INTEGRATION IN $\mathbb{R}^n$

1. COVERINGS

Definition 1.1. Let X be a metric space. A collection $\mathcal{U} = \{U_i\}_{i \in I}$ of open subsets of X is said to be a *covering* (or sometimes just *cover*) of X if

$$X = \bigcup_{i \in I} U_i.$$

A *subcover* of $\mathcal{U}$ is given by a subset $J \subset I$ such that $X = \bigcup_{i \in J} U_i$.

Definition 1.2. Let X be a metric space, and $\mathcal{U} = \{U_i\}_{i \in I}$ a covering of X, then $\lambda \in \mathbb{R}$, $\lambda > 0$ is said to be a *Lebesgue number* for $\mathcal{U}$ if, for all $x \in X$, there is an $i \in I$ with

$$B(x, \lambda) = \{y \in X : d(x, y) < \lambda\} \subset U_i.$$

(Note that i depends on x).

Given a metric space X and a cover $\mathcal{U} = \{U_i\}_{i \in I}$, it is not necessarily the case that a Lebesgue number for the cover exists, as the following example shows: Let $X = (0, 1)$ and for each $i \in \mathbb{N}$, let $U_i = (1/i, 1)$. Then $X = \bigcup_{i \geq 1} U_i$, but this cover has no Lebesgue number. To see this, suppose that $\lambda > 0$ is given. Choose $n \in \mathbb{N}$ such that $\frac{1}{n} < \lambda$. Then $(0, \frac{2}{n}) \subset B(\frac{1}{n}, \lambda)$ does not lie in any U_i. For sequentially compact metric spaces, on the other hand, the existence of a Lebesgue number is guaranteed for every cover by the following lemma.

Lemma 1.3. *(Lebesgue covering lemma): Let X be a sequentially compact metric space, and let $\mathcal{U} = \{U_i\}_{i \in I}$ be a covering of X. Then $\mathcal{U} = \{U_i\}_{i \in I}$ has a Lebesgue number $\lambda > 0$.*

Proof. We proceed by contradiction. If no such λ exists, then for each $n \in \mathbb{N}$, there must be a point a_n such that the ball $B(a_n, \frac{1}{n})$ does not lie in any U_i. Let $(a_{n_k})_{k \in \mathbb{N}}$ be a convergent subsequence of $(a_n)_{n \in \mathbb{N}}$, with limit $x_0 \in X$. Since $\mathcal{U}$ is a covering, there is an $i_0 \in I$ such that $x_0 \in U_{i_0}$. But then since U_{i_0} is open, there is a $\delta > 0$ such that $B(x_0, \delta) \subset U_{i_0}$. Since $a_{n_k} \to x_0$ we may find an $N_1 > 0$ such that $d(x_0, a_{n_k}) < \delta/2$ for all $k > N_1$. Set $N = \max\{2/\delta, N_1\}$, so that for $k > N$ we have $\frac{1}{n_k} \leq \frac{1}{k} < \delta/2$ and $d(a_{n_k}, x_0) < \delta/2$. It follows from the triangle inequality that for all $k > N$ we have

$$B(a_{n_k}, \frac{1}{n_k}) \subset B(x_0, \delta) \subset U_{i_0},$$

contradiction the definition of the a_ns. □

Definition 1.4. A metric space X is said to be *compact* if every covering has a finite subcover (that is, the indexing set J of the subcover is a finite set).

Lemma 1.5. *Every sequentially compact metric space is compact.*

Proof. Let X be a sequentially compact metric space, and suppose for the sake of contradiction that there is a covering $\mathcal{U} = \{U_i\}_{i \in I}$ of X which has no finite subcover. Using the previous lemma, $\mathcal{U}$ has a Lebesgue number λ. Let $a_1 \in X$ be an arbitrary point. Then $B(a_1, \lambda) \subset U_1$ for some $U_1 \in \mathcal{U}$, by the definition of a Lebesgue number. Since $\mathcal{U}$ has no finite subcover, $X \neq U_1$, and so there is an $a_2 \in X$ such that $a_2 \notin U_1$. Again by the definition of the Lebesgue number, there is an open set $U_2 \in \mathcal{U}$ such that $B(a_2, \lambda) \subset U_2$. Continuing in this way, we obtain a sequence $(a_n)_{n \in \mathbb{N}}$, and a sequence of open sets $U_n \in \mathcal{U}$ such that $B(a_n, \lambda) \subset U_n$ and $a_n \notin U_1 \cup U_2 \cup \ldots \cup U_{n-1}$.

We now show that this sequence cannot contain a convergent subsequence, for suppose $(a_{n_k})_{k \in \mathbb{N}}$ converges. Then necessarily $(a_{n_k})_{k \in \mathbb{N}}$ is a Cauchy sequence, and so there is an $N > 0$ such that for all $k, l > N$ we have $d(a_{n_k}, a_{n_l}) < \lambda$. But then picking $k > N$ we must have

$$a_{n_{k+1}} \in B(a_{n_k}, \lambda) \subset U_{n_k},$$

which contradicts the fact that $a_{n_{k+1}} \notin \bigcup_{j=1}^{n_{k+1}-1} U_j$. Thus the sequence $(a_n)_{n \in \mathbb{N}}$ contains no convergent subsequence, and so X is not sequentially compact. □

In fact, the converse to this lemma is true – a compact metric space must be sequentially compact.

Lemma 1.6. *A compact metric space is sequentially compact.*

Proof. Let X be compact, and suppose that $(a_n)_{n \in \mathbb{N}}$ is a sequence in X. We wish to show that $(a_n)_{n \in \mathbb{N}}$ has a convergent subsequence. First observe that if, for some point $x \in X$, every ball about x contains infinitely many terms of the sequence, then that point is the limit of a subsequence and we are done (*Exercise: fill in the details here*). Thus we may suppose that for every point $x \in X$ there is a ball $B(x, \delta_x)$ (for some $\delta_x > 0$ depending on x) which contains only finitely many terms of the sequence $(a_n)_{n \in \mathbb{N}}$. Then $\{B(x, \delta_x)\}_{x \in X}$ is clearly an open covering of X, and so, since X is compact, it has a finite subcover, say $B(x_1, \delta_1), B(x_2, \delta_2), \ldots, B(x_m, \delta_m)$. But then since $X = \bigcup_{j=1}^{m} B(x_j, \delta_j)$, and each of the balls $B(x_j, \delta_j)$ contains finitely many terms of the sequence, we contradict the fact that there are infinitely many terms in the sequence, and we are done. □

Finally, I want to point out that we can in fact give three different equivalent conditions for a metric space to be compact. The third condition captures the difference between a complete metric space and a compact metric space (recall that we have shown that any compact metric space is complete, or see below) and is in a sense just a distillation of what we need in order to use the "lion hunting" argument. We say that a metric space X is *totally bounded* if for every $\varepsilon > 0$ there is a finite set of points $x_1, x_2, \ldots x_m \in X$ such that

$$X = \bigcup_{i=1}^{m} B(x_i, \varepsilon).$$

Proposition 1.7. *Let X be a metric space. Then the follows properties are equivalent.*

1. *X is compact;*
2. *X is sequentially compact;*
3. *X is complete and totally bounded.*

Proof. Since we have shown that the first two are equivalent already, it is enough to show that (3) implies either (1) or (2) and vice versa. In one direction, since a sequentially compact space is clearly complete (since a Cauchy sequence converges if and only if it has a convergent subsequence), we just need to show that a compact metric space is totally bounded. But this follows directly from the definitions by considering the covering $X = \bigcup_{x \in X} B(x, \varepsilon)$. □

Exercise 1.8. Show the converse, *i.e.* that a complete totally bounded metric space is sequentially compact.

Remark 1.9. In the case of subsets X of $\mathbb{R}^n$, we also have the equivalent condition that X is closed and bounded (since these sets have been shown to be exactly the sequentially compact subsets of $\mathbb{R}^n$). It is worthwhile checking that you can show such sets are automatically complete and totally bounded – the main point being to see why boundedness and total boundedness are the same property in this case. The argument that shows sequential compactness is implied by completeness and total boundedness is essentially an abstraction of the "lion hunting" argument used to show that closed and bounded are sufficient for sequential compactness in $\mathbb{R}^n$. (It is easy to see that boundedness is not a useful property for a general metric space, and total boundedness is its more interesting substitute for arbitrary metric spaces).

Finally we recall some basic notions about open and closed sets. Given a subset A of a metric space X we define its *closure* to be

$$\bar{A} = \{x \in X : x \text{ is the limit of a sequence in } A\}.$$

Then $\bar{A}$ is a closed set, and is the smallest closed set containing A (that is, it is the intersection of the closed sets in X which contain A). The *interior* of A, is

$$\mathring{A} = \{x \in A : \exists \delta > 0 \text{ such that } B(x, \delta) \subset A\}$$

thus $\mathring{A}$ is the union of all open sets contained in A. The *boundary* of A is defined to be $\partial A = \bar{A} \backslash \mathring{A}$, that is, it is the set of points $x \in X$ for which any open ball centered at x contains both points in A and points not in A.

2. RIEMANN INTEGRATION IN $\mathbb{R}^n$

We begin with some basic definitions.

Definition 2.1. An *interval* in $\mathbb{R}$ is a bounded connected subset of $\mathbb{R}$, thus (by the intermediate value theorem), it is a set $I \subset \mathbb{R}$ of the from $[a, b]$, $(a, b]$, $[a, b)$, or (a, b) where $a \leq b$ are real numbers known as the endpoints of I. Define the *content* of I to be $\nu(I) = b - a$ (thus the content of a rectangle is the same for an open, closed or half-open interval).

A *rectangle* A in $\mathbb{R}^n$ is a product of intervals, that is, a subset of $\mathbb{R}^n$ of the form $I_1 \times I_2 \times \ldots I_n$, where I_k are intervals in $\mathbb{R}$, $(1 \leq k \leq n)$. Define the content of a rectangle to be $\nu(A) = \nu(I_1)\nu(I_2)\ldots\nu(I_n)$.

The next definition is the crucial one for the development of the integral:

Definition 2.2. Let I be an interval with endpoints $a, b \in \mathbb{R}$ where $a < b$. A partition of I is a sequence $P = (t_0, t_1, \ldots, t_m)$ where $a = t_0 < t_1 < \ldots < t_m = b$. The partition allows us to write I as the union of intervals $[t_i, t_{i+1}] \cap I$. We call these the subintervals of the partition, and write $J \in P$ when J is a subinterval of P. For a rectangle $A = I_1 \times I_2 \times \ldots \times I_n \subset \mathbb{R}^n$ a partition is a sequence $\mathcal{P} = (P_1, P_2, \ldots, P_n)$ where each P_k is a partition of the intervals I_k. If $J_1, J_2, \ldots J_n$ are subintervals of the partitions $P_1, P_2, \ldots P_n$ respectively, we say that $R = J_1 \times J_2 \times \ldots \times J_n$ is a *subrectangle* of the partition $\mathcal{P}$ and write $R \in \mathcal{P}$. Clearly the collection of subrectangles of $\mathcal{P}$ determine $\mathcal{P}$ and vice versa. Indeed if $\{\pi_k \colon \mathbb{R}^n \to \mathbb{R}\}_{1 \leq k \leq n}$ are the coordinate functions (that is, $\pi_k(x_1, x_2, \ldots, x_n) = x_k$), the sequence P_k is determined by the endpoints of the intervals $\pi_k(R)$ where R runs over the subrectangles of the partition.

Notice that if follows immediately from the definitions that if $\mathcal{P}$ is a partition of a rectangle A, then

$$\nu(A) = \sum_{R \in \mathcal{P}} \nu(P).$$

Given two partitions $\mathcal{P}$ and $\mathcal{Q}$ of A, we say that $\mathcal{P}$ *refines* $\mathcal{Q}$ if every subrectangle of P is contained in a subrectangle of Q. It follows that if R is a subrectangle of $\mathcal{Q}$, the subrectangles of P contained in Q give a partition of R. Notice also that if A is a rectangle with partition $\mathcal{P}$, then if $B \subset A$ is a rectangle inside A, the intersection of the subrectangles of $\mathcal{P}$ with B yield a partition of B. By abuse of notation we will denote this partition by $\mathcal{P} \cap B$, the partition of B *induced* by $\mathcal{P}$.

Lemma 2.3. *(Refinement lemma): Let A be a rectangle, and let $\{R_1, R_2, \ldots, R_k\}$ be any finite collection of rectangles contained in A. Then we may find a partition $\mathcal{P}$ of A such that for any subrectangle S of $\mathcal{P}$ and any rectangle R_j, either $S \cap \mathring{R}_j = \emptyset$ or $S \subset \bar{R}_j$.*

Proof. For each k, let P_k be the set of endpoints of the intervals $\pi_k(R_j)$ arranged in increasing order, and let $\mathcal{P} = (P_1, P_2, \ldots, P_n)$. It is then easy to check that this partition satisfies the conclusion of the lemma. □

For any rectangle $A \subset \mathbb{R}$ let $\mathcal{B}(A)$ be the set of bounded real-valued functions on A. Let $f \colon A \to \mathbb{R}$ in $\mathcal{B}(A)$. Set $M_A(f) = \sup\{f(x) : x \in A\}$ and $m_A(f) = \inf\{f(x) : x \in A\}$. For $\mathcal{P}$ a partition of A, we set

$$U(f, \mathcal{P}) = \sum_{R \in \mathcal{P}} M_R(f)\nu(R);$$
$$L(f, \mathcal{P}) = \sum_{R \in \mathcal{P}} m_R(f)\nu(R).$$

Clearly $L(f, \mathcal{P}) \leq U(f, \mathcal{P}$, however more importantly we have the following lemma:

Lemma 2.4. *Let A be a rectangle, and $f \in \mathcal{B}(A)$. Then if $\mathcal{P}$ and $\mathcal{Q}$ are partitions of A such that $\mathcal{P}$ refines $\mathcal{Q}$, we have*

$$L(f, \mathcal{Q}) \leq L(f, \mathcal{P}) \leq U(f, \mathcal{P}) \leq U(f, \mathcal{Q}).$$

Proof. The middle inequality was observed above. To show the left-hand inequality, suppose that $R \in \mathcal{Q}$. Then let $\mathcal{T}(R)$ be the collection of subrectangles of $\mathcal{P}$

contained in R. Then clearly we have

$$\begin{aligned}\sum_{R\in\mathcal{Q}} m_R(f)\nu(R) &= \sum_{R\in\mathcal{Q}}(\sum_{S\in\mathcal{T}(R)} m_R(f)\nu(S))\\ &\leq \sum_{R\in\mathcal{Q}}(\sum_{S\in\mathcal{T}(R)} m_S(f)\nu(S))\\ &= \sum_{S\in\mathcal{P}} m_S(f)\nu(S),\end{aligned}$$

and hence $L(f,\mathcal{Q}) \leq L(f,\mathcal{P})$. The remaining inequality is established in exactly the same fashion. □

Corollary 2.5. *Let $f \in \mathcal{B}(A)$ where A is a rectangle in $\mathbb{R}^n$, and let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of A. Then*

$$L(f,\mathcal{Q}) \leq U(f,\mathcal{P}),$$

Proof. Applying the refinement lemma to the the collection of rectangles $R \cap S$ where R is a subrectangle of $\mathcal{P}$ and S is a subrectangle of $\mathcal{Q}$, we obtain a partition $\mathcal{P}'$ which refines both the partitions $\mathcal{P}$ and $\mathcal{Q}$. If follows that

$$L(f,\mathcal{Q}) \leq L(f,\mathcal{P}') \leq U(f,\mathcal{P}') \leq U(f,\mathcal{P}),$$

and the claim is established. □

It follows immediately from the corollary that if $f \in \mathcal{B}(A)$ then the numbers

$$\sup_{\mathcal{P}} L(f,\mathcal{P}); \qquad \inf_{\mathcal{P}} U(f,\mathcal{P}),$$

exist. The lower Riemann integral of $f \in \mathcal{B}(A)$ is defined to be

$$\underline{\int}_A f = \sup_{\mathcal{P}} L(f,\mathcal{P}),$$

and the upper Riemann integral of f is

$$\overline{\int}_A f = \inf_{\mathcal{P}} U(f,\mathcal{P}).$$

Definition 2.6. Let A be a rectangle, and $f \in \mathcal{B}(A)$. We say that f is *Riemann integrable* if

$$\underline{\int}_A f = \overline{\int}_A f,$$

and denote their common value by $\int_A f$. We denote the set of Riemann integrable functions on A by $\mathcal{R}(A)$.

Finally we establish a useful criterion for a function to be integrable:

Lemma 2.7. *Let A be a rectangle and $f \in \mathcal{B}(A)$. Then $f \in \mathcal{R}(A)$ if and only if, for each $\varepsilon > 0$ there is a partition $\mathcal{P}$ of A such that*

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon.$$

Proof. First note that one implication is trivial from the definitions. Suppose now that f is Riemann integrable. Then we may find partitions $\mathcal{P}$ and $\mathcal{Q}$ such that

$$\int_A f - \varepsilon/2 < L(f,\mathcal{P}); \qquad U(f,\mathcal{Q}) < \int_A f + \varepsilon/2.$$

Picking $\mathcal{P}'$ to be a common refinement of $\mathcal{P}$ and $\mathcal{Q}$ as in the previous lemma, and applying Lemma 2.4 we see that $\mathcal{P}'$ has the required property. □

As an example of how we may use this criterion, we show that the integral is additive:

Lemma 2.8. *Let $A \subset \mathbb{R}^n$ be a rectangle and suppose that $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ are integrable. Then $f+g$ is integrable and moreover*

$$\int_A (f+g) = \int_A f + \int_A g.$$

Proof. Let $\varepsilon > 0$ be given. Since f is integrable there is a partition of A such that $U(f,\mathcal{P}) - L(f\mathcal{P}) < \varepsilon/2$, and similarly there is a partition $\mathcal{Q}$ such that $U(g,\mathcal{Q}) - L(g,\mathcal{Q}) < \varepsilon/2$. Using the refinement lemma (Lemma 2.3) for the rectangles $S \cap T$ where S is a subrectangle of $\mathcal{P}$ and T is a subrectangle of $\mathcal{Q}$ we obtain a partition $\mathcal{R}$ refining both. Now given any rectangle $R \subset A$ we have

$$M_R(f+g) \leq M_R(f) + M_R(g); \qquad m_R(f+g) \geq m_R(f) + m_R(g),$$

and hence it follows that $U(f+g,\mathcal{R}) \leq U(f,\mathcal{R}) + U(g,\mathcal{R})$ and $L(f+g,\mathcal{R}) \geq L(f,\mathcal{R}) + L(g,\mathcal{R})$. Then using Lemma 2.4 we have

$$\begin{aligned} U(f+g,\mathcal{R}) - L(f+g,\mathcal{R}) &\leq (U(f,\mathcal{R}) - L(f,\mathcal{R})) + (U(g,\mathcal{R}) - L(g,\mathcal{R}) \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon, \end{aligned}$$

and hence $f+g$ is integrable. The additivity follows immediately from the inequality $L(f+g,\mathcal{R}) \leq \int_A f+g \leq U(f+g,\mathcal{R})$, and the corresponding inequalities for f and g. □

Remark 2.9. Since it is straightforward to check that $\int_A$ is compatible with scalar multiplication, it follows that the set $\mathcal{R}(A)$ of bounded integrable functions on A is a vector subspace of the space of bounded functions, and moreover that $\int_A$ is a linear map from $\mathcal{R}(A)$ to $\mathbb{R}$.

3. MEASURE AND CONTENT ZERO

Given a subset of A of $\mathbb{R}^n$ we say that A has *measure zero* if for every $\varepsilon > 0$ we can find a sequence of rectangle $R_1, R_2, \ldots,$ such that

$$A \subset \bigcup_{i \geq 1} R_i, \text{ and } \sum_{i \geq 1} \nu(R_i) < \varepsilon.$$

Similarly, we say that A has *content zero* if for every $\varepsilon > 0$ we can find a finite set of rectangles $R_1, R_2, \ldots R_m$ such that $A \subset \bigcup_{i=1}^m R_i$ and $\sum_{i=1}^m \nu(R_i) < \varepsilon$.

Remark 3.1. It will be useful to notice that we could require in both the definition of content zero and measure zero that the rectangles R_i all be open, or that they all be closed. (*Check this*). Notice also that if B is a subset of $A \subset \mathbb{R}^n$ and A has measure zero (or content zero) then B does also.

Clearly, a set of content zero has measure zero, but the converse is not true, as we will see below. First make some basic observations about sets of measure zero. Clearly the finite union of sets of content zero has content zero, however we can do somewhat better for sets of measure zero.

Lemma 3.2. *Let $A_1, A_2, \ldots$ be subsets of $\mathbb{R}^n$ of measure zero. Then their union $A = \bigcup_{i\geq 1} A_i$ is a set of measure zero. Moreover, any countable subset of $\mathbb{R}^n$ has measure zero.*

Proof. Let $\varepsilon > 0$ be given. Since each A_i has measure zero, we may find rectangle R^i_j such that $A_i \subset \bigcup_{j\geq 1} R^i_j$ and $\sum_{j\geq 1} \nu(R^i_j) < \varepsilon/2^i$. Now the collection $\{R^i_j\}_{i,j\in\mathbb{N}}$ is countable set of rectangles, (since the product $\mathbb{N}\times\mathbb{N}$ is countable) so we may label the rectangles R^i_j as $S_1, S_2, \ldots$. Then we have

$$A = \bigcup_{i\geq 1} A_i \subset \bigcup_{i,j\geq 1} R^i_j = \bigcup_{k\geq 1} S_k,$$

and

$$\sum_{k\geq 1} \nu(S_k) = \sum_{i,j\geq 1} \nu(R^i_j) = \sum_{i\geq 1}(\sum_{j\geq 1} \nu(R^i_j)) < \sum_{i\geq 1} \varepsilon/2^i = \varepsilon,$$

and hence A has measure zero as required. The moreover part follows once we observe that a point has measure zero (in fact it clearly has content zero). □

This lemma implies in particular that since the rationals $\mathbb{Q}$ are countable, they are a set of measure zero, even though they are dense in $\mathbb{R}$.

We now want to understand what the difference between measure and content zero. Recall that for A a subset of $\mathbb{R}^n$ we denote its closure by $\bar{A}$ (see the end of section 1).

Lemma 3.3. *Suppose that $A \subset \mathbb{R}^n$.*

(1) *If A has content zero, then $\bar{A}$, the closure of A, is compact and has content zero.*
(2) *If A is compact, then A has content zero if and only if it has measure zero.*

Proof. For the first part, notice that if A has content zero, then given any $\varepsilon > 0$ we may find closed rectangles $R_1, R_2, \ldots, R_m$ such that $A \subset \bigcup_{1\leq i\leq m} R_i$ and $\sum_{i=1}^m \nu(R_i) < \varepsilon$. Since the rectangles are closed, and the finite union of closed sets is closed, it follows immediately that $\bar{A} \subset \bigcup_{1\leq i\leq m} R_i$, and hence $\bar{A}$ has content zero. Moreover, since any rectangle is bounded, and the finite union of bounded sets is bounded, it follows that $\bar{A}$ is bounded. Since $\bar{A}$ is also closed, $\bar{A}$ is a compact subset of $\mathbb{R}^n$.

For the second part, suppose that A is compact and has measure zero. Then given $\varepsilon > 0$ there are open rectangles $R_1, R_2, \ldots$ such that $A \subset \bigcup_{i\geq 1} R_i$ and $\sum_{i\geq 1} \nu(R_i) < \varepsilon$. Since A is compact, it follows that there is some finite subset of these rectangles $R_{i_1}, R_{i_2}, \ldots, R_{i_m}$ such that $A \subset \bigcup_{k=1}^m \nu(R_{i_k})$, and hence A has content zero. □

It follows from the lemma that a set has content zero if and only if its closure is compact and has measure (or content) zero, which in $\mathbb{R}^n$ is equivalent to requiring its closure to be bounded and have measure zero.

Finally we want to show that not all subsets of $\mathbb{R}^n$ have measure zero!

Lemma 3.4. *Let A be a rectangle in $\mathbb{R}^n$. Then A does not have measure zero.*

Proof. Clearly we may assume that A is close. Suppose that A has measure zero, then by the previous lemma it has content zero. Hence there exist closed rectangles $R_1, R_2, \dots R_m$ covering A with $\sum_{j=1}^{m} \nu(R_j) < \frac{1}{2}\nu(A)$. Using the refinement lemma we may find a partition $\mathcal{P}$ of A such that each subrectangle of the partition lies in some R_j. But then since the content of a rectangle is the sum of the contents of the subrectangles in any partition we must have

$$\nu(A) = \sum_{S\in\mathcal{P}} \nu(S) = \sum_{j=1}^{m}(\sum_{S\in\mathcal{P};S\subset R_j} \nu(S)) < \sum_{j=1}^{m} \nu(R_j) < \frac{1}{2}\nu(A),$$

and we have a contradiction. □

Remark 3.5. Notice that this gives a proof that the real numbers are uncountable.

4. Oscillation of bounded functions and Integrability

The goal of this section is to establish which functions are Riemann integrable. A good preparatory exercise is to show that any continuous function on a closed rectangle is integrable – a result which follows quickly from the fact that such a function is uniformly continuous.

Definition 4.1. Let $A \subset \mathbb{R}^n$, and $f\colon A \to \mathbb{R}$ a bounded function on A. Then for each $x \in A$ and $\delta > 0$ set

$$M(f, x, \delta) = \sup\{f(y) : y \in A, \|x - y\| < \delta\};$$
$$m(f, x, \delta) = \inf\{f(y) : y \in A, \|x - y\| < \delta\}.$$

It is clear that the difference $M(f, x, \delta) - m(f, x, \delta)$ decreases as δ decreases, hence the limit

$$\lim_{\delta\to 0}[M(f, x, \delta) - m(f, x, \delta)],$$

exists. We denote it by $o(f, x)$ – the *oscillation* of f at x.

Remark 4.2. It is easy to see that a bounded function $f\colon A \to \mathbb{R}$ is continuous at a point $x \in A$ if an only if $o(f, x) = 0$. In establishing our criterion for integrability, we will need to consider points where we only know $o(f, x) < \varepsilon$ for some positive constant ε. The advantage of considering the set of such points is given by the next lemma.

Lemma 4.3. *Let $A \subset \mathbb{R}^n$ be a closed set, and let $f\colon A \to \mathbb{R}$ be a bounded function. Then for any $\varepsilon > 0$ the set of points*

$$B_\varepsilon = \{x \in A : o(f, x) \geq \varepsilon\}$$

is closed.

Proof. We must show that $\mathbb{R}^n - B_\varepsilon$ is open. Consider a point $x \in \mathbb{R}^n - B_\varepsilon$. If $x \notin A$, then since A is closed there is a $\delta > 0$ with $B(x, \delta) \subset \mathbb{R}^n - A \subset \mathbb{R}^n - B_\varepsilon$, so $\mathbb{R}^n - B$ contains an open ball about x. On the other hand, if $x \in A$, then we must have $o(f, x) < \varepsilon$, and so there is a $\delta > 0$ such that

$$M(f, x, \delta) - m(f, x, \delta) < \varepsilon.$$

Now suppose that $y \in B(x, \delta/2)$. Then since, by the triangle inequality we have $B(y, \delta/2) \subset B(x, \delta)$, it follows directly from the definitions that $M(f, y, \delta/2) - m(f, y, \delta/2) < \varepsilon$, and so $B(x, \delta/2) \subset \mathbb{R}^n - B_\varepsilon$, and $\mathbb{R}^n - B_\varepsilon$ is open as required. □

On the other hand, functions with small oscillation are not much worse (for the purposes of integration) than continuous functions in the following sense:

Lemma 4.4. *Suppose that $A \subset \mathbb{R}^n$ is a rectangle, and $f\colon A \to \mathbb{R}$ is a bounded function. Then if for all $x \in A$ we have $o(f,x) < \varepsilon$ for some $\varepsilon > 0$, we can find a partition P of A such that*

$$U(f,P) - L(f,P) < \varepsilon\nu(A).$$

Proof. For each $x \in A$ we may choose $\delta_x > 0$ such that

$$M(f,x,\delta_x) - m(f,x,\delta_x) < \varepsilon.$$

Now $\{B(x,\delta_x)\}_{x\in A}$ is an open covering of A, hence since A is closed and bounded in $\mathbb{R}^n$, it has a Lebesgue number $\lambda > 0$. Choose a partition P of A such that each subrectangle R has

$$\operatorname{diam}(R) = \sup\{\|x-y\| : x, y \in A\} < \lambda.$$

(one can ensure this by, for example, insisting that the side lengths of the subrectangles are all less than λ/n). Then it follows from the definition of the Lebesgue number that each subrectangle R is contained in some ball $B(x,\delta_x)$ and so $M_R(f) - m_R(f) < \varepsilon$. Thus we have

$$U(f,P) - L(f,P) = \sum_{R\in P}(M_R(f) - m_R(f))\nu(R) < \varepsilon\sum_{R\in P}\nu(R) = \varepsilon\nu(A).$$

□

Remark 4.5. Observe that this lemma already shows that any continuous function on a rectangle is Riemann integrable – if $f\colon A \to \mathbb{R}$ is continuous, then $o(f,x) = 0$ for all $x \in A$, and so the lemma ensures that we can find partitions whose upper and lower sum differ by arbitrarily small amounts, and hence by Lemma 2.7 f is integrable.

Theorem 4.6. *Let A be a closed rectangle in $\mathbb{R}^n$, and $f\colon A \to \mathbb{R}$ a bounded function. Then f is Riemann integrable if and only if the set*

$$B = \{x \in A : f \text{ is not continuous at } x\}$$

has measure zero.

Proof. Suppose first that the set B has measure zero. Then observe that, for all $\varepsilon > 0$ we have

$$B_\varepsilon = \{x \in A : o(f,x) \geq \varepsilon\} \subset B,$$

and hence, since by Lemma 4.3 B_ε is closed about bounded, B_ε has content zero. Thus we may find a finite set of open rectangles $R_1, R_2, \ldots, R_m$ such that $B_\varepsilon \subset \bigcup_{i=1}^m R_i$ and $\sum_{i=1}^m \nu(R_i) < \varepsilon$. By the refinement lemma we may pick a partition $\mathcal{P}$ of A such that the subrectangles of $\mathcal{P}$ are either contained in one of the R_i, or are disjoint from B_ε. Let $\mathcal{S}_1$ be the subrectangles which are disjoint from B_ε, and let $\mathcal{S}_2$ be the subrectangles which lie in some R_i. For each $S \in \mathcal{S}_1$ by Lemma 4.4 we can find a partition $\mathcal{P}_S$ of S such that

$$U(f,\mathcal{P}_S) - L(f,\mathcal{P}_S) < \varepsilon\nu(S).$$

Using the refinement lemma for the collection of rectangle

$$\{R : R \in \mathcal{P}_S \text{ for some } S \in \mathcal{S}_1\} \cup \{R_i : 1 \leq i \leq m\},$$

it follows that we may find a partition $\mathcal{Q}$ of A such that the induced partition of $S \in S_1$ refines $\mathcal{P}_S$. Let $M > 0$ be such that $|f(x)| < M$ for all $x \in A$. Then using Lemma 2.4 we see that

$$\begin{aligned}
U(f,\mathcal{Q}) - L(f,\mathcal{Q}) &= \sum_{T\in\mathcal{Q}:T\subset S\in\mathcal{S}_1} (M_T(f) - m_T(f))\nu(T) \\
&\quad + \sum_{T\in\mathcal{Q};T\subset R_j,1\leq j\leq m} (M_T(f) - m_T(f))\nu(T). \\
&= \sum_{S\in\mathcal{S}_1} (U(f,\mathcal{Q}\cap S) - L(f,\mathcal{Q}\cap S)) \\
&\quad + \sum_{T\in\mathcal{Q};T\subset R_j,1\leq j\leq m} (M_T(f) - m_T(f))\nu(T). \\
&\leq \sum_{S\in\mathcal{S}_1} (U(f,\mathcal{P}_S) - L(f,\mathcal{P}_s)) + \sum_{j=1}^{m} 2M\nu(R_j) \\
&\leq \sum_{S\in\mathcal{S}_1} \varepsilon\nu(S) + 2M\varepsilon \\
&\leq \varepsilon(\nu(A) + 2M).
\end{aligned}$$

Since ε was arbitrary, we see that f satisfies the condition of Lemma 2.7, and hence is Riemann integrable.

To see the converse, we show that $B_{1/n}$ has content zero. Since $B = \bigcup_{n\geq 1} B_{1/n}$ this will imply that B has measure zero. Suppose that f is Riemann integrable, and let $\varepsilon > 0$ be given. Then there is a partition $\mathcal{P}$ of A such that

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon/n$$

Then let $\mathcal{T}$ be the set of subrectanges of $\mathcal{P}$ whose interiors intersect B_ε. Then we have

$$\begin{aligned}
\frac{1}{n}\sum_{S\in\mathcal{T}} \nu(S) &\leq \sum_{S\in\mathcal{T}} (M_S(f) - m_S(f))\nu(S) \\
&\leq U(f,\mathcal{P}) - L(f,\mathcal{P}) \\
&\leq \varepsilon/n.
\end{aligned}$$

and thus $\sum_{S\in\mathcal{S}_1} \nu(S) < \varepsilon$. Therefore

$$B_{1/n} \subset \{\partial S : S \in \mathcal{P}, S \notin \mathcal{T}\} \cup \{S : S \in \mathcal{T}\}.$$

But the boundary of a rectangle is a union of $2n$ rectangles R with $\nu(R) = 0$ (because it is a product of intervals one of which is a point), we see that $B_{1/n}$ is contained in the union of finitely many rectangles the sum of whose content is less than ε, and hence $B_{1/n}$ has content zero as required. □

Remark 4.7. Observe that the set B of discontinuities of f need not be closed or have content zero. If $f\colon [0,1] \to \mathbb{R}$ is the function given by

$$f(x) = \begin{cases} 1/q & \text{if } x = p/q, \text{ with } p,q \text{ relatively prime,} \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

then the set of discontinuities of f is $\mathbb{Q}\cap[0,1]$, which is dense in $[0,1]$. Nevertheless, f is Riemann integrable (with integral zero).

So far, we have only defined the integral of a bounded function over a rectangle. We may now define integrals over larger class of subsets of $\mathbb{R}^n$, and at the same time give a class of subsets for which we have a good notion of volume. First suppose that S is a bounded subset of $\mathbb{R}^n$. Let $\mathbf{1}_S$ be the characteristic function of S, that is $\mathbf{1}_S\colon \mathbb{R}^n \to \mathbb{R}$ with

$$\mathbf{1}_S(x) = \begin{cases} 1, & x \in S \\ 0, & x \notin S \end{cases}$$

Then we may find a rectangle A containing S, and then the integral

$$\int_A \mathbf{1}_S,$$

if it exists, is clearly independent of the choice of A. We define the volume or content $\nu(S)$ of S to be this integral. It is easy to see that this definition extends our earlier definition of the content of a rectangle. The previous theorem gives us a characterization of the sets whose content we can define in this way: they are exactly the sets for which $\mathbf{1}_S$ is continuous almost everywhere. Since the points where $\mathbf{1}_S$ is discontinuous are exactly the points in the boundary ∂S of S, we see that $\mathbf{1}_S$ is integrable if and only if ∂A is bounded and has measure zero (since ∂A is closed this is equivalent to requiring ∂A to be bounded and have content zero). The following definition sumarizes this discussion:

Definition 4.8. A bounded subset S of $\mathbb{R}^n$ is said to be *Jordan measurable*, or have Jordan content, if ∂S is bounded and of measure zero. In this case we set

$$\nu(S) = \int_A \mathbf{1}_S,$$

where A is any rectangle containing S.

Remark 4.9. It is not the case that any bounded open or closed set is necessarily Jordan measurable – see the problem sets for an example. In order to ensure that all bounded open and closed sets have a good notion of volume, one must develop a more sophisticated notion of integral than the one we use here.

As a consistency check, we note that this definition agrees with the definition of $\nu(R)$ we already have in the case where R is a rectangle.

Lemma 4.10. *Let A be a closed rectangle and R a rectangle contained in A. Then*

$$\int_A \mathbf{1}_R = \nu(R)$$

(where the right-hand side is the product of the side lengths of the rectangle R).

Proof. Note that the boundary ∂R of R has content zero, and so given $\varepsilon > 0$ we may find open rectangles $T_1, T_2, \dots, T_m$ such that $\partial R \subset \bigcup_{j=1}^m T_j$ and $\sum_{j=1}^m \nu(T_j) < \varepsilon$. Choose a partition $\mathcal{P}$ of A refining the collection of rectangles $\{R, R_1, R_2, \dots, R_m\}$ so that for ever subrectangle S of $\mathcal{P}$ we have $S \subset \bar{R}_j$ for some j or $S \cap \partial R = \emptyset$, and either $S \subset \bar{R}$ or $S \cap \mathring{R} = \emptyset$. Thus for $S \in \mathcal{P}$ we have $M_S(\mathbf{1}_R) - m_S(\mathbf{1}_R)$ is zero if S does not intersect the boundary of R and is 1 otherwise. But then we have

$$U(\mathbf{1}_R, \mathcal{P}) - L(\mathbf{1}_R, \mathcal{P}) \leq \sum_{j=1}^m \nu(R_j) = \varepsilon.$$

Moreover we have

$$\begin{aligned} U(\mathbf{1}_R, \mathcal{P}) - \nu(R) &= U(f, \mathcal{P}) - \sum_{S \in \mathcal{P}; S \subset \bar{R}} \nu(S) \\ &\leq \sum_{S \in \mathcal{P}; S \subset \bar{R}_j} \nu(S). \\ &< \sum_{j=1}^{m} \nu(R_j) = \varepsilon. \end{aligned}$$

hence it follows that $\int_A \mathbf{1}_R = \nu(R)$ as required. □

We may also now readily extend our definition of integral to allow us to integrate over Jordan measurable sets.

Definition 4.11. Let $\mathcal{R}(\mathbb{R}^n)$ be the set of bounded functions on $\mathbb{R}^n$ which are equal to zero on the complement of a bounded set and which are continuous everywhere except on a set of measure zero. We say that $f \in \mathcal{R}(\mathbb{R}^n)$ is Riemann integrable, and for any Jordan measurable subset S of $\mathbb{R}^n$ we define

$$\int_S f = \int_A f.\mathbf{1}_S.$$

Here the right-hand side is defined, because the function $f.\mathbf{1}_S$ is bounded and continuous everywhere except on a set of measure zero. We also write $\int_{\mathbb{R}^n} f$ for the integral of f over any rectangle outside of which $f = 0$. Note that the set of Riemann integrable functions on $\mathbb{R}^n$ is clearly closed under addition and multiplication.

Remark 4.12. In order to check that this definition is independent of the choice of the large rectangle A, one needs to prove an analogue of Lemma 4.10. This is left as an exercise for the fabled enthusiastic reader.

If S is Jordan measurable we define $\mathcal{R}(S)$ to be the set of functions $f \colon S \to \mathbb{R}$ which are bounded and continuous except on a set of measure zero. Notice that if $f \in \mathcal{R}(S)$ then the function

$$g(x) = \begin{cases} f(x), & x \in S \\ 0, & x \notin S \end{cases}$$

is in $\mathcal{R}(\mathbb{R}^n)$, thus the previous definition also gives a notion of integral for functions in $\mathcal{R}(S)$.

5. FUBINI'S THEOREM

We have now established a theory of integration for a large class of functions. However, we still do not have any convenient method by which we can evaluate integrals. Our next major result shows that integrals may be computed iteratively, and hence we can reduce the problem of computing integrals in $\mathbb{R}^n$ to the same problem for $\mathbb{R}$. Since we have the Fundamental Theorem of Calculus in the single variable case, this is a significant improvement. We start by defining a class of functions for which the integral is trivial to compute.

Definition 5.1. Let A be a rectangle in $\mathbb{R}^n$. Then we have $\mathbf{1}_A \in \mathcal{R}(\mathbb{R}^n)$. A *simple* function on $\mathbb{R}^n$ is a linear combination of the characteristic functions of rectangles, that is f is a simple function if we may write

$$f = \lambda_1 \mathbf{1}_{A_1} + \lambda_2 \mathbf{1}_{A_2} + \ldots + \lambda_m \mathbf{1}_{A_m},$$

where A_i, $(1 \leq i \leq m)$ are rectangles and $\lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}$. Since the intersection of two rectangles is again a rectangle, and $\mathbf{1}_R.\mathbf{1}_S = \mathbf{1}_{R\cap S}$ the space of simple functions is in fact closed under addition and multiplication of functions, and hence we refer to the set of simple functions $\mathcal{S}(\mathbb{R}^n)$ as the algebra of simple functions.

Lemma 5.2. *Let $f\colon A \to \mathbb{R}^n$ be a simple function,*

$$f = \lambda_1 \mathbf{1}_{R_1} + \lambda_2 \mathbf{1}_{R_2} + \ldots + \lambda_m \mathbf{1}_{R_m},$$

where $R_k \subset A$ for each k $(1 \leq k \leq m)$. Then

$$\int f = \sum_{i=1}^{m} \lambda_i \nu(R_i).$$

Proof. By the linearity of the integral, we are reduced to showing that $\int_A \mathbf{1}_R = \nu(R)$ for a rectangle $R \subset A$, but this is exactly the statement of Lemma 4.10 □

It is easy to check the following result. If f, g are real-valued functions on a set X, we write $f \leq g$ if for every $x \in X$ we have $f(x) \leq g(x)$.

Lemma 5.3. *Let A be a closed rectangle, and $f\colon A \to \mathbb{R}$ a bounded function. Then we have*

$$\overline{\int_A} f = \inf\{\int_A \psi : \psi \in \mathcal{S}(A), f \leq \psi\},$$
$$\underline{\int_A} f = \sup\{\int_A \phi : \phi \in \mathcal{S}(A), \phi \leq f\}$$

Proof. It suffices to show the first of these equalities (as the second follows from the first applied to $-f$). Now suppose that $\mathcal{P}$ is a partition of A. Let $\mathcal{B}$ denote the rectangles which occur as the boundary of some $S \in \mathcal{P}$, so that $\nu(R) = 0$ for all $R \in \mathcal{B}$. Let

$$\psi = \sum_{S\in\mathcal{P}} M_S(f)\mathbf{1}_{\mathring{S}} + \sum_{R\in\mathcal{B}} \|f\|_\infty \mathbf{1}_R$$

Then $f \leq \psi$ and it follows from the previous lemma that $\int_A \psi = U(f, \mathcal{P})$. Hence we see that $\inf\{\int_A \psi : \psi \in \mathcal{S}(A), f \leq \psi\} \leq \overline{\int}_A f$. On the other hand, for any functions g_1, g_2 it follows immediately from the definitions that

$$\overline{\int_A} g_1 \leq \overline{\int_A} g_2,$$

thus clearly $\overline{\int}_A f \leq \inf\{\int_A \psi : \psi \in \mathcal{S}(A), f \leq \psi\}$. □

Corollary 5.4. *Let A be a rectangle. Then a function $f\colon A \to \mathbb{R}$ is integrable if and only if for every $\varepsilon > 0$ we may find simple functions $\phi\colon A \to \mathbb{R}, \psi\colon A \to \mathbb{R}$ such that $\phi \leq f \leq \psi$, and*

$$\int_A \psi - \int_A \phi < \varepsilon.$$

Proof. This is immediate from the previous lemma. □

We now state Fubini's theorem. Let $A \subset \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}$ be integrable. Then suppose we write $A = A_1 \times A_2$ where $A_1 \subset \mathbb{R}^p$ is a rectangle in $\mathbb{R}^p$ and $\mathbb{A}_2 \subset \mathbb{R}^q$ (where $p + q = n$). Given $x \in A_1$ we define a function $f_x\colon A_2 \to \mathbb{R}$ by $f_x(y) = f(x, y)$, and similarly for $y \in A_2$ we define $f_y\colon A_1 \to \mathbb{R}$ by $f_y(x) = f(x, y)$. Also, define the function $\underline{f}_1\colon A_1 \to \mathbb{R}$ by $\underline{f}_1(x) = \underline{\int}_{A_2} f_x$, and the function $\overline{f}_1\colon A_1 \to \mathbb{R}$ by $\overline{f}_1(x) = \overline{\int}_{A_2} f_x$.

Theorem 5.5. *(Fubini's theorem): Suppose that $A = A_1 \times A_2$ is a rectangle in $\mathbb{R}^n$ as above, and $f\colon A \to \mathbb{R}$ is Riemann integrable. Then the functions $\underline{f}_1$ and $\overline{f}_1$ are integrable, and*

$$\int_A f = \int_{A_1} \underline{f}_1 = \int_{x \in A_1} \Big(\underline{\int}_{y \in A_2} f_x(y)\Big)$$
$$= \int_{A_1} \overline{f}_1 = \int_{x \in A_1} \Big(\overline{\int}_{y \in A_2} f_x(y)\Big)$$

Proof. Note that the $\underline{\int}$ are necessary in the definition of $\underline{f}_1$ because, although f is Riemann integrable on A, this does not imply that f_x will be integrable on A_2 for all $x \in A_1$.

We prove the theorem in a number of steps. First consider the case of simple functions on A. Let R be a rectangle contained in A, and say $R = R_1 \times R_2$, where $R_1 \subset \mathbb{R}^p$ and $R_2 \subset \mathbb{R}^q$, and let $f = \mathbf{1}_R$. Then if $x \in A_1$ and $y \in A_2$ it is easy to see that

$$\mathbf{1}_R(x, y) = \mathbf{1}_{R_1}(x)\mathbf{1}_{R_2}(y). \tag{1}$$

Thus for $x \in A_1$, the function f_x is just $\mathbf{1}_{R_2}$ if $x \in R_1$, and zero otherwise. Similarly, we see that for any simple function ψ, the function $\psi_x \in \mathcal{S}(A_2)$ for all $x \in A_1$, and hence $\underline{\psi}_1 = \int_{A_2} \psi_x$ (that is, the lower Riemann integral in the definition of $\underline{\psi}_1$ can be replaced with the actual integral since ψ_x is integrable). Moreover, it is clear from Equation (1) that $\underline{\psi}_1 \in \mathcal{S}(A_1)$. if $\psi \in \mathcal{S}(A)$. We claim that in this case

$$\int_A \psi = \int_{A_1} \underline{\psi_1}.$$

Because both sides are linear in ψ (here we use the fact that for simple functions we can use the Riemann integral throughout – the lower Riemann integral is *not* necessarily linear) we need only check the above equation for functions of the form $\mathbf{1}_R$ as above. But for $f = \mathbf{1}_R$ it follows from Equation (1) and Lemma 4.10 that

$$\underline{f}_1 = \nu(R_2)\mathbf{1}_{R_1},$$

hence using Lemma 4.10 again we find

$$\int_{A_1} \underline{f}_1 = \int_{A_1} \nu(R_2)\mathbf{1}_{R_1} = \nu(R_1)\nu(R_2) = \nu(R) = \int_A \mathbf{1}_R$$

as required. Thus we have established $\int_A f = \int_{A_1} \underline{f}_1$ for all simple functions $f \in \mathcal{S}(A)$.

Now we consider, for an arbitrary bounded function $f\colon A \to \mathbb{R}$, the inequality

$$\underline{\int}_A f \leq \underline{\int}_{A_1} \underline{f}_1. \qquad (\star)$$

Observe that we already know ($\star$) holds (and is even an equality) when $f \in \mathcal{S}(A)$. Moreover, for any bounded f we have by Lemma 5.3

$$\underline{\int}_A f = \sup\{\int_A \psi : \psi \in \mathcal{S}(A), \psi \leq f\}$$

. But using what we have already established for simple functions, this is just

$$\sup\{\int_{A_1} \underline{\psi}_1 : \psi \in \mathcal{S}(A), \psi \leq f\},$$

and moreover, it follows from the above that if $\psi \in \mathcal{S}(A)$ and $\psi \leq f$, then $\underline{\psi}_1 \in \mathcal{S}(A_1)$ and $\underline{\psi}_1 \leq \underline{f}_1$. Using Lemma 5.4 the result follows. It follows from Lemma 5.3 that

$$\underline{\int}_A f \leq \underline{\int}_{A_1} \underline{f}_1.$$

and ($\star$) is established for any f. But now applying ($\star$) to $-f$ (since $\underline{(-f)}_1 = -\overline{f}_1$) we immediately obtain the inequality

$$\overline{\int}_{A_1} \overline{f}_1 \leq \overline{\int}_A f. \qquad (\dagger)$$

Now combining ($\star$) and ($\dagger$) in the case where f is integrable (so that the upper and lower Riemann integrals of f agree) we obtain

$$\int_A f = \underline{\int}_A f \leq \underline{\int}_{A_1} \underline{f}_1 \leq \overline{\int}_{A_1} \underline{f}_1 \leq \overline{\int}_{A_1} \overline{f}_1 \leq \overline{\int}_A f = \int_A f$$

where the second inequality follows because $\underline{f}_1 \leq \overline{f}_1$. But then each of the inequalities must be an equality, and hence $\underline{f}_1$ is integrable (and similarly $\overline{f}_1$ is integrable) and so the statement of the theorem follows. ☐

Remark 5.6. Exactly the same proof shows that we can integrate "in the opposite order", that is, we have

$$\int_A f = \int_{y \in A_2} \big(\underline{\int}_{x \in A_1} f_y(x)\big)$$

and all the other corresponding equalities hold. Thus applying the previous theorem repeatedly, we see that if $A \subset \mathbb{R}^n$ and $f\colon A \to \mathbb{R}$ is integrable, we may evaluate the integral $\int_A f$ by integrating each of the variables one at a time, in any order that we choose. This freedom to choose the order of integration is useful both theoretically and in practice.

We finish this section by proving the fundamental theorem of calculus, which gives us a method of computing integrals in the case of $\mathbb{R}$.

Lemma 5.7. *(Fundamental theorem of Calculus). Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function. Define $F\colon [a,b] \to \mathbb{R}$ by*

$$F(t) = \int_{[a,t]} f.$$

Then F is differentiable on (a,b) with and $DF(t) = f(t)$.

Proof. Fix $x_0 \in (a,b)$. Then since f is continuous as x_0, we may write $f(x) = f(x_0) + \varepsilon(x)$ where $|\varepsilon(x)| \to 0$ as $x \to x_0$. Suppose that $h > 0$, then we have

$$\begin{aligned}\frac{1}{h}(F(x_0+h) - F(x_0)) &= \frac{1}{h}(\int_{[a,x_0+h]} f - \int_{[a,x_0]} f)\\ &= \frac{1}{h}(\int_{[x_0,x_0+h]} f\\ &= \frac{1}{h}(\int_{[x_0,x_0+h]} (f(x_0)+\varepsilon)\\ &= f(x_0) + \frac{1}{h}(\int_{[x_0,x_0+h]} \varepsilon).\end{aligned}$$

Since $\varepsilon \to 0$ as $x \to x_0$, it follows from the monotonicity of the integral that this term tends to zero as h does. The case $h < 0$ follows in the same way. It follows that F is differentiable, with derivative f. □

Remark 5.8. Observe that this result, which is the standard version of the fundamental theorem of calculus, does not fully answer the quesetion of when functions possess an antiderivative. If f is a differentiable function on $\mathbb{R}$, then it is not necessarily the case that Df is continuous (consider for example $f(x) = x^2\sin(1/x)$). Moreover we know that the integral of a function makes sense so long as the function is continuous away from a set of measure zero, thus one might hope for a stronger connection between the class of functions which possess a derivative and the class of integrable functions.

6. Multilinear algebra

We begin by discussing the content (or n-volume) of n-dimensional versions of a parallelograms: Given a set of n vectors, $v_1, v_2, \ldots, v_n$, we may consider the n-dimensional parallelepiped $P(v_1, v_2, \ldots, v_n)$:

$$\{t_1v_1 + t_2v_2 + \ldots t_nv_n : t_i \in [0,1], \forall i, 1 \leq i \leq n\}.$$

We want to consider the function on n-tuples of vectors $c\colon (\mathbb{R}^n)^n \to \mathbb{R}$ given by $c(v_1, v_2, \ldots, v_n) = \nu(P(v_1, v_2, \ldots, v_n))$.

Example 6.1. We consider the smallest cases, when $n = 1$ or 2. For $n = 1$ there is almost nothing to do: any vector $v \in \mathbb{R}$ is of the form λe_1, and $V(v) = |\lambda|$. Now let $n = 2$, and suppose we have two vectors $v = (a,b)$ and $w = (c,d)$. Then (working for the moment informally) the area of the parallelogram that they span is $|ad-bc|$. (You can see this by cutting up the parallelogram they yield into two triangles say). Notice that in each of the cases we have examined, the content function was the absolute value of a polynomial function, which gives a "signed volume". It turns out that this function is in fact easier to study. For the $n = 2$, let S be this signed area. Then $S(v,w)$ has the following properties:

- $S(v,v) = 0$ for all $v \in \mathbb{R}^2$;
- $S(e_1, e_2) = 1$ where $e_1 = (1,0), e_2 = (0,1)$;
- $S(v_1+v_2, w) = S(v_1, w) + S(v_2, w)$, for all $v_1, v_2, w \in \mathbb{R}^2$;
- $S(v, w_1+w_2) = S(v, w_1) + S(v, w_2)$, for all $v.w_1, w_2 \in \mathbb{R}^2$.

Of course all of these can be checked directly from the formula, but they can also be explained as follows: the first property simply asserts that the degenerate parallelogram consisting of the line segment from 0 to v does not have any area. The second is simply our normalization of area (a choice of units, if you like). Finally, the last two properties follow from the fact that the area is the product of the length of one of the vectors, w say, times the (signed) length of the component of the other vector in the direction perpendicular to w. Since taking the signed length of this component is a linear function, we see that S is linear in the vector v.

Similar reasoning suggests that a signed volume in $\mathbb{R}^3$ should have similar properties. In fact it turns out that the generalization of these properties uniquely determine a function on n-tuples of vectors in $\mathbb{R}^n$. We formalize this with the following definition:

Definition 6.2. A function $A\colon (\mathbb{R}^n)^k \to \mathbb{R}$ is an *alternating k-multilinear map* if

- (*alternating*): $A(v_1, v_2, \dots, v_n) = 0$ whenever at least two of the v_i are equal;
- (*multilinear*): A is linear in each factor: given $v_1, v_2, \dots, v_{i-1}, v_{i+1}, \dots v_n$, and w_1, w_2 we have

$$A(v_1, v_2, \dots v_{i-1}, w_1 + w_2, v_{i+1}, \dots, v_n) = A(v_1, v_2, \dots, v_{i-1}, w_1, v_{i+1}, \dots, v_n) + A(v_1, v_2, \dots, v_{i-1}, w_2, v_{i+1}, \dots, v_n).$$

Notice that the set $\Lambda^k(\mathbb{R}^n)$ of such functions for a fixed integer k is a vector space. For convenience, we also define $\Lambda^0(\mathbb{R}^n) = \mathbb{R}$. These spaces are finite dimensional, as we will soon see.

Lemma 6.3. *Let $v_1, v_2, \dots v_k \in \mathbb{R}^n$ and suppose that $v_i = \sum_{j=1}^n a^i_j e_j$ ($1 \leq i \leq k$). Then if $A \in \Lambda^k(\mathbb{R}^n)$ we have*

$$A(v_1, v_2, \dots, v_k) = \sum_{j_1, j_2, \dots, j_k} (a^1_{j_1} a^2_{j_2} \dots a^k_{j_k}) A(e_{j_1}, e_{j_2}, \dots, e_{j_k}).$$

Proof. We show this by induction on k: for $k = 0$ there is nothing to prove. Suppose now that the result is known for $k-1$. Then the function $(v_1, v_2, \dots, v_{k-1}) \mapsto A(v_1, v_2, \dots, v_{k-1}, v_k)$ is an element of $\Lambda^{k-1}(\mathbb{R}^n)$, and so we obtain

$$A(v_1, v_2, \dots, v_k) = \sum_{j_1, j_2, \dots, j_k} (a^1_{j_1} a^2_{j_2} \dots a^k_{j_k}) A(e_{j_1}, e_{j_2}, \dots, e_{j_{k-1}} v_k).$$

But now, since A is linear in v_k we have

$$A(e_{j_1}, e_{j_2}, \dots, e_{j_{k-1}}, v_k) = \sum_{j_k=1}^n a^k_{j_k} A(e_{j_1}, e_{j_2}, \dots, e_{j_{k-1}}, e_{j_k}).$$

Substituting this into the above expression gives the required result. □

Notice that the proof of the last lemma only used the multilinear property, not the alternating property. Using the alternating property, we see that each of the terms in the formula given by the lemma vanishes unless all the e_{j_k} are distinct. But in fact more is true:

Lemma 6.4. *Let $v_1, v_2, \dots, v_k$ vectors in $\mathbb{R}^n$, and $A \in \Lambda^k(\mathbb{R}^n)$ we have*

$$A(v_1, v_2, \dots, v_{i-1}, v_i, v_{i+1} \dots, v_{j-1}, v_j, v_{j+1}, \dots v_k) = -A(v_1, v_2, \dots, v_{i-1}, v_j, v_{i+1} \dots, v_{j-1}, v_i, v_{j+1}, \dots v_k),$$

for every $1 \leq i < j \leq k$.

Proof. Given vectors $v_1, v_2, \dots, v_k$ consider a pair of indices $i < j$. Then

$$0 = A(v_1, v_2, \dots, v_{i-1}, v_i + v_j, v_{i+1} \dots, v_{j-1}, v_i + v_j, v_{j+1}, \dots v_k)$$

But then using the fact that A is linear in each of the ith and jth entries we find that

$$\begin{aligned} 0 =& A(v_1, v_2, \dots, v_{i-1}, v_i, v_{i+1} \dots, v_{j-1}, v_i, v_{j+1}, \dots v_k) \\ &+ A(v_1, v_2, \dots, v_{i-1}, v_i, v_{i+1} \dots, v_{j-1}, v_j, v_{j+1}, \dots v_k) \\ &+ A(v_1, v_2, \dots, v_{i-1}, v_j, v_{i+1} \dots, v_{j-1}, v_i, v_{j+1}, \dots v_k) \\ &+ A(v_1, v_2, \dots, v_{i-1}, v_j, v_{i+1} \dots, v_{j-1}, v_j, v_{j+1}, \dots v_k) \end{aligned}$$

Using the alternating property for the right-hand side, we see that the first and last of terms are zero, and so the lemma follows. □

A multilinear map satisfying the conclusion of the lemma is said to be *skew-symmetric*. It is easy to check that this property implies the alternating property (because the only number c with $c = -c$ is zero), so the space of alternating multilinear maps is the same as the space of skew-symmetric multilinear maps. Using Lemma 6.4, the formula in the preceeding lemma can be made more precise: each of the terms $A(e_{j_1}, e_{j_2}, \dots, e_{j_k})$ is zero unless all the indices $j_1, j_2, \dots, j_k$ are distinct, and then given a set of distinct indices $\{j_1, j_2, \dots, j_k\}$ if we swap any pair of them, the value of A simply changes sign. In particular, if $k = n$ then the indices $\{j_1, j_2, \dots, j_n\}$ must just be a reordering of the integers $\{1, 2, \dots, n\}$. For $n = 2$ we have shown that the signed area function is a skew-symmetric bilinear function ("bilinear" = 2-multilinear), but we do not as yet know if there are *any* such functions (apart from the zero function) for $n > 2$.

Definition 6.5. A *permutation* of $\{1, 2, \dots, n\}$ is a bijection

$$\sigma \colon \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}.$$

We denote the set of all permutations of the set $\{1, 2, \dots, k\}$ by S_k, the *symmetric group*. Note that the composition of two permutations is clearly a permutation. A permutation which interchanges two elements of the set $\{1, 2, \dots, k\}$ and leaves the remaining $k - 2$ unchanged is called a *transposition*.

Suppose that we have a subset $\{j_1, j_2, \dots, j_r\}$ of $\{1, 2, \dots, k\}$. We may define a permutation σ by setting

$$\sigma(i) = \begin{cases} j_{s+1}, & \text{if } i = j_s, s < r; \\ j_1 & i = j_r; \\ i & 0 \end{cases}$$

We say that σ is a *cycle* of length r, and write $\sigma = (j_1 j_2 \dots j_r)$. Thus a transposition is cycle of length 2. We say two cycles $\sigma_1 = (i_1 i_2 \dots i_t)$ and $\sigma_2 = (j_1 j_2, \dots j_r)$ are *disjoint* if the sets $\{i_1, i_2, \dots, i_t\}$ and $\{j_1, j_2, \dots, j_s\}$ are disjoint. Notice that if σ_1 and σ_2 are disjoint cycles, then they commute, *i.e.* $\sigma_1 \sigma_2 = \sigma_2 \sigma_1$. The next lemma shows that there is a kind of unique factorization of a permutation as a product of cycles.

Lemma 6.6. *Let $\sigma \in S_k$ be a permutation. Then there are disjoint cycles $\gamma_1, \gamma_2, \gamma_l$ such that*

$$\sigma = \gamma_1 \gamma_2 \dots \gamma_l$$

Proof. To find the cycles we introduce a relation on $\{1, 2, \ldots, k\}$ as follows: say that $i \sim j$ if for some integer $r \in \mathbb{Z}$ we have $\sigma^r(i) = j$ (where σ^0 is the identity permutation by definition, and if $n > 0$ we set $\sigma^{-n} = (\sigma^{-1})^n$). Now $i = \sigma^0(i)$, so that $\sim$ is reflexive. Next, if $\sigma^r(i) = j$, then $\sigma^{-r}(j) = i$ so that $\sim$ is symmetric. Finally, if $i \sim j$ and $j \sim p$, then there are integers r, s such that $\sigma^r(i) = j$ and $\sigma^s(j) = p$. But then $\sigma^{r+s}(i) = \sigma^s(\sigma^r(i)) = \sigma^s(j) = p$, and so $i \sim p$, that is, $\sim$ is transitive. Thus $\sim$ is an equivalence relation.

It follows that $\{1, 2, \ldots, k\}$ is the disjoint union of the equivalence classes of $\sim$. Now let $\mathcal{O}$ be an equivalence class. Then if $j \in \mathcal{O}$, it follows $\sigma^s(j) \in \mathcal{O}$ for all $s \in \mathbb{Z}$. Since $\{1, 2, \ldots, k\}$ is finite, $\mathcal{O}$ is also finite, and so at some point we must have $\sigma^s(j) \in \{j, \sigma(j), \sigma^2(j), \ldots, \sigma^{s-1}(j)\}$. Suppose that r is the first s for which this is true. Then we have $\sigma^r(j) = \sigma^p(j)$ for some $p < r$. But then $\sigma^{r-p}(j) = j$, and so if r is minimal, we must have $p = 0$, and hence $\sigma^r(j) = j$. It follows that $\mathcal{O} = \{j, \sigma(j), \ldots, \sigma^r(j)\}$. Let $\{\mathcal{O}_i\}_{1 \leq i \leq l}$ be the equivalence classes of $\sim$ and $j_t \in \mathcal{O}_t$ be an element of $\mathcal{O}_t$, $(1 \leq t \leq l)$, so that $\mathcal{O}_t = \{j_t, \sigma(j_t), \ldots, \sigma^{r_t - 1}(j)\}$ where $r_t = |\mathcal{O}_t|$. Then if γ_i is the cycle $(j_t, \sigma(j_t), \ldots, \sigma^{r_t-1}(j_t))$ we see immediately that $\sigma = \gamma_1 \gamma_2 \ldots \gamma_l$ as required. □

Example 6.7. The lemma is easiest to understand in an example: Suppose that $\sigma \in S_6$ is the permutation which sends $1 \mapsto 4$, $2 \mapsto 1$, $3 \mapsto 5$, $4 \mapsto 2$, $5 \mapsto 6$ and $6 \mapsto 3$. Then if we start with 1 we find that $1 \mapsto 4 \mapsto 2 \mapsto 1$, so that (142) is one of the cycles for σ. Picking 3 as the smallest term not in the cycle we already have, we find that $3 \mapsto 5 \mapsto 6 \mapsto 3$, and so σ is the product $(142)(356) = (356)(142)$.

The number of cycles in the expression for $\sigma \in S_k$ as a product of disjoint cycles (more precisely, the number of equivalence classes for $\sim$ in the previous lemma) is called the *cycle length* of σ, written $c(\sigma)$. We define a map $\varepsilon\colon S_k \to \{\pm 1\}$ by setting

$$\varepsilon(\sigma) = (-1)^{k - c(\sigma)}.$$

We call $\varepsilon(\sigma)$ the *sign* of σ. The existence of non-zero skew-symmetric multilinear functions follows from the fact that the function ε we have just defined is compatible with composition of permutations, in that if $\sigma_1, \sigma_2 \in S_k$ then

$$\varepsilon(\sigma_1 \circ \sigma_2) = \varepsilon(\sigma_1)\varepsilon(\sigma_2). \tag{2}$$

The following lemmas gives a proof of this basic fact.

Lemma 6.8. *Let $\sigma \in S_k$ be a permutation. Then there are transpostions $\tau_1, \tau_2, \ldots, \tau_m$ such that*

$$\sigma = \tau_1 \tau_2 \ldots \tau_m.$$

(where this product is taken to be the identity map from $\{1, 2, \ldots, n\}$ to itself if $m = 0$).

Proof. Since any permutation can written as a product of cycles, it is enough to show that any cycle can be written as a product of transpositions. Let $(j_1, j_2, \ldots, j_s)$ be a cycle. It is easy to check that

$$(j_1, j_2, \ldots, j_s) = (j_1 j_s)(j_1 j_{s-1}) \ldots (j_1, j_2)$$

□

Example 6.9. Let $\sigma = (1234) \in S_4$, the we have

$$(1234) = (14)(13)(12)$$

The general case works in exactly the same way.

Notice that a representation of a permutation as a product of transpositions, in contrast to the representation as a product of disjoint cycles, is far from unique (as a trivial example, if τ is a transposition, then $\tau^2 = 1$, so given any product of transpositions which equals σ, we can add on 2 copies of τ at the end to get a new product of transpositions equaling σ).

Lemma 6.10. *Let $\sigma \in S_k$ and let τ be a transposition. Then $\varepsilon(\sigma\tau) = \varepsilon(\sigma)\varepsilon(\tau)$.*

Proof. It is enough to show that $c(\sigma\tau) = c(\sigma) \pm 1$. Let $\tau = (ij)$ be the presentation of τ as a product of cycles, and let $\sigma = \gamma_1\gamma_2, \ldots \gamma_l$ be the presentation of σ as a product of cycles. . To see this, observe that there are two cases. Either the elements (ij) are in the same cycle γ_r, or they are in two different cycles γ_r, γ_s. In the first case one checks that $\gamma_r\tau$ splits into two disjoint cycles: suppose that $\gamma_r = (i, n_1, n_2, \ldots, n_p, j, m_1, m_2, \ldots, m_q)$, then $\gamma_r\tau = (i, n_1, n_2, \ldots, n_p)(j, m_1, m_2, \ldots, m_q)$. In the second case we have $\gamma_r = (i, n_1, n_2, \ldots, n_p)$, $\gamma_s = (j, m_1, m_2, \ldots, m_q)$ say, and so $\gamma_r\gamma_s\tau = (i, n_1, n_2, \ldots, n_p, j, m_1, m_2, \ldots, m_q)$, is a single cycle. The result follows. □

Combining the two lemmas, we see that ε is compatible with composition of permutations as claimed.

Lemma 6.11. *Let $A \in \Lambda^k(\mathbb{R}^n)$, and let $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$. Then for any $\sigma \in S_k$ we have*

$$A(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(k)}) = \varepsilon(\sigma)A(v_1, v_2, \ldots, v_k).$$

Proof. For σ a transposition, this follows from skew-symmetry. For an arbitrary permutation σ, write is as a product of transpositions, and use induction. □

We now consider again the expression we obtained in Lemma 6.3 for $A \in \Lambda^k(\mathbb{R}^n)$, an alternating multilinear function: for vectors $v_1, v_2, \ldots, v_k$ with $v_i = \sum_{j=1}^n a_j^i e_j$, ($a_j^i \in \mathbb{R}$) we have

$$A(v_1, v_2, \ldots, v_k) = \sum_{j_1, j_2, \ldots, j_k} a_{j_1}^1 a_{j_2}^2 \ldots a_{j_k}^k A(e_{j_1}, e_{j_2}, \ldots, e_{j_k}).$$

Each of the terms $A(e_{j_1}, e_{j_2}, \ldots e_{j_k})$ vanishes unless all the $\{j_1, j_2, \ldots, j_k\}$ are distinct, and moreover, in this case we may find $\sigma \in S_k$ such that the sequence $(\{j_{\sigma(1)}, j_{\sigma(2)}, \ldots, j_{\sigma(k)})$ is increasing. It follows using Lemma 6.11 that if we group together the nonzero terms in the above summation according to the k-element sequences $J = (j_1 < j_2 < \ldots < j_k)$, for $j_s \in \{1, 2, \ldots, n\}$, $(1 \leq s \leq k)$ we have

$$A(v_1, v_2, \ldots, v_k) = \sum_{J \subset \{1,2,\ldots,n\}} A(e_{j_1}, e_{j_2}, \ldots, e_{j_k})\big(\sum_{\sigma \in S_k} \varepsilon(\sigma) a_{j_{\sigma(1)}}^1 a_{j_{\sigma(2)}}^2 \ldots a_{j_{\sigma(k)}}^k\big)$$

Now let, for each J as above, let $D_J \colon (\mathbb{R}^n)^k \to \mathbb{R}$ be the function given by

$$D_J(v_1, v_2, \ldots, v_k) = \sum_{\sigma \in S_k} \varepsilon(\sigma) a_{j_{\sigma(1)}}^1 a_{j_{\sigma(2)}}^2 \ldots a_{j_{\sigma(k)}}^k,$$

so the previous equation shows that for an arbitrary $A \in \Lambda^k(\mathbb{R}^n)$,

$$A = \sum_{J \subset \{1,2,\ldots,n\}} A(e_{j_1}, e_{j_2}, \ldots, e_{j_k}) D_J. \tag{3}$$

Lemma 6.12. *For any J a k-element subset of $\{1, 2, \dots, n\}$ (where $k < n$), the function D_J is a nonzero element of $\Lambda^k(\mathbb{R}^n)$. Moreover the functions D_J as J runs over the sequences $(j_1 < j_2 < \ldots < j_k)$, for $j_s \in \{1, 2, \dots, n\}, (1 \leq s \leq k)$ form a basis of $\Lambda^k(\mathbb{R}^n)$.*

Proof. Let $J = (j_1 < j_2 < \ldots < j_k)$. It is clear that D_J is multilinear, so we need only check that it is alternating and nonzero. To see that it is alternating, let $\tau = (ij)$, and $v_i = \sum_{j=1}^n a^i_j e_j \in \mathbb{R}^n$ $(1 \leq i \leq k)$, then we have

$$\begin{aligned}
D_J(v_{\tau(1)}, v_{\tau(2)}, \dots, v_{\tau(k)}) &= \sum_{\sigma \in S_k} \varepsilon(\sigma) a^1_{j_{\sigma(\tau(1))}} \dots a^k_{j_{\sigma(\tau(k))}} \\
&= \sum_{\rho \in S_k} \varepsilon(\rho\tau) a^1_{j_{\rho(1)}} \dots a^k_{j_{\rho(k)}} \\
&= \varepsilon(\tau) \sum_{\rho \in S_k} \varepsilon(\rho) a^1_{j_{\rho(1)}} \dots a^k_{j_{\rho(k)}} \\
&= -D_J(v_1, v_2, \dots, v_k).
\end{aligned}$$

where in the second equality we set $\rho = \sigma\tau$. To see that the functions D_J form a basis of $\Lambda^k(\mathbb{R}^n)$, notice that if $(v_1, v_2, \dots, v_k) = (e_{i_1}, e_{i_2}, \dots, e_{i_k})$ where $I = (i_1, i_2, \dots, i_k)$ is an increasing sequence of elements of $\{1, 2, \dots, k\}$. then we have

$$D_J(e_{i_1}, \dots, e_{i_k}) = \begin{cases} 1 & \text{if } J = I \\ 0 & \text{if } I \neq J \end{cases}$$

This shows that the D_J are linearly independent, but Equation (3) already shows that they span $\Lambda^k(\mathbb{R}^n)$, so we are done. ☐

Suppose now that $k = n$, then we have shown that $\Lambda^n(\mathbb{R}^n)$ is one dimensional with basis vector $D = D_{(1,2,\dots,n)}$ (as $(1, 2, \dots, n)$ is the only n-term increasing sequence in $\{1, 2, \dots, n\}$). This allows us to define the determinant of a linear map.

If $\alpha \colon \mathbb{R}^n \to \mathbb{R}^n$ is linear, then the function $\alpha^*(D)$ given by

$$\alpha^*(D)(v_1, v_2, \dots, v_n) = D(\alpha(v_1), \alpha(v_2), \dots, \alpha(v_n))$$

is clearly an alternating multilinear map, *i.e.* $\alpha^*(D) \in \Lambda^n(\mathbb{R}^n)$. But since this vector space is one-dimensional, with basis $\{D\}$, we see that

$$\alpha^*(D) = \lambda D,$$

and using Equation (3) above we see that $\lambda = D(\alpha(e_1), \alpha(e_2), \dots, \alpha(e_n))$. Thus we define

$$\det \colon \text{End}(\mathbb{R}^n) \to \mathbb{R}$$

to be the map which sends $\alpha \mapsto D(\alpha(e_1), \alpha(e_2), \dots, \alpha(e_n))$, the *determinant* of α.

Lemma 6.13. *Let $\alpha \in End(\mathbb{R}^n)$, and let $A = (a_{ij})_{1 \leq i,j \leq n}$ be the matrix of α, then*

$$\det(\alpha) = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)}.$$

Moreover if $\alpha, \beta \in End(\mathbb{R}^n)$ then

$$\det(\alpha \circ \beta) = \det(\alpha). \det(\beta).$$

Proof. The first part is immediate from the columns of A are the vectors $\alpha(e_i)$, as i runs from 1 to n. The second part follows from the fact that $(\alpha \circ \beta)^*(D) = \beta^*(\alpha^*(D))$ so that

$$\det(\alpha \circ \beta) = \det(\alpha).\det(\beta).$$

□

Corollary 6.14. *A linear map* $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ *is invertible if and only if* $\det(\alpha) \neq 0$.

Proof. Suppose that $\alpha \circ \beta = Id$. Then since it is easy to see that $\det(Id) = 1$, the previous lemma shows that

$$\det(\alpha)\det(\beta) = 1,$$

and hence $\det(\alpha) \neq 0$. For the converse, suppose that α is not invertible. Then the vectors $\alpha(e_1), \alpha(e_2), \dots, \alpha(e_n)$ are linearly independent. Thus we may find an i such that

$$\alpha(e_i) = \sum_{j \neq i} \lambda_j \alpha(e_j).$$

But then

$$\begin{aligned}
\det(\alpha) &= D(\alpha(e_1), \alpha(e_2), \dots, \alpha(e_n)) \\
&= D(\alpha(e_1), \dots, \sum_{j \neq i} \lambda_j \alpha(e_j), \dots, \alpha(e_n)) \\
&= \sum_{j \neq i} \lambda_j D(\alpha(e_1), \dots \alpha(e_j), \dots \alpha(e_n)) \\
&= 0.
\end{aligned}$$

using the fact that D is alternating and multilinear (in the $j - th$ term in the last sum, $\alpha(e_j)$ occurs twice, and so D vanishes). □

7. Integration and Transformations of $\mathbb{R}^n$

Suppose that $\varphi\colon \mathbb{U} \to \mathbb{V}$ is a transformation between subsets of $\mathbb{R}^n$ (in many examples it will be a diffeomorphism between open sets in $\mathbb{R}^n$). Then φ defines a correspondence between subsets of $\mathbb{R}^n$. Given $A \subset \mathbb{R}^n$, the set of all $x \in \mathbb{R}^n$ for which $\varphi(x) \in A$ is denoted $\varphi^{-1}(A)$. (Notice that this does not require φ to be a bijection). Given a subset $A \subset \mathbb{R}^n$ and any function $f\colon A \to \mathbb{R}^n$, we can define a function $\varphi^*(f)\colon \varphi^{-1}(A) \to \mathbb{R}$ by setting $\varphi^*(f)(x) = f(\varphi(x))$. Clearly this gives a linear map between the spaces of bounded functions on the two sets, $\varphi^*\colon \mathcal{B}(A) \to \mathcal{B}(\varphi^{-1}(A))$.

Example 7.1. Let $P\colon [0, \infty) \times [0, 2\pi] \to \mathbb{R} - \{(0,0)\}$ be the transformation given by

$$(r, s) \mapsto (r\cos(s), r\sin(s)).$$

Then the pair (r, s) gives the polar coordinates of the point $P(r, s)$, and if $f\colon \mathbb{R}^2 - \{0\} \to \mathbb{R}$ is a bounded function, the function $P^*(f)$ is simply the function f thought of as a function of the polar variables, rather than the Cartesian variables. Informally, the question we wish to address is how to integrate f in terms of the polar coordinates rather than the Cartesian coordinates. The transformation P clearly distorts area, so we should not expect the integral of $\int_A f$ to be simply the integral $\int_{P^{-1}(A)} P^*(f)$. It turns out, however, that we can precisely express the way in which area is distorted by P, and obtain nevertheless an expression for $\int_A f$ in terms of $P^*(f)$ and $P^{-1}(A)$.

We want to understand the relation between integrating f over a set A and integrating the function $\varphi^*(f)$ over $\varphi^{-1}(A)$. Before we can do this however, we need to ensure that it makes sense to integrate these objects. For this we use our characterization of integrable functions: in order to know that $\varphi^*(f)$ is integrable when f is, it is enough to check that φ sends sets of measure zero to sets of measure zero. For this we use the following definition: if $R = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$ is a rectangle, then set $d(R) = \max\{|b_i - a_i| : 1 \leq i \leq n\}$, so that, $\nu(R) \leq d(R)^n$. We say that R is an n-cube is all the sides have the same length, that is if $b_i - a_i = b_j - a_j$ for all i, j. For an n-cube R we have $d(R) = \nu(A)^{\frac{1}{n}}$.

Exercise 7.2. Show that in the definition of measure zero, we may assume that the rectangles are in fact n-cubes.

Lemma 7.3. (1) *Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz transformation, that is, there is a $K > 0$ such that $\|\varphi(x) - \varphi(y)\| \leq K\|x - y\|$ for all $x, y \in \mathbb{R}^n$. Then if $A \subset \mathbb{R}^n$, has content zero so does $\varphi(A)$.*

(2) *If $f \in \mathcal{R}(\mathbb{R}^n)$ is Riemann integrable, and φ is invertible with φ^{-1} Lipschitz, then $\varphi^*(f)$ is Riemann integrable.*

Proof. Let A be a set of measure zero. Then given $\varepsilon > 0$ we may find n-cubes $R_1, R_2, \ldots$ such that $A \subset \bigcup_{i=1}^{\infty} R_i$, and $\sum_{i=1}^{\infty} \nu(R_i) < \varepsilon$. It follows from the definitions that we may find an n-cube S_i containing $\varphi(R_i)$ such that $d(S_i) \leq 2Kn^{1/2}d(R_i)$ (to see this, note that for all $x, y \in R$ we have $\|x - y\| \leq d(R)n^{1/2}$. But then we have

$$\varphi(A) \subset \bigcup_{i=1}^{\infty} \varphi(R_i) \subset \bigcup_{i=1}^{\infty} S_i,$$

and $\sum_{i=1}^{\infty} \nu(S_i) = \sum_{i=1}^{\infty} d(S_i)^n < 2^n \varepsilon K^n n^{n/2}$. Since ε was arbitrary, $\varphi(S)$ has measure zero as required.

For the second part of the lemma, recall that a function is in $\mathcal{R}(\mathbb{R}^n)$ if it is zero outside of a bounded set, and is continuous everywhere except on a set of measure zero. Clearly if φ^{-1} is Lipschitz, then $\varphi^*(f)$ is zero outside of a bounded set. Moreover, since a Lipschitz function is continuous, the discontinuities of $\varphi^*(f)$ are contained in the set $\varphi^{-1}(B)$ where B is the set on which f is discontinuous. If ϕ^{-1} is Lipschitz, then the first part of the lemma now ensures that this set has measure zero as required. □

Remark 7.4. The previous lemma also implies that the image of a set of content zero under a Lipschitz map φ has content zero, indeed this follows once we observe that since φ is continuous, it takes compact sets to compact sets.

We now want to establish the relation between the integrals of the functions f and $\varphi^*(f)$. Our goal is the formula:

$$\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^n} |\det(D\phi)|\phi^*(f), \tag{4}$$

for any $f \in \mathcal{R}(\mathbb{R}^n)$ and any diffeomorphism $\phi\colon \mathbb{R}^n \to \mathbb{R}^n$. Note that hidden in this equation is the assertion that the function $|\det(D\phi)|\phi^*(f)$ is integrable. But if $f \in \mathcal{R}(\mathbb{R}^n)$, then f is zero outside of some compact set, K say. Since ϕ is a diffeomorphism, $\|D\phi^{-1}\|$ is bounded on K, and hence by the Mean Value Inequality, ϕ^{-1} is Lipschitz on K. Since the points where f is discontinuous will be in K, the previous lemma shows that $\phi^*(f)$ is integrable. Since $|\det(D\phi)|$ is continuous, it is integrable, and so the product $|\det(D\phi)|\phi^*(f)$ is integrable as claimed.

We will establish this formula by first demonstrating special cases of it. Our first reduction is to observe that if establish Equation (4) for simple functions, then it must also hold for any integrable function.

Lemma 7.5. *Let $\phi\colon \mathbb{R}^n \to \mathbb{R}^n$, and suppose that Equation (4) holds for ϕ and any simple function. Then Equation (4) holds for all $f \in \mathcal{R}(\mathbb{R}^n)$.*

Proof. Suppose that $f \in \mathcal{R}(\mathbb{R}^n)$,Then we have

$$\begin{aligned}\underline{\int}_{\mathbb{R}^n} f &= \sup\{\int_{\mathbb{R}^n} \psi : \psi \in \mathcal{S}(\mathbb{R}^n) : \psi \leq f\}\\ &= \sup\{\int_{\mathbb{R}^n} |\det(D\phi)|\phi^*(\psi) : \psi \in \mathcal{S}(\mathbb{R}^n) : \psi \leq f\}.\end{aligned}$$

But now if $\psi \leq f$ it follows immediately that $|\det(D\phi)|\phi^*(\psi) \leq |\det(D\phi)|\phi^*(f)$, and hence since the lower Riemann integral respects inequalities,

$$\underline{\int}_{\mathbb{R}^n} |\det(D\phi)|\phi^*(\psi) \leq \underline{\int}_{\mathbb{R}^n} |\det(D\phi)|\phi^*(f),$$

and so we obtain

$$\underline{\int}_{\mathbb{R}^n} f \leq \underline{\int}_{\mathbb{R}^n} |D\phi|\phi^*(f).$$

Exactly the same argument with the upper integral shows that we also have

$$\overline{\int}_{\mathbb{R}^n} f \geq \overline{\int}_{\mathbb{R}^n} |\det(D\phi)|\phi^*(f),$$

and hence in the case that f is integrable, we obtain the desired result.

□

Since continuously differentiable function are locally closely approximated by their derivative, we next consider the case when $\varphi = T$, a linear map. Since T is Lipschitz, the previous lemma shows that T^* takes integrable functions to integrable functions. Moreover the derivative of T is just T itself. Thus in this case, the previous equation becomes:

$$\int_{\mathbb{R}^n} |\det(T)|T^*(f) = \int_{\mathbb{R}^n} f \tag{5}$$

Our strategy is to use the fact that the determinant is compatible with composition to reduce to the case of a few simple linear maps.

Definition 7.6. For $1 \leq i \leq n$ and $\lambda \in \mathbb{R}$ let $E_i(\lambda) \in \mathrm{End}(\mathbb{R}^n)$, be the map given by

$$E_i(\lambda)(\mu_1, \mu_2, \ldots, \mu_i, \ldots, \mu_n) = (\mu_1, \mu_2, \ldots, \lambda\mu_i, \ldots, \mu_n).$$

For $1 \leq i \neq j \leq n$ let $E_{ij}(\lambda) \in \mathrm{End}(\mathbb{R}^n)$ be the map given by

$$E_{ij}(\lambda)(\mu_1, \ldots, \mu_i, \ldots, \mu_j, \ldots, \mu_n) = (\mu_1, \ldots, \mu_i, \ldots, \mu_j + \lambda\mu_i, \ldots, \mu_n).$$

We call the linear maps $E_{ij}(\lambda), E_i(\lambda)$ *elementary transformations*. Notice that

$$E_{ij}(\lambda)^{-1} = E_{ij}(-\lambda); \quad E_i(\lambda)^{-1} = E_i(\lambda^{-1}),$$

so that the inverse of an elementary transformation is again an elementary transformation. Note also that $\det(E_{ij}(\lambda)) = 1$ and $\det(E_i(\lambda)) = \lambda$.

Lemma 7.7. *Any linear map $\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ can be expressed as a composition of elementary transformations.*

Proof. This is simply another way of expressing the fact that we can solve linear systems of equations using elementary row operations: the operation of adding a multiple of one row to another corresponds to composing with an $E_{ij}(\lambda)$, the operation of scaling a row is just composing with $E_i(\lambda)$. We show this explicitly in the case $n = 2$ in the next example. (Normally when defining row operations we also have an operation which interchanges two rows, however this can be achieved using a succession of the above operations: if we wish to intechange rows i and j, we can add row i to row j, and then subtract the resulting row j from row i, change the sign of this row i, and finally subtract this new row i from row j(!!).) □

Example 7.8. We consider the $n = 2$ case. If we start with an invertible matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where we assume for convenience that $a \neq 0$ then we may apply elementary transformations as follows:

$$E_1(a^{-1})A = \begin{pmatrix} 1 & b/a \\ c & d \end{pmatrix};$$

$$E_{12}(-c)E_1(a^{-1})A = \begin{pmatrix} 1 & b/a \\ 0 & d - bc/a \end{pmatrix};$$

$$E_2(a/(ad-bc))E_{12}(-c)E_1(a^{-1})A = \begin{pmatrix} 1 & b/a \\ 0 & 1 \end{pmatrix};$$

$$E_{21}(-b/a)E_2(a/(ad-bc))E_{12}(-c)E_1(a^{-1})A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Taking inverses we see that

$$A = E_1(a)E_{12}(c)E_2((ad-bc)/a)E_{21}(b/a).$$

Thus in order to prove that Equation (5) holds, we need only check it for T an elementary transformation.

Lemma 7.9. *Let T be an elementary transformation, and $f \in \mathcal{R}(\mathbb{R}^n)$ a Riemann integrable function. Then*

$$\int_{\mathbb{R}^n} |\det(T)| T^*(f) = \int_{\mathbb{R}^n} f.$$

Proof. As for Fubini's theorem, we first establish the result for simple functions. Suppose that A is a rectangle, say $A = [a_1, b_1] \times \ldots \times [a_n, b_n]$. Then $T^*(\mathbf{1}_A) = \mathbf{1}_{T^{-1}(A)}$. If $T = E_i(\lambda)$, we see immediately that

$$T^{-1}(A) = [a_1, b_1] \times \ldots [\lambda^{-1}a_i, \lambda^{-1}b_i] \times \ldots \times [a_n, b_n],$$

and so the result follows. Thus it remains to check the case where $T = E_{ij}(\lambda)$ for some $\lambda \in \mathbb{R}$ and $1 \leq i \neq j \leq n$. It is clear that this reduces to a calculation in $\mathbb{R}^2$, so we assume that $n = 2$ and $i = 1, j = 2$, and $A = [a_1, b_1] \times [a_2, b_2]$. Then write $x \in \mathbb{R}^2$ as (y, z). Set $f = \mathbf{1}_{T^{-1}(A)}$ so that $f_y\colon \mathbb{R} \to \mathbb{R}$ given by $f_y(t) = f(y, t)$ is such that

$$f_y(t) = \begin{cases} \mathbf{1}_{[a_2 - \lambda y, b_2 - \lambda y]} & \text{if } y \in [a_1, b_1] \\ 0 & \text{otherwise} \end{cases}$$

Thus $g(y) = \int f_y = [(b_2 - \lambda y) - (a_2 - \lambda y)]\mathbf{1}_{[a_1,b_1]} = (b_2 - a_2)\mathbf{1}_{[a_1,b_1]}$. Hence by Fubini's theorem we have

$$\nu(T^{-1}(A)) = \int_{\mathbb{R}^2} f = \int_{\mathbb{R}} g = (b_2 - a_2)\int_{\mathbb{R}} \mathbf{1}_{[a_1,b_1]} = \nu(A).$$

Since $\det(T) = 1$ this establishes the result for $\mathbf{1}_A$. As both sides of Equation (5) are linear in f, this immediately implies that Equation (5) holds for any simple function. Applying Lemma 7.5 we see that the lemma follows. □

Corollary 7.10. *Let T be an arbitrary invertible linear transformation. Then*

$$\int_{\mathbb{R}^n} |\det(T)|T^*(f) = \int_{\mathbb{R}^n} f.$$

Proof. Any linear map T can be written as a product of elementary transformations, say $T = E_1 E_2 \dots E_m$. Then the result follows by induction, using the fact that for any linear maps T_1, T_2 we have $(T_1 \circ T_2)^*(f) = T_2^*((T_1^*)(f))$, and $\det(T_1 T_2) = \det(T_1)\det(T_2)$. □

The inverse function theorem tells us that, at least locally, a diffeomorphism behaves like its derivative at a point, thus it is plausible that Equation (7) should hold. Moreover, by Lemma 7.5 and linearity, it is enough to show that Equation (7) holds for ϕ a diffeomorphism when f is the characteristic function of a rectangle. We begin with a lemma which ensures that, at least locally, the derivative controls the action of ϕ on rectangles. Let $C_r(x) = \{y \in \mathbb{R}^n : |y_i - x_i| < r, 1 \leq i \leq n\}$ be the n-cube centered at x with side lengths $2r$. Thus $\nu(C_r(x)) = (2r)^n$. Let $\|.\|_\infty$ denotes the maximum norm, $\|x\|_\infty = \max\{|x_i| : 1 \leq i \leq n\}$. Then $C_r(x) = \{y \in \mathbb{R}^n : \|y - x\|_\infty < r\}$, that is, $C_r(x)$ is the "open ball" of radius r for this norm. Note that $\|x\|_\infty \leq \|x\|$, and $\|x\| \leq \sqrt{n}\|x\|_\infty$

Definition 7.11. Let $\phi\colon U \to \mathbb{R}^n$ be a continuous homeomorphism, where U is an open subset of $\mathbb{R}^n$. We say that ϕ has *measure derivative* at $x \in \mathbb{R}^n$ if the limit

$$\lim_{d(C)\to 0} \nu(\phi(C))/\nu(C)$$

exists, where C runs over cubes containing x with $d(C) \to 0$. Denote this limit by $\Delta\phi(x)$.

Notice that our results so far show that, for a map ϕ of the form $x \mapsto z + T(x)$ where T is linear, we have $\Delta\phi(x) = \det(T)$ for all x.

Lemma 7.12. *Let $\phi\colon U \to \mathbb{R}^n$ be a diffeomorphism. Then for each $x_0 \in U$ we have*

$$\Delta\phi(x_0) = |\det(D\phi(x_0))|$$

Proof. Let $x_0 \in U$, and fix $\varepsilon > 0$. Since we have established the change of variables formula for a linear map, and the determinant is multiplicative, we may replace ϕ with the function

$$y \mapsto D\phi(x_0)^{-1}(\phi(y - x_0) - \phi(x_0)),$$

and hence assume that $0 = x_0 = \phi(x_0)$ and $D\phi(0) = Id$. Thus we must show that $\Delta\phi(0) = 1$. Since $D\phi$ is continuous on U, given $\varepsilon > 0$ we may find a $\delta > 0$ such that $\|D\phi(x) - Id\| < \varepsilon$ for all $x \in B(0,\delta)$. Then by the mean value inequality for the function $\phi - Id$ we have,

$$\|\phi(x) - x\| = \|(\phi(x) - x) - (\phi(0) - 0)\| \leq \varepsilon\|x\|, \qquad \forall x \in B(0,\delta).$$

Using the inequalities relating $\|.\|_\infty$ and $\|.\|$ we see that this moreover implies that $\|\phi(x) - x\|_\infty \leq \varepsilon\sqrt{n}\|x\|_\infty$ for all $x \in B(0,\delta)$. Now suppose that $C_r(x)$ is cube containing 0 such that $C_r(x) \subset B(0,\delta)$ (this holds for example when $r < \delta/\sqrt{n}$). Then for any $y \in C_r(x)$, since $\|.\|_\infty$ is a norm, and so satisfies the triangle inequality, we find that

$$\begin{aligned}\|\phi(y) - x\|_\infty &\leq \|\phi(y) - y\|_\infty + \|y - x\|_\infty \\ &\leq \varepsilon\sqrt{n}\|y\|_\infty + \|y - x\|_\infty \\ &< 2\sqrt{n}\varepsilon r + r\end{aligned}$$

and hence $\phi(y) \in C_{(1+2\sqrt{n}\varepsilon)r}(x)$. Thus we see that

$$\phi(C_r(x)) \subset C_{(1+2\sqrt{n}\varepsilon)r}(x),$$

Now taking ν of these sets we find that

$$\nu(\phi(C_r(x)))/\nu(C_r(x)) \leq \nu(C_{(1+2\sqrt{n}\varepsilon)r}(x))/\nu(C_r(x)) = (1 + 2\sqrt{n}\varepsilon)^n.$$

Since ε is arbitrary, it follows that

$$\limsup_{r\to 0} \nu(\phi(C_r(x)))/\nu(C_r(x)) \leq 1,$$

as $C_r(x)$ runs through cubes containing 0 with side-lengths tending to zero.

On the other hand, considering ϕ^{-1} in the same manner, we find that given any $\varepsilon > 0$ there is a $\delta > 0$ such that for all n-cubes $C_r(x)$ centered at x containing 0, with diameter less than δ we have $\phi^{-1}(C_r(x)) \subset C_{(1+2\sqrt{n}\varepsilon)r}(x)$, and so $C_r(x) \subset \phi(C_{(1+2\sqrt{n}\varepsilon)r}(x))$. Thus setting $s = (1 + 2\sqrt{n}\varepsilon)r$, we have $\phi(C_s(x)) \supset C_{s(1+2\sqrt{n}\varepsilon)^{-1}}(x)$, whenever s is sufficiently small. Thus as above we get that

$$\nu(\phi(C_s(x)))/\nu(C_s(x)) \geq \nu(C_{s(1+2\sqrt{n}\varepsilon)^{-1}}(x))/\nu(C_s(x)) = (1 + 2\sqrt{n}\varepsilon)^{-n}.$$

Thus we find

$$\liminf_{s\to 0} \nu(\phi(C_s(x)))/\nu(C_s(x)) \geq 1.$$

and so we must have $\lim_{r\to 0} \nu(\phi(C_r(x)))/\nu(C_r(x)) = 1$ as required. □

Lemma 7.13. *Suppose that $\phi\colon U \to \mathbb{R}^n$ is a diffeomorphism from an open set $U \subset \mathbb{R}^n$ and A is an n-cube such that for all $x \in A$ we have $\Delta\phi(x) < M$. Then*

$$\nu(\phi(C))/\nu(C) \leq M$$

for all cubes C contained in A.

Proof. Suppose that we have a cube $C \subset A$ for which $\nu(\phi(C))/\nu(C) > K + \varepsilon$ some $\varepsilon > 0$. Then partitioning C into 2^n cubes of side-lengths half the side-length of C, there must be at least one of the subcubes, C_1 say, for which $\nu(\phi(C_1))/\nu(C_1) > M + \varepsilon$. Subdividing repeatedly, we obtain a nested sequence of cubes $(C_k)_{k\in\mathbb{N}}$ whose diameter tends 0. Let $x \in C$ be their intersection. Then clearly

$$\Delta\phi(x) = \lim_{k\to\infty} \nu(\phi(C_k))/\nu(C_k) \geq M + \varepsilon,$$

which is a contradiction. □

Lemma 7.14. *Let $\phi\colon U \to \mathbb{R}^n$ be a diffeomorphism defined on an open set $U \subset \mathbb{R}^n$. Then if A is rectangle, and $A \subset \phi(U)$, we have*

$$\nu(A) = \int_{\phi^{-1}(A)} \Delta\phi$$

Proof. Clearly we may assume that A is closed, and hence compact. Since ϕ^{-1} is continuous, $\phi^{-1}(A)$ is also compact. and hence we may find a compact set K such that

$$\phi^{-1}(A) \subset \mathring{K} \subset K \subset U.$$

Let $\varepsilon > 0$ be given. By the definition of $\Delta\phi$, for each $x \in K$ there is an $r_x > 0$ such that $|\nu(\phi(C_r(x))/\nu(C_r(x)) - \Delta\phi(x)| < \varepsilon$ for all $r < r_x$. Since $\Delta\phi$ is continuous, and K is compact there is an $r > 0$ which works for all $x \in K$ simultaneously (consider the sets $U_r = \{x \in K : \exists r_x > r\}$ for $r > 0$). Similarly, $\Delta\phi$ is uniformly continuous on K, and so there is a $\delta > 0$ such that $|\Delta\phi(x) - \Delta\phi(y)| < \varepsilon$ for all $x, y \in K$ with $\|x - y\| < \delta$. Suppose also that $|\Delta\phi(x)| < M$ for all $x \in K$.

Cover $\phi^{-1}(A)$ with cubes $C_1, C_2, \ldots, C_k$, such that $C_i = C_{r_i}(x_i) \subset K$, with $r_i \in \mathbb{R}$, $r_i < \min\{r, \delta\}$, $x_i \in \phi^{-1}(A)$ (here we do not assume that the cubes are closed, just that they are disjoint and cover $\phi^{-1}(A)$). Since $\phi^{-1}(A)$ is Jordan measurable, its boundary has content zero, so we may assume that the total content of the cubes intersecting $\partial\phi^{-1}(A)$ is less that ε/M, and so since $\bigcup_{i=1}^k \phi(C_i) \supset A$, it follows from Lemma 7.13 that $0 \leq \sum_{i=1}^k \nu(\phi(C_i)) - \nu(A) < \varepsilon$.

Set

$$\chi_{\mathcal{C}} = \sum_{i=1}^k \nu(\phi(C_{r_i}(x_i)))/\nu(C_{r_i}(x_i))\mathbf{1}_{C_{r_i}(x_i)}.$$

Then clearly $\int \chi_{\mathcal{C}} = \nu(\phi(\bigcup_{i=1}^k C_{r_i}(x_i)))$. Moreover we have for $x \in \bigcup_{i=1}^k C_{r_i}(x_i)$, say $x \in C_{r_p}(x_p)$,

$$|\Delta\phi(x) - \chi_{\mathcal{C}}(x)| \leq |\Delta\phi(x) - \Delta\phi(x_p)| + |\Delta\phi(x_p) - \chi_{\mathcal{C}}(x)| < 2\varepsilon.$$

and therefore

$$|\int \chi_{\mathcal{C}} - \int \Delta\phi| = |\int (\chi_{\mathcal{C}} - \Delta\phi)| \leq \int |\chi_{\mathcal{C}} - \Delta\phi| < 2\varepsilon(\nu(A) + \varepsilon).$$

and so finally

$$\begin{aligned} |\nu(A) - \int \Delta\phi| &\leq |\nu(A) - \int \chi_{\mathcal{C}}| + |\int \chi_{\mathcal{C}} - \int \Delta\phi| \\ &\leq \varepsilon + 2\varepsilon(\nu(A) + \varepsilon) \\ &= \varepsilon(2\nu(A) + 1 + \varepsilon). \end{aligned}$$

Since ε was arbitrary, we are done. □

The general change of variables formula is now an easy consequence:

Theorem 7.15. *Let $\phi\colon U \to \mathbb{R}^n$ a diffeomorphism from an open set $\mathbf{U} \subset \mathbb{R}^n$, and let V be a Jordan measurable set such that $V \subset \phi^{-1}(U)$. Then*

$$\int_V f = \int_{\phi^{-1}(V)} (\Delta\phi)\phi^*(f).$$

Proof. We first establish that if $f \in \mathcal{R}(\mathbb{R}^n)$ is zero outside of $\phi(U)$, then

$$\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^n} (\Delta\phi)\phi^*(f), \tag{6}$$

(where the integrand on the right-hand side is defined to be zero outside U). For $f = \mathbf{1}_A$ where A is a rectangle the previous lemma establishes the result. By linearity, we immediately obtain the result for simple function, and applying Lemma

7.5 we obtain the result for all $f \in \mathcal{R}(\mathbb{R}^n)$. Next note that by definition we have

$$\int_V f = \int f \mathbf{1}_V,$$

and moreover $\phi^*(f.\mathbf{1}_V) = \phi^*(f).\phi^*(\mathbf{1}_V) = \phi^*(f).\mathbf{1}_{\phi^{-1}(V)}$. The equation in the statement of the theorem now follows by applying Equation 6 to the function $f.\mathbf{1}_V$. □

7.1. **Examples and applications of the Change of Variables Formula.** We give some examples of the uses of the Change of Variables formula and Fubini's theorem.

7.1.1. Compute the area of the set $A = \{(x,y) \in \mathbb{R}^2 : 1 \leq xy \leq 2, x^2 \leq y \leq 2x^2\}$.

In this example we are integrating an easy function (the constant function 1) over a complicated domain (the set A). Using the change of variables formula, we can transform this into the integral of a (manageable) function over an easy domain (a rectangle). Examining the conditions, we can try for new coordinates $u = xy$, and $v = y/x^2$, that is, consider the function $\phi\colon \mathbb{R}^2 - \{(0,y) : y \in \mathbb{R}\} \to \mathbb{R}^2$, where $\phi(x,y) = (xy, y/x^2)$, so that $\phi(A) = [1,2] \times [1,2]$. If we can show that ϕ is a diffeomorphism onto its image, then the change of variables formula shows that if ψ is its inverse, we have (since $\psi^{-1}(A) = \phi(A) = [1,2] \times [1,2]$)

$$\begin{aligned}\nu(A) &= \int \mathbf{1}_A = \int \psi^*(\mathbf{1}_A)|\det(D\psi)| \\ &= \int \mathbf{1}_{\psi^{-1}(A)}|\det(D\psi)| = \int_{\psi^{-1}(A)} |\det(D\psi)| \\ &= \int_{[1,2]\times[1,2]} |\det(D\psi)|\end{aligned}$$

But we can solve the equations $u = xy$, $v = y/x^2$ to express x, y as functions of u, v as follows:

$$x = (u/v)^{1/3}; \quad y = (u^2 v)^{1/3},$$

so that ϕ is a diffeomorphism away from where u and v are zero. We could calculate $\det(D\psi)$ directly from the above formulas, but we can also calculate $\det(D\phi)$, as they are inverses. We have

$$D\phi = \begin{pmatrix} y & x \\ -2y/x^3 & 1/x^2 \end{pmatrix},$$

and hence $\det(D\phi(x,y)) = 3y/x^2$. Thus $\det(D\psi(u,v)) = x^2/3y = \frac{1}{3}v^{-1}$.

It follows that

$$\nu(A) = \int_{[1,2]\times[1,2]} \frac{1}{3}v^{-1} = \int_{u\in[1,2]} (\int_{v\in[1,2]} \frac{1}{3}v^{-1}) = \frac{1}{3}\log(2).$$

7.1.2. Let $\Phi\colon (0,\infty)\times(0,2\pi) \to \mathbb{R}^2$ be the map given by $\Phi(s,t) = (s\cos(t), s\sin(t))$. Show that if $B_r = \{x \in \mathbb{R} : \|x\| \leq r\}$, and $f \in \mathcal{R}(\mathbb{R}^n)$ then

$$\int_{B_r} f = \int_{(0,r]\times(0,2\pi)} s\Phi^*(f)(s,t)$$

If $C_r = [-r,r] \times [-r,r]$ show that

$$\int_{C_r} e^{-x^2-y^2} = (\int_{[-r,r]} e^{-x^2})^2.$$

Prove that $\lim_{r\to\infty}\int_{B_r} e^{-x^2-y^2} = \lim_{r\to\infty}\int_{C_r} e^{-x^2-y^2}$, and hence deduce that

$$\int_{\mathbb{R}} e^{-x^2} = \lim_{r\to\infty}\int_{[-r,r]} e^{-x^2} = \sqrt{\pi}.$$

(here the first equality is the *definition* of the left hand side).

We have

$$D\Phi(s,t) = \begin{pmatrix} \cos(t) & -s\sin(t) \\ \sin(t) & s\cos(t) \end{pmatrix},$$

and so $\det(D\phi(x,y)) = s$. If B_r is the disc centered at the origin, then clearly $\Phi^{-1}(B_r) = [0,r]\times[0,2\pi]$. Moreover $\Phi^*(f)(s,t) = e^{-s^2}$, and so by the change of variables formula, we have

$$\int_{B_r} e^{-x^2-y^2} = \int_{[0,r]\times[0,2\pi]} se^{-s^2}$$

But then using Fubini's theorem, we can integrate over t and then s to find that

$$\begin{aligned}\int_{[0,r]\times[0,2\pi]} se^{-s^2} &= \int_{t\in[0,2\pi]} (\int_{s\in[0,r]} se^{-s^2}) \\ &= \int_{t\in[0,2\pi]} [-\frac{1}{2}e^{-s^2}]_0^r \\ &= \int_{t\in[0,2\pi]} \frac{1}{2}(1-e^{-r^2}) \\ &= \pi(1-e^{-r^2}).\end{aligned}$$

On the other hand, using Fubini's theorem we have

$$\begin{aligned}\int_{C_r} e^{-x^2-y^2} &= \int_{x\in[-r,r]} (\int_{y\in[-r,r]} e^{-x^2-y^2}) \\ &= \int_{x\in[-r,r]} e^{-x^2}(\int_{y\in[-r,r]} e^{-y^2}) \\ &= (\int_{t\in[-r,r]} e^{-t^2})^2,\end{aligned}$$

as required.

Finally, if C_r is the cube of side length $2r$ centered at the origin, clearly we have $B_r \subset C_r \subset B_{\sqrt{2}r}$, and so since $e^{-x^2-y^2}$ is positive we have

$$\pi(1-e^{-r^2}) = \int_{B_r} e^{-x^2-y^2} \leq \int_{C_r} e^{-x^2-y^2} \leq \int_{B_{\sqrt{2}r}} e^{-x^2-y^2} = \pi(1-e^{-2r^2}).$$

Thus it is clear taking the limit as $r\to\infty$ we have

$$\lim_{r\to\infty}\int_{C_r} e^{-x^2-y^2} = \lim_{r\to\infty}(\int_{t\in[-r,r]} e^{-t^2})^2 = \pi$$

Since $\int_{t\in[r,-r]} e^{-t^2}$ is clearly positive, it follows that

$$\lim_{r\to\infty}\int_{[-r,r]} e^{-x^2} = \sqrt{\pi}.$$

Remark 7.16. The observant may complain that Φ is *not* a diffeomorphism at $(s,t) = (0,t)$. However, the set $\{(0,t) : 0 \leq t \leq 2\pi]\}$ is a set of content zero, so that neither it (nor it's image under Φ, the origin) contribute to the value of the integrals in the change of variables, thus our slight carelessness does not result in a false calculation.

7.1.3. Calculate the 4-volume of the 4-ball $B_r = \{x \in \mathbb{R}^4 : \|x\| \leq r\}$.
We need to find coordinates to make the set B_r simple to describe. One possibility is the function $\Psi\colon \mathbb{R}^4 \to \mathbb{R}^4$ given by setting $\Psi(s, t_1, t_2, t_3)$ to be

$$t_4(\cos(t_1)\cos(t_3), \sin(t_1)\cos(t_3), \cos(t_2)\sin(t_3), \sin(t_2)\sin(t_3)).$$

You can check that Φ gives a diffeomorphism from

$$(0, 2\pi) \times (0, 2\pi) \times (0, \pi/2) \times (0, r) \to B_r \backslash S$$

where S is a set of content zero. Thus we need to compute the derivative of this transformation, and then integrate the determinant of this over the rectangle $(0, 2\pi) \times (0, 2\pi) \times (0, \pi/2) \times (0, r)$.

The derivative of $D\Phi$ is given by the matrix

$$\begin{pmatrix} -t_4\sin(t_1)\cos(t_3) & 0 & -t_4\cos(t_1)\sin(t_3) & \cos(t_1)\cos(t_3) \\ t_4\cos(t_1)\cos(t_3) & 0 & -t_4\sin(t_1)\sin(t_3) & \sin(t_1)\cos(t_3) \\ 0 & -t_4\sin(t_2)\sin(t_3) & t_4\cos(t_2)\cos(t_3) & \cos(t_2)\sin(t_3) \\ 0 & t_4\cos(t_2)\sin(t_3) & t_4\sin(t_2)\cos(t_3) & \sin(t_2)\sin(t_3) \end{pmatrix}$$

Now the determinant of this matrix is "D" of the column vectors, where D was the basis vector of $\Lambda^n(\mathbb{R}^n)$ we constructed. Since D is multilinear, we can take out scalars: for example in the first row we can take out $t_4\cos(t_3)$. Doing the same thing for the second and third row, we find that $\det(D\Phi)$ is

$$t_4^3\cos(t_3)\sin(t_3)\det\begin{pmatrix} -\sin(t_1) & 0 & -\cos(t_1)\sin(t_3) & \cos(t_1)\cos(t_3) \\ \cos(t_1) & 0 & -\sin(t_1)\sin(t_3) & \sin(t_1)\cos(t_3) \\ 0 & -\sin(t_2) & \cos(t_2)\cos(t_3) & \cos(t_2)\sin(t_3) \\ 0 & \cos(t_2) & \sin(t_2)\cos(t_3) & \sin(t_2)\sin(t_3) \end{pmatrix}$$

Now using the formula for D this expands to:

$$\begin{aligned} t_4^3\cos(t_3)\sin(t_3)\big(-\sin(t_1)\{&-\sin(t_1)\cos^2(t_2)\sin^2(t_3) - \sin(t_1)\sin^2(t_2)\cos(t_3)^2 \\ &- \sin(t_1)\cos^2(t_2)\cos^2(t_3) - \sin(t_1)\sin^2(t_2)\sin^2(t_3)\} \\ &- \cos(t_1)\sin(t_3)\{-\cos(t_1)\sin^2(t_2)\sin(t_3) - \cos(t_1)\cos^2(t_2)\sin(t_3)\} \\ &- \cos(t_1)\cos(t_3)\{-\cos(t_1)\sin^2(t_2)\cos(t_3) - \cos(t_1)\cos^2(t_2)\cos(t_3)\}\big) \end{aligned}$$

(we split up the sum over $\sigma \in S_4$ into groups according to $\sigma(1) = 1, 2, 3$ or 4). This is just

$$\begin{aligned} t_4^3\cos(t_3)\sin(t_3)\big(&\sin^2(t_1)\cos^2(t_2)\sin^2(t_3) + \sin^2(t_1)\sin^2(t_2)\cos^2(t_3) \\ &+ \sin^2(t_1)\cos^2(t_2)\cos^2(t_3) + \sin^2(t_1)\sin^2(t_2)\sin^2(t_3) \\ &+ \cos^2(t_1)\sin^2(t_2)\sin^2(t_3) + \cos^2(t_1)\cos^2(t_2)\sin^2(t_3) \\ &+ \cos^2(t_1)\sin^2(t_2)\cos^2(t_3) + \cos^2(t_1)\cos^2(t_2)\cos^2(t_3)\big) \end{aligned}$$

which (using the fact that $\cos^2(\theta)+\sin^2(\theta)=1$) is

$$t_4^3\cos(t_3)\sin(t_3)\big(\sin^2(t_1)\cos^2(t_2)+\sin^2(t_1)\sin^2(t_2) + \cos^2(t_1)\sin^2(t_2)+\cos^2(t_1)\cos^2(t_2)\big)$$

which in turn simpifies to $t_4^3\cos(t_3)\sin(t_3) = \frac{1}{2}t_4^3\sin(2t_3)$. It follows that the volume of the 4-ball is just

$$\begin{aligned}\int_{(0,2\pi)\times(0,2\pi)\times(0,\pi/2)\times(0,r)} \frac{1}{2}t_4^3\sin(2t_3) &= 4\pi^2(\frac{1}{8}r^4)\int_{t_3\in(0,\pi/2)}\sin(2t_3).\\ &= 4\pi^2(\frac{1}{8}r^4)[-\frac{1}{2}\cos(2t_3)]_0^{\pi/2}\\ &= \frac{1}{2}\pi^2r^4\end{aligned}$$

7.1.4. Finally we calculate volume of the n-dimensional ball.

The n-ball of radius r is the set

$$B_n(r)=\{x\in\mathbb{R}^n : \|x\|\le r\}.$$

You can check that the boundary of $B_n(r)$ has content zero, so that $B_n(r)$ is Jordan measureable. Let $V_n(r)$ be its volume. Considering the diffeomorphism $\rho_r\colon\mathbb{R}^n\to\mathbb{R}^n$ given by $x\mapsto rx$, we see easily that $V_n(r)=r^nV_n(1)$, so we are reduced to computing $V_n(1)$.

Suppose that we know $V_{n-1}(r)$. We calculate $V_n(1)$ using Fubini's theorem: Write $x\in\mathbb{R}^n$ as (x',x_n) where $x'\in\mathbb{R}^{n-1}$. Then if $x\in\mathbb{R}^n$, and we fix x_n, then $x\in B_n(1)$ exactly when $x'\in B_{n-1}(\sqrt{1-x_n^2})$. Therefore we have

$$\begin{aligned}\int \mathbf{1}_{B_n(1)} &= \int_{x_n\in[-1,1]}\int \mathbf{1}_{B_{n-1}(\sqrt{1-x_n^2})}\\ &= \int_{x_n\in[-1,1]} V_{n-1}(\sqrt{1-x_n^2})\\ &= 2V_{n-1}(1)\int_{x_n\in[0,1]}(1-x_n^2)^{n/2}.\end{aligned}$$

To compute this last integral, we use the one-dimensional change of variables formula: let $\sin(t)=\sqrt{1-x_n^2}$.

$\cos(t)dt = -x_n(1-x_n^2)^{-1/2}dx_n = \cos(t)/\sin(t)dx_n$

Then the last integral becomes:

$$2V_{n-1}(1)\int_{x_n\in[0,1]}(1-x_n^2)^{n/2} = 2V_{n-1}(1)\int_{t\in[0,\pi/2]}\sin^{n+1}(t).$$

Finally, we can compute this last integral by induction on n, using integration by parts with the functions $\sin(t)$ and $sin^n(t)$: Let the integral be $I_n=\int_{[0,\pi/2]}\sin^{n+1}(t)$. Then

$$\begin{aligned}\int_{t\in[0,\pi/2]}\sin^{n+1}(t) &= [-\cos(t)\sin^n(t)]_0^{\pi/2} + \int_{[0,\pi/2]} n\sin^{n-1}(t)\cos^2(t)\\ &= n\int_{[0,\pi/2]}\sin^{n-1}(t) - n\int_{[0,\pi/2]}\sin^{n+1}(t)\end{aligned}$$

and hence we find that

$$(n+1)I_n = nI_{n-2}.$$

Since it is easy to see that $I_0 = 1$ and $I_1 = \pi/4$, we see that

$$I_n = \begin{cases} \frac{2k}{2k+1}\frac{2k-2}{2k-1}\dots\frac{2}{3}, & \text{if } n = 2k; \\ \frac{2k+1}{2k+2}\frac{2k-1}{2k}\dots\frac{3}{4}\pi/4, & \text{if } n = 2k+1. \end{cases}$$

we conclude that

$$V_n(r) = \begin{cases} \pi^k r^{2n}/k!, & \text{if } n = 2k; \\ 2^n k!\pi^k r^{2k+1}/n!, & \text{if } n = 2k+1. \end{cases}$$

8. Differential forms

8.1. **The exterior product and pullback.** We want to put a product on the space of alternating multilinear functions. To do this, we use the basis $\{D_J\}$ for $\Lambda^k(\mathbb{R}^n)$ which we defined previously. We start by recall what this basis is, in slightly different fashion from our original definition. First recall that $\Lambda^n(\mathbb{R}^n)$ is one dimensional, spanned by the function D, where

$$D(v_1, v_2, \dots, v_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) a^1_{\sigma(1)} a^2_{\sigma(2)} \dots a^n_{\sigma(n)}.$$

where the scalars a^i_j are given by $v_i = \sum_{j=1}^n a^i_j e_j$, *i.e.* the coordinates of the v_i with respect to the standard basis of $\mathbb{R}^n$.

Given a k-tuple of distinct integers $(j_1, j_2, \dots, j_k)$ where $1 \leq j_1, j_2, \dots, j_k \leq n$, we can define an element D_J of $\Lambda^k(\mathbb{R}^n)$ as follows: we set for vectors $w_1, w_2, \dots, w_k \in \mathbb{R}^n$

$$D_J(w_1, w_2, \dots, w_k) = D(v_1, v_2, \dots, v_n),$$

where

$$v_i = \begin{cases} w_r & \text{if } i = j_r; \\ e_i & \text{otherwise} \end{cases}$$

It then follows that for J an increasing k-tuple (that is, $j_1 < j_2 < \dots < j_k$) the function D_J coincides with the basis function denoted D_J in Section 6 (to see this, you can use Equation (3) of that section). Clearly if we reorder the elements of the k-tuple J, since D is alternating the function D_J will change only by a sign.

Example 8.1. Suppose that $n = 3$ and we have $J = (1, 3)$. Then $D_J = D_{13}$ has

$$D_{13}(v, w) = D(v, e_2, w).$$

Moreover we have, $D_{31}(v, w) = D(w, e_2, v) = -D(v, e_2, w) = -D_{31}(v, w)$

We can now define the product on alternating multilinear forms, which is called the *exterior* or *wedge* product. This is a bilinear map

$$\wedge \colon \Lambda^k(\mathbb{R}^n) \times \Lambda^l(\mathbb{R}^n) \to \Lambda^{k+l}(\mathbb{R}^n)$$

Definition 8.2. Let $I = (i_1, i_2, \dots i_k)$ and $J = (j_1, j_2, \dots j_l)$ be tuples. Define

$$K = (i_1, i_2, \dots, i_k, j_1, j_2, \dots, j_l)$$

and set

$$D_I \wedge D_J = \begin{cases} D_K, & \text{if all the } i_r \text{ and } j_s \text{ are distinct;} \\ 0, & \text{otherwise.} \end{cases}$$

In other words, the wedge product is zero unless the tuples I and J are disjoint (*i.e.* have no term in common) and in that case the product is the alternating function corresponding to the concatenation K of the two tuples.

Notice that it follows immediately from the definition that if $I = (i_1, i_2, \ldots, i_k)$ is a k-tuple of distinct integers between 1 and n, then

$$Di_1 \wedge D_{i_2} \wedge \ldots \wedge D_{i_k} = D_I.$$

Since we know that the functions D_J form a basis of $\Lambda^k(\mathbb{R}^n)$ we can define $\wedge$ on all of $\Lambda^k(\mathbb{R}^n)$ by extending linearly, so that $\wedge$ distributes over addition, *i.e.* so that

$$(\sum_I a_I D_I) \wedge (\sum_J b_J D_J) = \sum_{I,J} a_I b_J D_I \wedge D_J.$$

where the a_I and b_J are scalars.

We next define pullback by a linear map for alternating multilinear functions. We have already seen this for $\Lambda^n(\mathbb{R}^n)$ in our definition of the determinant. Suppose that $\alpha\colon \mathbb{R}^m \to \mathbb{R}^n$. Then if $A \in \Lambda^k(\mathbb{R}^n)$, we define $\alpha^*(A)$ by setting, for $v_1, v_2, \ldots, v_k \in \mathbb{R}^m$,

$$\alpha^*(A)(v_1, v_2, \ldots, v_k) = A(\alpha(v_1), \alpha(v_2), \ldots, \alpha(v_k)).$$

Since α is linear, it is easy to see that $\alpha^*(A)$ lies in $\Lambda^k(\mathbb{R}^m)$ if A lies in $\Lambda^k(\mathbb{R}^n)$, and moreover α^* is evidently linear, so we obtain a linear map

$$\alpha^*\colon \Lambda^k(\mathbb{R}^n) \to \Lambda^k(\mathbb{R}^m).$$

Notice that α^* goes "the other way" to α, as a result it is called the *pullback* associated to α.

Example 8.3. Recall that $\Lambda^k(\mathbb{R}^k)$ is one-dimensional, and using the standard basis of $\mathbb{R}^k$, we have defined a basis vector denoted D in the Section 6, which here for clarity we will denote by $D_{\mathbb{R}^k}$. Given a k-tuple J of integers between 1 and n, we may define $\pi_J\colon \mathbb{R}^n \to \mathbb{R}^k$ by setting for $x = (x_1, x_2, \ldots, x_n)$

$$\pi_J(x) = (x_{j_1}, x_{j_2}, \ldots, x_{j_k}).$$

Then it follows from the definitions that $D_J = \pi_J^*(D_{\mathbb{R}^k})$. Thus we can recover our basis of $\Lambda^k(\mathbb{R}^n)$ from the one alternating multilinear function $D_{\mathbb{R}^k}$ and the family of maps π_J.

The following lemma shows that pullbacks are easy to compute with, because they are compatible with the wedge product.

Lemma 8.4. *Let $\alpha\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear map, and $A \in \Lambda^k(\mathbb{R}^n)$, $B \in \Lambda^l(\mathbb{R}^n)$. Then we have*

$$\alpha^*(A \wedge B) = \alpha^*(A) \wedge \alpha^*(B).$$

Proof. We need only check this for $A = D_I$ and $B = D_J$ where I is a k-tuple and J is an l-tuple. We omit the details of the proof (which would require some more detailed discussion of $\Lambda^k(\mathbb{R}^n)$). □

We will write $\Lambda^*(\mathbb{R}^n)$ for the space consisting of linear combinations of alternating multilinear functions of any degree (this is simply a convenient shorthand, allowing us to write the wedge product say as a map $\Lambda^*(\mathbb{R}^n) \times \Lambda^*(\mathbb{R}^n) \to \Lambda^*(\mathbb{R}^n)$).

8.2. **Smooth differential forms.** We now define differential forms, which will be the objects we can sensibly integrate over surfaces in $\mathbb{R}^n$.

Definition 8.5. A differential k-form on an open subset U of $\mathbb{R}^n$ assigns to each point $x \in U$ an alternating k-multilinear function in smoothly varying way, that is, a differential k-form is a smooth function $\omega\colon U \to \Lambda^k(\mathbb{R}^n)$. Thus ω may be written in the form

$$\omega(x) = \sum_I f_I(x) D_I,$$

where I runs over all subsets of $\{1, 2, \dots, n\}$ of size k, and the condition that ω be smooth is simply that each $f_I\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function. We write $\Omega^k(U)$ for the set (in fact vector space) of differential k-forms on U.

Since $\Lambda^*(\mathbb{R}^n)$ is an algebra under the $\wedge$ product, the corresponding space of all linear combinations of differential forms $\Omega^*(U)$ is also: given α and β in $\Omega^*(\mathbb{R}^n)$ we set

$$(\alpha \wedge \beta)(x) = \alpha(x) \wedge \beta(x).$$

Note that since $\Lambda^0(\mathbb{R}^n) = \mathbb{R}$, the space of zero forms $\Omega^0(U)$ is just the space of smooth functions on U. Since the derivative of a smooth function $f\colon U \to \mathbb{R}$ is a smooth map which assigns to each point $x \in U$ a linear map $\mathbb{R}^n \to \mathbb{R}$, (and any such linear map is alternating), we see that the derivative can be thought of as a map $d\colon \Omega^0(U) \to \Omega^1(U)$. Let $x_i\colon \mathbb{R}^n \to \mathbb{R}$ be the function giving the i-th coordinate of a point $x \in \mathbb{R}^n$. Then we have $dx_i(y) = D_i$ for all $y \in \mathbb{R}^n$, that is, dx_i is the constant 1-form taking the value D_i. Moreover, if $f \in \Omega^0(\mathbb{R}^n)$ then

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \ldots + \frac{\partial f}{\partial x_n} dx_n \in \Omega^1(U).$$

We want to extend d to a map which sends k-forms to $(k+1)$-forms.

Definition 8.6. Let $d^k\colon \Omega^k(U) \to \Omega^{k+1}(U)$ be the linear map defined by

$$d^k(f_I dx_I) = df_I \wedge dx_I.$$

for each k-subset $I \subset \{1, 2, \dots, n\}$.

(When there is no possibility for confusion, we usually drop the superscript k. We denote the derivative by d when dealing with real-valued functions, where it coincides with d^0, while we will write D for the derivative of functions taking values in $\mathbb{R}^n$.) The map d is known as the *exterior derivative*. Its basic properties of the map d are as follows:

Lemma 8.7. *If $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^l(U)$ then*

(1)
$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta.$$

(2) $d \circ d(\alpha) = 0$.

Proof. For the first statement, it suffices to check the case where $\alpha = f_I dx_I$ and $\beta = g_J dx_J$. Then

$$\begin{aligned} d(\alpha \wedge \beta) &= d(f_I g_J dx_I \wedge dx_J) \\ &= (f_I dg_J + g_J df_I) \wedge dx_i \wedge dx_J \\ &= (d(f_I)dx_I) \wedge (g_J dx_J) + f_I(dg_J) \wedge dx_I \wedge dx_J \\ &= (d(f_I)dx_I) \wedge (g_J dx_J + (-1)^k f_I dx_I \wedge (dg_J) \wedge dx_J \\ &= d\alpha \wedge (\beta) + (-1)^k \alpha \wedge d\beta. \end{aligned}$$

For the second part, we again need only check on the forms $f_I dx_I$. But then

$$\begin{aligned} d \circ d(f_I dx_I) &= d(df_I \wedge dx_I) \\ &= (d^2(f_I) \wedge dx_I) - (df_I) \wedge d(dx_I) \\ &= (d^2(f_I)) \wedge dx_I, \end{aligned}$$

since $d(dx_I) = 0$ by definition. Thus we are reduced to checking that $d(df_I) = 0$. But now

$$\begin{aligned} d(df_I) &= d(\sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i) \\ &= \sum_{i=1}^n (\sum_{j=1}^n \frac{\partial}{\partial x_j}(\frac{\partial f}{\partial x_i}) dx_j) \wedge dx_i \\ &= \sum_{1 \leq i,j \leq n} (\frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial^2 f}{\partial x_j \partial x_i}) dx_i \wedge dx_j. \end{aligned}$$

But this last expression vanishes by the symmetry of mixed partial derivatives (thus we needed the f_I to be at least $\mathcal{C}^{(2)}$). □

Remark 8.8. The operator $d \colon \Omega^k(U) \to \Omega^{k+1}(U)$ is determined by the conditions:

(1) d is linear, that is $d(\alpha + \beta) = d\alpha + d\beta$,
(2) d is an antiderivation: $d(\alpha \wedge \beta) = d(\alpha) \wedge \beta + (-1)^k \alpha \wedge d\beta$ for $\alpha \in \Omega^k(U)$.
(3) $d \circ d \colon \Omega^k(U) \to \Omega^{k+2}(U)$ is zero.
(4) On $\Omega^0(U)$, the space of smooth functions, d is the derivative.

One of our goals in this section is to try and understand what it means to integrate over a surface – for example a sphere inside $\mathbb{R}^3$. Indeed notice that we don't yet have a way to define the area of a curved surface like the sphere. It will turn out that finding the area of a sphere in $\mathbb{R}^3$ will require will need to integrate a smooth 2-form over the sphere. To do this we need to understand how forms behave under smooth maps.

Let $\psi : \mathbb{R}^k \to \mathbb{R}^n$ be a smooth map. Recall that $\Omega^0(U)$ is simply the set of smooth functions on U an open subset of $\mathbb{R}^n$. Now given $f \in \Omega^0(U)$ the map ψ allows us to define an element of $\Omega^0(\psi^{-1}(U))$ by setting $\psi^*(f)(x) = f(\psi(x))$. We want to generalize this to obtain a map

$$\psi^* \colon \Omega^p(U) \to \Omega^p(\psi^{-1}(U)).$$

Thus we need to smoothly assign to a point $x \in \psi^{-1}(U)$ an alternating p-multilinear function on $\mathbb{R}^m$. Now $\omega(\phi(x))$ is an alternating p-multilinear function on $\mathbb{R}^n$, which unfortunately is not the vector space on which we need an alternating multilinear function. However, $D\psi(x)$ gives us a linear map $\mathbb{R}^m \to \mathbb{R}^n$ and so we can obtain an alternating multilinear function on the correct vector space by using $D\psi(x)$ to pullback the function $\omega(\psi(x))$ to $\mathbb{R}^m$:

Definition 8.9. Given $\psi\colon \mathbb{R}^k \to \mathbb{R}^n$ and $\alpha \in \Omega^p(U)$ let $\psi^*(\alpha)$ be given at $t \in \psi^{-1}(U)$ by setting, $\psi^*(\alpha)(t) = (D\psi(t))^*(\alpha(\psi(t)))$, that is,

$$\psi^*(\alpha)(t)(v_1, v_2, \ldots, v_p) = \alpha(\psi(t))(D\psi(t)(v_1), D\psi(t)(v_2), \ldots, D\psi(t)(v_p)),$$

$v_1, v_2, \ldots v_p \in \mathbb{R}^k$. Notice that we need ψ to be at least differentiable in order to be able to pull back at all. Since ψ is infinitely differentiable, it is straightforward to check (though elaborate to write out explicitly) that $\psi^*(\alpha)$ is a smooth form if α is (see the example below).

We now show that ψ^* is has all the compatibilities we could want.

Lemma 8.10. *Let $\alpha, \beta \in \Omega^*(U)$, and $\psi : \mathbb{R}^k \to \mathbb{R}^n$ be a smooth map. Then we have*

(1) $\psi^*(\alpha \wedge \beta) = \psi^*(\alpha) \wedge \psi^*(\beta)$
(2) $d(\psi^*(\alpha)) = \psi^*(d\alpha)$.

Proof. For the first part notice that the wedge product is defined pointwise, and so the result follows immediately from the corresponding statement for pullback of alternating multilinear functions by liner maps.

For the second part, we first prove it for functions, *i.e.* for $f \in \Omega^0(U)$. In this case for $x \in \psi^{-1}(U)$ and $v \in \mathbb{R}^k$, we have

$$d(\psi^*(f))(x)(v) = d(f \circ \psi)(x)(v) = df(\psi(x)) \circ d\psi(x)(v) = \psi^*(df)(x)(v).$$

by the chain rule. Now given a form $\alpha = f dx_I$ where $I = (i_1 < i_2 < \ldots < i_k)$, we see that if $\psi = (\psi_1, \psi_2, \ldots, \psi_n)$ we have

$$\begin{aligned} d\psi^*(f dx_I) &= d(\psi^*(f)\psi^*(dx_{i_1}) \wedge \ldots \wedge \psi^*(dx_{i_k})) \\ &= d(\psi^*(f) d\psi_{i_1} \wedge \ldots \wedge d\psi_{i_k}), \end{aligned}$$

using the result for 0-forms, since $\psi^*(x_i) = \psi_i$. But then using Lemma 8.7 and the result for the 0-form f we see that

$$\begin{aligned} d(\psi^*(f) d\psi_{i_1} \wedge \ldots \wedge d\psi_{i_k}) &= d(\psi^*(f)) \wedge d\psi_{i_1} \wedge \ldots \wedge d\psi_{i_k} \\ &= \psi^*(df) \wedge d\psi_{i_1} \wedge \ldots \wedge d\psi_{i_k} \\ &= \psi^*(df \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k}), \end{aligned}$$

where in the last equality we once again used the result for 0-forms. □

Example 8.11. Consider the special case where $\psi\colon \mathbb{R}^n \to \mathbb{R}^n$. Then if U is an open subset of $\mathbb{R}^n$ and $\alpha \in \Omega^n(U)$, we want to compute what $\psi^*(\alpha)$ is. We may write $\alpha = f dx_1 \wedge dx_2 \wedge \ldots dx_n$ where f is a smooth function on U. But then

$$\begin{aligned} \psi^*(\alpha)(x) &= f(\psi(x))\psi^*(dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n) \\ &= f(\psi(x))\text{det}(D\psi(x)) dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n. \end{aligned}$$

where the second equality follows from our definition of det and the definition of pullback. The close relation of this expression to the change of variables formula is what will allow us to define the integral of a k-form on an orientable k-surface.

Example 8.12. Suppose that $\psi\colon \mathbb{R}^2 \to \mathbb{R}^3$ so that

$$\psi(t_1, t_2) = (\psi_1(t_1, t_2), \psi_2(t_1, t_2), \psi_3(t_1, t_2)),$$

and α be the 2-form $dx_1 \wedge dx_3$. Then using the above properties of the pullback ψ^* we see that

$$\begin{aligned}\psi^*(dx_1 \wedge dx_3) &= \psi^*(dx_1) \wedge \psi^*(dx_3)\\ &= d(\psi^*(x_1)) \wedge d(\psi^*(x_3)\\ &= d\psi_1 \wedge d\psi_3\\ &= (\frac{\partial \psi_1}{\partial t_1}dt_1 + \frac{\partial \psi_1}{\partial t_2}dt_2) \wedge (\frac{\partial \psi_3}{\partial t_1}dt_1 + \frac{\partial \psi_3}{\partial t_2}dt_2)\\ &= (\frac{\partial \psi_1}{\partial t_1}\frac{\partial \psi_3}{\partial t_2} - \frac{\partial \psi_3}{\partial t_1}\frac{\partial \psi_1}{\partial t_2})dt_1 \wedge dt_2.\end{aligned}$$

Example 8.13. We define a 2-form ν on $\mathbb{R}^3$ as follows: given $x \in \mathbb{R}^3$ let

$$\nu(x)(v, w) = \det \begin{pmatrix} x_1 & v_1 & w_1 \\ x_2 & v_2 & w_2 \\ x_3 & v_3 & w_3 \end{pmatrix}$$

Expanding this by the first column we see that with respect to the forms $\{dx_i \wedge dx_j : 1 \leq i, j \leq 3\}$ the form is

$$\nu(x) = x_1 dx_2 \wedge dx_3 - x_2 dx_1 \wedge dx_3 + x_3 dx_1 \wedge dx_2,$$

then the same kind of calculation as the previous example shows that $\psi^*(\omega)$ is

$$\begin{aligned}\Big(\psi_1(\frac{\partial \psi_2}{\partial t_1}\frac{\partial \psi_3}{\partial t_2} - \frac{\partial \psi_3}{\partial t_1}\frac{\partial \psi_2}{\partial t_2}) - \psi_2(\frac{\partial \psi_1}{\partial t_1}\frac{\partial \psi_3}{\partial t_2} - \frac{\partial \psi_3}{\partial t_1}\frac{\partial \psi_1}{\partial t_2})&\\ + \psi_3(\frac{\partial \psi_1}{\partial t_1}\frac{\partial \psi_2}{\partial t_2} - \frac{\partial \psi_2}{\partial t_1}\frac{\partial \psi_1}{\partial t_2})\Big)dt_1 \wedge dt_2.&\end{aligned}$$

Hence if, say, $\psi(t_1, t_2) = (\cos(t_2)\sin(t_1), \sin(t_2)\sin(t_1), \cos(t_1))$ then we find

$$\psi^*(\nu) = \sin(t_1) dt_1 \wedge dt_2.$$

9. Integration of forms

If U is an Jordan measurable subset of $\mathbb{R}^n$ and α is a differential n-form on U which is zero outside a rectangle, then there is a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$ such that $\alpha = f dx_1 \wedge \ldots \wedge dx_n$. We define

$$\int_U \alpha = \int_U f,$$

i.e. the integral of f over U. Suppose that $\psi\colon V \to U$ is a diffeomorphism between open subsets of $\mathbb{R}^n$ then

$$\begin{aligned}\int_V \psi^*(\alpha) &= \int_V f(\psi(x)) \det(D\psi(x)) dx_1 \wedge \ldots \wedge dx_n\\ &= \int_V f(\psi(x)) \det(D\psi(x)) = \pm \int_U f\\ &= \pm \int_U \alpha,\end{aligned}$$

using the change of variables formula in the second line. Thus we see that

$$\int_V \psi^*(\alpha) = \pm \int_U \alpha \tag{7}$$

with the sign $\pm$ according as ψ preserves or reserves the orientation of $\mathbb{R}^n$.

Definition 9.1. Let I^k be the unit cube $[0,1]^k \subset \mathbb{R}^k$. A *parametrized surface* (S,ψ) is a smooth map $\psi\colon I^k \to \mathbb{R}^n$, such that $\psi(I^k) = S$. We say that ψ is a parametrization of S. We make a similar definition for a surface parametrized by the open cube $\mathring{I}^k = (0,1)^k$.

The way the integral of a k-form on $\mathbb{R}^k$ behaves under change of coordinates suggests that we should be able to define a consistent notion of the integral of a differential k-form over a parametrized k-surface (once we are careful about signs).

Definition 9.2. Let (S,ψ) be a parametrized surface, and let $\omega \in \Omega^k(\mathbb{R}^n)$ be a smooth k-form. Then we define

$$\int_S \omega = \int_{I^k} \psi^*(\omega).$$

One can then show, using Equation (7), that the definition of the integral is independent of the parametrization ψ provided the derivative of ψ has full rank at each point of C^k (again up to sign). Thus once we make a choice of signs, (this choice is known as an orientation for the surface) the exclusion of the parametrization ψ from the notation $\int_S \omega$ is justified.

Remark 9.3. This lst definition is one of the main reasons for considering differential forms. Notice that if we had a *function* rather than differential form, we would not be able to integrate it over a surface in a sensible way: given a parametrized surface (S,ψ) we could define

$$\int_{S,\psi} f = \int_{I^k} \psi^*(f),$$

but then by changing the parametrization we will obtain a completely different answer – the difference in the case of differential forms is that pullback of forms and the change of variables formula for integration match up perfectly (once we are careful about signs).

We end with a brief discussion of the calculation of area for surfaces. A surface which happened to lie inside a plane in $\mathbb{R}^3$ could easily be assigned an area, simply by identifying the plane with $\mathbb{R}^2$ and using our notion of area in $\mathbb{R}^2$. Clearly however, many surfaces in $\mathbb{R}^3$ do not lie in any such plane. Consider for example the sphere $S = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1\}$, which clearly does not lie in any plane. Suppose nevertheless we took a very small patch P of the sphere, and picked a point $x \in P$. Then there is a well-defined tangent space H_x to the sphere at the point (indeed it is just the plane perpendicular to the vector x), so that we can project the patch P to this plane H_x and take the area of the resulting subset of H_x (by identifying H_x with $\mathbb{R}^2$ as proposed above). If the patch P was small enough, then the area we compute should be close to the area of the patch itself. Thus by cutting S up into many small pieces, computing the area of the projection of each piece to a tangent plane, and adding the result, we obtain what should be an approximation to the area of the sphere S. Thus letting the patches we use get smaller and smaller we can hope that there is a well defined limit we approach, and then take this limit as the rigorous defintion of the area of the sphere.

But what would that limit be? Given $x \in S$, the operation of projecting to the tangent plane and taking (signed) area defines for us an element of $\Lambda^2(\mathbb{R}^3)$, and so this defines a 2-form on S. In fact it is not too hard to see that the form ν in question is just the one considered in Example (8.13). Then one can show that the limit of the procedure described tends to the integral

$$\int_S \nu.$$

The form ν described above is called the *area form* (or in general *volume form*).

Let's make this all explicit for a surface S in $\mathbb{R}^3$ of the form $\{x \in \mathbb{R}^3 : f(x) = 0\}$, where $f\colon \mathbb{R}^3 \to \mathbb{R}$ is a smooth function, and we assume that $df \neq 0$ for all x with $f(x) = 0$. In this situation the tangent space at a point $x \in S$ is just the plane given by $df(x)(v) = 0$ (recall that $df(x) \in \Lambda^1(\mathbb{R}^3)$ is a nonzero linear map from $\mathbb{R}^3$ to $\mathbb{R}$, so its kernel is a plane in $\mathbb{R}^3$). There is a slight ambiguity here, in that we normally picture the tangent plane sitting on the surface at the point $x \in S$, while the equation $df(x)(v) = 0$ defines a plane through the origin, in fact one can either think of the tangent space as passing through the origin in which case it is given by the equation $df(x)(v) = 0$, or as the plane passing through the point x, in which case it is given by the equation $df(x)(v) = df(x)(x)$.

The linear map $df(x)$ can be thought of as a vector in $\mathbb{R}^3$ (with associated linear map $\mathbb{R}^3 \to \mathbb{R}$ given by taking the dot product $v \mapsto df(x) \cdot v$). Thus the 2-form ν given by projecting to the tangent space, and then taking signed area can be expressed, just as for the sphere in Example 8.13, as

$$\nu(x)(v, w) = \|df(x)\|^{-1} \det \begin{pmatrix} \frac{\partial f}{\partial x_1} & v_1 & w_1 \\ \frac{\partial f}{\partial x_2} & v_2 & w_2 \\ \frac{\partial f}{\partial x_3} & v_3 & w_3 \end{pmatrix}.$$

Now suppose that we have a parametrization $\psi\colon I^2 \to S$, and write $t = (t_1, t_2)$ for a point in I^2. From the definition of pullback we see that

$$\psi^*(\nu)(t) = \|df(x)\|^{-1} \det \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial \psi_1}{\partial t_1} & \frac{\partial \psi_1}{\partial t_2} \\ \frac{\partial f}{\partial x_2} & \frac{\partial \psi_2}{\partial t_1} & \frac{\partial \psi_2}{\partial t_2} \\ \frac{\partial f}{\partial x_3} & \frac{\partial \psi_3}{\partial t_1} & \frac{\partial \psi_3}{\partial t_2} \end{pmatrix} dt_1 \wedge dt_2$$

Now consider the matrix in the above formula. This can be written as

$$B = B(t) = \begin{pmatrix} df(\psi(t)) & D\psi(t) \end{pmatrix},$$

that is, it is the matrix whose first column is df and whose second and third column together give the matrix of $D\psi(t)$. Since by defintion $f(\psi(t_1, t_2)) = 0$, it follows by the chain rule that

$$\sum_{j=1}^{3} \frac{\partial f}{\partial x_j}(\psi(t)) \frac{\partial \psi_j}{\partial t_i}(t) = 0,$$

for $i = 1, 2$. Using this, we find that

$$B^t.B = \begin{pmatrix} \|df(\psi(t))\|^2 & 0 \\ 0 & D\psi^t(t) D\psi(t) \end{pmatrix},$$

and thus $\det(B^tB) = \|df(\psi(t))\|^2 \det(D\psi^t(t)D\psi(t))$. On the other hand, we also have

$$\det(B^t(t).B(t)) = \det(B^t(t))\deg(B(t)) = \det(B(t))^2$$

and so we find that, if we can ensure $\det(B) > 0$, then

$$\psi^*(\nu)(t) = \sqrt{\det(D\psi(t)^t D\psi(t)}dt_1 \wedge dt_2.$$

Hence the area of the surface S is given by

$$\int_{I^2} \psi^*(\nu) = \int_{I^2} \sqrt{D\psi(t)^t D\psi(t)}.$$

Example 9.4. The parametrization of the sphere $\psi\colon I^2 \to \mathbb{R}^3$ given by

$$(t_1, t_2) \mapsto (\cos(2\pi t_1)\sin(\pi t_2), \sin(2\pi t_1)\sin(\pi t_2), \cos(\pi t_2))$$

has

$$D\psi(t) = \begin{pmatrix} -2\pi\sin(2\pi t_1)\sin(\pi t_2) & \pi\cos(2\pi t_1)\cos(\pi t_2) \\ 2\pi\cos(2\pi t_1)\sin(\pi t_2) & \pi\sin(2\pi t_1)\cos(\pi t_2) \\ 0 & -\pi\sin(\pi t_2) \end{pmatrix}$$

Thus we can compute $D\psi(t)^t D\psi(t)$ to be

$$\begin{pmatrix} 4\pi^2\sin^2(\pi t_2) & 0 \\ 0 & \pi^2 \end{pmatrix}$$

The square root of the determinant of this matrix is $2\pi^2\sin(\pi t_2)$, and integrating this over the square $[0,1]^2$ we get

$$\text{area}(S) = \int_{[0,1]^2} 2\pi^2\sin(\pi t_2) = 4\pi$$

10. Summary of differential forms

We summarize here the results on differential forms that we use in the study of integration on surfaces.

(1) The wedge product: Given $A \in \Lambda^k(\mathbb{R}^n)$ and $B \in \Lambda^l(\mathbb{R}^n)$ we define an alternating multilinear function $A \wedge B \in \Lambda^{k+l}(\mathbb{R}^n)$ using the basis functions D_J.

(2) Definition of differential forms: For an open set $U \subset \mathbb{R}^n$ we define $\Omega^k(U)$ to be the space of smooth (infinitely differentiable) functions $\omega\colon U \to \Lambda^k(\mathbb{R}^n)$.

(3) The pullback: Given a linear map $T\colon \mathbb{R}^m \to \mathbb{R}^n$ there is an associated linear map $T^*\colon \Lambda^k(\mathbb{R}^n) \to \Lambda^k(\mathbb{R}^m)$ called the *pullback* by T. When $n = m = k$ this gives the determinant of T. For $\psi\colon U \to V$ a smooth map between subsets $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$, we can define a pullback $\psi^*\colon \Omega^k(V) \to \Omega^k(U)$, using the derivative $D\psi(x)$ to pullback the alternating multilinear function $\omega(\psi(x))$ for each $x \in U$.

(4) The exterior derivative: Given a k-form $\omega \in \Omega^k(U)$ we define a $(k+1)$-form $d\omega$ by setting, for $\omega = f_I dx_I$,

$$d\omega = df_I \wedge dx_I,$$

and extending linearly. Then d has the following properties:

(a) $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta$.

(b) $d \circ d(\alpha) = 0$.

(c) $\psi^*(\alpha \wedge \beta) = \psi^*(\alpha) \wedge \psi^*(\beta)$

(d) $d(\psi^*(\alpha)) = \psi^*(d\alpha)$.

(5) The integral of a n-form over a cube: If $\omega \in \Omega^n(A)$ is a smooth n-form defined on a rectangle $A \subset \mathbb{R}^n$, we may write $\omega = f dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n$ where f is a smooth function. We define

$$\int_A \omega = \int_A f.$$

(6) The integral of a k-form over a k-surface: If $S \subset \mathbb{R}^n$ is a parametrizable k-surface, and $\psi \colon I^k \to \mathbb{R}^n$ is a parametrization (where I^k is a cube in $\mathbb{R}^k$), then for $\omega \in \Omega^k(\mathbb{R}^n)$ we define

$$\int_S \omega = \int_{I^k} \psi^*(\omega)$$

where the form $\psi^*(\omega) \in \Omega^k(I^k)$, and so the right-hand side has already been defined.