MULTIVARIABLE FUNCTIONS AND PARTIAL DERIVATIVES

A. HAVENS

Contents

0. Functions of Several Variables

§ 0.1. Functions of Two or More Variables

Definition. A real-valued function of two variables, or a real-valued *bivariate function*, is a rule for assigning a real number to any ordered pair (x, y) of real numbers in some set $D \subseteq \mathbb{R}^2$. We often label such functions by a symbol, such as f, and write $f(x, y)$ for the value of f with input (x, y). The inputs x and y are called *independent variables*. The set $D = \mathrm{Dom}(f)$ is called the *domain of* f. The set of all values f attains over D is called the *range of* f or *image of* D *by* f:

$$\mathrm{Range}(f) = f(D) = \{z \in \mathbb{R} \,|\, z = f(x, y), (x, y) \in D\}\,.$$

One may sometimes specify function labels and domain by writing things like "$f : D \to \mathbb{R}$", or "$g : E \to \mathbb{R}$", where D and E are known subsets of $\mathbb{R}^2$. This is meant to emphasize the interpretation of the function as a *map* from a region or subset of the plane to the real numbers. If no domain is specified, one should assume that $\mathrm{Dom}(f)$ is the "largest set possible" for the specified rule, meaning one includes any ordered pair (x, y) for which the rule gives a well defined value $f(x, y)$.

Example. The function $f(x, y) = \sqrt{x^2 + y^2}$ is a bivariate function which may be interpreted as returning, for a given point (x, y), its distance from the origin $(0, 0)$ in rectangular coordinates on $\mathbb{R}^2$. It is well defined for all points, since the expression $x^2 + y^2 \geq 0$ for all (x, y), and $\sqrt{t}$ is well defined for any nonnegative real numbers t. Thus the domain is $\mathrm{Dom}(f) = \mathbb{R}^2$. The range is all nonnegative real numbers, since for any given nonnegative real d, one can find points satisfying

$$d = \sqrt{x^2 + y^2}\,.$$

Indeed, we can say then that the *pre-image* of the value d is the set

$$f^{-1}(\{d\}) := \{(x, y) \,|\, x^2 + y^2 = d^2\}\,,$$

which is just the origin-centered circle of radius d or $(0, 0)$, if $d > 0$, or $d = 0$ respectively. Thus, the image/range of f is

$$f(\mathbb{R}^2) = \mathbb{R}_{\geq 0} = [0, \infty)\,.$$

Example. The domain of the function $f(x, y) = \arctan(y/x)$ is the set of all ordered pairs (x, y) with $x \neq 0$, i.e.,

$$\mathrm{Dom}\Big(\arctan(y/x)\Big) = \mathbb{R}^2 - \{(x, y) \,|\, x = 0\}\,.$$

Exercise 0.1. Can you give a geometric interpretation of the apparent discontinuity of $z = \arctan(y/x)$ along the y axis? (Hint: think about what $\arctan(y/x)$ means geometrically. If stuck, examine figure 4 in section 1.3, where the function is revisited.)

Exercise 0.2. State and sketch the natural domains of the following functions:

(a) $f(x, y) = \sqrt{36 - 4x^2 - 9y^2}$,

(b) $g(x, y) = \sqrt{\cos(x - y) - \cos(x + y)}$,

(c) $w(u, v) = \sin\Big(u \arcsin(v)\Big)$

(d) $k(\varphi, \theta) = \sec\Big[\ln(2 + \cos\varphi + \sin\theta)\Big]$

(e) (Challenge) $h(x, y) = (xy)^{\ln(e - y - x^2)}$.

Definition. A function of n variables is a rule f for assigning a value $f(x_1, \ldots, x_n)$ to a collection of n variables, which may be given as elements of a subset $D \subseteq \mathbb{R}^n$. Thus, $f : D \to \mathbb{R}$ is a real-valued map from ordered n-tuples of real numbers taken from the domain D.

Example 0.1. The function $F(x, y, z) = \frac{GMm}{x^2+y^2+z^2} = \frac{GMm}{\|\mathbf{r}\|^2}$ represents the magnitude of the force a central body of mass M at $(0, 0, 0)$ exerts on a smaller object of mass m positioned at $(x, y, z) \in \mathbb{R}^3$, where G is a constant, called the *universal gravitational constant*. The force is attractive, directed along a line segment connecting to the two bodies. Thus, to properly describe the gravitational force, we'd need to construct a *vector field*. This idea will be described later in the course.

What are the *level sets*, $F^{-1}(\{k\})$, of the gravitational force? Since objects each of mass m at equal distances should experience the same attractive force towards the central mass, we should expect radially symmetric surfaces as our level sets, i.e., we should expect spheres! Indeed, $k = F(\mathbf{r}) = \frac{GMm}{\|\mathbf{r}\|^2} \implies \|\mathbf{r}\|^2 = \frac{GMm}{k}$, whence the level set for a force of magnitude k is a sphere of radius $\sqrt{GMm/k}$.

Exercise 0.3. Write out appropriate set theoretic definitions of image and pre-image for an n variable function $f(x_1, \dots, x_n)$.

Exercise 0.4. Describe the natural domain of the function $f(x, y, z) = \frac{1}{x^2+y^2-z^2-1}$ as a subset of $\mathbb{R}^3$. What sort of subset is the pre-image $f^{-1}(\{1\})$?

---

§ 0.2. Graphs of Multivariate Functions

Definition. The graph of a bivariate function $f : D \to \mathbb{R}$ is the locus of points $(x, y, z) \in \mathbb{R}^3$ such that $z = f(x, y)$:

$$\mathcal{G}_f := \{(x, y, z) \in \mathbb{R}^3 \,|\, z = f(x, y), (x, y) \in D\}\,.$$

For "nice enough" bivariate functions f, the graph carves out a surface in 3-space, the shadow of which is the image of D under the embedding of $\mathbb{R}^2$ as the xy-plane in $\mathbb{R}^3$. This allows one to visualize much of the geometry of the graph and use it to study the function $f(x, y)$ by treating it as a height function for a surface over the image of D in the xy-plane.

Example. Consider the function $f(x, y) = 4 - \frac{1}{4}(x^2 + y^2)$. To understand the graph of $z = f(x, y)$, we can study *trace curves*. The vertical trace curves are curves made by intersecting the graph with planes of either constant x or y.

Clearly, if $y = k$ is constant, the equation $z = 4 - \frac{1}{4}(x^2 + k^2)$ gives a downward opening parabola in the plane $y = k$, with vertex at $(0, k, 4 - k^2/4)$. For larger $|k|$, the vertex has lower z height, and for $k = 0$ we get a parabola in the xz-plane with equation $z = 4 - x^4/4$ and the maximum height vertex at $(0, 0, 4)$.

By symmetry, we have a familiar story in planes $x = k$ with parabolae whose vertices are $(k, 0, 4 - k^2/4)$, and the maximum height vertex is also at $(0, 0, 4)$.

Finally, we study the *horizontal traces*, which correspond to constant heights. For constant $z = k$, we get the equation

$$k = 4 - \frac{1}{4}(x^2 + y^2) \implies 16 - 4k = x^2 + y^2\,,$$

which describes a circle of radius $2\sqrt{4 - k}$.

The surface is thus a downward opening circular paraboloid, as pictured in figure 1.

Unfortunately, functions in greater than 3 variables are not so readily amenable to such a visualization. We can still define a graph for a function of many variables:

Definition. The graph of a multivariate function $f : D \to \mathbb{R}$ of n variables is the locus of points $(x_1 \dots, x_n, x_{n+1}) \in \mathbb{R}^{n+1}$ such that $x_{n+1} = f(x_1, \dots, x_n)$:

$$\mathcal{G}_f := \{(x_1 \dots, x_n, x_{n+1}) \in \mathbb{R}^{n+1} \,|\, x_{n+1} = f(x_1, \dots, x_n), (x_1 \dots, x_n) \in D\}\,.$$

Observe that the graph of an n-variable function is thus a geometric subset of $(n+1)$-dimensional Euclidean space $\mathbb{R}^{n+1}$. For "nice enough" functions, the graph carves out a locally connected n-dimensional subset of $\mathbb{R}^{n+1}$; such a set is sometimes called a *hypersurface*.

Before we examine more graphs, we'll describe an important tool which aids in visualizing functions and constructing graphs.

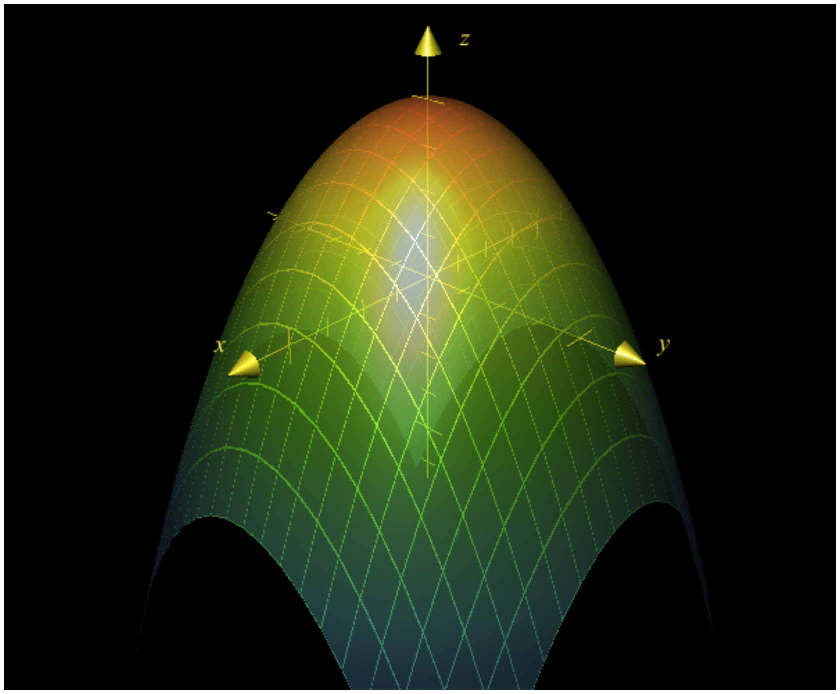


FIGURE 1. The graph of the paraboloid given by $z = f(x, y) = 4 - \frac{1}{4}(x^2 + y^2)$. Vertical trace curves form the pictured mesh over the surface.

§ 0.3. Contours and Level Sets

In the example above where we studied traces to understand the graph of a paraboloid. For a multivariable function $f(x, y)$, the horizontal traces of $z = f(x, y)$ are often the most useful ones: they capture the families of curves along which the function's value is constant. We view the traces as living in $\mathbb{R}^3$, but one can get a good understanding of how a function's values change by plotting the shadows of the traces in the xy-plane, and recording the information of which heights correspond to such a curve. This is how *contour maps* are made, which can tell a hiker or land surveyor about the terrain.

Definition. The *level curves* of a function $f(x, y)$, also called the *contours*, are the sets given as the pre-images of a single value in the range of f:

$$f^{-1}(\{k\}) := \{(x, y) \in D \,|\, k = f(x, y)\}\,.$$

For "sufficiently nice" functions, these sets describe (possibly disconnected) plane curves, with the exceptions of extreme values which give collections of points. For example, for the function $f(x, y) = 4 - \frac{1}{4}(x^2 + y^2)$, all the contours were circles except the contour for $k = 4$, which is a single point: $f^{-1}(\{4\}) = (0, 0)$, corresponding to the maximum value $f(0, 0) = 4$.

By considering vertically evenly spaced families of horizontal traces, one can generate a family of contours which captures the steepness of a graph. Fix an increment Δz, and an initial height k_0. Then generate a family of heights $k_n = k_0 + n\Delta z$, $n = 0, 1, \ldots m$ and consider the collection of level curves for the levels k_n. If the distance in the (x, y) plane between level curves for levels k_n and $k_{n\pm 1}$ is large near a point P on the k_n level curve, then the graph is not very steep there. However, if the level curves are close together near P, then the graph is steeper near P. Can you figure out how to determine the steepest direction from the level curves?

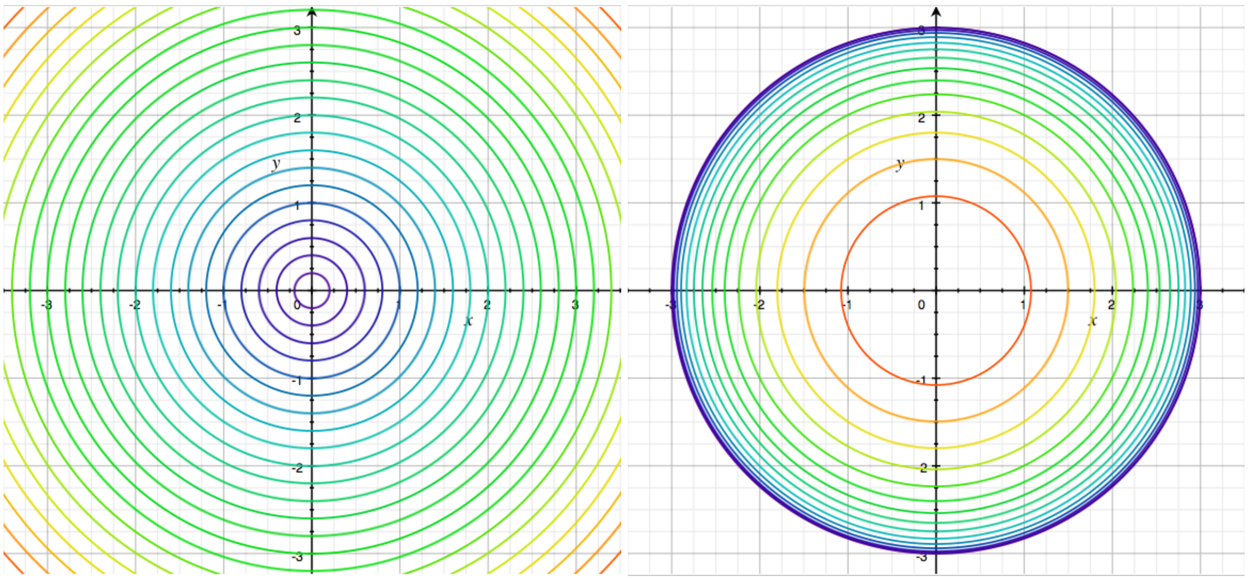
Example. Consider the two functions $f(x, y) = \sqrt{x^2 + y^2}$ and $g(x, y) = \sqrt{9 - x^2 - y^2}$. Observe that the domains are $\mathrm{Dom}(f) = \mathbb{R}^2$ and

$$\mathrm{Dom}(g) = \{(x, y) \in \mathbb{R}^2 \,|\, 0 \le x^2 + y^2 \le 9\} = \{\mathbf{r} : \|\mathbf{r}\| \le 3\} =: D_3\,.$$

The level curves are algebraically given by

$$f^{-1}(\{k\}) = \{(x,y)\in\mathbb{R}^2\,|\,x^2+y^2=k^2\}\,,\quad g^{-1}(\{k\}) = \{(x,y)\in D_3\,|\,x^2+y^2=9-k^2\}\,.$$

Both describe families of circles, but the circles given as level curves of f increase in radius as k grows, and are evenly spaced, where as the circles given as level curves of g shrink in radius as k ranges from 0 to 3, and become more tightly spaced as k approaches 3. Thus, the steepness of the graph of f is constant as we move along rays away from the origin, but for g the slope is steepest near the boundary $r=\sqrt{x^2+y^2}=3$. The level curves for each are pictured below in figure 2.



(a) (b)

Figure 2. (A) – The level curves for $f(x,y)=\sqrt{x^2+y^2}$ (B) – The level curves for $g(x,y)=\sqrt{9-x^2-y^2}$. Warmer colors indicate higher k value in both figures.

Of course, now we can attempt to understand the graphs themselves. The graph of $f(x,y)$ is just a cone: the level curves are just curves of constant distance from $(0,0)$, and so the z-traces are these concentric circles each lifted to a height equal to its radius. The graph of $g(x,y)$ is of the upper hemisphere of a radius 3 sphere centered at $(0,0,0)\in\mathbb{R}^3$: observe that $z=\sqrt{9-x^2-y^2} \implies x^2+y^2+z^2=9,\ z\geq 0$.

We can also define a notion similar to level curves for an n-variable function $f: D\to\mathbb{R}$:

Definition. The set given by the pre-image of a value $k\in f(D)$ is called the *level set with level* k, and is written

$$f^{-1}(\{k\}) := \{(x_1,\dots,x_n)\in D\,|\,f(x_1,\dots,x_n)=k\}\,.$$

For a "sufficiently nice" three variable function $f(x,y,z)$, the level sets are surfaces with implicit equations $k=f(x,y,z)$, except at extrema, where one may have collections of points and curves.

Exercise 0.5. Let $a\geq b>0$ be real constants. Give Cartesian or polar equations for the level curves of the following surfaces in terms of a, b, and $z=k$. Where relevant, determine any qualitative differences between the regimes $a>b$, $a=b$ and $a<b$. Sketch a sufficient family of level curves to capture the major features of each of the surfaces, and attempt to sketch the surfaces using a view which captures the essential features. You may use a graphing calculator or computer as an aid, but you must show the relevant algebra in obtaining the equations of the contours.

(a) $z = \sqrt{x^2 + y^2 + a^2}$

(b) $z = \sqrt{1 - b^2x^2 - a^2y^2}$

(c) $z = \sin(xy)$

(d) $z = ax^3 - 3bxy^2$

(e) $xz = 1 - \sqrt{x^2 + y^2}$

(f) $r^4 - (1+2xz)r^2 + (xz)^2 = 0$, where $r^2 = x^2 + y^2$ (Hint: work in polar/cylindrical coordinates).

Exercise 0.6. (*Challenge*: Try this without a computer, first!) Consider $z = \alpha x + \sqrt{x^2 + y^2}$. Suppose $0 < |\alpha| < 1$. What are the level curves? What about for $\alpha = 0$, $\alpha = 1$ and $\alpha > 1$? Sketch level curves and a surface for each scenario. (Hint: try writing things in polar coordinates; see also the discussion in section 5.4 of the notes on *Curvature, Natural Frames and Acceleration for Space Curves* and problem 23 of those notes.)

§ 0.4. Real-Valued Functions of Vector Inputs

It is often convenient to regard a multivariate function as a map from a set of vectors to the real numbers. In this sense, we can view multivariable functions as *scalar fields* over some domain whose elements are position vectors. E.g., the distance function from the origin for the plane can be written as the scalar field

$$f(\mathbf{r}) = \|\mathbf{r}\| = \sqrt{\mathbf{r}\cdot\mathbf{r}}\,.$$

Sometimes a multivariable function becomes easier to understand geometrically by writing it in terms of vector operations such as the dot product and computing magnitudes.

Example. Consider $f(x, y) = ax + by$ for nonzero constants a and b. The graph is a plane, but how do a and b control the plane? If we rewrite f as $f(x, y) = \mathbf{a}\cdot\mathbf{r}$ where $\mathbf{a} = a\hat{\mathbf{\imath}} + b\hat{\mathbf{\jmath}}$, then it is clear that the height $z = f(x, y)$ above the xy plane in $\mathbb{R}^3$ increases most rapidly in the direction of $\mathbf{a}$, and decreases most rapidly in the direction of $-\mathbf{a}$. The contours at height k are necessarily the lines $ax + by = k$, which are precisely the lines perpendicular to $\mathbf{a}$ (observe that such a line may be parameterized as $\mathbf{r}(t) = t(b\hat{\mathbf{\imath}} - a\hat{\mathbf{\jmath}}) + (k/b)\hat{\mathbf{\jmath}}$, which has velocity orthogonal to $\mathbf{a}$.) Of course, if we allow either $a = 0$ or $b = 0$, we have the case of planes whose levels are either horizontal or vertical lines respectively.

It will often be convenient to write definitions for functions in 3 or more variables using vector notation. For $\mathbb{R}^3$ we use the ordered, right-handed basis $(\hat{\mathbf{\imath}}, \hat{\mathbf{\jmath}}, \hat{\mathbf{k}})$, so a point $(x, y, z) \in \mathbb{R}^3$ corresponds to a position vector $x\hat{\mathbf{\imath}} + y\hat{\mathbf{\jmath}} + z\hat{\mathbf{k}} = \langle x, y, z\rangle$. For $\mathbb{R}^n$ with $n \geq 4$, we use $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_n)$ as the basis. Occasionally, we'll write a vector $\mathbf{r} = x_1\hat{\mathbf{e}}_1 + \dots x_n\hat{\mathbf{e}}_n$ and view it as a vector both in $\mathbb{R}^n$ and in $\mathbb{R}^{n+1}$, where the additional basis element $\hat{\mathbf{e}}_{n+1}$ spans the axis perpendicular to our choice of embedded $\mathbb{R}^n$. This is convenient, e.g., when considering the graph of an n-variable function $f(\mathbf{r})$, the definition of which can now be written

$$\mathcal{G}_f = \{\mathbf{x} \in \mathbb{R}^{n+1} \,|\, \mathbf{x} = \mathbf{r} + f(\mathbf{r})\hat{\mathbf{e}}_{n+1}, \mathbf{r} \in \mathrm{Dom}(f)\}\,.$$

---

§ 0.5. Limits

We review here the definitions of limits and continuity. For examples, see the lecture slides on *Limits and Continuity for Multivariate Functions* from February 13, 2020.

Definition. Given a function of two variables $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^2$ such that D contains points arbitrarily close to a point (a, b), we say that the limit of $f(x, y)$ as (x, y) approaches (a, b) exists and has value L if and only if for every real number $\varepsilon > 0$ there exists a real number $\delta > 0$ such that

$$|f(x, y) - L| < \varepsilon$$

whenever

$$0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta\,.$$

We then write
$$\lim_{(x,y)\to(a,b)} f(x,y) = L\,.$$

Thus, to say that L is the limit of $f(x,y)$ as (x,y) approaches (a,b) we require that for any given positive "error" $\varepsilon > 0$, we can find a bound $\delta > 0$ on the distance of an input (x,y) from (a,b) which ensures that the output falls within the error tolerance around L (that is, $f(x,y)$ is no more than ε away from L).

Another way to understand this is that for any given $\varepsilon > 0$ defining an *open metric neighborhood* $(L-\varepsilon, L+\varepsilon)$ of L on the number line $\mathbb{R}$, we can ensure there is a well defined $\delta(\varepsilon)$ such that the image of any (possibly *punctured*) *open disk* of radius $r < \delta$ centered at (a,b) is contained in the ε-neighborhood.

Recall, for functions of a single variable, one has notions of *left and right one-sided limits*:
$$\lim_{x\to a^-} f(x) \quad \text{and} \quad \lim_{x\to a^+} f(x)\,.$$

But in $\mathbb{R}^2$ there's not merely left and right to worry about; one can approach the point (a,b) along myriad different *paths*! The whole limit $\lim_{(x,y)\to(a,b)} f(x,y) = L$ if and only if the limits along all paths agree and equal L. To write a limit along a path, we can parameterize the path as some vector valued function $\mathbf{r}(t)$ with $\mathbf{r}(1) = \langle a,b\rangle$, and then we can write
$$\lim_{t\to 1^-} f(\mathbf{r}(t)) = L$$
if for any $\varepsilon > 0$, there is a $\delta > 0$ such that $|f(\mathbf{r}(t)) - L| < \varepsilon$ whenever $1-\delta < t < 1$. Similarly we may define a "right" limit along $\mathbf{r}(t)$, $\lim_{t\to 1^+} f(\mathbf{r}(t))$ if $\mathbf{r}(t)$ exists and describes a continuous path for $t > 1$. The two sided limit along the path is then defined in the natural way:
$$\lim_{t\to 1} f(\mathbf{r}(t)) = L \iff \forall\varepsilon > 0\ \exists\delta > 0 :$$
$$|f(\mathbf{r}(t)) - L| < \varepsilon \text{ whenever } 0 < |1-t| < \delta\,.$$

Using paths gives a way to prove *non-existence of a limit*: if the limits along different paths approaching a point (a,b) do not agree, then $\lim_{(x,y)\to(a,b)} f(x,y)$ does not exist.

Definition. A function of two variables $f : D \to \mathbb{R}$ is *continuous at a point* $(x_0,y_0) \in D$ if and only if
$$f(x_0,y_0) = \lim_{(x,y)\to(x_0,y_0)} f(x,y)\,,$$
i.e., the function is defined at (x_0,y_0), its limit exists as (x,y) approaches (x_0,y_0), and the function's value there is equal to the value of the limit.

A function is said to be *continuous throughout its domain*, or simply is called *continuous*, if it is continuous at every point (x_0,y_0) of its domain.

Fact: There is an alternate *topological characterization of continuity*[1]: a function $f : D \to \mathbb{R}$ is continuous throughout D if and only if the pre-image of any open interval $(a,b) = \{t : a < t < b\} \subseteq \mathbb{R}$ is an open subset of the domain. In this context, an open set $E \subset \mathbb{R}^2$ is one for which around every point $p \in E$, there is some open disk centered at p contained fully in E, and an open subset of D is a set which can be made as the intersection of an open set in $\mathbb{R}^2$ and D. For technical reasons, the empty set and the whole of the domain D are considered open subsets of the domain D.

Exercise 0.7. Prove the above fact about continuity and open sets.

[1]Topology studies the properties of geometric objects that remain invariant under continuous maps and continuous deformations, as well as classifications of objects up to equivalences built from continuous constructions. However, one needs a broad notion of continuity to study spaces more general than those in which calculus is performed. Thus, the subject of topology is founded on the notion of *a topology on a set*, which is a formal way of endowing a set with a enough structure to discuss continuity and other properties that make the set into "a space". A topology describes which subsets of the set are called open; open sets must satisfy certain axioms that constitute the defining properties of a topology. Closed sets are then defined in a manner complimentary to open sets. Thus, the concepts of open and closed sets are inherently topological.

Polynomials in two variables are continuous on all of $\mathbb{R}^2$. Recall a polynomial in two variables is a function of the form

$$p(x,y) = \sum_{i=0}^{m}\sum_{j=0}^{n} a_{ij}x^i y^j = a_{00} + a_{10}x + a_{01}y + a_{11}xy + a_{21}x^2y + \ldots + a_{mn}x^m y^n\,.$$

Rational functions are also continuous on their domains. Rational functions of two variables are just quotients of two variable polynomials $R(x,y) = p(x,y)/q(x,y)$. Observe that $\mathrm{Dom}\big(p(x,y)/q(x,y)\big) = \{(x,y)\in\mathbb{R}^2 : q(x,y)\neq 0\}$.

We now graduate to functions of 3 or more variables. For a function $f : D \to \mathbb{R}$ of several variables, regard the input $(x_1, x_2, \ldots, x_n) \in D \subseteq \mathbb{R}^n$ as a vector $\mathbf{r} = \langle x_1, x_2, \ldots, x_n\rangle$.

Definition. Given a function $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^n$, we say that the limit of $f(\mathbf{r})$ as $\mathbf{r}$ approaches $\mathbf{a}$ exists and has value L if and only if for every real number $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|f(\mathbf{r}) - L| < \varepsilon$$

whenever

$$0 < \|\mathbf{r} - \mathbf{a}\| < \delta\,.$$

We then write

$$\lim_{\mathbf{r}\to\mathbf{a}} f(\mathbf{r}) = L\,.$$

Definition. A function of many variables $f : D \to \mathbb{R}$ is continuous at a point $\mathbf{r}_0 \in D \subseteq \mathbb{R}^n$ if and only if

$$f(\mathbf{r}_0) = \lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})\,,$$

i.e., the function is defined at $\mathbf{r}_0$, its limit exists as $\mathbf{r}$ approaches $\mathbf{r}_0$, and the function's value there is equal to the value of the limit.

The function is said to be continuous throughout its domain if it is continuous for every point $\mathbf{r}_0 \in D$.

As before there is a topological reframing of the definition: a function $f : D \to \mathbb{R}$ is continuous throughout its domain if and only if the pre-images of open sets of $\mathbb{R}$ are open subsets of the domain (possibly empty, or all of the domain). The definition of openness involves being able to find an *open ball* around every point.

The open δ-balls appearing in the limit definition are neighborhoods of the approached point, lying in the pre-image of an ε-neighborhood. Thus, we can rephrase the limit definition as follows: $\lim_{\mathbf{r}\to\mathbf{r}_0} f(\mathbf{r})$ exists and equals L if and only if for *any* small open neighborhood $\mathscr{U}$ of L, we can always find a suitable open neighborhood $\mathscr{N}$ of $\mathbf{r}_0$ for which $f(\mathscr{N}) \subseteq \mathscr{U}$. This is a *topological characterization of limits.*

1. Partial Derivatives

§ 1.1. Partial Derivatives of Bivariate Functions

Consider a bivariate function $f : D \to \mathbb{R}$, and assume f is continuous. We will use the geometry of the graph to study how the function changes with respect to changes in the input variables. Let $z = f(x, y)$ be the height of the surface of the graph of f. Consider the planes $x = x_0$ and $y = y_0$, which intersect the graph surface in a pair of curves

$$\mathcal{C}_1 : \mathbf{c}_1(x) = x\hat{\mathbf{\imath}} + y_0\hat{\mathbf{\jmath}} + f(x, y_0)\hat{\mathbf{k}}\,,$$

$$\mathcal{C}_2 : \mathbf{c}_2(y) = x_0\hat{\mathbf{\imath}} + y\hat{\mathbf{\jmath}} + f(x_0, y)\hat{\mathbf{k}}\,.$$

If these curves are regular at the point $P\big(x_0, y_0, f(x_0, y_0)\big)$, then there are lines tangent to each of $\mathcal{C}_1$ and $\mathcal{C}_2$ at P, contained respectively in the planes $y = y_0$ and $x = x_0$. Thus at such a point P on the graph of f, we can assign slopes representing the instantaneous rate of change of the height $z = f(x, y)$ along directions parallel to either the x-axis or the y-axis.

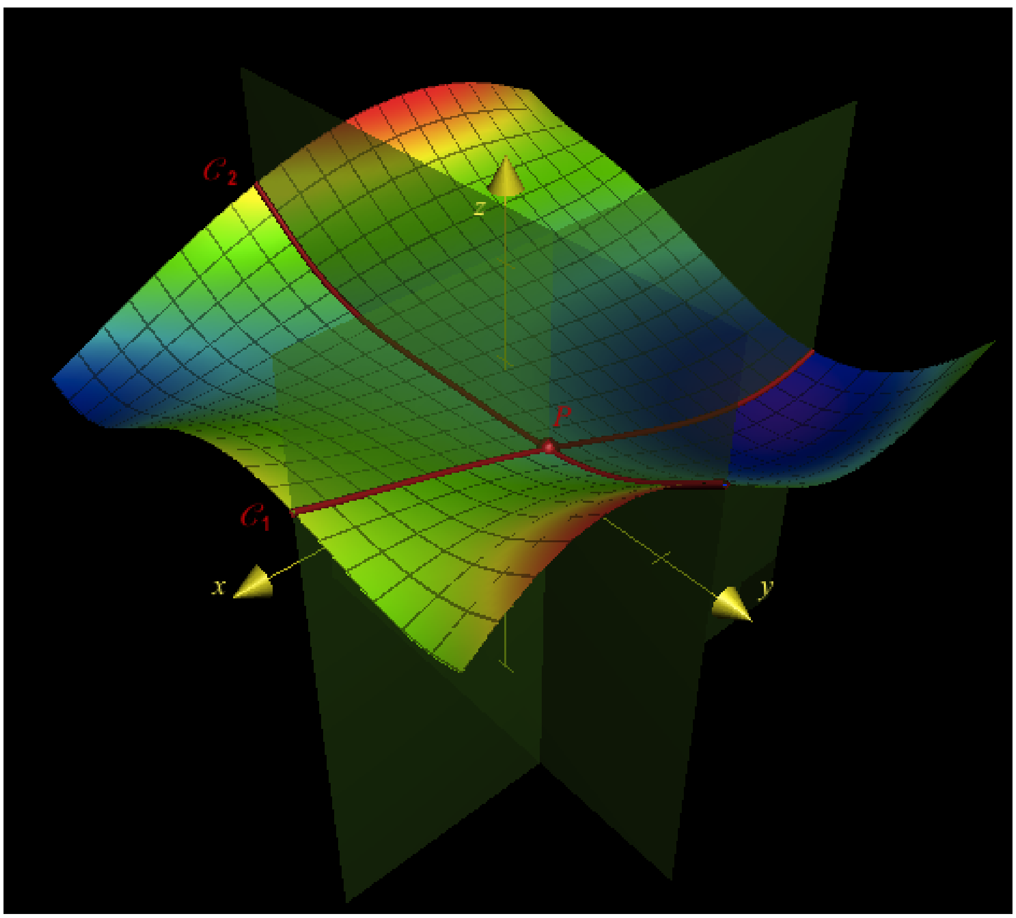


FIGURE 3. Curves $\mathcal{C}_1$ and $\mathcal{C}_2$ on the graph of a function along planes of constant y and x respectively.

Definition. The partial derivative of the two variable function $f(x, y)$ at a point (x_0, y_0) with respect to x, denoted variously

$$\frac{\partial f}{\partial x}\Big|_{(x_0,y_0)} = \partial_x f(x_0, y_0) = f_x(x_0, y_0) = D_x f(x_0, y_0)$$

is the value of the slope of the tangent line to the curve $\mathcal{C}_1$ in the vertical plane $y = y_0$ at the point $P\big(x_0, y_0, f(x_0, y_0)\big)$, which is given by

$$\partial_x f(x_0, y_0) := \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}\,.$$

Similarly one defines

$$\frac{\partial f}{\partial y}\Big|_{(x_0,y_0)} = \partial_y f(x_0, y_0) = f_y(x_0, y_0) = D_y f(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}\,,$$

which is the slope of the tangent line to the curve $\mathcal{C}_2$ in the vertical plane $x = x_0$ at the point $P\big(x_0, y_0, f(x_0, y_0)\big)$.

Definition. The first order partial derivative functions, or simply, first partial derivatives, of $f(x, y)$ are the functions

$$f_x(x, y) = \lim_{h\to 0} \frac{f(x + h, y) - f(x, y)}{h}\,, \quad f_y(x, y) = \lim_{h\to 0} \frac{f(x, y + h) - f(x, y)}{h}\,.$$

It follows straightforwardly from the definitions that to compute the partial derivative functions, one only has to differentiate the function $f(x, y)$ with respect to the chosen variable, while treating the other variable as a constant. Partial derivatives obey the usual derivative rules, such as the power rule, product rule, quotient rule, and chain rule. We'll discuss the chain rule in detail soon. Now, we'll examine how some of the rules interact for partial derivatives, through examples.

Example. Compute the first order partial derivatives $f_x(x, y)$ and $f_y(x, y)$ for the function $f(x, y) = x^3y^2 + x\cos(xy)$.

Solution: When we consider the first term x^3y^2, though it is a product of variables, the partial derivative operator $\frac{\partial}{\partial x}$ sees only a constant times a power of x, so

$$\frac{\partial}{\partial x}\left(x^3y^2\right) = 3x^2y^2\,.$$

For the term $x\cos(xy)$, though the y is treated as a constant, we still employ a power rule and chain rule for x to get

$$\frac{\partial}{\partial x}\left(x\cos(xy)\right) = \cos(xy) - xy\sin(xy)\,.$$

Since the derivative of a sum of functions is still the sum of the derivatives of each function, we obtain

$$\begin{aligned} f_x(x, y) &= \frac{\partial}{\partial x}\big(x^3y^2 + x\cos(xy)\big) \\ &= \frac{\partial}{\partial x}\left(x^3y^2\right) + \frac{\partial}{\partial x}\left(x\cos(xy)\right) \\ &= 3x^2y^2 + \cos(xy) - xy\sin(xy)\,. \end{aligned}$$

For the y-partial one obtains, by similar reasoning

$$f_y(x, y) = 2x^3y - x^2\sin(xy)\,.$$

Exercise 1.1. Find the first order partial derivatives f_x and f_y, for $f(x, y) = \sqrt{x^2 + y^2}$.

Exercise 1.2. Verify the following derivative rules from the limit definitions, assuming the existence of derivatives as necessary:

(i.) For $f(x, y)$ and $g(x, y)$,

$$\partial_x\big(f(x, y) + g(x, y)\big) = \partial_x f(x, y) + \partial_x g(x, y)$$

(ii.) For $f(x, y)$ and $k(y)$,

$$\partial_x\big(k(y)f(x, y)\big) = k(y)\partial_x f(x, y)\,,$$

(iii.) For $f(x, y)$ and $g(x, y)$,

$$\partial_x\big(f(x, y)g(x, y)\big) = \big(\partial_x f(x, y)\big)g(x, y) + f(x, y)\big(\partial_x g(x, y)\big)$$

§ 1.2. Partial Derivatives for functions of Three or More Variables

For functions of three or more variables, one has analogous definitions. If $f(x,y,z)$ is a function of three variables, we can define three first order partials, one with respect to x, one with respect to y, and another to z:

$$\frac{\partial f}{\partial x} = \lim_{h\to 0} \frac{f(x+h,y,z)-f(x,y,z)}{h}\,,$$
$$\frac{\partial f}{\partial y} = \lim_{h\to 0} \frac{f(x,y+h,z)-f(x,y,z)}{h}\,,$$
$$\frac{\partial f}{\partial z} = \lim_{h\to 0} \frac{f(x,y,z+h)-f(x,y,z)}{h}\,.$$

Let $r = x\hat{\mathbf{\imath}} + y\hat{\mathbf{\jmath}} + z\hat{\mathbf{k}}$. Then one can rewrite the limits in terms of $\mathbf{r}$:

$$\frac{\partial f}{\partial x} = \lim_{h\to 0} \frac{f(\mathbf{r}+h\hat{\mathbf{\imath}})-f(\mathbf{r})}{h}\,,$$
$$\frac{\partial f}{\partial y} = \lim_{h\to 0} \frac{f(\mathbf{r}+h\hat{\mathbf{\jmath}})-f(\mathbf{r})}{h}\,,$$
$$\frac{\partial f}{\partial x} = \lim_{h\to 0} \frac{f(\mathbf{r}+h\hat{\mathbf{k}})-f(\mathbf{r})}{h}\,.$$

We'll later study *directional derivatives*, where we may replace the coordinate direction vectors $\hat{\mathbf{\imath}}$, $\hat{\mathbf{\jmath}}$, or $\hat{\mathbf{k}}$ with any unit vector.

We can of course define partial derivatives for any multivariate function of more than 3 variables by the same principle. Let $\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_n$ be the standard basis of $\mathbb{R}^n$.

Definition. For a function $f(\mathbf{r}) = f(x_1, x_2, \ldots, x_n)$, $\mathbf{r} = x_1\hat{\mathbf{e}}_1 + x_2\hat{\mathbf{e}}_2 + \ldots + x_n\hat{\mathbf{e}}_n$, define the ith partial derivative

$$\frac{\partial f}{\partial x_i}(\mathbf{r}) = \partial_{x_i} f(\mathbf{r}) = f_{x_i}(\mathbf{r}) = D_{x_i} f(\mathbf{r})$$

by the limit

$$\frac{\partial f}{\partial x_i}(\mathbf{r}) := \lim_{h\to 0} \frac{f(\mathbf{r}+h\hat{\mathbf{e}}_i)-f(\mathbf{r})}{h}\,.$$

Such partial derivatives may still be interpreted as having a value at a point given by the rate of change of the value of the function along a coordinate direction, with all other coordinates held constant. In fact, one can still describe a curve in a plane slicing the graph of f, though this graph now lives in $\mathbb{R}^{n+1}$ if f has n variables. Indeed, suppose we wish to compute $f_{x_i}(\mathbf{r}_0)$. If the corresponding position on the graph of f is $\mathbf{r}_0 + f(\mathbf{r}_0)\hat{\mathbf{e}}_{n+1}$ (regarding any $\mathbf{r} = x_1\hat{\mathbf{e}}_1 + x_2\hat{\mathbf{e}}_2 + \ldots + x_n\hat{\mathbf{e}}_n$ as a vector in $\mathbb{R}^{n+1}$), then the plane $\Pi_{i,\mathbf{r}_0}$ is the one containing the position $\mathbf{r}_0 + f(\mathbf{r}_0)\hat{\mathbf{e}}_{n+1}$, and parallel to the vectors $\hat{\mathbf{e}}_i$ and $\hat{\mathbf{e}}_{n+1}$, and the curve is given parametrically (in terms of a parameter h) by

$$\mathcal{C}_{i,\mathbf{r}_0} : \mathbf{c}_{i,\mathbf{r}_0}(h) = \mathbf{r}_0 + h\hat{\mathbf{e}}_i + f(\mathbf{r}_0 + h\hat{\mathbf{e}}_i)\hat{\mathbf{e}}_{n+1}\,.$$

We can then interpret the difference quotient definition of $\partial_{x_i} f(\mathbf{r}_0)$ as giving the slope of the tangent line to the curve $\mathcal{C}_{i,\mathbf{r}_0}$ in the plane $\Pi_{i,\mathbf{r}_0}$ at the point $\mathbf{c}_{i,\mathbf{r}_0}(0) = \mathbf{r}_0 + f(\mathbf{r}_0)\hat{\mathbf{e}}_{n+1} \in \mathcal{G}_f \subset \mathbb{R}^{n+1}$, i.e.,

$$\partial_{x_i} f(\mathbf{r}_0) = \left.\frac{\mathrm{d}\mathbf{c}_{i,\mathbf{r}_0}}{\mathrm{d}h}\right|_{h=0} \cdot \hat{\mathbf{e}}_{n+1}\,.$$

As before, such partial derivative functions may be computed by differentiating with respect to the desired variable while holding all others constant.

Example. Find the first order partial derivatives f_x, f_y, and f_z of $f(x,y,z) = \sqrt{1-x^2-y^2+z^2}$.

Solution:

$$\frac{\partial f}{\partial x} = \frac{-x}{\sqrt{1-x^2-y^2+z^2}}\,,\quad \frac{\partial f}{\partial y} = \frac{-y}{\sqrt{1-x^2-y^2+z^2}}\,,\quad \frac{\partial f}{\partial z} = \frac{+z}{\sqrt{1-x^2-y^2+z^2}}\,.$$

Exercise 1.3. Find the first order partial derivatives f_x, f_y, and f_z for $f(x, y, z) = yz\cos(x - z) - xz\sin(y - z) + xy\sin(z)$.

Exercise 1.4. Find the x and y partial derivatives of $z = \arcsin\left(y/\sqrt{x^2 + y^2}\right)$ by writing $\sin z = y/\sqrt{x^2 + y^2}$ and differentiating *implicitly*. Express the final answers as functions of x and y only.

---

§ 1.3. Higher Derivatives

One can define higher order partial derivatives as well, by repeatedly applying partial differential operators. For a bivariate function $f(x, y)$, one can construct four second partial derivative functions:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \partial_x^2 f = f_{xx}\,,$$

$$\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \partial_x \partial_y f = (f_x)_y = f_{xy}\,,$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \partial_y^2 f = f_{yy}\,,$$

$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \partial_x \partial_y f = (f_y)_x = f_{yx}\,.$$

Example. Compute the second order partial derivatives f_{xx}, f_{xy}, f_{yy}, and f_{yx}, for $f(x, y) = \arctan y/x$.

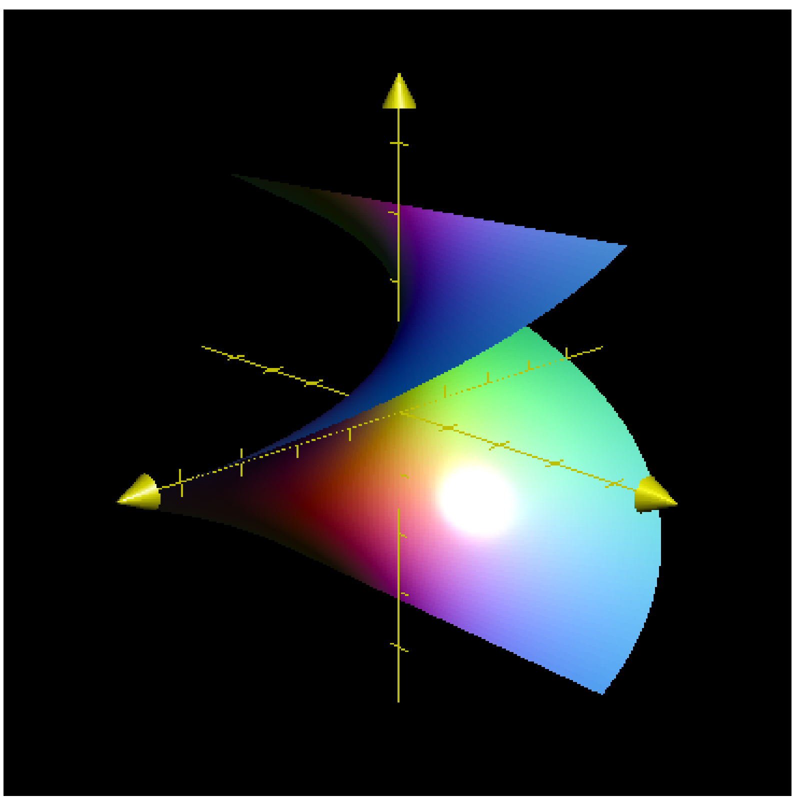


FIGURE 4. The surface $z = \arctan(y/x)$ is a portion of a *helicoid*.

Solution: Observe that the function is undefined along the line $x = 0$. Its graph is the portion of the *helicoid*[2] surface shown in figure 4.

The first partial derivatives are:

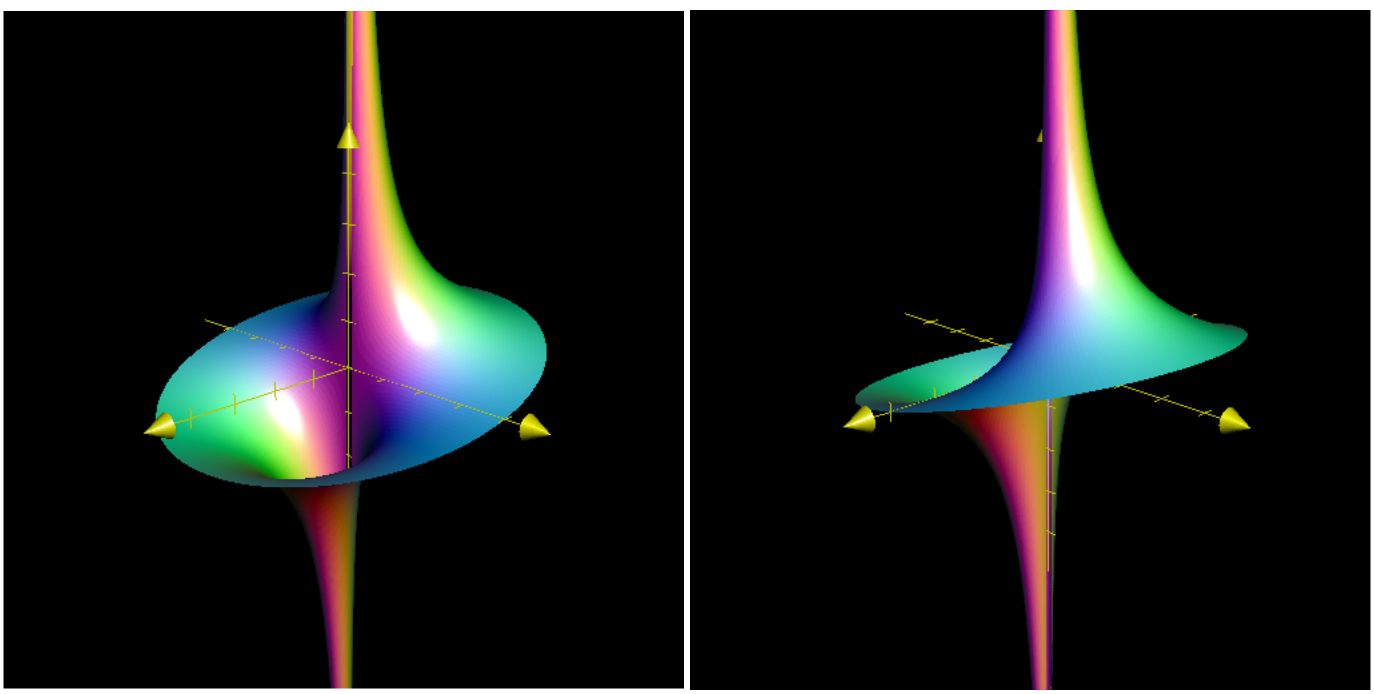
$$f_x(x,y) = \frac{-y}{x^2}\frac{1}{1+y^2/x^2} = \frac{-y}{x^2+y^2}\,,\ x \neq 0\,,$$
$$f_y(x,y) = \frac{1}{x}\frac{1}{1+y^2/x^2} = \frac{x}{x^2+y^2}\,,\ x \neq 0\,.$$

To compute the second partial derivatives, we merely differentiate the above functions with respect to either x or y:

$$f_{xx}(x,y) = \partial_x\left(\frac{-y}{x^2+y^2}\right) = \frac{2xy}{(x^2+y^2)^2}\,,\ x \neq 0\,,$$
$$f_{xy}(x,y) = \partial_y\left(\frac{-y}{x^2+y^2}\right) = \frac{-x^2-y^2+2y^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2},\ x \neq 0\,,$$
$$f_{yy}(x,y) = \partial_y\left(\frac{x}{x^2+y^2}\right) = \frac{-2xy}{(x^2+y^2)^2}\,,\ x \neq 0\,,$$
$$f_{yx}(x,y) = \partial_x\left(\frac{x}{x^2+y^2}\right) = \frac{x^2+y^2-2x^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2},\ x \neq 0\,.$$

Observe that $f_{xy}(x,y) = f_{yx}(x,y)$.

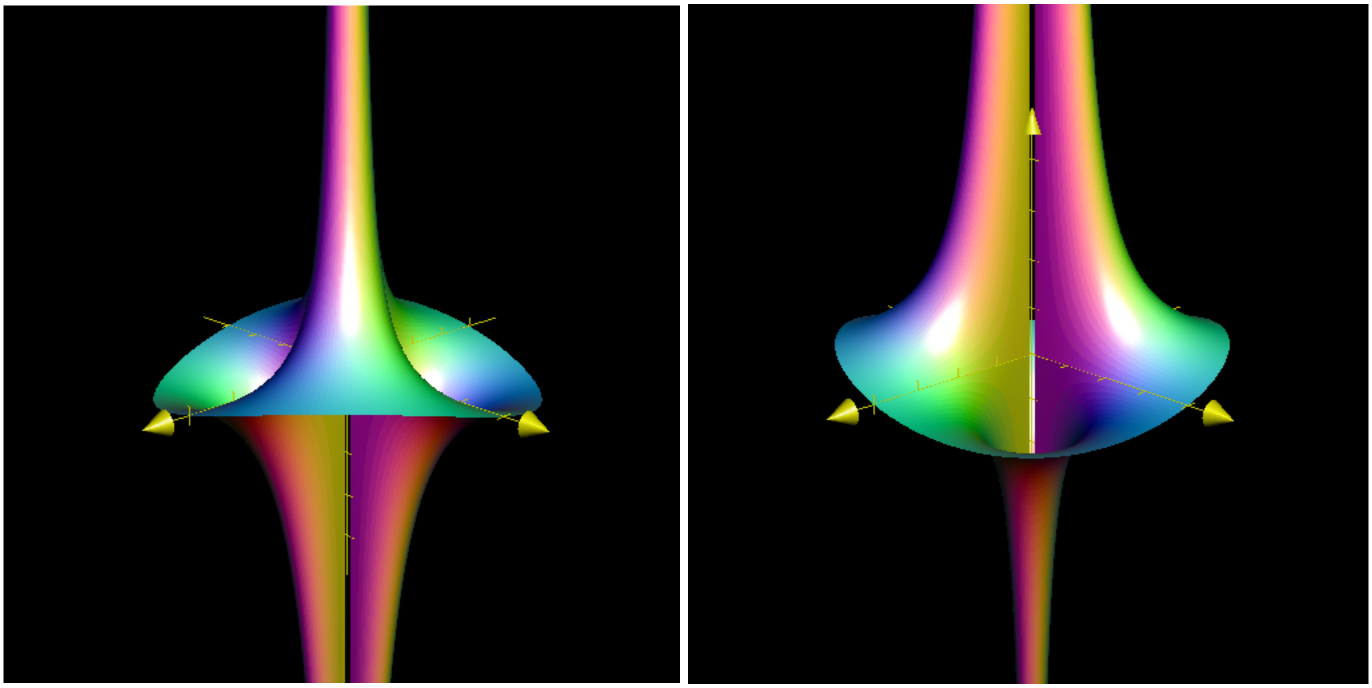
The graphs of the first and second partial derivative functions are shown below in figures 5, 6 and 7.



(a) (b)

FIGURE 5. (A) – The graph of $f_x(x,y) = -y/r^2$ for $f(x,y) = \arctan(y/x)$. (B) – The graph of $f_y(x,y) = x/r^2$ for $f(x,y) = \arctan(y/x)$.

[2] A helicoid is a surface swept out by revolving a line around an axis as you slide it along the axis. Stacking the graphs of functions $z_k = \arctan y/x + k\pi$ for $k \in \mathbb{Z}$, and filling in the z - axis and lines $x = 0, z = k\pi$ gives an entire helicoid. It can also be parameterized as the surface $\boldsymbol{\sigma}(u,v) = \langle u\cos v, u\sin v, v\rangle$, for $u \in \mathbb{R}$ and $v \in \mathbb{R}$.



(a) (b)

FIGURE 6. (A) – The graph of $f_{xx}(x,y)$ for $f(x,y) = \arctan(y/x)$. (B) – The graph of $f_{yy}(x,y)$ for $f(x,y) = \arctan(y/x)$.

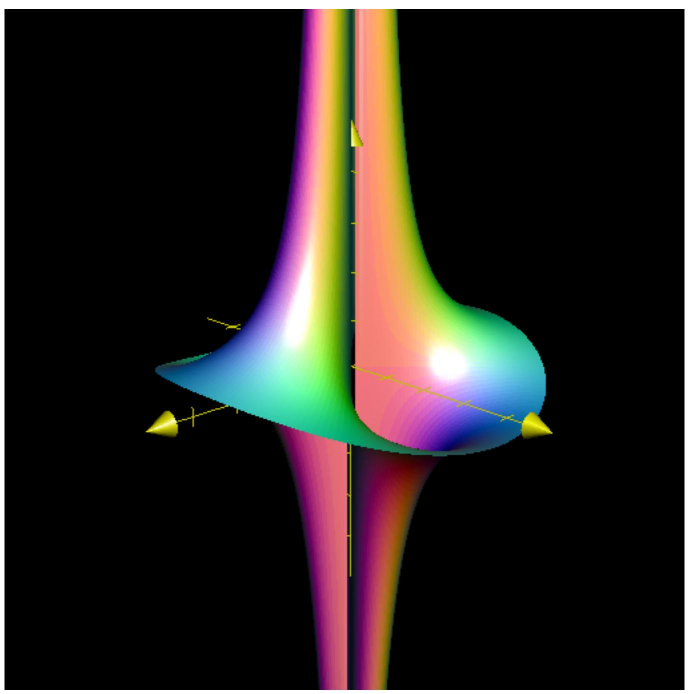


FIGURE 7. The graph of $f_{xy}(x,y) = f_{yx}(x,y)$ for $f(x,y) = \arctan(y/x)$.

Exercise 1.5. Rewrite the first and second partial derivatives of $f(x,y) = \arctan(y/x)$ in polar coordinates, and use the polar expressions to explain the symmetries visible in the above graphs of the partial derivative surfaces.

The equality of the mixed partial derivatives in the preceding example was not pure serendipity: the functions f_{xy} and f_{yx} are rational, and thus continuous on their domains. The following theorem, which has a long history of faulty proofs, states that we can expect such equality under suitable continuity conditions on the mixed partial derivatives:

Theorem (Clairaut-Schwarz Theorem)**.** *If the mixed partial derivative functions f_{xy} and f_{yx} are continuous on a disk D containing the point (x_0, y_0) in its interior, then $f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$.*

Here is an interpretation of the Clairaut-Schwarz theorem: recall that the partial derivative function $f_x(x, y)$ can be interpreted as the result of measuring slopes of tangent lines along curves parallel to the x-axis, cut in the graph by planes of constant y value. Then $f_{xy}(x, y)$ measures how those slopes change as we slide the cutting plane in the $\pm\hat{\mathbf{\jmath}}$ direction (i.e., parallel to the y-axis). The other mixed partial $f_{yx}(x, y)$ measures how the slopes of tangent lines along curves parallel to the y-axis, cut in the graph by planes of constant y value change as we slide the cutting planes in the $\pm\hat{\mathbf{\imath}}$ direction. Clairaut-Schwarz then says these must be equal at a point (x, y) if, at and around that point, both rates of change are well defined and continuous. One way to prove it is to consider a tiny square with sides parallel to the coordinate axes, and look at how the function changes along the edges of the square. One can form difference quotients whose limits as the square shrinks give second partial derivatives. Apply the mean value theorem and carefully examine the limits as the square shrinks!

Exercise 1.6. Compute the partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y\,\partial x}$, $\frac{\partial^2 f}{\partial y^2}$, and $\frac{\partial^2 f}{\partial x\,\partial y}$ for the following functions:

(a) $f(x, y) = \ln\sqrt{x^2 + xy + y^2}$

(b) $f(x, y) = e^{x\cos y}\sin(xy)$

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

(d) $f(x, y) = \int_{x^2+y^2}^{xy} e^{-t^2}\,\mathrm{d}t$

In each case, verify the equality of the mixed partials for the domains where the mixed partials are continuous.

Exercise 1.7. Let $f(x, y) = \ln\sqrt{\dfrac{1 - xy}{1 + xy}}$.

(a) Describe the natural domain of $f(x, y)$ as subset of $\mathbb{R}^2$, and sketch it.
(b) Describe level curves of $f(x, y)$ algebraically, and include a sketch of them.
(c) Compute f_x and f_y. Hint: use properties of logarithms to simplify before differentiating.
(d) Show that $f_{xy} = f_{yx}$ throughout the domain of f.

Exercise 1.8. Consider the function

$$f(x, y) = \lim_{(u,v)\to(x,y)} \frac{uv(v^2 - u^2)}{u^2 + v^2} = \begin{cases} \frac{xy(y^2-x^2)}{x^2+y^2} & \text{if } (x, y) \neq (0, 0)\,, \\ 0 & \text{if } (x, y) = (0, 0)\,. \end{cases}$$

Carefully compute all the second order partial derivatives f_{xx}, f_{xy}, f_{yy}, and f_{yx}. Show that they exist at $(x, y) = (0, 0)$, but are discontinuous, and show moreover that the mixed second partial derivatives at $(0, 0)$ are not equal.

Exercise 1.9. Compute the second order partial derivatives f_{xx}, f_{xy}, f_{xz}, f_{yy}, f_{yx}, f_{yz}, f_{zz}, f_{zx}, and f_{zy}, for $f(x, y) = \frac{xyz}{x^2+y^2+z^2}$. Make note of which pairs of mixed partials are equal.

§ 1.4. Partial Differential Equations

A partial differential equation is an equation giving a relationship between certain partial derivatives of a function. We briefly introduce some famous examples.

Laplace's equation in two variables is the *second order* partial differential equation

$$u_{xx} + u_{yy} = 0\,.$$

A function $u(x,y)$ which satisfies Laplace's equation is called *harmonic*. Laplace's equation is easily generalized to higher dimensions. There are analogues of Laplace's equation even on many complicated geometric spaces, such as *Riemannian manifolds*[3]. The three variable version of Laplace's equation is

$$u_{xx}(x,y,z) + u_{yy}(x,y,z) + u_{zz}(x,y,z) = \underbrace{\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) u(x,y,z)}_{\nabla^2 u(x,y,z)} = 0\,.$$

The expression $\nabla^2 u(x,y,z) = u_{xx}(x,y,z) + u_{yy}(x,y,z) + u_{zz}(x,y,z)$ is called the *Laplacian* of u; the Laplacian operator $\nabla^2 := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ appears in many partial differential equations. We may think of harmonic functions as those in the *kernel* of the Laplacian, i.e., as the functions on which the Laplacian operator vanishes.

Example. Show that $u(x,y) = e^{-x}\cos y$ is harmonic.

Solution: We merely compute the partial derivatives and check that u satisfies Laplace's equation:

$$u_x(x,y) = -e^{-x}\sin y\,, \quad u_{xx}(x,y) = e^{-x}\sin y\,,$$
$$u_y(x,y) = -e^{-x}\cos y\,, \quad u_{yy}(x,y) = -e^{-x}\sin y\,,$$
$$u_{xx} + u_{yy} = e^{-x}\sin y - e^{-x}\sin y = 0\,.$$

Thus $u(x,y) = e^{-x}\cos y$ is harmonic.

Exercise 1.10. Let $v(x,y) = (x^2 - y^2)e^{-y}\cos x - 2xye^{-y}\sin(x)$. Is v harmonic?

The *wave equation* furnishes another example of an important partial differential equation in the physical sciences. A function $u(x,t)$ satisfies the wave equation in one spatial dimension and one time variable t if

$$u_{tt} = a^2 u_{xx}\,,$$

where a is a positive constant representing the *speed of propagation* of the wave. This is called the 1+1 dimensional wave equation. The 3+1 dimensional wave equation (for a scalar wave propagating in $\mathbb{R}^3$ as time advances) can be expressed using the Laplacian:

$$\frac{\partial^2 u}{\partial t^2}(\mathbf{r},t) = a^2\nabla^2 u(\mathbf{r},t)\,,$$

where $\mathbf{r} = x\hat{\mathbf{\imath}} + y\hat{\mathbf{\jmath}} + z\hat{\mathbf{k}}$ is the spatial position and t is the time variable. One can also define a vector valued version of the wave equation, as is needed to study electromagnetic waves. To do so, one needs a *vector Laplacian operator*; we leave this digression for our future study of the calculus of vector fields.

Exercise 1.11. Let a be a positive constant. Show that $u(x,t) = \cos(x - at)$ satisfies the $(1+1)$D wave equation $u_{tt} = a^2 u_{xx}$.

[3] A Riemannian manifold is a space that looks locally like Euclidean space, together with a notion of something like a dot product. The spaces may be globally quite complex, requiring many patches that glue together, with nice conditions on how they overlap. Riemannian geometry gives a natural context in which to study intrinsic geometry, such as distances, curvature, and variational problems, on spaces that may be globally topologically unlike $\mathbb{R}^n$, except that they locally have the right structure to perform calculus.

Another famous partial differential equation is the *heat equation*, also called the *diffusion equation*. In one spatial variable x and one time variable t, the equation reads

$$u_t = \alpha u_{xx}\,,$$

where $\alpha > 0$ is the *thermal diffusivity* or simply the *diffusivity*. A solution function u is either a temperature function, or represents a concentration as a function of space and time, subject to a diffusion process. there are more elaborate heat and diffusion equations, and as with Laplace's equation, one can generalize them to higher dimensions and other spaces. For example, we may use the 3D Laplacian operator to write a heat/diffusion equation in three space variables and one time variable:

$$\frac{\partial}{\partial t}u(x,yz) = \alpha\nabla^2 u(x,y,z) = \alpha\left(u_{xx}(x,y,z) + u_{yy}(x,y,z) + u_{zz}(x,y,z)\right)\,.$$

Exercise 1.12. Let $u(x,t) = e^{-t/2}\sin x + \dfrac{e^{-x^2/2t}}{\sqrt{2t}}$.

Does the function $u(x,t)$ satisfy the heat equation for some constant $\alpha > 0$?

§ 1.5. The Chain Rule

Of the derivative rules, three are essential: linearity, the Leibniz rule for products, and the chain rule. We discuss now the chain rule, which concerns the relation between the derivative of a composite function, and the derivatives of the constituent functions building the composition.

For multivariable functions, we have several possibilities to consider. In the very simplest case, a differentiable single variable function f is composed with a bivariate function $g(x,y)$, and we can ask about the x and y partial derivatives of $f\big(g(x,y)\big)$. It should come as little surprise that in this case

$$\frac{\partial f}{\partial x}\bigg|_{g(x,y)} = \frac{\mathrm{d}f}{\mathrm{d}g}\bigg|_{g(x,y)}\frac{\partial g}{\partial x}\bigg|_{(x,y)} = f'\big(g(x,y)\big)\,g_x(x,y)\,.$$

This is of course something one encounters even when computing partial derivatives of simple examples, such as for a function like $f(x,y) = \sin(xy)$.

Exercise 1.13. Realize $\sin(xy)$ in the form $g\big(h(x,y)\big)$ for some g and h, and compute the first and second partial derivatives f_x, f_y, f_{xx}, $f_{xy} = f_{yx}$, and f_{yy}, writing your solutions so as to make the chain rule explicitly clear.

One can of course write down such a chain rule in any number of variables:

Proposition. *Let $f : E \to \mathbb{R}$ be a real differentiable function of one variable on a domain $E \subset \mathbb{R}$, and suppose $g(x_1,\dots,x_n) = g(\mathbf{r})$ is a function of $n \geq 2$ variables such that the first partials g_{x_i} exist and are continuous on $D \subseteq \mathbb{R}^n$, with image $g(D) \subseteq I$. Then the partials $\frac{\partial}{\partial x_i}f\big(g(\mathbf{r})\big)$ exist for all $\mathbf{r} \in D$ and are given by*

$$\frac{\partial f}{\partial x_i}\bigg|_{g(\mathbf{r})} = \frac{\mathrm{d}f}{\mathrm{d}g}\bigg|_{g(\mathbf{r})}\frac{\partial g}{\partial x_i}\bigg|_{\mathbf{r}} = f'\big(g(\mathbf{r})\big)\,g_{x_i}(\mathbf{r})\,.$$

In the next simplest scenario, a set of variables $x_1,\dots,x_n$ are determined as functions of a single parameter t, and then input into a function of multiple variables. This corresponds geometrically to asking about the change in the value of the multivariate function along a parameterized curve. We describe first the bivariate case:

Proposition. *Let $x(t)$ and $y(t)$ be differentiable functions of $t \in E \subset \mathbb{R}$, such that the images $(x(t),y(t))$ are contained in the domain $D \subseteq \mathbb{R}^2$ of a bivariate function $f : D \to \mathbb{R}$. Suppose further that the partials f_x and f_y exist and are continuous along the image curve. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t}f\big(x(t),y(t)\big) = f_x(x(t),y(t))\dot{x}(t) + f_y(x(t),y(t))\dot{y}(t)\,,$$

where a dot above x or y indicates the usual derivative with respect to t. Thinking of $z = f(x,y)$ as the height of the graph of the function above the xy plane in $\mathbb{R}^3$, one can write

$$\dot{z}(t) = \frac{\mathrm{d}}{\mathrm{d}t} f\big(x(t), y(t)\big) = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}\,,$$

where it is understood that $\partial_x f$ and $\partial_y f$ are evaluated at $\big(x(t), y(t)\big)$, and $\frac{\mathrm{d}x}{\mathrm{d}t} = \dot{x}$ and $\frac{\mathrm{d}y}{\mathrm{d}t} = \dot{y}$ are likewise being evaluated at t.

More generally:

Proposition. *If $\mathbf{r}(t) = \langle x_1(t), \dots x_n(t)\rangle$ is a differentiable curve with image contained in the domain D of a function $f(x_1, \dots, x_n)$, and all the first partial derivatives of f exist and are continuous along $\mathbf{r}(t)$, then*

$$\frac{\mathrm{d}}{\mathrm{d}t} f\big(\mathbf{r}(t)\big) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}\frac{\mathrm{d}x_k}{\mathrm{d}t}\Big|_{\mathbf{r}(t)} = \dot{x}_1 \partial_{x_1} f\big(\mathbf{r}(t)\big) + \ldots + \dot{x}_n \partial_{x_n} f\big(\mathbf{r}(t)\big)\,.$$

Example. Let $f(x,y) = x^2 + 4y^2$, and let $\mathbf{r}(t) = \cos(t)\hat{\imath} + \sin(t)\hat{\jmath}$ be the unit circle in the xy-plane. Find the derivative $z'(t)$ for $z = f(x,y)$, and interpret this derivative geometrically.

Solution: There are two routes of solution. One is to substitute the parametric equations of the curve (namely, the component functions $x(t) = \cos t$ and $y(t) = \sin t$) into $f(x,y)$, thus reducing the problem to a straightforward derivative from a first course in differential calculus. The other option is to employ the chain rule. We'll show both methods, beginning with the chain rule.

According to the proposition above, the derivative $\dot{z}(t)$ is given by

$$\begin{aligned}\dot{z}(t) &= \frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} \\ &= 2x(t)\dot{x}(t) + 8y(t)\dot{y}(t) \\ &= -2\cos t \sin t + 8 \sin t \cos t = 6 \sin t \cos t \\ &= 3 \sin 2t\,.\end{aligned}$$

Alternatively, we compute $z(t) = f(\cos t, \sin t) = \cos^2 t + 4\sin^2 t$, and

$$\dot{z}(t) = -2\cos t \sin t + 8 \sin t \cos t = 3 \sin 2t\,,$$

as before.

Geometrically, the image in $\mathbb{R}^3$ of $f(\mathbf{r}(t))$ on the graph is a loop on the elliptic paraboloid $z = x^2 + 4y^2$, and $\dot{z}$ is the rate of change of the height along this loop as the parameter t advances, see figure 8. Thinking of t as describing a particle, we can think of $\dot{z}$ as the vertical component of its velocity. The chain rule then tells us that this is computable as a *dot product*:

$$\dot{z}(t) = \langle f_x\big(\mathbf{r}(t)\big), f_y\big(\mathbf{r}(t)\big)\rangle \cdot \dot{r}(t)\,.$$

The vector $\Big\langle f_x\big(\mathbf{r}(t)\big), f_y\big(\mathbf{r}(t)\big)\Big\rangle$ is an example of what we will call a *gradient vector* (in this case, it's evaluated along the curve). We'll discuss gradients in greater depth in §3.2 of these notes.

Next we consider the case when we replace the variables of a bivariate function with bivariate functions, an application of which will be the study of derivatives of bivariate functions after a coordinate transformation.

Let u and v be variables, and in the uv-plane, let E be some domain such that we can define functions $g(u,v)$ and $h(u,v)$ whose first partials all exist and are all continuous throughout E. Then there is a multivariate transformation from E to the region $R = \big(x(E), y(E)\big) = \{(x,y) = \big(g(u,v), h(u,v)\big)\,|\,(u,v) \in E\} \subset \mathbb{R}^2$, defined by setting $x(u,v) = g(u,v)$ and $y(u,v) = h(u,v)$. Let D be the domain of a function $f(x,y)$ such that $R \subseteq D$, and suppose the first partials of f exist and are continuous throughout R. Then we have chain rules specifying the u and v partial derivatives of $f\big(x(u,v), y(u,v)\big)$:

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}\,,$$

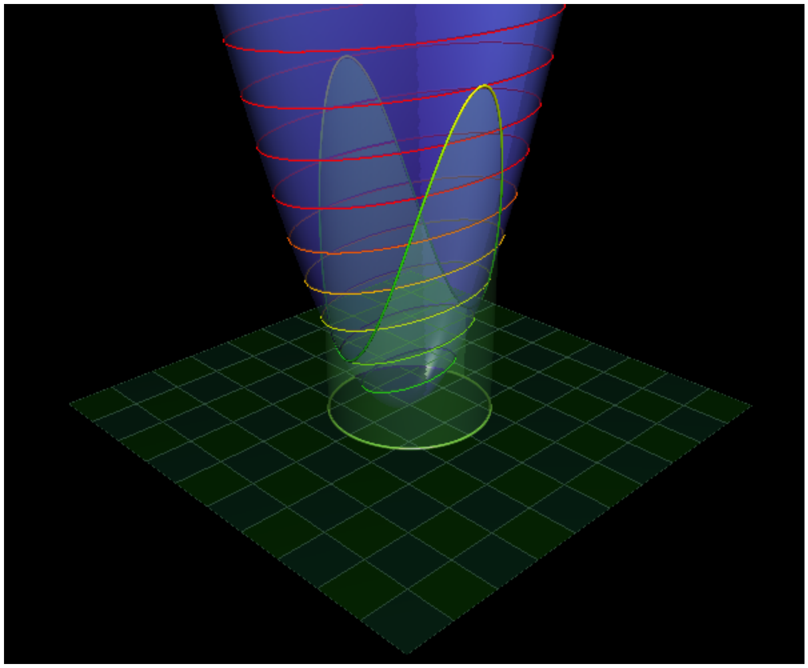


FIGURE 8. The curve of intersection of the unit right circular cylinder $x^2 + y^2 = 1$ and the elliptic paraboloid $z = x^2 + 4y^2$ is the curve $\mathbf{c}(t) = \cos t\hat{\imath} + \sin t\hat{\jmath} + f(\cos t, \sin t)\hat{\mathbf{k}} = \cos t\hat{\imath} + \sin t\hat{\jmath} + (1 + 3\sin^2 t)\hat{\mathbf{k}}$; the rate at which z changes around the loop is $\dot{z} = \dot{x}\,\partial_x z + \dot{y}\,\partial_y z = 3\sin 2t$ according to the chain rule.

$$\frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}\,.$$

A simple application is computing partial derivatives of a Cartesian bivariate function with respect to polar coordinate variables. Recall, the Cartesian variables (x, y) can be expressed as functions of the polar variables r and θ, via elementary trigonometry:

$$x = r\cos\theta\,,$$

$$y = r\sin\theta\,.$$

Example. Let $f(x, y) = 3x^2 - 2y^2$. Compute $\partial_r f$ and $\partial_\theta f$, and express the resulting functions in both in terms of polar variables and in terms of Cartesian variables.

Solution: At each step, we will express things using both coordinate systems, so that we can express the final answers in either coordinate system. First, observe that we have the four first partial derivatives of x and y with respect to r and θ:

$$\frac{\partial x}{\partial r} = \cos\theta = \frac{x}{r} = \frac{x}{\sqrt{x^2+y^2}}, \quad \frac{\partial x}{\partial \theta} = -r\sin\theta = -y$$

$$\frac{\partial y}{\partial r} = \sin\theta = \frac{y}{r} = \frac{y}{\sqrt{x^2+y^2}}\,, \quad \frac{\partial y}{\partial \theta} = r\cos\theta = x$$

Next, we compute the x and y partials of $f(x, y)$:

$$\frac{\partial f}{\partial x} = 6x = 6r\cos\theta\,,$$

$$\frac{\partial f}{\partial y} = -4y = -4r\sin\theta\,.$$

Putting these pieces together gives

$$\begin{aligned}\frac{\partial f}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = (6x)\left(\frac{x}{\sqrt{x^2+y^2}}\right) + (-4y)\left(\frac{y}{\sqrt{x^2+y^2}}\right)\\ &= \frac{6x^2-4y^2}{\sqrt{x^2+y^2}} = 6r\cos^2\theta - 4r\sin^2\theta\,,\\ \frac{\partial f}{\partial \theta} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = (6x)(-y) + (-4y)(x)\\ &= -10xy = -10r^2\cos\theta\sin\theta = -5r^2\sin(2\theta)\,.\end{aligned}$$

Observe that if we rewrite f in terms of polar coordinates, we have

$$f(r,\theta) = 3r^2\cos^2\theta - 2r^2\sin^2\theta\,,$$

from which we can directly compute $\partial_r f$ and $\partial_\theta f$ without the chain rule.

Remark. The computations of $\partial_u f$ and $\partial_v f$ from $\partial_x f$, $\partial_y f$, $\partial_u x$, $\partial_u y$, $\partial_v x$ and $\partial_v y$ may be neatly encoded by a matrix vector product. Observe that in general

$$\begin{bmatrix} \partial_u f \\ \partial_v f \end{bmatrix} = \begin{bmatrix} \partial_u x & \partial_u y \\ \partial_v x & \partial_v y \end{bmatrix} \begin{bmatrix} \partial_x f \\ \partial_y f \end{bmatrix}.$$

Note that the column vector that gets transformed by the matrix vector product is none other than the column vector form of the gradient of f introduced above. The *transpose* of the square matrix above,

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{bmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{bmatrix},$$

is called the *Jacobian matrix* of the transformation. The Jacobian evaluated at a point gives a matrix representing the linear map best approximating the coordinate transformation in a neighborhood of that point. The Jacobian determinant of a transformation is important in the theory of change of variables for multiple integrals. In a sense, the gradient, as a row vector, is also a Jacobian. We will refer to the notion of the derivative object giving the best linear approximation of a map as the *Jacobian derivative of the map.*

The appeal of the matrix expression of the chain rule is deeper than the mere convenience of the notation. When our multivariable chain rule arises from a change of coordinates, we can express things neatly in the language of Jacobians. Let $\mathbf{D}_{x,y}f = \begin{bmatrix} \partial_x f & \partial_y f \end{bmatrix}$ be the Jacobian derivative of f with respect to (x,y)-coordinates, and let $\mathbf{D}_{u,v}f = \begin{bmatrix} \partial_u f & \partial_v f \end{bmatrix}$ be the Jacobian derivative of f with respect to (u,v)-coordinates. Denote by $\mathbf{G}$ the transformation $(u,v)\mapsto(x,y)$, and $\mathbf{D}_{u,v}\mathbf{G} = \frac{\partial(x,y)}{\partial(u,v)}$. Then, using the transposes, the chain rule becomes

$$\mathbf{D}_{u,v}\big(f\circ\mathbf{G}\big)(u,v) = \mathbf{D}_{x,y}f\big(\mathbf{G}(u,v)\big)\circ\mathbf{D}_{u,v}\mathbf{G}(u,v)\,,$$

where "$\circ$" on the right hand side can be interpreted as composition of linear maps, which is just matrix multiplication (in this case, the row vector given by the (x,y)-gradient of f, evaluated at $\mathbf{G}(u,v) = \langle x(u,v), y(u,v)\rangle$, acts on the Jacobian matrix of the transformation $\mathbf{D}_{u,v}\mathbf{G} = \partial(x,y)/\partial(u,v)$, again evaluated for the ordered pair (u,v)). This allows us to rephrase the chain rule: *the Jacobian derivative of a composition of differentiable functions is the composition of their Jacobian derivative maps.*

We will now state a general chain rule for functions of many variables

Proposition. *Suppose $f(x_1,\dots x_n)$ has continuous partial derivatives $\partial_{x_i}f$ on a domain $D\subseteq\mathbb{R}^n$, and the variables $x_1,\dots,x_n$ are given as multivariate functions of variables $u_1,\dots,u_m$. If the partial derivatives $\frac{\partial x_i}{\partial u_j}$ exist and are continuous then*

$$\frac{\partial f}{\partial u_j} = \sum_{i=1}^n \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial u_j} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial u_j} + \ldots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial u_j}\,.$$

Exercise 1.14. Appropriately rephrase the above general chain rule in terms of Jacobian derivatives, in keeping with the philosophy that the chain rule should be expressible as "the Jacobian derivative of a composition of differentiable functions is the composition of their Jacobian derivative maps." In particular you should define Jacobians for maps involved, and write out what the matrix products look like in the general case. Be sure to see that their dimensions are compatible!

Example. Let $w = f(x, y, z, t)$ and suppose f_x, f_y, f_z and f_w all exist and are continuous on a set $E \subset \mathrm{Dom}(f) \subseteq \mathbb{R}^4$. Suppose further that x, y, z, and t are each functions of variables u and v defined on a set $U \subset \mathbb{R}^2$ such that the image of U is in E, and all necessary first partials exist and are continuous. Then

$$\frac{\partial w}{\partial u} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial u} + \frac{\partial w}{\partial t}\frac{\partial t}{\partial u} = w_x x_u + w_y y_u + w_z z_u + w_t t_u\,,$$

$$\frac{\partial w}{\partial v} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial v} + \frac{\partial w}{\partial t}\frac{\partial t}{\partial u} = w_x x_v + w_y y_v + w_z z_v + w_t t_v\,.$$

Example. It can be helpful to form a tree diagram to understand the nesting of variables. A tree in graph theory is a collection of vertices and edges connecting them, with no closed loops. For our variable trees, the vertices are labeled by the variables, and edges are labeled by partial derivatives. E.g., for the example above, one has

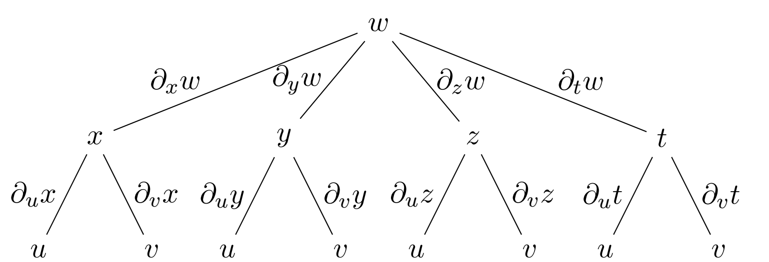


Figure 9

The terms of the chain rule sums are then found by taking products of the edge labels along paths from the root w to the ends of branches with leaves labeled by the appropriate variable. Thus, for $\partial_u w$, one follows all paths originating from w and ending in u, to collect the product terms which sum to give us the chain $\partial_x w\, \partial_u x + \partial_y w\, \partial_u y + \partial_z w\, \partial_u z + \partial_t w\, \partial_u t$.

If one has nested several levels of multivariable functions, then the tree may have more levels. For example, see the tree diagram for $f\Big(x\big(u(r,s,t), v(r,s,t)\big), y\big(u(r,s,t), v(r,s,t)\big)\Big)$ shown in figure 10.

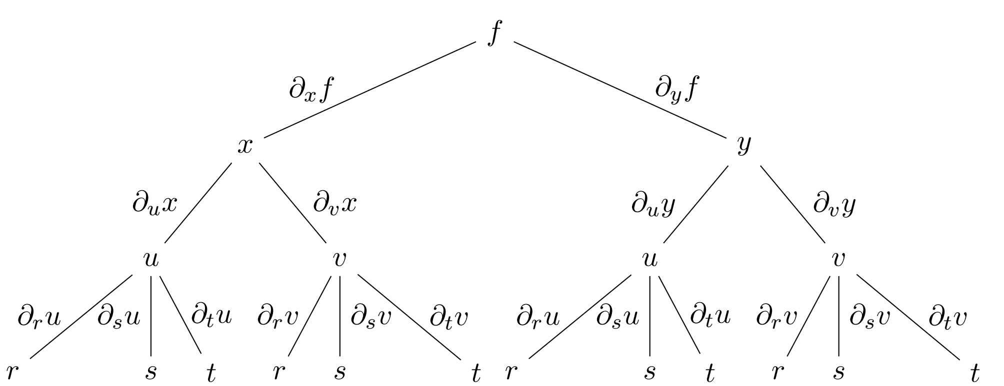


Figure 10

Exercise 1.15. Using the tree in figure 10, write out the chain rule expression for $\partial_s f$.

Example. Let $f(x,y) = xy - x^2 - y^2$, and $x(u,v,w) = ue^v \cos w$, $y(u,v,w) = e^{uv}\sin w$. Find f_w when $u = 1$, $v = -2$, and $w = \pi$. We can solve this problem via a tree as follows. The initial variable tree is shown in figure 11. In red are the branches we need to follow to form the appropriate chain rule.

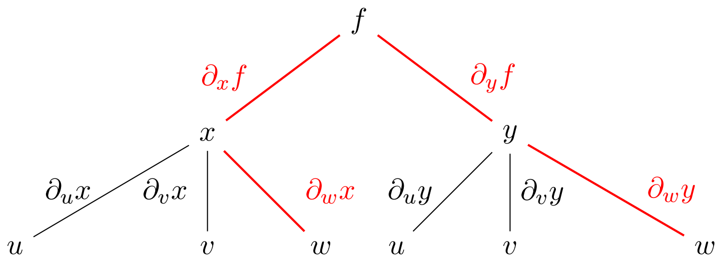


FIGURE 11. The chain rule gives $\partial_w f = \partial_x f\,\partial_w x + \partial_y f\,\partial_w y$.

We thus need to compute $\partial_x f$, $\partial_y f$, $\partial_w x$, $\partial_w y$, $x(1,-2,\pi)$, and $y(1,-2,\pi)$ in order to compute $\partial_w f = \partial_x f\,\partial_w x + \partial_y f\,\partial_w y$. We will redraw the tree, filling in information. It is helpful to alter the left-to-right order in which the leaves appear to make space to write out $\partial_w x$.

Rewriting our tree in terms of the functions and computed partials, we have:

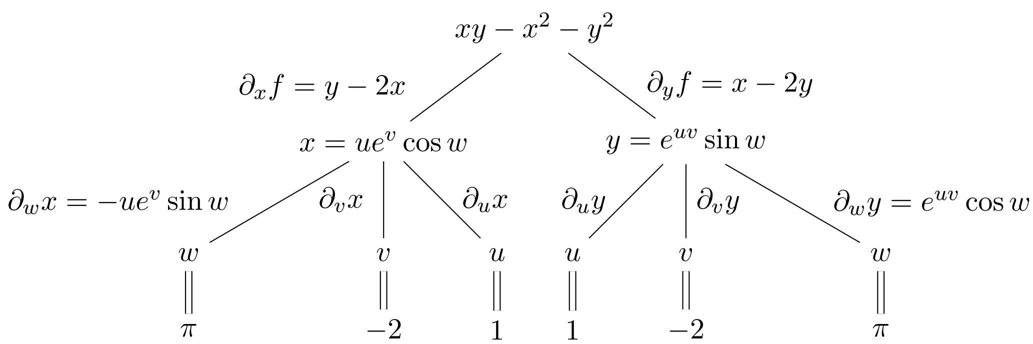


FIGURE 12. Filling in the tree with the necessary partial derivatives.

Evaluating then gives

$$\begin{aligned}
\frac{\partial f}{\partial w}(1,-2,\pi) &= \Big[\partial_x f\,\partial_w x + \partial_y f\,\partial_w y\Big]_{(u,v,w)=(1,-2,\pi)}\\
&= \Big[(y-2x)(-ue^v\sin w) + (x-2y)(e^{uv}\cos w)\Big]_{(u,v,w)=(1,-2,\pi)}\\
&= (0 - 2e^{-2})(0) + (-e^{-2} - 2(0))(-e^{-2})\\
&= \frac{1}{e^4}\,.
\end{aligned}$$

Consider now a bivariate function f whose variables x and y are given as bivariate functions of u and v, and assume all first and second partials exist and are continuous on appropriate domains. Then we can use the chain rule and product rules to compute expressions for the second partials

f_{uu}, f_{uv}, f_{vu} and f_{vv}. For example, to compute f_{uu}, one has

$$\begin{aligned}
\frac{\partial^2 f}{\partial u^2} &= \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial u}\right) = \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}\right) \\
&= \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial x}\right)\frac{\partial x}{\partial u} + \frac{\partial f}{\partial x}\frac{\partial^2 x}{\partial u^2} + \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial y}\right)\frac{\partial y}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial^2 y}{\partial u^2} \\
&= \frac{\partial^2 f}{\partial x^2}\left(\frac{\partial x}{\partial u}\right)^2 + \frac{\partial f}{\partial x}\frac{\partial^2 x}{\partial u^2} + 2\frac{\partial^2 f}{\partial y\,\partial x}\frac{\partial x}{\partial u}\frac{\partial y}{\partial u} + \frac{\partial^2 f}{\partial y^2}\left(\frac{\partial y}{\partial u}\right)^2 + \frac{\partial f}{\partial y}\frac{\partial^2 y}{\partial u^2}\,,
\end{aligned}$$

where we've applied the product and chain rules to expand it, and the Clairaut-Schwarz theorem to combine the mixed partial terms. All together, we have the following

Proposition. *If $f(x, y)$ has continuous first and second partial derivatives with respect to x and y, and x and y are given as functions of u and v with continuous first and second partial derivatives with respect to u and v, then $f\big(x(u, v), y(u, v)\big)$ has continuous first and second partial derivatives with respect to u and v, and*

$$\begin{aligned}
\frac{\partial^2 f}{\partial u^2} &= \frac{\partial^2 f}{\partial x^2}\left(\frac{\partial x}{\partial u}\right)^2 + \frac{\partial f}{\partial x}\frac{\partial^2 x}{\partial u^2} + 2\frac{\partial^2 f}{\partial y\,\partial x}\frac{\partial x}{\partial u}\frac{\partial y}{\partial u} + \frac{\partial^2 f}{\partial y^2}\left(\frac{\partial y}{\partial u}\right)^2 + \frac{\partial f}{\partial y}\frac{\partial^2 y}{\partial u^2} \\
\frac{\partial^2 f}{\partial u\,\partial v} &= \frac{\partial^2 f}{\partial v\,\partial u} = \frac{\partial^2 f}{\partial x^2}\frac{\partial x}{\partial v}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial x}\frac{\partial^2 x}{\partial v\,\partial u} + 2\frac{\partial^2 f}{\partial y\,\partial x}\frac{\partial y}{\partial v}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial^2 y}{\partial v\,\partial u} + \frac{\partial^2 f}{\partial y^2}\frac{\partial y}{\partial v}\frac{\partial y}{\partial u} \\
\frac{\partial^2 f}{\partial v^2} &= \frac{\partial^2 f}{\partial x^2}\left(\frac{\partial x}{\partial v}\right)^2 + \frac{\partial f}{\partial x}\frac{\partial^2 x}{\partial v^2} + 2\frac{\partial^2 f}{\partial y\,\partial x}\frac{\partial x}{\partial v}\frac{\partial y}{\partial v} + \frac{\partial^2 f}{\partial y^2}\left(\frac{\partial y}{\partial v}\right)^2 + \frac{\partial f}{\partial y}\frac{\partial^2 y}{\partial v^2}\,.
\end{aligned}$$

Exercise 1.16. Apply the chain rule and other applicable principles to get the remaining two formulae in the above proposition for $f_{uv} = f_{vu}$ and f_{vv}. Compute f_{uv} and f_{vu} separately and deduce their equality.

Exercise 1.17. Use the above proposition to re-express Laplace's equation $f_{xx} + f_{yy} = 0$ in polar coordinates. In particular, show that

$$f_{xx} + f_{yy} = f_{rr} + \tfrac{1}{r}f_r + \tfrac{1}{r^2}f_{\theta\theta}\,,$$

so that the polar form of Laplace's equation may be written as

$$\frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} = 0\,.$$

Exercise 1.18. Can you guess a formula for a spherically symmetric scalar wave propagating from a point with velocity a? That is, find any solution of the $(3 + 1)$D wave equation

$$\frac{\partial^2 u}{\partial t^2}(\mathbf{r}, t) = a^2\nabla^2 u(\mathbf{r}, t)$$

modeling a wave spherically symmetric wave (so in particular, u depends on $\varrho = \sqrt{x^2 + y^2 + z^2}$ rather than on x y and z independently). Hint: Building on the previous exercises, find an expression for the Laplacian in *spherical coordinates*. Look also at exercise 1.11 above.

---

§ 1.6. Implicit Differentiation

Recall, for a continuous multivariable function $f(x, y)$, an equation $f(x, y) = k$ for an appropriately chosen constant $k \in \mathbb{R}$ determines a curve. We may view this curve as a level curve of the graph of $z = f(x, y)$, at height $z = k$, or we may consider the equation to define an *implicit curve* in the plane $\mathbb{R}^2$. For example, the familiar unit circle can be expressed as the set of all points $(x, y) \in \mathbb{R}^2$ satisfying the equation $x^2 + y^2 = 1$, and it may also be viewed as the level curve of the paraboloid $= z = x^2 + y^2$ at height $z = 1$.

Implicit differentiation tackles the problem of computing the slope of a tangent line to such an implicit curve, by using the chain rule to compute $\frac{\mathrm{d}y}{\mathrm{d}x}$. One differentiates both sides of the equation $f(x,y)=k$ under the assumption that locally, y is a function of x:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}f\big(x,y(x)\big) &= \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}x}+\frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x}=0\\ \implies \frac{\mathrm{d}y}{\mathrm{d}x} &= -\frac{\partial f/\partial x}{\partial f/\partial y}\,.\end{aligned}$$

E.g., for the unit circle, using $z=x^2+y^2$, one has $y'(x)=-\partial_x z/\partial_y z=-2x/2y=-x/y$, which is of course geometrically sensible, as the slope of a tangent to a circle must be the negative reciprocal of the slope of the radial line, since the tangent line is perpendicular to the radius.

Similarly, for an *implicit surface* given by an equation $F(x,y,z)=k$, we can compute partial derivatives under the assumption that one of the variables, say, z, locally depends upon the other two, with the other two being independent there:

$$\frac{\partial}{\partial x}F\big(x,y,z(x,y)\big)=\frac{\partial F}{\partial x}\frac{\partial x}{\partial x}+\frac{\partial F}{\partial z}\frac{\partial z}{\partial x}=0 \implies \frac{\partial z}{\partial x}=-\frac{\partial F/\partial x}{\partial F/\partial z}\,,$$

and similarly

$$\frac{\partial z}{\partial y}=-\frac{\partial F/\partial y}{\partial F/\partial z}\,.$$

Example. Find $\partial_x z$ and $\partial_y z$ for the implicit surface $xy-xz+yz=1$.

Solution: According to the above discussion, we can differentiate $F(x,y,z)=xy-xz+yz$ to compute these partials. This yields

$$\begin{aligned}\frac{\partial z}{\partial x} &= -\frac{\partial F/\partial x}{\partial F/\partial z}=-\frac{y-z}{y-x}=\frac{y-z}{x-y}\,,\\ \frac{\partial z}{\partial y} &= -\frac{\partial F/\partial y}{\partial F/\partial z}=-\frac{x+z}{y-x}=\frac{x+z}{x-y}\,.\end{aligned}$$

Observe that we treat z as independent of x and y when computing the partials of F. However, the classical implicit differentiation approach is to assume z to be locally a function of x and y, and to compute via the chain rule. E.g., for $\frac{\partial z}{\partial x}$:

$$\begin{aligned}0=\frac{\partial}{\partial x}\big(xy-xz+yz\big) &\implies y-z-x\frac{\partial z}{\partial x}+y\frac{\partial z}{\partial x}=0\\ \implies y-z=(x-y)\frac{\partial z}{\partial x} &\implies \frac{\partial z}{\partial x}=\frac{y-z}{x-y}\,,\end{aligned}$$

Exercise 1.19. For the surface given implicitly by $r^4-(1+2xz)r^2+(xz)^2=0$, where $r^2=x^2+y^2$, use implicit differentiation to compute $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$ and $\frac{\partial r}{\partial z}$.

An important consideration is absent from our discussion above. We assumed that locally, z was a function of x and y, and so we could compute partial derivatives via the chain rule. But when is it okay to assume that an equation $F(x,y,z)=0$ implicitly defines a surface in such a way that z is locally a function of x and y? How do we find the points where this assumption is untenable?

We thus consider the *implicit function theorem*:

Theorem (Implicit Function Theorem for Trivariate Functions)**.** *Let $F(x,y,z)$ be a function such that $F(x_0,y_0,z_0)=0$, and suppose that the partials F_x, F_y and F_z are all continuous on a ball containing (x_0,y_0,z_0), and moreover $F_z(x_0,y_0,z_0)\neq 0$. Then there exists a neighborhood U of (x_0,y_0,z_0) and a function $f:D\to\mathbb{R}$, for some domain $D\subseteq\mathbb{R}^2$, such that*

$$\{(x,y,z)\in U\,|\,F(x,y,z)=0\}=\{(x,y,z)\,|\,z=f(x,y),(x,y)\in D\}\subset U\,,$$

i.e., the equation $F(x,y,z)=0$ in U implicitly defines a surface which is the graph of a bivariate function.

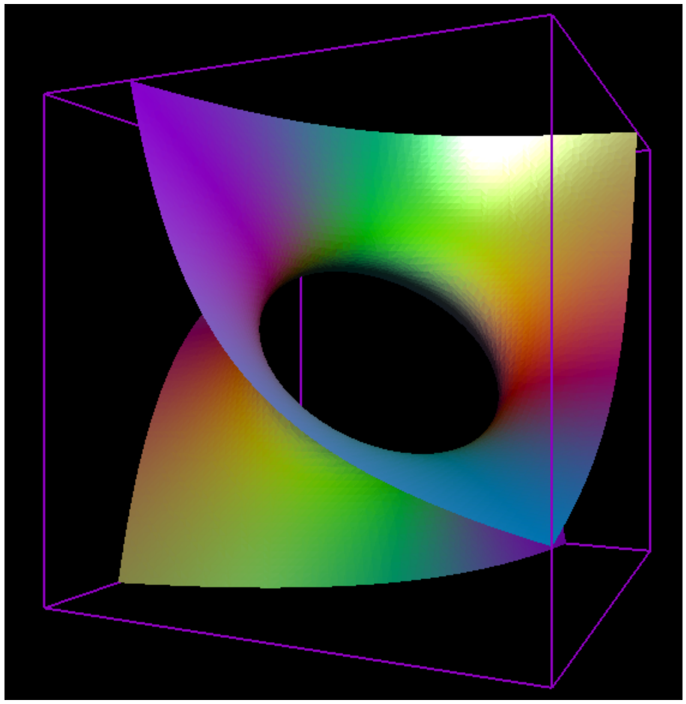


FIGURE 13. The surface defined by the equation $xy - xz + yz = 1$.

Certainly, nothing is special about z: one can ask instead that $F_x \neq 0$, and seek to express x as a function of y and z locally. One can of course make much more general statements, though we will leave such considerations for an advanced calculus course.

Exercise 1.20. Reconsider the function $F(x, y, z) = xy - xz + yz$ and the surface $F(x, y, z) = 1$.

(a) Check that F satisfies the conditions of the implicit function theorem at the point $(2, 1, 1)$, and verify that this point is on the surface defined by $F(x, y, z) = 1$.

(b) What is the local expression for z there? What is the domain D for which this local function is well defined?

(c) What happens to the surface at points where $F_z = 0$? Can you give an implicit description around such points using x or y as the dependent variable?

2. Tangent Planes, Linear Approximation, and differentiability

§ 2.1. The Tangent Plane to a Graph

Exercise 2.1. Consider, as in the previous section, the curves

$$\mathcal{C}_1 : \mathbf{c}_1(x) = x\hat{\mathbf{\imath}} + y_0\hat{\mathbf{\jmath}} + f(x, y_0)\hat{\mathbf{k}}\,,$$

$$\mathcal{C}_2 : \mathbf{c}_2(y) = x_0\hat{\mathbf{\imath}} + y\hat{\mathbf{\jmath}} + f(x_0, y)\hat{\mathbf{k}}\,.$$

lying in planes of constant y and x respectively, intersecting mutually on the surface $z = f(x, y)$ at the point $P(x_0, y_0, z_0)$, where $z_0 = f(x_0, y_0)$. Assume that both $f_x(x, y)$ and $f_y(x, y)$ exist and are continuous for the input (x_0, y_0). Argue that there is a unique plane containing both of the respective tangent lines to $\mathcal{C}_1$ and $\mathcal{C}_2$ at P, and compute its normal vector from the tangent vectors to the parameterized curves. This plane is called the *tangent plane to $z = f(x, y)$ at $P(x_0, y_0, z_0)$*, or simply *the tangent plane to the graph of f at P*. Show that the tangent plane has equation

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)\,.$$

Example. Let $f(x, y) = 4 - \frac{1}{4}(x^2 + y^2)$. We will find the tangent plane to the graph of f at the point $(2, 2, 2)$. The partials at $(x, y) = (2, 2)$ are

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

$$f_y(2, 2) = -\frac{y}{2}\bigg|_{(2,2)} = -1$$

Thus the tangent plane has equation

$$z - 2 = -(x - 2) - (y - 2) \implies x + y + z = 6\,.$$

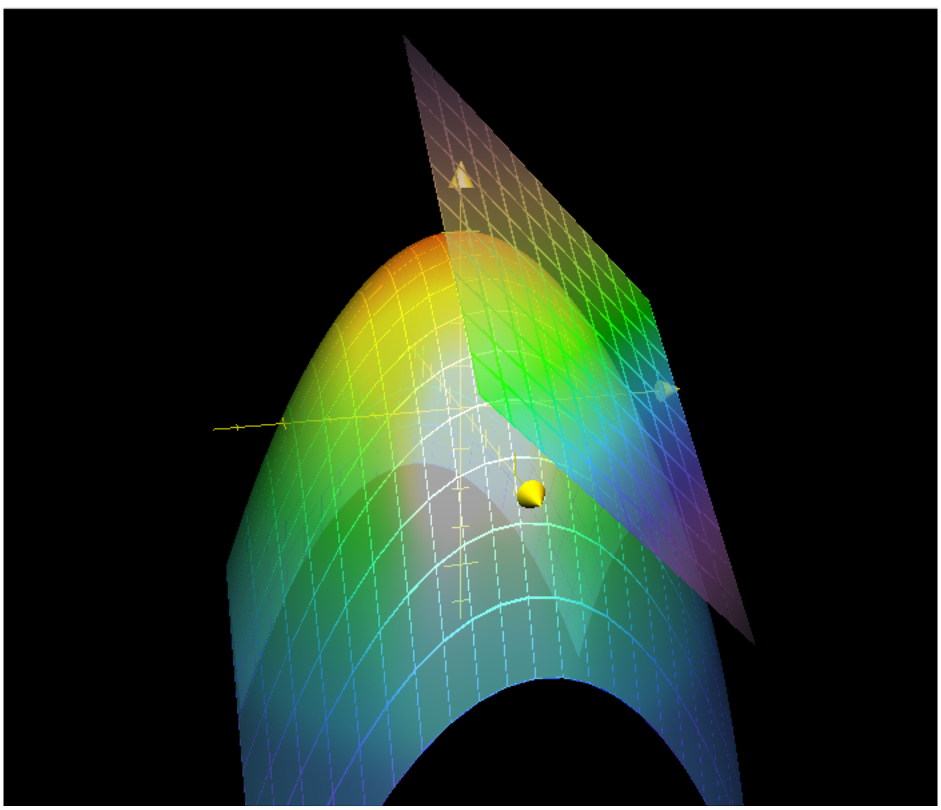


Figure 14. The tangent plane to $z = 4 - \frac{1}{4}(x^2 + y^2)$ at $(2, 2, 2)$.

§ 2.2. Linear Approximation

We can define a linear approximation from the tangent plane, much as one may use a tangent line to a differentiable single variable function to determine an approximation of the function near its point of tangency:

Definition. The linear approximation $L_{f,\mathbf{r}_0}(x,y)$of $f(x,y)$ near the point with position $\mathbf{r}_0 = \langle x_0, y_0\rangle$ is the value of the height z of the tangent plane to the point $P\big(x_0, y_0, f(x_0,y_0)\big)$:

$$L_{f,\mathbf{r}_0}(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)\,.$$

When it is clear what the function f is and which point $\mathbf{r}_0$ is to be the center of the approximation, one may simply write $L(x,y)$.

Example. For the function $f(x,y) = 4 - \frac{1}{4}(x^2+y^2)$, we have an approximation near $(2,2)$ given by

$$L(x,y) = 2 - (x-2) - (y-2) = 6 - x - y\,.$$

We can use this to compute $f(2.1, 1.8) \approx 6 - 2.1 - 1.8 = 2.1$. The actual value is $f(2.1, 1.8) = 2.0875$.

One can define linear approximations for functions of many variables analogously:

Definition. For a function $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^n$, the linear approximation $L_{f,\mathbf{r}_0}(\mathbf{r})$ to $f(\mathbf{r})$ at $\mathbf{r}_0 = \langle a_1, \ldots, a_n\rangle$ is

$$L_{f,\mathbf{r}_0}(\mathbf{r}) = f(\mathbf{r}_0) + \sum_{i=1}^{n} \partial_{x_i} f(\mathbf{r}_0)(x_i - a_i)\,.$$

Exercise 2.2. Compute the linear approximation to $f(x,y,z) = \sqrt{x^2+y^2+z^2+1}$ at $(1,1,1)$.

Exercise 2.3. For $f(x,y,z) = \sqrt{x^2+y^2+z^2}$, what is the linear approximation $L(x,y,z)$ at (x_0,y_0,z_0)? What happens to the error $f(x,y,z) - L(x,y,z)$ when you take $(x_0,y_0,z_0) = (0,0,0)$?

§ 2.3. Differentiability

Observe that one can apply the tangent plane formula and construct the linear approximation even if the partials f_x and f_y are *not continuous*; that is, the formulae are well defined so long as the limits

$$f_x(x_0,y_0) := \lim_{h\to 0} \frac{f(x_0+h, y_0) - f(x_0,y_0)}{h} \quad \text{and}$$

$$f_y(x_0,y_0) := \lim_{h\to 0} \frac{f(x_0, y_0+h) - f(x_0,y_0)}{h}$$

both exist. But the tangent plane and linear approximation may not have the desired geometric meaning if the partial derivatives are discontinuous.

Exercise 2.4. Consider the function

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

(a.) Calculate the partial derivative functions f_x and f_y. In particular, you should show that they exist at $(x,y) = (0,0)$ by computing

$$f_x(0,0) = \lim_{h\to 0}\frac{f(h,0) - f(0,0)}{h}\,, \quad \text{and} \quad f_y(0,0) = \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h}\,.$$

Explain why the partial derivatives fail to be continuous at the origin.

(b) Form the expression for $L_{f,\mathbf{0}}(x,y)$, and argue that it is not a good approximation.

(c) By rewriting f in polar coordinates, explain geometrically why the function fails to be well approximated by any plane at the origin.

Thus, we will define a notion of differentiability motivated by our geometric concerns: a function will fail to be differentiable at a point P if there is no good linear approximation at P, and so we may define differentiability in terms of the existence of such a linear approximation.

Definition. A bivariate function $f(\mathbf{r}) = f(x,y)$ is said to be *differentiable at* $\mathbf{r}_0 = \langle x_0, y_0\rangle$ if there exists a linear function $A(\mathbf{r})$ such that

$$\lim_{\mathbf{r}\to\mathbf{r}_0} \frac{f(\mathbf{r}) - f(\mathbf{r}_0) - A(\mathbf{r}-\mathbf{r}_0)}{\|\mathbf{r}-\mathbf{r}_0\|} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{\|\mathbf{h}\|}\big(f(\mathbf{r}_0+\mathbf{h}) - f(\mathbf{r}_0) - A(\mathbf{h})\big) = 0\,.$$

Equivalently, $f(x,y)$ is differentiable at (x_0,y_0) if the partial derivatives $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ both exist and there exist remainder functions ε_1 and ε_2 such that

$$f(x,y) = L_{f,\mathbf{r}_0}(x,y) + \varepsilon_1(x,y)\,(x-x_0) + \varepsilon_2(x,y)\,(y-y_0)\,,$$

and as $(x,y)\to(x_0,y_0)$, $(\varepsilon_1,\varepsilon_2)\to(0,0)$.

Writing $z = f(x,y)$, $\Delta z = z - z_0$, $\Delta x = x - x_0$ and $\Delta y = y - y_0$, one can rephrase the condition of differentiability as follows: f is differentiable at (x_0,y_0) if and only if

$$\Delta z = f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + \varepsilon_1\Delta x + \varepsilon_2\Delta y$$

for some pair of functions ε_1 and ε_2 both of which vanish in the limit as $(x,y)\to(x_0,y_0)$.

Proposition. *If a bivariate function is differentiable at* $\mathbf{r}_0 = \langle x_0,y_0\rangle$, *then the function*

$$L_{f,\mathbf{r}_0}(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

is the unique function such that the linear function $A(\mathbf{r}) := L(\mathbf{r}) - f(\mathbf{r}_0)$ *satisfies*

$$\lim_{\mathbf{r}\to\mathbf{r}_0} \frac{f(\mathbf{r}) - f(\mathbf{r}_0) - A(\mathbf{r}-\mathbf{r}_0)}{\|\mathbf{r}-\mathbf{r}_0\|} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{\|\mathbf{h}\|}\big(f(\mathbf{r}_0+\mathbf{h}) - f(\mathbf{r}_0) - A(\mathbf{h})\big) = 0\,.$$

More generally, for a multivariate function $f(\mathbf{r})$ with domain $D\subseteq\mathbb{R}^n$, one defines differentiability again in terms of the existence and effectiveness of a linear approximation:

Definition. The function $f(\mathbf{r}) = f(x_1,\ldots,x_n)$ is said to be *differentiable at* $\mathbf{r}_0$ if there exists a linear function $A(\mathbf{r})$ such that

$$\lim_{\mathbf{r}\to\mathbf{r}_0} \frac{f(\mathbf{r}) - f(\mathbf{r}_0) - A(\mathbf{r}-\mathbf{r}_0)}{\|\mathbf{r}-\mathbf{r}_0\|} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{\|\mathbf{h}\|}\big(f(\mathbf{r}_0+\mathbf{h}) - f(\mathbf{r}_0) - A(\mathbf{h})\big) = 0\,.$$

Exercise 2.5. Prove that a multivariate function $f(\mathbf{r})$ with domain $D\subseteq\mathbb{R}^n$ is differentiable at $\mathbf{r}_0 = \langle a_1,\ldots,a_n\rangle \in D$ if and only if

(i.) f is continuous at $\mathbf{r}_0$,
(ii.) all of the partial derivatives $\partial_{x_i} f$ exist at $\mathbf{r}_0$, and
(iii.) there exists a *remainder vector* $\boldsymbol{\varepsilon} = \langle \varepsilon_1, \ldots \varepsilon_n\rangle$ such that

$$f(\mathbf{r}) = L_{f,\mathbf{r}_0}(\mathbf{r}) + \boldsymbol{\varepsilon}\cdot(\mathbf{r}-\mathbf{r}_0)\,,$$

and as $\mathbf{r}\to\mathbf{r}_0$, $\boldsymbol{\varepsilon}\to\mathbf{0}$.

Deduce that if all of the first partials of f exist and are continuous at $\mathbf{r}_0$, then f is differentiable at $\mathbf{r}_0$. Can you find an example of a function g which is differentiable at a point $\mathbf{r}_0$, but not continuously differentiable there?

§ 2.4. The Total Differential

Definition. The *total differential* of a bivariate differentiable function $f(x,y)$ is

$$\mathrm{d}f = \frac{\partial f}{\partial x}\,\mathrm{d}x + \frac{\partial f}{\partial y}\,\mathrm{d}y\,.$$

The total differential can be thought of as a formal analogue to the increment formula

$$\Delta z = f_x\Delta x + f_y\Delta y\,,$$

giving an "infinitesimal version" of the linearization. It is an example of a *differential one-form*. We'll discuss one-forms again when we discuss line integrals. For now, the first utility of the total differential is as a means to estimate errors.

Example 2.1. Recall that the volume V of a right circular cone with base of radius r and height h is $V = \frac{\pi}{3}r^2h$. suppose you measure the radius and height of a cone to be $r = 10\,\mathrm{cm}$ and $h = 20\,\mathrm{cm}$ respectively. Suppose that the maximum error in each of your measurements is $0.4\,\mathrm{cm} = 4\,\mathrm{mm}$. Estimate the maximum error in the volume using differentials.

Solution: The volume differential is

$$\mathrm{d}V = V_r\,\mathrm{d}r + V_h\,\mathrm{d}h = \frac{2\pi}{3}rh\,\mathrm{d}r + \frac{\pi}{3}r^2\,\mathrm{d}h\,.$$

Using $\mathrm{d}r = 0.4\,\mathrm{cm} = \mathrm{d}h$, $r = 10\,\mathrm{cm}$, and $h = 20\,\mathrm{cm}$ gives an error of

$$\frac{2\pi}{3}(10\,\mathrm{cm})(20\,\mathrm{cm})(0.4\,\mathrm{cm}) + \frac{\pi}{3}(10\,\mathrm{cm})^2(0.4\,\mathrm{cm}) = \frac{200\pi}{3}\,\mathrm{cm}^3 \approx 209\,\mathrm{cm}^3\,.$$

Observe thus that a small error in length measurements leads to a potentially large error in volume measurement. Let us compare this error estimate to the real maximum error. The volume calculated from the measurements is

$$V_0 = \frac{2000\pi}{3}\,\mathrm{cm}^3\,.$$

The error is maximized in this case when the measurements given are smaller than the real lengths. Computing the real volume if the lengths are given by $r = 10.4\,\mathrm{cm}$ and $h = 20.4\,\mathrm{cm}$, one obtains

$$V = \frac{2206.464\pi}{3}\,\mathrm{cm}^3$$

The difference is then the real error: $V - V_0 = (206.464\pi)/3\,\mathrm{cm}^3 \approx 216\,\mathrm{cm}^3$.

3. The Gradient and Directional Derivatives

§ 3.1. The Directional Derivative

For a continuously differentiable bivariate function $f(x,y)$, consider the graph $\mathcal{G}_f = \{(x,y,z) \in \mathbb{R}^3 \,|\, z = f(x,y), (x,y) \in \mathrm{Dom}(f)\}$, and fix a point $\mathbf{r}_0 \in \mathrm{Dom}(f)$ at which we are interested in the rate of change of $z = f(x,y)$. Pick a direction vector $\hat{\mathbf{u}} \in \mathbb{S}^1$, and observe that $\hat{\mathbf{u}}$ and $\hat{\mathbf{k}}$ determine a plane $\Pi_{\hat{\mathbf{u}},\mathbf{r}_0}$ containing the point $\mathbf{p} = \mathbf{r}_0 + f(\mathbf{r}_0)\hat{\mathbf{k}} = \langle x_0, y_0, f(x_0,y_0)\rangle$, and this plane slices the surface $\mathcal{G}_f$ along some curve. In the plane $\Pi_{\hat{\mathbf{u}},\mathbf{r}_0}$, the variation of the curve as one displaces from $\mathbf{r}_0$ in the direction of $\pm\hat{\mathbf{u}}$ is purely in the z direction, and so it is natural to try to study the rate of change of z as one moves along the $\hat{\mathbf{u}}$ direction by a small displacement $h\hat{\mathbf{u}}$. The rate of change is what we define to be the directional derivative of f at $\mathbf{r}_0$ in the direction of $\hat{\mathbf{u}}$.

Definition 3.1. Given a unit vector $\hat{\mathbf{u}} \in \mathbb{S}^1$ and a bivariate function $f(x,y)$, the *directional derivative of f in the direction of $\hat{\mathbf{u}}$ at a point $\mathbf{r}_0 \in \mathrm{Dom}(f)$* is

$$D_{\hat{\mathbf{u}}} f(\mathbf{r}_0) = \lim_{h\to 0} \frac{f(\mathbf{r}_0 + h\hat{\mathbf{u}}) - f(\mathbf{r}_0)}{h}\,.$$

If the direction vector is given as $\hat{\mathbf{u}} = \cos\theta\hat{\mathbf{\imath}} + \sin\theta\hat{\mathbf{\jmath}}$ then one can write

$$D_{\hat{\mathbf{u}}} f(x_0,y_0) = D_{\langle\cos\theta,\sin\theta\rangle} f(x_0,y_0) = \lim_{h\to 0} \frac{f(x_0 + h\cos\theta, y_0 + h\sin\theta) - f(x_0,y_0)}{h}\,.$$

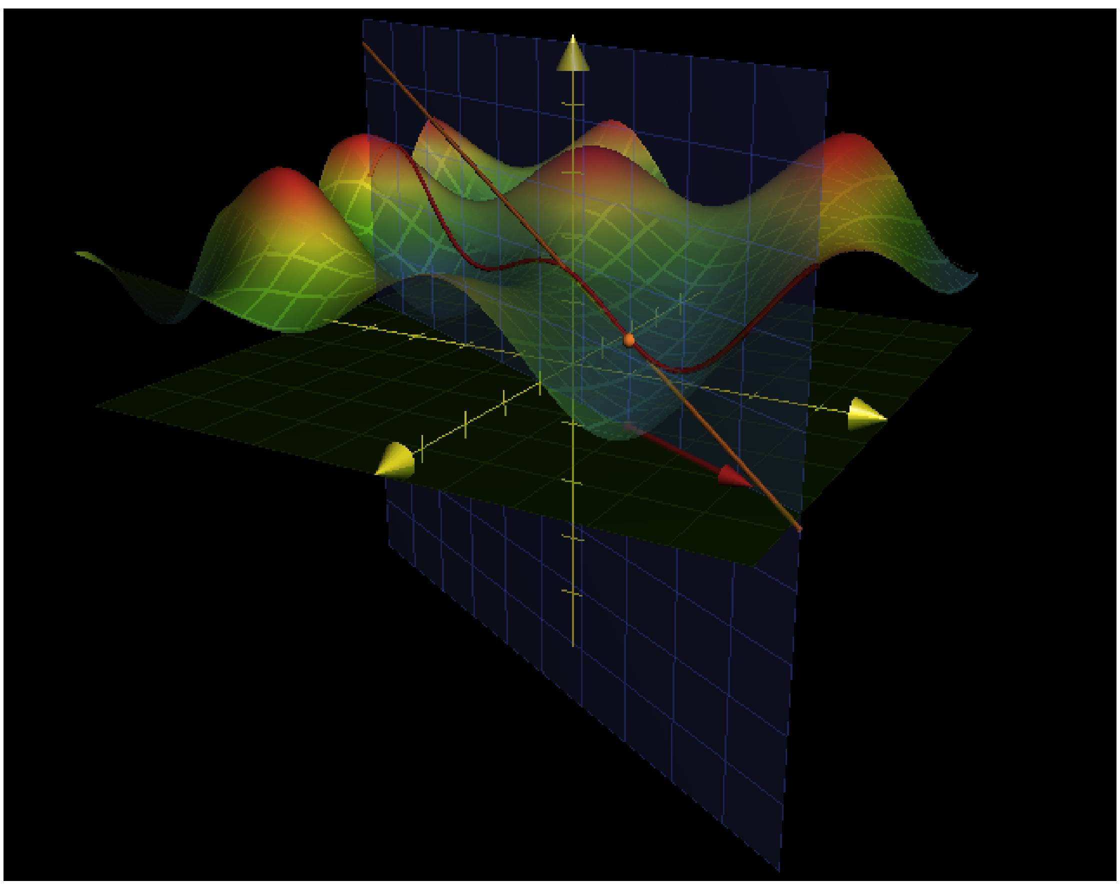


Figure 15. The directional derivative computes a slope to a curve of intersection of a vertical plane slicing the graph surface above a specified point and in the direction specified by a given unit vector.

We can derive an easy expression to compute the directional derivative in terms of $\hat{\mathbf{u}}$ from the chain rule.

Proposition. *The directional derivative of f at $\mathbf{r}_0$ in the direction of $\hat{\mathbf{u}}$ is given by*

$$D_{\hat{\mathbf{u}}}f(\mathbf{r}_0) = \langle f_x(\mathbf{r}_0), f_y(\mathbf{r}_0)\rangle \cdot \hat{\mathbf{u}}\,.$$

Proof. Let $g(h) = f(\mathbf{r}_0 + h\hat{\mathbf{u}})$, and write $\hat{\mathbf{u}} = \cos\theta\,\hat{\imath} + \sin\theta\,\hat{\jmath}$. Then

$$g'(0) = \lim_{h\to 0}\frac{f(\mathbf{r}_0 + h\hat{\mathbf{u}}) - f(\mathbf{r}_0)}{h} = D_{\hat{\mathbf{u}}}f(\mathbf{r}_0)\,.$$

On the other hand, by the chain rule,

$$\begin{aligned} g'(0) &= \frac{\mathrm{d}}{\mathrm{d}h}\Big(f(\mathbf{r}_0 + h\hat{\mathbf{u}})\Big)_{h=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}h}\Big(f(x_0 + h\cos\theta, y_0 + h\sin\theta)\Big)_{h=0} \\ &= \Big[\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}h} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}h}\Big]_{h=0} \\ &= f_x(x_0,y_0)\cos\theta + f_y(x_0,y_0)\sin\theta = \langle f_x(\mathbf{r}_0), f_y(\mathbf{r}_0)\rangle \cdot \hat{\mathbf{u}}\,. \end{aligned}$$

$\square$

Example. Find the directional derivative of $f(x,y) = 6 - 3x^2 - 2y^2$ at $(x,y) = (1,1)$ in the direction of $\hat{\mathbf{u}}$ where $\hat{\mathbf{u}}$ makes an angle of $\pi/3$ with the x-axis.

Solution: The direction vector we want is $\mathbf{u} = \cos(\pi/3)\,\hat{\imath} + \sin(\pi/3)\,\hat{\jmath} = \frac{1}{2}\hat{\imath} + \frac{\sqrt{3}}{2}\hat{\jmath}$. The partial derivatives at $(1,1)$ are

$$f_x(1,1) = -6\,, \quad \text{and} \quad f_y(1,1) = -4\,,$$

whence the directional derivative is

$$D_{\hat{\mathbf{u}}}f(1,1) = (-6\hat{\imath} - 4\hat{\jmath}) \cdot \left(\frac{1}{2}\hat{\imath} + \frac{\sqrt{3}}{2}\hat{\jmath}\right) = -3 - 2\sqrt{3}\,.$$

Example. Find the directional derivatives of $f(x,y) = \sin(xy)$ at $(\sqrt{\pi}/2, \sqrt{\pi}/3)$ in the directions of the vectors $\mathbf{v} = \hat{\imath} + \hat{\jmath}$ and $\mathbf{w} = 3\hat{\imath} - 4\hat{\jmath}$.

Solution:

Note that we must normalize the vectors $\mathbf{v}$ and $\mathbf{w}$ to obtain unit vectors:

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\hat{\imath} + \hat{\jmath}}{\sqrt{2}} = \frac{\sqrt{2}}{2}\hat{\imath} + \frac{\sqrt{2}}{2}\hat{\jmath}\,, \quad \hat{\mathbf{w}} = \frac{\mathbf{w}}{\|\mathbf{w}\|} = \frac{3\hat{\imath} - 4\hat{\jmath}}{\sqrt{25}} = \frac{3}{5}\hat{\imath} - \frac{4}{5}\hat{\jmath}\,.$$

We then compute the partial derivatives at $(\sqrt{\pi}/2, \sqrt{\pi}/3)$.

$$f_x(x,y) = y\cos(xy) \implies f_x(\sqrt{\pi}/2, \sqrt{\pi}/3) = \frac{\sqrt{\pi}}{3}\cos\left(\frac{\pi}{6}\right) = \frac{\sqrt{3\pi}}{6}\,,$$

$$f_y(x,y) = x\cos(xy) \implies f_y(\sqrt{\pi}/2, \sqrt{\pi}/3) = \frac{\sqrt{\pi}}{2}\cos\left(\frac{\pi}{6}\right) = \frac{\sqrt{3\pi}}{4}\,.$$

The desired directional derivatives are thus

$$\begin{aligned} D_{\hat{\mathbf{v}}}f(\sqrt{\pi}/2, \sqrt{\pi}/3) &= \langle f_x(\sqrt{\pi}/2, \sqrt{\pi}/3), f_y(\sqrt{\pi}/2, \sqrt{\pi}/3)\rangle \cdot \hat{\mathbf{v}} \\ &= \left(\frac{\sqrt{3\pi}}{6}\right)\left(\frac{\sqrt{2}}{2}\right) + \left(\frac{\sqrt{3\pi}}{4}\right)\left(\frac{\sqrt{2}}{2}\right) = \frac{5\sqrt{6\pi}}{24}\,, \\ D_{\hat{\mathbf{w}}}f(\sqrt{\pi}/2, \sqrt{\pi}/3) &= \langle f_x(\sqrt{\pi}/2, \sqrt{\pi}/3), f_y(\sqrt{\pi}/2, \sqrt{\pi}/3)\rangle \cdot \hat{\mathbf{w}} \\ &= \left(\frac{\sqrt{3\pi}}{6}\right)\left(\frac{3}{5}\right) - \left(\frac{\sqrt{3\pi}}{4}\right)\left(\frac{4}{5}\right) = -\frac{\sqrt{3\pi}}{10}\,. \end{aligned}$$

Exercise 3.1. Let $f(x,y)=\sqrt{4-x^2-y^2}$.

(a) Find the directional derivative of f at (x,y) in the direction of a vector making angle θ with $\hat{\mathbf{\imath}}$.

(b) At the point $(1,1)$, for what angle θ between $\hat{\mathbf{u}}$ and $\hat{\mathbf{\imath}}$ is the directional derivative largest? For an arbitrary but fixed point $(x_0,y_0)\in \mathrm{Dom}(f)$, determine the angle which maximizes the directional derivative in terms of x_0 and y_0.

(c) At $(1,1)$, in what directions $\pm\hat{\mathbf{u}}$ is $D_{\pm\hat{\mathbf{u}}}f(1,1)=0$? Give explicit unit vectors.

(c) For what (x,y) is the directional derivative 0 regardless of the direction $\hat{\mathbf{u}}$? What does this reveal about the geometry of the graph?

Exercise 3.2. Let $f(x,y)$ be a two-variable continuously differentiable function, $\mathcal{G}_f$ its graph, and $\Pi_{\hat{\mathbf{u}},\mathbf{r}_0}$ the vertical plane containing $\mathbf{p}=\mathbf{r}_0+f(\mathbf{r}_0)\hat{\mathbf{k}}$ and determined by a direction $\hat{\mathbf{u}}\in\mathbb{S}^1\subset\{z=0\}$. Find a parametrization $\mathbf{c}(t)$ of the curve which is the locus of the intersection $\mathcal{G}_f\cap\Pi_{\hat{\mathbf{u}},\mathbf{r}_0}$, so that at $t=0$ the position is $\mathbf{p}=\mathbf{r}_0+f(\mathbf{r}_0)\hat{\mathbf{k}}$, and compute the curvature of this curve at $\mathbf{p}$ using the chain rule for $f\big(\mathbf{c}(t)\big)$.

---

§ 3.2. The Gradient

The vector $\langle f_x(\mathbf{r}_0), f_y(\mathbf{r}_0)\rangle$ that appears in the formula for the directional derivative, as well as in the chain rule for the derivative of f along a curve $\mathbf{r}(t)$ is of great geometric importance, and so it is granted its own name.

Definition. The gradient vector of a bivariate function $f(x,y)$ at (x_0,y_0) is the vector

$$\nabla f(x_0,y_0) := f_x(x_0,y_0)\,\hat{\mathbf{\imath}} + f_y(x_0,y_0)\,\hat{\mathbf{\jmath}}\,.$$

The symbol ∇ is read as "nabla," though many prefer to call it "del". The symbols ∇f are read as "the gradient of f", "grad f", "nabla f", or "del f".

We can rewrite the directional derivative using the gradient:

$$D_{\hat{\mathbf{u}}}f(\mathbf{r}_0)=\nabla f(\mathbf{r}_0)\cdot\hat{\mathbf{u}}\,.$$

Proposition. *The gradient vector $\nabla f(\mathbf{r}_0)$ is orthogonal to the level curve through $\mathbf{r}_0=\langle x_0,y_0\rangle$, and gives the direction of steepest ascent on the surface of the graph $z=f(x,y)$, and the rate of change in this direction is $\|\nabla f(\mathbf{r}_0)\|$.*

Proof. By the directional derivative formula,

$$D_{\hat{\mathbf{u}}}f(\mathbf{r}_0)=\nabla f(\mathbf{r}_0)\cdot\hat{\mathbf{u}}=\|\nabla f(\mathbf{r}_0)\|\cdot\hat{\mathbf{u}}=\|\nabla f(\mathbf{r}_0)\|\cos\varphi\,,$$

where $\varphi\in[0,\pi]$ is the angle between $\nabla f(\mathbf{r}_0)$ and $\hat{\mathbf{u}}$. Thus, the direction $\hat{\mathbf{u}}$ maximizing $D_{\hat{\mathbf{u}}}f(\mathbf{r}_0)$ must maximize $\cos\varphi$, which implies $\varphi=0$. Thus, the maximizing direction is

$$\hat{\mathbf{u}}_{\max}=\nabla f(\mathbf{r}_0)/\|\nabla f(\mathbf{r}_0)\|\,,$$

and for this direction,

$$D_{\hat{\mathbf{u}}_{\max}}f(\mathbf{r}_0)=\|\nabla f(\mathbf{r}_0)\|\,.$$

On the other hand, the directions in which the value of f is constant are tangent to the level curves, which are sets where f has constant value. But these must be the directions along which the directional derivative is zero, which happens when $\cos\varphi=0 \implies \varphi=\pi/2$ or $3\pi/2$, i.e., when $\hat{\mathbf{u}}$ is perpendicular to $\nabla f(\mathbf{r}_0)$. Thus, the tangent vectors at $\mathbf{r}_0$ to the level curve through $\mathbf{r}_0$ are perpendicular to the gradient. □

The above proposition suggests an alternative definition of the gradient, which is "coordinate free" and serves as a reasonable definition in many variables as well:

Definition (The gradient as the vector of steepest ascent)**.** For $f(x_1, \dots, x_n)$ a multivariate function differentiable at the point P, the gradient of f at P is the unique vector $\nabla f(P)$ such that $D_{\hat{\mathbf{u}}} f(P)$ is maximized by choosing $\hat{\mathbf{u}} = \nabla f(P)/\|\nabla f(P)\|$, and

$$D_{\hat{\mathbf{u}}} f(P) = \|\nabla f(P)\|$$

gives the maximum rate of change of f at P. Observe that the minimum value of $D_{\hat{\mathbf{u}}} f(P)$ occurs for $\hat{\mathbf{u}} = -\nabla f(P)/\|\nabla f(P)\|$, and the minimum rate of change is $-\|\nabla f(P)\|$, and that $\nabla f(P)$ is orthogonal to the level set of f containing P.

Exercise 3.3. While exploring an exoplanet (alone and un-armed–what were you thinking‽) you've slid part way down a strangely smooth, deep hole. The alien terrain you are on is modeled locally (in a neighborhood around you spanning several dozen square kilometers) by the height function

$$z = f(x, y) = \ln\sqrt{16x^2 + 9y^2}\,,$$

where the height z is given in kilometers. Let $\hat{\imath}$ point eastward and $\hat{\jmath}$ point northward. Your current position is one eighth kilometers east, and one sixth kilometers south, relative to the origin of the (x, y) coordinate system given. You want to climb out of this strange crater to get away from the rumbling in the darkness below you.

(a) Find your current height relative to the $z = 0$ plane.

(b) Show that the level curves $z = k$ for constants k are ellipses, and explicitly determine the semi-major and semi-minor axis lengths in terms of the level constant k.

(c) In what direction(s) should you initially travel if you wish to stay at the current altitude?

(d) What happens if you travel in the direction of the vector $-(1/8)\hat{\imath} + (1/6)\hat{\jmath}$? Should you try this?

(e) In what direction should you travel if you wish to climb up (and hopefully out) as quickly as possible? Justify your choice mathematically.

(f) For each of the directions described in parts (c), (d), and (e), explicitly calculate the rate of change of your altitude along those directions.

§ 3.3. Tangent Spaces and Normal Vectors

Observe that we can rewrite the linear approximation using the gradient of f:

$$L_{f,\mathbf{r}_0}(\mathbf{r}) = f(\mathbf{r}_0) + \nabla f(\mathbf{r}_0) \cdot (\mathbf{r} - \mathbf{r}_0)\,,$$

and this formula works to define a linear approximation for n-variable f so long as $\nabla f(\mathbf{r}_0)$ is defined.

In particular, we can use gradients to describe the tangent spaces to a (hyper)surface given by a graph. But what of implicit surfaces?

An implicit surface can be viewed as specifying the surface as the level set of some function. That is, if the equation of the surface is given by $F(x, y, z) = k$ for some k, then the surface is precisely the level surface

$$F^{-1}(\{k\}) = \{(x, y, z) \,|\, F(x, y, z) = k\}\,.$$

But then, the gradient ∇F is always orthogonal to level sets of F, whence we can use the gradient vector at a point $P \in F^{-1}(\{k\})$ as a normal vector to the surface at P. We can use this normal vector as the normal to the tangent plane at P, and thus, obtain an equation for the tangent plane to a point P of the implicit surface with equation $F(x, y, z) = k$:

Proposition. *The equation of the tangent plane at $P(x_0, y_0, z_0)$ to the surface $F(x, y, z) = k$, assuming F is differentiable at P, is*

$$\nabla F(P) \cdot \langle x - x_0, y - y_0, z - z_0 \rangle = 0\,.$$

If we write $\mathbf{r}_0 = \langle x_0, y_0, z_0\rangle$*, and* $\mathbf{r} = \langle x, y, z\rangle$ *then this equation has the pleasing and easy to remember form*

$$\nabla F(\mathbf{r}_0)\cdot(\mathbf{r}-\mathbf{r}_0) = 0\,.$$

Example. For the function $F(x,y,z) = xy - xz + yz$, consider the implicit surface given by $F(x,y,z) = 1$ discussed in the section on implicit differentiation above. At the point $(2,3,-5)$ on the surface, the gradient of F is

$$\nabla F(2,3,-5) = \langle F_x, F_y, F_z\rangle\Big|_{(2,3,-5)} = \langle y-z, x+z, y-x\rangle\Big|_{(2,3,-5)} = \langle 8,-3,1\rangle\,,$$

and so the tangent plane to the surface at $(2,3,-5)$ has equation

$$\nabla f(2,3,-5)\cdot\langle x-2, y-3, z+5\rangle = 8(x-2) - 3(y-3) + (z+5) = 0\,,$$

which can be rewritten as

$$8x - 3y + z = 2\,.$$

Example. Any graph given by $z = f(x,y)$ can be rewritten as an implicit surface $F(x,y,z) = z - f(x,y) = 0$. If we apply the gradient to this F, treating z as an independent variable, we get

$$\nabla F(x,y,z) = \langle -f_x(x,y), -f_y(x,y), 1\rangle\,,$$

which gives a tangent plane equation at $x_0, y_0, z_0 = f(x_0,y_0)$ of

$$-f_x(x_0,y_0)(x-x_0) - f_y(x_0,y_0)(y-y_0) + (z-z_0) = 0 \implies z - z_0 = f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)\,,$$

which recovers the original tangent plane formula.

Exercise 3.4. Find the equation of the tangent plane to the hyperboloid $x^2 + y^2 - z^2 = 1$ at the point $(\sqrt{2}, \sqrt{3}, 2)$.

Exercise 3.5. Use gradients to demonstrate that the tangent plane to a sphere at a point is always perpendicular to the radius, and give a general formula for the tangent plane at $P(x_0, y_0, z_0)$ to an origin centered sphere containing the point P.

Exercise 3.6. Consider the surface implicitly defined by

$$\left(x^2 + y^2 + z^2 - \frac{5}{4}\right)^2 = 1 - 4x^2$$

.

(a) Find the equations of the traces for constant x, y, and z, and plot these families (you may use a computer, especially for the z traces). What is this surface?

(b) Find the heights for which the tangent planes are horizontal.

(c) There are horizontal tangent planes which intersect the surface along curves rather than in a single point. Sketch these curves.

(d) What do the self-intersections of the level curves in (c) tell us about the surface?

(e) Use techniques from single variable calculus to find the volume enclosed by this surface.

4. Extreme Values and Optimization

§ 4.1. Local extrema and critical points

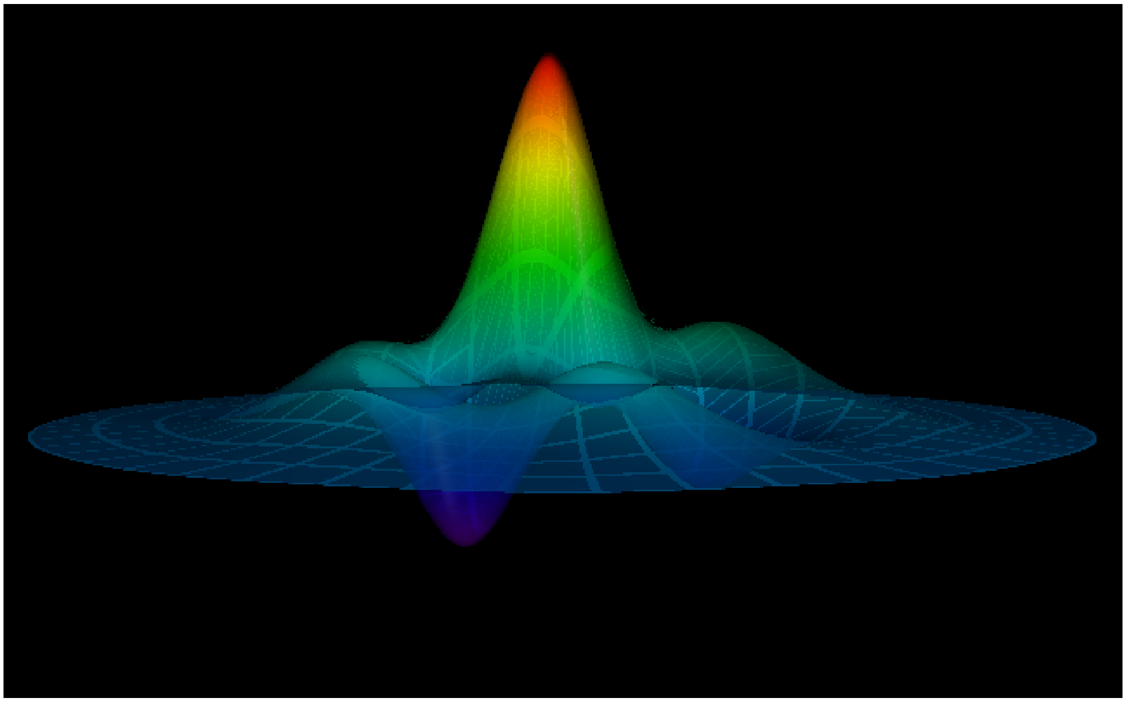


FIGURE 16. A graph surface revealing a function with a number of local extrema, as well as absolute extrema, over a disk domain D.

Consider the surface depicted in figure 16, given as the graph $z = f(x, y)$ for some bivariate function f defined over a domain D. This surface resembles a mountainous terrain, with a few mountain passes, and some depressions. Some of these features correspond to values of $f(x, y)$ which are *local extrema*. For example, a peak of the surface corresponds to some "critical" pair of an input and output for the function $f(x, y)$, for which the output value is larger than the values of the function for "nearby" inputs. We'll call such a value a "local maximum". There is peak in the picture which corresponds to a local maximum value that is also a "global" or "absolute" maximum, in that its z value is larger than all other z values visible. Let us formally define various types of extrema.

Definition. Let $f(\mathbf{r})$ be a multivariate function defined on a domain D, and let $\mathbf{r}_0 \in D$ be a particular point. Then

- We say that the value $f(\mathbf{r}_0)$ is a *local maximum value*, or simply a *local maximum* if there is some neighborhood U around $\mathbf{r}_0$ such that $f(\mathbf{r}_0) \geq f(\mathbf{r})$ for all $\mathbf{r} \in U$.
- We say that the value $f(\mathbf{r}_0)$ is a *global maximum value*, an *absolute maximum value* or simply an *absolute maximum* if $f(\mathbf{r}_0) \geq f(\mathbf{r})$ for all $\mathbf{r} \in D$.
- We say that the value $f(\mathbf{r}_0)$ is a *local minimum value*, or simply a *local minimum* if there is some neighborhood U around $\mathbf{r}_0$ such that $f(\mathbf{r}_0) \leq f(\mathbf{r})$ for all $\mathbf{r} \in U$.
- We say that the value $f(\mathbf{r}_0)$ is a *global minimum value*, an *absolute minimum value* or simply an *absolute minimum* if $f(\mathbf{r}_0) \leq f(\mathbf{r})$ for all $\mathbf{r} \in D$.

We say that a point is an *extremum* of f if it is a local or global maximum or minimum.

Some remarks are in order:

- For bivariate functions, the neighborhoods U can be taken to be small disks around the input $\mathbf{r}_0 = \langle x_0, y_0 \rangle$. More generally, the neighborhoods can be taken to be small balls: $U = \{\mathbf{r} : \|\mathbf{r} - \mathbf{r}_0\| \leq \delta\}$ for some sufficiently small real number $\delta > 0$.

- Note that absolute extrema in the interior of the domain are also local extrema, and that a given absolute extremum may not be the unique absolute extremum, if the absolute extreme value occurs at multiple points. Extreme values may occur on the boundary as well; we'll call these *boundary extreme values*, or *boundary extrema*. These will be discussed in section 4.3 below.
- For f differentiable at $\mathbf{r}_0$, if $f(\mathbf{r}_0)$ is a local extremum, one should expect the tangent plane to be horizontal there! That is, we expect the partial derivatives at $\mathbf{r}_0$ to be 0, else the gradient vector (respectively, its negative) tell us a direction to travel in to obtain a locally larger (respectively, smaller) value.
- We can imagine an extremum at a non-smooth point: just think about a cone like $z = \sqrt{x^2+y^2}$, which clearly has an absolute minimum value of 0 at $\mathbf{r}_0 = \mathbf{0}$, but is not differentiable there. There is also no well defined tangent plane at this point, nor a well defined gradient vector.

In light of the above remarks, we consider investigating the types of inputs which can produce local extrema.

Definition. A *critical point* $\mathbf{r}_0 \in \mathrm{Dom}(f)$ of a function $f(\mathbf{r})$ is a point at which the gradient is zero or fails to exist. The value $f(\mathbf{r}_0)$ is then called a *critical value*. We can also define the set of all critical points

$$\mathrm{crit}\,(f) := \{\mathbf{r}_0 \in \mathrm{Dom}(f) : \nabla f(\mathbf{r}_0) = \mathbf{0} \text{ or } \nabla f(\mathbf{r}_0) \text{ does not exist}\}\,.$$

Sometimes we will also use the term critical point to describe the location on the graph corresponding to the pairing of a critical input with the critical value it produces. It should be clear from context (e.g., if we refer to the graph itself) whether we mean the critical point as an input, or the location on the graph itself.

Theorem 4.1 (Fermat's Theorem on Critical Points)**.** *If a point $\mathbf{r}_0 \in \mathrm{Dom}(f)$ produces a local extremum of f, then $\mathbf{r}_0 \in \mathrm{crit}\,(f)$.*

Note that the converse is not true! For example, consider "mountain passes", like $z = x^2 - y^2$. The tangent plane at $(0,0,0)$ is horizontal, and the gradient is $\mathbf{0}$ there, but there are points $\mathbf{r}$ arbitrarily close to $\mathbf{0}$ in $\mathbb{R}^2$ for which $f(\mathbf{r})$ is either positive or negative (consider values along the x and y axes). Thus the point $(0,0,0)$ is neither a local maximum nor a local minimum. See example 4.3.

Example 4.1. Find all critical points of the function $f(x,y) = x^4 - 4xy + y^4$, and determine the corresponding critical values.

To begin, we compute the partial derivatives f_x and f_y:

$$f_x(x,y) = 4x^3 - 4y\,, \quad f_y(x,y) = 4y^3 - 4x\,.$$

Thus, the critical points are points (x,y) such that $x^3 = y$ and $y^3 = x$ simultaneously. Substituting y^3 for x in the first equation we obtain $y^9 = y$, which has real solutions precisely when $y = 0$ or $y = \pm 1$. Returning to the first equation, we see that $x = 0$ works when $y = 0$, and $x = y = \pm 1$ gives the remaining possible solutions. Thus

$$\mathrm{crit}\,(f) = \{(-1,-1),(0,0),(1,1)\}\,.$$

The corresponding values are

$$f(0,0) = 0, f(\pm 1, \pm 1) = -2\,.$$

How do we determine in general if a critical point corresponds to a local extremum? Intuitively, if we understand the concavity or curvature of the graph surface around a critical point, we can determine if there is an extremum or not. Thus, we will find a generalization the second derivative test to functions of 2-variables. Before we do, we build some geometric intuition by exploring some model cases.

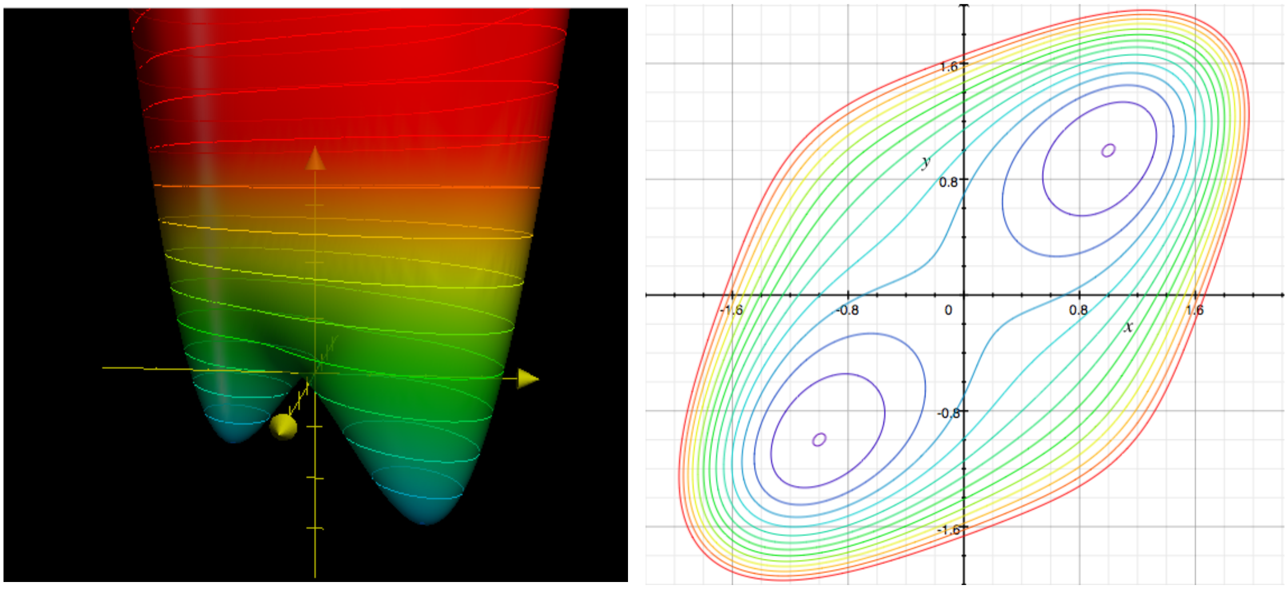


FIGURE 17. A view of the graph of $z = x^4 - 4xy + y^4$, and a selection of contours near the critical points.

Example 4.2. Let $f(x,y) = x^2 + y^2$ and $g(x,y) = 1 - x^2 - y^2$. Each has a single critical point at $(x_0, y_0) = (0,0)$. We can deduce that $(0,0)$ gives an absolute minimum value of 0 for f, while for g it produces an absolute maximum value of 1. Indeed, since $x^2 + y^2$ is a sum of squares, it is always nonnegative, and 0 is its minimum value. We can rewrite $g(x,y)$ in terms of f as $g(x,y) = 1 - f(x,y)$, which confirms that g has a maximum value of 1.

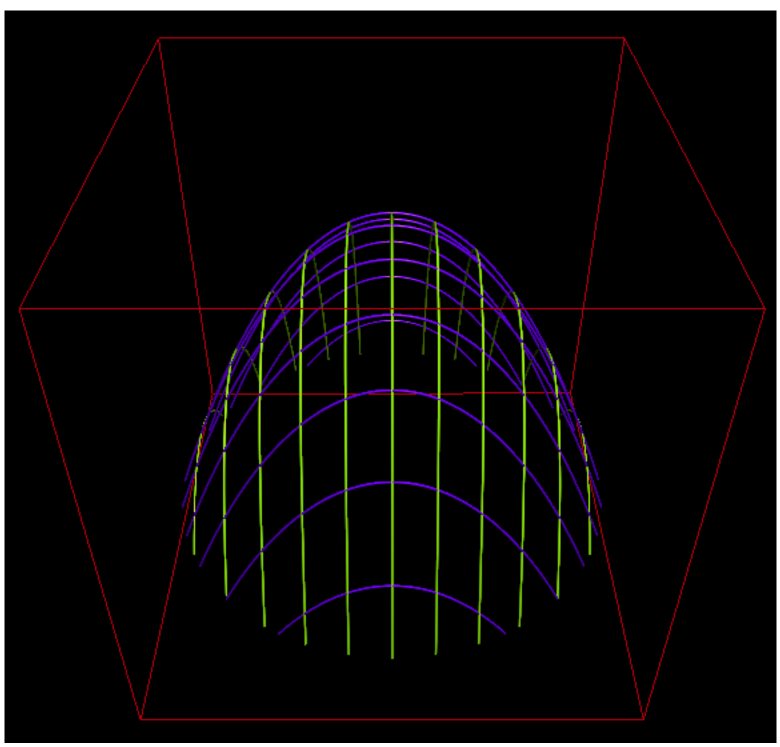


FIGURE 18. Traces of $z = 1 - x^2 - y^2$, which are all concave down.

An alternative approach is to examine the second derivatives. Observe that $f_{xx} = 2 = f_{yy}$, and $f_{xy} = f_{yx} = 0$. We can interpret $f_{xx}(x, y_0)$ as being the second derivative of the graph of the function $c_1(x) = f(x, y_0)$ in the plane $y = y_0$; that $f_{xx} = 2$ means that every such curve is concave up (i.e. convex in the positive z direction). We deduce similarly that every curve $c_2(y) = f(x_0, y)$ is concave up. Thus the surface $z = f(x, y)$ bends away from its horizontal tangent plane at $(0, 0, 0)$, and so this point is a global minimum point.

Similarly for $g(x, y)$, we have $g_{xx} = -2 = g_{yy}$, and $g_{xy} = g_{yx} = 0$, and we deduce that every trace curve in a vertical plane is concave down, and $(0, 0, 1)$ must be a global maximum point for the graph of $z = g(x, y) = 1 - f(x, y)$.

Example 4.3. Consider $h(x, y) = x^2 - y^2$ and $k(x, y) = 2xy$. There is again a single critical point at the origin of $\mathbb{R}^2$ for each surface, and the graphs $z = h(x, y)$ and $z = k(x, y)$ share a horizontal tangent plane of $z = 0$ at $(0, 0, 0)$. For each of h and k, there are inputs arbitrarily close to $(0, 0)$ for which outputs can be either positive or negative–that is, the graphs of each rise above and below the common tangent plane $z = 0$. Indeed, the tangent plane $z = 0$ intersects each graph in a pair of lines: setting $h(x, y) = 0$ gives $x^2 = y^2 \iff y = \pm x$, and similarly $k(x, y) = 0 \iff xy = 0 \iff x = 0$ or $y = 0$. These line pairings act as asymptotes for the hyperbolic level curves, and we can use this information to help construct the graphs. We see that $z = h(x, y)$ is a *saddle* with level curves given as hyperbolae $x^2 - y^2 = k$, while $z = k(x, y)$ is the same saddle rotated by $45°$.

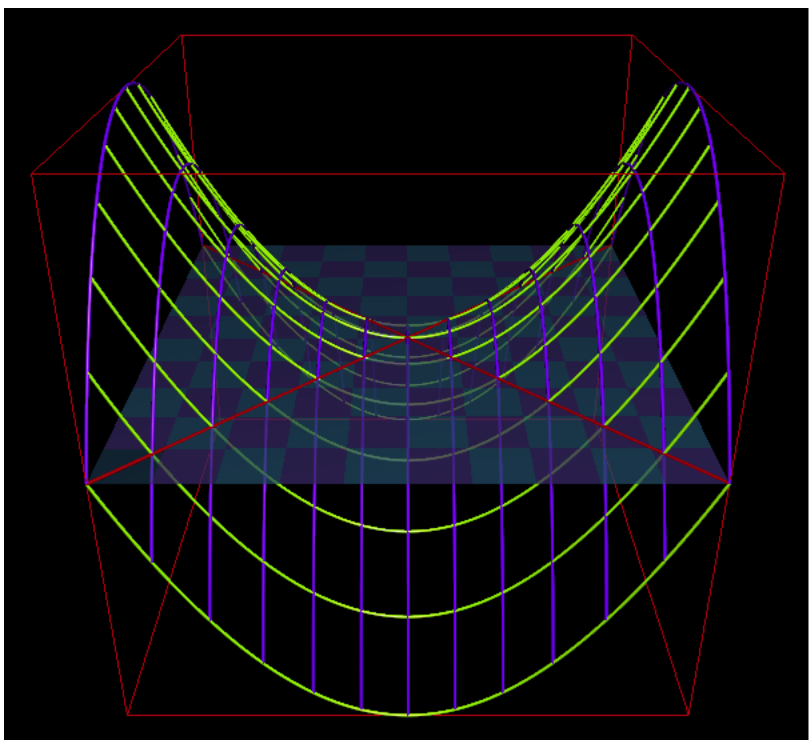


FIGURE 19. A saddle $z = x^2 - y^2$ showing lines of intersection with the tangent plane to the saddle point, and showing the opposite concavity of x and y traces.

As before, let us see if we can use second partial derivatives to analyze the concavity of traces and recover our conclusions from above. For $h(x, y)$ this will be mostly straightforward. Note that $h_{xx}(x, y) = 2$ but $h_{yy} = -2$. This tells us that traces in planes of constant y are concave up, while traces in planes of constant x are concave down. Note that $h_{xy} = 0$ identically.

The other saddle, $z = k(x, y)$ is just a rotation of the saddle $z = h(x, y)$. The interesting thing is, $k_{xx} = 0 = k_{yy}$ identically. Indeed, since $k_x(x, y) = 2y$ and $k_y(x, y) = 2x$, the only nonzero second derivatives are the mixed partials, both of which are identically 2. What this says is that the traces along planes of constant x and y are lines, but the lines slopes change as follows: as we sweep the plane $y = y_0$ through increasing values of y_0, the trace lines' slopes increase, since $k_{xy} = 2 > 0$.

Similarly, as we sweep planes $x = x_0$ through increasing values of x_0, the trace lines' slopes increase. The saddle, a hyperbolic paraboloid, can be swept out by lines in two ways! It is thus called doubly ruled.

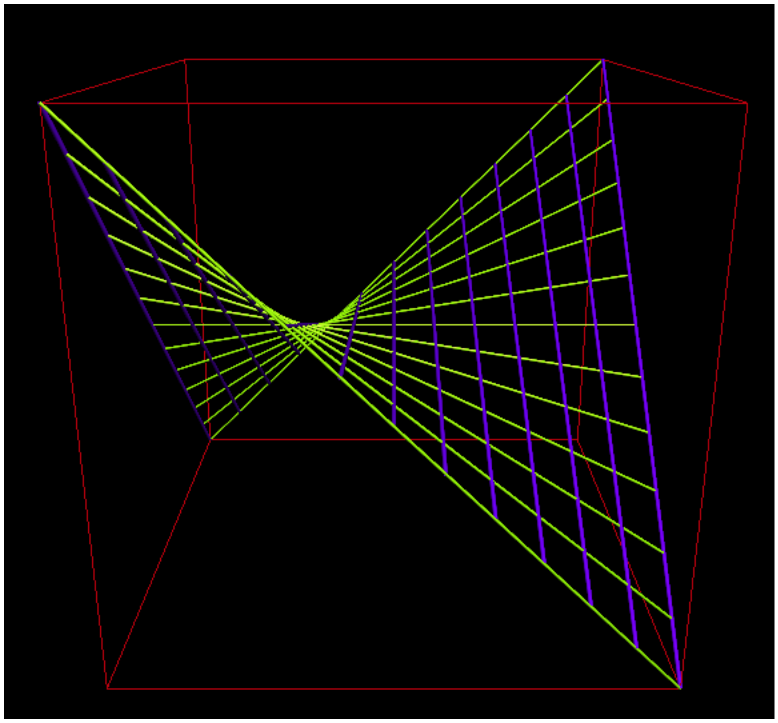


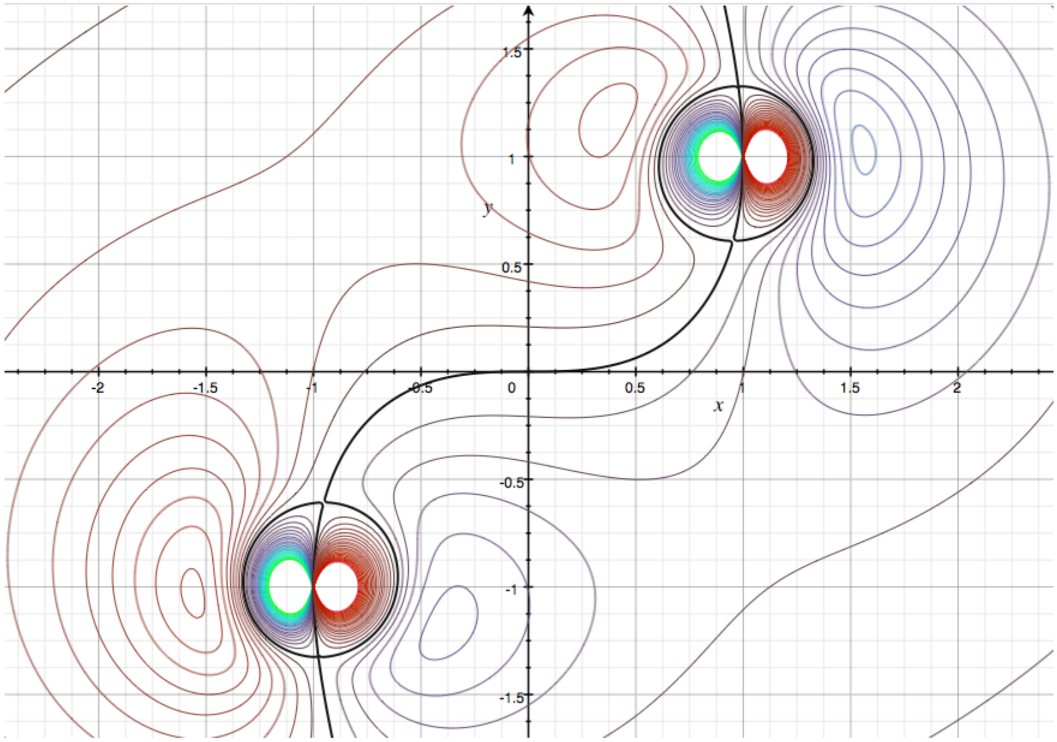
FIGURE 20. Traces of $z = 2xy$, which are all lines.

Observe that though k_{xy} was positive, the quantities $h_{xx}h_{yy} - (h_{xy})^2$ and $k_{xx}k_{yy} - (k_{xy})^2$ are both equal to -4. An interpretation of this will be illuminated in the next section.

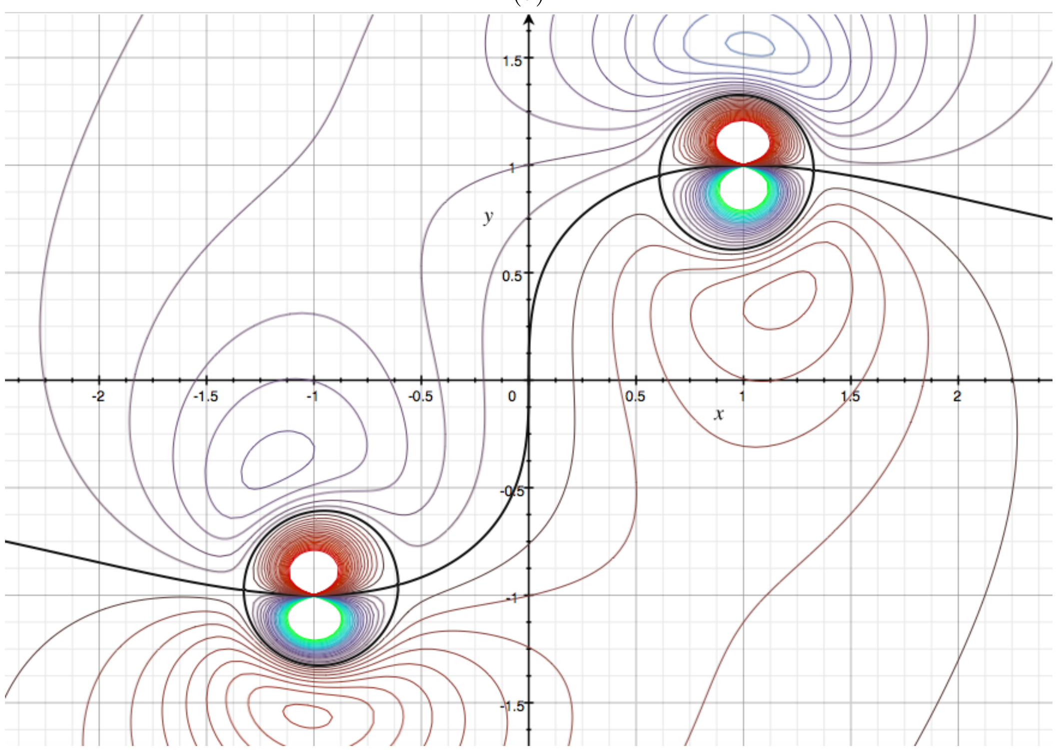
Example 4.4. Let $l(x, y) = (y - x)^2$ and let $m(x, y) = x^3 - 3xy^2$. We will see in these cases that the second derivatives don't tell the complete story. For $l(x, y)$, the critical locus is the whole line $y = x$, which corresponds to a global minimum value of 0. The graph is a parabolic cylinder, appearing like a trough with a whole line of minima! In particular, there is a direction in which the surface does not bend away from the tangent plane at a minimum, namely, along the vector $\frac{\sqrt{2}}{2}(\hat{\mathbf{\imath}} + \hat{\mathbf{\jmath}})$, even though $l_{xx} = 2 = l_{yy}$. But notice that $l_{xy} = l_{yx} = -2$, and so the quantity $l_{xx}l_{yy} - (l_{xy})^2$ is 0. In this case, the non-isolation of the critical points is related to the failure of the second derivatives to completely explain the behavior of the surface around its minima.

For $m(x, y)$, there is a unique critical point at the origin. However, $m_{xx}(0, 0) = 0 = m_{yy}(0, 0)$, and $m_{xy}(0, 0) = m_{yx}(0, 0) = 0$. The quantity $m_{xx}(\mathbf{0})m_{yy}(\mathbf{0}) - \big(m_{xy}(\mathbf{0})\big)^2 = 0$. In this case, one can check that the tangent plane intersects the graph in a collection of lines (how many?) and the surface has neither a local maximum nor a local minimum at $(0, 0, 0)$. The graph of $z = m(x, y)$ is often called a monkey saddle. Can you explain why?

Exercise 4.1. Suppose we are given the following contour plots (in figure 21) for the graphs of the first partial derivatives f_x and f_y of some function f. What information can be determined about crit (f) from these plots? Can we determine if critical points correspond to certain types of extrema?



(a)



(b)

FIGURE 21. (a) Contours for $k = f_x(x, y)$, with the bold black contour corresponding to the zero level. (b) Contours for $k = f_y(x, y)$, with the bold black contour corresponding to the zero level.

§ 4.2. The second derivative test

For this discussion we will work exclusively with bivariate functions.

To set up the second derivative test, first we define a matrix, called the *Hessian matrix*, which encodes all of the second derivatives. We saw in the above examples that the second derivatives, if nonzero, capture information about concavity along trace curves, and consequently, about the way a surface bends.

Definition. Let $f(x,y)$ be a function of two variables which is differentiable at (x_0,y_0), and assume that the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are also differentiable at (x_0,y_0). Then the *Hessian Matrix* $\mathrm{H}_f(x_0,y_0)$ is the 2×2 matrix whose entries are the second partial derivatives of f evaluated at (x_0,y_0):

$$\mathrm{H}_f(x_0,y_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(x_0,y_0) & \frac{\partial^2 f}{\partial x\,\partial y}(x_0,y_0) \\ \\ \frac{\partial^2 f}{\partial y\,\partial x}(x_0,y_0) & \frac{\partial^2 f}{\partial y^2}(x_0,y_0) \end{bmatrix}.$$

By Clairaut-Schwarz, this matrix is symmetric if the mixed second partials are continuous at (x_0,y_0), and we will assume this for the remainder of this section. The determinant $|\mathrm{H}_f(x_0,y_0)| = [f_{xx}f_{yy} - (f_{xy})^2]_{(x_0,y_0)}$ is called the *Hessian determinant*, or the *discriminant*.

Observe that the Hessian matrix has columns which are given by computing the gradients of the first partial derivatives, and rows which are given by computing partial derivatives of the gradient ∇f.

Our goal is to use the discriminant evaluated for a critical point $\mathbf{r}_0 \in \operatorname{crit}(f)$ to determine whether the critical point gives a local maximum, a local minimum, or neither.

Theorem. *Let $f(x,y)$ be a function with continuous first and second partial derivatives at a critical point $(x_0,y_0) \in \operatorname{crit}(f)$, and let $|\mathrm{H}| := |\mathrm{H}_f(x_0,y_0)| = [f_{xx}f_{yy} - (f_{xy})^2]_{(x_0,y_0)}$ be the discriminant. Then*

(i) *if $|\mathrm{H}| > 0$ and $f_{xx}(x_0,y_0) > 0$ then $f(x_0,y_0)$ is a local minimum value,*

(ii) *if $|\mathrm{H}| > 0$ and $f_{xx}(x_0,y_0) < 0$ then $f(x_0,y_0)$ is a local maximum value,*

(iii) *if $|\mathrm{H}| < 0$ then $f(x_0,y_0)$ is a saddle point, and thus neither a maximum nor a minimum,*

(iv) *if $|\mathrm{H}| = 0$ then the test is inconclusive, and the point $f(x_0,y_0)$ can exhibit any of the above behaviors.*

Example 4.5. For each of the model cases above in examples 4.2, 4.3, and 4.4 we can readily confirm the results of the test:

- For example 4.2: we have a local (and global) minimum value of 0 at $(0,0)$ for $f(x,y) = x^2+y^2$ and $|\mathrm{H}| = 4$ while $f_{xx}(0,0) = 2$, and for $g(x,y) = 1-x^2-y^2$ we have a local (and global) maximum value of 1 at $(0,0)$ with $|\mathrm{H}| = 4$ and $f_{xx}(0,0) = -2$. Thus f yields an example for criterion (i) in the above proposition, and g yields an example for criterion (ii).
- Both saddles $h(x,y) = x^2 - y^2$ and $k(x,y) = 2xy$ of example 4.3 have Hessian discriminant $|\mathrm{H}| < 0$ for the critical point $(0,0)$, and thus yield examples of criterion (iii).
- The functions $l(x,y) = (y-x)^2$ and $m(x,y) = x^3 - 3xy^2$ of example 4.4 yield cases under criterion (iv), since their Hessian discriminants are both 0. These are called *degenerate critical points*.

  Observe that after *perturbing* the function $m(x,y)$ by adding a small linear term $\varepsilon_1 x + \varepsilon_2 y$, the function has new critical points, and the second derivative test will work for these new critical points.

Exercise 4.2. Let ε_1 and ε_2 be any small positive constants, and define $\eta(x,y) = \varepsilon_1 x + \varepsilon_2 y$. Show that $m(x,y) + \eta(x,y)$ has two critical points, and classify them using the second derivative test.

Example 4.6. Reconsider the function $f(x,y) = x^4 - 4xy + y^4$ from example 4.1. Recall that $\mathrm{crit}\,(f) = \{(-1,-1),(0,0),(1,1)\}$. The Hessian is

$$H(x,y) = \begin{bmatrix} 12x & -4 \\ -4 & 12y \end{bmatrix},$$

and the corresponding discriminant is

$$|\mathrm{H}(x,y)| = 144xy - 16$$

$$|\mathrm{H}(\pm 1,\pm 1)| = 128 > 0, f_{xx}(\pm 1,\pm 1) = 12 < 0 \implies f(\pm 1,\pm 1) = -2 \text{ is a local minimum value,}$$
$$|\mathrm{H}(0,0)| = -16 < 0 \implies (0,0,0) \text{ is a saddle point}\,.$$

This is consistent with what we see in figure 17.

Exercise 4.3. Find and classify the critical points for the following functions:

(a) $f(x,y) = x^3 - 3xy^2 - 9x^2 - 6y^2 + 27$,

(b) $g(x,y) = \sin(x+y) + \cos(y-x)$,

(c) $h(x,y) = \dfrac{x}{1+y^2} + \dfrac{y}{1+x^2}$.

Exercise 4.4. Let $f(x,y) = xy + \cos(xy)$. Analyze the critical point locus of $f(x,y)$, and explain why one might appropriately say that the graph of $z = xy + \cos(xy)$ possesses infinitely many "saddle ridges". What can you say about extrema of f?

Exercise 4.5. Given a domain $\mathcal{D} \subset \mathbb{R}^2$, and a differentiable function of two variables $f : \mathcal{D} \to \mathbb{R}$, suppose $\mathbf{r} : I \to \mathcal{D}$ is a differentiable vector valued function defining a regular parametric curve $\mathbf{r}(t) = x(t)\hat{\imath} + y(t)\hat{\jmath}$ in the domain $\mathcal{D}$. Let $\mathbf{T}$ be the unit tangent vector of this curve and $D_{\mathbf{T}}$ the directional derivative operator in the direction of $\mathbf{T}$.

(a) Show that
$$\frac{\mathrm{d}}{\mathrm{d}t} f\big(\mathbf{r}(t)\big) = \|\dot{\mathbf{r}}(t)\| D_{\mathbf{T}} f\big(\mathbf{r}(t)\big)\,.$$

(b) Show that for a unit-speed parametrized curve $\mathbf{r}(s)$ and a continuously differentiable function f that
$$D^2_{\mathbf{r}'} f := D_{\mathbf{r}'}(D_{\mathbf{r}'} f) = \mathbf{r}' \cdot \mathrm{H}\mathbf{r}' = \frac{\partial^2 f}{\partial x^2}(x')^2 + 2\frac{\partial^2 f}{\partial x \partial y} x'y' + \frac{\partial^2 f}{\partial y^2}(y')^2\,,$$
where $\mathrm{H} = \mathrm{H}(f)$ is the Hessian matrix of f, and $\mathrm{H}\mathbf{r}'$ is the usual matrix-vector product.

(c) For a curve as in part (a) (not necessarily unit speed) compute
$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f\big(\mathbf{r}(t)\big)$$
in terms of partials of f and the derivatives $\dot{x}, \dot{y}, \ddot{x}, \ddot{y}$ using parts (a) and (b).

Exercise 4.6. Building off of the previous exercise (or assuming its results as needed), prove the second derivative test. Hints: consider directional derivatives for arbitrary $\mathbf{u} \in \mathbb{S}^1$, and show that if, e.g., the assumptions of (i) hold, then the curve of intersection of the graph and the plane through $\mathbf{r}_0$ and parallel to $\mathbf{u}$ is concave up; similarly use directional derivatives and the assumptions of (ii) and (iii) to assess the claims of the test. Finally, for (iv) produce and analyze examples with Hessian discriminant 0 exhibiting each type of behavior.

Exercise 4.7. This problem deals with the multivariable Taylor series. Consider a function of two variables defined on a domain $\mathcal{D} \subseteq \mathbb{R}^2$. Assume that f has continuous partial derivatives of all orders (a condition called *smoothness*; one often writes $f \in \mathcal{C}^\infty(\mathcal{D}, \mathbb{R})$ to indicate that it is in the class of smooth functions from $\mathcal{D}$ to $\mathbb{R}$.) The Taylor series of f centered at a point $(x_0, y_0) \in \mathcal{D}$ is

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{1}{n!\,m!} \frac{\partial^{n+m} f}{\partial x^n \partial y^m}\bigg|_{(x_0,y_0)} (x - x_0)^n (y - y_0)^m \,.$$

The k-th order Taylor polynomial of f is the truncated Taylor series:

$$T_{f,k}(x, y) = \sum\sum_{0 \le i+j \le k} \frac{1}{n!\,m!} \frac{\partial^{n+m} f}{\partial x^n \partial y^m}\bigg|_{(x_0,y_0)} (x - x_0)^n (y - y_0)^m$$

(*a*) Let $\nabla f(\mathbf{r}_0)$ denote the gradient vector of f evaluated at (x_0, y_0) and let $\mathrm{H}(\mathbf{r}_0)$ denote the Hessian matrix of f evaluated at $\mathbf{r}_0 = \langle x_0, y_0\rangle$. Show that the second order Taylor polynomial of f is

$$T_{f,2}(\mathbf{r}) = f(\mathbf{r}_0) + \nabla f(\mathbf{r}_0) \cdot (\mathbf{r} - \mathbf{r}_0) + \frac{1}{2}(\mathbf{r} - \mathbf{r}_0) \cdot (\mathrm{H}(\mathbf{r}_0)(\mathbf{r} - \mathbf{r}_0)) \,.$$

(*b*) Compute the second order Taylor polynomials of the following functions at the given points:

(*i*) $f(x, y) = e^{-x^2-y^2}$, $(x_0, y_0) = (0, 0)$,

(*ii*) $g(x, y) = \sin(xy)$, $(x_0, y_0) = (1/6, \pi)$.

(c) Note that $(0, 0)$ is a critical point for $f(x, y) = e^{-x^2-y^2}$. What can you say about the type of critical point from the second order Taylor polynomial?

(d) For $g(x, y) = \sin(xy)$, $(0, 0)$ is also a critical point. Compute the second order Taylor polynomial around $(0, 0)$ and determine the type of critical point. For the critical points $(\pi, 1/2)$, and $(3\pi, 1/2)$, compute 3rd order Taylor polynomials, and analyze the behavior of $g(x, y)$ around these points.

For exercises 4.8 and 4.9, the notations $\mathrm{H}_{x,y}(f)$, $\mathrm{H}_{u,v}(f)$ are used to emphasize which variables we use to compute partial derivatives, and to match with the Jacobian notation $\mathbf{D}_{x,y}(f)$, $\mathbf{D}_{u,v}(f)$ that was introduced in section 1.5.

Exercise 4.8. Use the notion of Jacobian derivatives $\mathbf{D}_{x,y} f = \begin{bmatrix} \partial_x f & \partial_y f \end{bmatrix}$ to show that the Hessian matrix may be defined as $\mathrm{H}_{x,y}(f) = \mathbf{D}_{x,y}(\mathbf{D}_{x,y} f)^{\mathrm{t}} = \mathbf{D}_{x,y}\nabla_{x,y} f$, where $(\mathbf{D}_{x,y} f)^{\mathrm{t}}$ is the transpose of the row vector, so $(\mathbf{D}_{x,y} f)^{\mathrm{t}} = \nabla_{x,y} f$.

In light of this alternate definition, sometimes the Hessian is denoted as $\mathbf{D}^2 f$ or $\nabla\nabla f$ by geometers (it should not be confused with $\nabla^2 f$, the Laplacian operator, which is a scalar valued function rather than a matrix).

Exercise 4.9. Let $\mathbf{G}$ be the coordinate transformation $(u, v) \mapsto (x, y)$ and $\mathbf{D}_{u,v}\mathbf{G} = \frac{\partial(x,y)}{\partial(u,v)}$ its Jacobian derivative matrix as defined in section 1.5 on the chain rule. Consider a function $f\big(x(u, v), y(u, v)\big)$ which is at least twice differentiable with respect to x and y, and assume x and y are each at least twice differentiable with respect to u and v.

Let $\mathbf{D}_{x,y} f = \begin{bmatrix} \partial_x f & \partial_y f \end{bmatrix}$ be the Jacobian derivative of f with respect to (x, y)-coordinates, and let $\mathbf{D}_{u,v} f = \begin{bmatrix} \partial_u f & \partial_v f \end{bmatrix}$ be the Jacobian derivative of f with respect to (u, v)-coordinates. Let $\mathrm{H}_{x,y} f$ denote the Hessian of f with respect to x and y coordinates, $\mathrm{H}_{u,v} f$ denote the Hessian of f with respect to u and v coordinates, and $\mathrm{H}_{u,v}\mathbf{G}$ denote the Hessian of the coordinate transformation $\mathbf{G}$, which can be regarded as a vector/block matrix whose 2 "entries" are the matrices $\mathrm{H}_{u,v} x$ and $\mathrm{H}_{u,v} y$.

(a) Show that

$$\mathrm{H}_{u,v}\big(f\circ\mathbf{G}(u,v)\big)=\Big[\mathbf{D}_{u,v}\mathbf{G}\circ\mathrm{H}_{x,y}(f)\circ\big(\mathbf{D}_{u,v}\mathbf{G}\big)^{\mathrm{t}}+(\mathbf{D}_{x,y}f)\circ\mathrm{H}_{u,v}(\mathbf{G})\Big]_{\langle x,y\rangle=\mathbf{G}(u,v)},$$

where for a matrix M, M^{t} is the transpose of M, and '$\circ$' denotes composition of linear maps (i.e. multiplication of the corresponding matrices), and

$$(\mathbf{D}_{x,y}f)\circ\mathrm{H}_{u,v}(\mathbf{G})=\partial_x f\mathrm{H}_{u,v}x+\partial_y f\mathrm{H}_{u,v}y$$

is just a linear combination of coordinate Hessians (added as matrices).

(b) Compute $H_{r,\theta}(f)$ for $f(x,y)=x^3-3xy^2$ using polar coordinates r,θ first by the above formula, and then by converting f into polar coordinates. What can you conclude about why the critical point $(0,0)$ is degenerate?

(c) Argue that for a critical point of f, the determinant of the Hessian does not change sign after a coordinate change. Thus, the sign of the discriminant is a coordinate invariant which detects something intrinsic about the graph of f at critical points (it turns out that it's related to curvature!)

---

§ 4.3. Optimization and the Extreme Value Theorem

Optimization problems are problems that involve producing global maximum and minimum values for a function, sometimes with constraints.

Example 4.7. Find the dimensions and volume of largest 3D rectangular box that can be fit between a plane and a hemisphere of radius $\sqrt{3}$ placed on the plane.

Intuition might suggest that the answer is half of a cube inscribed in a radius $\sqrt{3}$ sphere (for full sphere, the answer would be an inscribed cube as one can show by methods similar to those below). However, the goal of the example is to actually verify such intuition using calculus.

We can let our plane be the xy-plane in $\mathbb{R}^3$ with the hemisphere situated so that it is given as the graph of $z=\sqrt{3-x^2-y^2}$. By symmetry, we can argue that up to rotation, we may choose the box with sides parallel to coordinate planes.

Let (x,y,z) denote a corner of the box. We may choose the corner in the first octant, so x, y, and z are all positive. We may assume the other corners of the box are $(-x,y,z)$, $(x,-y,z)$, $(-x,-y,z)$, $(x,y,0)$, $(-x,y,0)$, $(-x,-y,0)$, and $(x,-y,0)$, so that the edge lengths are $2x$, $2y$ and z.

Thus, the function we wish to maximize is the volume $V(x,y,z)=4xyz$. Note that if the point (x,y,z) does not lie on the hemisphere, then we can certainly make a bigger volume by increasing z until (x,y,z) is on the hemisphere. Thus we can treat V as a two variable function by recognizing $z=z(x,y)=\sqrt{3-x^2-y^2}$ as a dependent variable.

Thus we want critical points of $V\big(x,y,z(x,y)\big)$:

$$\nabla V(x,y)=\langle V_x,V_y\rangle=\left\langle 4yz+4xy\frac{\partial z}{\partial x},4xz+4xy\frac{\partial z}{\partial y}\right\rangle=\left\langle 4yz+4xy\left(-\frac{x}{z}\right),4xz+4xy\left(-\frac{y}{z}\right)\right\rangle,$$

where we used implicit differentiation of $x^2+y^2+z^2=3$ to obtain $\partial_x z=-x/z$ and $\partial_y z=-y/z$. Setting $\nabla V(x,y)=\mathbf{0}$ gives the equations

$$0=4yz-4x^2y/z\implies yz^2=x^2y\implies z^2=x^2$$
$$0=4xz-4xy^2/z\implies xz^2=xy^2\implies z^2=y^2\,.$$

Since the point (x,y,z) denotes the corner where $x,y,z>0$, we deduce $x=y=z$, and since this point is on the sphere of radius $\sqrt{3}$, we have $3x^2=3$, so $x=y=z=1$. Thus the box is a half cube $[-1,1]\times[-1,1]\times[0,1]$, so the dimensions are 2 by 2 by 1, and the volume is $V=4$.

Observe that effectively, what we have done is compute an optimum of the function $V(x,y,z)=4xyz$ on the surface $z=\sqrt{3-x^2-y^2}$. For *the space region* between the plane and the hemisphere, we found a maximum value; a local minimum in this region happens at the only critical point of the (three variable) function $V(x,y,z)$, namely the origin, and also occurs whenever any of the coordinates takes a value of 0 . The global minimum value of -4 occurs at the boundary points $(-1, 1, 1)$ and $(1, -1, 1)$. We'll see shortly that global extrema for such a *compact* domain, like a solid ball or solid hemisphere, can either happen at interior critical points or at points along the boundary.

Exercise 4.10. Use calculus methods to prove that the minimum distance from a point $P(x_1, y_1, z_1)$ to a plane $ax + by + cz + d = 0$ is given by

$$D = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}\,,$$

and give the coordinates of the closest point. You should check your work using vector algebra methods.

We are interested in the following more general questions at the mathematical root of simple optimization problems:

- When is a function $f(x, y)$ or $F(x, y, z)$ guaranteed to have global extrema?
- How does one procedurally find global extrema, assuming they exist?

To answer these questions, we need a few topological preliminaries.

Definition 4.1. Let $E \subseteq \mathbb{R}^2$. A point $\mathbf{r}_0 \in E$ is called *a boundary point* of E if every open disk $B_\varepsilon(\mathbf{r}_0) = \{\mathbf{r} \in \mathbb{R}^2 : \|\mathbf{r} - \mathbf{r}_0\| < \varepsilon\}$ centered at $\mathbf{r}_0$ contains points both in E and in the complement of E. The *boundary of* E, denoted by ∂E is the set of al boundary points:

$$\partial E := \{\mathbf{r}_0 \in \mathbb{R}^2 : \text{ for all } \varepsilon > 0\,,\ B_\varepsilon(\mathbf{r}_0) \cap E \neq \varnothing \text{ and } B_\varepsilon(\mathbf{r}_0) \cap (\mathbb{R}^2 - E) \neq \varnothing\}\,.$$

Definition 4.2. The *interior* of $E \subseteq \mathbb{R}^2$ is the set of points of E which are *not* boundary points: $\operatorname{int} E := E - \partial E$.

Definition 4.3. A set $E \subseteq \mathbb{R}^2$ is called *closed* if it contains all of its boundary points: E closed $\iff \partial E \subseteq E$.

Definition 4.4. A set $E \subset \mathbb{R}^2$ is called *bounded* if there exists a disk D such that $E \subseteq D$.

Remark 4.1. If we replace open disks with open balls, the above definitions generalize to subsets of $\mathbb{R}^3$ or even $\mathbb{R}^n$.

Intuitively, boundary points are at the "edge" of the set; if the set is a contiguous region in $\mathbb{R}^2$, then the boundary is the collection of curves delineating the transition from "within" to "outside". Sets which are bounded intuitively don't "run off to infinity". Sets which are closed and bounded are often called *compact*; in the plane these are regions of finite area with boundaries that are (possibly several) closed curves.

Exercise 4.11. Draw pictures and indicate boundaries and interior for each of the following sets, and argue the corresponding claims:

(a) The "closed unit disk" $D = \{\mathbf{r} \in \mathbb{R}^2 : \|\mathbf{r}\| \leq 1\}$; D is closed and bounded by the above definitions.
(b) The region $E = \{(x, y) : xy \leq 1\}$; E is closed but unbounded by the above definitions.
(c) The "punctured disk" $D^* := D - \{\mathbf{0}\}$; D^* is bounded but is neither closed nor open (see the next exercise if you forgot the definition of open sets in $\mathbb{R}^2$).

Exercise 4.12. Recall, a set $E \subseteq \mathbb{R}^2$ is called *open* if around every point $\mathbf{r}_0 \in E$, there is an open disk $B_\varepsilon(\mathbf{r}_0)$ for some sufficiently small $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{r}_0) \subset E$. Prove the following using this definition for open sets and the above definitions for boundary points, interior points, and closed sets.

(a) The boundary ∂E is the complement in E of the interior: $\partial E = E - \operatorname{int} E$,

(b) The interior is the set of points which are everywhere surrounded by other interior points: $\mathbf{r}_0 \in \operatorname{int} E$ if and only if there exists a disk $B_\varepsilon(\mathbf{r}_0)$ for some $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{r}_0) \subset E$,

(c) A set is closed in $\mathbb{R}^2$ if and only if it is the complement in $\mathbb{R}^2$ of an open set: E closed if and only if there is an open set $U \subset \mathbb{R}^2$ such that $E = \mathbb{R}^2 - U$,

(d) A set $U \subseteq \mathbb{R}^2$ is open if and only if it equals its interior: U open if and only if $U = \operatorname{int} U$, i.e., if and only if $U \cap \partial U = \varnothing$,

The reason for introducing these topological ideas is that the question of the existence of absolute extrema depends upon topological properties of the domain and the function. Namely, we have the following version of the extreme value theorem:

Theorem 4.2 (Extreme Value Theorem for bivariate functions)**.** *A function $f(x, y)$ continuous on a closed and bounded (i.e., compact) domain $D \subset \mathbb{R}^2$ attains an absolute maximum value $f(\mathbf{r}_1)$ for some point $\mathbf{r}_1 \in D$ and an absolute minimum $f(\mathbf{r}_2)$ for some point $\mathbf{r}_2 \in D$.*

We won't prove this version of the extreme value theorem as it involves rigorously demonstrating the claims below about sequences in compact sets. We remark that its generalization holds: for appropriate definitions of compact and continuous, it is always true that a continuous $\mathbb{R}$-valued function defined on a compact domain K attains an absolute maximum value and an absolute minimum value for some inputs in K. We will use the abbreviation EVT to refer to any such result; context should make clear whether we are dealing with bivariate functions, trivariate functions, or some other case.

Though we won't prove the result, we make a few remarks about why topology comes up. Continuity is essentially a topological condition relating the domain and the function[4], and compactness is a topological condition on the domain itself. The intuition is that D being closed and bounded for continuous f prevents the function's values from "running away" indefinitely:

(i) because f is continuous, sequences of points in D that converge to a position in D produce convergent limits of values of f,

(ii) boundedness of D means no sequence of inputs can run off to infinity, with values of f becoming arbitrarily large or small,

(iii) because D is closed, sequences in D that converge must converge to points within D, where f is defined, so in particular sequences converging to boundary points yield definite limits of values of f;

(iv) compactness of D implies the boundary is itself compact, so our reasoning here and above extends to show that sequences of values of the function produced from convergent sequences within the boundary are also well behaved, and so in particular by reapplying EVT, there is a well defined *boundary extrema problem* whose solutions exist (though it may be difficult to find them),

(v) putting all these ideas together, there is no way for the value of f to increase or decrease indefinitely along any path or sequence in D, and so there must be some value which is largest, and some value which is smallest, and these may happen at interior critical points or somewhere along the boundary.

The main application of EVT is that, together with Fermat's theorem on critical points and local extrema, it suggests and guarantees the legitimacy of the following procedure to find absolute extrema.

[4]Recall that on page 6 of these notes continuity throughout a domain is rephrased in the context of open sets and pre-images. Point-set topology concerns itself with the minimum structures on sets necessary to define, analyze and infer continuity properties of functions; the first step is to create a coherent notion of open sets which defines "a topology" on the set of interest. Then concepts of connectedness, compactness, boundaries, and interiors are all definable as topological notions, determined as properties intrinsic to a set endowed with a given topology–that is, a set given a coherent notion of which subsets are to be regarded as open subsets.

Proposition 4.1 (Procedure to find global extrema)**.** *Let $f : D \to \mathbb{R}$ be a multivariate function which is continuous on D. If D is compact, then the following algorithm produces the global extreme values of f:*

(1) *Compute* crit f *and for each* $\mathbf{r} \in \operatorname{crit} f$ *compute the critical value* $f(\mathbf{r})$*,*
(2) *Compute* boundary extreme values*: find all extreme values of* f *on* ∂D*,*
(3) *Comparing values produced in steps 1 and 2, the global maximum value is the largest value obtained in either step, and the global minimum value is the smallest value obtained in either step.*

We remark that the hardest step is generally 2, as this involves constrained optimization. Since ∂D is itself compact when D is compact, one can try to rewrite the function f as a function with one less variable, and then apply the proposition again. In this way, one may recursively find boundary extrema. The following example illustrates this.

Example 4.8. Let $f(x, y) = 2x^2y^2 - x^2 - y^2 + 1$ and let S be the unit square $[0, 1] \times [0, 1] = \{(x, y) : 0 \le 1 \le x, 0 \le y \le 1\}$. Find the absolute maximum and minimum values of f, as well as all points where these values occur.

First we must compute the critical points within S and the corresponding critical values. The gradient of f is

$$\nabla f(x, y) = 2(2xy^2 - x)\hat{\imath} + 2(2x^2y - y)\hat{\jmath}\,.$$

Thus

$$\begin{aligned}\operatorname{crit}(f) &= \{(x, y) : 2xy^2 - x = 0 = 2x^2y - y\} \cap S \\ &= \{(0, 0\} \cup \{(x, y) : 2y^2 - 1 = 0 = 2x^2 - 1\} \cap S \\ &= \{(0, 0), (\sqrt{2}/2, \sqrt{2}/2)\}\,.\end{aligned}$$

The corresponding critical values are $f(0, 0) = 1$ and $f(\sqrt{2}/2, \sqrt{2}/2) = 1/2$.

We next examine the boundary of the square to seek potential boundary extrema. Note of course that $(0, 0)$ is a boundary point, but without second derivative testing or comparing to other boundary values, we don't yet have a conclusion as to what type of critical behavior occurs here. The boundary in this case consists of 4 line segments, each meeting two others at the squares corners. Label them as follows: $L_1 = [0, 1] \times \{0\}$ is the bottom edge, $L_2 = \{1\} \times [0, 1]$ is the right vertical edge, $L_3 = [0, 1] \times \{1\}$ is the top edge, and $L_4 = \{0\} \times [0, 1]$ is the left vertical edge.

Along L_1, the function restricts to the one-variable function $f(x, 0) = 1-x^2$, which is a parabolic curve which is decreasing as x increases. We conclude that along L_1, the maximum value occurs at $(0, 0)$, and the minimum along L_1 is $f(1, 0) = 0$, which occurs at $(1, 0)$.

Similarly, we'll find along each other edge a parabolic function which is either increasing or decreasing, and so for our function f, the boundary extrema occur at the corners of the square (boundaries of the boundary curves, if you will):

$$f(0, 0) = 1\,, \quad f(1, 0) = 0\,, \quad f(1, 1) = 1\,, \quad f(0, 1) = 0\,.$$

Comparing these values with the non-boundary critical value of $1/2$, we see that f has an absolute maximum value of 1 in S at the origin and also at the corner $(1, 1)$, and has an absolute minimum value of 0 at the remaining corners $(1, 0)$ and $(0, 1)$. The point $(\sqrt{2}/2, \sqrt{2}/2)$ is actually a saddle as the second derivative test easily confirms:

$$|H_f\big(\sqrt{2}/2, \sqrt{2}/2)\big)| = \Big[(4y^2 - 2)(4x^2 - 2) - (8xy)^2\Big]_{(\sqrt{2}/2, \sqrt{2}/2)} = 0 - 16 < 0\,.$$

Example 4.9. Find the absolute maximum and minimum values of $f(x, y, z) = 2x + y - 2z$ on the closed unit ball $B = \{\mathbf{r} : \|\mathbf{r}\| \le 1\}$.

Observe that f is linear, and hence its $3D$ gradient is never $\mathbf{0}$. Thus, the extreme values must occur on the boundary sphere $x^2 + y^2 + z^2 = 1$. We can implicitly differentiate f restricted to the boundary to obtain the (x, y) -gradient

$$\nabla f\big(x, y, z(x, y)\big) = \langle 2 - 2\partial_x z, 1 - 2\partial_y z\rangle = \langle 2 + 2x/z, 1 + 2y/z\rangle\,.$$

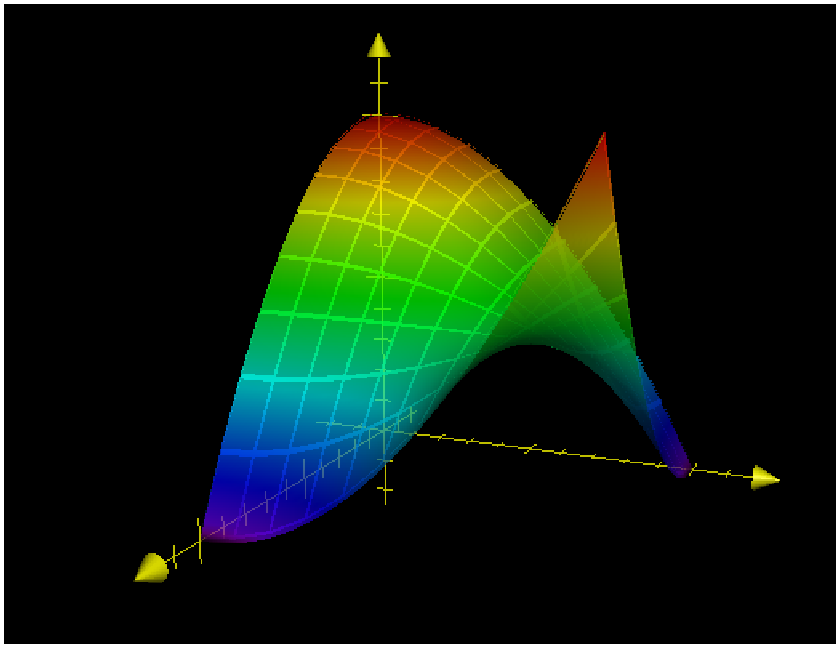


FIGURE 22. The function $f(x, y) = 2x^2y^2 - x^2 - y^2 + 1$ graphed over the unit square $[0, 1]^2$ has one saddle over the interior of the square, attains a boundary maximum value of 1 at opposite corners $(0, 0)$ and $(1, 1)$, and attains a boundary minimum value of 0 at the remaining opposite corners $(1, 0)$ and $(0, 1)$. The boundary extrema give the absolute extrema over the square in this case.

Then $\nabla f\big(x, y, z(x, y)\big) = \mathbf{0}$ if and only if, for $z \neq 0$, $x = -z = y/2$. Substituting into the equation of the sphere and solving, one has $x^2 + x^2/4 + x^2 = 1 \implies 9x^2 = 4$, so $x = \pm 2/3$, $y = \pm 1/3$ and $z = \mp 2/3$. The maximum is thus $f(2/3, 1/2. - 2/3) = 3$ and the minimum is $f(-2/3, -1/3, 2/3) = -3$.

Observe that these points are precisely points where the planes $f(x, y, z) = 2x + y - 2z = \pm 3$ are tangent to the sphere. In this case, had we realized that these optima occur where level surfaces of f are tangent to the constraint surface, we could have used elementary geometry of spheres and planes to locate these points: indeed they are given as the positive and negative unit vectors parallel to the gradient of f. In the next section we will exploit the relationship between tangencies of level sets and constrained extrema to give another method to solve such constrained optimization problems.

Exercise 4.13. Find the absolute maximum and minimum values of

$$f(x, y, z) = 9xyz - yx - 3(x + y)z + z$$

on the unit cube $[0, 1]^3 \subset \mathbb{R}^3$, and all points in $[0, 1]^3$ where these values occur.

Exercise 4.14. Find the points on the surface $S \subset \mathbb{R}^3$ with equation $xy+xz+yz = 1$ that are closest to the origin $(0, 0, 0)$, and explain why these give maxima of the function $f(x, y, z) = xy + xz + yz$ inside the *closed ball* whose radius is the minimum distance D from $(0, 0, 0)$ to the surface.

§ 4.4. Constrained Optimization and the method of Lagrange Multipliers

Before describing the method of Lagrange multipliers, we consider a simple constrained optimization problem: finding the highest point along a hiking trail by interpreting a map.

Example 4.10. Suppose you are looking at a trail map for a local mountain that you plan to hike. There's a trail that goes past some interesting ruins beneath a ridge-line and on the far side from summit. Since the trail does not take you to the summit, it is not automatic what point along the trail has the highest altitude. However, if the map is a contour map, then we claim that you can estimate or even know the exact point along the trail that has the highest altitude. For, as you follow the curve of the trail on the map, you periodically cross contour lines, assuming the trail is not itself a contour curve. As the trail crosses a contour, it is either ascending or descending. Once you know if the higher terrain is to your left or right, you should be able to determine whether you ascend or descend depending on how the contour and trail cross.

Exercise 4.15. If you are traveling so that higher terrain is to your left, then are you ascending or descending if the contour crosses the trail from right to left? Explain why, using the language of gradients and directional derivatives.

As we follow along, eventually and perhaps often, the trail goes from ascending to descending or descending to ascending, and correspondingly, on the map the trail goes from crossing contours left to right to crossing them right to left, or vice versa. We can argue by Rolle's theorem or the mean value theorem that there must be a critical point for the height function along the curve. But how does this relate to the geometry of the trail curve and contours on the map?

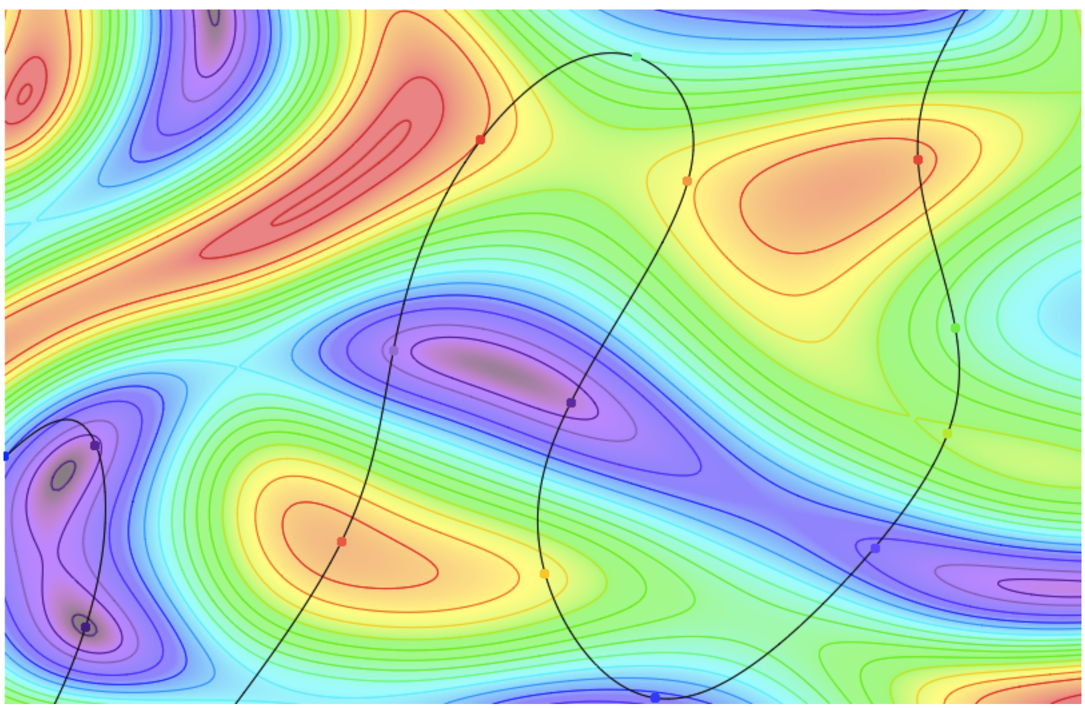


Figure 23. For continuously differentiable bivariate functions f and g, the critical points of the restriction of the height function $z = f(x, y)$ to the curve $g(x, y) = k$ occur at points where the curve $g(x, y) = k$ is tangent to some level curve of $f(x, y)$.

There are two possibilities depending on how the trail behaves and how the map is drawn. It is possible that there is a point where the trail is tangent to a contour. If this is the case, and for some stretch prior to that point, the trail is ascending, and thereafter it is descending, then clearly there is a local maximum altitude for the path attained at the point of tangency. However, the trail could be briefly tangent and then continue ascending (or descending if it initially was descending), or the map may simply not draw the level curve to which the trail is eventually tangent. However, provided the trail and contours are sufficiently smooth curves (at least differentiable), it should be clear that when the derivative of altitude changes sign there is some kind of tangency between the trail and a level curve.

Then to find the highest altitude, one simply finds the points where the trail is tangent to level curves, and then looks for the highest level curve where this happens. Note that if the curves aren't smooth, we also have to check any places where the trail has corners or cusps, or anywhere where a contour has a corner or cusp that meets the trail (even if the trail is smooth there).

This procedure can be turned into a mathematically rigorous way to find extrema of a two variable function $f(x, y)$ constrained along a curve given implicitly by $g(x, y) = k$ for some constant k. Think of $g(x, y) = k$ as describing the trail, and $z = f(x, y)$ being the altitude function. Since gradients are perpendicular to level curves, and the constraint curve $g(x, y) = k$ is just the level curve $g^{-1}(\{k\})$, we deduce that tangency points between the constraint curve and a level curve of f happen at a point $\mathbf{r}_0 = \langle x_0, y_0\rangle$ if and only if the gradients $\nabla f(\mathbf{r}_0)$ and $\nabla g(\mathbf{r}_0)$ are parallel. Thus, for some constant λ, such a point $\mathbf{r}_0$ satisfies the (nonlinear) system of equations

$$\nabla f(\mathbf{r}_0) = \lambda \nabla g(\mathbf{r}_0)\,,$$

$$g(\mathbf{r}_0) = k\,.$$

The constant λ is called the *Lagrange multiplier* associated to $\mathbf{r}_0$. Note that $\operatorname{crit} f$ is precisely the points satisfying the gradient condition when $\lambda = 0$, but it is possible that $\operatorname{crit} f$ is disjoint from the curve $g(\mathbf{r}) = k$. Different constrained critical points may correspond to different λ values.

Example 4.11. We will find the maximum and minimum values of $z = f(x, y) = x^2 + 4y^2$ along the unit circle $x^2 + y^2 = 1$. Note that we could do this via a parameterization; instead we will use $g(x, y) = x^2 + y^2 = 1$ as a constraint curve and apply the method of Lagrange multipliers. The equations we need to solve are

$$\nabla f(x, y) = \langle 2x, 8y\rangle = \lambda\langle 2x, 2y\rangle = \lambda \nabla g(x, y)\,,$$

$$x^2 + y^2 = 1\,.$$

From this we get

$$\begin{aligned} 2x &= 2\lambda x \\ 2y &= 8\lambda y \\ x^2 + y^2 &= 1\,. \end{aligned}$$

Note that $\lambda = 0$ allows the first two equations to be solved by $(0, 0)$, but this point is not on the circle $x^2 + y^2 = 1$. In fact, this corresponds to the unique solution to $\nabla f(x, y) = \mathbf{0}$; in general solutions to the Lagrangian equations in the case $\lambda = 0$ recover the critical points of f which also satisfy the constraint.

If we then assume $\lambda \neq 0$, it is clear that either $y = 0$, in which case $\lambda = 1$ and $x = \pm 1$, or $x = 0$ in which case $\lambda = 1/4$ and $y = \pm 1$.

The extrema of f along the unit circle thus happen above the axes: the maximum is $f(0, \pm 1) = 4$, and the minimum is $f(\pm 1, 0) = 1$. See figure 8 in the chain rule section (page 17) for a visualization of the function $f(x, y)$ evaluated along $x^2 + y^2 = 1$.

Exercise 4.16. Rework example 4.7 using the method Lagrange multipliers. Show that these dimensions also minimize the surface area of the open box for the fixed volume $V = 4$, again via Lagrange multipliers.

Exercise 4.17. Consider an ellipse defined by the equation $(x/a)^2 + (y/b)^2 = 1$ where $a > b > 0$ are the lengths of the semi-major and minor axes, respectively. Find the maximum area, in terms of a and b of a rectangle inscribed in the ellipse, and give the coordinates of its corners. Similarly find the maximum perimeter of an inscribed rectangle, and give the coordinates of its corners.

Exercise 4.18. Suppose you want to make an (open) cone out of paper. If you want the cone to have a volume of $4\pi/3$ then what would be the optimum radius and height to minimize the surface area of the cone? Recall that the area of an open cone with radius r and height h is $\mathcal{A}(r,h) = \pi r\sqrt{r^2+h^2}$.

Suppose we wanted to study optimization with two constraints. For example, perhaps we want to optimize a function $f(x,y,z)$ subject to constraints $g(x,y,z) = k$ and $h(x,y,z) = l$ for constants k and l. Geometrically, this corresponds to optimizing f along a curve again, this time realized as a curve of intersection of the two implicit surfaces provided by the constraints $g(x,y,z) = k$ and $h(x,y,z) = l$. At an optimum (either maximum or minimum) along the curve, the curve will be tangent to *a level surface* of f. But then the gradient of f must be perpendicular to the curve. But then it follows that the gradient of f is a linear combination of the gradients of g and h, which are also perpendicular to this curve of intersection. Thus, the two constraint Lagrangian equations are

$$\begin{aligned}\nabla f(x,y,z) &= \lambda\nabla g(x,y,z) + \mu\nabla h(x,y,z)\,,\\ g(x,y,z) &= k\,,\\ h(x,y,z) &= l\,,\end{aligned}$$

where λ and μ are both Lagrange multipliers.

Example 4.12. Consider the curve of intersection of the cylinder $x^2 + y^2 = 5$ and the plane $6x - 3y + 2z = 5$. Find the maximum straight line distance from $(0,0,0)$ to the curve, and give the points along the curve where this distance occurs.

We can let $f(\mathbf{r}) = \mathbf{r}\cdot\mathbf{r}$ be the square distance, for if we maximize distance we also maximize its square. Our constraints are the cylinder equation $g(x,y,z) = x^2 + y^2 = 5$ and the plane equation $h(x,y,z) = 6x - 3y + 2z = 5$. The Lagrangian equations are then

$$\begin{aligned}2x &= 2\lambda x + 6\mu\\ 2y &= 2\lambda y - 3\mu\,,\\ 2z &= 2\mu\,,\\ 5 &= x^2 + y^2\\ 5 &= 6x - 3y + 2z\,.\end{aligned}$$

The third equation tells us that $z = \mu$, whence from the first two equations we have $(1-\lambda)x = 3z = 2(\lambda - 1)y$. Then either $\lambda = 1$ or $x = -2y$. Note that $\lambda = 1$ then requires $z = 0$, which gives $x = (5+3y)/6$ from the equation of the plane. The two points we get from plugging this into the cylinder's equation correspond to the minimum square distance of 5 (we leave it as an exercise to find these two points and show this). For the maximum square distance, we then look at the case where $x = -2y$. Substituting into the cylinder equation first gives $4y^2 + y^2 = 5 \implies y = \pm 1$, whence $x = \mp 2$. Substituting these into the plane equation gives

$$h(\mp 2, \pm 1, z) = \mp 15 + 2z = 5 \implies z = 10 \text{ or } -5\,.$$

The corresponding square distances are

$$f(-2,1,10) = 105\,, \text{ and } f(2,-1,-5) = 30\,.$$

Thus the maximum distance from $(0,0,0)$ to the curve is $\sqrt{105}$, which occurs at the point $(-2,1,10)$.

Exercise 4.19. Find the minimum value of $f(x, y, z) = x^2 + 4y^2 + 9z^2$ on the intersection of the hyperboloids $4x^2 + y^2 - 9z^2 = 1$ and $9x^2 - 4y^2 - z^2 = 1$. Explain why there is no maximum value of f along this intersection locus.

Exercise 4.20. For the cylinders $x^2 + y^2 = 1$ and $y^2 + z^2 = 4/9$, find the minimum positive value of the x coordinate along the intersection curve of the cylinders, and locate all the points where this value occurs. Set up and solve the problem using the method of Lagrange multipliers with two constraints.

Note that without loss of generality, we may assume constraints have the form $g(\mathbf{r}) = 0$, as we may always arrange the equations of a constraint set with all terms on one side. The next theorem rephrases the idea of Lagrange multipliers for multiple constraints in the language of optimizing a single function, called a *Lagrangian.*

Theorem 4.3. *Let $f(\mathbf{r})$ be a differentiable multivariate function defined on a domain $\mathcal{D} \subseteq \mathbb{R}^n$, and suppose $g_1, \dots g_k$ are differentiable functions on $\mathcal{D}$ determining a set of $k < n$ constraint equations $\{g_i(\mathbf{r}) = 0\}$. Let $\Lambda : D \times \mathbb{R}^k \to \mathbb{R}$ be the* Lagrangian function *given by*

$$\Lambda(\mathbf{r}, \boldsymbol{\lambda}) = f(\mathbf{r}) - \boldsymbol{\lambda} \cdot \mathbf{G}(\mathbf{r})\,,$$

where $\mathbf{G}(\mathbf{r}) = \langle g_1(\mathbf{r}), \dots, g_k(\mathbf{r})\rangle$.

Then the absolute maximum and minimum values of $f(\mathbf{r})$ subject to the constraints $\{g_i(\mathbf{r}) = 0\}$, assuming they exist, occur at points $\mathbf{r}$ corresponding to points $(\mathbf{r}, \boldsymbol{\lambda}) \in \mathrm{crit}\,(\Lambda)$ such that the extreme values of $\Lambda(\mathbf{r}, \boldsymbol{\lambda})$ give the extreme values of $f(\mathbf{r})$. For a critical point $(\mathbf{r}_0, \boldsymbol{\lambda}_0) \in \mathrm{crit}\,(\Lambda)$, the vector $\boldsymbol{\lambda}_0$ gives the k Lagrange multipliers $\lambda_{i,0}$, $i = 1, \dots, k$, such that $\nabla f(\mathbf{r}_0) = \sum_{i=1}^k \lambda_{i,0}\, g_i(\mathbf{r}_0)$ holds.

Exercise 4.21. Prove the above theorem. Hint: first argue that the gradient of f must be a linear combination of the gradients of g, as we did above in the case of three variables and two constraints. Then show that the critical points of the Lagrangian correspond to the Lagrange multiplier equations and constraint equations, and that the corresponding values correspond to the constrained local extrema of f. In particular, you should be able to argue that there is a one-to-one correspondence between critical points of Λ and the collection of $\mathbf{r}$ such that $\mathbf{r}$ either solves the Lagrange multipliers system and constraint equations, or $\mathbf{r} \in \mathrm{crit}\, f$.

Exercise 4.22. Under what conditions on f, $\mathcal{D}$, and the constraints $\{g_i(\mathbf{r}) = 0\}$ can we infer the existence of maximum and minimum solutions to the constrained optimization problem? (Hint: consider the theorem above and answer the corresponding question about existence of absolute extrema for the Lagrangian function.)

5. Further Problems

Exercises 5.1**,** 5.2**, and** 5.3 **are cross-posted from the the notes on *Curvature, Natural Frames, and Acceleration for Plane and Space Curves*) and rely on definitions in those notes.**

Exercise 5.1. Let $f(\mathbf{r})$ be a differentiable function defined on a domain $\mathcal{D} \subseteq \mathbb{R}^2$. By $f(r,\theta)$ we mean f evaluated at the point with position $\mathbf{r} = r\hat{\mathbf{u}}_r(\theta) = r\cos(\theta)\hat{\imath} + r\sin(\theta)\hat{\jmath}$. Express the gradient of f in polar coordinates, meaning, describe the operator ∇ in terms of $\hat{\mathbf{u}}_r$ and $\hat{\mathbf{u}}_\theta$ by giving functions $u(r,\theta)$ and $v(r,\theta)$ such that

$$\nabla = u(r,\theta)\frac{\partial}{\partial r}\hat{\mathbf{u}}_r + v(r,\theta)\frac{\partial}{\partial \theta}\hat{\mathbf{u}}_\theta$$

and so that $\nabla f(r(x,y),\theta(x,y)) = \frac{\partial f}{\partial x}\hat{\imath} + \frac{\partial f}{\partial y}\hat{\jmath}$ for all points $\big(r(x,y),\theta(x,y)\big)_{\mathscr{P}} = (x,y)_{\mathscr{C}}$ in $\mathcal{D}$.

Exercise 5.2. Express the gradient operator in spherical coordinates (see the previous problem for the two dimensional, polar version of this problem.)

Exercise 5.3. Compute the gradients of the coordinate functions for spherical coordinates, i.e. compute $\nabla\varrho$, $\nabla\theta$ and $\nabla\varphi$. Express the answers in both the spherical frame and the rectangular frame.

Exercise 5.4. Recall the spherical coordinate system described in the notes *Curvature, Natural Frames, and Acceleration for Plane and Space Curves* (see pages 11-15). The transformation from the rectangular coordinates $(x,y,z)_{\mathscr{R}}$ on $\mathbb{R}^3$ to these spherical coordinates $(\varrho,\theta,\phi)_{\mathscr{S}}$ was given as

$$x = \varrho\cos\theta\cos\varphi\,,\quad y = \varrho\sin\theta\cos\varphi\,,\quad z = \varrho\sin\varphi\,,$$

where $\varrho \in [0,\infty)$, $\theta \in (-\pi,\pi]$, and $\varphi \in [-\pi/2,\pi/2]$.

(*a*) Compute the Jacobian matrices

$$\mathbf{D}_{\varrho,\theta,\varphi}\mathbf{G} = \frac{\partial(x,y,z)}{\partial(\varrho,\theta,\varphi)}\,,\quad \mathbf{D}_{x,y,z}\mathbf{X} = \frac{\partial(\varrho,\theta,\varphi)}{\partial(x,y,z)},$$

and verify that these matrices are inverses. (Note that they are 3×3 matrices.)

(*b*) Express the chain rule for a scalar function $f\big(x(\varrho,\theta,\varphi), y(\varrho,\theta,\varphi), z(\varrho,\theta,\varphi)\big)$ with respect to the spherical variables, using the Jacobians computed above.

(*c*) Use the chain rule from part (*b*) to compute the partials f_ϱ, f_θ and f_φ, where

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

(*d*) For f the function in part (*c*), compute $\frac{\mathrm{d}f}{\mathrm{d}t}$ along the curve

$$\mathbf{r}(t) = \big(2+\cos(3t)\big)\hat{\mathbf{u}}_r(2t) + \sin(3t)\hat{\mathbf{k}}\,,$$

where

$$\hat{\mathbf{u}}_r(2t) = \cos(2t)\hat{\imath} + \sin(2t)\hat{\jmath} = \frac{\big([x(t)]^2-[y(t)]^2\big)\hat{\imath} + 2x(t)y(t)\hat{\jmath}}{[x(t)]^2+[y(t)]^2}\,.$$

Exercise 5.5. This problem explores partial derivatives and directional derivatives of *multivariable vector-valued functions*. Let $\mathbf{v} : \mathcal{D} \to \mathbb{R}^n$ be a multivariable vector function over a domain $\mathcal{D}$ in $\mathbb{R}^m$. E.g. for a 2-dimensional vector-valued function from $\mathcal{D} \subset \mathbb{R}^2$ one has

$$\mathbf{v}(x, y) = v_1(x, y)\hat{\mathbf{\imath}} + v_2(x, y)\hat{\mathbf{\jmath}}\,,$$

where $v_i(x, y)$, $i = 1, 2$ are two-variable functions from $\mathcal{D}$ to $\mathbb{R}$. In this 2-dimensional case we define

$$\frac{\partial \mathbf{v}}{\partial x} := \frac{\partial v_1}{\partial x}\hat{\mathbf{\imath}} + \frac{\partial v_2}{\partial x}\hat{\mathbf{\jmath}}\,,$$

and analogously

$$\frac{\partial \mathbf{v}}{\partial y} := \frac{\partial v_1}{\partial y}\hat{\mathbf{\imath}} + \frac{\partial v_2}{\partial y}\hat{\mathbf{\jmath}}\,.$$

One can also define a notion of directional derivative of a vector-valued function along a unit vector: given $\mathbf{v} : \mathcal{D} \to \mathbb{R}^n$ and a unit vector $\hat{\mathbf{u}} \in \mathbb{R}^m$, define

$$D_{\hat{\mathbf{u}}}\mathbf{v}(\mathbf{x}) = \lim_{h\to 0} \tfrac{1}{h}\left[\mathbf{v}(\mathbf{x} + h\hat{\mathbf{u}}) - \mathbf{v}(\mathbf{x})\right]\,.$$

One can show that this can be calculated as

$$D_{\hat{\mathbf{u}}}\mathbf{v}(\mathbf{x}) = \sum_{i=1}^{n} \hat{\mathbf{u}} \cdot \nabla v_i(\mathbf{x})\hat{\mathbf{e}}_i = \big(\hat{\mathbf{u}} \cdot \nabla\big)\mathbf{v}(\mathbf{x})\,.$$

Here, $\hat{\mathbf{e}}_i$ are the coordinate basis vectors in $\mathbb{R}^n$, which are analogous to $\hat{\mathbf{\imath}}$, $\hat{\mathbf{\jmath}}$, and $\hat{\mathbf{k}}$ (namely, $\hat{\mathbf{e}}_i$ has entries equal to 0 for all coordinates other than the ith coordinate, which equals 1.)

Parts (*a*)–(*c*) focus on polar coordinates and two dimensions. Part (*d*) works in three dimensions, with spherical coordinates.

(*a*) Let $\hat{\mathbf{u}}_r(x, y) = \frac{1}{r}(x\hat{\mathbf{\imath}} + y\hat{\mathbf{\jmath}})$ and $\hat{\mathbf{u}}_\theta(x, y) = \frac{1}{r}(-y\hat{\mathbf{\imath}} + x\hat{\mathbf{\jmath}})$ where $r^2 = x^2 + y^2$, as in the treatment of the polar frame in the notes on *Curvature, Natural Frames, and Acceleration for Plane and Space Curves*. Compute all first and second partials with respect to x and y of $\hat{\mathbf{u}}_r$ and $\hat{\mathbf{u}}_\theta$.

(*b*) Justify the two dimensional case of the above formula for the directional derivative of a vector-valued function along a unit vector, i.e. use the limit definition of the directional derivative above to show that

$$D_{\hat{\mathbf{u}}}\mathbf{v}(x, y) = \Big(\hat{\mathbf{u}} \cdot \nabla v_1(x, y)\Big)\hat{\mathbf{\imath}} + \Big(\hat{\mathbf{u}} \cdot \nabla v_2(x, y)\Big)\hat{\mathbf{\jmath}}\,.$$

(*c*) Compute $D_{\hat{\mathbf{\imath}}}\hat{\mathbf{u}}_r$, $D_{\hat{\mathbf{\jmath}}}\hat{\mathbf{u}}_r$, $D_{\hat{\mathbf{\imath}}}\hat{\mathbf{u}}_\theta$, $D_{\hat{\mathbf{\jmath}}}\hat{\mathbf{u}}_\theta$, $D_{\hat{\mathbf{u}}_r}\hat{\mathbf{u}}_\theta$ and $D_{\hat{\mathbf{u}}_\theta}\hat{\mathbf{u}}_r$.

(*d*) Let $\big(\hat{\mathbf{u}}_\varrho, \hat{\mathbf{u}}_\theta, \hat{\mathbf{u}}_\varphi\big)$ be the spherical frame (as in the notes *Curvature, Natural Frames, and Acceleration for Plane and Space Curves*). Give $\hat{\mathbf{u}}_\varrho$, $\hat{\mathbf{u}}_\theta$, and $\hat{\mathbf{u}}_\varphi$ as vector-valued functions of x, y and z (the rectangular coordinates on $\mathbb{R}^3$, and compute $D_{\hat{\mathbf{\imath}}}\hat{\mathbf{u}}_\varrho$, $D_{\hat{\mathbf{\jmath}}}\hat{\mathbf{u}}_\varrho$, $D_{\hat{\mathbf{k}}}\hat{\mathbf{u}}_\varrho$, $D_{\hat{\mathbf{\imath}}}\hat{\mathbf{u}}_\theta$, $D_{\hat{\mathbf{\jmath}}}\hat{\mathbf{u}}_\theta$, $D_{\hat{\mathbf{k}}}\hat{\mathbf{u}}_\varphi$, $D_{\hat{\mathbf{\imath}}}\hat{\mathbf{u}}_\varphi$, $D_{\hat{\mathbf{\jmath}}}\hat{\mathbf{u}}_\varphi$, $D_{\hat{\mathbf{k}}}\hat{\mathbf{u}}_\varphi$, $D_{\hat{\mathbf{u}}_\varrho}\hat{\mathbf{u}}_\theta$, $D_{\hat{\mathbf{u}}_\varrho}\hat{\mathbf{u}}_\varphi$, $D_{\hat{\mathbf{u}}_\theta}\hat{\mathbf{u}}_\varrho$, $D_{\hat{\mathbf{u}}_\theta}\hat{\mathbf{u}}_\varphi$, $D_{\hat{\mathbf{u}}_\varphi}\hat{\mathbf{u}}_\varrho$, and $D_{\hat{\mathbf{u}}_\varphi}\hat{\mathbf{u}}_\varphi$.

(Part (*d*) is only recommended for a certain sort of student, who *really* enjoys/needs to take lots of partial derivatives, and finds it soothing to do so.)