MATH 425.1 ADVANCED MULTIVARIABLE CALCULUS

THE CLASSROOM TEACHING PART OF THE COURSE
CHAPTERS 1-6 IN THE BOOK

Contents

Date:

This first part of notes is on what has been covered in class, except for chapter 7 which I lumped to the second part as it is closest to the chapter 8 which will be covered online.

0. Introduction

This long introductory chapter starts in 0.1 with what is the course is about. We review the main ideas of calculus (0.2) and what is new in calculus when one considers functions of several variables: 2, 3 or more variables (0.3).

In the remainder of this chapter we cover more concrete material. In 0.4 we review the differentiation for several variables, the ideas and the notations. Then in 0.5 we explain integration from the point of view that will be used throughout the course.

0.1. **The nature of the course.** In the calculus of several variables course MATH233 we learned how calculus works for functions of several variables. However, we mostly concentrated on the functions of 2 or 3 variables since in these cases we could use the visual intuition. In this course we revisit the calculus of several variables MATH233 with some emphasis on doing everything carefully so that it works for functions of *any* number of variables. So, our geometry now now moves from the line $\mathbb{R}$, plane $\mathbb{R}^2$ and the space $\mathbb{R}^3$ to any any $\mathbb{R}^n$.

We will briefly review differentiation in this generality (the material in chapters 1-4 of the book), however our main focus will be on integration (chapters 5-8).

0.1.1. *Topics.* We will cover material from chapters 5-8. We will assume chapters 1-4. These and some other parts of our course have been covered in the Multivariable Calculus course MATH233.

- 5. Double and Triple Integrals
- 6. The Change of Variables Formula
- 7. Integrals over Paths and Surfaces
- 8. The Integral Theorems of Vector Analysis

0.2. **Calculus.** The main subjects of calculus are the *rate of change* (differentiation) and the *total* of a quantity which is spread over some region with a given density (integration).

The revolutionary idea that appears in both is that of *infinite processes* called *limits*.

The infinite processes (i.e., limits) are actually necessary in order to make sense of some quantities of obvious interest like (1) speed of an object at a given moment or (2) area under a curve. The abstract point of view on problems of this kind is called derivatives ("abstract rate of change") and integrals ("abstract areas").

While by definition such processes require an infinite amount of work, it turns out that often enough these processes are actually doable – however by thought rather than by work alone. The "thought" part is provided by mathematical methods of differentiation and integration. These allow us to efficiently calculate limits precisely in many situations (and approximately in other situations).

0.3. **Multidimensional calculus requires geometry.** From the calculus courses we know how to extend calculus from functions of one variable to functions of two, three or really any number of variables. One again considers the same two main ideas (differentiation and integration), however a new difficulty is added to the subject – the complexity of *geometry*.

In higher dimension, one has to consider more subtle geometric objects that we will call "*manifolds*" or even more generally "*spaces*"). The basic geometric objects in the one dimensional line $\mathbb{R} = \mathbb{R}^1$ are the intervals. In the 2 dimensional plane $\mathbb{R}^2$ we find *regions* and their boundaries which are called *curves*. In dimension 3 we find *solids* in $\mathbb{R}^3$ and their boundaries are called *surfaces*. In $\mathbb{R}^n$ for an arbitrary dimension n there are higher dimensional manifolds which are useful in many real life applications but we have difficulty visualizing these for $n > 3$.

0.3.1. *The idea of* $\boxed{\textit{space}}$. A key aspect of this course is a development of an abstract idea of space which extends our experience in low dimensions to higher dimensional situations. Our basic experience of space is with a three dimensional space. We can easily go to a lesser number of dimensions by neglecting some directions in the three dimensional space, placing cutlery on a table involves the two directions in the plane of the table. Riding a train effectively uses one dimension. However, we are secretly aware of 4 dimensions since the position of any event is described by 3 spatial directions and one time direction. So, when one considers how objects in our 3 dimensional space evolve in time we are really thinking of a four dimensional setting.[(1)]

A simpler example of higher dimensional spaces are the *spaces of data*. What I mean by data is a sequence of numbers describing various aspects of some object. For instance any strategy that involves choosing several parameters, say n parameters, requires choosing a point in the n-dimensional space $\mathbb{R}^n$.

Example. If baking a cake involves deciding on the amount of seven ingredients as well as temperature and how long we bake, then our baking strategy consists of 9 numbers, so we can think of it as a point in $\mathbb{R}^9$. So, a choice of how to bake can be thought of as a choice of point in $\mathbb{R}^9$. However, we are not likely to use this geometric point of view since in this business we have much experience (personal or collective, such as a recipe). For more complicated problems geometry may be essential.

0.4. **Differentiation for several variables.** Here we review differentiation first for functions of one variable and then for functions of an arbitrary number of variables.

For functions of one variable differentiation

[1] The contemporary physics uses more dimensions. The best description of physics that we have at the moment is *String theory*, it claims that our space is really 11 dimensional, beyond the 4 familiar directions the remaining ones are more difficult to observe but are seen as necessary in order to explain why the 4 familiar dimensions behave as they do.

The rate of change of a function $f(x)$ of one variable at $x = a$ is a number called the *derivative* $f'(a)$ of $f(x)$ at a. However, for a function $f(x_1, ..., x_n)$ of several variables the rate of change at a point $a = (a_1, ..., a_n)$ is a function called the *differential* $df(a)$ of the function f at a. This function records the all rates change of f when one moves away from the point a in various directions and with different speeds.

0.4.1. *The rate of change* $(D_v f)(a)$ *when one is moving with a constant velocity vector* v. Because we are in a higher dimensional setting one can move from a given point $a = (a_1, ..., a_n)$ in $\mathbb{R}^n$ in many directions and in each of these directions one can move with various speeds. The data of a direction and speed is called *velocity* and it is represented by a vector v in $\mathbb{R}^n$. When one is moving from the point $a = (a_1, ..., a_n)$ at the constant velocity $v = \langle v_1, ..., v_n \rangle$, the position at time t is described by

$$x(t) = \ a + tv = \ (a_1, ..., a_n) + t\langle v_1, ..., v_n \rangle = \ (a_1 + tv_1, ..., a_n + tv_n).$$

As one is moving on the line $x(t) = a + tv$ in $\mathbb{R}^n$, the value of the function $f(x(t)) = f(a + tv)$ changes. The change of function from time $t = 0$ zero to time $t = h$ is $f(a+hv) - f(a) = f(a_1 + hv_1, ..., a_n + hv_n) - f(a_1, ..., a_n)$. The *average rate of change* in this period is $\frac{1}{h}[f(a_1 + hv_1, ..., a_n + hv_n) - f(a_1, ..., a_n)]$.

Example. Imagine you are traveling in a train and as you look through the window you see mountains. The height h of the piece of the mountain which is straight in front of you when you are at a position $x = (x_1, x_2, x_3)$ depends on this position. So, it is a function $h(x)$ of x. The rate at which the height in front of you changes depends on how fast you are traveling, and also on the direction in which the train is traveling. □

We are really interested in the *instantaneous rate of change* at time $t = 0$, i.e., at the moment when one is at the position $a = (a_1, ..., a_n)$ and one is moving at constant velocity v on the path $x(t) = a + tv$. We denote this rate of change $(D_v f)(a)$. (Here "D" could stand for "directional", reminding us that at the moment we are just considering how f changes in a given direction.)

This turns out to be a more abstract idea and we will use the average rate of change $\frac{1}{h}[f(a + hv) - f(a)]$ to understand it. When the time period h is small then the average rate of change $\frac{1}{h}[f(a + hv) - f(a)]$ in this period can be used as an approximation of the instantaneous rate of change at the beginning of the period.

This approximation by average rate of change gets better as the time period h gets smaller. So, we can think of the instantaneous rate of change as the number approached by these approximations as h goes to 0, This gives us a formula for the instantaneous rate of change

$$(D_v f)(a) \ \stackrel{\text{def}}{=} \ \lim_{h \to 0} \ \frac{f(a + hv) - f(a)}{h} \ = \ \lim_{h \to 0} \ \frac{f(a_1 + hv_1, ..., a_n + hv_n) - f(a_1, ..., a_n)}{h}.$$

Remark. This formula is complicated because it involves a limit. However, this complexity is not avoidable since this formula is the *only* way we can even make sense of the instantaneous rate of change.

Example. If we only have one variable x then the velocity vector v is just a number. If we choose v to be 1 (we are moving to the right on the x-axis with the unit speed. In this case then $(D_{v=1}f)(a) \stackrel{\text{def}}{=} \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$ is just the derivative $f'(a)$.

If in this 1-dimensional situation we move with the velocity which is a number v then the rate of change is v times larger: $(D_v f)(a) = vf'(a)$ (since one is now differentiating $f(a+vt)$ instead of $f(v+t)$).

0.4.2. *The differential $d_a f$.* Since in higher dimension the rate of change of a function depends on the velocity vector v we have a large amount of information (the rates of change for each velocity vector v). We will package this information in two ways.

The first is completely formal – we just think of all this information as a single function called the *differential* of the function f at the point a and denoted $d_a f$ or $df(a)$. The second is practical – it turns out that in order to know all directional derivatives it suffices to know the n partial derivatives of f at a, and these numbers we put together into a vector called the *gradient vector* $(\nabla f)(a)$ (see 0.4.3 below).

We replace the usual derivative $f'(a)$ which is one number $f'(a)$, by a *function* called the *differential* $df(a)$ of the functions $f(x_1, ..., x_n)$ at $a_1, ..., a_n)$. This functions sends vectors to numbers, the value at a vector v is the rate of change $(D_v f)(a)$ of the function $f(x_1, ..., x_n)$ when is moving from the point $(a_1, ..., a_n)$ at velocity v:

$$[df(a)](v) \;=\; (D_v f)(a) \;\stackrel{\text{def}}{=}\; \lim_{h\to 0} \frac{f(a_1 + hv_1, ..., a_n + hv_n) - f(a_1, ..., a_n)}{h}.$$

So, $df(a)$ is a function on the space $\mathbb{R}^n$ of n-dimensional vectors, with values in real numbers.

0.4.3. *Gradient $\nabla_a f$.* The standard vectors $e_1, ..., e_n$ in $\mathbb{R}^n$ are $e_1 = (1, 0, ..., 0)$, $e_2 = (0, 1, 0, ..., 0)$, $e_3 = (0.0, 1, 0, ..., 0)$, ... $e_n = (0, 0, ..., 0, 1)$. So the i^{th} standard vector e_i is in the direction of the i^{th} coordinate axis – it has all entries 0 except that the i^{th} entry is 1.

The rate of change of a function $f(x)$ in the direction of the i^{th} coordinate axis is

$$D_{e_i} f(a) \;\stackrel{\text{def}}{=}\; \lim_{h\to 0} \frac{f(a + he_i) - f(a)}{h} \;=\; \lim_{h\to 0} \frac{f(a_1, ..., a_{i-1}, a_i + h, a_{i+1}, ..., a_n) - f(a_1, ..., a_n)}{h}.$$

This is a familiar object which we call the *partial derivative* of f with respect to the i^{th} variable x_i, denoted $\frac{\partial f}{\partial x_i}(a)$.

The *gradient vector* for the function $f(x)$ at the point $a = (a_1, ..., a_n)$ is the vector consisting of all partial derivatives:

$$\nabla_a(f) \;=\; \nabla f(a) \;\stackrel{\text{def}}{=}\; \langle \frac{\partial f}{\partial x_1}(a), ..., \frac{\partial f}{\partial x_n}(a) \rangle.$$

Lemma. The rate of change of f at a and with velocity vector v is the dot product of the gradient vector with v :

$$D_v f(a) \;=\; \nabla f(a) \cdot v.$$

□

Remark. So, the gradient is a very efficient way to capture all information of about rates of change of a function.

0.5. **The abstract idea of integration.** Here we consider what it means to integrate a function f over some space X.

At this point we use an imprecise phrase "space", meaning roughly a "geometric object" such as curve, surface or other objects that one can find in $\mathbb{R}^3$, $\mathbb{R}^4$, ... Because the idea of integration that we describe here is used in a variety of settings (some vastly more complicated than what one finds in any $\mathbb{R}^n$), we will leave it a little vague what we mean by a "space X". [This will be explained in the rest of the course.]

The goal is to calculate the *total amount* of a quantity Q which is spread over some space X in a way described by the density function $f(x)$ on X. Let us denote this total quantity $Q(X)$. We also denote it $\int_X f$ and call it the *integral of f over X*.

0.5.1. *The strategy for computing $Q(X)$.* It is obvious – the total amount $Q(X)$ is a sum of contributions $Q(P)$ from various pieces P of X. (We call contributions from *small* pieces the *local* contributions.) Here is how to make this strategy work:

Steps:

- S0. We need to understand the *size* of small pieces P of the space X. We denote this size by $Vol(P_i)$.
- S1. We approximate a (possibly complicated) space X by covering as much of it as we can by a number of small pieces $P_1, ..., P_N$ which are easier to understand. This gives an approximation of the total amount of the quantity Q in X by the sum
$$Q(X) \sim Q(P_1) + Q(P_2) + \cdots + Q(P_N)$$
of amounts in all pieces. (A short hand is the above sum is $\sum_{i=1}^{N} Q(P_i)$.)
- S2. We approximate the density function $f(x)$ on the i^{th} piece P_i by a constant function (i.e., a number). This number we choose to be $f(c_i)$ at some point c_i that we choose in the piece P_i.

- S3. We approximate the contribution from the i^{th} piece $Q(P_i)$ by $Vol(P_i)\cdot f(c_i)$. [This would be correct if f we constant on P_i.]
- S4. We approximate the total amount $Q(X)$ of the quantity Q in X by the sum of approximations in all pieces

$$Q(X) \sim \sum_1^N Q(P_i) \sim \sum_1^N Vol(P_i)\cdot f(c_i).$$

- S5. We find the *precise value* of the total $Q(X)$. This is the number approached by approximations as they get better and better

$$Q(X) \;=\; \lim \sum_1^N Vol(P_i)\cdot f(c_i)$$

0.5.2. *How to make approximations better?* To strategy to is to make the pieces P_i *smaller*. This helps in two ways:

- Using smaller pieces one can can cover more of X (this diminishes the error we made in the step S1 above).
- Using smaller pieces P_i diminishes the error we made in the step S2 above, i.e., the approximation of $Q(P_i)$ by $Vol(P_i)\cdot f(c_i)$ gets better. The reason is that when the piece P_i gets smaller then the function f varies less in P_i so it is more alike a constant function.

So, $Q(X)$ is the limit of approximations $\sum_1^N Vol(P_i)\cdot f(c_i)$ as we refine the subdivision of X into pieces $P_1, ..., P_N$. The refinements are made by making the pieces smaller and smaller. (This of course means that we will need more and more of these pieces to cover X.) This is our formula

$$Q(X) \;=\; \lim_{\substack{\text{as one refines the subdivision} \\ \text{by making pieces smaller}}} \sum_{i=1}^N Vol(P_i)\cdot f(c_i).$$

0.5.3. *The integration notation.* We will now forget that the meaning of the function f above was the density of a quantity Q spread over X. The above formula for the total of Q is then viewed as an operation on functions on X called *integration*.

We define the *integral of a function f over a space X* by

$$\int_X f \;\stackrel{\text{def}}{=}\; \lim_{\substack{\text{as one refines the subdivision} \\ \text{of } X \text{ by making pieces smaller}}} \sum_{i=1}^N Vol(P_i)\cdot f(c_i).$$

A more detailed notation is $\int_X f\, dV$ or $\int_X f(x)\, dV(x)$.

Here, the integral sign $\int$ is in the shape of letter S to remind us that this is a limit of approximations which are sums. The expression $f(x)\ dV$ reminds us that the terms in the sum are products $f(c_i)Vol(P_I)$ for small pieces P_i of X and a chosen point c_i in P_i. In particular "dV" stands for the volume of a small piece of the space X.

Remark. Again, the meaning of this complicated formula is that

- the integral is the totality of something;
- it is computed by gathering all *local* contributions, i.e., contributions from small pieces of the space X.[(2)]

0.5.4. *Practical computation of integrals.* The above definition of of an integral is a beauty on the philosophical level and it has been immensely useful for *applying* integration in many settings. However, it is clearly complicated because it:

- (i) even a single approximation can be difficult to calculate if it involves many pieces;
- (ii) the definition involves infinitely many calculations of approximations that become more and more accurate.

In dimension one we have two tricks for computing integrals: the *Fundamental Theorem of Calculus* and *Change of variable formula*. These course just deals with extending these two methods to higher dimension.

Remarks. Here are some more thoughts on difficulty of integrating by the above formula.

(0) One way you see the complexity is that when we are computing an approximation we are allowed much choice both in: (i) choosing small pieces that "almost" fill the space X, and in (ii) choosing a point c in each such piece.

(1) The meaning of the above definition of the integral as the above limit is that – however we make these choices – the approximations will always approach the same number called $\int_X f$. So, the value of the integral is independent of all possible choices that we make.

(2) Not all sequences of numbers have a limit. Therefore, not all functions have integrals. If for a given function f on X the integral exists (i.e., the sequence of approximations converges) we say that the function f is *integrable* on X.

0.6. **Integration of functions of several variables.** Now we will become more concrete and explain what are the spaces X over which we will integrate. The first (and simpler) class of such spaces are called *regions* in spaces $\mathbb{R}^n$.

[2] Here, "local contributions" are all $Vol(P_i)\cdot f(c_i)$ for $1 \le i \le N$, and "gathering" means that we add them up.

Understanding integration over regions is a basis for calculation of integrals over a more general class of spaces called the *submanifolds* of $\mathbb{R}^n$. (I will sometimes use an ad hoc informal terminology and call submanifolds the "shapes" in $\mathbb{R}^n$s.)

0.6.1. *Regions in* $\mathbb{R}^n$. A region R in $\mathbb{R}^n$ means a subset which is described by inequalities.

Integration over regions is covered in chapter 5 of the book. Here are some examples of regions.

Example. Boxes $B = B\left(\begin{smallmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{smallmatrix}\right)$. If one restricts each of the coordinates $x_1, ..., x_n$ to some interval, say $a_1 \leq x_1 \leq b_1$, $a_2 \leq x_2 \leq b_2$, ..., $a_n \leq x_n \leq b_n$; one obtains an n-dimensional box which we may denote $B = B\left(\begin{smallmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{smallmatrix}\right)$ since it depends on the numbers $a_1, ..., a_n, b_1, ..., b_n$, We define the n-dimensional volume Vol^n so that the the volume of the n-dimensional box $B = B\left(\begin{smallmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{smallmatrix}\right)$ is the product of its sides $(b_1 - a_1)(b_2 - a_2) \cdot \cdot \cdot (b_n - a_n)$.

Example. Balls Another example would be an *open ball* $B_r(a)$ in $\mathbb{R}^n$ with center $a = (a_1, ..., a_n) \in \mathbb{R}^n$ and radius r. It is given by all points $x = (x_1, ..., x_n) \in \mathbb{R}^n$ with $\sum_1^n (x_i - a_i)^2 < r^2$.

The version where the inequalities are not strict is called the *closed ball* $\overline{B}_r(a)$. It is given by all points $x \in \mathbb{R}^n$ with $\sum_1^n (x_i - a_i)^2 \leq r^2$. (The difference is that the closed ball also contains the *sphere* $S_r(a)$ in $\mathbb{R}^n$ with center $a = (a_1, ..., a_n) \in \mathbb{R}^n$. It consists all points $x \in \mathbb{R}^n$ with $\sum_1^n (x_i - a_i)^2 = r^2$.

Subexamples. In the case when the dimension is one, the open ball $B_r(a)$ is the open interval $(a - r, a + r)$ around a real number a. The closed ball $\overline{B}_r(a)$ is the closed interval $[a - r, a + r]$.

In the plane $\mathbb{R}^2$, an open ball $B_r(a)$ is a disc with its boundary circle removed. It becomes a closed ball when we do include the boundary circle. In $\mathbb{R}^3$ ...

0.6.2. *Submanifolds in* $\mathbb{R}^n$. These are geometric objects that lie in $\mathbb{R}^n$ that are defined by equations or a combination of equations and inequalities. We will learn to integrate over submanifolds in chapter **??**.

Example. 1. The sphere $S_r(a)$ in $\mathbb{R}^n$ with the center $a = (a_1, ..., a_n)$ and radius r is given by a single equation $\sum_1^n (x_i - a_i)^2 = r^2$.

Example. 2. In $\mathbb{R}^2$ one uses a single equation $F(x, y) = 0$ to describe a curve in $\mathbb{R}^2$. For instance if $F(x, y) = y - f(x)$ the meaning of equation $F(x, y) = 0$ is the $y = f(x)$, so the curve is the graph of the function $f(x)$.

Example. 3. In $\mathbb{R}^3$ a single equation $F(x, y, z) = 0$ describes a surface in $\mathbb{R}^3$. For instance if $F(x, y) = z - f(x, y)$ the surface is the graph of the function $f(x, y)$.

However, if in $\mathbb{R}^3$ one uses two equations one obtains a curve. For instance equation $x^2 + y^2 + z^2 = 25$ gives a sphere and the equation $x = 4$ gives a plane. The two equations together give a *circle* in $\mathbb{R}^3$ – it has center at $(4, 0, 0)$, it lies is the plane $x = 4$ and its radius is 3. [You should convince yourself of this!]

0.6.3. *The dimension of a region or a submanifold.* The dimension $\dim(X)$ of any "space" X is roughly the number of *independent directions* in X. We will use this vague definition without explaining when two directions are deemed independent. (The precise version uses the notion of *independent vectors* from linear algebra.)

Example. $\dim(\mathbb{R}^n)$ is n because the coordinate axis provide n independent directions in $\mathbb{R}^n$ one can move that one can move in.

Equations. The effect of imposing an equation is that it usually *decreases* the dimension by 1.

This principle is illustrated in examples 1-3 in 0.6.2 above,

Inequalities. Imposing inequalities usually does not change the dimension.

Example. Let us consider this intuitively in an example. For instance a closed ball in $\mathbb{R}^2$ is a disc given by inequality $(x_1 - a_1)^2 + (x_2 - a_2)^2 \leq r^2$. We say that it has two dimensions because from most points in this closed disc we can move within the disc in two "independent directions". However, if instead we use the corresponding equation $(x_1 - a_1)^2 + (x_2 - a_2)^2 = r^2$ we get a circle which has only one dimension since we can only move on the circle is in one "independent direction" (the opposite of a given direction is not viewed as "independent"),

Example. A region in $\mathbb{R}^n$ still has dimension n as inequalities do not diminish the number of directions.

0.6.4. *Tricks for calculation of integrals in higher dimension.* Integration in higher dimension is more difficult because geometry is more complicated than in dimension one. The heart of the course deals with the following three methods.

(1) **The "projection method", i.e., "iterated integrals".** This approach attempts to reduce the calculation to objects of lower dimension (meaning to functions of fewer variables).

The idea is that for regions R that of a special form (roughly, regions between graphs of two functions), the integral over R can be calculated as an *iterated integral*. By repeating this idea one ends with *repeated* integrate over intervals, i.e., a higher dimensional integration is reduced to (repeated!) one dimensional integration.

This is explained in the chapter 5 of the book.

(2) **Change of variables.** In dimension one this is the second most useful tool for integration. However, in higher dimension it will be of fundamental importance.

In dimension one, i.e., for functions of one variable all regions are intervals. So. the main problem is to choose the change of variables that simplifies the integral. Only then does one pay attention to how the change variable has transformed the original interval of integration to a new one.

In $\mathbb{R}^n$ some regions are more complicated than others. A change of variables transforms integration region R to another region R' which may have a drastically different shape. So, the first concern may be that the new region is simpler than the original one.

The change of variables relates R' and R by providing a correspondence of points of R' a and R. Such correspondence is called a *mapping*. The key aspect of a mapping is its "*stretching factor*" which measures how the mapping affects ("stretches") the volume of a small box. We will see that this stretching factor can be computed as the *Jacobian*) of the mapping.

This is explained in chapter 6 of the book. The Jacobians are best understood in terms of the notion of *differential forms* (Section 8.5 in the book).

(3) **The Fundamental Theorem of calculus.** In dimension one this is basic method for calculation of integrals. It says that the integral of a derivative of f on interval $[a, b]$ is the change of f on the boundary points of the interval:

$$\int_a^b f'(x)\, dx \;=\; f(b) - f(a).$$

When we integrate over a region R in $\mathbb{R}^n$, the same principle continuous to hold. However, now the boundary ∂R of the region R is a nontrivial geometric shape (for instance the boundary of a ball is a sphere). Moreover, one has to consider what is the useful notion of the derivative over R in this situation.

This material is treated in chapter 8.

5. Integration over regions in $\mathbb{R}^n$ (Chapter 5.)

5.1. **Integrals over regions.** In 0.6 we have explained a general idea of of integration over abstract "spaces". Here we consider how this idea of integration works in the particular case when the space we are integrating over is a region R in some $\mathbb{R}^n$.

5.1.1. In 0.6.1 we have introduced the notion of a *region* R in some $\mathbb{R}^n$ as a subset which is described by inequalities.

The integral $\int_R f$ of a function f over a region R was defined as a limit of its approximations of the form

$$\sum_1^N \ Vol(P_i)\ f(c_i)$$

where

- (i) one approximates the region R by small pieces $P_1, ..., P_N$ whose size $Vol(P_i)$ can be computed (or approximated);
- (ii) on each piece P_i one approximates the function f by a constant function – the value $f(c_i)$ of the function f at some chosen point c_i in P_i.

What is special about regions $R \subseteq \mathbb{R}^n$ is that one has convenient ways to approximate a region by small pieces, one chooses the pieces to be *boxes* (see) or more generally *paralellopipeds* which just means *slanted boxes* (see 5.1.3).

5.1.2. *Boxes.* These were our first examples of regions (see 0.6.1). A choice of numbers $a_1, ..., a_n, b_1, ..., b_n$ defines the box $B = B\left(\begin{smallmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{smallmatrix}\right)$ which consists of all points $x = (x_1, ..., x_n) \in \mathbb{R}^n$ such that for $i = 1, ..., n$ the i^{th} coordinate is restricted to the interval $[a_i, b_i]$, i.e., $a_i \leq x_i \leq b_i$.

We have also defined the *n-dimensional volume* $Vol^n = Vol$ of such box as the product of lengths of its sides

$$Vol^n[B\left(\begin{smallmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{smallmatrix}\right)] \ \stackrel{\text{def}}{=} \ (b_1 - a_1)(b_2 - a_2)\cdot\cdot\cdot(b_n - a_n) \ = \ \prod_{i=1}^n \ b_i - a_i.$$

Example. In low dimensions these boxes are familiar shapes.

- A 1-dimensional box $B\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$ is just the interval $[a, b]$ and the 1-dimensional volume is length.
- Two dimensional boxes $B\left(\begin{smallmatrix} a_1 & a_2 \\ b_1 & b_2 \end{smallmatrix}\right)$ are rectangles and the 2-dimensional volume is area.
- The three dimensional boxes $B\left(\begin{smallmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{smallmatrix}\right)$ are the usual boxes and the 3-dimensional volume is the usual volume of a box.
- Beyond that we can not draw the higher dimensional boxes but we do have some intuition based on the lower dimensional cases.

We see that the notion of the n-dimensional volume of an n-dimensional box (meaning the *size* of a box) is the obvious generalization of the usual ways of measuring size of intervals, rectangles and ordinary boxes.

5.1.3. *Paralellopipeds, i.e., slanted boxes.*

Remark. These will be used in the chapter 6 on Change of Variables.

For an n-tuple of vectors $v = (v_1, ..., v_n)$ with $v_i \in \mathbb{R}^n$, we define the *parallelopiped* $P(v_1, ..., v_n)$ as the set of all linear combinations $c_1v_1 + \cdots + c_nv_n$ of of vectors v_i with coefficient numbers c_i between 0 and 1 :

$$P(v_1, ..., v_n) \stackrel{\text{def}}{=} \{c_1v_1 + \cdots + c_nv_n;\ 0 \leq\ c_1, ..., c_n \leq 1\}.$$

Example. Recall the *basic* or *standard* vectors $e_1, ..., e_n$ in $\mathbb{R}^n$ defined by $e_1 = (1, 0, ..., 0)$, $e_2 = (0, 1, ..., 0)$ and so on until $e_n = (0, ..., 0, 1)$. Notice that if we multiply each e_i by a number $b_i \geq 0$, then the parallelopiped $Pt(b_1e_1, ..., b_ne_n)$ is the box $B\left(\begin{smallmatrix} 0 & \cdots & 0 \\ b_1 & \cdots & b_n \end{smallmatrix}\right)$ consisting of all $x \in \mathbb{R}^n$ with $0 \leq x_i \leq b_i$. □

Lemma. The volume of the parallelopiped $P(v_1, .., v_n)$ is

$$|\det(v_1, .., v_n)|$$

the absolute value of the determinant $\det(v_1, .., v_n)$ of the nxn matrix $(v_1, ..., v_n)$ with columns v_i.

Proof. Since this is a statement in linear algebra we will not check this statement in general, i.e., for all choices of vectors $v_1, ..., v_n$. We will only consider two classes of examples:

(**A.**) In dimension 3 this determinant formula for the volume of a slanted box is a standard topic in MATH233. This also implies the cases of dimensions that are ≤ 3.[(3)]

(**B.**) An example in any dimension is the case when the vectors v_i are multiples of vectors e_i. Then $P(b_1e_1, ..., b_ne_n)$ is the box $B\left(\begin{smallmatrix} 0 & \cdots & 0 \\ b_1 & \cdots & b_n \end{smallmatrix}\right)$. So we know that the volume is $(b_1 - 0)\cdot\cdot(b_n - 0) = b_1\cdot\cdot b_n$. This is the same as the determinant

$$\det(b_1e_1, ..., b_ne_n) = \det\begin{pmatrix} b_1 & \cdots & 0 & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \cdots & & \vdots \\ 0 & \cdots & 0 & b_n \end{pmatrix} = b_1\cdot\cdot b_n. \quad \square$$

[3] For $n = 1$ the formula is obvious. Say, in dimension $n = 2$ $Vol^2([P(v_1, v_2)])$ is the area of the parallelogram $P(v_1, v_2)]$ with sides given by vectors v_1, v_2. This is the same as the volume of the parallelopiped $P(v_1, v_2, e_3)$ in dimension 3 since its base is the parallelogram $P(v_1, v_2)]$ and the height is 1. So, it is given by $\det(v_1, v_2, e_3)$ which is (by expansion in the 3rd row or column) the same as $\det(v_1, v_2)$.

5.1.4. *Beyond boxes.* However, once we define integrals using boxes we will know in principle how to compute the volume of any region $\mathcal{B}$ since $Vol^n(\mathcal{B})$ will be defined as $\int_{\mathcal{B}} 1$, the integral of the constant function 1.

Then we will be able to calculate integrals $\int_R f$ by subdividing the space R into small pieces $\mathcal{B}_s$ of any shape convenient for a given situation (so $\mathcal{B}_s$ will not necessarily be "in the shape of a box"). we will use the same strategy as above and the formula for the corresponding approximation will be a sum of terms $Vol^n(\mathcal{B}_s)f(c_s)$ which will now makes sense.

The reason that initially we can use only boxes is that at the moment we only know how to define the volume of a box.

However, once we define integrals using boxes we will know in principle how to compute the volume of any region B since $Vol^n(B)$ will be defined as $\int_B 1$, the integral of the constant function 1.

Remark. For now one can ignore this more general subdivision strategy, it will be used later in chapter 6.

5.1.5. *Approximation of regions by boxes.* When R is a region in $\mathbb{R}^n$ there is a standard way to "almost fill up R by small boxes".

We say that a region R in $\mathbb{R}^n$ is *bounded* if it lies in some (sufficiently large) box $B = B\left(\begin{smallmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{smallmatrix}\right)$.

(1) For such region R we start by creating the subdivision of the box B on the "scale $1/N$" for any $N = 1, 2, 3, ...$ For this we divide each of the intervals $[a_i, b_i]$ (for $i = 1, ..., n$) that form the sides of the box into N subintervals. This is done by choosing points $a_i = \alpha^0 < \alpha^1 < \cdots < \alpha^n = b_i$. More precisely, since we have to make this choice for all i between 1 and n, we will add index i to the points α^s in order to remember that these are the division points in the i^{th} direction. So, we have $a_i = \alpha_i^0 < \alpha_i^1 < \cdots < \alpha_i^n = b_i$. The simplest way is to choose all the subintervals of the same length, then in the i^{th} direction this length will be $(b_i - a_i)/N$.

(2) These subdivisions of the intervals $I_i \stackrel{\text{def}}{=} [a_i, b_i]$ into smaller intervals $I_i^s \stackrel{\text{def}}{=} [\alpha_i^{s-1}, \alpha_i^s]$ for $s = 1, ..., N$, creates a subdivision of our box B with sides $I_i = [a_i, b_i]$ into smaller boxes whose $i\theta$ side will be one of the subintervals I_i^s.

Since we have to choose subintervals in each of n directions, the choice will be given by an n-tuple of integers $s = (s_1, ..., s_n)$ with each s_i between 1 and N. We can denote the corresponding "small box" by $B_s = B_{s_1,...,s_n}$, its sides are the subintervals $I_1^{s_1}, I_2^{s_2}, ..., I_n^{s_n}$.

(3) Now the "small boxes" that we are going to use in the N^{th} approximation of the integral will be some of the boxes $B_{s_1,...,s_n}$ with $1 \leq s_1, .., s_n \leq N$ – the ones that lie in the region R.

5.1.6. *Conclusion.* We have now a workable strategy of calculating the integral $\int_R f$ over a region R.

We choose a box B that contains R. to get an N^{th} approximation of the integral we subdivide B into smaller boxes B_s on "scale" $1/N$, and we throw away the ones that do not lie in our region R.

Next, we choose a point c_s in each of the remaining boxes B_s and we get an N^{th} approximation which is the sum of $f(c_s)Vol^n(B_s)$ over the remaining boxes. Finally, the integral is the limit of N^{th} approximations as $N \to \infty$.

Notation. Since the volume dV of a small box $B \begin{pmatrix} x_1' & \cdots & x_n' \\ x_1'' & \cdots & x_n'' \end{pmatrix}$ is $(x_1'' - x_1'') \cdots (x_n'' - x_n'')$ we will denote it $dV = dx_1 \cdots dx_n$ where dx_i stands for a small change $x_i'' - x_i''$ of the i^{th} variable.

5.2. **Existence of integrals.** It is a priori not clear that the integral of a given function f over a region R in $\mathbb{R}^n$ exists , i.e., that the above limit exists.[4]

In this respect we are saved by the following observation (we will make it more precise and explain it later).

Theorem. If the region R in $\mathbb{R}^n$ is closed and bounded and the function f is continuous on R then the integral $\int_R f$ exists. □

Remarks. (0) We have defined "bounded" above. A region is said to be *closed* if it contains all its boundary points. For instance a closed ball is a closed region but an open ball s not. So, we see that a region is closed when its description by inequalities only uses the non-strict inequalities.

(1) This existence result for integrals is very strong because in particular it asserts that all possible choices of approximations converge to the same number $\int_R f$.

We will occasionally make use of this ability to tailor the choices of approximations as it suits us.

5.3. **Computation of higher dimensional integrals by iterated integrals ("projection method").** The idea is that often for a region R in some $\mathbb{R}^n$ one can choose one

[4] Remember that some sequences of numbers do not have a limit, i.e., they do not approach a single number, for instance $\lim_{n\to\infty} \frac{n+1}{n} = \lim_{n\to\infty} 1 + \frac{1}{n} = 1$ but $\lim_{n\to\infty} (-1)^n$ does not exist because the sequence alternates between ± 1 and does not approach a single number.

5.3.1. *Projections.* The projections that we are going to use are from $\mathbb{R}^n_{x_1,...,x_n}$ to $\mathbb{R}^{n-1}_{x_1,...,x_{n-1}}$. To a point $a \in \mathbb{R}^n_{x_1,...,x_n}$ we associate a point $pr(a) \in \mathbb{R}^{n-1}_{x_1,...,x_{n-1}}$ simply by erasing the last coordinate, i.e., $pr(a_1, ..., a_n) = (a_1, ..., a_{n-1})$.

The projection $pr(R)$ of a subset $R \subseteq \mathbb{R}^n_{x_1,...,x_n}$ is the image of R under the projection, i.e., the set of all points $pr(a)$ where $a \in R$.

Example. The following will be our running example. For the disc D given in $\mathbb{R}^2_{x,y}$ by radius r and center at (a, b) (i.e., by $(x-a)^2 + (y-b)^2 \leq r^2$), the projection $pr(D)$ of D to the x-axis, is the interval $[a-r, a+r]$ on the x-axis (i.e., by $(x-a)^2 \leq r^2$). □

The *fiber* R_b of R at a point $b = (b_1, ..., b_{n-1})$ in $\mathbb{R}^{n-1}_{x_1,...,x_{n-1}}$ is the set of all points $a \in \mathbb{R}^n_{x_1,...,x_n}$ that project to b. So, these are the points $a = (a_1, ..., a_{n-1}, a_n)$ in R such that $pr(a) = (a_1, ..., a_{n-1})$ equals $b = (b_1, ..., b_{n-1})$. Then $a = (b_1, ..., b_{n-1}, a_n)$.

So, the fiber R_b of R at b can be viewed as consisting of all numbers a_n such that $(b_1, ..., b_{n-1}, a_n)$ lies in R.

Example. Let us choose a point x_0 in the interval $[a-r, a+r]$. Then the fiber D_{x_0} of the disc D at x_0 consists of all numbers y_0 such that (x_0, y_0) lies in the disc D, i.e., $(x_0 - a)^2 + (y_0 - b)^2 \leq r^2$.

To understand the fiber D_{x_0} recall that the boundary of the disc D is the circle C given by $(x-a)^2 + (y-b)^2 = r^2$. We will first find y_0's such that $(x_0 - a)^2 + (y_0 - b)^2 = r^2$. For this we solve for y_0 and we get $y_0 = b \pm \sqrt{r^2 - (x_0 - a)^2}$.

Now we see that in order that $(x_0 - a)^2 + (y_0 - b)^2 \leq r^2$ we need

$$b - \sqrt{r^2 - (x_0 - a)^2} \leq y_0 \leq b + \sqrt{r^2 - (x_0 - a)^2}.$$

So, the fiber D_{x_0} of the disc D at the point x_0 on the x axis is the interval

$$D_{x_0} = [b - \sqrt{r^2 - (x_0 - a)^2}, b + \sqrt{r^2 - (x_0 - a)^2}].$$

5.3.2. *Projection method.* This method applies for a region R in some $\mathbb{R}^n$ if for each b in the image $pr(R)$ of R under the projection, the fiber R_b of R at b is an interval.

The end points of this interval depend on b so they are some functions $g_\pm(b)$ of b. Then the interval is $[g_-(b), g_+(b)]$.

This means that the region R lies above its projection $pr(R)$ and between the graphs of two functions $x_n = g_\pm(b_1, ..., b_{n-1})$.

Theorem. Suppose that the fibers of the projection of the region R are intervals $[g_-(b), g_+(b)]$ given by tw=o functions $g_\pm(b)$. In other words that the region R is the

part of $\mathbb{R}^n$ that lies above the projection $pr(R)$ and between the graphs of two functions $x_n = g_{\pm}(b_1, ..., b_{n-1})$. Then

$$\int_R f(x_1, ..., x_n)\, dV^n(x_1, ..., x_n) \;=\; \int_{pr(R)} [\int_{x_n=g_-(b_1,...,b_{n-1}}^{x_n=g_+(b_1,...,b_{n-1}} f(x_1, ..., x_n)\, dx_n]\; dV^{n-1}(x_1, ..., x_{n-1}).$$

Example. We finish the running example by wring an integral over a disc as an iteration of integrals over intervals:

$$\int_D f(x,y)\, dV \;=\; \int_{pr(D)} [\int_{y=g_-(y)}^{y=g_+(y)} f(x,y)\, dy]\; dV^1(x) \;=\; \int_{a-r}^{a+r} \int_{b-\sqrt{r^2-(x-a)^2}}^{y=b+\sqrt{r^2-(x-a)^2}} f(x,y)\, dy\; dx.$$

Remarks. (0) The effect of the above formula is that we have replaced integration over a region R in $\mathbb{R}^n$ by an integration over a region $pr(R)$ in $\mathbb{R}^{n-1}$. So, one has decreased the dimension by one at the price of calculating an additional ordinary integrals, i.e., an integral over an interval.

(1) Instead of always using the above projection along the last variable, for any $i \in \{1, ..., n\}$ the method can use the projection $pr_i(x_1, ..., x_n) =)x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$ along the i^{th} variable.

(2) One can try to repeat this method in order to reduce an integral over a region in $\mathbb{R}^n$ to smaller and smaller dimension until it becomes an n-fold iteration of ordinary integrals, i.e., integrals over intervals.

(3) Sometimes no projection pr_i works, i.e., the fibers are not intervals. Then one cuts the region R into nicer pieces for which the method works.

(4) It may also happen that one several projections works (say pr_2, pr_5, pr_7). Then one uses the projection that seems to give the easiest integral.

5.3.3. *Proof of the projection theorem 5.3.2.* The theorem states the equality

$$\int_R f(x_1, ..., x_n)\, dV^n(x_1, ..., x_n) \;=\; \int_{pr(R)} [\int_{x_n=g_-(x_1,...,x_{n-1})}^{x_n=g_+(x_1,...,x_{n-1})} f(x_1, ..., x_n)\, dx_n]\; dV^{n-1}(x_1, ..., x_{n-1}).$$

The LHS means that we are gathering local contributions over pieces of R to the total amount of some quantity Q spread over R with density f. The RHS means that we organize this gathering operation in the following way:

The first step is that at each point $(b_1, ..., x_{n-1})$ in the region $pr(R)$ we gather all contributions that in R and above the point $(x_1, ..., x_{n-1})$, i.e., that can be found in the fiber $R_{(x_1,...,x_{n-1})}$. This fiber is an interval $[g_-(x_1, ..., x_{n-1}), g_+(x_1, ..., x_{n-1})]$ so the effect of the gathering over this interval is the integral

$$\int_{x_n=g_-(x_1,...,n-1)}^{x_n=g_+(x_1,...,x_{n-1})} f(x_1, ..., x_n)\; dx_n$$

over this interval. In this integration variables $x_1, ..., x_{n-1}$ are constants.

In the second step we gather the contributions from all fibers $R_{(x_1,...,x_{n-1})}$, i.e., at all points $(x_1, ..., x_{n-1})$ in $pr(R)$. This gathering procedure is the integral over $pr(R)$. □

5.4. **Change of order of integration.**

6. **Change of Variables Formula (Chapter 6.)**

6.0.1. *Mappings.* A change of variables in an integral $\int_R f = \int \cdots \int_R f(x_1, ..., x_n)\, dx_1 \cdots dx_n$ over some region R in $\mathbb{R}^n$ means that (as in the 1-dimensional case) we will view coordinates x_i of points in R as functions $x_i = T_i(u) = T_i(u_1, ..., u_n)$ of coordinates $u = (u_1, ..., u_n)$ in some other region R' in $\mathbb{R}^n$.

The geometrically point of view on such change of variables is that it constitutes a *mapping* T from the region R' to the region R. (The functions T_i are the component functions of the mapping T.) Since mappings are a special class of the general notion of functions we start with an introduction to language of functions in 6.1.2 and then we consider mappings in 6.2.

6.0.2. *Mappings and 1-1 correspondences.* We want to equate integrals over two regions R and R' that are related by a mapping T from R' to R. For the integrals to be the same, the mapping T has to provide a very strong relation between the regions R' and R. What we need is essentially that anything done in R can be repeated in R' using T and vice versa. We can say that such relation makes "R and R' contain the same information" or that it makes "R and R' into two equivalent ways of viewing the same thing".

We call such mappings "*1-1 correspondences*" T from R' to R. They are introduced in 6.1.4 in the generality of arbitrary functions. Then we show later in 6.3.5 that integrals can be compared when the mapping T is a 1-1 correspondence and even when it is only "close" to a 1-1 correspondence (as long as the error appears on the set of volume zero).

6.0.3. *The stretching factor.* So far we have noticed that having a map T from R' to R which is (close to) a 1-1 correspondence allows us to relate integrals in R' and R. However, this comparison involves a "stretching factor" function which tells us how the mapping T distorts the volume.

The calculation of the stretching factor (see 6.5), involves an extension of the notion of the rate of change (i.e., differentiation) to mappings. The notion of the *differential of a mapping* is introduced (and calculated) in 6.3.

6A. Differentiation of mappings from $\mathbb{R}^m$ to $\mathbb{R}^n$

6.1. **Functions.**

6.1.1. *Sets.* A *set* is a mathematical word for a collection of things. The theory of sets is very simple, it systematically talks of what we can do with collections of objects.

Its basic advantage organizational and notational: the set theoretic notations allow us to replace some parts of sentences with a few simple symbols.

A deeper usefulness comes from the fact that all mathematical objects can be described in terms of sets – as systems of related sets. The point is that we are able to describe highly

sophisticated (and therefore somewhat abstract and confusing) mathematical objects precisely, using the language of sets. There is also a disadvantage – presenting mathematical objects in terms of sets may seem obscure to uninitiated.

6.1.2. *Functions.* A function f from a set A to a set B is a rule which assigns to each element $a \in A$ certain (unique) element of B, we denote this element $f(a)$. One can think of f as a "motion" that moves elements of A into B.

For psychological reasons, in other to distinguish between different kinds of functions, functions will also be called *maps* or *mappings* or *transforms*.

Example. In calculus we get familiar with with functions f from an interval $[a, b]$ to the set $\mathbb{R}$ of real numbers. They are usually described by a formula for the values $f(s)$ at points s in the interval $[a, b]$.

6.1.3. *Composition of functions.* For a configuration of functions $A \xrightarrow{f} B \xrightarrow{g} C$ such that the second function starts where the first one has arrived, one defines the *composition function* $A \xrightarrow{g \circ f} C$ from the first set to the last one, by sending any element $a \in A$ to $(g \circ f)(a) \stackrel{\text{def}}{=} g(f(a))$, So, one first applies to $a \in A$ the "closer" motion f to get an element $f(a)$ of B, and then one applies the "farther" motion g to the result $f(a)$, and this produces $g(f(a))$. (Here "closer" and "farther" refer to the symbol $(g \circ f)(a)$ – in this symbol f is closer to a and g is farther away.)

6.1.4. *1-1 correspondences.* A function $f : A \to B$ is said to be a *1-1 correspondence* (also called *bijection*) if it

gives a complete translation between elements of A and B so that now sets A and B contain "the same information", i.e., anything done in A can be redone in B and vice versa.

This amounts to requiring that for each $b \in B$ there exists precisely one $a \in A$ such that $f(a) = b$. then we say that that elements $a \in a$ and b in B correspond under the function f.

We will now analyze this concept in details.

A function $f : A \to B$ is said to be *1-1* (or an *injection*) if it distinguishes elements of A. This means that if for $a, b \in A$ we have $a \neq b$ then $f(a) \neq f(b)$. Equivalently,that $f(a) = f(b)$ implies that $a = b$. (The meaning of both formulations is that the only way the images $f(a)$ and $f(b)$ can be the same is if the original elements a, b are the same.)

A function $f : A \to B$ is said to be *onto* (or a *surjection*) if it hits every element of B, i.e., for each $b \in B$ there exists some $a \in A$ such that $f(a) = b$.

Lemma. f is a 1-1 correspondence iff it is onto and 1-1.

Proof. If f is onto then for each $b \in B$ there exists some $a \in A$ such that $f(a) = b$. If f 1-1 then such b is unique. □

6.2. **Functions in calculus (mappings).**

6.2.1. *From $\mathbb{R}^n$ to $\mathbb{R}$.* In multivariable calculus one studies functions f from a subset R of some $\mathbb{R}^n$ to the set $\mathbb{R}$ of real numbers.

Example. When we use coordinates $x_1, ..., x_n$ on $\mathbb{R}^n$ then each coordinate x_i is a function from $\mathbb{R}^n$ to $\mathbb{R}$. □

We will sometimes denote by $\mathbb{R}^n_{x_1,...,x_n}$ the space $\mathbb{R}^n$ with coordinates x_i ("a copy of $\mathbb{R}^n$ with coordinates x_i").

A point p in $\mathbb{R}^n$ is an n-tuple of numbers $(p_1, ..., p_n)$ (here, p_i is the value of the i^{th} coordinate function x_i on the point p). So, we usually describe the values $f(p) = f(p_1, ..., p_n)$ of the function f on a point $p \in \mathbb{R}^n$ by a *formula* $f(x_1, ..., x_n)$ written its of the coordinate functions $x_1, ..., x_n$. To get the value $f(p)$ at a point p we need to plug in p_i for x_i in the formula for $f(x_1, ..., x_n)$. For this reason we say that f is a function of n variables $x_1, ..., x_n$.

6.2.2. *From $\mathbb{R}^m$ to $\mathbb{R}^n$ (mappings between regions).* We will now start to use more complicated functions $T : S \to R$ where S is a subset of some $\mathbb{R}^m$ and R is a subset of some $\mathbb{R}^n$. To emphasize that we will think of such functions "geometrically" we will call them *mappings* (or just *maps*).

We write formulas for such mappings T in terms of coordinates $u_1, ..., u_m$ on $\mathbb{R}^m$ and $x_1, ..., x_n$ on $\mathbb{R}^n$.

The value $T(s)$ of the mapping T on the point $s \in S$ is a point in $\mathbb{R}^n$, so it is an n-tuple of numbers and these numbers are the coordinates $x_i(T(s))$ of the point $T(s)$. The way these coordinates depend on the point s constitutes a function from S to $\mathbb{R}$, this is the composition of functions $x_i \circ T$ (the value $(x_i \circ T)(s)$ of the composition $x_i \circ T$ is $x_i(T(s))$).

We will denote these functions by $T_i \stackrel{\text{def}}{=} x_i \circ T : S \to \mathbb{R}$. Since $T(s) = \big(x_1(T(s), ..., x_n(T(s))\big) = \big(T_1(s), ..., T_n(s)\big)$ we call the functions T_i the *component functions* of T. Therefore, we can think of the mapping $T : S \to \mathbb{R}^n$ as an n-tuple of its component functions T_i.

It will be convenient to write this n-tuple vertically as a column vector $T = \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}$.

Remark. Finally, we will recall that a point s in S lies in $\mathbb{R}^m$ so it is an n-tuple of numbers $s = (s_1, ..., s_m)$ with $s_i = u_i(s)$. Then we can write the mapping T in more details:

$$T(s) \;=\; \begin{pmatrix} T_1(s) \\ \vdots \\ T_n(s) \end{pmatrix} \;=\; \begin{pmatrix} T_1(s_1,...,s_n) \\ T_2(s_1,...,s_n) \\ \vdots \\ T_n(s_1,...,s_n) \end{pmatrix}.$$

When we want to emphasize that a point in S depends on m coordinate functions $u_1, ..., u_m$ we rewrite the above formula for T as $T(u_1, ..., u_m) \;=\; \begin{pmatrix} T_1(u_1,...,u_m) \\ T_2(u_1,...,u_m) \\ \vdots \\ T_n(u_1,...,u_m) \end{pmatrix}$.

Remark. One of shorthands in calculus is to denote the component functions $T_i = x_i \circ T$ just by x_i. This is of course confusing (since the symbol x_i is already used to denote a coordinate function on $\mathbb{R}^n$) but it is also convenient for writing simple formulas. Then the above formula for T is now written as $T(u_1, ..., u_m) \;=\; \begin{pmatrix} x_1(u_1,...,u_m) \\ x_2(u_1,...,u_m) \\ \vdots \\ x_n(u_1,...,u_m) \end{pmatrix}$.

Example. [*Polar coordinates.*] This is a standard example of a mapping T from $\mathbb{R}^2_{r,\theta}$ to $\mathbb{R}^2_{x,y}$. So, the variables x_1, x_2 in the target of T are now called x, y and the variables u_1, u_2 in the source are now denoted r, θ. The mapping $T(r, \theta)$ is given by the usual relation of Cartesian and polar coordinates

$$T(r,\theta) \;=\; \begin{pmatrix} x(r,\theta) \\ y(r,\theta) \end{pmatrix} \;=\; \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}. \quad \square$$

(The component functions of T are $T_1(r, \theta) = x(r, \theta) = r\cos(\theta)$ and $T_1(r, \theta) = y(r, \theta) = r\sin(\theta)$.)

Then the mapping T sends a point with *Cartesian coordinates* (r, θ) to the point $T(r, \theta)$ which has *polar coordinates* r and θ. $\square$

6.3. **The differential of a mapping T from $\mathbb{R}^m$ to $\mathbb{R}^n$.** The differential $dT(a)$ will be a way of recording the rates of change $(D_vT)(a)$ of T at the point a and with respect to all velocity vectors. We find that $(D_vT)(a)$ is calculated in terms of the matrix $(\nabla T)(a)$ which is given by all partial derivatives.

6.3.1. *The differential of a function f from $\mathbb{R}^m$ to $\mathbb{R}$ (recollections from 0.4).* The rate of change of a function $f(x)$ of one variable at $x = a$ is a number called the *derivative* $f'(a)$ of $f(x)$ at a.

We have already extended the idea of the rate of change (in 0.4) to functions $f : \mathbb{R}^n \to \mathbb{R}$ from $\mathbb{R}^n$ to the real numbers. The rate of change of such function at a point $a = (a_1, ..., a_n)$ is called the *differential of f at a* and is denoted $df(a)$ (or d_af). Because one can move from the point a in many directions – described by vectors $v \in \mathbb{R}^n$ – the differential is not just a number but a <u>new function</u> from $\mathbb{R}^n$ to $\mathbb{R}$. The value at a vector $v \in \mathbb{R}^n$ is the

rate of change of the function f when one is moving from the point a with the constant velocity vector v :

$$(d_a f)v \ \stackrel{\text{def}}{=}\ (D_v f)(a) \ \stackrel{\text{def}}{=}\ \lim_{h\to 0}\ \frac{f(a+hv)-f(a)}{h}.$$

This new function is in some sense very simple: it can be described by a single vector $\nabla_a f$ (called the *gradient vector for f at a*) using the dot product

$$(d_a f)v \ = \ \nabla_a f \cdot v.$$

The gradient vector is explicitly known – its components are the partial derivatives at a

$$\nabla_a f = \ \big(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), ..., \frac{\partial f}{\partial x_n}(a)\big).$$

6.3.2. *The rate of change $(D_v T)(a)$ at the point a when one is moving from the point with a constant velocity vector v.* For a mapping T from $\mathbb{R}^m$ to $\mathbb{R}^n$ the rate of change at a point $a \in \mathbb{R}^m$ and with respect to the velocity velocity vector v is defined in the same way

$$(D_v T)(a) \ \stackrel{\text{def}}{=}\ \lim_{h\to 0}\ \frac{T(a+hv)-T(a)}{h}.$$

Here, $T(a+hv)-T(a)$ is the change of values of T when one moves in time h from the point a to the point $a+hv$. Notice that since values of T are in $\mathbb{R}^n$ we are subtracting vectors in $\mathbb{R}^n$ and the change is again a vector in $\mathbb{R}^n$. Next, the vector $\frac{1}{h}(T(a+hv)-T(a))$ is the *average rate* of change of values of T from a to $a+hv$. Finally, the limit $h \to 0$ says that we make the time interval $[0,h]$ very small, so as h approaches 0, the average rate of change approaches the instantaneous rate of change at a.

Notice that in the end $(D_v T)(a)$ is a vector in $\mathbb{R}^n$. We will now find a simple formula for this rate of change vector.

In order to write a formula for the rate of change $(D_v T)(a)$ of the mapping T with respect to a velocity vector v, we recall how one writes the mapping T in terms of its component function T_i as $T = \begin{pmatrix} T_1(a) \\ T_2(a) \\ \vdots \\ T_n(a) \end{pmatrix}$.

Lemma. The rate of change of the mapping T is a vector whose components are the rates of change of the component functions T_i of T :

$$(D_v T)(a) \ = \ \begin{pmatrix} (D_v T_1)(a) \\ (D_v T_2)(a) \\ \vdots \\ (D_v T_n)(a) \end{pmatrix} \ = \ \begin{pmatrix} (\nabla T_1)(a) \cdot v \\ (\nabla T_2)(a) \cdot v \\ \vdots \\ (\nabla T_n)(a) \cdot v \end{pmatrix}.$$

Proof.

$$(D_vT)(a) = \lim_{h\to 0} \frac{T(a+hv)-T(a)}{h} = \lim_{h\to 0} \frac{1}{h}\Big(\begin{pmatrix} T_1(a+ht) \\ T_2(a+ht) \\ \vdots \\ T_n(a+ht) \end{pmatrix} - \begin{pmatrix} T_1(a) \\ T_2(a) \\ \vdots \\ T_n(a) \end{pmatrix} \Big)$$

$$= \lim_{h\to 0} \begin{pmatrix} \frac{1}{h}[T_1(a+ht)-T_1(a)] \\ \frac{1}{h}[T_2(a+ht)-T_2(a)] \\ \vdots \\ \frac{1}{h}[T_n(a+ht)-T_n(a)] \end{pmatrix} = \lim_{h\to 0} \begin{pmatrix} \frac{1}{h}[T_1(a+ht)-T_1(a)] \\ \frac{1}{h}[T_2(a+ht)-T_2(a)] \\ \vdots \\ \frac{1}{h}[T_n(a+ht)-T_n(a)] \end{pmatrix} = \begin{pmatrix} (D_vT_1)(a) \\ (D_vT_2)(a) \\ \vdots \\ (D_vT_n)(a) \end{pmatrix}.$$

□

6.3.3. *The differential d_aT of a mapping T.* Again, we package the information of the rates of change of T at a with respect to each velocity vector v into a single function d_aT called the differential of the mapping T at the point a. This functions is (again, as for the functions f from $\mathbb{R}^n$ to real numbers,) a function on vectors v in $\mathbb{R}^m$. The value of the differential at a vector $v \in \mathbb{R}^m$ is the rate of change $(D_vT)(a)$ of T when we are moving away from a at velocity v

$$[dT(a)](v) \;=\; (D_vT)(a) \;\stackrel{\text{def}}{=}\; \lim_{h\to 0} \frac{T(a_1+hv_1,...,a_n+hv_n)-T(a_1,...,a_n)}{h}.$$

6.3.4. *The matrix $\nabla T(a)$ of the differential d_aT.* We define the nxm matrix $\nabla T(a)$ (i.e., with n rows and m columns), so that the entry in the i^{th} row and the j^{th} column is the partial derivative $\frac{\partial T_i}{\partial u_j}(a)$ at the point a of the i^{th} component function T_i with respect to the j^{th} coordinate u_j, i.e.,

$$\nabla T(a) \;\stackrel{\text{def}}{=}\; \begin{pmatrix} \frac{\partial T_1}{\partial u_1}(a) & \frac{\partial T_1}{\partial u_2}(a) & \cdots & \frac{\partial T_1}{\partial u_m}(a) \\ \frac{\partial T_2}{\partial u_1}(a) & \frac{\partial T_2}{\partial u_2}(a) & \cdots & \frac{\partial T_2}{\partial u_m}(a) \\ \vdots & & & \\ \frac{\partial T_1}{\partial u_1}(a) & \frac{\partial T_2}{\partial u_2}(a) & \cdots & \frac{\partial T_n}{\partial u_m}(a) \end{pmatrix}.$$

Notice that the i^{th} row is the gradient $\nabla T_i(a) = \begin{pmatrix} \frac{\partial T_i}{\partial u_1}(a) & \frac{\partial T_i}{\partial u_2}(a) & \cdots & \frac{\partial T_i}{\partial u_m}(a) \end{pmatrix}$ of the i^{th} coordinate function T_i. So we can symbolically write this as a column vector whose entries are gradient vectors:

$$\nabla T(a) \;=\; \begin{pmatrix} \nabla T_1(a) \\ \nabla T_2(a) \\ \vdots \\ \nabla T_n(a) \end{pmatrix}.$$

Lemma. $\nabla T(a)$ is the matrix corresponding to the operator $dT(a)$ in the sense that for any vector v (viewed as a column vector $v = \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$) in $\mathbb{R}^m$, the value of the differential $dT(a)$ on v is given by the product of a matrix and a vector

$$[(dT)(a)v \;=\; (\nabla T)(a)\; v.$$

Proof. When one recalls how matrices multiply with vectors, then the i^{th} entry of the product $(\nabla T)(a)\; v$ of a matrix $(\nabla T)(a)$ and a vector v, is just the dot product of the i^{th} row $(\nabla T_i)(a)$ of the matrix $\nabla T(a)$ with the vector v. This gives a formula for the product

$$(\nabla T)(a)\; v \;=\; \begin{pmatrix} (\nabla T_1)(a) \cdot v \\ (\nabla T_2)(a) \cdot v \\ \vdots \\ (\nabla T_n)(a) \cdot v \end{pmatrix}.$$

However, this is exactly the above formula for the rate of change $D_v T(a)$ (which is the same as $(D_v T)(a)$). □

6.3.5. *The Jacobian of a mapping.* When for a mapping T from $\mathbb{R}^m$ to $\mathbb{R}^n$ the dimensions are the same, i.e., $m = n$, then $(\nabla T)(a)$ is a square matrix and we can take its determinant. We call it the *Jacobian* of the mapping T at a and we denote it

$$\frac{\partial(x_1, ..., x_n)}{\partial(u_1, ..., u_n)} \;\stackrel{\text{def}}{=}\; \det[(\nabla T)(a)] \;=\; \det \begin{pmatrix} \frac{\partial T_1}{\partial u_1}(a) & \frac{\partial T_1}{\partial u_2}(a) & \cdots & \frac{\partial T_1}{\partial u_m}(a) \\ \frac{\partial T_2}{\partial u_1}(a) & \frac{\partial T_2}{\partial u_2}(a) & \cdots & \frac{\partial T_2}{\partial u_m}(a) \\ \vdots & & & \\ \frac{\partial T_1}{\partial u_1}(a) & \frac{\partial T_2}{\partial u_2}(a) & \cdots & \frac{\partial T_n}{\partial u_m}(a) \end{pmatrix}.$$

Remark. Recall that one sometimes simplifies the notation for the component functions T_i of$T$ to just x_i. Then the Jacobian is

$$\det \begin{pmatrix} \frac{\partial x_1}{\partial u_1}(a) & \frac{\partial x_1}{\partial u_2}(a) & \cdots & \frac{\partial x_1}{\partial u_m}(a) \\ \frac{\partial x_2}{\partial u_1}(a) & \frac{\partial x_2}{\partial u_2}(a) & \cdots & \frac{\partial x_2}{\partial u_m}(a) \\ \vdots & & & \\ \frac{\partial x_1}{\partial u_1}(a) & \frac{\partial x_2}{\partial u_2}(a) & \cdots & \frac{\partial x_n}{\partial u_m}(a) \end{pmatrix}$$

and this is the reason why it is denoted $\frac{\partial(x_1,...,x_n)}{\partial(u_1,...,u_n)}$.

6B. Change of variables: 1-1 correspondences of regions

6.4. **Change of variables as a mapping between regions.** We consider an integral $\int_R f \;=\; \int_R f(x_1, ..., x_n)\; dx_1 \cdots dx_n$ over some region R in $\mathbb{R}^n$. So, region R lies in $\mathbb{R}^n_{x_1,...,x_n}$, a copy of $\mathbb{R}^n$ where we use coordinates $x_1, ..., x_n$.

A change of variables means that we will view coordinates x_i of points in R as functions $x_i = T_i(u)$ of coordinates $u = (u_1, ..., u_n)$ in some other region R' in $\mathbb{R}^n_{u_1,...,u_n}$.

After we have chosen the change of variables functions $T_i(u)$ we still need to know the region R' in $\mathbb{R}^n_{u_1,...,u_n}$ that corresponds to our original region R in $\mathbb{R}^n_{x_1,...,x_n}$.

For this it is convenient to think of the n-tuple of functions $T_1(u), ..., T_n(u)$ from $\mathbb{R}^n_{u_1,...,u_n}$ to real numbers as a single function T from $\mathbb{R}^n_{u_1,...,u_n}$ to $\mathbb{R}^n_{x_1,...,x_n}$. The value $T(u) \in \mathbb{R}^n_{x_1,...,x_n}$ is an n-tuple of numbers which we will denote as a column vector

$$T(u) \;=\; \begin{pmatrix} T_1(u) \\ T_2(u) \\ \vdots \\ T_n(u) \end{pmatrix} \;=\; \begin{pmatrix} T_1(u_1,...,u_n) \\ T_2(u_1,...,u_n) \\ \vdots \\ T_n(u_1,...,u_n) \end{pmatrix}.$$

While the mapping T is presented here by a formula, we will need to think of it geometrically – as a motion of points from $\mathbb{R}^n_{u_1,...,u_n}$ to $\mathbb{R}^n_{x_1,...,x_n}$. Indeed for any point $b = (b_1, ..., b_n)$ in $\mathbb{R}^n_{u_1,...,u_n}$, the map produces a new point denoted $T(b) = \begin{pmatrix} T_1(b) \\ \vdots \\ T_n(b) \end{pmatrix}$ in $\mathbb{R}^n_{x_1,...,x_n}$, with coordinates $T_i(b) = T_i(b_1, ..., b_n)$. So, a map T as above is just a way of "pushing" points b in $\mathbb{R}^n_{u_1,...,u_n}$ to points $T(b)$ in $\mathbb{R}^n_{x_1,...,x_n}$.

6.4.1. *The (almost) 1-1 correspondences.* In order to translate integration over R in coordinates x_i into an integral in coordinates u_j we need to know a region R' in $\mathbb{R}^n_{u_1,...,u_n}$ that "corresponds" to R under the change of coordinates given by the mapping T.

Here, "corresponds" should mean that

- T maps the region R' to the region R (i.e., any point $b \in R'$ is sent to the point $T(b)$ which lies in R);
- Map T from R' to R is a *1-1 correspondence* from R' to R, i.e., any point $a \in R$ is hit by precisely one point a' in R' in the sense that $a = T(a')$.

We see that "T is a 1-1 correspondence from R' to R" means that T can be used to translate any information about R' into information about R and vice versa.

Here is an example of the change of variables to polar coordinates.

Example. Consider the region R in the xy plane $\mathbb{R}^2_{x,y}$ described by lying in the wedge between angles θ_1 and θ_2 and between the circles of radii r_1 and r_2 (centered at origin). (Here $\theta_1 \leq \theta_2$ and $r_1 \leq r_2$.)

We will use the change of variables to polar coordinates. It is given by the mapping $T = \begin{pmatrix} x(r,\theta) \\ y(r,\theta) \end{pmatrix} \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}$, Under this mapping our region R in the x, y plane corresponds to The region R' in the r, θ plane $\mathbb{R}^2_{r,\theta}$ given by $r_1 \leq r \leq r_2$ and $\theta_1 \leq \theta \leq \theta_2$ is a rectangle. The mapping T takes R' to R (because it maps the point in R' with Cartesian coordinates (r, θ) to a point in $\mathbb{R}^2_{x,y}$ with polar coordinates r, θ.)

Is this a 1-1 correspondence? First, it is onto, i.e., by applying T to all points in R' we do hit all points in R. However, the 1-1 property is here more subtle because if θ' and θ differ by a multiple of 2π then T is the same on (r, θ) and (R, θ'). Therefore, $T : R\rangle \to R$ is a 1-1 correspondence iff $\theta_2 - \theta_1 < 2\pi$. □

6.4.2. *Volume zero spaces do not matter for integrals.* This is the observation that if we remove from a region R in $\mathbb{R}^n$ a subset A for which the volume $Vol^n(A)$ is zero, the integral does not change.

When $n = 1$ this means that we can remove several points form an interval (Vol^1 is length and the length of a point is 0). When $n = 2$ we can remove several curves from a 2-dimensional region in the plane (Vol^2 is are and the area of a curve is 0). When $n = 3$ we can remove several surfaces from a 3-dimensional region in space (Vol^3 is the usual volume and the volume of a surface is 0).

The reason is that the integral $\int_R f$ is calculated approximately as a sum of terms $Vol^n(B) \cdot f(c)$ for small boxes B in R (and a choice of a point c in B). So, integral over n-dimensional regions R is based on calculating the n-dimensional volume of pieces of R. Therefore, a subset A of R that has volume zero will not affect the integral.

Example. The polar coordinates change of variables (the example in 6.4.1) is valid even in the case when $\theta_2 - \theta_1 = 2\pi$! (For instance in the case $\theta_1 = 0$ and $\theta_2 = 2\pi$ which is often of interest!)

In this case $T : R' \to R$ is not a 1-correspondence because T has the same values at $\theta = 0$ and at $\theta = 2\pi$. However, the failure of being a 1-1 correspondence is concentrated to the segment where $\theta = \theta_2$ (and $r_1 \leq r \leq r_2$). Since this segment has measure zero it does not influence the integral! □

This example shows that

- In order that the mapping $T : R' \to R$ can be used for a change of variable in integration it suffices that T is "close" to being a 1-1 correspondence in the sense that T becomes a 1-1 correspondence after removing from R and R' subsets of volume zero.

6.5. **Recollections on the 1-dimensional case.** The change of variable is a standard method for computing 1-dimensional integrals $\int_a^b f(x)\, dx$. The first step is to consider x as a function $x = x(u)$ of another variable u. Then formally for indefinite integrals the change of variables formula comes from $dx = x'(u)\, du$. This gives

$$\int f(x)\, dx \;=\; \int f(x(u))\, dx(u) \;=\; \int f(x(u))\, x'(u)\, du.$$

However, for definite integrals we also need to change boundaries of integration

$$\int_a^b f(x)\,dx \;=\; \int_{a'}^{b'} f(x(u))\,x'(u)\,du$$

where the new boundaries of integration a', b' are chosen so that

- $x(a') = a$ and $x(b') = b$ and
- as u travels the interval $[a', b']$ starting from a' and ending at b', its image $x(u)$ travels the interval $[a, b]$ from a to b and it visits each point in $[ab]$ precisely once.

These conditions ensure that the map $u \mapsto x(u)$ really gives a strong relation (a "1-1 correspondence") between intervals $[a', b']$ and $[a, b]$ so that the integrals can be compared.

Remark. When calculating a 1-dimensional definite integral using the change of variable, one can usually avoid finding the new boundaries of integration. Rather,

- one usually calculates a formula for the indefinite integral $\int f(x(u))\,x'(u)\,du$;
- this formula is a function of u and next we translate it into a function of x (by expressing in this formula u in terms of x). Then
- one calculates the definite integral using the original boundaries of integration.

However, the indefinite integrals in several variables need not be as useful. For this reason we usually have to calculate the boundaries of integration in several variables.

6C. Change of variables: The stretching function of a mapping is its Jacobian

6.5.1. *The stretching function.* Consider a mapping $T : R' \to R$ which is a 1-1 correspondence. In order to compare integrals over R and R' we will define the *stretching function* $\mathcal{J}$ on $\mathcal{R}'$. It measures how the mapping T distorts the volume.

First for a box B in $\mathcal{R}'$ we define the *average stretching factor* over the box B as

$$\overline{\mathcal{J}}(B) \;\stackrel{\text{def}}{=}\; \frac{Vol^n(T(B))}{Vol^n(B)},$$

the ratio of the volumes of the image $T(B)$ and of the of the volume of the original box B. Then we define the value $\mathcal{J}(b)$ of the function $\mathcal{J}$ at a point $b \in \mathcal{R}'$ as the limit of average stretching factors $\overline{\mathcal{J}}(B)$ over boxes B that contain the point b as the boxes become smaller and smaller (so that the average stretching of the volume over a box approximates better and better the stretching of the volume at b)

$$\mathcal{J}(b) \;\stackrel{\text{def}}{=}\; \lim_{\to \text{ boxes } B \ni b \text{ getting small}} \overline{\mathcal{J}}(B).$$

While the average stretching $\overline{\mathcal{J}}(B)$ over a given box may be difficult to calculate (it involves integration) the stretching function will be easy to compute (this involves differentiation). We will do this in 6.5.3).

6.5.2. *Comparison of integrals over R and R'.*

Lemma. If $T : R' \to R$ is a 1-1 correspondence (up to volume zero) then it gives a change of variable formula $\int_R f = \int_{R'} f \circ T \cdot \mathcal{J}$, i.e.,

$$\int_R f(x_1, ..., x_n)\, dx_1 \cdots dx_n = \int_{R'} f(T(u_1, ..., u_n)) \cdot \mathcal{J}(u_1, ..., u_n)\, du_1 \cdots du_n.$$

Proof. **I. Approximations of regions R' and R.** Let us approximate R' by a bunch of small boxes $B_1, ..., B_S$. The image $T(B_i)$ of the box B_i under the mapping T is defined as the set $\{T(b);\ b \in B_s\}$ of all values $T(b)$ of the mapping T on points b in the box B_s. It is a subset of R obtained my moving all points in B by T. Because mapping T is a 1-1 correspondence between R' and R it translates the the approximation of the region R' by the boxes $B_1, ..., B_S$. into an approximation of the region R by the images $T(B_1), ..., T(B_S)$ of these boxes under T.

II. Approximations of integrals over R via R'. We can use this approximation of R to approximate the integral $\int_R f$ by the sum $\sum_{s=1^S} f(c_s) Vol^n(T(B_s))$ where we choose one point c_s in each $T(B_s)$.

In order to reduce the computation in R to a computation in R we notice that each c_S is of the form c'_s for a unique point c'_s in B_s. We get

$$\int_R f \sim \sum_{s=1^S} f(T(c'_s))\ Vol^n(T(B_s)).$$

III. The average stretching factor $\overline{\mathcal{J}}$. In order for this sum to be related to an integral over R', we would like to have factors $Vol^n(B_s)$ rather than $Vol^n(T(B_s))$. So, we rewrite it as

$$\int_R f \sim \sum_{s=1^S} f(T(c'_s))\ \overline{\mathcal{J}}(B_s)\ Vol^n(B_s) \ \text{ for } \ \overline{\mathcal{J}}(B_s) \stackrel{\text{def}}{=} \frac{Vol^n(T(B_s))}{Vol^n(B_s)}.$$

Here, the *average stretching factor* factor $\overline{\mathcal{J}}(B_s)$ for T on B_s measures how the mapping T distorts the volume on the box B_s.

III. The stretching function $\mathcal{J}$. The last step is to turn the average stretching factors for boxes $\overline{\mathcal{J}}$ into an actual function $\mathcal{J}$ on $\mathcal{R}'$. The the definition of $\mathcal{J}$ above, the average stretching $\overline{\mathcal{J}}(B_s)$ over a box B_s is(well) approximated by the value $\mathcal{J}(c'_s)$ of the function $\mathcal{J}$ at a point in the small box B_s. So, we have

$$\int_R f \sim \sum_{s=1^S} f(T(c'_s))\ \overline{\mathcal{J}}(B_s)\ Vol^n(B_s) \sim \sum_{s=1^S} f(T(c'_s))\ \mathcal{J}(c'_s)\ Vol^n(B_s) \sim \int_{R'} f \circ T \cdot \mathcal{J}.$$

Now, $\int_R f = \int_{R'} f \circ T \cdot \mathcal{J}$ since both integrals are approximated by the same numbers! □

6.5.3. *The stretching function of a mapping is its Jacobian determinant.*

Proposition. The stretching factor of a mapping T from $\mathbb{R}^n$ to $\mathbb{R}^n$ at a point b is the absolute value $\left|\frac{\partial(x_1,...,x_n)}{\partial(u_1,...,u_n)}\right|$ of the determinant of the matrix $(\nabla T)(b)$ of the matrix of the differential $dT(b)$ of the mapping at b.

$$\mathcal{J}_T(b) = \left|\frac{\partial(x_1,...,x_n)}{\partial(u_1,...,u_n)}\right|. \quad \square$$

The proof will be divided into four steps. One first reduces the calculation of the stretching factor from the mapping T to its differential dT (step I). Then one one computes the stretching factor of the differential using the fact that it is a fairly some mapping – a linear mapping, so it is determined by its matrix. The computation of stretching factors for linear mappings (step III) is based on the volume formula for slanted boxes (established in step II).

I. Linear approximation. Remember that a function of one variable $f(x)$ has a linear approximation near a point a, by the linear function $y = f(a) + f'(a)(x-a)$. The graph of this approximation is the tangent line at $x = a$ to the graph of $y = f(x)$.

One can say it also in terms of the change of function $f(x)-f(a)$ – it has an approximation by the function $y = f'(a)(x-a)$ which is called the differential of f at a.

The same applies to a function $f(x_1,...,x_n)$ of n variables, the change $f(x) - f(a)$ – which in this case means $f(x_1,...,x_n) - f(a_1,...,a_n)$ – has a linear approximation by the differential $df(a)$ (which is a function from $\mathbb{R}^n$ to $\mathbb{R}$), applied to the vector $x - a = (x_1,...,x_n) - (a_1,...,a_n)$ in $\mathbb{R}^n$.

Finally, the same works for the mappings $T : \mathbb{R}^m \to \mathbb{R}^n$. For $u = (u_1,...,u_m)$ near a point $b = (b_1,...,b_m)$ in $\mathbb{R}^m$, the change of function vector $T(u) - T(b)$ in $\mathbb{R}^n$, is approximated by $[(dT)(b)](u-b)$, the differential $dT(b)$ of the mapping T at the point $b \in \mathbb{R}^m$ (this differential is a new function from $\mathbb{R}^m$ to $\mathbb{R}^n$), applied to the vector $u - b$ in $\mathbb{R}^m$.[5]

---

[5] Actually, there is nothing new here, the mapping is a column vector $T = \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}$, of functions from $\mathbb{R}^n$ to $\mathbb{R}$ so, approximating $T(u_1,...,u_m)$ means approximating each of the functions $T_i(u_1,...,u_m)$. This is done by the differential $dT(a)$ because this differential is again such column vector of the differentials of component functions

$dT(a) = \begin{pmatrix} dT_1(a) \\ \vdots \\ dT_n(a) \end{pmatrix}$. So, the new approximation of $T(u)$ reduces to the known approximations of component functions.

Lemma. The stretching factor $\mathcal{J}_T(b)$ for a mapping T at a point b is the same as the stretching factor $\mathcal{J}_{dT}(0)$ for the differential $dT(b)$ (which is another function $\mathbb{R}^m \to \mathbb{R}^n$) at the point $0 \in \mathbb{R}^m$.

Proof. The point is that in order for $T(u)$ to be as close to the value of its linear approximation $T(b) + [(dT)(b)](u-b)$ as we want, it suffices that u is sufficiently close to b; say that u is in a sufficiently small box B around b. Since the function $T(u)$ and its linear approximation are very close on B the images of B under T and its linear approximations are "almost the same". Therefore, their volumes are "almost the same". This implies that the average stretching factors on B are arbitrarily close for the function and its linear approximation. Now in the limit as the box B shrinks around the point b, we get that the stretching factor $\mathcal{J}_T(b)$ for T at b is the same as f the stretching factor $\mathcal{J}_{T(b)+dT(b)}$ for the linear approximation function $T(b) + dT(b)$.

Finally, the linear approximation $T(b) + dT(b)(x-u)$ is a composition of three simple function, a translation $u \mapsto u-b$ by vector b, the differential $dT(b)$ and the translation by the vector $T(b)$. Since the translations obviously preserve the volume, the stretching factors are the same: $[\mathcal{J}_{T(b)+dT(b)(x-u)](b)} = \mathcal{J}_{dT(b)}(0)$. So, $\mathcal{J}_T(b) = \mathcal{J}_{dT(b)}(0)$. □

xxx **II. Paralellopipeds, i.e., slanted boxes.** For an n-tuple of vectors $v = (v_1, ..., v_n)$ with $v_i \in \mathbb{R}^n$, we define the *parallelopiped* $P(v_1, ..., v_n)$ as the set of all linear combinations $c_1v_1 + \cdots + c_nv_n$ of of vectors v_i with coefficient numbers c_i between 0 and 1 :

$$P(v_1, ..., v_n) \stackrel{\text{def}}{=} \{c_1v_1 + \cdots + c_nv_n;\ 0 \leq\ c_1, ..., c_n \leq 1\}.$$

Example. In $\mathbb{R}^n$ we have *basic vectors* $e_1, ..., e_n$ where $e_1 = (1, 0, ..., 0)$, $e_2 = (0, 1, ..., 0)$ and so on until $e_n = (0, ..., 0, 1)$. Notice that if we multiply each e_i by a number $b_i \geq 0$, then the parallelopiped $Pt(b_1e_1, ..., b_ne_n)$ is the box $B\left(\begin{smallmatrix} 0 \\ b_1,...,b_n \end{smallmatrix}^{\cdots *0}\right)$ consisting of all $x \in \mathbb{R}^n$ with $0 \leq x_i \leq b_i$. □

Lemma. The volume of the parallelopiped $P(v_1, .., v_n)$ is $|\det(v_1, .., v_n)|$ of the nxn, the absolute value of the determinant $\det(v_1, .., v_n)$ of the nxn matrix $(v_1, ..., v_n)$ with columns v_i.

Proof. We will not check this statement in general, i.e., for all choices of vectors $v_1, ..., v_n$.

In dimension 3 this determinant formula for the volume of a parallelopiped is well known (see MATH233). This implies the cases of all dimensions ≤ 3.[6]

An example in any dimension is the case when the vectors $_i$ are multiples of vectors e_i. Then $P(b_1e_1, ..., b_ne_n) = B\left(\begin{smallmatrix} 0 \\ b_1,...,b_n \end{smallmatrix}^{\cdots *0}\right)$ the volume is $(b_1 - 0)\cdot\cdot(b_n - 0) = b_1\cdot\cdot b_n$. This

[6] For $n = 1$ the formula is obvious. Say, in dimension $n = 2$ $Vol^2([P(v_1, v_2)])$ is the area of the parallelogram $P(v_1, v_2)]$ with sides given by vectors v_1, v_2. This is the same as the volume of the parallelopiped $P(v_1, v_2, e_3)$ in dimension 3 since its base is the parallelogram $P(v_1, v_2)]$ and the height is 1. SO, it is given by $\det(v_1, v_2, e_3)$ which is (by expansion in the 3rd row or column) the same as $\det(v_1, v_2)$.

is the same as the determinant

$$\det(b_1e_1, ..., b_ne_n) = \det \begin{pmatrix} b_1 & \cdots & 0 & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \cdots & & \vdots \\ 0 & \cdots & 0 & b_n \end{pmatrix} = b_1 \cdot \cdot b_n.$$

□yyy

III. The stretching factor of a linear map is the determinant. Now consider an Nxn matrix A and the multiplication operator, i.e., the mapping $\widetilde{A} : \mathbb{R}^n \to \mathbb{R}^n$ given by multiplying vectors with the matrix A

$$\widetilde{A}(v) \stackrel{\text{def}}{=} A\ v.$$

One can easily check that for any vector u, v and numbers a, b one has

$$\widetilde{A}(au + bv) = a\widetilde{A}u + b\widetilde{A}v.$$

(Because multiplication with a matrix is —em distributive, i.e., $A(au+bv) = aAu+bAv$.) Maps with this property are called linear.

Lemma. The stretching factor of the mapping $\widetilde{A}$ at the vector 0 in $\mathbb{R}^n$ is the absolute value of the determinant of the matrix A

$$\mathcal{J}_{\widetilde{A}}(0 = |\det(A)|.$$

Actually, for any box B the average stretching factor for $\widetilde{A}$ over B is the same – it is $|\det(A)|$.

Proof. We can choose boxes B in $\mathbb{R}^n$ that contain vector 0 as $B\begin{pmatrix} 0 & \cdots & 0 \\ b_1 & \cdots & b_n \end{pmatrix} = P(b_1e_1, ..., b_ne_n)$ for positive b_i's. Now notice that the $\widetilde{A}$-image of a parallelopiped $P(v_1, ..., v_n)$ generated by vectors v_i is the parallelopiped $P(\widetilde{A}v_1, ..., \widetilde{A}v_n)$ generated by vectors $\widetilde{A}v_i = Av_i$.

Therefore, average stretching factor for the box $B = P(b_1e_1, ..., b_ne_n)$ is

$$\overline{\mathcal{J}}(B) = Vol^n[\widetilde{A}(B)]/Vol^n(B) = Vol^n[P(Ab_1e_1, ..., Ab_ne_n))/Vol^n(P(b_1, ..., b_ne_n)) = |\det\begin{pmatrix} Ab_1e_1 & \cdots & Ab_ne_n \end{pmatrix})|/$$

However, the matrix $\begin{pmatrix} Ab_1e_1 & \cdots & Ab_ne_n \end{pmatrix})$ with columns $\mathcal{A}b_ie_i$ which are products of the matrix A with vectors b_ie_i, can be thought of as the product of matrices $A\begin{pmatrix} b_1e_1 & \cdots & b_ne_n \end{pmatrix})$. So, its determinant is $det(A) \cdot \det[\begin{pmatrix} b_1e_1 & \cdots & b_ne_n \end{pmatrix}]$. Therefore. the average stretching factor for $\widetilde{A}$ over the box B is $|\det(A)|$.

Since all average stretching factor for $\widetilde{A}$ over boxes are $|\det(A)|$, the limit of such as the box shrink is also $|\det(A)|$. □

IV. Proof of the formula for the stretching function. By step (**I**) we know that the value $\mathcal{J}_T(b)$ of the stretching function does not change if we replace T by its differential $dT(b)$ and vector b by vector 0.

Moreover, we have a formula for the differential $dT(b)$ – it is the operator $\widetilde{\nabla T(b)}$, i.e., the multiplication of vectors by the matrix $\nabla T(b)$ of the differential $dT(b)$. Finally, by step (**III**) the stretching factor of the operator $\widetilde{\nabla T(b)}$ is the absolute value of the determinant of the matrix $\nabla T(b)$. $\square$