

## Important definitions and results

### 1. IMPLICIT AND INVERSE FUNCTION THEOREMS

A single equation in  $n$  unknowns can always be written

$$f(x_1, \dots, x_n) = 0$$

for some function  $f : \mathbf{R}^n \rightarrow \mathbf{R}$ . When we talk about ‘solving’ such an equation, we usually mean solving for one of the variables, say  $x_1$ , in terms of the others. If  $f$  is linear, this is straightforward. That is,

$$a_1x_1 + \dots + a_nx_n = 0 \quad \text{implies that} \quad x_1 = -\frac{a_2x_2 + \dots + a_nx_n}{a_1},$$

provided  $a_1 \neq 0$ . Moreover,  $x_1$  depends quite nicely (i.e. linearly) on  $x_2, \dots, x_n$ . If  $f$  is non-linear it’s much harder, practically speaking, to solve the equation. However, the Implicit Function Theorem gives a criterion for when one can solve it at least in principle. Roughly speaking, the theorem says that if we have one point  $\mathbf{a} = (a_1, \dots, a_n)$  such that  $f(\mathbf{a}) = 0$ , and the linearization of  $f$  exists and is ‘good’ at  $\mathbf{a}$ , then we can solve  $f(x_1, \dots, x_n) = 0$  for  $x_1$  in terms of  $(x_2, \dots, x_n)$  at least near  $\mathbf{a}$ . Just as importantly, the theorem says that  $x_1$  ‘varies nicely’ with  $(x_2, \dots, x_n)$ . This last fact is the basis for many numerical algorithms that approximate solutions to non-linear equations.

**Definition 1.1.** A neighborhood of a point  $\mathbf{a} \in \mathbf{R}^n$  is an open set  $U \subset \mathbf{R}^n$  that contains  $\mathbf{a}$ .

**Theorem 1.2** (Implicit Function Theorem, scalar-valued case). Suppose that  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is  $C^1$  and that  $\mathbf{a} = (a_1, \dots, a_n) \in \mathbf{R}^n$  satisfies

- $f(\mathbf{a}) = 0$ ;
- $\frac{\partial f}{\partial x_1}(\mathbf{a}) \neq 0$ ;

Then there are neighborhoods  $I \subset \mathbf{R}$  of  $a_1$  and  $U \subset \mathbf{R}^{n-1}$  of  $(a_2, \dots, a_n)$  such that for any  $(x_2, \dots, x_n) \in U$ , there exists a unique point  $x_1 = \phi(x_2, \dots, x_n) \in I$  satisfying

$$f(x_1, \dots, x_n) = 0.$$

Moreover, the ‘solution function’  $\phi : U \rightarrow I$  is also  $C^1$ .

Several remarks are in order here:

- Since  $\phi$  is  $C^1$ , one can differentiate both sides of  $f(\phi(x_2, \dots, x_n), x_2, \dots, x_n) = 0$  to compute the partial derivatives of  $\phi$ . The result is

$$\frac{\partial \phi}{\partial x_j} = -\frac{\partial f / \partial x_j}{\partial f / \partial x_1}.$$

- Even when the hypothesis  $\frac{\partial f}{\partial x_1}(\mathbf{a}) \neq 0$  fails, it’s sometimes still possible to solve  $f(x_1, \dots, x_n) = 0$  for  $x_1$ . However, the previous remark almost guarantees that in this case the solution function won’t vary as nicely (i.e. differentiably) with  $(x_2, \dots, x_n)$ .
- Geometrically speaking the Implicit Function Theorem tells us that ‘locally’ (i.e. near  $\mathbf{a}$ ) the level set  $\{f = 0\}$  (Shifrin writes this set as  $f^{-1}(0)$ ) can also be expressed as the graph  $\{x_1 = \phi(x_2, \dots, x_n)\}$  of a  $C^1$  function.
- Of course, one could always try to solve  $f(x_1, \dots, x_n) = 0$  for some variable other than the first one. If you don’t care which variable you solve for, then instead of  $\frac{\partial f}{\partial x_1}(\mathbf{a}) \neq 0$ , you can get by with the weaker condition  $\nabla f(\mathbf{a}) \neq \mathbf{0}$ .

- Similarly, the Implicit Function Theorem applies to equations  $f(\mathbf{x}) = b$  with a constant  $b \in \mathbf{R}$  different from 0.

Taken together, the last three remarks give us a reasonable way to say what we mean by a ‘smooth  $n - 1$  dimensional subset of  $\mathbf{R}^n$ ’.

**Definition 1.3.** Let  $U \subset \mathbf{R}^n$  be open,  $f : U \rightarrow \mathbf{R}$  be a  $C^1$  function and  $b \in \mathbf{R}$  be a scalar. If  $\nabla f(\mathbf{a}) \neq \mathbf{0}$  for every  $\mathbf{a} \in \{f = b\}$ , then the level set  $\{f = b\}$  is called a smooth (implicitly defined) hypersurface in  $U$ .

In particular, if  $n = 2$ , then  $\{f = b\}$  is a smooth curve and if  $n = 3$ , then  $\{f = b\}$  is a smooth surface.

As in linear algebra one is often interested in solving more than one equation simultaneously. Typically, the best case scenario occurs when one has the same number of equations as unknowns. In linear algebra, this is usually written

$$A\mathbf{x} = \mathbf{b}$$

where  $A$  is an  $n \times n$  matrix,  $\mathbf{x} \in \mathbf{R}^n$  is a vector of unknowns and  $\mathbf{b} \in \mathbf{R}^n$  is a given vector. If the matrix  $A$  is invertible, then the system has the unique solution  $\mathbf{x} = A^{-1}\mathbf{b}$ . If not (i.e.  $A$  is singular), then either there are no solutions or infinitely many solutions. At any rate, the outcome is worse.

In the non-linear case, it’s better to take a more functional point of view. Even in the linear case, one can do this as follows: let  $T : \mathbf{R}^n \rightarrow \mathbf{R}^n$  be the linear transformation given by  $T(\mathbf{x}) = A\mathbf{x}$ , and then rewrite the system as  $T(\mathbf{x}) = \mathbf{b}$ . Invertibility of  $A$  means that  $T$  has an inverse function  $T^{-1}(\mathbf{y}) := A^{-1}\mathbf{y}$ . The unique solution of  $T(\mathbf{x}) = \mathbf{b}$  is then given by  $\mathbf{x} = T^{-1}(\mathbf{b})$ . Since  $T^{-1}$  is linear, this means that the solution  $\mathbf{x}$  varies quite nicely with the given data  $\mathbf{b}$ .

All this is to help motivate the following.

**Theorem 1.4** (Inverse Function Theorem). Let  $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$  be  $C^1$ , and suppose that  $\mathbf{a}, \mathbf{b} \in \mathbf{R}^n$  satisfy

- $F(\mathbf{a}) = \mathbf{b}$ ;
- $DF(\mathbf{a})$  is non-singular.

Then there are neighborhoods  $U, V \subset \mathbf{R}^n$  of  $\mathbf{a}$  and  $\mathbf{b}$ , respectively, such that  $F : U \rightarrow V$  is bijective and the inverse function  $F^{-1} : V \rightarrow U$  is  $C^1$ .

To restate things a bit less precisely, for any  $\mathbf{y} \in V$ , the non-linear system of equations  $F(\mathbf{x}) = \mathbf{y}$  has a unique solution  $\mathbf{x} \in U$  (i.e.  $\mathbf{x} = F^{-1}(\mathbf{y})$ ), and this solution ‘varies nicely’ with  $\mathbf{y}$ .

The conclusion of the Inverse Function Theorem is often summarized by saying that  $F$  is a ‘local diffeomorphism’ near  $\mathbf{a}$ .

**Definition 1.5.** Let  $U, V \subset \mathbf{R}^n$  be open. A diffeomorphism  $F : U \rightarrow V$  is an invertible function such that both  $F$  and  $F^{-1} : V \rightarrow U$  are  $C^1$ .

More generally, a function  $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is a local diffeomorphism near  $\mathbf{a} \in \mathbf{R}^n$  if there are neighborhoods  $U \subset \mathbf{R}^n$  of  $\mathbf{a}$  and  $V \subset \mathbf{R}^n$  of  $F(\mathbf{a})$  such that  $F : U \rightarrow V$  is a diffeomorphism. In this case, we call  $F^{-1} : V \rightarrow U$  is a local inverse for  $F$ .

Note that applying the chain rule to the equation  $F^{-1}(F(\mathbf{x})) = \mathbf{x}$  gives us the following formula for the derivative of the inverse function:

$$D(F^{-1})(F(\mathbf{x})) = (DF(\mathbf{x}))^{-1}$$

for all  $\mathbf{x} \in V$  (in particular, for  $\mathbf{x} = \mathbf{a}$ ).

The proof of the Inverse Function Theorem is rather long, but the idea behind it is quite nice and quite practical (in computer algorithm terms). Namely, starting with the assumption  $F(\mathbf{a}) = \mathbf{b}$ , one wants to solve  $F(\mathbf{x}) = \mathbf{y}$  for  $\mathbf{y}$  close to  $\mathbf{b}$ . If we cheat a bit and replace  $F$  with its linear approximation at  $\mathbf{a}$ , then we get the ‘linearized equation’

$$F(\mathbf{a}) + DF(\mathbf{a})(\mathbf{x} - \mathbf{a}) = \mathbf{y}$$

which (since  $DF(\mathbf{a})$  is invertible) is much easier to solve:

$$\mathbf{x} = \mathbf{a} + DF(\mathbf{a})^{-1}(\mathbf{y} - F(\mathbf{a})).$$

Note that since  $\mathbf{y} - F(\mathbf{a})$  is small, the resulting value of  $\mathbf{x}$  is close to  $\mathbf{a}$ . And since the linear approximation of  $F$  is a good approximation of  $F$  for  $\mathbf{x}$  near  $\mathbf{a}$ , then it’s not too crazy to imagine that the value of  $\mathbf{x}$  we just computed is a good approximation of the solution of  $F(\mathbf{x}) = \mathbf{y}$ . In particular, we hope that it’s a better guess at a solution than just taking  $\mathbf{x} = \mathbf{a}$ .

This heuristic idea is known as Newton’s method. The practical power of it is that it can be iterated to produce a sequence of better and better approximations of the solution of  $F(\mathbf{x}) = \mathbf{y}$ . That is, one begins with a decent guess  $\mathbf{x}_0 = \mathbf{a}$  of a solution of  $F(\mathbf{x}) = \mathbf{y}$  and gets a better guess  $\mathbf{x}_1$  by solving the linearized equation

$$F(\mathbf{x}_0) + DF(\mathbf{x}_0)(\mathbf{x}_1 - \mathbf{x}_0) = \mathbf{y}.$$

And then one repeats the process, linearizing  $F(\mathbf{x}) = \mathbf{y}$  about  $\mathbf{x}_1$  and solving the equation

$$F(\mathbf{x}_1) + DF(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) = \mathbf{y}$$

for another (and hopefully even better) approximation  $\mathbf{x} = \mathbf{x}_2$  at a solution of  $F(\mathbf{x}) = \mathbf{y}$ . Etc. It can be shown that under reasonable circumstances, the resulting sequence of guesses  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$  converges extremely rapidly to the desired solution, and this can be further developed into a complete proof of the Inverse Function Theorem.

I close this section by returning to a more general version of the Implicit Function Theorem. This time I consider the situation in which one is given  $m$  equations in  $n$  unknowns,

$$\begin{aligned} f_1(x_1, \dots, x_n) &= b_1 \\ &\vdots \\ f_m(x_1, \dots, x_n) &= b_m \end{aligned}$$

and asked to solve for some of the unknowns in terms of the others. If our intuition from linear algebra is worth anything, then this is usually only reasonable in the situation where we have more unknowns than equations, i.e. when  $m < n$ . In the linear case, one has an  $m \times n$  matrix  $A$  and uses row reduction to ‘solve’  $A\mathbf{x} = \mathbf{b}$ , i.e. to express pivot variables in terms of free variables. The typical outcome is that  $A$  reduces to a matrix of the form

$$[I \quad *]$$

where  $I$  is the  $m \times m$  identity matrix and  $*$  is some undistinguished complementary  $m \times (n - m)$  matrix. In this case, the first  $m$  variables are the pivot variables and the last  $n - m$  variables are the free variables. The pivot columns of  $A$  form a basis for  $\text{col } A$ , hence in this case, we also have the first  $m$  columns of  $A$  form a *non-singular* square matrix. This helps explain the hypothesis in the following.

**Theorem 1.6** (Implicit Function Theorem (general case)). *Suppose that  $m < n$  and  $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$  is  $C^1$ . If  $\mathbf{a} \in \mathbf{R}^n$  satisfies*

- $F(\mathbf{a}) = \mathbf{b}$
- *The  $m \times m$  matrix  $\left[ \frac{\partial F}{\partial x_1}(\mathbf{a}) \quad \dots \quad \frac{\partial F}{\partial x_m}(\mathbf{a}) \right]$  is invertible.*

*Then there are neighborhoods  $U \subset \mathbf{R}^{n-m}$  of  $(a_{m+1}, \dots, a_n)$  and  $V \subset \mathbf{R}^m$  of  $(a_1, \dots, a_m)$  such that for any  $(x_{m+1}, \dots, x_n) \in U$ , there exists a unique point  $(x_1, \dots, x_m) = \Phi(x_{m+1}, \dots, x_n) \in V$  such that  $F(x_1, \dots, x_n) = \mathbf{b}$ . Moreover, the solution function  $\Phi$  is  $C^1$ .*

Actually, a clever trick shows that this version of the Implicit Function Theorem is really just a restatement of the Inverse Function Theorem. Full details of the equivalence (not to mention a proof of the Inverse Function Theorem) may be found in Shifrin 6.2, but the starting point is to observe that if  $F = (f_1, \dots, f_m)$  is a function from  $\mathbf{R}^n$  to  $\mathbf{R}^m$ , and  $m < n$ , then one can define a function from  $G : \mathbf{R}^n \rightarrow \mathbf{R}^n$  by setting

$$G(x) = G(x_1, \dots, x_n) = (F(\mathbf{x}), x_{m+1}, \dots, x_n) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}), x_{m+1}, \dots, x_n).$$

That is, one makes the last  $n - m$  independent variables serve double-duty as the last  $n - m$  components of  $G$ . Then applying the Inverse Function Theorem to  $G$  turns out to be equivalent to applying the Implicit Function Theorem to  $F$ .

## 2. DETERMINANTS

Any basis  $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbf{R}^n$  of  $\mathbf{R}^n$  determines an  $n$ -dimensional version of a parallelogram (the professional term is *parallelotope*):

$$P := \{\mathbf{x} \in \mathbf{R}^n : \mathbf{x} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n \text{ where } 0 \leq c_j \leq 1\}.$$

The ‘determinant’ of  $(\mathbf{v}_1, \dots, \mathbf{v}_n)$  is a sort of signed volume of  $P$ .

**Definition 2.1.** An  $n$ -dimensional determinant is a scalar-valued function  $\det : \mathbf{R}^n \times \dots \times \mathbf{R}^n \rightarrow \mathbf{R}$  on the  $n$ -fold product of  $\mathbf{R}^n$  with itself such that the following properties hold

- *det is multilinear:* i.e.  $\det(\mathbf{v}_1, \dots, \mathbf{v}_n)$  is linear with respect to each argument  $\mathbf{v}_j$ .
- *det is alternating:* i.e. if  $\mathbf{v}_j = \mathbf{v}_k$  for some  $j \neq k$ , then  $\det(\mathbf{v}_1, \dots, \mathbf{v}_n) = 0$ .
- *det is normalized:* i.e.  $\det(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1$ .

Several remarks are in order. By regarding  $\mathbf{v}_1, \dots, \mathbf{v}_n$  as rows of a matrix  $A$ , one can think of the  $n$ -dimensional determinant as a function on  $n \times n$  matrices. This is the more typical way to approach determinants. Shifrin regards the vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  as *columns* rather than rows of  $A$ , and this is more correct for many purposes, but it will turn out that (remarkably)  $\det A^T = \det A$  so that it really doesn’t matter.

By the multilinearity property, we have  $\det(-\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = -1$ , so a determinant function will assign negative volume to some parallelotopes. This is the price of having things be linear. Of course, one can replace  $\det$  with  $|\det|$  and eliminate the sign. But it also turns out (albeit not so immediately) that the sign of the determinant carries a lot of useful information.

At this point it’s not clear that determinant functions of all dimensions actually exists. When  $n = 1$ , it’s straightforward to check that the identity function  $\det(v_{11}) = v_{11}$  is the one and only determinant function. It’s a bit more work to check that

$$\det(\mathbf{v}_1, \mathbf{v}_2) = v_{11}v_{22} - v_{12}v_{21}$$

is the unique 2-dimensional determinant function. It will turn out that there’s a unique determinant function in every dimension, though it’s much harder to give formulas. Instead, we first establish uniqueness by showing how to compute a determinant using row operations and then work inductively to prove existence.

**Proposition 2.2.** Suppose that  $\det : \mathcal{M}_{n \times n} \rightarrow \mathbf{R}$  is an  $n$ -dimensional determinant (regarded as a function on the rows of an  $n \times n$  matrix). Suppose that  $A, B \in \mathcal{M}_{n \times n}$  are matrices.

- If  $B$  is obtained by multiplying a single row of  $A$  by  $c \in \mathbf{R}$ , then  $\det B = c \det A$ ;
- If  $B$  is obtained by switching two different rows of  $A$ , then  $\det B = -\det A$ ;
- If  $B$  is obtained by adding a multiple of one row of  $A$  to a different row of  $A$ , then  $\det B = \det A$ .

Since any  $n \times n$  matrix is row equivalent to either the identity matrix or a matrix with at least one row of zeroes, the Proposition 2.2 gives an effective means to *compute*  $\det A$  for any  $A \in \mathcal{M}_{n \times n}$  matrix. It is not clear (yet) that the result is independent of the row operations one uses, so the proposition does not directly imply that there actually is an  $n$ -dimensional determinant. However, if we use the proposition to compute the determinant of a matrix, then the result will *not* depend on which determinant function we use, so the proposition does imply that if an  $n$ -dimensional determinant function exists, then it is unique.

**Corollary 2.3.** *There exists at most one  $n$ -dimensional determinant function  $\det : \mathcal{M}_{n \times n} \rightarrow \mathbf{R}$ , and for any  $A \in \mathcal{M}_{n \times n}$  we have  $\det A = 0$  if and only if  $A$  is singular.*

Another consequence of Proposition 2.2 is that one can evaluate determinants of many matrices at a glance.

**Corollary 2.4.** *If  $\det : \mathcal{M}_{n \times n} \rightarrow \mathbf{R}$  is an  $n$ -dimensional determinant and  $A \in \mathcal{M}_{n \times n}$  is upper (or lower) triangular, then  $\det A = a_{11} \dots a_{nn}$  is the product of the (main) diagonal entries of  $A$ .*

Recall now that performing an elementary row operation on a matrix  $A$  to get a matrix  $\tilde{A}$  is the same as multiplying  $A$  by the elementary matrix  $E$  corresponding to that row operation, i.e.  $\tilde{A} = EA$ . By the Proposition 2.2, one has  $\det \tilde{A} = c \det A$  for some non-zero scalar  $c$ . It is straightforward to check for each the three types of elementary row operations that the corresponding elementary matrices satisfy

$$\det E = \det E^T = c.$$

This fact leads to two important (and highly non-obvious) properties of determinants

**Theorem 2.5.** *Let  $\det : \mathcal{M}_{n \times n} \rightarrow \mathbf{R}$  be an  $n \times n$  determinant function. Then for any  $A, B \in \mathcal{M}_{n \times n}$ , we have*

- $\det(AB) = (\det A)(\det B)$ ;
- $\det A^T = \det A$ .

Applying the first conclusion of this theorem to the formula  $AA^{-1} = I$  leads immediately to a formula for the determinant of an inverse matrix

$$\det A^{-1} = \frac{1}{\det A}$$

The fact that  $\det A = \det A^T$  means that we can think of the determinant as an alternating multilinear function of the *columns* instead of the rows of  $A$ . The result will be the same either way. In particular one can use row operations, column operations, or a combination of both to evaluate determinants.

While it's usually best to use row reduction and Proposition 2.2 to compute determinants, there is an alternative recursive procedure for finding determinants that has some theoretical value.

**Definition 2.6.** *Given  $A \in \mathcal{M}_{n \times n}$  and indices,  $i, j \in \{1, \dots, n\}$ , the  $ij$ -minor of  $A$  is the matrix  $A_{ij} \in \mathcal{M}_{(n-1) \times (n-1)}$  obtained by deleting the  $i$ th row and  $j$ th column from  $A$ . The  $ij$ -cofactor of  $A$  is the quantity*

$$c_{ij} := (-1)^{i+j} \det A_{ij}.$$

**Theorem 2.7** (Cofactor Expansion). *Suppose that  $\det_{n-1}$  is an  $n - 1$ -dimensional determinant function. Then for any  $j \in \{1, \dots, n\}$  there exists an  $n$ -dimensional determinant given for any  $A \in \mathcal{M}_{n \times n}$  by*

$$\det_n A := \sum_{i=1}^n a_{ij} c_{ij},$$

where  $a_{ij}$ ,  $c_{ij}$  are the  $ij$ -entry and  $ij$ -cofactor of  $A$ , respectively. The matrix  $(c_{ij}) \in \mathcal{M}_{n \times n}$  is called the cofactor matrix of  $A$ .

Since we know that there is a 1-dimensional determinant, one can use induction and Corollary 2.3 to prove

**Corollary 2.8.** *For every  $n \geq 1$ , there exists a unique determinant  $\det = \det_n : \mathcal{M}_{n \times n} \rightarrow \mathbf{R}$ .*

Uniqueness of determinants implies that in Theorem 2.7, it doesn't matter which column  $j$  one chooses for the cofactor expansion of  $\det_n$ . Since  $\det A^T = \det A$ , one has a corresponding cofactor expansion of  $\det_n$  in which one fixes the row  $i$  rather than the column  $j$ .

Cofactor expansion leads to various interesting (albeit not especially practical) formulas for solutions to linear systems and inverses of matrices. I state here only the formula for the inverse of a matrix.

**Corollary 2.9** (Cramer's Rule for matrix inverses). *Suppose that  $A \in \mathcal{M}_{n \times n}$  is non-singular. Then*

$$A^{-1} = \frac{C^T}{\det A}$$

where  $C$  is the cofactor matrix for  $A$ .

Note the easy to overlook transpose in the formula. I always seem to forget about it.

## 3. RECTANGLES, PARTITIONS AND STEP FUNCTIONS

**Definition 3.1.** A rectangle is a compact set  $R \subset \mathbf{R}^n$  of the form  $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$  where each closed interval  $[a_j, b_j] \subset \mathbf{R}$  has finite but non-zero length. That is,

$$R = \{(x_1, \dots, x_n) \in \mathbf{R}^n : x_j \in I_j\}.$$

The volume of  $R$  is the product  $\text{Vol } R := (b_1 - a_1) \cdots (b_n - a_n)$  of the lengths of the intervals  $[a_j, b_j]$ .

Note that as is typical in this subject, we use the word ‘rectangle’ for  $R \subset \mathbf{R}^n$  regardless of whether  $n = 1$  (in which case  $R$  is simply a closed interval),  $n = 2$  (in which case  $R$  is a rectangle in the usual sense) or  $n = 39$  (dimension-specific naming conventions for rectangles stop with  $n = 28\dots^1$ ). In what follows, we will often want to decompose a given rectangle  $R$  into a grid of smaller rectangles.

One can always subdivide a given rectangle  $R$  into a collection of smaller rectangles  $Q$ . We will be interested in doing this so that the smaller rectangles are arranged in a ‘grid’ pattern. Specifically, when  $R = [a, b] \subset \mathbf{R}$  is an interval, a partition of  $R$  is a collection  $\mathcal{I}$  of subintervals  $[a_{j-1}, a_j]$  corresponding to an increasing sequence  $a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b$  such that  $\mathcal{I} = \{[a_{j-1}, a_j] : 1 \leq j \leq n\}$ . Alternatively (but equivalently),  $\mathcal{I}$  is a collection of subintervals  $I \subset [a, b]$  such that

- $\bigcup_{I \in \mathcal{I}} I = [a, b]$ ; and
- different intervals  $I_1, I_2 \in \mathcal{I}$  satisfy  $\text{int} I_1 \cap \text{int} I_2 = \emptyset$ .

Now if  $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbf{R}^n$  is a higher dimensional rectangle, then a partition of  $R$  is again a collection  $\mathcal{P}$  of non-overlapping subrectangles  $Q \subset R$  that cover  $R$ . But we further require that these rectangles ‘form a grid’, i.e. there are partitions  $\mathcal{I}_j$  of  $[a_j, b_j]$  such that  $Q \in \mathcal{P}$  if and only if  $Q = I_1 \times \cdots \times I_n$  for intervals  $I_1 \in \mathcal{I}_1, \dots, I_n \in \mathcal{I}_n$ .

The point behind partitions is that we can make them ‘smaller and smaller.’ More precisely,

**Definition 3.2.** Let  $\mathcal{P}, \mathcal{P}'$  be partitions of the same rectangle  $R$ . We say that  $\mathcal{P}'$  refines  $\mathcal{P}$  if for any rectangle  $Q' \in \mathcal{P}'$ , there exists a rectangle  $Q \in \mathcal{P}$  such that  $Q' \subset Q$ .

The next two propositions are basic facts about partitions which we will use a lot. Both are ‘obvious’ from an intuitive standpoint. Nevertheless, it is somewhat tricky to give an uncluttered formal justification.

**Proposition 3.3.** Let  $\mathcal{P}$  be a partition of a rectangle  $R \subset \mathbf{R}^n$ . Then

$$\sum_{Q \in \mathcal{P}} \text{Vol } Q = \text{Vol } R.$$

If, moreover,  $\mathcal{P}'$  is another partition of  $R$ , then there is a third partition  $\mathcal{P}''$  of  $R$  that refines both  $\mathcal{P}$  and  $\mathcal{P}'$ .

To any subset  $S \subset \mathbf{R}^n$ , one can associate an *indicator function*  $\mathbf{1}_S : \mathbf{R}^n \rightarrow \mathbf{R}$  given by  $\mathbf{1}_S(x) = 1$  for  $x \in S$  and  $\mathbf{1}_S(x) = 0$  otherwise. More generally, if  $R \subset \mathbf{R}^n$  is a rectangle with a partition  $\mathcal{P}$  we will call a function  $f : R \rightarrow \mathbf{R}$  a *step function subordinate to  $\mathcal{P}$*  if  $f$

<sup>1</sup>You understand I’m kidding, right?



is constant on each rectangle  $Q \in \mathcal{P}$ . That is, there are constants  $f_Q \in \mathbf{R}$  such that  $f$  is a linear combination

$$f = \sum_{Q \in \mathcal{P}} f_Q \mathbf{1}_Q$$

of indicator functions. Note that, strictly speaking, when we say ‘piecewise constant’ we are disregarding overlapping edges between different rectangles. This little transgression makes no difference, so we will not repent of it in what follows. On ‘base times height’ grounds, it is reasonable that for rectangles  $Q \subset \mathbf{R}^n$  we should have  $\int_R \mathbf{1}_Q dV = \text{Vol } Q$ . So we will just define things that way. More generally,

**Definition 3.4.** *Let  $f : R \rightarrow \mathbf{R}$  be a step function on a rectangle  $R \subset \mathbf{R}^n$  subordinate to a partition  $\mathcal{P}$  of  $R$ . Then the integral of  $f$  over  $R$  is the quantity*

$$\int_R f dV = \sum_{Q \in \mathcal{P}} f_Q \text{Vol } Q.$$

An important thing to note is that if  $f : R \rightarrow \mathbf{R}$  is a step function subordinate to  $\mathcal{P}$  and  $\mathcal{P}'$  refines  $\mathcal{P}$ , then  $f$  is also subordinate to  $\mathcal{P}'$ . This opens up the possibility of evaluating the integral of  $f$  using either partition. It is therefore important to know that our definition of integral for a step function is independent of the underlying partition.

**Proposition 3.5.** *Let  $f : R \rightarrow \mathbf{R}$  be a step function on a rectangle  $R \subset \mathbf{R}^n$ . If  $f$  is subordinate to two partitions  $\mathcal{P}$  and  $\mathcal{P}'$  of  $R$ , then*

$$\sum_{Q \in \mathcal{P}} f_Q \text{Vol } Q = \sum_{Q' \in \mathcal{P}'} f_{Q'} \text{Vol } Q'$$

*I.e. the value of  $\int_R f dV$  does not depend on which partition one chooses to compute it.*

Finally, if  $f, g : R \rightarrow \mathbf{R}$  are step functions on the same rectangle, then by choosing a common refinement of the two partitions on which  $f$  and  $g$  are based, we can assume that  $f$  and  $g$  are subordinate to the same partition  $\mathcal{P}$  of  $R$ . Hence, for example,

$$f + g = \sum_{Q \in \mathcal{P}} (f_Q + g_Q) \mathbf{1}_Q$$

is also a step function. More generally, we have

**Proposition 3.6.** *Let  $f, g : R \rightarrow \mathbf{R}$  be step functions on a rectangle  $R \subset \mathbf{R}^n$ . Then  $f + g$ ,  $f - g$ ,  $fg$ ,  $\max\{f, g\}$ , and  $\min\{f, g\}$  are all step functions on  $R$ . Moreover,*

- if  $f \leq g$ , then  $\int_R f dV \leq \int_R g dV$ ;
- $\int_R cf dV = c \int_R f dV$  for all  $c \in \mathbf{R}$ ; and
- $\int_R (f + g) dV = \int_R f dV + \int_R g dV$ .

## 4. RIEMANN INTEGRABLE FUNCTIONS

In this section we will go on to define the integral of more general functions by approximating them from above and below with step functions.

**Definition 4.1.** Let  $R \subset \mathbf{R}^n$  be a rectangle and  $f : R \rightarrow \mathbf{R}$  be a bounded function. The upper integral of  $f$  on  $R$  is the quantity

$$\overline{\int}_R f dV := \inf \left\{ \int_R h dV : h \geq f \text{ is a step function on } R \right\}.$$

The lower integral of  $f$  on  $R$  is the quantity

$$\underline{\int}_R f dV := \sup \left\{ \int_R g dV : g \leq f \text{ is a step function on } R \right\}.$$

If the upper and lower integrals are the same, then we say that  $f$  is Riemann integrable on  $R$ . Either quantity is then called the Riemann integral of  $f$  on  $R$  and is denoted simply

$$\int_R f dV.$$

Typically, we will write simply ‘integrable’ in place of ‘Riemann integrable’ here. While there are other notions of integrability (most notably ‘Lebesgue integrability’), we will not consider them here.

The lower integral of a given function is never greater than the upper integral. Hence we have the following convenient criterion for Riemann integrability.

**Proposition 4.2.** A bounded function  $f : R \rightarrow \mathbf{R}$  is Riemann integrable on a rectangle  $R \subset \mathbf{R}^n$  if for any  $\epsilon > 0$  there exist step functions  $g \leq f \leq h$  on  $R$  such that

$$\int_R (h - g) dV < \epsilon.$$

Integration interacts well with basic arithmetic, etc of functions.

**Theorem 4.3.** Let  $f, g : R \rightarrow \mathbf{R}$  be integrable functions on a rectangle  $R \subset \mathbf{R}^n$ . Then  $f + g$ ,  $f - g$ ,  $fg$ ,  $\max\{f, g\}$ ,  $\min\{f, g\}$  are also integrable on  $R$ . Moreover,

- $f \geq g$  implies  $\int_R f dV \geq \int_R g dV$ ;
- $\int_R (f + g) dV = \int_R f dV + \int_R g dV$ ;
- for any  $c \in \mathbf{R}$ , we have  $\int_R cf dV = c \int_R f dV$ .

The following theorem gives us our most important set of integrable functions.

**Theorem 4.4.** Any continuous function  $f : R \rightarrow \mathbf{R}$  on a rectangle  $R \subset \mathbf{R}^n$  is integrable.

The definition of Riemann integral, while an effective theoretical tool, is not a very practical means for actually computing integrals. In order to compute, we have a result that allows one to integrate functions ‘one variable at a time.’

**Theorem 4.5** (Fubini’s Theorem). Let  $R \subset \mathbf{R}^n$  be a rectangle of the form  $R' \times [a_n, b_n]$  where  $R' \subset \mathbf{R}^{n-1}$  and  $[a_n, b_n] \subset \mathbf{R}$ . Suppose that  $f : R \rightarrow \mathbf{R}$  is an integrable function and that moreover

- For each fixed  $x_n \in [a_n, b_n]$ , the function  $f(\cdot, x_n) : R' \rightarrow \mathbf{R}$  is integrable, and
- The integral  $F(x_n) := \int_{R'} f(x', x_n) dV(x')$  is itself an integrable function on  $[a_n, b_n]$ .

Then

$$\int_R f(x) dV(x) dx = \int_{a_n}^{b_n} F(x_n) dx_n = \int_{a_n}^{b_n} \left( \int_{R'} f(x', x_n) dV(x') \right) dx_n.$$

**4.1. Integrating over regions.** Often we will want to integrate functions over bounded sets  $\Omega \subset \mathbf{R}^n$  that are not rectangles. This can usually be done by choosing a rectangle  $R$  that contains  $\Omega$  and then ‘extending’ the given function by declaring it’s value to be 0 at all points outside  $\Omega$ . Deciding when this actually works requires some more discussion.

**Definition 4.6.** Let  $R \subset \mathbf{R}^n$  be a rectangle and  $S \subset R$  be any subset. We say that  $S$  has zero volume if for any  $\epsilon > 0$  there exists a partition  $\mathcal{P}$  of  $R$  such that

$$\sum_{\substack{Q \in \mathcal{P} \\ Q \cap S \neq \emptyset}} \text{Vol } Q < \epsilon.$$

With this definition in hand, we can generalize the fact that continuous functions are integrable.

**Theorem 4.7.** If  $f : R \rightarrow \mathbf{R}$  is a bounded function on a rectangle  $R \subset \mathbf{R}^n$  and the subset of  $R$  where  $f$  fails to be continuous has zero volume, then  $f$  is integrable.

In order to see the utility of this fact, we need one more definition.

**Definition 4.8.** Let  $R \subset \mathbf{R}^n$  be a rectangle and  $U \subset R$  be an open set whose boundary  $\partial U$  has volume zero. Then we call the compact set  $\Omega := \overline{U}$  a region in  $\mathbf{R}^n$ . The volume of  $\Omega$  is the quantity

$$\text{Vol } \Omega := \int_R \mathbf{1}_\Omega dV.$$

More generally, if  $f : \Omega \rightarrow \mathbf{R}$  is a continuous function, then we set

$$\int_\Omega f dV = \int_R \tilde{f} dV,$$

where  $\tilde{f} : R \rightarrow \mathbf{R}$  is the function given by  $\tilde{f}(x) = f(x)$  if  $x \in \Omega$ , and  $\tilde{f}(x) = 0$  if  $x \notin \Omega$ .

Note that this definition is made possible by the previous theorem. That is, the functions  $\mathbf{1}_\Omega$  and  $\tilde{f}$  fail to be continuous only at points in  $\partial\Omega$  which, by assumption, has volume zero. There is one remaining loose end: how do we know when the boundary  $\partial U$  of an open set  $U$  has zero volume? This is hard to determine in general. However, the following will suffice for our purposes.

**Theorem 4.9.** Let  $R = R' \times [a_n, b_n]$  be a rectangle in  $\mathbf{R}^n$  and  $K' \subset R'$  be a compact set. If  $f : K' \rightarrow [a_n, b_n]$ , is continuous then the graph of  $f$

$$\{(x', f(x')) \in R : x' \in K'\}$$

has zero volume.

The moral here is that we are now justified in trying to integrate continuous functions over any compact set  $\Omega = \overline{U}$  where  $U$  is open and  $\partial U$  can be decomposed into finitely many pieces, each of which is the graph of a continuous function. Let us end by illustrating with an example: the unit ball  $\Omega = \overline{B_3(0, 1)} \subset \mathbf{R}^3$ , which has boundary

$$\partial\Omega := \{(x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 = 1\} = \{(x, y, z) \in \mathbf{R}^3 : z = \pm\sqrt{1 - x^2 - y^2}\}.$$

Since both functions  $-\sqrt{1-x^2-y^2}$  and  $\sqrt{1-x^2-y^2}$  are continuous it follows that  $\partial\Omega$  has volume zero. That is,  $\overline{B_3(0,1)}$  is a region, and we are therefore entitled to try to integrate any continuous function  $f : B_3(0,1) \rightarrow \mathbf{R}$ .

## 5. CHANGE OF VARIABLES IN INTEGRALS

Recall that if  $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$  is a function and  $S \subset \mathbf{R}^n$  is a subset of the domain of  $F$ , then the *image* of  $S$  by  $F$  is the set

$$F(S) := \{F(x) \in \mathbf{R}^m : x \in S\}.$$

Our first result explains how linear transformations affect the volumes of sets.

**Theorem 5.1.** *Suppose that  $\Omega \subset \mathbf{R}^n$  is a region and  $T : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is a linear transformation with standard matrix  $A \in \mathcal{M}_{n \times n}$ . Then  $T(\Omega)$  is also a region and*

$$\text{Vol } T(\Omega) = |\det A| \text{Vol } (\Omega).$$

Note in particular that if  $\Omega = [0, 1]^n$  is the unit cube in  $\mathbf{R}^n$  then  $T(\Omega)$  is the parallelotope defined by the columns of  $A$  and the theorem tells us that the volume of this parallelotope is  $|\det A|$ . This is consistent with the idea that the determinant of a matrix should be the signed volume of the parallelotope determined by columns of the matrix.

Given what we know, the hardest part to proving Theorem 5.1 is showing that  $T(\Omega)$  is actually a region. In particular, one must show that the image of a compact set by a linear transformation is compact. After that, one can quickly reduce the proof to the special case where  $\Omega$  is a rectangle and  $A$  is an elementary matrix. The case when  $A$  corresponds to adding a multiple of one row to another is the most interesting and is dealt with by using Fubini's Theorem.

There are (at least) two ways that one might seek to generalize Theorem 5.1. The harder of the two is to replace the linear transformation  $T$  with a more general non-linear transformation  $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$ . An easier generalization, which will be superseded by Theorem 5.5 below, is as follows.

**Theorem 5.2.** *Suppose that  $\Omega \subset \mathbf{R}^n$  is a region,  $T : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is linear with standard matrix  $A \in \mathcal{M}_{n \times n}$  and  $f : T(\Omega) \rightarrow \mathbf{R}$  is a continuous scalar-valued function on the image of  $\Omega$ . Then*

$$\int_{T(\Omega)} f \, dV = |\det A| \int_{\Omega} f \circ T \, dV.$$

Note that if  $f$  is the constant function 1, then we recover Theorem 5.1 from this one. The proof of Theorem 5.2 proceeds by partitioning  $\Omega$  up into rectangles on which the continuous function  $f$  is nearly constant and then applying Theorem 5.1.

Now I explain what sorts of non-linear transformations will be suitable replacements for the linear transformation  $T$  in this context.

**Definition 5.3.** *Let  $U, V \subset \mathbf{R}^n$  be open sets. A diffeomorphism between  $U$  and  $V$  is a  $C^1$  mapping  $\Phi : U \rightarrow V$  with a  $C^1$  inverse  $\Phi^{-1} : V \rightarrow U$ .*

Here is the harder generalization of Theorem 5.1. The general idea is that if  $\mathcal{P}$  is a partition of  $\Omega$  by small rectangles  $Q$ , then on each such  $Q$ , the mapping  $\Phi$  is approximately linear (plus a constant) with matrix given by  $D\Phi$ . Hence the image of  $Q$  by  $\Phi$  is approximately (a translate of) the image of  $Q$  by  $D\Phi$ , so Theorem 5.4 follows from applying Theorem 5.1 to the image of  $Q$  by  $D\Phi$ .

**Theorem 5.4.** *Suppose that  $\Phi : U \rightarrow V$  is a diffeomorphism between open sets  $U, V \subset \mathbf{R}^n$  and that  $\Omega \subset U$  is a region. Then  $\Phi(\Omega)$  is also a region and*

$$\text{Vol } \Phi(\Omega) = \int_{\Omega} |\det D\Phi| dV.$$

From here it is a fairly small (logical) step to the ultimate generalization of Theorem 5.1.

**Theorem 5.5** (Change of Variables Theorem). *Suppose that  $\Phi : U \rightarrow V$  is a diffeomorphism between open sets  $U, V \subset \mathbf{R}^n$  and that  $\Omega \subset U$  is a region. Then for any continuous scalar-valued function  $f : \Omega \rightarrow \mathbf{R}$ ,*

$$\int_{\Phi(\Omega)} f(y) dV(y) = \int_{\Omega} f \circ \Phi(x) |\det D\Phi(x)| dV(x).$$

I will conclude by working out what the Change of Variables Theorem says for some standard ‘change of coordinate’ mappings  $\Phi$ .

**Polar coordinates in  $\mathbf{R}^2$ :** One sometimes describes a point  $(x, y) \in \mathbf{R}^2$  in terms of the length  $r$  of the line segment joining  $(x, y)$  to  $(0, 0)$  and the angle  $\theta$  that this line segment makes with the positive  $x$ -axis.  $r$  and  $\theta$  are called the polar coordinates of  $(x, y)$  and they are related by

$$(x, y) = \Phi(r, \theta) := (r \cos \theta, r \sin \theta).$$

Note that  $\Phi$  is a diffeomorphism from  $(0, \infty) \times (-\pi, \pi)$  onto all of  $\mathbf{R}^2$  except the negative  $x$ -axis. One (ideally you!) computes that

$$\det D\Phi = r.$$

Hence if  $f : \Omega \rightarrow \mathbf{R}$  is a continuous function on a region  $\Omega \subset \mathbf{R}^2$ , then the change of variables theorem says that

$$\iint_{\Omega} f(x, y) dx dy = \iint_{\Phi^{-1}(\Omega)} f \circ \Phi(r, \theta) r dr d\theta$$

The trick in applying this formula is usually to figure out what  $\Phi^{-1}(\Omega)$  is, i.e. to infer the limits on  $r$  and  $\theta$  from those on  $x$  and  $y$ .

**Cylindrical coordinates in  $\mathbf{R}^3$ :** The most direct way to extend polar coordinates to  $(x, y, z) \in \mathbf{R}^3$  is to let  $(r, \theta)$  determine the first two coordinates  $(x, y)$  as before but let  $z$  represent itself. So one has a mapping  $(x, y, z) = \Phi(r, \theta, z) := (r \cos \theta, r \sin \theta, z)$  which is a diffeomorphism as long as  $r$  is restricted to be positive and  $\theta$  is limited to some interval of length smaller than  $2\pi$ . Not surprisingly, one has again that

$$\det D\Phi = r.$$

**Spherical coordinates in  $\mathbf{R}^3$ :** Another way to extend polar coordinates to  $(x, y, z) \in \mathbf{R}^3$  is to let  $\rho$  denote the distance from  $(x, y, z)$  to  $(0, 0, 0)$ ,  $\theta$  denote (as in polar coordinates) the angle from the positive  $x$ -axis to the line segment joining  $(0, 0)$  to  $(x, y)$ , and  $\phi$  denote the angle from the positive  $z$ -axis to the line segment joining  $(0, 0, 0)$  to  $(x, y, z)$ . In effect,  $\theta$  and  $\phi$  serve as longitude and latitude of the point  $(x, y, z)$ , except that latitude 0 occurs at the north pole. Anyhow, in this case, the change of coordinates map is

$$(x, y, z) = \Phi(\rho, \theta, \phi) := (\rho \sin \phi \cos \theta, \rho \sin \phi \sin \theta, \rho \cos \phi).$$

From this formula, one computes that

$$\det D\Phi = \rho^2 \sin \phi.$$

## 6. INTEGRATION ON CURVES

(by Christian Gorski)

We begin by recalling some basic terminology associated to parametrized curves in  $\mathbf{R}^n$ .

**Definition 6.1.** A (**parametrized**) **curve** in  $\mathbf{R}^n$  is a continuous function  $\gamma : [a, b] \rightarrow \mathbf{R}^n$  defined on a closed interval  $[a, b] \subset \mathbb{R}$ .

- If the derivative  $\gamma'$  of  $\gamma$  exists and is continuous at all points in  $[a, b]$ , then we say that  $\gamma$  is  $\mathcal{C}^1$ . We may refer to  $\gamma'(t)$  as the **tangent vector** to  $\gamma$  at the point  $\gamma(t)$ .
- If additionally,  $\gamma' \neq \mathbf{0}$  at all points in  $[a, b]$ , then we say that  $\gamma$  is **smooth**.
- $\gamma$  is **piecewise  $\mathcal{C}^1$**  (or **smooth**) if there exist finitely many parameters  $a = a_0 < a_1 < \dots < a_k = b \in \mathbb{R}$  such that  $\gamma$  is  $\mathcal{C}^1$  (respectively, smooth) on  $[a_{i-1}, a_i]$  for all  $1 \leq i \leq k$ .
- $\gamma$  is **simple** if  $(\gamma(s) = \gamma(t)) \Rightarrow (s = t \text{ or } \{s, t\} = \{a, b\})$  (that is,  $\gamma$  is one-to-one, except that its endpoints may be equal).
- $\gamma$  is **closed** if  $\gamma(a) = \gamma(b)$ .

**Definition 6.2.** Let  $\gamma : [a, b] \rightarrow \mathbf{R}^n$  be a  $\mathcal{C}^1$  curve. We say  $\tilde{\gamma} : [\tilde{a}, \tilde{b}] \rightarrow \mathbf{R}^n$  is a **reparametrization** of  $\gamma$  if there exists a diffeomorphism  $\Phi : [\tilde{a}, \tilde{b}] \rightarrow [a, b]$  such that  $\tilde{\gamma} = \gamma \circ \Phi$ . We say that  $\Phi$  is **orientation-preserving** if it is increasing, and we say that it is **orientation-reversing** if it is decreasing.

Note in particular that if  $\tilde{\gamma}$  is a reparametrization of  $\gamma$ , then the two curves have the same image in  $\mathbf{R}^n$ .

$$\tilde{\gamma}([\tilde{a}, \tilde{b}]) = \gamma(\Phi([\tilde{a}, \tilde{b}])) = \gamma([a, b]).$$

This is pertinent, since one is often more interested in the image of a curve than in the particular parametrization used to describe it. Moreover, it follows from the inverse function theorem (for one variable  $\mathcal{C}^1$  functions) that the converse is true under decent circumstances.

**Proposition 6.3.** If  $\gamma : [a, b] \rightarrow \mathbf{R}^n$  and  $\tilde{\gamma} : [\tilde{a}, \tilde{b}] \rightarrow \mathbf{R}^n$  are simple smooth curves with identical images  $\gamma([a, b]) = \tilde{\gamma}([\tilde{a}, \tilde{b}])$ , then  $\tilde{\gamma}$  is a reparametrization of  $\gamma$ .

**6.1. Arc-length integrals.** Now we introduce the first of two ways of integrating over curves: integration with respect to arc-length.

**Definition 6.4.** Given a  $\mathcal{C}^1$  curve  $\gamma : [a, b] \rightarrow \mathbf{R}^n$  and a continuous function  $f : \mathbf{R}^n \rightarrow \mathbb{R}$ , defined on some neighborhood of  $\gamma([a, b])$ , the **integral of  $f$  with respect to arc-length** on  $\gamma$  is the quantity

$$\int_{\gamma} f ds := \int_a^b f(\gamma(t)) \|\gamma'(t)\| dt.$$

In particular, the **length** of  $\gamma$  is the quantity

$$\text{Length}(\gamma) = \int_a^b \|\gamma'(t)\| dt = \int_{\gamma} 1 ds.$$

One sign that this definition is reasonable is that it is ‘parametrization independent’.

**Theorem 6.5.** Let  $\gamma : [a, b] \rightarrow \mathbf{R}^n$  be a  $\mathcal{C}^1$  curve and  $f : \mathbf{R}^n \rightarrow \mathbb{R}$  be continuous. Let  $\tilde{\gamma}$  be a reparametrization of  $\gamma$ . Then

$$\int_{\gamma} f ds = \int_{\tilde{\gamma}} f ds$$



Note that the arc-length integral is the same, regardless of whether the reparametrization reverses or preserves orientation. This will not be the case for the other type of integration on curves that we will consider.

## 6.2. Covectors.

**Definition 6.6.** A **covector** on  $\mathbb{R}^n$  is a linear function  $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ .

It is natural to think of a covector  $\omega$  in terms of its standard  $1 \times n$  matrix, i.e. as a row vector. In particular, for each  $1 \leq j \leq n$ , we have a covector  $dx_j : \mathbb{R}^n \rightarrow \mathbb{R}$  given corresponding to the row vector with  $j$ th entry 1 and all other entries equal to 0; i.e. for each  $\mathbf{x} \in \mathbb{R}^n$

$$dx_j(\mathbf{x}) = x_j$$

is the  $j$ th coordinate of  $\mathbf{x}$ . We will call  $dx_1, \dots, dx_n$  the “standard basis covectors.”

The set of all covectors  $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$  is denoted  $(\mathbb{R}^n)^*$ .

**Proposition 6.7.** Any covector  $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$  may be written uniquely as

$$\omega = \sum_{i=1}^n \omega_i dx_i$$

where each  $\omega_i \in \mathbb{R}$  is a scalar.

**Remark 6.8.** Covectors are also called “linear forms” or “linear functionals.”

- For  $\omega \in (\mathbb{R}^n)^*$ ,  $\mathbf{x} \in \mathbb{R}^n$ , physicists and mathematicians often write  $\langle \omega | \mathbf{x} \rangle := \omega(\mathbf{x})$ . This happens in e.g. quantum mechanics.

In keeping with the convention that vector means column vector and covector means row vector, it will be convenient here to use matrix product notation  $\omega \mathbf{x}$  to denote  $\omega$  applied to  $\mathbf{x}$ .

- Using the dot product can confuse a covector with a vector;

$$\omega(\mathbf{x}) = \sum \omega_i x_i = \begin{bmatrix} \omega_1 & \cdots & \omega_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

This is extremely common outside (and sometimes even inside) mathematics.

## 6.3. Line Integrals.

**Definition 6.9.** A **differential 1-form** on an open set  $U \subset \mathbb{R}^n$  is a function

$$\omega : U \rightarrow (\mathbb{R}^n)^*.$$

More concretely, a differential 1-form is an expression of the form  $\omega = \sum_{j=1}^n \omega_j dx_j$ , where  $\omega_j : U \rightarrow \mathbb{R}$  are scalar-value coefficient functions. We will say that  $\omega$  is continuous (or differentiable or  $C^1$ , etc) if the coefficient functions  $\omega_j$  are continuous (or differentiable or  $C^1$ , etc). When we don't need to stress the domain  $U$ , we will refer to  $\omega$  as a ‘1-form on  $\mathbb{R}^n$ ’.

**Definition 6.10.** Let  $U \subset \mathbb{R}^n$  be open,  $\omega$  be a continuous 1-form on  $U$  and  $\gamma : [a, b] \rightarrow U$  be a  $C^1$  parametrized curve. Then **the line integral of  $\omega$  over  $\gamma$**  is the quantity

$$\int_{\gamma} \omega := \int_a^b \omega(\gamma(t)) \gamma'(t) dt$$

The integral of a differential 1-form is unchanged by *orientation-preserving* reparametrization of curves. Unlike arc-length integrals, however, orientation-reversing reparametrizations change the sign of the integral.

**Theorem 6.11.** *Let  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  be a  $\mathcal{C}^1$  curve and  $\tilde{\gamma} : [c, d] \rightarrow \mathbb{R}^n$  be a reparametrization. Then, for any continuous 1-form  $\omega$  on  $\mathbb{R}^n$ , we have*

$$\int_{\tilde{\gamma}} \omega = \int_{\gamma} \omega$$

if  $\tilde{\gamma}$  and  $\gamma$  have the same orientation, and

$$\int_{\tilde{\gamma}} \omega = - \int_{\gamma} \omega$$

if  $\tilde{\gamma}$  and  $\gamma$  have opposite orientations.

The main advantage of line integrals over arc-length integrals is that they lead to (two) natural generalizations of the fundamental theorem of Calculus.

**Definition 6.12.** *If  $f : U \rightarrow \mathbb{R}$  is a  $\mathcal{C}^1$  function on an open set  $U \subset \mathbb{R}^n$ , then the 1-form*

$$df := \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i$$

is called the **differential** of  $f$ .

**Definition 6.13.** *A 1-form  $\omega : U \rightarrow (\mathbb{R}^n)^*$  on an open set  $U \subset \mathbb{R}^n$  is called **exact** if there exists a  $\mathcal{C}^1$  function  $f : U \rightarrow \mathbb{R}$  such that  $\omega = df$ . In this case, the function  $f$  is called a **potential** for  $\omega$ .*

The potential of an exact 1-form is not quite unique: if  $f$  is a potential for some 1-form  $\omega$ , we can obtain another potential  $\tilde{f}$  for  $\omega$  by simply adding a constant—i.e. by defining

$$\tilde{f}(x) = f(x) + c$$

for some constant  $c \in \mathbb{R}$ .

**Theorem 6.14.** *(The Fundamental Theorem of Calculus for Curves) Let  $U \subset \mathbb{R}^n$  be an open set, and let  $[a, b] \subset \mathbb{R}$  be a compact interval. If  $\gamma : [a, b] \rightarrow U$  is a  $\mathcal{C}^1$  curve and  $f : U \rightarrow \mathbb{R}$  is a  $\mathcal{C}^1$  function, then*

$$\int_{\gamma} df = f(\gamma(b)) - f(\gamma(a)).$$

Note that this theorem implies that the line integral of a differential over a curve is *path-independent*; that is, the integral depends only on the endpoints and orientation of the curve, and not the path itself. In particular, the integral of an exact 1-form over a closed curve must vanish. These facts characterize exactness.

**Theorem 6.15.** *Let  $\omega$  be a continuous 1-form defined on an open set  $U \subset \mathbb{R}^n$ . Then the following statements are equivalent:*

- (1)  $\omega$  is exact.
- (2)  $\int_{\gamma} \omega$  is path independent for any curve  $\gamma$ .
- (3)  $\int_{\gamma} \omega = 0$  for any closed curve  $\gamma$ .

**Definition 6.16.** A  $C^1$  1-form  $\omega = \sum \omega_i dx_i$  defined on  $U \subset \mathbb{R}^n$  is **closed** if

$$\frac{\partial \omega_i}{\partial x_j} = \frac{\partial \omega_j}{\partial x_i} \text{ for all } 1 \leq i, j \leq n.$$

Note that, since mixed partial derivatives of  $C^2$  functions commute, closed-ness is a (fairly easy to verify) necessary condition for exact-ness.

**Proposition 6.17.** A  $C^1$  and exact 1-form is closed.

It turns out that, locally speaking, the converse is true. The next theorem helps make this precise. To state it, we need one more definition:

**Definition 6.18.** A subset  $U \subset \mathbb{R}^n$  is **star-shaped** if there exists  $\mathbf{a} \in U$  such that, for any  $\mathbf{x} \in U$ , the line segment from  $\mathbf{a}$  to  $\mathbf{x}$  is contained in  $U$ .

**Theorem 6.19.** Let  $U \subset \mathbb{R}^n$  be open and star-shaped, and let  $\omega$  be a  $C^1$  1-form on  $U$ . Then  $\omega$  is exact if and only if  $\omega$  is closed.

Balls, rectangles and (more generally) convex sets are all instances of star-shaped sets. However, the set  $\mathbb{R}^2 - \{\mathbf{0}\}$  is not star-shaped; this allows for the interesting example, in the following definition, of a 1-form that is closed but not exact:

**Definition 6.20.** Let  $\gamma : [a, b] \rightarrow \mathbb{R}^2 - \{\mathbf{0}\}$  be a  $C^1$  closed curve, and set

$$\omega = \frac{-ydx + xdy}{x^2 + y^2}.$$

Then the **winding number** (about the origin) of  $\gamma$  is the quantity  $\frac{1}{2\pi} \int_{\gamma} \omega$ .

One computes that the line integral of  $\omega$  around the unit circle with counterclockwise orientation is  $2\pi$ . Since the circle is closed, it follows that  $\omega$  is not exact. As we saw in class, for any curve  $\gamma : [a, b] \rightarrow \mathbb{R}^2 - \{\mathbf{0}\}$ , closed or not, the integral  $\int_{\gamma} \omega$  measures the angle that  $\gamma$  turns about zero. Hence the winding number is an integer equal to the number of times the curve winds about the origin.

**6.4. Green's Theorem.** The second version of the fundamental theorem of Calculus involving line integrals involves integration over two dimensional regions. Curves appear as the *boundaries* of such regions.

**Definition 6.21.** Let  $\omega = \omega_1 dx_1 + \omega_2 dx_2$  be a  $C^1$  1-form on an open subset  $U$  of  $\mathbb{R}^2$ . Then the **exterior derivative** of  $\omega$  is the expression

$$d\omega := \left( \frac{\partial \omega_2}{\partial x_1} - \frac{\partial \omega_1}{\partial x_2} \right) dx_1 dx_2.$$

Note that  $\omega$  is a closed 1-form if and only if  $d\omega = 0$ . In particular,  $d(df) = 0$  for every  $C^2$  function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ .

**Theorem 6.22** (Green's Theorem). Let  $\Omega \subset \mathbb{R}^2$  be a region such that  $\partial\Omega$  is a finite union of piecewise  $C^1$  simple closed curves; let  $\omega$  be a  $C^1$  1-form on an open set  $U \supset \Omega$ . Then

$$\int_{\partial\Omega} \omega = \int_{\Omega} d\omega$$

where  $\partial\Omega$  is oriented positively relative to  $\omega$ —i.e. each curve in  $\partial\Omega$  is parametrized so that  $\Omega$  lies counterclockwise from (i.e. to the left of) the parametrization's tangent vector at each point in the curve.

Green's Theorem has many interesting consequences. One of them is an interesting formula for computing area of two dimensional regions with line integrals.

**Corollary 6.23.** *Let  $\Omega \subset \mathbf{R}^2$  be as in the Green's theorem. Then*

$$\text{Vol } \Omega = \frac{1}{2} \int_{\partial\Omega} x dy - y dx.$$

**6.5. Postscript: pullbacks of 1-forms.** One can conveniently restate the definition of line integral by defining what it means to 'compose' a 1-form with a function.

**Definition 6.24.** *Let  $U \subset \mathbf{R}^k$ ,  $V \subset \mathbf{R}^n$  be open. Given a  $\mathcal{C}^1$  function  $\Phi : U \rightarrow V$  and a 1-form  $\omega = \sum_{i=1}^n \omega_i dx_i$  on  $V$ , the **pullback**  $\Phi^*\omega$  of  $\omega$  by  $\Phi$  is the 1-form on  $U$  given by*

$$\Phi^*\omega := \sum_{i=1}^n (\omega_i \circ \Phi) d\Phi_i$$

where  $\Phi_i$  is the  $i^{\text{th}}$  component of  $\Phi$ .

Suppose in particular that  $\gamma : [a, b] \rightarrow \mathbf{R}^n$  is a  $\mathcal{C}^1$  curve and  $\omega = \sum_{i=1}^n \omega_i dx_i$  is a  $\mathcal{C}^1$  1-form on  $\mathbf{R}^n$ . Then

$$\gamma^*\omega = \sum_i (\omega_i \circ \gamma) d\gamma_i = \sum_i (\omega_i \circ \gamma) \gamma'_i(t) dt = \omega(\gamma(t)) \gamma'(t) dt$$

and therefore  $\int_\gamma \omega = \int_a^b \gamma^*\omega$ .

## 7. CROSS PRODUCT

**Definition 7.1.** *The cross product of two vectors  $\mathbf{v}, \mathbf{w} \in \mathbf{R}^2$  is*

$$\mathbf{v} \times \mathbf{w} := \det(\mathbf{v}, \mathbf{w}).$$

*The cross product of two vectors  $\mathbf{v}, \mathbf{w} \in \mathbf{R}^3$  is*

$$\mathbf{v} \times \mathbf{w} := \begin{vmatrix} \mathbf{e}_1 & v_1 & w_1 \\ \mathbf{e}_2 & v_2 & w_2 \\ \mathbf{e}_3 & v_3 & w_3 \end{vmatrix} := \mathbf{e}_1 \begin{vmatrix} v_2 & w_2 \\ v_3 & w_3 \end{vmatrix} - \mathbf{e}_2 \begin{vmatrix} v_1 & w_1 \\ v_3 & w_3 \end{vmatrix} + \mathbf{e}_3 \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} = \begin{bmatrix} v_2 w_3 - w_2 v_3 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{bmatrix}.$$

Since cross product in  $\mathbf{R}^2$  is just another way of saying 'determinant', I dwell from here on out online on cross product in  $\mathbf{R}^3$ . Some basic properties of cross product, all descending straightforwardly from the definition and from properties of determinants, are as follows.

**Proposition 7.2.** *For any  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{R}^3$  and any scalar  $c \in \mathbf{R}$ , we have*

- (1)  $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \det(\mathbf{u}, \mathbf{v}, \mathbf{w})$ ;
- (2)  $(c\mathbf{v}) \times \mathbf{w} = c(\mathbf{v} \times \mathbf{w})$ ;
- (3)  $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}$ ;
- (4)  $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$ ;
- (5)  $\mathbf{v} \times \mathbf{v} = \mathbf{0}$ ;
- (6)  $\mathbf{v} \times \mathbf{w}$  is orthogonal to  $\mathbf{v}$  and  $\mathbf{w}$ ;

Taking  $\mathbf{u} = \mathbf{v} \times \mathbf{w}$  in the first item of the proposition, we see that on the one hand the volume of the three dimensional parallelotope  $\Omega$  determined by  $\mathbf{v}, \mathbf{w}, \mathbf{v} \times \mathbf{w}$  is given by

$$\text{Vol } \Omega = \|\mathbf{v} \times \mathbf{w}\|^2$$

But on the other hand, the last item of the proposition tells us that  $\mathbf{v} \times \mathbf{w}$  is perpendicular to the two dimensional ‘base’ parallelogram determined by  $\mathbf{v}$  and  $\mathbf{w}$ . Our intuition and experience suggest that  $\text{Vol} \Omega$  should therefore be equal to the area of this base times the length of the remaining vector  $\|\mathbf{v} \times \mathbf{w}\|$ . We are thus led to the following definition.

**Definition 7.3.** For any  $\mathbf{v}, \mathbf{w} \in \mathbf{R}^3$ , the area of the parallelogram

$$P := \{s\mathbf{v} + t\mathbf{w} \in \mathbf{R}^3 : 0 \leq s, t \leq 1\}$$

is the quantity

$$\text{Area } P := \|\mathbf{v} \times \mathbf{w}\|.$$

The area of the base parallelogram  $P$  should also just be base times height, i.e.  $\text{Area } P = \|\mathbf{v}\| \|\mathbf{w}\| |\sin \theta|$  where  $\theta$  is the angle from  $\mathbf{v}$  to  $\mathbf{w}$ . This turns out to be consistent with the formula from first semester  $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos \theta$  that we used to define  $\cos \theta$ .

**Theorem 7.4.** For any  $\mathbf{v}, \mathbf{w} \in \mathbf{R}^3$  we have

$$\|\mathbf{v} \times \mathbf{w}\|^2 + (\mathbf{v} \cdot \mathbf{w})^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2.$$

There are a number of other identities involving cross products which I omit here. If you’re curious, section 1.5 in Shifrin’s book and the accompanying exercises give a few of these. I will close with another definition (which you are not responsible for!) that one arrives at via thought experiment similar to the one I just used to motivate Definition 7.3.

**Definition 7.5.** Suppose that  $k \leq n$  and that  $P \subset \mathbf{R}^n$  is the  $k$ -dimensional polytope determined by vectors  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbf{R}^n$ . The  $k$ -dimensional volume of  $P$  is the quantity

$$\text{Vol}_k P := \sqrt{\det(V^T V)} = \left| \begin{array}{cccc} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 & \dots & \mathbf{v}_1 \cdot \mathbf{v}_k \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 & \dots & \mathbf{v}_2 \cdot \mathbf{v}_k \\ \vdots & \vdots & & \vdots \\ \mathbf{v}_k \cdot \mathbf{v}_1 & \mathbf{v}_k \cdot \mathbf{v}_2 & \dots & \mathbf{v}_k \cdot \mathbf{v}_k \end{array} \right|^{1/2},$$

where  $V$  is the  $n \times k$  matrix with columns  $\mathbf{v}_1, \dots, \mathbf{v}_k$ .

The matrix  $V^T V$  in this definition is known as a *Gram matrix*. Note that it is symmetric. As the definition implies, its determinant is provably always non-negative, though I do not give the argument here. Often, this is discussed along with the so-called ‘spectral theorem’ for symmetric matrices (i.e. self-adjoint operators, for the illuminati among us).

When  $n = 3$  and  $k = 2$ , Definition 7.5 gives the same result as Definition 7.3. This isn’t obvious, but if you try to check it, you’ll find it follows quickly from Theorem 7.4.

## 8. INTEGRATING OVER SURFACES

**Definition 8.1.** Given  $k \leq n$ , let  $U \subset \mathbf{R}^k$  be open and  $G : U \rightarrow \mathbf{R}^n$  be  $C^1$ . Suppose that  $G$  is injective and that for each  $x \in U$  we have  $\dim \text{col } DG(x) = k$ . Then we call  $G(U)$  a parametrized  $k$ -dimensional submanifold of  $\mathbf{R}^n$ . For each  $x \in U$ , we call the column space of  $DG(x)$  the tangent space to  $G(U)$  at the point  $G(x)$ .

Note that the tangent space at  $G(x)$  does not usually contain  $G(x)$ , though it is generally drawn that way in pictures. The point is that the tangent space is a subspace of  $\mathbf{R}^n$  that consists of those displacements *which when added to  $G(x)$  remain close to  $G(U)$* . To turn the displacements into actual *points* near  $G(U)$  one must therefore translate by  $G(x)$ .

We will call two dimensional submanifolds *surfaces* in  $\mathbf{R}^n$ , and here we will stick to ambient dimension  $n = 3$ .

**Definition 8.2.** Let  $U \subset \mathbf{R}^k$  be open and  $G : U \rightarrow \mathbf{R}^n$  be a smooth parametrized  $k$ -dimensional submanifold. If  $\tilde{U} \subset \mathbf{R}^k$  is another open set and  $\Phi : \tilde{U} \rightarrow U$  is a diffeomorphism, then we call  $\tilde{G} = G \circ \Phi : \tilde{U} \rightarrow \mathbf{R}^n$  a **reparametrization** of  $G(U)$ .

**Definition 8.3.** Let  $U \subset \mathbf{R}^2$  be open, and let  $G : U \rightarrow \mathbf{R}^3$ ,  $(u, v) \mapsto (x, y, z)$  be a parametrized surface. If  $\Omega \subset U$  is a region, we define the **surface area** of  $G(\Omega)$  to be the quantity

$$\int_{G(\Omega)} dS := \iint_{\Omega} \left\| \frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v} \right\| du dv.$$

Moreover, if  $f : \mathbf{R}^3 \rightarrow \mathbf{R}$  is continuous, then we define the **integral of  $f$  over  $\Omega$  with respect to surface area** to be the quantity

$$\int_{G(\Omega)} f dS := \iint_{\Omega} f \circ G \left\| \frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v} \right\| du dv.$$

The notation ‘ $dS$ ’ in this definition is basically shorthand for  $\left\| \frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v} \right\| du dv$ . It is often called the *surface area element* for  $G(U)$ . As with arc-length, surface area integrals are unchanged by reparametrization.

**Theorem 8.4.** Let  $U, \tilde{U} \subset \mathbf{R}^2$  be open, with a diffeomorphism  $\Phi : \tilde{U} \rightarrow U$ . Let  $\Omega \subset U$ ,  $\tilde{\Omega} \subset \tilde{U}$  be regions such that  $\Omega = \Phi(\tilde{\Omega})$ . If  $G : U \rightarrow \mathbf{R}^3$  is a parametrized smooth surface and  $\tilde{G} = G \circ \Phi : \tilde{U} \rightarrow \mathbf{R}^3$  is a reparametrization of  $G(U)$ , then

$$\iint_{\tilde{\Omega}} \left\| \frac{\partial \tilde{G}}{\partial \tilde{u}} \times \frac{\partial \tilde{G}}{\partial \tilde{v}} \right\| d\tilde{u} d\tilde{v} = \iint_{\Omega} \left\| \frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v} \right\| du dv.$$

Now we come to the classical versions of the fundamental theorem of calculus as it relates to surface integrals. One is more ‘two dimensional’ concerning a smooth surface in  $\mathbf{R}^3$  bounded by a smooth closed curve. The other is more ‘two dimensional’ concerning a region in  $\mathbf{R}^3$  bounded by a smooth surface. Both involve smooth vector fields on  $\mathbf{R}^3$ , i.e.  $C^1$  functions  $F : \mathbf{R}^3 \rightarrow \mathbf{R}^3$ , which one thinks of as giving a vector  $F(x) = F_1(x)\mathbf{e}_1 + F_2(x)\mathbf{e}_2 + F_3(x)\mathbf{e}_3$  emanating from the point  $x$ .

**Definition 8.5.** Let  $F : \mathbf{R}^3 \rightarrow \mathbf{R}^3$  be a smooth vector field,  $G : U \rightarrow \mathbf{R}^3$  be a smooth surface oriented by the normal vector field  $N : G(U) \rightarrow \mathbf{R}^3$  and  $\Omega \subset U$  be a region such that  $F$  is defined about  $G(\Omega)$ .

- The divergence of  $F$  is the (continuous) scalar field  $\nabla \cdot F := \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3}$ .
- The curl of  $F$  is the (continuous) vector field  $\nabla \times F := \begin{bmatrix} \frac{\partial F_2}{\partial x_3} - \frac{\partial F_3}{\partial x_2} \\ \frac{\partial F_3}{\partial x_1} - \frac{\partial F_1}{\partial x_3} \\ \frac{\partial F_1}{\partial x_2} - \frac{\partial F_2}{\partial x_1} \end{bmatrix}$ .
- The flux of  $F$  through  $G(\Omega)$  is the quantity  $\int_{G(\Omega)} F \cdot N dS = \iint_{\Omega} \det \begin{bmatrix} F \circ G & \frac{\partial G}{\partial u} & \frac{\partial G}{\partial v} \end{bmatrix} du dv$ .

The most direct generalization of Green’s Theorem to surfaces in  $\mathbf{R}^3$  is

**Theorem 8.6** (Classical Stokes Theorem). Let  $G : U \rightarrow \mathbf{R}^3$  be a smooth parametrized surface and  $\Omega \subset U$  be a region with  $\partial\Omega$  equal to a finite union of piecewise smooth closed

curves. Then for any smooth vector field  $F : \mathbf{R}^3 \rightarrow \mathbf{R}^3$  defined about the surface  $G(\Omega)$ , we have

$$\int_{G(\partial\Omega)} F \cdot T \, ds = \int_{G(\Omega)} (\nabla \times F) \cdot N \, dS.$$

Here  $N$  is the positive unit normal vector field on  $G(\Omega)$  and  $T$  is the positive unit tangent vector to  $G(\partial\Omega)$ .

The notion of ‘positive’ for  $N$  and  $T$  is relative to the parametrization  $G$ . Namely  $N$  is chosen to be a positive multiple of  $\frac{\partial G}{\partial u} \times \frac{\partial G}{\partial v}$  and  $T$  a positive multiple of  $(\gamma \circ G)'$  where  $\gamma : \mathbf{R} \rightarrow \mathbf{R}^2$  smoothly parametrizes  $\partial\Omega$  so that  $\Omega$  lies to the left of  $\gamma'$ .

The following is known alternately as Gauss’ Theorem or the Divergence Theorem.

**Theorem 8.7.** *Let  $U \subset \mathbf{R}^3$  be an open set such that  $\Omega = \bar{U} := U \cup \partial U$  is a region and  $\partial U$  is a finite union of piecewise smooth surfaces. Then for any smooth vector field  $F : \mathbf{R}^3 \rightarrow \mathbf{R}^3$  defined about  $\Omega$ , we have*

$$\int_{\partial\Omega} F \cdot N \, dS = \int_{\Omega} \nabla \cdot F \, dV,$$

where  $N$  is the outward unit normal vector to  $\partial\Omega$ .

## 9. SOME MULTI-LINEAR ALGEBRA

I introduced the cross-product above as a means of assigning an ‘area’ to the parallelogram determined by two vectors in  $\mathbf{R}^3$ . In this section I give an alternative point of view on this issue in greater generality, namely I introduce a way of assigning a ‘size’ to the parallelogram determined by  $k$  vectors in  $\mathbf{R}^n$ .

**Definition 9.1.** *Let  $0 < k \leq n$  be integers. An alternating  $k$ -linear form on  $\mathbf{R}^n$  is a function  $\omega : (\mathbf{R}^n)^k \rightarrow \mathbf{R}$  such that for all  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbf{R}^n$ ,*

- $\omega(\mathbf{v}_1, \dots, \mathbf{v}_k)$  is linear with respect to each argument  $\mathbf{v}_j$ ;
- $i \neq j$  implies that  $\omega(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_k) = -\omega(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_k)$ .

For the sake of brevity here, I will write ‘linear  $k$ -form’ instead of ‘alternating linear  $k$ -form’. Beware, though, that this conflicts with fairly standard terminology used elsewhere in the mathematical literature.

**Example 9.2.** *Covectors are linear 1-forms on  $\mathbf{R}^n$  because the second condition in the definition is vacuous when  $k = 1$ . The ( $n$ -dimensional) determinant function  $\det : (\mathbf{R}^n)^n \rightarrow \mathbf{R}$  is a linear  $n$ -form. Indeed, the definition of  $k$ -form is identical to the definition of determinant except that there’s no ‘normalization hypothesis’.*

*Cross product gives a means to build linear 2-forms on  $\mathbf{R}^3$ . Namely, given a vector  $\mathbf{u} \in \mathbf{R}^3$ , the function*

$$\omega(\mathbf{v}, \mathbf{w}) = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \det(\mathbf{u}, \mathbf{v}, \mathbf{w}).$$

*is a linear 2-form on  $\mathbf{R}^3$ . From a mathematical standpoint (as we shall see), this is why flux integrals arise.*

**Proposition 9.3.** *Fix  $k \leq n$  and let  $\alpha, \beta$  be linear  $k$ -forms and  $c \in \mathbf{R}$  be a scalar. Then  $\alpha + \beta$  and  $c\alpha$  are also linear  $k$ -forms.*

In other words the set of all linear  $k$ -forms on  $\mathbf{R}^n$  is a vector space. We’ll see below that it’s finite dimensional.

The main way to construct linear  $k$ -forms is via a kind of multiplication operation.

**Definition 9.4.** Let  $\alpha_1, \dots, \alpha_k : \mathbf{R}^n \rightarrow \mathbf{R}$  be covectors on  $\mathbf{R}^n$ . Then the wedge product of  $\alpha_1, \dots, \alpha_k$  is the function  $\alpha_1 \wedge \dots \wedge \alpha_k : (\mathbf{R}^n)^k \rightarrow \mathbf{R}$  given by

$$\alpha_1 \wedge \dots \wedge \alpha_k(\mathbf{v}_1, \dots, \mathbf{v}_k) := \det \begin{bmatrix} \alpha_1(\mathbf{v}_1) & \dots & \alpha_1(\mathbf{v}_k) \\ \vdots & \ddots & \vdots \\ \alpha_k(\mathbf{v}_1) & \dots & \alpha_k(\mathbf{v}_k) \end{bmatrix}$$

**Proposition 9.5.** The wedge product of  $k$  covectors on  $\mathbf{R}^n$  is a linear  $k$ -form on  $\mathbf{R}^n$ .

**Example 9.6.** For any  $n > 0$ , the wedge product

$$dx_1 \wedge \dots \wedge dx_n$$

is the ( $n$ -dimensional) determinant.

If  $\mathbf{u} \in \mathbf{R}^3$  is a given vector and  $\omega(\mathbf{v}, \mathbf{w}) = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$  is the associated linear 2-form, then one can check that

$$\omega = u_1 dx_2 \wedge dx_3 + u_2 dx_3 \wedge dx_1 + u_3 dx_1 \wedge dx_2.$$

The lack of a minus sign in the second term comes from the fact that I put  $dx_3$  before  $dx_1$  in the wedge product.

The reason for the word ‘product’ in the term ‘wedge product’ is that many of the usual properties of multiplication apply. For instance, the properties of determinants imply that the distributive law holds:

$$(\alpha_1 + \beta_1) \wedge \alpha_2 \wedge \dots \wedge \alpha_k = \alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_k + \beta_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_k.$$

As with cross-products, though, we have *skew-commutativity*:

$$\dots \wedge \alpha_i \wedge \dots \wedge \alpha_j \wedge \dots = -(\dots \wedge \alpha_j \wedge \dots \wedge \alpha_i \wedge \dots)$$

This implies among other things that if  $\alpha_i = \alpha_j$  when  $i \neq j$ , then the entire wedge product vanishes.

Shifrin uses a nice piece of short-hand. If  $I = (i_1, \dots, i_k)$  is a set of indices  $i_j \in \{1, \dots, k\}$  in  $\mathbf{R}^n$ , then  $dx_I$  denotes the linear  $k$ -form

$$dx_{i_1} \wedge \dots \wedge dx_{i_k}.$$

The  $k$ -tuple  $I$  is called a *multi-index of length  $k$* . So for example, in Shifrin’s notation, we have that

$$\det = dx_{(1, \dots, n)}.$$

**Theorem 9.7.** Any linear  $k$ -form  $\omega$  on  $\mathbf{R}^n$  can be expressed in exactly one way as follows:

$$\omega = \sum_I \omega_I dx_I$$

where  $\omega_I \in \mathbf{R}$  are scalars and the sum is taken over all strictly increasing multi-indices  $I$  of length  $k$ .

The big takeaway here is that *all* linear  $k$ -forms can be built up from covectors. By counting the terms in the sum, one finds that for any fixed  $k \leq n$ , the vector space of linear  $k$ -forms has dimension

$$\binom{n}{k} := \frac{n!}{(n-k)!k!}.$$



It now makes sense to denote the set of all linear  $k$ -forms on  $\mathbf{R}^n$  by  $\bigwedge^k(\mathbf{R}^{n*})$ .

Note that in the particular case  $k = n$ , Theorem 9.7 says that  $\omega$  is a constant multiple of the  $n$ -form  $dx_1 \wedge \cdots \wedge dx_n$ . The value of this constant is given by the next result of this section. It can be proven using cofactor expansion plus induction or alternatively using the definition of determinant.

**Theorem 9.8.** *Let  $A$  be an  $n \times n$  matrix with entries  $a_{ij} \in \mathbf{R}$  and  $\alpha_1, \dots, \alpha_n : \mathbf{R}^n \rightarrow \mathbf{R}$  be the covectors with coefficients taken from the columns of  $A$ :*

$$\alpha_i = \sum_{j=1}^n a_{ij} dx_j.$$

Then

$$\alpha_1 \wedge \cdots \wedge \alpha_n = (\det A) dx_1 \wedge \cdots \wedge dx_n.$$

Another consequence of Theorem 9.7 is that we now have a way of taking wedge products of higher degree forms. Namely, if  $\omega$  is a linear  $k$ -form and  $\eta$  a linear  $\ell$ -form on  $\mathbf{R}^n$ , then we can use the theorem to write

$$\omega = \sum_{|I|=k}^l \omega_I dx_I \quad \text{and} \quad \eta = \sum_{|J|=\ell}^l \eta_J dx_J.$$

Then we define

$$\omega \wedge \eta := \sum_{|I|=k}^l \sum_{|J|=\ell}^l \omega_I \eta_J dx_I \wedge dx_J,$$

where

$$dx_I \wedge dx_J := dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}.$$

This extended notion of wedge product behaves in many ways like ordinary multiplication.

**Proposition 9.9.** *Let  $\omega$ ,  $\eta$ , and  $\nu$  be linear  $\ell$ ,  $k$ , and  $m$ -forms on  $\mathbf{R}^n$ . Then*

- $(\omega \wedge \eta) \wedge \nu = \omega \wedge (\eta \wedge \nu)$ ;
- $\omega \wedge \eta = (-1)^{\ell k} \eta \wedge \omega$ ;
- (assuming  $k = m$ )  $\omega \wedge (\eta + \nu) = \omega \wedge \eta + \omega \wedge \nu$ .

## 10. DIFFERENTIAL $k$ -FORMS

**Definition 10.1.** *Let  $0 < k \leq n$  be integers. A differential  $k$ -form on an open set  $U \subset \mathbf{R}^n$  is a function  $\omega : U \rightarrow \bigwedge^k(\mathbf{R}^{n*})$ . More concretely,*

$$\omega = \sum_I^l \omega_I dx_I$$

where the sum is taken over strictly increasing multi-indices of length  $k$  and the coefficients  $\omega_I : U \rightarrow \mathbf{R}$  are scalar-valued functions. We say that  $\omega$  is  $C^1$  (or continuous, etc) if the coefficients  $\omega_I$  are each  $C^1$  (or continuous, etc).

Let us extend this definition slightly to include  $k = 0$ . A *differential 0-form* on  $\mathbf{R}^n$  is just a scalar-valued function  $f : \mathbf{R}^n \rightarrow \mathbf{R}$

As with linear  $k$ -forms, one can add differential  $k$ -forms to each other, multiply them by scalars, and wedge them with differential  $\ell$ -forms. The same algebraic properties apply. We

will be interesting transforming differential  $k$ -forms in two new ways. Note that from here on, I will write simply ‘ $k$ -form’, dropping the ‘differential’.

**10.1. Exterior derivative.** We first extend the notion of differential from functions to  $k$ -forms.

**Definition 10.2.** Let  $U \subset \mathbf{R}^n$  be open and  $\omega = \sum_I' \omega_I dx_I$  be a  $C^1$   $k$ -form on  $U$ . The exterior derivative of  $\omega$  is the  $k+1$ -form

$$d\omega := \sum_I' d\omega_I \wedge dx_I.$$

Exterior derivative is a hard thing to motivate on visual grounds. Basically, the definition is what it needs to be to make the generalized Stokes’ Theorem (we’ll get to it) work. On the positive side, it has a fairly simple definition, it’s easy to compute in examples, and it has some nice algebraic properties to recommend it.

**Proposition 10.3.** Let  $\omega$  and  $\eta$  be  $C^1$   $k$  and  $\ell$  forms, respectively, on  $\mathbf{R}^n$ .

- If  $k = \ell$ , then  $d(\eta + \omega) = d\eta + d\omega$ .
- $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta$ .
- If  $\omega$  is  $C^2$ , then  $d^2\omega := d(d\omega) = 0$ .

The fact that  $d^2 = 0$  is somewhat analogous to the geometric fact that when some curve in  $\mathbf{R}^3$  is the boundary of a surface, then the curve is closed. I.e. the ‘boundary of the boundary curve’ is empty. Though it might seem like nothing more than a curiosity here, the ‘boundary of the boundary is zero’ phenomena occurs frequently in mathematics and is actually quite important.

**10.2. Pullback.** Next we extend the notion of composition of functions. Suppose we have a  $C^1$  function  $G : \mathbf{R}^n \rightarrow \mathbf{R}^m$  and that  $\omega$  is a  $k$ -form on  $\mathbf{R}^m$ . If  $x \in \mathbf{R}^n$  lies in the domain of  $G$ , and  $\mathbf{v}_1, \dots, \mathbf{v}_k$  are displacement vectors ‘at  $x$ ’, then as we learned in the first semester, we can ‘push forward’ these vectors to get displacement vectors  $\mathbf{w}_1, \dots, \mathbf{w}_k$  at the point  $y = G(x) \in \mathbf{R}^m$ . Namely, we let  $\mathbf{w}_j := DG(x)\mathbf{v}_j$  be the image of  $\mathbf{v}_j$  by the derivative of  $G$  at  $x$ . By this mechanism we can use  $\omega$  as if it were a  $k$ -form at  $x$  instead of at  $y$ .

**Theorem 10.4.** Let  $G : \mathbf{R}^n \rightarrow \mathbf{R}^m$  be a  $C^1$  mapping and  $\omega$  be a  $k$ -form on  $\mathbf{R}^m$ . Then there is a  $k$ -form  $G^*\omega$  on  $\mathbf{R}^n$  given at each point  $x \in \mathbf{R}^n$  and for any displacements  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbf{R}^n$  by

$$G^*\omega(x)(\mathbf{v}_1, \dots, \mathbf{v}_k) := \omega(G(x))(DG(x)\mathbf{v}_1, \dots, DG(x)\mathbf{v}_k).$$

If moreover,  $\omega = \sum_I' \omega_I dy_I$ , then

$$G^*\omega := \sum_I' (\omega_I \circ G) dG_I,$$

where for each multi-index  $I = (i_1, \dots, i_k)$ ,

$$dG_I := G^*dy_I := dG_{i_1} \wedge \dots \wedge dG_{i_k}.$$

Of course, I’m being a bit careless with domains in the statement of this theorem. The form  $G^*\omega$  is only defined at those  $x \in \mathbf{R}^n$  that belong to the domain of  $G$  and whose images  $y = G(x)$  belong to the domain of  $\omega$ .

**Definition 10.5.** The form  $G^*\omega$  in Theorem 10.4 is called the pullback of  $\omega$  by  $G$ .

One can verify in a fairly straightforward fashion that pullback of forms interacts well with all the other operations we have introduced.

**Proposition 10.6.** Let  $G : \mathbf{R}^n \rightarrow \mathbf{R}^m$  be  $C^1$  and  $\omega$  and  $\eta$  be  $k$  and  $\ell$  forms, respectively, on  $\mathbf{R}^m$ . Then

- (assuming  $k = \ell$ )  $G^*(\omega + \eta) = G^*\omega + G^*\eta$ ;
- $G^*(\omega \wedge \eta) = G^*\omega \wedge G^*\eta$ ;
- (assuming  $\omega$  is  $C^1$ )  $G^*(d\omega) = dG^*\omega$ ;
- if  $F : \mathbf{R}^p \rightarrow \mathbf{R}^n$  is another  $C^1$  function, then  $(G \circ F)^*\omega = F^*G^*\omega$ .

The final fact I record here follows from Theorem 9.8 and suggests that pullback might be connected with change of variables.

**Corollary 10.7.** Suppose that  $G : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is  $C^1$ . Then writing  $y = G(x)$ , we have

$$G^*(dy_1 \wedge \cdots \wedge dy_n) = (\det DG) dx_1 \wedge \cdots \wedge dx_n.$$

**10.3. Integrating over parametrized subsets of  $\mathbf{R}^n$ .** The idea here is that  $k$ -forms on  $\mathbf{R}^n$  should be integrated over  $k$ -dimensional subsets of  $\mathbf{R}^n$  with the result being some number that measures the size or (perhaps better) value of the subset. When  $k = n$ , this idea just reduces to ordinary integration.

**Definition 10.8.** Let  $\omega = f dx_1 \wedge \cdots \wedge dx_n$  be a continuous  $n$ -form on  $\mathbf{R}^n$  and  $\Omega \subset \mathbf{R}^n$  be a region. Then the integral of  $\omega$  over  $\Omega$  is the quantity

$$\int_{\Omega} \omega := \int_{\Omega} f dV.$$

When  $k < n$ , we reduce to the above situation by parametrizing the set and pulling back the form.

**Definition 10.9.** Suppose that  $\omega$  is a continuous  $k$ -form on  $\mathbf{R}^n$ , that  $G : \mathbf{R}^k \rightarrow \mathbf{R}^n$  is a  $C^1$  function and that  $\Omega \subset \mathbf{R}^k$  is a region in the domain of  $G$ . Then the integral of  $\omega$  over the parametrized subset  $G(\Omega) \subset \mathbf{R}^n$  is the quantity

$$\int_{G(\Omega)} \omega := \int_{\Omega} G^*\omega.$$

**10.4.  $k$ -dimensional manifolds with boundary.** Our definition of smooth parametrized  $k$ -dimensional submanifold has a serious limitation: many interesting  $k$ -dimensional subsets of  $\mathbf{R}^n$  are difficult to describe with a single parametrization. We need to allow for different parametrizations of different parts of the manifold.

**Definition 10.10.** A subset  $M \subset \mathbf{R}^n$  is a smooth  $k$ -dimensional manifold if for each  $y \in M$ , there exists a neighborhood  $W \subset \mathbf{R}^n$  of  $y$  such that  $M \cap W$  is the image  $G(U)$  of a smooth parametrized  $k$ -dimensional submanifold  $G : U_{open} \subset \mathbf{R}^k \rightarrow \mathbf{R}^n$ .

The parametrization  $G : U \rightarrow M \cap W$  is said to be a coordinate chart for  $M$ . If  $G, \tilde{G}$  are two coordinate charts for  $M$ , then the composition  $G_j^{-1} \circ G_i : \mathbf{R}^k \rightarrow \mathbf{R}^k$  is called a transition function.

Actually, the definition should include another technical hypothesis to prevent the manifold  $M$  from wrapping around to ‘land on itself’ at some point. Shifrin discusses this issue in section 8.5 if you’d like to hear more, but I will ignore it here. Another issue, one I won’t avoid, is that of orientation. A parametrized manifold is always orientable (whatever this means), but a manifold as I have just defined it need not be. For instance, the Moebius strip in  $\mathbf{R}^3$  is a two dimensional manifold that requires at least two coordinate charts, and it is not orientable.

**Definition 10.11.** A  $C^1$  mapping  $\Phi : \mathbf{R}^k \rightarrow \mathbf{R}^k$  is orientation-preserving if  $\det D\Phi(x) > 0$  for every  $x$  in the domain of  $\Phi$  and orientation-reversing if  $\det D\Phi(x) < 0$  for every  $x$  in the domain of  $\Phi$ .

**Definition 10.12.** A  $k$ -dimensional manifold  $M \subset \mathbf{R}^n$  is orientable if there exist coordinate charts  $G_i : U_i \subset \mathbf{R}^k \rightarrow \mathbf{R}^n$ ,  $1 \leq i \leq \ell$  whose images  $G(U_i)$  cover  $M$  (i.e.  $M \subset \bigcup_{i=1}^{\ell} G(U_i)$ ) and that each ‘transition function’  $G_j^{-1} \circ G_i$  is orientation-preserving.

Note that this definition is vacuous when there is only one chart (i.e. when  $M$  is a parametrized  $k$ -dimensional submanifold). In class I ‘oriented’ curves and surfaces in rather different terms, using tangent (for curves) or normal (for surfaces) vectors. Nevertheless, it can be shown that these notions of orientation are consistent with each other. For instance, if  $M \subset \mathbf{R}^3$  is a surface, and  $G_1 : U_1 \rightarrow \mathbf{R}^n$ ,  $G_2 : U_2 \rightarrow \mathbf{R}^n$  are charts for  $M$ , then each chart gives us a normal vector at each point on its image  $G_j(U_j) \subset M$ . The condition on  $\det D(G_j^{-1} \circ G_i)(x)$  implies that these normal vectors are positive multiples of each other wherever the images  $G_1(U_1), G_2(U_2)$  overlap. At any rate, any orientable submanifold  $M \subset \mathbf{R}^n$  can be given exactly two orientations. If  $G_1, \dots, G_\ell$  is one of them, then the other one is given by any collection of charts  $\tilde{G}_1, \dots, \tilde{G}_m$  whose images cover  $M$  and such that transitions  $\tilde{G}_j^{-1} \circ G_i$  between charts in different collections are all orientation-reversing.

Since a  $k$ -dimensional manifold  $M$  isn’t necessarily the image of a single coordinate chart, there remains some question about what it means to integrate a continuous  $k$ -form  $\omega$  over a reasonable compact subset  $\Omega \subset M$ . *In practice* one divides up  $\Omega$  into pieces, each contained in the image of a single coordinate chart, integrates over each piece separately and then adds up the results. In order to take care of sign ambiguities, one must fix an orientation for  $M$  and make sure all coordinate charts involved are compatible with that orientation.<sup>2</sup> Practice and theory diverge, however, when it comes to integrating forms on manifolds. For theoretical purposes it’s better to cut the manifold into pieces with functions rather than ‘scissors’.

**Definition 10.13.** Let  $M \subset \mathbf{R}^n$  be a  $k$ -dimensional manifold,  $\Omega \subset \mathbf{R}^n$  be a compact set and  $W_j \subset \mathbf{R}^n$ ,  $j = 1, \dots, m$  be open sets that cover  $\Omega$ . A partition of unity for  $\Omega$  relative to  $W_1, \dots, W_m$  is a collection of  $C^1$  functions  $\psi_j : \mathbf{R}^n \rightarrow \mathbf{R}$  such that

- For each  $j = 1, \dots, m$ , the function  $\psi_j$  vanishes outside some compact subset of  $W_j$ ;
- $\sum_j \psi_j(x) = 1$  for each  $x \in \Omega$ .

It turns out given  $\Omega$  and  $W_1, \dots, W_m$  that there always exists a partition of unity. These allow us (in theory) to integrate  $k$ -forms as follows: suppose for each  $j = 1, \dots, m$  that

<sup>2</sup>Integrating a  $k$ -form over a non-orientable manifold isn’t really a well-defined thing even though it’s still perfectly reasonable to ask about the volume of such a manifold—this is a big difference between integration of forms and integration with respect to volume.

$G_j : U_j \subset \mathbf{R}^k \rightarrow \mathbf{R}^n$  are (correctly oriented) coordinate chart for the  $M \cap W_j$  with images  $G(U_j) \subset W_j$  covering  $\Omega$ . Then we have  $\omega = \sum \psi_j \omega_j$  at each point  $x \in U_j$ . Since  $\psi_j$  vanishes outside  $W_j$ , we can treat  $\psi_j \omega_j$  as a form on  $M \cap W_j$  only. Therefore it makes sense to define

$$\int_{\Omega} \omega = \sum_{j=1}^m \int_{\Omega \cap W_j} \psi_j \omega_j = \int_{\Omega_j} G_j^*(\psi_j \omega_j)$$

where  $\Omega_j = G_j^{-1}(\Omega \cap W_j) \subset \mathbf{R}^k$ . I should add the (non-trivial) requirement that  $G_j^{-1}(\Omega)$  is a region in  $\mathbf{R}^k$  so that the final integral makes sense, but lets not worry too much about that here. Dissecting a form with a partition of unity is quite compatible with exterior differentiation: if  $\omega = d\eta$  is exact, then because  $\sum \psi_j \equiv 1$  on  $\Omega$ , we find that on  $\Omega$

$$\sum_{j=1}^m d(\psi_j \eta) = \sum_{j=1}^m (d\psi_j \wedge \eta + \psi_j d\eta) = d\left(\sum_{j=1}^m \psi_j\right) \wedge \eta + \sum_{j=1}^m \psi_j d\eta = \sum_{j=1}^m \psi_j d\eta$$

because  $d(1) = 0$ .

Stokes Theorem concerns integrating forms over compact sets with ‘nice’ boundaries. To make nice precise, I introduce the following notation. If  $U \subset \mathbf{R}^k$  is open, then I set

$$U^+ := \{x = (x_1, \dots, x_k) \in U : x_k \geq 0\}$$

**Definition 10.14.** *Let  $M \subset \mathbf{R}^n$  be a  $k$ -dimensional manifold. A compact subset  $Y \subset M$  is a  $k$ -dimensional manifold-with-boundary if for every  $y \in Y$  there is a neighborhood  $W \ni y$  and a chart  $G : U_{\text{open}} \subset \mathbf{R}^k \rightarrow M \cap W$  such that  $Y \cap W$  is equal to either  $G(U)$  or  $G(U^+)$ . In the latter case, all points  $y = G(x_1, \dots, x_{k-1}, 0)$  that are the images of points  $x \in U$  with vanishing last coordinate are called boundary points of  $Y$ . The boundary  $\partial Y$  of  $Y$  is the set of all boundary points of  $Y$*

Note that the present notion of (manifold) boundary is more restrictive than that of (topological) boundary that I introduced last semester. For instance, if  $Y$  is the upper half of the unit sphere in  $\mathbf{R}^2$ , then every point in  $Y$  is a topological boundary point, but the manifold boundary of  $Y$  consists only of those points  $y \in Y$  with  $y_n = 0$ . Even so, manifold boundary points are always topological boundary points and in particular contained in  $Y$  because  $Y$  is compact.

It is important to further note that the boundary  $\partial Y$  of a  $k$ -dimensional manifold-with-boundary  $Y$  is itself a  $k - 1$ -dimensional manifold-with-boundary: if  $G : U_{\text{open}} \rightarrow M \cap W$  is a boundary chart with  $G(x) = y$ , then we can associate a  $(k-1)$ -dimensional ‘boundary’ chart  $G_{\partial} : U_{\partial} \rightarrow \partial Y \cap W$  given by

$$G_{\partial}(x_1, \dots, x_{k-1}) := G(x_1, \dots, x_{k-1}, 0)$$

for all  $(x_1, \dots, x_{k-1})$  such that  $(x_1, \dots, x_{k-1}, 0) \in U$ . For Stokes Theorem (I’m getting there), there is a correct way to orient  $\partial Y$  relative to  $Y$ . We call this the *boundary orientation*. To make a long story short, the boundary charts  $G_{\partial}$  that I just defined have the correct orientation when  $k$  is even and the incorrect orientation when  $K$  is odd. When  $Y \subset \mathbf{R}^3$  (as in all homework problems), then there are only two cases (three if you count  $k = 1$ ). If  $k = 3$ , then  $Y$  is a region in  $\mathbf{R}^3$ , and  $\partial Y$  is a surface. The correct orientation for  $\partial Y$  is the one that uses a normal vector pointing outside of  $Y$ . If  $k = 2$ , then  $Y$  is a surface oriented at each point by a choice of normal vector  $N$ , and  $\partial Y$  is a closed curve oriented by a choice of tangent vector  $T$ . Then  $\partial Y$  has the boundary orientation relative to  $Y$  if when you imagine

yourself standing up (relative to  $N$ ) and facing forward (relative to  $T$ ) at some point  $p \in \partial Y$ , then the surface  $Y$  lies to your left.

**Theorem 10.15** (Generalized Stokes Theorem). *Let  $Y \subset \mathbf{R}^n$  be a smooth compact  $k$ -dimensional manifold with boundary and  $\omega$  be a  $C^1$   $(k - 1)$ -form on  $\mathbf{R}^n$  whose domain includes  $Y$ . Then*

$$\int_{\partial Y} \omega = \int_Y d\omega,$$

where  $\partial Y$  has the boundary orientation induced by  $Y$ .