

MULTIVARIABLE ANALYSIS

DANIEL S. FREED

What follows are lecture notes from an advanced undergraduate course given at the University of Texas at Austin in Spring, 2019. The notes are rough in many places, so use at your own risk!

CONTENTS

Lecture 1: Affine geometry	4
Basic definitions	4
Affine analogs of vector space concepts	6
Ceva's theorem	8
Lecture 2: Parallelism, coordinates, and symmetry	10
Parallelism and another classical theorem	10
Bases in vector spaces	11
Linear symmetry groups	12
Affine coordinates	13
The tangent space to affine space	15
Lecture 3. Normed linear spaces	16
Basic definitions	16
Norms on $\mathbb{R}^n$	17
Lecture 4. More on normed linear spaces	20
Banach spaces	20
Examples of Banach spaces	20
Equivalent norms	22
Lecture 5. Continuous linear maps; differentiability	24
Continuous linear maps	24
Shapes and functions	26
Setting for calculus	27
Continuity and differentiability	29
Lecture 6: Computation of the differential	31
Differentiability and continuity	31
Functions of one variable	31
Computation of the differential	32
The operator d and explicit computation	34
Lecture 7: Further properties of the differential	36
Chain rule	36
Mean value inequality	39

Lecture 8: Differentials and local extrema; inner products	41
Continuous differentiability	41
Local extrema and critical points	43
Real inner product spaces	45
Lecture 9: Inner product spaces; gradients	47
More on inner product spaces	47
Gradient	49
Lecture 10: Differential of the length function; contraction mappings	51
The first variation formula	51
Fixed points and contractions	56
Lecture 11: Inverse function theorem	58
Invertibility is an open condition	58
C^k functions	59
Inverse function theorem	60
Lecture 12: Examples and computations	64
Characterization of affine maps	64
Examples of the inverse function theorem	64
Sample computations	66
Lecture 13: Implicit function theorem; second differential	68
Motivation and examples	68
Implicit function theorem	71
Introduction to the second differential	72
Lecture 14: The second differential; symmetric bilinear forms	74
Main theorems about the second differential	74
Symmetric bilinear forms	77
Lecture 15: Quadratic approximation and the second variation formula	80
Symmetric bilinear forms and inner products	80
The second derivative test and quadratic approximation	81
Second variation formula	83
Lecture 16: Introduction to curvature	85
Introduction	85
Curvature of a curve	86
Curvature of a surface in Euclidean 3-space	92
Lecture 17: Fundamental theorem of ordinary differential equations	96
Setup and Motivation	96
The Riemann integral	97
An application of the contraction mapping fixed point theorem	99
Lecture 18: More on ODE; introduction to differential forms	102
Uniqueness and global solutions	102
Additional topics	104
Motivating differential forms, part one	105
Lecture 19: Integration along curves and surfaces; universal properties	108
Motivating differential forms, part two	108
Bonus material: integration of 1-forms	111

Characterization by a universal property	115
Lecture 20: Tensor products, tensor algebras, and exterior algebras	119
Tensor products of vector spaces	119
Tensor algebra	122
Exterior algebra	124
Lecture 21: More on the exterior algebra; determinants	126
More about the exterior algebra: $\mathbb{Z}$ -grading and commutativity	126
Direct sums and tensor products	128
Finite dimensional exterior algebras and determinants	131
Lecture 22: Orientation and signed volume	134
Orientations	134
Duality and exterior algebras	136
Signed volume	137
Lecture 23: The Cartan exterior differential	139
Introduction	139
Exterior d	139
Lecture 24: Pullback of differential forms; forms on bases	144
Pullbacks	144
Bases of vector spaces and differential forms	146
Bases of affine spaces and differential forms	148
Frames on Euclidean space	149
Lecture 25: Curvature of curves and surfaces	153
Curvature of plane curves	153
Curvature of surfaces	155
Lectures 26–28: Integration	159
Integration of differential forms	161

Lecture 1: Affine geometry

Basic definitions

Definition 1.1. Let V be a vector space. An *affine space* A over V is a set A with a simply transitive action of V .

Elements of A are called *points*; elements of V are called *vectors*. The result of the action of a vector $\xi \in V$ on a point $p \in A$ is written $p + \xi \in A$. We call V the *tangent space* to A , and will explain the nomenclature in the next lecture; see (2.33).

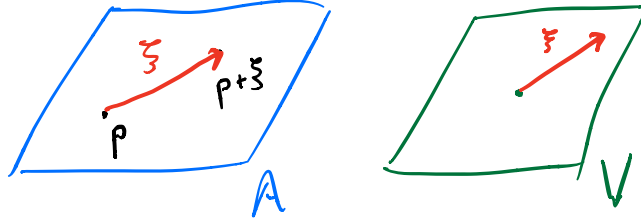


FIGURE 1. An affine space A over a vector space V

Remark 1.2 (Data and conditions). A vector space $(V, 0, +, *)$ over a field F consists of four pieces of data: a set V , a distinguished element $0 \in V$, a function $+: V \times V \rightarrow V$, and a function $*: F \times V \rightarrow V$. These data satisfy several axioms or conditions, including the fact that 0 is an identity for $+$; the existence of inverses; and commutativity, associativity, and distributivity axioms. An affine space $(A, +)$ over V provides two additional pieces of data: a set A and a function

$$(1.3) \quad +: A \times V \rightarrow A.$$

(This overloading of the symbol $+$ should not cause trouble.) There are additional axioms regarding the new $+$, here encoded in the phrase ‘simply transitive action’. As usual, symbols like ‘ V ’ and ‘ A ’ invoke all constituents of the relevant structure.

(1.4) Simple transitivity. A vector space $(V, 0, +, *)$ determines an abelian group $(V, 0, +)$. It is this abelian group which acts simply transitively on A . Simple transitivity is the statement that the map

$$(1.5) \quad \begin{aligned} A \times V &\longrightarrow A \times A \\ p, \xi &\longmapsto p, p + \xi \end{aligned}$$

is a bijection. Hence given $p_0, p_1 \in A$ there exists a unique $\xi \in V$ such that $p_1 = p_0 + \xi$. We write $\xi = p_1 - p_0$.

Remark 1.6. In this course we will almost exclusively consider $F = \mathbb{R}$. At the moment there is no topology on either V or A , and indeed much of the affine geometry we discuss in the first few lectures is valid over any field. Soon we will introduce an additional structure on V —a norm—which will induce a topology and allow us to discuss limits, completeness, compactness, and other topological properties which we use to develop analysis.

Remark 1.7. Affine space is the arena for *flat* geometry. The geometry of flat space was studied by Euclid and his contemporaries, but in Euclidean geometry there is more structure: distance and angle. The second half of the word ‘geometry’ invokes ‘measurement’, albeit measurement of the earth (‘geo’), which presumably was the (flawed) model for this part of ancient mathematics. We will discuss geometric structures in affine geometry, such as a Euclidean structure, but for now there is no measurement in affine geometry. (However, see (1.15) below.) Therefore ‘affine geometry’ is an oxymoron: the earth is not flat and measurement is not possible in bare affine space.

The distinction between points and vectors is more obvious in curved spaces, but even in flat geometry they play very different roles. For example, we might model time by an affine space A over a one-dimensional vector space V . Points of A represent instants of time, whereas elements in the group V represent intervals of time. Notice that it makes sense to add intervals of time, say $3 \text{ hours} + 4 \text{ hours} = 7 \text{ hours}$, whereas $3:00 \text{ AM} + 4:00 \text{ PM}$ does not make good sense. Then again, that the difference $4:00 \text{ PM} - 3:00 \text{ PM}$ (the same day!) is the time interval 1 hour is part of our intuition about time. Similarly, in a flat model of the Earth we do not try to make sense of the sum of Chicago and New York, but their difference as a displacement vector does make sense.

(1.8) *Points vs. vectors.*

A vector space V has a canonical (trivial) affine space over it defined by setting $A = V$ and letting (1.3) be vector addition. One can loosely describe this as “forgetting the zero vector”.

(1.9) *Vector spaces as affine spaces.*

Definition 1.10. Let A be affine over a vector space V and B affine over a vector space W . Then a function $f: A \rightarrow B$ is an *affine map* if there exists a linear map $T: V \rightarrow W$ such that

$$(1.11) \quad f(p + \xi) = f(p) + T\xi$$

for all $p \in A$, $\xi \in V$. The linear map T is called the *differential* of f , and we write $T = df$.

It is easy to verify that T is unique, if it exists. We will soon study non-affine maps $A \rightarrow B$, and such a map is differentiable if for each p there exists a linear T_p —its differential at p —such that (1.11) holds up to a controlled error term. An affine map has a constant differential.

Affine geometry is the study of affine spaces and affine maps between them.

The function $\tau_{\xi_0} A \rightarrow A$ which results from (1.3) by freezing $\xi_0 \in V$ is called *translation by ξ_0* . It is an affine automorphism of A . On the other hand, if we freeze $p_0 \in A$ then we obtain an isomorphism of affine spaces

$$(1.13) \quad \begin{aligned} \theta_{p_0}: V &\longrightarrow A \\ \xi &\longmapsto p + \xi \end{aligned}$$

Given a second point $p_1 \in A$ the *transition function* $\theta_{p_1}^{-1} \circ \theta_{p_0} : V \rightarrow V$ is translation by $p_0 - p_1$. In other words, we can identify an affine space with its tangent space up to a translation.

(1.12) Structural affine isomorphisms. If A is affine over V , then we denote the group of affine automorphisms $A \rightarrow A$ as $\text{Aut}(A)$, and the group of linear automorphisms of V as $\text{Aut}(V)$. There is a short exact sequence of groups

$$(1.14) \quad 1 \longrightarrow V \xrightarrow{\tau} \text{Aut}(A) \xrightarrow{d} \text{Aut}(V) \longrightarrow 1$$

Thus, $d \circ \tau$ is the constant map onto id_V and moreover $\ker d = \tau(V)$; in other words, an affine map f satisfies $df = \text{id}_V$ if and only if f is a translation. The term ‘short exact sequence’ includes the injectivity of τ (translations by distinct vectors are distinct affine automorphisms) and the surjectivity of d (a linear automorphism may be realized as an affine automorphism which fixes a point $p_0 \in A$).

Let A be affine over V , and fix $p \in A$, $\lambda \in \mathbb{R}$. The *homothety with center p and magnification λ* is the affine transformation

$$(1.16) \quad \begin{aligned} h_{p,\lambda} : A &\longrightarrow A \\ p + \xi &\longmapsto p + \lambda\xi \end{aligned}$$

If $\lambda \neq 1$, then $h_{p,\lambda}$ has a unique fixed point. Also, if $f \in \text{Aut}(A)$ satisfies $df = \lambda \text{id}_V$ for $\lambda \neq 1$, then $f = h_{p,\lambda}$ for some (unique) $p \in A$.

(1.15) Homotheties and the affine ratio. Now suppose $\dim V = 1$ and $p_0, p_1, p_2 \in A$ satisfy $p_0 \neq p_1$. Then there exists a unique $\lambda \in \mathbb{R}$ such that $h_{p_0,\lambda}(p_1) = p_2$. We write

$$(1.17) \quad \lambda = \frac{p_0 p_2}{p_0 p_1}.$$

Alternatively, $p_2 - p_0 = \lambda(p_1 - p_0)$ holds in V . This is a form of measurement in affine geometry. (We remark that there is a corresponding *cross-ratio* of four points on a *projective* line.)

Affine analogs of vector space concepts

The linear geometry notions of vector addition, dimension, linear subspace, containment of linear subspaces, generation of linear subspaces, linear independence, span, and basis all have counterparts in affine geometry.

(1.18) Weighted averages of points. Let $p_0, \dots, p_n$ be points in an affine space A , and let $\lambda^0, \dots, \lambda^n \in \mathbb{R}$ be real numbers which satisfy

$$(1.19) \quad \lambda^0 + \dots + \lambda^n = 1.$$

Then define

$$(1.20) \quad \lambda^i p_i = \lambda^0 p_0 + \cdots + \lambda^n p_n \in A$$

as follows. Let V be the tangent space to A . Choose $q \in A$ and for $i = 0, \dots, n$ choose $\xi_i \in V$ such that $p_i = q + \xi_i$. Then define

$$(1.21) \quad \lambda^i p_i := q + \lambda^i \xi_i.$$

An easy check shows that Definition 1.23 is independent of the choice of $q \in A$. Weighted averages of points are the affine analog of addition of vectors in a linear space.

Remark 1.22. The equality in (1.20) is the *Einstein summation convention*: an index in an expression, or term in an expression, which appears precisely twice—once as a superscript and once as a subscript—is summed over.

Definition 1.23. Let A be an affine space over a vector space V .

- (1) If $\dim V = n \in \mathbb{Z}^{\geq 0}$ is finite, then set $\dim A = n$; if not, we say A is infinite dimensional and write $\dim A = \infty$.
- (2) A subset $A_0 \subset A$ is an *affine subspace* if there exists a linear subspace $V_0 \subset V$ such that A_0 is an orbit of the action of V_0 on A by translations.
- (3) Affine subspaces $A_0, A_1 \subset A$ are *parallel* if their tangent spaces $V_0, V_1 \subset V$ satisfy either $V_0 \subset V_1$ or $V_1 \subset V_0$. We write $A_0 \parallel A_1$.
- (4) Let $S \subset A$ be a subset. The *affine span* $\mathcal{A}(S) \subset A$ is the smallest affine subspace of A which contains S .
- (5) Points $p_0, \dots, p_n \in A$ are *in general position* if $\dim \mathcal{A}(p_0, \dots, p_n) = n$.
- (6) Points $p_0, \dots, p_n \in A$ *span* A if $\mathcal{A}(p_0, \dots, p_n) = A$.
- (7) Points $p_0, \dots, p_n \in A$ form an *affine basis* of A if they are in general position and span A .

Basic theorems in linear algebra have affine analogs. For example, $p_0, \dots, p_n$ form an affine basis iff for every $q \in A$ there exist unique $\lambda^i \in \mathbb{R}$ such that $q = \lambda^i p_i$ and $\lambda^0 + \cdots + \lambda^n = 1$. The λ^i are called the *barycentric coordinates* of q . Also, all affine bases have the same cardinality.

(1.24) Convex subsets and convex hulls. The weighted average $\lambda^i p_i$ is said to be a *convex combination* of the $p_0, \dots, p_n$ if each $\lambda^i \geq 0$. A subset $C \subset A$ is *convex* if every convex combination of points of C lies in C . For $S \subset A$ an arbitrary subset, we denote by $\Delta(S)$ the smallest convex subset of A which contains S ; it is called the *convex hull* of S . The convex hull $\Delta(p_0, \dots, p_n)$ of points $p_0, \dots, p_n$ in general position is called an *n-simplex*. A 1-simplex is a *line segment*.

(1.25) Affine ratio revisited. As in (1.15) suppose $p_0 \neq p_1$ and p_2 are collinear points in an affine space A . (*Collinear* means $\dim \mathcal{A}(p_0, p_1, p_2) = 1$.) Hence p_0, p_1 is an affine basis of $L = \mathcal{A}(p_0, p_1)$ and $p_2 \in L$. Hence there exists a unique μ such that $p_2 = (1 - \mu)p_0 + \mu p_1$. Then $\mu = p_0 p_2 / p_0 p_1$ is the affine ratio (1.17).

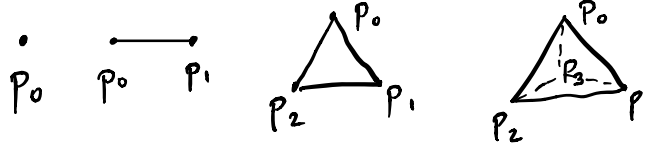


FIGURE 2. Low dimensional simplices

Ceva's theorem

We conclude this lecture with the following theorem in plane geometry, which was published by Giovanni Ceva in 1678 though it may have been known earlier.

Theorem 1.26 (Ceva). *Let A be an affine space and p_0, p_1, p_2 three points in general position. Choose $q_0 \in \mathcal{A}(p_1, p_2)$, $q_1 \in \mathcal{A}(p_2, p_0)$, and $q_2 \in \mathcal{A}(p_0, p_1)$. Then the three lines $\mathcal{A}(p_0, q_0)$, $\mathcal{A}(p_1, q_1)$, $\mathcal{A}(p_2, q_0)$ are concurrent or parallel if and only if*

$$(1.27) \quad \frac{q_2 p_0}{q_2 p_1} \frac{q_0 p_1}{q_0 p_2} \frac{q_1 p_2}{q_1 p_0} = -1.$$

Three or more lines are *concurrent* if they share a common point.

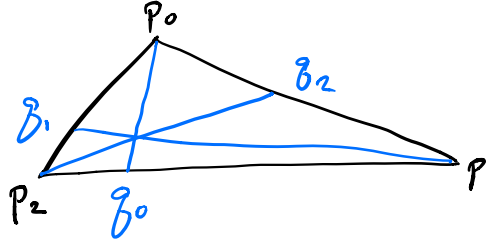


FIGURE 3. Three concurrent cevians

Proof. Choose $\lambda, \mu, \nu \in \mathbb{R}$ such that

$$(1.28) \quad \begin{aligned} q_2 &= \lambda p_0 + (1 - \lambda) p_1 \\ q_0 &= \mu p_1 + (1 - \mu) p_2 \\ q_1 &= (1 - \nu) p_0 + \nu p_2 \end{aligned}$$

Then $q_2 p_0 / q_2 p_1 = -\lambda / (1 - \lambda)$ and the negative of the product in (1.27) is

$$(1.29) \quad x = \frac{\lambda}{1 - \lambda} \frac{\mu}{1 - \mu} \frac{\nu}{1 - \nu}.$$

A general point on the cevian line $\mathcal{A}(p_i, q_i)$ is parametrized by $t_i \in \mathbb{R}$ and is, for $i = 0, 1, 2$

$$(1.30) \quad \begin{aligned} &(1 - t_0) p_0 + t_0 \mu p_1 + t_0 (1 - \mu) p_2 \\ &t_1 (1 - \nu) p_0 + (1 - t_1) p_1 + t_1 \nu p_2 \\ &t_2 \lambda p_0 + t_2 (1 - \lambda) p_1 + (1 - t_2) p_2 \end{aligned}$$

The three cevians are parallel iff

$$(1.31) \quad \begin{aligned} 1 - \mu + \mu\nu &= 0 \\ 1 - \nu + \nu\lambda &= 0 \\ 1 - \lambda + \lambda\mu &= 0, \end{aligned}$$

in which case $x = 1$. If they are concurrent at $a^0p_0 + a^1p_1 + a^2p_2$, then

$$(1.32) \quad x = \frac{a^0}{a^1} \frac{a^1}{a^2} \frac{a^2}{a^0} = 1.$$

Conversely, if $x = 1$ then

$$(1.33) \quad \lambda(1 - \mu + \mu\nu) = (1 - \mu)(1 - \nu + \nu\lambda).$$

Hence from the first equation in (1.31) if the lines $\mathcal{A}(p_0, q_0)$ and $\mathcal{A}(p_1, q_1)$ are parallel, then $1 - \mu + \mu\nu = 0$, and since in that case $\mu \neq 1$ we deduce $1 - \nu + \nu\lambda = 0$, which is equivalent to the lines $\mathcal{A}(p_1, q_1)$ and $\mathcal{A}(p_2, q_2)$ being parallel. If the lines $\mathcal{A}(p_0, q_0)$ and $\mathcal{A}(p_1, q_1)$ are not parallel, then they intersect at the point c which equals each of the first two expressions in (1.30) with

$$(1.34) \quad t_0 = \frac{\nu}{1 - \mu + \mu\nu}, \quad t_1 = \frac{1 - \mu}{1 - \mu + \mu\nu}.$$

From (1.33) we have $t_1 = \lambda/(1 - \nu + \nu\lambda)$ and setting $t_2 = (1 - \nu)/(1 - \nu + \nu\lambda)$ we see from the last expression in (1.30) that $c \in \mathcal{A}(p_2, q_2)$. $\square$

Lecture 2: Parallelism, coordinates, and symmetry

Parallelism and another classical theorem

(2.1) Global parallelism; Euclid's axiom. Let A be an affine space over a vector space V . Recall Definition 1.23(3) of parallel affine subspaces of A . Affine geometry is the geometry of *global* parallelism in the following sense. Suppose $A_0 \subset A$ is an affine subspace with tangent space $V_0 \subset V$. Then A_0 is an orbit of the V_0 -action on A by translations, and every other V_0 -orbit is parallel to A_0 . The collection of orbits is a *foliation* of A by parallel affine subspaces: a single affine subspace gives rise to the entire collection. Another manifestation is *Euclid's parallel postulate*, which in a general form asserts that given $A_0 \subset A$ an affine subspace and $p \in A$ there exists a unique affine subspace $A'_0 \subset A$ with tangent V_0 which contains p . It is obtained from A_0 by translation. Euclid studied the case of a line in a plane: $\dim A = 2$ and $\dim A_0 = 1$.

Remark 2.2. Children usually encounter Euclid as a means of learning mathematical rigor as well as geometry. For us the foundations of affine geometry rest on linear algebra, which in turn rests on other mathematical developments in the past few centuries: the theory of sets, of fields, etc.

(2.3) Parallelism is an affine property. An *affine property* is one preserved under affine isomorphism. Parallelism is even preserved under arbitrary affine maps.

Proposition 2.4. *Let $f: A \rightarrow B$ be an affine map and $A_0 \parallel A_1$ parallel subspaces. Then their images are parallel: $f(A_0) \parallel f(A_1)$.*

Proof. Let V, W be the tangent spaces to A, B and $V_0, V_1 \subset V$ be the tangent spaces to A_0, A_1 . Assume $V_0 \subset V_1$. (Simply change notation if instead $V_1 \subset V_0$.) Then $df(V_0) \subset df(V_1)$, where $df: V \rightarrow W$ is the differential of f . It is easy to check that $df(V_i)$ is the tangent space to $f(A_i)$, $i = 0, 1$, and so the conclusion follows. $\square$

(2.5) Homotheties and parallelism. We leave the reader to prove that homotheties map an affine subspace to a parallel affine subspace.

Proposition 2.6. *Let A be an affine space, $h: A \rightarrow A$ a homothety, and $A_0 \subset A$ an affine subspace. Then $A_0 \parallel h(A_0)$.* $\square$

(2.7) Pappus' theorem. The following is attributed to Pappus of Alexandria, who was a leading geometer in the 4th century BCE.

Theorem 2.8 (Pappus). *Let $L, L' \subset A$ be two lines in an affine plane. Fix points $p_1, p_2, p_3 \in L$ and $p'_1, p'_2, p'_3 \in L'$. Assume $\mathcal{A}(p_1, p'_2) \parallel \mathcal{A}(p'_1, p_2)$ and $\mathcal{A}(p_2, p'_3) \parallel \mathcal{A}(p'_2, p_3)$. Then $\mathcal{A}(p_3, p'_1) \parallel \mathcal{A}(p'_3, p_1)$.*

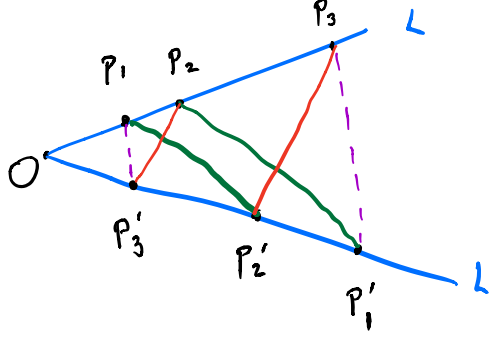


FIGURE 4. Pappus configuration

Proof. Assume L, L' are not parallel, so intersect in a point $O \in A$, as in Figure 4. (If they are parallel, then replace homotheties with translations in the following argument.) Let $f, g: A \rightarrow A$ be homotheties centered at O such that $f(p_1) = p_2$ and $g(p_2) = p_3$. Then from the assumed parallelisms and Proposition 2.6 we conclude $f(p'_2) = p'_1$ and $g(p'_3) = p'_2$. Hence $gf(p_1) = p_3$ and $fg(p'_3) = p'_1$. But the compositions gf and fg are equal homotheties centered at O , and so Proposition 2.6 yields the desired parallelism $\mathcal{A}(p_3, p'_1) \parallel \mathcal{A}(p'_3, p_1)$. $\square$

Bases in vector spaces

(2.9) *Review.* We begin with the standard definitions analogous to Definition 1.23(5)–(7).

Definition 2.10. Let V be a vector space.

- (1) A subset $S \subset V$ is *linearly independent* if whenever $c^1, \dots, c^n \in \mathbb{R}$ and $\xi_1, \dots, \xi_n \in V$ satisfy $c^i \xi_i = 0$ we have $c^i = 0$, $i = 1, \dots, n$.
- (2) A subset $S \subset V$ *spans* V if for every $\eta \in V$ there exist $c^1, \dots, c^n \in \mathbb{R}$ and $\xi_1, \dots, \xi_n \in V$ such that $\eta = c^i \xi_i$.
- (3) Vectors $\xi_1, \dots, \xi_n$ form a *basis* of V if they are linearly independent and span V .

Hence $\xi_1, \dots, \xi_n$ form a basis if the c^i in (2) exist and are unique. It is a theorem that any two bases have the cardinality, a nonnegative integer¹ we write as $\dim V$.

(2.11) *Standard model.* For each $n \in \mathbb{N}$ we define $\mathbb{R}^n = (\mathbb{R}^n, 0, +, *)$ as the standard n -dimensional vector space. As a set it consists of all ordered n -tuples of real numbers:

$$(2.12) \quad \mathbb{R}^n = \{(\xi^1, \dots, \xi^n) : \xi^i \in \mathbb{R}\}.$$

¹The zero vector space consisting of only the zero vector has dimension zero.

The zero vector is $0 = (0, \dots, 0)$. Vector addition $(\xi^1, \dots, \xi^n) + (\eta^1, \dots, \eta^n) = (\xi^1 + \eta^1, \dots, \xi^n + \eta^n)$ is defined component-wise, as is scalar multiplication $c \cdot (\xi^1, \dots, \xi^n) = (c\xi^1, \dots, c\xi^n)$. The *standard basis* of $\mathbb{R}^n$ is

$$(2.13) \quad e_1 = (1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1).$$

(2.14) Bases revisited. We recast Definition 2.10(3) as an explicit isomorphism from the model vector space to an abstract vector space.

Definition 2.15. Let V be a vector space. A *basis* of V is an isomorphism $b: \mathbb{R}^n \rightarrow V$ for some n . Denote the set of bases of V as $\mathcal{B}(V)$.

The basis $\xi_1, \dots, \xi_n$ in the previous sense is $b(e_1), \dots, b(e_n)$. If no basis exists, then V is infinite dimensional, in which case $\mathcal{B}(V)$ is empty. There is no *canonical*, or *natural*, basis of a finite dimensional vector space. We do not formalize that assertion,² but rather elucidate it in examples.

Remark 2.16. To illustrate the lack of canonical bases, consider the following three situations. First, define V as the space of solutions to the system of linear equations

$$(2.17) \quad \begin{aligned} \xi^1 + \xi^2 &= 0 \\ \xi^1 - 2\xi^2 &= 0 \end{aligned}$$

Although $V \subset \mathbb{R}^2$ and $\mathbb{R}^2$ has its canonical basis, there is no distinguished nonzero vector in V . Second, let V be the space of functions $f: \mathbb{R} \rightarrow \mathbb{R}$ which satisfy the ordinary differential equation $\ddot{f} + f = 0$. Then V is 2-dimensional, but there is no natural (ordered) basis. Finally—and here we rely on your intuition—let $S \subset \mathbb{A}^3$ be the sphere defined by the equation $(x^1)^2 + (x^2)^2 + (x^3)^2 = 1$. (We define the standard affine space $\mathbb{A}^3$ below in §(2.24).) We will eventually define the notion of a *smooth manifold*, prove that S is an example, and define the tangent space $T_p S$ at $p \in S$ to be a subspace of $\mathbb{R}^3$. At $p = (x^1, x^2, x^3)$ it is the subspace of vectors $\xi = (\xi^1, \xi^2, \xi^3)$ which satisfy

$$(2.18) \quad x^1 \xi^1 + x^2 \xi^2 + x^3 \xi^3 = 0.$$

There is no natural basis. In fact, if there were we would find a (presumably smoothly varying) nonzero vector field on the sphere, but that contradicts the “hairy ball theorem”.

Linear symmetry groups

(2.19) Automorphism groups. Quite generally, a symmetry of a mathematical object X is an invertible map $g: X \rightarrow X$ which preserves the “structure” of X . Such maps form a group $\text{Aut}(X)$ which acts on X on the left. In the case of a vector space V we obtain the group $\text{Aut}(V)$ of invertible *linear* symmetries of V . For the model space $V = \mathbb{R}^n$ we use the notation $\text{Aut}(\mathbb{R}^n) = \text{GL}_n \mathbb{R}$ and the term *general linear group*.

²But we could do so easily. Fix $n \in \mathbb{N}$ and let Vect_n denote the category of n -dimensional vector spaces and invertible linear maps between them. If there were canonical bases, then for each $V \in \text{Vect}_n$ there would be a morphism $b_V: \mathbb{R}^n \rightarrow V$ such that for all morphisms $f: V \rightarrow V'$ in Vect_n we would have $b_{V'} = f \circ b_V$. Nonsense!

(2.20) Matrices. A rectangular array of numbers $M = (M_j^i)$, $1 \leq i \leq m$, $1 \leq j \leq n$, is called an $m \times n$ *matrix*. The index i is the row index; the index j is the column index. If $N = (N_\ell^k)$, $1 \leq k \leq p$, $1 \leq \ell \leq q$ is another matrix, then if $n = p$ the matrix product $MN = (M_j^i N_\ell^j)$ is defined. (By the summation convention for each fixed i, ℓ we sum over j .) A vector $\xi = (\xi^i) \in \mathbb{R}^n$ is a column vector. If $T: \mathbb{R}^m \rightarrow \mathbb{R}^n$ is a linear transformation, it is represented by the matrix (T_j^i) defined by $Te_j = T_j^i e_i$, where the first e_j lies in the domain $\mathbb{R}^m$ and the second e_i lies in the codomain $\mathbb{R}^n$. In particular, an element $g \in \text{GL}_n \mathbb{R}$ is an invertible $n \times n$ matrix $g = (g_j^i)$, and the action on $\xi \in \mathbb{R}^n$ is the matrix product $g\xi = (g_j^i \xi^j)$.

(2.21) Bases, symmetry, structure. Let V be an n -dimensional vector space. There are two groups which act naturally on the set $\mathcal{B}(V)$ of bases $b: \mathbb{R}^n \xrightarrow{\cong} V$. First, $\text{GL}_n \mathbb{R}$ acts on the *right* by precomposition: if $g: \mathbb{R}^n \xrightarrow{\cong} \mathbb{R}^n$, then $b \circ g$ is another basis. Furthermore, that action is simply transitive. This is a situation we encounter often, and there is a special term used.

Definition 2.22. Let G be a group. A (*right*) G -*torsor* T is a set equipped with a simply transitive right G -action.

There is also a notion of a left G -torsor. We defined simple transitivity in the context of affine spaces (1.4), and in fact an affine space is a torsor over its tangent space. (Since the latter is an abelian group, left and right actions are equivalent.) In our current situation we see that the set $\mathcal{B}(V)$ of bases is a right $\text{GL}_n \mathbb{R}$ -torsor.

The other natural group acting on $\mathcal{B}(V)$ is $\text{Aut}(V)$, which acts on the *left* by postcomposition: $\phi: V \xrightarrow{\cong} V$ acts on $b: \mathbb{R}^n \xrightarrow{\cong} V$ to produce the basis $\phi \circ b$. Intuitively, the $\text{Aut}(V)$ -action is by symmetries of V , whereas the $\text{GL}_n \mathbb{R}$ -action is “internal”, reflecting the linear structure of V as encoded by the torsor of bases.

Remark 2.23. In geometry quite generally left actions are by symmetries whereas right actions are structural. We can say that the group $\text{GL}_n \mathbb{R}$ defines the *symmetry type* of general linear geometry, whereas for a specific vector space V the group $\text{Aut}(V)$ acts as symmetries on that vector space. There are other n -dimensional linear geometries, such as the geometry of an inner product space, and their symmetry type is defined by a pair (G, ρ) in which G is a group and $\rho: G \rightarrow \text{GL}_n \mathbb{R}$ a homomorphism. In the case just mentioned, $G = \text{O}_n$ is the orthogonal group and ρ the inclusion.

Affine coordinates

(2.24) *Standard model of affine space.* There is a standard³ affine space $\mathbb{A}^n = (\mathbb{A}^n, +)$ over the standard vector space $\mathbb{R}^n$ defined in (2.11). The underlying set is the same as that of $\mathbb{R}^n$, namely

$$(2.26) \quad \mathbb{A}^n = \{(x^1, \dots, x^n) : x^i \in \mathbb{R}\}.$$

The $\mathbb{R}^n$ -action is defined as component-wise addition: $(x^i) + (\xi^i) = (x^i + \xi^i)$ for $x = (x^1, \dots, x^n) \in \mathbb{A}^n$, $\xi = (\xi^1, \dots, \xi^n) \in \mathbb{R}^n$. The functions $x^i : \mathbb{A}^n \rightarrow \mathbb{R}$ are the *standard affine coordinate functions*. The group of affine symmetries $g : \mathbb{A}^n \xrightarrow{\cong} \mathbb{A}^n$ of $\mathbb{A}^n$ is denoted Aff_n .

(2.27) *Coordinates on affine space.* To do computations in geometry we introduce coordinates. For example, we used barycentric coordinates in an affine plane to prove Ceva's Theorem 1.26. In affine geometry it is often convenient to use affine coordinates.

Definition 2.28. Let A be an n -dimensional affine space. An *affine coordinate system* on A is an affine isomorphism $x : A \xrightarrow{\cong} \mathbb{A}^n$. The coordinate system is *centered* at $x^{-1}(0)$. The *coordinate functions* $x^i : A \rightarrow \mathbb{R}$ are defined by composition with the standard affine coordinate functions.

Remark 2.29. Whereas a basis of a vector space is a map out of the model space, a coordinate system is a map into the model space. The former is a *parametrization* by the model whereas the latter uses the model to navigate around an abstract space.

Let V be the tangent space to A and

$$(2.30) \quad dx : V \longrightarrow \mathbb{R}^n$$

the differential of x . Since x is an affine isomorphism, it follows that dx is a linear isomorphism.⁴ We write $dx = (dx^1, \dots, dx^n)$ for linear maps $dx^i : V \rightarrow \mathbb{R}$. Recall that the collection of linear maps $V \rightarrow \mathbb{R}$ is the *dual space* V^* to V . The elements $dx^1, \dots, dx^n$, which are the differentials of the coordinate functions, form a basis of V^* . Also, the inverse $(dx)^{-1} : \mathbb{R}^n \rightarrow V$ is a basis of V , and we write $\partial/\partial x^j \in V$ for the j^{th} element of the basis.⁵ In other words,

$$(2.31) \quad dx^i\left(\frac{\partial}{\partial x^j}\right) = \delta_j^i := \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

The n^2 equations in (2.31) express that dx^i and $\partial/\partial x^j$ are dual bases of V^* and V .

³This standard model is adapted to affine coordinates. A standard model for barycentric coordinates, as defined after Definition 1.23, is the affine subspace

$$(2.25) \quad \{(\xi^0, \dots, \xi^n) \in \mathbb{R}^{n+1} : \xi^0 + \dots + \xi^n = 1\}$$

of $\mathbb{R}^{n+1}$ whose tangent space is the linear subspace $\xi^0 + \dots + \xi^n = 0$.

⁴This uses the *chain rule*, which I forgot to put into Lecture 1. Namely, the differential of a composition of affine maps is the composition of the differentials. Hence, if $\phi : \mathbb{A}^n \xrightarrow{\cong} A$ is the inverse to x , then $d\phi$ is the inverse to dx .

⁵Note that since j is a superscript in the denominator, it counts as a subscript for index conventions.

The tangent space to affine space

Let A be an affine space over a vector space V .

(2.32) *V as a group of translations.* Our definition of an affine space is global. We regard V as a group under vector addition and define an affine space as a V -torsor. This global role is used to construct parallel affine spaces, for example; see [\(2.1\)](#).

(2.33) *V as the tangent space.* There is a local interpretation which justifies the nomenclature ‘tangent space’. For this we will ask the reader’s good will since we use concepts (limits, derivative) not yet introduces, but with which (s)he is surely familiar. Hence suppose A is affine over V , fix $\epsilon > 0$, and suppose $\gamma: (-\epsilon, \epsilon) \rightarrow A$ is a *parametrized curve* or *motion* in A . Intuitively, the function γ expresses position as a function of time. The initial position is $\gamma(0) \in A$. The initial velocity, if γ is differentiable at time 0, is defined as the limit of difference quotients:

$$(2.34) \quad \dot{\gamma}(0) = \lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t}.$$

All we need here is the formal structure of the difference quotient. The numerator is the difference of two points of A , so a vector in V . We are then instructed to scalar multiply this vector by $1/t$, so obtain another vector in V . In other words, the difference quotient defines a function $(-\epsilon, \epsilon) \setminus \{0\} \rightarrow V$, which is a parametrized curve of vectors. The limit, if it exists, is then also a vector in V . Therefore, the vector space V plays the role of the tangent space to the affine space A at the point $\gamma(0)$. But $\gamma(0)$ can be any point of A , so V is the tangent space at every point. In other words, an affine space has a constant tangent space. (By contrast, a curved space—smooth manifold—can have a variable tangent space; see [Remark 2.16](#).)

Lecture 3. Normed linear spaces

(3.1) Topology. In order to deal with curved smooth shapes, initially sitting in affine space, we need to be able to take limits, such as the one in (2.34) which defines the tangent vector to a parametrized curve. To take limits we need some notion of “closeness”, which is what a topology affords. A general topology can lead to difficulties, for example non-uniqueness of limits if the topology is non-Hausdorff. The topology defined by a metric is quite nice in many respects, and it is a natural one to use on affine space. Furthermore, a distance function on affine space is in our intuition a structure inherited from a length function on its tangent vector space: to measure the distance between Chicago and New York we compute the length of the displacement vector. Hence in this lecture we begin to study length functions, or *norms*, on linear spaces.

Basic definitions

Definition 3.2. Let V be a real vector space. A *norm* on V is a function

$$(3.3) \quad \rho: V \longrightarrow \mathbb{R}^{\geq 0}$$

such that for all $\xi, \eta \in V$ and $c \in \mathbb{R}$ we have

- (1) $\rho(\xi) = 0$ if and only if $\xi = 0$,
- (2) $\rho(c\xi) = |c|\rho(\xi)$,
- (3) $\rho(\xi + \eta) \leq \rho(\xi) + \rho(\eta)$.

The pair (V, ρ) is called a *normed linear space*

Property (3) is called the *triangle inequality*. We often use the notation $\|\xi\| = \rho(\xi)$ for the norm.

(3.4) Induced metric on affine space. Recall the definition of a metric space.

Definition 3.5. A metric space (X, d) is a set X together with a function

$$(3.6) \quad d: X \times X \longrightarrow \mathbb{R}^{\geq 0}$$

such that for all $x, y, z \in X$ we have

- (1) $d(x, y) = 0$ if and only if $x = y$,
- (2) $d(x, y) = d(y, x)$,
- (3) $d(x, y) \leq d(x, z) + d(z, y)$.

Again property (3) is called the *triangle inequality*.

Let (V, ρ) be a normed linear space and A an affine space over V . Then A inherits a metric $d_\rho: A \times A \rightarrow \mathbb{R}^{\geq 0}$ defined by

$$(3.7) \quad d_\rho(p, q) = \rho(q - p), \quad p, q \in A,$$

where recall from (1.4) that $\xi = q - p \in V$ is the unique vector such that $q = p + \xi$. We leave the reader to verify that properties (1)–(3) of ρ imply properties (1)–(3) of d_ρ .

Remark 3.8. A norm on a vector space simultaneously induces a metric on all affine spaces. It is one instance of a structure on a group G simultaneously inducing a structure on all G -torsors. Note in our situation that the vector space V , viewed as the trivial affine space over V (see (1.9)), has a metric as well.

Remark 3.9. A normed vector space has a notion of length, but not a notion of angle. There is another structure—an inner product space—which gives rise to a geometry with both length and angle.

Norms on $\mathbb{R}^n$

Fix $n \in \mathbb{N}$.

(3.10) Euclidean norm. We begin with the most familiar norm, derived from the Pythagorean formula.

Proposition 3.11. *The function $(\xi^1, \dots, \xi^n) \mapsto \sqrt{(\xi^1)^2 + \dots + (\xi^n)^2}$ is a norm on $\mathbb{R}^n$.*

We write $\|\xi\|$ for the norm of $\xi = (\xi^1, \dots, \xi^n)$ in this norm.

Proof. The only nontrivial assertion is the triangle inequality. Expanding out the formulas we see that $\|\xi + \eta\|^2 \leq (\|\xi\| + \|\eta\|)^2$ follows from the *Cauchy-Schwarz inequality*

$$(3.12) \quad \sum_i \xi^i \eta^i \leq \sqrt{\sum_i (\xi^i)^2} \sqrt{\sum_i (\eta^i)^2},$$

where $\xi, \eta \in \mathbb{R}^n$. To prove that consider the real-valued quadratic function $q(t) = \|\xi + t\eta\|^2$ of $t \in \mathbb{R}$. There is at most a single root of q —if there exists $t \in \mathbb{R}$ such that $\xi + t\eta = 0$, in which case such a t is unique—and so the discriminant of q is nonpositive. The latter assertion is equivalent to the square of (3.12). $\square$

Remark 3.13. The left hand side of (3.12) is the standard inner product on $\mathbb{R}^n$

$$(3.14) \quad \langle \xi, \eta \rangle = \sum_i \xi^i \eta^i, \quad \xi, \eta \in \mathbb{R}^n.$$

The induced norm $\|\xi\| = \sqrt{\langle \xi, \xi \rangle}$ is the Euclidean norm in Proposition 3.11. There is also an induced notion of angle θ between nonzero vectors ξ, η , namely

$$(3.15) \quad \cos \theta = \frac{\langle \xi, \eta \rangle}{\|\xi\| \|\eta\|}.$$

The general p -norm we consider next only comes from an inner product for $p = 2$.

(3.16) p -norms. For a real number $p \geq 1$ define the p -norm

$$(3.17) \quad \|\xi\|_p := (|\xi^1|^p + \cdots + |\xi^n|^p)^{1/p}.$$

We also define

$$(3.18) \quad \|\xi\|_\infty := \max_i |\xi^i|.$$

Theorem 3.19. For all $1 \leq p \leq \infty$ the function $\xi \mapsto \|\xi\|_p$ is a norm.

The theorem is easy for $p = 1$ and $p = \infty$, and properties (1) and (2) of Definition 3.2 are easy for all p . The triangle inequality for $1 < p < \infty$ follows from the next three lemmas.⁶

Lemma 3.20 (Young's inequality). Suppose $1 < p, q < \infty$ and $1/p + 1/q = 1$. Then for all $x, y \geq 0$ we have

$$(3.21) \quad xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

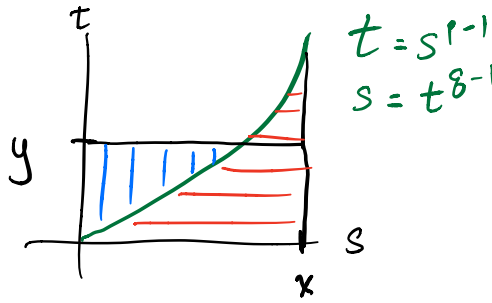


FIGURE 5. Proof of Young's inequality

Proof. Assume $y \leq x^{p-1}$; the proof is similar if the opposite inequality holds. In Figure 5 the blue area is

$$(3.22) \quad \int_0^y dt t^{q-1} = \frac{y^q}{q}$$

⁶These proofs follow those in B. Simon, *A Comprehensive Course in Analysis, Part 1*.

and the red area is

$$(3.23) \quad \int_0^x ds s^{p-1} = \frac{x^p}{p}.$$

The inequality follows from the fact that the area of the rectangle with vertices $(0, 0)$, $(x, 0)$, (x, y) , $(0, y)$ is bounded above by the area of the shaded region. $\square$

Lemma 3.24 (Hölder inequality). *Suppose $1 < p, q < \infty$ and $1/p + 1/q = 1$. Then for all $\xi, \eta \in \mathbb{R}^n$,*

$$(3.25) \quad \left| \sum_i \xi^i \eta^i \right| \leq \|\xi\|_p \|\eta\|_q.$$

Proof. If ξ or η is nonzero, then the statement is trivial, so assume both are nonzero. By scaling both sides it suffices to assume $\|\xi\|_p = \|\eta\|_q = 1$. By Young's inequality we have

$$(3.26) \quad |\xi^i \eta^i| \leq \frac{|\xi^i|^p}{p} + \frac{|\eta^i|^q}{q}$$

for each $i = 1, \dots, n$. Now sum over i . $\square$

Lemma 3.27 (Minkowski inequality). *Suppose $1 < p < \infty$. Then for all $\xi, \eta \in \mathbb{R}^n$,*

$$(3.28) \quad \|\xi + \eta\|_p \leq \|\xi\|_p + \|\eta\|_p.$$

This is precisely the triangle inequality for the p -norm.

Proof. If $\xi + \eta = 0$ the statement is trivial, so assume $\xi + \eta \neq 0$. Then

$$(3.29) \quad \begin{aligned} \sum |\xi^i + \eta^i|^p &= \sum |\xi^i + \eta^i| |\xi^i + \eta^i|^{p-1} \\ &\leq \sum |\xi^i| |\xi^i + \eta^i|^{p-1} + |\eta^i| |\xi^i + \eta^i|^{p-1} \\ &\leq (\|\xi\|_p + \|\eta\|_p) \left(\sum |\xi^i + \eta^i|^{(p-1)q} \right)^{1/q}. \end{aligned}$$

At the last stage we apply the Hölder inequality. Now use $(p-1)q = p$ to deduce

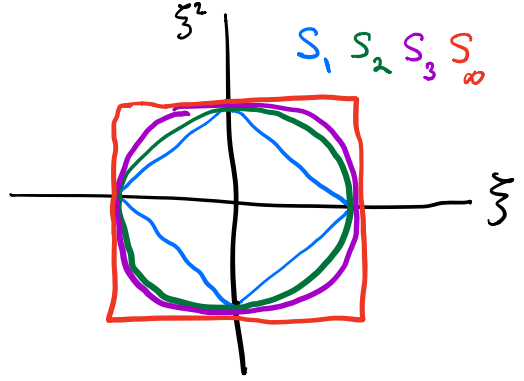
$$(3.30) \quad \|\xi + \eta\|_p^p \leq (\|\xi\|_p + \|\eta\|_p) \frac{\|\xi + \eta\|_p^p}{\|\xi + \eta\|_p},$$

from which (3.28) follows. $\square$

(3.31) Unit spheres. Because of the homogeneity of a norm (Definition 3.2(2)), its unit sphere

$$(3.32) \quad S_p = \{\xi \in \mathbb{R}^n : \|\xi\|_p = 1\}$$

contains all of the information. We depict the unit spheres for various p in Figure 6.

FIGURE 6. Unit sphere of the p -norm on $\mathbb{R}^2$

Lecture 4. More on normed linear spaces

Banach spaces

Let V be a normed linear space. Then as in (3.4) and Remark 3.8 there is an induced metric space structure on V . Hence it makes sense to talk about convergent sequences and about Cauchy sequences. Thus a sequence $\xi: \mathbb{N} \rightarrow V$ is a *Cauchy sequence* if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that if $m, n \geq N$, then $\|\xi_m - \xi_n\| < \epsilon$. (In the sequel we write a sequence as $(\xi_n) \subset V$.) A metric space is *complete* if every Cauchy sequence converges. (The limit is unique since every metric space is Hausdorff)

Definition 4.1. A *Banach space* is a complete normed linear space.

It follows easily that an affine space over a Banach space is a complete metric space. We will develop calculus in this setting; completeness is important for many basic theorems. We prove below () that every finite dimensional normed linear space is complete, hence is a Banach space. First, we discuss some infinite dimensional examples.

Examples of Banach spaces

Lemma 4.2. Let (V, ρ) be a normed linear space. Then $\rho: V \rightarrow \mathbb{R}^{\geq 0}$ is (uniformly) continuous.

Proof. It follows from two applications of the triangle inequality that

$$(4.3) \quad |\rho(\xi) - \rho(\eta)| \leq \rho(\xi - \eta), \quad \text{for all } \xi, \eta \in V.$$

This proves that ρ is Lipschitz continuous, so in particular is (uniformly) continuous. $\square$

A function $\xi: S \rightarrow W$ into a normed linear space is *bounded* if there exists $C > 0$ such that $\|\xi(s)\|_W \leq C$ for all $s \in S$.

Theorem 4.4. *Let S be a set, W a Banach space, and V the vector space*

$$(4.5) \quad V = \{\xi: S \rightarrow W : \xi \text{ is bounded}\}.$$

Then the function

$$(4.6) \quad \|\xi\|_V = \sup_{s \in S} \|\xi(s)\|_W.$$

is a complete norm on V .

The vector space structure on V is pointwise addition: $(\xi_1 + \xi_2)(s) = \xi_1(s) + \xi_2(s)$. Note that the sup in (4.6) is not necessarily a max, for example if $S = (0, 1)$, $W = \mathbb{R}$, and ξ is the inclusion.

Proof. We first verify that (4.6) defines a norm. The first two properties in Definition 3.2 are straightforward to verify, so we address the triangle inequality. Let $\xi, \eta \in V$. Given $\epsilon > 0$ choose $s \in S$ such that $\|\xi(s) + \eta(s)\|_W \geq \|\xi + \eta\|_V - \epsilon$. Then

$$(4.7) \quad \|\xi + \eta\|_V \leq \|\xi(s) + \eta(s)\|_W + \epsilon \leq \|\xi(s)\|_W + \|\eta(s)\|_W + \epsilon \leq \|\xi\|_V + \|\eta\|_V + \epsilon,$$

and taking $\epsilon \rightarrow 0$ we obtain the triangle inequality.

Next, we prove that the norm is complete. Suppose $(\xi_n) \subset V$ is a Cauchy sequence. Then for each $s \in S$ $(\xi_n(s)) \subset W$ is Cauchy, and since W is complete there exists $\xi(s) \in W$ such that $\xi_n(s) \rightarrow \xi(s)$. We claim that $\xi: S \rightarrow W$ is bounded and that $\xi_n \rightarrow \xi$ in V . To check the first claim, choose $N \in \mathbb{N}$ such that if $m, n \geq N$ then $\|\xi_m - \xi_n\|_V < 1$. Choose $C > 0$ such that $\|\xi_N(s)\|_W \leq C$ for all $s \in S$. Then since the norm is continuous (Lemma 4.2), for all $s \in S$ we have

$$(4.8) \quad \|\xi(s)\|_W = \lim_{m \rightarrow \infty} \|\xi_m(s)\|_W \leq \|\xi_N(s)\|_W + \lim_{m \rightarrow \infty} \|\xi_m(s) - \xi_N(s)\|_W \leq C + 1.$$

So ξ is bounded. Now given $\epsilon > 0$ choose $N \in \mathbb{N}$ such that if $m, n \geq N$ then $\|\xi_m - \xi_n\|_V < \epsilon$. Then for all $s \in S$

$$(4.9) \quad \|\xi(s) - \xi_n(s)\|_W = \lim_{m \rightarrow \infty} \|\xi_m(s) - \xi_n(s)\|_W \leq \epsilon,$$

since for fixed m we have $\|\xi_m(s) - \xi_n(s)\|_W \leq \|\xi_m - \xi_n\|_V < \epsilon$. Therefore $\|\xi - \xi_n\|_V < \epsilon$, from which $\xi_n \rightarrow \xi$ in V . $\square$

(4.10) *Example of an incomplete normed linear space.* Let $V^0 = C^0([0, 1], \mathbb{R})$ denote the vector space of continuous functions $f: [0, 1] \rightarrow \mathbb{R}$. Since $[0, 1]$ is compact, every continuous function is bounded, so V^0 is a subspace of the vector space (4.5) with $S = [0, 1]$ and $W = \mathbb{R}$. In this case the sup norm (4.6) is a max, since S is compact and a continuous function on a compact space realizes its supremum. Furthermore, the subspace of continuous functions is closed in the space of bounded functions, since if $f_n \rightarrow f$ in the max norm the convergence is uniform and a uniform limit of continuous functions is continuous. It follows from Theorem 4.4 and the fact that a closed subspace of a complete metric space is complete that V^0 is a Banach space in the max norm. On the other hand, the L^1 norm

$$(4.11) \quad \|f\|_1 = \int_0^1 dx |f(x)|, \quad f \in V^0,$$

is incomplete. For example, define

$$(4.12) \quad f_n(x) = \begin{cases} 0, & 0 \leq x \leq 1/2 - 1/n; \\ \frac{n}{2}(x - \frac{1}{2} + \frac{1}{n}), & 1/2 - 1/n \leq x \leq 1/2 + 1/n \\ 1, & 1/n + 1/2 \leq x \leq 1. \end{cases}$$

Then $(f_n) \subset V^0$ is Cauchy in the L^1 norm but does not converge.

Equivalent norms

Let X be a set and $d_1, d_2: X \times X \rightarrow \mathbb{R}^{>0}$ metric on X . Recall that d_1 and d_2 are *equivalent metrics* if there exists $C > 0$ such that

$$(4.13) \quad \frac{1}{C}d_1(x, y) \leq d_2(x, y) \leq Cd_1(x, y), \quad \text{for all } x, y \in X.$$

Equivalent metrics determine the same open sets, so the same topology on X . They also determine the same set of Cauchy sequences in X , so d_1 is complete if and only if d_2 is complete.

The definition of equivalent norms on a vector space V is designed so that the associated metrics on any affine space over V are equivalent.

Definition 4.14. Let V be a real vector space and $\rho_1, \rho_2: V \rightarrow \mathbb{R}^{\geq 0}$ norms. Then ρ_1 is equivalent to ρ_2 if there exists $C > 0$ such that

$$(4.15) \quad \frac{1}{C}\rho_1(\xi) \leq \rho_2(\xi) \leq C\rho_1(\xi), \quad \text{for all } \xi \in V.$$

Equivalence is an equivalence relation on the set of norms.

(4.16) *Example of inequivalent norms.* Let V denote the vector space of finitely supported functions $\xi: \mathbb{N} \rightarrow \mathbb{R}$, i.e., sequences $(x_i) \subset \mathbb{R}$ such that $x_i = 0$ for all but finitely many i . Then the norms

$$(4.17) \quad \begin{aligned} \rho_1(x_1, x_2, \dots) &= |x^1| + |x^2| + \dots \\ \rho_\infty(x_1, x_2, \dots) &= \max_i |x_i| \end{aligned}$$

are not equivalent, since the sequence $(\xi_n) \subset V$ defined by

$$(4.18) \quad \begin{aligned} \xi_1 &= (1, 0, 0, 0, 0, \dots) \\ \xi_2 &= \left(\frac{1}{2}, \frac{1}{2}, 0, 0, 0, \dots\right), \\ \xi_3 &= \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \dots\right) \end{aligned}$$

converges to 0 with respect to ρ_∞ but lies on the unit sphere with respect to ρ_1 .

(4.19) *All norms in finite dimensions are equivalent.* On a finite dimensional linear space there is a unique choice of topology compatible with the linear structure, where compatibility here means a topology defined by a norm. The same applies with ‘affine’ replacing ‘linear’.

Theorem 4.20. *Let V be a finite dimensional real vector space. Then any two norms on V are equivalent.*

Proof. It suffices to take $V = \mathbb{R}^n$ and prove that an arbitrary norm $\rho: \mathbb{R}^n \rightarrow \mathbb{R}^{>0}$ is equivalent to the 1-norm

$$(4.21) \quad \|(\xi^1, \dots, \xi^n)\|_1 = |\xi^1| + \dots + |\xi^n|.$$

Let $e_1, \dots, e_n$ be the standard basis (2.13) of $\mathbb{R}^n$ and set $C = \max_i \rho(e_i)$. Then for any $\xi = \xi^i e_i \in \mathbb{R}^n$,

$$(4.22) \quad \rho(\xi) = \rho(\xi^i e_i) \leq |\xi^i| \rho(e_i) \leq C \|\xi\|_1.$$

To prove the opposite inequality we apply the Heine-Borel theorem to conclude that the unit sphere in the 1-norm is compact.⁷ Inequality (4.22) implies that ρ is (uniformly) continuous in the 1-norm, and so there exists $\delta > 0$ such that $\rho \geq \delta$ on the unit sphere in the 1-norm. Then for any $\xi \in \mathbb{R}^n$

$$(4.23) \quad \rho(\xi) = \|\xi\|_1 \rho\left(\frac{\xi}{\|\xi\|_1}\right) \geq \delta \|\xi\|_1. \quad \square$$

Corollary 4.24. *A finite dimensional normed linear space is complete.*

Proof. By the Heine-Borel theorem, the statement is true for $\mathbb{R}^n$ with the 2-norm. $\square$

⁷The Heine-Borel theorem, which characterizes compact subsets of $\mathbb{R}^n$ as those which are closed and bounded, is usually proved using the 2-norm, not the 1-norm. To avoid circular reasoning, on the homework you prove directly that the 2-norm and 1-norm are equivalent.

Lecture 5. Continuous linear maps; differentiability

(5.1) Remark on “spaces”. There are several different types of spaces we have encountered already, and at first it may be confusing to keep them apart. So let’s review. On the one hand we have the general notion of a *topological space*. This is quite general, and in this course we will only seriously engage with *metric spaces*. On a topological space we have notions of open set, convergent sequences, continuous maps, etc. Open sets give a qualitative notion of “closeness”, and other notions are derived in those terms. On a metric space we can make measurements—the metric is a distance function—and so closeness becomes more concrete, as do definitions of convergence, continuity, etc. And we have additional notions, such as a Cauchy sequence. A topological space is *metrizable* if the topology is the topology of a metric (which is not specified). In the homework you may have run into a non-metrizable topological space: the moduli space of ordered triples of points on an affine line (if you allow arbitrary coincidences of points). If we topologize a space of infinitely differentiable functions, we will also encounter non-metrizable spaces. However, in this course (except perhaps for an occasional homework problem) we will always use metric spaces, not more general topological spaces.

Another type of space is a *vector space*. This belongs to algebra: it has no topology—the underlying set is discrete—and it has an algebraic structure. It is an abelian group under vector addition, and it has an additional algebraic operation: scalar multiplication. So vector spaces play very different roles than do metric spaces. Definition 3.2 is a marriage of the two kinds of space: a *normed linear space* (NLS) is simultaneously a metric space and a vector space, and the definition enforces compatibilities between the two structures. (To wit, property (1) of Definition 3.2 relates the norm to the zero vector, property (2) relates the norm to scalar multiplication, and property (3) relates the norm to vector addition.) Figure 7 depicts the relationship between the different types of space.

As stated earlier, *affine spaces* are the arena for flat geometry. An affine space over a normed linear space is a metric space (3.16), but of a very particular sort. It provides the setting for calculus, which we begin to develop in this lecture.

Continuous linear maps

Definition 5.2. Let V, W be normed linear spaces. A linear map $T: V \rightarrow W$ is *bounded* if there exists $C > 0$ such that

$$(5.3) \quad \|T\xi\|_W \leq C\|\xi\|_V, \quad \text{for all } \xi \in V.$$

Thus T is bounded if and only if $T(S(V)) \subset W$ is bounded if and only if $T(D(V)) \subset W$ is bounded, where $S(V) \subset V$ is the unit sphere and $D(V) \subset V$ the closed unit ball (both centered at zero).

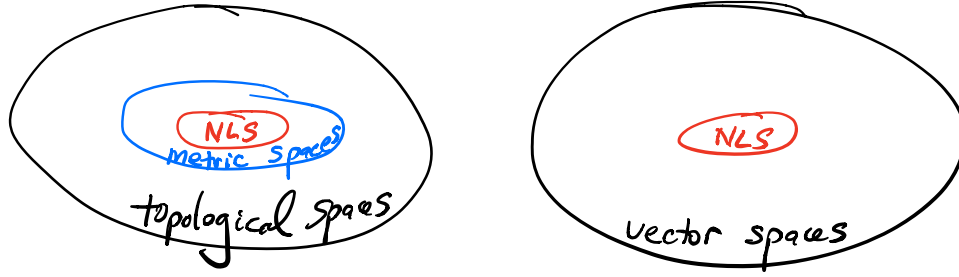


FIGURE 7. Schematic diagram of four different types of space

Theorem 5.4. *Let $T: V \rightarrow W$ be a linear map between normed linear spaces. The following are equivalent:*

- (i) T is bounded
- (ii) T is uniformly continuous
- (iii) T is continuous at $0 \in V$

Proof. To see (i) implies (ii), if T satisfies (5.3) then for all $\xi, \eta \in V$ we have

$$(5.5) \quad \|T\xi - T\eta\|_W \leq C\|\xi - \eta\|_V,$$

from which uniform continuity of T follows immediately.⁸ The implication (ii) implies (iii) is obvious. To prove (iii) implies (i), if T is not bounded then choose $\xi_n \in V$ such that $\|\xi_n\|_V = 1$ and $\|T\xi_n\|_W > n$. Then $\eta_n := \xi_n/n$ satisfies $\lim_{n \rightarrow \infty} \eta_n = 0$ but $\|T\eta_n\|_W > 1$, so the sequence $(T\eta_n)$ does not converge to $0 \in W$. Hence T is not continuous at $0 \in V$. $\square$

Definition 5.6. Let V, W be normed linear spaces. Define

$$(5.7) \quad \text{Hom}(V, W) = \{T: V \rightarrow W \text{ such that } T \text{ is continuous and linear}\}$$

and set $V^* = \text{Hom}(V, \mathbb{R})$.

The expression

$$(5.8) \quad \|T\| := \inf\{C \in \mathbb{R}^{>0} : \|T\xi\|_W \leq C\|\xi\|_V \text{ for all } \xi \in V\}$$

defines a norm on $\text{Hom}(V, W)$.

Example 5.9. Let $V = C^0([0, 1], \mathbb{R})$ be the vector space of continuous functions $[0, 1] \rightarrow \mathbb{R}$ endowed with the sup norm (4.6). Consider the two linear functionals

$$(5.10) \quad \begin{aligned} T_1(f) &= f(1/2), \\ T_2(f) &= f'(1/2). \end{aligned}$$

⁸The estimate (5.5) is called *Lipschitz continuity* with *Lipschitz constant* C .

The first is bounded, so belongs to V^* . The second is only defined on the subspace $V' \subset V$ of continuous functions which are differentiable at $1/2$, and it is unbounded on that subspace: consider the sequence of functions $f_n(x) = \sin(2\pi nx)$ in V' . So $T_2: V' \rightarrow \mathbb{R}$ is not continuous, and it does not extend to a linear map with domain V .

The following shows that any linear map with finite dimensional domain is continuous.

Theorem 5.11. *Let V, W be normed linear spaces and assume V is finite dimensional. Then any linear map $T: V \rightarrow W$ is bounded.*

Proof. Choose a basis $b: \mathbb{R}^n \xrightarrow{\cong} V$, and use the 1-norm (4.21) on $\mathbb{R}^n$. Then Theorem 4.20 implies that b is continuous. So we are reduced to proving that a linear map $S: \mathbb{R}^n \rightarrow W$ is continuous with respect to the 1-norm on $\mathbb{R}^n$. Let $C = \max_i \|S(e_i)\|_W$, where recall the standard basis (2.13) of $\mathbb{R}^n$. Then for any $\xi = \xi^i e_i \in \mathbb{R}^n$ we have

$$(5.12) \quad \|T\xi\|_W \leq |\xi^i| \|Te_i\|_W \leq C(|\xi^1| + \cdots + |\xi^n|) = C\|\xi\|_1. \quad \square$$

Shapes and functions

(5.13) *Expressions and shapes.* Consider the following three expressions:

$$(5.14) \quad \begin{aligned} &(\cos(t), \sin(t)) \\ &x^2 + y^2 = 1 \\ &\sqrt{1 - x^2} \end{aligned}$$

Each evokes a shape: a circle. (The last may evoke only an arc of a circle.) Let us articulate those evocations in the language of sets and functions: for each we define sets X, Y , a function $f: X \rightarrow Y$, and tell how the shape appears as a subset of either Y , X , or $X \times Y$.

(5.15) *Image.* Set $X = \mathbb{R}$, $Y = \mathbb{A}^2$, and

$$(5.16) \quad \begin{aligned} f: \mathbb{R} &\longrightarrow \mathbb{A}^2 \\ t &\longmapsto (\cos(t), \sin(t)) \end{aligned}$$

Then the circle is the *image* $f(X) \subset Y$ of the function f , a subset of the codomain. This image construction parametrizes a shape. Ideally we would have f a bijection onto its image, which is equivalent to f injective. For (5.16) we can achieve that by restricting to a subset of the domain, but we cannot choose that subset to be open.

(5.17) *Preimage.* Set $X = \mathbb{A}^2$, $Y = \mathbb{R}$, and

$$(5.18) \quad \begin{aligned} f: \mathbb{A}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto x^2 + y^2 \end{aligned}$$

Then the circle is the *preimage* $f^{-1}(1) \subset X$, a subset of the domain. A variation of the preimage construction expresses a subset of X as the preimage of a subset of Y which is not a singleton.

(5.19) *Graph.* Set $X = (-1, 1)$, $Y = \mathbb{R}$, and

$$(5.20) \quad \begin{aligned} f: (-1, 1) &\longrightarrow \mathbb{R} \\ x &\longmapsto \sqrt{1 - x^2} \end{aligned}$$

Then an open half-circle is the graph $\Gamma_f \subset X \times Y$, the subset of the Cartesian product defined by

$$(5.21) \quad \Gamma_f = \{(x, f(x)) : x \in X\}.$$

A subset of $X \times Y$ is called a *relation*, and it is a function if the restriction of the projection $X \times Y \rightarrow X$ is a bijection. In other words, from a formal point of view the graph Γ_f is the function f .

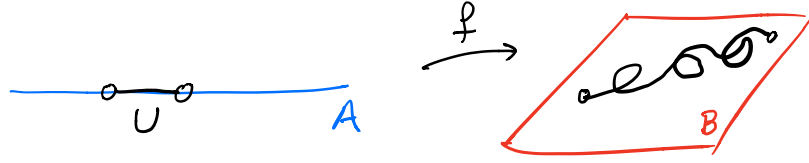
Remark 5.22. The three methods of associating shapes to functions and *visa versa* are quite universal and hold in any mathematical context, such as (say) algebraic geometry. In this course we are interested in *smooth* shapes, so apply these ideas in a setting where we can develop a theory of differentiation. It is to this setting that we now turn.

Setting for calculus

(5.23) *Standard data.* For the next several lectures we work with the following *standard data*.

$$(5.24) \quad \begin{array}{ll} V, W & \text{normed linear spaces} \\ A, B & \text{affine spaces over } V, W \text{ respectively} \\ U \subset A & \text{open subset} \\ f: U \longrightarrow B & \text{function} \end{array}$$

We illustrate with a few examples.

FIGURE 8. A motion in the affine space B

Example 5.25 (motion). If $\dim A = 1$, and $U \subset A$ is an open interval, then we can regard f as describing a parametrized curve in B . More poetically, it is the data of the *motion* (of a particle, say) in B . So A plays the role of time, B plays the role of space, $U \subset A$ is an open interval of time, and f describes position as a function of time. The norm on V gives a measurement of time intervals, and can be thought of as specifying units, such as seconds or hours. A *compatible* affine coordinate $t: A \xrightarrow{\cong} \mathbb{A}^1$ has differential a linear function $dt: V \xrightarrow{\cong} \mathbb{R}$ whose absolute value is the norm. Such a function exists and is unique up to translation and time-reversal (reflection). See Figure 8. The material in a first real analysis course pertains to these functions of a single variable.

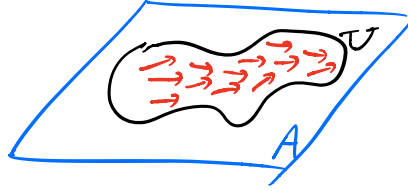


FIGURE 9. A vector field on an open subset of affine space

Example 5.26 (vector field). Suppose $B = W = V$, so that we consider a function $f: U \rightarrow V$. Then to each point $p \in U$ the function f assigns a vector $f(p) \in V$. Figure 9 depicts the function by drawing $f(p)$ as a vector emanating from p . It is regarded here as an infinitesimal displacement, rather than an actual displacement, and in this context f is called a *vector field*. (Recall the discussion in (2.32) and (2.33) of the two roles of V ; here we use the interpretation in (2.33).)

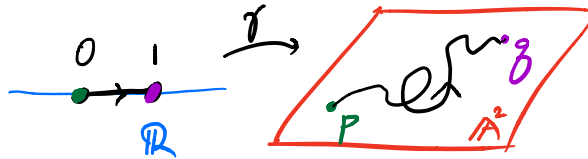


FIGURE 10. A continuously differentiable parametrized curve with fixed endpoints

Example 5.27 (Length of a curve). This example illustrates why we develop calculus allowing functions on infinite dimensional spaces. Fix $p, q \in \mathbb{A}^2$. Let

$$(5.28) \quad A = \{\gamma: [0, 1] \rightarrow \mathbb{A}^2 \text{ such that } \gamma \text{ is continuously differentiable, } \gamma(0) = p, \gamma(1) = q\}.$$

Then A is an infinite dimensional affine space over the infinite dimensional vector space

$$(5.29) \quad V = \{\xi: [0, 1] \rightarrow \mathbb{R}^2 \text{ such that } \xi \text{ is continuously differentiable, } \xi(0) = \xi(1) = 0\}.$$

The function $f: A \rightarrow \mathbb{R}$ defined by

$$(5.30) \quad f(\gamma) = \int_0^1 dt \|\dot{\gamma}(t)\|_{\mathbb{R}^2}$$

computes the length of γ if we use the standard Pythagorean norm (3.17) with $p = 2$. A typical problem in the *calculus of variations* is to minimize f . As in finite dimensions we solve it by computing the critical points of f , which means we must learn how to differentiate a function of infinitely many variables, as well as the theory behind the differentiation.

Continuity and differentiability

(5.31) *Recollection of continuity.* Fix an instantiation of standard data (5.24) and a point $p \in U$. Since A is a metric space, the subset $U \subset A$ inherits a metric space. Then f is a function between metric spaces, and there is a standard definition of continuity of f at p . Recall that for any $\delta > 0$ the open ball of radius δ about p is denoted $B_\delta(p)$.

Definition 5.32. f is *continuous at p* if for all $\epsilon > 0$ there exists $\delta > 0$ such that $B_\delta(p) \subset U$ and $f(B_\delta(p)) \subset B_\epsilon(f(p))$.

Our hypothesis that $U \subset A$ is open guarantees that $B_\delta(p) \subset U$ for δ sufficiently small. We can restate the condition in language adapted to affine space: if $\xi \in V$ satisfies $\|\xi\|_V < \delta$, then $p + \xi \in U$ and

$$(5.33) \quad \|f(p + \xi) - f(p)\|_W < \epsilon.$$

A heuristic interpretation of continuity at p is that f is well-approximated near p by the constant function with value $f(p)$. The estimate (5.33) quantifies this heuristic statement.

(5.34) *Definition of differentiability.* Heuristically, the function f is differentiable at p if it is well-approximated near p by an affine function

$$(5.35) \quad \alpha_p(p + \xi) = f(p) + T\xi, \quad \xi \in V,$$

for some $T \in \text{Hom}(V, W)$.

Definition 5.36. Fix standard data (5.24) and a point $p \in U$. Then f is *differentiable at p* if there exists $T \in \text{Hom}(V, W)$ such that for all $\epsilon > 0$ there exists $\delta > 0$ such that if $\|\xi\|_V < \delta$, then $p + \xi \in U$ and

$$(5.37) \quad \|f(p + \xi) - [f(p) + T\xi]\|_W \leq \epsilon \|\xi\|_V.$$

Example 5.38. To illustrate, consider $U = A = B = \mathbb{R}$ and $f(x) = x^3$ at $x = 2$. Define the affine approximation to f near $x = 2$ as $\alpha(2 + h) = 8 + 12h$. To check that this is a good affine approximation we must estimate

$$(5.39) \quad |f(2 + h) - \alpha(2 + h)| = |[8 + 12h + 6h^2 + h^3] - [8 + 12h]| = |h| |6h + h^2|,$$

which, if $|h|$ is sufficiently small, is less than any given $\epsilon > 0$ times $|h|$.

Lemma 5.40. If $T, T' \in \text{Hom}(V, W)$ satisfy Definition 5.36, then $T' = T$.

Proof. Fix $\eta \in V$ with $\|\eta\|_V = 1$ and suppose $\epsilon > 0$ is given. Choose $\delta, \delta' > 0$ as in Definition 5.36 for T, T' , respectively, and fix $0 < t < \min(\delta, \delta')$. Then

$$(5.41) \quad \begin{aligned} \|T'(t\eta) - T(t\eta)\|_W &= \| [f(p + t\eta) - f(p) - T'(t\eta)] - [f(p + t\eta) - f(p) - T(t\eta)] \|_W \\ &\leq 2\epsilon t, \end{aligned}$$

from which $\|T'\eta - T\eta\|_W \leq 2\epsilon$. Since this is true for all $\epsilon > 0$ we conclude $T'\eta = T\eta$. Apply linearity to conclude $T' = T$. $\square$

Definition 5.42. If f is differentiable at p , we call the unique continuous linear map T in Definition 5.36 the *differential* of f at p and use the notation $T = df_p$.

Remark 5.43. If $f: A \rightarrow B$ is an affine map (Definition 1.10), then f is differentiable for all $p \in A$ and df_p is independent of p . In other words, an affine map has a constant differential.

Remark 5.44. If $f: U \rightarrow B$ is differentiable at all $p \in U$, then the differential is a map

$$(5.45) \quad df: U \longrightarrow \text{Hom}(V, W).$$

The map df is an example of standard data (5.24) (spell it out!), and so we can consider whether df is differentiable using Definition 5.36.

In the next lecture we investigate how to compute the differential df_p .

Lecture 6: Computation of the differential

In this lecture we continue to work in the context of standard data (5.23). Henceforth we drop the subscripts ‘ V ’ and ‘ W ’ on the norms, since it is clear from the context which we mean. We also use the operator norm (5.8) on $\text{Hom}(V, W)$ without explicit labeling.

Differentiability and continuity

A differentiable function is continuous, as we now prove.

Theorem 6.1. *Suppose f is differentiable at $p \in U$. Then f is continuous at p .*

Proof. Let $C = \|df_p\|$ be the operator norm of the differential at p . Apply Definition 5.36 with $\epsilon = 1$ to produce $\delta_0 > 0$ such that if $\|\xi\| < \delta_0$, then (5.37) is satisfied. The triangle inequality implies

$$(6.2) \quad \begin{aligned} \|f(p + \xi) - f(p)\| &\leq \|f(p + \xi) - f(p) - df_p(\xi)\| + \|df_p(\xi)\| \\ &< (1 + C)\|\xi\|. \end{aligned}$$

Given $\epsilon > 0$ choose $\delta = \min(\delta_0, 1/(1 + C))$ to satisfy Definition 5.32 of continuity at p . $\square$

If the differential of f exists at all points of U , then we can inquire about the continuity of the differential as a map (5.45).

Definition 6.3. If f is differentiable on U and $df: U \rightarrow \text{Hom}(V, W)$ is continuous, then we say f is *continuously differentiable*.

Functions of one variable

A special case of our general context (5.23) is the situation studied in a first analysis course. Then $A = \mathbb{R}$ is the real line and $U \subset \mathbb{R}$ may as well be connected, in which case it is an open interval (a, b) for some real numbers $a < b$. Then $g: (a, b) \rightarrow B$ is a function of one variable. The simplest situation is $B = \mathbb{R}$, so one function of one variable; if $B = \mathbb{A}^m$, then $g = (g^1, \dots, g^m)$ is m functions of one variable. It is easier in terms of notation to take the codomain B to be an affine space over an arbitrary normed linear space W , and we need this generality later anyhow. Recall (Example 5.25) that we can interpret g as describing a motion in B .

For functions of one variable we define the derivative to be the limit of difference quotients. We foreshadowed the following in (2.33).

Definition 6.4. We say g is *old style differentiable* at $t_0 \in (a, b)$ if

$$(6.5) \quad \lim_{h \rightarrow 0} \frac{g(t_0 + h) - g(t_0)}{h}$$

exists, in which case we notate the limit as $g'(t_0) \in W$.

In (6.5) the numerator is the displacement vector between two points of B , and it is scalar multiplied by $1/t$.

Proposition 6.6. *If $g: (a, b) \rightarrow B$ is old style differentiable at $t_0 \in (a, b)$, then it is differentiable at t_0 and*

$$(6.7) \quad dg_{t_0}(h) = hg'(t_0), \quad h \in \mathbb{R}.$$

Any linear function $\mathbb{R} \rightarrow W$ is determined by its value at 1, which is a vector in W . The statement is that for dg_{t_0} that vector is $g'(t_0)$. We leave the reader to formulate and prove the converse to Proposition 6.6.

Proof. Given $\epsilon > 0$ use the existence of (6.5) to choose $\delta > 0$ such that $(t_0 - \delta, t_0 + \delta) \subset (a, b)$ and if $0 < |h| < \delta$ then

$$(6.8) \quad \left\| \frac{g(t_0 + h) - g(t_0)}{h} - g'(t_0) \right\| < \epsilon.$$

Now multiply through by $|h|$ to deduce the estimate in Definition 5.36. (If $h = 0$ that estimate is trivial.) $\square$

Computation of the differential

We say a motion $\gamma: (a, b) \rightarrow A$ has *constant velocity* if it is differentiable and $\gamma'(t)$ is independent of t . In that case γ extends to an affine map $\mathbb{R} \rightarrow A$. Given p, ξ there is a unique constant velocity motion $t \mapsto p + t\xi$ with initial position p and velocity ξ .

Now return to our standard data (5.23) and fix $p \in U$ and $\xi \in V$. Our task is to compute $df_p(\xi) \in W$, assuming f is differentiable at p . The idea is to use the “tea kettle principle”⁹ to reduce to the derivative of a function of one variable, since in that case the differential is computed by the limit of a difference quotient (6.5), and then we have all the techniques and formulas of one-variable calculus available. Let

$$(6.9) \quad \begin{aligned} \gamma: (-r, r) &\longrightarrow U \\ t &\longmapsto p + t\xi \end{aligned}$$

be the indicated constant velocity motion, where $r > 0$ is chosen sufficiently small so that the image lies in the open set $U \subset A$.

⁹A mathematician is asked to move a tea kettle from the stove to the sink, which is readily accomplished. The next day the same mathematician is asked to move the tea kettle from the counter to the sink. Solution: move the tea kettle to the stove, thereby reducing the problem to one previously solved.

Theorem 6.10. *If f is differentiable at p , then $f \circ \gamma$ is old style differentiable at 0 and*

$$(6.11) \quad df_p(\xi) = (f \circ \gamma)'(0).$$

Figure 11 depicts the situation in the theorem. In the next lecture we prove a generalization in which γ need not be a constant velocity motion; it need only have initial position p and *initial* velocity ξ .

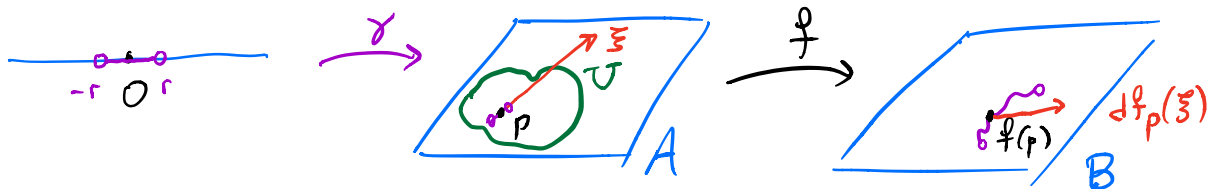


FIGURE 11. Computing the differential

Proof. We may assume $\xi \neq 0$. Since f is differentiable at p , given $\epsilon > 0$ choose $\delta > 0$ so that if $\eta \in V$ satisfies $\|\eta\| < \delta$, then $p + \eta \in U$ and

$$(6.12) \quad \|f(p + \eta) - f(p) - df_p(\eta)\| \leq \epsilon \frac{\|\eta\|}{\|\xi\|}.$$

Then for $0 < |t| < \delta/\|\xi\|$,

$$(6.13) \quad \left\| \frac{f(p + t\xi) - f(p)}{t} - df_p(\xi) \right\| \leq \epsilon.$$

This proves the limit of the difference quotient exists and equals $df_p(\xi)$. □

Definition 6.14. We call

$$(6.15) \quad \left. \frac{d}{dt} \right|_{t=0} f(p + t\xi)$$

the *directional derivative of f at p in the direction ξ* and denote it as $\xi f(p)$.

Thus if f is differentiable in U , then given ξ we can differentiate at every point in the direction ξ (using the global parallelism of affine space) to obtain a function

$$(6.16) \quad \xi f: U \rightarrow \mathbb{R}.$$

Remark 6.17. Theorem 6.10 asserts that if f is differentiable at p , then all directional derivatives at p exist. In the next lecture we prove a converse statement—if directional derivatives exist then f is differentiable—but with restrictions: we assume the domain is finite dimensional and that directional derivatives exist in a neighborhood of p .

Now suppose the domain U is an open subset of the standard affine space $A = \mathbb{A}^n$ for some $n \in \mathbb{Z}^{>0}$. Recall (2.24) the standard affine coordinate functions $x^i: \mathbb{A}^n \rightarrow \mathbb{R}$. In this situation we denote the standard basis elements of the vector space $\mathbb{R}^n$ of translations as

$$(6.18) \quad \frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}.$$

The notation is set up so that the directional derivative in the direction of a basis element

$$(6.19) \quad \frac{\partial}{\partial x^j} f = \frac{\partial f}{\partial x^j}: U \rightarrow \mathbb{R}$$

is the *partial derivative* in the j^{th} coordinate direction. If the codomain $B = \mathbb{A}^m$ is also a standard finite dimensional affine space, then we write $f = (f^1, \dots, f^m)$ for functions $f^i: U \rightarrow \mathbb{R}$, and then at each $p \in U$ obtain a matrix¹⁰

$$(6.20) \quad \left(\frac{\partial f^i}{\partial x^j}(p) \right)$$

of partial derivatives. It is the matrix which represents the linear map $df_p: \mathbb{R}^n \rightarrow \mathbb{R}^m$ in the standard bases.

The operator d and explicit computation

To compute the differential explicitly we observe that the operator d obeys the usual rules of differentiation, as follows from Theorem 6.10 and standard theorems of one-variable calculus. Namely,

- (1) d is linear: $d(f_1 + f_2) = df_1 + df_2$
- (2) d obeys the Leibniz rule: $d(f_1 \cdot f_2) = df_1 \cdot f_2 + f_1 \cdot df_2$

Notice that we do not exchange the order of the product, which is a good habit since for non-commutative products, as of matrix-valued functions, the same formula applies and one cannot permute factors. Then, after the application of d , we can collect terms and permute factors as allowed. The other basic rule for computing d is the chain rule, which we prove in the next lecture, though of course we already know it for functions of one variable. Using these rules we have a good algorithmic technique and can compute without thinking.

As an example we take $U = A = B = \mathbb{A}^2$, label the standard affine coordinates (r, θ) in the domain and (x, y) in the codomain, and define a function $f: \mathbb{A}_{(r, \theta)}^2 \rightarrow \mathbb{A}_{(x, y)}^2$ by the formulas

$$(6.21) \quad \begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned}$$

¹⁰The superscript j in the denominator is an overall subscript, so i is a superscript and j a subscript. As a matrix i is the row number and j the column number.

We could have written $f(r, \theta) = (r \cos \theta, r \sin \theta)$, but (6.21) is set up for easy computation without thinking, and there are fewer symbols: ‘ f ’ does not appear. So simply follow your nose and apply d :

$$(6.22) \quad \begin{aligned} dx &= dr \cos \theta + r d(\cos \theta) \\ &= \cos \theta dr - r \sin \theta d\theta \end{aligned}$$

The equality $d(\cos \theta) = -\sin \theta d\theta$ follows from the chain rule applied to the composition

$$(6.23) \quad \mathbb{A}^2 \xrightarrow{\theta} \mathbb{R} \xrightarrow{\cos} \mathbb{R},$$

but one gets used to computing without thinking through these justifications. (Do think through them at the beginning!) In the end, applying d to (6.21), we obtain the equations

$$(6.24) \quad \begin{aligned} dx &= \cos \theta dr - r \sin \theta d\theta \\ dy &= \sin \theta dr + r \cos \theta d\theta \end{aligned}$$

Recall from Remark 5.43 that the differentials $dr, d\theta: \mathbb{A}^2 \rightarrow (\mathbb{R}^2)^*$ of the affine functions $r, \theta: \mathbb{A}^2 \rightarrow \mathbb{R}$ are constant on $\mathbb{A}^2$, and they form a basis of $(\mathbb{R}^2)^*$. As in (6.18) the dual basis of $\mathbb{R}^2$ is denoted $\partial/\partial r, \partial/\partial \theta$. Evaluate (6.24) on $\partial/\partial r$ to see that the image of the vector $\partial/\partial r$ under the differential of f at (r, θ) is the vector

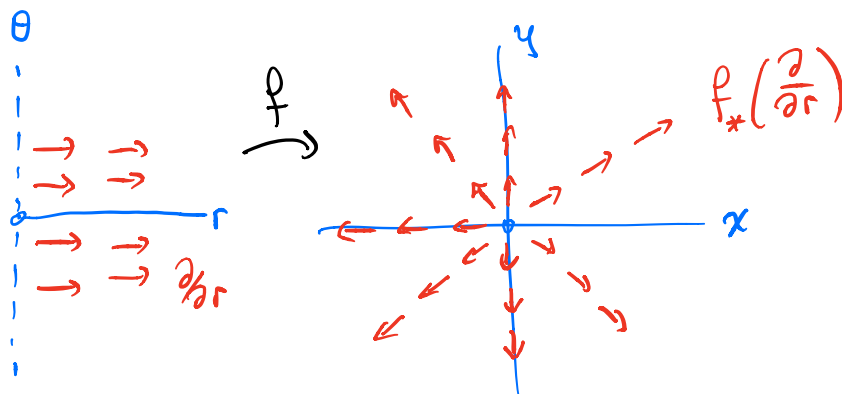
$$(6.25) \quad \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y},$$

and the image of the vector $\partial/\partial \theta$ under the differential of f at (r, θ) is the vector

$$(6.26) \quad -r \sin \theta \frac{\partial}{\partial x} + r \cos \theta \frac{\partial}{\partial y}.$$

Remark 6.27. It is worth contemplating this example in some detail to extract some general lessons. We might be tempted to take the image of the (constant) vector *field* $\partial/\partial r$ under df to construct a vector field on $\mathbb{A}^2$. But that is not possible. Observe that $f(0, \theta) = (0, 0)$ for all $\theta \in \mathbb{R}$, so to define the value of the supposed image vector field at $(0, 0)$ in the codomain we have many choices of which preimage point to use. And (6.25) shows that the vector we obtain is *not* independent of the choice of θ . So there is no well-defined image vector field. If restrict the domain of f to $r > 0$, then each $(x, y) \neq (0, 0)$ in the codomain has a collection of preimage points (r, θ) in which any two have the same value of r and values of θ differing by an integer multiple of 2π . Put differently, the preimage is a $\mathbb{Z}$ -torsor (Definition 2.22) for the action $n: (r, \theta) \rightarrow (r, \theta + 2\pi n)$ of $\mathbb{Z}$ on $\mathbb{A}_{(r, \theta)}^2$. Now formula (6.25) shows that the image vector is independent of the choice of preimage, and so there is a well-defined image vector *field*. We depict the image of $\partial/\partial r$ in Figure 12.

Another observation is that the transpose of the differential, which for our general data is a map $df_p^*: W^* \rightarrow V^*$ or $df^*: U \rightarrow \text{Hom}(W^*, V^*)$, is what is globally defined always and is what one computes directly. That is one interpretation of (6.24): the right hand side at each (r, θ) is the value of $df_{(r, \theta)}^*$ on dx, dy , respectively.

FIGURE 12. Image of the vector field $\partial/\partial r$

Lecture 7: Further properties of the differential

Chain rule

The chain rule can be summarized in the slogan “the affine approximation to a composition of functions is the composition of the affine approximations”. The precise statement is as follows.

Theorem 7.1. *Let V, W, X be normed vector spaces; A, B, C affine spaces over V, W, X , respectively; $U \subset A$, $U' \subset B$ open sets; $f: U \rightarrow U'$, $g: U' \rightarrow C$ functions; and $p \in U$. Assume f is differentiable at p and g is differentiable at $f(p)$. Then $g \circ f$ is differentiable at p and*

$$(7.2) \quad d(g \circ f)_p = dg_{f(p)} \circ df_p.$$

Recall that df_p is a continuous (bounded) linear map $V \rightarrow W$ and $dg_{f(p)}$ a continuous linear map $W \rightarrow X$, so (7.2) is an equation of continuous linear maps $V \rightarrow X$.

For convenience, denote $q = f(p)$.

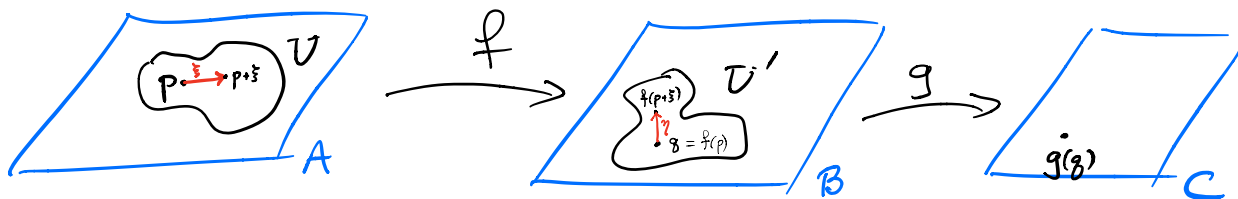


FIGURE 13. Composition of functions

Proof. The differentiability hypotheses imply that given $\epsilon_1, \epsilon_2 > 0$ there exist $\delta_1, \delta_2 > 0$ such that if $\xi \in V$, $\eta \in W$ satisfy $\|\xi\| < \delta_1$, $\|\eta\| < \delta_2$, then $p + \xi \in U$, $q + \eta \in U'$, and

$$(7.3) \quad \begin{aligned} \|f(p + \xi) - f(p) - df_p(\xi)\| &\leq \epsilon_1 \|\xi\| \\ \|g(q + \eta) - g(q) - dg_q(\eta)\| &\leq \epsilon_2 \|\eta\| \end{aligned}$$

Set

$$(7.4) \quad \delta = \min\left(\delta_1, \frac{\delta_2}{\epsilon_1 + \|df_p\|}\right).$$

Fix $\xi \in V$ with $\|\xi\| < \delta$ and set $\eta = f(p + \xi) - f(p)$. Then (7.3) implies

$$(7.5) \quad \|\eta - df_p(\xi)\| \leq \epsilon_1 \|\xi\|,$$

and the triangle inequality and (7.5) imply

$$(7.6) \quad \|\eta\| \leq \|\eta - df_p(\xi)\| + \|df_p(\xi)\| \leq (\epsilon_1 + \|df_p\|)\|\xi\| < \delta.$$

Then

$$(7.7) \quad \begin{aligned} \|g(f(p + \xi)) - g(f(p)) - dg_q(df_p(\xi))\| &\leq \|g(q + \eta) - g(q) - dg_q(\eta)\| + \|dg_q(\eta - df_p(\xi))\| \\ &\leq \epsilon_2 \|\eta\| + \|dg_q\| \epsilon_1 \|\xi\| \\ &\leq (\epsilon_2(\epsilon_1 + \|df_p\|) + \epsilon_1 \|dg_q\|)\|\xi\| \end{aligned}$$

We use this estimate to prove the theorem. Given $\epsilon > 0$ choose

$$(7.8) \quad \begin{aligned} \epsilon_1 &= \frac{1}{2} \frac{\epsilon}{\|dg_q\|} \\ \epsilon_2 &= \frac{1}{2} \frac{\epsilon}{\epsilon_1 + \|df_p\|} \end{aligned}$$

Then pick δ_1, δ_2 as before (7.3) and define δ by (7.4). So if $\xi \in V$ satisfies $\|\xi\| < \delta$, then

$$(7.9) \quad \|g(f(p + \xi)) - g(f(p)) - dg_q(df_p(\xi))\| \leq \epsilon \|\xi\|.$$

This proves that $g \circ f$ is differentiable at p with differential $dg_q \circ df_p$. □

Remark 7.10. The affine approximation to f at p is

$$(7.11) \quad \begin{aligned} A &\longrightarrow B \\ p + \xi &\longmapsto f(p) + df_p(\xi) \end{aligned}$$

and the affine approximation to g at $f(p)$ is

$$(7.12) \quad \begin{aligned} B &\longrightarrow C \\ f(p) + \eta &\longmapsto g(f(p)) + dg_{f(p)}(\eta) \end{aligned}$$

The composite of (7.11) and (7.12) is the affine approximation

$$(7.13) \quad \begin{aligned} A &\longrightarrow C \\ p + \xi &\longmapsto g(f(p)) + dg_{f(p)} \circ df_p(\xi) \end{aligned}$$

to $g \circ f$ at p , using the chain rule (7.2).

Remark 7.14. If $A = \mathbb{A}_{(x^1, \dots, x^n)}^n$, $B = \mathbb{A}_{(y^1, \dots, y^m)}^m$, and $C = \mathbb{A}_{(z^1, \dots, z^\ell)}^\ell$, then the function f is written

$$(7.15) \quad y^j = y^j(x^1, \dots, x^n), \quad j = 1, \dots, m,$$

and the function g is written

$$(7.16) \quad z^k = z^k(y^1, \dots, y^m), \quad k = 1, \dots, \ell.$$

So df is represented by the $m \times n$ matrix $(\partial y^j / \partial x^i)$ and dg by the $\ell \times m$ matrix $(\partial z^k / \partial y^j)$; cf. (6.20). The chain rule implies that $d(g \circ f)$ is represented by the product of the matrices. However, unless you are doing multiple explicit computations you will find it easier to compute as in (6.24).

(7.17) Directional derivatives revisited. We continue with standard data (5.23). Recall from Definition 6.14 that the directional derivative is defined as the derivative along an affine motion with prescribed initial position and initial velocity. The following corollary of the Chain Rule Theorem 7.1 tells that we can use any motion with the correct initial conditions.

Corollary 7.18. *Assume f is differentiable at $p \in U$. Fix $a > 0$ and let $\gamma: (-a, a) \rightarrow U$ be a curve such that $\gamma(0) = p$ and γ is differentiable at 0. Denote $\xi = \dot{\gamma}(0) \in V$. Then*

$$(7.19) \quad df_p(\xi) = \left. \frac{d}{dt} \right|_{t=0} f(\gamma(t)).$$

Proof. Use the relation (6.7) between the differential and the derivative of a function of one variable together with the chain rule (7.2):

$$(7.20) \quad \begin{aligned} \left. \frac{d}{dt} \right|_{t=0} f(\gamma(t)) &= d(f \circ \gamma)_0(1) \\ &= df_p(d\gamma_0(1)) \\ &= df_p(\gamma'(0)) \\ &= df_p(\xi). \end{aligned}$$

□

Remark 7.21. When we move to calculus on smooth curved spaces, such as the surface of a sphere, then there is no canonical motion with given initial position and initial velocity, as there is in affine space. Corollary 7.18 is crucial in that context.

Mean value inequality

For a motion in an affine space of dimension at least two there is no mean value theorem, which would state that the average velocity over the motion is realized as the instantaneous velocity at some particular time. Figure 14 depicts a helical motion which illustrates this point. On the other hand, there is an inequality which holds: if the speed of a motion is bounded by $C > 0$, and the total time is T , then the total distance traveled is $\leq CT$.

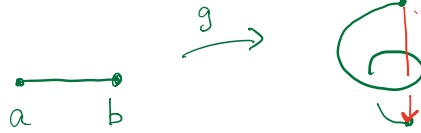


FIGURE 14. Helical motion in 3-space

Theorem 7.22. Suppose $g: [a, b] \rightarrow B$ is a motion in an affine space B over a normed linear space W . Assume g is differentiable on (a, b) and continuous on $[a, b]$. Assume there exists $C > 0$ such that $\|g'(t)\| \leq C$ for all $t \in (a, b)$. Then

$$(7.23) \quad \|g(b) - g(a)\| \leq C(b - a).$$

In the proof we give some wiggle room to make estimates. The technique of freeing up one variable (b) and bootstrapping from the knowledge of the theorem for special values (b near a) is a common one and worth contemplating carefully. We use a variation in the proof of Theorem 7.29.

Proof. Suppose $\epsilon > 0$. Define

$$(7.24) \quad I = \{t \in [a, b] \text{ such that } \|g(t) - g(a)\| \leq (C + \epsilon)(t - a) + \epsilon\}.$$

Since g is continuous at a , there exists $\delta > 0$ such that if $a \leq t < a + \delta$ then

$$(7.25) \quad \|g(t) - g(a)\| < \epsilon < (C + \epsilon)(t - a) + \epsilon.$$

Hence $[a, a + \delta) \subset I$, so in particular I is nonempty. Let $t_0 = \sup I$. If $t_0 < b$ then $g'(t_0)$ exists and there exists $\delta' > 0$ such that if $t_0 < t < t_0 + \delta'$ we have

$$(7.26) \quad \left\| \frac{g(t) - g(t_0)}{t - t_0} - g'(t_0) \right\| < \epsilon.$$

Then¹¹

$$(7.27) \quad \begin{aligned} \|g(t) - g(a)\| &\leq \|g(t) - g(t_0)\| + \|g(t_0) - g(a)\| \\ &\leq (C + \epsilon)(t - t_0) + (C + \epsilon)(t_0 - a) + \epsilon \\ &= (C + \epsilon)(t - a) + \epsilon, \end{aligned}$$

from which $t \in I$. This contradicts $t_0 = \sup I$, hence also the assumption $t_0 < b$. Therefore, $b \in I$. $\square$

¹¹Observe that $t_0 \in I$: take a sequence $t_n \nearrow t_0$ and take the limit of the inequality in (7.24), using the fact that g and the norm are both continuous.

(7.28) *Local and global constancy.* Return to our standard data (5.23) and assume $df_p = 0$ for all $p \in U$. This is an *infinitesimal* hypothesis—a constraint on the differential of the function. The existence of a good affine approximation allows us to pass from the infinitesimal hypothesis to a *local* conclusion, that is, a conclusion about the behavior of f in a neighborhood of the point where the hypothesis holds. In this case that is all points of U , and the conclusion is that f is *locally constant*, i.e., about every $p \in U$ there exists a neighborhood $U_p \subset U$ such that f is constant on U_p . The following theorem includes a topological hypothesis on U —connectivity—to allow passage from local to *global*, hence in total from infinitesimal to global.

Theorem 7.29. *Assume U is connected, f is differentiable, and $df_p = 0$ for all p . Then f is constant.*

Proof. Fix $p_0 \in U$ and define

$$(7.30) \quad S = \{p \in U \text{ such that } f(p) = f(p_0)\}.$$

We prove that (i) S is nonempty, (ii) S is closed, (iii) S is open. It then follows that $S = U$ since U is connected, and so f is constant. For (i) we simply observe $p_0 \in S$, and (ii) follows from the continuity of f (Theorem 6.1). For (iii) suppose $p \in S$ and choose $\delta > 0$ so that $B_\delta(p) \subset U$. Suppose $\xi \in V$ satisfies $\|\xi\| < \delta$. Define

$$(7.31) \quad \begin{aligned} g: (-1, 1) &\longrightarrow U \\ t &\longmapsto f(p + t\xi) \end{aligned}$$

Then $g'(t) = df_{p+t\xi}(\xi) = 0$ for all $t \in (-1, 1)$. Thus for any $C > 0$ Theorem 7.22 implies

$$(7.32) \quad \|f(p + \xi) - f(p)\| = \|g(1) - g(0)\| \leq C,$$

from which we conclude $B_\delta(p) \subset S$. Hence S is open as claimed. □

Lecture 8: Differentials and local extrema; inner products

Continuous differentiability

We continue in our standard setup (5.23). If f is differentiable on U , then the differential is a map

$$(8.1) \quad df: U \longrightarrow \text{Hom}(V, W).$$

Recall that $\text{Hom}(V, W)$ has a norm (5.8)—the operator norm—and so it makes sense to talk about the continuity of df .

Definition 8.2. The function f is *continuously differentiable*, or is a C^1 function, if df exists and is continuous.

We write $C^1(U; B)$ for the space of C^1 functions on U with values in B . It is an affine space over $C^1(U; W)$.

Proposition 8.3. Suppose A is finite dimensional and f is differentiable. Then f is C^1 if and only if all directional derivatives are continuous.

It suffices that the directional derivatives along a basis of V be continuous. In Theorem 8.6 we prove that differentiability follows from the existence and continuity of the directional derivatives.

Proof. If f is C^1 and $\xi \in V$, then the directional derivative is the composition

$$(8.4) \quad \xi f: U \xrightarrow{df} \text{Hom}(V, W) \xrightarrow{\text{ev}_\xi} W$$

of continuous functions, where ev is evaluation. (This direction does not require A to be finite dimensional.) Conversely, assume all directional derivatives $e_i f$ are continuous at $p \in U$ for a basis $e_1, \dots, e_n$ of V . Then given $\epsilon > 0$ for each $i = 1, \dots, n$ choose $\delta_i > 0$ such that if $\|\xi\| < \delta_i$ then $p + \xi \in U$ and $\|e_i f(p + \xi) - e_i f(p)\| < \epsilon$. Then for $\eta = \eta^i e_i$ if $\|\xi\| < \delta = \min_i \delta_i$ we have

$$(8.5) \quad \|df_{p+\xi}(\eta^i e_i) - df_p(\eta^i e_i)\| \leq |\eta^i| \|e_i f(p + \xi) - e_i f(p)\| \leq \epsilon \sum |\eta^i|.$$

Since all norms in finite dimensions are equivalent (Theorem 4.20), we can use the 1-norm $\|\eta\| = \sum |\eta^i|$ relative to the basis $e_1, \dots, e_n$, and then (8.5) gives the desired estimate for the operator norm: $\|df_{p+\xi} - df_p\| < \epsilon$. $\square$

Next we strengthen Proposition 8.3 by dropping the hypothesis that f is differentiable, and instead deduce the differentiability of f from the existence and continuity of the partial derivatives. The following is then a converse of Theorem 6.10, but with additional hypotheses.

Theorem 8.6. *Let V, W be normed linear spaces with V finite dimensional. Let A, B be affine over V, W . Let $U \subset V$ be an open subset and $f: U \rightarrow B$ a function. Assume the directional derivatives $\xi f: U \rightarrow V$ exist and are continuous for ξ running over a basis of V . Then f is continuously differentiable.*

As is evident from the proof to conclude that f is differentiable at a point $p \in U$ we need only assume the directional derivatives are continuous at p .

Proof. We may as well assume $A = \mathbb{A}^n$; use the notation $e_i = \partial/\partial x^i$, $i = 1, \dots, n$, for the standard basis of $V = \mathbb{R}^n$; and use the 1-norm (4.21) on $\mathbb{R}^n$. Fix $p \in U$. Given $\epsilon > 0$ choose $\delta > 0$ such that $B_\delta(p) \subset U$ and

$$(8.7) \quad \left\| \frac{\partial f}{\partial x^i}(q) - \frac{\partial f}{\partial x^i}(p) \right\| < \epsilon, \quad q \in B_\delta(p), \quad i = 1, \dots, n.$$

Suppose $\xi = \xi^i e_i \in V$ satisfies $\|\xi\| < \delta$. For $i = 1, \dots, n$ define the function

$$(8.8) \quad \begin{aligned} g_i: [0, \xi^i] &\longrightarrow W \\ t &\longmapsto f(p + \xi^1 e_1 + \dots + \xi^{i-1} e_{i-1} + t e_i) \\ &\quad - f(p + \xi^1 e_1 + \dots + \xi^{i-1} e_{i-1}) - t \frac{\partial f}{\partial x^i}(p) \end{aligned}$$

Then

$$(8.9) \quad g'_i(t) = \frac{\partial f}{\partial x^i}(p + \xi^1 e_1 + \dots + \xi^{i-1} e_{i-1} + t e_i) - \frac{\partial f}{\partial x^i}(p),$$

and so by (8.7) we conclude $\|g'_i(t)\| < \epsilon$ for all i . The mean value inequality Theorem 7.22 implies

$$(8.10) \quad \|g_i(\xi^i) - g_i(0)\| \leq \epsilon |\xi^i|, \quad i = 1, \dots, n.$$

Now estimate using the triangle inequality:

$$(8.11) \quad \|f(p + \xi) - f(p) - \xi^i \frac{\partial f}{\partial x^i}(p)\| \leq \|g_1(\xi^1)\| + \dots + \|g_n(\xi_n)\| \leq \epsilon(|\xi^1| + \dots + |\xi^n|) = \epsilon \|\xi\|.$$

The function

$$(8.12) \quad \begin{aligned} T_p: \mathbb{R}^n &\longrightarrow W \\ \xi^i e_i &\longmapsto \xi^i \frac{\partial f}{\partial x^i}(p) \end{aligned}$$

is linear, and (8.11) shows that f is differentiable at p with $df_p = T_p$. Then Proposition 8.3 implies that df is continuous. $\square$

Local extrema and critical points

(8.13) Local behavior of a real-valued function. An extremum of a function $f: U \rightarrow \mathbb{R}$ is a maximum or a minimum. Our next result is another instance of the derivative determining local behavior of a function, especially if the differential is “generic”. In this case for $p \in U$ the differential is a linear function $df_p: V \rightarrow \mathbb{R}$ and the generic behavior is that df_p be nonzero, which is equivalent to df_p surjective. In that case f is well-approximated near p by a nonconstant affine function, and so in a neighborhood of p takes values which are both larger and smaller than $f(p)$, i.e., f does not have a local extremum at p . Theorem 8.16 below is the contrapositive statement.

Example 8.14. If $df_p = 0$ then we cannot deduce the local behavior of f near p without further information. For example, consider $f_1, f_2, f_3: (-1, 1) \rightarrow \mathbb{R}$ defined by $f_1 = x^2$, $f_2 = -x^2$, and $f_3 = x^3$ near $x = 0$.

Definition 8.15. Let V be a linear space, A an affine space over V , $U \subset A$ an open set, and $f: U \rightarrow \mathbb{R}$ a function. Then f has a *local minimum* at $p \in U$ if there exists $\delta > 0$ such that $B_\delta(p) \subset U$ and for all $q \in B_\delta(p)$ we have $f(q) \geq f(p)$. Similarly, f has a *local maximum* at $p \in U$ if there exists $\delta > 0$ such that $B_\delta(p) \subset U$ and for all $q \in B_\delta(p)$ we have $f(q) \leq f(p)$. We say f has a *local extremum* at $p \in U$ if either holds.

Theorem 8.16. Let V be a normed linear space, A an affine space over V , $U \subset A$ an open set, and $f: U \rightarrow \mathbb{R}$ a function. Suppose f has a local extremum at $p \in U$ and f is differentiable at p . Then $df_p = 0$.

Proof. For any $\xi \in V$ consider the function $g(t) = f(p + t\xi)$, defined for $t \in (-\delta, \delta)$ for δ sufficiently small. Then g has a local extremum at $t = 0$. A standard theorem in one-variable calculus, proved by examining the sign of the difference quotient, asserts that $g'(0) = 0$. It follows from Theorem 6.10 that $df_p(\xi) = 0$. Since this is true for all $\xi \in V$ we conclude $df_p = 0$. $\square$

(8.17) Critical points and critical values. We introduce terminology which applies to functions on curved manifolds as well as in the flat situation we consider here.

Definition 8.18. Let V, W be normed linear spaces, A, B affine spaces over V, W , $U \subset A$ an open set, and $f: U \rightarrow B$ a differentiable function.

- (1) $p \in U$ is a *regular point* of f if $df_p: V \rightarrow W$ is surjective;
- (2) $p \in U$ is a *critical point* of f if it is not a regular point, i.e., if df_p fails to be surjective;
- (3) $b \in B$ is a *regular value* of f if each $p \in f^{-1}(b)$ is a regular point;
- (4) $b \in B$ is a *critical value* of f if it is not a regular value.

Notice the regular and critical *points* lie in the domain whereas the regular and critical *values* lie in the codomain. If $B = \mathbb{R}$ then df_p is surjective if and only if it is nonzero. Theorem 8.16 asserts that if a differentiable function has a local extremum at $p \in U$ then p is a critical point.

(8.19) Smooth curved shapes and regular values. Our next result gives the theoretical underpinning for Lagrange multipliers, which occur in a “constrained” extremum problem. The constraint means that we consider a function on a non-open subset of affine space, and we want to be able to differentiate, so we want that subset to be smooth in some sense. Later in the course we will develop some foundations for smooth manifolds, which are “smooth curved shapes”, but for now we proceed intuitively.

Let’s begin with a simple example. Consider the function $g: \mathbb{A}_{(x,y)}^2 \rightarrow \mathbb{R}$ defined by $g = x^2 - y^2$. The curves $g^{-1}(c)$, $c \in \mathbb{R}$, fill out the domain $\mathbb{A}^2$. For $c \neq 0$ these curves are smooth hyperbolas, but for $c = 0$ we obtain two intersecting lines, and at the intersection point we can observe intuitively that $g^{-1}(0)$ is not “smooth”. Check Definition 8.18 in this example: the only critical point of g is $(0, 0) \in \mathbb{A}^2$ and therefore the only critical value is $0 \in \mathbb{R}$. The inverse image is smooth for regular values. This is a general phenomenon, which we present here to motivate considering constraints only at regular values.

(8.20) Lagrange multipliers. Let V, W, X be normed linear spaces; A, B, C affine spaces over V, W, X ; $U \subset A$ an open set; $f: U \rightarrow B$ and $g: U \rightarrow C$ differentiable functions; and $c \in C$ a fixed element. For a constrained max-min problem we take $B = \mathbb{R}$ so that f is a real-valued function, and then we seek to extremize f on the subset $g^{-1}(c) \subset U$. If c is a regular value of g , then we have just motivated the idea that $g^{-1}(c) \subset U$ is smooth in some sense, and in particular has a tangent space at any $p \in g^{-1}(c)$ which is $\ker dg_p \subset V$. In that case we expect from Theorem 8.16 that if f has a local extremum on $g^{-1}(c)$ at p , then df_p vanishes on $\ker dg_p$. We cannot make a theorem out of that expectation until we develop some theory of curved smooth shapes. Instead, we prove the following theorem, which expresses this unproved criterion in terms of Lagrange multipliers.

Theorem 8.21. *Assume that either X is finite dimensional or V and X are Banach spaces. Then the following conditions on $p \in U$ are equivalent.*

- (i) *If $\xi \in V$ and $dg_p(\xi) = 0$ then $df_p(\xi) = 0$;*
- (ii) *There exists $\lambda \in \text{Hom}(X, W)$ such that*

$$(8.22) \quad df_p = \lambda \circ dg_p.$$

If f is real-valued, so $W = \mathbb{R}$, then $\lambda: X \rightarrow \mathbb{R}$ is a linear functional on X . If there is a finite set of independent constraints, so $C = \mathbb{A}^k$ for some $k \in \mathbb{Z}^{>0}$, then we can write $g = (g^1, \dots, g^k)$ and $\lambda = (\lambda_1, \dots, \lambda_k)$ for $\lambda_i \in \mathbb{R}$. In this case (8.22) reduces to the familiar equation

$$(8.23) \quad df_p = \lambda_1 dg_p^1 + \dots + \lambda_k dg_p^k.$$

Proof. The implication (ii) $\Rightarrow$ (i) is immediate. Assume (i) and let $K = \ker dg_p$. Then $K \subset V$ is a closed linear subspace. The hypothesis (8.22) implies that df_p factors through the quotient V/K .

So we obtain the following diagram

$$(8.24) \quad \begin{array}{ccc} V/K & \xrightarrow{\overline{df_p}} & W \\ & \searrow \cong & \\ & \overline{dg_p} & \searrow \\ & & X \end{array}$$

of linear maps. Since $K = \ker dg_p$ the map $\overline{dg_p}$ is a bijection, so it has a linear inverse. The hypotheses ensure that this inverse is continuous: if X is finite dimensional this is Theorem 5.11; and if V, X are Banach spaces, then so is V/K since K is closed, and now continuity follows from the open mapping theorem.¹² $\square$

Real inner product spaces

Let A be an affine space over a real vector space V . Geometry in A is the geometry of global parallelism, and as we have seen in Lecture 1 there are nontrivial theorems in that context. There is a large symmetry group, the group $\text{Aut}(A)$ of affine automorphisms of A , discussed in (1.12). Affine geometry is the study of properties of affine space invariant under affine automorphisms. We get a richer flat geometry—that is, a larger set of geometric concepts—if we have a smaller symmetry group. One way to cut down the symmetry group is to introduce a translation-invariant geometric structure on A , which amounts to a geometric structure on V transported around A by translation. We introduce the most common such structure here—an inner product on V —which induces a norm as well as a notion of angle between nonzero vectors. An affine space over an inner product space is called a Euclidean space, and in Euclidean geometry we have distance between points and angles between intersecting lines. There are derived notions of area and volume as well, and as you know from your first encounter with geometry the Euclidean context offers a much richer story than does affine geometry.

(8.25) *Geometric structures on affine space.*

Definition 8.26. Let V be a real vector space. An *inner product* on V is a function

$$(8.27) \quad \langle -, - \rangle: V \times V \rightarrow \mathbb{R}$$

such that for all $\xi, \xi_1, \xi_2, \xi_3 \in V$ and $c \in \mathbb{R}$ we have

- (i) $\langle \xi_1 + c\xi_2, \xi_3 \rangle = \langle \xi_1, \xi_3 \rangle + c\langle \xi_2, \xi_3 \rangle$;
- (ii) $\langle \xi_1, \xi_2 \rangle = \langle \xi_2, \xi_1 \rangle$; and
- (iii) if $\xi \neq 0$ then $\langle \xi, \xi \rangle > 0$.

Theorem 8.28 (Cauchy-Schwarz inequality). *Let V be an inner product space and $\xi_1, \xi_2 \in V$. Then we have*

$$(8.29) \quad |\langle \xi_1, \xi_2 \rangle| \leq \sqrt{\langle \xi_1, \xi_1 \rangle} \sqrt{\langle \xi_2, \xi_2 \rangle}.$$

¹²I don't expect that you have seen that theorem, but included that case for completeness.

Proof. Define the quadratic function $q: \mathbb{R} \rightarrow \mathbb{R}$ by

$$(8.30) \quad q(t) = \langle \xi_1 + t\xi_2, \xi_1 + t\xi_2 \rangle = \langle \xi_2, \xi_2 \rangle t^2 + 2\langle \xi_1, \xi_2 \rangle t + \langle \xi_1, \xi_1 \rangle.$$

Since $q(t) \geq 0$ the polynomial $q = at^2 + bt + c$ has at most one real root. Therefore its discriminant $b^2 - 4ac$ is nonpositive, which is equivalent to (8.29). $\square$

The derived concepts of length and angle follow.

Corollary 8.31.

- (1) The function $\xi \mapsto \sqrt{\langle \xi, \xi \rangle}$ is a norm on V ; the value on ξ is denoted ' $\|\xi\|$ ' as usual.
- (2) If ξ_1, ξ_2 are nonzero, then the formula

$$(8.32) \quad \cos \theta = \frac{\langle \xi_1, \xi_2 \rangle}{\|\xi_1\| \|\xi_2\|}$$

determines θ up to $\theta \mapsto \theta + 2\pi$ and $\theta \mapsto -\theta$.

Definition 8.33. Let V be a real inner product space. If the associated norm is complete, then we call V a *real Hilbert space*.

Any finite dimensional inner product space is complete.

Lecture 9: Inner product spaces; gradients

More on inner product spaces

(9.1) Remark on Cauchy-Schwarz inequality. The statement of Theorem 8.28 should include conditions for equality in (8.29), which occurs if $\xi_1 = 0$ or $\xi_2 = 0$ or ξ_1 is proportional to ξ_2 .

Example 9.2. Let $V = C^0([0, 1], \mathbb{R})$ be the vector space of continuous functions $f: [0, 1] \rightarrow \mathbb{R}$. The formula

$$(9.3) \quad \langle f_1, f_2 \rangle = \int_0^1 dx f_1(x) f_2(x)$$

defines an inner product, as is easy to verify. As an example of the Cauchy-Schwarz inequality, let $\underline{1}$ denote the constant function with value 1 and let $f \in V$ be arbitrary. Then

$$(9.4) \quad \langle \underline{1}, f \rangle^2 \leq \|\underline{1}\|^2 \|f\|^2$$

is the inequality

$$(9.5) \quad \left[\int_0^1 dx f(x) \right]^2 \leq \int_0^1 dx f(x)^2,$$

and equality holds if and only if f is a constant function. This is useful for Problem #8 on Homework #4.

Proposition 9.6 (parallelogram law). *Let V be a real inner product space and $\xi_1, \xi_2 \in V$. Then*

$$(9.7) \quad \|\xi_1 + \xi_2\|^2 + \|\xi_1 - \xi_2\|^2 = 2\|\xi_1\|^2 + 2\|\xi_2\|^2.$$

This says that in a parallelogram the sum of the squares of the diagonals equals the sum of the squares of the sides.

Proof.

$$(9.8) \quad \begin{aligned} \|\xi_1 + \xi_2\|^2 + \|\xi_1 - \xi_2\|^2 &= \|\xi_1\|^2 + 2\langle \xi_1, \xi_2 \rangle + \|\xi_2\|^2 + \|\xi_1\|^2 - 2\langle \xi_1, \xi_2 \rangle + \|\xi_2\|^2 \\ &= 2\|\xi_1\|^2 + 2\|\xi_2\|^2. \end{aligned}$$

□

Remark 9.9. Equation (9.7) can be formulated in any normed linear space, but it only holds if the norm comes from an inner product.

(9.10) Inner products and dual spaces. Let V be a real inner product space. The inner product provides a continuous linear map

$$(9.11) \quad \begin{aligned} \theta: V &\longrightarrow V^* \\ \xi &\longmapsto (\eta \mapsto \langle \xi, \eta \rangle). \end{aligned}$$

Continuity follows from the Cauchy-Schwarz inequality: $|\theta(\xi)(\eta)| \leq \|\xi\| \|\eta\|$ for all η , from which we may deduce $\|\theta(\xi)\| = \|\xi\|$ by setting $\eta = \xi$, and finally $\|\theta\| = 1$.

Lemma 9.12. θ is injective

Proof. If $\theta(\xi) = 0$, then $0 = \theta(\xi)(\xi) = \langle \xi, \xi \rangle$, from which $\xi = 0$. □

If V is finite dimensional, then it follows that θ is an isomorphism. That also holds in the infinite dimensional case if V is complete.

Theorem 9.13 (Riesz). *If V is a real Hilbert space, then θ is an isomorphism.*

I leave you to work out the proof as an extended exercise on the homework.

If θ is an isomorphism, we use it to transport the inner product on V to an inner product on V^* . Namely, define

$$(9.14) \quad \langle \alpha_1, \alpha_2 \rangle_{V^*} = \langle \theta^{-1}(\alpha_1), \theta^{-1}(\alpha_2) \rangle_V, \quad \alpha_1, \alpha_2 \in V^*.$$

(9.15) Formulas. Suppose V is finite dimensional and $e_1, \dots, e_n$ is a basis of V . Define

$$(9.16) \quad g_{ij} = \langle e_i, e_j \rangle, \quad i, j = 1, \dots, n.$$

Note that $g_{ii} > 0$ and $g_{ij} = g_{ji}$.

Definition 9.17. We say $e_1, \dots, e_n$ is an *orthonormal basis* of V if

$$(9.18) \quad g_{ij} = \begin{cases} 1, & i = j; \\ 0, & \text{otherwise.} \end{cases}$$

Equivalently, an orthonormal basis is an isometry $\mathbb{R}^n \rightarrow V$, where $\mathbb{R}^n$ has its standard inner product.

Let $e^1, \dots, e^n$ be the dual basis of V^* :

$$(9.19) \quad e^i(e_j) = \delta_j^i = \begin{cases} 1, & i = j; \\ 0, & \text{otherwise.} \end{cases}$$

Then $\theta(e_j)(e_i) = g_{ji}$ from which $\theta(e_j) = g_{ji}e^i$. In other words, (g_{ji}) is the matrix of θ relative to the dual bases of V and V^* . For the transported inner product (9.14) on the dual we define

$$(9.20) \quad g^{ij} = \langle e^i, e^j \rangle.$$

The reader can check that (g^{ij}) is the matrix representing θ^{-1} , and also $g^{ij}g_{jk} = \delta_k^i$.

Gradient

(9.21) Definition. Let V be a real Hilbert space (finite or infinite dimensional) and A an affine space over V . We call an affine space over an inner product a *Euclidean space*. Let $U \subset A$ be an open subset and $f: U \rightarrow \mathbb{R}$ a differentiable function. Thus for each $p \in U$ the differential is a continuous linear map $df_p \in V^*$ whose value on $\xi \in V$ is the directional derivative $\xi f(p)$; see Theorem 6.10. Use the isomorphism θ in (9.11) to produce a unique vector $\nabla f_p \in V$ such that

$$(9.22) \quad \xi f(p) = df_p(\xi) = \langle \nabla f_p, \xi \rangle, \quad \xi \in V.$$

Definition 9.23. $\nabla f_p \in V$ is the *gradient* of f at p .

In the infinite dimensional case the existence relies on completeness (Theorem 9.13), though the gradient may exist even if V is not complete. In any case the gradient, if it exists, is unique. The gradient is equivalent information to the differential. The differential as a functional, rather than a vector, is more primitive; we use the notion of angle and well as length to define the gradient vector. If we let p vary, then the gradient is a vector *field* $\nabla f: U \rightarrow V$.

(9.24) Geometric meaning. Let $S(V) = \{\xi \in V : \|\xi\| = 1\}$ be the unit sphere in V .

Proposition 9.25. In the situation of (9.21) assume $\nabla f_p \neq 0$ at some $p \in U$. Then the function

$$(9.26) \quad \begin{aligned} F: S(V) &\longrightarrow \mathbb{R} \\ \xi &\longmapsto \xi f(p) \end{aligned}$$

has a unique maximum at $\nabla f_p / \|\nabla f_p\|$ and a unique minimum at $-\nabla f_p / \|\nabla f_p\|$.

In other words, the gradient points in the direction of maximal increase of f and its magnitude is the rate of increase; the negative gradient points in the direction of maximal decrease of f and its magnitude is the rate of decrease.

Proof. The assertions follow immediately from the Cauchy-Schwarz inequality

$$(9.27) \quad |F(\xi)| \leq \|\nabla f_p\|$$

and the condition (9.1) that equality hold if and only if ξ is proportional to ∇f_p .

We can also give an argument using Lagrange multipliers (Theorem 8.21), where we extremize F under the constraint that $\xi \in g^{-1}(1/2)$, where

$$(9.28) \quad \begin{aligned} g: V &\longrightarrow \mathbb{R} \\ \xi &\longmapsto \frac{1}{2} \langle \xi, \xi \rangle. \end{aligned}$$

Then for $\dot{\xi} \in V$ we have

$$(9.29) \quad \begin{aligned} dF_{\xi}(\dot{\xi}) &= \langle \nabla f_p, \dot{\xi} \rangle \\ dg_{\xi}(\dot{\xi}) &= \langle \xi, \dot{\xi} \rangle. \end{aligned}$$

The Lagrange condition for an extreme point is then $\nabla f_p = \lambda \xi$ for some $\lambda \in \mathbb{R}$. □

(9.30) *Differential and gradient in affine coordinates.* Suppose now A is finite dimensional and $x^1, \dots, x^n: A \rightarrow \mathbb{R}$ form an affine coordinate system on the Euclidean space A . Recall that the differentials $dx^1, \dots, dx^n$ form a basis of V^* and that the dual basis of V is denoted $\partial/\partial x^1, \dots, \partial/\partial x^n$. (Since x^i is an affine function, its differential dx_p^i is independent of p .) Since

$$(9.31) \quad df \left(\frac{\partial}{\partial x^i} \right) = \frac{\partial}{\partial x^i} f = \frac{\partial f}{\partial x^i},$$

we have

$$(9.32) \quad df = \frac{\partial f}{\partial x^i} dx^i.$$

Following (9.16) and (9.20) define

$$(9.33) \quad \begin{aligned} g_{ij} &= \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle \\ g^{ij} &= \langle dx^i, dx^j \rangle. \end{aligned}$$

Then I leave you to deduce

$$(9.34) \quad \nabla f = g^{ij} \frac{\partial f}{\partial x^j} \frac{\partial}{\partial x^i}.$$

Lecture 10: Differential of the length function; contraction mappings

The first variation formula

(10.1) *Introduction.* The first topic in this lecture is a worked out example differentiating a function on an affine space A over an infinite dimensional vector space X . This particular function f is the length of a parametrized curve in a Euclidean space. We first compute its directional derivatives, which as always is a one-variable computation. Their existence does not guarantee differentiability; note that Theorem 8.6 is proved only for finite dimensional domains. Whereas directional derivatives are independent of a norm on X , differentiability depends on the norm: for some norms f is differentiable and for others not. The formula for the differential is called the *first variation formula*, usually in the context of Riemannian geometry where it is derived for a curved smooth manifold. Here we consider paths in a flat Euclidean space. In the course of our computations and estimates we use some results from one-variable analysis, such as Taylor's theorem, exchange of limits, and properties of the Riemann integral. Hopefully this example serves to illustrate how such theorems are used in practice.

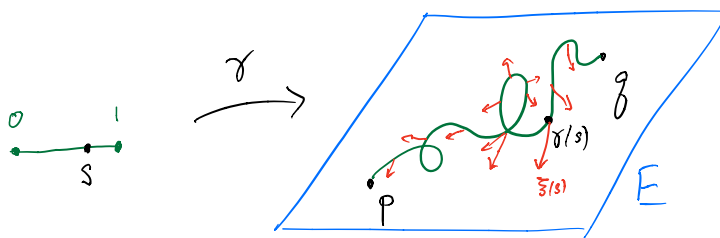


FIGURE 15. A parametrized curve in Euclidean space

(10.2) *Setup.* Let V be a finite dimensional real inner product space and E an affine space over V , a Euclidean space. Fix points $p, q \in E$. Define

$$(10.3) \quad A = \{ \gamma: [0, 1] \longrightarrow E \text{ such that } \gamma(0) = p, \gamma(1) = q, \gamma \in C^2([0, 1], E) \}.$$

This is the space of twice continuously differentiable parametrized paths from p to q . It is an affine space over the vector space

$$(10.4) \quad X = \{ \xi: [0, 1] \longrightarrow V \text{ such that } \xi(0) = \xi(1) = 0, \xi \in C^2([0, 1], V) \}.$$

Define the subset

$$(10.5) \quad U = \{\gamma \in A \text{ such that } \dot{\gamma}(s) \neq 0 \text{ for all } s \in [0, 1]\}.$$

of parametrized curves with nonzero velocity at all times. The length function is

$$(10.6) \quad \begin{aligned} f: U &\longrightarrow \mathbb{R} \\ \gamma &\longmapsto \int_0^1 ds \|\dot{\gamma}(s)\| \end{aligned}$$

Remark 10.7. Notice that we have not yet introduced a norm on X , so no topology on A . Therefore, we cannot say that $U \subset A$ is open. We will introduce a norm on X below in (10.26), and indeed U is open in that norm topology. For now, we give some possible values for the norm of $\xi \in X$:

$$(10.8) \quad \max_{s \in [0,1]} \|\xi(s)\| \quad \text{or} \quad \max_{s \in [0,1]} \|\dot{\xi}(s)\| \quad \text{or} \quad \int_0^1 ds (\|\xi(s)\| + \|\ddot{\xi}(s)\|) \quad \text{or} \quad \dots$$

(10.9) Variations of γ . Fix $\gamma \in U$ and $\xi \in X$. Recall from the end of Lecture 2 that vectors in the tangent space X to A play dual roles: they label global automorphisms of A —translations—and they also label represent parallel vector fields on A . In this context we typically think of the latter, and we obtain $\xi \in X$ as a tangent vector to A at γ from a *variation* of γ , which is a function

$$(10.10) \quad \Gamma: (-a, a) \times [0, 1] \longrightarrow E$$

for some $a > 0$ such that $\Gamma(t, 0) = p$ and $\Gamma(t, 1) = q$ for all $t \in (-a, a)$. Then $\xi(s) = \frac{d}{dt} \big|_{t=0} \Gamma(t, s)$. Given ξ there is a canonical variation in affine space which uses ξ to translate, namely $\Gamma(t, s) = \gamma(s) + t\xi(s)$, and it is defined for all $t \in \mathbb{R}$.

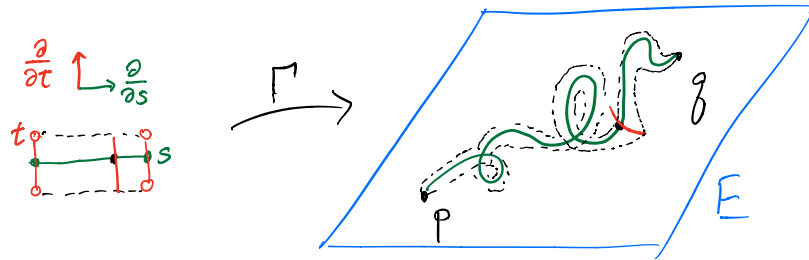


FIGURE 16. A variation of a path

(10.11) Directional derivative. The directional derivative of f at γ in the direction ξ is the derivative of a function of one variable; it does not require a norm on X . For readability we drop the argument ‘ s ’ of the functions γ, ξ and their derivatives.

$$\begin{aligned}
 \xi f(\gamma) &= \left. \frac{d}{dt} \right|_{t=0} f(\gamma + t\xi) \\
 &= \left. \frac{d}{dt} \right|_{t=0} \int_0^1 ds \langle \dot{\gamma} + t\dot{\xi}, \dot{\gamma} + t\dot{\xi} \rangle^{1/2} \\
 &= \int_0^1 ds \left. \frac{d}{dt} \right|_{t=0} \langle \dot{\gamma} + t\dot{\xi}, \dot{\gamma} + t\dot{\xi} \rangle^{1/2} \\
 (10.12) \quad &= \int_0^1 ds \frac{\langle \dot{\gamma}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}} \\
 &= \int_0^1 ds \frac{d}{ds} \frac{\langle \dot{\gamma}, \xi \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}} - \int_0^1 ds \left\langle \frac{d}{ds} \left(\frac{\dot{\gamma}}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}} \right), \xi \right\rangle \\
 &= - \int_0^1 ds \left\langle \frac{d}{ds} \left(\frac{\dot{\gamma}}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}} \right), \xi \right\rangle
 \end{aligned}$$

Several comments are in order. First, the length function (10.6) makes sense if V has any norm, not necessarily a norm associated to an inner product. However, a general norm is not necessarily differentiable: consider, for example, the absolute value as a norm on $\mathbb{R}$. Next, the passage from line 2 to line 3 is the interchange of two limits, the derivative in t and the integral over s . That interchange is justified in general when one limit is *uniform* over the other, and often compactness is used to prove uniformity. In this case the derivative in t is uniform over $s \in [0, 1]$. Finally, the Leibniz rule in the penultimate line is used to execute integration by parts, a powerful technique. We will use both the expression in the fourth line and the final expression. Notice that each is a linear function of ξ .

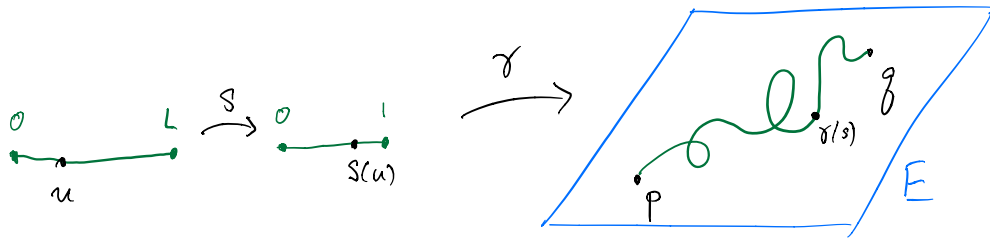


FIGURE 17. Changing the clock

(10.13) Reparametrization. The appearance of $\dot{\gamma}/\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}$ in (10.12) may trigger the thought that the length of a parametrized curve is invariant under reparametrization: if we run a certain distance that distance does not depend on our speed as long as we are always traveling in the same direction (guaranteed if the speed is nonzero). Reparametrization is a function

$$(10.14) \quad s: [0, L] \longrightarrow [0, 1]$$

which we require to be a C^2 monotone increasing bijection. The chain rule implies

$$(10.15) \quad \frac{d}{du} \gamma(s(u)) = \frac{ds}{du} \frac{d\gamma}{ds},$$

and the condition that this be a unit norm vector for all times on the u -clock is

$$(10.16) \quad \frac{ds}{du} = \frac{1}{\|\dot{\gamma}(s)\|}.$$

We will prove a theorem which implies that (10.16) has a unique solution $s = s(u)$ such that $s(0) = 0$. Then the u -clock is a *unit speed parametrization* of γ . If we assume that the s -clock is a unit speed parametrization (in which case its domain should be $[0, L]$ for some L not necessarily equal to one), then the first factor of the integrand in the last expression of (10.12) is the acceleration.

(10.17) Extremum. Recall that if a function f has an extremum at γ then all directional derivatives of f at γ vanish. The following lemma, which sometimes has the name du Bois-Reymond attached, applies.

Lemma 10.18. *Let $\eta: [0, 1] \rightarrow V$ be a continuous function and suppose that for all continuous $\xi: [0, 1] \rightarrow V$ of compact support we have*

$$(10.19) \quad \int_0^1 ds \langle \eta(s), \xi(s) \rangle = 0.$$

Then $\eta(s) = 0$ for all $s \in [0, 1]$.

Proof. If $\eta(s_0) \neq 0$ for some s_0 , let $f: [0, 1] \rightarrow \mathbb{R}$ be a nonnegative compactly supported function such that $f(s_0) > 0$. Then for $\xi = f\eta$ the integral in (10.19) is positive, which contradicts the hypothesis. $\square$

In our situation, if we assume a unit parametrization and use the last expression in (10.12), then we conclude that if all directional derivatives vanish at γ that γ has zero acceleration, i.e., it is a constant velocity motion. That means the path it traces out is an affine line segment. We have not proved that this straight line path minimizes length; the directional derivative condition is necessary, not sufficient. We will revisit this issue after studying higher differentials.

(10.20) Differentiability of the length function. Theorem 6.10 and the computation (10.12) imply that if f is differentiable at γ with respect to some norm on X , then the differential must be the linear function

$$(10.21) \quad T(\xi) = \int_0^1 ds \frac{\langle \dot{\gamma}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}}.$$

To prove differentiability we must estimate the quantity

$$\begin{aligned}
 (10.22) \quad Q &= |f(\gamma + \xi) - f(\gamma) - T(\xi)| \\
 &= \left| \int_0^1 ds \left\{ \langle \dot{\gamma} + \dot{\xi}, \dot{\gamma} + \dot{\xi} \rangle^{1/2} - \langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2} - \int_0^1 ds \frac{\langle \dot{\gamma}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}} \right\} \right| \\
 &= \left| \int_0^1 ds \left\{ \langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2} \left[1 + \frac{2\langle \dot{\gamma}, \dot{\xi} \rangle + \langle \dot{\xi}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle} \right]^{1/2} - \langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2} - \int_0^1 ds \frac{\langle \dot{\gamma}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}} \right\} \right|
 \end{aligned}$$

for all $\xi \in X$. To estimate this we use Taylor's theorem to see that if $|x|$ is sufficiently small for $x \in \mathbb{R}$, then there exists a constant M such that

$$(10.23) \quad \left| (1+x)^{1/2} - 1 - \frac{x}{2} \right| \leq Mx^2.$$

Apply (10.23) and the triangle inequality to (10.22) with

$$(10.24) \quad x = \frac{2\langle \dot{\gamma}, \dot{\xi} \rangle + \langle \dot{\xi}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle}$$

to conclude

$$(10.25) \quad Q \leq \int_0^1 ds \left\{ M \frac{[2\langle \dot{\gamma}, \dot{\xi} \rangle + \langle \dot{\xi}, \dot{\xi} \rangle]^2}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{3/2}} + \frac{1}{2} \frac{\langle \dot{\xi}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}} \right\}.$$

Since $\|\dot{\gamma}\|: [0, 1] \rightarrow \mathbb{R}^{>0}$ is continuous and $[0, 1]$ is compact, there exist constants $C, c > 0$ such that $c \leq \|\dot{\gamma}(s)\| \leq C$ for all $s \in [0, 1]$. Now it is clear that a convenient norm on X is

$$(10.26) \quad \|\xi\|_X := \max_{s \in [0, 1]} \|\dot{\xi}(s)\|_V.$$

Then if $\|\xi\|_X < \delta$ for some $\delta > 0$, then we apply the Cauchy-Schwarz and triangle inequalities to (10.25) to estimate

$$\begin{aligned}
 (10.27) \quad Q &\leq \left\{ \frac{M(2C\|\xi\| + \delta\|\xi\|)^2}{c^3} + \frac{\delta\|\xi\|}{2c} \right\} \\
 &\leq K\delta\|\xi\|
 \end{aligned}$$

for some $K > 0$.

Theorem 10.28. *The function f is differentiable at $\gamma \in U$ with differential (10.21).*

Proof. Given $\epsilon > 0$ choose $\delta = \epsilon/K$. Then if $\|\xi\|_X < \delta$ the estimates above show that

$$(10.29) \quad |f(\gamma + \xi) - f(\gamma) - T(\xi)| \leq \epsilon\|\xi\|_X. \quad \square$$

Fixed points and contractions

(10.30) Introduction. Our next big goal is the inverse function theorem, which we prove in the next lecture. It says, roughly, that if the differential of a function is invertible, then the function is locally invertible. The proof is constructive in the sense that we construct a local inverse. The technique for doing so is widely used. Namely we set up an iterative procedure which better and better approximates the inverse, so that iterating infinitely often we converge to the inverse. In this lecture we illustrate such iterative procedures in simple cases and then prove the general *contraction mapping fixed point theorem* which applies in many situations. We will also use it to construct solutions to ordinary differential equations.

Example 10.31. Problem: Find a positive real number x so that $e^{-x} = x$. Define

$$(10.32) \quad \begin{aligned} \phi: (0, \infty) &\longrightarrow (0, \infty) \\ x &\longmapsto e^{-x} \end{aligned}$$

Then we seek a fixed point of ϕ , i.e., a solution to $\phi(x) = x$. As an initial guess put $x_0 = 1$. Then $x_1 = \phi(x_0) = e^{-1}$ is not equal to x_0 . So we iterate:

$$(10.33) \quad \begin{aligned} x_2 &= \phi(x_1) = e^{-e^{-1}} \\ x_3 &= \phi(x_2) = e^{-e^{-e^{-1}}} \\ x_4 &= \phi(x_3) = e^{-e^{-e^{-e^{-1}}}} \\ x_5 &= \phi(x_4) = e^{-e^{-e^{-e^{-e^{-1}}}}} \\ x_6 &= \phi(x_5) = e^{-e^{-e^{-e^{-e^{-e^{-1}}}}}} \\ &\vdots \end{aligned}$$

If we imagine going on infinitely, then $x = e^{-[e^{\cdot} \cdot]}$ satisfies $x = e^{-x}$. Iterating the map ϕ shrinks the space $X = (0, \infty)$, as we see in Figure 18.

Example 10.34. Problem: Compute a good approximation to $\sqrt{2} = 1.41421356237\dots$. We posit an initial reasonable approximation, such as $x_0 = 1.4$, and then $x = x_0 + \xi$ is a square root of 2 iff x is a fixed point of the function

$$(10.35) \quad \phi(\xi) = \frac{2 - x_0^2 - \xi^2}{2x_0},$$

defined on the whole real line. We set up an iteration in ξ with seed $x_0 = 0$ and for $n \in \mathbb{Z}^{\geq 0}$ define $\xi_{n+1} = \phi(\xi)$. Then setting $x_n = x_0 + \xi_n$ we compute the first few iterations as

$$(10.36) \quad \begin{array}{ll} \xi_1 = 0.014285700 & x_1 = 1.414285700 \\ \xi_2 = 0.014212827 & x_2 = 1.414212827 \\ \xi_3 = 0.014213569 & x_3 = 1.414213569 \end{array}$$

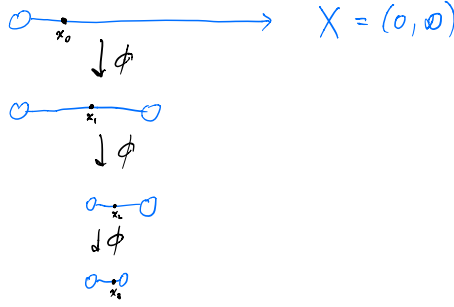


FIGURE 18. Iterating a contraction mapping

(10.37) Contraction fixed point theorem. The abstract setting for these iterations is a complete metric space. Completeness is used to produce a limit of the iteration.

Theorem 10.38. Let (X, d) be a complete metric space and $\phi: X \rightarrow X$ a function. Suppose there exists $C \in \mathbb{R}$ with $0 < C < 1$ such that

$$(10.39) \quad d(\phi(x_1), \phi(x_2)) \leq C d(x_1, x_2)$$

for all $x_1, x_2 \in X$. Then there exists a unique $x \in X$ such that $\phi(x) = x$.

The inequality (10.39) implies that ϕ is continuous, even uniformly continuous, and this stronger form of continuity is called *Lipschitz* continuity. If the Lipschitz constant C satisfies $C < 1$ as here, then ϕ is called a *contraction*. Thus the name ‘contraction mapping fixed point theorem’ for Theorem 10.38. I want to emphasize that the existence proof below is constructive: a simple iterative procedure produces the fixed point. Note too that the “seed” to the iteration is an arbitrary point of X .

Proof. The uniqueness is immediate from (10.39) applied to any two fixed points. Let $x_0 \in X$ and define the sequence $(x_n)_{n=0,1,\dots} \subset X$ by

$$(10.40) \quad x_{n+1} = \phi(x_n).$$

We claim that (x_n) is Cauchy. For if $n > m \geq 0$, then

$$\begin{aligned} d(x_m, x_n) &= d(\phi^m(x_0), \phi^n(x_0)) \\ &\leq C^m d(x_0, \phi^{n-m}(x_0)) \\ &\leq C^m (d(x_0, \phi(x_0)) + d(\phi(x_0), \phi^2(x_0)) + \cdots + d(\phi^{n-m-1}(x_0), \phi^{n-m}(x_0))) \\ &\leq C^m (\delta + C\delta + \cdots + C^{n-m-1}\delta) \\ &\leq C^m \frac{\delta}{1-C}, \end{aligned} \quad (10.41)$$

where $\delta = d(x_0, \phi(x_0))$. Since $C^m \rightarrow 0$ as $m \rightarrow \infty$ this proves the claim. By completeness there exists $x \in X$ such that $\lim_{n \rightarrow \infty} x_n = x$. Since ϕ is continuous, take $n \rightarrow \infty$ in (10.40) to deduce $x = \phi(x)$. $\square$

Lecture 11: Inverse function theorem

Invertibility is an open condition

(11.1) Finite dimensions. The slogan in the title is an important one, and we will first prove it in finite dimensions, where we can take advantage of the determinant of a square matrix. Let $M_n\mathbb{R}$ denote the vector space of $n \times n$ real matrices. It has the usual vector space topology (Theorem 4.20). Recall that $GL_n\mathbb{R} \subset M_n\mathbb{R}$ is the subset of invertible matrices; it forms a group under matrix multiplication.

Theorem 11.2. $GL_n\mathbb{R} \subset M_n\mathbb{R}$ is open

Proof. The determinant function

$$(11.3) \quad \det: M_n\mathbb{R} \longrightarrow \mathbb{R}$$

$$(A_j^i) \longmapsto \sum_{\sigma \in \text{Sym}_n} (-1)^{|\sigma|} A_{\sigma(1)}^1 A_{\sigma(2)}^2 \cdots A_{\sigma(n)}^n$$

is a degree n homogeneous polynomial in the entries of the matrix, and as such is a continuous function. (Notation: Sym_n is the group of permutations and $|\sigma| \in \{0, 1\}$ is the sign of the permutation σ .) Let $\mathbb{R}^\times \subset \mathbb{R}$ be the open subset of nonzero real numbers. Then by the continuity of $\det$, $GL_n\mathbb{R} = \det^{-1}(\mathbb{R}^\times)$ is also open. $\square$

It follows that for any two finite dimensional real vector spaces V, W of the same dimension, the subset $\text{Iso}(V, W) \subset \text{Hom}(V, W)$ of invertible linear maps is an open subset.

(11.4) Infinite dimensions. We now extend to infinite dimensions while giving an alternative argument in finite dimensions.

Definition 11.5. Let V, W be normed linear spaces. We say a continuous linear map $T: V \rightarrow W$ is *invertible* if there exists $T' \in \text{Hom}(W, V)$ such that $T'T = \text{id}_V$ and $TT' = \text{id}_W$. In that case we write $T' = T^{-1}$. Denote the subset of invertible maps as $\text{Iso}(V, W) \subset \text{Hom}(V, W)$.

A normed linear space is the marriage of a vector space and a topological space, and this is the natural notion of isomorphism in that union. So T is both an algebraic linear isomorphism and a homeomorphism of topological spaces.

Remark 11.6. If V, W are complete, i.e., are Banach spaces, then an algebraic inverse is always continuous. This is a corollary of the *open mapping theorem*, which asserts that a surjective continuous linear map $T: V \rightarrow W$ is open, i.e., $T(U) \subset W$ is open if $U \subset V$ is open.

Theorem 11.7. *Let V, W be Banach spaces. Then $\text{Iso}(V, W) \subset \text{Hom}(V, W)$ is open. More precisely, if $T \in \text{Iso}(V, W)$ then $B_r(T) \subset \text{Iso}(V, W)$ for $r = \|T^{-1}\|^{-1}$.*

We use the usual metric space notation $B_r(T)$ for the open ball of radius r about T .

Proof. Suppose $A \in \text{Hom}(V, W)$ with $\|A\| < r$. We must show that $T + A$ is invertible. Set

$$(11.8) \quad S_N = \sum_{n=0}^N (-1)^n (T^{-1}A)^n, \quad N \in \mathbb{Z}^{\geq 0}.$$

Then (S_N) is a sequence in $\text{Hom}(V, W)$ and setting $\delta = \|T^{-1}\|\|A\| < 1$ we estimate

$$(11.9) \quad \left\| \sum_{n=M}^N (-1)^n (T^{-1}A)^n \right\| \leq \sum_{n=M}^N (\|T^{-1}\|\|A\|)^n \leq \frac{\delta^M}{1-\delta}.$$

It follows that (S_N) is Cauchy, and since $\text{Hom}(V, W)$ is complete that $S_N \rightarrow S$ for a unique $S \in \text{Hom}(V, W)$. We claim that ST^{-1} is the inverse of $T + A$. First, observe that ST^{-1} is the composition of continuous maps, so is continuous. Now compute

$$(11.10) \quad \begin{aligned} S_N T^{-1} (T + A) &= \sum_{n=0}^N (-1)^n (T^{-1}A)^n T^{-1} (T + A) \\ &= \sum_{n=0}^N (-1)^n [(T^{-1}A)^n + (T^{-1}A)^{n+1}] \\ &= \text{id}_V + (-1)^N (T^{-1}A)^{N+1} \end{aligned}$$

using the telescoping of the sum in the middle line. Since $\|(T^{-1}A)^{N+1}\| \leq \delta^{N+1} \rightarrow 0$ as $N \rightarrow \infty$, we conclude $(ST^{-1})(T + A) = \text{id}_V$. A similar argument proves $(T + A)(ST^{-1}) = \text{id}_W$. $\square$

C^k functions

(11.11) Higher differentials. Assume the standard data (5.23). Assuming the function f is differentiable, its differential is a function

$$(11.12) \quad df: U \longrightarrow \text{Hom}(V, W).$$

This is another instance of the standard data with U, A, V as before but now $B = W = \text{Hom}(V, W)$. Hence we can ask if df is differentiable, in which case its differential is a function

$$(11.13) \quad d^2 f = d(df): U \longrightarrow \text{Hom}(V, \text{Hom}(V, W)).$$

We can clearly iterate k times for any $k \in \mathbb{Z}^{>0}$.

Definition 11.14. Assume the standard data.

- (1) f is C^1 if it is differentiable and the differential (11.12) is continuous;
- (2) For $k \in \mathbb{Z}^{>0}$ we say f is C^k if it is C^{k-1} and $d^{k-1}f$ is C^1 .
- (3) f is C^∞ if f is C^k for all positive integers k .

In (2) we implicitly use mathematical induction.

Remark 11.15. We often say f is *smooth* as a shorthand for saying f is C^k , but the precise value of k depends on the context. In many instances, and this tends to be my usage, ‘ f is smooth’ is synonymous with ‘ f is C^∞ ’.

(11.16) *Inversion is smooth.* We will not prove the following theorem in these notes; I asked you to prove the C^1 version on homework.

Theorem 11.17. *Let V, W be Banach spaces. Then the inversion map*

$$(11.18) \quad \begin{aligned} \iota: \operatorname{Iso}(V, W) &\longrightarrow \operatorname{Iso}(W, V) \\ T &\longmapsto T^{-1} \end{aligned}$$

is C^∞ .

In Theorem 11.22 we only use continuity of ι , which is fairly straightforward to prove by estimating $\|S_N T^{-1} - T^{-1}\|$ in terms of $\|A\|$. In Corollary 11.35 we use the full statement of Theorem 11.17.

Remark 11.19. As shown in the proof of Theorem 11.7 inversion is an *analytic* function: it is locally defined by a convergent power series. In general, analytic functions are C^∞ .

Inverse function theorem

(11.20) *Global, local, and infinitesimal.* In geometry it is important to keep track of the scope of one’s considerations. In most contexts there is a global vs. local distinction. ‘Global’ refers to the entire space whereas ‘local’ refers to a neighborhood of a point. For example, the topological spaces S^1 and $\mathbb{A}^1$ are locally homeomorphic but not globally homeomorphic. By contrast $\mathbb{A}^1$ and $\mathbb{A}^2$ are not locally homeomorphic, although that is a deep theorem. In “smooth” contexts one has an even smaller scope: infinitesimal. This is tangential information which depends on only derivative information at one point. (The number of derivatives can be arbitrary, even infinite. We speak of the k^{th} order infinitesimal neighborhood of a point if we use only $\leq k$ derivatives.)

(11.21) *Inverse function theorem.* The following basic theorem passes from infinitesimal information to local information. In this lecture we state it and give a proof. In the next lecture we will explain some basic applications.

Theorem 11.22. *Let A, B be affine spaces modeled on Banach spaces V, W . Let $U \subset A$ be an open set and $f: U \rightarrow B$ a C^1 function. Suppose for some $p \in U$ that $df_p: V \rightarrow W$ is invertible. Then there exist open sets $U' \subset U$, $V' \subset B$ and a function $g: V' \rightarrow U'$ which is inverse to the restriction of f to U' . Furthermore, g is C^1 and $dg_{f(p)} = df_p^{-1}$.*

Notice that the last equation in the theorem follows immediately by differentiating $g \circ f = \text{id}_{U'}$ and using the chain rule; it holds at all points in U' . We say that f is a C^1 local diffeomorphism at p . In Corollary 11.35 below we prove a version for C^k functions.

The proof proceeds by first translating and composing with the inverse of df_p to replace f by a nonlinear mapping on an open neighborhood of the origin in V . Next, a simple device turns solving the inverse problem for f into a fixed point problem, which we show has a unique solution using the contraction mapping fixed point theorem (10.37). Finally, we prove that the inverse is continuously differentiable.

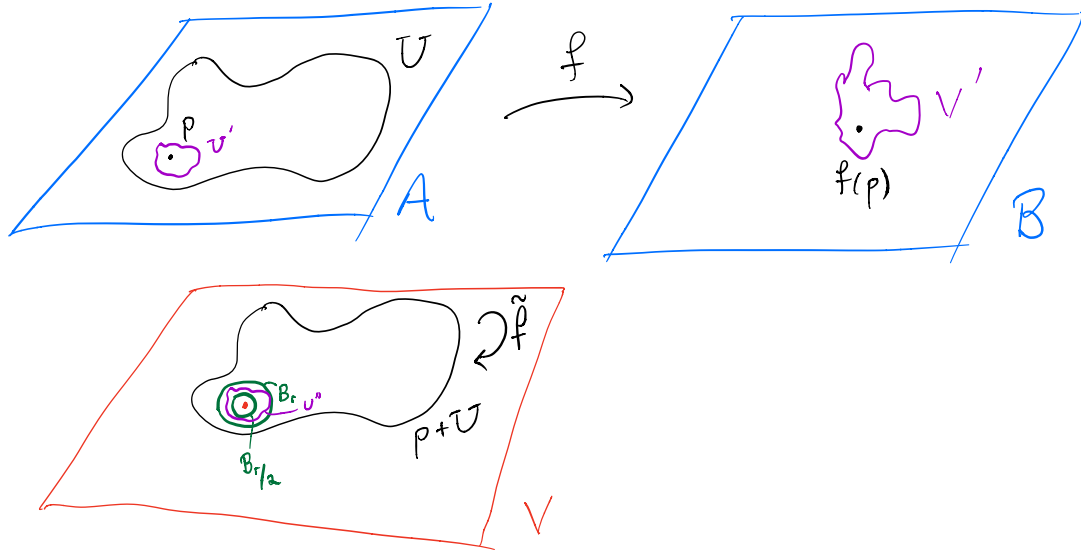


FIGURE 19. Open sets in proof of inverse function theorem

Proof. Define

$$(11.23) \quad \tilde{f}(\xi) = df_p^{-1}(f(p + \xi) - f(p))$$

where $\xi \in \{\xi \in V : p + \xi \in U\}$. The latter is an open neighborhood of $0 \in V$, and $\tilde{f}$ maps this neighborhood into another open neighborhood of $0 \in V$. Set

$$(11.24) \quad \phi(\xi) = \xi - \tilde{f}(\xi).$$

Then $\phi(0) = 0$ and $d\phi_0 = 0$. Since $d\phi$ is continuous we can choose $r > 0$ such that $\|d\phi_\xi\| < 1/2$ for $\xi \in \overline{B}_r$, where $\overline{B}_r$ is the closed ball of radius r about $0 \in V$. Then a corollary¹³ of the mean

¹³From Homework #4: Let A be an affine space over a normed linear space V . A set $U \subset A$ is *convex* if whenever $p_0, p_1 \in U$, then $tp_0 + (1-t)p_1 \in U$ for all $t \in [0, 1]$. Suppose $U \subset A$ is open and convex, B is an affine space

value inequality Theorem 7.22 implies that $\|\phi(\xi)\| \leq \frac{1}{2}\|\xi\|$ if $\|\xi\| \in \overline{B}_r$, i.e., $\phi(\overline{B}_r) \subset \overline{B}_{r/2}$. Observe that if $\xi \in \overline{B}_r$ then $d\tilde{f}_\xi$ is invertible (Theorem 11.7).

For $\eta \in \overline{B}_{r/2}$ define

$$(11.25) \quad \phi^\eta(\xi) = \eta + \xi - \tilde{f}(\xi) = \eta + \phi(\xi), \quad \xi \in \overline{B}_r.$$

Observe $\phi^\eta(\xi) = \xi$ if and only if $\tilde{f}(\xi) = \eta$. Now $\phi^\eta(\overline{B}_r) \subset \overline{B}_r$ since $\|\phi^\eta(\xi)\| \leq \|\eta\| + \|\phi(\xi)\| \leq r/2 + r/2 = r$. Also, for $\xi_1, \xi_2 \in \overline{B}_r$ we have from the same corollary of the mean value inequality that

$$(11.26) \quad \|\phi^\eta(\xi_2) - \phi^\eta(\xi_1)\| = \|\phi(\xi_2) - \phi(\xi_1)\| \leq \frac{1}{2}\|\xi_2 - \xi_1\|.$$

Thus ϕ^η is a contraction on $\overline{B}_r$, and as the latter is a complete metric space there is a unique fixed point $\xi \in \overline{B}_r$. The fixed point is the unique solution ξ to the equation $\tilde{f}(\xi) = \eta$ in $\overline{B}_r$. This unique solution defines a map $\tilde{g}: \overline{B}_{r/2} \rightarrow \overline{B}_r$. Set $U'' = B_r \cap \tilde{f}^{-1}(B_{r/2})$, where B_r is the *open* ball of radius r about $0 \in V$. Then the restriction of $\tilde{f}$ to U'' is inverse to the restriction of $\tilde{g}$ to $B_{r/2}$. To undo (11.23) we set $U' = p + U''$, $V' = f(p) + df_p(B_{r/2})$, and

$$(11.27) \quad g(q) = p + \tilde{g}(df_p^{-1}(q - f(p))).$$

It remains to prove that $\tilde{g}$, and therefore also g , is smooth.

We first show that $\tilde{g}$ is continuous. Suppose $\eta_1, \eta_2 \in B_{r/2}$ and set $\xi_i = \tilde{g}(\eta_i)$, $i = 1, 2$. Then since $\phi^\eta(\xi_i) = \xi_i$ we have from (11.25) that

$$(11.28) \quad \|\xi_2 - \xi_1\| \leq \|\eta_2 - \eta_1\| + \|\phi(\xi_2) - \phi(\xi_1)\| \leq \|\eta_2 - \eta_1\| + \frac{1}{2}\|\xi_2 - \xi_1\|,$$

from which

$$(11.29) \quad \|\tilde{g}(\eta_2) - \tilde{g}(\eta_1)\| = \|\xi_2 - \xi_1\| \leq 2\|\eta_2 - \eta_1\|.$$

This proves that $\tilde{g}$ is Lipschitz continuous with Lipschitz constant 2.

Next we show that $\tilde{g}$ is differentiable at $\eta_1 \in B_{r/2}$. If also $\eta_2 \in B_{r/2}$ and $\xi_i = \tilde{g}(\eta_i)$, then

$$(11.30) \quad \begin{aligned} \left\| \tilde{g}(\eta_2) - \tilde{g}(\eta_1) - d\tilde{f}_{\tilde{g}(\eta_1)}^{-1}(\eta_2 - \eta_1) \right\| &= \left\| \xi_2 - \xi_1 - d\tilde{f}_{\xi_1}^{-1}(\eta_2 - \eta_1) \right\| \\ &\leq \|d\tilde{f}_{\xi_1}^{-1}\| \left\| d\tilde{f}_{\xi_1}(\xi_2 - \xi_1) - (\eta_2 - \eta_1) \right\| \\ &= \|d\tilde{f}_{\xi_1}^{-1}\| \left\| \tilde{f}(\xi_2) - \tilde{f}(\xi_1) - d\tilde{f}_{\xi_1}(\xi_2 - \xi_1) \right\|. \end{aligned}$$

over a normed linear space W , and $f: U \rightarrow B$ is differentiable. Assume there exists $C \in \mathbb{R}^{>0}$ such that $\|df_p\| < C$ for all $p \in U$. Prove that if $p_0, p_1 \in U$, then

$$\|f(p_1) - f(p_0)\| \leq C\|\xi\|, \quad p_1 = p_0 + \xi.$$

We use Theorem 11.7 to guarantee the existence of $d\tilde{f}_{\tilde{g}(\eta_1)}^{-1}$. Since $\tilde{f}$ is differentiable at ξ_1 , given $\epsilon > 0$ there exists $\delta > 0$ such that if $\|\xi_2 - \xi_1\| < \delta$ then

$$(11.31) \quad \left\| \tilde{f}(\xi_2) - \tilde{f}(\xi_1) - d\tilde{f}_{\xi_1}(\xi_2 - \xi_1) \right\| \leq \frac{\epsilon}{2\|d\tilde{f}_{\xi_1}^{-1}\|} \|\xi_2 - \xi_1\|.$$

Then if $\|\eta_2 - \eta_1\| < \delta/2$, by (11.29) we have $\|\xi_2 - \xi_1\| < \delta$ and so

$$(11.32) \quad \left\| \tilde{g}(\eta_2) - \tilde{g}(\eta_1) - d\tilde{f}_{\tilde{g}(\eta_1)}^{-1}(\eta_2 - \eta_1) \right\| \leq \frac{\epsilon}{2} \|\xi_2 - \xi_1\| \leq \epsilon \|\eta_2 - \eta_1\|,$$

which proves that $\tilde{g}$ is differentiable with

$$(11.33) \quad d\tilde{g}_\eta = d\tilde{f}_{\tilde{g}(\eta)}^{-1}.$$

To prove that $\tilde{g}$ is C^1 we must show that $\eta \mapsto d\tilde{g}_\eta$ is continuous. For that write $d\tilde{g}$ in (11.33) as the composition

$$(11.34) \quad \begin{array}{ccccccc} d\tilde{g}: B_{r/2} & \xrightarrow{g} & U'' & \xrightarrow{d\tilde{f}} & \text{Iso}(V) & \xrightarrow{\iota} & \text{Iso}(V) \\ \eta & \longmapsto & g(\eta) & \longrightarrow & d\tilde{f}_{g(\eta)} & \longmapsto & d\tilde{f}_{g(\eta)}^{-1} \end{array}$$

Since $d\tilde{f}$ is continuous, the inverse map ι is continuous (Theorem 11.17), and g is continuous it follows that $d\tilde{g}$ is also continuous. Hence $\tilde{g}$ is C^1 . $\square$

Corollary 11.35. *If in Theorem 11.22 the function f is C^k for some $k \in \mathbb{Z}^{>0} \cup \{\infty\}$, then the inverse g is also C^k .*

Proof. We proceed by induction on k finite. We already proved the case $k = 1$. If the corollary is true for $k - 1$, then assuming f is C^k we have in (11.34) that g and $d\tilde{f}$ are C^{k-1} . Since ι is also C^{k-1} , it follows that $d\tilde{g}$ is C^{k-1} and hence that g is C^k . $\square$

Lecture 12: Examples and computations

Characterization of affine maps

We work with the standard data (5.23). Recall Definition 1.10 of an affine map between affine spaces.

Theorem 12.1. *Assume U is connected. Then f is the restriction of an affine map $A \rightarrow B$ if and only if $df: U \rightarrow \text{Hom}(V, W)$ is constant.*

Proof. If f extends to an affine map, then the equality (1.11) shows that the estimate (5.37) in the definition of the differential is satisfied for all ϵ, δ . Conversely, suppose the differential has constant value $T \in \text{Hom}(V, W)$. Fix $p \in U$ and define the neighborhood $U' \subset V$ of zero such that $U = p + U'$. Define

$$(12.2) \quad \begin{aligned} \tilde{f}: U' &\longrightarrow W \\ \xi &\longmapsto f(p + \xi) - f(p) - T\xi \end{aligned}$$

Then $d\tilde{f}_\xi = df_{p+\xi} - T = 0$ for all $\xi \in U'$. It follows from Theorem 7.29 that $\tilde{f}$ is constant, and evaluating at $\xi = 0$ we conclude the constant is zero. $\square$

Examples of the inverse function theorem

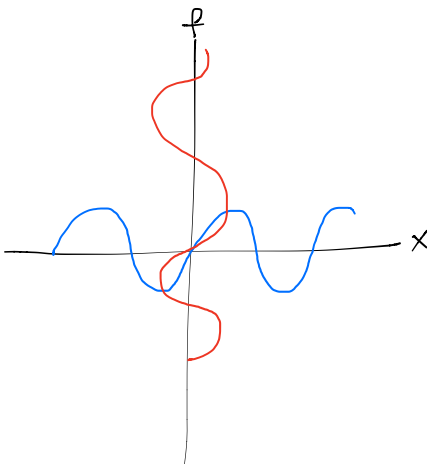


FIGURE 20. Inverse function theorem in one dimension

Example 12.3. We begin with an example in one dimension. Here $U = A = V = B = W = \mathbb{R}$ and $f: \mathbb{R} \rightarrow \mathbb{R}$ is $x \mapsto \sin x$. Then $df_x = f'(x)dx$ is invertible if and only if $f'(x) \neq 0$, which occurs if and only if $x \notin \pi/2 + \pi\mathbb{Z}$. The inverse function theorem, which is much more elementary in this one-variable case, implies that f is locally invertible near such x . The graph of the local inverse is obtained by reflecting the graph of f in the line $f = x$. That reflection is a relation which, locally near x with $f'(x) \neq 0$, is a function.

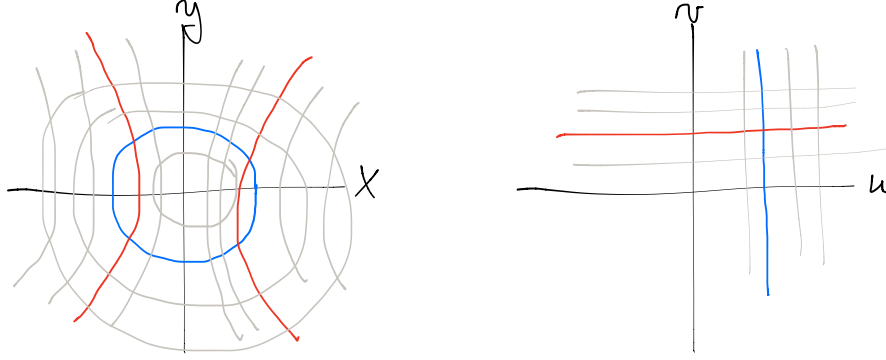


FIGURE 21. Local non-affine change of coordinates

Example 12.4. In this example we look at how the inverse function theorem can be used to produce nonlinear coordinate systems. Consider the function $f: \mathbb{A}_{x,y}^2 \rightarrow \mathbb{A}_{u,v}^2$ defined by the equations

$$(12.5) \quad \begin{aligned} u &= x^2 + y^2 \\ v &= x^2 - y^2 \end{aligned}$$

We would like to use the functions u, v as coordinates on the affine space $\mathbb{A}_{x,y}^2$. They are not global coordinates: the functions u and v agree at points (x, y) and $(-x, -y)$ for all x, y , so those points cannot be distinguished by (u, v) . In other words, (12.5) cannot be inverted globally. The inverse function theory tells for which (x_0, y_0) we can invert locally, and so regard u, v as local coordinates near (x_0, y_0) . To find out which (x_0, y_0) we compute

$$(12.6) \quad \begin{aligned} du &= 2(x dx + y dy) \\ dv &= 2(x dx - y dy) \end{aligned}$$

In other words, with respect to bases $\partial/\partial x, \partial/\partial y$ in the domain and $\partial/\partial u, \partial/\partial v$ in the codomain, the matrix of $df_{(x_0, y_0)}$ is $2 \begin{pmatrix} x_0 & y_0 \\ x_0 & -y_0 \end{pmatrix}$. This is invertible if both x_0 and y_0 are nonzero. The inverse function theorem implies that locally near such (x_0, y_0) we can solve for x, y in terms of u, v . In the u, v coordinate system concentric circles with center the origin (in the x, y plane) are straightened out into the lines $u = \text{constant}$.

Remark 12.7. In these lectures we have often used *global* affine coordinates on affine space (2.27). In the previous example, the non-affine functions u, v are *local* coordinates. Some concepts in affine geometry, such as parallelism, are not invariant under non-affine changes of coordinates.

But many concepts of calculus, such as differentiability of functions, are. A *smooth manifold* is a space glued together from open subsets of affine spaces with gluing maps which are local non-affine changes of coordinates. Hence parallelism does not make sense on a smooth manifold (without extra structure), whereas differentiability of functions does.

Sample computations

You need to have facility with computations and be sure that you understand how to interpret the formulas you write. So here are a few examples.

Example 12.8. Let $U \subset \mathbb{E}_{x,y,z}^3$ be the upper open hemisphere defined by $x^2 + y^2 + z^2 < 1$, $z > 0$. Let $B \subset \mathbb{E}^3$ be the affine subspace $z = 2$, and use affine coordinates $x', y': B \rightarrow \mathbb{R}$ defined as the restrictions of $x, y: \mathbb{E}^3 \rightarrow \mathbb{R}$. Let $f: U \rightarrow B$ be radial projection from the origin. Then you can work out that

$$(12.9) \quad \begin{aligned} x' &= \frac{2x}{z} \\ y' &= \frac{2y}{z} \end{aligned}$$

It follows that

$$(12.10) \quad \begin{aligned} dx' &= \frac{2}{z} dx - \frac{2x}{z^2} dz \\ dy' &= \frac{2}{z} dy - \frac{2y}{z^2} dz \end{aligned}$$

Spend some time interpreting these formulas. Give a geometric interpretation in words of the signs of the various terms. As $z \rightarrow 0$ the values of x', y' change more and more rapidly with changes in x, y, z . Do the relative rates make sense to you? Why does x' change more rapidly with a change in very small z than a change in x for very small z ?

Example 12.11. Suppose $f = f(r, \theta)$ is a function in polar coordinates. Our task here is to compute a formula for the gradient ∇f . First, let's spell out the context. Let $U \subset \mathbb{E}_{x,y}^2$ be the open subset of Euclidean space which omits the ray $x \leq 0$, $y = 0$. Then $r, \theta: U \rightarrow \mathbb{R}$ are well-defined C^∞ functions, where we must make a choice to define θ , say that it takes values in $(-\pi, \pi)$. In fact, the map $(r, \theta): U \rightarrow \mathbb{A}_{r,\theta}^2$ is a bijection onto the open subset $r > 0$, $-\pi < \theta < \pi$, the map is C^∞ , and its inverse is also C^∞ . The inverse is described by the standard equations

$$(12.12) \quad \begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta \end{aligned}$$

from which we compute

$$(12.13) \quad \begin{aligned} dx &= \cos \theta dr - r \sin \theta d\theta \\ dy &= \sin \theta dr + r \cos \theta d\theta \end{aligned}$$

and

$$(12.14) \quad \begin{aligned} \frac{\partial}{\partial r} &= \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} \\ \frac{\partial}{\partial \theta} &= -r \sin \theta \frac{\partial}{\partial x} + r \cos \theta \frac{\partial}{\partial y} \end{aligned}$$

The vectors $\partial/\partial x, \partial/\partial y$ are orthonormal with respect to the standard inner product, and so we deduce

$$(12.15) \quad \left\langle \frac{\partial}{\partial r}, \frac{\partial}{\partial r} \right\rangle = 1, \quad \left\langle \frac{\partial}{\partial r}, \frac{\partial}{\partial \theta} \right\rangle = 0, \quad \left\langle \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \theta} \right\rangle = r^2.$$

Now if $f: U \rightarrow \mathbb{R}$ is a differentiable function, then

$$(12.16) \quad df = \frac{\partial f}{\partial r} dr + \frac{\partial f}{\partial \theta} d\theta,$$

and from (9.22) we deduce

$$(12.17) \quad \nabla f = \frac{\partial f}{\partial r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \frac{\partial}{\partial \theta}.$$

We can also deduce (12.17) from (9.34).

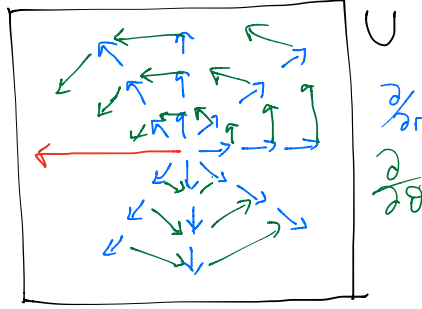


FIGURE 22. The vector fields $\partial/\partial r$ and $\partial/\partial \theta$

Lecture 13: Implicit function theorem; second differential

Motivation and examples

(13.1) Equations with parameters. The implicit function theorem, which is a corollary of the inverse function theorem, concerns equations with parameters. In other words, we seek to solve $f_a(x) = y$ for x given y and a . We view a as the parameter, so we have a family of equations parametrized by a . If there is a solution x_0 at a parameter value a_0 for fixed y_0 , then the implicit function theorem gives a sufficient condition for the existence of solutions near x_0 to the equation with parameter value near a_0 and right hand side near y_0 . For fixed right hand side the solution x is a function of a which is as smooth as the function f . We make all of this more precise below. The equations in question can be in finite or infinite dimensions, the latter occurring for example in the study of differential equations. The “equations” can also be of a more geometric nature, as we illustrate in Example 13.8.

(13.2) Continuity method. It may happen that we have a connected space of parameters a and that we can solve our problem for some a_0 easily and would like to prove the existence of a solution for a different parameter value a_1 . The continuity method consists in proving that the subset S of parameters for which solutions exist is both open and closed. Since $a_0 \in S$ it follows that S is the entire connected space of parameters, and in particular there is a solution for parameter value a_1 . The implicit function theorem is often the tool used to prove that S is open. Closedness of S involves proving that a sequence of solutions converges to a solution, and that often requires estimates to control the convergence.

Example 13.3. Consider the function

$$(13.4) \quad f(a, x) = f_a(x) = x^5 + ax^4 - 1, \quad a, x \in \mathbb{R}.$$

We seek real roots x of this quintic for fixed parameter value a . If $a = 0$ then $x = 1$ is a solution, and we’d like to prove that there is a neighborhood $(-\epsilon, \epsilon)$ of $a = 0$ and a function $\phi: (-\epsilon, \epsilon) \rightarrow \mathbb{R}$ such that $\phi(a)$ is a root of f_a , i.e., $f_a(\phi(a)) = 0$. The proof of the implicit function theorem proceeds by reducing to the inverse function theorem. In this example we have one function of two variables, and for the inverse function theorem we need an equal number of functions and variables. Hence introduce

$$(13.5) \quad \begin{aligned} F: \mathbb{A}^2 &\longrightarrow \mathbb{A}^2 \\ (a, x) &\longmapsto (a, x^5 + ax^4 - 1) \end{aligned}$$

Then in matrix form we have

$$(13.6) \quad dF_{(a,x)} = \begin{pmatrix} 1 & 0 \\ x^4 & 5x^4 + 4ax^3 \end{pmatrix}.$$

Notice that the right hand side of (13.6) is a triangular matrix, so is invertible if and only if the lower right entry, which is $\partial f / \partial x$, is nonzero. In particular, $dF_{(0,1)} = \begin{pmatrix} 1 & 0 \\ 1 & 5 \end{pmatrix}$ is invertible. The inverse function theorem gives a local inverse function G . Write $G(a, 0) = (a, \phi(a))$. Then $F(a, \phi(a)) = (a, 0)$, and so $f_a(\phi(a)) = 0$ as desired. The function ϕ is differentiable and the derivative can be computed by differentiating the equation

$$(13.7) \quad \phi(a)^5 + a\phi(a)^4 - 1 = 0.$$

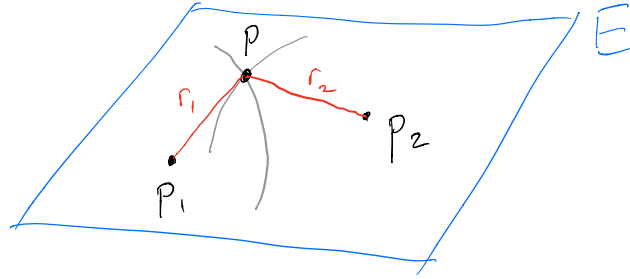


FIGURE 23. Intersection of two circles

Example 13.8. Let E be a Euclidean plane, so an affine space over a 2-dimensional real inner product space V . The problem is to show that the intersection point of two circles is a smooth function of the centers and the radii, and then to compute its differential. We first express the situation in the language of sets and functions. The pair of centers is a point (p_1, p_2) of the set $E^{\times 2} \setminus \Delta$, where $E^{\times 2} = E \times E$ is the Cartesian product and

$$(13.9) \quad \Delta = \{(p, p) : p \in E\} \subset E^{\times 2}$$

is the diagonal: we do not allow the centers to coincide. The radii r_1, r_2 are positive real numbers. Consider the function

$$(13.10) \quad \begin{aligned} f: (E^{\times 2} \setminus \Delta) \times E &\longrightarrow \mathbb{A}^2 \\ (p_1, p_2) \times p &\longmapsto (d(p, p_1)^2, d(p, p_2)^2) \end{aligned}$$

Then p lies on the intersection of the circles of radius r_i with center p_i , $i = 1, 2$, if and only if $f(p_1, p_2; p) = (r_1^2, r_2^2)$. (We use the squares of the distances to simplify the computations below.)

As in (13.5) form the function

$$(13.11) \quad \begin{aligned} F: (E^{\times 2} \setminus \Delta) \times E &\longrightarrow (E^{\times 2} \setminus \Delta) \times \mathbb{A}^2 \\ (p_1, p_2) \times p &\longmapsto (p_1, p_2) \times (d(p, p_1)^2, d(p, p_2)^2) \end{aligned}$$

The differential is triangular, as in (13.6), and so invertibility of $dF_{(p_1, p_2) \times p}$ is equivalent to invertibility of the partial differential

$$(13.12) \quad df_{(p_1, p_2) \times p}^2: V \rightarrow \mathbb{R}^2,$$

defined as $df_{(p_1, p_2) \times p}^2(\xi) = df_{(p_1, p_2) \times p}(0, \xi)$, $\xi \in V$. We compute

$$(13.13) \quad df_{(p_1, p_2) \times p}^2(\xi) = (2\langle \xi, p - p_1 \rangle, 2\langle \xi, p - p_2 \rangle).$$

It is an easy linear algebra exercise to prove from this formula that (13.12) is an isomorphism if and only if $p - p_1, p - p_2$ form a basis of V , which only fails if p lies on the affine line containing p_1, p_2 . (In that case, the two circles of radii $r_1 = d(p, p_1)$ and $r_2 = d(p, p_2)$, respectively, are tangent at p .) Away from that situation the inverse function theorem shows that F is a local diffeomorphism from a neighborhood of some $(\hat{p}_1, \hat{p}_2) \times \hat{p}$ to a neighborhood of $F(\hat{p}_1, \hat{p}_2; \hat{p}) = (\hat{p}_2, \hat{p}_2) \times (\hat{r}_1, \hat{r}_2)$. Let G be a local inverse. Then the intersection point as a function of centers and radii is

$$(13.14) \quad p(p_1, p_2, r_1, r_2) := G(p_1, p_2; r_1, r_2)^2,$$

where the subscript denotes the E -component. Note $p(\hat{p}_1, \hat{p}_2; \hat{r}_1, \hat{r}_2) = \hat{p}$.

Compute the differential by applying the operator d to the equations

$$(13.15) \quad \begin{aligned} \langle p - p_1, p - p_1 \rangle &= r_1^2 \\ \langle p - p_2, p - p_2 \rangle &= r_2^2 \end{aligned}$$

In these equations p, p_1, p_2 are E -valued functions on $(E^{\times 2} \setminus \Delta) \times E$, so their differentials are $\text{Hom}(V^{\oplus 2} \oplus V, V)$ -valued. Differentiating (13.15), canceling the twos, and rearranging we have

$$(13.16) \quad \begin{aligned} \langle dp, p - p_1 \rangle &= \langle dp_1, p - p_1 \rangle + r_1 dr_1 \\ \langle dp, p - p_2 \rangle &= \langle dp_2, p - p_2 \rangle + r_2 dr_2 \end{aligned}$$

These are equalities of linear functionals in the dual space to $V^{\oplus 2} \oplus V \oplus \mathbb{R}$. By the same linear algebra lemma mentioned above, if $p - p_1$ and $p - p_2$ form a basis of V then these equations determine dp .

Remark 13.17. One can and should get used to computing by “following your nose”, so effortlessly pass from (13.15) to (13.16) without worrying where everything lives. At the same time, on request one should be able to tell exactly what each symbol means and where the mathematical object it stands for lives.

Remark 13.18. Less polemically, it is worth playing around with (13.16) to see how the formulas work. Try varying one of the four variables p_1, p_2, r_1, r_2 at a time. How does dp reflect how the intersection of the circles changes?

Remark 13.19. As mentioned above the second partial differential (13.13) fails to be invertible at values $\hat{p}_1, \hat{p}_2, \hat{r}_1, \hat{r}_2$ for which the two circles are tangent. Fixing the radii $\hat{r}_1, \hat{r}_2$, in every neighborhood of $\hat{p}_1, \hat{p}_2$ there are centers p_1, p_2 for which the circles do not intersect, and therefore we cannot locally solve for the intersection point as a function of parameters.

Implicit function theorem

Theorem 13.20. Let A_1, A_2, B be affine spaces modeled on the Banach spaces V_1, V_2, W . Suppose $U_1 \subset A_1$ and $U_2 \subset A_2$ are open, $f: U_1 \times U_2 \rightarrow B$ is a C^k function for some $k \in \mathbb{Z}^{\geq 1} \cup \{\infty\}$, $(\hat{p}_1, \hat{p}_2) \in U_1 \times U_2$, and $f(\hat{p}_1, \hat{p}_2) = \hat{q}$. Assume that the second partial differential

$$(13.21) \quad (df)_{(\hat{p}_1, \hat{p}_2)}^2: V_2 \longrightarrow W$$

is invertible. Then there exists a neighborhood $U'_1 \subset U_1$ of $\hat{p}_1$ and a C^k function $\phi: U'_1 \rightarrow U_2$ such that

$$(13.22) \quad f(p_1, \phi(p_1)) = \hat{q}, \quad p_1 \in U'_1,$$

and $\phi(\hat{p}_1) = \hat{p}_2$. Furthermore,

$$(13.23) \quad d\phi_{p_1} = - \left[(df)_{(p_1, \phi(p_1))}^2 \right]^{-1} \circ (df)_{(p_1, \phi(p_1))}^1, \quad p_1 \in U'_1.$$

At any $(p_1, p_2) \in U_1 \times U_2$ the differential of f is a continuous linear map

$$(13.24) \quad df_{(p_1, p_2)}: V_1 \oplus V_2 \longrightarrow W;$$

the partial differentials $df_{(p_1, p_2)}^1$ and $df_{(p_1, p_2)}^2$ are the restrictions to $V_1 \oplus 0$ and $0 \oplus V_2$, respectively. In terms of (13.1) the variable p_1 is the parameter and we solve for p_2 in terms of p_1 . We can also vary $\hat{q}$, as in Example 13.8; see Corollary 13.29 below.

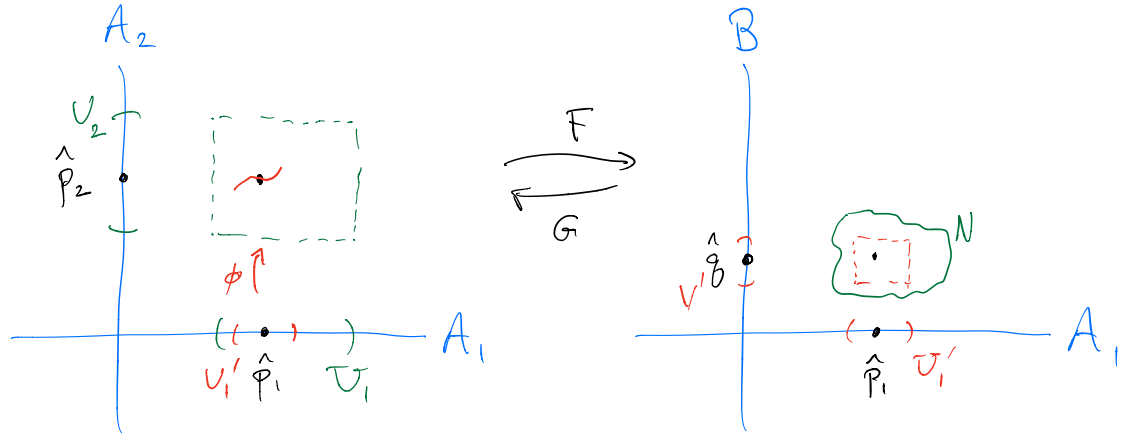


FIGURE 24. Reducing the implicit function theorem to the inverse function theorem

Proof. Consider the function

$$(13.25) \quad \begin{aligned} F: U_1 \times U_2 &\longrightarrow A_1 \times B \\ (p_1, p_2) &\longmapsto (p_1, f(p_1, p_2)) \end{aligned}$$

Then $F(\hat{p}_1, \hat{p}_2) = (\hat{p}_1, \hat{q})$. We claim $dF_{(\hat{p}_1, \hat{p}_2)}$ is invertible. Compute

$$(13.26) \quad dF_{(p_1, p_2)}(\xi_1, \xi_2) = (\xi_1, df_{(p_1, p_2)}^1(\xi_1) + df_{(p_1, p_2)}^2(\xi_2)), \quad \xi_1 \in V_1, \quad \xi_2 \in V_2.$$

Given $(\xi_1, \eta) \in V_1 \times W$, choose $\xi_2 \in V_2$ such that $df_{(\hat{p}_1, \hat{p}_2)}^2(\xi_2) = \eta - df_{(\hat{p}_1, \hat{p}_2)}^1(\xi_1)$, which we can do since $df_{(\hat{p}_1, \hat{p}_2)}^2$ is assumed surjective. Then $dF_{(\hat{p}_1, \hat{p}_2)}(\xi_1, \xi_2) = (\xi_1, \eta)$, which proves surjectivity. The proof that $dF_{(\hat{p}_1, \hat{p}_2)}$ is injective is similar. The Inverse Function Theorem 11.22 provides an open set $N \subset U_1 \times B$ and a local inverse $G: N \rightarrow U_1 \times U_2$ to F . Choose $U'_1 \subset U_1$ containing $\hat{p}_1$ and a neighborhood $V' \subset B$ of $\hat{q}$ so that $U'_1 \times V' \subset N$. Then define

$$(13.27) \quad \begin{array}{ccccccc} \phi: U'_1 & \longrightarrow & N & \xrightarrow{G} & U_1 \times U_2 & \xrightarrow{\text{project}} & U_2 \\ p_1 & \longmapsto & (p_1, q) & \longrightarrow & G(p_1, q) & \longmapsto & G(p_1, q)^2 \end{array}$$

Since G is C^k and the first and last functions are C^∞ , the composition is C^k . To derive (13.23) differentiate (13.22):

$$(13.28) \quad df_{(p_1, \phi(p_1))}^1 + df_{(p_1, \phi(p_1))}^2 \circ d\phi_{p_1} = 0.$$

□

The proof of the following is immediate from the previous proof.

Corollary 13.29. *In the situation of Theorem 13.20 there exists a neighborhood V' of $\hat{q}$ in B and a C^k function $\Phi: U'_1 \times V' \rightarrow U_2$ such that*

$$(13.30) \quad f(p_1, \Phi(p_1, q)) = q, \quad p_1 \in U'_1, \quad q \in V',$$

and $\Phi(\hat{p}_1, \hat{q}) = \hat{p}_2$.

Simply set $\Phi(p_1, q) = G(p_1, q)^2$.

Introduction to the second differential

Resume with the standard data (5.23). Assume that f is differentiable; the differential is a function $df: U \rightarrow \text{Hom}(V, W)$.

Definition 13.31. *f is twice differentiable at $p \in U$ if df is differentiable at p .*

In that case the second differential is a continuous linear map

$$(13.32) \quad d^2f_p = d(df)_p: V \longrightarrow \text{Hom}(V, W),$$

so a map $V \times V \rightarrow W$. We claim it is continuous and bilinear. First, if $\xi \in V$ then by definition $d^2f_p(\xi)(-) \in \text{Hom}(V, W)$ is a continuous linear function (of the argument, indicated as ‘-’). On the other hand, $d^2f_p(-)(\xi)$ is the composition

$$(13.33) \quad V \xrightarrow{d^2f_p} \text{Hom}(V, W) \xrightarrow{\text{ev}_\xi} W$$

of continuous linear maps, so is a continuous linear map. Here ev_ξ is evaluation of a linear functional on ξ .

The main result about higher differentials is the following.

Theorem 13.34. *If f is twice differentiable at $p \in U$, then d^2f_p is a continuous symmetric bilinear form:*

$$(13.35) \quad d^2f_p(\xi_1)(\xi_2) = d^2f_p(\xi_2)(\xi_1), \quad \xi_1, \xi_2 \in V.$$

We give the proof in the next lecture. For now we give a formula for the second differential, assuming it exists, in terms of iterated directional derivatives. (Recall the analog Theorem 6.10.)

Proposition 13.36. *If f is C^1 and is twice differentiable at p , then*

$$(13.37) \quad d^2f_p(\xi_1)(\xi_2) = \xi_1 \xi_2 f(p), \quad \xi_1, \xi_2 \in V.$$

The right hand side is the directional derivative of the function $\xi_2 f: U \rightarrow B$ at p in the direction ξ_1 .

Proof. Apply (6.11) twice:

$$(13.38) \quad d^2f_p(\xi_1)(-) = d(df)_p(\xi_1) = (\xi_1 \cdot df)(p) \in \text{Hom}(V, W),$$

and so

$$(13.39) \quad d^2f_p(\xi_1)(\xi_2) = \text{ev}_{\xi_2}(\xi_1 \cdot df)(p) = (\xi_1 \cdot \text{ev}_{\xi_2} df)(p) = \xi_1 \xi_2 f(p) \in W. \quad \square$$

Lecture 14: The second differential; symmetric bilinear forms

Main theorems about the second differential

(14.1) *Introduction.* We continue to work in our standard setting (5.23). One of the main theorems for the first differential, Theorem 8.6, asserts that if the domain U of the function f is finite dimensional, and if all directional derivatives exist and are continuous at a point $p \in U$, then f is differentiable at p . (If, furthermore, the directional derivatives are continuous on U , then df is continuous on U .) One of our tasks is to prove an analog of Theorem 8.6 for the second differential, where now the hypothesis concerns iterated directional derivatives. This is Theorem 14.24 below. But to begin we prove Theorem 13.34, which has no analog for the first differential.

(14.2) *Symmetry of the second differential.* Recall that if f is twice differentiable at p , then the second differential is a continuous bilinear map

$$(14.3) \quad df_p^2: V \times V \longrightarrow W.$$

Theorem 13.34 asserts that (14.3) is a *symmetric* function.

Proof of Theorem 13.34. Fix $\epsilon > 0$. Since df is differentiable at p we can and do choose $\delta > 0$ such that if $\xi \in V$ satisfies $\|\xi\| < \delta$, then $p + \xi \in U$ and

$$(14.4) \quad \|df_{p+\xi} - df_p - d^2f_p(\xi)\| \leq \epsilon \|\xi\|.$$

Note that the norm on the left is the operator norm in $\text{Hom}(V, W)$. Suppose $\xi, \xi_2 \in V$ satisfy $\|\xi\|, \|\xi_2\| < \delta/2$. Apply (14.4) with $\xi + \xi_2$ to obtain

$$(14.5) \quad \|df_{p+\xi+\xi_2} - df_p - d^2f_p(\xi + \xi_2)\| \leq \epsilon \|\xi + \xi_2\|.$$

Subtract (14.5) and (14.4) and apply the triangle inequality:

$$(14.6) \quad \|df_{p+\xi+\xi_2} - df_{p+\xi} - d^2f_p(\xi_2)\| \leq \epsilon(\|\xi\| + \|\xi + \xi_2\|) \leq 2\epsilon(\|\xi\| + \|\xi_2\|).$$

For fixed $\xi_2 \in V$ with $\|\xi_2\| < \delta/2$ define the function $g: B_{\delta/2}(0) \rightarrow W$ on the ball of radius $\delta/2$ about the origin in V by

$$(14.7) \quad \phi(\xi) = f(p + \xi + \xi_2) - f(p + \xi) - d^2f_p(\xi_2)(\xi).$$

Then (14.6) implies

$$(14.8) \quad \|d\phi_\xi\| \leq 2\epsilon(\|\xi\| + \|\xi_2\|), \quad \|\xi\| < \delta/2.$$

Now fix $\xi_1 \in V$ with $\|\xi_1\| = \|\xi_2\|$. Then a corollary¹⁴ of the mean value inequality Theorem 7.22, applied on the ball of radius $\delta/2$, yields

$$(14.9) \quad \|\phi(\xi_1) - \phi(0)\| \leq 2\epsilon(\|\xi_1\| + \|\xi_2\|)\|\xi_1\| = 4\epsilon\|\xi_1\|\|\xi_2\|.$$

Note the first line of

$$(14.10) \quad \begin{aligned} \phi(\xi_1) - \phi(0) &= f(p + \xi_1 + \xi_2) - f(p + \xi_1) - f(p + \xi_2) + f(p) \\ &\quad - d^2f_p(\xi_2)(\xi_1) \end{aligned}$$

is symmetric in ξ_1, ξ_2 . Repeat the argument exchanging ξ_1, ξ_2 . Then subtract and use the triangle inequality to conclude from (14.9) that

$$(14.11) \quad \|d^2f_p(\xi_1)(\xi_2) - d^2f_p(\xi_2)(\xi_1)\| \leq 8\epsilon\|\xi_1\|\|\xi_2\|.$$

We have proved (14.11) for any vectors $\xi_1, \xi_2 \in V$ with $\|\xi_1\| = \|\xi_2\| < \delta/2$. Multiplying ξ_1, ξ_2 by arbitrary scalars $c_1, c_2 > 0$ and using linearity and homogeneity of the norm we conclude that (14.11) holds for all $\xi_1, \xi_2 \in V$. As it also holds for all $\epsilon > 0$ we can take the limit as $\epsilon \rightarrow 0$ to deduce the desired equality (13.35) for all $\xi_1, \xi_2 \in V$. $\square$

(14.12) Equality of mixed partials. Suppose $A = \mathbb{A}^n$ is standard affine space with affine coordinates $x^1, \dots, x^n$. The directional derivative of f in the direction of the coordinate vector $\partial/\partial x^i$ is the partial derivative $\partial f/\partial x^i: U \rightarrow W$. Assuming f is twice differentiable, Theorem 13.34 and Proposition 13.36 imply

$$(14.13) \quad \frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i}$$

for all $1 \leq i, j \leq n$. Equation (14.13) goes by the slogan “equality of mixed partials”.

¹⁴From Homework #4: Let A be an affine space over a normed linear space V . A set $U \subset A$ is *convex* if whenever $p_0, p_1 \in U$, then $tp_0 + (1-t)p_1 \in U$ for all $t \in [0, 1]$. Suppose $U \subset A$ is open and convex, B is an affine space over a normed linear space W , and $f: U \rightarrow B$ is differentiable. Assume there exists $C \in \mathbb{R}^{>0}$ such that $\|df_p\| < C$ for all $p \in U$. Prove that if $p_0, p_1 \in U$, then

$$\|f(p_1) - f(p_0)\| \leq C\|\xi\|, \quad p_1 = p_0 + \xi.$$

(14.14) *Iterated directional derivatives and the second differential.* As preparation for Theorem 14.24 we prove two lemmas.

Lemma 14.15. *Assume V is finite dimensional, $g: U \rightarrow B$ a function, and for a basis $e_1, \dots, e_n$ of V the directional derivatives $e_i g: U \rightarrow W$ exist and are bounded on U . Then g is continuous.*

Proof. All norms on V are equivalent, so as in previous proofs the 1-norm is most convenient: for $\xi = \xi^i e_i \in V$ define

$$(14.16) \quad \|\xi\| = \sum_i |\xi^i|.$$

Since each of the finitely many $e_i g$ is bounded, there exists $C > 0$ such that $\|(e_i g)(q)\| \leq C$ for all $q \in U$, $i = 1, \dots, n$. Then if $p, p + \xi \in U$ we have

$$(14.17) \quad \begin{aligned} \|g(p + \xi) - g(p)\| &= \left\| \sum_{i=1}^n g(p + \xi^1 e_1 + \dots + \xi^i e_i) - g(p + \xi^1 e_1 + \dots + \xi^{i-1} e_{i-1}) \right\| \\ &\leq \sum_{i=1}^n \|g(p + \xi^1 e_1 + \dots + \xi^i e_i) - g(p + \xi^1 e_1 + \dots + \xi^{i-1} e_{i-1})\| \\ &\leq \sum_{i=1}^n C |\xi^i| \\ &= C \|\xi\| \end{aligned}$$

The third line follows from Theorem 7.22 applied to the function

$$(14.18) \quad t \longmapsto g(p + \xi^1 e_1 + \dots + \xi^{i-1} e_{i-1} + t e_i).$$

Inequality (14.17) implies that g is (Lipschitz) continuous. □

Lemma 14.19. *Suppose V is finite dimensional, W is a normed linear space, $U \subset A$ an open subset of an affine space over V , and $p \in U$. Let $e_1, \dots, e_n$ be a basis of V and $\xi \in V$ a fixed vector. Then for a function $T: U \rightarrow \text{Hom}(V, W)$ the directional derivative $\xi T(p) \in \text{Hom}(V, W)$ exists if and only if the directional derivative $\xi(T(e_j))(p)$ exists for all j . In that case*

$$(14.20) \quad (\xi T)(e_j)(p) = \xi(T(e_j))(p).$$

Proof. The function $T(e_j)$ is the composition

$$(14.21) \quad U \xrightarrow{T} \text{Hom}(V, W) \xrightarrow{\text{ev}_{e_j}} W$$

so the implication in the forward direction follows from the chain rule, which yields (14.20). Conversely, assuming the directional derivative $\xi(T(e_j))(p)$ exists for all j , introduce the dual basis $e^1, \dots, e^n$ of V^* and write

$$(14.22) \quad T = T(e_j) e^j.$$

Then

$$(14.23) \quad \xi T(p) = \frac{d}{dt} \Big|_{t=0} T(p + t\xi) = \frac{d}{dt} \Big|_{t=0} T(e_j)(p + t\xi)e^j = (\xi T(e_j))(p)e^j$$

exists, and again (14.20) holds. $\square$

Theorem 14.24. *Assume V is finite dimensional and for a basis $e_1, \dots, e_n$ of V the iterated directional derivatives $e_i e_j f: U \rightarrow W$ exist and are continuous at $p \in U$. Then f is twice differentiable at p .*

Proof of Theorem 14.24. Since the n^2 second directional derivatives $e_i e_j f$ are continuous at p , there is an open neighborhood $U' \subset U$ of p on which they are all bounded. It follows from Lemma 14.15 that each $e_j f$ is continuous on U' . Then Theorem 8.6 implies that f is differentiable on U' and $df(e_j) = e_j f$. Lemma 14.19 implies that the directional derivatives $e_i df$ of df exist and are continuous at p . Now another application of Theorem 8.6 proves that f is twice differentiable at p . $\square$

Symmetric bilinear forms

(14.25) *Introduction.* Symmetric bilinear forms arise in many places, and so we prove some basic structure theory for a real-valued form B on a finite dimensional vector space V . The basic result is that B can be diagonalized. However, there is no meaning to the magnitude of the “diagonal entries”; only the signs have geometric significance. On the other hand, if we introduce an inner product on V as well, then the “ratio” of B with the inner product is a self-adjoint operator, and its eigenvalues do have geometric significance: they are the critical values of a function associated to B .

(14.26) *Basic definitions.* The following terms apply in infinite dimensions as well. Let V be a normed linear space and

$$(14.27) \quad B: V \times V \rightarrow \mathbb{R}$$

a continuous symmetric bilinear form. It gives rise to a continuous linear map

$$(14.28) \quad \begin{aligned} T_B: V &\longrightarrow V^* \\ \xi &\longmapsto (\eta \mapsto B(\xi, \eta)) \end{aligned}$$

Definition 14.29.

- (i) The *kernel* of B is the closed subspace $\text{Ker } B = \text{Ker } T_B \subset V$.
- (ii) B is *nondegenerate* if $\text{Ker } B = 0$.

- (iii) B is *positive definite* if $B(\xi, \xi) > 0$ for all $\xi \neq 0$. B is *negative definite* if $-B$ is positive definite. If either holds we say B is *definite*.
- (iv) The *null cone*¹⁵ of B is the set of vectors

$$(14.30) \quad \mathcal{N}_B = \{\xi \in V : B(\xi, \xi) = 0\}.$$

So $\xi \in \text{Ker } B$ if and only if $B(\xi, \eta) = 0$ for all $\eta \in V$. If B is nondegenerate and V is finite dimensional, then T_B is an isomorphism. A form is definite if and only if its null cone consists only of the zero vector. Recall (Definition 8.26) that a positive definite symmetric bilinear form is an inner product.

Remark 14.31. The bilinear form (14.27) induces a continuous symmetric bilinear form

$$(14.32) \quad \overline{B}: \frac{V}{\text{Ker } B} \times \frac{V}{\text{Ker } B} \longrightarrow \mathbb{R}$$

on the quotient space $V/\text{Ker } B$, since it is constant on cosets of the kernel. The form $\overline{B}$ is nondegenerate. In this way the theory of symmetric bilinear forms can be reduced to the nondegenerate case.

Example 14.33. Let $V = \mathbb{R}^2$ and define

$$(14.34) \quad B((\xi^1, \xi^2), (\eta^1, \eta^2)) = \xi^1 \eta^1 - \xi^2 \eta^2.$$

This form is nondegenerate, but is neither positive nor negative definite. Its null cone is the union of two lines. Those two lines are two points on $\mathbb{RP}^1 = \mathbb{P}(\mathbb{R}^2)$. Deleting those points we are left with two components, one consisting of lines in $\mathbb{R}^2$ on which B is positive definite and the other consisting of lines in $\mathbb{R}^2$ on which B is negative definite.

Suppose V is finite dimensional and B a symmetric bilinear form on V . One numerical invariant is $\dim \text{Ker } B$. The next result gives two more.

(14.35) *Numerical invariants of a bilinear form on a finite dimensional vector space.*

Proposition 14.36. *Suppose $P, P' \subset V$ are maximal subspaces on which B is positive definite. Then $\dim P' = \dim P$.*

Proof. Since $P \cap \text{Ker } B = P' \cap \text{Ker } B = \emptyset$ we may replace V by $V/\text{Ker } B$ and so from the beginning can assume B is nondegenerate. Suppose $\dim P' < \dim P$. Since B is positive definite on P and B is nondegenerate, it follows that $V = P \oplus P^\perp$, where $P^\perp = \{\zeta \in V : B(\xi, \zeta) = 0 \text{ for all } \xi \in P\}$. Let $\pi: V \rightarrow P$ be the projection with kernel $P^\perp$. Then $\pi(P') \subset P$ is proper—i.e., not equal to P —so we can and do choose a nonzero vector $\eta \in P$ such that $B(\eta, \pi(P')) = 0$. It follows that $B(\eta, P') = 0$, and so $B(\xi' + \eta, \xi' + \eta) = B(\xi', \xi') + B(\eta, \eta) > 0$ for all $\xi' \in P'$. Hence P' is not maximal. $\square$

¹⁵A cone in a real vector space is a subset invariant under scalar multiplication by positive real numbers.

The same result applies to subspaces on which B is negative definite. (Apply Proposition 14.36 to $-B$.) Define $b_0(B)$, $b_+(B)$, $b_-(B)$ to be the dimension of the kernel of B , the dimension of a maximal subspace on which B is positive definite, and the dimension of a maximal subspace on which B is negative definite, respectively. These are the numerical invariants of B . To repeat: there is a canonical subspace $K = \text{Ker } B$ and we can choose (noncanonical) subspaces P, N to be maximal subspaces on which B is positive/negative definite. Clearly the only vector lying in any two of these is the zero vector. Also, the span $K + P + N \subset V$ equals V . (Evaluate $B(\xi, \xi)$ on a nonzero vector not contained in the span to deduce that P or N is not maximal.) Hence $V = K \oplus P \oplus N$. Note that B is an inner product on P , and $-B$ is an inner product on N . Choose a basis $e_1, \dots, e_n$ of V consisting of a basis of K followed by an orthonormal basis of P followed by an orthonormal basis of N . Then

$$(14.37) \quad B(e_i, e_j) = \begin{cases} 0, & i \neq j; \\ 0, +1, -1, & i = j, \end{cases}$$

where the value of $B(e_i, e_i)$ is determined by which subspace e_i lives in.

Remark 14.38. The difference $b_+(B) - b_-(B)$ is called the *signature* of B . The number $b_-(B)$ is called the *index* of B .

Lecture 15: Quadratic approximation and the second variation formula

Symmetric bilinear forms and inner products

Let V be a finite dimensional real vector space and $B: V \times V \rightarrow \mathbb{R}$ a symmetric bilinear form. In (14.35) we proved that there is a basis $e_1, \dots, e_n$ of V in which B is diagonal, in the sense that $B(e_i, e_j) = 0$ if $i \neq j$. By scaling the basis elements we can arrange that the diagonal entries $B(e_i, e_i)$ are either 0, +1, or -1; see (14.37).

(15.1) *The associated self-adjoint operator.* Now suppose V also has an inner product $\langle -, - \rangle$. Define a linear operator $S_B: V \rightarrow V$ by

$$(15.2) \quad B(\xi_1, \xi_2) = \langle \xi_1, S_B(\xi_2) \rangle, \quad \xi_1, \xi_2 \in V.$$

The symmetry of B implies that S_B is *self-adjoint* in the sense that

$$(15.3) \quad \langle \xi_1, S_B(\xi_2) \rangle = \langle S_B(\xi_1), \xi_2 \rangle, \quad \xi_1, \xi_2 \in V.$$

(15.4) *Diagonalization.* The operator S_B is diagonalizable, as we now prove.

Theorem 15.5. *Let V be a finite dimensional real inner product space and $S: V \rightarrow V$ a self-adjoint operator. Then S has a (nonzero) eigenvector.*

Of course, not every linear operator has an eigenvector: a nontrivial rotation in the plane does not fix any line. The following proof is essentially a reprise of the second proof of Proposition 9.25.

Proof. Consider the functions $f, g: V \rightarrow \mathbb{R}$ defined by

$$(15.6) \quad \begin{aligned} f(\xi) &= \frac{1}{2} \langle \xi, S(\xi) \rangle, \\ g(\xi) &= \frac{1}{2} \langle \xi, \xi \rangle. \end{aligned}$$

Let $S(V) = g^{-1}(1) \subset V$ be the unit sphere. Since $S(V)$ is compact, f has a maximum on $S(V)$, say at $e_1 \in S(V)$. The Lagrange multiplier criterion implies that there exists $\lambda_1 \in \mathbb{R}$ such that $df_{e_1} = \lambda_1 dg_{e_1}$, in other words $S(e_1) = \lambda_1 e_1$. $\square$

Corollary 15.7. *In the situation of Theorem 15.5 the operator S is diagonalizable.*

Proof. Let V_1 be the orthogonal complement to the eigenvector e_1 . The self-adjointness (15.3) implies that if $\xi_2 \in V_1$, then $S(\xi_2) \in V_1$ and the restriction of S to V_1 is self-adjoint. Theorem 15.5 produces an eigenvector e_2 of this restriction with eigenvalue $\lambda_2 \leq \lambda_1$. Repeat the argument a total of $\dim V$ times to produce an orthonormal basis $e_1, \dots, e_n$ of eigenvectors with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. $\square$

It follows immediately from (15.2) that B is diagonalized as well:

$$(15.8) \quad B(e_i, e_j) = \begin{cases} 0, & i \neq j; \\ \lambda_i, & i = j, \end{cases}$$

The inner product gives meaning to the diagonal entries; compare (14.37).

The second derivative test and quadratic approximation

(15.9) *Introduction.* Resume our standard setup (5.23) with $B = \mathbb{R}$. Let $p_0 \in U$ be a critical point of f , i.e., $df_{p_0} = 0$. Then we would like to say that the function

$$(15.10) \quad p_0 + \xi \longmapsto f(p_0) + \frac{1}{2}d^2f_{p_0}(\xi, \xi)$$

is a good approximation to f near p_0 . But this is not necessarily true.

Example 15.11. Take $U = A = \mathbb{R}$ and consider the five functions

$$(15.12) \quad \begin{aligned} f_1(x) &= +x^2 \\ f_2(x) &= -x^2 \\ f_3(x) &= x^3 \\ f_4(x) &= +x^4 \\ f_5(x) &= -x^4 \end{aligned}$$

Each function has a critical point at $x = 0$. The quadratic approximation (15.10) works—in fact is exact—for f_1, f_2 . For f_3, f_4, f_5 the quadratic approximation is a constant function and that does not predict the local behavior: $x = 0$ is an inflection point of f_3 , a local minimum of f_4 , and a local maximum of f_5 .

(15.13) *Second derivative test for local extrema.* The quadratic approximation is guaranteed to be good if $d^2f_{p_0}$ is nondegenerate and if A is finite dimensional. We first prove a special case, when $d^2f_{p_0}$ is positive definite. The analogous theorem for $d^2f_{p_0}$ negative definite follows by applying the following to $-f$.

Theorem 15.14. *Suppose A is finite dimensional, f is a C^1 function, $p_0 \in U$ is a critical point, f is twice differentiable at p_0 , and $d^2f_{p_0}$ is positive definite. Then f has a strict local minimum at p_0 .*

The strictness means that there exists a neighborhood $U' \subset U$ of p_0 such that $f(p) > f(p_0)$ for all $p \in U' \setminus \{p_0\}$.

Remark 15.15. The following proof works for A infinite dimensional if the given norm on V is equivalent to the norm (15.16) defined by the second differential. In finite dimensions all norms are equivalent, and so we can and do use (15.16) as the norm in the definition of differentiability.

Proof. Since d^2f is positive definite,

$$(15.16) \quad \|\xi\| = \sqrt{d^2f_{p_0}(\xi, \xi)}, \quad \xi \in V,$$

is a norm on V . The twice differentiability of f at p_0 is the assertion: given $\epsilon > 0$ there exists $\delta > 0$ such that if $\xi \in V$ satisfies $\|\xi\| < \delta$, then $p_0 + \xi \in U$ and

$$(15.17) \quad |df_{p_0+\xi} - df_{p_0} - d^2f_{p_0}(\xi)| \leq \epsilon \|\xi\|.$$

Fix $0 < \epsilon < 1$ and choose $\delta > 0$ so that (15.17) holds if $\|\xi\| < \delta$. Now fix $\xi_0 \in B_\delta(0)$ and set $g(t) = f(p_0 + t\xi_0)$, $t \in [0, 1]$. Then $g'(t) = df_{p_0+t\xi_0}(\xi_0)$. Evaluate the linear functionals in (15.17) on ξ_0 and use the fact that $df_{p_0} = 0$ to conclude

$$(15.18) \quad (1 - \epsilon)t\|\xi_0\|^2 \leq g'(t) \leq (1 + \epsilon)t\|\xi_0\|^2.$$

This is an inequality of real-valued functions of t . Integrating we conclude that

$$(15.19) \quad \frac{1 - \epsilon}{2} \|\xi_0\|^2 \leq g(1) - g(0) \leq \frac{1 + \epsilon}{2} \|\xi_0\|^2,$$

which is

$$(15.20) \quad f(p_0) + \frac{1 - \epsilon}{2} \|\xi_0\|^2 \leq f(p_0 + \xi_0) \leq f(p_0) + \frac{1 + \epsilon}{2} \|\xi_0\|^2.$$

In particular, $f(p_0 + \xi_0) \geq f(p_0)$ and $f(p_0 + \xi_0) > f(p_0)$ if $\xi_0 \neq 0$. Therefore, f has a strict local minimum at p_0 . $\square$

(15.21) More general quadratic approximations. The inequalities (15.20) sandwich the function f between two quadratic functions, an approximation valid in a neighborhood of the critical point p_0 with positive definite second differential. More generally, suppose p_0 is a critical point of f with nondegenerate second differential. Choose a decomposition $V = P \oplus N$ such that $d^2f_{p_0}$ is positive definite on P and negative definite on N . Define the norm

$$(15.22) \quad \|\xi' + \xi''\| = \sqrt{d^2f_{p_0}(\xi', \xi')} + \sqrt{-d^2f_{p_0}(\xi'', \xi'')}, \quad \xi' + \xi'' \in P \oplus N,$$

on V .

Theorem 15.23. *Suppose A is finite dimensional, f is a C^1 function, $p_0 \in U$ is a critical point, f is twice differentiable at p_0 , and $d^2f_{p_0}$ is nondegenerate. Then in a neighborhood $U' \subset U$ of p_0 we have*

$$(15.24) \quad f(p_0) + \frac{1-\epsilon}{2}\|\xi'\|^2 - \frac{1+\epsilon}{2}\|\xi''\|^2 \leq f(p_0 + \xi' + \xi'') \leq f(p_0) + \frac{1+\epsilon}{2}\|\xi'\|^2 - \frac{1-\epsilon}{2}\|\xi''\|^2,$$

in terms of the norm (15.22), for $\xi' + \xi'' \in V = P \oplus N$.

Proof. We only need a small modification of the proof of Theorem 15.14. Namely, (15.17) implies that for $\xi = \xi' + \xi''$ of norm less than δ and $0 \leq t \leq 1$ we have

$$(15.25) \quad \left| df_{p_0+t\xi'+t\xi''}(\xi' + \xi'') - t\|\xi'\|^2 + t\|\xi''\|^2 \right| \leq \epsilon t(\|\xi'\|^2 + \|\xi''\|^2),$$

and so writing $\xi_0 = \xi'_0 + \xi''_0$ we replace (15.18) with the inequalities

$$(15.26) \quad (1-\epsilon)t\|\xi'_0\|^2 - (1+\epsilon)t\|\xi''_0\|^2 \leq g'(t) \leq (1+\epsilon)t\|\xi'_0\|^2 - (1-\epsilon)t\|\xi''_0\|^2.$$

The inequalities (15.24) follow by integrating (15.26). □

Second variation formula

(15.27) *Recalling the setup.* In Lecture 10 we computed the first variation formula, that is, the differential of the length function. The setup is that V is a finite dimensional real inner product space, E an affine space over V , and $p, q \in E$ fixed points. Parametrized paths from p to q form an affine space

$$(15.28) \quad A = \{\gamma: [0, 1] \longrightarrow E \text{ such that } \gamma(0) = p, \gamma(1) = q, \gamma \in C^2([0, 1], E)\}$$

whose tangent space is the vector space

$$(15.29) \quad X = \{\xi: [0, 1] \longrightarrow V \text{ such that } \xi(0) = \xi(1) = 0, \xi \in C^2([0, 1], V)\}$$

with norm

$$(15.30) \quad \|\xi\| = \max_{s \in [0, 1]} \|\dot{\xi}(s)\|_V,$$

where the dot denotes d/ds . Set

$$(15.31) \quad U = \{\gamma \in A \text{ such that } \dot{\gamma}(s) \neq 0 \text{ for all } s \in [0, 1]\}.$$

and define the length function

$$(15.32) \quad \begin{aligned} f: U &\longrightarrow \mathbb{R} \\ \gamma &\longmapsto \int_0^1 ds \|\dot{\gamma}(s)\| \end{aligned}$$

In Theorem 10.28 we proved that f is differentiable with differential

$$(15.33) \quad df_\gamma(\xi) = \xi f(\gamma) = \int_0^1 ds \frac{\langle \dot{\gamma}, \dot{\xi} \rangle}{\langle \dot{\gamma}, \dot{\gamma} \rangle^{1/2}}.$$

Furthermore, γ is a critical point if it is a constant velocity motion, or a reparametrization of a constant velocity motion.

(15.34) The second directional derivative. Assume γ is a unit speed, so $\langle \dot{\gamma}, \dot{\gamma} \rangle = 1$ and $\ddot{\gamma} = 0$. We do not prove that f is twice differentiable at γ , but content ourselves with computing the iterated second directional derivative. Fix $\xi_1, \xi_2 \in X$. Then commuting differentiation and integration, as in (10.12) and with the same justification, we have

$$(15.35) \quad \begin{aligned} \xi_1 \xi_2 f(\gamma) &= \left. \frac{d}{dt} \right|_{t=0} \xi_2 f(\gamma + t\xi_1) \\ &= \left. \frac{d}{dt} \right|_{t=0} \int_0^1 ds \frac{\langle \dot{\gamma} + t\dot{\xi}_1, \dot{\xi}_2 \rangle}{\langle \dot{\gamma} + t\dot{\xi}_1, \dot{\gamma} + t\dot{\xi}_1 \rangle^{1/2}} \\ &= \int_0^1 ds \left\{ \langle \dot{\xi}_1, \dot{\xi}_2 \rangle - \frac{1}{2} \langle \dot{\gamma}, \dot{\gamma} \rangle^{-3/2} 2 \langle \dot{\gamma}, \dot{\xi}_1 \rangle \langle \dot{\gamma}, \dot{\xi}_2 \rangle \right\} \\ &= \int_0^1 \left\{ \langle \dot{\xi}_1, \dot{\xi}_2 \rangle - \langle \dot{\gamma}, \dot{\xi}_1 \rangle \langle \dot{\gamma}, \dot{\xi}_2 \rangle \right\}. \end{aligned}$$

This is a symmetric bilinear form in the variables ξ_1, ξ_2 , as it should be. It has an infinite dimensional kernel due to reparametrization invariance. Namely, if $\rho_1: [0, 1] \rightarrow \mathbb{R}$ is a C^2 function with $\rho_1(0) = \rho_1(1) = 0$, then for $\xi_1(s) = \rho_1(s)\dot{\gamma}$ we have $\dot{\xi}_1 = \dot{\rho}_1\dot{\gamma}$ and (15.35) vanishes for all $\xi_2 \in X$. We claim that (15.35) is positive semidefinite. Namely, in general we write $\xi \in X$ as

$$(15.36) \quad \xi(s) = \rho(s)\dot{\gamma} + \eta, \quad \langle \eta(s), \dot{\gamma} \rangle = 0,$$

for $\eta: [0, 1] \rightarrow V$ with $\eta(0) = \eta(1) = 0$. Differentiating the constraint we find $\langle \dot{\eta}, \dot{\gamma} \rangle = 0$. Setting $\xi_1 = \xi_2 = \rho\dot{\gamma} + \eta$ we compute

$$(15.37) \quad \xi \xi f(\gamma) = \int_0^1 \|\dot{\eta}\|^2$$

which is nonnegative.

Remark 15.38. Theorem 15.14 does not apply since the second differential is only semidefinite, not definite, and the domain is infinite dimensional. We can work modulo the kernel to obtain a positive definite form, but it is not equivalent to (15.30) (on the quotient), so we would need further argument to prove that—spoiler alert!—the shortest distance between two points in Euclidean space is a straight line segment.

Lecture 16: Introduction to curvature

Introduction

In (15.21) we described how the second differential forms part of an affine quadratic approximation to a function. (There, in Theorem 15.23, we considered the approximation centered at a critical point, but a similar idea—the second-order Taylor expansion—works with arbitrary center to the approximation.) The first-order, or affine linear, approximation to a function (5.34), has an analog for “smooth curved shapes”: the tangent space. Of course, this prompts us to ask:

Question 16.1. What is the mathematical definition of a ‘smooth curved shape’?

We do not answer that question in this lecture, but for 1-dimensional curves and 2-dimensional surfaces in affine space we work with a provisional model. The proper term for a ‘smooth curved shape’ is a ‘smooth manifold’.

Remark 16.2. These smooth manifolds are embedded in an affine space, and we use the ambient affine space to be able to work with the curved shapes. There are also more abstract smooth manifolds which do not appear embedded in affine space. For example, let V be a finite dimensional real vector space. Then the Grassmannian $Gr_k(V)$ of k -dimensional subspaces of V can be made into a smooth manifold.

A smooth manifold—abstract or embedded—has a well-defined tangent space at each point. This is the first-order approximation to the manifold at a point. The issue at hand is then:

Question 16.3. Is there a well-defined second-order approximation to a smooth manifold centered at any point?

We investigate this question in this lecture, in part as motivation for the further general developments we undertake in future lectures. Lacking the proper apparatus of definitions and infrastructure, our treatment in this lecture is necessarily somewhat heuristic. Since Question 16.3 is local, we do not worry about global issues and are content to use a local description of a smooth manifold. Our other motivation is to introduce first notions of curvature, here in the most classical contexts.

Remark 16.4. Recall from (14.35) that a symmetric bilinear form on a vector space has only simple numerical invariants, whereas a symmetric bilinear form on an inner product space has an eigenspace decomposition (15.8). The nonlinear situation is parallel: there is no well-defined second-order approximation to a smooth manifold embedded in affine space, whereas there is if the

manifold is embedded in Euclidean space. Said better: armed with ruler and compass we can make measurements which define curvature.

Curvature of a curve

(16.5) Smooth curves in affine space. Let V be a normed vector space and A an affine space over V . (For the first part of the discussion we do not restrict A to be 2-dimensional.) The following is a provisional definition.

Definition 16.6. Let $a < b$ be real numbers. A smooth map $\gamma: (a, b) \rightarrow A$ is an *injective immersion* if γ is injective and $\frac{d\gamma}{dt}(t) \neq 0$ for all $t \in (a, b)$.

The image $C \subset A$ of an injective immersion is our model for a smooth curve.

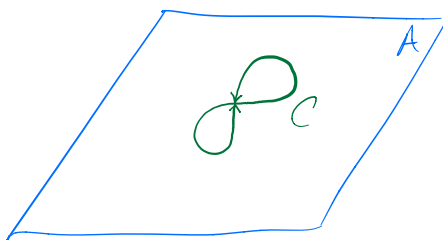


FIGURE 25. An injective immersion which is not an embedding

Remark 16.7. It is desirable to characterize those subsets $C \subset A$ which are smooth curves. The existence (locally) of parametrizing injective immersions is not quite enough, since we want to disallow subsets such as those depicted in Figure 25. That planar subset is the image of an injective immersion, but the immersion is not a homeomorphism onto the image (in the subset topology induced from A .) An injective immersion which is a homeomorphism onto its image is called an *embedding* and the image is called a *submanifold*.

Since we are interested in curves, and not motions, we want notions which do not depend on the parametrization.

Definition 16.8. Let $\gamma: (a, b) \rightarrow A$ be an injective immersion. A *reparametrization* of γ is a pair $a' < b'$ of real numbers and an injective map $t: (a', b') \rightarrow (a, b)$ such that $\frac{dt}{ds}(s) \neq 0$ for all $s \in (a', b')$.

The function t is either monotonic increasing or monotonic decreasing. Think of t as a change of clock, a map from the s -clock to the t -clock. The composition $\gamma \circ t: (a', b') \rightarrow A$ is the reparametrized curve, a new motion which traces out the same curve C according to the s -clock. Fix $p \in C$ and suppose $p = \gamma(t_0)$ and $t_0 = t(s_0)$ for $t_0 \in (a, b)$ and $s_0 \in (a', b')$; see Figure 26.

(16.9) First-order approximation. As stated above, there is a well-defined first-order approximation, which in this case is the tangent line. The affine linear approximation to γ at $t = t_0$ is the function $L: \mathbb{R} \rightarrow A$ given by

$$(16.10) \quad L(t) = p + t \frac{d\gamma}{dt}(t_0).$$

The affine linear approximation to $\gamma \circ t$ at $s = s_0$ is

$$(16.11) \quad \begin{aligned} L'(s) &= p + s \frac{d(\gamma \circ t)}{ds}(s_0) \\ &= p + s \frac{dt}{ds}(s_0) \frac{d\gamma}{dt}(t_0). \end{aligned}$$

The velocity vectors in the s - and t -clocks are proportional (with constant of proportionality $dt/ds(0)$), and so the images of L and L' agree: they are the same affine line in A . This *affine tangent line* is the first-order approximation to C at p .

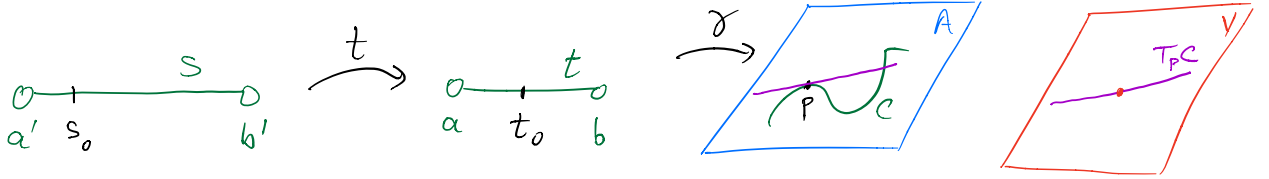


FIGURE 26. The tangent line is independent of the parametrization

Remark 16.12. The *tangent line* $T_p C \subset V$ is defined as the *linear* subspace of V which is the image of dL or, equivalently, the image of dL' . The tangent construction is a map

$$(16.13) \quad C \rightarrow \mathbb{P}V$$

where $\mathbb{P}V$ is the set of all 1-dimensional subspaces of V . Sometimes (16.13) is called the *Gauss map* of C .

(16.14) Second-order approximation. The second-order Taylor series of γ at $t = t_0$ is the affine quadratic function

$$(16.15) \quad Q(t) = p + t \frac{d\gamma}{dt}(t_0) + \frac{t^2}{2} \left\{ \frac{d^2\gamma}{dt^2}(t_0) \right\},$$

whereas the second-order Taylor series of $\gamma \circ t$ at $s = s_0$ is

$$(16.16) \quad Q'(s) = p + s \frac{dt}{ds}(s_0) \frac{d\gamma}{dt}(t_0) + \frac{s^2}{2} \left\{ \frac{d^2t}{ds^2}(s_0) \frac{d\gamma}{dt}(t_0) + \frac{dt}{ds}(s_0) \frac{d^2\gamma}{dt^2}(t_0) \right\}.$$

Notice that the coefficient of the quadratic term in Q' , which is the acceleration in the s -clock, is a linear combination of the acceleration in the t -clock and the velocity in the t -clock. The image of Q' is not in general equal to the image of Q . Said differently, there is not a clear quadratic shape which approximates C near p .

Remark 16.17. In fact, the invariant theory referenced in Remark 16.4 shows that quadratic shapes in affine space only have discrete numerical invariants, no continuous geometric invariants.

(16.18) Unit speed parametrization. In view of our failure to find a well-defined second-order shape approximating a curve in affine space, we move to the Euclidean situation. Thus let $\langle -, - \rangle$ be an inner product on V and assume the norm is that induced by the inner product. Let E be an affine space over V , a Euclidean space. There is a distinguished parametrization of a curve $C \subset E$.

Definition 16.19. An injective immersion $f: (a', b') \rightarrow E$ is *unit speed* if $\left\| \frac{df}{ds}(s) \right\| = 1$ for all $s \in (a', b')$. In that case we say that f is a *unit speed parametrization* of its image.

Given an arbitrary injective immersion $\gamma: (a, b) \rightarrow E$ we ask for existence and uniqueness of reparametrizations $t: (a', b') \rightarrow (a, b)$ such that $\gamma \circ t$ is unit speed. Observe first that reparametrization by a translation (in time) preserves the unit speed condition, so we eliminate the translation symmetry by fixing $t_0 \in (a, b)$ and $s_0 \in \mathbb{R}$ and demand that

$$(16.20) \quad t(s_0) = t_0.$$

We do not fix a', b' , however. The unit speed condition on $\gamma \circ t$ is satisfied if and only if

$$(16.21) \quad \frac{dt}{ds}(s) = \frac{\pm 1}{\left\| \frac{d\gamma}{dt}(t(s)) \right\|}$$

for all s . (The immersion condition ensures that the right hand side makes sense.) We interpret the right hand side as a vector field $F(t)\partial/\partial t$ on (a, b) , where

$$(16.22) \quad F(t) = \frac{\pm 1}{\left\| \frac{d\gamma}{dt}(t) \right\|}.$$

For each sign the *ordinary differential equation* tells that $t: (a', b') \rightarrow (a, b)$ is a motion in (a, b) whose velocity at time s is the vector $F(t(s))\partial/\partial t$. Intuitively, there should be a unique solution with initial condition (16.20) and the solution should extend to a maximal s -time interval (a', b') .

Problem 16.23. Prove that each ordinary differential equation in (16.21) has a unique maximal solution with initial condition (16.20).

We take up the general theory of ordinary differential equations in the next few lectures. For the purposes of this lecture we assume the existence of unit speed parametrizations, and that after eliminating translations there are two unit speed parametrizations, one in each direction.

Remark 16.24. Let $\mathbb{E}^1$ denote the Euclidean line and consider the group Euc_1 of Euclidean symmetries of $\mathbb{E}^1$. It has a normal subgroup $\mathbb{R}$ of translations with quotient of order two: an element of Euc_1 which projects to the nonidentity element is a reflection on $\mathbb{E}^1$. Suppose $f: (a', b') \rightarrow E$ is a unit speed motion and $\varphi \in \text{Euc}_1$. Then $f \circ \varphi$ is also unit speed. Furthermore, every unit speed parametrization of the curve has this form, so the unit speed parametrizations form a right torsor for Euc_1 . (See Definition 2.22.)

(16.25) Quadratic approximation to a Euclidean curve and curvature. A motion in affine space has a very large group of reparametrizations (Definition 16.8), whereas the group of unit speed reparametrizations of a unit speed motion in Euclidean space is relatively small. This gives the possibility to construct interesting geometric invariants. This is a general principle: the smaller the symmetry group the more invariants.

Suppose the function t in Figure 26 a unit speed reparametrization. Then the velocities at position $p \in C$ in the two parametrizations either agree or are opposite and the accelerations are equal. Hence there are exactly two quadratic approximations (16.15) of a unit speed parametrization; they differ only in the sign of the linear term. In particular, the image of that affine quadratic function is independent of the unit speed parametrization.

Definition 16.26. Let $C \subset E$ be a curve in Euclidean space. The image of the second-order Taylor series of a unit speed parametrization is the *osculating parabola*.

The osculating parabola degenerates to the affine tangent line at points where the acceleration of a unit speed parametrization vanishes. The osculating parabola is the “quadratic shape”, or *affine quadric*, which best approximates C near a given point. There is a single parameter—the norm of the acceleration in a unit speed parametrization—which tells the shape of the parabola: it vanishes if the parabola degenerates to the affine tangent line and in general it tells how much the curve is “bending” (instantaneously at a point).

Definition 16.27. Let $C \subset E$ be a curve in Euclidean space and $p \in C$. The *curvature* of C at p is $\left\| \frac{d^2\gamma}{dt^2}(t_0) \right\|$ for a unit speed parametrization γ with $\gamma(t_0) = p$.

This curvature is a function $C \rightarrow \mathbb{R}^{\geq 0}$.

(16.28) A lemma about acceleration. In the case of a plane curve ($\dim E = 2$) we refine the curvature to a real-valued curvature which has a sign; see Definition 16.35 below. As a preliminary we observe the following.

Lemma 16.29. Suppose $\gamma: (a', b') \rightarrow E$ is a unit speed parametrization. Then the acceleration is perpendicular to the velocity.

After all, a nonzero tangential component of acceleration changes the speed.

Proof. Differentiate the unit speed condition

$$(16.30) \quad \left\langle \frac{d\gamma}{ds}, \frac{d\gamma}{ds} \right\rangle = 1$$

with respect to s . □

(16.31) *Curvature of a Euclidean plane curve.* Now assume $\dim E = \dim V = 2$. The *normal space* $N_p C \subset V$ to C at $p \in C$ is defined to be the orthogonal complement of $T_p C \subset V$. In this case it is a line.

Definition 16.32. A *coorientation* of C is a smooth choice of unit normal vector field $e_0: C \rightarrow V$.

So e_0 is a smooth function and $e_0(p) \in N_p C$ for all $p \in C$.

Remark 16.33. We have not defined the notion of a smooth function on C , so this definition—as befits this lecture—is not built on a solid infrastructure in this presentation. Certainly smoothness means that $e_0 \circ \gamma: (a, b) \rightarrow V$ is smooth for any parametrization γ , and the chain rule implies that if it is smooth for one choice of γ then it is smooth for all choices. (All we use in this lecture is C^1 , not C^∞ .) But we still must prove that such a smooth function exists, and in fact there are precisely two of them. This is a good exercise in the implicit function theorem, and should have been on your homework...

Remark 16.34. A coorientation is a coherent orientation of the normal spaces $N_p C$. Intuitively, it determines a “side” of C inside the plane E . We will discuss orientations of vector spaces in a future lecture. Here, in the Euclidean situation, there is a unique oriented unit vector in an oriented 1-dimensional inner product space.

Definition 16.35. Let γ be a unit speed parametrization of a cooriented smooth curve $C \subset E$ in a Euclidean plane. The *curvature* $\kappa: C \rightarrow \mathbb{R}$ is the unique function so that

$$(16.36) \quad \frac{d^2 \gamma}{ds^2} = \kappa e_0.$$

Lemma 16.29 shows that the acceleration is a multiple of e_0 , and the discussion in (16.25) shows that it is independent of the unit speed parametrization. The curvature is positive if C is “curving towards e_0 ” and negative if C is curving away from e_0 .

Remark 16.37. Acceleration has units L/T^2 , where L = length and T = time. A unit speed parametrization implies the existence of a constant which relates length and time, so setting $L = T$ we see that the left hand side of (16.36) has units $1/L$. Since e_0 is a unit vector it is dimensionless, from which we conclude that the curvature κ has units $1/L$. The same is true for the curvature in Definition 16.27.

Example 16.38. Let C be the circle of radius $r > 0$ centered at the origin in the Euclidean plane $\mathbb{E}_{x,y}^2$. A unit speed parametrization (of the circle minus a point) is

$$(16.39) \quad \begin{aligned} x &= r \cos(s/r) \\ y &= r \sin(s/r), \end{aligned}$$

where $0 < s < 2\pi r$. Then the acceleration is

$$(16.40) \quad \begin{aligned} \ddot{x} &= -\frac{1}{r} \cos(s/r) \\ \ddot{y} &= -\frac{1}{r} \sin(s/r), \end{aligned}$$

and so the curvature is the constant function $\pm 1/r$; the sign depends on the coorientation.

(16.41) *Moving frames and curvature as an infinitesimal rotation.* We give a few closely related reformulations of the curvature of a cooriented plane curve. This will be useful when we come to surfaces.

As before let $C \subset E$ be a cooriented plane curve with normal vector field $e_0: C \rightarrow V$. We complete e_0 to an orthonormal basis $\{e_0, e_1\}$ of V at each point of C by defining e_1 to be the velocity vector of a unit speed parametrization of C . There are two choices for e_1 which differ by a sign. Any choice gives a *moving frame*¹⁶ along C . Let $\mathcal{B}_O(V)$ denote the set of orthonormal bases of V . (It can be made into an (abstract) smooth manifold; its shape is the union of two circles.) Then $\{e_0, e_1\}: C \rightarrow \mathcal{B}_O(V)$. Composing with a unit speed parametrization we can take the derivative with respect to the unit speed parameter s . Since for any indices $i, j = 0, 1$ the inner product $\langle e_i, e_j \rangle$ is a constant function of s (equal to 0 or 1), we have

$$(16.42) \quad \left\langle \frac{de_i}{ds}, e_j \right\rangle + \left\langle e_i, \frac{de_j}{ds} \right\rangle = 0.$$

(Compare Lemma 16.29.) This leads to the skew-symmetry of the matrix-valued function in the matrix equation

$$(16.43) \quad \frac{d}{ds} \begin{pmatrix} e_0 & e_1 \end{pmatrix} = \begin{pmatrix} e_0 & e_1 \end{pmatrix} \begin{pmatrix} 0 & \kappa \\ -\kappa & 0 \end{pmatrix}.$$

A skew-symmetric matrix A “generates” rotations in the precise sense that $t \mapsto \exp(tA)$ is a group homomorphism from $\mathbb{R}$ to the group of 2×2 rotation matrices. In that sense the curvature is an infinitesimal rotation.

(16.44) *Interlude on directional derivatives.* In Definition 6.14 we defined the directional derivative of a function on an affine space A over a normed linear space V at a point $p \in A$ in the direction $\xi \in V$ by differentiating along the constant velocity motion with initial position p and velocity ξ . Then in Corollary 7.18 we proved that we can compute the directional derivative by differentiating along any motion with initial position p and initial velocity ξ . Thus if we have a function defined only along a curve $C \subset A$, then given $p \in C$ and $\xi \in T_p C$ we can give a plausible definition of the directional derivative at p in the direction ξ by differentiating along a parametrization of C . But this does not match our previous context since the function is not assumed to be

¹⁶Moving frames were introduced and used extensively and effectively by the great French geometer Elie Cartan. He uses the French term *repère mobile*.

defined in an open set. To safely apply our previous theory we need to prove that the function locally extends to open sets in A ; then the previous theory shows that the directional derivative is well-defined and does not depend on the extension. Again this is an application of the implicit function theorem, which can be deployed to locally “straighten out” the curve C .

As a matter of notation, if η is a vector field we use ‘ $D_\xi\eta$ ’ for the directional derivative of η in the direction ξ .

(16.45) *Another formula for the curvature of a cooriented plane curve.* With this understood, in (16.43) we write

$$(16.46) \quad \frac{de_0}{ds} = D_{e_1}e_0 = -\kappa e_1,$$

and so

$$(16.47) \quad \kappa = -\langle D_{e_1}e_0, e_1 \rangle.$$

Curvature of a surface in Euclidean 3-space

(16.48) *Definition of a surface.* Let V be a 3-dimensional real inner product space and E an affine space over V , so a 3-dimensional Euclidean space.

Definition 16.49. A subset $\Sigma \subset E$ is a *surface* if for every $p \in \Sigma$ there exists $r > 0$, a 2-dimensional affine space A , and open set $U \subset A$, and an injective immersion $f: U \rightarrow E$ such that $B_p(r) \cap \Sigma = f(U)$.

Here $B_p(r) \subset E$ is the open ball of radius r about p . The map f is an immersion if df_a is injective for all $a \in U$. The definition does not use the inner product on V , so applies to any affine space.

Remark 16.50. We can give a similar definition for a *submanifold* of any dimension lying in an affine space of any dimension. In particular, the definition applies to curves and avoids the problem flagged in Figure 25. We used Definition 16.6 to emphasize the parametrization with an eye towards the nearly unique unit speed parametrization introduced subsequently. For higher dimensional submanifolds, such as two-dimensional surfaces, there is no analogous canonical family of parametrizations.

(16.51) *The tangent space.* At each point p of a surface $\Sigma \subset E$ there is a well-defined two-dimensional tangent space $T_p\Sigma \subset V$. It is equal to the image of df_a for any local injective immersion f parametrizing Σ with $f(a) = p$. We must prove that this image is independent of the choice of f , something we do not undertake here.

Remark 16.52. One strategy is to give a definition of $T_p\Sigma$ which does not rely on a choice of f . For example, we can define $T_p\Sigma$ as the set of initial velocities of all motions $\gamma: (-\epsilon, \epsilon) \rightarrow E$ such that $\gamma(0) = p$ and $\gamma(t) \in \Sigma$ for all $t \in (-\epsilon, \epsilon)$. Then we must prove that $T_p\Sigma \subset V$ is a 2-dimensional linear subspace.

The normal line $N_p\Sigma$ is the orthogonal complement to $T_p\Sigma$. A *coorientation* of a surface is defined as in Definition 16.32 as a smooth normal vector field $e_0: \Sigma \rightarrow V$.

Remark 16.53. Locally a surface has two coorientations. But globally there may not be a coorientation; the classic example with no global coorientation is a Möbius band embedded in space.

(16.54) *The fundamental forms and curvature of normal sections.* We define two symmetric bilinear forms on each tangent space $T_p\Sigma$. The *first fundamental form* is the inner product on V restricted to the subspace $T_p\Sigma$:

$$(16.55) \quad \begin{aligned} I_p: T_p\Sigma \times T_p\Sigma &\longrightarrow \mathbb{R} \\ \xi_1, \xi_2 &\longmapsto \langle \xi_1, \xi_2 \rangle \end{aligned}$$

The second fundamental form is defined in terms of directional derivatives, motivated by (16.47):

$$(16.56) \quad \begin{aligned} II_p: T_p\Sigma \times T_p\Sigma &\longrightarrow \mathbb{R} \\ \xi_1, \xi_2 &\longmapsto -\langle D_{\xi_1} e_0, \xi_2 \rangle \end{aligned}$$

In fact, (16.21) gives us the interpretation of II_p on the diagonal. Namely, if $\xi \in T_p\Sigma$ is a nonzero vector, then e_0, ξ span a 2-dimensional subspace $W_\xi \subset V$ which intersects $T_p\Sigma$ in the span of ξ . The affine subspace $p + W_\xi \subset E$, which is a Euclidean plane, intersects Σ near p in a plane curve C_ξ with coorientation e_0 . Then the curvature of the plane curve $C_\xi \subset p + W_\xi$ at p is

$$(16.57) \quad \kappa(p; \xi) = \frac{II_p(\xi, \xi)}{I_p(\xi, \xi)}.$$

The intersection $\Sigma \cap (p + W_\xi)$ is called a *normal section* of Σ . The curvatures of normal sections were studied by Euler in 1760.

We stated, but have not yet proved, that the second fundamental form II_p is symmetric.

Proposition 16.58. *Let $\xi_1, \xi_2 \in T_p\Sigma$. Then*

$$(16.59) \quad -\langle D_{\xi_1} e_0, \xi_2 \rangle = \langle D_{\xi_1} \xi_2, e_0 \rangle = \langle D_{\xi_2} \xi_1, e_0 \rangle = -\langle D_{\xi_2} e_0, \xi_1 \rangle.$$

In particular, $II_p(\xi_1, \xi_2) = II_p(\xi_2, \xi_1)$.

To make sense of the two inner expressions we must extend ξ_1, ξ_2 to vector fields on Σ in a neighborhood of p . The normal component of the directional derivative is independent of the extension.

Proof. Choose a local parametrization $f: U \rightarrow E$ as in Definition 16.49, where recall $U \subset A$ for a 2-dimensional affine space A over a normed linear space X and $f(a) = p$ for some $a \in U$. There are unique $\eta_1, \eta_2 \in X$ such that $df_a(\eta_i) = \xi_i$, $i = 1, 2$. Extend ξ_i to the vector fields on the image of f which are the images of the constant vector fields η_i on U , and write e_0, ξ_1, ξ_2 as functions $U \rightarrow V$. Then we have

$$\begin{aligned}
 (16.60) \quad -\langle D_{\xi_1} e_0, \xi_2 \rangle &= -\langle \eta_1 \cdot e_0, \xi_2 \rangle \\
 &= -\eta_1 \langle e_0, \xi_2 \rangle + \langle e_0, \eta_1 \cdot \xi_2 \rangle \\
 &= \langle e_0, \eta_1 \cdot \xi_2 \rangle = \langle e_0, D_{\xi_1} \xi_2 \rangle \\
 &= \langle e_0, \eta_1 \cdot df(\eta_2) \rangle \\
 &= \langle e_0, d^2 f_a(\eta_2, \eta_1) \rangle
 \end{aligned}$$

The result now follows from the symmetry of $d^2 f_a$ (Theorem 13.34). $\square$

(16.61) *Shape operator and curvature.* As in (15.1) we construct a self-adjoint operator—the *shape operator*—from the first and second fundamental forms:

$$(16.62) \quad \Pi_p(\xi_1, \xi_2) = \mathbf{I}_p(\xi_1, S_p(\xi_2)), \quad \xi_1, \xi_2 \in T_p \Sigma.$$

Then S_p is diagonalizable (Corollary 15.7). Let the eigenvalues be $\lambda_1 \geq \lambda_2$. The normal curvature (16.57) is a function on $\mathbb{P}T_p \Sigma$ with maximum value λ_1 and minimum value λ_2 . (The projective space $\mathbb{P}T_p \Sigma$ is the space of lines in $T_p \Sigma$.) There are several possibilities, which we now enumerate.

If $\lambda_1 = \lambda_2$ the normal curvature is constant and we say that p is an *umbilic point*. If $\lambda_1 = \lambda_2 = 0$ then Σ is *flat* at p and is approximated to second order by its affine tangent plane. If $\lambda = \lambda_1 = \lambda_2 \neq 0$ then Σ is approximated to second order by a sphere of radius $1/\lambda$ passing through p with the same affine tangent plane as Σ .

If $\lambda_1 \neq \lambda_2$ then there is a decomposition $T_p \Sigma = L_1 \oplus L_2$ into an orthogonal sum of lines, where $S_p = \lambda_i$ on L_i . The L_i are called the *principal directions* at p and the λ_i the *principal curvatures*. It follows from (16.57) that the normal curvature on a line L which makes an angle θ with L_1 is

$$(16.63) \quad \lambda_1 \cos^2 \theta + \lambda_2 \sin^2 \theta,$$

a result known as *Euler's formula*.

If $\lambda_1 \geq 0 \geq \lambda_2$ and $\lambda_1 \neq \lambda_2$, then there are two points of $\mathbb{P}T_p \Sigma$ at which the normal curvature vanishes. These are called *asymptotic directions*.

In every case we define the *mean curvature* and *Gauss curvature* by the formulas

$$(16.64) \quad H = \frac{\lambda_1 + \lambda_2}{2}$$

$$K = \lambda_1 \lambda_2.$$

Gauss' *Theorema Egregium* (translation: Amazing Theorem!) states that K is *intrinsic* in the sense that it can be computed from the first fundamental form in a neighborhood of p without using the second fundamental form. This is truly a remarkable theorem and is the beginning of modern Riemannian geometry.

(16.65) *Second order approximation.* Finally, we can tell the second order approximation to Σ at p . Introduce an orthonormal system of affine coordinates x, y, z on E centered at p such that the x - and y -axes line in the affine tangent space to Σ at p and align with the lines L_1, L_2 ; the z -axis is the normal affine line at p . (At an umbilic point we are free to choose the x - and y -axes as arbitrary orthogonal lines in the affine tangent space.) Then the osculating quadric surface Q_p is described by the equation

$$(16.66) \quad z = \frac{\lambda_1}{2} x^2 + \frac{\lambda_2}{2} y^2.$$

It is nondegenerate if both λ_1 and λ_2 are nonzero, in which case it approximates well the surface near p . (Compare Example 15.11 and Theorem 15.23.)

Lecture 17: Fundamental theorem of ordinary differential equations

Setup and Motivation

(17.1) Vector fields and integral curves. Let V be a normed linear space and A an affine space over V . (We will require completeness of V to produce solutions to ordinary differential equations, but for the general setup it is not necessary.) Let $U \subset A$ be an open subset and $I \subset \mathbb{R}$ an open interval.

Definition 17.2.

- (1) A *vector field* on U is a continuous function $\xi: U \rightarrow V$.
- (2) A *time-varying vector field* on U is a continuous function $\xi: I \times U \rightarrow V$.
- (3) An *integral curve* of ξ is a C^1 function $f: J \rightarrow U$ on a subinterval $J \subset I$ such that

$$(17.3) \quad f'(t) = \xi(t, f(t)).$$

Equation (17.3) is the ordinary differential equation we seek to solve subject to the initial condition

$$(17.4) \quad f(t_0) = p_0$$

for some $t_0 \in I$ and $p_0 \in U$. We sometimes say that a vector field as in Definition 17.2(1) which does not depend on time is *static*. In Definition 17.2 we discuss continuous vector fields; as with other functions, vector fields may be C^k for some $k \in \mathbb{Z}^{>0} \cup \{\infty\}$, may be distributional, etc.

(17.5) Examples. Systems of ordinary differential equations can be put in the form (17.3). We give some examples for a single real-valued function.

Example 17.6. The equation

$$(17.7) \quad f'(t) = f(t)^3 + f(t)$$

for a function $f: J \rightarrow \mathbb{R}$ on some interval $J \subset \mathbb{R}$ has the form (17.3) where $U = A = \mathbb{R}$ with standard coordinate x and $\xi = (x^3 + x)\partial/\partial x$ is a static vector field.

Example 17.8. The static linear vector field $\xi = x\partial/\partial x$ on $\mathbb{R}$ has integral curves $x(t) = x_0 e^t$, where $f(0) = x_0$. For each initial position x_0 the integral curve exists for all time, i.e., has domain $\mathbb{R}$.

Example 17.9. For the quadratic vector field $\xi = x^2 \partial/\partial x$ the integral curve

$$(17.10) \quad x(t) = \frac{1}{1/x_0 - t}$$

with initial position $x_0 \neq 0$ blows up at finite (positive or negative) time. This illustrates why we allow subintervals in Definition 17.2(3): we do not expect to solve (17.3) on the entire time interval I . There is a maximal interval on which the solution is defined; for $x_0 > 0$ it is $(-\infty, 1/x_0)$.

Example 17.11. Let $g: \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function and $\xi = g(t) \partial/\partial x$ the time-varying vector field on $\mathbb{R}$ which at each time t is translation-invariant. Then the solution to (17.3) is

$$(17.12) \quad f(t) = \int_0^t g(s) ds + c$$

where $c = f(0)$ is the initial condition. The integral in (17.12) is the Riemann integral. In the generality of Definition 17.2 we need the Riemann integral for V -valued functions, which we take up next.

The Riemann integral

We give a brief treatment and rely on the reader to prove elementary properties (such as the additivity $\int_a^b + \int_b^c = \int_a^c$ and the fact that the integral is a linear operator).

Let $a < b$ be real numbers and V a Banach space. Suppose $F: [a, b] \rightarrow V$ is continuous. We define the integral $\int_a^b F(t) dt \in V$ as a limit of Riemann sums. Fix $n \in \mathbb{Z}^{>0}$. Set

$$(17.13) \quad t_i = a + i \frac{b-a}{n}, \quad i = 1, \dots, n,$$

and

$$(17.14) \quad I_n = \sum_{i=1}^n F(t_i)(t_i - t_{i-1}) = \left(\sum_{i=1}^n F(t_i) \right) \frac{b-a}{n}.$$

Lemma 17.15. *The sequence $(I_n) \subset V$ is Cauchy.*

Proof. Since $[a, b]$ is compact, the function F is uniformly continuous. Hence given $\epsilon > 0$ choose $\delta > 0$ such that if $|t' - t| < \delta$ then $\|F(t') - F(t)\| < \epsilon/(b-a)$. Choose a positive integer N such that $1/N < \delta$. For $m, n \geq N$ we have partitions (17.13) of $[a, b]$, say $s_1, \dots, s_m$ and $t_1, \dots, t_n$. Make new partitions, each of cardinality mn , by repeating each s_i n times and each t_j m times. Then the distance between corresponding elements of these new partitions is at most $1/N$, so is less than δ , and therefore

$$(17.16) \quad \begin{aligned} \|I_n - I_m\| &= \left\| \left(\sum_{j=1}^n m F(t_j) \right) \frac{b-a}{mn} - \left(\sum_{i=1}^m n F(s_i) \right) \frac{b-a}{mn} \right\| \\ &\leq mn \frac{\epsilon}{b-a} \frac{b-a}{mn} = \epsilon. \end{aligned}$$

□

Since V is complete the sequence I_n converges.

Definition 17.17.

$$(17.18) \quad \int_a^b F(s) ds = \lim_{n \rightarrow \infty} I_n.$$

Proposition 17.19. *If F satisfies $\|F(s)\| \leq M$ for some $M > 0$ and all $s \in [a, b]$, then*

$$(17.20) \quad \left\| \int_a^b F(s) ds \right\| \leq M(b-a).$$

Proof. For each positive integer n we use the triangle inequality to estimate (17.14):

$$(17.21) \quad \|I_n\| \leq nM \frac{b-a}{n} = M(b-a).$$

Since the norm is continuous, the estimate holds for $\lim_{n \rightarrow \infty}$ as well. □

Proposition 17.22. *Fix $t_0 \in [a, b]$. Then the function $f: (a, b) \rightarrow V$ defined by*

$$(17.23) \quad f(t) = \int_{t_0}^t F(s) ds$$

is differentiable with derivative $f'(t) = F(t)$.

Proof. Fix $t \in (a, b)$. For $0 < |h| < \min(b-t, t-a)$ we have

$$(17.24) \quad \frac{f(t+h) - f(t)}{h} = F(t) + \frac{1}{h} \int_t^{t+h} (F(s) - F(t)) ds.$$

Using uniform continuity of F as above, given $\epsilon > 0$ choose $\delta > 0$ such that if $|s - t| < \delta$ implies $\|F(s) - F(t)\| < \epsilon$. Then if $0 < |h| < \delta$ we have

$$(17.25) \quad \left\| \frac{1}{h} \int_t^{t+h} (F(s) - F(t)) ds \right\| \leq \frac{1}{|h|} \epsilon |h| = \epsilon.$$

This implies

$$(17.26) \quad \lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h} = F(t). \quad \square$$

An application of the contraction mapping fixed point theorem

(17.27) *Local uniform Lipschitz continuity.* We need a stronger form of continuity.

Definition 17.28. Let $\xi: I \times U \rightarrow V$ be a time-varying vector field as in Definition 17.2. Then ξ is *locally uniformly Lipschitz in the second variable* if for all $(t_0, p_0) \in I \times U$ there exists a neighborhood $I' \times U' \subset I \times U$ and a real number $C > 0$ such that

$$(17.29) \quad \|\xi(t, q_1) - \xi(t, q_0)\| \leq C\|q_1 - q_0\|, \quad t \in I', \quad q_0, q_1 \in U'.$$

Both norms in (17.29) are in V . The Lipschitz estimate (17.29) in the second variable holds locally in space locally uniformly in time. We use this property to make the crucial estimate for the contraction property below.

The mean value inequality¹⁷ implies the following criterion for local uniform Lipschitz continuity.

Proposition 17.30. *Let $\xi: I \times U \rightarrow V$ be a continuous time-varying vector field. Assume the second partial differential $d\xi^2: I \times U \rightarrow \text{Hom}(V, V)$ exists and is continuous. Then ξ is locally uniformly Lipschitz in the second variable.*

Proof. Given $(t_0, p_0) \in I \times U$ choose a neighborhood $I' \times U'$ such that U' is convex and

$$(17.31) \quad \|d\xi_{(t,p)}^2\| \leq C, \quad (t, p) \in I' \times U',$$

where $C = 2\|d\xi_{(t_0, p_0)}^2\|$. Now apply the mean value inequality to the restriction of ξ to $\{t\} \times U'$. $\square$

(17.32) *The main theorem.* In brief, a local solution to (17.3) subject to the initial condition (17.4) exists and is unique. We take up more global existence and uniqueness theorems in the next lecture. We remark that completeness enters twice in the proof: for the existence of the integral and for the contraction mapping fixed point theorem.

As a first move we rewrite the differential equation (17.3) as the integral equation

$$(17.33) \quad f(t) = p_0 + \int_{t_0}^t ds \xi(s, f(s)),$$

obtained by integrating (17.3) and applying Proposition 17.22 and the initial condition (17.4). Given f the right hand side defines a new function. Restricting the domain and codomain of f suitably this defines a contraction on the function space.

¹⁷From Homework #4: Let A be an affine space over a normed linear space V . A set $U \subset A$ is *convex* if whenever $p_0, p_1 \in U$, then $tp_0 + (1-t)p_1 \in U$ for all $t \in [0, 1]$. Suppose $U \subset A$ is open and convex, B is an affine space over a normed linear space W , and $f: U \rightarrow B$ is differentiable. Assume there exists $C \in \mathbb{R}^{>0}$ such that $\|df_p\| < C$ for all $p \in U$. Prove that if $p_0, p_1 \in U$, then

$$\|f(p_1) - f(p_0)\| \leq C\|\xi\|, \quad p_1 = p_0 + \xi.$$

Theorem 17.34. *Let V be a Banach space, A an affine space over V , $U \subset A$ an open set, $\xi: I \times U \rightarrow V$ a continuous time-varying vector field which is locally uniformly Lipschitz in the second variable, and $(t_0, p_0) \in I \times U$. Then there exists a neighborhood $B \subset U$ of p_0 and $J \subset I$ an open interval containing t_0 such that there exists a unique integral curve $f: J \rightarrow B$ of ξ with $f(t_0) = p_0$.*

Proof. Choose $I' \times U' \subset I \times U$ a neighborhood of (t_0, p_0) such that F restricted to $I' \times U'$ is uniformly Lipschitz in the second variable with Lipschitz constant $C > 0$. Then choose a neighborhood $I'' \times U'' \subset I' \times U'$ such that

$$(17.35) \quad \|\xi(t, p)\| \leq M, \quad t \in I'', \quad p \in U'',$$

where $M = 2\|\xi(t_0, p_0)\|$. Choose $r > 0$ so that $\overline{B_r(p_0)} \subset U''$. Set $B = B_r(p_0)$. Finally, choose $\delta > 0$ such that $J = (t_0 - \delta, t_0 + \delta) \subset I''$ and

$$(17.36) \quad \delta < \min\left(\frac{r}{M}, \frac{1}{C}\right).$$

Now let X be the vector space of bounded continuous functions $J \rightarrow V$, and endow X with the sup norm. Then X is a Banach space. (See Theorem 4.4 and Problem #1 on Homework #3.) The affine space of bounded continuous functions $J \rightarrow A$ is then a complete metric space, as is the closed subspace $\overline{B_r(f_0)}$, where $f_0: J \rightarrow A$ is the constant function with value p_0 . (Note $B = B_r(p_0)$ is a ball in A whereas $B_r(f_0)$ is a ball in the space of bounded continuous functions $J \rightarrow A$.)

For $f \in \overline{B_r(f_0)}$ define $Kf: J \rightarrow A$ by

$$(17.37) \quad Kf(t) = p_0 + \int_{t_0}^t ds \xi(s, f(s)), \quad t \in J.$$

Then $Kf(t) \in \overline{B_r(f_0)}$ since by Proposition 17.19 we have

$$(17.38) \quad \|Kf(t) - p_0\| = \left\| \int_{t_0}^t ds \xi(s, f(s)) \right\| \leq |t - t_0| M \leq \delta M,$$

and $\delta M < r$ by (17.36). Taking the sup over t we conclude that $Kf \in \overline{B_r(f_0)}$, i.e.,

$$(17.39) \quad K: \overline{B_r(f_0)} \longrightarrow \overline{B_r(f_0)}.$$

To see that K is a contraction, for $f_1, f_2 \in \overline{B_r(f_0)}$ we estimate for $t \in J$ that

$$(17.40) \quad \begin{aligned} \|Kf_2(t) - Kf_1(t)\| &= \left\| \int_{t_0}^t ds \{ \xi(s, f_2(s)) - \xi(s, f_1(s)) \} \right\| \\ &\leq \int_{t_0}^t ds \|\xi(s, f_2(s)) - \xi(s, f_1(s))\| \\ &\leq |t - t_0| C \|f_2 - f_1\|_{\sup} \\ &\leq \delta C \|f_2 - f_1\|_{\sup}. \end{aligned}$$

The same then holds for the sup over $t \in J$. Now (17.36) implies that (17.39) is a contraction, and so there is a unique fixed point $f \in \overline{B_r(f_0)}$, i.e., a unique solution to (17.33), or equivalently a unique solution to the pair of equations (17.3) and (17.4). $\square$

Lecture 18: More on ODE; introduction to differential forms

Uniqueness and global solutions

(18.1) *Recollection of the setup.* Let V be a Banach space, A an affine space over V , and $U \subset A$ an open subset. Then a time-varying vector field on U over a time interval $I \subset \mathbb{R}$ is a continuous function $\xi: I \times U \rightarrow V$. The basic existence theorem, Theorem 17.34, asserts that if ξ is locally uniformly Lipschitz in the second variable, then for any $(t_0, p_0) \in I \times U$ there exists an interval $J \subset I$ containing t_0 and a neighborhood $B \subset U$ of p_0 such that there is a unique function $f: J \rightarrow B$ which satisfies

$$(18.2) \quad \begin{aligned} f'(t) &= \xi(t, f(t)) \\ f(t_0) &= p_0 \end{aligned}$$

(18.3) *Uniqueness of solutions.* The following result shows that any two solutions with codomain U agree on their common domain.

Proposition 18.4. *Let $f_1: J_1 \rightarrow U$ and $f_2: J_2 \rightarrow U$ be solutions to (18.2), where $J_1, J_2 \subset I$ are subintervals containing t_0 . Then $f_1(t) = f_2(t)$ for all $t \in J_1 \cap J_2$.*

Proof. Suppose that $f_1(t') \neq f_2(t')$ for some $t' \in J_1 \cap J_2$. Assume $t' > t_0$; if $t' < t_0$ the argument we give applies with minimal change. Set $T = \{t \in J_1 \cap J_2 : t > t_0 \text{ and } f_1(t) \neq f_2(t)\}$. Then T is open, nonempty, and bounded below. Hence it does not contain $t_1 = \inf T$, so $f_1(t_1) = f_2(t_1) = p_1$ for some $p_1 \in U$. Apply Theorem 17.34 with initial condition (t_1, p_1) to obtain a unique solution $g: (t_1 - \delta, t_1 + \delta) \rightarrow \overline{B_r(p_1)}$ with $g(t_1) = p_1$, for some $\delta, r > 0$. Assume, by shrinking δ if necessary, that both f_1 and f_2 map $(t_1 - \delta, t_1 + \delta)$ into $\overline{B_r(p_1)}$. The uniqueness statement in Theorem 17.34 implies that $g = f_1 = f_2$ on the domain $(t_1 - \delta, t_1 + \delta)$. Then $t_1 + \delta/2 \in T$, which contradicts $t_1 = \inf T$. $\square$

Corollary 18.5. *The uniqueness in Theorem 17.34 holds for solutions $f: J \rightarrow U$ to (18.2).*

Observe that the uniqueness in the theorem is for functions with codomain B , whereas here it is more general: for functions with codomain U .

(18.6) *Remark on functions.* As a preliminary to discussing global solutions, we remind that the graph Γ_f of a function $f: X \rightarrow Y$ is a subset of $X \times Y$. Indeed, that subset is formally the function. The graph satisfies the property that for each $x \in X$ there exists a unique $y \in Y$ such that $(x, y) \in \Gamma_f$. It makes sense to take the union of graphs of functions, but in general such a union is not the graph of a function. For convenience we notate such unions as the union of functions.

(18.7) *Global solutions.* Returning to the ordinary differential equation (18.2), define

$$(18.8) \quad \mathcal{F} = \{f: J \rightarrow U : t_0 \in J \text{ and } f \text{ solves (18.2) on } J\}.$$

Set

$$(18.9) \quad f_{\max} = \bigcup_{f \in \mathcal{F}} f.$$

The union is taken in $I \times U$.

Theorem 18.10. *The union $f_{\max}$ is a function with domain an open interval $J_{\max} \subset I$. Then $f_{\max}: J_{\max} \rightarrow U$ is the unique maximal solution to (18.2).*

Proof. If $(t, p_1), (t, p_2) \in f_{\max}$, then there exist solutions $f_1: J_1 \rightarrow U$ and $f_2: J_2 \rightarrow U$ such that $t \in J_1 \cap J_2$, $f_1(t) = p_1$, and $f_2(t) = p_2$. Proposition 18.4 implies $p_1 = p_2$. Hence $f_{\max}$ is a function. Define its domain to be $J_{\max} \subset I$. For any $t \in J_{\max}$ there exists a solution $f: J \rightarrow U$ with $t \in J$, and since $f_{\max} = f$ restricted to J it follows that $f_{\max}$ solves (18.2) at t , hence is a solution for all $t \in J_{\max}$. Furthermore, since J is open it contains a neighborhood of t , and since $J_{\max} \supset J$ it follows that $J_{\max}$ is open. Also, $J_{\max}$ is the union of intervals containing t_0 , so each $t \in J_{\max}$ is connected by a continuous path to t_0 . It follows that $J_{\max}$ is connected, hence is an open interval. The solution $f_{\max}$ is maximal by definition, and the uniqueness of solutions on $J_{\max}$ is Corollary 18.5. $\square$

(18.11) *Unit speed parametrization revisited.* In (16.18) we encountered the question (Problem 16.23) of finding a unit speed parametrization to a curve, and used that problem as one motivation to develop the basic theory of ordinary differential equations. Having done so, we revisit that problem.

Let V be an inner product space and E an affine space over V . Suppose $\gamma: (a, b) \rightarrow E$ is a C^1 injective immersion (Definition 16.6) defined on some open interval $(a, b) \in \mathbb{R}$. (We do not use injectivity of γ in what follows.) We seek a solution $t: J \rightarrow (a, b)$ of the ordinary differential equation (16.21) (with the $+$ sign), where $J \subset \mathbb{R}$ is an open interval. This can be formulated in our standard form (18.2) for the static vector field

$$(18.12) \quad \frac{1}{\left\| \frac{d\gamma}{dt}(t) \right\|} \frac{\partial}{\partial t}$$

on (a, b) . Choose¹⁸ the initial condition $t(0) = t_0$ for some $t_0 \in (a, b)$. Let $t: J_{\max} \rightarrow (a, b)$ be the maximal solution constructed in Theorem 18.10.

Proposition 18.13. $(a, b) \subset t(J_{\max})$.

¹⁸The choice of the $+$ sign and the time $s_0 = 0$ “breaks” the Euc_1 symmetry described in Remark 16.24.

Then $t: J_{\max} = (a', b') \rightarrow (a, b)$ is the desired reparametrization.

Proof. If not, let $S = \{x \in (a, b) : x \notin t(J_{\max})\}$. Then S is nonempty and bounded below. Let $x_0 = \inf S$. Then for some $\delta > 0$ there is a unique integral curve $\tilde{t}: (-\delta, \delta) \rightarrow U$ of (18.12) with $\tilde{t}(0) = x_0$. Since $\tilde{t}(-\delta/2) \in t(J_{\max})$ there exists a unique $s_0 \in J_{\max}$ such that $\tilde{t}(-\delta/2) = t(s_0)$. Paste the solutions $t, \tilde{t}$ to a solution

$$(18.14) \quad \begin{aligned} J_{\max} \cup (s_0, s_0 + 3\delta/2) &\longrightarrow U \\ s &\longmapsto \begin{cases} t(s), & s \in J_{\max}; \\ \tilde{t}(s - s_0 - \delta/2), & s \in (s_0, s_0 + 3\delta/2). \end{cases} \end{aligned}$$

The uniqueness theorem Corollary 18.5 ensures that this is a well-defined solution, which contradicts the maximality of $J_{\max}$. $\square$

Additional topics

(18.15) *Regret.* With additional time and its dual, energy, we would have liked to expound on some important topics:

- higher order ordinary differential equations
- smooth dependence of solutions on parameters
- theory of linear equations

Alas, we content ourselves with one example to illustrate how to reduce higher-order ordinary differential equations to first-order equations.

Example 18.16. Consider the equation

$$(18.17) \quad f''(t) = f(t)f'(t)^2 - 3tf(t)^3$$

for a function $f: J \rightarrow \mathbb{R}$, where $J \subset \mathbb{R}$ is some interval. The technique is to introduce a second function $g: J \rightarrow \mathbb{R}$ and to convert (18.17) to a system of first-order ODEs:

$$(18.18) \quad \begin{aligned} f'(t) &= g(t) \\ g'(t) &= f(t)g(t)^2 - 3tf(t)^3 \end{aligned}$$

This now has our standard form: it is the equation for an integral curve of the time-varying vector field

$$(18.19) \quad \xi = g \frac{\partial}{\partial f} + (fg^2 - 3tf^3) \frac{\partial}{\partial g}$$

on the standard 2-dimensional affine space $\mathbb{A}^2$ with coordinates f, g . (You may be more comfortable substituting ‘ x, y ’ for ‘ f, g ’!)

Motivating differential forms, part one

(18.20) *The differential of a real-valued function.* Let V be a normed linear space, A an affine space over V , $U \subset A$ an open set, and $f: U \rightarrow \mathbb{R}$ a smooth function. Then the differential $df: U \rightarrow V^*$ assigns a continuous linear functional $V \rightarrow \mathbb{R}$ to each point $p \in U$.

Definition 18.21. A *differential 1-form on U* , or simply *1-form*, is a smooth function $\alpha: U \rightarrow V^*$.

Introduce the notation

$$(18.22) \quad \begin{aligned} \Omega^0(U) &= \{f: U \rightarrow \mathbb{R}\} \\ \Omega^1(U) &= \{\alpha: U \rightarrow V^*\} \end{aligned}$$

where implicitly we mean smooth functions f and smooth 1-forms α . These are vector spaces under pointwise addition and scalar multiplication. The differential is a linear map

$$(18.23) \quad \Omega^0(U) \xrightarrow{d} \Omega^1(U)$$

The notation suggests that there are vector spaces $\Omega^q(U)$ for all $q \in \mathbb{Z}^{\geq 0}$ and, perhaps, that there is an extension of the differential to a linear map

$$(18.24) \quad \Omega^q(U) \xrightarrow{d} \Omega^{q+1}(U)$$

for each q . This is true, as we shall see presently. Our task here is to motivate such an extension.

Example 18.25. Let $A = \mathbb{A}^n$ be affine space with standard coordinates $x^1, \dots, x^n$. Then any 1-form on an open set $U \subset \mathbb{A}^n$ can be written

$$(18.26) \quad \alpha = \alpha_i(x^1, \dots, x^n) dx^i$$

for some smooth functions $\alpha_i \in \Omega^0(U)$, $i = 1, \dots, n$. In particular, if $f \in \Omega^0(U)$, then (see (9.32))

$$(18.27) \quad df = \frac{\partial f}{\partial x^i} dx^i.$$

(18.28) *Prescribing the differential.* In the general situation of (18.20) we ask the following: given $\alpha \in \Omega^1(U)$ does there exist $f \in \Omega^0(U)$ such that

$$(18.29) \quad df = \alpha?$$

In other words, can we prescribe the differential of a function arbitrarily? The uniqueness aspect of (18.29) is straightforward. Since this is an affine equation, the difference $f_1 - f_0$ between two

solutions satisfies the linear equation $dg = 0$. We proved in Theorem 7.29 that any such g is locally constant. Therefore, the affine space of solutions to (18.29), if nonempty, has tangent space the space of locally constant functions. But we have not shown existence. We remark that (18.29) is a first-order linear *partial* differential equation. The word ‘partial’ refers to the partial derivatives which appear when we write (18.29) in coordinates, as opposed to the ordinary derivatives which appear in the ordinary differential equation (18.2).

Remark 18.30. We can already observe that the existence and uniqueness theory of (18.29) is tied to the topology of U . For example, if U is an interval in $\mathbb{R}$, then there is a 1-dimensional vector space of locally constant functions, whereas if U is the union of k disjoint open intervals then the space of locally constant functions has dimension k . In other words, (18.29) detects the *connectivity* of U .

(18.31) *A necessary condition.* For concreteness and ease of notation specialize to $\mathbb{A}^2$ with standard coordinates x, y . Write

$$(18.32) \quad \alpha = P(x, y)dx + Q(x, y)dy$$

for functions $P, Q: U \rightarrow \mathbb{R}$. Then (18.29) is the system of equations

$$(18.33) \quad \begin{aligned} \frac{\partial f}{\partial x} &= P \\ \frac{\partial f}{\partial y} &= Q \end{aligned}$$

We can immediately see an obstruction to existence. For if f is C^2 , then by Theorem 13.34

$$(18.34) \quad \frac{\partial Q}{\partial x} = \frac{\partial}{\partial x} \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} \frac{\partial f}{\partial x} = \frac{\partial P}{\partial y}.$$

Equation (18.34) is a necessary condition for existence. It is not satisfied by every 1-form, for example not by $\alpha = xdy - ydx$.

(18.35) *The 2-form obstruction.* Since the necessary condition (18.34) involves first derivatives of the coefficients of α , we are motivated to express it directly in terms of a derivative of α . That is precisely what we contemplated in (18.24), but as of yet we have not defined that operator. Nonetheless, assuming the most basic properties—that d is linear and obeys a Leibniz rule—we compute from (18.32) as follows, simply juxtaposing 1-forms to indicate some as-of-yet-not-defined multiplication:

$$(18.36) \quad \begin{aligned} d\alpha &= d(Pdx + Qdy) \\ &= \left(\frac{\partial P}{\partial x} dx + \frac{\partial P}{\partial y} dy \right) dx + P d^2x \\ &\quad \left(\frac{\partial Q}{\partial x} dx + \frac{\partial Q}{\partial y} dy \right) dy + Q d^2y \end{aligned}$$

As desired, we see the relevant derivatives $\partial P/\partial y$ and $\partial Q/\partial x$ appearing, but there are 4 extraneous terms. They will be set to zero if we stipulate the following rules:

$$(18.37) \quad dx \wedge dy = -dy \wedge dx$$

$$(18.38) \quad d^2f = 0$$

for all functions f , in particular for the coordinate functions x and y . Here, in view of the odd commutativity rule (18.37) we change notation and write ‘ $\wedge$ ’ for the product of 1-forms. With this understood, (18.36) reduces to

$$(18.39) \quad d\alpha = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy.$$

The necessary condition (18.34) for a solution to (18.33) is now the equation

$$(18.40) \quad d\alpha = 0.$$

Said differently, $d\alpha$ is an obstruction to solving the equation (18.29).

Remark 18.41. Quite generally, the necessary condition (18.40) for a solution to (18.29) follows immediately by applying d and using the rule (18.38).

(18.42) *The road ahead.* This discussion reinforces our desire to define $\Omega^2(U)$ and the extension (18.24) of d , which then gives a sequence of linear maps

$$(18.43) \quad \Omega^0(U) \xrightarrow{d} \Omega^1(U) \xrightarrow{d} \Omega^2(U)$$

such that the composition is zero. Furthermore, in view of the definitions (18.22) we might anticipate constructing a new vector space, $\bigwedge^2 V^*$ so that

$$(18.44) \quad \Omega^2(U) = \{\omega: U \rightarrow \bigwedge^2 V^*\}.$$

We also have to incorporate the wedge product into our theory, which should be a pointwise operation on 1-forms, so a bilinear map

$$(18.45) \quad \wedge: V^* \times V^* \longrightarrow \bigwedge^2 V^*.$$

Anticipating higher degrees, we will construct the *exterior algebra* $\bigwedge V^*$ whose multiplication $\wedge$ satisfies the odd commutativity rule (18.37). We begin that task in the next lecture, after an additional motivation.

Lecture 19: Integration along curves and surfaces; universal properties

Motivating differential forms, part two

(19.1) *A Riemann sum for 1-forms.* Let V be a normed linear space, A an affine space over V , and $U \subset A$ an open set. Suppose $C \subset U$ is a “compact curve with boundary”. We have not defined this, so as a small variation of the treatment in (16.5) we consider C as the image of a C^1 injective immersion $\gamma: [a, b] \rightarrow U$, where $a < b$ are real numbers.¹⁹ Our goal is to define an integral which does not depend on the parametrization, so we will not use γ in any essential way. Let $\alpha: U \rightarrow V^*$ be a continuous 1-form on U . We claim that 1-forms are the natural objects to integrate over curves. To see why, approximate C by a piecewise affine curve as follows: choose points $p_1, p_2, \dots, p_{n+1}$ “in order”²⁰ along C . Define $\xi_i = p_{i+1} - p_i \in V$ to be the displacement vector from p_i to p_{i+1} . Then define the “Riemann sum”

$$(19.2) \quad \sum_{i=1}^n \alpha_{p_i}(\xi_i).$$

In essence this approximates C by a union of affine line segments $p_i p_{i+1}$ and replaces α on each line segment by the constant 1-form with value α_{p_i} . The formula is obtained from the stipulation that the integral of a constant 1-form along an affine line segment is the pairing of the 1-form with the displacement vector. In other words, to define $\int_C \alpha$ we use the pairing

$$(19.3) \quad V^* \times V \longrightarrow \mathbb{R}$$

on a piecewise constant approximation of the integrand and a piecewise constant approximation of the region of integration.

Remark 19.4. Most definitions of integration begin with the notion of an integral of a piecewise constant quantity and then introduce a limiting process.

(19.5) *Orientation.* Notice that (19.2) changes sign if we traverse the curve in the opposite direction. For then we sum in the opposite order—this is irrelevant since addition in $\mathbb{R}$ is commutative—but also each ξ_i is replaced by $-\xi_i$, and this does change the sign of the sum. Therefore, we need one

¹⁹By contrast with Definition 16.6, here the domain is not open, so we must say what γ is C^1 means. Rather than relying on 1-sided derivatives, which are special in dimension one, we say that $\gamma: [a, b] \rightarrow U$ is C^1 if it is the restriction of a C^1 function defined on some open set $U \subset \mathbb{R}$ which contains $[a, b]$.

²⁰Use γ as a crutch: choose $t_1 = a < t_2 < \dots < t_{n+1} = b$ in $[a, b]$ and set $p_i = \gamma(t_i)$.

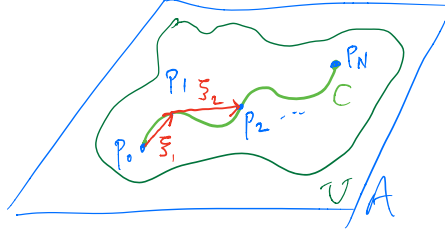


FIGURE 27. Integration of a 1-form along a curve

more piece of data for the integral to be well-defined: an *orientation* of C . In this 1-dimensional case an orientation is a “choice of direction” on the curve. In general, to integrate differential forms in any dimension we need to orient the region over which we are integrating. We develop an algebraic theory of orientations in a future lecture.

(19.6) 2-forms and surfaces. Now let $\Sigma \subset U$ be a surface (Definition 16.49). We illustrate that, modulo the question of orienting Σ , the natural object to integrate over Σ is a differential 2-form. We gave a different motivation for introducing 2-forms in (18.35). We indicated in (18.42) that there is a vector space $\bigwedge^2 V^*$ constructed algebraically from V and that a 2-form is a continuous map $\omega: U \rightarrow \bigwedge^2 V^*$. Now we must see why it is natural to integrate a 2-form over the surface.

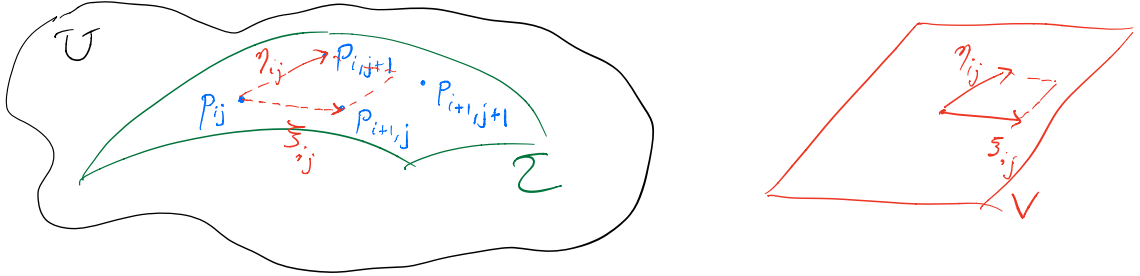


FIGURE 28. Piecewise parallelogram approximation of a surface

Following the strategy of (19.1) choose a grid of points p_{ij} in Σ ($1 \leq i, j \leq n+1$) as depicted in Figure 28. For example, we can suppose that we have a global parametrization by an injective immersion from a square, and we take the p_{ij} to be the image of a lattice in the square. Then for each $i, j \in \{1, \dots, n\}$ approximate a piece of Σ by the “affine parallelogram” with vertices

$$(19.7) \quad p_{ij}, \quad p_{ij} + \xi_{ij}, \quad p_{ij} + \eta_{ij}, \quad p_{ij} + \xi_{ij} + \eta_{ij},$$

where

$$(19.8) \quad \begin{aligned} \xi_{ij} &= p_{i+1,j} - p_{ij} \\ \eta_{ij} &= p_{i,j+1} - p_{ij} \end{aligned}$$

The 2-dimensional version of the Riemann sum (19.2) is

$$(19.9) \quad \sum_{i,j=1}^n \omega_{p_{ij}}(\xi_{ij} \wedge \eta_{ij}),$$

where the wedge product $\xi_{ij} \wedge \eta_{ij}$ of vectors represents the linear parallelogram in V spanned by ξ_{ij} and η_{ij} . This product lives in a vector space $\bigwedge^2 V$, as we now explain, and the pairing in (19.9) is

$$(19.10) \quad \bigwedge^2 V^* \times \bigwedge^2 V \longrightarrow \mathbb{R},$$

analogous to (18.3).

Remark 19.11. Just as the integral of a 1-form along a curve requires an orientation of the curve, so too does the integral of a 2-form over a surface. An orientation in two dimensions is heuristically a coherent sense of rotation, which here determines the ordering of the sides ξ_{ij}, η_{ij} of each parallelogram.

(19.12) Wedge products of vectors and parallelograms. Recall that the wedge product of *covectors*—elements of the dual space V^* —appeared in (18.45). We now explain how, intuitively, the wedge product of *vectors*

$$(19.13) \quad \begin{aligned} \wedge : V \times V &\longrightarrow \bigwedge^2 V \\ \xi_1, \xi_2 &\longmapsto \xi_1 \wedge \xi_2 \end{aligned}$$

represents the parallelogram spanned by two vectors in the vector space V . The product is bilinear and skew-symmetric in the sense that it satisfies the analog of (18.37) for vectors:

$$(19.14) \quad \xi_1 \wedge \xi_2 = -\xi_2 \wedge \xi_1, \quad \xi_1, \xi_2 \in V.$$

One consequence of the skew-symmetry is that if ξ_2 is a multiple of ξ_1 , then $\xi_1 \wedge \xi_2 = 0$. Then the parallelogram degenerates and so zero makes sense as its algebraic representative. Now suppose that ξ_1 and ξ_2 are linearly independent, so they span a 2-dimensional subspace $V' \subset V$. Change basis in V' by setting

$$(19.15) \quad \begin{aligned} \eta_1 &= A_1^1 \xi_1 + A_1^2 \xi_2 \\ \eta_2 &= A_2^1 \xi_1 + A_2^2 \xi_2 \end{aligned}$$

Then using (19.14), and its consequence $\xi_1 \wedge \xi_1 = \xi_2 \wedge \xi_2 = 0$, we find

$$(19.16) \quad \eta_1 \wedge \eta_2 = (A_1^1 A_2^2 - A_2^1 A_1^2) \xi_1 \wedge \xi_2 = \det(A) \xi_1 \wedge \xi_2,$$

where $A = (A_j^i)$ is the 2×2 matrix of coefficients in (19.15). This makes sense: if we have a notion of area on the 2-dimensional vector space V' , then the ratio of areas of the parallelogram $\|(\eta_1, \eta_2)$ and the parallelogram $\|(\xi_1, \xi_2)$ is the *absolute value* of the determinant of A . A 2-dimensional vector space does not come with a notion of area, yet the *ratio* of areas is defined. Furthermore, this *is* the ratio of areas for any choice of “area structure” on V' . Furthermore, the sign of the determinant tells the ratio of orientations of $\|(\eta_1, \eta_2)$ and $\|(\xi_1, \xi_2)$. Here again the ratio of orientations makes sense even though the individual orientations are not defined without a further choice.

Remark 19.17. The restriction of a 2-form $\omega \in \bigwedge^2 V^*$ to V' gives a notion of signed area—an orientation and a notion of area—as long as that restriction is nonzero. In that sense the sum (19.9) is a sum of signed areas, just as (19.2) is a sum of signed lengths.

Remark 19.18. If $\xi_1, \xi_2 \in V$ are linearly independent, then this discussion motivates representing the 2-dimensional subspace $V' = \mathbb{R}\{\xi_1, \xi_2\} \subset V$ as the line $\mathbb{R}(\xi_1 \wedge \xi_2) \subset \bigwedge^2 V$ spanned by their wedge product. The computation (19.16) shows that this line is independent of the choice of basis. So the construction factors to a map

$$(19.19) \quad Gr_2(V) \longrightarrow \mathbb{P}(\bigwedge^2 V)$$

from the Grassmannian of 2-dimensional subspaces of V to the projective space of the (yet-to-be-defined) vector space $\bigwedge^2 V$. The map (19.19) is the *Plücker embedding*.

(19.20) Summary of motivation. Our two motivations for differential forms and exterior algebra—(i) the obstruction to solving (18.29) and (ii) integration in dimension one (19.2) and dimension two (19.9)—are meant to give a geometric picture of the abstract algebra to which we turn shortly. The wedge products (18.45) and (19.13) to which we are led are characterized by skew-symmetry (18.37) and (19.14). We did not define the codomains—the second exterior power—and for our geometric applications it is only the skew-symmetry which is relevant, not any detailed construction. In other words, we take the second exterior power to be the *universal* codomain for a skew-symmetric product. It is this universality which we study later in this lecture. When we come to the wedge product, we construct the entire exterior *algebra* at once and then extract the homogeneous pieces from the whole algebra. The geometric heuristics give a picture of both the wedge product of vectors (parallelograms and higher dimensional parallelepipeds) and of covectors (signed area and higher dimensional signed volume), and they provide a counterpoint to the seemingly formal nature of the algebra to follow.

Bonus material: integration of 1-forms

(19.21) Plan. As an antidote to the preceding heuristics, in this section we give a rigorous treatment of the sum (19.2) as an approximation to an integral $\int_C \alpha$. The idea is to use a parametrization $\gamma: [a, b] \rightarrow U$ of C by a C^1 injective immersion, to “pull back” the 1-form α to a 1-form $f(t)dt$ on $[a, b]$, and then to compare the sum (19.2) to the Riemann sum (17.14) for the integral $\int_a^b f(t)dt$. To make that comparison we use the *uniform differentiability* of γ . So we begin with preliminary general discussions on pullback and uniform differentiability.

(19.22) Pullbacks of 1-forms. Let V, V' be normed linear spaces; A, A' affine spaces over V, V' , respectively; $U \subset A, U' \subset A'$ open sets; and $\varphi: U' \rightarrow U$ a C^1 map. Suppose first that $f: U \rightarrow \mathbb{R}$ is a continuous function. Then the *pullback* function $\varphi^*f: U' \rightarrow \mathbb{R}$ is the composition

$$(19.23) \quad \varphi^*f = f \circ \varphi.$$

The pullback of a 1-form is a bit more complicated. Namely, at each $p' \in U'$ the differential of φ is a continuous linear map

$$(19.24) \quad d\varphi_{p'}: V' \longrightarrow V.$$

Its dual is a continuous linear map

$$(19.25) \quad d\varphi_{p'}^*: V^* \longrightarrow (V')^*$$

defined by

$$(19.26) \quad d\varphi_{p'}^*(\theta)(\xi') = \theta(d\varphi_{p'}(\xi')), \quad \theta \in V^*, \quad \xi' \in V'.$$

Then if $\alpha: U \rightarrow V^*$ is a continuous 1-form on U , its pullback $\varphi^*\alpha: U' \rightarrow (V')^*$ is a continuous 1-form on U' defined by

$$(19.27) \quad d\varphi^*(\alpha)_{p'}(\xi') = \alpha_{\varphi(p')}(d\varphi_{p'}(\xi')), \quad p' \in U', \quad \xi' \in V'.$$

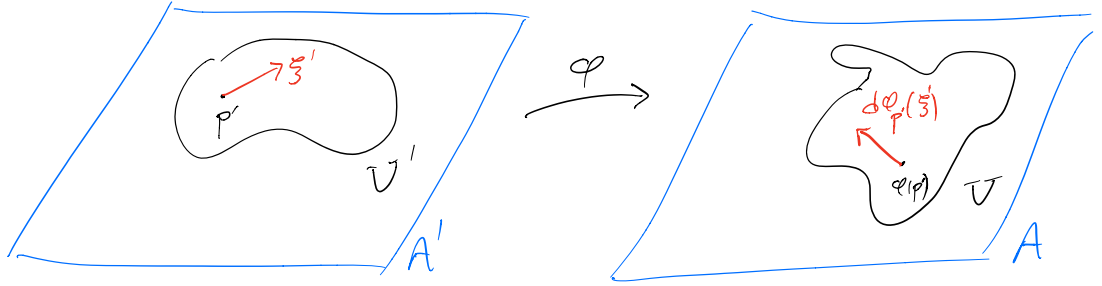


FIGURE 29. Pullback of a 1-form

Example 19.28. In practice, pullback is computed by blind substitution. Consider the 1-form

$$(19.29) \quad \alpha = y \, dx - x \, dy$$

on $\mathbb{A}_{x,y}^2$, and let $\gamma: \mathbb{R}_t \rightarrow \mathbb{A}_{x,y}^2$ be the map defined by

$$(19.30) \quad \begin{aligned} x &= t \\ y &= t^2 \end{aligned}$$

The left hand sides of (19.30) are shorthand for the pullbacks γ^*x, γ^*y of the coordinate functions. In a similar vein, blindly applying d we obtain the pullbacks $\gamma^*(dx), \gamma^*(dy)$ of their differentials:

$$(19.31) \quad \begin{aligned} dx &= dt \\ dy &= 2t \, dt \end{aligned}$$

Then the pullback $\gamma^*\alpha$ is computed by plugging (19.30) and (19.31) into (19.29):

$$(19.32) \quad \alpha = t^2 dt - t(2t \, dt) = -t^2 dt.$$

(19.33) *Uniform differentiability.* Compare the following to the definition of uniform continuity.

Definition 19.34. Let V, W be normed linear spaces; A, B affine spaces over V, W , respectively; $U \subset A$ an open set; $C \subset U$ a subset of U ; and $f: U \rightarrow B$ a C^1 function. Then $f|_C$ is *uniformly differentiable* if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $p \in C$, $\xi \in V$ such that $\|\xi\| < \delta$ and $p + \xi \in C$ the estimate

$$(19.35) \quad \|f(p + \xi) - f(p) - df_p(\xi)\| \leq \epsilon \|\xi\|$$

holds.

As usual, compactness implies uniformity.

Theorem 19.36. *In the situation of Definition 19.34 assume that W is complete and C is compact. Then $f|_C$ is uniformly differentiable.*

Proof. Since C is compact, the continuous function $df|_C: C \rightarrow \text{Hom}(V, W)$ is *uniformly* continuous. Hence given $\epsilon > 0$ choose $\delta > 0$ such that if $p \in C$, $\xi \in V$, $\|\xi\| < \delta$, and $p + \xi \in C$, then

$$(19.37) \quad \|df_{p+\xi} - df_p\| < \epsilon.$$

Define

$$(19.38) \quad \begin{aligned} g: [0, 1] &\longrightarrow B \\ t &\longmapsto f(p + t\xi) \end{aligned}$$

Then $g'(t) = df_{p+t\xi}(\xi)$. Applying the Riemann integral (Definition 17.17) and its basic properties, we deduce

$$(19.39) \quad \begin{aligned} \|f(p + \xi) - f(p) - df_p(\xi)\| &= \left\| \int_0^1 dt \{g'(t) - df_p(\xi)\} \right\| \\ &= \left\| \int_0^1 dt \{df_{p+t\xi}(\xi) - df_p(\xi)\} \right\| \\ &\leq \int_0^1 dt \|df_{p+t\xi}(\xi) - df_p(\xi)\| \\ &\leq \int_0^1 dt \|df_{p+t\xi} - df_p\| \|\xi\| \\ &\leq \int_0^1 dt \epsilon \|\xi\| \\ &= \epsilon \|\xi\|. \end{aligned}$$

□

(19.40) *Integral of a 1-form along a curve.* Resume the setup of (19.21), assuming that the underlying vector space V is complete. We also implicitly assume that γ is consistent with a chosen orientation of C . From (19.27) we compute

$$(19.41) \quad \gamma^* \alpha(t) = \alpha_{\gamma(t)}(\dot{\gamma}(t)) dt.$$

For $n \in \mathbb{Z}^{>0}$, as in (17.13) set

$$(19.42) \quad t_i = a + (i-1) \frac{b-a}{n}, \quad i = 1, \dots, n+1,$$

and let $p_i = \gamma(t_i)$. Let J_n denote the sum (19.2) and define the Riemann sum

$$(19.43) \quad I_n = \left(\sum_{i=1}^n \alpha_{\gamma(t_i)}(\dot{\gamma}(t_i)) \right) \frac{b-a}{n},$$

as in (17.14).

Proposition 19.44. *For $\epsilon > 0$ there exists $N \in \mathbb{Z}^{>0}$ such that if $n \geq N$ then $|J_n - I_n| < \epsilon$.*

Proof. The continuous function α is bounded on the compact set $\gamma([a, b])$, say $\|\alpha_{\gamma(t)}\| \leq M$ for all $t \in [a, b]$. By Theorem 19.36 γ is uniformly differentiable, so given $\epsilon > 0$ we can and do choose $\delta > 0$ so that if $|t' - t| < \delta$ and $t, t' \in [a, b]$, then

$$(19.45) \quad \|\gamma(t') - \gamma(t) - \dot{\gamma}(t)(t' - t)\| \leq \frac{\epsilon}{M(b-a)} |t' - t|.$$

Choose $N \in \mathbb{Z}^{>0}$ such that $\frac{b-a}{N} < \delta$. We estimate using (19.35) that if $n \geq N$ then

$$(19.46) \quad \begin{aligned} \left\| \sum_i \alpha_{\gamma(t_i)} \left(\gamma(t_{i+1}) - \gamma(t_i) - \dot{\gamma}(t_i) \frac{b-a}{n} \right) \right\| &\leq \sum_{i=1}^n \left\| \alpha_{\gamma(t_i)} \left(\gamma(t_{i+1}) - \gamma(t_i) - \dot{\gamma}(t_i) \frac{b-a}{n} \right) \right\| \\ &\leq \sum_{i=1}^n \|\alpha_{\gamma(t_i)}\| \frac{\epsilon}{M(b-a)} \frac{b-a}{n} \\ &\leq \epsilon. \end{aligned}$$

□

From Lemma 17.15 and Definition 17.17 we immediately deduce the following.

Corollary 19.47. $\lim_{n \rightarrow \infty} J_n$ exists and equals $\int_a^b f(t) dt$, where $\gamma^* \alpha = f(t) dt$.

This proves that the sums (19.2) converge, as long as the points p_i are suitably chosen, and so justifies calling the limit in Corollary 19.47 the integral $\int_C \alpha$.

Characterization by a universal property

(19.48) *Introduction.* One of the triumphs of mathematics is its conceptualization of structure. (This is only one of many triumphs.) Consider, for example, the definition/theorem about the real numbers: $\mathbb{R}$ is the complete ordered field. The three words ‘complete’, ‘ordered’, and ‘field’ are historically hard-won. Infinities had to be tamed in order to arrive at the first of these, and essential structures had to be extracted to formulate the latter two. Imagine a different, unlikely version of history in which these concepts are known, but the real numbers do not yet exist. Then we can wish them into existence by defining $\mathbb{R}$ as the complete ordered field. We could go on to prove all the properties we need from these three words, and in practice this is what one does in a first course in analysis. But one must prove existence by exhibiting a construction. By contrast, uniqueness, indicated by that innocuous looking ‘the’ which precedes the three magic words, is automatic. But what uniqueness? After all, there are two common constructions of the real numbers: Dedekind cuts and limits of Cauchy sequences. These construct unequal sets R_1, R_2 that we call $\mathbb{R}$. The uniqueness is the strongest possible: there are unique inverse maps $R_1 \xrightarrow{\sim} R_2$ which preserve the complete ordered field structures.²¹

(19.49) *The vector space generated by a set.* As a warmup to illustrate characterization by a universal property, suppose S is a set. We want to find a vector space F “generated” by S . Intuitively it should contain S and all elements and relations forced by the structure of a vector space, but should not contain more. So, for example, F has a zero vector, and it contains all finite linear combinations of elements of S . We could try to construct F along these lines, but instead we encode our specifications as follows.

Definition 19.50. Let S be a set. A pair (F, i) consisting of a vector space F and a map of sets $i: S \rightarrow F$ is a *free vector space generated by S* if for every (W, f) consisting of a vector space W and a map of sets $f: S \rightarrow W$ there exists a unique linear map $T: F \rightarrow W$ such that $f = T \circ i$.

The definition is summarized by the diagram

$$(19.51) \quad \begin{array}{ccc} S & \xrightarrow{i} & F \\ & \searrow f & \swarrow \exists! T \\ & & V \end{array}$$

The dashed line indicates that T is output whereas i and f are inputs. The symbol $\exists!$ indicates the existence of a unique map T . The two ways of traveling from S to V are assumed equal—one says “the diagram commutes”—which is the condition $f = T \circ i$. Notice that in the definition we use the article ‘a’ in front of ‘free vector space generated by S ’; uniqueness is a theorem (Theorem 19.57 below). Also, we do not assume that i is injective; that is also a theorem. Intuitively, the existence of T ensures that F is big enough and the uniqueness of T ensures that F is not too big. The

²¹ 0 and 1 are preserved, from which it follows that the map matches the copies of $\mathbb{Q}$ sitting in R_1 and R_2 . The requirement that it preserves completeness, say least upper bounds, determines the rest.

word ‘free’ evokes this middle ground. We say that (F, i) is *universal* among all pairs (W, f) , and Definition 19.50 spells out the precise universal property.

The power of the universal property is illustrated by the theorems which follow.

Example 19.52. If $S = \emptyset$ is the empty set, then the only choice for F is the zero vector space.

Example 19.53. If $S = \{s\}$ is a singleton, then define $F = \mathbb{R}$ and $i_k: S \rightarrow \mathbb{R}$ by $i_k(s) = k$, $k \in \mathbb{Z}$. The factorization problem (19.51) is solved uniquely by the linear map

$$(19.54) \quad T_k(x) = \frac{x}{k} f(s), \quad x \in \mathbb{R},$$

so long as $k \neq 0$. Hence $(\mathbb{R}, i_k)$ is a free vector space generated by S for all $k \in \mathbb{Z}^{\neq 0}$.

Example 19.55. Similarly, if $S = \{s_1, \dots, s_n\}$ is a finite set, define $F = \mathbb{R}^n$ and $i: S \rightarrow \mathbb{R}^n$ by $i(s_j) = e_j$, where $e_1, \dots, e_n$ is the standard basis of $\mathbb{R}^n$. The factorization is solved uniquely by

$$(19.56) \quad T(\xi^1, \dots, \xi^n) = \xi^j f(s_j).$$

Theorem 19.57. Let $(F_1, i_1), (F_2, i_2)$ be free vector spaces generated by a set S . Then there is a unique linear isomorphism $\varphi: F_1 \rightarrow F_2$ such that $i_2 = \varphi \circ i_1$.

Call a map $\varphi: F_1 \rightarrow F_2$ which satisfies $i_2 = \varphi \circ i_1$ a *morphism* from (F_1, i_1) to (F_2, i_2) . Then Theorem 19.57 asserts that any two solutions to the universal problem (19.51) are *unique up to unique isomorphism*. This is the strongest form of uniqueness for a problem whose answer is a set.²² In this circumstance we speak of *the* free vector space generated by S and introduce a special notation $F(S)$; the map i is implicit.

Proof. Apply (19.51) four times:

- (1) Use the universal property for (F_1, i_1) to construct $\varphi: F_1 \rightarrow F_2$:

$$(19.58) \quad \begin{array}{ccc} S & \xrightarrow{i_1} & F_1 \\ & \searrow i_2 & \swarrow \varphi \\ & F_2 & \end{array}$$

- (2) Use the universal property for (F_1, i_1) to construct $\psi: F_2 \rightarrow F_1$:

$$(19.59) \quad \begin{array}{ccc} S & \xrightarrow{i_1} & F_1 \\ & \searrow i_2 & \swarrow \psi \\ & F_2 & \end{array}$$

²²The pairs (F, i) which solve the universal problem (19.51) are the objects of a category in which there is a unique isomorphism between any two objects. This is the technical meaning of “unique up to unique isomorphism”.

(3) Use uniqueness in the universal property for (F_1, i_1) to prove $\psi \circ \varphi = \text{id}_{F_1}$:

$$(19.60) \quad \begin{array}{c} \begin{array}{ccc} & F_1 & \\ i_1 \nearrow & \downarrow \varphi & \searrow \text{id}_{F_1} \\ S & \xrightarrow{i_2} & F_2 \\ i_1 \searrow & \downarrow \psi & \nearrow \\ & F_1 & \end{array} \end{array}$$

(4) Use uniqueness in the universal property for (F_2, i_2) to prove $\varphi \circ \psi = \text{id}_{F_2}$:

$$(19.61) \quad \begin{array}{c} \begin{array}{ccc} & F_2 & \\ i_2 \nearrow & \downarrow \psi & \searrow \text{id}_{F_2} \\ S & \xrightarrow{i_1} & F_1 \\ i_2 \searrow & \downarrow \varphi & \nearrow \\ & F_2 & \end{array} \end{array}$$

Hence φ and ψ are inverse isomorphisms. $\square$

A *basis* of a vector space F is a subset $B \subset F$ such that every $\eta \in F$ can be written uniquely as a linear combination

$$(19.62) \quad \eta = c^1 \xi_1 + \cdots + c^n \xi_n$$

for a finite subset $\{\xi_1, \dots, \xi_n\} \subset B$ and scalars $c^1, \dots, c^n$.

Theorem 19.63. *Let (F, i) be the free vector space generated by a set S . Then $i(S) \subset F$ is a basis.*

Proof. Let $F' \subset F$ be the *span* of $i(S)$ and $i': S \rightarrow F'$ the inclusion. Then (F', i') satisfies the universal property, as follows from the existence and uniqueness of T in the diagram

$$(19.64) \quad \begin{array}{ccc} & F' & \\ i' \nearrow & \downarrow j & \\ S & \xrightarrow{i} & F \\ f \searrow & \downarrow T & \\ & W & \end{array}$$

Then the uniqueness Theorem 19.57 implies that the inclusion j is an isomorphism, so $F' = F$. This proves the existence of (19.62) for each η .

Next, we claim that the image of the restriction of i to every finite subset $S' \subset S$ is a linearly independent set in F ; this is equivalent to the uniqueness of (19.62) for each η . Suppose $S' \subset S$ has

cardinality n . Example 19.55 shows that the map $i': S' \rightarrow F(S')$ is injective. Use the universal property of $(F(S'), i')$ and (F, i) to construct T' and T in the diagram

$$(19.65) \quad \begin{array}{ccc} S' & \xrightarrow{i'} & F(S') \\ j \downarrow & \nearrow f & \uparrow T \\ S & \xrightarrow{i} & F \end{array} \quad \begin{array}{c} T' \\ \downarrow \\ T' \end{array}$$

where

$$(19.66) \quad f(s) = \begin{cases} i'(s'), & s' \in S'; \\ 0, & s \notin S'. \end{cases}$$

The diagram commutes. Now an argument similar to (3) in the proof of Theorem 19.57 shows $T \circ T' = \text{id}_{F(S')}$. In particular, T' is injective, which is the claim. $\square$

(19.67) Existence. We have still not proved existence of a free vector space generated by an infinite set. I leave that to the problem set.

Lecture 20: Tensor products, tensor algebras, and exterior algebras

(20.1) *The base field.* Throughout this lecture the base field can be arbitrary, though our applications of this algebra in this class only use vector spaces over the real numbers. A few cautions are necessary. When a field has characteristic 2, there is a difference between skew-symmetric and alternating maps; see (20.39). Also, the $\mathbb{Z}$ -grading on the tensor and exterior algebras using the action by invertible scalars. Over the reals or complexes that argument works directly; a modification (Remark 20.31) works in general.

Tensor products of vector spaces

The tensor product is the codomain for the universal bilinear map. We characterize it by a universal property which captures that universality.

Definition 20.2. Let V', V'' be vector spaces. A *tensor product* (X, b) is a vector space X and a bilinear map $b: V' \times V'' \rightarrow X$ such that for all pairs (W, B) of a vector space W and a bilinear map $B: V' \times V'' \rightarrow W$, there exists a unique linear map $T: X \rightarrow W$ such that $B = T \circ b$.

The definition is encoded in the commutative diagram

$$(20.3) \quad \begin{array}{ccc} V' \times V'' & \xrightarrow{b} & X \\ & \searrow B & \swarrow \exists! T \\ & W & \end{array}$$

The argument in Theorem 19.57 proves uniqueness up to unique isomorphism.

Theorem 20.4. *There exists a tensor product (X, b) of vector spaces V' and V'' .*

Proof. Let $F(V' \times V'')$ be the free vector space generated by $V' \times V''$. Let $R(V' \times V'')$ be the subspace generated by vectors

$$(20.5) \quad \begin{aligned} & (c'\xi'_1 + c^2\xi'_2, \xi'') - c^1(\xi'_1, \xi'') - c^2(\xi'_2, \xi''), \\ & (\xi', c^1\xi''_1 + c^2\xi''_2) - c^1(\xi', \xi''_1) - c^2(\xi', \xi''_2), \end{aligned}$$

for all choices of $\xi', \xi'_1, \xi'_2 \in V'$, $\xi'', \xi''_1, \xi''_2 \in V''$, and $c^1, c^2 \in \mathbb{R}$. Define the quotient vector space

$$(20.6) \quad X = F(V' \times V'') / R(V' \times V'')$$

and the composition

$$(20.7) \quad b: V' \times V'' \xrightarrow{i} F(V' \times V'') \longrightarrow X,$$

where i is the map in Definition 19.50 of the free vector space and the second map is the natural quotient map. (We can define the quotient vector space as satisfying a universal property with respect to that quotient map.) The relations (20.5) imply that b is bilinear. If (W, B) is as in Definition 20.2, then in the diagram

$$(20.8) \quad \begin{array}{ccccc} V' \times V'' & \xrightarrow{i} & F(V' \times V'') & \longrightarrow & X \\ & \searrow B & \downarrow \tilde{T} & \nearrow T & \\ & & W & & \end{array}$$

the unique map $\tilde{T}$ is the one in the universal property of the free vector space, and then the unique map T which completes the right triangle exists because of the universal property of the quotient. $\square$

(20.9) Notation for tensor product. The standard notation is $X = V' \otimes V''$ and $b(\xi', \xi'') = \xi' \otimes \xi''$ for $\xi' \in V'$ and $\xi'' \in V''$. Since the tensor product is unique up to unique isomorphism, we speak of *the* tensor product.

Theorem 20.10. *Let V', V'' be vector spaces with bases S', S'' .*

(1) *Every vector in $V' \otimes V''$ has a unique expression as a finite sum*

$$(20.11) \quad \sum_i \eta'_i \otimes \xi''_i, \quad \eta'_i \in V', \quad \xi''_i \in S''.$$

(2) *The set*

$$(20.12) \quad S = \{\xi' \otimes \xi'' : \xi' \in S', \xi'' \in S''\}$$

is a basis of $V' \otimes V''$.

Proof. Let $X \subset V' \otimes V''$ be the subspace of vectors (20.11). Define the bilinear map $b: V' \times V'' \rightarrow X$ by $b(\eta', \xi'') = \eta' \otimes \xi''$ for $\eta' \in V'$ and $\xi'' \in S''$. Since S'' is a basis of V'' this suffices to define the bilinear map b . Then (X, b) satisfies the universal property of the tensor product: construct a factorization in (20.3) using $V' \times V''$ and then restrict to the subspace X . By the uniqueness of the tensor product, the inclusion map is an isomorphism $X = V' \otimes V''$. This proves existence in (1). If there is not uniqueness, then for some vectors $\eta'_i \in V', \xi''_i \in S'', i = 1, \dots, N$, we have

$$(20.13) \quad \sum_{i=1}^N \eta'_i \otimes \xi''_i = 0.$$

Let $L: V' \rightarrow \mathbb{R}$ be a linear functional such that $L(\eta'_1) = 1$. Then

$$(20.14) \quad \begin{aligned} B: V' \times V'' &\longrightarrow V'' \\ \eta' , \eta'' &\longmapsto L(\eta')\eta'' \end{aligned}$$

is a bilinear map which sends equation (20.13) to the nontrivial linear relation $\sum L(\eta'_i)\xi''_i = 0$ among basis elements in S'' , which is absurd. This completes the proof of (1).

Assertion (2) is an immediate corollary: expand each $\eta'_i \in V'$ in (20.11) in the basis S' to write any vector in $V' \otimes V''$ uniquely as a linear combination of a finite subset of elements of S . $\square$

(20.15) *Commutativity and associativity of tensor product.* The tensor product satisfies commutative and associative “laws”, but rather than equalities of elements of a set these are isomorphisms between vector spaces. We write them on decomposable vectors for vector spaces V', V'', V''' as

$$(20.16) \quad \begin{aligned} V' \otimes V'' &\longrightarrow V'' \otimes V' \\ \xi' \otimes \xi'' &\longmapsto \xi'' \otimes \xi' \end{aligned}$$

and

$$(20.17) \quad \begin{aligned} (V' \otimes V'') \otimes V''' &\longrightarrow V' \otimes (V'' \otimes V''') \\ (\xi' \otimes \xi'') \otimes \xi''' &\longmapsto \xi' \otimes (\xi'' \otimes \xi''') \end{aligned}$$

These isomorphism satisfy “equations” known as the pentagon and hexagon diagrams, and tell that vector spaces with tensor product form a *symmetric monoidal category*, which is a “higher” version of an abelian group. We will not pursue this idea here, but will implicitly use the associativity. In particular, we use the notation ‘ $V' \otimes V'' \otimes V'''$ ’ for either²³ of the vector spaces in (20.17).

Remark 20.18. One has to work a bit to prove that tensor products of arbitrary finite collections of vector spaces are unambiguously defined, independent of ordering and of putting in parentheses.

Notation 20.19. For a vector space V write

$$(20.20) \quad \begin{aligned} \otimes^0 V &= \mathbb{R} \\ \otimes^1 V &= V \\ \otimes^2 V &= V \otimes V \\ \otimes^3 V &= V \otimes V \otimes V, \\ &\text{etc.} \end{aligned}$$

²³The correct categorical notion is that of a colimit or limit of the map (20.17)

Tensor algebra

(20.21) Algebras: basic definitions. A vector space has two operations: vector addition and scalar multiplication. An algebra has another binary operation called multiplication.

Definition 20.22.

- (1) An *algebra* is a vector space A , a bilinear map $m: A \times A \rightarrow A$, and an element $1 \in A$ such that multiplication m is associative and 1 is a unit for m . Write $a_1 a_2 = m(a_1, a_2)$ for the product of $a_1, a_2 \in A$.
- (2) A *homomorphism* $\varphi: A \rightarrow B$ of algebras is a linear map which preserves units and multiplication: $\varphi(1) = 1$ and $\varphi(a_1 a_2) = \varphi(a_1) \varphi(a_2)$ for all $a_1, a_2 \in A$.
- (3) A *subalgebra* of an algebra A is a linear subspace $A' \subset A$ which contains 1 and such that $a'_1 a'_2 \in A'$ for all $a'_1, a'_2 \in A'$.
- (4) A *2-sided ideal* $I \subset A$ is a linear subspace such that $AI = I$ and $IA = I$.
- (5) A $\mathbb{Z}$ -grading of an algebra A is a direct sum decomposition $A = \bigoplus_{k \in \mathbb{Z}} A^k$ such that $A^{k_1} A^{k_2} \subset A^{k_1 + k_2}$ for all $k_1, k_2 \in \mathbb{Z}$.
- (6) An algebra A is *commutative* if

$$(20.23) \quad a_1 a_2 = a_2 a_1, \quad a_1, a_2 \in A.$$

A $\mathbb{Z}$ -graded algebra is *(super)commutative* if

$$(20.24) \quad a_1 a_2 = (-1)^{k_1 k_2} a_2 a_1, \quad a_1 \in A^{k_1}, \quad a_2 \in A^{k_2}.$$

- (7) If A is a $\mathbb{Z}$ -graded algebra and $a \in A^k$, $k \in \mathbb{Z}^{>0}$, then a is *decomposable* if it is expressible as a product $a = a_1 \cdots a_k$ for $a_1, \dots, a_k \in A^1$. If not, we say a is *indecomposable*.

The associative law and identity law are $(a_1 a_2) a_3 = a_1 (a_2 a_3)$ and $1a = a1 = a$ for all $a, a_1, a_2, a_3 \in A$. If $I \subset A$ is a 2-sided ideal, then the quotient vector space A/I inherits a product, since the bilinear map $m: A \times A \rightarrow A$ factors to a bilinear map $\bar{m}: A/I \times A/I \rightarrow A/I$. For a $\mathbb{Z}$ -graded algebra we often use the notation $A^\bullet$ to signify the grading. A $\mathbb{Z}$ -grading is a structure, not a condition. Elements in the summand A^k are called *homogeneous of degree k* . An element which is not homogeneous is called *inhomogeneous*. The sign in (20.24) is referred to as the *Koszul sign rule*. We motivated it in (18.37) in the context of differential forms.

Example 20.25. The polynomial algebra $\mathbb{R}[x]$ is $\mathbb{Z}$ -graded but is not commutative as a $\mathbb{Z}$ -graded algebra; it is commutative as an ungraded algebra. The algebra of 2×2 matrices is noncommutative.

(20.26) The tensor algebra. The tensor algebra is the free algebra generated by a vector space; that is, there are no relations beyond those in Definition 20.22(1).

Definition 20.27. Let V be a vector space. A *tensor algebra* (A, i) over V is an algebra A and a linear map $i: V \rightarrow A$ such that for all pairs (B, T) consisting of an algebra B and a linear map $T: V \rightarrow B$ there exists a unique algebra homomorphism $\varphi: A \rightarrow B$ such that $T = \varphi \circ i$.

A commutative diagram encodes the universal property:

$$(20.28) \quad \begin{array}{ccc} V & \xrightarrow{i} & A \\ & \searrow T & \swarrow \exists! \varphi \\ & B & \end{array}$$

By the usual argument a tensor algebra is unique up to unique isomorphism, if it exists. It does, but before we construct it we deduce consequences of the universal property.

Theorem 20.29. *Let V be a vector space and (A, i) a tensor algebra of V .*

- (1) *The linear map i is injective.*
- (2) *There exists a canonical $\mathbb{Z}$ -grading on A .*

The meaning of ‘canonical’ is that if $L: V \rightarrow W$ is a linear map of vector spaces, then the induced algebra homomorphism $\varphi: A \rightarrow B$ between choices of tensor algebras for V and W , respectively, preserves the $\mathbb{Z}$ -gradings in the sense that $\varphi(A^k) \subset B^k$ for all $k \in \mathbb{Z}$.

Proof. Suppose $\xi \in V$ is nonzero and $i(\xi) = 0$. Define the algebra $\mathbb{R} \oplus \mathbb{R}\xi$ to have $\xi^2 = 0$. Choose a linear map $\pi: V \rightarrow \mathbb{R} \oplus \mathbb{R}\xi$ which is the identity on $\mathbb{R}\xi$. Define φ by the universal property

$$(20.30) \quad \begin{array}{ccc} V & \xrightarrow{i} & A \\ & \searrow \pi & \swarrow \varphi \\ & \mathbb{R} \oplus \mathbb{R}\xi & \end{array}$$

Then $0 \neq \xi = \pi(\xi) = \varphi i(\xi) = 0$. This proves i is injective.

For (2), if $\lambda \in \mathbb{R}$ is nonzero, let $T_\lambda: V \rightarrow V$ denote scalar multiplication by λ . The universal property (20.28) implies the existence of a unique extension to $\varphi_\lambda: A \rightarrow A$, an algebra homomorphism. Suppose $\lambda \neq 1$ and let $A^k \subset A$ denote the eigenspace of φ_λ with eigenvalue λ^k . Then $A' = \bigoplus_{k \in \mathbb{Z}} A^k \subset A$ is a subalgebra: it contains $1 \in A^0$ and is closed under multiplication. It comes to us as a $\mathbb{Z}$ -graded algebra. Also, $i(V) \subset A^1$. Then A' with the factored linear map $i: V \rightarrow A'$ satisfies the universal property: apply the universal property of A and restrict the resulting map to $A' \subset A$. So the inclusion $A' \hookrightarrow A$ is an isomorphism. We leave the reader to prove that the $\mathbb{Z}$ -grading is canonical. $\square$

Remark 20.31. We can consider all $\lambda \in \mathbb{R}^\times$ at once and then we are decomposing A under a representation of this multiplicative group. The argument fails over a finite field F . In that, or even the general, case we can extend scalars to the ring $F[x, x^{-1}]$ and consider the linear operator multiplication by x to construct the $\mathbb{Z}$ -grading on the tensor algebra. (In this argument we work with modules over a ring rather than vector spaces over a field.)

Remark 20.32. The tensor algebra is not commutative in either sense of Definition 20.22(6).

Theorem 20.33. *Let V be a vector space. Then a tensor algebra over V exists.*

Proof. Define

$$(20.34) \quad A = \bigoplus_{k=0}^{\infty} \otimes^k V,$$

where we use the notation introduced in (20.20). Apply Theorem 20.10(2), extended to tensor products of k vector spaces, to see that it suffices to define multiplication on decomposable vectors. Set

$$(20.35) \quad (\xi_1 \otimes \cdots \otimes \xi_{k_1})(\eta_1 \otimes \cdots \otimes \eta_{k_2}) = \xi_1 \otimes \cdots \otimes \xi_{k_1} \otimes \eta_1 \otimes \cdots \otimes \eta_{k_2},$$

$$\xi_1, \dots, \xi_{k_1}, \eta_1, \dots, \eta_{k_2} \in V.$$

If B is an algebra and $T: V \rightarrow B$ a linear map, define $\varphi: A \rightarrow B$ by

$$(20.36) \quad \varphi(\xi_1 \otimes \cdots \otimes \xi_k) = T(\xi_1) \cdots T(\xi_k), \quad \xi_1, \dots, \xi_k \in V,$$

and extend to be linear. It follows that φ is an algebra homomorphism, and that property forces (20.36), so it is unique. This proves the universal property. $\square$

The explicit construction implies both that the components of the tensor algebra A in negative degree vanish and that for $k \in \mathbb{Z}^{>0}$ the component in degree k is generated by the image of the k -linear map

$$(20.37) \quad \begin{aligned} V \times \cdots \times V &\longrightarrow A \\ \xi_1, \dots, \xi_k &\longmapsto i(\xi_1) \cdots i(\xi_k). \end{aligned}$$

Notation 20.38. We denote the $\mathbb{Z}$ -graded tensor algebra as ‘ $\otimes^\bullet V$ ’.

Exterior algebra

(20.39) *Alternation and skew-symmetry.* The characteristic property of the exterior algebra is the skew-symmetry of the product on vectors, as we motivated in (18.37) and (19.14). Skew-symmetry of the wedge product (19.14) on vectors is implied by the alternating property

$$(20.40) \quad \xi \wedge \xi = 0, \quad \xi \in V.$$

For if $\xi_1, \xi_2 \in V$, then expand $(\xi_1 + \xi_2) \wedge (\xi_1 + \xi_2)$ and use (20.40) three times to deduce (19.14). The converse is also true over fields of characteristic not equal to 2, such as the real or complex numbers.

Definition 20.41. Let V be a vector space. An *exterior algebra* (E, j) over V is an algebra E and a linear map $j: V \rightarrow E$ satisfying $j(\xi)^2 = 0$ for all $\xi \in V$ such that for all pairs (B, T) consisting of an algebra B and a linear map $T: V \rightarrow B$ satisfying $T(\xi)^2 = 0$ for all $\xi \in V$, there exists a unique algebra homomorphism $\varphi: E \rightarrow B$ such that $T = \varphi \circ j$.

Uniqueness up to unique isomorphism follows from the universal property. We prove existence by constructing the exterior algebra as a quotient of the tensor algebra.

Theorem 20.42. *Let V be a vector space. Then an exterior algebra over V exists.*

Proof. Let $\otimes V$ be the tensor algebra of V . Define $Q(V) \subset \otimes^2 V$ as

$$(20.43) \quad Q(V) = \{\xi \otimes \xi : \xi \in V\},$$

and let $I(V) \subset \otimes V$ be the 2-sided ideal generated²⁴ by $Q(V)$. Since $\otimes^{\geq 2} V \subset V$ is a 2-sided ideal, then so too is the intersection $I(V) \cap \otimes^{\geq 2} V$. But $Q(V) \subset I(V) \cap \otimes^{\geq 2} V \subset I(V)$, and since $I(V)$ is generated by $Q(V)$ it follows that $I(V) \subset \otimes^{\geq 2} V$.

Define $E = \otimes V / I(V)$ and let $j: V \xrightarrow{i} \otimes V \xrightarrow{q} E$ be the composition of the inclusion $\otimes^1 V \hookrightarrow \otimes V$ and the quotient map. We claim (E, j) is an exterior algebra. To prove the universal property, given (B, T) as in Definition 20.41, construct $\tilde{\varphi}$ in

$$(20.44) \quad \begin{array}{ccccc} V & \xrightarrow{i} & \otimes V & \xrightarrow{q} & E \\ & \searrow T & \downarrow \tilde{\varphi} & \swarrow \varphi & \\ & & B & & \end{array}$$

using the universal property for the tensor algebra. It factors through E since $qi(\xi)^2 = 0$ for all $\xi \in V$, which implies $\tilde{\varphi}(Q(V)) = 0$, and then finally $\tilde{\varphi}(I(V)) = 0$. Uniqueness of φ follows immediately from uniqueness of $\tilde{\varphi}$. $\square$

Notation 20.45. We denote the exterior algebra as ‘ $\bigwedge V$ ’ and use ‘ $\wedge$ ’ for the product.

Example 20.46. Let $V = 0$ be the zero vector space. Then $\otimes V = \mathbb{R}$ and $\bigwedge V = \mathbb{R}$. We leave the reader to deduce these assertions from the universal properties.

Example 20.47. Let $V = L$ be a line, i.e., a 1-dimensional vector space. Then we claim the algebra $E = \mathbb{R} \oplus L$ with $\ell^2 = 0$ for all $\ell \in L$ and the obvious inclusion $j: L \hookrightarrow E$ is an exterior algebra over L . Again this is a straightforward consequence of the universal property. Notice that $\otimes L$ is infinite dimensional.

²⁴ $I(V)$ is the intersection of all 2-sided ideals containing $Q(V)$.

Lecture 21: More on the exterior algebra; determinants

(21.1) *Context and remark.* As in the previous lecture, with a small modification (Remark 20.31) our account works over any field, but we write the ground field as the real numbers in view of our application to differential forms. Also, the vector space V on which we build the tensor and exterior algebras need not be finite dimensional. Later in this lecture we specialize to the finite dimensional case, and in particular discuss the determinant line.

Our treatment here leans heavily on universal properties to demonstrate their power. But that does involve a bit more abstraction than is necessary for the finite dimensional case that we use, so you may want to read lightly the general arguments and be confident about facility with computations and the geometric applications before focusing excessively on the algebraic aspects.

More about the exterior algebra: $\mathbb{Z}$ -grading and commutativity

(21.2) *The tensor algebra maps to the exterior algebra.* Our construction in the proof of Theorem 20.42 expresses the exterior algebra $\bigwedge V$ of a vector space V as a quotient of the tensor algebra $\bigotimes V$ by an ideal $I(V)$. In fact, it follows easily from the universal property of the tensor algebra (Definition 20.27) that there is such a homomorphism.

Proposition 21.3. *Let V be a vector space. Then there is a canonical homomorphism*

$$(21.4) \quad q: \bigotimes V \rightarrow \bigwedge V.$$

Proof. Factor the linear map $j: V \rightarrow \bigwedge V$ through the tensor algebra. □

(21.5) *Induced maps.* Let V, V' be vector spaces and $T: V' \rightarrow V$ a linear map. In the proof of Theorem 20.29 we used that T induces an algebra homomorphism on tensor algebras. We spell out the similar statement on exterior algebras. Namely, let i, i' denote the inclusions of V, V' into their exterior algebras. Then construct the algebra homomorphism $\bigwedge T$ in the diagram

$$(21.6) \quad \begin{array}{ccc} V' & \xrightarrow{i'} & \bigwedge V' \\ T \downarrow & & \downarrow \bigwedge T \\ V & \xrightarrow{i} & \bigwedge V \end{array}$$

via the universal property of the exterior algebra $\bigwedge V'$ applied to $i \circ T$; see Definition 20.41. We leave the reader to use the uniqueness in the universal property to prove that for a sequence $V'' \xrightarrow{T'} V' \xrightarrow{T} V$ of linear maps, the induced algebra homomorphisms on the exterior algebras satisfy $\bigwedge(T' \circ T) = \bigwedge T' \circ \bigwedge T$.

(21.7) $\mathbb{Z}$ -gradings. As in Theorem 20.29(2) for the tensor algebra, the exterior algebra admits a canonical $\mathbb{Z}$ -grading. The proof is similar.

Theorem 21.8. *Let V be a vector space. Then the exterior algebra $\bigwedge V$ is $\mathbb{Z}$ -graded. Furthermore, the homomorphism q preserves the $\mathbb{Z}$ -gradings. Also, if $T: V' \rightarrow V$ is a linear map, then the induced algebra homomorphism $\bigwedge T: \bigwedge V' \rightarrow \bigwedge V$ preserves the $\mathbb{Z}$ -gradings.*

Proof. For $\lambda \in \mathbb{R}$ let $T_\lambda: V \rightarrow V$ denote scalar multiplication by λ . The $\mathbb{Z}$ -grading on $\bigwedge V$ is constructed by choosing $\lambda \neq 0, 1$ and setting $\bigwedge^k V$ to be the eigenspace for $\bigwedge T_\lambda$ with eigenvalue λ^k . The subspace $j(V) \subset \bigwedge V$ is contained in $\bigwedge^1 V$. We leave the reader to check that the inclusion of V into $\bigoplus_{k=0}^\infty \bigwedge^k V$ satisfies the universal property, from which it follows that the inclusion of this direct sum into $\bigwedge V$ is an isomorphism.

The remaining assertions follow from the fact that scalar multiplication commutes with all linear maps. $\square$

(21.9) The $\mathbb{Z}$ -graded ideal. We can also deduce the $\mathbb{Z}$ -grading from the construction in the proof of Theorem 20.42. Recall the subset $Q(V) \subset \bigotimes^2 V$ defined in (20.43) and the 2-sided ideal $I(V) \subset \bigotimes V$ which it generates. Define $I^k(V) = I(V) \cap \bigotimes^k V$. The proof of Theorem 20.42 shows that $I^k(V) = 0$ for $k < 2$.

Theorem 21.10. *The ideal $I(V)$ is $\mathbb{Z}$ -graded in the sense that $I(V) = \bigoplus_{k=2}^\infty I^k(V)$.*

Proof. Let $\bigotimes T_\lambda: \bigotimes V \rightarrow \bigotimes V$ denote the homomorphism induced on the tensor algebra by scalar multiplication. We claim $\bigotimes T_\lambda$ maps $I(V)$ into itself. If $\lambda \neq 0$, then $T_{\lambda^{-1}}$ clearly preserves Q , and so $(\bigotimes T_{\lambda^{-1}})(I(V))$ is a 2-sided ideal containing Q . Since $I(V)$ is the smallest 2-sided ideal containing Q , we have $I(V) \subset (\bigotimes T_{\lambda^{-1}})(I(V))$, or equivalently $(\bigotimes T_\lambda)(I(V)) \subset I(V)$. The λ^k -eigenspace of the restriction of $\bigotimes T_\lambda$ to $I(V)$, for $\lambda \neq 1$, is by definition $I^k(V)$. Clearly $\bigoplus_{k=2}^\infty I^k(V) \subset I(V)$. On the

other hand, if $a \in I(V) \subset \bigotimes V$, then we can uniquely express $a = \sum_{k=2}^\infty a_k$ where $a_k \in \bigotimes^k V$ and only finitely many a_k are nonzero. Now for each $k \in \mathbb{Z}^{\geq 2}$

$$(21.11) \quad I(V) / I(V) \cap \bigotimes^{>k} V \subset \bigotimes V / \bigotimes^{>k} V$$

is a 2-sided ideal. Apply (21.11) for $k = 2$ to deduce $a_2 \in I(V)$, and then induct on k to show each $a_k \in I(V)$. $\square$

Corollary 21.12. *As a vector space, $\bigwedge^k V \cong \bigotimes^k V / I^k(V)$. In particular, we have $\bigwedge^0 V = \mathbb{R}$ and $\bigwedge^1 V = V$.*

The first statement follows from the definition of the $\mathbb{Z}$ -grading as an eigenspace decomposition for scalar multiplication. The last assertion follows since $I^k(V) = 0$ for $k = 0, 1$.

Corollary 21.13. *The image of the alternating k -linear map*

$$(21.14) \quad \begin{aligned} V \times \cdots \times V &\longrightarrow \bigwedge^k V \\ \xi_1, \dots, \xi_k &\longmapsto \xi_1 \wedge \cdots \wedge \xi_k \end{aligned}$$

generates $\bigwedge^k V$.

This follows from the corresponding statement (20.37) for the tensor algebra, which in turn follows from the explicit construction.

Remark 21.15. A k -linear map is *alternating* if it vanishes when two arguments are equal. In fact, (21.14) is the universal alternating k -linear map, in the same sense that the tensor product of two vector spaces is the universal bilinear map (Definition 20.2). That follows from the fact that (20.37) (with codomain $\bigotimes^k V$) is the universal k -linear map, which we could prove by developing the ideas in (20.15).

(21.16) *Commutativity.* Recall from Definition 20.22(6) that for a $\mathbb{Z}$ -graded algebra the definition of commutativity has a sign, the Koszul sign.

Theorem 21.17. *The $\mathbb{Z}$ -graded exterior algebra $\bigwedge^\bullet V$ over a vector space V is commutative.*

Proof. Corollary 21.13 implies that it suffices to check (20.24) for decomposable vectors. Let k, k' be positive integers and let $\xi_1, \dots, \xi_k, \xi'_1, \dots, \xi'_{k'}$ be vectors in V . Set $X = \xi_1 \wedge \cdots \wedge \xi_k$ and $X' = \xi'_1 \wedge \cdots \wedge \xi'_{k'}$. Then

$$(21.18) \quad \begin{aligned} X \wedge X' &= \xi_1 \wedge \cdots \wedge \xi_k \wedge \xi'_1 \wedge \cdots \wedge \xi'_{k'} \\ &= (-1)^{kk'} \xi'_1 \wedge \cdots \wedge \xi'_{k'} \wedge \xi_1 \wedge \cdots \wedge \xi_k \\ &= (-1)^{kk'} X' \wedge X, \end{aligned}$$

since in reordering we move k vectors past k' vectors, for a total of kk' transpositions. Each gives a minus sign according to the defining property of the exterior algebra; see (20.39). $\square$

Direct sums and tensor products

(21.19) *Direct sum via a universal property.* We have used the direct sum of vector spaces many times already in these lectures. Now we make explicit its universal property.

Definition 21.20. Let V_1, V_2 be vector spaces. A *direct sum* (S, i_1, i_2) is a diagram $V_1 \xrightarrow{i_1} S \xleftarrow{i_2} V_2$ which is universal in the sense that for any linear maps $j_i: V_i \rightarrow W$ there exists a unique linear map $T: S \rightarrow W$ which makes the diagram

$$(21.21) \quad \begin{array}{ccccc} V_1 & \xrightarrow{i_1} & S & \xleftarrow{i_2} & V_2 \\ & \searrow j_1 & \downarrow T & \swarrow j_2 & \\ & & W & & \end{array}$$

commute.

The direct sum is unique up to unique isomorphism and is denoted $V_1 \oplus V_2$.

(21.22) Tensor product of algebras. Let A_1, A_2 be algebras. In particular, they are vector spaces and so we can form the tensor product vector space $A_1 \otimes A_2$, as in Definition 20.2. To endow $A_1 \otimes A_2$ with an algebra structure, as in Definition 20.22(1), we must specify a bilinear map $m: A_1 \otimes A_2 \rightarrow A_1 \otimes A_2$ and a unit $1 \in A_1 \otimes A_2$. We can take $m = m_1 \otimes m_2$ and $1 = 1_1 \otimes 1_2$, where $m_i, 1_i$ define the algebra structure of A_i , $i = 1, 2$. This is the correct definition for ungraded algebras, but for $\mathbb{Z}$ -graded algebras there is a (Koszul) sign.

Definition 21.23. Let $A_1^\bullet, A_2^\bullet$ be $\mathbb{Z}$ -graded algebras with units $1_1, 1_2$. Endow the vector space $A_1 \otimes A_2$ with the $\mathbb{Z}$ -grading

$$(21.24) \quad (A_1 \otimes A_2)^k = \bigoplus_{\substack{k_1, k_2 \in \mathbb{Z} \\ k_1 + k_2 = k}} A_1^{k_1} \otimes A_2^{k_2}, \quad k \in \mathbb{Z}.$$

Endow $A_1 \otimes A_2$ with a $\mathbb{Z}$ -graded algebra structure: define the unit $1_1 \otimes 1_2$ and multiplication

$$(21.25) \quad (a_1 \otimes a_2)(a'_1 \otimes a'_2) = (-1)^{k_2 k'_1} a_1 a'_1 \otimes a_2 a'_2,$$

for $a_1, a'_1 \in A_1$, $a_2, a'_2 \in A_2$, where a_2 is homogeneous of degree k_2 and a'_1 is homogeneous of degree k'_1 .

Since multiplication is bilinear, it suffices to define it on decomposable vectors which are tensor products of homogeneous vectors.

Example 21.26. Let L_1, L_2 be lines. Recall the exterior algebras $\bigwedge^\bullet L_i = \mathbb{R} \oplus L_i$, $i = 1, 2$, as in Example 20.47. Let $\ell_i \in L_i$, $i = 1, 2$, be basis elements. The tensor product of the exterior algebras is the $\mathbb{Z}$ -graded algebra

$$(21.27) \quad (\bigwedge L_1 \otimes \bigwedge L_2)^\bullet = \mathbb{R} \oplus (L_1 \oplus L_2) \oplus (L_1 \otimes L_2)$$

supported in degrees 0, 1, 2. In this algebra, $\ell_1 \ell_2 = -\ell_2 \ell_1$ by (21.25).

(21.28) *Exterior algebra of a direct sum.* The passage from a vector space to its exterior algebra is an exponentiation in the sense that sums go over to products. (The same holds for tensor algebras.)

Theorem 21.29. *Let V_1, V_2 be vector spaces. Then there is a canonical isomorphism*

$$(21.30) \quad \bigwedge^\bullet (V_1 \oplus V_2) \xrightarrow{\cong} (\bigwedge V_1 \otimes \bigwedge V_2)^\bullet$$

of $\mathbb{Z}$ -graded algebras.

Proof. Let $j_i: V_i \rightarrow \bigwedge V_i$, $i = 1, 2$, be the inclusion into the exterior algebra. Define algebra homomorphisms $\psi_i: \bigwedge V_i \rightarrow \bigwedge V_1 \otimes \bigwedge V_2$, $i = 1, 2$, by

$$(21.31) \quad \begin{aligned} \psi_1(X_1) &= X_1 \otimes 1, & X_1 &\in \bigwedge V_1, \\ \psi_2(X_2) &= 1 \otimes X_2, & X_2 &\in \bigwedge V_2. \end{aligned}$$

Consider the diagram

$$(21.32) \quad \begin{array}{ccccc} V_1 & \xrightarrow{i_1} & V_1 \oplus V_2 & \xleftarrow{i_2} & V_2 \\ j_1 \downarrow & & j \downarrow & & \downarrow j_2 \\ \bigwedge V_1 & \xrightarrow{\psi_1} & \bigwedge V_1 \otimes \bigwedge V_2 & \xleftarrow{\psi_2} & \bigwedge V_2 \\ & \searrow \varphi_1 & \downarrow \varphi \quad T & \swarrow \varphi_2 & \\ & & B & & \end{array}$$

First, use the universal property of the direct sum (Definition 21.20) to construct j . Then we claim that the pair $((\bigwedge V_1 \otimes \bigwedge V_2)^\bullet, j)$ is an exterior algebra of $V_1 \oplus V_2$, where multiplication is defined on the tensor product with the Koszul sign (21.25). To check the universal property in Definition 20.41 suppose B is an algebra and $T: V_1 \oplus V_2 \rightarrow B$ in (21.32) satisfies $T(\xi_1 + \xi_2)^2 = 0$ for all $\xi_1 \in V_1$, $\xi_2 \in V_2$. Apply the universal property of the exterior algebra to $T \circ i_1$ to produce the (unique) algebra homomorphism φ_1 and to $T \circ i_2$ to produce the (unique) algebra homomorphism φ_2 . Finally, define the (unique) linear map $\varphi: \bigwedge V_1 \otimes \bigwedge V_2 \rightarrow B$ by applying the universal property of the tensor product (Definition 20.2) to the bilinear map

$$(21.33) \quad \begin{aligned} \bigwedge V_1 \times \bigwedge V_2 &\longrightarrow B \\ X_1, X_2 &\longmapsto \varphi_1(X_1) \varphi_2(X_2) \end{aligned}$$

We leave the reader to check that φ is an algebra homomorphism; it suffices to check on the tensor products of decomposable vectors in $\bigwedge V_1$ and $\bigwedge V_2$. $\square$

Finite dimensional exterior algebras and determinants

(21.34) *Applications of Theorem 21.29.* First, in Example 21.26 we see that (21.27) is isomorphic to $\bigwedge^\bullet(L_1 \oplus L_2)$. By induction on the dimension of a finite dimensional vector space, we leave the reader to deduce the following from Theorem 21.29.

Theorem 21.35. *Let n be a positive integer and suppose V is a vector space of dimension n .*

- (1) $\bigwedge^k V = 0$ if $k < 0$ or $k > n$.
- (2) $\dim \bigwedge^k V = \binom{n}{k}$, $0 \leq k \leq n$.
- (3) If $e_1, \dots, e_n$ is a basis of V , then for $1 \leq k \leq n$

$$(21.36) \quad \{e_{i_1} \wedge \cdots \wedge e_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$$

is a basis of $\bigwedge^k V$.

In (2) the expression $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient. In (3) the indices run over all strictly increasing ordered subsets of $\{1, \dots, n\}$ of cardinality k . If we write $V = L_1 \oplus \cdots \oplus L_n$ as a sum of lines, then one can prove by induction on $\dim V$ that

$$(21.37) \quad \bigwedge^k V \cong \bigoplus_{1 \leq i_1 < \cdots < i_k \leq n} L_{i_1} \otimes \cdots \otimes L_{i_k},$$

generalizing (21.27).

Corollary 21.38. *Let V be a vector space and $S = \{\xi_1, \dots, \xi_k\} \subset V$. Then S is linearly independent if and only if $\xi_1 \wedge \cdots \wedge \xi_k \in \bigwedge^k V$ is nonzero.*

Proof. Let $V' \subset V$ be the span of S ; then $\xi_1 \wedge \cdots \wedge \xi_k \in \bigwedge^k V'$. If S is not linearly independent, then $\dim V' < k$ in which case Theorem 21.35(1) implies $\bigwedge^k V' = 0$. Conversely, if S is a basis of V' , then Theorem 21.35(3) implies that $\xi_1 \wedge \cdots \wedge \xi_k$ is a basis of $\bigwedge^k V'$, hence is nonzero. $\square$

(21.39) *The determinant line.* According to Theorem 21.35(2) the top exterior power of a finite dimensional vector space is 1-dimensional.

Definition 21.40. Let n be a positive integer and suppose V is a vector space of dimension n . The *determinant line* of V is the 1-dimensional vector space

$$(21.41) \quad \text{Det } V = \bigwedge^n V.$$

The reason for the name will be apparent shortly. The following is a special case of Corollary 21.38.

Proposition 21.42. *Let V have dimension n . Then $\xi_1, \dots, \xi_n \in V$ is a basis if and only if the wedge product $\xi_1 \wedge \cdots \wedge \xi_n \in \text{Det } V$ is nonzero.*

(21.43) Determinant of a linear map. Suppose V, V' are finite dimensional vector spaces and $T: V' \rightarrow V$ a linear map.

Definition 21.44.

- (1) If $\dim V' \neq \dim V$, then define $\det T: \text{Det } V' \rightarrow \text{Det } V$ to be the zero map.
- (2) If $\dim V' = \dim V$, then define $\det T = \bigwedge^n T: \text{Det } V' \rightarrow \text{Det } V$.

Recall that $\bigwedge^n T$ is the map induced by T on the n^{th} exterior power; see **(21.5)**.

Proposition 21.45. $\det T \neq 0$ if and only if T is invertible.

Proof. If $\dim V' \neq \dim V$, then T is not invertible and $\det T = 0$. Assume $\dim V' = \dim V$ and $\xi'_1, \dots, \xi'_n \in V'$ is a basis. Proposition **21.42** implies that $\xi'_1 \wedge \dots \wedge \xi'_n$ is a basis of $\text{Det } V'$. By Definition **21.44**(2),

$$(21.46) \quad (\det T)(\xi'_1 \wedge \dots \wedge \xi'_n) = T\xi'_1 \wedge \dots \wedge T\xi'_n.$$

This is nonzero if and only if $T\xi'_1, \dots, T\xi'_n$ is a basis of V if and only if T is invertible. $\square$

We can give another proof using the composition law at the end of **(21.5)**. Namely, if T is invertible, then $\text{id}_{\text{Det } V'} = \det(T^{-1} \circ T) = \det T^{-1} \circ \det T$, which implies $\det T$ is nonzero.

(21.47) The numerical determinant. If V is finite dimensional and $T: V \rightarrow V$ is a linear operator, then $\det T: \text{Det } V \rightarrow \text{Det } V$ is a linear operator on a line, so is scalar multiplication. We identify the scalar with the numerical determinant of the linear operator T . The following computations show that this agrees with the usual determinant of matrices.

Example 21.48 (2-dimensional determinant). Suppose $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is represented by the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. This means that for the standard basis e_1, e_2 we have

$$(21.49) \quad \begin{aligned} Te_1 &= ae_1 + ce_2 \\ Te_2 &= be_1 + de_2 \end{aligned}$$

Hence

$$(21.50) \quad \begin{aligned} (\det T)(e_1 \wedge e_2) &= Te_1 \wedge Te_2 \\ &= (ae_1 + ce_2) \wedge (be_1 + de_2) \\ &= (ad - bc) e_1 \wedge e_2. \end{aligned}$$

Example 21.51 (n -dimensional determinant). Let n be a positive integer and $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ a linear operator. Write

$$(21.52) \quad Te_j = T_j^i e_i, \quad j = 1, \dots, n,$$

for n^2 numbers $T_j^i \in \mathbb{R}$. Then

$$\begin{aligned}
 (\det T)(e_1 \wedge \cdots \wedge e_n) &= Te_1 \wedge \cdots \wedge Te_n \\
 &= (T_1^{i_1} e_{i_1}) \wedge \cdots \wedge (T_n^{i_n} e_{i_n}) \\
 (21.53) \qquad &= \left\{ \sum_{\sigma \in \text{Sym}_n} \epsilon(\sigma) T_1^{\sigma(1)} \cdots T_n^{\sigma(n)} \right\} e_1 \wedge \cdots \wedge e_n,
 \end{aligned}$$

where Sym_n is the permutation group of $\{1, \dots, n\}$ and $\epsilon(\sigma) = \pm 1$ is the sign of the permutation. The homogeneous polynomial of degree n in braces is the usual expression for the determinant of a matrix.

Standard properties of the determinant, including the formula (21.53), are easily derived from the definition using exterior algebras.

Lecture 22: Orientation and signed volume

(22.1) *Introduction.* The problem of computing lengths, areas, and volumes dates from the beginnings of geometry in ancient Babylonia around 5000 years ago. Our treatment is more modern, but does not start from first principles. Rather, we tell the data needed to introduce a notion of n -dimensional volume in an n -dimensional real vector space. We do so in terms of the exterior algebra, and what naturally emerges is a notion of *signed* volume. In a later lecture we disentangle the sign; here we begin with the sign, which is the structure of an *orientation*. Along the way we also discuss duality in exterior algebra, and so realize abstract vectors in the exterior algebra of the dual space as alternating multilinear functionals.

Orientations

(22.2) *Intuition: dimension 1.* The notion of an orientation is familiar in low dimensions. There are two directions to traverse a curve, and an orientation is a choice between them. The linear version is a sense of direction on a 1-dimensional vector space L . This can be expressed as a choice of nonzero vector $e \in L$, but then if we multiply e by a *positive* scalar the resulting vector points in the same direction. So an orientation is a choice of nonzero vector up to positive multiple. If we consider the space $L \setminus \{0\}$ of nonzero vectors, then there are two components in the usual (norm) topology. An orientation $\mathfrak{o}$ is a choice of one component; see Figure 30.



FIGURE 30. Orientation of a line

(22.3) *Intuition: dimension 2.* Let V be a 2-dimensional vector space. Then an orientation is a sense of direction of rotation, a choice of clockwise vs. counterclockwise. A basis e_1, e_2 of V determines an orientation: the direction of rotation from e_1 to e_2 ; see Figure 31. If e'_1, e'_2 is another basis, and $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ the change-of-basis matrix defined by

$$(22.4) \quad e'_j = T_j^i e_i,$$

then e'_1, e'_2 and e_1, e_2 determine the same orientation if and only if $\det T > 0$. Recall the map

$$(22.5) \quad \begin{aligned} \mathcal{B}(V) &\longrightarrow \text{Det } V \setminus \{0\} \\ e_1, e_2 &\longmapsto e_1 \wedge e_2 \end{aligned}$$

where $\mathcal{B}(V)$ is the set of bases of V . The foregoing tells that two bases have the same orientation map if and only if they map to the same component of $\text{Det } V \setminus \{0\}$. Therefore, an orientation of V is equivalent to an orientation of $\text{Det } V$, a choice of component of $\text{Det } V \setminus \{0\}$, as in (22.2).

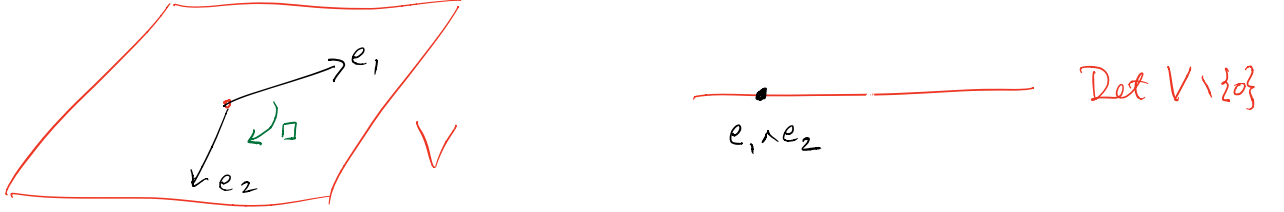


FIGURE 31. Orientation of a plane

(22.6) *Intuition: dimension 3.* An orientation of a 3-dimensional vector space is a choice of right-hand rule vs. a left-hand rule. So if e_1, e_2, e_3 is a basis of V , and the right-hand rule orientation is chosen, then the basis is positively oriented if when we take our right hand and point the fingers towards e_1 and rotate so that they curl in the direction of e_2 , then our thumb and e_3 should be on the same side of the e_1 - e_2 plane.

Definition 22.7.

- (1) Let V be a finite dimensional real vector space. An *orientation* of V is a choice $\mathfrak{o}$ of component of $\text{Det } V \setminus \{0\}$.
- (2) Suppose V, V' are finite dimensional, $\dim V = \dim V'$, and $\mathfrak{o}, \mathfrak{o}'$ are orientations of V, V' . An invertible linear map $T: V' \rightarrow V$ *preserves orientation* if $(\det T)(\mathfrak{o}') = \mathfrak{o}$. If instead $(\det T)(\mathfrak{o}') = -\mathfrak{o}$, then we say that T *reverses orientation*.

Remark 22.8. If $V = V'$, then $\det T \in \mathbb{R}^{\neq 0}$, and T preserves orientation if and only if $\det T > 0$. Notice that we do not need to choose an orientation on V to determine whether an automorphism $T: V \rightarrow V$ preserves or reverses orientation.

(22.9) *Bases and orientation.* If $\dim V = n$, then as in (22.5) there is a map

$$(22.10) \quad \begin{aligned} \mathcal{B}(V) &\longrightarrow \text{Det } V \setminus \{0\} \\ e_1, \dots, e_n &\longmapsto e_1 \wedge \dots \wedge e_n \end{aligned}$$

which takes a basis to a nonzero point of the determinant line. In this way a basis determines an orientation. Moreover, the inverse images of the two components of $\text{Det } V \setminus \{0\}$ partition the bases into two equivalence classes, the orbits of the action of the group $\text{GL}_n^+(\mathbb{R})$ of invertible $n \times n$ matrices with positive determinant.

Duality and exterior algebras

Let n be a positive integer and V an n -dimensional real vector space. The exterior powers of V and V^* form an array

$$(22.11) \quad \begin{array}{ccccccc} \mathbb{R} & V^* & \bigwedge^2 V^* & \cdots & \bigwedge^n V^* = \text{Det } V^* \\ \mathbb{R} & V & \bigwedge^2 V & \cdots & \bigwedge^n V = \text{Det } V \end{array}$$

There is a natural duality between the vector spaces in each column.

Proposition 22.12. *For all $k \in \mathbb{Z}^{>0}$ the pairing*

$$(22.13) \quad \begin{array}{c} \bigwedge^k V^* \times \bigwedge^k V \longrightarrow \mathbb{R} \\ \otimes^1 \wedge \cdots \wedge \theta^k, \xi_1 \wedge \cdots \wedge \xi_k \longmapsto \det(\theta^i(\xi_j))_{i,j} \end{array}$$

is nondegenerate

We have only specified the pairing on decomposable vectors in the exterior powers; it extends to all vectors using bilinearity. This determinant pairing identifies $\bigwedge^k V^*$ as the dual space to $\bigwedge^k V$. Compose with the alternating k -linear map (21.14) to identify $\bigwedge^k V^*$ as the space of k -linear alternating functions $V \times \cdots \times V \rightarrow \mathbb{R}$. Namely, if $\alpha \in \bigwedge^k V^*$, define

$$(22.14) \quad \begin{array}{c} \hat{\alpha}: V \times \cdots \times V \longrightarrow \mathbb{R} \\ \xi_1, \dots, \xi_k \longmapsto \langle \alpha, \xi_1 \wedge \cdots \wedge \xi_k \rangle, \end{array}$$

using the determinant pairing (22.13). We usually omit the carrot over α in (22.14) and simply identify the k -form α with this alternating k -linear map.

Proof. Let $e_1, \dots, e_n$ be a basis of V and $e^1, \dots, e^n$ the dual basis of V^* . Then by Theorem 21.35 we obtain a bases of $\bigwedge^k V$ and $\bigwedge^k V^*$. Introduce the multi-index notation $I = (i_1 \cdots i_k)$ for an increasing set of indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. Suppose $X = X^I e_I \in \bigwedge^k V$ lies in the kernel of (22.13). Writing the determinant pairing (22.13) as $\langle -, - \rangle$ we have

$$(22.15) \quad 0 = \langle e^J, X \rangle = X^J$$

for all multi-indices J . This implies $X = 0$ and proves nondegeneracy. $\square$

Example 22.16. For $k = 2$, if $\theta^1 \wedge \theta^2$ is a decomposable 2-form, the product of 1-forms $\theta^1, \theta^2 \in V^*$, then

$$(22.17) \quad (\theta^1 \wedge \theta^2)(\xi_1, \xi_2) = \theta^1(\xi_1)\theta^2(\xi_2) - \theta^1(\xi_2)\theta^2(\xi_1), \quad \xi_1, \xi_2 \in V.$$

This is an oft-used formula.

Signed volume

Definition 22.18. Let V be a finite dimensional real vector space. A *volume form* is a nonzero vector in $\text{Det } V^*$.

If A is affine over V , then a volume form on V determines a translation-invariant volume form on A . In linear geometry a volume form gives a notion of signed volume to parallelepipeds.

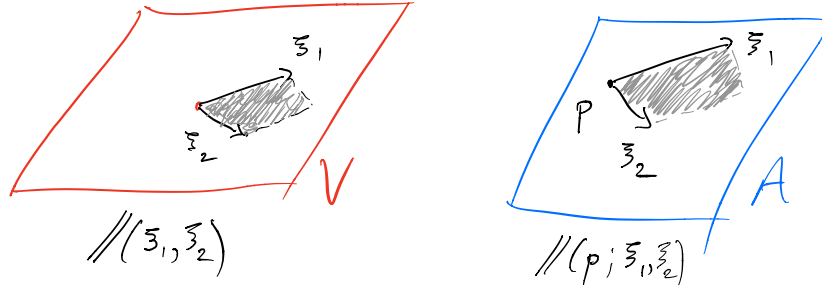


FIGURE 32. Parallelepipeds in a vector space and an affine space

Definition 22.19. Let V be a real vector space and A an affine space over V . A *k-dimensional parallelepiped* in V is the set of vectors

$$(22.20) \quad //(\xi_1, \dots, \xi_k) = \{t^i \xi_i : 0 \leq t^i \leq 1\} \subset V$$

for vectors $\xi_1, \dots, \xi_k \in V$. A *k-dimensional parallelepiped* in A is the set of points

$$(22.21) \quad //(p; \xi_1, \dots, \xi_k) = \{p + t^i \xi_i : 0 \leq t^i \leq 1\} \subset A$$

for a point $p \in A$ and vectors $\xi_1, \dots, \xi_k \in V$.

The parallelepiped is *nondegenerate* if $\xi_1, \dots, \xi_k$ are linearly independent; otherwise it is *degenerate*. There are many expressions for the same parallelepiped. In the vector case we can permute the vectors, and in the affine case we can also change the choice of vertex.

(22.22) Oriented parallelepipeds. If the parallelepiped is nondegenerate, then a choice of presentation which orders the vectors, as in (22.20) and (22.21), induces an orientation on the subspace spanned by the vectors, which we consider to be an orientation on the parallelepiped. Hence we speak of the *oriented parallelepiped* $//(\xi_1, \dots, \xi_k)$.

(22.23) Signed volume and volume. Let $\omega \in \bigwedge^k V^*$. Then for all k -dimensional subspaces $W \subset V$, the restriction of ω to W is either zero or is a volume form on W . If $\xi_1, \dots, \xi_k$ is a linearly independent set in V , define the *signed volume* of the oriented parallelepiped spanned as

$$(22.24) \quad \omega(\xi_1 \wedge \dots \wedge \xi_k).$$

Define the volume as

$$(22.25) \quad \text{Vol}(\llbracket(\xi_1, \dots, \xi_k)\rrbracket) = |\omega(\xi_1 \wedge \dots \wedge \xi_k)|.$$

Note the special case $\dim V = n$ and $\omega \in \text{Det } V^*$, which gives a notion of (signed) n -dimensional volume.

Remark 22.26. In a later lecture we give meaning to $|\omega|$, and in fact to a line $|\text{Det } V^*|$ of *densities* on V which give a notion of volume without defining signed volume.

(22.27) Standard choices. The vector space $\mathbb{R}^n$ has a standard orientation in which the standard basis $e_1, \dots, e_n$ is positively oriented. It has a standard volume form $e_1 \wedge \dots \wedge e_n$. The reader should check for $n = 1, 2, 3$ that (22.24) and (22.25) reproduce standard formulas for length, area, and volume.

Lecture 23: The Cartan exterior differential

Introduction

In this lecture we work in finite dimensions; we leave infinite dimensional Banach spaces to the homework. We have had several lectures on exterior algebra and some associated geometric concepts: orientation and volume. Now we return to calculus and define the exterior differential, which extends the differential on functions defined at the beginning of the semester. Recall that in (18.20) we motivated the introduction of the exterior differential on 1-forms, and computed a formula in two dimensions in (18.39). Now that we have constructed the exterior algebra, we are in a position to give a systematic treatment.

Exterior d

(23.1) Differential forms. Let V be a finite dimensional normed real vector space and A an affine space over V . Let $U \subset A$ be an open subset.

Definition 23.2. A (differential) k -form on U , $k \in \mathbb{Z}^{\geq 0}$, is a function $\alpha: U \rightarrow \bigwedge^k V^*$. The space of smooth k -forms on U is denoted $\Omega^k(U)$.

Differential forms of arbitrary degree form a $\mathbb{Z}$ -graded vector space

$$(23.3) \quad \Omega^\bullet(U) = \bigoplus_{k=0}^{\infty} \Omega^k(U).$$

The operation of exterior multiplication is defined on differential forms pointwise, and with it $\Omega^\bullet(U)$ is a commutative $\mathbb{Z}$ -graded algebra; see Definition 20.22.

Example 23.4. Let $\mathbb{A}_{x,y,z}^3$ be standard affine 3-space with coordinate functions x, y, z . Let $U \subset \mathbb{A}^3$ be open. At each point $p \in U$ the differentials dx_p, dy_p, dz_p form a basis of $(\mathbb{R}^3)^*$. (Since $x, y, z: U \rightarrow \mathbb{R}$ are affine functions, the differentials are constant, so the basis is independent of p .) Therefore, by Theorem 21.35(3) we can write any element of $\Omega^2(U)$ as a linear combination

$$(23.5) \quad f(x, y, z)dx \wedge dy + g(x, y, z)dx \wedge dz + h(x, y, z)dy \wedge dz$$

for functions $f, g, h: U \rightarrow \mathbb{R}$.

(23.6) *The main theorem.* The exterior differential is characterized by a few basic properties, which were mostly flagged in Lecture 18.

Theorem 23.7. *There exists a unique map $d: \Omega^\bullet(U) \rightarrow \Omega^\bullet(U)$ of degree +1 such that*

- (i) d is linear,
- (ii) $d(\alpha_1 \wedge \alpha_2) = d\alpha_1 \wedge \alpha_2 + (-1)^k \alpha_1 \wedge d\alpha_2$, $\alpha_1 \in \Omega^{k_1}(U)$, $\alpha_2 \in \Omega^\bullet(U)$,
- (iii) $d^2 = 0$,
- (iv) d agrees with the usual differential on $\Omega^0(U)$.

That d is homogeneous of degree +1 means that if $\alpha \in \Omega^k(U)$ is homogeneous of degree k , then $d\alpha \in \Omega^{k+1}(U)$ is homogeneous of degree $k+1$. We prove Theorem 23.7 below but first illustrate with some explicit computations.

Example 23.8. Consider the differential form

$$(23.9) \quad \alpha = e^{2x} - x^2 y^2 dy - x dx \wedge dy$$

on $\mathbb{A}_{x,y}^2$. Then using the rules in Theorem 23.7 we compute

$$(23.10) \quad \begin{aligned} d\alpha &= d(e^{2x}) - d(x^2 y^2 dy) - d(x dx \wedge dy) \\ &= 2e^{2x} dx - d(x^2 y^2) \wedge dy - dx \wedge dx \wedge dy \\ &= 2e^{2x} dx - 2xy^2 dx \wedge dy. \end{aligned}$$

In this example both α and $d\alpha$ are inhomogeneous.

Example 23.11. Consider

$$(23.12) \quad \alpha = x dy \wedge dz + y dx \wedge dz + z dx \wedge dy$$

in $\Omega^2(\mathbb{A}_{x,y,z}^3)$. Then

$$(23.13) \quad d\alpha = 3 dx \wedge dy \wedge dz.$$

Example 23.14. On an open set $U \subset \mathbb{A}_{x^1, \dots, x^n}^n$ consider the general smooth $(n-1)$ -form

$$(23.15) \quad \alpha = f^1 dx^2 \wedge \cdots \wedge dx^n - f^2 dx^1 \wedge dx^3 \wedge \cdots \wedge dx^n + \cdots,$$

where $f^i: U \rightarrow \mathbb{R}$ are smooth functions. Then

$$(23.16) \quad d\alpha = \left(\sum_i \frac{\partial f^i}{\partial x^i} \right) dx^1 \wedge \cdots \wedge dx^n.$$

Proof of Theorem 23.7. Let $x^1, \dots, x^n: A \rightarrow \mathbb{R}$ be affine coordinates on A , and restrict them to functions on U . It suffices to define d on k -forms for all $k \in \mathbb{Z}^{>0}$ and then use (i) to extend uniquely by linearity, since every differential form is a finite sum of homogeneous forms. By Theorem 21.35(3) we can write $\alpha \in \Omega^k(U)$ uniquely as

$$(23.17) \quad \alpha = \sum_{1 \leq i_1 < \dots < i_k \leq n} f_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k}$$

for functions $f_{i_1 \dots i_k}: U \rightarrow \mathbb{R}$. Then if d exists satisfying (i)–(iv) we compute

$$(23.18) \quad \begin{aligned} d\alpha &= d \left(\sum_{1 \leq i_1 < \dots < i_k \leq n} f_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k} \right) \\ &\stackrel{(i)}{=} \sum_{1 \leq i_1 < \dots < i_k \leq n} d(f_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k}) \\ &\stackrel{(ii)}{=} \sum_{1 \leq i_1 < \dots < i_k \leq n} df_{i_1 \dots i_k} \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\ &\quad + f_{i_1 \dots i_k} d^2 x^{i_1} \wedge \dots \wedge dx^{i_k} \\ &\quad - f_{i_1 \dots i_k} dx^{i_1} \wedge d^2 x^{i_2} \wedge dx^{i_3} \dots \wedge dx^{i_k} \\ &\quad + \dots \\ &\stackrel{(iii)}{=} \sum_{1 \leq i_1 < \dots < i_k \leq n} df_{i_1 \dots i_k} \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\ &\stackrel{(iv)}{=} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\partial f_{i_1 \dots i_k}}{\partial x^j} dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k}. \end{aligned}$$

Therefore, if d exists and satisfies (i)–(iv) it must be given by the formula (23.18). This proves the uniqueness. To prove existence we *define* d on k -forms by formula (23.18) and check (i)–(iv). Property (i) is easy. As a variation, for (ii) use increasing multi-indices, as in the proof of Proposition 22.12. So as not to have index wars let us call the forms $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^\ell(U)$. (By linearity it suffices to take the second form homogeneous as well.) Write

$$(23.19) \quad \begin{aligned} \alpha &= \alpha_I dx^I \\ \beta &= \beta_J dx^J \end{aligned}$$

for functions $\alpha_I, \beta_J: U \rightarrow \mathbb{R}$. Then

$$(23.20) \quad \begin{aligned} d(\alpha \wedge \beta) &= d(\alpha_I \beta_J dx^I \wedge dx^J) \\ &= \alpha_J \frac{\partial \alpha_I}{\partial x^i} dx^i \wedge dx^I \wedge dx^J + \alpha_I \frac{\partial \alpha_J}{\partial x^j} dx^j \wedge dx^I \wedge dx^J \\ &= \left(\frac{\partial \alpha_I}{\partial x^i} dx^i \wedge dx^I \right) \wedge (\alpha_J dx^J) + (-1)^k (\alpha_I dx^I) \wedge \left(\frac{\partial \alpha_J}{\partial x^j} dx^j \wedge dx^J \right) \\ &= d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta. \end{aligned}$$

The sign comes from commuting the 1-form dx^j past the k -form dx^I . For (iii) compute

$$\begin{aligned}
 d^2\alpha &= d \left(\sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\partial f_{i_1 \dots i_k}}{\partial x^j} dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \right) \\
 &= \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\partial^2 f_{i_1 \dots i_k}}{\partial x^\ell \partial x^j} dx^\ell \wedge dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
 &= \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{1}{2} \left(\frac{\partial^2 f_{i_1 \dots i_k}}{\partial x^\ell \partial x^j} + \frac{\partial^2 f_{i_1 \dots i_k}}{\partial x^j \partial x^\ell} \right) dx^\ell \wedge dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
 (23.21) \quad &= \frac{1}{2} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\partial^2 f_{i_1 \dots i_k}}{\partial x^\ell \partial x^j} dx^\ell \wedge dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
 &\quad + \frac{1}{2} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\partial^2 f_{i_1 \dots i_k}}{\partial x^\ell \partial x^j} dx^j \wedge dx^\ell \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
 &= \frac{1}{2} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\partial^2 f_{i_1 \dots i_k}}{\partial x^\ell \partial x^j} dx^\ell \wedge dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
 &\quad - \frac{1}{2} \sum_{1 \leq i_1 < \dots < i_k \leq n} \frac{\partial^2 f_{i_1 \dots i_k}}{\partial x^\ell \partial x^j} dx^\ell \wedge dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_k} \\
 &= 0.
 \end{aligned}$$

In the fourth equality we exchange the dummy indices j and ℓ ; in the penultimate equality we use $dx^j \wedge dx^\ell = -dx^\ell \wedge dx^j$. The main point is the third equality, which expresses that second partials are symmetric; this contrasts the skew-symmetry of the wedge product. Finally, the definition (23.18) reduces to (9.32) for $k = 0$, which proves (iv). $\square$

(23.22) *Expression for d in terms of directional derivatives.* We continue with $U \subset A$ an open subset of a finite dimensional affine space A over a normed linear space V . Recall that for $f \in \Omega^0(U)$ and $\xi \in V$ we have

$$(23.23) \quad df(\xi) = \xi f,$$

an equality of functions on U . The following generalizes (23.23) for all k . In this theorem we treat a differential k -form as an alternating k -linear function on vectors; see Proposition 22.12.

Theorem 23.24. *Let $k \in \mathbb{Z}^{\geq 0}$, $\alpha \in \Omega^k(U)$, and $\xi_1, \dots, \xi_{k+1} \in V$. Then*

$$\begin{aligned}
 d\alpha(\xi_1, \dots, \xi_{k+1}) &= \xi_1 \alpha(\xi_2, \dots, \xi_{k+1}) - \xi_2 \alpha(\xi_1, \xi_3, \dots, \xi_{k+1}) \\
 (23.25) \quad &\quad + \dots + (-1)^k \xi_{k+1} \alpha(\xi_1, \dots, \xi_k).
 \end{aligned}$$

Each term is the directional derivative of the k -form α evaluated on k vectors. Note the oft-used special case $k = 1$ in which (23.25) reduces to

$$(23.26) \quad d\alpha(\xi_1, \xi_2) = \xi_1 \alpha(\xi_2) - \xi_2 \alpha(\xi_1).$$

Proof. Choose affine coordinates $x^1, \dots, x^n$ as in the proof of Theorem 23.7. By multilinearity suffices to verify (23.25) when each

$$(23.27) \quad \xi_j = \partial/\partial x^{i_j}$$

is a basis vector for some $1 \leq i_1, \dots, i_{k+1} \leq n$. If any two indices are equal, then it is straightforward to check that both sides of (23.25) vanish using the alternating property of differential forms. Furthermore, by the skew-symmetry property of differential forms it suffices to assume that $1 \leq i_1 < \dots < i_{k+1} \leq n$. Write $\alpha = \alpha_I dx^I$ as a sum over increasing indices of length k . Then with (23.27) the right hand side of (23.25) is

$$(23.28) \quad \sum_j (-1)^j \frac{\partial}{\partial x^{i_j}} \alpha_{i_1 \dots \hat{i}_j \dots i_{k+1}}.$$

Using the pairing (22.13) we compute the left hand side as

$$(23.29) \quad \left\langle \sum_{j,I} \frac{\partial \alpha_I}{\partial x^j} dx^j \wedge dx^I, \frac{\partial}{\partial x^{i_1}} \wedge \dots \wedge \frac{\partial}{\partial x^{i_{k+1}}} \right\rangle = \sum_j (-1)^j \frac{\partial}{\partial x^{i_j}} \alpha_{i_1 \dots \hat{i}_j \dots i_{k+1}}.$$

The terms in the first sum which contribute are those in which the multi-index jI of length $k+1$ is a permutation of $i_1, \dots, i_{k+1}$; the sign is the determinant of the permutation matrix. $\square$

Lecture 24: Pullback of differential forms; forms on bases

Pullbacks

(24.1) *Definition.* We already defined the pullback of 0-forms (functions) and 1-forms in (19.22). Now we extend to forms of all degrees.

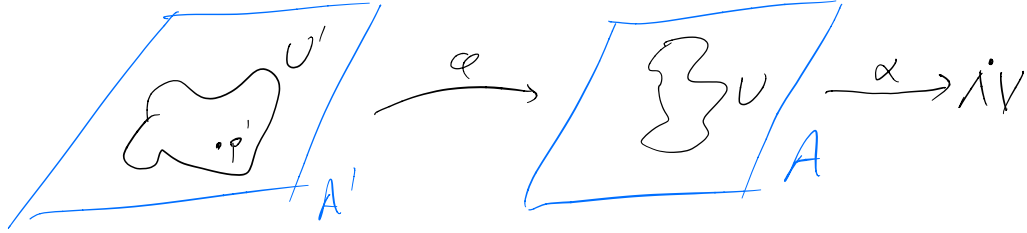


FIGURE 33. Pullback of differential forms

Let V, V' be real vector spaces, which for this definition may be infinite dimensional. Let A, A' be affine over V, V' and $U \subset A, U' \subset A'$ open subsets. Finally, let $\varphi: U' \rightarrow U$ be a C^1 map and $p' \in U'$. Then the differential and its dual give linear maps

$$\begin{aligned}
 (24.2) \quad & V' \xrightarrow{d\varphi_{p'}} V \\
 & (V')^* \xleftarrow{d\varphi_{p'}^*} V^* \\
 & \bigwedge^\bullet (V')^* \xleftarrow{\bigwedge d\varphi_{p'}^*} \bigwedge^\bullet V^*
 \end{aligned}$$

where the last map is the induced map (21.5) on the exterior algebra.

Definition 24.3. Let $\alpha \in \Omega^\bullet(U)$ be a differential form. The *pullback* $\varphi^*\alpha \in \Omega^\bullet(U')$ is

$$(24.4) \quad (\varphi^*\alpha)_{p'} = \bigwedge d\varphi_{p'}^*(\alpha_{\varphi(p')}), \quad p' \in U'.$$

(24.5) *Pullbacks and products.* The following is an immediate consequence of the fact that $\bigwedge d\varphi_{p'}^*$ is an algebra homomorphism.

Proposition 24.6. If $\alpha, \beta \in \Omega^\bullet(U)$, then

$$(24.7) \quad \varphi^*(\alpha \wedge \beta) = \varphi^*\alpha \wedge \varphi^*\beta.$$

(24.8) Computation. Definition 24.3 is not applied directly in computations; rather one plugs and chugs as illustrated in Example 19.28. Here is another illustration.

Example 24.9. Consider the 2-form α in (23.12). Consider the map

$$(24.10) \quad \begin{aligned} \varphi: (0, \pi) \times (0, 2\pi) &\longrightarrow \mathbb{A}_{x,y,z}^3 \\ \phi, \theta &\longmapsto \sin \phi \cos \theta, \sin \phi \sin \theta, \cos \phi \end{aligned}$$

This embeds an open rectangle in $\mathbb{A}_{\phi,\theta}^2$ into the unit sphere in $\mathbb{A}^3$. To compute we write

$$(24.11) \quad \begin{aligned} x &= \sin \phi \cos \theta \\ y &= \sin \phi \sin \theta \\ z &= \cos \phi \end{aligned}$$

and then apply d :

$$(24.12) \quad \begin{aligned} dx &= \cos \phi \cos \theta d\phi - \sin \phi \sin \theta d\theta \\ dy &= \cos \phi \sin \theta d\phi + \sin \phi \cos \theta d\theta \\ dz &= -\sin \phi d\phi. \end{aligned}$$

Substitute (24.11) and (24.12) into (23.12) and use the rules of exterior algebra to deduce

$$(24.13) \quad \varphi^* \alpha = \sin \phi d\phi \wedge d\theta.$$

In computations it is customary to omit ‘ φ^* ’ in (24.13).

(24.14) Pullback and d . Just as pullback commutes with product (Proposition 24.6), it also commutes with exterior d .

Proposition 24.15. Assume A, A' are finite dimensional. Let $\alpha \in \Omega^\bullet(U)$. Then

$$(24.16) \quad d\varphi^* \alpha = \varphi^* d\alpha.$$

Proof. Choose affine coordinates $x^1, \dots, x^n$ on A and $y^1, \dots, y^m$ on A' . Express the map φ as n functions of m variables:

$$(24.17) \quad \begin{aligned} x^1 &= x^1(y^1, \dots, y^m) \\ x^2 &= x^2(y^1, \dots, y^m) \\ &\vdots \\ x^n &= x^n(y^1, \dots, y^m) \end{aligned}$$

Differentiating we have

$$(24.18) \quad dx^i = \frac{\partial x^i}{\partial y^a} dy^a.$$

The pullback is effected by substituting (24.17) and (24.18) into the differential form, as in Example 24.9. For simplicity of notation we write the proof for $k = 1$. So

$$(24.19) \quad \alpha = f_i dx^i = f_i(x^1, \dots, x^n) dx^i$$

for some functions $f_i: U \rightarrow \mathbb{R}$ and

$$(24.20) \quad \varphi^* \alpha = f_i(x^1(y^1, \dots, y^m), \dots, x^n(y^1, \dots, y^m)) \frac{\partial x^i}{\partial y^a} dy^a,$$

which we write more concisely as

$$(24.21) \quad \varphi^* \alpha = \varphi^*(f_i) \frac{\partial x^i}{\partial y^a} dy^a.$$

Then from (23.18) we compute

$$(24.22) \quad \begin{aligned} d\varphi^* \alpha &= \left(\frac{\partial(\varphi^* f_i)}{\partial y^b} \frac{\partial x^i}{\partial y^a} + \varphi^*(f_i) \frac{\partial^2 x^i}{\partial y^b \partial y^a} \right) dy^b \wedge dy^a \\ &= \frac{\partial(f_i \circ \varphi)}{\partial y^b} \frac{\partial x^i}{\partial y^a} dy^b \wedge dy^a \\ &= \varphi^* \left(\frac{\partial f_i}{\partial x^j} \right) \frac{\partial x^j}{\partial y^b} \frac{\partial x^i}{\partial y^a} dy^b \wedge dy^a \end{aligned}$$

In the second equality we use the symmetry of second partials and the skew-symmetry of wedge product, as in (23.21); in the third equality we use the usual chain rule. On the other hand

$$(24.23) \quad \begin{aligned} \varphi^* d\alpha &= \varphi^* \left(\frac{\partial f_i}{\partial x^j} dx^j \wedge dx^i \right) \\ &= \varphi^* \left(\frac{\partial f_i}{\partial x^j} \right) \frac{\partial x^j}{\partial y^b} \frac{\partial x^i}{\partial y^a} dy^b \wedge dy^a \end{aligned}$$

The equality of (24.22) and (24.23) completes the proof. $\square$

Bases of vector spaces and differential forms

(24.24) *Bases of a vector space.* Fix $n \in \mathbb{Z}^{>0}$ and let V be an n -dimensional real normed vector space. Recall (Definition 2.15) that

$$(24.25) \quad \mathcal{B}(V) = \{b: \mathbb{R}^n \rightarrow V : b \text{ is invertible} \}$$

is the set of bases of V ; it is a right $\mathrm{GL}_n \mathbb{R}$ -torsor. Also, $\mathcal{B}(V)$ is an open subset of the vector space $\mathrm{Hom}(\mathbb{R}^n, V)$ (recall Theorem 11.2), so we can apply our calculus of differential forms to $\mathcal{B}(V)$.

(24.26) *The Maurer-Cartan 1-forms.* Define the 1-forms Θ_j^i , $1 \leq i, j \leq n$, as follows. At $b \in \mathcal{B}(V)$ it is the linear functional defined by the matrix equation

$$(24.27) \quad b^{-1}\dot{b} = \left(\Theta_j^i|_b(\dot{b}) \right), \quad \dot{b} \in \text{Hom}(\mathbb{R}^n, V).$$

Each side of (24.27) is a linear map $\mathbb{R}^n \rightarrow \mathbb{R}^n$, i.e., an $n \times n$ real matrix. The row index is i and the column index is j .

Remark 24.28. The definition (24.27) yields an immediate and extremely important geometric interpretation. Namely, the form Θ_j^i at the basis b tells the rate at which the j^{th} basis vector of b is moving towards the i^{th} basis vector of b . It is a linear functional of the tangent direction $\dot{b}$.

(24.29) *Structure equations I.* The exterior differential of Θ_j^i obey an important structure equation.

Theorem 24.30. *For each $1 \leq i, j \leq n$ the differential form $\Theta_j^i \in \Omega^1(\mathcal{B}(V))$ satisfies*

$$(24.31) \quad d\Theta_j^i + \Theta_k^i \wedge \Theta_j^k = 0.$$

The index k in (24.31) is summed over. There are n^2 equations; each right hand side is a sum of n terms. If we write $\Theta = (\Theta_j^i)$ as a matrix of 1-forms, then (24.27) is the equation

$$(24.32) \quad \Theta = b^{-1}db,$$

and (24.31) can be written as a product of matrices:

$$(24.33) \quad d\Theta + \Theta \wedge \Theta = 0.$$

The matrix product is computed by substituting wedge product for multiplication. Equation (24.33) is called the *Maurer-Cartan equation*.

Proof. First differentiate $b^{-1}b = \text{id}_{\mathbb{R}^n}$ to conclude

$$(24.34) \quad d(b^{-1}) = -b^{-1}db b^{-1}.$$

Then differentiate (24.32) to find

$$(24.35) \quad d\Theta = -b^{-1}db b^{-1} \wedge db = -\Theta \wedge \Theta. \quad \square$$

Bases of affine spaces and differential forms

(24.36) Frames. Let V be an n -dimensional real normed vector space and A an affine space over V . We define bases, or frames, of A to include a point as well as a basis of V .

Definition 24.37. The space of *frames* in A is the Cartesian product

$$(24.38) \quad \mathcal{B}(A) = A \times \mathcal{B}(V).$$

So a frame is a pair (p, b) in which $p \in A$ and $b \in \mathcal{B}(V)$. We encountered such pairs when defining parallelepipeds (Definition 22.19). The space $\mathcal{B}(A)$ is an open subset of the affine space $A \times \text{Hom}(\mathbb{R}^n, V)$, so we have the foundations to use differential forms on it.

(24.39) The “soldering” form. The 1-forms Θ_j^i are defined on $\mathcal{B}(A)$, formally by pullback via the projection $\mathcal{B}(A) \rightarrow \mathcal{B}(V)$. There are new 1-forms θ^i on $\mathcal{B}(A)$, $1 \leq i \leq n$, as well. At a point $(p, b) \in \mathcal{B}(A)$ write the basis of V as an ordered n -tuple $e_1, \dots, e_n$ of vectors in V . Then on a tangent vector

$$(24.40) \quad (\dot{p}; \dot{e}) = (\dot{p}; \dot{e}_1, \dots, \dot{e}_n) \in V \times V^{\times n}$$

we have

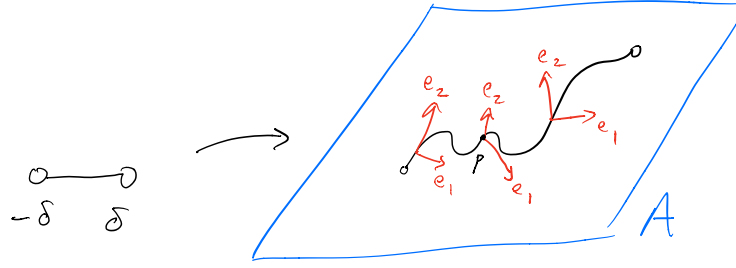
$$(24.41) \quad \dot{p} = \theta^i(\dot{p}; \dot{e})e_i.$$

More succinctly, we can regard $\theta = \theta^i \hat{e}_i$ as an $\mathbb{R}^n$ -valued 1-form on $\mathcal{B}(A)$. Then

$$(24.42) \quad \theta_{(p,b)}(\dot{p}, \dot{b}) = b^{-1}(\dot{p}).$$

Notice that θ is translation-invariant—it does not depend on p —and also it does not depend on $\dot{b}$, only on $\dot{p}$. In other words, a motion in $\mathcal{B}(A)$ is a curve in A together with a moving frame, or moving basis, along the curve. The forms Θ_j^i depend on the moving frame, whereas θ^i only depends on the underlying curve of points.

Remark 24.43. As in Remark 24.28 there is an intuitive and very important picture of the forms θ^i which follows from the definition (24.41). Namely, given a motion $\gamma: (-\delta, \delta) \rightarrow A$ and an initial basis $e_1, \dots, e_n$ at $\gamma(0)$, then θ^i at $(\gamma(0); e_1, \dots, e_n)$ in the direction $\dot{\gamma}(0)$ is the i^{th} component of $\dot{\gamma}(0)$ relative to the basis $e_1, \dots, e_n$.

FIGURE 34. A moving frame: a curve in $\mathcal{B}(A)$

(24.44) *Structure equations II.* Since pullback commutes with d and products, the structure equations (24.31) also hold on $\mathcal{B}(A)$. There is a new structure equation for the soldering forms.

Theorem 24.45. *For each $1 \leq i \leq n$ the differential form $\theta^i \in \Omega^1(\mathcal{B}(A))$ satisfies*

$$(24.46) \quad d\theta^i + \Theta_j^i \wedge \theta^j = 0.$$

We can write (24.46) in matrix form as

$$(24.47) \quad d\theta + \Theta \wedge \theta = 0.$$

The matrix product is computed as in (24.33) with wedge product in place of multiplication. Equations (24.33) and (24.47) are another instance of the *Maurer-Cartan equations*.

Proof. Evaluate (24.47) at (p, b) on the tangent vectors $(\dot{p}', \dot{b}')$ and $(\dot{p}'', \dot{b}'')$. For the first term we use (23.26):

$$(24.48) \quad \begin{aligned} d\theta_{(p,b)}((\dot{p}', \dot{b}'), (\dot{p}'', \dot{b}'')) &= (\dot{p}', \dot{b}') \cdot b^{-1}(\dot{p}'') - (\dot{p}'', \dot{b}'') \cdot b^{-1}(\dot{p}') \\ &= -b^{-1}\dot{b}'b^{-1}(\dot{p}'') + b^{-1}\dot{b}''b^{-1}(\dot{p}'), \end{aligned}$$

where ‘ $\cdot$ ’ denotes the directional derivative and we use (24.34) to differentiate b^{-1} . For the second term we use (22.17):

$$(24.49) \quad \begin{aligned} (\Theta \wedge \theta)_{(p,b)}((\dot{p}', \dot{b}'), (\dot{p}'', \dot{b}'')) &= \Theta_{(p,b)}(\dot{p}', \dot{b}') \theta_{(p,b)}(\dot{p}'', \dot{b}'') - \Theta_{(p,b)}(\dot{p}'', \dot{b}'') \theta_{(p,b)}(\dot{p}', \dot{b}') \\ &= b^{-1}\dot{b}'b^{-1}(\dot{p}'') - b^{-1}\dot{b}''b^{-1}(\dot{p}'). \end{aligned}$$

The sum of (24.48) and (24.49) vanishes, as required. $\square$

Frames on Euclidean space

(24.50) *Calculus on curved spaces.* In the rest of this lecture and the next we compute with differential forms on curved spaces. As we have neither developed the theory of smooth curved spaces, i.e., smooth manifolds, nor the theory of differential forms on them, you should take this material as heuristic. At the same time, you should be able to make your way through the computations without the theory, and doing so now will help you learn the theory later. The goal is to illustrate the utility of differential forms in a geometric setting. In the next lecture we apply the equations developed in this lecture to the curvature of curves and surfaces, including a proof of Gauss' beautiful *Theorema Egregium*, which was mentioned in **(16.61)**.

(24.51) *Warmup: differential forms on the circle.* Differential forms on the circle are a small variation of differential forms on the affine line, and they are essentially a special case of differential forms on bases treated below: orthonormal bases on a 2-dimensional inner product space.

Let $\mathbb{A}_x^1$ be the standard affine line, and let $\mathbb{Z} \subset \mathbb{R}$ act by translation: the integer $n \in \mathbb{Z}$ translates

$$(24.52) \quad \begin{aligned} \mathbb{A}^1 &\longrightarrow \mathbb{A}^1 \\ x &\longmapsto x + n \end{aligned}$$

The quotient $\mathbb{A}^1/\mathbb{Z}$ is the circle. Let $\pi: \mathbb{A}^1 \rightarrow \mathbb{A}^1/\mathbb{Z}$ be the quotient map. We treat differential forms $\alpha \in \Omega^\bullet(\mathbb{A}^1/\mathbb{Z})$ via the pullback $\pi^*\alpha \in \Omega^\bullet(\mathbb{A}^1)$. The pullback map is injective, so we do not lose information. If $f \in \Omega^0(\mathbb{A}^1/\mathbb{Z})$, then $\tilde{f} = \pi^*f: \mathbb{A}^1 \rightarrow \mathbb{R}$ is a *periodic* function: $\tilde{f}(x+n) = \tilde{f}(x)$ for all $x \in \mathbb{A}^1$, $n \in \mathbb{Z}$. Similarly, the pullback $\tilde{\alpha} = \pi^*\alpha$ of a 1-form $\alpha \in \Omega^1(\mathbb{A}^1/\mathbb{Z})$ has the form $\tilde{g}(x)dx$, where $\tilde{g}: \mathbb{A}^1 \rightarrow \mathbb{R}$ is a periodic function.

Let $\Theta \in \Omega^1(\mathbb{A}^1/\mathbb{Z})$ be the 1-form with $\pi^*\Theta = dx$.

Lemma 24.53. *There does not exist $f \in \Omega^0(\mathbb{A}^1/\mathbb{Z})$ such that $df = \Theta$.*

Proof. If so, then its pullback $\tilde{f} = \pi^*f$ to $\mathbb{A}^1$ is a periodic function such that $d\tilde{f} = dx$. But then $\tilde{f} = x + c$ for some $c \in \mathbb{R}$, and this is not periodic. $\square$

(24.54) *Inner product spaces and orthonormal bases.* Now suppose V is an n -dimensional real vector space endowed with an inner product $\langle -, - \rangle_V$. Recall that the subset of orthonormal bases is (Definition 9.17)

$$(24.55) \quad \mathcal{B}_O(V) = \{b \in \mathcal{B}(V) : b \text{ is an isometry} \}.$$

The isometry condition may be expressed as

$$(24.56) \quad \langle be, be' \rangle_V = \langle e, e' \rangle_{\mathbb{R}^n}, \quad e, e' \in \mathbb{R}^n.$$

Recall that the adjoint $b^*: V \rightarrow \mathbb{R}^n$ is characterized by

$$(24.57) \quad \langle be, \xi \rangle_V = \langle e, b^*\xi \rangle_{\mathbb{R}^n}, \quad e \in \mathbb{R}^n, \quad \xi \in V.$$

Proposition 24.58.

- (1) Let $b \in \mathcal{B}(V)$. Then $b \in \mathcal{B}_O(V)$ if and only if $b^*b = \text{id}_{\mathbb{R}^n}$.
- (2) Suppose $b_t: (-\delta, \delta) \rightarrow \text{Hom}(\mathbb{R}^n, V)$ is a smooth curve with $b_t \in \mathcal{B}_O(V)$ for all $t \in (-\delta, \delta)$. Then $b_0^{-1}\dot{b}_0: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is skew-symmetric.

Proof. The first assertion follows immediately from (24.56) and (24.57). For the second, differentiate $b_t^*b_t = \text{id}_{\mathbb{R}^n}$ with respect to t :

$$\begin{aligned}
 0 &= \left. \frac{d}{dt} \right|_{t=0} (b_t^*b_t) \\
 &= \dot{b}^*b + b^*\dot{b} \\
 &= \dot{b}^*b^{*-1} + b^{-1}\dot{b} \\
 &= (b^{-1}\dot{b})^* + (b^{-1}\dot{b}).
 \end{aligned}
 \tag{24.59}$$

□

(24.60) Tangent space to orthonormal bases. We treat $\mathcal{B}_O(V) \subset \text{Hom}(\mathbb{R}^n, V)$ as a smooth subset on which we can do calculus. It is a higher dimensional version of a surface in 3-space, and one can give a formal definition as in (16.48). Furthermore, one can use the implicit function theorem to show that $\mathcal{B}_O(V)$ satisfies the definition: there are injective immersions which locally parametrize it. We will implicitly assume this and simply use Proposition 24.58(2) to motivate the following definition of the tangent space:

$$T_b\mathcal{B}_O(V) = \{\dot{b} \in \text{Hom}(\mathbb{R}^n, V) : b^{-1}\dot{b} \text{ is skew-symmetric} \}.
 \tag{24.61}$$

This is a subspace of dimension $n(n-1)/2$. Furthermore, we freely use the calculus of differential forms, following the rules established in the previous few lectures for open subsets of affine space.

(24.62) Structure equations III. Since $\mathcal{B}_O(V) \subset \mathcal{B}(V)$ we can restrict the 1-forms Θ_j^i to obtain 1-forms on $\mathcal{B}_O(V)$. We notate them the same, but need to be conscious of which space we are working over. There is an additional structure equation.

Theorem 24.63. For each $1 \leq i, j \leq n$ the differential form $\Theta_j^i \in \Omega^1(\mathcal{B}_O(V))$ satisfies

$$\Theta_j^i + \Theta_i^j = 0,
 \tag{24.64}$$

$$d\Theta_j^i + \Theta_k^i \wedge \Theta_j^k = 0.
 \tag{24.65}$$

Proof. The first equation follows from Proposition 24.58(2). The second is the restriction of (24.31) to $\mathcal{B}_O(V)$ and the fact that pullback commutes with wedge products and with d ; see (24.7) and (24.16). □

(24.66) *Orthonormal moving frames on Euclidean space.* Finally, if V has an inner product, then A is a Euclidean space and we define the subspace

$$(24.67) \quad \mathcal{B}_O(A) = A \times \mathcal{B}_O(V) \subset \mathcal{B}(A)$$

to be the set of pairs $(p; e_1, \dots, e_n)$ in which p is a point of A and $e_1, \dots, e_n$ an orthonormal basis of V , thought of as living in the tangent space to A at p . The forms θ^i, Θ_j^i pull back to $\mathcal{B}_O(A)$ and satisfy the structure equations (24.46), (24.64), and (24.65), which we gather here:

$$(24.68) \quad d\theta^i + \Theta_j^i \wedge \theta^j = 0,$$

$$(24.69) \quad \Theta_j^i + \Theta_i^j = 0,$$

$$(24.70) \quad d\Theta_j^i + \Theta_k^i \wedge \Theta_j^k = 0.$$

Lecture 25: Curvature of curves and surfaces

(25.1) *Introduction.* The comments in (24.50) remain in effect. Namely, in this lecture we take the mathematical license to compute with differential forms on curved manifolds with impunity. We introduce smooth manifolds of orthonormal bases adapted to a plane curve and then to a surface in space. The restrictions of the differential 1-forms θ, Θ introduced in Lecture 24 determine the curvature, and the structure equations (24.68)–(24.70) quickly lead to deep theorems about curvature, of which we only skim the surface here. We will use the material in Lecture 16, which the reader may want to review before proceeding further.

Curvature of plane curves

(25.2) *Tangent, normal, and coorientation.* Let E be a Euclidean plane, that is, an affine space over a 2-dimensional real inner product space V . Suppose $C \subset E$ is a curve; see Remark 16.7 and the better Definition 16.49 for surfaces. (The reader can easily adapt the latter to plane curves.) Recall that at each $p \in C$ there is an orthogonal decomposition

$$(25.3) \quad V = T_p C \oplus N_p C$$

as the direct sum of the tangent line and the normal line. Suppose $e_2: C \rightarrow V$ is a coorientation, as in Definition 16.32. That is, $e_2(p)$ is a unit vector in the normal line $N_p C$ and it varies smoothly with p . We use the symbol ‘ C ’ for the curve together with its coorientation. Recall from (24.66) the space $\mathcal{B}_O(E)$ of oriented frames in E .

Definition 25.4. The space of *adapted orthonormal frames* is

$$(25.5) \quad \mathcal{B}_O(C) = \{(p; e_1, e_2) \in \mathcal{B}_O(E) : p \in C, e_2 \text{ is the coorientation}\}.$$

For each $p \in C$ there are two choices for e_1 , namely the two unit vectors in the tangent line $T_p C$.

Remark 25.6. It is natural to decompose spaces of frames via the projection to Euclidean space with forgets the frame. So, for example, if we restrict $\mathcal{B}_O(E)$ to $C \subset E$ we have the projection

$$(25.7) \quad \begin{aligned} \mathcal{B}_O(E) \Big|_C &\longrightarrow C \\ (p; e_1, e_2) &\longmapsto p \end{aligned}$$

Note that $\mathcal{B}_O(C) = C \times \mathcal{B}_O(V)$ is a product. Also, $\mathcal{B}_O(V)$ is diffeomorphic to the disjoint union of two circles, so locally $\mathcal{B}_O(E)|_C$ is diffeomorphic to the disjoint union of two cylinders, as depicted in Figure 35. The adapted orthonormal frames $\mathcal{B}_O(C)$ sits inside as two copies of C , one in each cylinder. The projection

$$(25.8) \quad \begin{aligned} \pi: \mathcal{B}_O(C) &\longrightarrow C \\ (p; e_1, e_2) &\longmapsto p \end{aligned}$$

is a *double cover*; each fiber $\pi^{-1}(p)$ consists of two points.

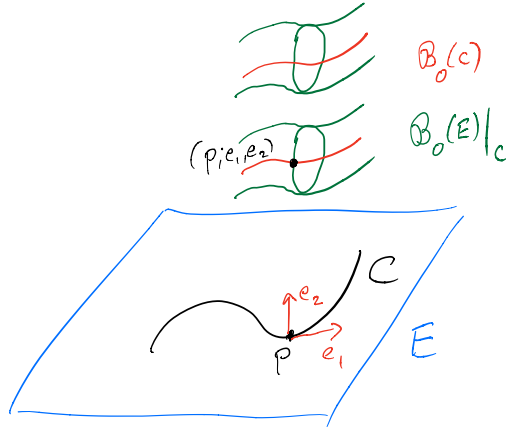


FIGURE 35. Adapted frames $\mathcal{B}_O(C)$ sitting in $\mathcal{B}_O(E)|_C$

(25.9) Curvature. Let $i: \mathcal{B}_O(C) \hookrightarrow \mathcal{B}_O(E)$ denote the inclusion. Recall the curvature $\kappa: C \rightarrow \mathbb{R}$ (Definition 16.35) and the expression (16.43) for it.²⁵ The restrictions to $\mathcal{B}_O(C)$ of the 1-forms θ^1, θ^2 and $\Theta_1^2 = -\Theta_2^1$ satisfy the following.

Proposition 25.10.

- (1) $i^*\theta^2 = 0$
- (2) $i^*\Theta_1^2 = (\pi^*\kappa)(i^*\theta^1)$

If we have no previous knowledge of curvature, then we can use (2) to define it. This shows how much information is packed into the 1-forms.

Remark 25.11. Recall the geometric intuition for the form Θ_1^2 , as told in Remark 24.28. Namely, it is the rate of turning of the basis vector e_1 towards the basis vector e_2 . Evaluate (2) on $(p; e_1, e_2) \in \mathcal{B}_O(C)$ to see that this rate of turning along the curve at p is the curvature $\kappa(p)$. This is precisely the interpretation given in (16.41).

²⁵In Lecture 16 we used e_0 for the unit normal instead of e_2 .

Proof. The manifold $\mathcal{B}_O(C)$ is 1-dimensional, and at $(p; e_1, e_2)$ the lift of the vector e_1 is a basis of its tangent space. But $\theta^2(e_1) = 0$ since θ^1, θ^2 is the dual basis to e_1, e_2 ; see the definition (24.41). This proves (1).

For (2) fix $p \in C$ and let $\gamma: (-\delta, \delta) \rightarrow E$ be a unit speed parametrization of a neighborhood of p in C with $\gamma(0) = p$. Lift to a motion on $\mathcal{B}_O(C)$:

$$(25.12) \quad \begin{aligned} \hat{\gamma}: (-\delta, \delta) &\longrightarrow \mathcal{B}_O(C) \\ s &\longmapsto (\gamma(s); \dot{\gamma}(s), e_2(\gamma(s))) \end{aligned}$$

Then $\hat{\gamma}^*\theta^1 = ds$ as 1-forms on $(-\delta, \delta)$ since by definition $\hat{\gamma}^*\theta^1(d/ds) = \theta^1(d\hat{\gamma}(d/ds))$,

$$(25.13) \quad d\hat{\gamma}\left(\frac{d}{ds}\right) = (\dot{\gamma}(s); \ddot{\gamma}(s), *),$$

and the e_1 -component of $\dot{\gamma}(s)$ is e_1 since $\dot{\gamma}(s) = e_1$. (It is straightforward to determine ‘ $*$ ’ but we do not need that formula.) It also follows from (25.13) that

$$(25.14) \quad \hat{\gamma}^*\Theta_1^2 = \langle \ddot{\gamma}(s), e_2(\gamma(s)) \rangle ds;$$

use the definition (24.27) and the fact that e_1, e_2 is orthonormal. It follows from (16.43) that $\langle \ddot{\gamma}(s), e_2(\gamma(s)) \rangle = \kappa(s)$, which implies (2). $\square$

Curvature of surfaces

(25.15) Tangent, normal, and coorientation. Let E be a 3-dimensional Euclidean space, an affine space over a 3-dimensional real inner product space V . Let $\Sigma \subset E$ be a surface, as in Definition 16.49. At each $p \in \Sigma$ there is an orthogonal decomposition

$$(25.16) \quad V = T_p\Sigma \oplus N_p\Sigma$$

in which the tangent space $T_p\Sigma$ is 2-dimensional and the normal space $N_p\Sigma$ is a line. Let $e_3: \Sigma \rightarrow V$ be a coorientation; $e_3(p) \in N_p\Sigma$ for all $p \in \Sigma$.

Definition 25.17. The space of *adapted orthonormal frames* is

$$(25.18) \quad \mathcal{B}_O(\Sigma) = \{(p; e_1, e_2, e_3) \in \mathcal{B}_O(E) : p \in \Sigma, e_3 \text{ is the coorientation}\}.$$

At each $p \in \Sigma$ the fiber of the projection

$$(25.19) \quad \begin{aligned} \pi: \mathcal{B}_O(\Sigma) &\longrightarrow \Sigma \\ (p; e_1, e_2, e_3) &\longmapsto p \end{aligned}$$

is the set of orthonormal bases of the tangent space $T_p\Sigma$. Geometrically it is the disjoint union of two circles, one for each orientation of $T_p\Sigma$. Since Σ is 2-dimensional, it follows that $\mathcal{B}_O(\Sigma)$ is 3-dimensional.

(25.20) *The 1-forms.* There are 6 linearly independent 1-forms on $\mathcal{B}_O(E)$, namely $\theta^1, \theta^2, \theta^3$ and $\Theta_1^2, \Theta_1^3, \Theta_2^3$. In fact, $\mathcal{B}_O(E)$ is 6-dimensional and these forms give a basis of the dual to its tangent space at each point. Since $\mathcal{B}_O(\Sigma) \subset \mathcal{B}_O(E)$ is 3-dimensional, we expect several linear relations among the restriction of these forms. For convenience we enumerate the structure equations (24.68)–(24.70):

$$(25.21) \quad d\theta^1 + \Theta_2^1 \wedge \theta^2 + \Theta_3^1 \wedge \theta^3 = 0$$

$$(25.22) \quad d\theta^2 + \Theta_1^2 \wedge \theta^1 + \Theta_3^2 \wedge \theta^3 = 0$$

$$(25.23) \quad d\theta^3 + \Theta_1^3 \wedge \theta^1 + \Theta_2^3 \wedge \theta^2 = 0$$

$$(25.24) \quad d\Theta_1^2 + \Theta_3^2 \wedge \Theta_1^3 = 0$$

$$(25.25) \quad d\Theta_1^3 + \Theta_2^3 \wedge \Theta_1^2 = 0$$

$$(25.26) \quad d\Theta_2^3 + \Theta_1^3 \wedge \Theta_2^1 = 0$$

and also the skew-symmetry equations

$$(25.27) \quad \Theta_1^2 + \Theta_2^1 = 0$$

$$(25.28) \quad \Theta_1^3 + \Theta_3^1 = 0$$

$$(25.29) \quad \Theta_2^3 + \Theta_3^2 = 0$$

Let $i: \mathcal{B}_O(\Sigma) \hookrightarrow \mathcal{B}_O(E)$ be the inclusion map. Introduce the notation

$$(25.30) \quad \begin{aligned} \bar{\theta}^i &= i^* \theta^i \\ \bar{\Theta}_j^i &= i^* \Theta_j^i \end{aligned}$$

for the restrictions of the forms to the space of adapted orthonormal frames.

Proposition 25.31.

- (1) $\bar{\theta}^3 = 0$
- (2) $d\bar{\Theta}_1^2 = (\pi^* \tilde{K}) \bar{\theta}^1 \wedge \bar{\theta}^2$ for some function $\tilde{K}: \Sigma \rightarrow \mathbb{R}$.

Proof. The proof of Proposition 25.10(1) applies to prove (1). An immediate consequence from (25.21) and (25.22) is

$$(25.32) \quad \begin{aligned} d\bar{\theta}^1 + \bar{\Theta}_2^1 \wedge \bar{\theta}^2 &= 0 \\ d\bar{\theta}^2 + \bar{\Theta}_1^2 \wedge \bar{\theta}^1 &= 0 \end{aligned}$$

Apply d to each of these equations to deduce

$$(25.33) \quad d\bar{\Theta}_1^2 \wedge \bar{\theta}^1 = \bar{\Theta}_1^2 \wedge \bar{\theta}^2 = 0.$$

We let the reader verify that $\bar{\theta}^1, \bar{\theta}^2, \bar{\Theta}_1^2$ form a basis of the dual space to the tangent space of $\mathcal{B}_O(\Sigma)$ at each point, and so

$$(25.34) \quad d\bar{\Theta}_1^2 = \tilde{K} \bar{\theta}^1 \wedge \bar{\theta}^2 + f \bar{\theta}^1 \wedge \bar{\Theta}_1^2 + g \bar{\theta}^2 \wedge \bar{\Theta}_1^2$$

for some functions $f, g: \mathcal{B}_O(\Sigma) \rightarrow \mathbb{R}$. It follows from (25.33) that $f = g = 0$. It remains to prove that $\tilde{K}$ is constant on the fibers of π in (25.19).

First, apply d to $d\bar{\Theta}_1^2 = \tilde{K} \bar{\theta}^1 \wedge \bar{\theta}^2$ to deduce

$$(25.35) \quad 0 = d\tilde{K} \wedge \bar{\theta}^1 \wedge \bar{\theta}^2 + \tilde{K} d\bar{\theta}^1 \wedge \bar{\theta}^2 - \tilde{K} \bar{\theta}^1 \wedge d\bar{\theta}^2 = d\tilde{K} \wedge \bar{\theta}^1 \wedge \bar{\theta}^2.$$

It follows that $d\tilde{K} = a\bar{\theta}^1 + b\bar{\theta}^2$ for some functions $a, b: \mathcal{B}_O(\Sigma) \rightarrow \mathbb{R}$, and in particular $d\tilde{K}(\zeta) = 0$ if ζ is a tangent vector along the fiber. (You will need to recall the definition of θ^i to make that conclusion.) This proves that $\tilde{K}$ is *locally* constant along the fibers of π . Since the fibers of π are not connected—they are diffeomorphic to the union of two circles—we need an additional argument to see that the constant is the same on each component.

Consider the diffeomorphism

$$(25.36) \quad \begin{aligned} \varphi: \mathcal{B}_O(E) &\longrightarrow \mathcal{B}_O(E) \\ (p; e_1, e_2, e_3) &\longmapsto (p; -e_1, e_2, e_3) \end{aligned}$$

It follows from the definition that

$$(25.37) \quad \begin{aligned} \varphi^* \theta^1 &= -\theta^1 \\ \varphi^* \theta^2 &= \theta^2 \\ \varphi^* \Theta_1^2 &= -\Theta_1^2 \end{aligned}$$

Furthermore, φ preserves $\mathcal{B}_O(\Sigma) \subset \mathcal{B}_O(E)$, so the analogous equations hold for the restrictions to $\mathcal{B}_O(\Sigma)$. Notice that φ exchanges the components in each fiber of π . Applying (25.37) we have

$$(25.38) \quad \varphi^* d\bar{\Theta}_1^2 = d\varphi^* \bar{\Theta}_1^2 = -d\bar{\Theta}_1^2$$

and

$$(25.39) \quad \varphi^*(\tilde{K} \bar{\theta}^1 \wedge \bar{\theta}^2) = \varphi^* \tilde{K} \varphi^* \bar{\theta}^1 \wedge \varphi^* \bar{\theta}^2 = -\varphi^* \tilde{K} \bar{\theta}^1 \wedge \bar{\theta}^2.$$

Hence $\varphi^* \tilde{K} = \tilde{K}$, which proves that $\tilde{K}$ is constant on fibers of π , so descends to a function on Σ . $\square$

Remark 25.40. The important point is that $\tilde{K}$ is *intrinsic*, that is, can be computed from the geometry of Σ without using its embedding into E . First, we observe that $\mathcal{B}_O(\Sigma)$ can be defined as the space of orthonormal frames of each tangent space to Σ , so once we know Σ as a smooth manifold, independent of its embedding into E , then we can construct $\mathcal{B}_O(\Sigma)$. The forms $\bar{\theta}^1, \bar{\theta}^2$ can be defined directly on $\mathcal{B}_O(\Sigma)$. You proved on homework that $\bar{\Theta}_1^2$ is uniquely determined by the structure equations (25.32) and the skew-symmetry (25.27). Then Proposition 25.31 determines the function $\tilde{K}$.

(25.41) *Identifying $\tilde{K}$ with minus Gauss curvature.* So far we have just followed our nose with the structure equations and out popped a function $\tilde{K}: \Sigma \rightarrow \mathbb{R}$. Now we want to recognize it as minus the Gauss curvature, defined in (16.64). That definition uses the *extrinsic* geometry of the second fundamental form (16.56) and the shape operator defined in (16.61).

First, we claim that if ζ is a tangent vector along a fiber of π , then $\bar{\Theta}_1^3(\zeta) = \bar{\Theta}_2^3(\zeta) = 0$. This follows since ζ is an infinitesimal rotation of the e_1, e_2 plane which fixes e_3 , and $\bar{\Theta}_j^3$ tells the rate of turning of e_j towards e_3 . Hence there are functions $h_{ij}: \mathcal{B}_O(\Sigma) \rightarrow \mathbb{R}$ such that

$$(25.42) \quad \begin{aligned} \bar{\Theta}_1^3 &= h_{11}\bar{\theta}^1 + h_{12}\bar{\theta}^2 \\ \bar{\Theta}_2^3 &= h_{21}\bar{\theta}^1 + h_{22}\bar{\theta}^2 \end{aligned}$$

Structure equation (25.23), restricted to $\mathcal{B}_O(\Sigma) \subset \mathcal{B}_O(E)$, implies $h_{12} = h_{21}$. In fact, h_{ij} is the matrix of the second fundamental form in the given basis.

Proposition 25.43. $h_{ij}(p; e_1, e_2, e_3) = \Pi_p(e_i, e_j)$.

Proof. Apply Proposition 16.58:

$$(25.44) \quad \Pi_p(e_i, e_j) = -\langle D_{e_i} e_3, e_j \rangle = -\bar{\Theta}_3^j(e_i) = \bar{\Theta}_j^3(e_i).$$

□

Theorem 25.45 (Gauss). *The Gauss curvature K is intrinsic.*

Proof. We prove that $K = -\tilde{K}$, and we explained in Remark 25.40 that $\tilde{K}$ is intrinsic. Apply structure equation (25.24), restricted to $\mathcal{B}_O(\Sigma)$, to find

$$(25.46) \quad \begin{aligned} 0 &= d\bar{\Theta}_1^2 + \bar{\Theta}_3^2 \wedge \bar{\Theta}_1^3 \\ &= d\bar{\Theta}_1^2 - \bar{\Theta}_2^3 \wedge \bar{\Theta}_1^3 \\ &= \tilde{K} \bar{\theta}^1 \wedge \bar{\theta}^2 + \det(h) \bar{\theta}^1 \wedge \bar{\theta}^2. \end{aligned}$$

Then $\tilde{K} = -K$ follows from (16.64).

□

Lectures 26–28: Integration

(26.1) Summary. The last several lectures covered integration theory. I posted Chapter 3 from Spivak's *Calculus on Manifolds* as a reference for integration of functions in standard affine space $\mathbb{A}^n$. This is a multivariable version of the Riemann integral in one variable. It was developed by Camille Jordan at the end of the 19th century. It is superseded by the Lebesgue theory, which you learn in real analysis courses, so I did not give a complete treatment of the Jordan theory. In these notes I will summarize some of the points I made in lecture which are not in Spivak's book.

(26.2) Highlights from Spivak. Fix $n \in \mathbb{Z}^{>0}$. A *box* is a Cartesian product of closed intervals:

$$(26.3) \quad B = \bigtimes_{i=1}^n [a^i, b^i] \subset \mathbb{A}^n, \quad a^i, b^i \in \mathbb{R}, \quad a^i < b^i.$$

The volume of B is given by the standard formula $\text{Vol}(B) = \prod_{i=1}^n (b^i - a^i)$.

If $f: B \rightarrow \mathbb{R}$ is a bounded function, then as in Riemann's theory we consider partitions of B ; squeeze f on each subbox between its inf and its sup; make a lower sum for the integral of f using the piecewise constant function of infs, and make an upper sum for the integral of f using the piecewise constant function of sups; define the lower and upper integral as the sup, respectively inf, over all partitions of lower and upper sums; and define f to be integrable if they agree. We proved that a function f is integrable if and only if its locus of discontinuity has *measure zero* in B .

We proved Fubini's theorem, which tells how to compute the integral over B of an integrable function as an iterated integral over lower dimensional boxes.

From there we introduced partitions of unity. This is an important tool in many contexts. Here we used it to define the integral of a compactly supported function on open sets and also to prove the change of variables formula, which we state as follows.

Theorem 26.4 (change of variables). *Let $U, U' \subset \mathbb{A}^n$ be open sets and $f: U \rightarrow \mathbb{R}$ a bounded function of compact support which is integrable. Suppose $\varphi: U' \rightarrow U$ is a C^1 diffeomorphism. Then $\varphi^*f: U' \rightarrow \mathbb{R}$ is integrable and*

$$(26.5) \quad \int_U f = \int_{U'} \varphi^*f |\det d\varphi|.$$

The determinant factor is the continuous function which is the composition

$$(26.6) \quad U' \xrightarrow{d\varphi} \text{Hom}(\mathbb{R}^n, \mathbb{R}^n) \xrightarrow{\det} \mathbb{R} \xrightarrow{|\cdot|} \mathbb{R}$$

It tells the instantaneous stretching factor on volumes of the diffeomorphism φ .

Example 26.7. The special case $n = 1$ of (26.5) is not the usual change of variables formula you first learned in calculus, which is for definite integrals. Thus if you substitute

$$(26.8) \quad \begin{aligned} x &= -2y \\ dx &= -2dy \end{aligned}$$

you obtain formulas such as

$$(26.9) \quad \int_2^4 x^2 dx = \int_{-1}^{-2} (4y^2)(-2dy).$$

On the other hand, Theorem 26.4 addresses the integral over a subset, which here we apply to a closed subset, to obtain instead

$$(26.10) \quad \int_{[2,4]} x^2 |dx| = \int_{[-2,-1]} 4y^2 2|dy|.$$

In (26.5) we did not write the standard measure on $\mathbb{A}^n$, which here we render in standard affine coordinates $x^1, \dots, x^n$ as $|dx^1 \cdots dx^n|$. The absolute value is consonant with the change of variables formula.

Remark 26.11. The change of variables formula (26.9) treats $x^2 dx$ as a differential 1-form on $[2, 4]$, whereas (26.10) is for the integral of the *density* $x^2 |dx|$. The integration theory discussed up to now is for densities, not differential forms. We tell a bit about integration of differential forms below, but for completeness we first define densities.

(26.12) Interlude: densities. Let V be an n -dimensional real vector space.

Definition 26.13. The *line of densities* of V is

$$(26.14) \quad |\mathrm{Det} V^*| = \{\mu: \mathcal{B}(V) \rightarrow \mathbb{R} : \mu(b \cdot g) = |\det g| \mu(b) \text{ for all } b \in \mathcal{B}(V), g \in \mathrm{GL}_n \mathbb{R}\}.$$

Let $|\mathrm{Det} V^*|_+ \subset |\mathrm{Det} V^*|$ be the ray of positive functions. A *(positive) density* is an element of $|\mathrm{Det} V^*|_+$.

Recall that $\mathcal{B}(V)$ is a right $\mathrm{GL}_n \mathbb{R}$ -torsor, that is, the group of invertible $n \times n$ matrices acts simply transitively on $\mathcal{B}(V)$ by right composition. A density μ is a volume function on parallelepipeds in V . (Recall Definition 22.19.) If A is affine on V , then μ defines a translation-invariant density on A , in particular a translation-invariant volume function on parallelepipeds in A .

More generally, we can consider variable densities

$$(26.15) \quad \mu: U \rightarrow |\mathrm{Det} V^*|$$

defined on an open set $U \subset A$. The product of a function and a density is a density, so for example in (26.10) $x^2 |dx|$ is a variable density on $[2, 4]$. The integral discussed in (26.2) can be viewed as an integral of a variable density.

Integration of differential forms

(26.16) *Reduction to integration of functions.* Let V be an n -dimensional real normed vector space, A an affine space over V , and $U \subset A$ an open set. Let $\omega: U \rightarrow \text{Det } V^*$ be a differential n -form, and assume ω has compact support. We would like to define the integral of ω over U . Naturally, we do so by reducing to a problem already solved, the integral of compactly supported functions on $\mathbb{A}^n$. Hence let

$$(26.17) \quad \psi: \mathbb{A}_{x^1, \dots, x^n}^n \xrightarrow{\cong} A$$

be an affine isomorphism; its inverse is an affine coordinate system on A . Pull back ω to a compactly supported n -form on $\mathbb{A}^n$, which necessarily has the form

$$(26.18) \quad \psi^* \omega = f dx^1 \wedge \cdots \wedge dx^n$$

for some $f: \psi^{-1}(U) \rightarrow \mathbb{R}$. We would like to define

$$(26.19) \quad \int_U \omega = \int_{\psi^{-1}(U)} f,$$

but of course we must check that this definition is independent of ψ .

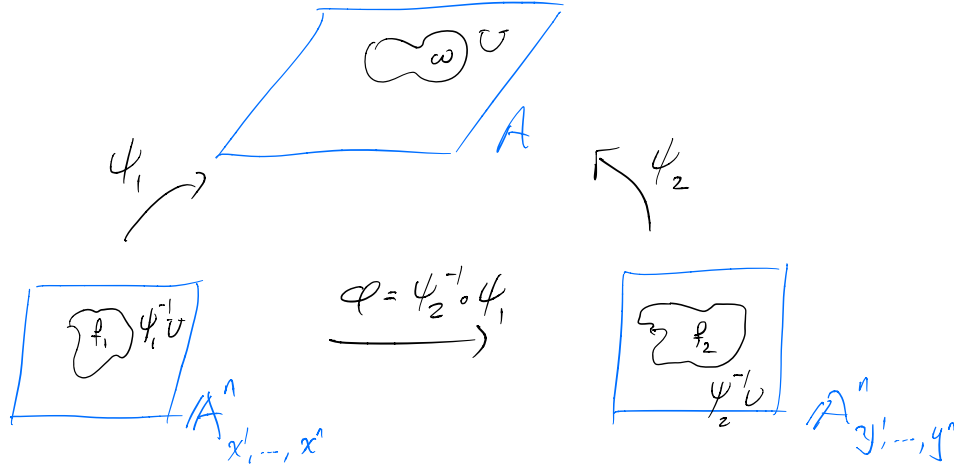


FIGURE 36. A differential form in affine coordinates

(26.20) *Checking independence of choices.* As in Figure 36 suppose we have two affine coordinate systems. Write

$$(26.21) \quad \begin{aligned} \psi_1^* \omega &= f_1 dx^1 \wedge \cdots \wedge dx^n \\ \psi_2^* \omega &= f_2 dy^1 \wedge \cdots \wedge dy^n \end{aligned}$$

Write the affine function φ as $y^i = y^i(x^1, \dots, x^n)$. Then from

$$(26.22) \quad dy^1 \wedge \dots \wedge dy^n = (\det d\varphi) dx^1 \wedge \dots \wedge dx^n$$

we have

$$(26.23) \quad \varphi^* \psi_2^* \omega = \varphi^* (f_2 dy^1 \wedge \dots \wedge dy^n) = \varphi^* f_2 (\det d\varphi) dx^1 \wedge \dots \wedge dx^n,$$

from which we conclude

$$(26.24) \quad f_1 = \varphi^* f_2 (\det d\varphi).$$

Then the change of variables formula (26.5) implies the desired equality $\int_{\psi_2^{-1}U} f_2 = \int_{\psi_1^{-1}U} f_1$ if and only if $\det d\varphi > 0$. By Remark 22.8 this holds if and only if $d\varphi$ is orientation-preserving at each point.

(26.25) Orientations. So now we see that we must orient A , which amounts to orienting V (as in Definition 22.7), and then we demand that the affine isomorphism (26.17) be *orientation-preserving* in the sense that its differential is an orientation-preserving linear map $\mathbb{R}^n \rightarrow V$, i.e., is an oriented basis of V . This defines the integral of a differential form.

Remark 26.26. We remind that in Lecture 19 we gave a motivation for introducing differential forms based on integration. You may want to review that now in light of the theory we have developed.

(26.27) Integration over curved manifolds. We do not have a formal theory of curved manifolds, but nonetheless you should be able to compute. Just as in affine space, to integrate a differential form you need an orientation. I illustrate via an example.

Example 26.28. We use the differential 2-form α defined in Example 23.11. To integrate it over the unit sphere S in $\mathbb{A}^3$ we use the parametrization (24.10) in Example 24.9. It is a diffeomorphism of an open rectangle in $\mathbb{A}^2$ onto an open subset of S whose complement has measure zero. Omission of a set of measure zero does not affect the value of the integral. Equation (24.13) asserts

$$(26.29) \quad \varphi^* \alpha = \sin \phi d\phi \wedge d\theta,$$

and we identify the 2-form $d\phi \wedge d\theta$ on $\mathbb{A}_{\phi, \theta}^2$ with the constant density $|d\phi d\theta|$. Thus

$$(26.30) \quad \int_S \alpha = \int_{(0, \pi) \times (0, 2\pi)} \sin \phi |d\phi d\theta| = 4\pi.$$

An important point is that just as in (26.25) we need to use an orientation-preserving parametrization, so to assert that (24.10) is orientation-preserving is to choose an orientation of S , that is, a coherent orientation of each tangent space.

(26.31) Stokes' theorem. We stated, but did not have time to prove, a basic theorem, Stokes' theorem. If A is an oriented n -dimensional affine space, $U \subset A$ open, α a compactly supported $(n-1)$ -form on U , and $R \subset U$ a closed subset with “smooth” boundary ∂R . (Again, regrettably, we have not defined this notion carefully in this course, but you will learn this if you take the Differential Topology prelim course.) We must orient ∂R , which we do by the “ONF rule”.²⁶ Then

$$(26.32) \quad \int_R d\alpha = \int_{\partial R} \alpha.$$

Example 26.33. Let $R \subset \mathbb{A}^3$ be the closed unit ball with boundary $\partial R = S$. Stokes' theorem implies that the integral computed in (26.30) equals

$$(26.34) \quad \int_R d\alpha = \int_R 3 \, dx \wedge dy \wedge dz = \int_R 3 \, |dx \, dy \, dz| = 4\pi.$$

The computation of $d\alpha$ is (23.13).

²⁶ONF = one never forgets = outward normal first