1 Graphs of Scalar Functions

Domain and range of scalar functions, visualizing and describing the surface $z = f(x, y)$ in 3D space using level curves and cross sections, generalizing level curves and cross sections to higher dimension.

1.1 Scalar Functions

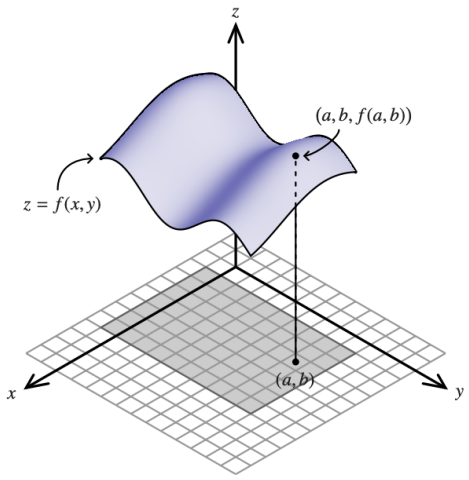
A **function** $f : A \to B$ associates with each element $a \in A$ a unique element $f(a) \in B$ called the **image** of a under f. The set A is the **domain** of f and is denoted $D(f)$, the set B is the **codomain** of f. The subset of B consisting of all $f(a)$ is the **range** of f and is denoted $R(f)$. A **scalar function** $f(x_1, \ldots, x_n)$ of n variables is a function whose domain is a subset of $\mathbb{R}^n$ and whose range is a subset of $\mathbb{R}$.

For example, f which maps points (x, y) in $\mathbb{R}^2$ to a real scalar $f(x, y) \in \mathbb{R}$, and is often denoted $z = f(x, y)$. We also often use $f(x, y)$ to refer to the function with independent variables x, y rather than the value of the function f at the point (x, y). It is convenient to view points in $\mathbb{R}^n$ as vectors $\vec{x}$ in $\mathbb{R}^n$ to make use of results from linear algebra. Note,

1. To determine the range of a scalar function, try to see if the function can take on any positive value in $\mathbb{R}$ by setting all the variables but one to dummy values which simplify the function (such as 0). Then, do the same for negative values in $\mathbb{R}$ and also for 0.
2. To determine the range, we might also try completing the square. For example, $x^2 + bx = x^2 + bx + \frac{b^2}{4} - \frac{b^2}{4} = (x + \frac{b}{2})^2 - \frac{b^2}{4}$ then solve the resulting inequality.
3. Recall open sets from CO 342. If we find that the domain of $f(x, y)$ is x, y such that $x^2 + y^2 < 1$, then the domain is an open circle of radius 1 centered at the origin. We can sketch the domain in 2D space, since $f(x, y)$ is a scalar function whose domain is a subset of $\mathbb{R}^2$.

1.2 Geometric Interpretation

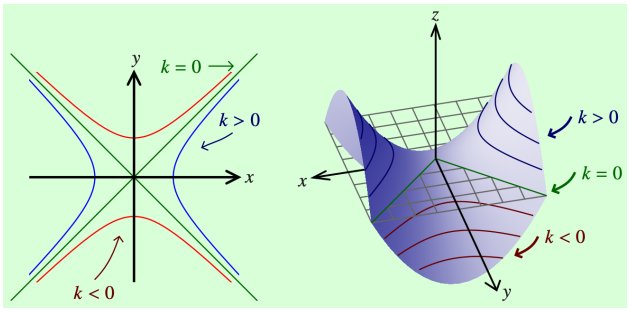
The **graph** of a function $f(x, y)$ is the set of all points $(a, b, f(a, b))$ in $\mathbb{R}^3$ such that $(a, b) \in D(f)$. We can think of $f(a, b)$ as representing the height of the graph $z = f(x, y)$ above (or below if negative) the xy plane at the point $(x, , y) = (a, b)$. This definition generalizes for scalar functions whose domain is a subset of $\mathbb{R}^n$.



In general, when $f(x,y) = c_1x + c_2y + c_3$, where c_1, c_2, c_3 are real constants, the graph $z = f(x,y)$ is a plane in $\mathbb{R}^3$. However, $z = f(x,y)$ can be quite complicated.

The **level curves** of a function $f(x,y)$ are the curves $f(x,y) = k$ where k is a constant in the range of f. That is, for each k, we have a separate level curve. If $f(x,y) = 2x - 3y + 1$, then $R(f) = \mathbb{R}$, so the level curves are $2x - 3y + 1 = k$ for $k \in \mathbb{R}$, which gives a family of parallel lines. Observe that $f(x,y) = k$ is the intersection of $z = f(x,y)$ in $\mathbb{R}^3$ and the horizontal plane $z = k$ in $\mathbb{R}^3$. Thus, in our family of level curves, each value of k represents the height of that level curve above the xy plane. For example, the level curve from $f(x,y) = 1$ gives the intersection of $z = f(x,y)$ and $z = 3$ in $\mathbb{R}^3$, which is height 3 above the xy plane in $\mathbb{R}^3$. The family of level curves is often called a contour map or a topographic map. A level curve which is just a single point is an **exceptional level curve**.

From the level curves of f, we can use them to sketch the surface $z = f(x,y)$. Take each level curve $f(x,y) = k$ and draw it in the $z = k$ plane in $\mathbb{R}^3$ (the plane k above the xy plane). The elevation of the earth's surface above sea level is described by an equation $z = h(x,y)$ where x is the latitude and y is the longitude of the position. A contour map shows the curves of constant elevation $h(x,y) = k$ which are precisely the level curves of h. As k increases, if the level curves get closer to each other, we can say that the surface is getting steeper. To determine the "shape" of each level curve, we can try plotting for various values of k and taking a guess. For most simple functions, the curves will generally all be one of: lines, circles, parabolas, hyperbolas, ellipses. When the level curves are hyperbolas, the surface is likely a saddle surface. When the level curves are straight lines, the surface might be a parabolic cylinder.



A **cross section** of a surface $z = f(x,y)$ is the intersection of $z = f(x,y)$ with a vertical plane. For sketching the graph of a surface $z = f(x,y)$, we consider the family of cross sections $f(c,y)$ and $f(x,d)$ formed by intersecting the surface with the vertical planes $x = c$ (the yz plane at $x = c$) and $y = d$ (the xz plane at $y = d$), for constants c, d. Thus, we draw curves in the yz plane by setting $x = c$ and curves in the xz plane by setting $y = d$. Note that cross sections for f with the planes $x = c$ might be different from cross sections for f with the planes $y = d$ (for example, parabolas vs lines). Choose c, d from points in the domain.

We can generalize the idea of level curves for functions of more than two variables. Introducing an extra variable introduces a new dimension. A **level surface** of a scalar function $f(x,y,z)$ is defined by $f(x,y,z) = k$ for $k \in R(f)$. A **level set** of a scalar function $f(\vec{x}), \vec{x} \in \mathbb{R}^n$ is defined by $\{\vec{x} \in \mathbb{R}^n | f(\vec{x}) = k, \text{ for } k \in R(f)\}$. For example, if $f(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2$, the level sets of $f(\vec{x}) = k$ in $\mathbb{R}^n$ for $k > 0$ are called $(n-1)$-spheres, denoted S^{n-1}.

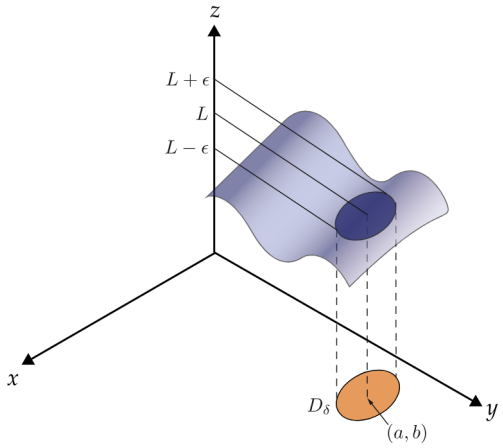
Hiker question: For the hiker question of the unit 1 Mobius assignment, the hiker is walking across cross sections since he is "climbing" up and down the mountain, thus the part he is walking is alone an intersection of the mountain with a vertical plane. If he were walking around a mountain, then it would be level curves. The equation of the line the hiker is walking until he reaches the mountain is obtained by solving for b in $y = x + b$ using his starting position. The equation of the cross section is obtained by plugging in $y = x + b$ into the equation of the mountain curve, giving a function of x. Then, the x value of the max elevation can be found by equating the derivative of the equation of the cross section (wrt x) to 0. Plug this value back into the equation of the cross section to find the max elevation. Note the max elevation reached by the hiker is not necessarily the highest point in the entire mountain.

2 Limits

Limits of scalar functions and their properties, finding a limit if it exists or proving a limit does not exist.

An **open interval** is defined as $(-r, r) = \{x : |x| < r\}$ where $r \in \mathbb{R}$. An $r-$**neighbourhood** of a point $(a, b) \in \mathbb{R}^2$ is a set $N_r(a, b) = \{(x, y) \in \mathbb{R}^2 | ||(x, y) - (a, b)|| < r\}$. Recall that $||(x, y) - (a, b)||$ is the Euclidean distance in $\mathbb{R}^2$.

Recall that $\forall \epsilon > 0, \exists \delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $0 < |x - a| < \delta$ and $\lim_{x\to a} f(x) = L$ iff $\lim_{x\to a^-} f(x) = L = \lim_{x\to a^+} f(x)$. Similarly, for scalar functions $f(x, y)$, assume $f(x, y)$ is defined in a neighbourhood of (a, b), except possibly at (a, b). Then, if $\forall \epsilon > 0, \exists \delta > 0$ such that $0 < ||(x, y) - (a, b)|| < \delta$ implies $|f(x, y) - L| < \epsilon$, then $\lim_{(x,y)\to(a,b)} f(x, y) = L$. This limit definition refers only to the distance between (x, y) and (a, b) and there is no notion of approaching a limit from a certain direction, since we can approach from infinitely many directions.



2.1 Basic Theorems

If $\lim_{(x,y)\to(a,b)} f(x, y)$ and $\lim_{(x,y)\to(a,b)} g(x, y)$ both exist, then,

1. **Addition**: $\lim_{(x,y)\to(a,b)}[f(x, y) + g(x, y)] = \lim_{(x,y)\to(a,b)} f(x, y) + \lim_{(x,y)\to(a,b)} g(x, y)$
2. **Multiplication**: $\lim_{(x,y)\to(a,b)}[f(x, y)g(x, y)] = [\lim_{(x,y)\to(a,b)} f(x, y)][\lim_{(x,y)\to(a,b)} g(x, y)]$
3. **Division**: $\lim_{(x,y)\to(a,b)} \frac{f(x,y)}{g(x,y)} = \frac{\lim_{(x,y)\to(a,b)} f(x,y)}{\lim_{(x,y)\to(a,b)} g(x,y)}$, provided $\lim_{(x,y)\to(a,b)} g(x, y) \neq 0$

Proof of 1: Let $\epsilon > 0, \lim_{(x,y)\to(a,b)} f(x, y) = L_1, \lim_{(x,y)\to(a,b)} g(x, y) = L_2$. Then, since $\frac{\epsilon}{2} > 0$, $\exists \delta > 0$ s.t.,

$$0 < ||(x, y) - (a, b)|| < \delta \implies |f(x, y) - L_1| < \frac{\epsilon}{2} \text{ and } |f(x, y) - L_2| < \frac{\epsilon}{2}$$

Thus, $|f(x, y) + g(x, y) - (L_1 + L_2)| \leq |f(x, y) - L_1| + |g(x, y) - L_2| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ by triangle inequality.

Theorem: If $\lim_{(x,y)\to(a,b)} f(x, y)$ exists, then the limit is unique. This can be easily proved using the addition rule above, taking $g(x, y) = -f(x, y)$.

2.2 Proving a Limit Does Not Exist

Recall to show a limit does not exist for a function of one variable, we showed that the left-sided limit did not equal the right-sided limit, or the limit from one of the two sides did not even exist (eg, endpoints). For multivariable functions, we show the limit does not exist by showing that the limit when approaching along two distinct smooth curves are different. For example, take the limit along the line $y = x$ and the limit along the line $y = 0$ and the limit does not exist if these two values do not match. We can approach the limit along infinitely many lines/smooth curves simultaneously by introducing an arbitrary coefficient m. If the value of the limit depends on m, then it is not unique and hence, the limit does not exist. For example, approach the limit along lines $y = mx$. This leads to $\lim_{(x,y)\to(a,b)} f(x, y) = \lim_{x\to a} f(x, mx)$ and to evaluate this, we can consider the two sided limit at a, like in Math 137. However, even if the limit produced by this result does not depend on m, this does not mean the limit exists, because the line $x = a$ also goes through (a, b) but is not included in $y = mx$, and the limit if we approach along $x = a$ might give a different result. So, we try $\lim_{(x,y)\to(a,b)} f(x, y) = \lim_{(x,y)\to(a,b)} f(a, y) = \lim_{y\to b} f(a, y)$. However, even if all lines give the same result, this does not mean the limit exists and we might want to try curves next such as $y = mx^2$. Make sure all lines and curves actually approach the limit. Testing the line $y = 1$ for a limit that approaches $(0, 0)$ is pointless. Also, this method of testing different lines and curves cannot show that a limit exists, since there is an infinite number of ways to approach the limit and we cannot test that they all match.

2.3 Proving a Limit Exists

Squeeze Theorem: If there exists a function $B(x, y)$ such that $|f(x, y) - L| \leq B(x, y)$ for all $(x, y) \neq (a, b)$ in some neighbourhood of (a, b) and $\lim_{(x,y)\to(a,b)} B(x, y) = 0$, then $\lim_{(x,y)\to(a,b)} f(x, y) = L$.

Proof: Let $\epsilon > 0$. Then, $\exists\delta > 0$ such that if $0 < ||(x, y) - (a, b)|| < \delta$, then $|B(x, y)| < \epsilon$. Thus, $|f(x, y) - L| \leq B(x, y) = |B(x, y)| < \epsilon$, so by the definition of a limit, $\lim_{(x,y)\to(a,b)} f(x, y) = L$. Note that instead of using two functions to lower and upper bound $f(x, y)$, we instead upper bound $|f(x, y) - L|$. Also, note that we need a candidate L for this to work (or, it would be given when asked to prove the limit equals a certain L). When working with fractions, it is often useful to manipulate the numerator to match the denominator (or, match up to some constant factor), so that they cancel out and we are left with a nice function for $B(x, y)$ which we can evaluate the limit for directly.

General approach: When asked to determine whether $\lim_{(x,y)\to(a,b)} f(x, y)$ exists and if so, find its value, first try direct substitution if $f(x, y)$ is continuous at (a, b). Then, try to factor and cancel if $f(x, y)$ is continuous at (a, b). Then, try approaching along some combination of the paths $x = a, y = b, y = x - a + b$ to try to prove that the limit does not exist. Then, try approaching along $y = b + m(x - a)^k$ for some clever choices of m, k. If we are still stuck, we have a good candidate for the Squeeze Theorem now (the value found when approaching from the various curves). If this still does not work, we try: epsilon-delta proof, rearranging terms, etc.

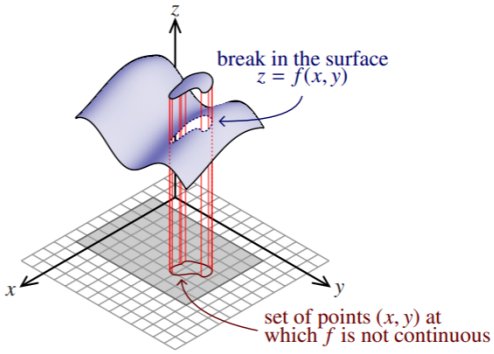
Advice: Since there is no L'Hopital's rule for multivariate functions, it can be difficult to evaluate some limits which are of the form $\frac{0}{0}$, etc. A clever choice of $B(x, y)$ is to make it a univariate function which we can apply L'Hopital's rule on. For example, to prove that $\lim_{(x,y)\to(0,0)} x\ln\left(x^2 + 2y^2\right) = 0$, we can notice that the closer the value is to 0 (and less than 1), the larger the absolute value of its ln. For example, $|\ln(0.01)| > |\ln(0.1)|$. Thus, $|x\ln\left(x^2 + 2y^2\right)| \leq |x\ln\left(x^2\right)|$. We now take $B(x, y) = x\ln\left(x^2\right)$. However, since it is not a function of y, $\lim_{(x,y)\to(0,0)} B(x, y) = \lim_{x\to 0} x\ln\left(x^2\right)$ and we can make use of the normal tricks from Math 137.

3 Continuous Functions

Determining continuity using basic and composite functions, together with various continuity theorems.

Recall that a function of one variable $f(x)$ is continuous at $x = a$ iff: f is defined at $x = a$, $\lim_{x\to a} f(x)$ exists ($\lim_{x\to a^-} f(x)$ and $\lim_{x\to a^+} f(x)$ both exist and are equal), and $\lim_{x\to a} f(x) = f(a)$.

For a function of two variables, $f(x, y)$ is **continuous** at (a, b) iff $\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b)$. There are three requirements: $\lim_{(x,y)\to(a,b)} f(x, y)$ exists, f is defined at (a, b), and $\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b)$. To check continuity, we must check all three of these requirements and to disprove continuity, we show at least one of these three fail. An easy way to do this is to find a curve such that the limit if we approach along this curve is not $f(a, b)$. If f is continuous at every point in a set $D \subseteq \mathbb{R}^2$, then we say f is continuous on D. Visually, a discontinuous surface $z = f(x, y)$ has breaks/holes,



3.1 Continuity Theorems

We may assume the following basic functions are continuous on their domains,

1. $f(x, y) = k$ for some constant k
2. $f(x, y) = x^n, f(x, y) = y^n$
3. $\ln(\cdot)$
4. $e^{(\cdot)}$
5. Trigonometric functions $\sin(\cdot), \cos(\cdot)$, etc.
6. Inverse trigonometric functions $\arcsin(\cdot)$, etc.
7. Absolute value $|\cdot|$
8. Coordinate functions $f(x, y) = x, f(x, y) = y$

If $f(x, y), g(x, y)$ are scalar functions and $(x, y) \in D(f) \cap D(g)$, then:

1. The sum $f + g$ is defined by $(f + g)(x, y) = f(x, y) + g(x, y)$
2. The product fg is defined by $(fg)(x, y) = f(x, y)g(x, y)$
3. The quotient $\frac{f}{g}$ is defined by $(\frac{f}{g})(x, y) = \frac{f(x,y)}{g(x,y)}$, if $g(x, y) \neq 0$

For scalar functions $g(t), f(x,y)$, the **composite function** $g \circ f$ is defined by $(g \circ f)(x,y) = g(f(x,y))$ for all $(x,y) \in D(f)$ for which $f(x,y) \in D(g)$. When composing multivariable functions, it is very important to make sure the range of the inner function is a subset of the domain of the outer function. For example, if $f(x,y) = \sin(x+y)$ and $h(t) = \frac{1}{t}$, the composition $(h \circ f)$ is not defined because 0 is in the range of f but not in the domain of h.

We have the following continuity theorems,

1. Theorem 1: If f, g are both continuous at (a,b), then $f+g$ and fg are both continuous at (a,b)
2. Theorem 2: If f, g are both continuous at (a,b) and $g(a,b) \neq 0$, then $\frac{f}{g}$ is continuous at (a,b)
3. Theorem 3: If $f(x,y)$ is continuous at (a,b) and $g(t)$ is continuous at $f(a,b)$, then $g \circ f$ is continuous at (a,b)

The first two theorems are easily proven using the limit theorems and the third can be proven using delta-epsilon.

Now, we may prove continuity using composition of basic functions. For example, to show that $f(x,y) = \sin\big(6x^2y\big)$ is continuous for all $(x,y) \in \mathbb{R}^2$, we can apply Theorem 1 to the constant function and power functions to show that $g(x,y) = 6x^2y$ is continuous for all $(x,y) \in \mathbb{R}^2$, then by Theorem 3, $f(x,y)$ is continuous for all $(x,y) \in \mathbb{R}^2$. Now, we can prove continuity "by inspection". However, sometimes the continuity theorems do not apply and we have to go back to the definition of continuity to determine whether or not the function is continuous.

Overall, we can easily evaluate $\lim_{(x,y)\to(a,b)} f(x,y)$ if f is continuous at (a,b) (which can be proven using the continuity theorems) by evaluating $f(a,b)$. This is useful for applying the Squeeze Theorem, since we need to prove that $\lim_{(x,y)\to(a,b)} B(x,y) = 0$ and we want to ideally evaluate this limit by inspection, so we try to set up the inequality so that $B(x,y)$ is continuous at (a,b).

4 The Linear Approximation

Generalizing linear approximations to functions of n variables.

4.1 Partial Derivatives

A scalar function $f(x, y)$ can be naturally differentiated in two ways:

1. Treating y as a constant and differentiating with respect to x to obtain $\frac{\partial f}{\partial x}$
2. Treating x as a constant and differentiating with respect to y to obtain $\frac{\partial f}{\partial y}$

These two derivatives are called the (first) **partial derivatives** of f. Formally, provided the limits exist,

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

Sometimes, it is convenient to use operator notation $D_1 f, D_2 f$ for the partial derivatives, where $D_1 f$ means to differentiate f with respect to the variable in the first position, holding the other fixed. So, if the independent variables are x, y, then $D_1 f = \frac{\partial f}{\partial x} = f_x$ (and it is also common to use $\frac{\partial f}{\partial x}$ to mean $\frac{\partial f}{\partial x}(x, y)$. To determine whether $\frac{\partial f}{\partial x}(a, b)$ exists, first find the partial derivative then see if you can plug in (a, b) without any problems. If there are problems, then try the $\lim_{h\to 0}$ first principles definition. If the function changes definition at (a, b) (for example, piece wise), then we must use the first principles definition as well, since we are not able to find the partial derivative directly. If we get an invalid value (such as $\pm\infty$), then we conclude $\frac{\partial f}{\partial x}(a, b)$ does not exist.

As well, note that partial derivatives are not defined at boundaries. For example, if $x \in [-1, 1]$ and $y \in [-1, 1]$ is the domain, then $f_x(-1, 1)$ is undefined. This is because from the first principles limit definition, since the limit of h from 0^- does not exist, then the two sided limit $h \to 0$ does not exist.

For functions of one variable, differentiability implies continuity. Although this is also true for multivariable functions as seen later, it is possible for partial derivatives to exist at (a, b) but not have continuity or differentiability at (a, b). For example,

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

has $f_x(0, 0) = f_y(0, 0) = 0$ but $f(x, y)$ is not continuous at $(0, 0)$ and this is because the limit does not exist (approach along $y = mx$). We will see later that indeed differentiability does imply continuity even for multivariable functions however partial derivatives does not imply differentiability. Also, there is no L'Hopital's rule for multivariable functions.

Generalization: For a scalar function of n variables $f(\vec{x})$ where $\vec{x} \in \mathbb{R}^n$, to take the partial derivative of f with respect to its ith variable, we hold all the other variables constant and differentiate with respect to the ith variable. The first principles definition is analogous. Eg, for $f(x, y, z)$, then $f_y = \lim_{h\to 0} \frac{f(x,y+h,z)-f(x,y,z)}{h}$.

4.2 Higher-Order Partial Derivatives

Since the partial derivatives of a scalar function of two variables are also scalar functions of two variables, we can take the partial derivatives of the partial derivatives of any scalar function. There are then four possible **second partial derivatives** of f for $f(x, y)$ which are:

1. $\frac{\partial^2 f}{\partial x^2} = f_{xx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$: differentiate $\frac{\partial f}{\partial x}$ with respect to x with y fixed
2. $\frac{\partial^2 f}{\partial y \partial x} = f_{xy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$: differentiate $\frac{\partial f}{\partial x}$ with respect to y with x fixed
3. $\frac{\partial^2 f}{\partial x \partial y} = f_{yx} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$: differentiate $\frac{\partial f}{\partial y}$ with respect to x with y fixed
4. $\frac{\partial^2 f}{\partial y^2} = f_{yy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$: differentiate $\frac{\partial f}{\partial y}$ with respect to y with x fixed

It is often convenient to use the operator notation: $D_1^2 f = f_{xx}, D_2 D_1 f = f_{xy}, D_1 D_2 f = f_{yx}, D_2^2 f = f_{yy}$. The **Hessian matrix** of $f(x, y)$, denoted by $Hf(x, y)$ is defined as

$$Hf(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

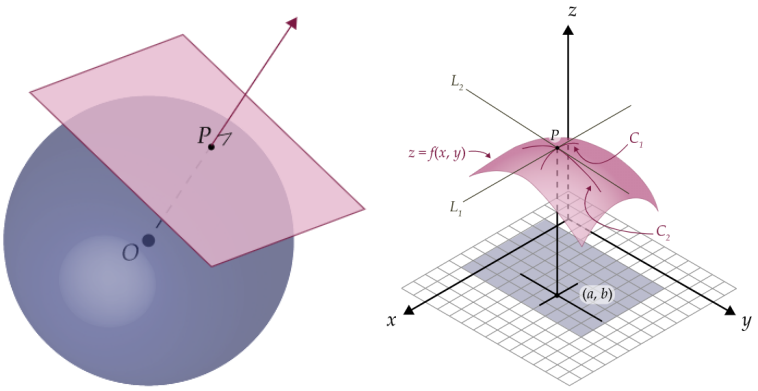
Clairaut's Theorem: If f_{xy} and f_{yx} are defined in some neighborhood of (a, b) and are both continuous at (a, b), then $f_{xy}(a, b) = f_{yx}(a, b)$. This can be generalized to functions of n variables, where the second partial derivatives satisfy $f_{ij}(x_1, \ldots, x_n) = f_{ji}(x_1, \ldots, x_n)$ for all pairs i, j. If the second partial derivatives are defined in some neighbourhood and continuous, then the Hessian matrix is symmetric but otherwise, is not necessarily symmetric.

We can take higher-order partial derivatives in the expected way. In particular, $f(x, y)$ has eight third partial derivatives: f_{xxx}, f_{xxy}, etc. Clairaut's Theorem also extends to higher-order partial derivatives: if the higher-order partial derivatives are defined in a neighborhood of a point (a, b) and are continuous at (a, b), then they are equal regardless of the order in which they are taken. For example, if the partial derivatives of f satisfy Clairaut's Theorem, then $f_{xxy}(a, b) = f_{xyx}(a, b) = f_{yyx}(a, b)$.

If the kth partial derivatives of $f(x_1, \ldots, x_n)$ are continuous, then we write $f \in C^k$ (f is in class C^k). So, $f(x, y) \in C^2$ means f has continuous second partial derivatives, and therefore by Clairaut's Theorem, $f_{xy} = f_{yx}$.

4.3 The Tangent Plane

The surface of a sphere has a tangent plane at each point P, namely, the plane through P that is orthogonal to the line joining P and the center O. The tangent plane at P can be thought of as the plane which best approximates the surface of the sphere near P.



This can be generalized to a surface defined by an equation $z = f(x, y)$. Let C_1 be the cross section $y = b$ of the surface, that is, C_1 is given by $z = f(x, b)$. Then, $\frac{\partial f}{\partial x}(a, b)$ is the slope of the tangent line L_1 of C_1

at the point $P(a, b, f(a, b))$. Similarly, let C_2 be the cross section $x = a$ of the surface, that is, C_2 is given by $z = f(a, y)$. Then, $\frac{\partial f}{\partial y}(a, b)$ is the slope of the tangent line L_2 of C_2 at the point $P(a, b, f(a, b))$. This is illustrated above. Thus, the tangent plane to the surface $z = f(x, y)$ at the point $P(a, b, f(a, b))$ is the plane containing the tangent lines L_1, L_2.

Recall from Math 136 the equation of a plane through the point (x_1, y_1, z_1) whose normal vector is $\vec{n} = (a, b, c)$ is $a(x - x_1) + b(y - y_1) + c(z - z_1) = 0$. So, any (non-vertical) plane through the point $P(a, b, f(a, b))$ has an equation of the form $z = f(a, b) + m(x - a) + n(y - b)$ where m, n are constants. The intercept of this plane with the vertical plane $y = b$ is the line $z = f(a, b) + m(x - a)$. We require this line coincide with L_1, so the slope m of the line must equal the slope $\frac{\partial f}{\partial x}(a, b)$ of the line L_1. Similarly, n must equal the slope $\frac{\partial f}{\partial y}(a, b)$ of L_2. Thus, the tangent plane to $z = f(x, y)$ at the point $(a, b, f(a, b))$ is,

$$z = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$$

Overall, this equation was derived from solving for m, n, knowing from above that the tangent plane must contain the tangent lines L_1, L_2 which involve the partial derivatives of x, y respectively. To see if the tangent plane passes through a point (x_1, y_1, z_1), we can just plug these values into the equation of the tangent plane and see if it resolves.

4.4 Linear Approximation

For the 1D case and function $f(x)$, the tangent line can be used to approximate the graph of the function near the point of tangency. Recall that the equation of the tangent line to $y = f(x)$ at the point $(a, f(a))$ is $y = f(a) + f'(a)(x - a)$. The function L_a defined by $L_a(x) = f(a) + f'(a)(x - a)$ is the **linearization** of f at a since $L_z(x) \approx f(x)$ for x sufficiently close to a (this is called the **linear approximation** of f at a).

For the 2D case and the multivariable differentiable function $f(x, y)$, the tangent plane can be used to approximate the surface $z = f(x, y)$ near the point of tangency $P(a, b, f(a, b))$.

For a function $f(x, y)$, the linearization $L_{(a,b)}(x, y)$ of f at (a, b) is defined as the equation of the tangent plane at P,

$$L_{(a,b)}(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$$

The approximation $f(x, y) \approx L_{(a,b)}(x, y)$ is the linear approximation of $f(x, y)$ at (a, b). As we consider a smaller neighbourhood of the point of tangency P, the surface $z = f(x, y)$ and $(L_{(a,b)}(x, y)$ will become very close, just as in the 1D case.

An example of the application is to approximate $\sqrt{(0.95)^3 + (1.98)^3}$ by choosing a function (in this case, $f(x, y) = \sqrt{x^3 + y^3}$) and point of tangency (in this case, $(a, b) = (1, 2)$). We choose nice numbers close to $(0.95, 1.98)$ so that the calculation is easy. Then, we use $L_{(1,2)}(x, y)$ to approximate $\sqrt{(0.95)^3 + (1.98)^3}$.

Next, suppose we know $f(a, b)$ and want to calculate $f(x, y)$ at a nearby point. Let $\Delta x = x - a, \Delta y = y - b, \Delta f = f(x, y) - f(a, b)$. Then, the linear approximation is $f(x, y) \approx f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$. Rearranging,

$$\Delta f \approx \frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b)\Delta y$$

This equation is called the **increment form** of the linear approximation and gives an approximation for Δf in $f(x, y)$ due to a change $(\Delta x, \Delta y)$ from the point (a, b). An example of its usage is to estimate the change in the area of an isosceles triangle, given the changes to the base height (the one unique length) and the equal angles (the two angles which are equal). We can consider the area a multivariable function, then use the increment form above.

4.5 Linear Approximation in Higher Dimensions

Consider the $\mathbb{R}^3$ and a function $f(x, y, z)$. Similar to the case of a function of two variables, define the linearization of f at $\vec{a} = (a, b, c)$ by,

$$L_{\vec{a}}(x, y, z) = f(\vec{a}) + f_x(\vec{a})(x - a) + f_y(\vec{a})(y - b) + f_z(\vec{a})(z - c)$$

The three terms can be represented by the dot product of the vectors $(x - a, y - b, z - c) = (x, y, c) - (a, b, c)$ and $\nabla f(\vec{a}) = (f_x(\vec{a}), f_y(\vec{a}), f_z(\vec{a}))$. The vector $\nabla f(\vec{a})$ of partial derivatives is the **gradient** of f at $\vec{a}$. This is assuming that $f(x, y, z)$ has partial derivatives at $\vec{a} \in \mathbb{R}^3$.

Suppose that $f(\vec{x}), \vec{x} \in \mathbb{R}^3$ has partial derivatives at $\vec{a} \in \mathbb{R}^3$, the linearization of f at $\vec{a}$ is defined by,

$$L_{\vec{a}}(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a})$$

And the linear approximation of f at $\vec{a}$ is,

$$f(\vec{x}) \approx f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a})$$

An example of usage is to estimate $4.99 \times 7.01 \times 9.99$ by choosing $f(x, y, z) = xyz$ and $\vec{a} = (5, 7, 10)$.

The two above equations generalize for a function of n variables $f(\vec{x}), \vec{x} \in \mathbb{R}^n$. For an arbitrary $\vec{a} \in \mathbb{R}^n$, we have $\Delta\vec{x} = \vec{x} - \vec{a} = (x_1 - a_1, \ldots, x_n - a_n)$ and the gradient of f at $\vec{a}$ is defined to be $\nabla f(\vec{a}) = (D_1 f(\vec{a}), \ldots, D_n f(\vec{a}))$. Then, the increment form of the linear approximation for $f(\vec{x})$ is,

$$\Delta f \approx \nabla f(\vec{a}) \cdot \Delta\vec{x}$$

Observe this works even when $n = 1$. For a function $g(t)$ of one variable, $\nabla g(a) = g'(a)$ and the increment form of the linear approximation is $\Delta g \approx \nabla g(a) \cdot \Delta x = g'(a)(x - a)$. For a function of two variables $f(x, y)$, we have $\nabla f(a, b) = (f_x(a, b), f_y(a, b))$ and the increment form of the linear approximation is $\Delta f \approx \nabla f(a, b) \cdot \Delta(x, y) = f_x(a, b)(x - a) + f_y(a, b)(y - b)$, the same as what we derived earlier.

5 Differentiable Functions

Extending the concept of differentiation to multivariable functions.

5.1 Differentiability

For single variable functions, $g(x)$ is differentiable at $x = a$ if $g'(a)$ exists. However, for multi variable functions, partial derivatives are necessary but not sufficient for differentiability. Informally, $g(x)$ is differentiable at $x = a$ if the graph of $g(x)$ has no cusps or jumps at $x = a$ and the linear approximation is a good approximation. Let $R_{1,a}(x) = g(x) - L_a(x) = g(x) - g(a) - g'(a)(x - a)$ be the error in the linear approximation. Then, we can say that if $g'(a)$ exists, then $\lim_{x\to a} \frac{|R_{1,a}(x)|}{|x-a|} = 0$.

This statement says that the error $R_{1,a}(x)$ tends to zero faster than the displacement $|x - a|$ and the tangent line is the unique line that satisfies this property. If we replace the tangent line $y = L_a(x)$ by any other straight line $y = g(a) + m(x - a)$ passing through the point $(a, g(a))$, then error will not satisfy the conclusion of the theorem. So, $\lim_{x\to a} \frac{|R_{1,a}(x)|}{|x-a|} = 0$ characterizes the tangent line at $(a, g(a))$ as the best straight line approximation to the graph $y = g(x)$ near $(a, g(a))$.

To define differentiability for a multivariable function $f(x, y)$, we match this definition. First, let the error in the linear approximation be $R_{1,(a,b)}(x, y) = f(x, y) - L_{(a,b)}(x, y)$. Then, $f(x, y)$ is **differentiable** at (a, b) iff,

$$\lim_{(x,y)\to(a,b)} \frac{|R_{1,(a,b)}(x, y)|}{||(x, y) - (a, b)||} = 0$$

where $||(x, y)-(a, b)||= \sqrt{(x - a)^2 + (y - b)^2}$. Moreover, the only tangent plane $z = f(a, b)+c(x-a)+d(y-b)$ through the point $(a, b, f(a, b))$ that has this property is $z = L_{(a,b)}(x, y)$. So, if $f(x, y)$ satisfies

$$\lim_{(x,y)\to(a,b)} \frac{|f(x, y) - f(a, b) - c(x - a) - d(y - b)|}{||(x, y) - (a, b)||} = 0$$

for some constants c, d, then $c = f_x(a, b), d = f_y(a, b)$. The proof for this follows by approaching the limit along $y = b$ and $x = a$, both limits which should give 0. This implies that the tangent plane gives the best linear approximation to the graph $z = f(x, y)$ near (a, b) and the linear approximation is a good approximation iff f is differentiable at (a, b). This is why partial derivatives are necessary for differentiability (but not sufficient). For example, consider $f(x, y) = \sqrt{|xy|}$ at $(0, 0)$. Both partial derivatives at $(0, 0)$ exist but the above limit does not evaluate to 0.

We now have a formal definition of the **tangent plane**. Consider a function $f(x, y)$ that is differentiable at (a, b). The tangent plane of the surface $z = f(x, y)$ at $(a, b, f(a, b))$ is the graph of the linearization. That is, the tangent plane is given by,

$$z = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$$

And, at the point $(a, b, f(a, b))$, the surface $z = f(x, y)$ is **smooth**.

5.2 Differentiability and Continuity

Recall that for a single variable function, if $g(x)$ is differentiable at $x = a$, then g is continuous at a. Similarly, **differentiability implies continuity** for multivariable functions as well; if $f(x, y)$ is differentiable at (a, b),

then f is continuous at (a,b) (but the converse is not true, as seen earlier). We can prove this by re-writing $R_{1,(a,b)}(x,y) = f(x,y) - L_{(a,b)}(x,y)$ as,

$$f(x,y) = f(a,b) + \nabla f(a,b) \cdot (x-a, y-b) + R_{1,(a,b)}(x,y)$$

And for $(x,y) \neq (a,b)$,

$$R_{1,(a,b)}(x,y) = \frac{R_{1,(a,b)}(x,y)}{||(x,y)-(a,b)||}||(x,y)-(a,b)||$$

Since f is differentiable, then $\lim_{(x,y)\to(a,b)} R_{1,(a,b)}(x,y) = 0$, so it follows from the first equation that $\lim_{(x,y)\to(a,b)} f(x,y) = f(a,b)$ and by definition, f is continuous at (a,b).

Theorem: If the partial derivatives $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ are both continuous at (a,b), then $f(x,y)$ is differentiable at (a,b). However, the converse is not true. This lets us easily prove a function is differentiable at a certain point.

Proof: Since f_x, f_y are continuous, then f_x, f_y exist in some neighbourhood $B(a,b)$. For $(x,y) \in B(a,b)$, we can write,

$$f(x,y) - f(a,b) = f(x,y) - f(a,y) + f(a,y) - f(a,b)$$

Next, recall from single variable calculus the **Mean Value Theorem**: If $f(x)$ is continuous on the closed interval $[x_1, x_2]$ and f is differentiable on the open interval (x_1, x_2), then there exists $x_0 \in (x_1, x_2)$ such that $f(x_1) - f(x_1) = f'(x_0)(x_2 - x_1)$. So, $\exists \bar{x}$ between a and x such that $f(x,y) - f(a,y) = f_x(\bar{x}, y)(x-a)$. We can apply MVT here because y is held fixed and the partial derivatives are assumed to exist. Let $A = f_x(\bar{x}, y) - f_x(a,b)$. Then, we can re-write as,

$$f(x,y) - f(a,y) = f_x(a,b)(x-a) + A(x-a)$$

Similarly, there is a $\bar{y}$ between b and y, and if $B = f_y(a,\bar{y}) - f_y(a,b)$,

$$f(a,y) - f(a,b) = f_y(a,b)(y-b) + B(y-b)$$

Substituting these two equations in to the first and re-arranging gives,

$$R_{1,(a,b)}(x,y) = A(x-a) + B(y-b)$$

By the triangle inequality,

$$\frac{|R_{1,(a,b)}(x,y)|}{||(x,y)-(a,b)||} \leq \frac{|A||x-a|}{\sqrt{(x-a)^2+(y-b)^2}} + \frac{|B||y-b|}{\sqrt{(x-a)^2+(y-b)^2}} \leq |A|+|B|$$

Note that as $(x,y) \to (a,b)$, then $(\bar{x}, y) \to (a,b)$ and $(a, \bar{y}) \to (a,b)$. Since f_x, f_y are continuous at (a,b), then from the equations for A, B, $\lim_{(x,y)\to(a,b)} A = \lim_{(x,y)\to(a,b)} B = 0$. Therefore, by the Squeeze Theorem with $L = 0, B(x,y) = |A|+|B|$, we have

$$\lim_{(x,y)\to(a,b)} \frac{|R_{1,(a,b)}(x,y)|}{||(x,y)-(a,b)||} = 0$$

so f is differentiable at (a,b) by definition.

From this result, we know now that if $f(x,y) \in C^2$ at (a,b) (recall $f(x,y) \in C^k$ means that f has continuous kth partial derivatives and that the mixed higher-order partial derivatives are equal regardless of the order in which they are taken by Clairaut's Theorem), then:

1. f_x, f_y are differentiable at (a,b) since the partial derivatives of f_x (namely, f_{xx}, f_{xy} are continuous at (a,b)) and similarly for f_y

2. f is differentiable at (a, b) since the partial derivatives are both continuous at (a, b)
3. f, f_x, f_y are all continuous at (a, b) since differentiability implies continuity

We also now have an interesting chain of implications between differentiability and continuity.

1. If f_x, f_y are continuous at (a, b), then f is differentiable at (a, b), which then implies that f is continuous at (a, b).
2. If f_x, f_y are continuous at (a, b), then $f_x(a, b)$ and $f_y(a, b)$ exist and we also know that if f is differentiable at (a, b), then by definition, $f_x(a, b), f_y(a, b)$ must exist.
3. If f is not continuous at (a, b), then f_x, f_y cannot be both continuous at (a, b).

Generalization: For scalar functions of n variables, a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at a point $\vec{a} = (a_1, \ldots, a_n)$ iff,

$$\lim_{\vec{x} \to \vec{a}} \frac{|f(\vec{x}) - f(\vec{a}) - L_{\vec{a}}|(\vec{x} - \vec{a})}{||\vec{x} - \vec{a}||} = 0$$

where $L : \mathbb{R}^n \to \mathbb{R}$ is a linear transformation. And, if $f(x_1, \ldots, x_n)$ is differentiable at $\vec{a} = (a_1, \ldots, a_n)$, then f is continuous at $\vec{a}$. Lastly, if $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ are continuous at $\vec{a} = (a_1, \ldots, a_n)$, then $f(x_1, \ldots, x_n)$ is differentiable at $\vec{a}$.

5.3 Linear Approximation Revisited

For $f(x, y) = f(a, b) + \nabla f(a, b) \cdot (x - a, y - b) + R_{1,(a,b)}(x, y)$, we see that if the error term is small, then $f(x, y) \approx f(a, b) + \nabla f(a, b) \cdot (x - a, y - b)$. In general, we have no information about $R_{1,(a,b)}(x, y)$ and it is unclear whether the approximation is reasonable. However, we know now that if the partial derivatives of f are continuous at (a, b), then f is differentiable and hence $\lim_{(x,y)\to(a,b)} \frac{|R_{1,(a,b)}(x,y)|}{||(x,y)-(a,b)||} = 0$.

In this case, the approximation is reasonable for (x, y) sufficiently close to (a, b) and $L_{(a,b)}(x, y)$ is a **good approximation** of $f(x, y)$ near (a, b). Approximation is a recurring theme in calculus and the following equation is of fundamental importance,

$$f(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}) + R_{1,\vec{a}}(\vec{x})$$

6 The Chain Rule

Chain rule generalized for multivariable functions and composite functions.

6.1 Basic Chain Rule

Recall in single-variable calculus, the chain rule was used to take derivatives of composite functions: if $T = f(x)$ and $x(t)$ are differentiable functions, then the derivative of the composition $T(t) = f(x(t))$ is, $T'(t) = f'(x(t))x'(t)$. In Leibniz form, $\frac{dT}{dt} = \frac{dT}{dx}\frac{dx}{dt}$, where the first T is used as a composite of t and the second T is used as a function of x (abuse of notation).

Let $T = f(x, y)$ and $x = x(t), y = y(t)$. Informally, a derivation of the chain rule to find the derivative of $T(t) = f(x(t), y(y))$ is to note that in a time change Δt, x, y change by $\Delta x = x(t + \Delta t) - x(t)$ and $\Delta y = y(t + \Delta t) - y(t)$. By the increment form of the linear approximation, the change in T corresponding to changes $\Delta x, \Delta y$ is approximated by $\Delta t \approx \frac{\partial T}{\partial x}\Delta x + \frac{\partial T}{\partial y}\Delta y$ for $\Delta x, \Delta y$ sufficiently small. Dividing by Δt and letting $\Delta t \to 0$, then use the definition of the derivative to get $\frac{dT}{dt}$ on the left side. Assuming that $T(x, y)$ is differentiable at (x, y), then as $\Delta x, \Delta y \to 0$, the error in the linear approximation tends to 0 so the approximation becomes increasingly accurate, leading to,

$$\frac{dT}{dt} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt}$$

This is again an abuse of notation. T is used as a composite function of t on the left and as a function of x, y on the right.

Chain Rule: Without an abuse of notation, let $G(t) = f(x(t), y(t))$ and $a = x(t_0), b = y(t_0)$. If f is differentiable at (a, b) and $x'(t_0), y'(t_0)$ exist, then $G'(t_0)$ exists and is given by,

$$G'(t_0) = f_x(a, b)x'(t_0) + f_y(a, b)y'(t_0)$$

We require f is differentiable. A weaker hypothesis of $f_x(a, b), f_y(a, b)$ existing is not enough. For example, let $f(x, y) = (xy)^{\frac{1}{3}}, x(t) = t, y(t) = t^2, G(t) = f(x(t), y(t))$. f is not differentiable at $(0, 0)$ and the Chain Rule fails (the above equality does not hold), even though both f_x, f_y are defined. However, sometimes it is useful to use stronger hypotheses in the Chain Rule. For example, assuming that f has continuous partial derivatives at (a, b) and $x'(t), y'(t)$ are both continuous at t_0. Then, we can obtain a stronger conclusion that $G'(t)$ is continuous at t_0.

Proof: By definition, provided the limit exists, $\star$ $G'(t_0) = \lim_{t\to t_0} \frac{G(t)-G(t_0)}{t-t_0}$. By definition of $G(t)$,

$$G(t) - G(t_0) = f(x(t), y(t)) - f(x(t_0), y(t_0))$$

Since f is differentiable,

$$f(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) + R_{1,(a,b)}(x, y)$$

where $\lim_{(x,y)\to(a,b)} \frac{|R_{1,(a,b)}(x,y)|}{\sqrt{(x-a)^2+(y-b)^2}} = 0$, Since $a = x(t_0), b = y(t_0)$, it follows that,

$$\star\star \quad \frac{G(t) - G(t_0)}{t - t_0} = f_x(a, b)\frac{x(t) - x(t_0)}{t - t_0} + f_y(a, b)\frac{y(t) - y(t_0)}{t - t_0} + \frac{R_{1,(a,b)}(x(t), y(t))}{t - t_0}$$

Next, define,

$$E(x, y) = \begin{cases} \frac{R_{1,(a,b)}(x,y)}{\sqrt{(x-a)^2+(y-b)^2}} & ; \text{if } (x, y) \neq (a, b) \\ 0 & ; \text{if } (x, y) = (a, b) \end{cases}$$

Since $\lim_{(x,y)\to(a,b)} \frac{|R_{1,(a,b)}(x,y)|}{\sqrt{(x-a)^2+(y-b)^2}} = 0 = E(a,b)$, E is continuous at (a,b). As well, for all (x,y)

$$R_{1,(a,b)}(x,y) = E(x,y)\sqrt{(x-a)^2+(y-b)^2}$$

Since $a = x(t_0), b = y(t_0)$,

$$\frac{|R_{1,(a,b)}(x(t),y(t))}{|t-t_0|} = |E(x(t),y(t))|\sqrt{\left(\frac{x(t)-x(t_0)}{t-t_0}\right)^2+\left(\frac{y(t)-y(t_0)}{t-t_0}\right)^2}$$

Since $x'(t_0), y'(t_0)$ exist and E is continuous at (a,b) and $E(a,b) = 0$,

$$\lim_{t\to t_0} \frac{|R_{1,(a,b)}(x(t),y(t))}{|t-t_0|} = E(x(t_0),y(t_0))\sqrt{(x'(t_0))^2+(y'(t_0))^2} = 0$$

Therefore, it follows from $\star$ and $\star\star$ that $G'(t_0)$ exists and is given by the desired Chain Rule formula.

We can use the dot product to rewrite the Chain Rule into a vector form. In particular, if $T(t) = f(x(t),y(t))$ where $f(x,y), x(t), y(t)$ are differentiable, then,

$$\frac{dT}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)\cdot\left(\frac{dx}{dt}, \frac{dy}{dt}\right) = \nabla f \cdot \frac{d\vec{x}}{dt}$$

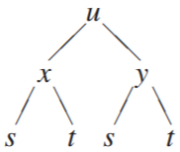
So, $\frac{dT}{dt}(t) = \frac{d}{dt}f(\vec{x}(t)) = \nabla f(\vec{x}(t))\cdot\frac{d\vec{x}}{dt}(t)$ where $\vec{x}(t) = (x(t),y(t))$. In this vector form, the Chain Rule holds for any differentiable function $f(\vec{x}), \vec{x}\in\mathbb{R}^n$. For example, $T = f(x,y,z)$ representing temperature or some other quantity in 3-space.

6.2 Extensions of the Basic Chain Rule

Suppose $x = x(s,t), y = y(s,t)$ have first order partial derivatives at (s,t) and let $u = f(x,y)$, where f is differentiable at $(x,y) = (x(s,t),y(s,t))$. Then, u is a composite function of two independent variables s,t and we can write a chain rule for each of,

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s}$$
$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t}$$

This relationship can be easily visualized using a tree representing the chain of dependence,



We call s,t the **independent variables**, x,y are the **intermediate variables** and u is the **dependent variable**. Logically, the chain rule makes sense from the point of view of rate of change because from the tree, we see u is dependent on x,y which are each dependent on s,t. So, the rate of change of u with respect to s is the sum of the rate of change with respect to its x and y components. The term $\frac{\partial u}{\partial x}\frac{\partial x}{\partial t}$ calculates the rate of change of u with respect to t that affects u through x. As well, $\frac{\partial u}{\partial x}$ means: regard u as the given function of x,y and differentiate with respect to x, holding y fixed. And, $\frac{\partial u}{\partial s}$ means: regard u as the composite function of s,t and differentiate with respect to s, holding t fixed.

A more precise form of the Chain Rule which displays the functional dependence is: suppose g is the composite function of $f(x, y)$ and $x(s, t), y(s, t)$. That is, $g(s, t) = f(x(s, t), y(s, t))$. Then,

$$\frac{\partial g}{\partial s}(s,t) = \frac{\partial f}{\partial x}(x(s,t), y(s,t))\frac{\partial x}{\partial s}(s,t) + \frac{\partial f}{\partial y}(x(s,t), y(s,t))\frac{\partial y}{\partial s}(s,t)$$

The general algorithm to write the Chain Rule from a tree dependence diagram is,

1. Take all possible paths from the differentiated variable to the differentiating variable
2. For each link $-$ between two "nodes" in the tree in a given path, differentiate the "parent" variable with respect to the "child" variable, being careful to consider if this is a derivative (functions of one variable) or partial derivative (multivariable). For example, if u is the parent variable and v is the child variable, we want either $\frac{du}{dv}$ or $\frac{\partial u}{\partial v}$. Then, multiply all such derivatives in that path.
3. Add the products from step 2 together to complete the Chain Rule.

For notation, to specify which variables are held constant, we can write it in the subscript. For example, if differentiating z with respect to t with u, v constant, we can write $\left(\frac{\partial z}{\partial t}\right)_{u,v}$.

Generalization: The Chain Rule can be generalized to functions having any number of independent variables: Let $g = f(x_1, \ldots, x_m)$ be a differentiable function of m independent variables and for $1 \leq i \leq m$, let $x_i = x_i(t_1, \ldots, t_n)$ be a differentiable function of n independent variables. Then for $1 \leq j \leq n$,

$$\frac{\partial g}{\partial t_j} = \frac{\partial g}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial g}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \cdots + \frac{\partial g}{\partial x_m}\frac{\partial x_m}{\partial t_j}$$

6.3 Chain Rule for Second Partial Derivatives

Often, it is useful to calculate second derivatives of composite functions using the Chain Rule. For example, when working with partial differential equations which involve second derivatives (eg Laplace's equation $u_{xx} + u_{yy} = 0$. Instead of requiring that the function is differentiable, we now require it is twice differentiable (or in certain cases, a stronger assumption, that the function is in C^2; the second order partial derivatives not only exist but are continuous).

The tree for chain of dependence can also be drawn to compute the second order partial derivatives from the first partial derivatives. The same tree is used, however, if the top (root) variable of the original tree were z, then the root variable of this tree will be z' (or equivalently, a partial derivative in the multivariable case).

See course notes for examples.

---

7 Directional Derivatives and the Gradient Vector

Generalizing partial derivatives by considering the rate of change in a direction specified by an arbitrary unit vector (directional derivatives).

7.1 Directional Derivatives

The partial derivative f_x can be thought of as the rate of change of f in the x-direction. Similarly, f_y can be thought of as the rate of change of f in the y-direction. We wish to define a derivative which gives the rate of change of f in a direction specified by the unit vector $\vec{u} = (u_1, u_2)$ (ie, $||\vec{u}|| = 1$) from a given point (a, b). If L is the line passing through (a, b) in the direction $\vec{u}$, then L has vector equation $(x, y) = (a, b) + s\vec{u} = (a + su_1, b + su_2)$ for $s \in \mathbb{R}$. At points on the line L, $f(x, y)$ has value $f(a + su_1, b + su_2)$ and this defines a function of one variable s

Directional Derivative: The directional derivative of $f(x, y)$ at (a, b) in the direction of a unit vector $\vec{u} = (u_1, u_2)$ is,

$$D_{\vec{u}} f(a, b) = \frac{d}{ds} f(a + su_1, b + su_2)\Big|_{s=0}$$

That is, the rate of change of f at (a, b) in the direction of $\vec{u}$ is the derivative of $f(a + su_1, b + su_2)$ with respect to s evaluated at $s = 0$.

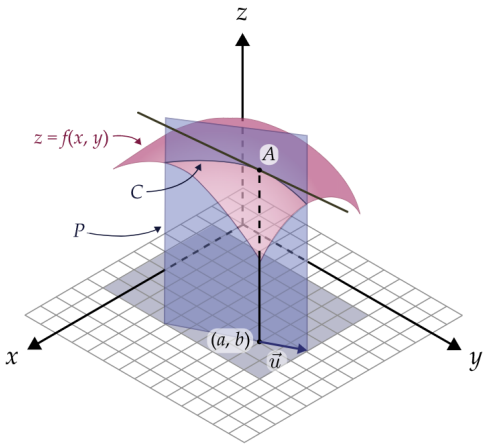
Directional Derivative (DD) Theorem: If $f(x, y)$ is differentiable at (a, b) and $\vec{u} = (u_1, u_2)$ is a unit vector, then $D_{\vec{u}} f(a, b) = \nabla f(a, b) \cdot \vec{u}$ (dot product).

Proof: Since f is differentiable at (a, b), we can apply the chain rule to get,

$$\begin{aligned}
D_{\vec{u}} f(a, b) &= \frac{d}{ds} f(a + su_1, b + su_2)\Big|_{s=0} \\
&= \left[D_1 f(a + su_1, b + su_2) \frac{d}{ds}(a + su_1) + D_2 f(a + su_1, b + su_2) \frac{d}{ds}(b + su_2) \right]\Big|_{s=0} \\
&= [D_1 f(a + su_1, b + su_2)u_1 + D_2 f(a + su_1, b + su_2)u_2]|_{s=0} \\
&= D_1 f(a, b) + D_2 f(a, b)u_2 \\
&= \nabla f(a, b) \cdot (u_1, u_2)
\end{aligned}$$

When using the DD Theorem, if the given vector is not a unit vector, it must first be normalized. If f is not differentiable at (a, b), we must apply the definition of the directional derivative to find it. This DD Theorem is only a convenient shortcut for f differentiable at (a, b). Also, note that if $\vec{u}$ is $(1, 0)$ or $(0, 1)$, then the directional derivative is equal to the partial derivatives f_x or f_y. The definition of the directional derivative and DD Theorem is also easily generalized to higher dimensions.

Geometric Representation: The directional derivative represents a rate of change with respect to distance s along the line L. Geometrically, it is the slope of the tangent to the cross-section C at the point A (the vertical plane P cuts the surface $z = f(x, y)$ along the curve C),



7.2 Gradient Vector in Two Dimensions

In general, the directional derivative has infinitely many values corresponding to all possible directions $\vec{u}$ at (a, b). Using the dot product property $\vec{x} \cdot \vec{y} = ||\vec{x}||||\vec{y}||\cos\theta$, where θ is the angle between $\vec{x}, \vec{y}$, we can find the direction in which the directional derivative is maximized.

Greatest Rate of Change (GRC) Theorem: If $f(x, y)$ is differentiable at (a, b) and $\nabla f(a, b) \neq (0, 0)$, then the largest value of $D_{\vec{u}} f(a, b)$ is $||\nabla f(a, b)||$ and occurs when $\vec{u}$ is in the direction of $\nabla f(a, b)$.

Proof: By the DD Theorem, since f is differentiable at (a, b) and $\vec{u}$ is a unit vector,

$$D_{\vec{u}} f(a, b) = \nabla f(a, b) \cdot \vec{u} = ||\nabla f(a, b)||||\vec{u}||\cos\theta = ||\nabla f(a, b)||\cos\theta$$

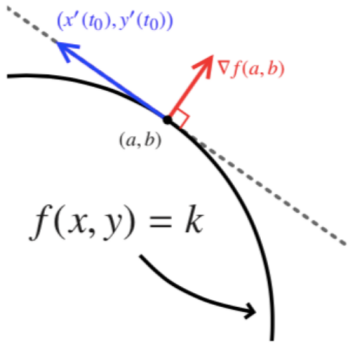
This is maximized when $\cos\theta = 1$, ie $\theta = 0$. So, the largest value of $D_{\vec{u}} f(a, b)$ is $||\nabla f(a, b)||$ and occurs when $\vec{u}$ is in the direction of the gradient vector $\nabla f(a, b)$. We can also note this is minimized when $\cos\theta = -1$, so the minimal value of the directional derivative is in the direction of $-\nabla f(a, b)$. That is why for gradient descent, we go in the direction of $-\nabla f(\vec{a})$ at each step to find the local minimum. Lastly, the directional derivative is 0 when $\cos\theta = 0$, when $\vec{u}$ is perpendicular to $\nabla f(a, b)$.

The GRC Theorem also applies to any dimension. That is, if $f(\vec{x}), \vec{x} \in \mathbb{R}^n$ is differentiable at $\vec{a}$ and $\vec{u} \in \mathbb{R}^n$ is a unit vector, then the largest value of $D_{\vec{u}} f(\vec{a})$ is $||\nabla f(\vec{a})||$ and occurs when $\vec{u}$ is in the direction of $\nabla f(\vec{a})$.

Orthogonality Theorem: If $f(x, y) \in C^1$ in a neighbourhood of (a, b) and $\nabla f(a, b) \neq (0, 0)$, then $\nabla f(a, b)$ is orthogonal to the level curve $f(x, y) = k$ through (a, b).

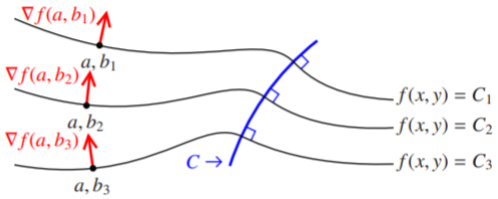
Proof: Since $\nabla f(a, b) \neq (0, 0)$, by the Implicit Function Theorem, the level curve $f(x, y) = k$ can be described by parametric equations $x = x(t), y = y(t)$ for $t \in I$ for an interval I where $x(t), y(t)$ are differentiable. Hence, the level curve may be written as $f(x(t), y(t)) = k$, $t \in I$. Suppose $a = x(t_0), b = y(t_0)$ for some $t_0 \in I$. Since f is differentiable, we can take the derivative of this equation with respect to t using the Chain Rule to get,

$$f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t) = 0$$



Setting $t = t_0$, $\nabla f(a,b) \cdot (x'(t_0), y'(t_0)) = 0$. Thus, $\nabla f(a,b)$ is orthogonal to $(x'(t_0), y'(t_0))$ which is tangent to the level curve. In words, the direction of the gradient of f is orthogonal to the level curves of f. One use of this theorem is to prove that the level curves of functions f, g intersect orthogonally by showing that $\nabla f \cdot \nabla g = 0$.

The gradient of f at (x,y), $\nabla f(x,y) = (f_x(x,y), f_y(x,y))$ associates a vector with each point of the domain of f and is referred to as a **vector field**. It is represented graphically by drawing $\nabla f(a,b)$ as a vector emanating from the corresponding point (a,b). The vector field has important geometric properties: it gives the direction in which the function has its largest rate of change and also gives the direction that is orthogonal to the level curves of the function.



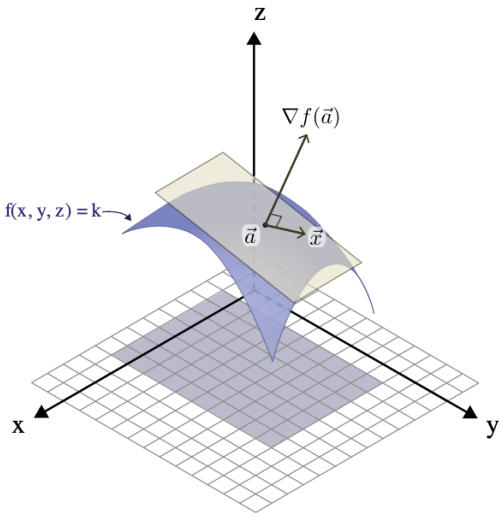
If the level curves are contour lines, then a curve such as C, which intersects the level curves orthogonally, would define a curve of steepest ascent on the surface. Note that if we know all the gradient vectors are in the same direction for any (x,y), then by the Orthogonality Theorem, since ∇f is always orthogonal to the level curves of f, then the level curves are of the form parallel straight lines.

7.3 Gradient Vector in Three Dimensions

Consider the level surfaces in $\mathbb{R}^3$ for the function $w = f(x,y,z)$, $f(x,y,z) = k$ where $k \in R(f)$. By the GRC Theorem, $\nabla f(a,b,c)$ gives the direction of the largest rate of change of f. And, the Orthogonality Theorem can be applied in 3D as well.

Orthogonality Theorem in Three Dimensions: If $f(x,y,z) \in C^1$ in a neighbourhood of (a,b,c) and $\nabla f(a,b,c) \neq (0,0,0)$, then $\nabla f(a,b,c)$ is orthogonal to the level surface $f(x,y,z) = k$ through (a,b,c).

This theorem gives a quick way of finding the equation of the tangent plane of a surface in $\mathbb{R}^3$ given by $f(x,y,z) = k$. If $\vec{x} \in \mathbb{R}^3$ is an arbitrary point on the tangent plane to the surface $f(x,y,z) = k$ at the point $\vec{a} \in \mathbb{R}^3$, then the vector $\vec{x} - \vec{a}$ lies on the tangent plane and by the theorem above, is orthogonal to $\nabla f(\vec{a})$, meaning that $\nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}) = 0$. Since this equation is satisfied for all $\vec{x}$ in the tangent plane, it is the **equation of the tangent plane**. In component form, $f_x(a,b,c)(x-a) + f_y(a,b,c)(y-b) + f_z(a,b,c)(z-c) = 0$, where $\vec{x} = (x,y,z)$ and $\vec{a} = (a,b,c)$.


z
$\nabla f(\vec{a})$
f(x, y, z) = k
$\vec{a}$
$\vec{x}$
x
y

8 Taylor Polynomials and Taylor's Theorem

Extending Taylor polynomials to higher dimensions.

8.1 Taylor Polynomials of Degree 2

Recall from Math 137 that the **Taylor polynomial** of a function $f(x)$ is a polynomial approximation which uses the derivatives of f at a point a to approximate f in a neighbourhood around a. For example, the second derivative f'' plays an important role in approximating f because it determines whether f is concave up or down, so it can improve the accuracy of the linear approximation by defining a quadratic approximation which is the second degree Taylor polynomial.

For a function of one variable $f(x)$, the Taylor polynomial of degree 2 at point a is denoted $P_{2,a}(x)$ and is defined as,

$$P_{2,a}(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$$

This is the sum of the linear approximation $L_a(x) = f(a) + f'(a)(x-a)$ and the quadratic term. The coefficient $\frac{1}{2}f''(a)$ is determined by requiring that the second derivative of $P_{2,a}(x)$ equals the second derivative of f at a: $P_{2,a}(a)'' = f''(a)$ and indeed for higher orders i, the fractional term is always $\frac{1}{i!}$ (to cancel out the chain rule when differentiating a polynomial).

Now, suppose $f(x,y)$ has continuous second partial derivatives at (a,b). The Taylor polynomial of f of degree 2 at (a,b) is denoted $P_{2,(a,b)}(x,y)$ and is obtained by adding appropriate second degree terms in $(x-a)$ and $(y-b)$ to the linear approximation $L_{(a,b)}(x,y)$. The general second order form is,

$$P_{2,(a,b)}(x,y) = L_{(a,b)}(x,y) + A(x-a)^2 + B(x-a)(y-b) + C(y-b)^2$$

where A, B, C are constants. We can solve for A, B, C by setting the second partial derivatives of $P_{2(a,b)}$ to equal the second partial derivatives of f at (a,b), in the same way we set the derivatives of a Taylor polynomial of a single variable to the derivatives of the function it is approximating. That is, we require,

$$\frac{\partial^2 P_{2,(a,b)}}{\partial x^2} = 2A = \frac{\partial^2 f}{\partial x^2}(a,b) \text{ and } \frac{\partial^2 P_{2,(a,b)}}{\partial x \partial y} = B = \frac{\partial^2 f}{\partial x \partial y}(a,b) \text{ and } \frac{\partial^2 P_{2,(a,b)}}{\partial y^2} = 2C = \frac{\partial^2 f}{\partial y^2}(a,b)$$

Note that $L_{(a,b)}(x,y)$ does not contribute to the second partial derivatives since it is of first degree in x, y.

Therefore, the second degree Taylor polynomial $P_{2,(a,b)}$ of $f(x,y)$ at (a,b) is,

$$\begin{aligned} P_{2,(a,b)}(x,y) = {} & f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ & + \frac{1}{2}\big[f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2\big] \end{aligned}$$

In general, $P_{2,(a,b)}(x,y) \approx f(x,y)$ for (x,y) sufficiently close to (a,b) with better accuracy than the linear approximation.

For example, to approximate $\sqrt{0.95^3 + 1.98^3}$ using the Taylor polynomial of degree 2, we let $f(x,y) = \sqrt{x^3+y^3}$ and $(a,b) = (1,2)$. Find $\nabla f(1,2)$ and the Hessian matrix $Hf(1,2)$ (which is a matrix containing second order partial derivatives, see earlier for the definition). Then, find $P_{2,(1,2)}(x,y)$, which will approximate $f(x,y)$ near $(1,2)$.

8.2 Taylor's Formula with Second Degree Remainder

Recall Taylor's formula from Math 137: for single variable functions $f(x)$, if $f''(x)$ exists on $[a, x]$, then there exists a number c between a and x such that,

$$f(x) = f(a) + f'(a)(x - a) + R_{1,a}(x)$$

where the error in the linear approximation $L_a(x) = f(a) + f'(a)(x - a)$, the **Taylor remainder**, is,

$$R_{1,a}(x) = \frac{1}{2}f''(c)(x - a)^2$$

We do not know the value of c but since it lies between a, x, we can find an upper bound on the error. If f has continuous second derivatives on an interval $[a - \delta, a + \delta]$ centered on a, then f'' is bounded on this interval; there exists a number B such that $|f''(x)| \leq B$ for all $x \in [a - \delta, a + \delta]$. Thus,

$$|f(x) - L_a(x)| = |R_{1,a}(x)| = \left|\frac{1}{2}f''(c)(x - a)^2\right| = \frac{1}{2}|f''(c)|(x - a)^2 \leq \frac{1}{2}B(x - a)^2$$

for all $x \in [a - \delta, a + \delta]$. Knowing $f''(x)$, we can find a value for B.

Taylor's formula For the two variable case, if $f(x, y) \in C^2$ (continuous second partial derivatives) in some neighbourhood $N(a, b)$ of (a, b), then for all $(x, y) \in N(a, b)$, there exists a point (c, d) on the line segment joining (a, b) and (x, y) such that,

$$f(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) + R_{1,(a,b)}(x, y)$$

where the Taylor remainder is

$$R_{1,(a,b)}(x, y) = \frac{1}{2}\left[f_{xx}(c, d)(x - a)^2 + 2f_{xy}(c, d)(x - a)(y - b) + f_{yy}(c, d)(y - b)^2\right]$$

Proof: By considering only points on the line segment L joining (a, b) and (x, y), we can reduce $f(x, y)$ to a function $g(t)$ of one variable. We parameter the line segment L by $L(t) = (a + t(x - a), b + t(y - b))$, $0 \leq t \leq 1$. For simplicity, let $h = x - a, k = y - b$. Taylor's formula can be re-written as,

$$f(x, y) = f(a, b) + f_x(a, b)h + f_y(a, b)k + R_{1,(a,b)}(x, y)$$

where

$$R_{1,(a,b)}(x, y) = \frac{1}{2}\left[f_{xx}(c, d)h^2 + 2f_{xy}(c, d)hk + f_{yy}(c, d)k^2\right]$$

Let $g(t) = f(L(t))$ for $0 \leq t \leq 1$. By hypothesis, f has continuous second partial derivatives, so we can apply the Chain Rule to conclude that g' and g'' are continuous and given by,

$$\begin{aligned} g'(t) &= f_x(L(t))h + f_y(L(t))k \\ g''(t) &= f_{xx}(L(t))h^2 + 2f_{xy}(L(t))hk + f_{yy}(L(t))k^2 \end{aligned}$$

for $0 \leq t \leq 1$. Since g'' is continuous on $[0, 1]$, we can apply Taylor's formula, so there exists a $\bar{c} \in (0, 1)$ such that,

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(\bar{c})$$

Using the above equations, we can calculate each term in this equation,

$$\begin{aligned} g(1) &= f((a, b) + [(x, y) - (a, b)]) = f(x, y) \\ g(0) &= f(a, b) \\ g'(0) &= f_x(a, b)h + f_y(a, b)k \end{aligned}$$

If we let $(c,d) = L(\bar{c})$, then $\frac{1}{2}g''(\bar{c}) = R_{1,(a,b)}(x,y)$ and substituting in the equations for $g(1), g(0), g'(0)$, the result follows. A common way to apply Taylor's formula to find an upper bound on the error is to apply the triangle inequality,

$$\begin{aligned}|R_{1,(a,b)}(x,y)| &= \left|\frac{1}{2}\left[f_{xx}(c,d)(x-a)^2 + 2f_{xy}(c,d)(x-a)(y-b) + f_{yy}(c,d)(y-b)^2\right]\right| \\ &\leq \frac{1}{2}[|f_{xx}(c,d)|(x-a)^2 + 2|f_{xy}(c,d)||(x-a)||(y-b)| + |f_{yy}(c,d)|(y-b)^2]\end{aligned}$$

And find upper bounds for each of $|f_{xx}(c,d)|, |f_{xy}(c,d)|, |f_{yy}(c,d)|$ for (c,d) on the line between (a,b) and (x,y). Taylor's formula tells us that the error is upper bounded by this value.

A common inequality to use here is: $2|x-a||y-b| \leq (x-a)^2 + (y-b)^2$.

Corollary: The error term is dependent on the magnitude of the displacement $||(x,y)-(a,b)||^2$. If $f(x,y) \in C^2$ in some closed neighbourhood $N(a,b)$ of (a,b), then there exists a positive constant M such that for all $(x,y) \in N(a,b)$,

$$|R_{1,(a,b)}(x,y) \leq M||(x,y)-(a,b)||^2$$

8.3 Generalizations of the Taylor Polynomial

Multi index: If $f \in C^k$ is a function of n variables $x_1, \ldots, x_n$, we can write a kth order partial derivative of f as,

$$\partial^\alpha f = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_2}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}}\left(\frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}}\left(\cdots\left(\frac{\partial^{\alpha_n} f}{\partial x_n^{\alpha_n}}\right)\right)\right)$$

where α is a multi-index; that is, $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_i \in \mathbb{N}$. The sum $\alpha_1 + \cdots + \alpha + n = k$ is called the **order** of α and is sometimes denoted $|\alpha|$. We also define $\alpha! = \alpha_1!\,\alpha_2!\cdots\alpha_n!$. So, if α is a multi index of order k, $\alpha^\alpha f$ is a partial derivative of order k of f. We interpret $\partial^\alpha f$ as first taking the α_nth order partial derivative of f wrt x_n, then taking the α_{n-1}th order partial derivative of that wrt x_{n-1}, etc. By Clairaut's theorem, the order of the variables is interchangeable.

We can also use the multi index notation as follows: let $\vec{x} = (x_1, \ldots, x_n), \vec{a} = (a_1, \ldots, a_n), \alpha = (\alpha_1, \ldots, \alpha_n)$. Then,

$$(\vec{x} - \vec{a})^\alpha = (x_1 - a_1)^{\alpha_1}(x_2 - a_2)^{\alpha_2}\cdots(x_n - a_n)^{\alpha_n}$$

For example, if $\vec{x} = (x,y,z), \vec{a} = (-5,3,6), \alpha = (3,2,4)$, then $(\vec{x} - \vec{a})^\alpha = (x+5)^3(y-3)^2(z-7)^4$.

Higher degree Taylor polynomial: Using multi index notation, the kth degree Taylor polynomial of $f(x,y)$ is,

$$P_{k,(a,b)}(x,y) = \sum_{|\alpha| \leq k} \partial^\alpha f(a,b)\frac{[(x,y)-(a,b)]^\alpha}{\alpha!}$$

We are summing over all mixed partial derivatives whose order is $\leq k$. For example, if $k=2$, then we would sum with $\alpha = (0,0), \alpha = (1,0), \alpha = (0,1), \alpha = (2,0), \ldots$ and if we put it all together, we get the same $P_{2,(a,b)}(x,y)$ as described earlier.

Taylor's Theorem of Higher Order: For order k, if $f(x,y) \in C^{k+1}$ in some neighbourhood $N(a,b)$ of (a,b), then for all $(x,y) \in N(a,b)$, there exists a point (c,d) on the line segment between (a,b) and (x,y) such that,

$$f(x,y) = P_{k,(a,b)}(x,y) + R_{k,(a,b)}(x,y)$$

where,

$$R_{k,(a,b)}(x,y) = \sum_{|\alpha| = k+1} \partial^\alpha f(c,d)\frac{[(x,y)-(a,b)]^\alpha}{\alpha!}$$

Corollary: From Taylor's Theorem of order k, it follows that when a function $f(x,y) \in C^k$ is close to the point (a,b), then the Taylor polynomial is a good approximation for f at (a,b). Specifically, if $f(x,y) \in C^k$ in some neighbourhood of (a,b), then,

$$\lim_{(x,y)\to(a,b)} \frac{|f(x,y) - P_{k,(a,b)}(x,y)|}{||(x,y)-(a,b)||^k} = 0$$

The numerator of this fraction is the difference between f and its kth order Taylor polynomial and the denominator is the displacement between (x,y) and (a,b).

Corollary: If $f(x,y) \in C^{k+1}$ in some closed neighbourhood $N(a,b)$ of (a,b), then there exists a constant $M > 0$ such that,

$$|f(x,y) - P_{k,(a,b)}(x,y)| \leq M||(x,y)-(a,b)||^{k+1}$$

for all $(x,y) \in N(a,b)$. This corollary only tells that M exists but does not tell us how to find it.

Further Generalization: The Taylor polynomial of degree k for functions of n variables $f(\vec{x}), \vec{x} \in \mathbb{R}^n$ is,

$$P_{k,\vec{a}}(\vec{x}) = \sum_{|\alpha|\leq k} \partial^\alpha f(\vec{a}) \frac{(\vec{x}-\vec{a})^\alpha}{\alpha!}$$

---

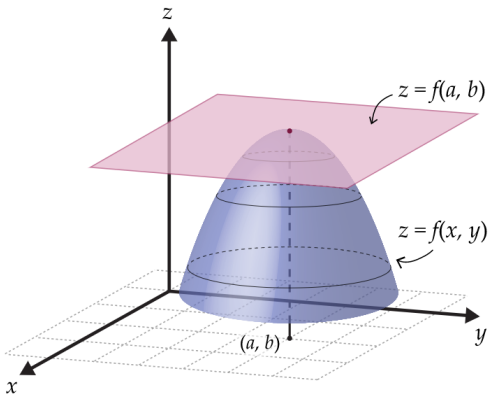
9 Critical Points

Generalizing critical points for multivariable functions, second partial derivative test to classify critical points as local maxima, local minima, or saddle points, quadratic forms.

9.1 Local Extrema and Critical Points

Let $f(x)$ be real-valued. Recall $c \in D(f)$ is a **critical value** of f if either $f'(c) = 0$ or $f'(c)$ is undefined. Critical values are often used to find local extrema (maxima or minima) of $f(x)$.

For functions of two variables, $(a, b) \in D(f)$ is a **local maximum point of** f if $f(x, y) \leq f(a, b)$ for all (x, y) in some neighbourhood of (a, b). And if $\geq$, then (a, b) is a **local minimum point**. Geometrically, if (a, b) is a local max/min point of f, then it is a local max/min for the cross section $f(x, b)$ and the cross section $f(a, y)$; (a, b) is a critical point of the cross sections $f(x, b)$ and $f(a, y)$.



Theorem: If (a, b) is a local max or min point of f, then each partial derivative is either equal to 0 or does not exist.

Proof: Let $g(x) = f(x, b)$. If (a, b) is a local max/min of f, then $x = a$ is a local max/min of g and hence either $g'(a) = 0$ or $g'(a)$ does not exist. So, either $f_x(a, b) = 0$ or $f_x(a, b)$ does not exist. A similar argument holds for f_y.

Critical points: A point $(a, b) \in D(f)$ is a critical point of f if $\frac{\partial f}{\partial x}(a, b) = 0$ or does not exist and $\frac{\partial f}{\partial y}(a, b) = 0$ or does not exist. All local max/min are critical points but not all critical points are local max/min (could be saddle points).

Saddle point: A critical point (a, b) of $f(x, y)$ is a saddle point of f if in every neighbourhood of (a, b), there exist points (x_1, y_1) and (x_2, y_2) such that,

$$f(x_1, y_1) > f(a, b) > f(x_2, y_2)$$

So, to find local extrema for $f(x, y)$, we first find the critical points of f then determine whether they are local maxima, minima, or saddle points. Note that $f_x(x, y) = 0, f_y(x, y)$ form a system of equations which is generally non-linear and there are no general algorithms for solving such systems exactly; numerical methods for finding approximate solutions is required.

9.2 Second Derivative Test

Quadratic Forms: Recall from Math 235. A function Q of the form $Q(u,v) = a_{11}u^2 + 2a_{12}uv + a_{22}v^2$ where a_{11}, a_{12}, a_{22} are constants is a quadratic form on $\mathbb{R}^2$. We could use a 2 by 2 matrix to define the quadratic form on $\mathbb{R}^2$, since

$$Q(u,v) = \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

There are four main classes of quadratic forms on $\mathbb{R}^2$:

1. If $Q(u,v) > 0$ for all $(u,v) \neq (0,0)$, then $Q(u,v)$ is positive definite
2. If $Q(u,v) < 0$ for all $(u,v) \neq (0,0)$, then negative definite
3. If $Q(u,v) < 0$ for some (u,v) and $Q(w,z) > 0$ for some (w,z), then indefinite
4. Else, $Q(u,v)$ is semidefinite. There are two classes:
   (a) If $Q(u,v) \geq 0$ for all (u,v), then $Q(u,v)$ is positive semidefinite
   (b) If $Q(u,v) \leq 0$, then negative semidefinite

These classes could also be used to describe the associated symmetric matrix containing a_{11}, a_{12}, a_{22}. We can complete the square to easily classify a quadratic form. For example, for $Q(u,v) = u^2 + 6uv + 2v^2$, completing the square gives $Q(u,v) = (u+3v)^2 - 7v^2$. We can then easily see this is indefinite.

Theorem: A quadratic form with corresponding matrix A is:

1. positive definite if $\det(A) > 0$ and $a_{11} > 0$
2. negative definite if $\det(A) > 0$ and $a_{11} < 0$
3. indefinite if $\det(A) < 0$
4. semidefinite if $\det(A) = 0$

Recall that $\det(A) = a_{11}a_{22} - a_{12}^2$ for our matrix A filled in as above.

Recall the second degree Taylor polynomial approximation for $f(x)$,

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$$

for x close to a. So, if $x = a$ is a critical point of f, then $f'(a) = 0$ and the approximation can be re-arranged to give $f(x) - f(a) \approx \frac{1}{2}f''(a)(x-a)^2$. The **second derivative test** is then:

1. If $f''(a) > 0$, then $f(x) - f(a) > 0$ for x sufficiently close to a and $x = a$ is a local min point
2. Same as above but for $<$, and $x = a$ is a local max point
3. If $f''(a) = 0$, we can conclude nothing

For a function of two variables, if $f(x,y) \in C^2$, then $f(x,y) \approx P_{2,(a,b)}(x,y)$ for (x,y) sufficiently close to (a,b). If (a,b) is a critical point of f such that $f_x(a,b) = 0 = f_y(a,b)$, then the approximation can be re-arranged to,

$$f(x,y) - f(a,b) \approx \frac{1}{2}[f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2]$$

for (x,y) sufficiently close to (a,b). The sign of the expression on the right will then (approximately) determine the sign of $f(x,y) - f(a,b)$ and hence, whether (a,b) is a local max, min, or saddle point. Now, let $u = x - a, v = y - b$. The expression on the right is a quadratic form. The reasoning above leads to the second partial derivatives test.

Second Partial Derivatives Test: Recall the Hessian matrix of f at (a,b) is,

$$Hf(a,b) = \begin{bmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{xy}(a,b) & f_{yy}(a,b) \end{bmatrix}$$

Suppose $f(x,y) \in C^2$ in some neighbourhood of (a,b) and $f_x(a,b) = 0 = f_y(a,b)$. Then,

1. If $Hf(a,b)$ is positive definite, then (a,b) is a local min point of f
2. If $Hf(a,b)$ is negative definite, then local max
3. If $Hf(a,b)$ is indefinite, then saddle point
4. If $Hf(a,b)$ is semidefinite, then the test is inconclusive

In this test, the requirement that $g''(a) > 0$ which implies a local min is replaced by the requirement that $Hf(a,b)$ is positive definite, etc. To easily classify $Hf(a,b)$, we consider the sign of $\det(Hf(a,b)) = f_{xx}(a,b)f_{yy}(a,b) - f_{xy}^2(a,b)$ and the sign of $f_{xx}(a,b)$. Note that another way of classifying a Hessian matrix is by finding its eigenvalues. In particular, a symmetric matrix is positive definite if all eigenvalues are positive, negative definite if all eigenvalues are negative, and indefinite if it has both positive and negative eigenvalues.

Proof: See pages 105 to 108 in the textbook.

Degenerate Critical Points: If $Hf(a,b)$ is semidefinite, the Second Partial Derivatives test is inconslusive and the critical point (a,b) may be a local max, min, or a saddle point. We call such a point **degenerate**. To further classify it, we need to investigate the sign of $f(x,y) - f(a,b)$ in a small neighbourhood of (a,b); if $f(x,y) - f(a,b)$ is always negative in a small neighbourhood of (a,b), then (a,b) is a local min, if always positive then local max, if both negative and positive, then saddle point.

Generalizations: Consider the Hessian matrix of a function f of n variables at $\vec{a}$. It would be an $n \times n$ symmetric matrix and can be classified as positive definite, negative definite, indefinite, or semidefinite by considering the associated quadratic form in $\mathbb{R}^n$,

$$Q(\vec{u}) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a}) u_i u_j$$

The second derivative test in $\mathbb{R}^2$ also holds in $\mathbb{R}^n$ and can be justified intuitively using the second degree Taylor polynomial approximation $f(\vec{x}) \approx P_{2,\vec{a}}(\vec{x})$ which leads to,

$$f(\vec{x}) - f(\vec{a}) \approx \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a,b)(x_i - a_i)(x_j - a_j)$$

Level Curves Near Critical Points: Next, consider a function $f \in C^2$. If $\nabla f(a,b) \neq (0,0)$, by the Orthogonality Theorem, the level curve of f through (a,b) is a smooth curve (sufficiently close to (a,b)). Also, by continuity, $\nabla f(x,y) \neq (0,0)$ for all (x,y) in some neighbourhood of (a,b). Thus, if $\nabla f(a,b) \neq (0,0)$, there will be some neighbourhood of (a,b) in which the level curves of f are smooth non-intersecting curves. It can then be shown that to approximate the level curves of f near a critical point, we can sketch the level curves of the associated quadratic form $Q(u,v)$.

Convex Functions 1D: for functions of one variable, we define a twice differentiable function $f(x)$ to be **convex** if $f''(x) \geq 0$ for all x and **strictly convex** if $f''(x) > 0$ for all x. Convex means "concave up". If $f(x) \in C^2$ and is strictly convex, then,

1. f lies above any tangent line. Formally, $f(x) > L_a(x) = f(a) + f'(x)(x-a)$ for all $x \neq a$ for any $a \in \mathbb{R}$. This follows from Taylor's Theorem, since $f(x) = L_a(x) + \frac{f''(c)}{2}(x-a)^2$ where c is between a and x.

2. Any secant line lies above f. Formally, for $a < b$, $f(x) < f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$ for $x \in (a,b)$. The proof is: let $g(x) = f(x) - \left[f(a) + \frac{f(b)-f(a)}{b-a}(x-a)\right]$. Then, $g(a) = g(b) = 0$ and $g''(x) = f''(x) > 0$ and we want to show that $g(x) < 0$ for all $x \in (a,b)$. By the MVT, $\frac{f(b)-f(a)}{b-a} = f'(c)$ for some $c \in (a,b)$ and $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a} = f'(x) - f'(c)$, so $g'(c) = 0$. Since $g''(x) > 0$, then $g'(x)$ is strictly increasing, so $g'(x) < 0$ on $[a,c)$ and $g'(x) > 0$ on $(c,b]$, implying that $g'(x)$ is strictly decreasing on $[a,c]$ and strictly increasing on $[c,b]$. Since $g(a) = g(b) = 0$, then $g(x) < 0$ on $(a,c]$ and on $[c,b)$, so $g(x) < 0$ on (a,b) as required.

Convex Functions 2D: suppose $f(x,y)$ has continuous second partial derivatives. Then, f is convex if $Hf(x,y)$ is positive semi-definite for all (x,y) and strictly convex if $Hf(x,y)$ is positive definite for all (x,y). So, f is strictly convex if $f_{xx} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ for all (x,y). Similar to the 1D case, if $f(x,y)$ has continuous second partial derivatives and is strictly convex, then,

1. f lies above the tangent plane. Formally, $f(x,y) > L_{(a,b)}(x,y)$ for all $(x,y) \neq (a,b)$. This follows from Taylor's Theorem.

2. The cross-section of the graph of f above the line segment from (a_1,a_2) to (b_1,b_2) lies below the secant line. Formally, $f(a_1 + t(b_1-a_1), a_2 + t(b_2-a_2)) < f(a_1,a_2) + t[f(b_1,b_2) - f(a_1,a_2)]$ for $0 < t < 1$ and $(a_1,a_2) \neq (b_1,b_2)$. The proof: we parameterize the line segment L from (a_1,a_2) to (b_1,b_2) by,

$$L(t) = (a_1 + t(b_1-a_1), a_2 + t(b_2-a_2)), 0 \leq t \leq 1$$

Let $h = b_1 - a_2, k = b_2 - a_2$. Let $g(t) = f(L(t)), 0 \leq t \leq 1$. Since f has continuous second partial derivatives, we apply the Chain Rule to conclude that $g'(t) = f_x(L(t))h + f_y(L(t))k$ and $g''(t) = f_{xx}(L(t))h^2 + 2f_{xy}(L(t))hk + f_{yy}(L(t))k^2$ are both continuous for $0 \leq t \leq 1$. Since $f_{xx}(L(t)) > 0$ and $f_{xx}(L(t))f_{yy}(L(t)) - f_{xy}(L(t))^2 > 0$ for all t, then $g''(t) > 0$. So, by point 2 of the 1D section on convex functions,

$$g(t) < g(0) + \frac{g(1)-g(0)}{1-0}(t-0)$$

for $0 < t < 1$. Therefore, $f(a_1 + t(b_1-a_1), a + 2 + t(b_2-a_2)) < f(a_1,a_2) + t[f(b_1,b_2) - f(a_1,a_2)]$ for $0 < t < 1$ as required.

Theorem: If $f(x,y) \in C^2$ is convex, then every critical point (c,d) satisfies $f(x,y) \geq f(c,d)$ for all $(x,y) \neq (c,d)$. In other words, if $f \in C^2$ is convex, then all its critical points minimize f. If $f(x,y) \in C^2$ is strictly convex and has a critical point (c,d), then $f(x,y) > f(c,d)$ for all $(x,y) \neq (c,d)$ and f has no other critical point. In other words, if $f \in C^2$ is strictly convex and has a critical point, then the critical point is unique and minimizes f. So, a strictly convex function has at most one critical point which if it exists, must be a minimum.

Proof: We will prove the second part of this theorem, the first part proceeds by a similar argument. Since $f(x,y) \in C^2$, then f has continuous second partial derivatives so since (c,d) is a critical point, then $f_x(c,d) = f_y(c,d) = 0$. Thus, $L_{(c,d)}(x,y) = f(c,d)$. So, $f(x,y) > L_{(c,d)}(x,y) = f(c,d)$ for all $(x,y) \neq (c,d)$. Next, suppose for contradiction f has a second critical point (c_1,d_1) such that $(c_1,d_1) \neq (c,d)$. By similar reasoning to above, $f(x,y) > f(c_1,d_1)$ for all $(x,y) \neq (c_1,d_1)$ but then $f(c,d) > f(c_1,d_1)$ and $f(c_1,c_2) > f(c,d)$, contradiction.

Note that a convex function $f \in C^2$ might have more than one critical points and an example is $f(x,y) = (x-y)^2$, which has critical points anywhere $x = y$. As well, it is possible that $f \in C^2$ has a unique critical point which is a minimum but f is not strictly convex. For example, $f(x,y) = e^{-(x^2+y^2)^{-1}}$ if $(x,y) \neq (0,0)$ and 0 otherwise has a unique critical point at $(0,0)$ which is a minimum but f is not strictly convex.

---

10 Optimization Problems

Absolute max/min for multivariable functions, open and closed sets, algorithm for extreme values, Method of Lagrange Multipliers.

10.1 Extreme Value Theorem

Recall from single variable calculus, for a real valued function $f(x)$ and interval $I \subseteq \mathbb{R}$, a point $x = c \in I$ is the absolute max point of f on I if $f(x) \leq f(c)$ for all $x \in I$. Similarly for absolute min point $f(x) \geq f(c)$. As well, recall the 1D Extreme Value Theorem which tells us any continuous function has an absolute max and min on any closed interval: if $f(x)$ is continuous on a finite closed interval I, then there exists $c_1, c_2 \in I$ such that $f(c_1) \leq f(x) \leq f(c_2)$ for all $x \in I$. We require all three parts of the hypothesis: continuity, finite and closed interval; counterexamples exist if one does not hold.

For functions $f(x, y)$ of two variables and a set $S \subseteq \mathbb{R}^2$, a point $(a, b) \in S$ is an **absolute maximum point** of f on S if $f(x, y) \leq f(a, b)$ for all $(x, y) \in S$. $f(a, b)$ is the **absolute maximum value** of f on S. Similarly for absolute min.

Topological Aspects: Before generalizing the Extreme Value Theorem for 2D, we need to generalize the concept of a finite closed interval to sets in $\mathbb{R}^2$ by discussing some topological aspects of $\mathbb{R}^2$.

1. A set $S \subset \mathbb{R}^2$ is **bounded** iff it is contained in some neighborhood of the origin. In other words, every point in S must have finite distance from the origin. For example, the set of points in a circle $\{(x, y) | x^2 + y^2 < 1\}$ is bounded since it is contained in the neighborhood of radius 2 around the origin.
2. Given a set $S \subseteq \mathbb{R}^2$, a point $(a, b) \in \mathbb{R}^2$ is a **boundary point** of S iff every neighborhood of (a, b) contains at least one point in S and one point not in S. Recall from CO 342. Intuitively, these are points on the "edge" of S. Note S does not need to be bounded for it to have boundary points. For example, the boundary of $\{(x, y) \in \mathbb{R}^2 | x \geq 0\}$ is the y axis.
3. The set $B(S)$ of all boundary points of S is called the **boundary** of S.
4. A set $S \subseteq \mathbb{R}^2$ is **closed** if S contains all its boundary points. For example, since the y axis is in $S = \{(x, y) \in \mathbb{R}^2 | x \geq 0\}$, then S is closed. The concept of a "closed set" in $\mathbb{R}^2$ generalizes the idea of a closed interval in $\mathbb{R}$.

Extreme Value Theorem (EVT) for functions of two variables: if $f(x, y)$ is continuous on a closed and bounded set $S \subset \mathbb{R}^2$, then there exist points $(a, b), (c, d) \in S$ such that, for all $(x, y) \in S$,

$$f(a, b) \leq f(x, y) \leq f(c, d)$$

So, $f(x, y)$ always has an absolute max and min on any closed and bounded set. Note $f(x, y)$ may have an absolute max or min on a set $S \subseteq \mathbb{R}^2$ even if the conditions of EVT are not satisfied.

Algorithm: to find absolute max/min values for $f(x, y)$ on some $S \subset \mathbb{R}^2$ which is closed and bounded, where f is continuous on S,

1. Find all critical points of f contained in S
2. Evaluate f at each point
3. Find the max/min values of f on the boundary $B(S)$
4. The max value of f on S is the largest value of the function found in steps 2 and 3. The min value is the smallest.

This is analogous to the Closed Interval Method for finding max/min for single variable functions on closed intervals. Overall, the max/min occurs either at a critical point in S or on the boundary of S. The absolute max/min may occur at more than one point in S and also, this algorithm does not require determining whether the critical points are local max or min.

To analyze the boundary, it is helpful to describe it in parametric form. For example, the boundary of the unit circle $x^2 + y^2 = 1$ is, $x = \cos t, y = \sin t, 0 \leq t \leq 2\pi$ and $g(t) = f(\cos t, \sin t)$ considers all points of f on the boundary of the circle. See page $113 - 114$ of the textbook for examples.

10.2 Optimization with Constraints

Suppose we want to find the max/min of a function $f(x, y)$ subject to a constraint $g(x, y) = k$, where $g \in C^1$. Geometrically, we want to find the max/min value of $f(x, y)$ on the level curve $g(x, y) = k$. If $f(x, y)$ has a local max/min at (a, b) relative to nearby points on the curve $g(x, y) = k$ and $\nabla g(a, b) \neq (0, 0)$, then by the Implicit Function Theorem, $g(x, y) = k$ can be described by parameteric equations,

$$x = p(t) \qquad y = q(t)$$

where p, q are differentiable and $(a, b) = (p(t_0), q(t_0))$ for some t_0. Let $u(t) = f(p(t), q(t))$. This function gives values of f on the constraint curve and hence has a local max or min at t_0. As well, $u'(t_0) = 0$. Assuming f is differentiable, by the Chain Rule,

$$u'(t) = f_x(p(t), q(t))p'(t) + f_y(p(t), q(t))q'(t)$$

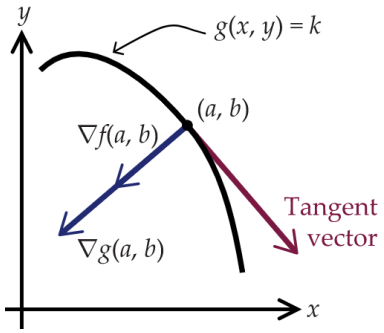
Evaluating at t_0 and using the fact that $u'(t_0) = 0$,

$$0 = f_x(a, b)p'(t_0) + f_y(a, b)q'(t_0) \implies \nabla f(a, b) \cdot (p'(t_0), q'(t_0)) = 0$$

Recall if $\nabla g(a, b)$ is non-zero, it is orthogonal to the level curve $g(x, y) = k$ at (a, b). So, since $(p'(t_0), q'(t_0))$ is the tangent vector to the constraint curve,

$$\nabla g(a, b) \cdot (p'(t_0), q'(t_0)) = 0$$

Thus, $\nabla f(a, b)$ and $\nabla g(a, b)$ are scalar multiples of each other: there exists a constant λ such that $\nabla f(a, b) = \lambda \nabla g(a, b)$.



Method of Lagrange Multipliers: Suppose $f(x, y)$ is differentiable and $g \in C^1$. To find the max/min value of f subject to the constraint $g(x, y) = k$, evaluate $f(x, y)$ at all points (a, b) which satisfy one of the following conditions:

1. $\nabla f(a, b) = \lambda \nabla g(a, b)$ and $g(a, b) = k$
2. $\nabla g(a, b) = (0, 0)$ and $g(a, b) = k$

3. (a, b) is an endpoint of the curve $g(x, y) = k$

The max/min value of f is the largest/smallest value of f obtained at the points found in conditions $1 - 3$. For condition 1, we have to solve the system of 3 equations in 3 unknowns,

$$f_x(x, y) = \lambda g_x(x, y) \qquad f_y(x, y) = \lambda g_y(x, y) \qquad g(x, y) = k$$

for x, y. This should be done systematically to find all possible points. The variable λ is the **Lagrange multiplier** and is not required for our purposes but is very useful in other applications. Note that if the curve $g(x, y) = k$ is unbounded, we must consider $\lim_{||(x,y)||\to\infty} f(x, y)$ for (x, y) satisfying $g(x, y) = k$.

This algorithm can be generalized to functions of three variables $f(x, y, z)$. To find the max/min of a differentiable function $f(x, y, z)$ subject to $g(x, y, z) = k$ such that $g \in C^1$, we evaluate $f(x, y, z)$ at all points (a, b, c) that satisfy one of the following:

1. $\nabla f(a, b, c) = \lambda \nabla g(a, b, c)$ and $g(a, b, c) = k$
2. $\nabla g(a, b, c) = (0, 0, 0)$ and $g(a, b, c) = k$
3. (a, b, c) is an endpoint of the curve $g(x, y, z) = k$

The max/min value of $f(x, y, z)$ is the largest/smallest value of f obtained from the points in conditions $1 - 3$. Note if condition 1 in the algorithm holds, it follows that the level surface $f(x, y, z) = f(a, b, c)$ and the constraint surface $g(x, y, z) = k$ are tangent at the point (a, b, c) since their normals coincide.

Generalization: This method can also generalize to functions of n variables $f(\vec{x}), \vec{x} \in \mathbb{R}^n$ and r constraints of the form $g_i(\vec{x}) = 0$ for $1 \leq i \leq r$. To find the possible max/min points of f subject to these r constraints, we find all points $\vec{a}$ such that,

$$\nabla f(\vec{a}) = \lambda_1 \nabla g_1(\vec{a}) + \cdots + \lambda_r \nabla g_r(\vec{a}) \text{ and } g_i(\vec{a}) = 0, 1 \leq i \leq r$$

The scalars $\lambda_1, \ldots, \lambda_r$ are the Lagrange multipliers. When $r = 1$ and $n = 2$ or 3, this reduces to one of the previous algorithms.

Overall, this method allows optimization to be solved without explicit parameterization in terms of the constraints. Solving optimization with constraints can help us find max/min on the boundaries of sets in the EVT algorithm.

---

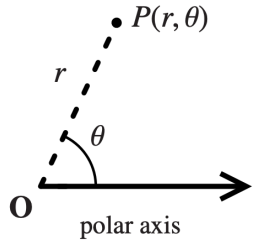
11 Coordinate Systems

Polar coordinates, cylindrical coordinates, spherical coordinates, converting to and from Cartesian coordinates, graphing curves.

11.1 Polar Coordinates

A **coordinates system** is a system for representing the location of a point in a space by an ordered n-tuple. The elements of the n-tuple are the **coordinates** of the point. A common technique for solving multiple integrals which will be seen later is to choose a coordinate system which simplifies the calculations. The choice of coordinates often depends on the type of symmetry involved in the problem.

Polar coordinates is a 2D coordinate system. As in all coordinate systems, we must have a frame of reference. In a plane, we choose the point O called the **pole** (or origin). From O, we draw a ray called the **polar axis**, generally drawn horizontally to the right to match the positive x-axis in Cartesian coordinates. The position of any point P in the plane is represented by the ordered pair (r, θ), where $r \geq 0$ is the length of the line OP and θ is the angle between the polar axis and OP. r, θ are the polar coordinates of P,



The following conventions will be used: angle θ is considered positive if measured in the ccw direction from the polar axis and negative if measured in the cw direction, the origin O is represented by $(0, \theta)$ for any θ, the value of r is non-negative to coincide with the interpretation of r as distance. Polar coordinates are suited for solving problems in which there is symmetry about the pole. A point (r, θ) in polar coordinates can have infinitely many representations,

$$(r, \theta) = (r, \theta + 2\pi k), \quad k \in \mathbb{Z}$$

Often we restrict the range of θ such as $0 \leq \theta < 2\pi$ to get a unique representation.

If O is placed at the origin of the Cartesian plane and the polar axis aligns with the positive x-axis, then for a point (x, y) in Cartesian coordinates and a point (r, θ) in polar coordinates,

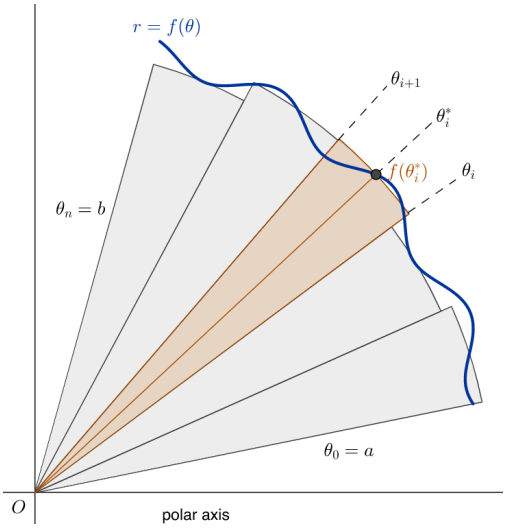
$$x = r\cos\theta \qquad y = r\sin\theta \qquad r = \sqrt{x^2 + y^2} \qquad \tan\theta = \frac{y}{x} \qquad r^2 = x^2 + y^2$$

Note that $\tan\theta = \frac{y}{x}$ does not uniquely determine θ, and we need to be careful to choose the θ which lies in the correct quadrant.

Graphing: The graph of an explicitly defined polar equation is of the form $r = f(\theta)$, or $\theta = f(r)$. Alternatively, we can define a polar equation implicitly as $f(r, \theta) = 0$. This is the curve consisting of all points that have at least one polar representation (r, θ) which satisfies the equation of the curve. For example, the polar equation $r = 1$ can be sketched as a circle of radius 1 around O. The equation $r = e^{\theta}$ gives a logarithmic spiral which often appears in nature. An easy way to sketch is to first sketch as if in Cartesian coordinates in the $r\theta$-plane (but only for $r \geq 0$) and use this graph to plot the points in the xy-plane. For example, for the polar equation $r = 1 + \sin\theta$, we first sketch in a plane where the x-axis is θ and y-axis is r. This would let us see that as θ increases from 0 to $\frac{\pi}{2}$, the radius increases from 1 to 2, etc.

We can use the five equations above to convert polar equations to Cartesian coordinates and vice versa. For example, $r = \cos\theta$ becomes $(x - \frac{1}{2})^2 + y^2 = \frac{1}{4}$ using the equations $r^2 = x^2 + y^2$ and $x = r\cos\theta$. And the equation of the curve $(x^2 + y^2)^{\frac{3}{2}} = 2xy$ can be converted to $r = \sin 2\theta$.

Area under a curve: Polar coordinates can also be used to compute areas between curves by using sectors of a circle rather than rectangles. Recall if θ_1, θ_2 where $\theta_2 > \theta_1$ are two angles in a circle of radius r, the area between them is $\frac{1}{2}r^2(\theta_2 - \theta_1)$. Now, to compute the region below,



we divide the region bounded by $\theta = a$ and $\theta = b$ and $r = f(\theta)$ into subregions $\theta_0, \ldots, \theta_n$ of equal difference $\Delta\theta$. In each subregion bounded by θ_i, θ_{i+1} for $0 \leq i < n$, choose some point θ_i^* and form the sector between θ_i, θ_{i+1} with radius $f(\theta_i^*)$. The area of this sector is $\frac{1}{2}[f(\theta_i^*)]^2\Delta\theta$ and the area is approximately the sum of all these sectors. Letting the number of sectors (subdivisions) go to ∞ and hence each $\Delta\theta$ goes to 0, the area is,

$$\lim_{||\Delta\theta||\to 0} \sum_{i=0}^{n-1} \frac{1}{2}[f(\theta_i^*)]^2\Delta\theta = \int_a^b \frac{1}{2}[f(\theta)]^2 d\theta$$

Using this result, we can prove that the area inside the circle $r = a$ is $\int_0^{2\pi} \frac{1}{2}a^2 d\theta = \frac{1}{2}a^2(2\pi - 0) = \pi a^2$, as expected.

Algorithm for areas between two curves: To find the area between two curves in polar coordinates, we use the same method we used to do this using Cartesian coordinates: first find the points of intersections, then graph the curves and split the desired region into easily integrable regions, then integrate. For example, we can use this algorithm to find the area inside $r = 2\sin 2\theta$ but outside $r = 1$. It is easier to find points of intersection by graphing, since points in polar coordinates do not have unique expressions so setting two equations equal may miss points. See example in page 132.

Polar coordinates are best used when there is symmetry about the origin in 2D.

11.2 Cylindrical Coordinates

We can extend polar coordinates to 3D by introducing another axis, the **axis of symmetry**, through the pole perpendicular to the polar plane. Then, any point P is represented by clindrical coordinates (r, θ, z) where r, θ are in polar coordinates and z is the height above (or below) the polar plane. As in polar coordinates, we have the restrictions $r \geq 0, 0 \leq \theta < 2\pi$ (or $-\pi < \theta \leq \pi$).

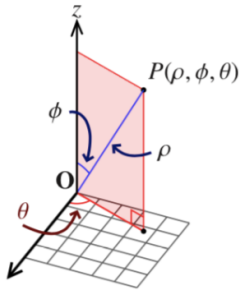
If we place the pole at the origin of 3D Cartesian space, polar axis along the positive x-axis as in polar coordinates, and the axis of symmetry along the z-axis, then a point (r, θ, z) can be related to a point (x, y, z) in Cartesian coordinates by the same five equations as earlier described, plus $z = z$. We can use these to transform equations of curves between coordinate systems. For example, $z = r^2 \cos\theta$ is $z = x\sqrt{x^2 + y^2}$.

As with functions $z = f(x, y)$, the graphs of functions $z = f(r, \theta)$ or more generally $f(r, \theta, z) = 0$ are surfaces in $\mathbb{R}^3$. For example, the graph of $r = 1$ in cylindrical coordinates is the infinite cylinder of radius 1 centered at O.

Cylindrical coordinates are useful when there is symmetry about the z-axis in 3D. It is sometimes useful to line the polar axis and axis of symmetry along different axes.

11.3 Spherical Coordinates

This is another extension of polar coordinates to 3D. Again, we use the pole O and polar axis and draw an axis z perpendicular to the polar plane. Then, a point P can be represented using spherical coordinates (ρ, ϕ, θ) where $\rho \geq 0$ is the length of the line OP, θ is the same angle as in polar and cylindrical coordinates, and ϕ is the angle between the positive z-axis and line OP.



θ gives the orientation of P around the z-axis and ϕ should indicate the "tilt" of the point with the z-axis so we restrict $0 \leq \phi \leq \pi$. Also, $\rho \geq 0, 0 \leq \theta < 2\pi$ (or $-\pi < \theta \leq \pi$). A point (x, y, z) in Cartesian coordinates and a point (ρ, ϕ, θ) in spherical coordinates are related as,

$$x = \rho\sin\phi\cos\theta \qquad y = \rho\sin\phi\sin\theta \qquad z = \rho\cos\phi \qquad \rho = \sqrt{x^2 + y^2 + z^2} \qquad \tan\theta = \frac{y}{x}$$

$$\cos\phi = \frac{z}{\sqrt{x^2 + y^2 + z^2}} \qquad x^2 + y^2 + z^2 = \rho^2$$

θ determines which quadrant the point is in (its rotation around the z-axis) and ϕ determines whether the point will be above or below the xy-plane. We can use these seven equations to convert between coordinate systems. For example, $z = x^2 + y^2$ is $\cos\phi = \rho\sin^2\phi$ for $0 \leq \phi \leq \frac{\pi}{2}$. Note ϕ in the range $\frac{\pi}{2} < \phi \leq \pi$ is left out because $z \geq 0$ in the original equation.

As with cylindrical coordinates, the graph of a function $f(\rho, \phi, \theta) = 0$ in spherical coordinates gives a surface in $\mathbb{R}^3$. For example, $\rho = 2$ is the graph of all points 2 units from the origin in the 3D Cartesian space (a sphere of radius 2).

This coordinate system is best used when there is symmetry about the origin in 3D.

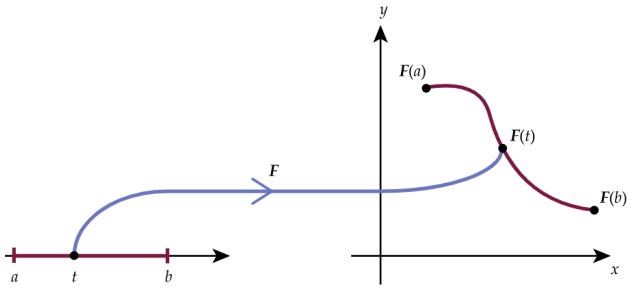
---

12 Mappings

Mappings of $\mathbb{R}^2$ into $\mathbb{R}^2$, geometric interpretation of mappings as transformations of regions of space, using matrices to define a linear approximation of a mapping.

12.1 Geometry of Mappings

A function whose domain is a subset of $\mathbb{R}^n$ and whose codoamin is $\mathbb{R}^m$ is a **vector-valued** function. An example is parametric equations $x = f(t), y = g(t)$ for a curve in $\mathbb{R}^2$. These two scalar equations can be written as a vector equation, $(x, y) = F(t) = (f(t), g(t))$, where F is a vector-valued function mapping a subset of $\mathbb{R}$ to a subset of $\mathbb{R}^2$.



A vector-valued function whose domain is a subset of $\mathbb{R}^n$ and whose codomain is also a subset of $\mathbb{R}^n$ is a **mapping** (or transformation).

A pair of equations $u = f(x, y), v = g(x, y)$ associates with each point $(x, y) \in \mathbb{R}^2$ a single point $(u, v) \in \mathbb{R}^2$ and thus defines a vector valued function $(u, v) = F(x, y) = (f(x, y), g(x, y))$. The scalar functions f, g are called the **component functions** of the mapping. Mappings will be used later to perform a change of variables in multiple integrals.

In general, if a mapping F from $\mathbb{R}^2$ to $\mathbb{R}^2$ acts on a subset S in its domain (eg a curve C), it will determine a set $F(S)$ in its range, called the **image of S under F**. For any linear mapping $(u, v) = F(x, y)$, the image of a straight line in the xy-plane is a straight line in the uv-plane. This is not necessarily the case with a non-linear mapping (eg a mapping from polar coordinates to Cartesian coordinates). To determine images of a nice set S (such as a square), we can first find the image of each of the boundary lines. See page 142 for an example.

12.2 Linear Approximation of a Mapping

Suppose a mapping F defined by $u = f(x, y), v = g(x, y)$ has continuous partial derivatives (meaning, f, g have continuous partial derivatives). The point (a, b) in the xy-plane is the point (c, d) in the uv-plane, where $c = f(a, b), d = g(a, b)$. We want to approximate the image $(c+\Delta u, d+\Delta v)$ of a nearby point $(a+\Delta x, b+\Delta y)$. Using the linear approximation for f, g separately,

$$\Delta u \approx \frac{\partial f}{\partial x}(a,b)\Delta x + \frac{\partial f}{\partial y}(a,b)\Delta y \qquad \Delta v \approx \frac{\partial g}{\partial x}(a,b)\Delta x + \frac{\partial g}{\partial y}(a,b)\Delta y$$

for sufficiently small $\Delta x, \Delta y$. In matrix form,

$$\begin{bmatrix}\Delta u \\ \Delta v\end{bmatrix} \approx \begin{bmatrix}\frac{\partial f}{\partial x}(a,b) & \frac{\partial f}{\partial y}(a,b) \\ \frac{\partial g}{\partial x}(a,b) & \frac{\partial g}{\partial y}(a,b)\end{bmatrix}\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix}$$

This motivates the following definition.

The **derivative matrix** of a mapping defined by $F(x,y) = (f(x,y), g(x,y))$ is denoted DF and defined by,

$$DF = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial x} \end{bmatrix}$$

If we introduce the column vectors $\Delta\vec{u} = \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}, \Delta\vec{x} = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$, then the **increment form of the linear approximation for mappings** becomes,

$$\Delta\vec{u} \approx DF(a,b)\Delta\vec{x}$$

for $\Delta\vec{x}$ sufficiently small. Thus, the **linear approximation for mappings** is,

$$F(x,y) \approx F(a,b) + DF(a,b)\Delta\vec{x}$$

Geometrically, the derivative matrix $DF(a,b)$ acts as a linear mapping on the displacement vector $\Delta\vec{x}$ to give an approximation of the image $\Delta\vec{u}$ of the displacement under F.

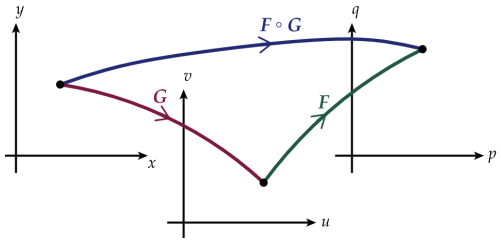
Generalization: A mapping F from $\mathbb{R}^n$ to $\mathbb{R}^m$ is defined by a set of m component functions $u_1 = f_i(x_1, \ldots, x_n)$ for $1 \leq i \leq m$. In vector notation, $\vec{u} = F(\vec{x}) = (f_1(\vec{x}), \ldots, f_m(\vec{x}))$ for $\vec{x} \in \mathbb{R}^n$. Assuming F has continuous partial derivatives, then the derivative matrix of F, denoted $DF(\vec{x})$, is the $m \times n$ matrix where the entry in the jth row (indexing from 1) and k column is $\frac{\partial f_j}{\partial x_k}$. As expected, the linear approximation for F at $\vec{a}$ is $F(\vec{x}) \approx F(\vec{a}) + DF(\vec{a})\Delta\vec{x}$, where $\Delta\vec{u} \in \mathbb{R}^m, \Delta\vec{x} \in \mathbb{R}^n$, and the ith entry in $\Delta\vec{u}$ is Δu_i and the jth entry in $\Delta\vec{x}$ is Δx_j, for $1 \leq i \leq m, 1 \leq j \leq n$.

12.3 Composite Mappings and the Chain Rule

Consider successive mappings F, G of $\mathbb{R}^2$ into $\mathbb{R}^2$ defined by,

$$F : \begin{cases} p = p(u,v) \\ q = q(u,v) \end{cases} \qquad G : \begin{cases} u = u(x,y) \\ v = v(x,y) \end{cases}$$

The composite mapping $F \circ G$ is defined by $p = p(u(x,y), v(x,y))$ and $q = q(u(x,y), v(x,y))$.



which maps the xy-plane onto the pq-plane.

Chain Rule in Matrix Form for Mappings: Let F, G be mappings from $\mathbb{R}^2$ to $\mathbb{R}^2$. If G has continuous partial derivatives at (x,y) and F has continuous partial derivatives at $(u,v) = G(x,y)$, then the composite mapping $F \circ G$ has continuous partial derivatives at (x,y) and,

$$D(F \circ G)(x,y) = DF(u,v)DG(x,y)$$

Proof: Can be proven by using the matrix definitions of $DF(u,v), DG(x,y)$ and chain rules on partial derivatives. For example, $\frac{\partial p}{\partial x} = \frac{\partial p}{\partial u}\frac{\partial u}{\partial x}$.

---

13 Jacobians and Inverse Mappings

Let F be a mapping from a set D_{xy} onto a set D_{uv}. If there exists a mapping F^{-1} (the **inverse mapping of** F) which maps D_{uv} onto D_{xy} such that $(x,y) = F^{-1}(u,v)$ iff $(u,v) = F(x,y)$, then F is invertible on D_{xy}. Note $(F^{-1} \circ F)(x,y) = (x,y)$ for all $(x,y) \in D_{xy}$ and $(F \circ F^{-1})(u,v) = (u,v)$ for all $(u,v) \in D_{uv}$.

A mapping F from $\mathbb{R}^2$ to $\mathbb{R}^2$ is **one-to-one/injective** on a set D_{xy} iff $F(a,b) = F(c,d)$ implies $(a,b) = (c,d)$ for all $(a,b),(c,d) \in D_{xy}$. That is, each element of the domain is sent to a unique element of the codomain.

Theorem: Let F be a mapping from D_{xy} onto D_{uv}. If F is one-to-one on D_{xy}, then F is invertible on D_{xy}.

Recall if $f'(x) > 0$ for all $x \in [a,b]$, then f is one-to-one on $[a,b]$ and hence has an inverse on $[a,b]$. For a mapping F, there is also a relationship between the derivative matrix DF of F and F being invertible.

Theorem: Let F be a mapping from D_{xy} to D_{uv}. If F has continuous partial derivatives at $\vec{x} \in D_{xy}$ and there exists an inverse mapping F^{-1} of F which has continuous partial derivatives at $\vec{u} = F(\vec{x}) \in D_{uv}$, then,

$$DF^{-1}(\vec{u})DF(\vec{x}) = I$$

In words, if a mapping F is invertible, then its derivative matrix is invertible and its inverse matrix is the derivative matrix of the inverse map. This gives a way of finding the inverse of a mapping $DF(\vec{x})$, since $(DF)^{-1}(\vec{x}) = DF^{-1}(\vec{u})$.

Proof: By the Chain Rule in matrix form, then since $(F^{-1} \circ F)(\vec{x}) = \vec{x}$,

$$DF^{-1}(\vec{u})DF(\vec{x}) = D(F^{-1} \circ F)(\vec{x}) = D\vec{x} = \begin{bmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial y} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

where D is the derivative matrix of the identity function.

The **Jacobian** of a mapping $(u,v) = F(x,y) = (u(x,y), v(x,y))$ is denoted $\frac{\partial(u,v)}{\partial(x,y)}$ and is defined by,

$$\frac{\partial(u,v)}{\partial(x,y)} = \det[DF(x,y)] = \det \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}$$

The derivative matrix is also sometimes called the Jacobian matrix. As well, if F is one-to-one and maps $(x,y) \to (u,v)$, then the inverse mapping F^{-1} maps $(u,v) \to (x,y)$ and the Jacobian of the inverse mapping is,

$$\frac{\partial(x,y)}{\partial(u,v)} = \det\left[DF^{-1}(u,v)\right] = \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

Theorem: Consider a mapping $(u,v) = F(x,y) = (f(x,y), g(x,y))$ which maps a subset D_{xy} onto a subset D_{uv}. Suppose f, g have continuous partial derivatives on D_{xy}. If F has an inverse mapping F^{-1} with continuous partial derivatives on D_{uv}, then the Jacobian of F, $\frac{\partial(u,v)}{\partial(x,y)}$ is non-zero. This follows from the earlier theorem because a square matrix has an inverse iff its determinant is non-zero. However, the converse is not true (non-zero Jacobian does not imply there is an inverse). For example, $(u,v) = F(x,y) = (e^x \cos y, e^x \sin y)$.

Theorem: Recall $\det(AB) = \det A \det B$ for all $n \times n$ matrices A, B. If F is a mapping from D_{xy} to D_{uv} with continuous partial derivatives at $\vec{x} \in D_{xy}$ and there is an inverse mapping F^{-1} of F with continuous partial derivatives at $\vec{u} = F(\vec{x}) \in D_{uv}$, then since $\det I = 1$,

$$\frac{\partial(x,y)}{\partial(u,v)} = \frac{1}{\frac{\partial(u,v)}{\partial(x,y)}}$$

Inverse Mapping Theorem: If a mapping $(u, v) = F(x, y)$ has continuous partial derivatives in some neighborhood of (a, b) and $\frac{\partial(u,v)}{\partial(x,y)} \neq 0$ at (a, b), then there is a neighborhood of (a, b) in which F has an inverse mapping $(x, y) = F^{-1}(u, v)$ which has continuous partial derivatives. In other words, if we restrict our attention to a neighborhood of a point (a, b) in which the mapping F has continuous partial derivatives and a non-zero Jacobian, then F has an inverse mapping with continuous partial derivatives in a neighborhood of (a, b). Proof beyond scope of this course.

Generalized Jacobian: For a mapping defined by $\vec{u} = F(\vec{x}) = (f_1(\vec{x}), \ldots, f_n(\vec{x}))$ where $\vec{u} = (u_1, \ldots, u_n)$ and $\vec{x} = (x_1, \ldots, x_n)$, the Jacobian of F is,

$$\frac{\partial(u_1, \ldots, u_n)}{\partial(x_1, \ldots, x_n)} = \det[DF(x, y)] = \det \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

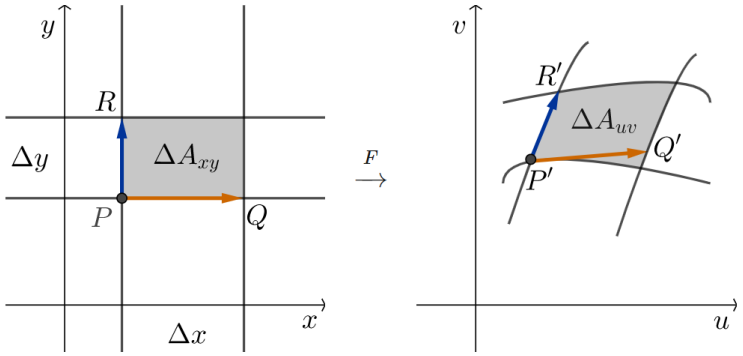
The inverse property of the Jacobian also generalizes:

$$\frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)} = \frac{1}{\frac{\partial(u_1, \ldots, u_n)}{\partial(x_1, \ldots, x_n)}}$$

where $\frac{\partial(x_1,\ldots,x_n)}{\partial(u_1,\ldots,u_n)}$ is the Jacobian of the inverse mapping of F.

13.1 Geometric Interpretation of the Jacobian

In 2D: Recall from linear algebra that the area of a parallelogram defined by the vectors $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ is given by $\left|\det \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}\right|$. Given a parallelogram induced by vectors $\vec{PQ}$ and $\vec{PR}$ in the xy-plane, the mapping F sends it to some other region of space in the uv-plane and this other region may or may not be a parallelogram,



The image under F of the rectangle defined by $\vec{PQ}, \vec{PR}$ can be approximated as a parallelogram defined by the vectors $\vec{P'Q'}, \vec{P'R'}$. We can use the linear approximation to approximate $\vec{P'Q'}, \vec{P'R'}$. Since $\vec{PQ} = \begin{bmatrix} \Delta x \\ 0 \end{bmatrix}, \vec{PR} = \begin{bmatrix} 0 \\ \Delta y \end{bmatrix}$, we obtain,

$$\vec{P'Q'} \approx \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} \Delta x \\ 0 \end{bmatrix} = \begin{bmatrix} u_x \Delta x \\ v_x \Delta x \end{bmatrix} \qquad \vec{P'R'} \approx \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} 0 \\ \Delta y \end{bmatrix} = \begin{bmatrix} u_y \Delta y \\ v_y \Delta y \end{bmatrix}$$

for $\Delta x, \Delta y$ sufficiently small. Note the partial derivatives are evaluated at P and $\Delta A_{uv} = \Delta x \Delta y$, so,

$$\Delta A_{uv} \approx \left|\det \begin{bmatrix} u_x \Delta x & u_y \Delta y \\ v_x \Delta x & v_y \Delta y \end{bmatrix}\right| = \left|\det \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}\right| \Delta x \Delta y$$

since $\Delta x, \Delta y$ are positive. Then by the definition of the Jacobian,

$$\Delta A_{uv} \approx \left|\frac{\partial(u,v)}{\partial(x,y)}\right| \Delta A_{xy}$$

where the Jacobian is evaluated at P. Thus, the Jacobian of a mapping F describes the extent to which F increases or decreases areas. It can be thought of as a scaling factor for (very small) areas that are mapped by F. The relations derived above are increasingly accurate as $\Delta x, \Delta y$ approach 0.

Note that for a linear mapping $(u, v) = F(x, y) = (ax+by, cx+dy)$ where a, b, c, d are constants, the derivative matrix is,

$$DF(x,y) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and thus the linear approximation is exact by Taylor's Theorem since all second partial derivatives are zero. So for a linear mapping, the approximation given above is exact and,

$$\Delta A_{uv} = \left|\frac{\partial(u,v)}{\partial(x,y)}\right| \Delta A_{xy}$$

In 3D: The results using the linear approximation and Jacobian above also hold in 3D. Consider a mapping $(u, v, w) = F(x, y, z) = (f(x, y, z), g(x, y, z), h(x, y, z))$. The rectangular volume in xyz-space denoted ΔV_{xyz} is equal to $\Delta x \Delta y \Delta z$ and the approximate volume of ΔV_{uvw} is, $\Delta V_{uvw} \approx \left|\frac{\partial(u,v,w)}{\partial(x,y,z)}\right| \Delta V_{xyz}$ where $\frac{\partial(u,v,w)}{\partial(x,y,z)}$ is the Jacobian of the mapping F evaluated at (x, y, z).

13.2 Constructing Mappings

Later when performing change of variables in double and triple integrals, it will be very important to invent an invertible mapping which transforms one region to another, simpler region. For example, a parallelogram into a square, an ellipse into a unit circle, etc. See pages 162, 163, 164 for examples.

---

14 Double Integrals

14.1 Definition

Recall in single variable calculus, the definite integral is defined as a limit of Riemann sums,

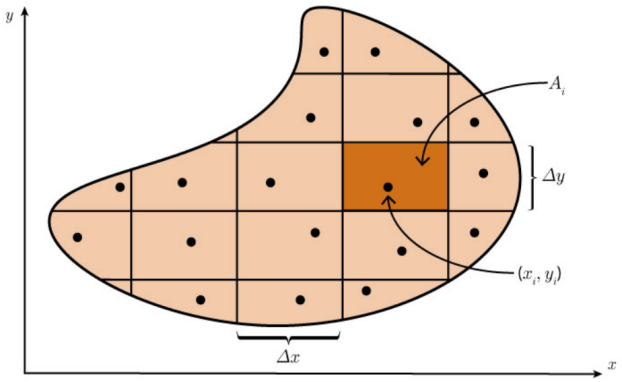
$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)\Delta x_i$$

where Δx_i is the length of the ith subinterval in some partition of the interval $[a,b]$ and x_i is some reference point in the ith subinterval. Besides calculating areas under curves, the single integral can also be used for finding the mass of rods, volumes of revolution, etc. Double integrals can be used to find the volume of more complicated regions.

Let D be a closed and bounded set in $\mathbb{R}^2$ whose boundary is a piecewise smooth closed curve. Let $f(x,y)$ be a function which is bounded on D (there exists M such that $|f(x,y)|\leq M$ for all $(x,y)\in D$. Subdivide D using straight lines parallel to the axes, forming a partition P of D and label the n rectangles that lie completely in D, in some specific order, and denote their areas ΔA_i, $i=1,\ldots,n$. Choose a point (x_i,y_i) in the ith rectangle and form the Riemann sum,

$$\sum_{i=1}^{n} f(x_i,y_i)\Delta A_i = \sum_{i=1}^{n} f(x_i,y_i)\Delta x_i \Delta y_i$$

Visualized,



Integrable functions: Let $D\subset\mathbb{R}^2$ be closed and bounded and let P be a partition of D as described above and let $|\Delta P|$ be the length of the longest side of all rectangles in the partition P. A function $f(x,y)$ which is bounded on D is **integrable** on D if all Riemann sums approach the same value as $|\Delta P|\to 0$.

Double integral: If $f(x,y)$ is integrable on a closed bounded set D, then the double integral of f on D is,

$$\iint_D f(x,y)\,dA = \lim_{\Delta P\to 0}\sum_{i=1}^{n} f(x_i,y_i)\Delta A_i$$

If f is continuous on D, it can be shown that f is integrable on D (the double integral exists). However, functions which are continuous on D may still be integrable on D. For example, if f is piece-wise continuous and continuous on D except at points which lie on a curve C, then f is integrable. Proofs of these results beyond the scope of this course.

Double integral as area: The simplest interpretation of the double integral symbol $\iint_D$ is when we specialize f to be the constant function with value 1 ($f(x,y) = 1$ for all $(x,y) \in D$). Then, the Riemann sum $\sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$ simply sums the areas of all rectangles in D and the double integral defines the area $A(D) = \iint_D 1 \, dA$ of the set D.

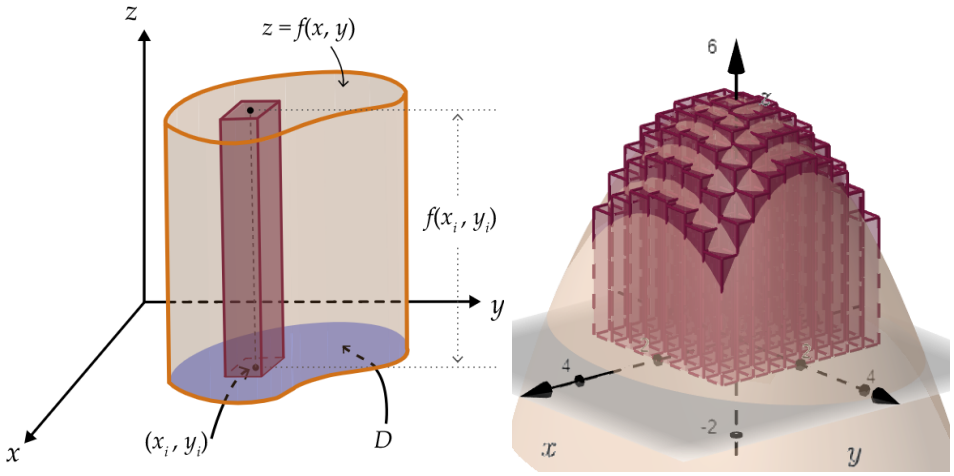
Double integral as volume: If $f(x,y) \geq 0$ for all $(x,y) \in D$, then the double integral can be interpreted as the volume $V(S)$ of the region defined by,

$$S = \{(x,y,z) | 0 \leq z \leq f(x,y), (x,y) \in D\}$$

which represents the solid below the surface $z = f(x,y)$ and above the set D in the xy-plane. To see why this is case, note that the partition P of D decomposes the solid S into vertical "columns". The height of the column above the ith rectangle is approximately $f(x_i, y_i)$ and so its volume is approximately $f(x_i, y_i)\Delta A_i$. The Riemann sum $\sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$ thus approximates the volume $V(S)$,

$$V(S) \approx \sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$$

Visualized,



As $|\Delta P| \to 0$, the partition becomes increasingly fine, so the error in the approximation tends to zero. Thus,

$$V(S) = \iint_D f(x,y) \, dA$$

We will see later triple integrals can also be interpreted as volumes.

Double integral as mass: Consider a thin flat plate of metal whose density varies with position. Since the plate is thin, the varying density can be described by an "area density" (a function $f(x,y)$ that gives the mass per unit area at position (x,y)). Ie, the mass of a small rectangle of area ΔA_i located at position (x_i, y_i) is approximately,

$$\Delta M_i \approx f(x_i, y_i)\Delta A_i$$

The Riemann sum corresponding to a partition P of D approximates the total mass M of the plate D and the double integral (being the limit of the sum) represents the total mass,

$$M = \iint_D f(x,y) \, dA$$

Double integral as probability: Let $f(x,y)$ be the pdf of a continuous 2D random variable (X,Y). The probability that $(X,Y) \in D$, a given subset of $\mathbb{R}^2$ is,

$$P((X,Y) \in D) = \iint_D f(x,y)\, dA$$

Double integral as average value: Recall for a function of one variable $f(x)$, the average value of f over an interval $[a,b]$ is,

$$\frac{1}{b-a}\int_a^b f(x)\, dx$$

Similarly, for $f(x,y)$, the average value of f over a closed and bounded subset D of $\mathbb{R}^2$ is,

$$\frac{1}{A(D)}\iint_D f(x,y)\, dA$$

14.2 Properties

The basic properties of single integrals can be generalized to double integrals. Let $D \subset \mathbb{R}^2$ be a closed and bounded set and let f, g be two integrable functions on D.

1. For any constant c,

$$\iint_D (f+g)\, dA = \iint_D f\, dA + \iint_D g\, dA \qquad \iint_D cf\, dA = c\iint_D f\, dA$$

2. If f, g are such that $f(x,y) \leq g(x,y)$ for all $(x,y) \in D$, then,

$$\iint_D f\, dA \leq \iint_D g\, dA$$

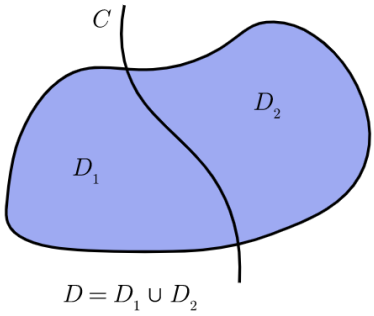
   This property can be used to estimate a double integral that is difficult (or impossible) to evaluate exactly.

3. Absolute value also works as expected,

$$\left|\iint_D f\, dA\right| \leq \iint_D |f|\, dA$$

4. Decomposition property. If D is decomposed into two closed and bounded subsets D_1, D_2 by a piecewise smooth curve C, then,

$$\iint_D f\, dA = \iint_{D_1} f\, dA + \iint_{D_2} f\, dA$$



   This property is useful when dealing with complicated regions of integration and discontinuous integrands.

14.3 Iterated Integrals

For sufficiently simple functions and regions of integration, double integrals can be evaluated exactly using methods which work for single integrals. The idea is to write the double integral as a succession of two single integrals, called an **integrated integral**. We will derive a method for doing this using the interpretation of the double integral as volume.

Theorem: Let $D \subset \mathbb{R}^2$ such that $y_l(x) \leq y \leq y_u(x)$ and $x_l \leq x \leq x_u$ where $y_l(x), y_u(x)$ are continuous for $x_l \leq x \leq x_u$. If $f(x,y)$ is continuous on D, then,

$$\iint_D f(x,y)\, dA = \int_{x_l}^{x_u} \int_{y_l(x)}^{y_u(x)} f(x,y)\, dy\, dx$$

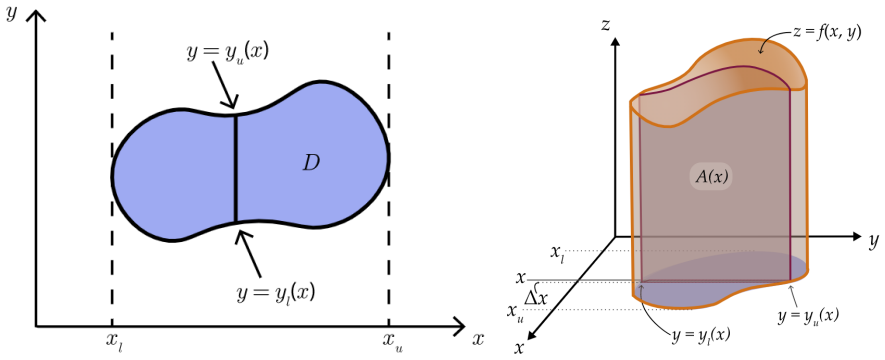
Note that the inner integral must be evaluated first. We are integrating with respect to y while holding x constant (**partial integration**).

Proof: Outside of the scope of this course. But to see intuitively why this works for f such that $f(x,y) \geq 0$ for all $(x,y) \in D$, let V denote the volume of the solid above D and below the surface $z = f(x,y)$, then,

$$V = \iint_D f(x,y)\, dA$$

So, denoting the cross-sectional area of the solid for any fixed value of x as $A(x)$,

$$V = \int_{x_l}^{x_u} A(x)\, dx \qquad A(x) = \int_{y_l(x)}^{y_u(x)} f(x,y)\, dy$$



Equating the two expressions for V,

$$\iint_D f(x,y)\, dA = \int_{x_l}^{x_u} \int_{y_l(x)}^{y_u(x)} f(x,y)\, dy\, dx$$

Theorem: Similarly to above, suppose D can be described by $x_l(y) \leq x \leq x_u(y)$ and $y_l \leq y \leq y_u$, where y_l, y_u are constants and $x_l(y), x_u(y)$ are continuous functions of y on the interval $[y_l, y_u]$. Then, the iterated integral can be written in the order "x first, then y",

$$\iint_D f(x,y)\, dA = \int_{y_l}^{y_u} \int_{x_l(y)}^{x_u(y)} f(x,y)\, dx\, dy$$

Depending on the shape of the region D and the form of the integrand $f(x,y)$, it may be easier (or even essential) to use one order of integration instead of the other (express upper and lower bounds of y in terms of x or vice versa). Note that for notation, $\int_a^b \int_c^d f(x,y)\, dx\, dy$ is equivalent to $\int_a^b dy \int_c^d f(x,y)\, dx$ (moving differentials to avoid confusion about the order of integration).

14.4 Change of Variable Theorem

A well-chose mapping F from $\mathbb{R}^2$ to $\mathbb{R}^2$ can be used to simplify a double integral $\iint_{D_{xy}} G(x,y)\,dA$ either by changing the integrand $G(x,y)$ or by deforming the set D_{xy} in the xy-plane into a simpler shape D_{uv} in the uv-plane. The latter is called a **change of variables**. Note it is convenient to replace dA with $dx\,dy$ when working in the xy-plane and $du\,dv$ when working in the uv-plane.

Recall the Jacobian can be used to create a formula which describes how areas are related under a mapping F given by $(x,y) = F(u,v) = (f(u,v), g(u,v))$,

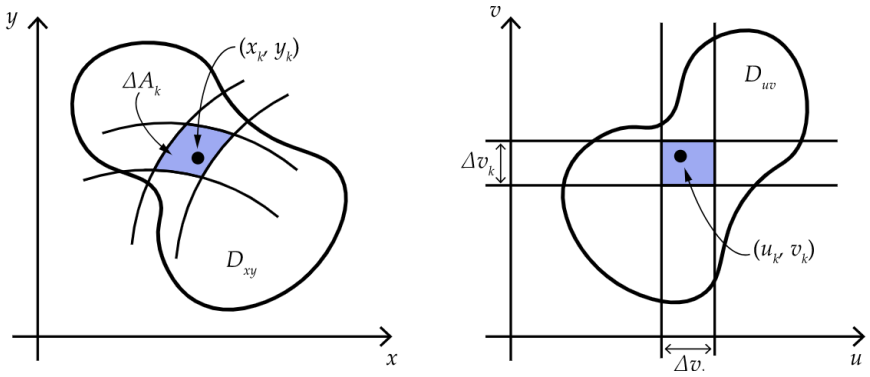
$$\Delta A_{xy} \approx \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \Delta A_{uv}$$

for $\Delta u, \Delta v$ sufficiently small where the Jacobian is evaluated at a point in the region. This idea inspires the following theorem.

Change of Variables Theorem: Let D_{uv}, D_{xy} be closed bounded sets whose boundary is a piecewise smooth closed curve. Let $(x,y) = F(u,v) = (f(u,v), g(u,v))$ be a one-to-one mapping of D_{uv} onto D_{xy} with $f, g \in C^1$ and $\frac{\partial(x,y)}{\partial(u,v)} \neq 0$ except for possibly on a finite collection of piecewise smooth curves in D_{uv}. If $G(x,y)$ is continuous on D_{xy}, then,

$$\iint_{D_{xy}} G(x,y)\,dx\,dy = \iint_{D_{uv}} G(f(u,v), g(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv$$

Proof beyond scope of this course. This thoerem tells us that under certain reasonable conditions, we can change variables in such a way as to simplify a double integral and use the Jacobian as the local scaling factor. The overall idea behind this theorem is to transform the region so that you can create an iterated integral which is easy to evaluate.



In the picture above, we consider a partition P of D_{uv} into rectangles using straight lines parallel to the coordinate axes. The images of these lines under the given transformation will generally be two families of curves which define a partition P^* of D_{xy} into elements of the area which are approximately parallelograms,

$$\begin{aligned}
\iint_{D_{xy}} G(x,y)\,dx\,dy &= \lim_{\Delta P^* \to 0} \sum_{i=1}^{n} G(x_i, y_i) \Delta A_i \\
&= \lim_{\Delta P^* \to 0} \sum_{i=1}^{n} G(f(u_i,v_i), g(u_i,v_i)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right|_{(u_i,v_i)} \Delta A_i \\
&= \iint_{D_{uv}} G(f(u,v), g(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| dA
\end{aligned}$$

This is not a real proof, since we are using an approximation in the second line.

Double integrals in polar coordinates: If the boundary of the region is a circle centered on the origin or a circle that passes through the origin, it will often help to transform from polar to Cartesian coordinates, $(x, y) = F(r, \theta) = (r\cos\theta, r\sin\theta)$, which has the Jacobian $\frac{\partial(x,y)}{\partial(r,\theta)} = r$. We must restrict $r > 0$ so that the mapping is one-to-one and the Jacobian is non-zero in order to apply the Change of Variables Theorem. We can make this restriction even if the origin is in the region as the integral over a single point is 0.

Overall, to solve double integrals over a region D, the goal is always to write the integral as an iterated integral. If D does not allow us to write an iterated integral, we can try to change variables using the Change of Variables Theorem or split D into a disjoint union using the Decomposition Theorem. If D allows us to write an iterated integral but we are still stuck, we can try to change the order of integration or rewrite D to set up a different iterated integral or perform a change of variables to set up a different iterated integral.

---

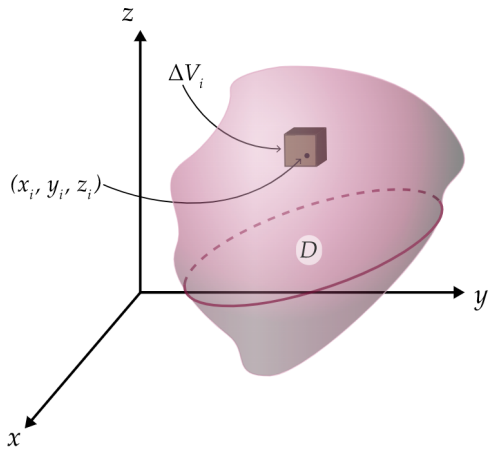
15 Triple Integrals

15.1 Definition

Let D be a closed bounded set in $\mathbb{R}^3$ whose boundary consists of a finite number of surface elements which are smooth except possibly at isolated points. Let $f(x, y, z)$ be a function bounded on D. Subdivide D by means of three families of planes which are parallel to the xy, yz, xz planes respectively, forming a partition P of D. Then, label the N rectangular blocks that lie completely in D in some specific order, and denote their volumes by $\Delta V_i, i = 1, \ldots, n$. Choose an arbitrary point (x_i, y_i, z_i) in the ith block, $i = 1, \ldots, n$, and form the Riemann sum,

$$\sum_{i=1}^{n} f(x_i, y_i, z_i)\Delta V_i$$

Visually,



Let ΔP denote the maximum of the dimensions of all rectangular blocks in the partition P.

Integrable function: A function $f(x, y, z)$ bounded on a closed bounded set $D \subset \mathbb{R}^3$ is **integrable** on D iff all Riemann sums approach the same value as $\Delta P \to 0$. This is a very similar definition to the 2D case.

Triple integral: If $f(x, y, z)$ is integrable on a closed bounded set D, then the triple integral of f over D is,

$$\iiint_D f(x, y, z)\, dV = \lim_{\Delta P \to 0} \sum_{i=1}^{n} f(x_i, y_i, z_i)\Delta V_i$$

If f is continuous on D, it can be shown that f is integrable on D. However, functions which are discontinuous in D may still be integrable on D. For example, if f is continuous on D except at points which lie on a surface or curve in D, then f is integrable on D. Proofs beyond the scope of this course.

The triple integral symbol $\iiint_D$ stands for the limit of a Riemann sum and is itself a mathematically defined object with many interpretations, depending on the meaning assigned to the integrand $f(x, y, z)$. The dV should remind us of the volume of a rectangular block in a partition of D.

Triple integral as volume: Suppose $f(x, y, z) = 1$ for all $(x, y, z) \in D$. Then, the Riemann sum $\sum_{i=1}^{n} f(x_i, y_i, z_i)\Delta V_i$ simply sums the volumes of all rectangular blocks in D and the triple integral over D defines the volume $V(D)$ in the set,

$$V(D) = \iiint_D 1\, dV$$

Triple integral as mass: Consider a star whose density varies with position. Let D denote the subset of $\mathbb{R}^2$ occupied by the star. Let $f(x, y, z)$ denote the density (mass per unit volume) at position (x, y, z). The mass of a small rectangular block located within the star at position (x_i, y_i, z_i) will be approximately,

$$\Delta M_i \approx f(x_i, y_i, z_i)\Delta V_i$$

Thus, the Riemann sum corresponding to a partition P of D will approximate the total mass M of the star, and the triple integral represents the total mass,

$$M = \iiint_D f(x, y, z)\, dV$$

Average value of a function: Let $D \subset \mathbb{R}^3$ be closed and bounded with volume $V(D) \neq 0$ and let $f(x, y, z)$ be a bounded and integrable function on D. The average value of f over D is defined by,

$$f_{avg} = \frac{1}{V(D)} \iiint_D f(x, y, z)\, dV$$

This is very similar to the average value interpretation for double integrals, except replacing area by volume.

15.2 Properties

Let $D \subset \mathbb{R}^3$ be a closed and bounded set and f, g be two integrable functions on D. Then,

1. Linearity: let c be a constant. Then,

$$\iiint_D (f + g)\, dV = \iiint_D f\, dV + \iiint_D g\, dV \qquad \text{and} \qquad \iiint_D cf\, dV = c \iiint_D f\, dV$$

2. Basic inequality: suppose f, g are such that $f(x, y, z) \leq g(x, y, z)$ for all $(x, y, z) \in D$. Then,

$$\iiint_D f\, dV \leq \iiint_D g\, dV$$

   This property is often useful to estimate a triple integral that cannot be evaluated exactly.

3. Absolute value inequality:

$$\left| \iiint_D f\, dV \right| \leq \iiint_D |f|\, dV$$

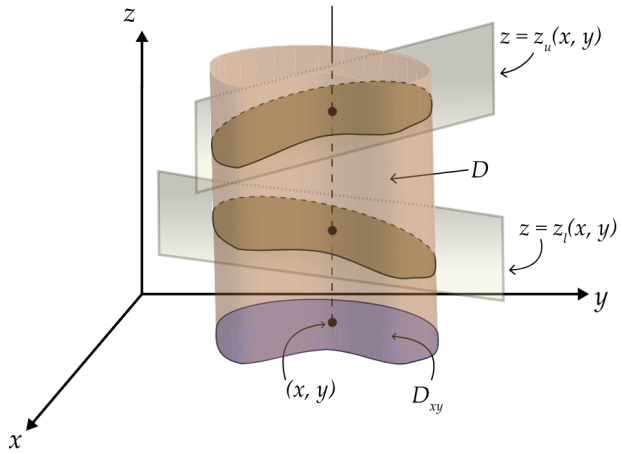
4. Decomposition property: if D is decomposed into two closed and bounded subsets D_1, D_2 by a piecewise smooth surface C, then,

$$\iiint_D f\, dV = \iiint_{D_1} f\, dV + \iiint_{D_2} f\, dV$$

   This property is useful for dealing with complicated regions of integration and discontinuous integrands.

15.3 Iterated Integrals

Consider a set $D \subseteq \mathbb{R}^3$ described by inequalities of the form $z_l(x, y) \leq z \leq z_u(x, y)$ and $(x, y) \in D_{xy}$. Here, D_{xy} is a closed bounded subset of $\mathbb{R}^2$ whose boundary is a piecewise smooth closed curve and z_l, z_u are continuous functions on D_{xy}. We can think of D as being the 3D region with bottom surface $z = z_l(x, y)$, top surface $z = z_u(x, y)$, where the extent is defined by the 2D set D_{xy}. Visually,



To write a triple integral as an iterated integral, we take an arbitrary point $(x, y) \in D_{xy}$, then integrate $f(x, y, z)$ with respect to z from $z_l(x, y)$ to $z_u(x, y)$ and integrate the result over D_{xy}, as a double integral. This procedure essentially sums over all the rectangle blocks in a partition of D and hence gives the triple integral over D.

Iterated integrals: Let D, z_l, z_u, D_{xy} as described above. If $f(x, y, z)$ is continuous on D, then,

$$\iiint_D f(x, y, z)\, dV = \iint_{D_{xy}} \int_{z_l(x,y)}^{z_u(x,y)} f(x, y, z)\, dz dA$$

As with double integrals, we are doing partial integration. The inner integral of the iterated integral is evaluated holding x, y constant and with respect to z. As well, it is not essential to integrate with respect to z first. Suppose D is described by the inequalities $x_l(y, z) \leq x \leq x_u(y, z)$ with $(y, z) \in D_{yz}$. Then,

$$\iiint_D f(x, y, z)\, dV = \iint_{D_{yz}} \int_{x_l(y,z)}^{x_u(y,z)} f(x, y, z)\, dx dA$$

If D is described by the inequalities $y_l(x, z) \leq y \leq y_u(x, z)$ with $(x, z) \in D_{xz}$, then,

$$\iiint_D f(x, y, z)\, dV = \iint_{D_{xz}} \int_{y_l(x,z)}^{y_u(x,z)} f(x, y, z)\, dy dA$$

In total, there are six ways to write a triple integral as an iterated integral, since fixing each of three variables as the first (inner most integral), there are two ways to iterate over the remaining two.

15.4 Change of Variables Theorem

A mapping $F : \mathbb{R}^3 \to \mathbb{R}^3$ can be used to simplify a triple integral $\iiint_{D_{xyz}} G(x, y, z)\, dV$ either by changing the integrand $G(x, y, z)$ or by deforming the set D_{xyz} in xyz-space into a simpler shape D_{uvw} in uvw-space, thereby simplifying the limits of integration. It is convenient to replace dV with $dxdydz$ if in xyz-space and $dudvdw$ if in uvw-space.

Change of Variable Theorem: Let $x = f(u, v, w), y = g(u, v, w), z = h(u, v, w)$ be a one-to-one mapping of D_{uvw} onto D_{xyz} with f, g, h having continuous partials, and

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} \neq 0$$

on D_{uvw}. If $G(x,y,z)$ is continuous on D_{xyz}, then,

$$\iiint_{D_{xyz}} G(x,y,z)\,dV = \iiint_{D_{uvw}} G(f(u,v,w),g(u,v,w),h(u,v,w))\left|\frac{\partial(x,y,z)}{\partial(u,v,w)}\right| dV$$

Proof is beyond the scope of this course, but the volume transformation formula using the Jacobian gives intuitive meaning to this theorem, as in the case of the double integral.

Recall for double integrals, if there is symmetry about the origin, it is helpful to evaluate the double integral using polar coordinates. Similarly, if we have symmetry about the z-axis or the origin in $\mathbb{R}^3$, it may be helpful to use our mappings to cylindrical or spherical coordinates.

Triple integrals in cylindrical coordinates: Recall that a mapping from Cartesian to cylindrical coordinates is,

$$x = r\cos\theta \qquad y = r\sin\theta \qquad z = z$$

with $r \geq 0, 0 \leq \theta < 2\pi$. We can verify that the Jacobian is $\frac{\partial(x,y,z)}{\partial(r,\theta,z)} = r$. Since we need a non-zero Jacobian, we must restrict $r > 0$. So, by the Change of Variable Theorem,

$$\iiint_{D_{xyz}} G(x,y,z)\,dxdydz = \iiint_{D_{r\theta z}} G(r\cos\theta, r\sin\theta, z)r\,drd\theta dz$$

Triple integrals in spherical coordinates: Recall that a mapping from Cartesian to spherical coordinates is,

$$x = \rho\sin\phi\cos\theta \qquad y = \rho\sin\phi\sin\theta \qquad z = \rho\cos\phi$$

with $\rho \geq 0, 0 \leq \phi \leq \pi, 0 \leq \theta < 2\pi$. The Jacobian is $\left|\frac{\partial(x,y,z)}{\partial(\rho,\theta,\phi)}\right| = \rho^2\sin\phi$. To ensure a non-zero Jacobian (and one-to-one mapping), we must restrict $\rho > 0$ and $0 < \phi < \pi$. This means we are not just removing one point but the entire z-axis. However, this does not affect our result as the triple integral over the z-axis is 0. The Change of Variable Theorem gives,

$$\iiint_{D_{xyz}} G(x,y,z)\,dxdydz = \iiint_{D_{r\theta\phi}} G(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)\rho^2\sin\phi\,d\rho d\theta d\phi$$

For example, if we wanted to evaluate a triple integral over D which is a spherical shell between the spheres of radius a, b centered on the origin where $a < b$, this is much simpler to do with spherical coordinates than x, y, z.

Special note: Remember you can find a volume using both triple and double integrals and using double integrals is often easier (eg finding the volume in an ellipsoid).

---

16 Other

16.1 Implicitly Defined Functions

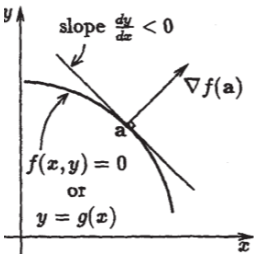
Implicit Differentiation: An equation of the form $f(x,y) = 0$ defines a relationship between the two variables x, y. If $y = g(x)$ is a solution of $f(x,y) = 0$ (ie, $f(x, g(x)) = 0$) for all x in some interval I, then g is **defined implicitly** by the equation $f(x,y) = 0$. In general, given an equation of the form $f(x,y) = 0$, it is not possible to solve for y in terms of x to obtain the function $g(x)$ explicitly. However, it is easy to calculate derivatives of g by differentiating $f(x, g(x)) = 0$ with respect to x, a process called **implicit differentiation**. In this way, one can find the linear approximation and second degree Taylor polynomial of g at a suitable reference point.

For example, the equation $y^3 - y + x = 0$ defines y implicitly as a function of $x, y = g(x)$ with $g(0) = 1$. We can then differentiate $y^3 - y + x = 0$ wrt x, treating y as a function of x to solve for $\frac{dy}{dx}$ at certain points of (x,y); $3y^2 \frac{dy}{dx} - \frac{dy}{dx} + 1 = 0$.

If $y = g(x)$ is defined implicitly by the equation $f(x,y) = 0$, where f has continuous partials, one can derive a formula for $g'(x)$ in terms of the partial derivatives of f. We have $f(x, g(x)) = 0$ for all x in some interval. So, $f(x, g(x))$ is the zero function and $\frac{d}{dx} f(x, g(x)) = 0$. Using the chain rule to expand this derivative, $f_x(x, g(x)) + f_y(x, g(x)) g'(x) = 0$ so if $f_y(x, g(x)) \neq 0$, then $g'(x) = -\frac{f_x(x, g(x))}{g_y(x, g(x))}$. In Leibniz notation, this equation reads,

$$\frac{dy}{dx} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}}$$

A simple geometrical explanation for the negative sign is the following diagram, where $\frac{\partial f}{\partial x} > 0$ and $\frac{\partial f}{\partial y} > 0$ based on the direction of the gradient vector but $\frac{dy}{dx} < 0$ based on the slope of the tangent line,



Generalization: A function $g(x,y)$ can be defined implicitly by an equation of the form $f(x,y,z) = 0$. One can use implicit differentiation to calculate the partial derivatives of g. We assume that f has continuous partial derivatives.

Implicit Function Theorem: For a function $f(x,y)$, the **null set** of f, denoted $N(f)$, is the set of points (x,y) which satisfy $f(x,y) = 0$. This set is simply the level curve of f corresponding to the constant value 0. Even if f is a "well-behaved" function, it is difficult to make any general statements about $N(f)$.

In general, for a given x-value, the equation $f(x,y) = 0$ does not have a unique solution for y and may not have any solution. However, apart from a few exceptional points, for each point $(a,b) \in N(f)$, there is a neighbourhood of (a,b) such that when restricted to this neighbourhood, $N(f)$ is the graph of a differentiable function $y = g(x)$. And the function $y = g(x)$ represents the unique solution of the equation $f(x,y) = 0$ in this neighbourhood. We can locate the exceptional points where this is not true by studying the gradient vector ∇f.

If a level set $f(x,y)=0$ is a smooth curve and (a,b) lies on the curve (ie, $f(a,b)=0$), then $\nabla f(a,b)$ is normal to the tangent line to the curve at (a,b). Thus, at the exceptional points at which the tangent line is vertical, $\nabla f=(f_x,0)$, ie $f_y=0$. If $f_y(a,b)\neq 0$, then the level set $f(x,y)=0$ is the graph of the function $y=g(x)$ in some neighbourhood of (a,b) or equivalently, the graph $f(x,y)=0$ has a unique solution $y=g(x)$.

Implicit Function Theorem: This theorem formalizes the ideas above. Let $f(x,y)\in C^1$ in a neighbourhood of (a,b). If $f(a,b)=0$ and $f_y(a,b)\neq 0$, then there exists a neighbourhood of (a,b) in which the equation $f(x,y)=0$ has a unique solution for y in terms of $x, y=g(x)$, where g has a continuous derivative. This also works if x,y are interchanged. Specifically, if $f_x(a,b)\neq 0$, then $f(x,y)=0$ has a unique solution for $x, x=h(y)$.

Corollary: If $f(x,y)\in C^1$ and $f(a,b)=0, \nabla f(a,b)\neq(0,0)$, then near the point (a,b), the equation $f(x,y)=0$ describes a smooth curve, whose tangent line at (a,b) is orthogonal to $\nabla f(a,b)$. If $f_y(a,b)\neq 0$, then the curve can be written uniquely in the form $y=g(x)$ and if $f_x(a,b)\neq 0$, it can be written uniquely in the form $x=h(y)$.

Generalization: The above can be applied to an equation of the form $f(x,y,z)=0$. The geometric interpretation is in terms of surfaces in $\mathbb{R}^3$.

Implicit Function Theorem: Let $f(x,y,z)\in C^1$ in a neighbourhood of (a,b,c). If $f(a,b,c)=0$ and $f_z(a,b,c)\neq 0$, then there exists a neighbourhood of (a,b,c) in which the equation $f(x,y,z)=0$ has a unique solution for z in terms of x and y, $z=g(x,y)$, where $g\in C^1$.

Corollary: If $f(x,y,z)$ has continuous partial derivatives and $f(a,b,c)=0, \nabla f(a,b,c)\neq(0,0,0)$, then near the point (a,b,c), the equation $f(x,y,z)=0$ describes a smooth surface in $\mathbb{R}^3$ whose tangent plane a (a,b,c) is orthogonal to $\nabla f(a,b,c)$.

If $f_z(a,b,c)\neq 0$, then the surface can be described uniquely in the form $z=g(x,y)$ near the point (a,b,c). In general however, the equations $f(x,y,z)=0$ will not be the graph $z=g(x,y)$ of one function g.

For example, $f(x,y,z)=x^2+y^2+z^2-1=0$ represents a sphere and can thus be described by the graphs of two functions $z=\sqrt{1-x^2-y^2}$ and $z=-\sqrt{1-x^2-y^2}$.

Note that when applying the Implicit Function Theorem, it is easy to remember which partial derivative of f must be non-zero; it is the partial derivative with respect to the variable for which one wishes to solve.

16.2 Quick Maffs

1. The derivative of $|x|$ is $\frac{|x|}{x}$

16.3 Common Shapes

1. Circle: $(x-h)^2+(y-k)^2=r^2$, centered at (h,k) with radius r
2. Parabola: $y=ax^2+bx+c=a(x-h)^2+k$, min/max at (h,k)
3. Horizontal Hyperbola: $\frac{(x-h)^2}{a^2}-\frac{(y-k)^2}{b^2}=1$, centered at (h,k) with asymptotes $y=\frac{b}{a}(x-h)+k$ and $y=-\frac{b}{a}(x-h)+k$
4. Vertical Hyperbola: $\frac{(y-k)^2}{a^2}-\frac{(x-h)^2}{b^2}=1$, same center and asymptotes as above
5. Ellipse: $\frac{(x-h)^2}{a^2}+\frac{(y-k)^2}{b^2}=r^2$, centered at (h,k) with length of axis $2a$ in the x direction and $2b$ in the y direction. If $a=b$, this is a circle.

16.4 Inequalities

1. Triangle inequality $|x+y| \leq |x|+|y|$
2. Triangle inequality variation $|x-y|^2 \leq (|x|+|y|)^2$
3. When using Squeeze Theorem, the most common ones are the triangle inequality and cosine inequality $(2|x||y| \leq x^2+y^2)$
4. A common mistake is to say that if $c>0$, then $|a+c| \leq a+c$
5. $|a| = \sqrt{a^2}$
6. If $c>0$, then $a < a+c$. A common usage is $|x| = \sqrt{x^2} \leq \sqrt{x^2+y^2}$
7. $x < x^2$ is false for $|x| < 1$
8. $2|x||y| \leq x^2+y^2$ for all x, y. Can be proven starting from $0 \leq (|x|-|y|)^2$.

These are mostly obvious, it is spotting when to use them that can sometimes be difficult.

16.5 Inputting Math in Mobius

1. Use `exp(x)` for e^x
2. Use `Pi` for π
3. Use `infinity` and `-infinity` for $\infty, -\infty$
4. Use `I` for i
5. Use `theta, alpha, beta` for θ, α, β
6. Use `M[p]` for subscripts M_p
7. Use `sqrt(x)` for $\sqrt{x}$
8. Use `abs(-4)` for $|-4|$
9. Use `(-infinity, 5]` for $(-\infty, 5]$
10. For scientific notation, use `7e1` for 70
11. Multiplication must be explicitly written with an asterisk, $2x$ is `2*x`. Use the carat ^ for exponents and forward slash / for fractions with brackets around the numerator and denominator.
12. If a question has more than one answer, separate using commas
13. Use `<5, 3>` for vectors $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$
14. Use `<<1, 0> | <-1, 2>>` for matrices $\begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}$; note that you enter by column, not by row