21-268: Multi-dimensional Calculus
Spring 2012

Department of Mathematical-Sciences,
Carnegie Mellon University.

Preface

These notes were typed in 2012 when *21-268: Multi-dimensional Calculus* was taught by *Russell Schwab* in the Department of Mathematical-Sciences at Carnegie-Mellon university. The notes were typed by *Christopher Almost*.

The LaTeX source for these notes is currently publicly hosted at GitLab: https://gitlab.com/gi1242/cmu-math-268. If you are interested in modifying these notes, please contact the current maintainer to discuss the best method.

Contents

0 Linear algebra review

1. (a) Recall that λ and v are respectively an eigenvalue and associated eigenvector of the linear map T if… *If $v \neq 0$ and $Tv = \lambda v$.*
   (b) Implicitly, assumptions were made in (a) about the domain and range of T. What were they? *They must be the same vector space.*
2. Find the eigenvalues of $A = \left[\begin{smallmatrix} -2 & 5 \\ 0 & 4 \end{smallmatrix}\right]$. *They are -2 and 4 because A is upper triangular and those are the diagonal elements.*
3. Let $p \in \mathbb{R}^n$ be a fixed, non-zero, vector.
   (a) What space should v be in for $p \cdot v$ to make sense? *The dual space of $\mathbb{R}^n$, which is canonically isomorphic to $\mathbb{R}^n$.*
   (b) What is the solution space of all v with $p \cdot v = 0$? Is it a vector space? Why? What is the dimension, and why? *It is the hyperplane of all vectors perpendicular to p, a vector space of dimension $n-1$.*
4. (a) To which *diagonal matrix*, say B, is A in question 2 similar (conjugate)? *It is similar to $B = \left[\begin{smallmatrix} -2 & 0 \\ 0 & 4 \end{smallmatrix}\right]$ because its eigenvalues are distinct and hence A is diagonalizable.*
   (b) What does it mean for two matrices to be similar (conjugate)? *A is similar to B if there is an invertible matrix M (a change of basis) such that $A = M^{-1}BM$.*
5. Suppose D is similar (conjugate) to

$$C = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

   Evaluate $\det(D)$ and $\operatorname{tr}(D)$. *They are -2 and 2, respectively, because those are the answers for C, and both* det *and* tr *are invariant for conjugation.*
6. Does it make sense to talk about eigenvectors of

$$E = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}.$$

   Why or why not? *No, because the domain of E is $\mathbb{R}^4$ but its range is $\mathbb{R}^3$, which are not the same space.*

1 Functions of several real variables

1.1 Examples of multi-dimensional functions

We know how to do calculus for functions $f : \mathbb{R} \to \mathbb{R}$ or $f : [a, b] \to \mathbb{R}$ of a single real variable. Here are some examples of functions with more inputs and outputs than just one.

1.1.1 Examples.

1. Temperature in a physical space (e.g. this room). The input is the Cartesian coordinate of the point in the room, $x = (x_1, x_2, x_3)$, and the output is $T(x)$, the temperature. Here $T : \mathbb{R}^3 \to \mathbb{R}$ or $T : \Omega \subseteq \mathbb{R}^3 \to \mathbb{R}$.
2. Topographical height on a map or on the Earth's surface. The input is the coordinates $x = (x_1, x_2)$, or longitude and latitude, and the output is the height, $h(x)$. Here $h : \mathbb{R}^2 \to \mathbb{R}$ or $h : S^2 \to \mathbb{R}$, where S^2 is the 2-dimensional sphere.
3. Electrostatic potential Φ and electric field E. The input is the coordinates, x, of a point in a physical space and the outputs are $\Phi : \mathbb{R}^3 \to \mathbb{R}$, the scalar electric potential (voltage) at x, and $E : \mathbb{R}^3 \to \mathbb{R}^3$, the vector indicating force on a particle due to electric field. We will see that, typically, $E = D\Phi$, the *gradient* of Φ, a topic of this course.
4. From economics, the utility function U. The inputs are the amounts of each good you could consume (e.g. hours at a pinball machine, number of apples, number of oranges, amount of fizzy beverage, etc.) The ouput is the scalar quantity that is the amount of "satisfaction" derived from that particular choice of consumption bundle. Here $U : \mathbb{R}^n_+ \to \mathbb{R}$, where n is the number of goods under consideration. $\mathbb{R}_+ = [0, \infty)$ is the collection of non-negative real numbers.

1.2 Notation and basic definitions

We need to define some notation that makes analogy with the idea of a function of a single variable $f : [a, b] \to \mathbb{R}$ in 1-dimensional calculus. The *underlying space* for this course is $\mathbb{R}^n$, n-tuples of real numbers, with *canonical basis* $\{e_1, \ldots, e_n\}$, where e_i is the n-vector with a 1 in the i^{th} position and zeros everywhere else.

The *inner product* on $\mathbb{R}^n$ is defined by

$$\langle x, y \rangle := x \cdot y = x_1 y_1 + \cdots + x_n y_n.$$

This defines a *distance* on $\mathbb{R}^n$ defined by $d(x, y)^2 := \langle x - y, x - y \rangle$, the squared length of $x - y$. We will also write $|x - y| := d(x, y)$ (absolute value bars). So the *norm* on $\mathbb{R}^n$ is defined by $|x|^2 = \langle x, x \rangle$. Yet otherwise said,

$$|x - y| = d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

An *ε-neighbourhood* of $x \in \mathbb{R}^n$ is $B_\varepsilon(x) := \{y \in \mathbb{R}^n : d(x, y) < \varepsilon\}$, the ball centred at x with radius ε, *not* including the boundary (i.e. the *open ball*).

A subset $\Omega \subseteq \mathbb{R}^n$ is an *open set* if for all $x \in \Omega$, there is some neighbourhood of x also contained in Ω (i.e. there is $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq \Omega$). This generalizes the idea that if $x \in (a, b)$ then there is always "room" between a and x and between x and b. A subset $\Omega \subseteq \mathbb{R}^n$ is *closed set* if its complement $\Omega^c := \mathbb{R}^n \setminus \Omega$ is an open set. In symbols, for all $y \notin \Omega$ there is some $\varepsilon > 0$ such that $B_\varepsilon(y) \cap \Omega = \varnothing$.

1.2.1 Example. $\Omega := \{x \in \mathbb{R}^2 : |x_1| \leq 1 \text{ and } |x_2| \leq 1\}$ is closed. It is the closed box/square $[-1,1]\times[-1,1]$. For any y outside of Ω (so $|y_1| > 1$ or $|y_2| > 1$) there is $\varepsilon > 0$ such that $B_\varepsilon(y)$ does not intersect the square.

A subset $\Omega \subseteq \mathbb{R}^n$ is *bounded* if there is some $R > 0$ such that $\Omega \subseteq B_R(0)$. (Note that you should be able to tell that R is a scalar from the context: the only collection in this course for which we will define a total ordering "$<$" is $\mathbb{R}$.)

1.2.2 Examples.

1. $\Omega := \{x \in \mathbb{R}^n : |x_i| \leq 1 \text{ for all } i = 1,\dots,n\}$ is bounded. Taking $R := \sqrt{8}$ shows this.
2. $\Omega := \{x \in \mathbb{R}^2 : x_1 x_2 < 1\}$ is the region "between" two hyperbolas in $\mathbb{R}^2$. Ω is not bounded. Indeed, it contains the line $x_2 = -x_1$.
3. Vector subspaces of $\mathbb{R}^n$ are not bounded sets. In general they are not open sets either.

A subset $\Omega \subseteq \mathbb{R}^n$ is *connected* (actually *path-connected*) if, for all $x, y \in \Omega$, there is a path γ from x to y. (A *path* γ is a continuous function $\gamma : [a,b] \to \mathbb{R}^n$, and γ is said to *go from x to y* if $\gamma(a) = x$ and $\gamma(b) = y$.) For open sets, it suffices to replace "path γ" in the definition of connected with "an ordered collection of unbroken straight line segments starting at x and ending with y."

We will say that Ω is a *domain* if it is open and (path-)connected. Domains are the sets over which we will do multi-variable calculus.

1.2.3 Example. Consider the following function $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ defined by

$$f(x_1, x_2) := \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2}.$$

What does the graph of f look like? Is f "well-behaved" near $x = 0$?

We have several tools at our disposal. The *level curves* of f are the sets $\{x \in \mathbb{R}^2 \setminus \{0\} : f(x) = c\}$ for various choices of c. The level set for $c = 0$ is the pair of lines $x_1 = x_2$ and $x_1 = -x_2$ (not including zero). The level sets for $c = 1$ and -1 are the lines $x_2 = 0$ and $x_1 = 0$, respectively. In polar coordinates,

$$f(r,\theta) = \frac{(r\cos\theta)^2 - (r\sin\theta)^2}{r^2} = \cos(2\theta).$$

Interestingly, this formula does not depend on the radius.

1.3 Limits and continuity

This section follows K§2.4. We saw that the function from Example 1.2.3 was not so well-behaved for its argument near zero, in the sense that the values the function takes on a circle centred at the origin are $\cos(2\theta)$, irrespective of the radius. Continuity is the property of a function, say f, that the values $f(x)$ can be made as close to $f(x_0)$ as we like by taking x close to x_0. The precise definition is as follows. Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a function.

1.3.1 Definition. We say that $\lim_{x\to x_0} f(x) = c$, that f has a *limit* of c at $x_0 \in \Omega$ or x_0 on the boundary of Ω, if for all $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - c| < \varepsilon$ for every $x \in \Omega$ such that $0 < |x - x_0| < \delta$. Note that we do not require that this hold at $x = x_0$. This is the δ-ε definition of limit. We say that f is continuous at $x_0 \in \Omega$ if $\lim_{x\to x_0} f(x) = f(x_0)$.

If the domain of f is $\mathbb{R}$ then there are only two directions from which to approach a point x_0 – from the left or from the right. In $\mathbb{R}^n$ there are many ways to approach a point, and more than just the $2n$ ways along the coordinate directions.

1.3.2 Example. Let f be as in Example 1.2.3 and define

$$g(x) = \begin{cases} f(x), & x \neq 0 \\ 0, & x = 0. \end{cases}$$

Let's try $x \to x_0$ for a few different 1-dimensional approaches. Let

$$\gamma(s) = (x_1(s), x_2(s))$$

for the following choices of γ.

1. If $\gamma(s) = (s, 0)$ then
$$\lim_{s\to 0} g(\gamma(s)) = \lim_{s\to 0} \frac{s^2}{s^2} = 1.$$
2. If $\gamma(s) = (0, s)$ then
$$\lim_{s\to 0} g(\gamma(s)) = \lim_{s\to 0} \frac{-s^2}{s^2} = -1.$$
3. If $\gamma(s) = (s, s)$ then
$$\lim_{s\to 0} g(\gamma(s)) = \lim_{s\to 0} \frac{0}{2s^2} = 0.$$
4. If $\gamma(s) = (s\cos s, s\sin s)$ then $\lim_{s\to 0} g(\gamma(s)) = \lim_{s\to 0} \cos(2s) = 1$.
5. Think of a function γ for which $\lim_{s\to 0} g(\gamma(s))$ does not exist.

1.3.3 Exercises.

1. Let $\gamma(s) = (s\cos(\frac{1}{s}), s\sin(\frac{1}{s}))$ and $s(0) = 0$. Describe the path γ. Is is continuous at $s = 0$? *Yes, it is continuous at $s = 0$ because*
$$\lim_{s\to 0} |\gamma(s)| = \lim_{s\to 0} \sqrt{s^2\cos^2(\tfrac{1}{s}), s^2\sin^2(\tfrac{1}{s})} = \lim_{s\to 0} s = 0 = \gamma(0).$$
The path γ spirals around $(0, 0)$, making infinitely many complete turns around it before reaching zero.
2. What happens with $f(x) = (x_1^2 - x_2^2)/(x_1^2 + x_2^2)$ under $s \mapsto f(\gamma(s))$? Does $\lim_{s\to 0} f(\gamma(s))$ exist? *The limit does not exist because $f(\gamma(s)) = \cos(\frac{2}{s})$, which is very badly behaved near $s = 0$.*

The key point is that 1-dimensional limits existing for various ways of taking x to x_0 (e.g. $\gamma(s) \to x_0$ as $s \to 0$) do not imply that the limit exists according to the $\mathbb{R}^n$ definition. Recall that if $f : \Omega_1 \to \Omega_2$ and $g : \Omega_2 \to \mathbb{R}^m$ then $g \circ f : \Omega_1 \to \mathbb{R}^m : x \mapsto g(f(x))$ is the *composition*.

1.3.4 Theorem. *If f is continuous at x_0 and g is continuous at $y_0 = f(x_0)$ then $g \circ f$ is continuous at x_0.*

A similar statement, appropriately modified, holds for limits.

$$\text{If } \lim_{x\to x_0} f(x) = c \text{ and } \lim_{y\to c} g(y) = \ell \text{ then } \lim_{x\to x_0} g \circ f(x) = \ell.$$

1.3.5 Corollary. *If $\lim_{x\to x_0} f(x) = c$ and $\gamma : [0, \varepsilon] \to \Omega$ is any path with values in Ω such that $\lim_{s\to 0} \gamma(s) = x_0$ then $\lim_{s\to 0} f(\gamma(s))$ exists and is equal to c.*

Going back to Example 1.3.2, g is not continuous at $x_0 = 0$ because the limit $\lim_{x\to 0} g(x)$ does not even exist by the corollary (because there are multiple different limits of $g(\gamma(s))$ for different choices of γ) and therefore there is no hope of $\lim_{x\to 0} g(x) = 0 = g(0)$.

2 Differentiation of multi-dimensional functions

2.1 Best affine approximation and partial derivatives

Recall that, when $f : \mathbb{R} \to \mathbb{R}$ is differentiable,

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

This implies that $c = f'(x_0)$ is the unique number such that

$$\lim_{x\to x_0} \left| \frac{f(x) - f(x_0) - c(x - x_0)}{x - x_0} \right| = 0$$

That is, the best *affine approximation* of f at x_0 is $\ell_{x_0}(x) := f(x_0) + c(x - x_0)$.

What should we do for $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$? Could we use an affine approximation to f? This is related to the material in K§2.6 and K§2.7. Recall that a linear map $\mathbb{R}^n \to \mathbb{R}$ is characterized by $p \in \mathbb{R}^n$, namely $x \mapsto p \cdot x$ is linear, and this gives all linear maps. The "best" linear approximation for f at x_0 had better at least match at x_0, i.e. we should take $\ell_{x_0}(x_0) = f(x_0)$. It follows that

$$\ell_{x_0}(x) := f(x_0) + p \cdot (x - x_0)$$

for some $p \in \mathbb{R}^n$. But which one? The best approximation will correspond to the p such that $f(x) - \ell_{x_0}(x)$ goes to zero "faster than" $(x - x_0)$ as $x \to x_0$. Our goal is hence the unique p such that

$$\lim_{x\to x_0} \frac{|f(x) - (f(x_0) + p \cdot (x - x_0))|}{|x - x_0|} = 0. \tag{1}$$

But how do we do this? Note that $x \mapsto p \cdot x$ is a linear map $\mathbb{R}^n \to \mathbb{R}$. To know p, we need to know p_i for $i = 1, \ldots, n$. But $p \cdot e_i = p_i$, so we need only look in the directions of the canonical basis. Whence, for small $h \in \mathbb{R}$,

$$\ell_{x_0}(x_0 + he_i) = f(x_0) + p \cdot (x_0 + he_i - x_0) = f(x_0) + hp_i,$$

From the definition of the derivative, with $x = x_0 + he_i$,

$$\begin{aligned} 0 &= \lim_{x \to x_0} \frac{|f(x) - (f(x_0) + p \cdot (x - x_0))|}{|x - x_0|} \\ &= \lim_{h \to 0} \frac{|f(x_0 + he_i) - (f(x_0) + hp \cdot e_i)|}{h} \\ &= \lim_{h \to 0} \left| \frac{f(x_0 + he_i) - f(x_0)}{h} - p_i \right| \end{aligned}$$

It follows that

$$p_i = \lim_{h \to 0} \frac{f(x_0 + he_i) - f(x_0)}{h}$$

This is the 1-dimensional derivative of the function $g(h) := f(x_0 + he_i)$. Think of g as the restriction of f to the 1-dimensional space $x_0 + \operatorname{span}(\{e_i\})$.

2.1.1 Theorem. *If f has ℓ_{x_0} as its best linear approximation at x_0 in the sense of* (1) *then $p_i = \lim_{h \to 0} \frac{1}{h}(f(x_0 + he_i) - f(x_0))$.*

2.1.2 Definition. The *partial derivative* of $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ at x_0 in the x_i component is $g'(0)$, where $g(h) := f(x_0 + he_i)$. It is denoted

$$\frac{\partial f}{\partial x_i}(x_0) = \lim_{h \to 0} \frac{f(x_0 + he_i) - f(x_0)}{h}.$$

How to compute the partial derivative? The function g only records changes in f the direction of e_i, so from the point of view of g the x_j for $j \neq i$ are constants. Apply the 1-dimensional rules assuming that x_j for $j \neq i$ are constant. All this material "matches" the material in K§2.5, K§2.6, and K§2.7.

2.1.3 Example. If $f(x) := x_1^2 \sin(x_2) + x_1 x_2 + x_2^2$ then $\frac{\partial f}{\partial x_1} = 2x_1 \sin(x_2) + x_2$ and $\frac{\partial f}{\partial x_2} = x_1^2 \cos(x_2) + x_1 + 2x_2$.

The best affine approximation to $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ can be represented by an $m \times n$ matrix with respect to the canonical bases as follows. Note that $f(x)$ is a vector in $\mathbb{R}^m$, so think of $f = (f_1, \ldots, f_m)^T$, where $f_i : \Omega \to \mathbb{R}$. We require $\ell_{x_0}(x) := f(x_0) + A(x - x_0)$, where

$$\lim_{x \to x_0} \frac{|f(x) - (f(x_0) + A(x - x_0))|}{|x - x_0|} = 0. \tag{2}$$

If this holds then A is unique and $A = (a_{ij})_{i=1,\dots,m,j=1,\dots,n}$, where $a_{ij} = [Ae_i]_j$ is the i^{th} component of the vector $Ae_j \in \mathbb{R}^m$. To isolate a_{ij} we perform the same basic procedure as in the $\mathbb{R}^n \to \mathbb{R}$ case, but applied to the i^{th} entries of ℓ_{x_0} and f.

$$\begin{aligned}[f(x_0+he_j)-f(x_0)-A(x_0+he_j-x_0)]_i &= [f(x_0+he_j)-f(x_0)-hAe_j]_i \\ &= f_i(x_0+he_j)-f_i(x_0)-[hAe_j]_i\end{aligned}$$

But this gives exactly the partial derivative of f_i in the x_j direction, i.e.

$$a_{ij} = \frac{\partial f_i}{\partial x_j} = \lim_{h\to 0}\frac{f_i(x_0+he_j)-f_i(x_0)}{h}$$

for $i = 1,\dots,m$ and $j = 1,\dots,n$.

2.2 Linearization and the Jacobian matrix

We say that $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is *total differentiable* or *full differentiable* at x_0 if there is an $m \times n$ matrix A such that (2) holds. We write $Df(x_0) = A$. This matrix is also known as the *Jacobian* of f at x_0.

2.2.1 Theorem. *If f is full differentiable at x_0 then $\frac{\partial f_i}{\partial x_j}(x_0)$ exists and is equal to $[Df(x_0)]_{ij}$, the $(ij)^{th}$ entry of A.*

But when is A given by the partials?

2.2.2 Theorem. *Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $x_0 \in \Omega$, and $\varepsilon > 0$ be such that $B_\varepsilon(x_0) \subseteq \Omega$. If $\frac{\partial f_i}{\partial x_j}$ all exist and are continuous on $B_\varepsilon(x_0)$ then f is full differentiable at x_0 and $[Df(x_0)]_{ij} = \frac{\partial f_i}{\partial x_j}(x_0)$.*

Note that $|\cdot|_{\mathbb{R}^m}$ controls the norms $|\cdot|_{\mathbb{R}}$ of each of the components of $f \in \mathbb{R}^m$ and also of $x \in \mathbb{R}^n$. So (2) is very strong. The information about $f_i(x_0 + he_j)$ as functions $\mathbb{R} \to \mathbb{R}$ is not enough to get the full derivative. But the extra requirement that $\frac{\partial f_i}{\partial x_j}$ exist and are continuous on a whole neighbourhood is enough.

2.2.3 Example. $\lim_{x\to 0}\cos(\frac{1}{x})$ does not exist. If the limit did exists, say was c, then $c \in [-1,1]$ because that is the range of the function cosine. Suppose that $c \neq 1$. Let $\varepsilon := \frac{1}{2}|1-c| > 0$. Then $\varepsilon > 0$ and we will show that there is no $\delta > 0$ such that $|\cos(\frac{1}{x})-c| < \varepsilon$ for all $|x| < \delta$. Let $x_n := \frac{1}{2\pi n}$. Then $x_n \to 0$ as $n \to \infty$ and

$$|\cos(\tfrac{1}{x_n})-c| = |\cos(2\pi n)-c| = |1-c| > \frac{1}{2}|1-c| = \varepsilon.$$

Therefore, no matter which $\delta > 0$ we try, there is n large enough that $|x_n| < \delta$, and $|\cos(\frac{1}{x_n})-c| > \varepsilon$. Therefore c is not the limit, so either 1 is the limit or the limit does not exist. But repeating the same argument with the assumption that $c \neq -1$ and $x_n := \frac{1}{\pi(2n+1)}$ shows that the limit cannot be 1, so the limit does not exist.

We have seen that the existence of a "best affine approximation" implies that the partial derivatives exist. That is to say, given $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x_0 \in \Omega$, if there is a unique $p \in \mathbb{R}^n$ such that

$$\lim_{x \to x_0} \frac{|f(x) - f(x_0) - p \cdot (x - x_0)|}{|x - x_0|} = 0.$$

(this is (1) again), then f has partial derivatives in all directions at x_0. We have seen that $\ell_{x_0}(x) := f(x_0) + p \cdot (x - x_0)$ is the best affine approximation to f at x_0. Functions f for which a best affine approximation exist are said to be *differentiable*. Not all functions are differentiable.

Taking $x = x_0 + he_j$ and plugging it into (1),

$$\begin{aligned}
0 &= \lim_{x \to x_0} \frac{|f(x) - \ell_{x_0}(x)|}{|x - x_0|} \\
&= \lim_{h \to 0} \frac{|f(x_0 + he_j) - f(x_0) - p \cdot (he_j)|}{|he_j|} && x - x_0 = he_j \\
&= \lim_{h \to 0} \frac{|f(x_0 + he_j) - f(x_0) - hp \cdot e_j|}{|h|} && |e_j| = 1 \\
&= \lim_{h \to 0} \left| \frac{f(x_0 + he_j) - f(x_0) - hp \cdot e_j}{h} \right| && \text{ratio of real numbers} \\
0 &= \lim_{h \to 0} \frac{f(x_0 + he_j) - f(x_0)}{h} - p \cdot e_j \\
p \cdot e_j &= \lim_{h \to 0} \frac{f(x_0 + he_j) - f(x_0)}{h}
\end{aligned}$$

This tells us both that the limit on the right exists and that it is equal to $p \cdot e_j = p_j$. We write $Df(x_0) = p$ and $\frac{\partial f}{\partial x_j}(x_0) = p_j$. We call Df the *gradient* of f at x_0 (but only when f maps to $\mathbb{R}$, in which case Df is an n-vector). The best affine approximation to f at x_0 can hence be rewritten as

$$\begin{aligned}
\ell_{x_0}(x) &= f(x_0) + Df(x_0) \cdot (x - x_0) \\
&= f(x_0) + \frac{\partial f}{\partial x_1}(x_0)(x_1 - x_{01}) + \cdots + \frac{\partial f}{\partial x_n}(x_0)(x_n - x_{0n}).
\end{aligned}$$

2.2.4 Example. Is $f(x) := |x| = \sqrt{x_1^2 + \cdots + x_n^2}$ differentiable at $x = 0$? It is not, because if it were then $\frac{\partial f}{\partial x_1}(0)$ would exist, but

$$\left| \frac{f(0 + he_1) - f(0)}{h} \right| = \frac{\sqrt{h^2}}{h} = \pm 1$$

where the $\pm$ depends on the sign of h. Therefore the limit as $h \to 0$ does not exist, so the limit defining $\frac{\partial f}{\partial x_1}(0)$ does not exist. This implies that f is not differentiable. Similarly, $g(x) = x_1^2 + |x_2|$ is not differentiable at $x = 0$ because $\frac{\partial f}{\partial x_2}(0)$ does not exist (though in this case $\frac{\partial f}{\partial x_1}(0)$ does exist).

In which direction does f increase the fastest from $f(x_0)$? Hint: first think of which direction ℓ_{x_0} increases the fastest. Since f and ℓ_{x_0} are "very close" in a neighbourhood of x_0, the answers should be the same.

The *tangent plane* to the graph of $f : \mathbb{R}^n \to \mathbb{R}$ at x_0 is the graph of the best affine approximation ℓ_{x_0}. It is the unique hyperplane that touches the graph of f at x_0 and which satisfies $\text{error}(x-x_0)/|x-x_0| \to 0$ as $x \to x_0$.

2.3 Rules of Differentiation

As with 1-dimensional derivatives, there are a few rules that the derivative satisfies.

1. If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $c \in \mathbb{R}$ then
$$D(cf)(x_0) = cDf(x_0).$$
2. If $f, g : \Omega \to \mathbb{R}$ then
$$D(gf)(x_0) = g(x_0)Df(x_0) + f(x_0)Dg(x_0).$$
3. If $f, g : \Omega \to \mathbb{R}^m$ then
$$D(f+g)(x_0) = Df(x_0) + Dg(x_0).$$
4. If $g : \Omega_1 \subseteq \mathbb{R}^\ell \to \Omega_2 \subseteq \mathbb{R}^m$ and $f : \Omega_2 \to \mathbb{R}^n$ then
$$D(f \circ g)(x_0) = Df(g(x_0))Dg(x_0).$$

The last rule is the *chain rule*. It is one of the best reasons for using the "best affine approximation" definition of the derivative.

$$\begin{aligned}
g(x) &= g(x_0) + Dg(x_0)(x-x_0) + \text{err}_g(x-x_0) \\
f(g(x)) &= f(g(x_0)) + Df(g(x_0))(g(x)-g(x_0)) + \text{err}_f(g(x)-g(x_0)) \\
&= f(g(x_0)) + Df(g(x_0))[Dg(x_0)(x-x_0) \\
&\qquad + \text{err}_g(x-x_0)] + \text{err}_f(g(x)-g(x_0)) \\
&= f(g(x_0)) + Df(g(x_0))Dg(x_0)(x-x_0) \\
&\qquad + [Df(g(x_0))\text{err}_g(x-x_0) + \text{err}_f(g(x)-g(x_0))]
\end{aligned}$$

Furthermore, because the errors in the best affine approximation go to zero faster than $|x-x_0|$ as $x \to x_0$,

$$\begin{aligned}
&\lim_{x\to x_0} \frac{|Df(g(x_0))\text{err}_g(x-x_0) + \text{err}_f(g(x)-g(x_0))|}{|x-x_0|} \\
&= \lim_{x\to x_0} |Df(g(x_0))| \frac{\text{err}_g(x-x_0)}{|x-x_0|} + \lim_{x\to x_0} \frac{\text{err}_f(g(x)-g(x_0))}{|g(x)-g(x_0)|} \frac{|g(x)-g(x_0)|}{|x-x_0|} \\
&= 0
\end{aligned}$$

Make the previous computation precise as an exercise. Note that we have used the fact that g is continuous at x_0 (to know that $g(x) \to g(x_0)$) and also that g is

differentiable at x_0 (to know that the ratio $|g(x)-g(x_0)|/|x-x_0|$ is bounded). This shows that $Df(g(x_0))Dg(x_0)$ is the correct matrix for the best affine approximation of $f \circ g$ at x_0. Check that the shapes are in fact correct for multiplying these matrices together.

2.3.1 Exercise. Suppose $f : \mathbb{R}^n \to \mathbb{R} : x \mapsto x + q \cdot x$. What is $Df(x_0)$ for any x_0? *It is always q because f is affine, so it is its own best affine approximation. Think of* $f(x) = (c + q \cdot x_0) + q \cdot (x - x_0)$.

3 Geometric applications

Given a *direction* v (so $|v| = 1$), how does f change at x_0 in the straight line through x_0 in the direction v, i.e. in the affine space $x_0 + \text{span}(\{v\})$? This is a 1-dimensional problem. Let $g(s) = x_0 + sv$ and use the chain rule on the composition $f(g(s))$. Since $Dg(0) = v$ and $f(g(0)) = f(x_0)$, the derivative is $Df(x_0) \cdot v$. This is the *directional derivative* of f at x_0 in the direction v.

3.1 Curves and paths

Notation. $C^1(\Omega)$ is the collection of functions $f : \Omega \to \mathbb{R}^m$ such that $Df(x)$ exists and is continuous for all $x \in \Omega$. Typically Ω is an open connected set, i.e. a domain. Such functions are said to be *continuously differentiable*.

A *path* is a map $\gamma : [a, b] \to \mathbb{R}^n$. The terminology of continuity and differentiability carry over from what we have been discussing for the past few weeks. We will write $\gamma' := \dot{\gamma} := D\gamma$ for paths, when appropriate, and we refer to this n-vector as the *velocity vector*. *curve* is the image left behind after tracing a path. More precisely, it is the image of a path, $\Gamma = \gamma([a, b]) = \{\gamma(t) : t \in [a, b]\}$. Note the difference between γ and Γ.

We say that Γ is *differentiable* at $x_0 \in \Gamma$ if there exists a continuously differentiable path $\gamma : [a, b] \to \Gamma$ and $t_0 \in [a, b]$ such that both $\gamma(t_0) = x_0$ and $\gamma'(t_0) \neq 0$. The *tangent line* is the unique line with direction $\gamma'(t_0)$ and which touches (i.e. is tangent to) Γ at x_0.

3.1.1 Examples.

1. Let $\gamma : [0, 4\pi] \to \mathbb{R}^2 : t \mapsto (t \cos t, t \sin t)$. The partial derivatives are

$$\gamma_1'(t) = \cos t - t \sin t \quad \text{and} \quad \gamma_2'(t) = \sin t + t \cos t,$$

   which are both differentiable and never both zero. Indeed, the sum of their squares is $1 + t^2$. Hence the image of this path is a differentiable curve.
   [Insert sketch of the spiral.]
2. Consider now the path $[-1, 1] \to \mathbb{R}^2 : t \mapsto (t^4, t^2)$. It is differentiable at all $t \in (-1, 1)$, but the image curve Γ is not differentiable at zero. It has a cusp at the origin.

[Insert image. Note that $\gamma_2 = \sqrt{|\gamma_1|}$.]

To prove that the curve Γ is not differentiable at $(0,0)$, we must show that any possible candidate $\gamma : [a,b] \to \Gamma$ that is continuously differentiable fails the requirement that $\gamma'(t_0) \neq 0$. By "shifting", we may always assume without loss of generality that $t_0 = 0$. Observe that all points $x \in \Gamma$ satisfy $x_2 = \sqrt{|x_1|}$. Therefore $\gamma(t) = (c(t), \sqrt{|c(t)|})$ for some $c : [a,b] \to \mathbb{R}$. We are requiring that γ is continuously differentiable, so there is a best affine approximation at zero, call it $v = \gamma'(0)$, such that

$$\begin{aligned}
0 &= \lim_{t\to 0} \frac{|\gamma(t) - \gamma(0) - v(t-0)|}{|t|} \\
&= \lim_{t\to 0} \frac{\sqrt{(c(t) - tv_1)^2 + (\sqrt{|c(t)|} - tv_2)^2}}{|t|} \\
&\geq \lim_{t\to 0} \frac{\sqrt{(\sqrt{|c(t)|} - tv_2)^2}}{|t|} \\
&= \lim_{t\to 0} \frac{|t| \left| \frac{\sqrt{|c(t)|}}{t} - v_2 \right|}{|t|} \\
&= \lim_{t\to 0} \left| \frac{\sqrt{|c(t)|}}{t} - v_2 \right|
\end{aligned}$$

Therefore we may conclude

$$\begin{aligned}
v_2 &= \lim_{t\to 0} \frac{\sqrt{|c(t)|}}{t} \\
v_2^2 &= \lim_{t\to 0} \frac{|c(t)|}{t^2} && x \mapsto x^2 \text{ is continuous} \\
\lim_{t\to 0} t v_2^2 &= \lim_{t\to 0} \frac{|c(t)|}{t} \\
0 &= \lim_{t\to 0} \frac{|c(t)|}{t} \\
0 &= \lim_{t\to 0} \frac{|c(t) - 0|}{t}
\end{aligned}$$

Since γ is continuously differentiable, $c'(0)$ exists, and by the limit above, it must be zero. Can it be the case that $c'(0) = 0$ but $(\sqrt{|c(t)|})'(0) \neq 0$? Again, since γ is continuously differentiable, $\sqrt{|c(t)|}$ is differentiable at $t = 0$. The rest of the proof is an exercise.

3.2 Paths on surfaces

What is a surface? Intuitively, a surface is a 2-dimensional object inside of a 3-dimensional space, e.g. a piece of paper in this room, the surface of a ball (a sphere), etc. More precisely, $S \subseteq \mathbb{R}^{n+1}$ is a *level surface* if it can be written as

$S = \{x \in \mathbb{R}^{n+1} : F(x) = c\}$, where $F : \mathbb{R}^{n+1} \to \mathbb{R}$ is continuously differentiable and c is a fixed constant, and $DF(x) \neq 0$ for all x near S. In this case we say the dimension of S is n. Note that $F : \mathbb{R}^{n+1} \to \mathbb{R}$ actually describing an n-dimensional surface is special. We must be careful with F, because not every such function gives rise to an n-dimensional surface.

3.2.1 Examples.

1. Justification for $DF \neq 0$. Let
$$F(x) = \begin{cases} 0, & |x| \leq 1 \\ (|x|-1)^2, & |x| > 1. \end{cases}$$
Then F is continuously differentiable. (Check this carefully for $|x| = 1$.)
[Insert diagram of the 3-d picture of F.]
Note $\{x : F(x) = 0\} = \{x : |x| \leq 1\}$ is not "1-dimensional" in the usual sense of the term. $DF(x) = 0$ for all x in this level set, which is why we exclude this case.
2. Now let $F(x) = |x|$. Then $DF(x) = x/|x|$ for all $n \neq 0$ and F is not differentiable at zero. For $c > 0$, the level surface associated with F and c is the circle of radius c when $n = 2$ and the sphere of radius c when $n = 3$.
3. If $f : \mathbb{R}^2 \to \mathbb{R}$ then the *graph* of f is $\{x \in \mathbb{R}^3 : x_3 = f(x_1, x_2)\}$. It can be realized as a 2-dimensional surface by taking $F : \mathbb{R}^3 \to \mathbb{R}$ with $F(x) = f(x_1, x_2) - x_3$ and $c = 0$. Note that $DF = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, -1)$ is never zero. For example, if $f(x) = x_1^2 - x_2^2$ then $F(x) = x_1^2 - x_2^2 - x_3$.
4. Another way to find curves in $\mathbb{R}^2$ other than images of paths is level surfaces of $F : \mathbb{R}^2 \to \mathbb{R}$, i.e. a 1-dimensional level surface is basically a curve. For example, if $F(x) = x_1^2 - s_2^2$ then the level surfaces are a tool for understanding the graph of F. At $c = 0$ we have $x_1^2 - x_2^2 = 0$, so $x_1 = \pm x_2$. For $c = 1$ we get $x_1^2 - x_2^2 = 1$, a hyperbola, and for $c = -1$, we get $x_1^2 - x_2^2 = -1$, another hyperbola.
[Insert diagram of level sets and of the 3-d picture.]

3.2.2 Example. Let $\gamma : [0,1] \to \mathbb{R}^3 : t \mapsto (t\cos(4\pi t), t\sin(4\pi t), \sqrt{1-t^2})$, and note that $|\gamma(t)| = 1$ for all t. This means that γ happens to always take values on the sphere $S = \{x \in \mathbb{R}^3 : |x| = 1\}$. The path of γ starts at the north pole at $t = 0$ and spirals down in an anti-clockwise direction until it reaches the equator at $t = 1$ after making two full revolutions.

In general, if γ is a path on a level surface $S = \{x : F(x) = c\}$ then, necessarily, $F(\gamma(t)) = c$ for every t in the domain of γ. A special case is a path on a graph of a function $f : \mathbb{R}^n \to \mathbb{R}$. Recall that the graph of f is an n-dimensional surface in $\mathbb{R}^{n+1}$. If $\gamma : [a,b] \to \mathbb{R}^{n+1}$ is on the graph of f then it must be the case that $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t), f(\gamma(t)))$, i.e. the $(n+1)^{\text{st}}$ component of γ is related to the first n components by f.

What is the analog of the best affine approximation for surfaces? If the level surface is the graph of a function f, then we have seen that the tangent plane at a

point x_0 is exactly the graph of the best affine approximation of f at x_0. Basically all directional derivatives are encoded in the "slope" of the tangent plane.

We determined the partial derivatives by reducing to the 1-dimensional case. For each direction e_i, we consider the function $g_i(h) = f(x_0 + he_i)$, so that the partial derivatives are given by $\frac{\partial f}{\partial x_i}(x_0) = g'(0)$. Note that g_i is a path on the graph of f that passes through x_0! Furthermore, $g'(0) = Df(x_0) \cdot e_i$. This is good material with which to formulate a definition. We say that the *tangent plane* to the graph of f is the unique plane which touches the graph at $(x_0, f(x_0))$ and contains the velocity vectors, $\dot{\gamma}_i(0)$, of each of the paths $\gamma_i(s) := (x_0 + se_i, f(x_0 + se_i))$.

Now let $S = \{x \in \mathbb{R}^{n+1} : F(x) = c\}$ be a level surface. We say that the *tangent plane* (or *tangent hyperplane*) to S at $x_0 \in S$ is the unique n-dimensional plane that contains x_0 and $x_0 + \dot{\gamma}(0)$ for every continuously differentiable path γ contained in S with $\gamma(0) = x_0$. Recall that γ is a path in S if $\gamma : [a, b] \to \mathbb{R}^{n+1}$ and $\gamma(t) \in S$ for all $t \in [a, b]$. Equivalently, if $F(\gamma(t)) = c$ for all $t \in [a, b]$. From this latter equation (and the chain rule) we will be able to calculate the equation of the tangent plane. Differentiating at $t = 0$,

$$0 = DF(\gamma(0)) \cdot \dot{\gamma}(0) = DF(x_0) \cdot \dot{\gamma}(0)$$

But how many choices are there for $\dot{\gamma}(0)$ as γ ranges over the collection of appropriate paths? The tangent plane should contain all the directions solving $DF(x_0) \cdot p = 0$, or equivalently and more precisely, the tangent plane is the affine space $x_0 + (\mathrm{span}(\{DF(x_0)\}))^{\perp}$. Yet otherwise said, x is in the tangent plane to S at $x_0 \in S$ if and only if $DF(x_0) \cdot (x - x_0) = 0$.

3.2.3 Example. Let S be the unit sphere in $\mathbb{R}^3$, $S = \{x \in \mathbb{R}^3 : |x| = 1\}$. The tangent plane at $(1, 0, 0)$ is computed as follows. In this case $F(x) = |x|$, so (compute as an exercise) $DF(x) = x/|x|$, and since $x_0 = (1, 0, 0)$, $DF(x_0) = (1, 0, 0)$.

[Insert diagram?]

The tangent plane is the unique plane perpendicular to $(1, 0, 0)$ containing $(1, 0, 0)$. (It is a coincidence of the sphere that $x_0 = DF(x_0)$.) More explicitly, the tangent plane is $\{(1, s, t) : s, t \in \mathbb{R}\}$.

4 Higher derivatives and Taylor's theorem

4.1 Higher derivatives

In the 1-dimensional case, if $f : \mathbb{R} \to \mathbb{R}$ is smooth enough then $f' : \mathbb{R} \to \mathbb{R}$ and the derivative operation can be repeated over and over. When $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, we compute Df via partial derivatives, which are each 1-dimensional derivatives. If f is smooth enough, then $\frac{\partial f}{\partial x_i}$ is differentiable, so we can take further partial derivatives. Unlike the 1-dimensional case, there is not a unique "second derivative"; there are n^2 combinations of second partial derivatives of f. We write $\frac{\partial^2 f}{\partial x_j \partial x_i}$ for the *second partial derivative* of f with respect to x_i first and then with respect

to x_j. This notation can be extended in the obvious way to k partial derivatives with respect to a list of k indices $i_1, \dots, i_k$ (in that specific order):

$$\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}.$$

An obvious question is, "Does the order really matter?"

4.1.1 Example (Important warning). Let

$$f(x) = \begin{cases} x_1 x_2 \left(\frac{x_1^2 - x_2^2}{x_1^2 + x_2^2} \right), & x \neq 0 \\ 0, & x = 0. \end{cases}$$

Away from $x = 0$, f is nice and we can figure out the derivatives using the usual rules.

$$\frac{\partial f}{\partial x_1} = \cdots \quad \text{etc.}$$

We will see when we fill in this example that, at some x_0,

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} \neq \frac{\partial^2 f}{\partial x_2 \partial x_1}.$$

"Mind blown." —Avia

4.1.2 Theorem. *If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ and both $\frac{\partial f}{\partial x_i}$ and $\frac{\partial f}{\partial x_j}$ are differentiable and if $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$ both exist and are continuous then they are equal.*

PROOF: The case $\Omega \subseteq \mathbb{R}^2$ is the only important one for the main idea. Let $x \in \Omega$ be fixed. We will show that $\frac{\partial^2 f}{\partial x_1 \partial x_2}(x)$ and $\frac{\partial^2 f}{\partial x_2 \partial x_1}(x)$ are equal to the limit of the same quantity $\lim_{y \to 0} \frac{S(y)}{y_1 y_2}$.

Assume that y is small enough that $x + y \in \Omega$. Let

$$S(y) := f(x + y) - f(x_1 + y_1, x_2) - f(x_1, x_2 + y_2) + f(x).$$

We want to isolate two separate 1-dimensional function and apply the mean value theorem for derivatives. Towards this goal, let

$$g(s) := f(s, x_2 + y_2) - f(s, x_2) \quad \text{and} \quad h(s) := f(x_1 + y_1, s) - f(x_1, s).$$

Then

$$g(x_1 + y_1) - g(x_1) = S(y) = h(x_2 + y_2) - h(x_2).$$

The mean value theorem applies, so there is $\bar{x}_1$ between x_1 and $x_1 + y_1$ such that

$$S(y) = g(x_1 + y_1) - g(x_1) = g'(\bar{x}_1) y_1 = y_1 \left(\frac{\partial f}{\partial x_1}(\bar{x}_1, x_2 + y_2) - \frac{\partial f}{\partial x_1}(\bar{x}_1, x_2) \right).$$

Apply it again, this time to the function $s \mapsto \frac{\partial f}{\partial x_1}(\bar{x}_1, s)$, to obtain $\bar{x}_2$ between x_2 and $x_2 + y_2$ such that

$$\frac{\partial f}{\partial x_1}(\bar{x}_1, x_2 + y_2) - \frac{\partial f}{\partial x_1}(\bar{x}_1, x_2) = y_2 \frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}_1, \bar{x}_2).$$

Unravelling this, $S(y) = y_1 y_2 \frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{x}_1, \bar{x}_2)$. Applying the same reasoning to h, it can be shown that $S(y) = y_1 y_2 \frac{\partial^2 f}{\partial x_1 \partial x_2}(\hat{x}_1, \hat{x}_2)$ for some $\hat{x}_i$ between x_i and $x_i + y_i$ for $i = 1, 2$. Now we appeal to the continuity of the second partial derivatives to conclude

$$\frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1, x_2) = \lim_{y \to 0} \frac{S(y)}{y_1 y_2} = \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1, x_2).$$

Note of course that $\bar{x}_i \to x_i$ and $\hat{x}_i \to x_i$ as $y_i \to 0$ for $i = 1, 2$. □

Notation. $C^k(\Omega) := \{f : \Omega \to \mathbb{R} \mid \text{all partials of } f \text{ of order up to and including } k \text{ exist and are continuous}\}$. Such functions are said to be *k-times continuously differentiable*.

Recall from 1-dimensional calculus that if f is better than C^1 then we can make better than affine approximations to f. This holds for arbitrarily high orders if f has that many derivatives. In particular, if $f \in C^k$ then there is a function $R_k(x)$ such that $R_k(x) \to 0$ as $x \to x_0$ and

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \ldots + \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^k + R_k(x)(x - x_0)^k.$$

How can we do this for $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$? For simplicity, let's consider $n = 2$. We should try to apply what we already know to the single variable functions $g_1(s) := f(x_0 + se_1)$ and $g_2(s) := f(x_0 + se_2)$. Naively, we take the partials of g_i and add them up. Is is true that

$$f(x) \approx f(x_0) + \frac{\partial f}{\partial x_1}(x_0)(x_1 - x_{01}) + \frac{1}{2} \frac{\partial^2 f}{\partial x_1^2}(x_0)(x_1 - x_{01})^2 + \frac{\partial f}{\partial x_2}(x_0)(x_2 - x_{02}) + \frac{1}{2} \frac{\partial^2 f}{\partial x_2^2}(x_0)(x_2 - x_{02})^2 ?$$

We are missing the cross terms! Obviously terms of the form $x_1 x_2$ will feature in this approximation, in general. The proof of Theorem 4.1.2 tells us which extra terms to include, namely $\frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) x_1 x_2$. We assume that f is nice enough that the mixed partials are equal.

In fact, for $f \in C^2$, the second order approximation is

$$f(x) \approx f(x_0) + Df(x_0)\cdot(x-x_0) + \frac{1}{2}(x-x_0)^T D^2 f(x_0)\cdot(x-x_0)$$

$$\text{where } D^2 f(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_0) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) & \frac{\partial^2 f}{\partial x_2^2}(x_0) \end{bmatrix}$$

Note that $D^2 f(x_0)$ is a symmetric matrix.

4.1.3 Theorem (Taylor's theorem for order 3). *Let $f \in C^3(\Omega)$ and $x_0 \in \Omega$ be given. There are functions $R_{i,j,k} : \Omega \to \mathbb{R}$, $i,j,k = 1,\dots,n$, such that $R_{i,j,k}(x) \to 0$ as $x \to x_0$ and*

$$\begin{aligned} f(x) = f(x_0) + Df(x_0)\cdot(x-x_0) + \frac{1}{2}(x-x_0)\cdot D^2 f(x_0)(x-x_0) \\ + \frac{1}{6}\sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(x_0)(x_i - x_{0i})(x_j - x_{0j})(x_k - x_{0k}) \\ + \frac{1}{6}\sum_{i,j,k=1}^{n} R_{i,j,k}(x)(x_i - x_{0i})(x_j - x_{0j})(x_k - x_{0k}) \end{aligned}$$

Aside on tangent spaces and tangent planes:

There is a key distinction between tangent spaces and tangent planes. Let $S = \{x \in \mathbb{R}^{n+1} : F(x) = 0\}$ be an n-dimensional surface in $\mathbb{R}^{n+1}$. Recall that the tangent space at x_0 is $T = \ker(DF(x_0)) = \{p \in \mathbb{R}^{n+1} : DF(x_0)\cdot p = 0\}$. It is a linear subspace of $\mathbb{R}^{n+1}$ of codimension 1. By definition, $DF(x_0)/|DF(x_0)|$ is a vector of length 1 that is perpendicular to T. Note that T contains the velocity vectors $\dot\gamma(0)$ for all paths γ with $\gamma(0) = x_0$. But T most likely does not touch T at x_0! The tangent plane is the translation of the tangent space, $x_0 + T = \{v = x_0 + p : p \in T\}$.

[Insert diagram of generic S in $\mathbb{R}^3$ with the various planes.]

Why is $DF(x_0)$ perpendicular to the surface at x_0? If γ is a path on S then $F(\gamma(s)) = 0$ for all s (by definition), so $D(F(\gamma(s))) = 0$, and the chain rule implies that, at $s = 0$, $DF(\gamma(0))\cdot\dot\gamma(0) = 0$. Hence $DF(x_0)\cdot\dot\gamma(0) = 0$, so every vector in the tangent plane is perpendicular to $DF(x_0)$.

Consider the difference between the graph of $f : \mathbb{R}^{n+1} \to \mathbb{R}$ and the level surface associated with F. The tangent plane to the graph of f at $x_0 \in \mathbb{R}^n$ is the graph of $x \mapsto f(x_0) + Df(x_0)\cdot(x - x_0)$. The function that defines a level surface that is the same as the graph of f is $F(x) = f(x_1,\dots,x_n) - x_{n+1}$. For $x_0 \in \mathbb{R}^n$, the corresponding point on $S = \{x \in \mathbb{R}^{n+1} : F(x) = 0\}$ is $(x_0, f(x_0)) \in \mathbb{R}^{n+1}$. Then $DF(x_0, f(x_0)) = (Df(x_0), -1)$, so the equation defining the tangent plane T is

$$\frac{\partial f}{\partial x_1}(x_0)p_1 + \cdots + \frac{\partial f}{\partial x_n}(x_0)p_n - p_{n+1} = 0.$$

This gives the same tangent plane.

4.1.4 Examples.

1. Give a second order polynomial approximation to $f(x) = \sin(x_1 x_2)$ at $x_0 = (1, \pi/2)$. Note that f is nice as it is the composition of smooth functions, so we must calculate Df and $D^2 f$.

$$Df(x) = (x_2 \cos(x_1 x_2), x_1 \cos(x_1 x_2))$$
$$D^2 f(x) = \begin{pmatrix} -x_2^2 \sin(x_1 x_2) & \cos(x_1 x_2) - x_1 x_2 \sin(x_1 x_2) \\ \cos(x_1 x_2) - x_1 x_2 \sin(x_1 x_2) & -x_1^2 \sin(x_1 x_2) \end{pmatrix}$$

so $Df(x_0) = (0, 0)$ and

$$D^2 f(x_0) = \begin{pmatrix} -\pi^2/4 & -\pi/2 \\ -\pi/2 & -1 \end{pmatrix}.$$

Therefore

$$\begin{aligned} \sin(x_1 x_2) &\approx 1 + \frac{1}{2} \begin{pmatrix} x_1 - 1 \\ x_2 - \pi/2 \end{pmatrix}^T \begin{pmatrix} -\pi^2/4 & -\pi/2 \\ -\pi/2 & -1 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 - \pi/2 \end{pmatrix} \\ &= 1 - \frac{\pi^2}{8}(x_1 - 1)^2 - \frac{\pi}{2}(x_1 - 1)(x_2 - \pi/2) - \frac{1}{2}(x_2 - \pi/2)^2 \end{aligned}$$

As long as we are will to admit a local error of order $|x - x_0|^2$ (small), f is approximately equal to this quadratic in the neighbourhood of $(1, \pi/2)$.

4.1.5 Example (Comments on Homework 3, Question 0.7).
Given $f(x) = x \cdot Ax$, find a path $\gamma : [-1, 1] \to \mathbb{R}^3$ such that $g = f \circ \gamma$ attains a local minimum at $t_0 = 1/2$. To do this, compute the eigenvalues and eigenvectors of A. As A is symmetric, there is a basis of $\mathbb{R}^3$ consisting of eigenvectors, $\{v_1, v_2, v_3\}$. Without loss of generality, assume that $\lambda_1 > 0$. The easiest path will be $\gamma(t) = (t - t_0)v_1$. In this case,

$$f(\gamma(t)) = (t - t_0)v_1 \cdot A((t - t_0)v_1) = (t - t_0)^2 \lambda_1 |v_1|^2 > 0$$

for all $t \neq t_0$, and it is zero at $t = t_0$.

4.1.6 Example. Use the second order Taylor expansion to approximate

$$\frac{(3.98 - 1)^2}{(5.97 - 3)^2}.$$

To do this, we use $f(x) = (x_1-1)^2/(x_2-3)^2$, $x_0 = (4,6)$, and $x = (3.98, 5.97)$. For the second order expansion, we need all partials up to second order.

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= 2(x_1-1)(x_2-3)^{-2} \\
\frac{\partial f}{\partial x_2} &= -2(x_1-1)^2(x_2-3)^{-3} \\
\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} &= -4(x_1-1)(x_2-3)^{-3} \\
\frac{\partial^2 f}{\partial x_1^2} &= 6(x_1-1)(x_2-3)^{-4} \\
\frac{\partial^2 f}{\partial x_2^2} &= 2(x_2-3)^{-2}
\end{aligned}$$

From these and the Taylor expansion,

$$\begin{aligned}
f(x) &\approx f(x_0) + Df(x_0)\cdot(x-x_0) + \frac{1}{2}(x-x_0)\cdot D^2 f(x_0)(x-x_0) \\
&= 1 + \frac{2}{3}(-0.02) - \frac{2}{3}(-0.03) + \frac{1}{2}\frac{2}{9}(-0.02)^2 + \frac{1}{2}\frac{2}{3}(-0.03)^2 \\
&\quad - \frac{4}{9}(-0.02)(-0.03) \\
&= 1.00674
\end{aligned}$$

The exact value is 1.00675. Note that $f(x) = \text{polynomial} + R(x)$, by Taylor's theorem, where $R(x)/|x-x_0|^2 \to 0$ as $x \to x_0$. Therefore, for this example, we expect the difference between $f(x)$ and the approximation to be within $C \cdot 10^{-4}$.

4.1.7 Examples.

1. Evaluate $\lim_{x\to 0}\left((1+x_1)^{x_2}-1\right)/\sqrt{x_1^2+x_2^2}$. Plugging in gives the indeterminate form $0/0$, so we must be more clever. Let $f(x) = (1+x_1)^{x_2} = \exp(x_2\log(1+x_1))$ and apply Taylor's theorem at $x_0 = 0$.

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= \frac{x_2}{1+x_1} f(x) \\
\frac{\partial f}{\partial x_2} &= \log(1+x_1) f(x) \\
\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} &= \frac{1}{1+x_1} f(x) + \frac{x_2 \log(1+x_1)}{1+x_1} f(x) \\
\frac{\partial^2 f}{\partial x_1^2} &= \frac{-x_2}{(1+x_1)^2} f(x) + \frac{x_2^2}{1+x_1} f(x) \\
\frac{\partial^2 f}{\partial x_2^2} &= \left(\log(1+x_1)\right)^2 f(x)
\end{aligned}$$

We obtain $f(x) \approx 1 + 0 + \frac{1}{2}x\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)x + R(x) = 1 + \frac{1}{2}x_1x_2 + R(x)$, where $R(x)/|x|^2 \to 0$ as $x \to 0$. This implies that the limit is zero.

2. Evaluate $\lim_{x\to 0}\big((1+x_1)^{x_2} - (1+x_1x_2)\big)/(x_1^2+x_2^2)$. For this example, the denominator is $|x|^2$, so the limit is still zero.

4.2 Extrema and the second derivative test

We say that $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ has a *local minimum* at x_0 if $f(x) \geq x_0$ for all $x \in B_\varepsilon(x_0) \cap \Omega$ for some $\varepsilon > 0$. If $f(x) > x_0$ (strict inequality) for $x \neq x_0$ then we say f has a *strict local minimum* at x_0. We say that f has a (strict) local maximum if $-f$ has a (strict) local minimum. Let's figure out what local minima of f say about Df and D^2f. Note that we may always focus on minima without loss of generality because differentiation is linear.

Assume first that $f \in C^1(\Omega)$ and $x_0 \in \Omega$ (so that it is an interior point). Along the directions of the canonical basis, the 1-dimensional functions $g_i(s) := f(x_0 + se_i)$ all attain local minima at $s = 0$. This observation allows us to prove the following theorem.

4.2.1 Theorem. *If $f \in C^1(\Omega)$ and $x_0 \in \Omega$ is a local minimum of f then $Df(x_0) = 0$.*

PROOF: $\frac{\partial f}{\partial x_i}(x_0) = \frac{d}{ds}g(s)|_{s=0} = g_i'(0) = 0$ for $i = 1, \ldots, n$. □

If $f \in C^2$ then the previous theorem applies, of course, and if $f(x_0) = 0$ then Taylor's theorem tells us that

$$\begin{aligned} f(x) &= f(x_0) + Df(x_0) + \frac{1}{2}(x-x_0)\cdot D^2f(x_0)(x-x_0) + R(x) \\ &= \frac{1}{2}(x-x_0)\cdot D^2f(x_0)(x-x_0) + R(x) \end{aligned}$$

The pure quadratic $(x-x_0)\cdot D^2f(x_0)(x-x_0)$ has a local minimum at x_0 if and only if all of the eigenvalues of $D^2f(x_0)$ are all non-negative.

4.2.2 Theorem. *If $f \in C^2(\Omega)$ and $x_0 \in \Omega$ is a local minimum of f then for all $w \in \mathbb{R}^n$, $w \cdot D^2f(x_0)w \geq 0$. Yet otherwise said, $D^2f(x_0)$ is a non-negative definite matrix.*

We note that the condition, $w \cdot D^2f(x_0)w \geq 0$ for all $w \in \mathbb{R}^n$, is equivalent to all of the eigenvalues of $D^2f(x_0)$ being non-negative.

PROOF: We wish to show the inequality $w \cdot D^2f(x_0)w \geq 0$ for each w. To this end, let $w \in \mathbb{R}^n$ be a generic fixed element with $w \neq 0$. Because x_0 was assumed to be interior to Ω, there is some $\epsilon > 0$ such that $B_\epsilon(x_0) \subset \Omega$. Thus for $t \in \mathbb{R}$ small

enough (smaller than $\epsilon/|w|$), $x_0 + tw \in \Omega$ and we can use Taylor's Theorem with $x = x_0 + tw$

$$f(x_0 + tw) = f(x_0) + Df(x_0) + \frac{1}{2}(tw) \cdot D^2 f(x_0)(tw) + R(x_0 + tw),$$

and recall the very useful fact that $R(x_0 + tw)/|tw|^2 \to 0$ as $|tw| \to 0$.

The inequality is nearly staring us in the face once we rearrange terms and use the local minimum property. Indeed, for t small enough we have

$$0 \leq f(x_0 + tw) - f(x_0)$$

and hence (recall $Df(x_0) = 0$)

$$0 \leq f(x_0 + tw) - f(x_0) = \frac{1}{2} t^2 w \cdot D^2 f(x_0) w + R(x_0 + tw).$$

Again rearranging, we see that we have obtained the inequality we want but with the remainder as a error:

$$-R(x_0 + tw)/t^2 \leq w \cdot D^2 f(x_0) w.$$

Taking the limit as $t \to 0$ and using the property of the remainder function, we conclude

$$0 \leq w \cdot D^2 f(x_0) w.$$

Since w was generic we conclude the theorem. □

Now we ask a slightly different question. We know that a local minimum forces a sign on the matrix $D^2 f$. However, can we use information about $D^2 f$ to determine if a given point is a local minimum? The answer is yes.

4.2.3 Theorem. *If $x_0 \in \Omega$ (i.e. x_0 is an interior point of Ω), $Df(x_0) = 0$, and for all $w \in \mathbb{R}^n$, $w \cdot D^2 f(x_0) w > 0$, then f has a strict local minimum at x_0.*

4.2.4 Definition. We say that $f \in C^2(\Omega)$ has a *saddle point* at $x_0 \in \Omega$ if $Df(x_0) = 0$ and $D^2 f(x_0)$ has at least one positive and one negative eigenvalue.

Saddle points are named for the simplest saddle point, the one at the origin of the graph of $f(x) = x_1^2 - x_2^2$, which looks like a saddle.

4.2.5 Theorem. *If $x_0 \in \Omega$ (i.e. x_0 is an interior point of Ω), $Df(x_0) = 0$, and $D^2 f(x_0)$ has at least one positive and one negative eigenvalue, then f has neither a local minimum nor a local maximum at x_0.*

PROOF (SKETCH): Let v_- and v_+ be eigenvectors associated with eigenvalues $\lambda_- < 0$ and $\lambda_+ > 0$, respectively. Use Taylor's theorem to investigate $f(x_0 + tv_-)$ and $f(x_0 + tv_+)$. The argument from last time gives a strict local minimum/maximum for these two functions, implying that f has neither a minimum nor a maximum at x_0. □

Remark.

1. If f has a local minimum at x_0 then $Df(x_0) = 0$ and $D^2f(x_0) \geq 0$.
2. If $Df(x_0) = 0$ and $D^2f(x_0) > 0$ then x_0 is a strict local minimum.
3. If f has a strict local minimum at x_0 then it is not necessarily the case that $Df^2(x_0) > 0$, i.e. there may be $w \neq 0$ such that $w \cdot D^2f(x_0)w = 0$.
4. If $D^2f(x_0)$ has eigenvalues of differing signs then x_0 is neither a local minimum nor a local maximum.

4.2.6 Examples.

1. Let $f(x) = e^{x_1^2}e^{-x_2^2}$. Where can f has a local minimum? Can f ever have a strict local minimum? Does f have a global minimum?

   We calculate $Df(x) = (2x_1, -2x_2)f(x)$ and

   $$D^2f(x) = \begin{pmatrix} 2+4x_1^2 & -4x_1x_2 \\ -4x_1x_2 & -2+4x_2^2 \end{pmatrix} f(x).$$

   $Df(x) = 0$ if and only if $x = 0$, so the only possible point at which there can be a minimum is $x = 0$. At this point, $Df^2(0) = \left(\begin{smallmatrix} 2 & 0 \\ 0 & -2 \end{smallmatrix}\right)$, which has eigenvalues ± 2. Therefore f does not have a local minimum (or maximum) at $x = 0$. It is not hard to see that f does not attain a global minimum.
2. Let $f(x) = e^{x_1^2}(1 - e^{-x_2^2})$. Where can f has a local minimum? Can f ever have a strict local minimum? Does f have a global minimum?

   Again, $Df(x) = (2x_1e^{x_1^2}(1-e^{-x_2^2}), -2x_2e^{x_1^2}e^{-x_2^2})$ and

   $$D^2f(x) = \begin{pmatrix} (2+4x_1^2)e^{x_1^2}(1-e^{-x_2^2}) & 4x_1x_2e^{x_1^2}e^{-x_2^2} \\ 4x_1x_2e^{x_1^2}e^{-x_2^2} & (2-4x_2^2)e^{x_1^2}e^{-x_2^2} \end{pmatrix}.$$

   Now $Df(x) = 0$ implies $x_2 = 0$, but x_1 can be anything.

   $$D^2f(x_1, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 2e^{x_1^2} \end{pmatrix} \geq 0$$

   for any x_1, so all points on the line $x_2 = 0$ are local minima. None of these local minima are strict. Finally, f has a global minimum value of 0, which is attained at all points on the line $x_2 = 0$.
3. For $f(x) = x_1^4 + x_2^4$, $D^2f(0) = 0$, but f has a strict local minimum at $x = 0$.

4.3 Constrained extrema and Lagrange multipliers

Going back to 1-dimensional calculus, to solve the problem of finding the minimizer $x \in [a, b]$ of $f : [a, b] \to \mathbb{R}$, we usually go through the following process:

1. Find all solutions to $f'(x) = 0$ in (a, b) (the *critical points*).
2. Check which solutions of $f'(x) = 0$ have $f''(x) \geq 0$. (This step is optional.)
3. Compare the values of $f(x)$ for all x obtained and also $x = a, b$. The smallest gives the minimizer.

The points a and b are the boundary points of the set $[a,b]$. In higher dimensions the boundaries are much more interesting and delicate. For $\Omega \subseteq \mathbb{R}^n$, the *boundary* is $\partial\Omega := \{x \in \mathbb{R}^n : \text{for all } \varepsilon > 0,\ B_\varepsilon(x) \cap \Omega \neq \varnothing \text{ and } \neq B_\varepsilon(x)\}$.

In the multi-dimensional settings, given a domain Ω and $f : \overline{\Omega} \to \mathbb{R}$, we are concerned with finding the minimizer for the problem $\min_{x\in\overline{\Omega}} f(x)$. If Ω is bounded then $\overline{\Omega}$ is compact, so the minimum value is achieved for some $x_0 \in \overline{\Omega}$, i.e. $f(x_0) = \min_{x\in\overline{\Omega}} f(x)$. As before, we must check critical points and boundary points. Unlike in the 1-dimensional setting, the boundary is not simply two points; it is a whole curve!

For many (actually all) of our examples, $\overline{\Omega} = \{x \in \mathbb{R}^n : F(x) \leq 0\}$ for some function $F : \mathbb{R}^n \to \mathbb{R}$. In fact, taking $F(x)$ to be the (signed) distance from x to the boundary $\partial\Omega$ will do. Such Fs are not unique. Finding the minimum over the boundary is now reduced to finding the minimum of f on the set $\partial\Omega = \{x \in \mathbb{R}^n : F(x) = 0\}$, which leads to constrained optimization.

Suppose that the minimum value of f over $\overline{\Omega}$ occurs at $x_0 \in \partial\Omega$. If $\gamma : [a,b] \to \partial\Omega$ is a path in the boundary with $\gamma(0) = x_0$, then $t \mapsto f(\gamma(t))$ has a minimum value of $f(x_0)$ at $t = 0$. Therefore it's derivative with respect to t at $t = 0$ is zero. Hence by the chain rule, $0 = Df(x_0) \cdot \dot{\gamma}(0)$, and this works for any path. We already know that $\dot{\gamma}(0)$ is perpendicular to $DF(x_0)$ for every path in $\partial\Omega$. If the tangent space has codimension one (i.e. if $\Omega \subseteq \mathbb{R}^n$ has "full dimension") then it must be the case that $Df(x_0)$ and $DF(x_0)$ are parallel! Therefore there is $\lambda \in \mathbb{R}$ such that $Df(x_0) = \lambda DF(x_0)$. Rephrasing, if x_0 is a minimizer of f on $\partial\Omega$ then there must exist some $\lambda \in \mathbb{R}$ such that $Df(x_0) = \lambda DF(x_0)$. After introducing the new variable λ, there are $n+1$ unknowns, and the equality of vectors is n equations. But $F(x_0) = 0$ (the fact that $x_0 \in \partial\Omega$) gives another equation. Thus there is a hope of finding solutions! This is the method of the *Lagrange multiplier*.

1. Find all solutions to $Df(x_0) = 0$ for $x_0 \in \Omega$ (the *critical points*).
2. Check which solutions of $Df(x_0) = 0$ have $D^2 f(x_0) \geq 0$. (This step is optional.)
3. Solve for x_0 and λ with $Df(x_0) = \lambda DF(x_0)$ and $F(x_0) = 0$.
4. Compare the values of $f(x_0)$ for all x_0 obtained. The smallest gives the minimizer.

4.3.1 Examples.

1. Let $f(x) = x_1^2 - x_2^2$ and $\Omega = B_1(0)$. Then $\overline{\Omega} = \{x : |x|^2 - 1 \leq 0\}$ and $Df(x) = (2x_1, -2x_2)$.

   [Insert image.]

   $Df(x) = 0$ implies that $x = 0 \in \Omega$, and this is the only critical point. The system of equations obtained by introducing the Lagrange multiplier is

   $$\begin{aligned} 2x_1 &= \lambda 2x_1 \\ -2x_2 &= \lambda 2x_2 \\ 0 &= x_1^2 + x_2^2 - 1 \end{aligned}$$

   If $x_1 \neq 0$ then $\lambda = 1$ and hence $x_2 = 0$ and $x_1 = \pm 1$. Similarly, if $x_2 \neq 0$ then $\lambda = -1$ and hence $x_1 = 0$ and $x_2 = \pm 1$. There are hence five points

to compare, $(0,0)$, $(\pm 1,0)$, and $(0,\pm 1)$. The minimum of -1 occurs at both $(0,\pm 1)$.

2. Let $f(x) = x_1 x_2$ and $\Omega = B_1(0)$. Then $Df(x) = (x_2, x_1)$, so $x = 0$ is the only critical point in Ω. The equations resulting from introducing the Lagrange multiplier are

$$\begin{aligned} x_2 &= \lambda 2 x_1 \\ x_1 &= \lambda 2 x_2 \\ 0 &= x_1^2 + x_2^2 - 1. \end{aligned}$$

It follows that $\lambda^2 = \frac{1}{4}$ since x_1 and x_2 cannot both be zero. Therefore $\lambda = \pm\frac{1}{2}$, so $x_1 = \pm x_2$, and hence $(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}})$ are the four solutions. Of the five possibilities (one critical point and four points on the boundary of Ω), the minimum value is $-\frac{1}{2}$, which occurs at $(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ and $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

3. Let $\Gamma = \{(t, t+1) : t \in \mathbb{R}\}$, a line in $\mathbb{R}^2$, and let $f(x) = x_1^2 + x_2^2$. Show that f does not attain a maximum value on Γ and compute the minimum value of f on Γ. Recognize that Γ is the boundary of the region $\{x \in \mathbb{R}^2 : x_1 + 1 - x_2 \leq 0\}$. To find the max/min of f on Γ we use a Lagrange multiplier.

$$\begin{aligned} 2x_1 &= \lambda(1) \\ 2x_2 &= \lambda(-1) \\ 0 &= x_1 + 1 - x_2 \end{aligned}$$

Therefore $\lambda = -1$, $x_1 = 1/2$, and $x_2 = -1/2$. So $(1/2, -1/2)$ is the only possible critical point of f on Γ. The value of f at this point is $1/2$, and it is clear that f can take larger values than this on Γ (e.g. $f(0,1) = 1 > 1/2$). Therefore this point is the minimum, and there is no maximum.

5 Multi-dimensional integrals

Continuing our quest to extend all of the ideas from 1-dimensional calculus to multi-dimensional calculus, we turn to integrals. The integral, $\int_a^b f(x)\,dx$, of a continuous function $f : [a,b] \to \mathbb{R}$ can be defined to be the (signed) area between graph of f, $\{(x,y) : y = f(x)\}$ and the x-axis.

The obvious analog for $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is the "volume" of the region lying between the hyperplane $\mathbb{R}^n$, embedded in $\mathbb{R}^{n+1}$ as $\{x_{n+1} = 0\}$, and the graph of f, $\{x \in \mathbb{R}^{n+1} : x_{n+1} = f(x_1, \ldots, x_n)\}$. When $n = 2$ this can be visualized and corresponds to the usual idea of (signed) volume. For $n > 2$ this generalizes volume.

5.1 Double integrals

5.1.1 Example. The simplest case has $\Omega = [a,b] \times [c,d]$, a rectangle, and $f(x) = \alpha$, a constant. The region between the x_1-x_2 plane and the graph of f is the rectangular prism with volume $\alpha(b-a)(d-c)$.

5.1.2 Example. Let $\Omega = [0,1] \times [-1,1]$ and $f(x) = x_1^2 + x_2^2$.

[Insert diagram.]

Consider "slices" of the volume obtained by holding x_1 constant. When $x_1 = s \in [0,1]$ is held fixed, the height of the curve on the corresponding slice is given by the function $g(x_2) = s^2 + x_2^2$. The area of this slice is hence

$$\int_{-1}^{1} g(x_2)\,dx_2 = \int_{-1}^{1} (s^2 + x_2^2)\,dx_2 = 2s^2 + \frac{2}{3}.$$

If we then "add up" (i.e. integrate) over all slices, we can calculate the total volume to be

$$\int_0^1 \left(2x_1^2 + \frac{2}{3}\right) dx_1 = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}.$$

We can write this method succinctly as

$$\int_0^1 \int_{-1}^{1} x_1^2 + x_2^2\,dx_2 dx_1 = \frac{4}{3},$$

where we integrate first with respect to x_2 and then with respect to x_1. Would we have obtained the same answer if we integrated with respect to x_1 first?

For a general volume in $\mathbb{R}^3$, *Cavalieri's principle of volume* says that volume is equal to the integral of cross-sectional area, i.e. vol $= \int \text{area}(s)\,ds$, where s parameterizes the cross-sectional area. This principle works, of course, for signed area as well.

When $\Omega = [a,b] \times [c,d]$ is a rectangle, the integral of a function $f : \Omega \to \mathbb{R}$ may be computed as

$$\int_a^b \int_c^d f(x_1, x_2)\,dx_2 dx_1,$$

since, as a function of x_1, the cross-sectional area is given by $\int_c^d f(x_1, x_2)\,dx_2$. We will write $\iint_\Omega f(x)\,dA$ for this *double integral* of a continuous function $f : \Omega \to \mathbb{R}$. When written with explicit bounds, it is referred to as an *iterated integral*.

5.1.3 Theorem. *If $\Omega = [a,b] \times [c,d]$ and $f : \Omega \to \mathbb{R}$ is continuous then both $\int_a^b \left(\int_c^d f(x_1, x_2)\,dx_2\right) dx_1$ and $\int_c^d \left(\int_a^b f(x_1, x_2)\,dx_1\right) dx_2$ exist and they are equal to each other. In this case we say that the double integral, $\iint_\Omega f(x)\,dA$, exists and is equal to the common value of the iterated integrals.*

5.1.4 Example. Let $\Omega = [0, \pi/2]^2$ and $f(x) = \cos(x_1)\sin(x_2)$.

[Insert image.]

Using the theorem, since f is a continuous function,

$$\begin{aligned}\iint_\Omega f\,dA &= \int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} \cos(x_1)\sin(x_2)\,dx_2dx_1\\ &= \int_0^{\frac{\pi}{2}} \sin(x_2)\,dx_2 \int_0^{\frac{\pi}{2}} \cos(x_1)\,dx_1\\ &= 1\cdot 1 = 1\end{aligned}$$

5.1.5 Example. Let Ω be a metal plate sitting in the x-y plane as $[0,2]\times[0,1]$. If Ω has a non-constant density of mass modelled by $\rho(x,y) = ye^{xy}$ g/cm^2, then what is the total mass of Ω? The usual rule of "mass = density × area" only applies when the density is constant. Note that ρ is continuous, so for very small sub-rectangles it is approximately constant. Mass is additive, so we may compute the Riemann sums and take the limit to obtain that the mass is the integral of density.

$$\begin{aligned}\text{mass} &= \iint_\Omega \rho(x,y)\,dxdy\\ &= \int_0^1\int_0^2 ye^{xy}\,dxdy\\ &= \int_0^1 (e^{2y}-1)\,dy\\ &= \frac{1}{2}e^2 - \frac{3}{2}\end{aligned}$$

To compute the integral over regions that are not necessarily rectangular, we appeal to Cavalieri's principle. Suppose $\Omega = \{(x,y)\in\mathbb{R}^2 : \varphi_1(x)\le y\le\varphi_2(x)\}$ is the region between two graphs (we say that Ω is a *region of type I*). Then

$$\iint_\Omega f(x,y)\,dA = \int_a^b\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,dydx.$$

There should be nothing special about this particular orientation. If we can write $\Omega = \{(x,y)\in\mathbb{R}^2 : \psi_1(y)\le x\le\psi_2(y)\}$ then Ω is a *region of type II* and

$$\iint_\Omega f(x,y)\,dA = \int_a^b\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,dxdy.$$

There are regions that are neither type I nor type II, and it is a bit of an art to decompose them into pieces over which the integral can be computed.

5.1.6 Examples.

1. Let $f(x,y) = x+y$ and $\Omega = \{(x,y) : 0\le x\le 1/2, 0\le y\le x^2\}$.
   [Insert diagram, including coloured lines in both directions.]

This region is of both type I (as given) and type II, which can be seen by noting $\Omega = \{(x,y) : 0 \le y \le 1/4, \sqrt{y} \le x \le 1/2\}$. Hence

$$\iint_\Omega f\, dA = \int_0^{\frac{1}{2}} \int_0^{x^2} x + y\, dy dx = \int_0^{\frac{1}{2}} x^3 + \frac{1}{2}x^4\, dx$$
$$= \frac{3}{160} = \int_0^{\frac{1}{4}} \int_{\sqrt{y}}^{\frac{1}{2}} x + y\, dx dy$$

2. Evaluate $\int_0^1 \int_{x^3}^{x^2} xy\, dy dx$.
   [Insert diagram of Ω.]

$$\int_0^1 \int_{x^3}^{x^2} xy\, dy dx = \int_0^1 \frac{1}{2}x^5 - \frac{1}{2}x^7\, dx = \frac{1}{48}$$

3. As a type II integral, the above integral may be written $\int_0^1 \int_{\sqrt{y}}^{\sqrt[3]{y}} xy\, dx dy$.

5.2 Triple integrals

Consider two point objects P_i in $\mathbb{R}^3$ with masses m_i and locations (x_i, y_i, z_i), for $i = 1, 2$. Then the center of mass of the two points is the weighted average of the positions,

$$(\overline{x}, \overline{y}, \overline{z}) = \frac{1}{m_1 + m_2}(m_1x_1 + m_2x_2, m_1y_1 + m_2y_2, m_1z_1 + m_2z_2).$$

With n points the formulae are analogous, e.g.

$$\overline{x} = \frac{m_1x_1 + \cdots + m_nx_n}{m_1 + \cdots + m_n}.$$

If you had a rectangular prism $\Omega = [a,b] \times [c,d] \times [e,f]$ in $\mathbb{R}^3$ with non-constant mass density $\rho : \Omega \to \mathbb{R}$ then where is its centre of mass? If ρ is "nice enough" (e.g. continuous) then we can decompose each interval into N small pieces, hence decomposing Ω into N^3 small cubes, over which ρ is approximately constant. Over each small rectangle the mass is approximately the volume times the (nearly) constant density. The centre of mass of Ω is then the weighted sum of the N^3 small rectangles. In symbols,

$$\overline{x} \approx \frac{1}{\text{total mass}} \sum_{i,j,k=1}^{N} x_i \rho(x_i, y_j, z_k) \Delta x_i \Delta y_j \Delta z_k$$
$$\approx \frac{\sum_{i,j,k=1}^{N} x_i \rho(x_i, y_j, z_k) \Delta x_i \Delta y_j \Delta z_k}{\sum_{i,j,k=1}^{N} \rho(x_i, y_j, z_k) \Delta x_i \Delta y_j \Delta z_k}$$
$$\to \frac{\iiint_\Omega x\rho(x,y,z) dV}{\iiint_\Omega \rho\, dV}$$

as $N \to \infty$.

5.2.1 Definition. Let $\Omega = [a,b]\times[c,d]\times[e,f]$ and $\rho : \Omega \to \mathbb{R}$ be given. Fix $x_1,\dots,x_N$, $y_1,\dots,y_N$, $z_1,\dots,z_N$ be given and

$$S_N := \sum_{i,j,k=1}^{N} \rho(x_i^*, y_j^*, z_k^*)\Delta x_i \Delta y_j \Delta z_k$$

where $x_i^* \in [x_{i-1}, x_i]$, $y_j^* \in [y_{j-1}, y_j]$, and $z_k^* \in [z_{k-1}, z_k]$. If S_N has the same limit as $N \to \infty$ and $\Delta x_i, \Delta y_j, \Delta z_k \to 0$ which is independent of the choice of x_i^*, y_j^*, and z_k^* then we say that ρ is *Riemann integrable*. In this case the limit is denoted $\iiint_\Omega \rho\, dV = \iiint_\Omega \rho\, dx dy dz$, or with any other permutation of the variables.

The integral can be computed as any of the six ways of permuting the interated integrals, e.g.

$$\iiint_\Omega \rho\, dV = \int_a^b \int_c^d \int_e^f \rho(x,y,z)\, dz dy dx$$

When ρ is Riemann integral all choices of permutation give the same result.

Moving to more general domains, we say that Ω is *elementary region* if one variable is between two functions of the other two, one of the remaining variables is between two functions of the other, and the last is between constants. For example,

$$\Omega = \{(x,y,z) : \psi_1(x,y) \le z \le \psi_2(x,y), \varphi_1(x) \le y \le \varphi_2(x), a \le x \le b\}$$

[Insert diagrams of two and three dimensional cases.]

For that Ω, for example, we would then be able to compute

$$\iiint_\Omega \rho\, dV = \int_a^b \int_{\varphi_1(x)}^{\varphi_2(x)} \int_{\psi_1(x,y)}^{\psi_2(x,y)} \rho(x,y,z)\, dz dy dx.$$

5.2.2 Examples.

1. Let $B = \{x \in \mathbb{R}^3 : |x| \le 1\}$ be the ball in $\mathbb{R}^3$. It is an elementary region because it can be written "inside out" as

$$-\sqrt{1-x^2-y^2} \le z \le \sqrt{1-x^2-y^2},$$
$$-\sqrt{1-x^2} \le y \le \sqrt{1-x^2},$$
$$-1 \le x \le 1.$$

2. Compute the volume of the ball using a triple integral. If we use a constant

mass-density of 1 then the volume equals the mass. Hence

$$\begin{aligned}
\text{volume} &= \iiint_B 1\,dV \\
&= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} 1\,dz\,dy\,dx \\
&= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 2\sqrt{1-x^2-y^2}\,dy\,dx \\
&= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 2\sqrt{\sqrt{1-x^2}^2-y^2}\,dy\,dx \\
&= \int_{-1}^{1} 2\frac{1}{2}\pi(1-x^2)\,dx \qquad \text{area of half-circle} \\
&= \frac{4\pi}{3}
\end{aligned}$$

3. Let Ω be the region bounded by $x=0$, $y=0$, $z=2$, and $z=x^2+y^2$ in the first octant. Evaluate $\iiint_\Omega x\,dV$.
   [Insert diagram of Ω.]
   Clearly $\Omega = \{x \geq 0, y \geq 0, x^2+y^2 \leq z \leq 2\}$. Hence

$$\begin{aligned}
\iiint_\Omega x\,dV &= \int_0^{\sqrt{2}}\int_0^{\sqrt{2-x^2}}\int_{x^2+y^2}^{2} x\,dz\,dy\,dx \\
&= \int_0^{\sqrt{2}} x\int_0^{\sqrt{2-x^2}} 2-x^2-y^2\,dy\,dx \\
&= \int_0^{\sqrt{2}} x\left((2-x^2)\sqrt{2-x^2}-\frac{1}{3}\sqrt{2-x^2}^3\right)dx \\
&= \int_0^{\sqrt{2}} \frac{4}{3}x(2-x^2)^{3/2}\,dx = \frac{8\sqrt{2}}{15}
\end{aligned}$$

4. Evaluate the same integral with dx first. Rewrite Ω "inside out" as $\{0 \leq z \leq 2, 0 \leq y \leq \sqrt{z}, 0 \leq x \leq \sqrt{z-y^2}\}$.

$$\begin{aligned}
\iiint_\Omega x\,dV &= \int_0^2\int_0^{\sqrt{z}}\int_0^{\sqrt{z-y^2}} x\,dx\,dy\,dz \\
&= \int_0^2\int_0^{\sqrt{z}} \frac{1}{2}(z-y^2)\,dy\,dz \\
&= \int_0^2 \frac{1}{3}z^{3/2}\,dz = \frac{8\sqrt{2}}{15}
\end{aligned}$$

5.3 Riemann sums

As in the 1-dimensional setting, we will follow the course pioneered by Riemann and divide the domain, Ω, into many little rectangles and construct the integral as a limit of "Riemann sums." For now we will restrict ourselves to rectangles.

More formally, we can define the multiple integral as a limit of "Riemann sums" of the following form. Suppose Ω is a rectangle $[a_1, b_1] \times \cdots \times [a_n, b_n]$. Partition each interval $[a_i, b_i]$ into N small sub-intervals of length $(b_i - a_i)/N$ with points

$$x_i^k := a_1 + \frac{k(b_i - a_i)}{N}$$

for $k = 0, \ldots, N$.

[Insert image, same one as in the previous example, with sub-rectangles.]

Suppose $f : \Omega \to \mathbb{R}$ is continuous. For any $0 \le k_1, \ldots, k_n \le N$, volume lying over the small sub-rectangle

$$[x_1^{k_1}, x_1^{k_1+1}] \times \cdots \times [x_n^{k_n}, x_n^{k_n+1}]$$

is approximated well by

$$f(c^{k_1,\ldots,k_n})\Delta x_1^{k_1} \cdots \Delta x_n^{k_n} = f(c^{k_1,\ldots,k_n})\frac{b_1 - a_1}{N} \cdots \frac{b_n - a_n}{N},$$

where $c^{k_1,\ldots,k_n}$ is any point inside that rectangle. The *Riemann sum* is then

$$S(N) = \sum_{k_1,\ldots,k_n=0}^{N-1} f(c^{k_1,\ldots,k_n})\frac{(b_1 - a_1)\cdots(b_n - a_n)}{N^n}.$$

If it converges then we define $\int_\Omega f(x)\,dx$ to be equal to the limit.

Note that the same reasoning can be applied (with care) to functions f that are defined piece-wise to be continuous on a finite number of sub-regions. The integral is the sum of the integrals over the individual pieces.

The following three questions need to be addressed.

1. Over which Ω can we integrate "integrable" functions f?
2. Which f are Riemann integrable and which are not?
3. What are the properties of $\int_\Omega f\,dA$?

So far we know that it is possible to integrate continuous functions over finite unions of sets of type I and type II (and in particular, rectangles).

5.3.1 Properties of the multiple integral.

1. *If f is continuous then $\int_\Omega f\,dA$ exists.*
2. *If $\Omega = \Omega_1 \cup \Omega_2$ and Ω_1 and Ω_2 intersect only in their boundaries then*

$$\int_\Omega f\,dA = \int_{\Omega_1} f\,dA + \int_{\Omega_2} f\,dA.$$

3. *If $f_1 \le f_2$ on Ω then $\int_\Omega f_1\,dA \le \int_\Omega f_2\,dA$.*

4. *If $f \equiv \alpha$ is constant then $\int_\Omega f\, dA = \alpha\,\text{area}(\Omega)$.*
5. *The map $f \mapsto \int_\Omega f\, dA$ is a linear map on the space of integrable functions.*
6. $\left|\int_\Omega f\, dA\right| \le \int_\Omega |f|\, dA$.

Recall that for most of this class we have been concerned with domains, which are open, connected sets. We now restrict ourselves further to those domains whose boundaries can be realized as a finite union of smooth level surfaces. Using the same construction as in the rectangle case, it can be shown that if f is discontinuous only on a finite union of smooth level surfaces, then the integral of f can be shown to exist. What we will do, given $f : \Omega \to \mathbb{R}$, where Ω is nice in the above sense, is extend f to $\bar{f}$ on a rectangle containing Ω and use the previous construction. We should clearly take $\bar{f} \equiv 0$ outside of Ω, and $\bar{f} = f$ inside Ω. Note that $\bar{f}$ is generally discontinuous even if f is continuous on Ω. We hence define f to be *Riemann integrable* if $\int_R \bar{f}\, dA$ exists for all rectangles $R \supseteq \Omega$ and $\bar{f}$ extending f to R in the way described. All this said, we will typically work with regions that can be recognized as or decomposed as a finite union of regions of type I or type II.

5.4 Changes of coordinates

Evaluate $\iint_\Omega \log(x^2+y^2)\, dA$ where Ω is the region in the first quadrant bounded between the circles of radius 1 and 2. Clearly polar coordinates are the way to go. Let $x = r\cos\theta$ and $y = r\sin\theta$, and we will be able to write

$$\iint_\Omega \log(x^2+y^2) dA = \iint_{\tilde{\Omega}} \log(r^2) dA.$$

It is clear that $\tilde{\Omega} = [1,2]\times[0,\pi/2]$, but the question is, how does dA change?

A *change of coordinates* is a smooth, invertible function $F : \mathbb{R}^k \to \mathbb{R}^k$ with smooth inverse. Suppose $F : \tilde{\Omega} \to \Omega$, where $\tilde{\Omega}$ is to be thought of as the domain of the "u-coordinates" and Ω is the domain of the "x-coordinates."

[Insert diagram.]

$$\int_\Omega f(x) dx = \int_{\tilde{\Omega}} f(F(u))(???) du$$

When F is linear and $\tilde{\Omega}$ is a rectangle, $\Omega = F(\tilde{\Omega})$ is a parallelepiped. If $F(u) = Au$ for some invertible matrix A then the volume of Ω is $\det(A)$ times the volume of $\tilde{\Omega}$.

5.5 Polar coordinates

Lecture Friday was missed by Chris because Russell didn't tell him it was on.

5.6 Spherical Coordinates

$$F(\rho,\theta,\varphi) = (\rho\sin\theta\cos\varphi, \rho\sin\varphi\sin\theta, \rho\cos\varphi)$$
$$\iiint_\Omega f(x,y,z)dxdydz = \iiint_{\tilde{\Omega}} f(F(\rho,\theta,\varphi))|\det(DF)|d\rho d\varphi d\theta$$
$$\det(DF) = \rho^2\sin\varphi$$

5.6.1 Examples.

1. Compute the volume of the ball of radius R in $\mathbb{R}^2$ via spherical coordinates. Note that we have done this in Cartesian coordinates already.

$$\begin{aligned}\iiint_{\overline{B}_R(0)} dV &= \int_0^{2\pi}\int_0^{\pi}\int_0^{R} \rho^2\sin\varphi\, d\rho d\varphi d\theta \\ &= \int_0^{2\pi}\int_0^{\pi} \sin\varphi\frac{1}{3}R^3\, d\varphi d\theta \\ &= \int_0^{2\pi}\frac{2}{3}R^3\, d\theta = \frac{4\pi}{3}R^3\end{aligned}$$

2. Compute.

$$\begin{aligned}\iiint_{\overline{B}_1(0)} e^{(x^2+y^2+z^2)^{3/2}}\, dV &= \int_0^{2\pi}\int_0^{\pi}\int_0^{1} e^{\rho^3}\rho^2\sin\varphi\, d\rho d\varphi d\theta \\ &= \int_0^{2\pi}\int_0^{\pi}\frac{e-1}{3}\sin\varphi\, d\varphi d\theta \\ &= \frac{4\pi}{3}(e-1)\end{aligned}$$

3. Compute the moment of inertia about the z-axis of the region Ω bounded by $z = x^2+y^2$ and $x^2+y^2 = b^2$ and $z \geq 0$ with constant density ρ.

$$\begin{aligned}I_z &= \iiint_\Omega (\text{density})(\text{distance to } z\text{-axis})\, dV \\ &= \iiint_\Omega \rho(x^2+y^2)\, dxdydz \\ &= \rho\int_0^{2\pi}\int_0^{b}\int_0^{r^2} r^2 r\, dzdrd\theta \\ &= \rho\int_0^{2\pi}\int_0^{b} r^5\, drd\theta \\ &= \rho\int_0^{2\pi}\frac{1}{6}r^6\, d\theta = \frac{\pi\rho b^6}{3}\end{aligned}$$

Multiple integrals are contained in Kaplan, chapter 4. Chapter 5 of Kaplan contains all of the vector calculus that we will cover in this course, including line, path, and surface integrals.

6 Integrals over paths and surfaces

6.1 Line and path integrals

To motivate integrals along paths, consider how we would calculate the total mass of a long thin wire with non-constant density, or the total work done by a force along a trajectory. In both cases we want to integrate along a path $\gamma : [a, b] \to \mathbb{R}^3$. Let Γ_γ denote the corresponding curve in $\mathbb{R}^3$. For the mass calculation, Γ_γ is the wire, with some density function $\rho : \mathbb{R}^3 \to \mathbb{R}$, and for the work calculation, γ is the path of the object being acted upon by a force $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$.

In both cases, Γ_γ may be very complicated. We can approximate it with $\tilde{\Gamma}_N$, an N-segment piecewise affine approximation to Γ_γ by connecting, via straight line segments, a sequence of points $x_i \in \Gamma_\gamma$. Say $x_i = \gamma(t_i)$ for time points $a = t_0 < t_1 < \cdots < t_N = b$. If γ and ρ are nice enough and Δt_i is small enough, we can write the mass of the i^{th} segment as

$$m_i \approx \rho(\gamma(t_i^*))|\gamma(t_i) - \gamma(t_{i-1})|.$$

Therefore

$$\text{mass} \approx \sum_{i=1}^{N} \rho(\gamma(t_i^*))|\gamma(t_i) - \gamma(t_{i-1})|.$$

In the case of work, by summing the work done over each small straight line segment,

$$\text{work} \approx \sum_{i=1}^{N} \mathbf{F}(\gamma(t_i^*)) \cdot (\gamma(t_i) - \gamma(t_{i-1})).$$

When Γ_γ is C^1, so $\gamma \in C^1([a, b]; \mathbb{R})$, we can approximate via Taylor's theorem

$$\gamma(t_i) = \gamma(t_{i-1}) + \dot{\gamma}(t_i)(t_i - t_{i-1}) + \text{err}(t_i - t_{i-1})$$

where $\text{err}(t_i - t_{i-1})/|t_i - t_{i-1}| \to 0$ as $|t_i - t_{i-1}| \to 0$. Therefore

$$\begin{aligned} \text{mass} &\approx \sum_{i=1}^{N} \rho(\gamma(t_i^*))|\dot{\gamma}(t_i)|\Delta t_i + \rho(\gamma(t_i^*))\frac{\text{err}(\Delta t_i)}{\Delta t_i}\Delta t_i \\ &\to \int_a^b \rho(\gamma(t))|\dot{\gamma}(t)|\,dt \qquad \text{as } \Delta t \to 0. \end{aligned}$$

since the part involving the error converges to zero. Similarly, the work can be shown to be

$$\text{work} = \int_a^b \mathbf{F}(\gamma(t)) \cdot \dot{\gamma}(t)\,dt$$

6.1.1 Definition. The *path integral* of ρ along Γ_γ is

$$\int_{\Gamma_\gamma} \rho\, ds := \int_a^b \rho(\gamma(t))|\dot\gamma(t)|\, dt.$$

The *line integral* of $\mathbf{F}$ along Γ_γ is

$$\int_{\Gamma_\gamma} \mathbf{F}\cdot d\mathbf{s} := \int_a^b \mathbf{F}(\gamma(t))\cdot\dot\gamma(t)\, dt.$$

Remark. In $\mathbb{R}^3$ we can think of $\gamma(t) = (x(t), y(t), z(t))$. The "differential" of γ is given by $\dot x = \frac{dx}{dt}$, $\dot y = \frac{dy}{dt}$, and $\dot z = \frac{dz}{dt}$. Thus "$\dot\gamma(t)\,dt = dx + dy + dz$" where the latter needs to be interpreted as a vector quantity. In particular, dx, dy, and dz should be thought of as the standard basis in the vector space of differential forms. We will hence sometimes write $\mathbf{F}\cdot d\mathbf{s} = F_1 dx + F_2 dy + F_3 dz$.

6.1.2 Examples.

1. Compute $\int_{\Gamma_\gamma} \cos(z)dx + e^x dy + e^y dz$ where $\gamma(t) = (1, t, e^t)$ for $t \in [0, 2\pi]$. Here $\mathbf{F}(u) = (\cos(u_3), e^{u_1}, e^{u_2})$ and $\dot\gamma(t) = (0, 1, e^t)$, so
$$\int_{\Gamma_\gamma} \mathbf{F}\cdot d\mathbf{s} = \int_0^{2\pi} 0 + e^1 + e^{2t} dt = 2\pi e + \frac{1}{2}(e^{4\pi} - 1)$$
2. Let γ be any nice C^1 path (i.e. with $|\dot\gamma| \neq 0$) and let $f \in C^1$. Write $\int_{\Gamma_\gamma} \nabla f \cdot d\mathbf{s}$ in terms of f and γ alone.
$$\begin{aligned}\int_{\Gamma_\gamma} \nabla f \cdot d\mathbf{s} &= \int_a^b \nabla f(\gamma(t))\cdot\dot\gamma(t)\, dt \\ &= \int_a^b \frac{d}{dt} f(\gamma(t))\, dt \\ &= f(\gamma(b)) - f(\gamma(a))\end{aligned}$$
It follows that the integral of ∇f depends only on the values of f at the endpoints of γ. We say that ∇f is a *conservative vector field*.
3. Compute $\int_{\Gamma_\gamma} y dx + x dy$ where $\gamma(t) = (t^9, \sin^9(\pi t/2))$ for $t \in [0, 1]$. Clearly $\mathbf{F}(x, y) = \nabla(xy)$, so the result is $1^9 \sin^9(\pi/2) - 0 = 1$.
4. Let $\gamma_1(t) = (\cos t, \sin t, t/(2\pi))$, $\gamma_2(t) = (\cos t, -\sin t, t/(2\pi))$ and $\mathbf{F}(u) = (u_2, -u_1, 1)$. Evaluate $\int_{\Gamma_{\gamma_i}} \mathbf{F}\cdot d\mathbf{s}$ for $i = 1, 2$ and $t \in [0, 2\pi]$.
$$\begin{aligned}\mathbf{F}(\gamma_1(t))\cdot\dot\gamma_1(t) &= -\sin^2 t - \cos^2 t + \frac{1}{2\pi} = -1 + \frac{1}{2\pi} \\ \mathbf{F}(\gamma_2(t))\cdot\dot\gamma_2(t) &= \sin^2 t + \cos^2 t + \frac{1}{2\pi} = 1 + \frac{1}{2\pi}\end{aligned}$$

so the integrals are, respectively, $1-2\pi$ and $1+2\pi$. It follows in particular that there is no function f such that $\mathbf{F}=\nabla f$, because if there were then the integrals should have been equal by Example 2.

5. Let Γ be a C^1 curve realized by a C^1 path γ. What is the length of the curve Γ? If we imagine that Γ is a thin wire of unit density then its length is equal to its mass, so the length is $\int_{\Gamma_\gamma} 1\,ds = \int_a^b |\dot\gamma(t)|\,dt$. Does this agree with the formula from 1-dimensional calculus? (Hint: It does.)

How does $\int_{\Gamma_\gamma} \mathbf{F}\cdot d\mathbf{s}$ depend on γ? If Γ is a fixed curve, we say that $\gamma_i : [a_i, b_i] \to \Gamma$, $i = 1,2$ are parameterizations of Γ if they are both C^1 paths with non-zero velocity and their image is Γ. We say that γ_2 is a *reparameterization* of γ_1 if there is $h : [a_2, b_2] \to [a_1, b_1]$, increasing and bijective, such that $\gamma_2 = \gamma_1 \circ h$.

6.1.3 Theorem. *If γ_2 is a reparameterization of γ_1 then $\int_{\Gamma_{\gamma_1}} \mathbf{F}\cdot d\mathbf{s} = \int_{\Gamma_{\gamma_2}} \mathbf{F}\cdot d\mathbf{s}$.*

PROOF: Suppose $\gamma_2 = \gamma_1 \circ h$. Then $\dot\gamma_2 = (\dot\gamma_1 \circ h)\dot h$, so

$$\begin{aligned}\int_{\Gamma_{\gamma_2}} \mathbf{F}\cdot d\mathbf{s} &= \int_{a_2}^{b_2} \mathbf{F}(\gamma_2(t))\dot\gamma_2(t)dt \\ &= \int_{a_2}^{b_2} \mathbf{F}(\gamma_1(h(t)))\dot\gamma_1(h(t))\dot h(t)dt \\ &= \int_{a_1}^{b_1} \mathbf{F}(\gamma_1(t))\dot\gamma_1(t)dt = \int_{\Gamma_{\gamma_1}} \mathbf{F}\cdot d\mathbf{s}.\end{aligned}$$ □

We are no longer tied down to a specific parameterization of Γ. If we don't like the original then we can choose a more convenient reparameterization. This gives rise to the question of what it means for two paths to be equivalent. In the homework we will find out that it is always possible (because $|\dot\gamma(t)| \neq 0$) to choose a special parameterization $\gamma_u : [0, \text{length}(\Gamma)] \to \Gamma$, called the *unit-length parameterization*, with $|\gamma_u(t)| = 1$ for all $t \in [0, \text{length}(\Gamma)]$. The fact that $\int_\Gamma \mathbf{F}\cdot d\mathbf{s}$ is independent of the parameterization of Γ is an important philosophical point. Colloquially, the length of a racetrack doesn't depend on the speed at which you choose to drive it.

6.2 Parameterized surfaces

This section is a light introduction to differential geometry.

6.2.1 Examples.

1. Let $\Phi : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ be defined by

$$\Phi(\theta, \varphi) := (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi).$$

These are spherical coordinates with $\rho = 1$. For the correct choice of D we will obtain the unit sphere as the image of Φ. Taking $D = [0, 2\pi] \times [0, \pi]$ works, with the line $\{\varphi = 0\}$ mapped to the north pole, $\{\varphi = \pi/2\}$ to the equator, and $\{\varphi = \pi\}$ to the south pole. D is a nice 2-dimensional surface and the image of D under Φ is the unit sphere, another 2-dimensional surface.

2. Let $r < R$ be given and define $\Phi : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ by

$$\Phi(u, v) := ((R + r\cos v)\cos u, (R + r\cos v)\sin u, r\sin v).$$

With $D = [0, 2\pi] \times [0, \pi]$ as above, the image of Φ is the torus (donut) with radius R and tube radius r.

[Insert image.]

For a general $\Phi : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$ to define a surface, it must at least be the case that $\Phi \in C^1$ and $D\Phi$ has full rank. Then $S := \Phi(D)$ will define a nice 2-dimensional surface in $\mathbb{R}^3$. Contrast this with level surfaces. What is the tangent plane/space at $x_0 \in S$? It is an affine space in $\mathbb{R}^3$. From the definition of tangent space, $T(x_0)$ is the unique linear space such that all possible C^1 paths γ on S which pass through x_0, say $\gamma(0) = x_0$, satisfy $\dot\gamma(0) \in T(x_0)$. Suspend disbelief for a moment and suppose that Φ is invertible in a neighbourhood of x_0. Let γ be a C^1 path in S with $\gamma(0) = x_0$. Make a new path $\alpha(t) = \Phi^{-1}(\gamma(t))$, and suppose $\alpha(0) = y_0$. Assuming that D is a domain in $\mathbb{R}^2$, we know that $\dot\alpha(0)$ is a linear combination of $e_1, e_2 \in \mathbb{R}^2$. In fact, $e_1 = \frac{d}{ds}(y_0 + se_1)|_{s=0}$ and $e_2 = \frac{d}{ds}(y_0 + se_2)|_{s=0}$ are two standard paths passing through y_0 at $s = 0$ in D. The images of these paths in Φ should hence span $T(x_0)$ on S. Explicitly, $T(x_0)$ should be the spanned by $\frac{d}{ds}\Phi(y_0 + se_i)|_{s=0} = D\Phi(y_0)e_i$, i.e. by the columns of $D\Phi(y_0)$. Let Φ_u denote the first column and Φ_v denote the second column, so $T(x_0) = \operatorname{span}\{\Phi_u(y_0), \Phi_v(y_0)\}$, where $\Phi(y_0) = x_0$. This illustrates why it is important that $D\Phi$ have full rank. $T(x_0)$ is a 2-dimensional affine space in $\mathbb{R}^3$, so it should be possible to describe it using the normal vector. If $\mathbf{n}(x_0)$ is the normal vector to $T(x_0)$ in $\mathbb{R}^3$ then the canonical choice is $\Phi_u \times \Phi_v$, the cross product. The tangent plane to S at x_0 is $x_0 + \operatorname{span}\{\Phi_u, \Phi_v\}$, also written as $\mathbf{n}(x_0) \cdot (x - x_0) = 0$.

The definition above is perfectly correct for our uses, but the intuition can be horribly wrong. We assumed that Φ is locally invertible in a nice way.

6.2.2 Examples.

1. Let $\gamma : [a, b] \to S$ be a path on the unit sphere mapping to the equator with $\gamma(0) = (1, 0, 0)$, and let $\Phi(u, v) := (\sin v \cos u, \sin v \sin u, \cos v)$, which maps $[0, 2\pi) \times [0, \pi)$ onto S.

   [Insert image.]

   Does $\Phi^{-1} \circ \gamma$ really give a C^1 path in D that can be used to compute a velocity vector, $\dot\alpha$, in the tangent space of $\mathbb{R}^2$ at $y_0 = (0, \pi/2)$? In this case, no, because y_0 is on the boundary of D. Instead we would want to use another map $\tilde\Phi : \tilde D \to S$ such that the $\tilde y_0$ corresponding to $(1, 0, 0)$ is an interior point of $\tilde D$.

   [Insert image.]

2. Let S be the graph of $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ living in $\mathbb{R}^3$; $S = \{(u, v, f(u, v))\}$. Let $\Phi(u, v) = (u, v, f(u, v))$ and compute $T(x_0) = T(u_0, v_0, f(u_0, v_0))$ using this parameterization. Check that it gives the same tangent plane as we computed before.

$$\Phi_u(u_0, v_0) = \begin{pmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial u}(u_0, v_0) \end{pmatrix} \qquad \Phi_v(u_0, v_0) = \begin{pmatrix} 0 \\ 1 \\ \frac{\partial f}{\partial v}(u_0, v_0) \end{pmatrix}$$

so $\mathbf{n}(x_0) = \Phi_u \times \Phi_v = (-\frac{\partial f}{\partial u}, -\frac{\partial f}{\partial v}, 1)^T$. Therefore the tangent plane is the set of solutions to $\mathbf{n}(x_0) \cdot (x - x_0) = 0$, i.e.

$$\begin{aligned} x_3 &= f(u_0, v_0) + \frac{\partial f}{\partial u}(x_2 - u_0) + \frac{\partial f}{\partial v}(x_2 - v_0) \\ &= f(u_0, v_0) + \nabla f(u_0, v_0) \cdot (x_1 - u_0, x_2 - v_0) \end{aligned}$$

which is exactly the graph of the best affine approximation.

6.3 Surface integral

What about integrals and/or areas? As before, we can break the domain D into manageable pieces (rectangles), R_{ij}, and hence cover S with the "deformed" rectangles $\Phi(R_{ij})$. We will need to multiply by an appropriate scale factor to account for the deformation, and then proceed as usual. For area, we have $\text{area}(S) \approx \sum_{i,j} \text{area}(\Phi(R_{ij}))$. If Φ is linear then, since the R_{ij} are rectangles, this latter area is very easy to compute. Say $\Phi(x) = Ax$, where A is a 3×2 matrix. Then, from linear algebra, $\text{area}(\Phi(R_{ij}))$ is $|A_1 \times A_2|$ times the area of R_{ij}. Taking the limit as the number of rectangles goes to infinity, $\text{area}(S) = \iint_D |\Phi_u \times \Phi_v| \, dA$. This formula implicitly assumes that $|\Phi_u \times \Phi_v|$ is continuous (so that we may take the Riemann integral), which will be the case when Φ is C^1.

Similarly, this works to integrate a function $\rho : S \to \mathbb{R}$ or $F : S \to \mathbb{R}^3$, obtaining

$$\iint_S \rho \, dS := \iint_D (\rho \circ \Phi) |\Phi_u \times \Phi_v| \, dA$$

and

$$\iint_S \mathbf{F} \cdot d\mathbf{S} := \iint_D (\mathbf{F} \circ \Phi) \cdot (\Phi_u \times \Phi_v) \, dA.$$

7 Vector Calculus

There are some important relationships between area, volume, line, and surface integrals that bear the names of such eminent mathematicians as Green, Gauss, and Stokes.

7.1 Divergence and curl

A function $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ may be thought of as a *velocity field*, where the velocity of a particle at position x is given by $\mathbf{F}(x)$.

7.1.1 Definition. The *divergence* of $\mathbf{F}$ is defined to be the local, normalized, measurement of the tendency per unit time for the flow governed by $\mathbf{F}$ to expand a region of particles.

While intuitive, this definition gives no hint in how to compute it. By *local* we mean, for $x_0 \in \mathbb{R}^3$, we care only about the behavior close to x_0, i.e. in $B_r(x_0)$ for very small r. By *normalized* we mean that the expansion should be weighted by dividing by the test volume, $\Delta V/V$.

[Insert image of $B_r(x_0)$ with a small rectangle highlighted.]

Divide the surface of $B_r(x_0)$ into small "rectangles" R_{ij}.

[Insert image of the small rectangle with the normal vector highlighted.]

[Insert image of the deformed rectangular prism.]

$\Delta x = \mathbf{F} \cdot \mathbf{n} \Delta t$, so $\Delta V_{ij} = \mathbf{F} \cdot \mathbf{n}\,\text{area}(R_{ij})$. The change in volume per unit time (i.e. after dividing by Δt) is hence $\iint_{\partial B_r(x_0)} \mathbf{F} \cdot \mathbf{n}\, dS$. If $\mathbf{F}$ is C^1 and r is very small then by Taylor's theorem,

$$\mathbf{F}(y) = \mathbf{F}(x_0) + D\mathbf{F}(x_0)(y - x_0) + \text{err}(|y - x_0|)$$

for $y \in \partial B_r(x_0)$. To normalize we divide by $|B_r(x_0)|$ and obtain

$$\begin{aligned}\text{divergence} = \frac{1}{|B_r(x_0)|}\Bigg(&\iint_{\partial D_r(x_0)} \big((D\mathbf{F}(x_0)(y - x_0)\big) \cdot \mathbf{n}(y)\, dS(y) \\ &+ \underbrace{\iint_{\partial D_r(x_0)} \mathbf{F}(x_0) \cdot \mathbf{n}(y)\, dS(y)}_{=0} + \underbrace{\text{err}(r) \iint_{\partial D_r(x_0)} \mathbf{n}(y)\, dS(y)}_{=O(1/r)} \Bigg) \\ &= \text{by homework, last two terms die.}\end{aligned}$$

We can simplify further by noting that $\mathbf{n}(y) = (y - x_0)/r$ for $y \in \partial B_r(x_0)$. We know $y^T A y = \sum_{i,j} y_i y_j A_{ij}$, so

$$\big(D\mathbf{F}(x_0)(y - x_0)\big) \cdot \mathbf{n}(y) = \frac{1}{r} \sum_{i,j} (y - x_0)_i (y - x_0)_j [D\mathbf{F}(x_0)]_{ij}.$$

Furthermore,

$$\iint_{\partial B_r(x_0)} (y - x_0)_i (y - x_0)_j\, dS(y) = \begin{cases} \frac{1}{3} r^3\, \text{area}(\partial B_r(x_0)) & i = j \\ 0 & i \neq j. \end{cases}$$

Thus

$$\begin{aligned}\text{divergence} &= \frac{1}{|B_r(x_0)|}\iint_{\partial D_r(x_0)} \big((D\mathbf{F}(x_0)(y-x_0)\big)\cdot \mathbf{n}(y)\,dS(y)\\ &= \frac{3}{4\pi r^3}\frac{1}{r}\sum_{i=1}^{3}[D\mathbf{F}(x_0)]_{ii}\frac{1}{3}r^2(4\pi r^2)\\ &= \frac{\partial \mathbf{F}_1}{\partial x_1}(x_0)+\frac{\partial \mathbf{F}_2}{\partial x_2}(x_0)+\frac{\partial \mathbf{F}_3}{\partial x_3}(x_0)\end{aligned}$$

We will denote the divergence of $\mathbf{F}$ by $\operatorname{div}(\mathbf{F}) = \nabla\cdot\mathbf{F} = \operatorname{tr}(D\mathbf{F})$.

7.1.2 Definition. The *curl* of $\mathbf{F}$ is a local measurement of the tendency of $\mathbf{F}$ to cause a particle to undergo rotational motion.

Heuristically, in $\mathbb{R}^2$, rotation happens when movement to the right implies movement up, i.e. $\frac{\partial \mathbf{F}_2}{\partial x_1}$ is positive, and movement up implies movement left, i.e. $\frac{\partial \mathbf{F}_1}{\partial x_2}$ is negative. Whence a local measure of rotation in $\mathbb{R}^2$ is

$$\frac{\partial \mathbf{F}_2}{\partial x_1}-\frac{\partial \mathbf{F}_1}{\partial x_2} =: \operatorname{curl}_{\mathbb{R}^2}(\mathbf{F})(x_0).$$

In $\mathbb{R}^3$, angular tendency is best recorded as a vector quantity, i.e. as angular momentum. According to the right-hand rule, the rotation in the x_1-x_2, x_2-x_3, and x_2-x_3 planes should be recorded as

$$\begin{aligned}&\left(\frac{\partial \mathbf{F}_3}{\partial x_2}-\frac{\partial \mathbf{F}_2}{\partial x_3}\right)\mathbf{e}_1-\left(\frac{\partial \mathbf{F}_3}{\partial x_1}-\frac{\partial \mathbf{F}_1}{\partial x_3}\right)\mathbf{e}_2+\left(\frac{\partial \mathbf{F}_2}{\partial x_1}-\frac{\partial \mathbf{F}_1}{\partial x_2}\right)\mathbf{e}_3\\ &\qquad = \det\begin{pmatrix} e_1 & e_2 & e_3\\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3}\\ \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{F}_3\end{pmatrix} =: \operatorname{curl}(\mathbf{F})\end{aligned}$$

Last time we introduced operations on C^1 vector fields $\mathbf{F}: D\subseteq\mathbb{R}^3\to\mathbb{R}^3$. They were

$$\operatorname{div}(\mathbf{F})(x_0) = \operatorname{tr}(DF(x_0)) = (\nabla\cdot\mathbf{F})(x_0),$$

a scalar field, and

$$\operatorname{curl}(\mathbf{F})(x_0) = (\nabla\times\mathbf{F})(x_0),$$

a vector field, where $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ is the "grad" operator.

7.2 Orientation of boundaries

Let D be a domain in $\mathbb{R}^2$ and suppose its boundary is a C^1 curve $\Gamma = \partial D$, parameterized by a C^1 path $\sigma:[0,L]\to\Gamma$ with unit velocity (see homework). Up to a plus or minus, this standardizes the tangent spaces along Γ, $T(x_0) = \operatorname{span}\{\dot\sigma(t_0)\}$

when $\sigma(t_0) = x_0$, in the sense that $\dot\sigma$ is a unit length basis vector. The only choice is which direction to walk around Γ. We will say that σ gives ∂D a *positive orientation* if an ant walking around ∂D according to σ always has the interior of D to its left hand side. Another way to describe positive orientation is to require, at each point $x_0 \in \partial D$, that the infinitesimal motion is in the direction of $e_3 \times \mathbf{n}(x_0)$. Here we embed D in $\mathbb{R}^3$ via the identification of $\mathbb{R}^2$ in $\mathbb{R}^3$ as $\{x_3 = 0\}$ and $\mathbf{n}$ is the unit outward pointing normal of ∂D in this plane. It is perpendicular to the velocity vector of any path that parameterizes ∂D (a useful fact for computing it). We say that D is a *nice domain* or *Jordan domain* in $\mathbb{R}^2$ if it is bounded, (path-)connected, and it is the region contained between a finite number of simple, closed, piecewise C^1 curves $\Gamma_1, \dots, \Gamma_k$. (A closed curve Γ is simple and closed if it is parameterized by some $\gamma : [a,b] \to \Gamma$ for which $\gamma(s) = \gamma(t)$ if and only if $\{s,t\} = \{a,b\}$. Yet otherwise said, Γ is a closed curve with no self-intersections.)

7.2.1 Examples.

1. Image of a circular-ish blob. Yes.
2. Image of a beer bottle opener. Yes.
3. Image of a smiley face. Yes, ∂D has five curves making it up and some curves have multiple C^1 pieces.
4. Image of a plane minus a triangle. No, because it is not bounded.

Remark.

1. We will always assume that ∂D implies a positive orientation. $-\partial D$ means the negative orientation (not the usual meaning of minus of a set).
2. $-\partial D$ is parameterized by $\hat\sigma(t) := \sigma(L-t)$. We will use the hat notation for paths to mean the path traversed in the reverse direction, so that $\Gamma_{\hat\gamma} = -\Gamma_\gamma$.
3. $\int_{-\Gamma} \mathbf{F}\cdot ds = -\int_\Gamma \mathbf{F}\cdot ds$.

7.2.2 Theorem (Green's theorem).
Let $\mathbf{F}$ be a C^1 vector field defined on a bounded domain $\tilde D$ containing a nice domain D, with $\overline{D} \subset \tilde D$ (so D is compactly contained within $\tilde D$). Then

$$\int_{\partial D} \mathbf{F}\cdot d\mathbf{s} = \iint_D \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) dA.$$

Or, as it is typically seen, with $F = (P,Q)$,

$$\int_{\partial D} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA.$$

PROOF (SKETCH): We need to show that a certain line integral is equal to a certain double integral. We first argue that it suffices to do the calculation when D is a triangle with one side parallel to the x-axis and one side parallel to the y-axis. Indeed, any polygon can be built up from such triangles, and since D is a nice domain its boundary is piecewise C^1, and hence can be approximated well by polygonal curves. We need to show that the double integrals over these polygonal domains

converge to the double integral over D, and then do the calculation directly for a triangle.

[Insert image of a triangulated domain.]

The only edges in the triangulation of D_Δ which are not traversed twice in opposite directions after applying Green's theorem to each sub-triangle are the exterior edges, Γ_Δ. The theorem for triangles hence adds up to the theorem for D_Δ and $\partial D_\Delta = \Gamma_\Delta$. If we show that the errors in the integrals go to zero as the error in the polygonal approximation goes to zero, then we are done.

Assume that Γ is parameterized by $\gamma : [a,b] \to \Gamma$ and Γ_Δ is the polygonal curve corresponding to the time points $\{a = t_0, \ldots, t_N = b\}$. For a single segment,

[Insert image of a single segment $[\gamma(t_{i-1}), \gamma(t_i)]$].

The area error is not more than $\max(|\dot\gamma|^2)(\Delta t)^2$ and the length of the linear piece is about $\max(|\dot\gamma|)\Delta t$.

$$\begin{aligned}\left|\iint_{D_\Delta} G\,dxdy - \iint_D G\,dxdy\right| &\le \iint_{D\Delta D_\Delta} |G|\,dxdy \\ &\le \sum_{i=1}^{N} \max(|G|)(\text{max area of symmetric diff}) \\ &\le \Delta t \max(|G|)|\tilde D| \to 0 \quad \text{as } \Delta t \to 0\end{aligned}$$

Note that $\int_{\Gamma_\Delta} \mathbf{F}\cdot ds = \sum_{i=1}^N \mathbf{F}(\gamma(t_i))\cdot(\gamma(t_i)-\gamma(t_{i-1}))$ since Γ_Δ is a polygonal curve.

$$\begin{aligned}\left|\int_{\Gamma_\Delta} \mathbf{F}\cdot ds - \int_\Gamma \mathbf{F}\cdot ds\right| &\le \sum_{i=1}^N \left|\mathbf{F}(\gamma(t_i))\cdot(\gamma(t_i)-\gamma(t_{i-1})) - \int_{t_{i-1}}^{t_i} \mathbf{F}(\gamma(t))\dot\gamma(t)dt\right| \\ &\le \sum_{i=1}^N \int_{t_{i-1}}^{t_i} |\mathbf{F}(\gamma(t_i)) - \mathbf{F}(\gamma(t))||\dot\gamma(t)|dt \\ &\le \max(|D\mathbf{F}|)\text{length}(\Gamma)\Delta t\end{aligned}$$

Now we prove Green's theorem for a triangle with one edge parallel to each axis. Without loss of generality, the triangle has one vertex at the origin and may be described by the following curves.

$$\begin{aligned}\gamma_1(t) &= (0, b-t) && t\in[0,b] \\ \gamma_2(t) &= (t,0) && t\in[0,a] \\ \gamma_3(t) &= (a-t, bt/a) && t\in[0,a]\end{aligned}$$

Note that the above parameterization gives a positive orientation. Then

$$\begin{aligned}\iint_D \frac{\partial F_2}{\partial x}\,dA &= \int_0^b \int_0^{a-ay/b} \frac{\partial F_2}{\partial x}\,dxdy \\ &= \int_0^b F_2(a-ay/b, y) - F_2(0,y)\,dy\end{aligned}$$

$$\iint_D -\frac{\partial F_1}{\partial y}\,dA = \int_0^a \int_0^{b-bx/a} -\frac{\partial F_1}{\partial y}\,dy\,dx$$
$$= \int_0^a -F_1(x, b-bx/a) + F_1(x,0)dx$$

We compute the line integral by breaking it up as $\int_\Gamma = \int_{\Gamma_{\gamma_1}} + \int_{\Gamma_{\gamma_2}} + \int_{\Gamma_{\gamma_3}}$. Note that $\dot\gamma_1 = (0,-1)$, $\dot\gamma_2 = (1,0)$, and $\dot\gamma_3 = (-1, b/a)$.

$$\int_{\Gamma_{\gamma_1}} \mathbf{F}\cdot ds = \int_0^b \mathbf{F}(0, b-t)\cdot(0,-1)\,dt = \int_0^b -F_2(0,b-t)\,dt$$
$$\int_{\Gamma_{\gamma_2}} \mathbf{F}\cdot ds = \int_0^a \mathbf{F}(t,0)\cdot(1,0)\,dt = \int_0^a F_1(t,0)\,dt$$
$$\int_{\Gamma_{\gamma_3}} \mathbf{F}\cdot ds = \int_0^a \mathbf{F}(a-t, bt/a)\cdot(-1, b/a)\,dt$$
$$= \int_0^a -F_1(a-t, bt/a) + (b/a)F_2(a-t, bt/a)\,dt$$
$$= \int_0^a -F_1(t, b(a-t)/a)\,dt + \int_0^b F_2(a-as/b, s)\,ds$$

The sum of these integrals is equal to the result obtain for the double integral, so Green's theorem is proved. □

7.2.3 Examples.

1. Show that $\text{area}(D) = \frac{1}{2}\int_{\partial D} x\,dy - y\,dx$. Let $\mathbf{F}(x,y) = (-y,x)$ and $\dot\gamma\,dt = (dx,dy)$, so that $\mathbf{F}\cdot d\mathbf{s} = x\,dy - y\,dx$. Then, by Green's theorem,
$$\int_{\partial D} \mathbf{F}\cdot ds = \iint_D (1-(-1))\,dA = 2\,\text{area}(D).$$
2. Let Γ be the collection of four line segments connecting, in order, the points $(0,0)$, $(1,1)$, $(0,1)$, $(0,1)$, $(0,0)$. Evaluate $\int_\Gamma (y^4+x^3)\,dx + 2x^6\,dy$. Here $\mathbf{F}(x,y) = (y^4+x^3, 2x^6)$. We may use Green's theorem with $D = [0,1]^2$ (if it is convenient to do so) since $\Gamma = \partial D$. The line integral hence is equal to
$$\iint_D (12x^5 - 4y^3)\,dA = \int_0^1\int_0^1 (12x^5-4y^3)\,dx\,dy = 2-1 = 1.$$

7.3 Flux and Gauss' theorem

Given a vector field $\mathbf{F}:\mathbb{R}^2\to\mathbb{R}^2$, the *flux* is the rate of flow through a surface, which is 2-dimensions is the rate across a curve Γ bounding a region D. Let $\mathbf{n}$ be the outward normal vector. The amount going in across a small piece $\Delta\Gamma$ is

"$-\mathbf{F}\cdot\mathbf{n}\Delta\Gamma$," where the minus sign appears because $\mathbf{n}$ points outward. After taking Riemann sums, the total flux is seen to be $\int_\Gamma(\mathbf{F}\cdot\mathbf{n})\,ds$. With the same assumptions on the vector field and on the domain as in Green's theorem, we obtain Gauss' theorem,

$$\int_\Gamma(\mathbf{F}\cdot\mathbf{n})\,ds=\iint_D \operatorname{div}(\mathbf{F})\,dA.$$

The proof is straightforward. Assume that Γ has unit-length parameterization γ, with a positive orientation, so that $\mathbf{n}=(\dot\gamma_2,-\dot\gamma_1)$. Let $\mathbf{G}=(-F_2,F_1)$, so that

$$\int_\Gamma(\mathbf{F}\cdot\mathbf{n})\,ds=\int_\Gamma\mathbf{G}\cdot d\mathbf{s}=\iint_D\frac{\partial G_2}{\partial x}-\frac{\partial G_1}{\partial y}\,dA=\iint_D\operatorname{div}(\mathbf{F})\,dA$$

7.3.1 Example. Compute the flux of $\mathbf{F}(x,y)=(y^3,x^5)$ across the boundary of the unit square. By Gauss' theorem,

$$\int_\Gamma(\mathbf{F}\cdot\mathbf{n})\,ds=\iint_D\operatorname{div}(\mathbf{F})\,dA=\iint_D 0\,dA=0.$$

7.4 Stoke's theorem

Green's theorem may be written as

$$\iint_D\operatorname{curl}_{\mathbb{R}^2}(\mathbf{F})\,dA=\iint_D\frac{\partial F_2}{\partial x}-\frac{\partial F_1}{\partial y}\,dA=\int_{\partial D}\mathbf{F}\cdot d\mathbf{s}$$

Heuristically, how does rotation "add up" locally? Rotation in opposite directions cancels, so the net rotation in a region is what is happening around the boundary. What if particles are moving according to $\mathbf{F}$ in $\mathbb{R}^3$, but are constrained to lie on some 2-dimensional surface? Is there an analog to Green's theorem for this situation?

Consider $\Phi: D\subseteq\mathbb{R}^2_{u,v}\to\mathbb{R}^3$ defining a surface $S=\Phi(D)$ in $\mathbb{R}^3$. Perhaps it has a boundary ∂S. We need to define orientation on S and ∂S. For a single (2-dimensional) triangle sitting in $\mathbb{R}^3$ with a specified "normal direction" (there are two to choose from) both notions are clear. Pasting triangles together along common edges, with "compatible" normal directions, gives an orientation for the resulting surface and its boundary. Adding up the integrals over the boundaries of all the triangles will result in integrals over the common edges cancelling out, leaving only the integral over the boundary of the combined surface.

On any given small triangle (with specified normal direction $\mathbf{n}$), the contribution of the curl in the plane of that triangle is $\operatorname{curl}(\mathbf{F})\cdot\mathbf{n}$.

7.4.1 Theorem (Stoke's theorem). *Let $\tilde\Omega$ be bounded domain in $\mathbb{R}^3$ (so $\tilde\Omega$ is open, connected, and bounded) and let $S=\Phi(D)$ be a smooth, 2-dimensional, positively oriented surface in $\mathbb{R}^3$, compactly contained in $\tilde\Omega$. If $\mathbf{F}:\tilde\Omega\to\mathbb{R}^3$ is a C^1 vector field then*

$$\int_{\partial S}\mathbf{F}\cdot d\mathbf{s}=\iint_S\operatorname{curl}(\mathbf{F})\cdot d\mathbf{S}=\iint_D(\operatorname{curl}(\mathbf{F})\circ\Phi)\cdot(\Phi_u\times\Phi_v)\,dA.$$

For our purposes, $S = \Phi(D)$, where $D \subseteq \mathbb{R}^2_{u,v}$ and Φ is smooth, injective, and *orientation preserving*, i.e. $\det(D\Phi) > 0$. Then $\partial\Omega$ inherits orientation from ∂D via Φ. The same intuitive definition applies: $\partial\Omega$ has a positive orientation if a person walking along $\partial\Omega$ in the direction of the parameterization, with head pointing in the direction of the normal vector, the surface is on the left hand side.

7.4.2 Example. Let $\mathbf{F}(x,y,z) = (y,-x,e^{xz})$. Compute $\iint_\Omega \nabla\times\mathbf{F}\cdot d\mathbf{S}$ where Ω is the larger portion of the sphere of radius 2 which has been sliced in such a way that the boundary is the circle of radius 1 lying in the x-y plane, with orientation given by the usual outward pointing normal for the sphere (anti-clockwise in the plane). It would be a real pain to parameterize Ω and compute curl($\mathbf{F}$), but of course we have Stoke's theorem, which says that $\iint_\Omega \nabla\times\mathbf{F}\cdot d\mathbf{S} = \int_{\partial\Omega}\mathbf{F}\cdot d\mathbf{s}$. Note that $\gamma(t) := (\cos t, \sin t, 0)$ parameterizes the boundary, so $\dot\gamma(t) = (-\sin t, \cos t, 0)$, whence $\mathbf{F}(\gamma(t))\cdot\dot\gamma(t) = -\sin^2 t - \cos^2 t = -1$, so the integral is -2π.

7.5 Proof of Stoke's theorem for a graph

Recall that Stoke's theorem states that $\iint_\Omega \text{curl}(\mathbf{F})\cdot d\mathbf{S} = \int_{\partial\Omega}\mathbf{F}\cdot d\mathbf{s}$, where Ω is a surface and $\partial\Omega$ is the positively oriented boundary of Ω. We will prove it now for the case where $\Omega = \{(x,y,z) : z = f(x,y), (x,y)\in D\}$ is the graph of a C^1 function $f : D\subseteq\mathbb{R}^2\to\mathbb{R}$. We need to decide what is the positive orientation for $\partial\Omega$ and calculate curl($\mathbf{F}$)$\cdot d\mathbf{S}$ using the parameterization $\Phi(u,v) = (u,v,f(u,v))$. In this case $\Phi_u\times\Phi_v = (-\frac{\partial f}{\partial v}, -\frac{\partial f}{\partial v}, 1)$ and recall that

$$\text{curl}(\mathbf{F}) = \Big(\underbrace{\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}}_{G_1}, \underbrace{-\Big(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}\Big)}_{G_2}, \underbrace{\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}}_{G_3}\Big).$$

Whence

$$\iint_\Omega \text{curl}(\mathbf{F})\cdot d\mathbf{S} = \iint_D G_1\Big(-\frac{\partial f}{\partial u}\Big) + G_2\Big(-\frac{\partial f}{\partial v}\Big) + G_3\, dA$$

For the parameterization of $\partial\Omega$ we must use a positively oriented parameterization of ∂D in $\mathbb{R}^2$. Let σ be such an orientation and let $\gamma(t) = (\sigma_1(t), \sigma_2(t), f(\sigma(t)))$. Then

$$\dot\gamma_3(t) = \nabla f(\sigma(t))\cdot\dot\sigma(t) = \frac{\partial f}{\partial u}\dot\sigma_1(t) + \frac{\partial f}{\partial v}\sigma_2(t),$$

so

$$\begin{aligned}\int_{\partial\Omega}\mathbf{F}\cdot d\mathbf{s} &= \int_a^b \mathbf{F}(\gamma(t))\cdot\dot\gamma(t)\,dt\\ &= \int_a^b F_1\dot\sigma_1 + F_2\dot\sigma_2 + F_3\frac{\partial f}{\partial x}\dot\sigma_1 + F_3\frac{\partial f}{\partial y}\sigma_2\,dt\\ &= \int_a^b \underbrace{\Big(F_1 + F_3\frac{\partial f}{\partial x}, F_2 + F_3\frac{\partial f}{\partial y}\Big)}_{\mathbf{H}(\sigma(t))}\dot\sigma(t)\,dt = \int_{\partial D}\mathbf{H}\cdot d\mathbf{s}\end{aligned}$$

Now we apply Green's theorem.

$$\begin{aligned}
\frac{\partial}{\partial y}H_1(x,y) &= \frac{\partial}{\partial y}\left(F_1(x,y,f(x,y)) + F_3(x,y,f(x,y))\frac{\partial f}{\partial x}(x,y)\right)\\
\frac{\partial H_1}{\partial y} &= \frac{\partial F_1}{\partial y} + \frac{\partial F_1}{\partial z}\frac{\partial f}{\partial y} + \left(\frac{\partial F_3}{\partial y} + \frac{\partial F_3}{\partial z}\frac{\partial f}{\partial y}\right)\frac{\partial f}{\partial x} + F_3\frac{\partial^2 f}{\partial y \partial x}\\
\frac{\partial H_2}{\partial x} &= \frac{\partial F_2}{\partial x} + \frac{\partial F_2}{\partial z}\frac{\partial f}{\partial x} + \left(\frac{\partial F_3}{\partial x} + \frac{\partial F_3}{\partial z}\frac{\partial f}{\partial x}\right)\frac{\partial f}{\partial y} + F_3\frac{\partial^2 f}{\partial x \partial y}
\end{aligned}$$

Subtracting and applying Green's theorem,

$$\begin{aligned}
\int_{\partial D}\mathbf{H}\cdot d\mathbf{s} &= \iint_D \frac{\partial H_2}{\partial x} - \frac{\partial H_1}{\partial y}\,dA\\
&= \iint_D \frac{\partial F_2}{\partial x} + \frac{\partial F_2}{\partial z}\frac{\partial f}{\partial x} + \frac{\partial F_3}{\partial x}\frac{\partial f}{\partial y} - \frac{\partial F_1}{\partial y} - \frac{\partial F_1}{\partial z}\frac{\partial f}{\partial y} - \frac{\partial F_3}{\partial y}\frac{\partial f}{\partial x}\,dA\\
&= \iint_D \left(\frac{\partial F_2}{\partial z} - \frac{\partial F_3}{\partial y}\right)\frac{\partial f}{\partial x} + \left(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}\right)\frac{\partial f}{\partial y} + \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\,dA\\
&= \iint_D G_1\left(-\frac{\partial f}{\partial u}\right) + G_2\left(-\frac{\partial f}{\partial v}\right) + G_3\,dA\\
&= \iint_\Omega \mathrm{curl}(\mathbf{F})\cdot d\mathbf{S}.
\end{aligned}$$

This completes the proof of Stokes' Theorem for the case that Ω is the surface made by the graph of f over $D \subset \mathbb{R}^2$.

7.5.1 Example. Use Stoke's theorem to evaluate $\int_\Gamma -y^3dx + x^3dy - z^3dz$, where Γ is the curve in $\mathbb{R}^3$ made by intersection of the cylinder $\{x^2+y^2=1\}$ and the plane $\{x+y+z=1\}$, oriented anticlockwise in the x-y plane.

[Insert diagram.]

To apply Stoke's theorem we need to identify some Ω such that $\partial\Omega = \Gamma$ with the correct orientation. First let's compute curl($\mathbf{F}$) to see whether it is worth applying Stoke's theorem. $\mathbf{F} = (-y^3, x^3, -z^3)$, so curl($\mathbf{F}$) $= (0, 0, 3x^2+3y^2)$. Let's take Ω to be the graph of $z = 1-x-y$ over the unit circle in the x-y plane, i.e. $\Phi(u,v) = (u, v, 1-u-v)$. The vector normal is $(1,1,1)$, so

$$\int_\Gamma \mathbf{F}\cdot d\mathbf{s} = \iint_D (3x^2+3y^2)\,dA = 3\int_0^{2\pi}\int_0^1 r^2(r\,dr\,d\theta) = \frac{3\pi}{2}.$$

Recall that Stoke's theorem says that if you evaluate the surface integral the curl of a smooth vector field over a surface then this is equal to the line integral of the vector field over the positively oriented boundary of that surface.

7.5.2 Corollary. *If Ω_1 and Ω_2 are two nice surfaces with the same boundary Γ (and the same positive orientation) then*

$$\iint_{\Omega_1} \operatorname{curl}(\mathbf{F}) \cdot d\mathbf{S} = \iint_{\Omega_2} \operatorname{curl}(\mathbf{F}) \cdot d\mathbf{S}.$$

Note that $\mathbf{n}dS = d\mathbf{S}$. Indeed, assume that $\Omega = \Phi(D)$ is a parameterized surface and $\Phi_u \times \Phi_v$ is in the outward pointing normal direction at Ω at each point. The unit normal is hence

$$\mathbf{n}(x) = \frac{\Phi_u(x) \times \Phi_v(x)}{|\Phi_u(x) \times \Phi_v(x)|}.$$

Since $dS = |\Phi_u \times \Phi_v| du dv$, it follows that

$$\mathbf{n}dS = \frac{\Phi_u \times \Phi_v}{|\Phi_u \times \Phi_v|} |\Phi_u \times \Phi_v| du dv = (\Phi_u \times \Phi_v) du dv = d\mathbf{S}.$$

7.5.3 Examples.

1. Evaluate $\iint_\Omega (\nabla \times \mathbf{F}) \cdot \mathbf{n}\, dS$ where Ω is the portion of the unit sphere such that $x + y + z \geq 1$ also holds and $\mathbf{F} = (x, y, z) \times (1, 1, 1)$. By the corollary to Stoke's theorem, we can instead choose the "flat" surface with the same boundary as Ω and do the computation over it. This has the advantage of having a constant (rather than varying) normal vector. Let Ω_2 be the circle lying inside the plane $x + y + z = 1$ with boundary $\partial\Omega$. The normal vector to Ω_2 is $(1, 1, 1)/\sqrt{3}$, so the integral is equal to $\iint_{\Omega_2} \nabla \times \mathbf{F} \cdot d\mathbf{s}$. It can be shown that $\nabla \times \mathbf{F} = (-2, -2, -2)$, implying that the integral is equal to $-6/\sqrt{3}$ times the area of the circle.
2. Evaluate $\int_\Gamma \mathbf{F} \cdot d\mathbf{s}$ where $\Gamma \subseteq \mathbb{R}^2$ is the unit circle, $\Gamma = \{x^2 + y^2 = 1\}$, and

$$\mathbf{F}(x, y) = \left(\frac{y}{x^2 + y^2}, \frac{-x}{x^2 + y^2}, 1 \right).$$

Is it okay to naively use Stoke's theorem to try to compute $\iint_\Omega \operatorname{curl}(\mathbf{F}) \cdot d\mathbf{S}$ for some surface in $\mathbb{R}^3$ whose boundary is the unit circle? This would be fortuitous because $\operatorname{curl}(\mathbf{F}) = (0, 0, 0)$, so the integral would be zero by Stoke's theorem. However, the line integral is not too hard, so we can compute it via brute force. Let $\gamma(t) = (\cos t, \sin t)$, so $\dot{\gamma}(t) = (-\sin t, \cos t)$ and $\mathbf{F}(\gamma(t)) \cdot \dot{\gamma}(t) = -1$. Hence the line integral is actually -2π, which is not zero.

What happened? Stoke's theorem requires that $\mathbf{F} \in C^1(\tilde{\Omega})$ for some open domain in $\mathbb{R}^3$ with $\Omega \subset\subset \tilde{\Omega}$. Any such domain in $\mathbb{R}^3$ is going to lie over the point $(x, y) = (0, 0)$, which is a point of discontinuity of $\mathbf{F}$.

7.6 Gauss' theorem

Recall Homework 9, question 0.4, parts 1–3, which asks you show that, for a fixed constant vector $\mathbf{v}$,

$$\iint_{\partial B_r(x_0)} \mathbf{v} \cdot \mathbf{n}(y)\, dS(y) = 0.$$

Recall that Green's theorem identifies a trade off between difficulties: a lower dimensional integral is easier than a higher dimensional integral to compute, but the higher dimensional integral is of a differential operator applied to the vector field. Differential operators tend to simplify functions because they reduce the order of polynomials. The 3-dimensional analog is Gauss' theorem.

7.6.1 Theorem (Gauss' theorem). *Let $\mathbf{F} \in C^1(\tilde{W})$ be given, where $\tilde{W} \subseteq \mathbb{R}^3$ is a domain, and let $W \subset\subset \tilde{W}$ be a nice region with nice surface boundary ∂W. Then*

$$\iint_{\partial W} \mathbf{F} \cdot \mathbf{n}\, dS = \iiint_W \operatorname{div}(\mathbf{F})\, dV.$$

7.6.2 Examples.

1. Evaluate $\iint_\Omega \mathbf{F} \cdot d\mathbf{S}$, where $\mathbf{F} = (xy^2, x^2y, y)$ and Ω is the surface of the can $\{x^2 + y^2 = 1, -1 \le z \le 1\}$. Either we can evaluate three surface integrals (corresponding to the top, side, and bottom of the can), or try Gauss' theorem. Here $\operatorname{div}(\mathbf{F}) = y^2 + x^2 + 0$ ($= r^2$ in polar coördinates).

$$\begin{aligned}\iint_\Omega \mathbf{F} \cdot d\mathbf{S} &= \iint_\Omega \mathbf{F} \cdot \mathbf{n}\, dS \\ &= \iiint_{\text{can}} \operatorname{div}(\mathbf{F})\, dV \\ &= \int_{-1}^{1} \int_0^{2\pi} \int_0^1 r^2 (r\, dr\, d\theta\, dz) = \pi\end{aligned}$$

2. Compute the outward flux of $\mathbf{v} = (x^3, y^3, z^3)$ through the unit sphere. Recall that the total flux is the net volume rate of flow via $\mathbf{v}$ through the surface, which is $\iint_\Omega \mathbf{v} \cdot \mathbf{n}\, dS$. By Gauss' theorem, since $\operatorname{div}(\mathbf{v}) = 3(x^2 + y^2 + z^2)$, the flux is

$$\iiint_{\partial B_1(0)} 3(x^2 + y^2 + z^2)\, dV = \int_0^{2\pi} \int_0^{\pi} \int_0^1 3\rho^2 (\rho^2 \sin\psi\, d\rho\, d\psi\, d\theta) = \frac{12\pi}{5}$$

3. Show that an incompressible flow cannot point inward at all points on any given nice closed surface which bounds a nice region. In this case an *incompressible flow* is one for which there is no local volume change, i.e. for which the divergence is zero. Let $\mathbf{v} = \mathbf{v}(x, y, z)$ be such a flow and let W be a nice domain with nice boundary ∂W. By Gauss' theorem,

$$\iint_{\partial W} \mathbf{v} \cdot \mathbf{n}\, dS = \iiint_W \operatorname{div}(\mathbf{v})\, dV = 0.$$

If it were the case that $\mathbf{v}$ pointed in at all points of ∂W then it would be the case that $\mathbf{v} \cdot \mathbf{n} \le 0$ for all points on ∂W. If $\mathbf{v}$ is not the identically zero flow, then $\mathbf{v} \cdot \mathbf{n} < 0$ for some non-trivial portion of the surface. But then we would have $\iint_{\partial W} \mathbf{v} \cdot \mathbf{n}\, dS < 0$, which is a contradiction.

4. Let W be a nice region in $\mathbb{R}^3$ with a nice boundary. Prove that $\iint_{\partial W} \mathbf{r}\cdot\mathbf{n}\, dS = 3\,\mathrm{vol}(W)$, where $\mathbf{r} = (x, y, z)$. By Gauss' theorem the integral is equal to

$$\iiint_W \mathrm{div}(\mathbf{r})\, dV = 3 \iiint_W dV = 3\,\mathrm{vol}(W).$$

Index