = \frac{e^y}{1-xe^y}$, $\mathbf{a} = (0, 2)$
11. $\mathbf{f}(s, t) = (3\ln\frac{s}{t}, 2^s)$, $\mathbf{a} = (1, 1)$
12. $\mathbf{f}(x, y) = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$, $\mathbf{a} = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$
13. $\mathbf{f}(x, y) = (x^2, xy, y^2)$, $\mathbf{a} = (2, 1)$

In Exercises 14 through 19, compute the second partial derivatives and the mixed partial derivatives of the given function at the indicated point. Are the mixed partials equal?

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

15. $f(s,t) = s^{3/2}(s + 5\sin t)$, $(4,\pi)$

16. $f(x,y,z) = 10xyz$, $(1,0,2)$

17. $f(u,v,w) = \frac{u^2v+3w}{\sqrt{w}}$, $(-2,5,3)$

18. $f(x,y) = \frac{\sin x}{y^2}$, $(\frac{\pi}{3},1)$

19. $f(x,y) = xy^3e^{2x}$, $(\frac{1}{2},3)$

20. Let $f(x,y,z) = \frac{\cos x^2 y}{\sqrt{y^2+z^2}}$. Find $f_{xxx}$, $f_{zyx}$ and $f_{zxy}$.

21. Let $f(x,y) = \cos(y^2+1)\ln x$. Find $f_{xyy}$, $f_{yyx}$ and $f_{xyx}$.

22. Let $F(x,y,z) = xz\sin 2y + ye^z$.
    (a) Find all the mixed partial derivatives of $F$ and verify that Clairaut's Theorem holds.
    (b) Is $F_{yxz} = F_{yzx}$? Could you have known this without explicitly performing the calculation?
    (c) Is $F_{xyy} = F_{yzx}$?
    (d) Is $F_{yzy} = F_{zyy}$?

23. Let $F(x,y,z) = 6x^2y - yz^2 + xz + 4y^3 + 2z^2 - xyz$.
    (a) Find all the mixed partial derivatives of $F$ and verify that Clairaut's Theorem holds.
    (b) Is $F_{zyx} = F_{yzx}$? Could you have known this without explicitly performing the calculation?
    (c) Is $F_{xyx} = F_{yxx}$?
    (d) Is $F_{yzy} = F_{zzy}$?

24. Verify the equality of the mixed partial derivatives for $f(x,y,z) = \ln(x+2y+z^2)$.

25. For what real numbers $a$ and $b$ does the function $T(s,t) = e^{as}\sin(bt)$ satisfy the partial differential equation
$$T_{ss} - 4T_{tt} = 0\,?$$

26. For what real numbers $a$ and $b$, and $c$ does the function $f(x,y) = ax^2 + bxy + cy^2$ satisfy the partial differential equation
$$f_{xx} + f_{yy} = 0\,?$$

27. Let
$$f(x,y) = \begin{cases} \frac{x^3+xy^2}{x+y} & \text{if } x \neq -y \\ 0 & \text{if } x = -y. \end{cases}$$

(a) Find $f_x(0,0)$ and $f_y(0,0)$ by explicitly evaluating the appropriate limits.

(b) Compute the mixed partial derivatives at the point $(0,0)$. Is $f_{xy}(0,0) = f_{yx}(0,0)$? Reconcile your answer with Clairaut's Theorem.

28. Let $k$ be a real number. The partial differential equation

$$\frac{\partial f}{\partial t} - k\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) = 0$$

is known as the *heat equation*.

(a) Verify that the functions

$$\begin{aligned} f_1(x,y,t) &= e^{-kt}(\sin x), \quad \text{and} \\ f_2(x,y,t) &= e^{-kt}(\cos x + \cos y) \end{aligned}$$

both satisfy the heat equation.

(b) Are there any real numbers $a$ and $b$ which make $f(x,y,t) = x^a t^b$ satisfy the heat equation?

29. Let $S$ and $P$ be the sphere and plane defined, respectively, by

$$\begin{aligned} S &= \{(x,y,z) \in \mathbf{R}^3 \,|\, x^2 + y^2 + z^2 = 6\}, \\ P &= \{(x,y,z) \in \mathbf{R}^3 \,|\, x = 2\}. \end{aligned}$$

Find the tangent line to the curve where $S$ and $P$ intersect at the point $(2,1,1)$.

30. Let $f : \mathbf{R}^2 \to \mathbf{R}$ be defined by

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

(a) Show that the partial derivative $f_x(\mathbf{0})$ and $f_y(\mathbf{0})$ exist.

(b) Prove that $f$ is not continuous at $\mathbf{0}$.

31. Let $f : \mathbf{R}^2 \to \mathbf{R}$ be defined by

$$f(x,y) = \begin{cases} 2 & \text{if } xy \geq 0 \\ 0 & \text{if } xy < 0. \end{cases}$$

(a) Show that the partial derivatives $f_x(\mathbf{0})$ and $f_y(\mathbf{0})$ exist.

(b) Prove that f is not differentiable at $\mathbf{0}$.

32. Let

$$f(x,y) = \begin{cases} \frac{x^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

(a) Is $f$ continuous at $(0,0)$?
(b) Do $f_x(0,0)$ and $f_y(0,0)$ exist?

33. Let

$$F(x,y) = \begin{cases} 0 & \text{if } (x,y) = (0,0), \\ \frac{\sin x^2}{x} & \text{if } (x,y) \text{ is on the } x\text{-axis but not (0,0)} \\ \frac{\sin y^2}{y} & \text{if } (x,y) \text{ is on the } y\text{-axis but not (0,0)} \\ \frac{\sin x^2}{x} + \frac{\sin y^2}{y} & \text{otherwise} \end{cases}$$

(a) Prove $F$ is continuous at $(0,0)$.
(Hint: Use $\lim\limits_{t \to 0} \dfrac{\sin t^2}{t} = 0$. )

(b) Find $F_x(0,0)$ and $F_y(0,0)$.
(Hint: Use $\lim\limits_{t \to 0} \dfrac{\sin t^2}{t^2} = 1$. )

(c) Find $F_x(x,0)$, $F_y(x,0)$, $F_x(0,y)$, and $F_y(0,y)$.
(d) Find $F_{xy}(0,0)$ and $F_{yx}(0,0)$. Are they equal?

34. If $\mathbf{f}(x,y) = (ye^{2x}, \sin xy)$ and $\mathbf{a} = (1, \frac{\pi}{2})$, show that $D\mathbf{f}(\mathbf{a})(x,y) = (\pi e^2 x + e^2 y, 0)$.

35. If $\mathbf{f}(x,y) = (x^2 + y^2, 2y)$ and $\mathbf{a} = (0,0)$, show that $D\mathbf{f}(\mathbf{a})(x,y) = (0, 2y)$.

36. If $\mathbf{f}(x,y) = (\frac{1}{x}, \frac{1}{y^2})$ and $\mathbf{a} = (1,1)$, show that $D\mathbf{f}(\mathbf{a})(x,y) = (-x, -2y)$.

37. Suppose $D\mathbf{f}(\mathbf{a})(x,y) = (2x + y, x - 3y)$. Find $D\mathbf{f}(\mathbf{a})$.

38. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation such that

$$T(\mathbf{x}) = M\mathbf{x}$$

for some $m \times n$ matrix $M$. Show that for every $\mathbf{a} \in \mathbf{R}^n$, $DT(\mathbf{a}) = M$.

39. Show that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a} \in \mathbb{R}^n$, then the partial derivatives $\frac{\partial f}{\partial x_i}(\mathbf{a})$ exist.

40. Suppose that $F$, $T$, and $S$ are linear transformations such that $F = T \circ S$. Show that for every $\mathbf{a}$,

$$DF(\mathbf{a}) = DT(\mathbf{a})DS(\mathbf{a}).$$

41. For each of the following statements, decide if the statement is always true or sometimes false. Justify your answer.

(a) If all the partial derivatives of $F$ exist at $\mathbf{a}$, then $F$ is differentiable at $\mathbf{a}$.
(b) There exists continuous functions which are not differentiable.
(c) The function $f(x,y) = \frac{y}{x} + \frac{x}{y}$ is differentiable.
(d) If $f$ is a function from $\mathbf{R}^4$ to $\mathbf{R}^3$ and $\mathbf{a} \in \mathbf{R}^4$, then $Df(\mathbf{a})$ is a $3 \times 4$ matrix.
(e) Every differentiable function is continuous

# 8 The Chain Rule

The Chain Rule describes the derivative of a composition of functions as a composition of derivatives. The Chain Rule is one of the most important tools in differential calculus.

To begin with, recall the single variable Chain Rule:

**Theorem 15** (The Chain Rule for single variable functions)**.** Suppose that $f : \mathcal{D}^1 \to \mathbf{R}$ is differentiable at $a$ and that $g : \mathcal{E}^1 \to \mathbf{R}$ is differentiable at $f(a)$. Then

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

**Example 8.1.** *Let $g(x) = \sin x + x$ and let $f(t) = t^2$. Compute $\frac{d}{dt}(g \circ f)(\sqrt{\pi})$.*
In order to apply the Chain Rule, first compute the derivatives of $f$ and $g$:

$$\begin{aligned} g'(x) &= \cos x + 1 \\ f'(t) &= 2t. \end{aligned}$$

Next, evaluate theses derivatives:

$$\begin{aligned} g'(f(\sqrt{\pi})) &= g'(\pi) = \cos \pi + 1 = -1 + 1 = 0 \\ f'(\sqrt{\pi}) &= 2\sqrt{\pi}. \end{aligned}$$

Finally, apply the Chain Rule:

$$\frac{d}{dt}(g \circ f)(\sqrt{\pi}) = g'(f(\sqrt{\pi}))f'(\sqrt{\pi}) = 0.$$

## 8.1 The multivariable Chain Rule

The multivariable Chain Rule is similar to the single variable Chain Rule. We state the theorem in this section, and you may find the proof in Section 15.3 in the Appendix.

**Theorem 16** (The Chain Rule)**.** Suppose that $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ is differentiable at $\mathbf{a} \in \mathcal{D}^n$ and that $\mathbf{g} : \mathcal{E}^m \to \mathbf{R}^p$ is differentiable at $\mathbf{f}(\mathbf{a}) \in \mathcal{E}^m$. Then the composition $\mathbf{g} \circ \mathbf{f}$ is differentiable at $\mathbf{a}$, and

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{a}) = D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a}).$$

Observe that the dimensions of these matrices are compatible for multiplication: the matrix of partial derivatives of $\mathbf{g}$ is a $p \times m$ matrix, and the matrix of partial derivatives of $\mathbf{f}$ is an $m \times n$ matrix. Their product is the $p \times n$ matrix of partial derivatives of $\mathbf{g} \circ \mathbf{f} : \mathbf{R}^n \to \mathbf{R}^p$.

Because multiplication of matrices corresponds to composition of linear transformations, you can think of the Chain Rule as a statement simultaneously about multiplying matrices of partial derivatives and about composing total derivatives.

**Example 8.2.** *Let $f(s,t) = 2s + st^3$, and let $\mathbf{g}(y) = (e^y, 2y-1)$. Find the matrix of partial derivatives $D(\mathbf{g} \circ f)(1,0)$.*
We can check that the partial derivatives of $f$ and g are continuous everywhere, so both $f$ and g are differentiable at every point in their respective domains. Thus, Theorem 16 implies that g $\circ f$ is differentiable at $(1,0)$.
The Chain Rule implies that

$$D(\mathbf{g} \circ f)(1,0) = D\mathbf{g}(f(1,0))Df(1,0).$$

Substituting in $f(1,0) = 2$ yields

$$D(\mathbf{g} \circ f)(1,0) = D\mathbf{g}(2)Df(1,0).$$

Next, we compute the matrices of partial derivatives of $f$ and g:

$$Df(1,0) = [2+t^3, 3st^2]_{(1,0)} = [2,0]$$
$$D(\mathbf{g})(2) = \begin{bmatrix} e^y \\ 2 \end{bmatrix}_2 = \begin{bmatrix} e^2 \\ 2 \end{bmatrix}$$

Putting these together, we see that

$$D(\mathbf{g} \circ f)(1,0) = \begin{bmatrix} e^2 \\ 2 \end{bmatrix} [2,0] = \begin{bmatrix} 2e^2 & 0 \\ 4 & 0 \end{bmatrix}.$$

Thus, the associated total derivative of the composition is the linear map from $\mathbf{R}^2$ to $\mathbf{R}^2$ defined by

$$D(\mathbf{g} \circ f)(1,0)(x,y) = \begin{bmatrix} 2e^2 & 0 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2e^2x, 4x)$$

for all $(x,y) \in \mathbf{R}^2$.

Of course, in this example we could have first computed the composition

$$(\mathbf{g} \circ f)(s,t) = (e^{2s+st^3}, 4s + 2st^3 - 1).$$

In this case, we could compute $D(\mathbf{g} \circ f)(1,0)$ without using the Chain Rule, and you can check that we'd get the same answer as in the calculation above. However, the next example illustrates that sometimes the Chain Rule is the only recourse.

**Example 8.3.** *Suppose that* $\mathbf{f} : \mathbf{R}^2 \to \mathbf{R}^2$ *satisfies*

$$D\mathbf{f}(1,-1) = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{f}(1,-1) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

*If* $\mathbf{g} : \mathbf{R}^2 \to \mathbf{R}^3$ *is defined by* $\mathbf{g}(x,y) = (xy, x+xy, 2x)$*, find* $D(\mathbf{g} \circ \mathbf{f})(1,-1)$.
Although we cannot write down a formula for $\mathbf{g} \circ \mathbf{f}$, the Chain Rule implies that

$$D(\mathbf{g} \circ \mathbf{f})(1,-1) = D\mathbf{g}(\mathbf{f}(1,-1))D\mathbf{f}(1,-1).$$

First compute $D\mathbf{g}(\mathbf{f}(1,-1))$:

$$D\mathbf{g}(\mathbf{f}(1,-1)) = D\mathbf{g}(1,3) = \begin{bmatrix} y & x \\ 1+y & x \\ 2 & 0 \end{bmatrix}_{(1,3)} = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 2 & 0 \end{bmatrix}.$$

Thus,

$$\begin{aligned} D(\mathbf{g} \circ \mathbf{f})(1,-1) &= D\mathbf{g}(\mathbf{f}(1,-1))D\mathbf{f}(1,-1) \\ &= \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & -1 \\ 6 & -1 \\ 2 & 0 \end{bmatrix} \end{aligned}$$

You can check that $D(\mathbf{g} \circ \mathbf{f})(1,-1)$ has the expected dimensions: since the composition $\mathbf{g} \circ f$ is a function from $\mathbf{R}^2$ to $\mathbf{R}^3$, its derivative should be represented by a $3 \times 2$ matrix.

### 8.1.1 The matrix of partial derivatives of a composition

The Chain Rule describes the total derivative of the composition $\mathbf{g} \circ \mathbf{f}$, but in some circumstances you may only be interested in the derivative of a particular coordinate function. Alternatively, you might want to understand how the composition is affected by changing a single variable in the domain of $\mathbf{f}$. In order to answer these kinds of questions, it's helpful to take a closer look at the entries of the matrix of partial derivatives of a composition.

Suppose that $\mathbf{f}$ and $\mathbf{g}$ are functions satisfying the hypotheses of Theorem 16. Let's describe $\mathbf{f}$, $\mathbf{g}$ and $\mathbf{g} \circ \mathbf{f}$ in terms of their coordinate functions:

$$\begin{aligned} \mathbf{f}(x_1, x_2, ...x_n) &= (f_1(x_1, x_2, ...x_n), f_2(x_1, x_2, ...x_n), ...f_m(x_1, x_2, ...x_n)) \\ \mathbf{g}(y_1, y_2, ...y_m) &= (g_1(y_1, y_2, ...y_m), g_2(y_1, y_2, ...y_m), ...g_p(y_1, y_2, ...y_m)) \\ (\mathbf{g} \circ \mathbf{f})(x_1, x_2, ...x_n) &= (h_1(x_1, x_2, ...x_n), h_2(x_1, x_2, ...x_n), ...h_p(x_1, x_2, ...x_n)). \end{aligned}$$

Then the Chain Rule reads

$$D(\mathbf{g}\circ\mathbf{f})(\mathbf{a}) = \begin{bmatrix} \frac{\partial g_1}{\partial y_1}(\mathbf{f}(\mathbf{a})) & \frac{\partial g_1}{\partial y_2}(\mathbf{f}(\mathbf{a})) & \dots & \frac{\partial g_1}{\partial y_m}(\mathbf{f}(\mathbf{a})) \\ \frac{\partial g_2}{\partial y_1}(\mathbf{f}(\mathbf{a})) & \frac{\partial g_2}{\partial y_2}(\mathbf{f}(\mathbf{a})) & \dots & \frac{\partial g_2}{\partial y_m}(\mathbf{f}(\mathbf{a})) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial y_1}(\mathbf{f}(\mathbf{a})) & \dots & \dots & \frac{\partial g_p}{\partial y_m}(\mathbf{f}(\mathbf{a})) \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \dots & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{bmatrix}.$$

The $(i,j)$ entry of the matrix of partial derivatives of $\mathbf{g}\circ\mathbf{f}$ is the partial derivative of the $i^{th}$ coordinate function of the composition with respect to the $j^{th}$ variable in the domain of $\mathbf{f}$. The coordinate expansion above gives us a precise description of this entry:

$$(D(\mathbf{g}\circ\mathbf{f})(\mathbf{a}))_{ij} = \frac{\partial h_i}{\partial x_j}(\mathbf{a}) = \begin{bmatrix} \frac{\partial g_i}{\partial y_1}(\mathbf{f}(\mathbf{a})) \\ \frac{\partial g_i}{\partial y_2}(\mathbf{f}(\mathbf{a})) \\ \vdots \\ \frac{\partial g_i}{\partial y_m}(\mathbf{f}(\mathbf{a})) \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_j}(\mathbf{a}) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(\mathbf{a}) \end{bmatrix} = \sum_{k=1}^{m} \frac{\partial g_i}{\partial y_k}(\mathbf{f}(\mathbf{a}))\frac{\partial f_k}{\partial x_j}(\mathbf{a}).$$

**Example 8.4.** *Consider the functions* $\mathbf{f}$ *and* $\mathbf{g}$ *from Example 8.3. Denote their composition by* $\mathbf{h} = \mathbf{g}\circ\mathbf{f}$, *and write* $\mathbf{h}(x,y) = (h_1(x,y), h_2(x,y), h_3(x,y))$. *Find* $Dh_2(1,-1)$.
In Example 8.3 we computed the matrix of partial derivatives of $\mathbf{h}$ at $(1,-1)$:

$$D(\mathbf{g}\circ\mathbf{f})(1,-1) = D\mathbf{h}(1,-1) = \begin{bmatrix} 5 & -1 \\ 6 & -1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} Dh_1(1,-1) \\ Dh_2(1,-1) \\ Dh_3(1,-1) \end{bmatrix}.$$

Thus,

$$Dh_2(1,-1) = \begin{bmatrix} \frac{\partial h_2}{\partial x}(1,-1) & \frac{\partial h_2}{\partial y}(1,-1) \end{bmatrix} = \begin{bmatrix} 6 & -1 \end{bmatrix}.$$

### 8.1.2 Tree diagrams

In many Chain Rule problems, the key step is organizing the relationships between the variables, and a *tree diagram* is sometimes a helpful tool.

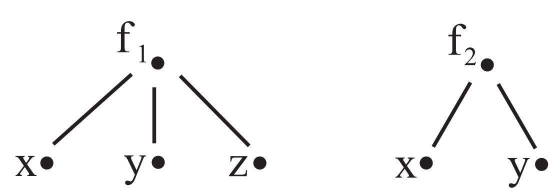
In a tree diagram for the function $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$, each coordinate function is represented by a dot. If $f_i$ depends on the variable $x_j$, then the dot representing $f_i$ is connected to a dot labeled $x_j$ which is placed below it.

For example, consider the function $\mathbf{f}(x,y,z) = (xyz, x+y)$.

If we write

$$f_1(x,y,z) = xyz$$
$$f_2(x,y,z) = x+y,$$

then the figure shows a tree diagram for $\mathbf{f}$.



In a composition $\mathbf{g} \circ \mathbf{f}$, the coordinate functions of $\mathbf{f}$ correspond to variables in the domain of $\mathbf{g}$. In order to draw a tree diagram for a composition, we stack the tree diagrams for each of the functions being composed.

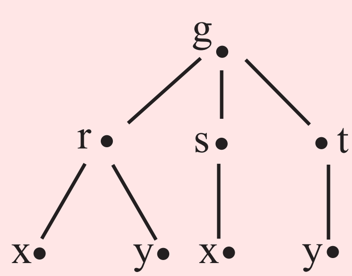
**Example 8.5.** *Given the functions*

$$\mathbf{f}(x, y) = \begin{bmatrix} x + y \\ x \\ y^2 \end{bmatrix} \text{ and } g(r, s, t) = rst,$$

*consider a tree diagram for the composition* $g \circ \mathbf{f}$.

In order to draw a tree diagram for $g \circ \mathbf{f}$, we identify the variables in the domain of $g$ with the coordinate functions of $\mathbf{f}$:

$$\begin{aligned} r(x, y) &= x + y \\ s(x, y) &= x \\ t(x, y) &= y^2 \end{aligned}$$




We can use the computation from Section 8.1.1 to write down an expression for $\frac{\partial g}{\partial x}(2, 3)$:

$$\frac{\partial g}{\partial x}(2, 3) = \frac{\partial g}{\partial r}(r(2, 3), s(2, 3), t(2, 3))\frac{\partial r}{\partial x}(2, 3) + \frac{\partial g}{\partial s}(r(2, 3), s(2, 3), t(2, 3))\frac{ds}{dx}(2, 3).$$

This describes $\frac{\partial g}{\partial x}(2, 3)$ as a sum of two terms, each of which has two factors, and we can see this structure in the tree diagram. The tree diagram has two distinct paths connecting $g$ to $x$, and each path has two edges.

In general, a tree diagram can provide a quick check for missing terms when multiplying matrices. Suppose that $h_i$ is a coordinate function for the composition $\mathbf{g_n} \circ \ldots \circ \mathbf{g_1}$, and suppose that $x_j$ is a variable in the domain of $\mathbf{g_1}$. The number of summands in $\frac{\partial h_i}{\partial x_j}$ is equal to the number of distinct paths connecting $h_i$ to $x_j$ in the tree diagram for $\mathbf{g_n} \circ \ldots \circ \mathbf{g_1}$. Furthermore, the number of edges in each path equals the number of factors in each summand.

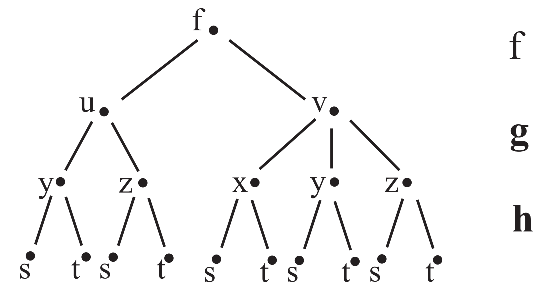
## 8.2 Worked Examples

**Example 8.6.** Let $f(u, v) = uv$ and let

$$\begin{aligned} u(x, y, z) &= yz \\ v(x, y, z) &= 3x + yz. \end{aligned}$$

If $x = st$, $y = s^2t^2$, and $z = s + t$, find the partial derivative of $f$ with respect to $t$ at $(2, -1)$.

We begin by recognizing that $u$ and $v$ can be viewed as the coordinate functions for some $\mathbf{g} : \mathbf{R}^3 \to \mathbf{R}^2$. Similarly, $x, y,$ and $z$ can be viewed as the coordinate functions for some $\mathbf{h} : \mathbf{R}^2 \to \mathbf{R}^3$.




$$\mathbf{g}(x,y,z) = (u(x,y,z,),v(x,y,z))$$
$$\mathbf{h}(s,t) = (x(s,t),y(s,t),z(s,t))$$

If we look at the formulae defining these functions, we get the tree diagram shown above.

Next, compute the matrices of partial derivatives $D\mathbf{h}(2,-1)$, $D\mathbf{g}(\mathbf{h}(2,-1))$, and $Df(\mathbf{g}(\mathbf{h}(2,-1)))$:

$$D\mathbf{h}(2,-1) = \begin{bmatrix} t & s \\ 2st^2 & 2s^2t \\ 1 & 1 \end{bmatrix}_{(2,-1)} = \begin{bmatrix} -1 & 2 \\ 4 & -8 \\ 1 & 1 \end{bmatrix}$$
$$D\mathbf{g}(\mathbf{h}(2,-1)) = \begin{bmatrix} 0 & z & y \\ 3 & z & y \end{bmatrix}_{(-2,4,1)} = \begin{bmatrix} 0 & 1 & 4 \\ 3 & 1 & 4 \end{bmatrix}$$
$$Df(\mathbf{g}(\mathbf{h}(2,-1))) = \begin{bmatrix} v & u \end{bmatrix}_{(4,-2)} = \begin{bmatrix} -2 & 4 \end{bmatrix}$$

Multiplying these together, we see that

$$Df(2,-1) = \begin{bmatrix} -2 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 & 4 \\ 3 & 1 & 4 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 4 & -8 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 16 \end{bmatrix}.$$

Thus, $\frac{\partial f}{\partial t}(2,-1) = 16$.

If we multiply the matrices of partial derivatives before evaluating at $(s,t) = (2,-1)$, we get

$$\begin{aligned}\frac{\partial f}{\partial t}(2,1) = {} & \frac{\partial f}{\partial u}\mathbf{g}(\mathbf{h}(2,1))\frac{\partial u}{\partial y}\mathbf{h}(2,1)\frac{\partial y}{\partial t}(2,1) \\ & + \frac{\partial f}{\partial u}\mathbf{g}(\mathbf{h}(2,1))\frac{\partial u}{\partial z}\mathbf{h}(2,1)\frac{\partial z}{\partial t}(2,1) \\ & + \frac{\partial f}{\partial v}\mathbf{g}(\mathbf{h}(2,1))\frac{\partial v}{\partial x}\mathbf{h}(2,1)\frac{\partial x}{\partial t}(2,1) \\ & + \frac{\partial f}{\partial v}\mathbf{g}(\mathbf{h}(2,1))\frac{\partial v}{\partial y}\mathbf{h}(2,1)\frac{\partial y}{\partial t}(2,1) \\ & + \frac{\partial f}{\partial v}\mathbf{g}(\mathbf{h}(2,1))\frac{\partial v}{\partial z}\mathbf{h}(2,1)\frac{\partial z}{\partial t}(2,1)\end{aligned}$$

This expression has five summands, each with three factors. In the tree diagram, this corresponds to the fact that there are five paths between $f$ and $t$, each with three edges.

## 8.3 Exercises

In Exercises 1 through 6, make a tree diagram for the composition and then use the Chain Rule to compute the derivative at the indicated point. If you find tree diagrams to be helpful, you may continue to use them for the rest of the problems in this chapter.

1. $z = x^2 + 2xy,\ \ x = rs + 3r,\ \ y = \sqrt{r^2 + s^2}$. Compute $\frac{\partial z}{\partial r}(1,0)$ and $\frac{\partial z}{\partial s}(1,0)$.
2. $w = qrs^5,\ \ q = t^2,\ \ r = t^3 + 3,\ \ s = t - 1$. Compute $\frac{dw}{dt}(-2)$.
3. $\mathbf{f}(x,y) = (2x + 3y^2, 2x^3 - 7y),\ \ g(r,s) = \frac{1}{2}r^2 + s^2$. Compute $D(g \circ f)(1,2)$.
4. $q = \tan^{-1}(x^2 + y^2),\ x = \sqrt{w^2 + 1},\ y = z^3 - w,\ w = st,\ z = (t - s)^2$.
   Compute $\frac{\partial q}{\partial s}(2,-1)$ and $\frac{\partial q}{\partial t}(2,-1)$.
5. $h(r,s,t) = r^2 - t^2 + rs,\ \ \mathbf{g}(x) = (x + 1, x^2, \sqrt{x^2 + 9})$. Compute $(h \circ g)'(-1)$.
6. $h(r,s,t) = r^2 - t^2 + rs,\ \ \mathbf{f}(x) = (x + 1, \sqrt{x^2 + 9})$. Compute $D(\mathbf{f} \circ h)(2,0,-3)$.

In Exercises 7 through 11 compute the matrix of partial derivatives of the composition $D(\mathbf{f} \circ \mathbf{g})$ in two ways. First, write out the composition $f \circ g$ and compute $D(\mathbf{f} \circ \mathbf{g})$ directly. Then compute $D(\mathbf{f} \circ \mathbf{g})$ using the Chain Rule.

7. $\mathbf{f}(x,y,z) = (\sin x \cos y + e^z, xy\ln(xyz) + xyz^2),\ \ \mathbf{g}(r,s) = (1/s, 1/r, s^2)$.
8. $\mathbf{f}(s,t) = (\sqrt{s^2 + t^2}, st)\ ,\ \ \mathbf{g}(x) = (\sqrt{x^2 + 1}, \sqrt{x^2 - 1})$, defined for $x > 1$.
9. $f(q,r,s) = q^3 + 2rs,\ \ \mathbf{g}(u) = (u - 2, u^2 + u + 1, \ln(u^2 + 1))$.
10. $\mathbf{f}(u) = (u - 2, u^2 + u + 1, \ln(u^2 + 1)),\ \ g(q,r,s) = q^3 + 2rs$.
11. $\mathbf{f}(t,u) = (17t + 4, 3u^2 + t),\ \ \mathbf{g}(x,y) = (\ln x^2, x^2 y)$.

In Exercise 12 through 15 use the given information to compute the indicated derivative.

12. Let $\mathbf{f} : \mathbf{R}^3 \to \mathbf{R}^2$ be defined by $\mathbf{f}(x,y,z) = (e^{x^2+y^2+z^2} - 2z, x^2 y - y^2 z)$ and $g : \mathbf{R}^2 \to \mathbf{R}$ has $Dg(e^2, 1) = [12, -17]$. Let $h = g \circ \mathbf{f}$. Compute $Dh(1,1,0)$.
13. Suppose $\mathbf{h} : \mathbf{R}^2 \to \mathbf{R}^3$ has $\mathbf{h}(1,2) = (4,5,6)$ and

    $$D\mathbf{h}(1,2) = \begin{bmatrix} 0 & 4 \\ 1 & 2 \\ -3 & 6 \end{bmatrix}.$$

    Let $\mathbf{g} : \mathbf{R}^3 \to \mathbf{R}^4$ be defined by $\mathbf{g}(r,s,t) = (r, s^2, t^3, -t^2 + 4rs)$. Compute $D(\mathbf{g} \circ \mathbf{h})(1,2)$.
14. Suppose $\mathbf{r} : \mathbf{R} \to \mathbf{R}^3$ satisfies $\mathbf{r}(0) = (1,2,5)$ and $\mathbf{r}'(0) = (-1,0,2)$. If $\mathbf{v} : \mathbf{R}^3 \to \mathbf{R}^2$ is defined by $\mathbf{v}(x,y,z) = (xyz, xy + yz + xz)$, find $D(\mathbf{v} \circ \mathbf{r})(0)$.

15. Let $f(r,s,t) = \ln\sqrt{r^2+s^2+t^2}$ and let $\mathbf{h}(x) = (\tan\frac{x}{3}, \cos x, \sin\frac{x}{6})$. Suppose that $\mathbf{g}(\sqrt{3},-1,1) = (1,\frac{3}{2},-1)$ and

$$D\mathbf{g}(\sqrt{3},-1,1) = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 3 \\ -1 & \frac{2}{3} & 0 \end{bmatrix}.$$

Compute $(f \circ \mathbf{g} \circ \mathbf{h})'(\pi)$.

16. (a) Let $\mathbf{f}(u,v) = (u^2v, uv^4, \cos(uv))$. Compute $D\mathbf{f}(3,0)$.
    (b) Let $\mathbf{g}(r,s,t) = (\sin^2(r+s), \exp(2st), t^2+r+3s)$. Compute $D(\mathbf{g}\circ\mathbf{f})(3,0)$
    (c) Let $\mathbf{h}(x,y,z) = (2xy+3yz+xz, \sqrt{x^2+y^2+3z^2}-2)$. Compute $D(\mathbf{h}\circ\mathbf{g}\circ\mathbf{f})(3,0)$.
    (d) Compute $D(\mathbf{h}\circ\mathbf{g}\circ\mathbf{f}\circ\mathbf{h}\circ\mathbf{g}\circ\mathbf{f})(3,0)$.

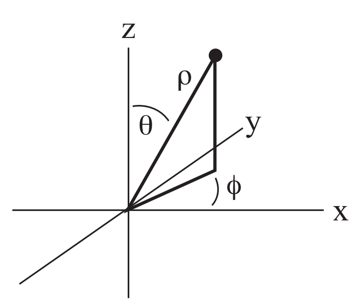
17. Let $\mathbf{f} : \mathbf{R}^3 \to \mathbf{R}^2$ be given by $\mathbf{f}(r,s,t) = (r^3s+t^2, rst)$. Let $\mathbf{g} : \mathbf{R}^2 \to \mathbf{R}^{10}$, and $\mathbf{h} : \mathbf{R}^{10} \to \mathbf{R}^4$ be two differentiable functions. Calculate $D(\mathbf{h}\circ\mathbf{g}\circ\mathbf{f})(0,62,0)$. Why doesn't it matter what **g** and **h** are?

18. Let $\mathbf{f} : \mathbf{R}^2 \to \mathbf{R}^2$ be defined by

$$\mathbf{f}(x,y) = \left((x^2+y^2)^{3/2}, \frac{\ln(x^2+y^2)}{(x^2+y^2)^2}\right)$$

    (a) Compute $D\mathbf{f}(1,5)$. Take the determinant to find out if $D\mathbf{f}(1,5)$ is invertible.
    (b) Show that **f** can be written as a composition of two functions $h : \mathbf{R}^2 \to \mathbf{R}$ and $\mathbf{g} : \mathbf{R} \to \mathbf{R}^2$.
    (c) Use the Chain Rule and linear algebra to explain how we could have known whether or not $D\mathbf{f}(1,5)$ is invertible without actually doing the computation in part (a).

19. The vector $\mathbf{v} \in \mathbf{R}^3$ has *spherical coordinates* $(\rho,\theta,\phi)$ where $\rho = \|\mathbf{v}\|$, the *polar angle* $\theta$ is the angle $\mathbf{v}$ makes with the $z$ axis, and the *azimuthal angle* $\phi$ is the angle that the projection of $\mathbf{v}$ into the $xy$ plane makes with the $x$ axis.




    (a) Compute the matrix of partial derivatives for the spherical coordinate transformation

$$x = \rho\sin\theta\cos\phi$$
$$y = \rho\sin\theta\sin\phi$$
$$z = \rho\cos\theta.$$

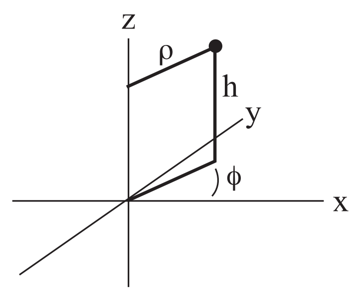
(b) Let $w = f(x, y, z)$. Use the Chain Rule and the result of part (a) to write down the partial derivatives of $w$ with respect to the spherical coordinates $\rho, \theta, \phi$.

(c) Consider the function $T = e^{xy^2} + 3z^3y$. Compute $\frac{\partial T}{\partial \rho}$, $\frac{\partial T}{\partial \theta}$, and $\frac{\partial T}{\partial \phi}$ at the point with spherical coordinates $\rho = 2,\ \theta = \frac{\pi}{4},\ \phi = \frac{3\pi}{2}$.

20. A rectangular box has width $4.0$ cm, length $6.0$ cm, and height $3.0$ cm. At what rate is the volume of the box changing if the width is decreasing at the rate $1.5$ cm/s, the length is increasing at the rate $2.5$ cm/s and the height is increasing at the rate $0.2$ cm/s. At what rate is the surface area changing?

21. Suppose the pressure of a gas is given by $P = \rho T$ where $\rho$ is the density, and $T$ is the temperature. Suppose that $T$ and $\rho$ depend on position according to the functions $T = e^{-(x^2+y^2+z^2)}$, $\rho = e^{-z}$. At what rate is the pressure changing for an object located at position $(2, 3, 1)$ moving with velocity $(1, 1, 1)$?

22. The vector $\mathbf{v} \in \mathbf{R}^3$ has *cylindrical coordinates* $(\rho, \phi, h)$ where the *radius* $\rho$ is the distance to the $z$ axis, the *angular position* $\phi$ is the angle that the projection of $\mathbf{v}$ into the $xy$ plane makes with the $x$ axis, and the *altitude* $h$ is simply the $z$ coordinate of $\mathbf{v}$.




(a) Compute the matrix of partial derivatives for the cylindrical coordinate transformation

$$
\begin{aligned}
x &= \rho \cos \phi \\
y &= \rho \sin \phi \\
z &= h.
\end{aligned}
$$

(b) Let $w = f(x, y, z)$. Use the chain rule and the result of part (a) to write down the partial derivatives of $w$ with respect to the cylindrical coordinates $\rho, \phi, h$.

(c) Consider the function $U = 17 \ln(y^2 + z^2) - 3x^2 - 2y^2$. Compute $\frac{\partial U}{\partial \rho}$, $\frac{\partial U}{\partial \phi}$, and $\frac{\partial U}{\partial h}$ at the point with cylindrical coordinates $\rho = 1,\ \phi = \frac{\pi}{2},\ h = 2$.

Exercises 23 - 25 are related to the proof of the Chain Rule, which may be found in the Appendix, Section 15.3.

23. Prove Lemma 1 from the proof of the Chain Rule.

24. Set $\mathbf{y} = \mathbf{s} - \mathbf{f}(\mathbf{a})$. Use Definition 7.4 from Section 7.4 to show how Equation 15.12 follows from the differentiability of $\mathbf{g}$ at $\mathbf{f}(\mathbf{a})$.

25. In order to complete second limit in the proof of the Chain Rule, we need to deal with the possibility that $\mathbf{s} = 0$ for some arbitrarily small value of $\mathbf{t}$. Define the function $\widetilde{\mathbf{R}} : \mathbf{R}^m \to \mathbf{R}$ by

$$\widetilde{\mathbf{R}}(\mathbf{x}) = \begin{cases} \frac{||\mathbf{R_g}(\mathbf{x})||}{||\mathbf{x}||} & \text{if } \mathbf{x} \neq \mathbf{0} \\ 0 & \text{if } \mathbf{x} = \mathbf{0} \end{cases}$$

(a) Show that $||\mathbf{R_g}(\mathbf{x})|| = ||\mathbf{x}||\widetilde{\mathbf{R}}(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{R}^m$.

(b) For $\mathbf{s}$ and $\mathbf{t}$ as in the the proof, show that

$$\lim_{\mathbf{t}\to\mathbf{0}} ||\widetilde{\mathbf{R}}(\mathbf{s})|| = 0.$$

(c) Use $\widetilde{\mathbf{R}}$ to prove the limit in Equation 15.20, without making any additional assumptions on $\mathbf{s}$.

# 9 Directional Derivatives and the Gradient

The partial derivative $\frac{\partial f}{\partial x_i}(\mathbf{a})$ describes the dependence of the function $f$ on the variable $x_i$ near the point $\mathbf{a}$. You can think of $\frac{\partial f}{\partial x_i}(\mathbf{a})$ as answering the question, "How does $f$ change if $\mathbf{x}$ moves away from $\mathbf{a}$ in the $\mathbf{e_i}$ direction?" In this section we discuss the related question of how $f$ changes if $\mathbf{a}$ moves in an arbitrary direction.

## 9.1 Directional derivatives

**Definition 9.1.** Given $f : \mathcal{D}^n \to \mathbf{R}$, let $\mathbf{a} \in \mathcal{D}^n$. Furthermore, let $\mathbf{v} \in \mathbf{R}^n$ and set $\mathbf{u} = \frac{\mathbf{v}}{||\mathbf{v}||}$ to be the unit vector in the direction of $\mathbf{v}$. The *directional derivative of $f$ at $\mathbf{a}$ in the direction* $\mathbf{v}$ is

$$D_{\mathbf{v}} f(\mathbf{a}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{u}) - f(\mathbf{a})}{h},$$

provided this limit exists.

When $\mathbf{u}$ is the standard basis vector $\mathbf{e_i}$, then Definition 9.1 recovers the definition of the $i^{th}$ partial derivative:

$$D_{\mathbf{e_i}} f(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}).$$

**Example 9.1.** *Let $f(x, y) = 3x + 2y$. Compute the directional derivative of $f$ at $(2, 1)$ in the direction $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.*

First compute the unit vector in the direction of $\mathbf{v}$:

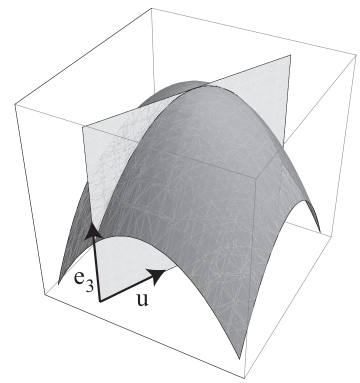
$$\mathbf{u} = \frac{\mathbf{v}}{||\mathbf{v}||} = \frac{1}{||(1,1)||}(1, 1) = \frac{1}{\sqrt{2}}\mathbf{e_1} + \frac{1}{\sqrt{2}}\mathbf{e_2}.$$

Next we compute the limit from Definition 9.1:

$$\begin{aligned}
D_{\mathbf{v}} f(2, 1) &= \lim_{h \to 0} \frac{f(2 + \frac{h}{\sqrt{2}}, 1 + \frac{h}{\sqrt{2}}) - f(2, 1)}{h} \\
&= \lim_{h \to 0} \frac{3(2 + \frac{h}{\sqrt{2}}) + 2(1 + \frac{h}{\sqrt{2}}) - \big(3(2) + 2(1)\big)}{h} \\
&= \lim_{h \to 0} \frac{(6 + \frac{3h}{\sqrt{2}} + 2 + \frac{2h}{\sqrt{2}}) - 8}{h} \\
&= \lim_{h \to 0} \frac{5h}{\sqrt{2}h} \\
&= \frac{5}{\sqrt{2}}.
\end{aligned}$$

Just as a partial derivative describes the slope of the tangent line to a curve in the graph of the function, the directional derivative also describes a slope.

Let $f : \mathbf{R}^2 \to \mathbf{R}$ be a function with the property that the directional derivative $D_{\mathbf{u}} f(\mathbf{a})$ exists. Then the plane spanned by $\mathbf{e_3}$ and $\mathbf{u}$ intersects $\Gamma_f$ in a curve, and $D_{\mathbf{u}} f(\mathbf{a})$ is the slope of the tangent line to this curve at the point $(\mathbf{a}, f(\mathbf{a}))$.




## 9.2 The gradient

**Definition 9.2.** Given $f : \mathcal{D}^n \to \mathbf{R}$, the *gradient of* $f$ *at* $\mathbf{a}$ is the vector

$$\nabla f(\mathbf{a}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) \\ \frac{\partial f}{\partial x_2}(\mathbf{a}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{bmatrix},$$

if all the partial derivatives at $\mathbf{a}$ exist.

Notice that the gradient of a scalar-valued function $f : \mathcal{D}^n \to \mathbf{R}$ is a vector in $\mathbf{R}^n$. Furthermore, $\nabla f(\mathbf{a}) = Df(\mathbf{a})^T$.

**Example 9.2.** *Find the gradient of the function* $f(x, y) = x^2 \sin(xy)$ *at* $(\pi, 0)$.
Compute the partial derivatives of $f$ to find the coordinates of the gradient:

$$\nabla f(\pi, 0) = \begin{bmatrix} f_x(\pi, 0) \\ f_y(\pi, 0) \end{bmatrix} = \begin{bmatrix} 2x\sin(xy) + x^2 y \cos(xy) \\ x^3 \cos(xy) \end{bmatrix}_{(\pi,0)} = \begin{bmatrix} 0 \\ \pi^3 \end{bmatrix}.$$

The gradient vector is useful for computing directional derivatives.

**Theorem 17.** Given $f : \mathcal{D}^n \to \mathbf{R}$, let $\mathbf{v} \in \mathbf{R}^n$ and let $\mathbf{a} \in \mathcal{D}^n$. If $f$ is differentiable at $\mathbf{a}$, then

$$D_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \frac{\mathbf{v}}{||\mathbf{v}||}.$$

As in Definition 9.1, it's important to note that $\frac{\mathbf{v}}{||\mathbf{v}||}$ is a unit vector. Problem 34 explores this requirement.

Before proving Theorem 17, we revisit Example 9.1.

**Example 9.3.** *As before, let* $f(x, y) = 3x + 2y$*. Use Theorem 17 to compute the directional derivative of* $f$ *at* $(2, 1)$ *in the direction* $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

This time, we begin by computing $\nabla f(2, 1)$. At every point $(x, y) \in \mathbf{R}^2$,

$$\nabla f(x, y) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

According to Theorem 17, for $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ we have

$$D_{\mathbf{v}} f(\mathbf{a}) = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{3}{\sqrt{2}} + \frac{2}{\sqrt{2}} = \frac{5}{\sqrt{2}},$$

which agrees with the limit we computed previously.

*Proof of Theorem 17.* Let $\mathbf{u} = \frac{\mathbf{v}}{||\mathbf{v}||}$ and define a new function $\mathbf{g} : \mathbf{R} \to \mathbf{R}^n$ by

$$\mathbf{g}(t) = \mathbf{a} + t\mathbf{u}.$$

In order to prove the theorem, we will evaluate the expression $\frac{d}{dt}(f \circ \mathbf{g})|_{t=0}$ in two different ways and compare the results.

First, note that $(f \circ \mathbf{g})(t) = f(\mathbf{a} + t\mathbf{u})$ is a real-valued function of one variable. Thus we may differentiate $f \circ \mathbf{g}$ using the definition of the derivative:

$$\frac{d}{dt}(f \circ \mathbf{g})(t) = \lim_{h \to 0} \frac{f(\mathbf{a} + (t + h)\mathbf{u}) - f(\mathbf{a} + t\mathbf{u})}{h}.$$

If we evaluate this limit at $t = 0$, we get

$$\frac{d}{dt}(f \circ \mathbf{g})|_{t=0} = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{u}) - f(\mathbf{a})}{h} = D_{\mathbf{v}} f(\mathbf{a}). \tag{9.1}$$

However, we can also compute the derivative of the composition $f \circ \mathbf{g}$ using the Chain Rule:

$$\begin{aligned} \frac{d}{dt}(f \circ \mathbf{g})|_{t=0} &= Df(\mathbf{g}(0))D\mathbf{g}(0) \\ &= Df(\mathbf{a})D\mathbf{g}(0) \\ &= \nabla f(\mathbf{a}) \cdot \mathbf{u} \end{aligned} \tag{9.2}$$

Combining Equations 9.1 and 9.2, we see that

$$D_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}.$$

□

## 9.3 Maximizing and minimizing rates of change

Suppose that for some function $f : \mathcal{D}^n \to \mathbf{R}$ and some point $\mathbf{a} \in \mathcal{D}^n$, the directional derivatives $D_{\mathbf{v}}f(\mathbf{a})$ exist for all $\mathbf{v} \in \mathbf{R}^n$. In this case, it's natural to ask which choice of $\mathbf{v}$ maximizes $D_{\mathbf{v}}f(\mathbf{a})$. Similarly, which choice of $\mathbf{v}$ minimizes $D_{\mathbf{v}}f(\mathbf{a})$? Theorem 17 provides an easy answer to this question.

Recall that the dot product satisfies the identity

$$\mathbf{a} \cdot \mathbf{b} = ||\mathbf{a}||\ ||\mathbf{b}|| \cos\theta, \tag{9.3}$$

where $\theta$ is the angle between $\mathbf{a}$ and $\mathbf{b}$.

If we use Theorem 17 to describe $D_{\mathbf{v}}f(\mathbf{a})$ as a dot product, Equation 9.3 shows that

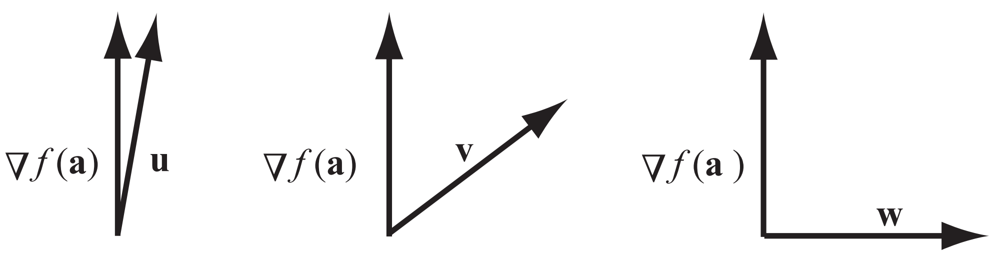
$$D_{\mathbf{v}}f(\mathbf{a}) = ||\nabla f(\mathbf{a})|| \cos\theta.$$

In this case, $\theta$ describes the angle between $\mathbf{v}$ and $\nabla f(\mathbf{a})$.

Thus, $D_{\mathbf{v}}f(\mathbf{a})$ is maximized when $\theta = 0$, which occurs when $\mathbf{v} = c\nabla f(\mathbf{a})$ for some $c > 0$. At a fixed point $\mathbf{a}$, the maximum possible direction derivative is

$$D_{\nabla f(\mathbf{a})}f(\mathbf{a}) = ||\nabla f(\mathbf{a})||.$$

Similarly, the directional derivative $D_{\mathbf{v}}f(\mathbf{a})$ is at its most negative when $\theta = \pi$, which occurs when $\mathbf{v} = -c\nabla f(\mathbf{a})$ for some $c > 0$.

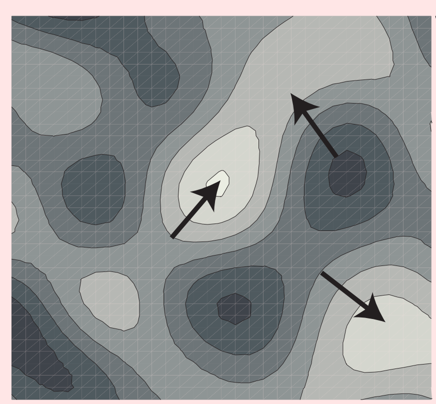



Here, $D_{\mathbf{u}}f(\mathbf{a}) > D_{\mathbf{v}}f(\mathbf{a}) > D_{\mathbf{w}}f(\mathbf{a}) = 0$.

**Example 9.4.** Suppose that a mountain is modeled by the graph of a function $f : \mathcal{D}^2 \to \mathbf{R}$, so that $f(x, y)$ gives the elevation of the mountain at the point with latitude and longitude coordinates $(x, y)$. The figure below shows a contour plot for the elevation function, where lighter shades correspond to higher altitudes.

If a hiker at the point $(a_1, a_2)$ wants to ascend as steeply as possible, then he should walk in the direction of the the gradient vector $\nabla f(a_1, a_2)$.

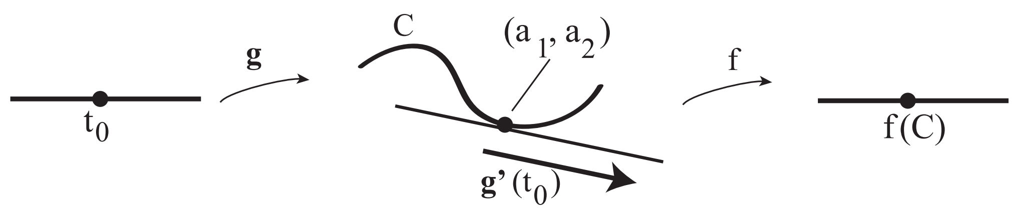
The figure shows the gradient vector at three different points on the contour plot. At each point, this arrow indicates the direction of steepest ascent.



In the figure in Example 9.4, observe that at each point, the gradient vector is orthogonal to the level curve through the point. This illustrates a general phenomenon which is described more carefully in Propositions 9.1 and 9.2.

**Proposition 9.1.** If $f : \mathcal{D}^2 \to \mathbf{R}$ is differentiable at $(a_1, a_2)$ and $\nabla f(a_1, a_2) \neq \mathbf{0}$, then $\nabla f(a_1, a_2)$ is orthogonal to the tangent line to the level set of $f$ which passes through $(a_1, a_2)$.

*Proof.* Let $C$ denote the level set $f^{-1}(f(a_1, a_2))$. Suppose that $\mathbf{g} : \mathcal{I} \to \mathbf{R}^2$ is a parameterized curve whose image is $C$ and which satisfies $\mathbf{g}(t_0) = (a_1, a_2)$ for $t_0 \in \mathcal{I}$. Then the tangent line to $C$ at $(a_1, a_2)$ has direction vector $\mathbf{g}'(t_0)$.




The composition $f \circ \mathbf{g}$ is a constant function, so its derivative is zero at all values of $t$:

$$0 = \frac{d}{dt}(f \circ \mathbf{g})(t) = D(f \circ \mathbf{g})(t).$$

On the other hand, we can also compute this derivative using the Chain Rule:

$$\begin{aligned} D(f \circ \mathbf{g})(t_0) &= Df(\mathbf{g}(t_0))D\mathbf{g}(t_0) \\ &= Df(a_1, a_2)\mathbf{g}'(t_0) \\ &= \nabla f(a_1, a_2) \cdot \mathbf{g}'(t_0). \end{aligned}$$

Comparing the two equations shows that

$$0 = \nabla f(a_1, a_2) \cdot \mathbf{g}'(t_0),$$

so the tangent line to $C$ at $\mathbf{t_0}$ is orthogonal to $\nabla f(a_1, a_2)$. □

In fact, a similar proof shows the following more general result:

**Proposition 9.2.** Suppose that $f : \mathcal{D}^n \to \mathbf{R}$ is differentiable at $\mathbf{a} \in \mathcal{D}^n$ and that $\nabla f(\mathbf{a}) \neq \mathbf{0}$. Denote by $S$ the level set $f^{-1}(f(\mathbf{a}))$. Then if $C$ is any curve in $S$ which passes through $\mathbf{a}$, the tangent line to $C$ at $\mathbf{a}$ is orthogonal to $\nabla f(\mathbf{a})$.

## 9.4 Worked examples

**Example 9.5.** *Suppose that the position of a bee in a room at time $t$ is given by the function*

$$\mathbf{r}(t) = \begin{bmatrix} 2t \\ 5-t \\ \sqrt{1+t} \end{bmatrix}.$$

*Furthermore, suppose that the air temperature is described by the function*

$$T(x,y,z) = e^{20-xyz}.$$

1. *At $t = 3$, is the bee flying towards warmer or cooler air?*
2. *In what direction should the bee fly to cool down the fastest at $t = 3$?*

The first question is answered by computing the directional derivative of $T$ in the direction of the bee's flight at $t = 3$. We begin by solving for this direction:

$$\mathbf{r}'(3) = \begin{bmatrix} 2 \\ -1 \\ \frac{1}{2\sqrt{1+t}} \end{bmatrix}\Big|_{t=3} = \begin{bmatrix} 2 \\ -1 \\ \frac{1}{4} \end{bmatrix}.$$

The unit vector in the same direction is

$$\mathbf{u} = \frac{4}{9}\begin{bmatrix} 2 \\ -1 \\ \frac{1}{4} \end{bmatrix}.$$

Next, we compute $\nabla T(\mathbf{r}(3))$:

$$\nabla T(\mathbf{r}(3)) = \begin{bmatrix} -yze^{20-xyz} \\ -xze^{20-xyz} \\ -xye^{20-xyz} \end{bmatrix}\Big|_{(6,2,2)} = \begin{bmatrix} -4e^{-4} \\ -12e^{-4} \\ -12e^{-4} \end{bmatrix}.$$

We can now compute the directional derivative using Theorem 17:

$$D_{\mathbf{r}'(3)}T(6,2,2) = \frac{4}{9}\begin{bmatrix} 2 \\ -1 \\ \frac{1}{4} \end{bmatrix} \cdot \begin{bmatrix} -4e^{-4} \\ -12e^{-4} \\ -12e^{-4} \end{bmatrix} = \frac{4}{9}e^{-4}.$$

Since this number is positive, the temperature is increasing in the bee's direction of flight at $t = 3$.

The second question requires finding the direction $\mathbf{u}$ for which the directional derivative $D_{\mathbf{u}}f(6,2,2)$ is as negative as possible. The quantity $\mathbf{u} \cdot \nabla T(6,2,2)$ is minimized when the angle between the two vectors is $\pi$, or equivalently, when $\mathbf{u} = -\nabla T(6,2,2)$. Thus, the temperature will decrease most quickly if the bee flies in the direction $\begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}$. (Notice that when we only care about the direction, we can rescale $-\nabla T(6,2,2)$ to simplify the coordinates.)

**Example 9.6.** *The sphere defined by $x^2 + y^2 + z^2 = 26$ intersects the plane $x + 7y + 7z = 50$ in a curve $C$. Find the tangent line to $C$ at the point $(1, 3, 4)$.*

We can use Proposition 9.2 to solve this problem. The sphere is the level set at height $26$ of the function $G(x, y, z) = x^2 + y^2 + z^2$, so the tangent line to $C$ is orthogonal to $\nabla G(1, 3, 4)$. On the other hand, the plane is the level set at height $50$ of the function $H(x, y, z) = x + 7y + 7z$, so the tangent line to $C$ is also orthogonal to $\nabla H(1, 3, 4)$.

The cross product of $\nabla G$ and $\nabla H$ is a direction vector for the tangent line to $C$:

$$\nabla G(1,3,4) \times \nabla H(1,3,4) = \begin{bmatrix} 2 \\ 6 \\ 8 \end{bmatrix} \times \begin{bmatrix} 1 \\ 7 \\ 7 \end{bmatrix} = \begin{bmatrix} -14 \\ -6 \\ 8 \end{bmatrix}$$

Therefore, we can parameterize the tangent line to $C$ at $(1, 3, 4)$ by

$$\mathbf{L}(t) = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} + t \begin{bmatrix} -7 \\ -3 \\ 4 \end{bmatrix}.$$

## 9.5 Exercises

In Exercises 1 through 4, use Definition 9.1 to compute the directional derivative $D_{\mathbf{v}}f(\mathbf{a})$ for the given function $f$, direction $\mathbf{v}$, and point $\mathbf{a}$.

1. $f(x, y) = 2xy$, $\mathbf{v} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$, $\mathbf{a} = (7, -2)$.

2. $f(x, y) = \frac{1}{xy}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $\mathbf{a} = (1, 2)$

3. $f(x, y, z) = 4xz - \frac{1}{y}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$, $\mathbf{a} = (0, 1, 1)$

4. $f(x, y, z) = x^2yz - xz^2 + 3y^2$, $\mathbf{v} = \begin{bmatrix} 5 \\ 0 \\ -12 \end{bmatrix}$, $\mathbf{a} = (-2, -1, 1)$

In Exercises 5 through 18, use Theorem 17 to compute the directional derivative $D_{\mathbf{v}}f(\mathbf{a})$ for the given function $f$, direction $\mathbf{v}$, and point $\mathbf{a}$.

5. $f(x, y) = 2xy$, $\mathbf{v} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$, $\mathbf{a} = (7, -2)$.

6. $f(x, y) = \frac{1}{xy}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $\mathbf{a} = (1, 2)$

7. $f(x,y,z) = 4xz - \frac{1}{y}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$, $\mathbf{a} = (0,1,1)$

8. $f(x,y,z) = x^2yz - xz^2 + 3y^2$, $\mathbf{v} = \begin{bmatrix} 5 \\ 0 \\ -12 \end{bmatrix}$, $\mathbf{a} = (-2,-1,1)$

9. $f(x,y) = 2x^7 - 7x^2y + 4xy^2 + y^6$, $\mathbf{v} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$, $\mathbf{a} = (-1,-1)$

10. $f(x,y) = e^{-x^2-y^2}$, $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\mathbf{a} = (0,1)$

11. $f(x,y) = xe^{xy}$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, $\mathbf{a} = (-2,0)$

12. $f(x,y) = \sin(x^2 - 5xy - 2y^2 + 3x + 3y)$, $\mathbf{v} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$, $\mathbf{a} = (1,-2)$

13. $f(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2}}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{a} = (2,1,-2)$

14. $f(x,y,z) = e^{xyz}\sin(xyz)$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 5 \end{bmatrix}$, $\mathbf{a} = (0,3,10)$

15. $f(x,y,z) = x\cos(yz) + y\cos(xz) + z\cos(yx)$, $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix}$, $\mathbf{a} = (0,2,\pi/4)$

16. $f(x,y,z) = \ln(x^2 + y^2 + z^2)$, $\mathbf{v} = \begin{bmatrix} -2 \\ 2 \\ 2 \end{bmatrix}$, $\mathbf{a} = (3,4,-3)$

17. $f(w,x,y,z) = 6w^2z - xy^3 - 3wy + 7x^2z^2$, $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{a} = (1,0,2,-5)$

18. $f(x_1,...,x_n) = \frac{1}{x_1^2+...+x_n^2}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, $\mathbf{a} = (1,1,...,1)$

19. Suppose the altitude in meters of some region of the Swiss Alps is approximately given by $A(x, y) = 4(x-y)^2 - (3x+y)^2 + 2000$.

    (a) How does the altitude change if a hiker located at the point $(1, -3)$ walks in the direction $30°$ south of east?

    (b) What if she walks in the direction $45°$ north of west?

20. Suppose you are swimming in Lake Lagunita, whose depth in meters is given by $d(x, y) = 3 - \frac{1}{13000}x^2 - \frac{1}{5000}y^2$. If you are at the point $(200, 50)$, in what direction should you swim if you want the depth to decrease as rapidly as possible?

21. An ant is at the point $(\pi/6, \pi/6)$ on a patch of earth whose height (in cm) is given by $h(x, y) = \sin(x - 2y) + 2\cos(x + 3y) + 4$.

    (a) In what direction should the ant face so that its ascent will be steep as possible when it starts walking forward?

    (b) In what direction should it face to remain at the same height?

22. A crow is at the point $(3, -1, 6)$ in a region of the sky where the humidity is given by $h(x, y, z) = (x + 2y + z)^2 - (y + z)^2 + 3z^2$

    (a) If it flies in the direction $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, is the humidity increasing or decreasing?

    (b) What if it flies in the direction $\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$?

23. The temperature in degrees Celsius of an oven is given by $T(x, y, z) = 200 - x^2 - y^2 - z^2$. A heat-resistant robotic fly is at the point $(-3, -1, 5)$. In what direction should it fly to heat up as quickly as possible?

24. Suppose that the air pressure (in kPa) of some region in the sky is a function of position (in km) given by $p(x, y, z) = 20e^{0.15(11-z)} + e^{-0.15}(\sin x + 2y)$ and that an airplane is at the point $(\pi, \pi/2, 12)$.

    (a) Find a direction in which the pressure will stay constant.

    (b) Now suppose that the same airplane is flying in the direction $\begin{bmatrix} -4 \\ 5 \\ 1 \end{bmatrix}$. Find the direction in which the pressure will stay constant which is closest to the direction the plane is already flying in.
    (Hint: Use orthogonal projection to a plane in $\mathbf{R}^3$.)

25. Let $C$ be the curve in $\mathbf{R}^2$ defined by the equation $\sin x + \cos(2y) = 1$. Find an equation for the line perpendicular to the tangent line to $C$ at the point $(\pi/6, \pi/6)$.

26. Let $C$ be the curve in $\mathbf{R}^2$ defined by $3x^3 - x^2y^2 + y^4 = 21$. Find an equation for the tangent line to $C$ at the point $(2, -1)$.

27. Let $C$ be the curve given by $4x^2 - xy + y^2 = 4$. At what points is the tangent line to $C$ parallel to the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$?

28. Let $S$ be the hyperboloid of two sheets given by $3x^2 - 2y^2 + z^2 = -5$.

    (a) Find two parametrized curves $C_1, C_2$ on $S$ through the point $(2, 3, -1)$ whose tangent lines $L_1, L_2$ are distinct.

    (b) Let $P$ be the plane containing the lines $L_1, L_2$. Find an equation for $P$.

    (c) Describe the relationship between the normal vector to $P$ and $\nabla f(2, 3, -1)$.

    (d) On one set of axes, sketch $S$ and $P$.

29. Let $S$ be the hyperbolic paraboloid given by $7x^2 - y^2 - 2z = 1$. Let $\mathbf{g}_1$ and $\mathbf{g}_2$ be parametrized curves whose images lie in $S$ and which pass through $\mathbf{a} = (-1, 2, 1)$. Recall that for $i = 1, 2$, the tangent line to $\mathbf{g}_i$ at the point $\mathbf{a}$ is parallel to $\mathbf{g}_i'(\mathbf{a})$. Find a vector parallel to $\mathbf{g}_1'(\mathbf{a}) \times \mathbf{g}_2'(\mathbf{a})$.

30. Let $C$ be the curve which is the intersection of the ellipsoid $2x^2+3y^2+z^2 = 9$ and the hyperboloid $-3x^2 + 6y^2 + z^2 = 7$. Find a parametric representation for the tangent line to $C$ at the point $(1, -1, 2)$.

31. Let $f(x, y, z) = x^2 + 3y^2 + 4z^2$, and let $E$ be the ellipsoid given by $f(x, y, z) = 8$. Let $P$ be the plane given by $P = \{(0, 0, 2) + s\mathbf{u} + t\mathbf{v} | s, t \in \mathbf{R}\}$ where

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}.$$

    Notice that $(1, 1, 1)$ lies in the intersection $E \cap P$.

    (a) What geometric relationships can you find between $\mathbf{u}$, $\mathbf{v}$, and $\nabla f(1, 1, 1)$? (Hint: Take dot products of pairs of these vectors.)

    (b) Use your answer from the first part to describe or sketch the intersection of $E$ and $P$.

32. Let $f : \mathbf{R}^n \to \mathbf{R}$ be a differentiable function, and let $\mathbf{a}, \mathbf{v} \in \mathbf{R}^n$. Determine whether each of the following statements is always true or sometimes false

   (a) $\nabla f(\mathbf{a})$ is a vector in $\mathbf{R}^n$.

   (b) $D_{\mathbf{v}} f(\mathbf{a})$ is a vector in $\mathbf{R}^n$.

   (c) $||\mathbf{v}|| D_{\mathbf{v}} f(\mathbf{a}) = Df(\mathbf{a})\mathbf{v}$.

   (d) $D_{-\mathbf{v}} f(\mathbf{a}) = -D_{\mathbf{v}} f(\mathbf{a})$.

   (e) If $g(x, y) = x^2 + 3xy - y^2 - 4x + 3y$, then there exists $\mathbf{w} \in \mathbf{R}^2$ so that $D_{\mathbf{w}} g(0, 0) = 5$.

33. Let $g(x, y, z) = f(x, y) - z$. Determine whether each of the following statements is always true or sometimes false.

   (a) The vector $\nabla f(x, y)$ is perpendicular to $\Gamma_g$ at the point $(x, y, f(x, y))$.

   (b) The vector $\nabla g(x, y, f(x, y))$ is perpendicular to $\Gamma_g$ at the point $(x, y, f(x, y))$.

   (c) The vector $\nabla g(x, y, z)$ is never perpendicular to $(0, 0, 1)$.

   (d) The vector $\begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ 1 \end{bmatrix}$ points in the direction in which $g$ increases the most rapidly.

34. Compute the following limits:

$$\lim_{t \to 0} \frac{(a + 2t)^2 - a^2}{t}$$

$$\lim_{t \to 0} \frac{(a - t)^2 - a^2}{t}$$

$$\lim_{t \to 0} \frac{(a + \frac{1}{2}t)^2 - a^2}{t}$$

   Use the results of your computations to explain why the definition of a directional derivative in the direction $\mathbf{v}$ uses the unit vector $\frac{\mathbf{v}}{|\mathbf{v}|}$.

35. Pick $\mathbf{v}, \mathbf{a} \in \mathbf{R}^n$ and construct functions $f$ and $g$ that satisfy the following properties:

   (a) $D_{\mathbf{v}} f(\mathbf{a})$ exists but $\nabla f(\mathbf{a})$ does not exist.

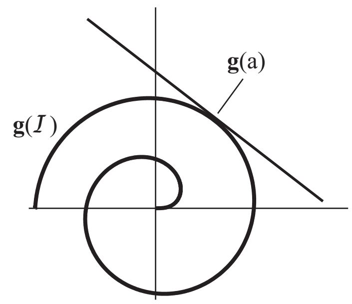
   (b) $D_{\mathbf{v}} g(\mathbf{a})$ does not exist but $\nabla g(\mathbf{a})$ exists.

# 10 Tangent Planes

In Section 4, we discussed tangent lines to space curves. If $C$ is the image of a parameterized curve $\mathbf{g} : \mathcal{I} \to \mathbf{R}^n$, then the tangent line to $C$ at $\mathbf{g}(a)$ is

$$\{\mathbf{g}(a) + t\mathbf{g}'(a) \mid t \in \mathbf{R}\}.$$

If the curve $C$ has a tangent line at every point, then we say that $C$ is a *differentiable curve*.




In this section we discuss a higher-dimenionsal analogue of the tangent line to a differentiable curve: the tangent plane to a differentiable surface.

## 10.1 Differentiable surfaces and tangent planes

We call surfaces which have tangent planes at every point *differentiable surfaces*. Loosely speaking, a differentiable surface is a subset of $\mathbf{R}^3$ that looks locally like a graph of a differentiable function of two variables.

There are two particularly important classes of differentiable surfaces:

- If $f : \mathcal{D}^2 \to \mathbf{R}$ is a differentiable function, the graph $\Gamma_f$ is a differentiable surface.
- If $F : \mathcal{D}^3 \to \mathbf{R}$ is a differentiable function, many level sets of $F$ are differentiable surfaces.

We will use the term *level surface* to indicate that a given level set is a differentiable surface. If you are interested in understanding which level sets are level surfaces, Section 16 in the Appendix discusses the technical definition.

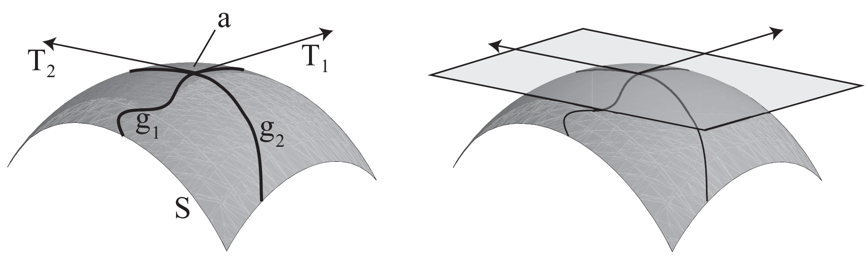
**Definition 10.1.** Let $S$ be a differentiable surface in $\mathbf{R}^3$ which contains the point $\mathbf{a}$. Denote by $\mathcal{C}(\mathbf{a})$ the set of parameterized curves $\mathbf{g} : [-1, 1] \to \mathbf{R}^3$ which satisfy the following:

- the image of $\mathbf{g}$ is a subset of $S$; and
- $\mathbf{g}(0) = \mathbf{a}$.

The *tangent plane to $S$ at* $\mathbf{a}$ is the set of all vectors of the form

$$\{\mathbf{g}(0) + \mathbf{g}'(0) \mid \mathbf{g} \in \mathcal{C}(\mathbf{a})\}.$$

The figure illustrates the idea of Definition 10.1. On the left, a surface $S$ is shown together with the images of two parameterized curves $\mathbf{g_1}$ and $\mathbf{g_2}$ in $\mathcal{C}(\mathbf{a})$. The tangent vector to the image of $\mathbf{g_i}$ at $\mathbf{a}$ is denoted by $T_i$. The tangent plane to $S$ at $\mathbf{a}$, which is shown on the right, is the plane containing $T_1$ and $T_2$.




If $\mathbf{a}$ is any point in $S$, then the set described in Definition 10.1 is in fact a plane. However, we won't prove this here. Instead, we'll focus on how to compute an equation for the tangent plane to a differentiable surface at a given point.

## 10.2 Tangent planes to level surfaces

**Theorem 18.** Suppose that $F : \mathcal{D}^3 \to \mathbf{R}$ has continuous partial derivatives $F_x(\mathbf{x})$, $F_y(\mathbf{x})$, and $F_z(\mathbf{x})$, and suppose that these functions are not identically $0$ in a neighborhood of $\mathbf{a}$. Let $S$ be the level surface of $F$ at height $F(\mathbf{a})$:

$$S = F^{-1}(F(\mathbf{a})).$$

Then the tangent plane to $S$ at $\mathbf{a} = (a_1, a_2, a_3)$ is described by the equation

$$0 = F_x(\mathbf{a})(x - a_1) + F_y(\mathbf{a})(y - a_2) + F_z(\mathbf{a})(z - a_3). \tag{10.1}$$

To prove Theorem 18, we'll show that the plane defined by Equation 10.1 agrees with the one described in Definition 10.1. But first, an example:

**Example 10.1.** *Let $S$ be the sphere*

$$S = \{(x, y, z) \in \mathbf{R}^3 \,|x^2 + y^2 + z^2 = 26\}.$$

*Find the tangent plane to $S$ at $(4, 1, 3)$.*
First, observe that $S = F^{-1}(26)$ for the function

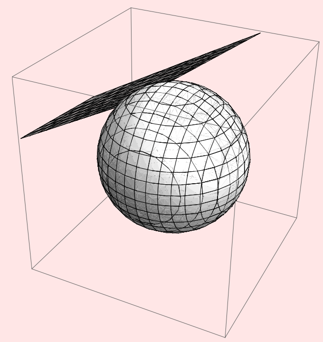
$$F(x, y, z) = x^2 + y^2 + z^2.$$

Since $F$ has continuous partial derivatives at every point in $\mathbf{R}^3$, we are free to apply the formula from Theorem 18.

First, we compute the gradient of $F$ at the point $(4, 1, 3)$:

$$\nabla F(4,1,3) = \begin{bmatrix} 2x \\ 2y \\ 2z \end{bmatrix}_{(4,1,3)} = \begin{bmatrix} 8 \\ 2 \\ 6 \end{bmatrix}.$$

Thus, Theorem 18 implies the tangent plane is given by the equation



$$0 = 8(x-4) + 2(y-1) + 6(z-3).$$

In the Worked Examples section, we compute the same result directly from the definition of the tangent plane.

*Proof of Theorem 18.* Let $\mathbf{g} : \mathcal{I} \to \mathbf{R}^3$ be a parameterized curve in $\mathcal{C}(\mathbf{a})$. We will study the composition $F \circ \mathbf{g}$.

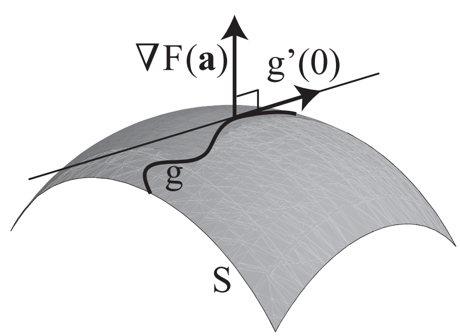
Since $S$ is a level set for $F$, the composition $F \circ \mathbf{g}$ is a constant function. Therefore,

$$\frac{d}{dt}(F \circ \mathbf{g})(t) = 0 \quad \text{for all } t.$$

We can also compute this derivative via the Chain Rule:

$$\frac{d}{dt}F(\mathbf{g}(0)) = DF(\mathbf{g}(0))\mathbf{g}'(0) = \nabla F(\mathbf{a}) \cdot \mathbf{g}'(0).$$

Combining these two results shows that




$$\nabla F(\mathbf{a}) \cdot \mathbf{g}'(0) = 0.$$

The vector $\mathbf{g}'(0)$ is a direction vector for the tangent line to $\mathbf{g}(\mathcal{I})$ at $\mathbf{a}$. Thus, $\nabla F(\mathbf{a})$ is orthogonal to the tangent line to the image of any $\mathbf{g} \in \mathcal{C}(\mathbf{a})$. This implies that $\nabla F(\mathbf{a})$ is a normal vector for the tangent plane to $S$ at $\mathbf{a}$.

The formula then follows from the standard equation of a plane involving a point and a normal vector. □

## 10.3 Tangent planes to graphs

When a surface $S$ is actually the graph of a differentiable function of two variables, we can find an equation for the tangent plane a different way.

**Proposition 10.1.** Suppose that $f : \mathcal{D}^2 \to \mathbf{R}$ is differentiable at $(a, b) \in \mathcal{D}^2$. Then the tangent plane to $\Gamma_f$ at $(a, b, f(a, b))$ is given by the equation

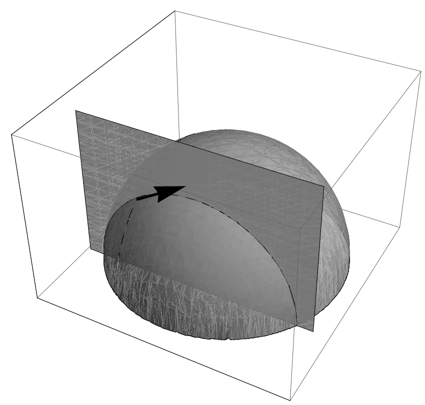
$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

Since $\Gamma_f$ is the level set at height $0$ of $F(x, y, z) = f(x, y) - z$, we could extract the formula in Proposition 10.1 from Theorem 18. However, the geometric interpretation of partial derivatives that we discussed in Section 7.3 gives us an independent proof.

*Proof of Proposition 10.1.* First, consider the intersection of $\Gamma_f$ with the plane $y = b$, and denote this curve by $C_b$. Then $C_b$ is the graph of the function $f(x, b)$, so the slope of the tangent line to $C_b$ at $(a, b, f(a, b))$ is $\frac{\partial f}{\partial x}(a, b)$.

This slope describes the change in $z$ with respect to unit change in $x$, so the direction vector for the tangent line is parallel to

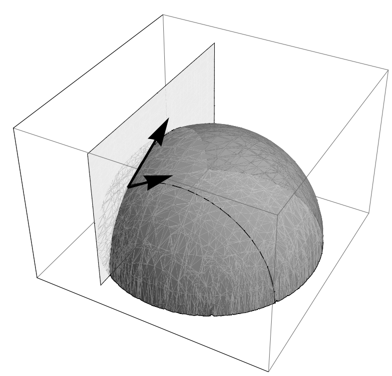
$$\mathbf{v_x} = \begin{bmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial x}(a, b) \end{bmatrix}.$$



The figure on the right shows the surface $\Gamma_f$, the plane $y = b$, and the vector $\mathbf{v_x}$.

Similarly, $\Gamma_f$ intersects the plane $x = a$ in a curve $C_a$. The tangent line to $C_a$ at $(a, b, f(a, b))$ has direction vector

$$\mathbf{v_y} = \begin{bmatrix} 0 \\ 1 \\ \frac{\partial f}{\partial y}(a, b) \end{bmatrix}.$$



The figure on the right shows both $\mathbf{v_x}$ and $\mathbf{v_y}$.

According to Definition 10.1 the vectors $\mathbf{v_y}$ and $\mathbf{v_x}$ lie in the tangent plane to $\Gamma_f$. They are linearly independent, so their cross product $\mathbf{n}$ is a normal vector for the tangent plane to $\Gamma_f$ at $(a, b, f(a, b))$.

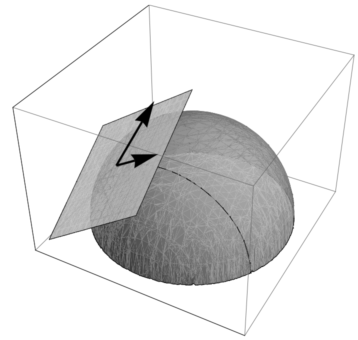
$$\mathbf{n} = \mathbf{v_y} \times \mathbf{v_x} = \begin{bmatrix} 0 \\ 1 \\ \frac{\partial f}{\partial y}(a, b) \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial x}(a, b) \end{bmatrix} = \begin{bmatrix} f_x(a, b) \\ f_y(a, b) \\ -1 \end{bmatrix}.$$

The plane through $(a, b, f(a, b))$ with normal vector $\mathbf{n}$ has the equation proposed by Proposition 10.1:

$$0 = \begin{bmatrix} x-a \\ y-b \\ z-f(a,b) \end{bmatrix} \cdot \begin{bmatrix} f_x(a,b) \\ f_y(a,b) \\ -1 \end{bmatrix},$$

or equivalently,

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).$$



□

## 10.4 Worked Examples

**Example 10.2.** *Use Definition 10.1 to find the tangent plane to the sphere*

$$S = \{(x,y,z) | x^2 + y^2 + z^2 = 26\}$$

*at the point* $(4, 1, 3)$.

Let $\mathbf{g}(t) = \big(x(t), y(t), z(t)\big)$ be a parameterized curve whose image lies in $S$, and suppose that $\mathbf{g}(0) = (4, 1, 3)$. Furthermore, assume that $\mathbf{g}'(0) \neq \mathbf{0}$. Then the tangent line to the image of $\mathbf{g}$ at $(4, 1, 3)$ is the set of vectors

$$\Big\{ \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} + t\mathbf{g}'(0) \Big\}.$$

By definition, the tangent plane contains this line.

Now consider the defining equation for $S$:

$$\big(x(t)\big)^2 + \big(y(t)\big)^2 + \big(z(t)\big)^2 = 26.$$

Differentiate both sides of this equations with respect to $t$ to get

$$2x(t)\frac{dx}{dt} + 2y(t)\frac{dy}{dt} + 2z(t)\frac{dz}{dt} = 0.$$

We can rewrite this equation as a dot product:

$$\begin{bmatrix} 2x(t) \\ 2y(t) \\ 2z(t) \end{bmatrix} \cdot \mathbf{g}'(t) = 2\mathbf{g}(t) \cdot \mathbf{g}'(t) = 0.$$

Evaluating this equation at $0$ shows that $\mathbf{g}'(0)$ is orthogonal to $\mathbf{v} = \begin{bmatrix} 8 \\ 2 \\ 6 \end{bmatrix}$. However, this relation holds for all $\mathbf{g}$ satisfying the properties above, so $\mathbf{v}$ is a normal vector to the tangent plane to $S$ at $(4, 1, 3)$.

Thus, we get an equation of the tangent plane:

$$\begin{bmatrix} 8 \\ 2 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} x-4 \\ y-1 \\ z-3 \end{bmatrix} = 0$$

## 10.5 Exercises

In Exercises 1 through 13 , find the tangent plane to the given level surface at the indicated point.

1. $x^2 + y^2 + z = 6$, $\mathbf{a} = (-1, 2, 1)$
2. $xy = z$, $\mathbf{a} = (2, 2, 4)$
3. $3x^2 + 2y^2 + z^2 = 15$, $\mathbf{a} = (1, 2, 2)$
4. $x - 2y + 3z = 7$, $\mathbf{a} = (0, -2, 1)$
5. $y^2 - z = 0$, $\mathbf{a} = (1, 0, 0)$
6. $yz \sin x^2 = 1$, $\mathbf{a} = (\sqrt{\frac{\pi}{6}}, 1, 2)$
7. $x + \ln\left(\frac{z}{y}\right) = 0$, $\mathbf{a} = (0, 2, 2)$
8. $z = e^x + y$, $\mathbf{a} = (1, 1, e^1 + 1)$
9. $x^2 - 2y^2 + 5xz = 7$, $\mathbf{a} = (-1, 0, -\frac{6}{5})$
10. $\tan x - 2x + 3y = e^z - 2\pi$, $\mathbf{a} = (\pi, \frac{e}{3}, 1)$
11. $e^z \sin(xy) = 1$, $\mathbf{a} = (\frac{1}{2}, \pi, 0)$
12. $e^{z^2} + x\cos y = x^2 - 1$, $\mathbf{a} = (1, \pi, 1)$
13. $x \ln(y + 2z) = 3$, $\mathbf{a} = (3, e - 2, 1)$

In Exercises 14 through 19, find the tangent plane to the graph of the given function $f$ at the point $(\mathbf{a}, f(\mathbf{a}))$.

14. $f(x, y) = x^2 + 4y^2$, $\mathbf{a} = (0, 3)$
15. $f(x, y) = \sin(x^2 + y^2), \mathbf{a} = (\sqrt{\pi}, 0)$
16. $f(y, z) = yz, \mathbf{a} = (2, 2)$
17. $f(y, z) = 2y + 3z - 2, \mathbf{a} = (1, -1)$
18. $f(x, z) = e^{xz}, \mathbf{a} = (0, 0, )$
19. $f(x, y) = e^x \sin(xy), \mathbf{a} = (\pi, 1)$
20. Find the point(s) on the differentiable surface

$$9x^2 + 45y^2 + 5z^2 = 45$$

where the tangent plane is parallel to the plane $x + 5y - 2z = 7$.

21. Find the point(s) on the differentiable surface

$$2x + 2\ln 2y + z^2 = 9$$

where the tangent plane is parallel to the plane $x + 2y + z = 7$.

22. Find the point(s) on the differentiable surface

$$x\cos y + e^z = 2$$

where the tangent plane has normal vector $\mathbf{n} = \begin{bmatrix} 1 \\ 0 \\ e \end{bmatrix}$.

23. Find the point(s) on the differentiable surface

$$-5\sin y + x^3 = z - 8$$

where the tangent plane has normal vector $\mathbf{n} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

24. Suppose that $F : \mathbf{R}^4 \to \mathbf{R}$ is defined by $F(w, x, y, z) = w^2 + x^2 + y^2 + z^2$.

    (a) Describe the level sets of $F^{-1}(c)$.
    When $c > 0$, the level set at height $c$ is an example of a *differentiable hypersurface*.

    (b) Generalize Theorem 18 to give a formula for the *tangent hyperplane* to a differentiable hypersurface. Explain why you think this is the right generalization.

    (c) Use (b) to find an equation for the tangent hyperplane to the level set of $F$ which contains the point $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

25. Define

$$F(x, y, z) = \ln(x^2y^2z^4 + 2)\frac{\cos e^{xyz}}{x^2 + yz + 1}.$$

Let $P$ denote the tangent plane to the level surface of $F$ at height $F(1, 1, 1) = (\ln 3)\frac{\cos e}{3}$. If $\mathbf{v}$ is a vector in $P$, what is $D_{\mathbf{v}}F(1, 1, 1)$?

# 11 Linearization and Taylor's Theorem

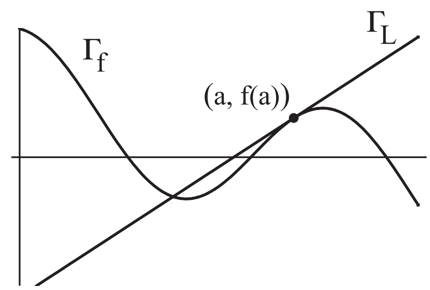
Sometimes one can study a function by considering the properties of a similar but simpler function. In particular, it's often helpful to approximate a function by a polynomial, since polynomials are relatively easy to work with. In this section we discuss how to find polynomial approximations to multivariable functions. This extends the ideas of Section 7.4, where a function was defined to be differentiable if its derivative provided a sufficiently good approximation.

## 11.1 Linearization

If the single variable function $f : \mathbf{R} \to \mathbf{R}$ is differentiable at $a$, then the linearization of $f$ at $a$ is the function

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

The graph of $L$ is the tangent line to the graph of $f$ at the point $(a, f(a))$.




The function $L$ is a good linear approximation to $f$ at $a$. When $x$ is near $a$, $|L(x) - f(x)|$ is small, although this difference may grow as $x$ moves away from $a$.

We can define an analogous object for a multivariable function $\mathbf{f}$.

**Definition 11.1.** Suppose that $\mathbf{f} : \mathcal{D}^n \to \mathbf{R}^m$ is differentiable at $\mathbf{a}$. The *linearization of* $\mathbf{f}$ *at* $\mathbf{a}$ is the function $\mathbf{L} : \mathbf{R}^n \to \mathbf{R}^m$ defined by

$$\mathbf{L}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

**Example 11.1.** *Find the linearization of* $\mathbf{f}(x, y) = (2x^3 - 6x^2y^3, 4x^2y^2 + y^3)$ *at the point* $(1, -1)$*. Use the linearization it to estimate* $\mathbf{f}(.9, -1.1)$*.*

We begin by computing the matrix of partial derivatives $D\mathbf{f}(1, -1)$:

$$D\mathbf{f}(1,-1) = \begin{bmatrix} 6x^2 - 12xy^3 & -18x^2y^2 \\ 8xy^2 & 8x^2y + 3y^2 \end{bmatrix}_{(1,-1)} = \begin{bmatrix} 18 & -18 \\ 8 & -5 \end{bmatrix}$$

Next, we evaluate $\mathbf{f}$ at $(1, -1)$:

$$\mathbf{f}(1,-1) = (2(1)^3 - 6(1)^2(-1)^3, 4(1)^2(-1)^2 + (-1)^3) = \begin{bmatrix} 8 \\ 3 \end{bmatrix}$$

Definition 11.1 implies that the linearization of $\mathbf{f}$ at $(1, -1)$ is

$$\begin{aligned}\mathbf{L}(x,y) &= \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 18 & -18 \\ 8 & -5 \end{bmatrix} \begin{bmatrix} x-1 \\ y+1 \end{bmatrix} \\ &= (-28 + 18x - 18y, -10 + 8x - 5y).\end{aligned}$$

In order to estimate $\mathbf{f}(.9, -1.1)$, we replace $\mathbf{f}$ by $\mathbf{L}$ and evaluate at $(.9, -1.1)$:

$$\mathbf{L}(.9, -1.1) = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 18 & -18 \\ 8 & -5 \end{bmatrix} \begin{bmatrix} -.1 \\ -.1 \end{bmatrix} = \begin{bmatrix} 8 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ -.3 \end{bmatrix} = \begin{bmatrix} 8 \\ 2.7 \end{bmatrix}.$$

With a calculator, one can verify that $(8, 2.7)$ is a respectable approximation of $\mathbf{f}(.9, -1.1)$:

$$f(.9, -1.1) = (7.9266, 2.5894).$$

The term "linearization" indicates that the linear behavior of $L$ and $f$ is the same; that is,

$$D\mathbf{L}(\mathbf{x}) = D\mathbf{f}(\mathbf{a}) \quad \text{for all } \mathbf{x} \in \mathbf{R}^n,$$

and the higher derivatives of $\mathbf{L}$ vanish. However, this *doesn't* mean that $\mathbf{L}$ is a linear transformation. (See Exercise 16.)

If $L$ is the linearization of a function $f : \mathbf{R}^2 \to \mathbf{R}$, then the graph of $L$ is a plane in $\mathbf{R}^3$. In fact, $\Gamma_L$ is the tangent plane to $\Gamma_f$ at $(\mathbf{a}, f(\mathbf{a}))$. This observation gives us a new way to think about the tangent plane: just as a tangent line approximates the graph of a differentiable single variable function, the tangent plane approximates the graph of a differentiable function of two variables.

## 11.2 Taylor's Theorem

Linearization lets us approximate a differentiable function by a degree one polynomial. However, we might prefer an approximation whose higher derivatives also agree with those of the original function. In single variable calculus, Taylor's theorem tells us how to find such an approximation.

Given a smooth function $f : \mathbf{R} \to \mathbf{R}$, we can define the $n^{th}$ degree Taylor polynomial $T_n(x)$:

$$T_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

For $0 \le i \le n$, $T_n$ satisfies $T_n^{(i)}(\mathbf{a}) = f^{(i)}(\mathbf{a})$. Furthermore, $T_n^{(j)} = 0$ for $j > n$.

In order to develop an analogous formula for a real-valued function of several variables, we introduce the *Hessian*. Just as the matrix of partial derivatives organizes the first derivatives of a multivariable function, the Hessian keeps track of all the second partial derivatives of a scalar-valued function of several variables.

**Definition 11.2.** Let $f : \mathcal{D}^n \to \mathbf{R}$ be a function whose second partial derivatives exist at $\mathbf{a}$. Then the *Hessian* of $f$ at $\mathbf{a}$ is the $n \times n$ matrix whose $(i, j)$ entry is $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$:

$$Hf(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \dots & \dots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{a}) \end{bmatrix}.$$

If $f : \mathbf{R} \to \mathbf{R}$ is a function of one variable, then the Hessian is just the ordinary second derivative of $f$.

**Example 11.2.** *Let $f(x, y, z) = xye^{2z}$. Compute $Hf(1, 2, 0)$.*
We begin by computing the partial derivatives of $f$:

$$\frac{\partial f}{\partial x} = ye^{2z} \qquad \frac{\partial f}{\partial y} = xe^{2z} \qquad \frac{\partial f}{\partial z} = 2xye^{2z}.$$

Next, compute the second partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} = 0 \qquad \frac{\partial^2 f}{\partial y^2} = 0 \qquad \frac{\partial^2 f}{\partial z^2} = 4xye^{2z}.$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} = e^{2z} \qquad \frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x} = 2ye^{2z} \qquad \frac{\partial^2 f}{\partial z \partial y} = \frac{\partial^2 f}{\partial y \partial z} = 2xe^{2z}.$$

Finally, use Definition 11.2 to find the Hessian of $f$ at $(1, 2, 0)$:

$$Hf(1, 2, 0) = \begin{bmatrix} 0 & e^{2z} & 2ye^{2z} \\ e^{2z} & 0 & 2xe^{2z} \\ 2ye^{2z} & 2xe^{2z} & 4xye^{2z} \end{bmatrix}_{(1,2,0)} = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 2 \\ 4 & 2 & 8 \end{bmatrix}.$$

**Definition 11.3.** If $f : \mathcal{D}^n \to \mathbf{R}$ is a function whose second partial derivatives are continuous at the point $\mathbf{a}$, then the *second-order Taylor approximation of $f$ at* $\mathbf{a}$ is the function

$$T_2(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T Hf(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

**Theorem 19** (Taylor's Theorem I)**.** Suppose that $f : \mathcal{D}^n \to \mathbf{R}$ is a function whose second partial derivatives are continuous at $\mathbf{a}$. If $T_2$ is the second-order Taylor approximation of $f$ at $\mathbf{a}$, then

$$\begin{aligned} T_2(\mathbf{a}) &= f(\mathbf{a}) \\ DT_2(\mathbf{a}) &= Df(\mathbf{a}) \\ HT_2(\mathbf{a}) &= Hf(\mathbf{a}). \end{aligned}$$

Furthermore, all third-order and higher partial derivatives of $T_2$ vanish.

In fact, Taylor approximations of any order $n$ exist for multivariable functions. Since they can be complicated to write down for large $n$, we'll focus on the second-order case.

**Example 11.3.** *Let $f$ be the function defined by $f(x, y, z) = xye^{2x}$. Find the second-order Taylor polynomial of $f$ at $(1, 2, 0)$.*

First, evaluate $f$ at the indicated point to find $f(1, 2, 0) = 2$. Note that we computed the partial derivatives and Hessian of $f$ in Example 11.2. Evaluating these at $(1, 2, 0)$, we see that

$$Df(1, 2, 0) = [2, 1, 4]$$

and

$$Hf(1, 2, 0) = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 2 \\ 4 & 2 & 8 \end{bmatrix}.$$

Putting these together with the formula from Definition 11.3, we have

$$T_2(x, y, z) = 2 + [2, 1, 4] \begin{bmatrix} x-1 \\ y-2 \\ z \end{bmatrix} + \frac{1}{2}[x-1, y-2, z] \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 2 \\ 4 & 2 & 8 \end{bmatrix} \begin{bmatrix} x-1 \\ y-2 \\ z \end{bmatrix},$$

which simplifies to the degree-two polynomial

$$T_2(x, y, z) = -4z + xy + 4xz + 2yz + 4z^2.$$

You can check that this polynomial satisfies the claims of Theorem 19.

Theorem 19 gives an explicit formula for a second-order approximation to a smooth function, and you can find the proof in Section 11.2.1. The next theorem gives some indication of how good this approximation is.

**Theorem 20.** [Taylor's Theorem II] Let $f : \mathcal{D}^n \to \mathbf{R}$ be a smooth function near $\mathbf{a}$, and let $T_2(\mathbf{x})$ be the second-order Taylor approximation of $f$ at $\mathbf{a}$. Then the remainder

$$R_2(\mathbf{x}) = f(\mathbf{x}) - T_2(\mathbf{x})$$

satisfies

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_2(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||^2} = 0.$$

Although we will not prove Theorem 20, you can verify it for a variety of special cases. (See the Worked Examples and Exercise 24.)

The function $R_2$ defined in Theorem 20 measures the difference between $f$ and $T_2$:

$$\begin{aligned} R_2(\mathbf{x}) &= f(\mathbf{x}) - T_2(\mathbf{x}) \\ &= f(\mathbf{x}) - f(\mathbf{a}) - Df(\mathbf{a})(\mathbf{x}-\mathbf{a}) - \frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a}). \end{aligned}$$

Compare this to the remainder $R_1(\mathbf{x})$ that appears in the definition of the derivative:

$$R_1(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{a}) - Df(\mathbf{a})(\mathbf{x}-\mathbf{a}).$$

The remainder function $R_1$ measures the difference between $f$ and its linearization at $\mathbf{a}$. By definition, $\mathbf{f}$ is differentiable at $\mathbf{a}$ if and only if

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_1(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||} = 0.$$

In this case, we say that the remainder $R_1$ vanishes *sublinearly*: faster than the linear term $||\mathbf{x}-\mathbf{a}||$.

In Theorem 20, the condition

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_2(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||^2} = 0$$

states that the remainder vanishes *subquadratically*: faster than the quadratic term $||\mathbf{x}-\mathbf{a}||^2$.

### 11.2.1 Proof of Taylor's Theorem

Taylor's Theorem consists of four separate statements:

1. $T_2(\mathbf{a}) = f(\mathbf{a})$;
2. $DT_2(\mathbf{a}) = Df(\mathbf{a})$;
3. $HT_2(\mathbf{a}) = Hf(\mathbf{a})$; and
4. All third-order and higher partial derivatives vanish.

We leave statements $1$, $2$, and $4$ as exercises and prove only the third statement here.

*Proof.* In order to show that $HT_2(\mathbf{a}) = Hf(\mathbf{a})$, we need to show that $T_2$ and $f$ have exactly the same second partial derivatives at $\mathbf{a}$. Observe that we can rewrite $T_2$ as

$$T_2(\mathbf{x}) = L(\mathbf{x}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a}),$$

where $L(\mathbf{x})$ is the linearization of $f$ at $\mathbf{a}$. Each term in $L(\mathbf{x})$ has degree less than two, so the second partial derivatives of $L(\mathbf{x})$ vanish.

Thus,

$$\frac{\partial^2}{\partial x_i \partial x_j} T_2(\mathbf{x}) = \frac{\partial^2}{\partial x_i \partial x_j}\big(\frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})\big).$$

We need to show that for all $i, j \in \{1, 2, ...n\}$,

$$\big(\frac{\partial^2}{\partial x_i \partial x_j}\big(\frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})\big)\big)(\mathbf{a}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}).$$

In order to facilitate the proof, express a and x in coordinates as

$$\begin{aligned}\mathbf{a} &= (a_1, a_2, ...a_n)\\ \mathbf{x} &= (x_1, x_2, ...x_n).\end{aligned}$$

Expand $(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})$ in terms of these coordinates (Exercise 21):

$$(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\big(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})\big)(x_i - a_i)(x_j - a_j). \tag{11.1}$$

Fix some $k \in \{1, 2, ....n\}$ and differentiate Equation 11.1 with respect to $x_k$:

$$\begin{aligned}\frac{\partial}{\partial x_k}[(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})] &= \frac{\partial}{\partial x_k}[\sum_{i=1}^{n}\sum_{j=1}^{n}\big(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})\big)(x_i - a_i)(x_j - a_j)]\\ &= \sum_{j=1}^{n}\big(\frac{\partial^2 f}{\partial x_k \partial x_j}(\mathbf{a})\big)(x_j - a_j) + \sum_{i=1}^{n}\big(\frac{\partial^2 f}{\partial x_i \partial x_k}(\mathbf{a})\big)(x_i - a_i).\end{aligned}$$

Now pick some $l \in \{1, 2, ...n\}$ and differentiate this equation with respect to $x_l$:

$$\begin{aligned}\frac{\partial^2}{\partial x_l \partial x_k}\big(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})\big) &= \frac{\partial}{\partial x_l}\big(\sum_{j=1}^{n}\big(\frac{\partial^2 f}{\partial x_k \partial x_j}(\mathbf{a})\big)(x_j - a_j) + \sum_{i=1}^{n}\big(\frac{\partial^2 f}{\partial x_i \partial x_k}(\mathbf{a})\big)(x_i - a_i)\big)\\ &= \frac{\partial^2 f}{\partial x_k \partial x_l}(\mathbf{a}) + \frac{\partial^2 f}{\partial x_l \partial x_k}(\mathbf{a})\\ &= 2\frac{\partial^2 f}{\partial x_k \partial x_l}(\mathbf{a}).\end{aligned}$$

The last equality follows from Clairaut's Theorem, which asserts the equality of continuous mixed partial derivatives. Dividing by 2 yields

$$\frac{\partial^2}{\partial x_k \partial x_l}\big(\frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})\big) = \frac{\partial^2 T_2}{\partial x_k \partial x_l}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_k \partial x_l}(\mathbf{a}).$$

Note that each second partial derivative of $T_2$ is a constant function. Thus for all $k$ and $l$, we have

$$\frac{\partial^2 T_2}{\partial x_k \partial x_l}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_k \partial x_l}(\mathbf{a})$$

This proves that

$$HT_2(\mathbf{a}) = Hf(\mathbf{a}).$$

□

## 11.3 Worked Examples

**Example 11.4.** *Find the second-order Taylor approximation for $f(x,y) = x^2y - 3xy + y^3$ at $(0,0)$. Then verify that Theorem 20 holds for this approximation by showing*

$$\lim_{(x,y)\to(0,0)} \frac{R_x(x,y)}{||(x,y)-(0,0)||^2} = \lim_{(x,y)\to(0,0)} \frac{f(x,y) - T_2(x,y)}{||(x,y)-(0,0)||^2} = 0.$$

We begin by computing the second-order Taylor approximation for $f$ at the origin. This requires us to find $f(0,0)$, $Df(0,0)$, and $Hf(0,0)$, so we compute the first and second partial derivatives of $f$ at $(0,0)$:

$$\begin{aligned}
\frac{\partial f}{\partial x}(0,0) &= (2xy - 3y)|_{(0,0)} = 0 \\
\frac{\partial f}{\partial y}(0,0) &= (x^2 - 3x + 3y^2)|_{(0,0)} = 0 \\
\frac{\partial^2 f}{\partial x^2}(0,0) &= (2y)|_{(0,0)} = 0 \\
\frac{\partial^2 f}{\partial y^2}(0,0) &= (6y)|_{(0,0)} = 0 \\
\frac{\partial^2 f}{\partial y \partial x}(0,0) &= \frac{\partial^2 f}{\partial x \partial y}(0,0) = (2x-3)|_{(0,0)} = -3.
\end{aligned}$$

Thus,

$$Df(0,0) = \begin{bmatrix} 0 & 0 \end{bmatrix} \text{ and } Hf(0,0) = \begin{bmatrix} 0 & -3 \\ -3 & 0 \end{bmatrix}.$$

Following Definition 11.3, we can compute $T_2$:

$$\begin{aligned} T_2(x,y) &= f(0,0) + Df(0,0)\begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2}[x,y]Hf(0,0)\begin{bmatrix} x \\ y \end{bmatrix} \\ &= 0 + [0,0]\begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2}[x,y]\begin{bmatrix} 0 & -3 \\ -3 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\ &= \frac{1}{2}[x,y]\begin{bmatrix} -3y \\ -3x \end{bmatrix} \\ &= -3xy \end{aligned}$$

Having shown $T_2(x,y) = -3xy$, we will show that the difference between $T_2$ and $f$ vanishes subquadratically as $(x,y) \to (0,0)$. Theorem 20 defines the remainder $R_2$ as the difference between $f$ and $T_2$:

$$\begin{aligned} R_2(x,y) &= f(x,y) - T_2(x,y) \\ &= x^2y - 3xy + y^3 - (-3xy) \\ &= x^2y + y^3. \end{aligned}$$

To verify Theorem 20 in this case, we need to show that

$$\lim_{(x,y)\to(0,0)} \frac{R_2(x,y)}{||(x,y)-(0,0)||^2} = 0.$$

Evaluate this limit:

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{R_2(x,y)}{||(x,y)-(0,0)||^2} &= \lim_{(x,y)\to(0,0)} \frac{x^2y+y^3}{x^2+y^2} \\ &= \lim_{(x,y)\to(0,0)} \frac{y(x^2+y^2)}{x^2+y^2} \\ &= \lim_{(x,y)\to(0,0)} y \\ &= 0. \end{aligned}$$

The last equality follows from the continuity of the polynomial $y$.

Since the remainder $R_2$ vanishes subquadratically, Theorem 20 holds for $f$ at $(0,0)$.

## 11.4 Exercises

In Exercises 1-8, compute the linearization of the given function at the indicated point.

1. $h(t) = t^3 + 2t^2 - t - 7,\ t = 1.$
2. $\mathbf{f}(x,y,z) = (x^2 \log y,\ z/x,\ y\sin(x^2-z^2)),\ \mathbf{a} = (2,1,-1).$
3. $\mathbf{g}(u,v) = \big((u^2+1)^{-3/2},\ (v^2+1)^{-5/2},\ uv^2\big),\ \mathbf{a} = (-1,5).$

4. $h(r,s,t) = r^9 + \cos st$, $\mathbf{a} = (-1,0,14)$.

5. $\mathbf{p}(u,v,w) = (17u + v^2w,\ 13uvw)$, $\mathbf{a} = (0,1,1)$.

6. $f(m,n) = \ln(m^2 + n^2 + 1)$, $\mathbf{a} = (7,3)$.

7. $q(w,x,y,z) = xy + zw^2 - 3yzw$, $\mathbf{a} = (2,1,-1,0)$.

8. $g(q,r,s) = \frac{q^2+2rs}{\sqrt{q^2+r^2+s^2}}$, $\mathbf{a} = (1,1,1)$

In Exercises 9-14, compute the second order Taylor approximation of the given function at the indicated point. Then use this result to give the desired approximation. (You do not need a calculator.)

9. $f(u,v) = u^2\cos v + v\sin u$, $\mathbf{a} = (0,0)$. Approximate $f(0.1, 0.2)$.

10. $h(t,u,v) = t^2 + 3uv - 11t^3uv^6 + 17t^{100}u^{32}v^{14}$, $\mathbf{a} = (0,0,0)$. Approximate $h(0.03, 0.09, -0.02)$.

11. $b(w,z) = w^{3/2} + z^{5/2}$, $\mathbf{a} = (1,4)$. Approximate $b(1.02, 3.96)$.

12. $g(x,y) = \ln(xy - 1)$, $\mathbf{a} = (2,1)$. Approximate $g(1.8, .9)$.

13. $p(x,y,z) = xy + zy^2$, $\mathbf{a} = (2,1,-1)$. Approximate $p(2.03, 1.05, -0.97)$.

14. $q(u,v) = uve^{-(u^2-9v^2)}$, $\mathbf{a} = (3,1)$. Approximate $q(3.03, 0.98)$.

15. Let $\mathbf{g} : \mathbf{R}^2 \to \mathbf{R}^3$ be given by $\mathbf{g}(x,y) = (x + y - 7, 4x - 12y + 38, -6y + 2x - 1)$.

    (a) Compute the linearization of $\mathbf{g}$ at the point $(1,0)$.

    (b) Compute the linearization of $\mathbf{g}$ at the point $(102, -96)$.

    (c) Given any point $\mathbf{a} \in \mathbf{R}^2$, write down the linearization of $\mathbf{g}$ at $\mathbf{a}$. What special property of $\mathbf{g}$ makes this so easy?

16. Let $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^m$ be a differentiable function, and pick some point $\mathbf{a} \in \mathbf{R}^n$. Consider the following translations:

$$\mathbf{h} : \mathbf{R}^n \to \mathbf{R}^n \text{ defined by } \mathbf{h}(\mathbf{x}) = \mathbf{x} - \mathbf{a};$$
$$\mathbf{g} : \mathbf{R}^m \to \mathbf{R}^m \text{ defined by } \mathbf{g}(\mathbf{x}) = \mathbf{x} + \mathbf{f}(\mathbf{a}).$$

    Show that the linearization of $\mathbf{f}$ at $\mathbf{a}$ is the function $L(\mathbf{x}) = (\mathbf{g} \circ D\mathbf{f}(\mathbf{a}) \circ \mathbf{h})(\mathbf{x})$.

17. Complete the proof of Theorem 19 by showing that the following statements are true:

    (a) $T_2(\mathbf{a}) = f(\mathbf{a})$

    (b) $DT_2(\mathbf{a}) = Df(\mathbf{a})$

(c) The third-order and higher partial derivatives of $T_2$ vanish.

18. Let $p(x, y) = 3 + x + 3xy - 7y^2$.

    (a) Describe all the polynomials whose second order Taylor approximation at $(0, 0)$ is $p(x, y)$.

    (b) Write $p(x, y)$ as a polynomial in powers of $x - 1$ and $y - 1$. Now describe all the polynomials whose second order Taylor approximation at $(1, 1)$ is $p(x, y)$.

    (c) Generalize the results above to describe all the polynomials whose second order Taylor approximation at $(a, b)$ is $p(x, y)$.

19. Verify Theorem 20 for all polynomials $p(x, y)$.

20. Consider the equation

$$xz^2 + y^2z^5 = 19. \tag{11.2}$$

The point $(3, 4, 1)$ is a solution to this equation, but there are many more solutions. If we treat $x$ and $y$ as independent variables, then in a neighborhood of $(3, 4, 1)$, Equation 11.2 determines $z$ as a function of $x$ and $y$. We say that this function $z(x, y)$ is defined *implicitly* by Equation 11.2. Writing an explicit formula for this function is difficult, but you can nevertheless compute the linearization of the implicitly-defined function $z(x, y)$ at the point $(3, 4)$.

    (a) Treating $x$ and $y$ as independent variables and treating $z$ as a function of $x$ and $y$, differentiate both sides of Equation (11.2) with respect to $x$. Solve for $\frac{\partial z}{\partial x}$ and evaluate this function at $(3, 4)$. This technique is called *implicit differentiation.*

    (b) Use implicit differentiation to compute $\frac{\partial z}{\partial y}(3, 4)$.

    (c) Write down the linearization $L(x, y)$ of $z(x, y)$ at $(3, 4)$.

    (d) Use $L$ to approximate the value $z(3.01, 4.02)$. Plug your approximate result into Equation 11.2 to see if it approximately satisfies the equation.

21. *Higher order Taylor expansions.*

Although it can be difficult to write down higher-order Taylor expansions in a compact form, it is possible to express them using summation notation. Suppose that $f : \mathcal{D}^n \to \mathbf{R}$ is a smooth function, and let $T_2$ be the second order Taylor approximation of $f$ at $\mathbf{a} = (a_1, a_2, ...a_n)$.

    (a) Show that

$$T_2(\mathbf{x}) = L(\mathbf{x}) + \frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})\ (x_i - a_i)(x_j - a_j),$$

where $L(\mathbf{x})$ is the linearization of $f$ around $\mathbf{a}$.

(b) Define

$$T_3(\mathbf{x}) = T_2(\mathbf{x}) + \frac{1}{3!}\sum_{i,j,k=1}^{n} \big(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{a})\big)(x_i - a_i)(x_j - a_j)(x_k - a_k).$$

Show that $T_3$ is the third order Taylor polynomial for $f$. That is, show that

$$T_3^{(i)}(\mathbf{a}) = \begin{cases} f^{(i)}(\mathbf{a}) & \text{if } 0 \leq i \leq 3 \\ 0 & i > 3. \end{cases}$$

(c) Guess an expression for the $4^{th}$ order Taylor polynomial of $f$.

22. *Secant line approximations.*

Let $g : \mathbf{R} \to \mathbf{R}$ be a differentiable function. A secant line to $\Gamma_g$ is line in $\mathbf{R}^2$ that contains two points on $\Gamma_g$. A *secant approximation* of $g$ is a function whose graph is a secant line to $\Gamma_g$.

(a) Find an explicit formula for the secant line to $\Gamma_g$ that contains the points $(a, g(a))$ and $(b, g(b))$. Use your answer to find an equation for the corresponding secant approximation $S : \mathbf{R} \to \mathbf{R}$.

(b) Suppose we want to approximate the function $g(x) = x^3$ near $x = 0$. Compute the secant line to $\Gamma_g$ for $a = -1$ and $b = 1$ and find an equation for the corresponding secant approximation $S$. Now compute the linearization $L$ of $g$ at $x = 0$.

(c) Suppose that we care about approximating $g$ on the interval $[-2, 2]$. For $x \in [-2, 2]$, what is the largest possible error in each approximation?

$$E_S = \max_{-2 \leq x \leq 2} |g(x) - S(x)|$$

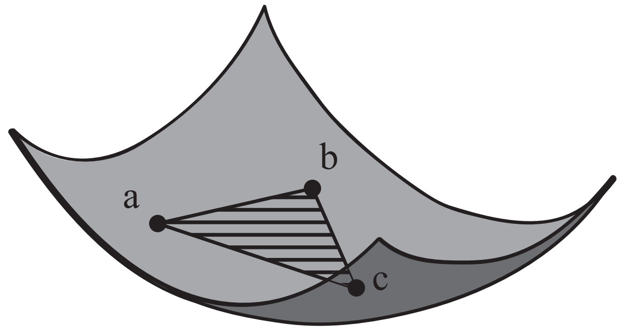
$$E_L = \max_{-2 \leq x \leq 2} |g(x) - L(x)|.$$

Which approximation is better on this interval?

(d) Finally, suppose we care more about the accuracy of our approximations on the interval $[-.001, .001]$. Compute the maximum errors on this smaller interval. Now which approximation is better?

23. *Secant plane approximations*

Let $f : \mathbf{R}^2 \to \mathbf{R}$ be a differentiable function, and let $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ be three points on $\Gamma_f$. Define the *secant plane* to $\Gamma_f$ as the plane in $\mathbf{R}^3$ which contains $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$.




(a) Let $f(x, y) = x^2 + 2y$. Find the linearization of $f$ at the point $(0, 0)$. Use your result to find equation for the tangent plane to $\Gamma_f$ at $(0, 0, 0)$.

(b) Find an equation for the secant plane to $\Gamma_f$ that contains the points $(1, 0)$, $(0, 1)$, $(-1, -1)$.

(c) Letting $\epsilon$ be a positive constant, find an equation for the secant plane to $\Gamma_f$ that contains the points $(\epsilon, 0)$, $(0, \epsilon)$, $(-\epsilon, -\epsilon)$.

(d) What happens to the equation for the secant plane as $\epsilon \to 0$? Compare this equation to the equation for the tangent plane from the first part of the exercise.

24. Consider the function $f(t) = t^{\frac{7}{3}}$.

(a) Compute the second order Taylor expansion $T_2(t)$ of $f(t)$ at $t = 0$. Compute the remainder $R_2(t)$.

(b) Show explicitly that $R_2$ satisfies $\lim_{t\to 0} \frac{R_2(t)}{t^2} = 0$, as is guaranteed by Taylor's theorem.

(c) Show that there is no constant $K > 0$ such that $|R_2(t)| < K|t|^3$ in a neighborhood of $t = 0$.

# 12 Extrema of Multivariable Functions

When studying the physical world, we are often interested in superlatives like "most", "biggest", "fastest", or "warmest". For example, a business model could try to predict the most profitable allocation of funds for advertising, product development, and distribution. Given a fixed quantity of building materials, a manufacturer might want to design a container with the largest possible volume.

In a mathematical model, answers to these questions often take the form of maxima or minima of related functions. In the next three sections we develop tools for finding the extreme points of real-valued differentiable functions.

## 12.1 Maxima and minima

**Definition 12.1.** Let $f$ be a real-valued function defined on $\mathcal{D}^n \subset \mathbf{R}^n$.
The function $\mathbf{f}$ has a *local maximum* at $\mathbf{a}$ if there exists some $\epsilon > 0$ such that

$$f(\mathbf{a}) \geq f(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathcal{D}^n \text{ satisfying } ||\mathbf{x} - \mathbf{a}|| < \epsilon.$$

Similarly, $\mathbf{f}$ has a *local minimum* at $\mathbf{a}$ if there exists some $\epsilon > 0$ such that

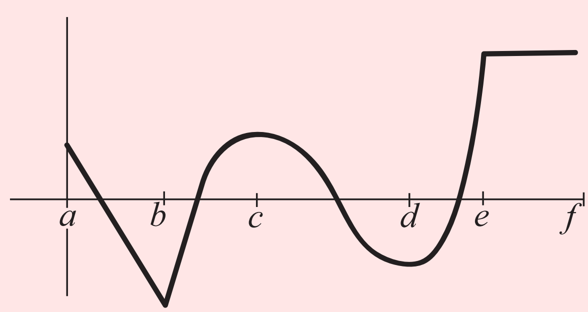
$$f(\mathbf{a}) \leq f(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathcal{D}^n \text{ satisfying } ||\mathbf{x} - \mathbf{a}|| < \epsilon.$$

The collection of all local maxima and minima for a function $f$ is called the set of *extrema of* $f$ .

If $f(\mathbf{a}) \geq f(\mathbf{x})$ for all $\mathbf{x}$ in the domain of $f$, then $f$ has a *global maximum*, or an *absolute maximum*, at $\mathbf{a}$. If $f$ has an absolute maximum at $\mathbf{a}$, we call $\mathbf{f}(\mathbf{a})$ the *maximum value* of $\mathbf{f}$. If $f(\mathbf{a}) \leq f(\mathbf{x})$ for all $\mathbf{x}$ in the domain of $f$, then $f$ has a *global minimum*, or an *absolute minimum*, at $\mathbf{a}$, and we call $\mathbf{f}(\mathbf{a})$ the *minimum value* of $f$.

**Example 12.1.** *Let $g : [a, f] \to \mathbf{R}$ be the function whose graph is shown below. Classify the extrema of $f$.*

The graph shows that $g$ has local maxima at $a$, $c$, and each point in the interval $[e, f]$. The function has local minima at $b$, $d$, and each point in $(e, f]$. The global minimum occurs at $b$. However, global extrema are not necessarily unique; $g$ attains a global maximum at each point in $[e, f]$.



## 12.2 The First Derivative Test for multivariable functions

If the differentiable function $f : \mathbf{R} \to \mathbf{R}$ has a local extremum at a point $a$, then $f'(a) = 0$. In this section we show an analogous statement holds for functions of several variables.

**Definition 12.2.** Given $f : \mathcal{D}^n \to \mathbf{R}$, the point $\mathbf{a} \in \mathcal{D}^n$ is a *critical point* of $f$ if one of the following holds:

- $\nabla f(\mathbf{a}) = \mathbf{0}$; or
- $f$ is not differentiable at $\mathbf{a}$.

This definition allows us to state the following important theorem:

**Theorem 21** (The First Derivative Test)**.** Suppose that $f : \mathbf{R}^n \to \mathbf{R}$ has a local extremum at $\mathbf{a} \in \mathbf{R}^n$. Then $\mathbf{a}$ is a critical point of $\mathbf{f}$.

Note that we've only stated the First Derivative Test for functions defined on $\mathbf{R}^n$. In Section 13 we'll give a version of this theorem for functions whose domains are proper subsets of some Euclidean space.

### 12.2.1 Proof of the First Derivative Test

We'll prove the First Derivative Test for local minima, and the proof for local maxima is similar. (See Exercise 12.)

*Proof.* Suppose that $f : \mathbf{R}^n \to \mathbf{R}$ has a local minimum at $\mathbf{a}$. If $f$ is not differentiable at $\mathbf{a}$, then $\mathbf{a}$ is a critical point and Theorem 21 holds. Therefore, suppose that $f$ is differentiable at $\mathbf{a}$.

Fix a vector $\mathbf{v} \in \mathbf{R}^n$ and let $\mathbf{x} = \mathbf{a} + t\mathbf{v}$ be the line in $\mathbf{R}^n$ with direction vector $\mathbf{v}$ that passes through $\mathbf{a}$.

Since $f$ is differentiable at $\mathbf{a}$, the directional derivative $D_{\mathbf{v}} f(\mathbf{a})$ exists:

$$\lim_{t \to 0^-} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} = D_{\mathbf{v}} f(\mathbf{a}) = \lim_{t \to 0^+} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t}. \tag{12.1}$$

Because $\mathbf{a}$ is a local minimum,

$$f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) \geq 0.$$

This implies that the numerator in both one-sided limits is non-negative. Thus, when $t$ approaches $0$ from below,

$$\frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} \leq 0.$$

On the other hand, when $t$ approaches $0$ from above,

$$\frac{f(\mathbf{a}+t\mathbf{v})-f(\mathbf{a})}{t} \geq 0.$$

Equation 12.1 therefore implies $D_{\mathbf{v}}f(\mathbf{a}) = 0$. This argument holds for all $\mathbf{v} \in \mathbf{R}^n$, so $D_{\mathbf{e_i}}f(\mathbf{a}) = 0$ for $i \in \{1, 2, ...n\}$. Thus, $\nabla f(\mathbf{a}) = \mathbf{0}$ as desired.

□

## 12.3 The Second Derivative Test for multivariable functions

The First Derivative test helps us identify the critical points of a differentiable function, and in this section we use second derivatives to help us classify them.

### 12.3.1 Saddle points

As with single variable functions, not every critical point of a multivariable function corresponds to a local extremum.

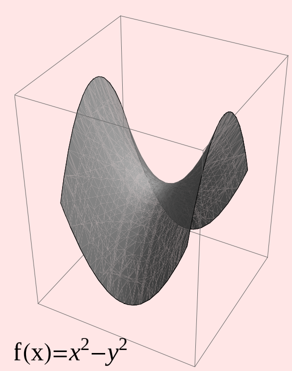
**Definition 12.3.** A function $f : \mathcal{D}^n \to \mathbf{R}$ has a *saddle point* at $\mathbf{a}$ if $f$ is differentiable at $\mathbf{a}$, $\nabla f(\mathbf{a}) = \mathbf{0}$, and $f$ has neither a local maximum nor a local minimum at $\mathbf{a}$.

**Example 12.2.** *Let $f(x, y) = x^2 - y^2$. Show that $f$ has a saddle point at $(0, 0)$.*
The function $f$ is a polynomial, so $f$ is certainly differentiable at $(0, 0)$. Next, we compute

$$\nabla f(0,0) = \begin{bmatrix} 2x \\ -2y \end{bmatrix}_{(0,0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

To complete the problem, we must show that $(0, 0)$ isn't an extremum of $f$.

- Moving away from the origin along the $x$-axis increases the value of $f$, so $f$ does not have a local maximum at $(0, 0)$.
- Moving away from the origin along the $y$-axis decreases the value of $f$, so $f$ does not have a local minimum at $(0, 0)$.



f(x)=$x^2-y^2$

This shows that $(0, 0)$ is a saddle point.

The term “saddle point” comes from the shape of the graph $\Gamma_f$ in Example 12.2. However, the next example shows that the definition also makes sense for functions of more than two variables.

**Example 12.3.** *Classify the critical points of the function* $F(x, y, z) = x^2 + y^2 - z^2$.
Since the polynomial $F$ is everywhere differentiable, the only critical points are points where the gradient vanishes. To find the critical points, compute the gradient of $F$ and solve the equation $\nabla F(\mathbf{x}) = \mathbf{0}$:

$$\nabla F(x, y, z) = \begin{bmatrix} 2x \\ 2y \\ -2z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The only solution is $(0, 0, 0)$, so $F$ has just one critical point. We claim $(0, 0, 0)$ is a saddle point.
First, show that $F$ does not have a local minimum at $(0, 0, 0)$. If we restrict $F$ to the $z$-axis, we get

$$F(0, 0, z) = -z^2.$$

Increasing $z$ decreases $F(0, 0, z)$, so the point $(0, 0, 0)$ is not a local minimum for $F$.
In order to show that $F$ does not have a local maximum at $(0, 0, 0)$, restrict $F$ to the $xy$-plane:

$$F(x, y, 0) = x^2 + y^2.$$

Increasing either $x$ or $y$ increases $F(x, y, 0)$, so $(0, 0, 0)$ is not a local maximum for $F$.
Since $F$ has neither a local minimum nor a local maximum at $(0, 0, 0)$, it follows that the critical point $(0, 0, 0)$ is a saddle point.

### 12.3.2 The Second Derivative Test

Section 11 defined the Hessian of a function $f$ at the point $\mathbf{a}$ as the matrix of second partial derivatives:

$$Hf(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \dots & \dots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{a}) \end{bmatrix}.$$

If the second partial derivatives of $f$ are continuous at $\mathbf{a}$, then Clairaut’s Theorem implies that the mixed partial derivtives are equal (Theorem 13 in Section 7):

$$\frac{\partial f^2}{\partial x_i \partial x_j}(\mathbf{a}) = \frac{\partial f^2}{\partial x_j \partial x_i}(\mathbf{a}).$$

In this case, the Hessian $Hf(\mathbf{a})$ is a symmetric matrix. Recall that any $n \times n$ symmetric matrix $M$ defines a quadratic form $Q : \mathbf{R}^n \to \mathbf{R}$, where

$$Q(\mathbf{x}) = \mathbf{x}^T M \mathbf{x}.$$

Furthermore, $M$ has $n$ real eigenvalues, and the signs of these eigenvalues classify the quadratic form $Q_M$:

**Theorem 22** (Proposition 26.1 in Levandosky's *Linear Algebra*)**.** Suppose that $Q_M$ is the quadratic form generated by the $n \times n$ symmetric matrix $M$:

- If all the eigenvalues of $M$ are positive, $Q_M$ is positive definite.
- If all the eigenvalues of $M$ are negative, $Q_M$ is negative definite.
- If $M$ has both positive and negative eigenvalues, $Q_M$ is indefinite.

This classification is the key to the multivariable Second Derivative Test.

**Theorem 23.** [The Second Derivative Test] Suppose that $f : \mathcal{D}^n \to \mathbf{R}$ is a function with the property that all the second partial derivatives $\frac{\partial f^2}{\partial x_i \partial x_j}(\mathbf{x})$ are continuous at $\mathbf{a}$, and suppose further that $\nabla f(\mathbf{a}) = \mathbf{0}$. Let $Q_{\mathbf{a}}$ denote the quadratic form generated by $Hf(\mathbf{a})$.
Then the following statements hold:

- If $Q_{\mathbf{a}}$ is positive definite, then $f$ has a local minimum at $\mathbf{a}$;
- If $Q_{\mathbf{a}}$ is negative definite, then $f$ has a local maximum at $\mathbf{a}$;
- If $Q_{\mathbf{a}}$ is indefinite, then $f$ has a saddle point at $\mathbf{a}$.

Notice that if $Q_{\mathbf{a}}$ is positive or negative semidefinite, then the Second Derivative Test does not classify the critical point $\mathbf{a}$.

The proof of the Second Derivative Test may be found in Section 15.4 in the Appendix.

**Example 12.4.** *Classify the critical points of* $F(x, y, z) = x^2 + y^2 + z^2 - 4x + 2y - 6z$.
We first note that the function $F$ is a polynomial, so all its derivatives are continuous and the First and Second Derivative Tests apply.
In order to find the critical points of $F$, we begin by computing the gradient.

$$\nabla F(x,y,z) = \begin{bmatrix} 2x-4 \\ 2y+2 \\ 2z-6 \end{bmatrix}$$

Setting $\nabla F(x,y,z) = \mathbf{0}$ shows that $(2,-1,3)$ is the unique critical point of $F$. Differentiating again, compute the Hessian of $F$ at $(2,-1,3)$:

$$HF(2,-1,3) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Since $HF(2,-1,3)$ is a diagonal matrix, the eigenvalues are simply the diagonal entries. It's easy to see that the associated quadratic form is positive definite, and the Second Derivative Test implies that $F$ has a local minimum at $(2,-1,3)$.

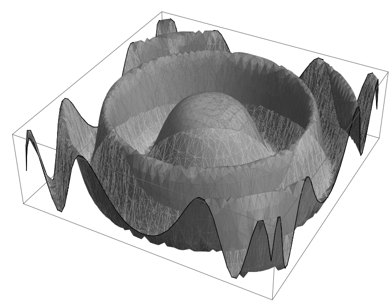
## 12.4 Worked Examples

**Example 12.5.** *Find the local and global extrema of $f(x,y) = \cos(x^2+y^2)$.*

First, we note that the domain of $f$ is all of $\mathbf{R}^2$, and that $f$ is a smooth function. This means implies that every local extremum is a solution to the equation $\nabla f(x,y) = \mathbf{0}$.

We compute

$$\nabla f(x,y) = \begin{bmatrix} -2x\sin(x^2+y^2) \\ -2y\sin(x^2+y^2) \end{bmatrix}.$$



We see one solution immediately:

$$\nabla f(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The origin is not the only critical point, however, as the gradient vanishes at any point which satisfies

$$x^2+y^2 = n\pi \quad \text{for } n \in \mathbb{Z}. \tag{12.2}$$

However, the Second Derivative Test is inconclusive when applied to the critical points of $f$. (See Exercise 16.) We can nevertheless complete the problem using an algebraic argument. Because the range of the cosine function is $[-1,1]$, the maximum value in the range of $f$ is no larger than $1$. Thus, $f$ attains an absolute maximum at any point $(x,y)$ such that $f(x,y) = 1$. Similarly, $f$ attains an absolute minimum at any point $(x,y)$ where $f(x,y) = -1$.

Observe that

$$f(x,y) = \begin{cases} 1 & \text{if } x^2+y^2 = n\pi, \text{where } n \text{ is an even integer} \\ -1 & \text{if } x^2+y^2 = n\pi, \text{where } n \text{ is an odd integer} \end{cases}$$

Since this description includes all of the critical points of $f$, we have a complete classification of the extrema of $f$.

## 12.5 Exercises

1. Can you construct a continuous function $f : \mathbf{R} \to \mathbf{R}$ with exactly $4$ local maxima and $2$ local minima? Why or why not?

2. For each of the following, provide a function $f : \mathcal{D}^n \to \mathbf{R}$ which satisfies the given condition:

   (a) Every point in $\mathcal{D}^n$ is a local maximum.

   (b) One local maximum is also a local minimum.

   (c) One absolute maximum is also an absolute minimum.

   (d) There are infinitely-many critical points.

   (e) No critical point is a local extremum.

   (f) There are twice as many local maxima as local minima.

3. Let $f(x, y, z) = xy + y^2 \sin(2x) + z^2$.

   (a) Show that $(0, 0, 0)$ is a critical point of $f$.

   (b) Compute $Hf(0, 0, 0)$.

   (c) Classify $(0, 0, 0)$ as a local minimum, a local maximum, or a saddle point.

4. Let $f(x, y) = x^2 - 4xy + 5y^2$.

   (a) Show that $(0, 0)$ is a critical point of $f$.

   (b) Compute $Hf(0, 0)$.

   (c) Classify $(0, 0)$ as a local minimum, a local maximum, or a saddle point.

5. Let $f(x, y) = (\cos x)(\ln y)$.

   (a) Show that $(\frac{\pi}{2}, 1)$ is a critical point of $f$.

   (b) Compute $Hf(\frac{\pi}{2}, 1)$.

   (c) Classify $(\frac{\pi}{2}, 1)$ as a local minimum, a local maximum, or a saddle point.

In Exercises 6 through 11, find the critical points of the given function and classify each one.

6. $f(x, y) = xy - \ln x - \frac{1}{y}$

7. $f(x, y) = 2x - 3y + \ln xy$

8. $f(x, y, z) = xy + xz + 2yz + \frac{1}{x}$

9. $f(x, y) = ye^x + x - 2y = 0$

10. $f(x, y, z) = xyz$

11. $f(s, t) = \cos s \sin t$

12. Prove the First Derivative Test for local maxima.

13. Under what conditions on the constant $k$ does the Second Derivative Test guarantee that the function
$$f(x, y) = x^2 + kxy + y^2$$
has a local minimum at $(0, 0)$?

14. Under what conditions on the constants $a$ and $b$ does the Second Derivative Test guarantee that the function
$$g(x, y, z) = ax^2 + 2axz + by^2 - 2byz + z^2$$
has a local maximum at $(0, 0, 0)$? A local minimum at $(0, 0, 0)$?

15. Under what conditions on the constants $a$ and $b$ does the Second Derivative Test guarantee that the function
$$f(x, y) = ax^2 - 2ax + xy - y + by^2$$
has a local minimum at $(1, 0)$?

16. Consider the function $f(x, y) = \cos(x^2 + y^2)$ studied in Example 12.5.
    (a) Describe the critical points of $f$.
    (b) Show that the Second Derivative Test is inconclusive at every critical point of $f$.

17. The graph of $g(x, y) = x^3 - 3xy^2$ is known as the "monkey saddle". Show that the origin is a saddle point for $g$.

18. Let $f(x, y) = x^2 + y^3$. Classify the critical point $(0, 0)$ as a local maximum, a local minimum, or a saddle point.

19. This problem will help you prove the third statement of the Second Derivative Test. Suppose that $f : \mathcal{D}^n \to \mathbf{R}$ is a differentiable function, and suppose that $\mathbf{a}$ is a critical point of $f$ such that $Hf(\mathbf{a})$ has both positive and negative eigenvalues.
    (a) Let $\mathbf{v}$ be an eigenvector for $Hf(\mathbf{a})$ whose eigenvalue is positive. Show that the restriction of $f$ to the line $\{\mathbf{a} + t\mathbf{v} \mid t \in \mathbf{R}\}$ has a local minimum at $\mathbf{a}$.
    (b) Let $\mathbf{w}$ be an eigenvector for $Hf(\mathbf{a})$ whose eigenvalue is negative. Show that the restriction of $f$ to the line $\{\mathbf{a} + s\mathbf{w} \mid s \in \mathbf{R}\}$ has a local maximum at $\mathbf{a}$.
    (c) Use the first two parts of this problem to deduce that $\mathbf{a}$ is a saddle point.

20. If $f : \mathbf{R}^n \to \mathbf{R}$ attains a local minimum at $\mathbf{a}$, then the restriction of $f$ to any line through $\mathbf{a}$ attains a local minimum at $\mathbf{a}$. In this problem, you will show that the converse is not true.

    Let $f(x, y) = (y - x^2)(y - 3x^2)$.
    (a) Show that $f$ has a critical point at $(0, 0)$ and compute $Hf(0, 0)$.

(b) Use the single variable Second Derivative Test to show that the restriction of $f$ to any line through the origin has a local minimum at $(0, 0)$.

(c) Show that $(0, 0)$ is not a local minimum for $f$.

# 13 Extrema II: Global Extrema

The previous section developed techniques for identifying and classifying local extrema. In this section, we focus on global extrema. We will pay particular attention to functions which are defined on proper subsets of some Euclidean space.

This topic has a lot of material associated to it, so we summarize the main points in Section 13.5, immediately before the Worked Examples.

## 13.1 Interior and boundary points

When a function is defined on a proper subset of $\mathbf{R}^n$, our techniques for finding extrema will distinguish between *interior* and *boundary* points of the domain. In this section, we make these definitions precise.

**Definition 13.1.** Let $S \subset \mathbf{R}^n$. The point $\mathbf{a} \in \mathbf{R}^n$ is a *boundary point of* $S$ if for every $\epsilon > 0$, the ball

$$B(\mathbf{a}, \epsilon) = \{\mathbf{x} \in \mathbf{R}^n \big| \, ||\mathbf{x} - \mathbf{a}|| < \epsilon\}$$

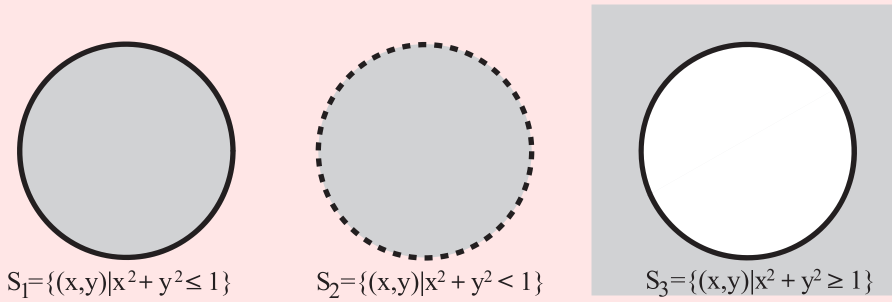
contains some point which is in $S$ and some point which is not in $S$.
The set of boundary points of $S$ is called the *boundary of* $S$.

**Example 13.1.** *Describe the boundary of the interval* $[0, 1)$ *in* $\mathbf{R}$.
The boundary points of $[0, 1)$ are the endpoints of the interval.
To see that $0$ is a boundary point, observe that any interval $(-\epsilon, \epsilon)$ contains some points which are not in $[0, 1)$ and some points which are in $[0, 1)$. Applying the same argument to the interval $(1 - \epsilon, 1 + \epsilon)$ shows that $1$ is a boundary point of $[0, 1)$.
Any other point lies in an interval completely contained in either $(-\infty, 0)$, $(0, 1)$ or $(1, \infty)$, so there are no other boundary points.

Example 13.1 demonstrates that a boundary point of $S$ may be in $S$ or in the complement of $S$.

**Definition 13.2.** Let $S \subset \mathbf{R}^n$. The point $\mathbf{a} \in \mathbf{R}^n$ is an *interior point of* $S$ if there exists $\epsilon > 0$ such that $S$ contains the ball $B(\mathbf{a}, \epsilon)$. The set of interior points of $S$ is called the *interior of* $S$.

**Example 13.2.** *Describe the boundary and interior of each of the three subset of* $\mathbf{R}^2$ *shown below.*




Each of these sets has the same boundary, the circle

$$\{(x,y) \mid x^2 + y^2 = 1\}.$$

The sets $S_1$ and $S_2$ have the same interior, the open disc

$$\{(x,y) \mid x^2 + y^2 < 1\}.$$

The interior of the set $S_3$ is the complement of the closed disc:

$$\{(x,y) \mid x^2 + y^2 > 1\}.$$

It's worth noting that any point in a set $S$ is either in the interior of $S$ or in the boundary of $S$. (See Exercise 1 ).

## 13.2 The First Derivative Test revisited

With the definitions from the previous section, we can now state a version of the First Derivative Test for functions defined on proper subsets of $\mathbf{R}^n$.

**Theorem 24** (The First Derivative Test)**.** Suppose that $f : \mathcal{D}^n \to \mathbf{R}$ has a local extremum at $\mathbf{a} \in \mathcal{D}^n$. Then $\mathbf{a}$ is a critical point of $\mathbf{f}$ or $\mathbf{a}$ is a boundary point of $\mathcal{D}^n$.

You should think about why the proof of the First Derivative Test from Section 12.2.1 covers this case, as well.

## 13.3 Existence of extrema

An important theorem in single variable calculus states that a continuous function defined on a closed interval always attains an absolute maximum and an absolute minimum on the interval. In order to generalize this statement to multivariable functions, we first need to introduce the multivariable analogue of a closed interval.

**Definition 13.3.** A set $S \subset \mathbf{R}^n$ is *closed* if $S$ contains all its boundary points.

**Definition 13.4.** A set $S$ in $\mathbf{R}^n$ is *bounded* if there is some $r < \infty$ such that

$$S \subset B(\mathbf{0}, r) = \{\mathbf{x} \mid ||\mathbf{x}|| < r\}.$$

Definition 13.4 is equivalent to the statement that a set is bounded if it's contained in some ball of finite radius.

**Example 13.3.** *Is the interval* $[0, 1)$ *closed in* $\mathbf{R}$*? Is it bounded?*
The interval $[0, 1)$ is not a closed subset of $\mathbf{R}$. Example 13.1 showed that the point $1$ is a boundary point of the interval $[0, 1)$, but $1 \notin [0, 1)$.
However, $[0, 1)$ is bounded, as it is completely contained in the ball $B(0, 1) = [-1, 1]$.

**Theorem 25.** Suppose that $f : \mathcal{D}^n \to \mathbf{R}$ is continuous. If $S$ is a closed and bounded subset of $\mathcal{D}^n$, then $f$ attains an absolute maximum and an absolute minimum on $S$.

This theorem is usually proved in an analysis course, and we won't give a proof here.

Theorems 24 and 25 can be combined to find the absolute extrema of a function whose domain is closed and bounded. Every absolute extremum is a local extremum, and every local extremum is a critical point or a boundary point. Evaluating a function at its local extrema and comparing these values identifies the absolute extrema.

## 13.4 Extrema on the boundary of a domain

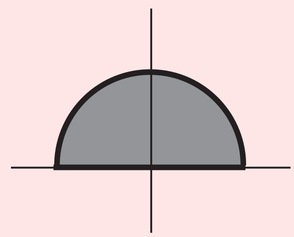
If $f$ is a continuous function defined on a closed and bounded domain, then Theorem 25 guarantees the existence of an absolute maximum and an absolute minimum. Assuming $f$ is sufficiently differentiable, the derivative tests are useful for identifying extrema in the interior of the domain of $f$, and in this section we begin to address the question of finding extrema that may lie on the boundary of the domain. This topic is discussed further in Section 14.

**Example 13.4.** *Let $S$ be the semidisc in $\mathbf{R}^2$ defined by the inequalities*

$$x \geq 0 \qquad y \leq \sqrt{1-x^2}$$

*If $f : S \to \mathbf{R}$ is defined by $f(x, y) = xy$, find the absolute maximum value of $f$.*
First, observe that $S$ is closed and bounded, so $f$ must attain an absolute maximum on $S$. Applying the First Derivative Test shows that $f$ has no critical points on the interior of $S$, so the absolute maximum must occur on the boundary of $S$.

The boundary of $S$ consists of a line segment and a semicircle, and it's easy to check that $f$ evaluates to $0$ at all points on the line segment. Thus, the absolute maximum must occur on the semicircle.



Parameterize the semicircle by

$$\mathbf{h}(t) = \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} \quad \text{for} \quad t \in [0, \pi].$$

In order to find maxima of $f$ on the semicircle, we look for maxima of the composition $f \circ \mathbf{h}$ on $[0, \pi]$.
Applying the single variable First Derivative Test, we set the first derivative of $f \circ \mathbf{h}$ equal to zero:

$$\frac{d}{dt}(f \circ \mathbf{h})(t) = \frac{d}{dt}\big((\cos t)(\sin t)\big) = \cos^2 t - \sin^2 t = 0.$$

The only solutions to this equation in the interval $[0, \pi]$ are $t = \frac{\pi}{4}$ and $t = \frac{3\pi}{4}$. Plugging in these value yields

$$(f \circ \mathbf{h})(\frac{\pi}{4}) = \frac{1}{2} \quad \text{and} \quad (f \circ \mathbf{h})(\frac{3\pi}{4}) = \frac{-1}{2}.$$

It's important to check the endpoints of the interval $[0, \pi]$ as well, but

$$(f \circ \mathbf{h})(0) = (f \circ \mathbf{h})(\pi) = 0.$$

Thus, the absolute maximum value of $f$ on $S$ is $\frac{1}{2}$.

Example 13.5 illustrates a useful technique for identifying extrema that occur on the boundary of a closed domain: describe the boundary of the domain as the image of some other function, and then try to find the extrema of the composition.

This technique relies on the following proposition:

**Proposition 13.1.** Let $D$ be a closed and bounded subset of $\mathbf{R}^n$, and denote the boundary of $D$ by $C$. Suppose that there exists some function $\mathbf{h} : \mathcal{E}^{n-1} \to \mathbf{R}^n$ such that the image of $\mathbf{h}$ is $C$.
If $g : D \to \mathbf{R}$ has an extremum at $\mathbf{h}(t_0) \in C$, then the composition $g \circ \mathbf{h}$ has an extremum at $t_0 \in \mathcal{E}^{n-1}$.

## 13.5 Strategy for finding extrema

In this section, we merge the previous results into a general strategy for finding extrema of multivariable functions. Although this approach won't work for every problem, it indicates the relevant considerations and may provide a helpful overview.

Suppose that you are trying to find the absolute extrema of a multivariable function $f$ on some set $S$.

1. First, decide if $S$ is closed and bounded. If so, then $f$ is guaranteed to have at least one absolute maximum and minimum on $S$.

2. Next, find the critical points of $f$ on the interior of $S$. There are two kinds of critical points:

   (a) points where $f$ is differentiable and the gradient of $f$ is $\mathbf{0}$;

   (b) points where $f$ is not differentiable.

3. If $\nabla f(\mathbf{a}) = \mathbf{0}$, compute the Hessian $Hf(\mathbf{a})$ and apply the Second Derivative Test. Although this isn't guaranteed to determine whether $\mathbf{a}$ is a local maximum, local minimum, or saddle point, it will be effective in many cases.

4. If $f$ fails to be differentiable at a finite number of points in the interior of $S$, evaluate $f$ at each one and compare these values to the values from Step 3.

5. To find extrema of $f$ on the boundary of a set $S \subset \mathbf{R}^2$, it may be possible to parameterize the boundary of $S$. In this case, the First Derivative Test may identify the points where the restriction of $f$ to the boundary has extrema. Evaluate $f$ at these points and compare these values to the values from Steps 3 and 4.

6. If parameterizing the boundary of $S$ is prohibitively difficult, the Lagrange Multipliers technique (which will be introduced in Section 14) offers another approach to finding extrema of $f$ on the boundary of $S$.

## 13.6 Worked Examples

**Example 13.5.** *Let $D$ be the elliptical disc defined by*

$$D = \{(x, y) \in \mathbf{R}^2 \big| 4x^2 + y^2 \leq \pi^2\},$$

*and let $g : D \to \mathbf{R}$ be defined by $g(x, y) = \sin(x)$. Find the maximum value of $g$, if it exists.*

First, observe that $D$ is closed and bounded. According to Theorem 25, this implies that $g$ attains an absolute maximum on $D$. However, we can show that $g$ has no critical points on the interior of the disc. (Try this!)

It follows that $g$ has an absolute maximum on the boundary of $D$. In order to identity the maxima, we view the boundary of $D$ as the image of a parameterized curve.

Define

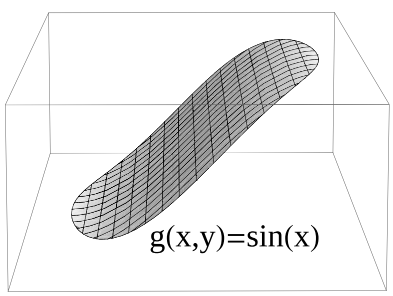
$$\mathbf{h}(t) = \begin{bmatrix} \frac{\pi}{2}\cos t \\ \pi \sin t \end{bmatrix} \quad \text{for} \quad t \in [0, 2\pi].$$

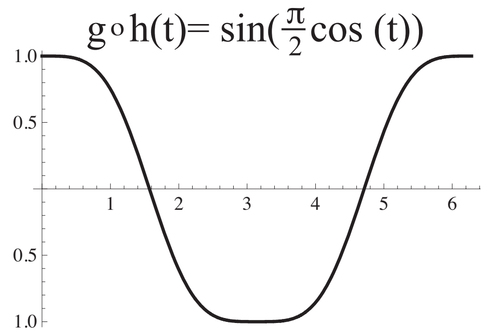
Writing $\mathbf{h}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$, we can see that the image of $\mathbf{h}$ satisfies the equation $4x^2 + y^2 = \pi^2$, so $\mathbf{h}$ parameterizes the boundary of $D$.

As suggested in Section 13.4, we'll find the extrema of $g$ on the boundary of $D$ by looking at the extrema of the composition

$$(g \circ \mathbf{h})(t) = \sin\left(\frac{\pi}{2}\cos t\right).$$

Maxima for $g$ on the boundary of $D$ correspond to maxima of $g \circ \mathbf{h}(t)$.




In order to find critical points of $g \circ \mathbf{h}$, set the derivative of $g \circ \mathbf{h}(t)$ equal to $0$:

$$\frac{d}{dt}(g \circ \mathbf{h})(t) = \frac{-\pi}{2}\sin t \cos(\frac{\pi}{2}\cos t) = 0.$$

Solving this equation yields the critical point $t = \pi$, and we also need to consider the boundary of the domain, which consists of the two endpoints $t = 0$ and $t = 2\pi$. Evaluate $g \circ \mathbf{h}(t)$ on each of these points:

$$g \circ \mathbf{h}(0) = 1 \qquad g \circ \mathbf{h}(\pi) = -1 \qquad g \circ \mathbf{h}(2\pi) = 1.$$

This shows that that $(g \circ \mathbf{h})(0) = (g \circ \mathbf{h})(2\pi) = 1$ is the maximum value of $g$ on the boundary of $D$, and therefore on the entire domain.

**Example 13.6.** *Let $f : \mathbf{R}^2 \to \mathbf{R}$ be defined by*

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

*Show that $f$ must attain an absolute maximum.*

At first glance, this example may seem troubling: the domain of $f$ is certainly not bounded, so we can't apply Theorem 25. Instead of attacking the problem directly, we'll separate the plane into two pieces. For any real number $a$, define

$$\mathcal{P}_a = \{(x, y) \mid x^2 + y^2 \leq a\} \text{ and } \mathcal{Q}_a = \{(x, y) \mid x^2 + y^2 > a\}.$$

Clearly, any point in $\mathbf{R}^2$ is in either $\mathcal{P}_a$ or $\mathcal{Q}_a$. Furthermore, $\mathcal{P}_a$ is closed and bounded for any choice of $a \geq 0$. Therefore $f$ must attain an absolute maximum on $\mathcal{P}_a$. In order to show that this is a true global maximum, we will chose a value of $a$ so that the range of $f$ restricted to $\mathcal{Q}_a$ is bounded from above.

Observe that the value of $f$ gets very small as $(x, y)$ moves away from the origin. More precisely, we have

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

Thus, if $(x, y) \in \mathcal{Q}(a)$, then

$$f(x, y) < \frac{1}{1 + a^2} + \frac{1}{a}.$$

Any choice of $a$ gives an upper bound for the value of $f$ on $\mathcal{Q}_a$.

If we set $a = 2$, for example, then any point in $\mathcal{Q}_2$ evaluates to less than $\frac{1}{5} + \frac{1}{2}$. On the other hand, $(1, 0) \in \mathcal{P}_2$ and $f(1, 0) = 1$, so the maximum value of $f$ on $\mathcal{P}_2$ is at least $1$. This shows that every point in $\mathcal{Q}_2$ evaluates to a number smaller than the maximum of $f$ on $\mathcal{P}_2$. Thus, the absolute maximum of $f$ on $\mathcal{P}_2$ is a maximum of $f$ on $\mathbf{R}^2$.

## 13.7 Exercises

1. Show that every point in a set $S$ is either a boundary point of $S$ or an interior point of $S$. Then, find an example of a set $S$ which is *not* equal to the union of its boundary and its interior.

2. For each of the sets $S_1$, $S_2$, and $S_3$ in Example 13.2, decide if $S_i$ is closed and/or bounded.

3. Which of the following are closed subsets of the given Euclidean space? Which are bounded?

   (a) $\{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 < 1\}$

   (b) $\{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 = 1\}$

   (c) The line $x + y = 3$ in $\mathbf{R}^2$

   (d) The strip $0 \leq x \leq 1$ in $\mathbf{R}^2$

   (e) $\mathbf{R}^2$

   (f) $\{(x, y) \in \mathbf{R}^2 \mid y = x^2 + 2\}$

   (g) $\{(x, y, z) \in \mathbf{R}^3 \mid x^2 + 4y^2 + 3z^2 = 1\}$

   (h) $\{(x, y, 0) \in \mathbf{R}^3 \mid -1 < x < 3, 3 \leq y \leq 5\}$

In Exercises 4 through 8, find the absolute extrema of the given function on the indicated closed and bounded region $R$.

4. $f(x, y) = x^2 + xy - 2y$, and $R = \{(x, y) \in \mathbf{R}^2 \mid |x| \leq 3, |y| \leq 3\}$.
5. $f(x, y) = x + 2y$, and $R$ is the triangular region in $\mathbf{R}^2$ with vertices $(1, 0), (3, 0)$, and $(1, 4)$.
6. $f(x, y) = x^2 + xy + y^2$, and $R = \{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 \leq 4\}$.
7. $f(x, y) = 2x^2 + y^2 - y + 3$, and $R = \{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 \leq 1\}$.
8. $f(x, y) = \sin x \cos y$, and $R = \{(x, y) \in \mathbf{R}^2 \mid 0 \leq x \leq 2\pi, 0 \leq y \leq 3\}$

9. Let $f(x, y) = x^3 + x^2 - 2xy + 3y^2$.
   (a) Find all the critical points of $f$.
   (b) Classify each critical point of $f$ as a local minimum, local maximum, or saddle point.
   (c) Does $f$ have any global extrema in $\mathbf{R}^2$?
10. Let $f(x, y) = x^2y + y^3 - 2xy$.
    (a) Find all the critical points of $f$.
    (b) Classify each critical point of $f$ as a local minimum, local maximum, or saddle point.
    (c) Does $f$ have any global extrema on $\mathbf{R}^2$?
11. Let $B$ be the closed ball defined by

$$B = \{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 \leq 24.\}$$

    Where does the function $f(x, y, z) = 2x + y - z$ attain its minimum and maximum values on $B$?
12. Suppose that in one day a student spends $x_1$ hours studying, $x_2$ hours sleeping and eating, and $x_3$ hours on entertainment. If the student's utility function is

$$U(x_1, x_2, x_3) = -\frac{1}{4}x_1^2 + \frac{1}{2}x_2^2 + x_1x_3 - x_2x_3,$$

    what lifestyle will maximize his utility?

In Exercises 13 through 15, find the critical points of the given function and classify each one as a local minimum, global minimum, local maximum, global maximum, or saddle point.

13. $f(x, y) = 3y^2 - 2y^3 - 3x^2 + 6xy$
14. $f(x, y, z) = x^4 + xy + y^3 - 2yz - 3$

15. $f(x, y) = x^2y - 4xy + y^2 + 10$

16. Show that largest rectangle with a fixed perimeter must be a square.

17. Find three positive numbers whose sum is $24$ and whose product is as large as possible.

18. Suppose you want to build a closed rectangular box which can hold $1$ liter of oil. What dimensions that will minimize the cost of the metal needed for the box?

19. Find the distance between the point $(0, -3, 3)$ and the closest point on the paraboloid
$$x = y^2 + z^2.$$

20. Find the distance between the point $(2, 1, 3)$ and the closest point on the plane
$$x + y + z = 1.$$

21. Let $S$ be the unit sphere in $\mathbf{R}^3$:
$$S = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}.$$
   (a) Find the distance between the point $(10, 5, 3)$ and the closest point on $S$.
   (b) Find the distance between the point $(-1, 1, 0)$ and the closest point on $S$.
   (c) Show that in parts 21a and 21b , the line connecting the given point and its closest point $S$ also contains the origin.

# 14 Lagrange Multipliers

In this section we develop an important technique for finding extrema of multivariable functions. In particular, we focus on *constrained optimization problems*. In such problems, we study the restriction of a function $f : \mathcal{D}^n \to \mathbf{R}$ to some set $C \subset \mathcal{D}^n$. We are interested in finding the extrema of the restricted function $f|_C$. For example, consider the following problems:

**Example 14.1.** *What is the largest value attained by $f(x, y) = 9x^2 - 2xy + 3y^2$ on the ellipse*

$$E = \{(x, y) \mid 3x^2 + y^2 = 12\}?$$

**Example 14.2.** *A rectangular box with dimensions $x \times y \times z$ has surface area*

$$S(x, y, z) = 2(xy + yz + xz).$$

*Among all boxes with volume $8$ cubic feet, there is a unique box of minimal surface area. What are its dimensions?*

The Lagrange Multipliers technique is a useful tool for solving such problems. We'll return to these two problems after explaining how to approach them.

## 14.1 Lagrange Multipliers

Each of the examples above involves a function which is defined on some Euclidean space. However, the solution to the problem requires finding an extremum of the function on some restricted subset of $\mathbf{R}^n$. In both examples, the *constrained domain* can be described as a level set of some other differentiable function:

**Example 14.1:** The ellipse is the level curve $g^{-1}(12)$ for the function $g(x, y) = 3x^2 + y^2$.

**Example 14.2:** If the volume of the box is given by the function $V(x, y, z) = xyz$, the solution must satisfy $(x, y, z) \in V^{-1}(8)$.

The Lagrange Multipliers Theorem is useful for identifying the extrema of a differentiable function on a constrained domain which is a level set of another differentiable function.

**Theorem 26.** [Lagrange Multipliers Theorem] Let $f$ and $g$ be real-valued functions with continuous partial derivatives on $\mathcal{D}^n$, and let $S = g^{-1}(c)$ be a level set for $g$. Suppose that the restriction of $f$ to $S$ has a local extremum at $\mathbf{a}$. If $\nabla g(\mathbf{a}) \neq \mathbf{0}$, then there exists $\lambda \in \mathbf{R}$ such that

$$\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a}).$$

The scalar $\lambda$ in Theorem 26 is called the *Lagrange multiplier*.

**Example 14.1:** First, observe that Example 1 has the right form for applying the Lagrange Multipliers Theorem. The function to maximize, $f$, and the constraint function, $g$, are both polynomials, so they have continuous partial derivatives. We noted above that $E = g^{-1}(12)$, so if $\mathbf{a}$ is a local maximum of $f$ on $E$ and $\nabla \mathbf{g}(\mathbf{a}) \neq \mathbf{0}$, then

$$\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$$

for some real number $\lambda$.
In order to apply the Lagrange Multipliers Theorem, compute the gradients of $f$ and $g$:

$$\nabla f(x, y) = \begin{bmatrix} 18x - 2y \\ -2x + 6y \end{bmatrix} \quad \text{and} \quad \nabla g(x, y) = \begin{bmatrix} 6x \\ 2y \end{bmatrix}.$$

The only solution to $\nabla g(x, y) = \mathbf{0}$ is the point $(0, 0)$, but this is not a point on $E$. Therefore, Theorem 26 implies that extrema of $f|_E$ satisfy

$$\begin{bmatrix} 18x - 2y \\ -2x + 6y \end{bmatrix} = \lambda \begin{bmatrix} 6x \\ 2y \end{bmatrix}.$$

Combining this with the constraint condition shows that any extremum of $f(x, y)$ on $g^{-1}(12)$ is a solution to the following system of equations:

$$\begin{aligned} 18x - 2y &= \lambda 6x \\ -2x + 6y &= \lambda 2y \\ 3x^2 + y^2 &= 12 \end{aligned}$$

To solve this system, eliminate $\lambda$ from the first two equations:

$$(6x)(-2x + 6y) = (18x - 2y)(2y)$$

Rearranging this equation yields $3x^2 = y^2$, so we solve the new system

$$\begin{aligned} 3x^2 - y^2 &= 0 \\ 3x^2 + y^2 &= 12 \end{aligned}$$

This system has four solutions:

$$\{(\sqrt{2}, \sqrt{6}), (-\sqrt{2}, \sqrt{6}), (\sqrt{2}, -\sqrt{6}), (-\sqrt{2}, -\sqrt{6})\}.$$

Since the ellipse $E$ is closed and bounded in $\mathbf{R}^2$, Theorem 25 from Section 13 ensures that $f$ attains an absolute maximum value on $E$. That means that $f$ attains an absolute maximum at one or more of these points. Evaluating $f$ at each of them shows that the restriction of $f$ to $E$ attains its maximum value at $(\sqrt{2}, -\sqrt{6})$ and at $(-\sqrt{2}, \sqrt{6})$.

**Example 14.2:** Recall the surface area and volume functions for a box with dimensions $x \times y \times z$:

$$\begin{aligned} S(x, y, z) &= 2(xy + yz + xz) \\ V(x, y, z) &= xyz. \end{aligned}$$

Both $S$ and $V$ have continuous partial derivatives, so Theorem 26 implies that at any extremum of $S$ on $V^{-1}(8)$ where $\nabla V \neq 0$, then

$$\nabla S(x, y, z) = \lambda \nabla V(x, y, z)$$

for some real number $\lambda$.
First, compute the gradients of $S$ and $V$:

$$\nabla S(x, y, z) = \begin{bmatrix} 2(y+z) \\ 2(x+z) \\ 2(x+y) \end{bmatrix} \text{ and } \nabla V(x, y, z) = \begin{bmatrix} yz \\ xz \\ xy \end{bmatrix}.$$

The statement of the problem asserted that a unique absolute minimum exists, so it must be a solution to the system of equations

$$\begin{aligned} 2(y+z) &= \lambda yz \\ 2(x+z) &= \lambda xz \\ 2(x+y) &= \lambda xy \\ xyz &= 8 \end{aligned}$$

Solving algebraically shows that the unique solution to this system is the point $(2, 2, 2)$. Since the problem asserted that $S$ attains an absolute minimum on $V^{-1}(8)$ the point $(2, 2, 2)$ is the only possible local extremum, it follows that a $2' \times 2' \times 2'$ box has smaller surface area than any other box with the same volume.

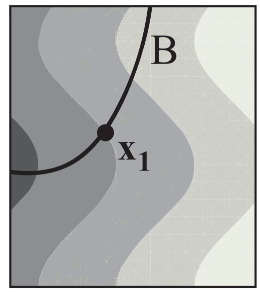
In Example 14.2, it's important that the problem guaranteed the existence of an absolute minimum; Theorem 26 provides a necessary criterion for a point to be an extremum, but it does not imply that any absolute or local extrema exist.

## 14.2 Why do Lagrange multpliers work?

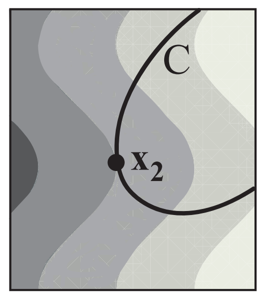
Before giving a rigorous proof of Theorem 26, we first present a heuristic argument.

The figures below show contour maps for a function $f : \mathbf{R}^2 \to \mathbf{R}$. In the maps, higher values are represented by darker colors. The curve $B$ is an example of a constrained set on which one might try to maximize $f$, so let's assume that $B$ is level set for some function $g_B : \mathbf{R}^2 \to \mathbf{R}$. Similarly, assume $C$ is a level set for some other function $g_C : \mathbf{R}^2 \to \mathbf{R}$.

In the first picture, $B$ crosses a level curve for $f$ at the point $\mathbf{x_1}$. Moving along $B$ to the right of $\mathbf{x_1}$ decreases the value of $f$ and moving along $B$ to the left of $\mathbf{x_1}$ increases the value of $f$. This shows that $\mathbf{x_1}$ is neither a local maximum nor a local minimum.




In the second picture, $C$ is tangent to a level curve for $f$ at $\mathbf{x_2}$. Moving away from $\mathbf{x_2}$ along $C$ decreases the value of $f$, so $\mathbf{x_2}$ is a local maximum for the restriction of $f$ to $C$. In this case, the gradient vector $\nabla f(\mathbf{x_2})$ is perpendicular to the tangent line to $C$ at $\mathbf{x_2}$. Since $\nabla g_C(\mathbf{x_2})$ is also perpendicular to this tangent line, $\nabla f(\mathbf{x_2})$ and $\nabla g_C(\mathbf{x_2})$ are parallel.

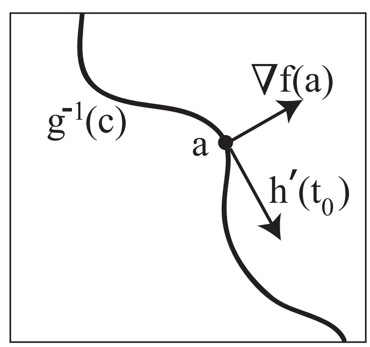



We now formalize this idea in a rigorous proof.

*Proof of the Lagrange Mulitpliers Theorem.* To begin, let $f, g : \mathcal{D}^2 \to \mathbf{R}$ be differentiable functions, and suppose that the restriction of $f$ to $g^{-1}(c)$ attains a maximum at $\mathbf{a}$.

Let $\mathbf{h} : \mathcal{I} \to \mathbf{R}^2$ be a parameterized curve with differentiable coordinate function, and suppose that further that

- the image of $\mathbf{h}$ lies in $g^{-1}(c)$;
- $\mathbf{h}(t_0) = \mathbf{a}$; and
- $\mathbf{h}'(t_0) \neq \mathbf{0}$.




By construction, $(f \circ \mathbf{h})(t)$ has a local maximum at $t_0$, so the first derivative condition for local extrema implies that

$$\frac{d}{dt}(f \circ \mathbf{h})(t_0) = 0. \tag{14.1}$$

We can compute the same derivative using the Chain Rule:

$$\begin{aligned}\frac{d}{dt}(f \circ \mathbf{h})(t_0) &= Df(\mathbf{h}(t_0))D\mathbf{h}(t_0) \\ &= \nabla f(\mathbf{a}) \cdot \mathbf{h}'(t_0) \end{aligned} \tag{14.2}$$

Combining Equations 14.1 and 14.2, we see that

$$\nabla f(\mathbf{a}) \cdot \mathbf{h}'(t_0) = 0.$$

Thus, $\nabla f(\mathbf{a})$ is orthogonal to $\mathbf{h}'(t_0)$.

On the other hand, recall from Section 9.3 that the tangent line to $g^{-1}(c)$ is orthogonal to the gradient vector $\nabla g(\mathbf{a})$. Since the image of $\mathbf{h}$ lies in $g^{-1}(c)$, this implies that $\mathbf{h}'(t_0)$ is orthogonal to $\nabla g(\mathbf{a})$.

The vectors $\nabla g(\mathbf{a})$ and $\nabla f(\mathbf{a})$ are both orthogonal to $\mathbf{h}'(t_0)$, and since they lie in $\mathbf{R}^2$, this implies that they are parallel to each other:

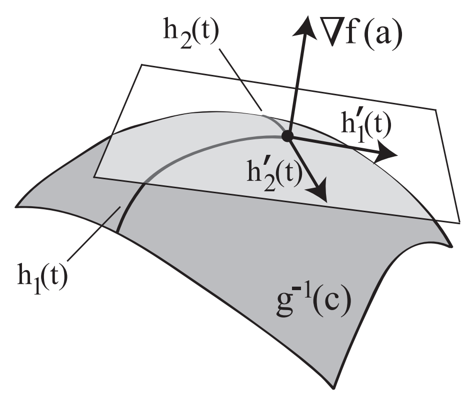
$$\nabla g(\mathbf{a}) = \lambda \nabla f(\mathbf{a}) \text{ for some } \lambda \in \mathbf{R}.$$

To complete the proof, we drop the assumption that $f$ and $g$ are defined on subsets of $\mathbf{R}^2$. Let $g : \mathbf{R}^{n+1} \to \mathbf{R}$. Since $\nabla g(\mathbf{a}) \neq \mathbf{0}$, the null space of $\nabla g(\mathbf{a})^T$ is an $n$-dimensional subspace $N \subset \mathbf{R}^{n+1}$.

Pick a basis $\{\mathbf{v}_1, ...\mathbf{v}_n\}$ for $N$. For $i \in \{1, 2, ...n\}$, let $\mathbf{h}_i(t)$ be a parameterized curve such that

- $\mathbf{h}_i(t_0) = \mathbf{a}$; and
- $\mathbf{h}_i'(t_0) = \mathbf{v}_i$.

By construction, $\nabla g(\mathbf{a})$ is orthogonal to $\mathbf{h_i}'(t_0) = 0$ for each $i$. Additionally, the argument above shows that $\nabla f(\mathbf{a}) \cdot \mathbf{h_i}'(t_0) = 0$ for each $i$, so $\nabla f(\mathbf{x_0})$ is also orthogonal to each $\mathbf{v}_i$. This proves that $\nabla g(\mathbf{a})$ and $\nabla f(\mathbf{a})$ are parallel.




□

## 14.3 Lagrange multipliers for multiple constraints

The Lagrange multiplier technique may also be used to solve optimization problems involving more than one constraint. This case arises if the constrained domain can be described as the intersection of levels sets of different differentiable functions.

**Theorem 27.** For $i \in \{1, 2, ...k\}$ with $k < n$, let $g_i : \mathbf{R}^n \to \mathbf{R}$ be a real-valued function with continuous partial derivatives and let $c_i \in \mathbf{R}$. Denote the intersection of the sets $g_i^{-1}(c_i)$ by $C$:

$$C = \{\mathbf{x} \in \mathbf{R}^n | \mathbf{x} \in g_i^{-1}(c_i) \text{ for all } i\}.$$

Let $f : \mathbf{R}^n \to \mathbf{R}$ be a function with continuous partial derivatives and suppose that the restriction of $f$ to $C$ has a local extremum at $\mathbf{a}$. If the vectors $\{\nabla g_i(\mathbf{a})\}_{i=1}^k$ are linearly independent, then

$$\nabla f(\mathbf{a}) = \sum_{i=1}^{k} \lambda_i \nabla g_i(\mathbf{a})$$

for some $\lambda_i \in \mathbf{R}$.

A special case of the proof of Theorem 27 is developed in Exercise 26.

## 14.4 Worked Examples

**Example 14.3.** *Let $S$ be the intersection of the cylinder $(x-1)^2+y^2 = 4$ and the plane $x+y+2z = 1$ in $\mathbf{R}^3$. What point on $S$ is closest to the origin?*

We want to minimize the function which gives the distance between a point in $\mathbf{R}^3$ and the origin. Instead of trying to minimize the function $d(x, y, z) = \sqrt{x^2 + y^2 + z^2}$, however, it is computationally simpler to minimize the *distance squared* function:

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

The set $S$ is defined to be the intersection of a plane and a cylinder, and before computing anything, we note that the question is well-posed: there *is* a nearest point to the origin, since the function $D$ is being evaluated on a closed and bounded set. Obsever that the cylinder and the plane can each be described as a level set of a differentiable function:

$$\begin{aligned} &\text{For } C(x, y, z) = (x-1)^2 + y^2 - 4, \text{ the cylinder is } C^{-1}(0); \\ &\text{For } P(x, y, z) = x + y + 2z - 1, \text{ the plane is } P^{-1}(0). \end{aligned}$$

Since the functions $D$, $C$, and $P$ are all differentiable, we can apply the general form of the Lagrange Mulipliers Theorem to this problem. Any point nearest to the origin must satisfy

$$\begin{bmatrix} 2x \\ 2y \\ 2z \end{bmatrix} = \lambda_1 \begin{bmatrix} 2(x-1) \\ 2y \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

for some $\lambda_1$ and $\lambda_2$ in $\mathbf{R}$.

Together with the constraint functions, this yields the system of equations

$$2x = 2\lambda_1(x-1) + \lambda_2 \tag{14.3}$$
$$2y = 2\lambda_1 y + \lambda_2 \tag{14.4}$$
$$2z = 2\lambda_2 \tag{14.5}$$
$$0 = (x-1)^2 + y^2 - 4 \tag{14.6}$$
$$0 = x + y + 2z - 1. \tag{14.7}$$

Substituting Equations 14.5 and 14.7 into Equations 14.3 and 14.4 eliminates the variables $\lambda_2$ and $z$.

$$2x = 2\lambda_1(x-1) + \frac{1}{2}(1-x-y) \tag{14.8}$$
$$2y = 2\lambda_1(y) + \frac{1}{2}(1-x-y) \tag{14.9}$$

After solving for $\lambda_1$ in each of these equations and equating the results, we are left with

$$\frac{4x-1+x+y}{4(x-1)} = \frac{4y-1+x+y}{4y}.$$

Simplifying this equation and recalling Equation 14.6 leaves us with two equations and two unknowns:

$$\begin{aligned} 0 &= x^2 - 2x + 1 - 4y - y^2 \\ 0 &= (x-1)^2 + y^2 - 4. \end{aligned} \tag{14.10}$$

A bit more algebra and the quadratic formula (try this!) show that there are two real solutions to this system:

$$(1+\sqrt{2\sqrt{3}}, -1+\sqrt{3}) \quad \text{and} \quad (1-\sqrt{2\sqrt{3}}, -1+\sqrt{3})$$

Finally, we plug these into the defining equation for the plane to get the third coordinate:

$$\left(1+\sqrt[4]{12}, -1+\sqrt{3}, \frac{1}{2}(1-\sqrt[4]{12}-\sqrt{3})\right) \quad \text{and} \quad \left(1-\sqrt[4]{12}, -1+\sqrt{3}, \frac{1}{2}(1+\sqrt[4]{12}-\sqrt{3})\right)$$

Evaluating the distance function on these two points shows that the second point is closer to the origin than the first. Although this may look daunting to check, the computation is simplified by a few observations. First, note that the two points have the same $y$ coordinate, so this won't distinguish the two points. When comparing $x^2+z^2$, it's enough to multiply out the terms and see which ones appear with different signs in the two expressions. You don't need a calculator!

## 14.5 Exercises

In Exercises 1 through 7, view the set $S$ either as a level set of a single function $g$ or as the intersection of functions $g_1$ and $g_2$. Identify the points on $S$ where the vectors $\{\nabla f, \nabla g\}$ or $\{\nabla f, \nabla g_1, \nabla g_2\}$are linearly dependent.

1. $f(x,y) = 2x + 3y, \quad S = \{(x,y) \in \mathbf{R}^2 \mid x^2 + 4y^2 = 16\}$.
2. $f(x,y) = -x^2 - y^2, \quad S = \{(x,y) \in \mathbf{R}^2 \mid 2x - 3y = 1\}$.
3. $f(x,y,z) = 2xy + 2yz + 2xz, \quad S = \{(x,y,z) \in \mathbf{R}^3 \mid xyz = 64\}$.
4. $f(x,y,z) = -2y + 5x^2 + 3z^2, \quad S = \{(x,y,z) \in \mathbf{R}^3 \mid x^2 + y^2 = 4z^2\}$.
5. $f(x,y,z) = x, \quad S = \{(x,y,z) \in \mathbf{R}^3 \mid y^2 + z^2 = 1 \text{ and } x + y - z = 0\}$.
6. $f(x,y,z) = x^2 + y^2 + z^2, \quad S = \{(x,y,z) \in \mathbf{R}^3 \mid 2x^2 + 3y^2 + 4z^2 = 1 \text{ and } y + z = 0\}$.
7. $f(x,y,z) = xy + xz, \quad S = \{(x,y,z) \in \mathbf{R}^3 \mid x - y - 2z = 0 \text{ and } x + y + z = 3\}$.

For Exercises 8 through 12, look at the indicated problem in Section 13. In each case, use Lagrange multipliers to find the absolute extrema of the given function on the boundary of the indicated domain.

8. Problem 4 9. Problem 5 10. Problem 6 11. Problem 7 12. Problem 8

13. Use Lagrange multipliers to find the points on the surface $x^2 - yz = 7$ that are closest to the origin. Prove that these points really are global minima.

    (Hint: When looking for minima of the distance function, it's easier to work with the distance squared function.)

14. A parallelogram of side lengths $x$ and $y$ and interior angle $\theta$ has area

$$A = xy \sin\theta.$$

    Find the maximum area of a parallelogram with fixed perimeter $\ell$ in two ways.

    (a) Write down the fixed perimeter condition and use it to eliminate one of the variables directly. Now you have an unconstrained optimization problem in two variables. Find the dimensions which yield the maximum area using techniques from Section 12.

    (b) Use Lagrange multipliers to solve the constrained three variable optimization.

15. The inequalities

$$x^2 + y^2 \leq 4$$
$$-1 \leq x - y + 4z \leq 3$$

    define a solid cylinder. Maximize the function $F(x,y,z) = 7x^2 + 2xy - y^2 + z^2$ over this region.

16. Let $a$ and $b$ be positive constants with $a > b$, and consider the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

(a) Write down, but do not solve, a system of equations that must be satisfied by any pair of points $(x_1, y_1)$ and $(x_2, y_2)$ which are maximally far apart on the ellipse.

(b) Show that with an appropriate choice of Lagrange multiplier(s), the points $(-a, 0)$ and $(a, 0)$ are solutions to the system in part (a).

17. Use Lagrange multipliers to prove that the distance $d$ from a point $(x_0, y_0, z_0)$ to the nearest point on the plane $Ax + By + Cz + D = 0$ is given by the formula

$$d = \frac{Ax_0 + By_0 + Cz_0 + D}{\sqrt{A^2 + B^2 + C^2}}.$$

18. Let $f(x, y) = y$ and let $g(x, y) = y^3 - x^2$.

(a) Show, without using calculus, that the restriction of $f$ to the level set $g^{-1}(0)$ attains a minimum at the point $(0, 0)$.

(b) Show that there is no $\lambda$ such that $\nabla f(0, 0) = \lambda \nabla g(0, 0)$.

(c) Explain why part (b) does not contradict the Lagrange Multipliers Theorem.

19. *Cauchy-Schwartz Inequality*

Let $\mathbf{r_1} = (x_1, y_1, z_1)$ and $\mathbf{r_2} = (x_2, y_2, z_2)$ be two vectors in $\mathbf{R}^3$ constrained to have fixed norms: $\|\mathbf{r_1}\| = a$, $\|\mathbf{r_2}\| = b$.

(a) Use Lagrange multipliers to show that the maximum value of the dot product $\mathbf{r_1} \cdot \mathbf{r_2}$ is attained when the vectors are parallel and that the minimum is attained when the vectors are anti-parallel.

(b) Use part (a) to prove the Cauchy Shwartz inequality:

$$|\mathbf{r_1} \cdot \mathbf{r_2}| \leq \|\mathbf{r_1}\| \|\mathbf{r_2}\|.$$

20. (a) Show that the product of three positive numbers $x, y, z$ constrained to have a fixed sum $S$ is maximal when $x, y$ and $z$ are all equal.

(b) Use the result of (a) to show that the geometric mean of three numbers is always less than or equal to the arithmetic mean:

$$(xyz)^{1/3} \leq \frac{x + y + z}{3}.$$

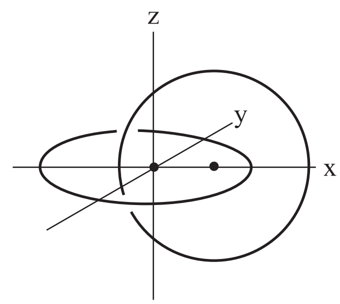
21. *Electrostatic Potential* Two charged particles in three-dimensional space have an associated electrostatic potential energy given by the *Coulomb potential*:

$$U(\mathbf{r_1}, \mathbf{r_2}) = \frac{q_1 q_2}{r}.$$

Here, $q_1$ and $q_2$ are fixed positive integers describing the charges of the particles, $\mathbf{r_i}$ describes the position of the $i^{th}$ particle, and $r = ||\mathbf{r_1} - \mathbf{r_2}||$ is the distance between the two particles. A *stationary state of the system* is a critical point of the energy function $U$; at a critical point, the system can not convert any potential energy into kinetic energy by making a small change, and the particles will remain at rest.

(a) Show that $U(\mathbf{r_1}, \mathbf{r_2})$ has no critical points.

(b) Now, suppose that the first charge is constrained to lie in the plane $z = 0$ on the circle of radius $3$ centered at the origin and that the second charge is constrained to lie in the plane $y = 0$ on the circle of radius $3$ centered at the point $(2, 0, 0)$. Find the stationary states of the system.




22. You are addicted to online shopping. A research psychologist comes up with the following model to describe your daily happiness as a function of the number of hours spent shopping for various types of products. Let $s$ be the number hours spent shopping for shoes, $t$ the number of hours researching travel prices, $g$ the number of hours reading reviews of electronic gadgets, and $h$ the number of hours reading forums about home appliances. Then your happiness is given by

$$H = -1153 + 6tg + 18sg + 12th - 18st + 36s^2$$

Of course, there are only $24$ hours in a day, so $s + t + g + h = 24$. Use Lagrange multipliers to plan your day.

23. *Hooke's Law*

A spring satisfies *Hooke's Law* if the potential energy stored in the spring satisfies:

$$E(x) = k(x - \ell_{rest})^2$$

where $x$ is the length of the stretched or compressed spring, $\ell$ is the rest length of the spring (i.e. the length of the spring when no external forces are applied to it), and $k$ is the spring constant determining how easily stretched the spring is.


Three springs are hooked up in series as indicated in the figure. The beginning of the first spring and the end of the last spring are fixed so that the total length of the spring configuration is $10$. The springs have constants $k_1 = 4, k_2 = 2, k_3 = 5$ and rest lengths $\ell_1 = 2, \ell_2 = 3, \ell_3 = 1$. The total energy $E$ of the system is the sum of the energies of each spring: $E = E_1 + E_2 + E_3$. Find the configuration with minimum total energy using Lagrange mulitipliers.

24. An auto manufacturing company can produce $F$ cars using labor $L$ and capital $K$ where $F(L, K) = \sqrt{LK}$. The costs incurred are given by $C(L, K) = wL + rK$. where the wage rate $w$ and the interest rate $r$ are constants. What is the most cost effective way for this company to build $10$ cars?

25. Fred lives two periods and wants to maximize his happiness. Fred's per period happiness, or *utility*, is described by $U(C) = \ln C$, where $C$ is his consumption in dollars for that period. Fred's total two-period utility is described by $F(C_1, C_2) = U(C_1) + \beta U(C_2)$, where $0 < \beta < 1$ is a factor describing how much Fred *discounts* his second period utility when making a consumption decision in the first period. Fred must obey two budget constraints when choosing $C_1$ and $C_2$. His first period consumption is at most equal to his per period wage $W$ and he saves the remaining amount $S$ for the second period, $C_1 + S = W$. In the second period, Fred consumes his wage and his savings plus interest, $C_2 = W + rS$, where $r > 1$ is the interest rate. How much should Fred save in period one in order to maximize his total two period utility?

26. For $i = 1, 2$, suppose that $g_i : \mathbf{R}^3 \to \mathbf{R}$ have continuous partial derivatives and let $c_i \in \mathbf{R}$. Let $C$ denote the intersection of the level sets of $g_i$ at height $c_i$:

$$C = \{\mathbf{x} \in \mathbf{R}^3 \mid g_1(\mathbf{x}) = c_1 \text{ and } g_2(\mathbf{x}) = c_2\}.$$

Let $f : \mathbf{R}^3 \to \mathbf{R}$ have continuous partial derivatives and suppose that $f|_C$ has a local extremum at $\mathbf{a}$.

(a) Let $\mathbf{h}(t) : [0, 1] \to C$ be a parameterized curve with $\mathbf{h}(\frac{1}{2}) = \mathbf{a}$. Show that

$$\nabla f(\mathbf{a}) \cdot \mathbf{h}'(\frac{1}{2}) = 0.$$

(b) Show that for $i = 1, 2$,

$$\nabla g_i(\mathbf{a}) \cdot \mathbf{h}'(\frac{1}{2}) = 0.$$

(c) Show that if $\{\nabla g_1(\mathbf{a}), \nabla g_2(\mathbf{a})\}$ is a linearly independent set, then

$$\nabla f(\mathbf{a}) = d_1 \nabla g_1(\mathbf{a}) + d_2 \nabla g_2(\mathbf{a})$$

for some scalars $d_1$ and $d_2$.

(d) Construct an example that shows why the assumption of linear independence in part (c) is necessary.

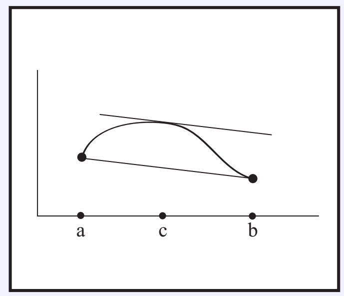
# 15 Appendix I: Proofs

## 15.1 Proof of Clairaut's Theorem

The proof of Clairaut's theorem relies on the Mean Value Theorem from single variable calculus, which we recall here:

**Theorem 28** (Mean Value Theorem)**.** If $g(x) : [a, b] \to \mathbf{R}$ is continuous on $[a, b]$ and differentiable on $(a, b)$, then there exists some $c \in (a, b)$ such that

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




This theorem is proved in any text on single variable calculus.

*Proof of Clairaut's Theorem.* Let $\mathcal{N}$ be a neighborhood of the point $\mathbf{a}$ with the property that $f_{x_i x_j}$ exists for all $\mathbf{x} \in \mathcal{N}$ and all $i, j \in \{1, 2, ...n\}$.

Fix some $h > 0$ which is small enough to ensure the following:

$$\begin{aligned} \mathbf{a} + h\mathbf{e_i} + h\mathbf{e_j} &\in \mathcal{N} \\ \mathbf{a} + h\mathbf{e_i} &\in \mathcal{N} \\ \mathbf{a} + h\mathbf{e_j} &\in \mathcal{N} \end{aligned}$$

In order to prove Clairaut's Theorem, we will study the quantity

$$H(h) = f(\mathbf{a} + h\mathbf{e_i} + h\mathbf{e_j}) - f(\mathbf{a} + h\mathbf{e_i}) - f(\mathbf{a} + h\mathbf{e_j}) + f(\mathbf{a}).$$

By writing $H(h)$ in two different ways, we will show that that

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \lim_{h \to 0} \frac{H(h)}{h^2} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}),$$

and this will prove the theorem.

Define a new function $g : \mathcal{N} \to \mathbf{R}$ by

$$g(\mathbf{x}) = f(\mathbf{x} + h\mathbf{e_j}) - f(\mathbf{x}).$$

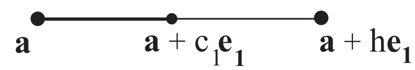
Notice that for $i \neq j$,

$$\frac{\partial g}{\partial x_i}(\mathbf{x}) = \frac{\partial f}{\partial x_i}(\mathbf{x} + h\mathbf{e_j}) - \frac{\partial f}{\partial x_i}(\mathbf{x}).$$

Consider the restriction of $g$ to the interval connecting $\mathbf{a}$ and $\mathbf{a} + h\mathbf{e_i}$. We can describe this restriction as

$$g(\mathbf{a} + t\mathbf{e_i}) \quad \text{for} \quad t \in [0, h].$$

Note that $g(\mathbf{a} + t\mathbf{e_i})$ is differentiable for $t \in (0, h)$. We can therefore apply the Mean Value Theorem to $g(\mathbf{a} + t\mathbf{e_i})$. This implies that there exists some $c_i \in (0, h)$ such that




$$\frac{g(\mathbf{a} + h\mathbf{e_i}) - g(\mathbf{a})}{h} = \frac{\partial g}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i}).$$

On the other hand, we can expand the numerator of the left-hand expression:

$$\begin{aligned} g(\mathbf{a} + h\mathbf{e_i}) - g(\mathbf{a}) &= f(\mathbf{a} + h\mathbf{e_i} + h\mathbf{e_j}) - f(\mathbf{a} + h\mathbf{e_i}) - f(\mathbf{a} + h\mathbf{e_j}) + f(\mathbf{a}) && (15.1) \\ &= H(h) && (15.2) \end{aligned}$$

This gives us the equation

$$\frac{H(h)}{h} = \frac{\partial g}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i}).$$

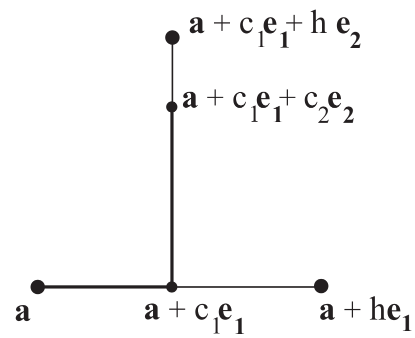
Equivalently, we get

$$H(h) = h\frac{\partial g}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i}) = h(\frac{\partial f}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i} + h\mathbf{e_j}) - \frac{\partial f}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i})) \tag{15.3}$$

Now study the function

$$\frac{\partial f}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i} + s\mathbf{e_j}).$$

This is a function of one variable which is defined for $s \in [0, h]$. We can therefore apply the Mean Value Theorem again.




The Mean Value Theorem implies that there exists some $c_j \in (0, h)$ such that

$$\frac{\frac{\partial f}{\partial x_i}(\mathbf{a} + c\mathbf{e_i} + h\mathbf{e_j}) - \frac{\partial f}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i})}{h} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(\mathbf{a} + c_i\mathbf{e_i} + c_j\mathbf{e_j}).$$

Combining this with Equation 15.3, we get

$$H(h) = h^2\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a} + c_i\mathbf{e_i} + c_j\mathbf{e_j}).$$

Up until now, the value of $h$ has been held constant. We now study the limit

$$\lim_{h\to 0}\frac{H(h)}{h^2} = \lim_{h\to 0}\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a} + c_i\mathbf{e_i} + c_j\mathbf{e_j}).$$

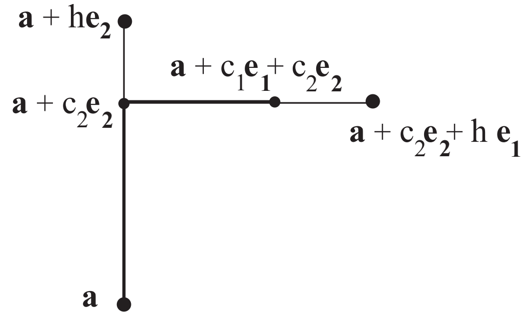
Since $0 < c_i < h$ and $0 < c_j < h$,

$$\lim_{h \to 0}(\mathbf{a} + c_i\mathbf{e_i} + c_j\mathbf{e_j}) = \mathbf{a}.$$

By hypothesis, the mixed partial derivative $\frac{\partial^2 f}{\partial x_j \partial x_i}$ is continuous at $\mathbf{a}$. This implies

$$\lim_{h \to 0} \frac{H(h)}{h^2} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}).$$

On the other hand, the entire computation could be rewritten with the roles of $i$ and $j$ exchanged. In this case we would get the equation




$$\lim_{h \to 0} \frac{H(h)}{h^2} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}).$$

This shows that $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})$ and proves Clairaut's Theorem. □

## 15.2 Proof of Theorem 14

In this section we show that if a function $\mathbf{f}$ has continuous partial derivatives in some neighborhood of the point $\mathbf{a}$, then $\mathbf{f}$ is differentiable at $\mathbf{a}$.

*Proof.* It suffices to prove Theorem 14 for a function $f : \mathcal{D}^n \to \mathbf{R}$. We need to show that

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

Equivalently, we can write $\mathbf{t} = \mathbf{x} - \mathbf{a}$ and rephrase the claim as

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

This will have the advantage of allowing us to use coordinates with no confusion:

$$\begin{aligned}
\mathbf{a} &= (a_1, a_2, ...a_n) \\
\mathbf{t} &= (t_1, t_2, ...t_n) \\
Df(\mathbf{a}) &= \big(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), ...\frac{\partial f}{\partial x_n}(\mathbf{a})\big).
\end{aligned}$$

Fix $\epsilon > 0$.

By hypothesis, each partial derivative $\frac{\partial f}{\partial x_i}$ is continuous in a neighborhood of $\mathbf{a}$. Therefore, there exist numbers $\delta_i > 0$ for $i \in \{1, 2, ...n\}$ such that

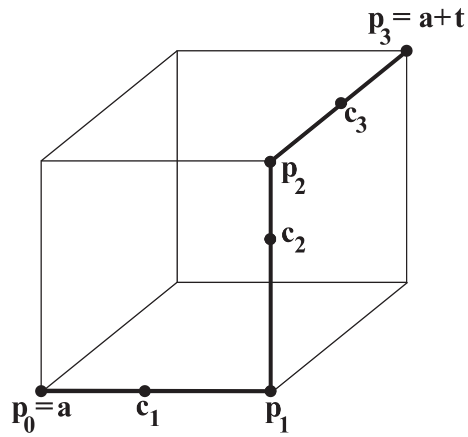
$$|\mathbf{t}| < \delta_i \quad \text{implies} \quad |\frac{\partial f}{\partial x_i}(\mathbf{a} + \mathbf{t}) - \frac{\partial f}{\partial x_i}(\mathbf{a})| < \frac{\epsilon}{n}.$$

Let $\delta = \min(\delta_1, \delta_2, ...\delta_n)$. Then for each coordinate function $f_i$,

$$|\mathbf{t}| < \delta \text{ implies } |\frac{\partial f}{\partial x_i}(\mathbf{a}+\mathbf{t}) - \frac{\partial f}{\partial x_i}(\mathbf{a})| < \frac{\epsilon}{n}. \tag{15.4}$$

In order to make use of the hypothesis that the partial derivatives are continuous, we will connect the points $\mathbf{a}$ and $\mathbf{a}+\mathbf{t}$ by a rectilinear path in which only one coordinate changes at a time.

This path consists of linear segments connecting the points $\mathbf{p_i}$:




$$\begin{aligned}
\mathbf{p_0} =& \mathbf{a}\\
\mathbf{p_1} =& (a_1+t_1, a_2, a_3, ...a_n),\\
\mathbf{p_2} =& (a_1+t_1, a_2+t_2, a_3, ...a_n),\\
&\vdots\\
\mathbf{p_{n-2}} =& (a_1+t_1, ...a_{n-2}+t_{n-2}, a_{n-1}, a_n),\\
\mathbf{p_{n-1}} =& (a_1+t_1, ...a_{n-2}+t_{n-2}, a_{n-1}+t_{n-1}, a_n),\\
\mathbf{p_n} =& \mathbf{a}+\mathbf{t}.
\end{aligned}$$

Observe that

$$\begin{aligned}
f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) &= f(\mathbf{p_n}) - f(\mathbf{p_0})\\
&= f(\mathbf{p_n}) - f(\mathbf{p_{n-1}}) + f(\mathbf{p_{n-1}}) - f(\mathbf{p_0})\\
&= f(\mathbf{p_n}) - f(\mathbf{p_{n-1}}) + f(\mathbf{p_{n-1}}) - f(\mathbf{p_{n-2}}) + f(\mathbf{p_{n-2}}) - f(\mathbf{p_0})\\
&\vdots\\
&= \sum_{i=1}^{n} f(\mathbf{p_i}) - f(\mathbf{p_{i-1}})
\end{aligned}$$

The line segment connecting $\mathbf{p_{i-1}}$ to $\mathbf{p_i}$ has length $t_i$, and the restriction of $f$ to this segment is a function which depends only on the single variable $x_i$. We can therefore use the single variable Mean Value Theorem to claim that for each $i$, there exists some $d_i \in (0, t_i)$ with the property that

$$\frac{f(\mathbf{p_i}) - f(\mathbf{p_{i-1}})}{t_i} = \frac{\partial f}{\partial x_i}(\mathbf{p_{i-1}} + d_i\mathbf{e_i}).$$

Applying the Mean Value Theorem to each segment in turn, we have that

$$f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) = \sum_{i=1}^{n} t_i \frac{\partial f}{\partial x_i}(\mathbf{p_{i-1}} + d_i\mathbf{e_i}).$$

If $|\mathbf{t}| < \delta$, then the bounds from Equation 15.4 imply

$$
\begin{aligned}
|f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{t}| &= \Big|\sum_{i=1}^{n} t_i \frac{\partial f}{\partial x_i}(\mathbf{p_{i-1}} + d_i\mathbf{e_i}) - \sum_{i=1}^{n} t_i \frac{\partial f}{\partial x_i}(\mathbf{a})\Big| \\
&\leq \sum_{i=1}^{n} |t_i| \Big|\big(\frac{\partial f}{\partial x_i}(\mathbf{p_{i-1}} + d_i\mathbf{e_i}) - \frac{\partial f}{\partial x_i}(\mathbf{a})\big)\Big| \\
&\leq \sum_{i=1}^{n} |t_i| \frac{\epsilon}{n} \\
&\leq \frac{\epsilon}{n} \sum_{i=1}^{n} |\mathbf{t}| \\
&= \epsilon |\mathbf{t}|.
\end{aligned}
$$

Thus

$$\frac{|f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{t}|}{|\mathbf{t}|} \leq \frac{\epsilon|\mathbf{t}|}{|\mathbf{t}|} = \epsilon.$$

This proves that

$$\lim_{\mathbf{t}\to 0} \frac{f(\mathbf{a}+\mathbf{t}) - f(\mathbf{a}) - Df(\mathbf{a})\mathbf{t}}{|\mathbf{t}|} = 0,$$

and therefore, that $f$ is differentiable at $\mathbf{a}$.

□

## 15.3 Proof of the Chain Rule

The proof of Theorem 16 is somewhat technical, but the strategy for the entire argument appears in the first step.

*Proof.* **Step 1**

We would like to prove that the function $\mathbf{g} \circ \mathbf{f}$ is differentiable at $\mathbf{a}$, so we begin by recalling the definition of the derivative from Section 7.4:

> A function $\mathbf{h} : \mathcal{D}^n \to \mathbf{R}^p$, a point $\mathbf{a}$, and the matrix of partial derivatives $D\mathbf{h}(\mathbf{a})$ together define a remainder function $\mathbf{R}(\mathbf{x})$:
>
> $$\mathbf{R}(\mathbf{x}) = \mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{a}) - \mathbf{L}(\mathbf{x} - \mathbf{a}).$$
>
> According to Definition 7.4, $\mathbf{h}$ is differentiable at $\mathbf{a}$ if and only if
>
> $$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\mathbf{R}(\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} = \mathbf{0}.$$

The statement of the Chain Rule provides a candidate for the matrix of partial derivatives of $\mathbf{g} \circ \mathbf{f}$ at $\mathbf{a}$:

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{a})(\mathbf{x}) = D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})(\mathbf{x}),$$

and the associated remainder function is

$$\mathbf{R}(\mathbf{x}) = (\mathbf{g} \circ \mathbf{f})(\mathbf{x}) - (\mathbf{g} \circ \mathbf{f})(\mathbf{a}) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}). \tag{15.5}$$

Thus, to prove the theorem it suffices to show that this remainder satisfies

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{||\mathbf{R}(\mathbf{x})||}{||\mathbf{x} - \mathbf{a}||} = 0. \tag{15.6}$$

This will prove both that $\mathbf{g} \circ \mathbf{f}$ is differentiable at $\mathbf{a}$ and that $D(\mathbf{g} \circ \mathbf{f})(\mathbf{a})$ is represented by the matrix $D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})$.

Although the proof may look long, the essential structure of the argument has already been presented: we will study the ratio $\frac{||\mathbf{R}(\mathbf{x})||}{||\mathbf{x}-\mathbf{a}||}$ and show that its limit as $\mathbf{x} \to \mathbf{a}$ is $0$. The rest of the proof itself is a careful computation of this limit.

**Step 2** We begin by making the substitution $\mathbf{t} = \mathbf{x} - \mathbf{a}$, so that Equation 15.5 becomes

$$\mathbf{R}(\mathbf{a} + \mathbf{t}) = (\mathbf{g} \circ \mathbf{f})(\mathbf{a} + \mathbf{t}) - (\mathbf{g} \circ \mathbf{f})(\mathbf{a}) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t} \tag{15.7}$$

$$= \mathbf{g}(\mathbf{f}(\mathbf{a} + \mathbf{t})) - \mathbf{g}(\mathbf{f}(\mathbf{a})) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t}. \tag{15.8}$$

The function $\mathbf{f}$ is differentiable at $\mathbf{a}$, which means that

$$\mathbf{f}(\mathbf{a} + \mathbf{t}) = \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})\mathbf{t} + \mathbf{R}_\mathbf{f}(\mathbf{a} + \mathbf{t}), \quad \text{where } \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{f}(\mathbf{a} + \mathbf{t})||}{||\mathbf{t}||} = 0. \tag{15.9}$$

Plugging Equation 15.9 into Equation 15.8, we have

$$\mathbf{R}(\mathbf{a} + \mathbf{t}) = \mathbf{g}[\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})\mathbf{t} + \mathbf{R}_\mathbf{f}(\mathbf{a} + \mathbf{t})] - \mathbf{g}(\mathbf{f}(\mathbf{a})) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t}. \tag{15.10}$$

Let $\mathbf{s} = D\mathbf{f}(\mathbf{a})\mathbf{t} + \mathbf{R}_\mathbf{f}(\mathbf{a} + \mathbf{t})$. Then Equation 15.10 becomes

$$\mathbf{R}(\mathbf{a} + \mathbf{t}) = \mathbf{g}[\mathbf{f}(\mathbf{a}) + \mathbf{s}] - \mathbf{g}(\mathbf{f}(\mathbf{a})) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t}. \tag{15.11}$$

Next, we use the fact that $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{a})$ to write

$$\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s}) = \mathbf{g}(\mathbf{f}(\mathbf{a}))+D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{s}+\mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s}), \quad \text{where } \lim_{\mathbf{s}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a}) + \mathbf{s})||}{||\mathbf{s}||} = 0. \tag{15.12}$$

Starting from Equation 15.11, we can rewrite the remainder $\mathbf{R}(\mathbf{a} + \mathbf{t})$ as a sum of two terms:

$$\mathbf{R}(\mathbf{a} + \mathbf{t}) = \mathbf{g}[\mathbf{f}(\mathbf{a}) + \mathbf{s}] - \mathbf{g}(\mathbf{f}(\mathbf{a})) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t} \tag{15.13}$$

$$= \mathbf{g}(\mathbf{f}(\mathbf{a})) + D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{s} + \mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a}) + \mathbf{s}) - \mathbf{g}(\mathbf{f}(\mathbf{a})) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t} \tag{15.14}$$

$$= D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{s} + \mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a}) + \mathbf{s}) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t} \tag{15.15}$$

$$= D\mathbf{g}(\mathbf{f}(\mathbf{a}))[D\mathbf{f}(\mathbf{a})\mathbf{t} + \mathbf{R}_\mathbf{f}(\mathbf{a} + \mathbf{t})] + \mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a}) + \mathbf{s}) - D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t} \tag{15.16}$$

$$= D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t} + D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{R}_\mathbf{f}(\mathbf{a} + \mathbf{t}) + \mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a}) + \mathbf{s})$$

$$- D\mathbf{g}(\mathbf{f}(\mathbf{a}))D\mathbf{f}(\mathbf{a})\mathbf{t} \tag{15.17}$$

$$= D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{R}_\mathbf{f}(\mathbf{a} + \mathbf{t}) + \mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a}) + \mathbf{s}). \tag{15.18}$$

Recall that our goal is to show that

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||}.$$

Applying the triangle inequality, we see that

$$\begin{aligned}\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} &= \lim_{\mathbf{t}\to\mathbf{0}} \frac{||D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{R}_\mathbf{f}(\mathbf{a}+\mathbf{t}) + \mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{t}||} \\ &\le \lim_{\mathbf{t}\to\mathbf{0}} \frac{||D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{R}_\mathbf{f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} + \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{t}||}, \end{aligned} \tag{15.19}$$

if both the right-hand limits exist.

In Steps 3 and 4, we will show that both limits on the right-hand side of this inequality exist and are equal to zero. This will complete the proof of the Chain Rule.

**Step 3**

In this step, we will show that

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{||D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{R}_\mathbf{f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} = 0.$$

**Lemma 1.** Given any $n \times m$ matrix $M$, there exists some scalar $c \geq 0$ such that for all $\mathbf{x} \in \mathbf{R}^m$,

$$||M\mathbf{x}|| \leq c||\mathbf{x}||.$$

*Proof.* The proof of this lemma is left as an exercise. (See Exercise 23.) ☐

Applying Lemma 1 to $||D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{R}_\mathbf{f}(\mathbf{a}+\mathbf{t})||$ shows that there exists some $c \in \mathbf{R}$ such that

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{||D\mathbf{g}(\mathbf{f}(\mathbf{a}))\mathbf{R}_\mathbf{f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} \le \lim_{\mathbf{t}\to\mathbf{0}} \frac{c||\mathbf{R}_\mathbf{f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} = c \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} = 0.$$

The last equality follows because $\mathbf{f}$ is differentiable (Equation 15.9).

**Step 4**

In this step we will show that the second limit from Equation 15.19 vanishes:

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{t}||} = 0. \tag{15.20}$$

When $\mathbf{s} \neq \mathbf{0}$, we have the following:

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{t}||} = \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{t}||}\frac{||\mathbf{s}||}{||\mathbf{s}||} = \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R}_\mathbf{g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{s}||}\frac{||\mathbf{s}||}{||\mathbf{t}||}. \tag{15.21}$$

The limit on the right has two factors: we will show that the limit of the first one is zero, and that the second one is bounded from above.

Since g is differentiable,

$$\lim_{\mathbf{s}\to\mathbf{0}} \frac{||\mathbf{R_g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{s}||} = 0.$$

However, recall the definition of s:

$$\lim_{\mathbf{t}\to\mathbf{0}} \mathbf{s} = \lim_{\mathbf{t}\to\mathbf{0}}(D\mathbf{f}(\mathbf{a})\mathbf{t} + \mathbf{R_f}(\mathbf{a}+\mathbf{t})) = 0.$$

This shows that

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R_g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{s}||} = \lim_{\mathbf{s}\to\mathbf{0}} \frac{||\mathbf{R_g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{s}||} = 0.$$

Next, we prove that $\frac{||\mathbf{s}||}{||\mathbf{t}||}$ is bounded from above. Use the definition of s to make a substitution:

$$\begin{aligned}
\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{s}||}{||\mathbf{t}||} &= \lim_{\mathbf{t}\to\mathbf{0}} \frac{||D\mathbf{f}(\mathbf{a})\mathbf{t} + \mathbf{R_f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} \\
&\leq \lim_{\mathbf{t}\to\mathbf{0}}\left(\frac{||D\mathbf{f}(\mathbf{a})\mathbf{t}||}{||\mathbf{t}||} + \frac{||\mathbf{R_f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||}\right) \\
&\leq \lim_{\mathbf{t}\to\mathbf{0}} \frac{||D\mathbf{f}(\mathbf{a})\mathbf{t}||}{||\mathbf{t}||} + \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R_f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||}.
\end{aligned}$$

The last inequality holds if both of the right-hand terms are finite.

Apply Lemma 1 to the matrix $D\mathbf{f}(\mathbf{a})$ to get

$$\begin{aligned}
\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{s}||}{||\mathbf{t}||} &\leq \lim_{\mathbf{t}\to\mathbf{0}} \frac{c||\mathbf{t}||}{||\mathbf{t}||} + \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R_f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} \\
&\leq \lim_{\mathbf{t}\to\mathbf{0}} c + \lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R_f}(\mathbf{a}+\mathbf{t})||}{||\mathbf{t}||} \\
&= c + 0 \qquad (15.22)
\end{aligned}$$

Returning to Equations 15.21, we can now see that

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{||\mathbf{R_g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{s}||}\frac{||\mathbf{s}||}{||\mathbf{t}||} \leq \lim_{\mathbf{s}\to\mathbf{0}} \frac{||\mathbf{R_g}(\mathbf{f}(\mathbf{a})+\mathbf{s})||}{||\mathbf{s}||}c = (0)c = 0.$$

This almost completes the proof of the Chain Rule.

In Step 4, we made the assumption that $\mathbf{s} \neq \mathbf{0}$ and we divided by $||\mathbf{s}||$. However, there is a "trick" that allows us to complete the proof without this assumption. See Exercise 25 in Section 9 for a description of how this is done.

## 15.4 Proof of the Second Derivative Test

Suppose that the function $f : \mathcal{D}^n \to \mathbf{R}$ has continuous second partial derivatives at $\mathbf{a}$, and suppose that $\nabla f(\mathbf{a}) = \mathbf{0}$. Let $Q_\mathbf{a}$ be the quadratic form generated by $Hf(\mathbf{a})$, and suppose that $Q_\mathbf{a}$ is positive definite. We will prove that $f$ has a local minimum at $\mathbf{a}$.

Recall Theorem 20 from Section 11, which states that the remainder associated to the second-order Taylor expansion of $f$ vanishes subquadratically. That is,

$$f(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \frac{1}{2}Q_{\mathbf{a}}(\mathbf{x}-\mathbf{a}) + R_2(\mathbf{x})$$

where

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_2(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||^2} = 0. \tag{15.23}$$

By hypothesis, $Df(\mathbf{a}) = \nabla f(\mathbf{a}) = \mathbf{0}$, so we can simplify this equation to

$$f(\mathbf{x}) = f(\mathbf{a}) + \frac{1}{2}Q_{\mathbf{a}}(\mathbf{x}-\mathbf{a}) + R_2(\mathbf{x}), \quad \text{where} \quad \lim_{\mathbf{x}\to\mathbf{a}} \frac{R_2(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||^2} = 0. \tag{15.24}$$

We will study the terms on the right-hand side of Equation 15.24 in order to show that if $\mathbf{x}$ is sufficiently close to $\mathbf{a}$, then $f(\mathbf{x}) \geq f(\mathbf{a})$.

The matrix $Hf(\mathbf{a})$ is symmetric, so Theorem 25.2 in Levandosky's *Linear Algebra* implies that there is a basis $\{\mathbf{v_1}, ...\mathbf{v_n}\}$ for $\mathbf{R}^n$ consisting of eigenvectors for $Hf(\mathbf{a})$:

$$Hf(\mathbf{a})\mathbf{v_i} = \lambda_i \mathbf{v_i}.$$

Since $Q_{\mathbf{a}}$ is positive definite, $\lambda_i > 0$ for all $i$. Replacing $\mathbf{v}$ with $\frac{\mathbf{v}}{||\mathbf{v}||}$ if necessary, we may assume that $||\mathbf{v_i}|| = 1$ for all $i$.

Let $\lambda_{min} = \min(\lambda_1, \lambda_2, ...\lambda_n)$, and set $\epsilon = \frac{\lambda_{min}}{4}$. Equation 15.24 implies that there exists some $\delta > 0$ such that

$$||\mathbf{x}-\mathbf{a}|| < \delta \text{ implies } \frac{|R_2(\mathbf{x})|}{||\mathbf{x}-\mathbf{a}||^2} < \epsilon = \frac{\lambda_{min}}{4}.$$

Let $\mathbf{b} \in \mathbf{R}^n$ satisfy $||\mathbf{b}-\mathbf{a}|| < \delta$. We claim that $f(\mathbf{b}) \geq f(\mathbf{a})$.

To see this, first express $\mathbf{b}-\mathbf{a}$ as a linear combination of the basis vectors $\mathbf{v_i}$:

$$\mathbf{b}-\mathbf{a} = \sum_{i=1}^{n} c_i \mathbf{v_i}.$$

Then

$$\begin{aligned}
|R_2(\mathbf{b})| &\leq \frac{\lambda_{min}}{4}||\mathbf{b}-\mathbf{a}||^2 \\
&= \frac{\lambda_{min}}{4}(\mathbf{b}-\mathbf{a})\cdot(\mathbf{b}-\mathbf{a}) \\
&\leq \frac{\lambda_{min}}{4}\sum_{i=1}^{n} c_i^2 \\
&\leq \frac{1}{4}\sum_{i=1}^{n} \lambda_i c_i^2 \\
&= \frac{1}{4}Q_{\mathbf{a}}(\mathbf{b}-\mathbf{a})
\end{aligned}$$

Since $|R_2(\mathbf{b})| \leq \frac{1}{4}Q_{\mathbf{a}}(\mathbf{b}-\mathbf{a})$ and $Q_{\mathbf{a}}(\mathbf{b}-\mathbf{a}) > 0$, it follows that

$$\frac{1}{2}Q_{\mathbf{a}}(\mathbf{b}-\mathbf{a}) + R_2(\mathbf{b}) = p > 0.$$

Plugging this into Equation 15.24, we have

$$\begin{aligned} f(\mathbf{b}) &= f(\mathbf{a}) + \frac{1}{2}Q_{\mathbf{a}}(\mathbf{b}-\mathbf{a}) + R(\mathbf{b}) \\ &= f(\mathbf{a}) + p, \text{ where } p > 0. \end{aligned}$$

This shows that $f(\mathbf{b}) > f(\mathbf{a})$. This argument holds for any $\mathbf{b}$ such that $||\mathbf{b}-\mathbf{a}|| < \delta$, so $f$ has a local minimum at $\mathbf{a}$.

The proof that $f$ has a local maximum at $\mathbf{a}$ when $Hf(\mathbf{a})$ is negative definite is similar, and the proof that $f$ has a saddle point when $Hf(\mathbf{a})$ is indefinite is left as Exercise 19. $\square$

# 16 Appendix II: Differentiable Surfaces

In Section 10 we described a differentiable surface as a subset of $\mathbf{R}^3$ that looks locally like a graph of a differentiable function of two variables. In this section we'll make this description more precise.

## 16.1 Open Sets and Differentiable Surfaces

Suppose that $S \subset \mathbf{R}^3$ is defined by

$$S = \{\mathbf{x} \in \mathbf{R}^3 | F(\mathbf{x}) = 0\}$$

for some real-valued function $F$. Fixing a point $\mathbf{a} \in S$, we would like to describe the part of $S$ near $\mathbf{a}$ as the graph of some differentiable function of two variables, even if that description doesn't work globally.

Recall that an *open set* in $\mathbf{R}^n$ is a subset of $\mathbf{R}^n$ with the property that each point in the set has a neighborhood that's also contained in the set.

**Definition 16.1.** Let $S \subset \mathbf{R}^3$ and suppose that there is a collection of open sets $U_i$ with the following properties:

- each point of $S$ lies in some $U_i$ ; and
- for each $i$, the intersection $S \cap U_i$ is the graph of a differentiable function from $\mathbf{R}^2$ to $\mathbf{R}$.

Then $S$ is a *differentiable surface*.

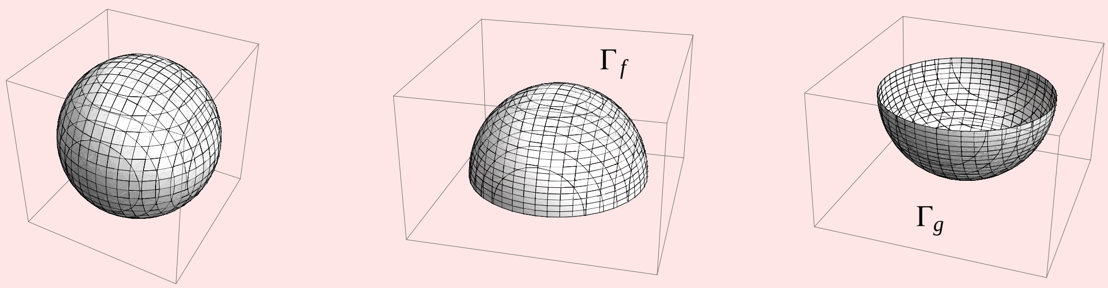
**Example 16.1.** Let $f : \mathcal{D}^2 \to \mathbf{R}$ be a differentiable function. In the open set $\mathbf{R}^3$, the entire surface $\Gamma_f$ is the graph of the differentiable function $f$, so $\Gamma_f$ is a differentiable surface.
This is a trivial example because it doesn't matter how we choose our open sets. In general, finding a collection of open sets which satisfies Definition 16.1 requires more care.

**Example 16.2.** *Let $S$ denote the unit sphere in $\mathbf{R}^3$:*

$$S = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}.$$

*Show that $S$ is a differentiable surface*
Although the sphere isn't the graph of any single function, it can be built piecewise from a collection of graphs.

We first consider the open set $U_1 = \{(x,y,z) \mid z > 0\}$. The intersection $S \cap U_1$ is the graph of the function
$$f(x,y) = \sqrt{1 - x^2 - y^2}$$
over the domain defined by $x^2 + y^2 < 1$.
Similarly, if we let $U_2 = \{(x,y,z) \mid z < 0\}$, then $S \cap U_2$ is the graph of the function
$$g(x,y) = -\sqrt{1 - x^2 - y^2}$$
over the domain $x^2 + y^2 < 1$.




In order to show that $S$ is a differentiable surface, we need to include the points of $S$ that lie on the $z = 0$ plane. Consider the open sets:
$$\begin{aligned} U_3 &= \{(x,y,z) \mid x > 0\} \\ U_4 &= \{(x,y,z) \mid x < 0\} \\ U_5 &= \{(x,y,z) \mid y > 0\} \\ U_6 &= \{(x,y,z) \mid y < 0\}. \end{aligned}$$
The intersection of $U_3$ with $S$ is the graph of the function
$$h(y,z) = \sqrt{1 - y^2 - z^2}$$
over the disc $y^2 + z^2 < 1$, and the intersection of $U_4$ with $S$ is the graph of the function
$$j(y,z) = -\sqrt{1 - y^2 - z^2}$$
over the disc $y^2 + z^2 < 1$.
The open sets $U_1, U_2, U_3$, and $U_4$ contain all the points of $S$ except $(0,1,0) \in U_5$ and $(0,-1,0) \in U_6$.
However, $S \cap U_5$ is the graph of
$$k(x,z) = \sqrt{1 - x^2 - z^2}.$$

Similarly, $S \cap U_6$ is the graph of

$$m(x,z) = -\sqrt{1-x^2-z^2}.$$

The open sets $\{U_i\}_{i=1}^6$ contain all the points of $S$. Since $S \cap U_i$ is the graph of a differentiable function, this proves that the unit sphere in $\mathbf{R}^3$ is a differentiable surface.

## 16.2 The Implicit Function Theorem

In Example 16.2, we proved that the sphere is a differentiable surface by explicitly solving for a set of differentiable functions whose graphs collectively compose $S$. However, this was a fairly involved process. In this section we introduce a special case of the powerful Implicit Function Theorem as a tool for showing that level sets are differentiable surfaces.

Suppose that $F : \mathcal{D}^{n+1} \to \mathbf{R}$ is defined in a neighborhood of the point $\mathbf{a}$. Denote the level set at height $F(\mathbf{a})$ by $S$:

$$S = F^{-1}(F(\mathbf{a})).$$

**Theorem 29** (Implicit Function Theorem)**.** Suppose that the derivative of $F : \mathcal{D}^{n+1} \to \mathbf{R}$ is continuous in a neighborhood of $\mathbf{a}$, and suppose $\frac{\partial F}{\partial x_i}(\mathbf{a}) \neq 0$ for some $i$. Then in a neighborhood of $\mathbf{a}$, $S$ is the graph of a differentiable function $g : \mathcal{D}^n \to \mathbf{R}$.

When $n = 2$, Theorem 29 can be used to identify a level set as a differentiable surface. When $n = 1$, Theorem 29 can be used to identify differentiable curves, which are curves that admit tangent lines.

**Example 16.3.** *Consider the unit circle in* $\mathbf{R}^2$*:*

$$C = \{(x,y) | x^2 + y^2 = 1\}.$$

*Use the Implicit Function Theorem to identify where $C$ can be described as the graph of a differentiable function.*

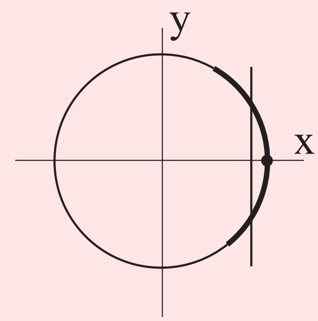
First, observe that the circle is the level set at height $1$ of the function $F(x,y) = x^2 + y^2$. Both partial derivatives of $F$ are continuous at every point in $\mathbf{R}^2$. The Implicit Function Theorem says that if $\frac{\partial F}{\partial y}(a,b) \neq 0$ for $(a,b)$ on $C$, then some neighborhood of $(a,b)$ in $C$ is the graph of a differentiable function $g : \mathcal{D}^1 \to \mathbf{R}$.

$$\text{Since } \frac{\partial F}{\partial y}(x,y) = 2y, \quad \frac{\partial F}{\partial y}(a,b) \neq 0 \text{ if and only if } b \neq 0.$$

In fact, for $(a,b)$ with $b \neq 0$, we can explicitly find $g$, much as we did in Example 16.2:

When $(a, b) \in C$ with $b \in (0, 1]$, $C$ is locally the graph of $g(x) = \sqrt{1 - x^2}$.
When $(a, b) \in C$ with $b \in [-1, 0)$, $C$ is locally the graph of $g(x) = -\sqrt{1 - x^2}$.
However, the partial derivative $\frac{\partial F}{\partial y}$ vanishes at $(1, 0)$ and $(-1, 0)$.

The figure shows that in any neighborhood of these points, the circle fails the vertical line test. This shows directly that $C$ is not the graph of a function of $x$ near $(1, 0)$ or $(-1, 0)$.




Reversing the roles of $x$ and $y$ in this example shows that away from the points $(0, 1)$ and $(0, -1)$, the circle can be described locally as the graph of a differentiable function of $y$.

## 16.3 Worked Examples

**Example 16.4.** *Consider the function*

$$F(x, y, z) = (3 - \sqrt{x^2 + y^2})^2 + z^2.$$

*Let $T$ be the level set of $F$ at height 1: $T = F^{-1}(1)$. Prove that $T$ is a differentiable surface.*

We will use the Implicit Function Theorem to show that $T$ is a differentiable surface, without explicitly solving for any of the functions which locally describe it as a graph. The surface $T$ is called a *torus*.

The function $F$ has continuous partial derivatives in a neighborhood of $T$:

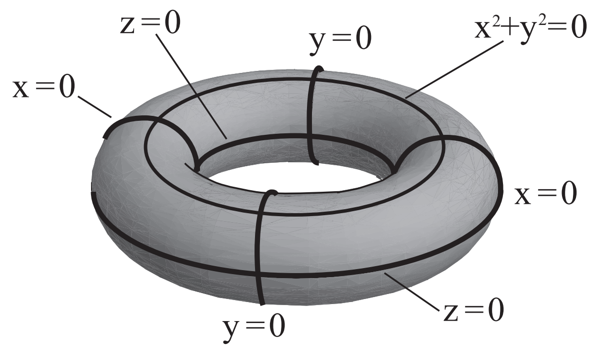
$$\begin{aligned} F_x(x, y, z) &= \frac{-2x(3 - \sqrt{x^2 + y^2})}{(\sqrt{x^2 + y^2})} \\ F_y(x, y, z) &= \frac{-2y(3 - \sqrt{x^2 + y^2})}{(\sqrt{x^2 + y^2})} \\ F_z(x, y, z) &= 2z \end{aligned}$$

The partial derivative $F_z$ vanishes only when $z = 0$, so the Implicit Function Theorem implies that away from points of the form $(x, y, 0)$, $T$ is a differentiable surface. More precisely, let

$$\begin{aligned} U_1 &= \{(x, y, z) \big| z > 0\} \\ U_2 &= \{(x, y, z) \big| z < 0\}. \end{aligned}$$

The Implicit Function Theorem implies that in $U_1$ and $U_2$, $T$ can be described locally as the graph of a differentiable function of $x$ and $y$.

It remains to show that $T$ is the graph of a differentiable function in a neighborhood of the points on $T$ of the form $(x, y, 0)$. The partial derivative $F_x$ vanishes when $x = 0$ and when $\sqrt{x^2 + y^2} = 3$. Thus, away from these points, $T$ can be described locally as the graph of a function of $y$ and $z$. Observe that the points on $T$ which satisfy $\sqrt{x^2 + y^2} = 3$ all have non-zero $z$ coordinate, so they are covered by the previous computation.




The partial derivatives $F_z$ and $F_x$ have shown us that away from points of the form $(0, y, 0)$, $T$ is a differentiable surface. To show that the entire torus is a differentiable surface, we turn to the partial derivative $F_y$. Since $F_y(0, y, 0) \neq 0$ for points on $T$, we see that in a neighborhood of $(0, \pm 2, 0)$, and $(0, \pm 4, 0)$, $T$ can be described as the graph of a function of $x$ and $z$.

At every point on the torus, at least one of the partial derivatives of $F$ is non-zero. Thus every point on $T$ has a neighborhood in which $T$ can be described as the graph of a differentiable function of two variables. This proves that $T$ is a differentiable surface.

# 17 Index of notation

$\nabla f(\mathbf{a})$- the gradient of the function $\mathbf{f}$ at the point $\mathbf{a}$
$B(\mathbf{a}, \delta_{\mathbf{a})}$- an open ball of radius $\delta_{\mathbf{a}}$ centered at the point $\mathbf{a} \in \mathbf{R}^n$
$\delta$- usually, a small positive number
$\frac{df}{dx}(a)$- the derivative of $f(x)$ at the point $a$
$\frac{d^2f}{dx^2}(a)$- the second derivative of $f$ at the point $a$
$\frac{\partial f}{\partial x}$- the partial derivative of $f$ with respect to the variable $x$
$\frac{\partial^2 f}{\partial x^2}$- the second partial derivative of $f$ with respect to the variable $x$
$\frac{\partial^2 f}{\partial x \partial y}$- the second partial derivative of $f$ with respect to the variables $x$ and $y$
$D\mathbf{f}(\mathbf{x})$- the total derivative of the function $\mathbf{f}$
$D\mathbf{f}(\mathbf{a})$- the total derivative of the function $\mathbf{f}$ at the point $\mathbf{a}$
$D_{\mathbf{v}}f(\mathbf{a})$- the directional derivative of the function $\mathbf{f}$ in the direction $\mathbf{v}$ at the point $\mathbf{a}$
$\mathcal{D}^n$, $\mathcal{E}^n$- a subset of $\mathbf{R}^n$
$\mathbf{e_i}$- the $i^{th}$ coordinate vector (equivalently, standard basis vector) in $\mathbf{R}^n$
$\epsilon$- usually, a small positive number
$f'(a)$- the derivative of $f$ at the point $a$
$f''(a)$- the second derivative of $f$ at the point $a$
$f^{(i)}$- the $i^{th}$ derivative of $f$
$f_{x_i}$- the partial derivative of $f$ with respect to the variable $x_i$
$f_x$- the partial derivative of $f$ with respect to the variable $x$
$f_{xy}$- the second partial derivative of $f$ with respect to the variables $x$ and $y$
$f_i$- the $i^{th}$ coordinate function of a vector-valued function $\mathbf{f}$
$\Gamma_f$- the graph of the function $f$
$Hf(\mathbf{a})$- the Hessian of the function $f$ at the point $\mathbf{a}$
$\mathcal{I}$- an interval in $\mathbf{R}$, oftens used to denote the domain of a paramaterized curve
$\mathbf{L}$- often used to denote the linearization of a function
$\mathbb{N}$- the natural numbers $\{1, 2, 3, ...\}$
$\mathbb{Q}$- the rational numbers
$T_2(\mathbf{x})$- the second order Taylor approximation of a function at some (specified) point
$\mathbb{Z}$- the integers $\{0, 1, -1, 2, -2, ...\}$

# 18 Index

# 19 Hints and answers for selected exercises

**Section 1**

1. (a) domain: $\{(x,y) \in \mathbf{R}^2 \,|xy \geq 0\}$; range: $\{z \in \mathbf{R} \,|z \geq 0\}$
(d) domain: $\mathbf{R}^3$; range: $\{z \in \mathbf{R} \,|z > 0\}$
(g) domain: $\{(x,y) \in \mathbf{R}^2 \,|y \neq 0\}$; range: $\mathbf{R}$

3. No. The pairs $(a_1, b_3)$ and $(a_1, b_2)$ violate the requirement that each point in the domain be associated to a unique point in the codomain.

5. Hint: Find two points on $T$ with the same $w$ and $x$ coordinates.

7. Every line through the origin intersects $S^n$ twice, so $S^n$ fails the line test.

9. (a) F; (c) T; (e) F

11. (IF) Suppose that each coordinate function $f_i : \mathbf{R}^n \to \mathbf{R}^m$ is odd. Then

$$\begin{aligned}
\mathbf{f}(-\mathbf{x}) &= \big(f_1(-\mathbf{x}), f_2(-\mathbf{x}), ...f_m(-\mathbf{x})\big) \\
&= \big(-f_1(\mathbf{x}), -f_2(\mathbf{x}), ... -f_m(\mathbf{x})\big) \\
&= -\big(f_1(\mathbf{x}), f_2(\mathbf{x}), ...f_m(\mathbf{x})\big) \\
&= -\mathbf{f}(\mathbf{x}).
\end{aligned}$$

25. (a) 3; (b) 6; (c) 1; (d) 2; (e) 4; (f) 5

**Section 2**

1. Hint: Let $\delta = \frac{\epsilon}{2}$.

3. Hint: Let $\delta = \sqrt{\epsilon}$.

5. Hint: Let $\delta = \frac{\epsilon}{2}$.

7. Hint: $2x^2 - 3x - 9 = 2(x-3)^2 + 9(x-3)$

9. Hint: $3x + x^2 + 2 = (x+1)^2 + (x+1)$

11. Hint: $x^3 + 2x^2 - 3 = (x-1)^3 + 5(x-1)^2 + 7(x-1)$

13. $\frac{1}{6}$

15. $2$

17. $0$

19. does not exist

21. $3$

23. does not exist

25. $12$

27. does not exist

29. does not exist

31. yes

33. no

## Section 3

1. $y = 3x - 2$

3. $y = x$

5. $y = -20x + 28$

7. Hint: compare the limit from Defintion 3.1 as $x$ approaches $0$ from above to the limit as $x$ approaches $0$ from below.

9. The graph of $f$ is a line, so $f(x) = c_1 x + c_2$ for $c_i \in \mathbf{R}$.

11. (a) Hint: Compute the limits from Defintion 3.1 as $x$ approaches $2$ from above to the limit as $x$ approaches $2$ from below.

13. $L = 1$.

15. $L = 0$.

17. $L = 0$.

19. $L = -\frac{37}{5}$.

21. does not exist ($L = \infty$).

## Section 4

1. $(2, 6) + t(4, 1)$

3. $(1, 1, 0) + t(1, 0, 1)$

5. $\mathbf{g}(t) = \big(2\cos t, 2\sin t, 0\big)$

7. $\mathbf{g}(t) = \big(\frac{\pi}{2}, 2\sin t, 2\cos t\big)$

9. (b) $(1, 0, 0)$ is not in the image of g.

11. $y = 2x$ , $y = -2x$.

13. $\mathbf{v} = (3t^2 + 1\ ,\ 4te^{2t^2})$, $\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{(3t^2+1\ ,\ 4te^{2t^2})}{\sqrt{9t^4+(6+16e^{4t^2})t^2+1}}$, $\mathbf{a} = (6t\ ,\ 4e^{2t^2} + 16t^2e^{2t^2})$.

15. $\mathbf{v} = (-4\sin 4t,\ 2\cos 2t,\ 2t)$, $\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{(-2\sin 4t,\ \cos 2t,\ 2t)}{\sqrt{4\sin^2 4t+\cos^2 2t+t^2}}$, $\mathbf{a} = (-16\cos 4t, -4\sin 2t, 2)$

17. (c) $g(t) = \ln(2t) - 3$

19. If $\mathbf{f}'(t)\cdot\mathbf{f}'(t) = \mathbf{0}$, then $x'(t)x(t) + y'(t)y(t) = 0$. This implies $0 = 2\big(x'(t)x(t) + y(t)y'(t)\big) = \frac{d}{dt}\big((x(t))^2 + (y(t))^2\big)$. But if the derivative of $(x(t))^2 + (y(t))^2$ is $0$, then $(x(t))^2 + (y(t))^2$ is a constant, so the image of $\mathbf{f}$ lies on a circle.

21. Let $\mathbf{r}(t) = \begin{bmatrix} r_1(t) \\ r_2(t) \\ r_3(t) \end{bmatrix}$, and let $\mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{bmatrix}$.

$$\begin{aligned}
\frac{d}{dt}\big(\mathbf{r}(t)\cdot\mathbf{s}(t)\big) &= \frac{d}{dt}\big(r_1(t)s_1(t) + r_2(t)s_2(t) + r_3(t)s_3(t)\big) \\
&= r_1'(t)s_1(t) + r_1(t)s_1'(t) + r_2'(t)s_2(t) + r_2(t)s_2'(t) + r_3'(t)s_3(t) + r_3(t)s_3'(t) \\
&= r_1'(t)s_1(t) + r_2'(t)s_2(t) + r_3'(t)s_3(t) + r_1(t)s_1'(t) + r_2(t)s_2'(t) + r_3(t)s_3'(t) \\
&= \mathbf{r}'(t)\mathbf{s}(t) + \mathbf{r}(t)\mathbf{s}'(t)
\end{aligned}$$

23. (a) $\mathbf{a}_T(1) = \frac{14}{5}(1,\ 2),\ \mathbf{a}_N(1) = \frac{18}{5}(2,\ -1)$

(b) $\mathbf{a}_T = 0,\ \mathbf{a}_N = (-\cos t,\ -\sin t)$

(c) $\mathbf{a}_T = (20t^3 + 30t)(1,\ 4),\ \mathbf{a}_N = 0$

**Section 5**

1. $0$

3. does not exist

5. does not exist

7. $0$

9. $0$

11. does not exist

13. $0$

15. $\frac{-2}{7}$

17. does not exist

19. $1$

21. $0$

23. $0$

25. $0$

27. $1$

31. no

33. yes

**Section 6**

1. Hint: use $\delta = \sqrt{\epsilon}$

3. Hint:
$$x^2 + y^2 - 5 = (x-1)^2 + (y-2)^2 + 2(x-1) + 4(y-2)$$

5. Hint:
$$x^2 + 2y - 6 = (x-2)^2 + 2(y-1) + 4(x-2)$$

7. Hint:
$$x^4 + y^4 \geq x^4$$

9. Hint:
$$xy + y - 4 = (x-1)(y-2) + 2(y - 2_+ 2(x-1)$$

**Section 7**

5.
$$Df(0,2) = \begin{bmatrix} 3 & 2 \\ 7 & -11 \end{bmatrix}$$

7.
$$Df(0,2) = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$$

9.
$$D\mathbf{f}(3,4,0,0) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & \frac{-6}{25^2} & \frac{-8}{25^2} & 0 \end{bmatrix}$$

11.
$$D\mathbf{f}(1,1) = \begin{bmatrix} 3 & -3 \\ 2\ln 2 & 0 \end{bmatrix}$$

13.
$$DF(\mathbf{a}) = \begin{bmatrix} 4 & 0 \\ 1 & 2 \\ 0 & 2 \end{bmatrix}$$

15. (before evaluating)

$$\begin{aligned} f_s s &= \frac{3}{4}s^{\frac{-1}{2}}(s + 5\sin t) + \frac{3}{2}s^{\frac{1}{2}} + \frac{3}{2}s^{\frac{1}{2}} \\ f_{st} &= f_{ts} = \frac{15}{2}s^{\frac{1}{2}}\cos t \\ f_{tt} &= -5s^{\frac{3}{2}}\sin t \end{aligned}$$

17. (before evaluating)

$$\begin{aligned} f_{uu} &= 2vw^{\frac{-1}{2}} \\ f_{vv} &= 0 \\ f_{ww} &= \frac{3}{4}u^2vw^{\frac{-5}{2}} + \frac{-3}{4}w^{\frac{-3}{2}} \\ f_{uw} &= f_{wu} = uvw^{\frac{-3}{2}} \\ f_{vu} &= f_{uv} = 2uw^{\frac{-1}{2}} \\ f_{wv} &= f_{vw} = \frac{-1}{2}u^2w^{\frac{-3}{2}} \end{aligned}$$

19.

$$\begin{aligned} f_{xx} &= 4y^3e^{2x} + 4xy^3e^{2x} \\ f_{xy} &= f_{yx} = 3y^2e^{2x} + 6xy^2e^{2x} \\ f_{yy} &= 6yxe^{2x} \end{aligned}$$

21.

$$\begin{aligned} f_{xyy} &= \frac{1}{x}\big( -4y^2\cos(y^2+1) - 2\sin(y^2+1)\big) \\ f_{yyx} &= \frac{1}{x}\big( -4y^2\cos(y^2+1) - 2\sin(y^2+1)\big) \\ f_{xyx} &= \frac{1}{x^2}\big(2y\sin(y^2+1)\big) \end{aligned}$$

23. (b) Yes, since Clairaut's Theorem implies $F_{zy} = F_{yz}$.

25. $b = 0$

27. (a) $f_{xy}(0,0) = 1$, $f_{yx}(0,0) = 0$

29.
$$\mathbf{L}(t) = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

35.
$$Df(0,0)(x,y) = \begin{bmatrix} 2x & 2y \\ 0 & 2 \end{bmatrix}_{(0,0)} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (0, 2y).$$

37.
$$D\mathbf{f}(\mathbf{a}) = \begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix}$$

37. (a) see Exercise 31; (b) see Example 7.9; (c) see Theorem 14; (d) see Definition 7.2; (e) see Definition 7.4.

## Section 8

1. $\frac{\partial z}{\partial r}(1,0) = 30$, $\frac{\partial z}{\partial s}(1,0) = 8$ 3. $D(g \circ \mathbf{f})(1,2) = [-116\ 336]$ 5. 3

7.
$$D\mathbf{f} \circ \mathbf{g} = \begin{bmatrix} \frac{1}{r^2} \sin\left(\frac{1}{s}\right) \sin\left(\frac{1}{r}\right) & -\frac{1}{s^2} \cos\left(\frac{1}{s}\right) \cos\left(\frac{1}{r}\right) + 2se^{s^2} \\ \frac{-1}{sr^2}\left(1 + \ln\left(\frac{s}{r}\right)\right) - \frac{s^3}{r^2} & \frac{1}{s^2 r}\left(1 - \ln\left(\frac{s}{r}\right)\right) + \frac{3s^2}{r} \end{bmatrix}$$

9.
$$D(f \circ \mathbf{g}) = 3(u-2)^2 + (4u+2)\ln(u^2+1) + \frac{4u^3 + 4u^2 + 4u}{u^2+1}$$

11.
$$D(\mathbf{f} \circ \mathbf{g}) = \begin{bmatrix} 34xy & 17x^2 \\ \frac{12}{x} \ln x^2 + 2xy & x^2 \end{bmatrix}$$

13.
$$D(\mathbf{g} \circ \mathbf{h})(1,2) = \begin{bmatrix} 0 & 4 \\ 10 & 20 \\ -324 & 648 \\ 52 & 40 \end{bmatrix}$$

19. (a) $\begin{bmatrix} \sin\theta\cos\phi & \rho\cos\theta\cos\phi & -\rho\sin\theta\sin\phi \\ \sin\theta\sin\phi & \rho\cos\theta\sin\phi & \rho\sin\theta\cos\phi \\ \cos\theta & -\rho\sin\theta & 0 \end{bmatrix}$.

(b)

$$\begin{aligned}\frac{\partial w}{\partial \rho} &= \frac{\partial f}{\partial x}\sin\theta\cos\phi + \frac{\partial f}{\partial y}\sin\theta\sin\phi + \frac{\partial f}{\partial z}\cos\theta \\ \frac{\partial w}{\partial \theta} &= \frac{\partial f}{\partial x}\rho\cos\theta\cos\phi + \frac{\partial f}{\partial y}\rho\cos\theta\sin\phi - \frac{\partial f}{\partial z}\rho\sin\theta \\ \frac{\partial w}{\partial \phi} &= -\frac{\partial f}{\partial x}\rho\sin\theta\sin\phi + \frac{\partial f}{\partial y}\rho\sin\theta\cos\phi\end{aligned}$$

(c) $\frac{\partial T}{\partial \rho}(2, \frac{\pi}{4}, \frac{3\pi}{2}) = -24$, $\frac{\partial T}{\partial \theta}(2, \frac{\pi}{4}, \frac{3\pi}{2}) = 24$, $\frac{\partial T}{\partial \phi}(2, \frac{\pi}{4}, \frac{3\pi}{2}) = 2\sqrt{2}$

**Section 9**

1. $\frac{-68}{5}$

3. $\frac{2}{3}$

5. $\frac{-68}{5}$

7. $\frac{2}{3}$

9. $\frac{-31}{5}$

11. $\frac{-7}{\sqrt{5}}$

13. $\frac{-\sqrt{3}}{81}$

15. $\frac{-4}{3}$

17. $\frac{-49}{2}$

19. (a) $16(\sqrt{3}+1)$ , (b) $-16\sqrt{2}$

21. (a) Hint: $\nabla h(\frac{\pi}{6}, \frac{\pi}{6})$;
(b) any direction orthogonal to $\nabla h(\frac{\pi}{6}, \frac{\pi}{6})$

23. $\begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}$

25. $y = -2x + \frac{\pi}{2}$

27. $(\pm\frac{1}{\sqrt{15}}, \mp\frac{7}{\sqrt{15}})$

**Section 10**

1. $-2x + 4y + z - 11 = 0$

3. $6x + 8y + 4z - 30 = 0$

5. $z = 0$

7. $x - \frac{1}{2}y - \frac{1}{2}z = 0$

9. $8x + 5z = -14$

11. $z = 0$

13. $x + \frac{3}{e}y + \frac{6}{e}z = 0$

15. $z = -2\sqrt{\pi}x + 2\pi$

17. $x = 2y + 3z - 2$

19. $z = -e^{\pi}x - \pi e^{\pi}y + 2\pi e^{\pi}$

21. $(4, \frac{1}{2}, 1)$

23. $(0, \frac{\pi}{2}, 3)$, $(0, \frac{3\pi}{2}, 13)$,$(0, \frac{5\pi}{2}, 3)$...

25. $0$

**Section 11**

1. $L(t) = 6t - 11$

3. $\mathbf{L}(u, v) = \begin{bmatrix} 2^{-\frac{3}{2}} + 3 \cdot 2^{-\frac{5}{2}}(u+1) \\ 26^{-\frac{5}{2}} - 25 \cdot 26^{-\frac{7}{2}}(v-5) \\ 25u - 10v + 50 \end{bmatrix}$

5. $\mathbf{L}(u, v, w) = \begin{bmatrix} 17u + 2v + w - 3 \\ 13u \end{bmatrix}$

7. $L(w, x, y, z) = 1 - x + y + 10z$

9.$T_2(u, v) = u^2 + uv,\ \ f(0.1, 0.2) \approx 0.03$

11. $T_2(w, z) = 33 + \frac{3}{2}(w-1) + 20(z-4) + \frac{3}{8}(w-1)^2 + \frac{15}{4}(z-4)^2$,
$b(1.02, 3.96) \approx 33.23615$

13. $T_2(x, y, z) = x + z$, $p(2.03, 1.05, -.97) \approx 1.060925$

15. (a) $\mathbf{L}(x, y) = (x + y - 7, 4x - 12y + 38, -6y + 2x - 1)$
(b) $\mathbf{L}(x, y) = (x + y - 7, 4x - 12y + 38, -6y + 2x - 1)$

19. Hint: Look at Example 11.4.

**Section 12**

1. Hint: Use the Mean Value Theorem.

3. (c) saddle point

5. (c) saddle point

7. $(\frac{-1}{2}, \frac{1}{3})$ is a local maximum.

9. $(\ln 2, \frac{-1}{2})$ is a saddle point.

11. The critical points are $(n\pi, \frac{2m+1}{2}\pi)$, $(\frac{2m+1}{2}\pi, n\pi)$ for $n, m \in \mathbb{Z}$.

13. $|k| < 2$

15. $4ab > 1$

17. Hint: Along the line $y = 0$, $g(x, 0) = x^3$.

**Section 13**

1. Hint: Find a set which doesn't contain all its boundary points.

3. closed: (b), (c), (d), (e), (f), (g); bounded: (a), (b), (g), (h)

5. $f(1,0) = 1$, $f(1,4) = 9$

7. $f(0, \frac{1}{2}) = \frac{11}{4}$, $f(\pm\frac{\sqrt{3}}{2}, \frac{-1}{2}) = \frac{21}{4}$

9. (b) $(0,0)$ is a local minimum, and $(\frac{-4}{9}, \frac{-4}{27})$ is a saddle point.
(c) No. (Consider the line $y = 0$. What happens as $x \to \pm\infty$?)

11. Hint: You can parameterize the sphere using the spherical coordinates from Exercise 19 in Section 8 and setting $\rho = \sqrt{24}$.

13. $(0,0)$ is a saddle point, and $(2,2)$ is a local maximum.

15. $(0,0)$ and $(4,0)$ are saddle points, and $(2,2)$ is a local minimum.

17. Hint: Maximize the function $xy(24 - x - y)$.

19. Hint: Minimize the square of the distance between $(0,-3,3)$ and points of the form $(y^2 + z^2, y, z)$.

**Section 14**

1. $(\frac{16}{5}, \frac{6}{5}), (\frac{-16}{5}, \frac{-6}{5})$

3. $(4,4,4)$

5. $(\sqrt{2}, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (-\sqrt{2}, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$

7. $(\frac{3}{2}, \frac{3}{2}, 0)$

13. First use Lagrange multipliers to find the points on $x^2 - yz = 7$ *and* which satisfy $x^2 + y^2 + z^2 \leq 1{,}000{,}000$ which are closest to the origin. (Note that such points must exist, since the ball is closed and bounded.) This yields $(\pm\sqrt{7}, 0, 0)$, and clearly any points outside the ball of radius $1{,}000{,}000$ are further away.

15. Hint: Use Lagrange mulitpliers to check for maxima on the three faces and two edges of the boundary. Make sure that the maxima you find actually lie on the cylinder.

17. Hint: Show that when $A$, $B$, and $C$ are non-zero, any nearest point satisfies

$$\frac{x - x_0}{A} = \frac{y - y_0}{B} = \frac{z - z_0}{C} = \lambda\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}.$$

19. (a) Hint: Show that $\mathbf{r_2}$ is a scalar multiple of $\mathbf{r_1}$.

21. (b) Show that for any stationary state, the $y$ and $z$ coordinates of both particles are $0$.

23. Hint: Minimize $E$ subject to the constraint $x + y + z = 6$.

25. Hint: Maximize $U(C_1, S) = \ln(C_1) + \beta \ln(W + rS)$ subject to the constraint $C_1 + S = W$. Remember that $W$, $r$, and $\beta$ are constants given to you by the problem.