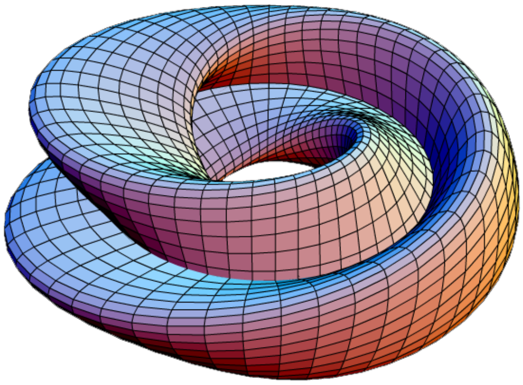

Differential Geometry

Differentials & Differential Forms, Integration on Manifolds and Constrained Optimization

Giuseppe Bassi

Academic Year 2024/2025

An introductory course

Abstract

These notes originate from a personal undertaking to consolidate and summarize the core material of a Differential Geometry course taught by Professor Alessandro Pigati at Bocconi University. They offer a concise re-elaboration of the main topics covered, rather than an exhaustive, textbook-level treatment. Hence, they are not intended to serve as an official reference and do not provide comprehensive coverage of the entire syllabus. I encourage readers to suggest improvements or report any errors they may encounter so that future editions can be improved.

Version 1.0 Date: September 25, 2025

Contents

1 Differentiable Manifolds

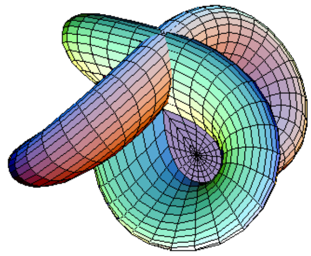

Figure 1: Morin surface

These notes focus on the theory of *differentiable manifolds*, geometric spaces that, while possibly curved or globally intricate (see Figure 1), locally resemble Euclidean spaces closely enough to allow the application of calculus. We will construct a rigorous setting for differentiability and integrability on such spaces, ultimately leading to core results like the method of *Lagrange multipliers* for constrained optimization, and *Stokes' Theorem* for the integration of differential forms. Before defining what a manifold is, we will recall two fundamental theorems, the *inverse function theorem* and the *implicit function theorem*.

1.1 Inverse and Implicit Function Theorems

Given two vector spaces V and W, with $\dim(V) = m$, $\dim(W) = n$, and a linear map $f : V \to W$ represented by a matrix $M \in \mathbb{R}^{n\times m}$ with respect to two given bases, we have the following:

i) f is surjective $\iff$ f has full rank $\iff$ M has a minor of dimension $n \times n$ with nonzero determinant. As a consequence, f is surjective only if $n \leq m$;

ii) f is injective $\iff$ f has zero null space $\iff$ M has a minor of dimension $m \times m$ with nonzero determinant. As a consequence, f is injective only if $n \geq m$;

iii) f is bijective $\iff$ f has full rank and zero null space $\iff$ M has nonzero determinant. As a consequence, f is bijective only if $n = m$. In such case, surjectivity, injectivity, bijectivity and invertibility coincide.

Example 1.1. Consider the matrices

$$A(x,y,z) = \begin{pmatrix} 2x+2 \\ 3y \\ z+2 \end{pmatrix}, \quad B(x,y,z) = \begin{pmatrix} 2x+2 & 3y & z+2 \end{pmatrix}, \quad C(x,y,z) = \begin{pmatrix} e^x \sin y & e^x \cos y & 0 \\ -e^x \cos y & e^x \sin y & 0 \\ 0 & 0 & z^2 \end{pmatrix}.$$

In this example, A, B and C are matrices corresponding to linear operators $\alpha : \mathbb{R} \to \mathbb{R}^3$, $\beta : \mathbb{R}^3 \to \mathbb{R}$ and $\gamma : \mathbb{R}^3 \to \mathbb{R}^3$, not linear functions. α can only be injective, which occurs when $(x, y, z) \neq (-1, 0, -2)$; β can only be surjective, which occurs when $(x, y, z) \neq (-1, 0, -2)$; γ is invertible (i.e. injective, surjective and bijective) if and only if $\det(C) \neq 0$, which occurs whenever $z \neq 0$.

As we just explained, a linear function $f : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if and only if has nonzero determinant. If f is nonlinear but preserve a C^1 regularity, global invertibility is not an easy task; however, we can still recover local invertibility by looking at the linearization of it. Morally, a C^1 function is locally invertible at x_0 if its linearization $Df(x_0)$ is:

Theorem 1.2 (Inverse Function Theorem). *Given a C^1 function $f : U \to \mathbb{R}^n$, with $U \subseteq \mathbb{R}^n$ open, if the differential $Df(x_0)$ is invertible [1] at a given point $x_0 \in U$, then there exists an open set $U' \subseteq U$ containing x_0 such that $f(U')$ is open and $f|_{U'}$ is a bijection from U' to $f(U')$ with C^1 inverse whose differential is given by*

$$Df^{-1}(y) = \left(Df(f^{-1}(y))\right)^{-1}$$

Theorem 1.3 (Implicit Function Theorem). *Let $U \subseteq \mathbb{R}^k$, $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^n$ be open neighborhoods of x_0, y_0 and $z_0 = F(x_0, y_0)$, respectively. Suppose*

$$F : U \times V \longrightarrow W, \quad (x, y) \longmapsto F(x, y)$$

is of class C^1 and that the partial Jacobian $D_yF(x_0, y_0)$ is invertible for some $(x_0, y_0) \in U \times V$. Then there exist (possibly smaller) open sets

$$U' \subseteq U, \quad V' \subseteq V, \quad W' \subseteq W$$

containing x_0, y_0, z_0 such that, for every $(x, z) \in U' \times W'$ the equation $F(x, y) = z$ has a ***unique*** *solution $y = g(x, z) \in V'$. Moreover, the assignment $(x, z) \mapsto g(x, z)$ is of class C^1 and satisfies $g(x_0, z_0) = y_0$.*

[1] Recall that $Df(x_0) : \mathbb{R}^n \to \mathbb{R}^n$ is the best linear approximation of f at x_0 (up to the constant $f(x_0)$), and thus it can be represented as a matrix multiplication $Df(x_0)[v] = \left[\frac{\partial f_i}{\partial x_j}(x_0)\right]_{0\le i,j\le n} \cdot [v]$. Therefore, $Df(x_0)$ is invertible if and only if $\det\left(\left[\frac{\partial f_i}{\partial x_j}(x_0)\right]_{0\le i,j\le n}\right) \neq 0$. Notice how we only need to look at the value of the determinant at x_0, even if the property holds in a neighbourhood U' of it. That's not surprising: if $\det\left(\left[\frac{\partial f_i}{\partial x_j}(x_0)\right]_{1\le i,j\le n}\right) \neq 0$, then the same property holds around x_0, by continuity.

Proof.
We only provide a sketch of the proof. The proof is a corollary of the *inverse function theorem*, which is based on the *Banach fixed-point theorem*. We define the function $\Phi : U \times V \to U \times W$ given by

$$\Phi(x, y) := (x, F(x, y)).$$

Then solving $F(x, y) = z$ is equivalent to inverting Φ. The Jacobian of Φ is

$$D\Phi(x, y) = \begin{pmatrix} I_k & 0 \\ D_x F(x, y) & D_y F(x, y) \end{pmatrix} \implies \det D\Phi(x, y) = \det D_y F(x, y) \neq 0.$$

Therefore, $D\Phi(x, y)$ is invertible, and the inverse function theorem applies to Φ: we find proper neighbourhoods for which F has a unique C^1 inverse G. From G it is possible to construct a C^1 implicit function $g : U' \times W' \to V'$ which solves $F(x, y) = z$, and its smoothness directly follows from that of G. $\square$

Example 1.4. Consider $F : \mathbb{R}_x \times \mathbb{R}_y \to \mathbb{R}$ given by

$$F(x, y) := \sin(y) + x^2 e^y.$$

The theorem guarantees the existence of a **unique** value y such that

$$\sin(y) + x^2 e^y = z,$$

for every (x, z) both sufficiently close to 0. For example, there exists a solution to

$$\sin(y) + 5e^y = \frac{1}{2}.$$

This is not surprising, and simple arguments of calculus I would have yielded the same result. What's new is how general (and *implicit*!) this theorem is, telling us not only that a solution exists, but also that it depends continuously on the other parameters. In particular, for the zero-level set $z = 0$, the equation

$$\sin(y) + x^2 e^y = 0$$

has a unique solution $y = g(x)$ near $(x, y) = (0, 0)$ which, Taylor-expanding, shows $g(x, 0) = -x^2 + O(x^4)$, so that locally $y \approx -x^2$.
Of course, one should first check the hypotheses (at the base point $(x_0, y_0) = (0, 0)$ and $z_0 = F(0, 0) = 0$): the Jacobian matrix $D_y F(x_0, y_0)$ simply becomes $\frac{\partial F}{\partial y}(0, 0) = \cos y + x^2 e^y \Big|_{(x,y)=(0,0)} = 1 \neq 0$, hence the theorem applies.

Example 1.5. Let us infer the existence of a solution (x, y) for the system

$$\begin{cases} x^2 + y^2 + a^2 = 3 \\ x - y + e^a = 1 \end{cases},$$

in a neighbourhood of $(x, y, a) = (1, 1, 1)$. Take the function

$$F : \mathbb{R}_a \times \mathbb{R}^2_{(x,y)} \to \mathbb{R}^2, \qquad F(a, (x, y)) = \begin{pmatrix} x^2 + y^2 + a^2 - 3 \\ x - y + e^a - 1 \end{pmatrix}.$$

Computing the Jacobian matrix with respect to (x, y) at $(1, 1, 1)$ yields

$$D_{(x,y)} F(1, 1, 1) = \begin{pmatrix} 2x & 2y \\ 1 & -1 \end{pmatrix}_{(x,y)=(1,1)} = \begin{pmatrix} 2 & 2 \\ 1 & -1 \end{pmatrix} \implies \det(D_{(x,y)} F(1, 1, 1)) = -4 \neq 0.$$

We then deduce the existence of a solution $(x, y) = g(a)$ in a neighbourhood of $(1, 1)$, depending smoothly on the parameter a.

1.2 Equivalent Definitions for Manifolds

A k-dimensional manifold $M \subseteq \mathbb{R}^n$ is, morally, a subset of $\mathbb{R}^n$ that can be locally approximated, at each point, by a k-dimensional vector subspace. In what follows, we present four different definitions of a manifold. We will first state them formally, then comment on each of them individually, and finally prove that they are all equivalent. Before doing so, let us recall two important notions:

Definition 1.6 (Homeomorphism). A homeomorphism between two open sets $U, V \subseteq \mathbb{R}^n$ is a bijective function $\psi : U \to V$ with ψ and ψ^{-1} of class C^0, i.e., continuous.[2]

Definition 1.7 (Diffeomorphism). A diffeomorphism between two open sets $U, V \subseteq \mathbb{R}^n$ is a bijective function $\Phi : U \to V$ with Φ and Φ^{-1} of class C^1.

Remark 1.8. Given a point p and a C^1 map $\Phi : \mathbb{R}^n \to \mathbb{R}^n$, then Φ is a local diffeomorphism of p with its image if and only if there exists an open set $U \subseteq \mathbb{R}^n$ where Φ is bijective and $D\varphi$ is invertible.

Definition 1.9 (Manifold). $M \subseteq \mathbb{R}^n$ is a manifold of dimension k if, for any $p \in M$, one of the following (equivalent) conditions occurs:[3]

i) There exists an open set $U \subseteq \mathbb{R}^n$ containing p and a C^1 map $\Phi : U \to \mathbb{R}^n$ such that:
   - i) Φ is a diffeomorphism with its image;
   - ii) $\Phi(M \cap U) = [\mathbb{R}^k \times \{0\}] \cap \Phi(U)$;

ii) There exists an open set $U \subseteq \mathbb{R}^n$ containing p and a C^1 map $h : U \to \mathbb{R}^{n-k}$ such that:
   - i) $Dh(p)$ is surjective (or, equivalently, has full rank);
   - ii) $M \cap U = \{x \in U : h(x) = 0\}$;

iii) There exists an open set $U \subseteq \mathbb{R}^n$ containing p, an open set $V \subseteq \mathbb{R}^k$ and a C^1 map $\psi : V \to U$ such that:
   - i) $D\psi(x)$ is injective (or, equivalently, has full rank) for all $x \in V$;
   - ii) ψ is a homeomorphism with its image;
   - iii) $M \cap U = \psi(V)$;

iv) There exists an open set $U \subseteq \mathbb{R}^n$ of the form $U = V \times W$, with $V \subseteq \mathbb{R}^k$ and $W \subseteq \mathbb{R}^{n-k}$ (open), and a C^1 map $\tilde{\psi} : V \to W$ such that:
   - i) $M \cap U = \{(x, \tilde{\psi}(x)) | x \in V\}$.

Remark 1.10. These four definitions are local properties, since we study properties in a small neighbourhood U of any given $p \in M$

Remark 1.11. If $M \subseteq \mathbb{R}^n$ is a k-dimensional manifold and $\mathcal{O} \subseteq \mathbb{R}^n$ is open, then $M \cap \mathcal{O}$ is also a k-dimensional manifold.

Remark 1.12. Be careful with the dimensions of the maps established by the definitions: $\Phi : \mathbb{R}^n \to \mathbb{R}^n$, $h : \mathbb{R}^n \to \mathbb{R}^{n-k}$, $\psi : \mathbb{R}^k \to \mathbb{R}^n$ and $\tilde{\psi} : \mathbb{R}^k \to \mathbb{R}^{n-k}$.

First Definition: Φ should be thought as a deformation acting on $M \cap U$, the portion of M close to p. We require that $M \cap U$ gets deformed into a portion, included in $\Phi(U)$, of a k-dimensional plane.

Example 1.13. A circumference $M = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is a manifold of dimension 1. Consider the point $p = (3/5, 4/5)$. We wish every point near p to be mapped, for example, onto the x-axis (a 1 dimensional plane). One might first consider the map $\Phi(x, y) = (x, 0)$; however, this map is not a C^1 diffeomorphism since it is not even invertible: distinct points in M could be sent to the same point. In addition, $\Phi(x, y)$ has second coordinate equal to zero even for points outside M, actually implying $[\mathbb{R} \times \{0\}] \cap \Phi(U) \supsetneq \Phi(M \cap U)$. The correct choice is

$$\Phi(x, y) = \big(x,\, y - \sqrt{1 - x^2}\big).$$

In fact, every point near p is mapped to a point whose second coordinate is 0. Moreover, this map is C^1 in a neighborhood of p and its inverse, $\Phi^{-1}(x, y) = \big(x, y + \sqrt{1 - x^2}\big)$, is also C^1.

How to find a suitable map? The following procedure works in most cases. First, write the manifold in the form

$$M = \{(x_1, \ldots, x_k, \ldots, x_n) \in \mathbb{R}^n : F(x_1, \ldots, x_k, \ldots, x_n) = 0\}.$$

[2] We require the continuity of the inverse in order to make this an equivalence relation between U and V, so that they share the same topological properties. Note that there exist continuous bijections whose inverses are not continuous. For instance, there is a continuous bijection from $[0, 1)$ onto the unit circle, but it does not admit a continuous inverse.

[3] As for the borderline cases $k = 0$ and $k = n$, these definitions are difficult to interpret. Conventionally, a 0-dimensional manifold is a discrete subset of $\mathbb{R}^n$, and a n-dimensional manifold is an open subset of $\mathbb{R}^n$. As for the empty set, it is a k-dimensional manifold for every $k = 0, \ldots, n$, for good reasons (related with the dimension of the tangent space).

Given a point $p \in M$, compute the Jacobian of F at p with respect to the last $n-k$ coordinates. If this Jacobian is invertible, then by the Implicit Function Theorem there exists a C^1 function $f : U \to V$, with $U \subseteq \mathbb{R}^k$ and $V \subseteq \mathbb{R}^{n-k}$, such that

$$F(x_1, \ldots, x_k, f(x_1, \ldots, x_k)) = 0.$$

In our example, where $F(x, y) = x^2 + y^2 - 1$, we have $f(x) = \sqrt{1-x^2}$ (if instead we had chosen $p = \big(3/5, -4/5\big)$, then $f(x) = -\sqrt{1-x^2}$). Next, define the map $\Phi : B_r(p) \to \mathbb{R}^n$, for some $r > 0$, by

$$\Phi(x_1, \ldots, x_n) = (x_1, \ldots, x_k) \times \Big((x_{k+1}, \ldots, x_n) - f(x_1, \ldots, x_k)\Big).$$

Since f is C^1, then Φ is a C^1 diffeomorphism. Although this definition might appear abstract, it is exactly what we did in our example.

Why do we require Φ as a C^1 diffeomorphism? We define Φ in this way to ensure that $\Phi(M \cap U)$ preserves all the geometric properties of $M \cap U$, while placing it in a more tractable vector space. Every condition imposed on Φ follows this guiding philosophy:

- **Invertibility:** without invertibility, distinct points in $M \cap U$ could be mapped to the same point in $\Phi(M \cap U)$, leading to a loss of information.
- **Continuity:** requiring Φ to be continuous with a continuous inverse (i.e., a homeomorphism) ensures that $M \cap U$ and $\Phi(M \cap U)$ are topologically equivalent (homeomorphic). This means the map preserves the notions of closeness and openness.
- **Continuous Differentiability:** requiring Φ to be a C^1 diffeomorphism guarantees that $M \cap U$ and $\Phi(M \cap U)$ are differentiably equivalent (diffeomorphic). This means the map preserves the notion of direction.

Remark 1.14. In other words, we call a manifold a geometric object that behaves well enough to be able to analyze it by locally mapping it to a diffeomorphic vector space.

Second Definition: Essentially, M is locally defined by a nonlinear system of $n-k$ independent equations (the zero set of a function from $\mathbb{R}^n$ to $\mathbb{R}^{n-k}$), leaving k degrees of freedom for its points. The surjectivity of $Dh(p)$ ensures that the $n-k$ equations defined by $h(x) = 0$ are non-redundant and regular in a differential sense.

This definition is less direct since it does not explicitly state how M is locally approximated by a vector subspace. Surprisingly, the two definitions are equivalent, so that this second is more technical but often easier to use to prove that M is a manifold.

Proof. Showing that the first definition implies the second is trivial: if there exists $\Phi : U \to \mathbb{R}^n$ as in the first definition, then one can define $h := (\Phi_{k+1}, \ldots, \Phi_n)$ by taking the last $n-k$ components of Φ. Then

$$\{x \in U : h(x) = 0\} = \{x \in U : \Phi(x) \in \mathbb{R}^k \times \{0\}\} = M \cap U, \quad \text{since} \quad [\mathbb{R}^k \times \{0\}] \cap \Phi(U) = \Phi(M \cap U).$$

Moreover, $Dh(p)$ is surjective because $D\Phi(p)$ is a bijective linear map (as Φ is a diffeomorphism). Proving the converse implication requires more work and will be done by these steps: $2 \to 4 \to 3 \to 1$. $\square$

Example 1.15. We have already seen that the n-dimensional sphere S^{n-1}, defined by $\|x\| = 1$, is an $(n-1)$-dimensional manifold according to the first definition. For the second definition, one simply takes $h(x) := \|x\|^2 - 1 = -1 + \sum_{i=1}^n x_i^2$, which vanishes precisely on the sphere, and checks that its differential is surjective. Since in this case $k = n-1$, the map $Dh(p) : \mathbb{R}^n \to \mathbb{R}$ is full rank if and only if it is nonzero—that is, h must have a nonzero gradient at every point $p \in M$. Since $\nabla h(p) = 2p \neq 0$, we are done.

Third Definition: Basically, M is locally defined as the image of a C^1 parametrization ψ, so that $M \cap U = \psi(V)$.

- **Injectivity of $D\psi(x)$:** We demand that the differential $D\psi(x)$ has full rank at every $x \in V$. Geometrically, this prevents the image $\psi(V)$ from "collapsing", it guarantees exactly k independent tangent directions at each point, confirming that $\psi(V)$ really looks like an open set in $\mathbb{R}^k$.
- **Homeomorphism of ψ:** Requiring ψ to be bijective with continuous inverse forbids local self-intersections or foldings, so that the k coordinates on V parameterize $\psi(V)$ unambiguously. Equivalently, it ensures the topology of $\psi(V)$ matches that of an open set in $\mathbb{R}^k$. An instructive counterexample is the lemniscate in Figure 2, whose parametrization fails to be a homeomorphism due to overlaps and a noncontinuous inverse.

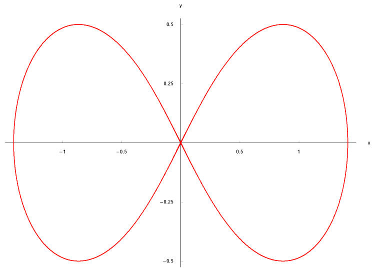

Figure 2: The lemniscate does not form a manifold because its parametrization overlaps itself. Its inverse is not continuous: one can approach a "double-point" by a sequence of parameters that map far apart in the domain, breaking the homeomorphism condition.

Fourth Definition: Essentially, M is locally defined as the graph of a C^1 map $\tilde{\psi} : V \to W$, where $V \subseteq \mathbb{R}^k$ and $W \subseteq \mathbb{R}^{n-k}$.

Example 1.16. The sphere $M = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ is a manifold of dimension 2. Consider the point $p = (3/13, 4/13, 12/13)$. Then $\tilde{\psi} : V \to W$, with $V \subseteq \mathbb{R}^2$, $W \subseteq \mathbb{R}$, defined $\tilde{\psi}(x, y) = \sqrt{1 - x^2 - y^2}$ is a proper parametrization, since $(x, y, \tilde{\psi}(x, y)) \in M$ locally.

We will now prove that $4 \implies 3$ and $2 \implies 4$.

Proof. To prove that (iv) implies (iii), just take $\psi(x) := (x, \tilde{\psi}(x))$. ψ is a homeomorphism: Its image is $M \cap U$ and ψ is manifestly injective; moreover, its inverse

$$\psi^{-1} : M \cap U \to V, \quad \psi^{-1}(x, y) = x$$

is just the projection onto the first factor. Therefore, it is continuous. Hence, ψ is a homeomorphism onto $M \cap U$. Moreover, $D\psi(x)$ has an $I_{k\times k}$ identity block in the first k rows, thus it has rank k and, therefore, it is injective.
As for the other implication, we only sketch the intuition. Taking $h : U \to \mathbb{R}^{n-k}$ as in the second definition, we know that there exists a $(n - k) \times (n - k)$ minor in the matrix $Dh(p)$ with nonzero determinant. We can then apply the implicit function theorem, which gives us a C^1 function in proper neighbourhoods which turns out to be the map $\tilde{\psi}$. □

Proof.

To conclude, we need to prove that the third definition implies the first one. It is a constructive proof. Given ψ as in the third definition, let $a_0 := \psi^{-1}(p) \in V$. We define a map

$$\Psi : V \times \mathbb{R}^{n-k} \to \mathbb{R}^n \quad \text{with} \quad \Psi(a, 0) = \psi(a) \quad \forall a \in V$$

such that $D\Psi(a_0, 0)$ is invertible. For instance, we can find $n - k$ vectors $v_1, \ldots, v_{n-k} \in \mathbb{R}^n$ which, together with the image of $D\psi(a_0)$, span the whole space $\mathbb{R}^n$, and we can simply take

$$\Psi(a, b) := \psi(a) + \sum_{i=1}^{n-k} b_i v_i.$$

Now the idea is to apply the inverse function theorem: up to shrinking the domain to a smaller $\tilde{U} \subseteq V \times \mathbb{R}^{n-k}$, Ψ is a diffeomorphism with its image, and we can simply take $\Phi := \Psi^{-1}$. Indeed, Ψ maps $[\mathbb{R}^k \times \{0\}] \cap \tilde{U}$ into M; with some work, it's possible to shrink $\tilde{U}$ properly so that those are the only points in the domain mapped to points in M. This concludes the proof. □

1.3 Tangent Space

Definition 1.17 (Tangent Space). Given a manifold M and a point $p \in M$, the tangent space to M at p is the collection of speeds of curves starting at p and traveling on M, that is,

$$T_pM = \{\gamma'(0) | \gamma : [0, \varepsilon] \to M, \gamma \in C^1, \gamma(0) = p, \varepsilon > 0\}.$$

Theorem 1.18 (Dimension of the tangent space). *The tangent space T_pM is a k-dimensional linear subspace of $\mathbb{R}^n$, for every $p \in M$.*

Proof.

Fix $p \in M$. We will show that T_pM is "trapped" between two k-dimensional linear subspaces A and B.[a] As a consequence, it will follow that $T_pM = A = B$.

- By definition of a manifold, there exists a C^1 map $h : U \to \mathbb{R}^{n-k}$ defined on an open set $U \ni p$, such that $M \cap U = \{x \in U : h(x) = 0\}$ and $Dh(p)$ is surjective. Take $s = \gamma'(0) \in T_pM$, with γ as in the definition of T_pM. If we can show that $Dh(p)[s] = 0$, it follows that $T_pM \subseteq \ker(Dh(p))$. Indeed, for all sufficiently small $t \in [0, \varepsilon]$, applying the chain rule at !,

$$h(\gamma(t)) \equiv 0 \implies 0 = \frac{\mathrm{d}(h \circ \gamma)}{\mathrm{d}t}(0) \overset{!}{=} Dh(\gamma(0))[\gamma'(0)] = Dh(p)[\gamma'(0)],$$

- By definition of a manifold, there exists a C^1 map $\psi : V \to U$ with $\psi(V) = M \cap U$, where $V \subseteq \mathbb{R}^k$ and $U \subseteq \mathbb{R}^n$ open and $p \in U$, with $D\psi(x_0)$ injective, and $\psi(x_0) = p$. Take $v \in \mathbb{R}^k$ and consider the function

$$\tau : [0, \varepsilon] \to V, \quad \tau(t) := x_0 + tv,$$

with ε small enough to make $\tau([0, \varepsilon]) \subseteq V$ and, consequently, $\psi \circ \tau([0, \varepsilon]) \subseteq M$. If we can show that $D\psi(x_0)[v] \in T_pM$, if follows that $\mathrm{Im}(D\psi(p)) \subseteq T_pM$. Indeed, by the chain rule (!),

$$\psi \circ \tau([0, \varepsilon]) \subseteq M \quad \text{and} \quad \psi(\tau(0)) = p \implies T_pM \ni \frac{\mathrm{d}(\psi \circ \tau)}{\mathrm{d}t}(0) \overset{!}{=} D\psi(\tau(0))[\tau'(0)] = D\psi(x_0)[v].$$

Since $Dh(p)$ is surjective it has rank $n - k$ and nullity $n - (n - k) = k$, so $\ker(Dh(p))$ is k-dimensional. Since $D\psi(x_0)$ is injective, then its rank is $k - 0 = k$, so $\mathrm{Im}(D\psi(p))$ is k-dimensional. Therefore, both $T_pM \subseteq \ker Dh(p)$ and $\mathrm{Im}(D\psi(x_0)) \subseteq T_pM$, and all these subspaces have dimension k, so we conclude:

$$T_pM = \mathrm{Im}(D\psi(x_0)) = \ker Dh(p).$$

□

[a] In fact, if $A \subseteq B$ are linear subspaces and both have dimension k, then necessarily $A = B$.

Definition 1.19 (Tangent Cone). Given a set $S \subseteq \mathbb{R}^n$ and $p \in S$ the tangent cone to S at p is

$$T_pS := \left\{ \lambda \lim_{i \to \infty} \frac{p^{(i)} - p}{|p^{(i)} - p|} \ , \ \lambda \geq 0 \right\},$$

where we look at all sequences of points $p^{(i)} \in S \setminus \{p\}$ such that $p^{(i)} \to p$ and $\frac{p^{(i)} - p}{|p^{(i)} - p|}$ also converges.

Remark 1.20. If p is isolated in S, we let $T_pS := \{0\}$.

Remark 1.21. In other words, we are looking at all possible "limit directions" along which we can approach a point p and we take all nonnegative multiples of these. This is a nice generalization of the tangent space which works even for sets which are not manifolds.

Remark 1.22. The tangent cone is a cone, i.e., it is invariant to dilatation. Therefore, for any $\mu > 0$, $\mu(T_pS) = T_pS$. Moreover, the tangent cone is always closed.

Remark 1.23. If S is convex, the tangent cone just defined turns out to be the closure of the tangent cone that one usually defines in convex analysis, which is $T_pC = \bigcup_{\lambda > 0} \lambda(C - p)$.

Theorem 1.24 (Tangent Cone on a Manifold). *The tangent cone on a k-dimensional manifold must be a k-dimensional vector space and it coincides with the tangent space.*

Proof.

Fix $p \in M$. We will employ the same "trapping" strategy used in the proof of Theorem 1.18.

- By definition of a manifold, there there exists a C^1 map $h : U \to \mathbb{R}^{n-k}$ defined on an open set $U \ni p$, such that $M \cap U = \{x \in U : h(x) = 0\}$ and $Dh(p)$ is surjective. Take a sequence of points $p^{(i)}$ as in the definition of tangent cone, and set $v := \lim_{i\to+\infty} \frac{p^{(i)}-p}{|p^{(i)}-p|}$. If we can show that $Dh(p)[\lambda v] = 0$ for every $\lambda \geq 0$, then $\lambda v \in \ker(Dh(p))$ and, consequently, $T_pM \subseteq \ker(Dh(p))$. Eventually, $p^{(i)} \in M \cap U$. Therefore, $h(p^{(i)}) = 0$ and $h(p) = 0$. Linearizing h we obtain

$$\cancel{h(p^{(i)})} = \cancel{h(p)} + Dh_p(p^{(i)} - p) + o(|p^{(i)} - p|) \iff Dh(p)\left[\frac{p^{(i)} - p}{|p^{(i)} - p|}\right] \to 0 \quad \text{as } i \to +\infty.$$

By continuity of $Dh(p)$, we deduce that $Dh(p)[\lambda v] = 0$.

- By definition of a manifold, there exists a C^1 map $\psi : V \to U$ with $\psi(V) = M \cap U$, where $V \subseteq \mathbb{R}^k$ and $U \subseteq \mathbb{R}^n$ open and $p \in U$, with $D\psi(x_0)$ injective, and $\psi(x_0) = p$. Fix $v \in \mathbb{R}^k \setminus \{0\}$ ($v = 0$ can be easily treated independently). If we can show that there exist $\lambda \geq 0$ and a sequence $p^{(i)}$ as in the definition of T_pM such that $\lambda \frac{p^{(i)}-p}{|p^{(i)}-p|} \to D\psi(x_0)[v]$, then $\operatorname{Im}(D\psi(x_0)) \subseteq T_pM$.

  Consider the function
$$\tau : [0, 1] \to V, \quad \tau(t) := x_0 + \varepsilon t v,$$
  with ε small enough to make $\tau([0, 1]) \subseteq V$ and, consequently, $\psi \circ \tau([0, 1]) \subseteq M$. Take
$$p^{(i)} := \psi \circ \tau(1/2^i)$$
  which correctly converges to $\psi(x_0) = p$. Linearizing, we have
$$p^{(i)} - p = \psi \circ \tau(1/2^i) - \psi \circ \tau(0) = D\psi(x_0)[\tau'(1/2^i)] + o(2^{-i}) = 2^{-i}\varepsilon D\psi(x_0)[v] + o(2^{-i}).$$
  Since $D\psi(x_0)[v] \neq 0$ by injectivity, then
$$\frac{p^{(i)} - p}{|p^{(i)} - p|} = \frac{2^{-i}\varepsilon D\psi(x_0)[v] + o(2^{-i})}{|2^{-i}\varepsilon D\psi(x_0)[v] + o(2^{-i})|} \to \frac{D\psi(x_0)[v]}{\|D\psi(x_0)[v]\|},$$
  so it's enough to take $\lambda := \|D\psi(x_0)[v]\| \geq 0$ and $(p^{(i)})_{i\in\mathbb{N}}$ defined as before.

The conclusion then follows exactly as in theorem 1.18. □

Remark 1.25. The converse is false . For instance, $S \subseteq \mathbb{R}^2$ given by the union $S = \{(x, y) : y = 0\} \cup \{(x, y) : y = x^2\}$ is not a 1-dimensional manifold, and yet T_pS is a 1-dimensional vector space at all $p \in S$. Another way this reverse implication may fail is that S could have "holes": for instance, we can take $S \subseteq \mathbb{R}$ given by **IMPROVE**

$$S := (-\infty, 0] \cup \left(\frac{1}{2}, 1\right) \cup \left(\frac{1}{8}, \frac{1}{4}\right) \cup \left(\frac{1}{32}, \frac{1}{16}\right) \cup \ldots,$$

which has $T_pS = \mathbb{R}$ for all $p \in S$ but is not a 1-dimensional manifold.

1.4 How to prove that a set is or isn't a manifold

2 Smooth Maps Between Manifolds

In the previous section we defined differentiable manifolds, i.e., manifolds with a C^1 regularity. Our study then naturally extends to C^1 functions between manifolds, which will labeled as *smooth*.

Definition 2.1 (Smooth map between manifolds). Given two manifolds $M \subseteq \mathbb{R}^m$ and $N \subseteq \mathbb{R}^n$, with M being a k-dimensional manifold, a map $f : M \to N$ is C^1 if, for any C^1 function $\psi : V \to M$ defined on an open set $V \subseteq \mathbb{R}^k$, the composition
$$f \circ \psi : V \to N$$
is C^1 (seen as a function to $\mathbb{R}^n$), i.e. the partial derivatives $\left(\frac{\partial f(x)}{\partial x_i}\right)_{i=1}^k$ exist and are continuous functions of $x \in \operatorname{Im}(\psi)$.

Example 2.2. The function $f(x, y) = \sqrt[3]{x^2 - x}$ fails to be C^1 whenever $x \in \{0, 1\}$. Let us restrict it as $\tilde{f} : M \to \mathbb{R}$, with $M = \{(x, y) \in \mathbb{R}^2 : x = y^3\}$. Let $\psi : \mathbb{R} \to M$ defined by $\psi(t) = (t^3, t)$. Then $(\tilde{f} \circ \psi)(t) = \sqrt[3]{t^6 - t^3} = t\sqrt[3]{t^3 - 1}$ is C^1 on $\mathbb{R} \setminus \{1\}$, so $\tilde{f}$ is C^1 on $M \setminus \{1, 1\}$.

Remark 2.3. If M is 0-dimensional, then any map $f : M \to N$ is C^1.

Remark 2.4. In the above definition, the manifold $N \subseteq \mathbb{R}^n$ plays no active role beyond fixing the codomain: one simply checks that $f \circ \psi$ is C^1 as a map into $\mathbb{R}^n$. Therefore, the next theorems will be defined, for simplicity, for functions $f : M \to \mathbb{R}^n$.

ADD...

Remark 2.5. Not every manifold admits a single global parametrization.For example, the sphere $S^n \subset \mathbb{R}^{n+1}$ requires at least two parametrizations to describe it. If there was a single homeomorphism $\psi : V \to S^n$ parametrizing S^n, then V would be compact (since S^n is compact and ψ^{-1} is continuous). But no non-empty open subset of $\mathbb{R}^n$ is compact. Hence S^n cannot be covered by one chart alone, and in practice one uses at least two (for instance stereographic projections from the north and south poles). When a manifold M is described as a union of a family of distinct parametrizations, i.e. $M = \bigcup_{i \in I} \psi_i(V_i)$ for a family $\psi_i : V_i \to M$, then $f : M \to \mathbb{R}^n$ is C^1 if and only if each composition $f \circ \psi_1$ is C^1.

In principle, we need to check the smoothness of $f \circ \psi$ for infinitely many maps ψ; luckily, it is usually enough to check only finitely many of them:

Theorem 2.6 (Smooth map between manifolds (2)). *A map $f : M \to \mathbb{R}^n$ is C^1 if and only if, for any $p \in M$, there exists a C^1 function $\tilde{f} : U \to \mathbb{R}^n$; with $U \ni p$ open, such that*

$$\tilde{f}|_{M \cap U} = f|_{M \cap U}.$$

Proof.

Assume that f admits a C^1 extension at any point of M. Given $\psi : V \to M$ of class C^1, we need to check that, given $x_0 \in V$ and setting $p := \psi(x_0)$, then $f \circ \psi(x_0) = f(p)$ is C^1. Invoking $\tilde{f} : U \to \mathbb{R}^n$, defined on a proper $U \ni p$, we have $f \circ \psi = \tilde{f} \circ \psi \quad$ on $\psi^{-1}(U) \subseteq V$. Since $\tilde{f} : U \to \mathbb{R}^n$ and $\psi|_{\psi^{-1}(U)}$ are C^1 between open sets in Euclidean spaces, their composition is C^1 as well, and therefore $f \circ \psi$ is C^1 at x_0 as well.
To prove the converse, let us use the fourth definition of manifold; up to a permutation of the coordinates, we can find an open set $U = V \times W$ containing p such that $M \cap U = \{(x, \tilde{\psi}(x)) \mid x \in V\}$ for some C^1 map $\tilde{\psi} : V \to W$, with $V \subseteq \mathbb{R}^k$ and $W \subseteq \mathbb{R}^{m-k}$. A possible extension $\tilde{f}$ is given by

$$\tilde{f}(x) := f(\pi(x), \tilde{\psi} \circ \pi(x)),$$

where $\pi(x) : U \to V$ given by $\pi(x) := (x_1, \dots, x_k)$ is the projection function. It's easy to see that $\tilde{f}$ is an extension of f and it's C^1 because it's a composition of C^1 functions. $\square$

Let us turn our attention to the differential properties of C^1 maps between manifolds.

Definition 2.7 (Differential of a smooth map). Given a C^1 map $f : M \to N$ between two manifolds, we define the differential of f at p as the linear map

$$Df(p) : T_pM \to T_{f(p)}N \qquad \text{given by} \qquad Df(p)[\gamma'(0)] := (f \circ \gamma)'(0),$$

with $\gamma : [0, \varepsilon] \to M$ C^1 and $\gamma(0) = p$.

Remark 2.8. If either M or N is 0-dimensional, so that either T_pM or $T_{f(p)}N$ is $\{0\}$, then $Df(p)$ is simply the zero map.

There are many problems with this definition:

- Is $f \circ \gamma$ really C^1? To answer this question, note that, for any $t_0 \in [0, \varepsilon]$, we can extend f locally near $\gamma(t_0)$ to a C^1 map $\tilde{f}$, defined on $U \ni \gamma(t_0)$, so that $f \circ \gamma = \tilde{f} \circ \gamma$ is C^1.
- Is $Df(p)$ well defined? In other words, does $Df(p)$ depend uniquely on $\gamma'(0)$ and not on γ? To answer this question, we again start by finding a local C^1 extension of f, namely $\tilde{f} : U \to \mathbb{R}^n$. Then, by the chain rule,
$$(f \circ \gamma)'(0) = (\tilde{f} \circ \gamma)'(0) = D\tilde{f}(\gamma(0))[\gamma'(0)],$$
whose value depends uniquely on $\gamma'(0)$ (here $\gamma : [0, \varepsilon] \to M$ has ε small enough, so that $\tilde{f} \circ \gamma$ is well defined).
- Is the differential linear? Yes, because we have just shown that, for every $v \in T_pM$, we have
$$Df(p)[v] = D\tilde{f}(p)[v],$$
so $Df(p)$ is obtained by restricting the linear map $D\tilde{f}(p)$ to the linear subspace T_pM.

ADD...

Example 2.9. We provide a concrete example that shows how to find the differential of a smooth map.

3 Differential Forms

In multivariable calculus, we often compute line, surface, or volume integrals using parametrizations and coordinate changes. But each time we do so, we must handle Jacobians, orientations, and index juggling by hand, and the expressions become messy and hard to interpret.
Differential forms provide a unified language for integration on manifolds. They allow us to write integrals in a *coordinate-free way*, making *changes of variables* and *orientation* systematic rather than ad hoc. More importantly, they are the natural objects to which Stokes' theorem applies, not just in $\mathbb{R}^3$, but on arbitrary smooth manifolds. Finally, differential forms provide one of the possible ways to deal with differential topology, which is the study of invariances of smooth manifolds and maps under continuous deformations.

3.1 Multilinear Alternating k-forms

In this chapter, we introduce multilinear alternating k-forms, which will become crucial to generalize the concept of infinitesimal volume in a *coordinate-free way*. Before defining them, it is worth to recall some basic properties of the *sign* of a permutation. Let $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ be a a permutation. For example, take $\sigma(\{1, 2, 3, 4\}) = \{2, 4, 3, 1\}$, in this order. This permutation can be obtained from the identity by applying two transpositions: $(1 \leftrightarrow 4) \circ (1 \leftrightarrow 2)$. Since an even number of transpositions is required, we say that σ is an *even* permutation, and we define its sign to be $\mathrm{sgn}(\sigma) = +1$. If an odd number of flips were needed, we would have $\mathrm{sgn}(\sigma) = -1$. Another way to compute the sign of a permutation is to compute the determinant of the corresponding permutation matrix. In the previous example, we obtain

$$\begin{vmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{vmatrix} = 1 = \mathrm{sgn}(\sigma).$$

Lastly, an important algebraic property of the sign function is that it is *multiplicative*: $\mathrm{sgn}(\sigma \circ \tau) = \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau)$ for any permutation σ and τ. This makes the sign function a group homomorphism from the symmetric group S_n to $\{\pm 1\}$.

Definition 3.1 (Multilinear Alternating k-form). A (multilinear) alternating k-form on a vector space V is a function

$$\alpha : \underbrace{V \times \dots \times V}_{k \text{ times}} \to \mathbb{R}$$

which is linear in each component and, flippling two entries, it takes opposite value. An alternating 0-form is a constant function.

Remark 3.2. We denote by $\Lambda^k V$ the (vector) space of alternating k-forms. Note that $\Lambda^1 V = V^*$, the dual space of V.

Remark 3.3. A multilinear k-form is alternating if and only if $\alpha(\dots, v, \dots, v, \dots) = 0$ for every v, in any position.

Remark 3.4. If σ is a permutation of $\{1, \dots, k\}$, then

$$\alpha(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \mathrm{sgn}(\sigma)\alpha(v_1, \dots, v_k).$$

Remark 3.5. Let $\alpha \in \Lambda^k V$ and fix a basis $\{e_1, \dots, e_n\}$ of V. For arbitrary vectors $v_1, \dots, v_k \in V$ write $v_j = \sum_{r=1}^n a_{jr} e_r$; by multilinearity,

$$\alpha(v_1, \dots, v_k) = \alpha\Big(\sum_{r=1}^n a_{1r} e_r, \ \dots, \ \sum_{r=1}^n a_{kr} e_r\Big) \underbrace{= \dots =}_{\text{multilinearity}} \sum_{r_1, \dots, r_k = 1}^n a_{1r_1} \cdots a_{kr_k} \, \alpha(e_{r_1}, \dots, e_{r_k}).$$

If $k \leq n$, then the k-form is a linear combination of $\frac{n!}{(n-k!)} = \binom{n}{k} \cdot k!$ monomials of degree k (see Example 3.6). If $k > n$, every alternating k-form is identically zero: indeed, the entries of α are k and the vectors e_i of the basis are $n < k$, so, by the pigeonhole principle, we always have two entries with the same value. Therefore, we are summing over terms which are all zero.

Example 3.6. Below we provide a list of examples of alternating k-forms to familiarize with them:

$$\begin{aligned} f(u, v) &= u_1 v_2 - u_2 v_1 = \det[u|v] && \text{an alternating bilinear form on } \mathbb{R}^2. \\ f(u, v) &= u_1 v_1 + u_2 v_2 = u \cdot v && \text{a bilinear form on } \mathbb{R}^2 \text{ which is not alternating} \\ f(v) &= v_1 + 2v_2 - v_3 + 5v_4 && \text{an alternating linear form on } \mathbb{R}^4 \end{aligned}$$

Definition 3.7 (Wedge Product). Given $k, \ell \geq 0$, the wedge (exterior) product is an operator $\wedge : \Lambda^k V \times \Lambda^\ell V \to \Lambda^{k+\ell} V$ given by

$$\boxed{(\alpha \wedge \beta)(v_1, \dots, v_{k+\ell}) := \sum_{\sigma \in S_{k+\ell}} \frac{\operatorname{sgn}(\sigma)}{k!\ell!} \alpha(v_{\sigma(1)}, \dots, v_{\sigma(k)}) \beta(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)})}.$$

Example 3.8. Below we provide a list of examples of wedge products of alternating forms:

- If α and β are 1-forms, then

$$(\alpha \wedge \beta)(x, y) = \frac{1}{1!1!} \cdot (\alpha(x)\beta(y) - \alpha(y)\beta(x)) = \alpha(x)\beta(y) - \alpha(y)\beta(x).$$

- If $\alpha \in \Lambda^1 V$ and $\beta \in \Lambda^2 V$, then

$$\begin{aligned}(\alpha \wedge \beta)(x, y_1, y_2) &= \frac{1}{1!2!} \cdot (\alpha(x)\beta(y_1, y_2) + \alpha(y_2)\beta(x, y_1) + \alpha(y_1)\beta(y_2, x) \\ &\qquad - \alpha(y_1)\beta(x, y_2) - \alpha(y_2)\beta(y_1, x) - \alpha(x)\beta(y_2, y_1)) \\ &= \alpha(x)\beta(y_1, y_2) + \alpha(y_2)\beta(x, y_1) + \alpha(y_1)\beta(y_2, x)\end{aligned}$$

- If $f \in \Lambda^2 V$ and $g \in \Lambda^2 V$, then

$$\begin{aligned}(f \wedge g)(a, b, c, d) = \frac{1}{2!2!} \cdot (&+ f(a, b)g(c, d) - f(d, a)g(b, c) + f(c, d)g(a, b) - f(b, c)g(d, a) \\ &- f(a, b)g(d, c) + f(c, a)g(b, d) - f(d, c)g(a, b) + f(b, d)g(c, a) \\ &+ f(a, c)g(d, b) - f(b, a)g(c, d) + f(d, b)g(a, c) - f(c, d)g(b, a) \\ &- f(a, c)g(b, d) + f(d, a)g(c, b) - f(b, d)g(a, c) + f(c, b)g(d, a) \\ &+ f(a, d)g(b, c) - f(c, a)g(d, b) + f(b, c)g(a, d) - f(d, b)g(c, a) \\ &- f(a, d)g(c, b) + f(b, a)g(d, c) - f(c, b)g(a, d) + f(d, c)g(b, a)).\end{aligned}$$

  In this example, it's visible the alternating property: exchanging two entries amounts to an additional flip to get every permutation, which changes the parity of the number of flips, and therefore the signs of every term. Moreover, using antisymmetry, we can recover a much more compact version as follows:

$$(f \wedge g)(a, b, c, d) = f(a, b)g(c, d) - f(a, c)g(b, d) + f(a, d)g(b, c) + f(b, c)g(a, d) - f(b, d)g(a, c) + f(c, d)g(a, b).$$

  The possibility to recover such compact linear combinations (with no denominators) is a general fact: the wedge product between an alternating k-form and an alternating ℓ-form always has $\binom{\ell+k}{k} = \binom{\ell+k}{\ell}$ terms (in this case, $\binom{2+2}{2} = 6$).

- We show multilinearity of $(f \wedge g)$ in the simple case $f \in \Lambda^1 V$ and $g \in \Lambda^2 V$, but the same reasoning works for any two alternating k-forms:

$$\begin{aligned}(f \wedge g)(a, \lambda x + \mu y, b) =& \frac{1}{2}(+f(a)g(\lambda x + \mu y, b) + f(b)g(a, \lambda x + \mu y) + f(\lambda x + \mu y)g(b, a) \\ &- f(\lambda x + \mu y)g(a, b) - f(b)g(\lambda x + \mu y, a) - f(a)g(b, \lambda x + \mu y)) \\ =& + \frac{\lambda}{2}(+f(a)g(x, b) + f(b)g(a, x) + f(x)g(b, a) - f(x)g(a, b) - f(b)g(x, a) - f(a)g(b, x)) \\ &+ \frac{\mu}{2}(+f(a)g(\mu, b) + f(b)g(a, \mu) + f(\mu)g(b, a) - f(\mu)g(a, b) - f(b)g(\mu, a) - f(a)g(b, \mu)) \\ =& \lambda(f \wedge g)(a, x, b) + \mu(f \wedge g)(a, y, b).\end{aligned}$$

IMPROVE

Remark 3.9. We have the following properties on the wedge product:

$$\begin{aligned}&\text{distributivity (1):} && (\lambda f + \mu g) \wedge h = \lambda f \wedge h + \mu g \wedge h \\ &\text{distributivity (2):} && f \wedge (\lambda g + \mu h) = \lambda f \wedge g + \mu f \wedge h \\ &\text{antisymmetry:} && f \wedge g = (-1)^{k\ell} g \wedge f, \quad f \in \Lambda^k V, g \in \Lambda^\ell V \\ &\text{associativity:} && (f \wedge g) \wedge h = f \wedge (g \wedge h) \\ &\text{iterated composition:} && \alpha_1 \wedge \dots \wedge \alpha_n(v_1, \dots, v_K) = \sum_{\sigma \in S_K} \frac{\operatorname{sgn}(\sigma)}{k_1! \dots k_n!} \prod_{i=1}^n \alpha_i\big(v_{\sigma(1+\sum_{j<i} k_j)}, \dots, v_{\sigma(\sum_{j \leq i} k_j)}\big)\end{aligned}$$

Definition 3.10 (Dual Basis). Given a basis $\{e_1, \dots, e_n\}$ of V, we can consider the collection of 1-forms $\{e_1^*, \dots, e_n^*\}$, where e_i^* is given by

$$e_i^*(a_1 e_1 + \dots + a_n e_n) := a_i.$$

Such collection is a basis for $V^* = \Lambda^1 V$, called *dual basis* of $\{e_1, \dots, e_n\}$.

Given indices $1 \leq i_1 < \dots i_k \leq n$, we have the identity

$$\boxed{e_{i_1}^* \wedge \dots \wedge e_{i_k}^*(v_1, \dots, v_k) = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) e_{i_1}^*(v_{\sigma(1)}) \dots e_{i_k}^*(v_{\sigma(k)}).}$$

Applying this property to $v = (e_{i_1'}, \dots, e_{i_k'})$, with $1 \leq i_1' < \dots < i_k' \leq n$, we see that

$$e_{i_1}^* \wedge \dots \wedge e_{i_k}^*(e_{i_1'}, \dots, e_{i_k'}) = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) e_{i_1}^*(e_{\sigma(i_1')}) \dots e_{i_k}^*(e_{\sigma(i_k')}) = \begin{cases} 1 & \text{if } i_1 = i_1', \dots, i_k = i_k', \\ 0 & \text{otherwise} \end{cases}. \tag{3.1}$$

Theorem 3.11. *If $0 < k \leq n$, the space $\Lambda^k V$ has dimension $\binom{n}{k}$, where $n = \dim(V)$.*

Proof.
Equation (3.1) implies that the collection $\{e_{i_1}^* \wedge \dots \wedge e_{i_k}^*\}$ is linearly independent as we vary the choice of indices (a total of $\binom{n}{k}$ vectors). To prove that such list is generating, we claim that for every $\alpha \in \Lambda^k V$, we have

$$\beta := \alpha - \sum_{i_1 < \dots < i_k} \alpha(e_{i_1}, \dots, e_{i_k}) \cdot (e_{i_1}^* \wedge \dots \wedge e_{i_k}^*) = 0 \quad \forall (v_i)_{i=1}^k \in V^k.$$

β is an alternating k-form and it clearly vanishes at $(v_1, \dots, v_k) = (e_{i_1}, \dots, e_{i_k})$ for every choice of $i_1 < \dots < i_k$. Since β is an alternating form, thanks to antisymmetry, it vanishes at $(v_1, \dots, v_k) = (e_{i_1}, \dots, e_{i_k})$ for any choice of indices, even non increasing, and even with repetitions. Then we can conclude, using multilinearity, that β vanishes everywhere on $\Lambda^k V$, which in turn implies that our list spans $\Lambda^k V$. $\square$

Theorem 3.11 clarifies why we care about dual bases. The reason is that any alternating k-form on $\mathbb{R}^n$ can be expressed as a wedge product of elements of the dual basis of $\mathbb{R}^n$. Given the central role of $\{e_1^*, \dots, e_n^*\}$, henceforth we will denote it as $\{dx_1, \dots, dx_n\}$. Later, we will see that there are clever reasons for such notation.

There is a nice way to compute the value of a k-form once decomposed as a wedge between elements of the dual basis, based on the determinant.

Example 3.12. $(dx \wedge dy \wedge dw)(a, b, c)$ is a 3-form on $\mathbb{R}^4$. It represents the oriented volume of the projection of the parallelepiped spanned by a, b, c on the oriented subspace with coordinates x, y, w. We have

$$\begin{aligned}(dz \wedge dy \wedge dw)(a, b, c) &= \det \begin{pmatrix} dz(a) & dz(b) & dz(c) \\ dy(a) & dy(b) & dy(c) \\ dw(a) & dw(b) & dw(c) \end{pmatrix} = \det \begin{pmatrix} a_3 & b_3 & c_3 \\ a_2 & b_2 & c_2 \\ a_4 & b_4 & c_4 \end{pmatrix} \\ &= a_3(b_2 c_2 - c_2 b_4) - c_3(a_2 c_4 - c_2 a_4) + c_3(a_2 c_4 - c_2 a_4).\end{aligned}$$

Similarly,

$$(dx \wedge dy)(a, b) = \det \begin{pmatrix} dx(a) & dx(b) \\ dy(a) & dy(b) \end{pmatrix} = \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = a_1 b_2 - b_1 a_2.$$

Moreover,

$$(dx \wedge dy \wedge dz)(e_{\sigma(1)}, e_{\sigma(2)}, e_{\sigma(3)}) = \det \begin{pmatrix} dx(e_{\sigma(1)}) & dx(e_{\sigma(2)}) & dx(e_{\sigma(3)}) \\ dy(e_{\sigma(1)}) & dy(e_{\sigma(2)}) & dy(e_{\sigma(3)}) \\ dz(e_{\sigma(1)}) & dz(e_{\sigma(2)}) & dz(e_{\sigma(3)}) \end{pmatrix} = \operatorname{sgn}(\sigma)$$

and

$$(dx \wedge dy \wedge dz)(a, a, b) = \det \begin{pmatrix} dx(a) & dx(a) & dx(b) \\ dy(a) & dy(a) & dy(b) \\ dz(a) & dz(a) & dy(b) \end{pmatrix} = 0,$$

since the matrix has two identical columns. Intuitively, $(dx \wedge dy \wedge dz)(a, a, b)$ is the volume of the parallelepiped spanned by a, a, b, but since a and a are not linearly independent, then the volume is zero.

3.2 Differential Forms

Definition 3.13 (Differential k-form). Given $U \subseteq \mathbb{R}^n$ open, a differential k-form of class C^1 is a C^1 map $\alpha : U \to \Lambda^k \mathbb{R}^n$.

Remark 3.14. We denote by $\Omega^k(U)$ the (vector) space of all differential k-forms on $U \subseteq \mathbb{R}^n$.

Example 3.15.

- $x_1^3\, dx_1 + x_1\, dx_2 + (e^{x_2} + \sin x_3)\, dx_3$ is a differential 1-form on $\mathbb{R}^3$, since it is a linear combination of all possible 1-forms in $\mathbb{R}^3$, each weighted by a C^1 function.
- $x_1\, dx_1 \wedge dx_2 + x_1 x_2\, dx_2 \wedge dx_3 + (x_3 - \cos x_1)\, dx_3 \wedge x_1$ is a differential 2-form on $\mathbb{R}^3$, since it is a linear combination of all possible 2-forms in $\mathbb{R}^3$, each weighted by a C^1 function. In general, a differential k form on $\mathbb{R}^n$, $n \geq k$, will be a linear combination of $\binom{n}{k}$ alternating k-forms.
- $(x_1 + x_2)\, dx_1 \wedge dx_2$ is a differential 2-form on $\mathbb{R}^2$, and indeed it has only one term; Note that the only differential k-form on $\mathbb{R}^n$, with $k > n$, is the constant 0, by pigeonhole principle (Remark 3.5).
- A differential 0-form is simply a C^1 function from U to $\mathbb{R}$.
- The general form of an element α of $\Omega^k(U)$ is

$$\alpha(x) = \sum_{i_1 < \cdots < i_k} a_{i_1,\ldots,i_k}(x) dx_{i_1} \wedge \cdots \wedge dx_{i_k} = \sum_I a_I(x) dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \tag{3.2}$$

Definition 3.16 (Exterior Derivative). Given $\alpha \in \Omega^k(U)$ in the form 3.2, we define the exterior derivative $d\alpha$ as the differential $(k+1)$-form of class C^0 given by

$$d\alpha(x) := \sum_{i_1 < \cdots < i_k} \underbrace{\sum_{\ell=1}^{n} \left(\frac{\partial a_{i_1,\ldots,i_k}(x)}{\partial x_\ell} dx_\ell \right)}_{da_I(x)} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k} = \sum_I da_I(x) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}.$$

Basically, we take the total derivative of each C^1 function.

Example 3.17. Let us compute the exterior derivative of $\alpha \in \Omega^2(\mathbb{R}^2)$ defined by $\alpha := \cos x_3\; dx_1 \wedge dx_2 - x_1 x_2\; dx_2 \wedge dx_3$.

$$\begin{aligned} d\alpha &= d(\cos x_3\; dx_1 \wedge dx_2 - x_1 x_2\; dx_2 \wedge dx_3) = d(\cos x_3) \wedge dx_1 \wedge dx_2 - d(x_1 x_2) \wedge dx_2 \wedge dx_3 \\ &= (-\sin x_3\; dx_3) \wedge dx_1 \wedge dx_2 - (x_2\; dx_1 + d_1\; dx_2) \wedge dx_2 \wedge dx_3 = -(\sin x_3 + x_2)\; dx_1 \wedge dx_2 \wedge dx_3. \end{aligned}$$

Theorem 3.18 (Leibniz Rule). *If $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^\ell(U)$, then*

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

In particular, if $k = 0$, then

$$\boxed{d(f\alpha) = (df) \wedge \alpha + f d\alpha}$$

Proof.

Writing a_I in place of $a_{i_1 < \cdots < i_k}$, dx_I in place of $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, b_J in place of $b_{j_1 < \cdots < j_\ell}$ and dx_J in place of $dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}$, we have

$$\begin{aligned} d(\alpha \wedge \beta) &= d\left(\sum_I a_I(x) dx_I \wedge \sum_J b_J(x) dx_J \right) = d\left(\sum_I \sum_J a_I(x) b_J(x) dx_I \wedge dx_J \right) \\ &= \sum_I \sum_J d(a_I b_J) dx_I \wedge dx_J = \sum_I \sum_J (da_I b_J + a_I db_J) dx_I \wedge dx_J \\ &= \sum_{I,J} da_I \wedge dx_I \wedge (b_J dx_J) + \sum_{I,J} (-1)^k (a_I dx_I) \wedge db_J \wedge dx_J = (d\alpha) \wedge \beta + (-1)^k \alpha \wedge d\beta. \end{aligned}$$

□

Theorem 3.19. *Given $\alpha \in \Omega^k(U)$ of class C^2, we have $d(d\alpha) = 0$.*

Proof.
With the same notation used in the previous proof, we see that

$$d(d\alpha) = \sum_I d(da_I) \wedge dx_I,$$

that is, the exterior derivative only affects the C^2 functions a_I and not the alternating 1-forms. If we are able to show that $d(da_I) = 0$ for every $a_I \in C^2$, then $d(d\alpha) = \sum_I 0 \wedge dx_I = 0$. Since a_I is a C^2 function, replace it with $f : U \to \mathbb{R}$ in C^2. We then compute:

$$d(df)(x) = d\left(\sum_{i=1}^n \frac{\partial f}{\partial x_i}\, dx_i\right)(x) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} dx_i \wedge dx_j \overset{!}{=} \sum_{i<j} \frac{\partial^2 f}{\partial x_i \partial x_j}\,(dx_i \wedge dx_j + dx_j \wedge dx_i) = 0,$$

where ! is Schwarz Theorem (notice that if $i = j$, then $dx_i \wedge dx_j = 0$) and the last passage follows from the antisymmetry of the wedge product: $dx_a \wedge dx_b = -dx_b \wedge dx_a$. □

A pullback is a change of variables. We transform a differential k-form in $\mathbb{R}^a$ to a differential k-form in $\mathbb{R}^b$ through a map $F : \mathbb{R}^b \to \mathbb{R}^a$:

Definition 3.20 (Pullback). Given two open sets $U \subseteq \mathbb{R}^n$ and $U' \subseteq \mathbb{R}^{n'}$, a C^2 map $F : U \to U'$ and a differential k-form α on U', the pullback of α by F is a differential k-form on U, denoted by $F^*\alpha$ and given by

$$\boxed{F^*\alpha(x)[v_1, \dots, v_k] := \alpha(F(x))[DF(x)[v_1], \dots, DF(x)[v_k]]}.$$

Example 3.21. Let us compute $F^*\alpha$, where $\alpha = \cos x_3\; dx_1 \wedge dx_2 - x_1x_2\; dx_2 \wedge dx_3$ in $\mathbb{R}^3$ and $F : \mathbb{R}^2 \to \mathbb{R}^3$ defined as $F(y_1, y_2) = y_1y_2, y_1 + y_2, y_1^2 - y_2^2)$.
We have

$$\begin{aligned}
F^*\alpha &= \cos(y_1^2 - y_2^2)d(y_1y_2) \wedge d(y_1 + y_2) - y_1y_2(y_1 + y_2) \wedge d(y_1^2 - y_2^2)\\
&= \cos(y_1^2 - y_2^2)(y_2dy_1 + y_1dy_2) \wedge (dy_1 + dy_2) - y_1y_2(y_1 + y_2)(dy_1 + dy_2) \wedge (2y_1dy_1 - 2y_2dy_2)\\
&= \cos(y_1^2 - y_2^2)(y_1 - y_1)dy_1 \wedge dy_2 - y_1y_2(y_1 + y_2)(-2y_2 - 2y_1)dy_1 \wedge dy_2\\
&= [(y_2 - y_2)\cos(y_1^2 - y_2^2) + 2y_1y_2(y_1 + y_2)^2]dy_1 \wedge dy_2.
\end{aligned}$$

Notice that $F^*\alpha$ is now a 2-form on $\mathbb{R}^2$. Thus $d(F^*\alpha)$ will be zero, as it is a 3-form on $\mathbb{R}^2$ (by Pigeonhole Principle).

Remark 3.22. For 1-forms $\alpha = dx_j$ on U', we can compute that

$$F^*\alpha(x)[v] = \alpha(F(x))[DF(x)[v]] = dx_j[DF(x)[v]] = DF_j(x)[v] = \sum_{i=1}^n \frac{\partial F_j(x)}{\partial x_i} v_i,$$

and hence

$$F^*dx_j(x) = \sum_{i=1}^n \frac{\partial F_j(x)}{\partial x_i} dx_i = dF_j(x)$$

Remark 3.23. For 0-forms (i.e., C^1 functions f), the exterior derivative coincides with out notion of differential:

$$\boxed{df(x)[v] = \sum_{i=1}^n \frac{\partial f(x)}{\partial x_i} dx_i[v] = \sum_{i=1}^n \frac{\partial f(x)}{\partial x_i} v_i = Df(x)[v]}$$

This is perfectly consistent with the previous notation $dx_i = e_i^*$ and actually justifies it.

Theorem 3.24 (Pullback with wedge product). *Given a C^2 map $F : U \to U'$ and two differential forms α, β on U', we have*

$$\boxed{F^*(\alpha \wedge \beta) = (F^*\alpha) \wedge (F^*\beta)}.$$

Proof.

First, note that, given $\alpha, \beta \in \Omega^k(U')$ and $\lambda, \mu \in \mathbb{R}$,

$$F^*(\lambda\alpha + \mu\beta) = \lambda F^*\alpha + \mu F^*\beta.$$

Therefore, it suffices to check the property for $\alpha = f dx_I$ and $\beta = g dx_J$, with $f, g \in C^1(U)$. A trivial computation shows that

$$F^*(\alpha \wedge \beta) = F^*\big(fg\, dx_I \wedge dx_J\big) = (f \circ F)(g \circ F)\, dF_I \wedge dF_j = (f \circ F)dF_I \wedge (g \circ F)dF_j = (F^*\alpha) \wedge (F^*\beta).$$

□

Theorem 3.25 (Pullback with Composition). *Given two C^2 maps $F : U \to U'$, $G : U' \to U''$ and a differential form α on U'', we have*

$$\boxed{(G \circ F)^*\alpha = F^*(G^*\alpha)}.$$

Indeed, given $x \in U$ and $v_1, \ldots, v_k$, we have

$$\begin{aligned}
F^*(G^*\alpha)(x)[v_1, \ldots, v_k] &= G^*\alpha(F(x))[DF(x)[v_1], \ldots, DF(x)[v_k]] \\
&= \alpha(G \circ F(x))\Big[\underbrace{DG(F(x))[DF(x)[v_1]]}_{D(G\circ F)(x)[v_1]}, \ldots, \underbrace{DG(F(x))[DF(x)[v_k]]}_{D(G\circ F)(x)[v_k]}\Big] \\
&= \alpha(G \circ F(x))\Big[D(G \circ F)[v_1], \ldots, D(G \circ F)[v_k]\Big] = (G \circ F)^*\alpha(x)[v_1, \ldots, v_k].
\end{aligned}$$

Intuitively, it's a double change of variables, reversing the order.

IMPROVE **Theorem 3.26 (Pullback and Exterior Derivative).** *Given $F : U \to U'$ of class C^2 and $\alpha \in \Omega^k(U')$, we have*

$$\boxed{d(F^*\alpha) = F^*(d\alpha)},$$

meaning pullback and exterior derivative commute.

Proof.
By linearity of both operations, it suffices to show this when α has the form $\alpha = f d_{i_1} \wedge \cdots \wedge dx_{i_k}$. We provide a sketch of the proof, which relies on a sequence of interesting identities:

$$\begin{gathered}
F^*(dx_i) = \sum_{j=1}^n \frac{\partial F_i}{\partial x_j} dx_j = dF_i \\
F^*\alpha = F^*(f)F^*(dx_{i_1}) \wedge \cdots \wedge F^*(dx_{i_j}))(f \circ F)dF_{i_1} \wedge \cdots \wedge dF_{i_k} \\
d(F^*\alpha) = d(f \circ F) \wedge dF_{i_1} \wedge \cdots \wedge dF_{i_k} \\
F^*(d\alpha) = F^*((df) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}) = F^*(df) \wedge dF_{i_1} \wedge \cdots \wedge dF_{i_k}.
\end{gathered}$$

To conclude, we simply note that

$$F^*(df)(x)[v] = df(F(x))[DF(x)[v]] = d(f \circ F)(x)[v],$$

since the exterior derivative of a 0-form is the usual differential of a function, giving the last equality by the chain rule. This implies that $F^*(df) = d(f \circ F)$, concluding the proof. □

3.3 Differential Forms on Manifolds

Previously, we defined differential k-forms in an *extrinsic* way, via Definition 3.13, relying on the ambient Euclidean space in which the manifold is embedded. We now turn to a more *intrinsic* perspective: differential forms will be defined directly on the manifold as **smooth fields of multilinear alternating forms on the tangent spaces**.

Definition 3.27 (Differential Forms (2)). A C^1 (resp. C^0) differential k-form α on a manifold M is a map $\alpha : M \to \Lambda^k T_pM$ such that, for every $p \in M$, there exists a C^1 (resp. C^0) differential form $\tilde{\alpha} \in \Omega^k(U)$, with $U \ni p$ open such that, for every $M \cap U$, $\alpha(x)$ is the restriction of $\tilde{\alpha}(x)$ to T_xM.[4]

[4] Another possible definition is the following: a C^1 (resp. C^0) differential k-form α on a manifold M is a map $\alpha : M \to \Lambda^k T_pM$

Intuition: the inclusion map $\iota\colon M \hookrightarrow \mathbb{R}^n$ sends each point $p \in M$ to itself in the ambient space $\mathbb{R}^n$. Therefore, we require that locally $\alpha = \iota^*\tilde{\alpha}$, i.e., α on M is obtained by restricting $\tilde{\alpha}(p)$ from $T_p\mathbb{R}^n \cong \mathbb{R}^n$ down to the subspace T_pM.

Example 3.28. Given the differential 3-form $\tilde{\alpha}(x) := dx_1 \wedge dx_2 \wedge dx_3$ on $\mathbb{R}^3$, its restriction α to S^2 (namely, $\alpha := \iota^*\tilde{\alpha}$ where $\iota\colon S^2 \to \mathbb{R}^3$ is the inclusion) is simply $\alpha = 0$, since for every $x \in S^2$ we have $\Lambda^3 T_x S^2 = 0$.
Intuitively, $\tilde{\alpha} = dx_1 \wedge dx_2 \wedge dx_3$ is the standard "unit" volume form in $\mathbb{R}^3$, a cube of volume 1. However, for the sphere S^2 the tangent space T_pS^2 is only 2-dimensional. Hence there is no room for nonzero infinitesimal 3-volumes: the pullback of any 3-form to a 2-dimensional space must vanish. Equivalently,

$$\iota^*(dx_1 \wedge dx_2 \wedge dx_3) \;=\; 0 \quad \text{on } S^2.$$

Definition 3.29 (Exterior Derivative (2)). Given a differential k-form α, we define the exterior derivative as the restriction of $d\tilde{\alpha}$, where $\tilde{\alpha}$ (which exists locally) is as in the previous definition. It can be proved that it is a well defined differential $k+1$-form of class C^0 on M.

4 Extrinsic Approach for Integration on Manifolds

In this section, we describe how one can define integration on smooth manifolds by viewing them as subsets of $\mathbb{R}^n$ and using the ambient Euclidean structure. This method, which relies on the k-dimensional Hausdorff measure (a measure theoretic tool), is referred to as an *extrinsic* approach, as it depends on how the manifold sits inside the surrounding space.
While it lacks the full generality and coordinate independence of the intrinsic theory developed via differential forms and volume elements, the extrinsic approach is intuitive, geometrically concrete, and often sufficient in practical applications.

4.1 Hausdorff Measure

We begin our discussion by rigorously defining the notion of area as a set measure. To streamline the exposition, we will take many key results from measure theory for granted and focus on the main ideas relevant to differential geometry. In particular, we introduce the Hausdorff measure: a generalization of the Lebesgue measure that allows us to assign a meaningful "size" to sets of arbitrary dimension, including fractals, curves, surfaces, and manifolds.

Definition 4.1 (Hausdorff Measure). Given a metric space $(\mathbb{R}^n, d)$, we define the k-dimensional Hausdorff Measure $\mathcal{H}^k : 2^{\mathbb{R}^n} \to [0, \infty]$ as

$$\boxed{\mathcal{H}^k(S) := \lim_{\delta \to 0^+} \inf \left\{ \sum_{i=1}^{\infty} \omega_k \left(\frac{\operatorname{diam}(E_i)}{2} \right)^k \,\middle|\, \{E_i\}_{i=1}^{\infty} \text{ is a } \delta\text{-cover of } S \right\}}.$$

- We also set $\mathcal{H}^k(\emptyset) = 0$.
- $\operatorname{diam}(E_i) = \sup\{d(x,y) \mid x, y \in E_i\}$.
- A δ-cover of S is a countable collection of sets $\{E_i\}$ such that $S \subseteq \bigcup_{i=1}^{\infty} E_i$ and $\operatorname{diam}(E_i) < \delta$ for every i.
- $\omega_k := \frac{\pi^{k/2}}{\Gamma\left(\frac{k}{2}+1\right)}$, with $\Gamma(t) := \displaystyle\int_0^{\infty} e^{-x} x^{t-1} \mathrm{d}x$, is the area of the k-dimensional unit ball, used as a reference. In particular, $\omega_0 = 1$, $\omega_1 = 2$, $\omega_2 = \pi$ and $\omega_3 = \frac{4\pi}{3}$.
- It can be shown that the infimum exists for every δ, and that the limit always exists since the interior function increases as we decrease $\delta > 0$.

Moreover, we have the following facts:

i) We have $\mathcal{H}^0(E) = \#E$, $\mathcal{H}^1(E) = \mathcal{L}^1(E)$ measures the length of curves, $\mathcal{H}^2(E) = \mathcal{L}^2(E)$ measures the area of surfaces, and, in general, $\mathcal{H}^n = \mathcal{L}^n$.

ii) A compact k-dimensional nonempty manifold has finite nonzero k-dimensional Hausdorff Measure.

such that, for every $p \in M$ and any parametrization $\psi : V \to M$, the pullback $\psi^*\alpha$ is a differential form of class C^1 on V. However, this definition is in general not equivalent to the previous one, since ψ is typically just C^1 and the pullback $\psi^*\alpha = \psi^*\iota^*\tilde{\alpha} = (\iota \circ \psi)^*\tilde{\alpha}$ is then typically just C^0.

iii) It is monotonic: if $S \subseteq S'$, then $\mathcal{H}^k(S) \leq \mathcal{H}^k(S')$;

iv) If $S \subseteq \bigcup_{i=1}^{\infty} S_i$ and $\{S_i\}_{i=1}^n$ is a family of disjoint sets, then $\mathcal{H}^k(S) \leq \sum_{i=1}^{\infty} \mathcal{H}^k(S_i)$, with equality if $S = \bigcup_{i=1}^{\infty} S_i$ and the sets S_i are Borel[5] (and disjoint);

v) For any Borel set $S \subseteq \mathbb{R}^n$ and $r > 0$, $a \in \mathbb{R}^n$, we have $\mathcal{H}^k(rS + a) = r^k \mathcal{H}^k(S)$.

If $k < n$, $\mathcal{H}^s$ is useful for measuring fractals or manifolds. For example, a smooth curve in $\mathbb{R}^2$ has zero Lebesgue measure but finite Hausdorff 1-measure. We can formalize this intuition with the following notion:

Definition 4.2 (Hausdorff Dimension). Given $S \subseteq \mathbb{R}^n$, the Hausdorff dimension of S is[6]

$$\dim_{\mathcal{H}}(S) := \inf\{k \geq 0 | \mathcal{H}^k(S) = 0\} = \sup\{k \geq 0 | \mathcal{H}^k(S) = \infty\}.$$
$$\dim_{\mathcal{H}}(S) = \sup\{k \geq 0 \mid \mathcal{H}^k(S) = \infty\} = \inf\{k \geq 0 \mid \mathcal{H}^k(E) = 0\}.$$

The Hausdorff Dimension of a k-dimensional manifold is k.

4.2 Integration on a Manifold

In this section we treat a k-dimensional manifold N as the image of a C^1 parametrization $f : M \to N$, where M is itself a k-dimensional manifold, possibly the cartesian product of open subsets of $\mathbb{R}$. As we will see, integrating over N amounts to integrating over M (pullback from N to M), up to a scalar factor called the Jacobian J_f, which quantifies how infinitesimal volumes are scaled under f. Morally, the formula is the following:

$$\int_N h \, \mathrm{d}V_N = \int_M (h \circ f) \, J_f \, \mathrm{d}V_M. \tag{4.1}$$

Definition 4.3 (Jacobian). Given a C^1 map $f : M \to N$ between two k-dimensional manifolds with $k \geq 1$, for any $p \in M$ we define the *Jacobian* of f at p as

$$J_f(p) := |\det(A)| \in [0, +\infty),$$

where A is a $k \times k$ matrix representing the differential $Df(p) : T_pM \to T_{f(p)}N$ with respect to an orthonormal basis for T_pM and an[7] orthonormal basis for $T_{f(p)}N$.

Remark 4.4. You might remember from Example 2.9 how long is to find the differential $Df(p) : T_pM \to T_{f(p)}N$ between two tangent spaces. Thankfully, it's enough to find the differential $Df(p) : T_pM \to \mathbb{R}^n$, where $\mathbb{R}^n$ is the space in which $T_{f(p)}N$ is embedded. In this case, given an orthonormal basis $\{v_1, \ldots, v_k\}$ of T_pM, the matrix representing the differential will be $B := \begin{pmatrix} Df(p)[v_1] & \ldots & Df(p)[v_k] \end{pmatrix}$, and the Jacobian is defined

$$J_f(p) = \sqrt{\det(B^T B)}.$$

To see why the two definitions are equivalent, one can take an orthonormal basis $\{w_1, \ldots, w_k\}$ of $T_{f(p)}N$ and A representing $Df(p) : T_pM \to T_{f(p)}N$ with respect to $(v_i)_{i \leq k}$ and $(w_i)_{i \leq k}$. Then

$$\sqrt{\det(B^T B)} = \sqrt{\det\left(A^T \begin{pmatrix} w_1^T \\ \vdots \\ w_k^T \end{pmatrix} \begin{pmatrix} w_1 & \ldots & w_k \end{pmatrix} A\right)} \overset{!}{=} \sqrt{\det(A^T A)} = \sqrt{(\det(A))^2} = |\det(A)|,$$

where ! is true by orthonormality.

[5] Borel sets are a wide family of non pathological sets which can be easily measured via Hausdorff (or Lebesgue) measure, or via probability measures. The Vitali set is an instance of a non Borel set, which is non measurable with respect to the Lebesgue measure, and yet still measurable with respect to the Hausdorff measure. Sets which are non measurable with respect to the Hausdorff measure do exist, but are out of the scope of these lecture notes.

[6] The interpretation is the following: for k too large, $\mathcal{H}^k(S) = 0$. In particular, $\mathcal{H}^k(S) = 0$ for any $k > n$. For k too small, $\mathcal{H}^k(S) = +\infty$. The dimension of S is the critical value $\dim_{\mathcal{H}}(S)$ which is the smallest value of k for which $\mathcal{H}^k(S) < +\infty$. For instance, the Sierpinki Triangle has Hausdorff dimension $\log(3)/\log(2) \approx 1.585$, meaning it is "more than a curve" but "less than a surface".

[7] $J_f(p)$ It is invariant to the choice of the orthonormal basis. This is because the change-of-basis matrix between two orthonormal bases is orthogonal, and thus has determinant equal to ± 1, and, therefore, $J_f(p)$ can only change in sign if different orthonormal bases are chosen. The determinant of an orthogonal basis P can be easily computed as $(\det(P))^2 = \det(PP^T) = \det(I) = 1$. Note, however, that our choice is computationally relevant. A good choice is to take an orthonormal basis $\{v_1, \ldots, v_k\}$ for T_pM such that the vectors

$$Df(p)[v_1], \ldots, Df(p)[v_k]$$

are orthogonal to each other, and such choice is always possible by the spectral theorem (actually, SVD).

Example 4.5. Taking for granted that $N = \{(x, y, z) \in \mathbb{R}^3 : z^2 = x^2 + y^2 - 1, |z| < 1\}$ is a 2-dimensional manifold, define a proper parametrization $\psi : M \to N$ (up to a negligible set) for it and compute its Jacobian.

Solution: There are at least two possible obvious parametrizations:

$$\psi : (0, 2\pi) \times (-1, 1) \to N \setminus S, \quad \psi \begin{pmatrix} \theta \\ t \end{pmatrix} = \begin{pmatrix} \sqrt{t^2+1}\cos\theta \\ \sqrt{t^2+1}\sin\theta \\ t \end{pmatrix}$$

and

$$\varphi : (0, 2\pi) \times (0, \sinh^{-1}(1)) \to N \setminus S', \quad \varphi \begin{pmatrix} \theta \\ \sigma \end{pmatrix} = \begin{pmatrix} \cos\theta\cosh\sigma \\ \sin\theta\cosh\sigma \\ \sinh\sigma \end{pmatrix},$$

both up to negligible sets S and S'. Let us focus on ψ. Since $T_{(\theta,t)}[(0, 2\pi) \times (-1, 1)] = \mathbb{R}^2$, an orthonormal basis for this tangent space is $\{(1, 0)^T, (0, 1)^T\}$. The differential of ψ is then represented by the matrix

$$B = \left(D\psi(\theta, t) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad D\psi(\theta, t) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{pmatrix} \frac{t}{\sqrt{t^2+1}}\cos\theta & -\sqrt{t^2+1}\sin\theta \\ \frac{t}{\sqrt{t^2+1}}\sin\theta & \sqrt{t^2+1}\cos\theta \\ 1 & 0 \end{pmatrix}.$$

We need to compute the determinant of $B^T B$. To simplify our computations, we recall the Cauchy-Binet formula: if $A \in \mathbb{R}^{m\times n}$, with $m \geq n$, then $\det(A^T A) = \sum_M \det(M)^2$, where M varies among all $n \times n$ minors of A. Thus

$$\begin{aligned} J_\psi(\theta, t) = \sqrt{\det(B^T B)} &= \sqrt{\begin{vmatrix} \frac{t\cos\theta}{\sqrt{t^2+1}} & -\sqrt{t^2+1}\sin\theta \\ \frac{t\sin\theta}{\sqrt{t^2+1}} & \sqrt{t^2+1}\cos\theta \end{vmatrix}^2 + \begin{vmatrix} \frac{t\cos\theta}{\sqrt{t^2+1}} & -\sqrt{t^2+1}\sin\theta \\ 1 & 0 \end{vmatrix}^2 + \begin{vmatrix} \frac{t\sin\theta}{\sqrt{t^2+1}} & +\sqrt{t^2+1}\cos\theta \\ 1 & 0 \end{vmatrix}} \\ &= \sqrt{t^2 + (t^2+1)\sin^2\theta + (t^2+1)\cos^2\theta} = \sqrt{2t^2+1}. \end{aligned}$$

The following theorem formalizes equation 4.1:

Theorem 4.6 (Area Formula). *Let $f : M \to N$ be a C^1 map between two k-dimensional manifolds with $k \geq 1$ and $h : N \to \mathbb{R}$. If $S \subseteq M$ is a Borel subset of M, then*

$$\boxed{\int_{f(S)} h(y)\mathrm{d}\mathcal{H}^k(y) = \int_S h(f(x))J_f(x)\mathrm{d}\mathcal{H}^k(x)}, \tag{4.2}$$

provided that h is integrable in $f(S)$ with respect to $\mathcal{H}^k(y)$. As a special case, if f is injective and $h = 1$, we obtain a formula for the area of $f(S)$:

$$\boxed{\mathcal{H}^k(f(S)) = \int_S J_f(x)\mathrm{d}\mathcal{H}^k(x)}.$$

Remark 4.7. Therefore, to compute the area of a manifold N it suffices to find one or more parametrizations the union of whose images is almost N (up to a negligible set), and then sum the contribution of each individual integral.

Remark 4.8. If f is not injective, then

$$\int_S J_f(x)\mathrm{d}\mathcal{H}^k(x) = \int_{f(S)} n(y)\mathrm{d}\mathcal{H}^k(y), \quad \text{with} \quad n(y) := \#\{x \in S : f(x) = y\}.$$

Remark 4.9. Typically $M \subseteq \mathbb{R}^k$ and, in this case, $\mathrm{d}\mathcal{H}^k(x) = \mathrm{d}x = \mathrm{d}x_1 \cdot \ldots \cdot \mathrm{d}x_k$.

Proof.

We provide a sketch of the proof in the case $h = 1$ and f is injective. The idea is to approximate M with a countable collection of k-dimensional simplexes, i.e., local portions of tangent spaces. Specifically, we consider a countable union $(x_i)_{i=1}^\infty$ of points in M and, for each of them, we approximate a portion

$S_i \subset M$ with a little set $x_i + S_i'$, where $S_i' \subset T_{x_i}M$ is a portion of tangent space. Therefore

$$\mathcal{H}^k(f(S)) \approx \sum_{i=1}^{\infty} \mathcal{H}^k(f(S_i)).$$

Each portion $f(S_i)$ is roughly $f(x_i) + Df(x_i)[S_i']$, where we approximated f itself with the linear map $Df(x_i) : T_{x_i}M \to T_{f(x_i)}N$. It is a general fact that a linear map $L_A : \mathbb{R}^k \to \mathbb{R}^k$ dilatates any set by $|\det(A)|$, where A is its matrix representation with respect to the canonical basis.[a] Therefore, the area of $f(x_i) + Df(x_i)[S_i']$ is

$$\mathcal{H}^k(f(x_i) + Df(x_i)[S_i']) = \mathcal{H}^k(Df(x_i)[S_i']) = J_f(x_i)\mathcal{H}^k(S'),$$

where $J_f(x_i)$ is the Jacobian factor, representing the dilatation due to the action of the differential operator. Finally,

$$\mathcal{H}^k(f(S)) \approx \sum_{i=1}^{\infty} \mathcal{H}^k(f(S_i)) \approx \sum_{i=1}^{\infty} \mathcal{H}^k(f(x_i) + Df(x_i)[S_i']) = \sum_{i=1}^{\infty} J_f(x_i)\mathcal{H}^k(S') \approx \int_M J_f(x) \mathrm{d}\mathcal{H}^k(x).$$

$\square$

[a] The reason is Corollary 6.7: we decompose A as $A = R'DR''$, with R' and R'' representing isometries (which do not change the area), and D with every entry stretching the correposponding direction by a factor equal to the eigenvalue in that entry. And we then recover the determinant easily.

Example 4.10. Following Example 4.5, we compute the integral of $h(x, y, z) = \frac{x^2+y^2}{\sqrt{2z^2+1}}$ on M. Applying equation 4.2 we get

$$\begin{aligned} \int_N h(y) \mathrm{d}\mathcal{H}^2(y) &= \int_{(0,2\pi)\times(-1,1)} h(\psi(\theta, t)) J_\psi(\theta, t) \mathrm{d}\mathcal{H}^2(\theta, t) \\ &= \int_0^{2\pi} \int_{-1}^{1} \frac{t^2 + 1}{\sqrt{2t^2 + 1}} \cdot \sqrt{2t^2 + 1} \mathrm{d}\theta \mathrm{d}t = 4\pi \int_0^1 t^2 + 1 \mathrm{d}t = \frac{16}{3}\pi. \end{aligned}$$

4.3 Charts and Immersions

CONTINUE...

In this optional section we will try to give framework to the results obtained so far.
A chart is a smooth map that provides local coordinates by means of a mapping, point by point, onto a subset of $\mathbb{R}^n$. In fact, no matter how curved the manifold may be, in order to understand it we need to flatten and stretch it into $\mathbb{R}^n$, since this is the only space we can fully handle. For instance, we flatten the globe into a map via Mercator Projection to fully understand where states are located. Of course, such a global flattening is not always possible, but it can always be done locally. By joining several charts together we obtain an atlas. The following is the formal definition of a manifold:

Definition 4.11 (Manifolds). A (smooth) n-manifold is a Hausdorff, second-countable topological space such that every point has a neighbourhood homeomorphic to $\mathbb{R}^n$.

Definition 4.12 (Chart). A *chart* is a pair (U, φ) with $U \subseteq M$ open and $\varphi : U \to \mathbb{R}^n$ a homeomorphism, providing local coordinates.

Given our notion of Φ, the local chart of M would be $\pi \circ \Phi|_{M\cap U}$, that is, the map Φ composed with a projection that maps $[\mathbb{R}^k \times \{0\}] \cap \Phi(U)$ to $\mathbb{R}^k$. Meanwhile, our notion of h corresponds with that of a *submersion*, since locally projects the ambient space into a lower dimensional space.

Definition 4.13 (Transition maps). Given two charts (U, φ) and (V, ψ) such that $U \cap V \neq \emptyset$, the *transition maps* are

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V) \qquad \text{and} \qquad \varphi \circ \psi^{-1} : \psi(U \cap V) \to \varphi(U \cap V).$$

Two charts are C^∞-compatible if the *transition maps* between them are C^∞ maps between open subsets of $\mathbb{R}^n$. A manifold is smooth if every pair of charts is C^∞ compatible. There exists a unique maximal smooth atlas, called a *smooth structure* on M.

Definition 4.14 (Atlas). An *atlas* is a collection of pairwise C^∞-compatible charts whose domains cover M.

Definition 4.15 (Submanifold). Given an ambient manifold $\tilde{M}$, a submanifold is a smooth manifold M together with an injective immersion $\iota : M \to \tilde{M}$. Identifying M with its image $\iota(M) \subset \tilde{M}$, we can consider M as a subset of $\tilde{M}$.

Example 4.16. Let $M = S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ and let $\tilde{M} = \mathbb{R}^3$. Consider the injective map $\iota : M \hookrightarrow \mathbb{R}^3$, defined by $\iota(p) = p$. The map ι is the natural **inclusion** (an identity **immersion**) that places the manifold M inside the ambient manifold $\tilde{M}$. Since $M \subseteq \mathbb{R}^3$ by definition, this inclusion is immediate. Moreover, ι is a homeomorphism, so it is also an **embedding**, and M is then *regular*. Now consider $f : \mathbb{R} \to \mathbb{R}^2$, $f(t) = (\cos t, \sin t)$. This map is an **immersion**, since it parametrizes a manifold in $\mathbb{R}^2$ using local coordinates and its differential is injective (never vanishing). However, f is not an embedding, because it is not injective and hence not a homeomorphism onto its image ($f(t) = f(t + 2\pi)$). Hence, the topology of $\mathbb{R}$ is different from that of a circle (the latter being compact). If instead we restrict the domain of f to $(0, 2\pi)$, we obtain an embedding: in this case, the immersed manifold is the circle with the point $(1, 0)$ removed.

5 Intrinsic Approach for Integration on Manifolds

In this section, we present an *intrinsic* approach to integration, showing how one can define the integration of differential forms on manifolds without any reference to an ambient space.

5.1 Oriented Manifolds and Oriented Maps

In a given n-dimensional vector space, it is always possible to partition the set of bases into exactly two equivalence classes, sometimes labeled as *positive orientation* $(+)$ and *negative orientation* $(-)$. We say that two bases B and B' have the same (resp. opposite) orientation, written $B \sim B'$, if the change-of-basis matrix from B' to B satisfies $\det(P) > 0$ (resp. $\det(P) < 0$).

Likewise, it is always possible to partition the set of isomorphisms (linear bijections) between vector spaces into exactly two equivalence classes. We say that a linear isomorphism $f : V \to W$ is *orientation preserving* or *projecting* or *proper* (resp. *orientation reversing*, or *reflecting* or *improper*) if $\det(f) > 0$ (resp. $\det(f) < 0$).[8] We denote f_* the operator that associates, to each orientation $\mathcal{O}$ of a basis $\{v_1, \dots, v_n\}$, the orientation induced by $\{f(v_1), \dots, f(v_n)\}$. In other words, indicating by ± 1 the two orientations of the bases in V, then $f_*(\pm 1) = \pm 1$, or $f_*(\pm 1) = \mp 1$.

As previously mentioned, a linear function can either be a *projecting isomorphism*, a *reflecting isomorphism* or non-invertible, depending on the sign of the determinant. As for locally invertible nonlinear functions ψ, their local behaviour is determined by their local linear approximation, i.e. $D\psi$. Again, a nonlinear parametrization $\psi : V \to M$ is *orientation preserving*, or simply *oriented*, if, for every $a \in V$, $\det\big(D\psi(a)\big) > 0$;[9] ψ is *orientation reversing* if, for every $a \in V$, $\det\big(D\psi(a)\big) < 0$. Finally, if the sign of $\det D\psi$ is mixed, then continuity forces a point where $\det D\psi = 0$. At that point ψ ceases to be a local diffeomorphism (hence fails to be locally injective) and so is not a valid parametrization.
In practice, when integrating we should avoid this third case (or at worst allow isolated zeros on negligible sets), because otherwise overlapping parameter-values or collapsed volumes destroy the straightforward change-of-variables formula and our control over the integral.

Remark 5.1. Given $B = \{v_1, v_2, \dots, v_n\}$, there are two simple ways to invert the orientation: the first is inverting the order of two vectors in the basis, as in $B' = \{v_2, v_1, v_3, \dots, v_n\}$ and the second is to change sign to a vector, as in $B'' = \{-v_1, v_2, \dots, v_n\}$.

Thanks to this remark, it's clear that orientating a basis amounts to choosing the proper order for the vectors composing the basis (swapping any two basis vectors reverses the orientation; note the analogy with differential k-forms). Now imagine to take a basis B_p of T_pM for each $p \in M$, where M is a manifold of dimension $k \geq 1$. For each basis chosen, consider its orientation $\mathcal{O}_p$. Intuitively, we say that $(M, \mathcal{O}_p)$ is an oriented manifold if the orientation (not the basis![10]) of each tangent space T_pM is a continuous choice, i.e., the orientation stays constant (always positive, or always negative) on a connected component of M. One might then ask how can we compare the orientations of bases in different vector spaces (T_pM and T_qM, with $p \neq q$). The following definition formalizes the intuition and answers this question:

[8] Of course, $\det(f) = 0$ is not contemplated, since a linear map f is an isomorphism if and only if $\det(f) \neq 0$. Orientation is well defined only for isomorphisms, since we should be able to carry out information from one space to the other, and to reverse the process via the inverse isomorphism.

[9] In practice, one has $D\psi(a)_*(\mathcal{O}_{\mathbb{R}^k}) = \mathcal{O}_{\psi(a)}$

[10] Using tools from differential topology, it's possible to prove that it is impossible to assign a basis to each tangent space in a continuous way on S^2 (even if S^n is orientable for every n)

Definition 5.2 (Oriented Manifold). $(M, \mathcal{O}_p)$ is an *oriented* manifold, with *orientation* $\mathcal{O}_p$, if, for any parametrization $\psi : V \to M$ with connected domain $V \subseteq \mathbb{R}^k$, we have

$$(D\psi(a)^{-1})_*(\mathcal{O}_{\psi(a)}) = (D\psi(b)^{-1})_*(\mathcal{O}_{\psi(b)})$$

for every $a, b \in V$, where $D\psi(a) : \mathbb{R}^k \to T_{\psi(a)}M$ (and the same for b).

Remark 5.3. By considering the inverse maps $\left([D\psi(a)]^{-1}\right)_*$ and $\left([D\psi(b)]^{-1}\right)_*$ we are transferring the orientation $\mathcal{O}_{\psi(a)}$ of the chosen basis of $T_{\psi(a)}M$ to $\mathbb{R}^k$, and similarly for $\mathcal{O}_{\psi(b)}$. Now that we have orientations on the same space $\mathbb{R}^k$, we can impose those orientations to be coherent, i.e., constant, for every point $\psi(a)$ on M (actually, for every point $\psi(a)$ on the same connected component of M, since we shouldn't bother of comparing points on different connected components of M).

Remark 5.4. A manifold is *orientable* if it admits an orientation and *non-orientable* otherwise. There are examples of manifolds which do not admit an orientation.

Example 5.5. On the unit sphere $S^2 \subseteq \mathbb{R}^3$, parametrized by

$$\psi : (0, \pi) \times (0, 2\pi) \to S^2, \qquad \psi(\theta) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

up to a negligible set, the differential $D\psi(\theta, \varphi)$ can be easily computed by noting that $T_{(\theta,\varphi)}((0, \pi) \times (0, 2\pi)) \cong \mathbb{R}^2$. Consequently, a basis for it is $\{(1, 0)^T, (0, 1)^T\}$. Since the differential $D\psi(\theta, \varphi) : T_{(\theta,\varphi)}\mathbb{R}^2 \to T_{\psi(\theta,\varphi)}S^2$ is a linear isomorphism, to find a basis for $T_{\psi(\theta,\varphi)}S^2$ is enough to find generating vectors for the span of $D\psi(\theta, \varphi)$. Thus we compute

$$\begin{aligned} D\psi(\theta, \varphi)(e_1) &= \frac{\partial\psi}{\partial\theta} = (\cos\theta, \cos\varphi, \cos\theta\sin\varphi, -\sin\theta) \\ D\psi(\theta, \varphi)(e_2) &= \frac{\partial\psi}{\partial\varphi} = (-\sin\theta\sin\varphi, \sin\theta\cos\varphi, 0). \end{aligned}$$

We now need to find a basis for the space spanned by those two vectors, for a fixed $(\theta, \varphi) \in (0, \pi) \times (0, 2\pi)$. Computations are easier since

$$\det[D\psi(\theta, \varphi)] = \sin\theta > 0 \quad \forall(\theta, \varphi) \in (0, \pi) \times (0, 2\pi).$$

As a consequence, $\{D\psi(\theta, \varphi)(e_1), D\psi(\theta, \varphi)(e_2)\}$ is a basis for $T_{\psi(\theta,\varphi)}S^2$ and it has positive orientation everywhere.
Hence for any two points $a = (\theta_a, \varphi_a)$ and $b = (\theta_b, \varphi_b)$ in U we have

$$(D\psi(a)^{-1})_*(\mathcal{O}_{\psi(a)}) \;=\; (D\psi(b)^{-1})_*(\mathcal{O}_{\psi(b)})$$

because pulling back the "positive" basis $\{D\psi(e_1), D\psi(e_2)\}$ always yields the standard orientation on $\mathbb{R}^2$. This shows that $(S^2, \mathcal{O}_p)$ is oriented, up to the poles. By selecting a different parametrization with the same property, we can show that the orientation is preserved also on the poles, concluding that S^2 is an oriented manifold.

Theorem 5.6. *If $M = \{x \in U : h(x) = 0\}$ is the zero set of a C^1 function $h : U \to \mathbb{R}^{n-k}$ and $Dh(p)$ is surjective for every $p \in M$, then M is orientable.*[11]

Proof.
We provide a sketch of the proof. To construct an orientation $\mathcal{O}_p$ for M, we need to find, for every $p \in M$, a basis $\{v_1, \ldots, v_k\}$ of T_pM which induces a proper orientation $\mathcal{O}_p$. Since

$$T_pM = \operatorname{span}\{\nabla h_1(p), \ldots, \nabla h_{n-k}(p)\}^\perp$$

and the vectors $\nabla h_i(p)$ are linearly independent ($Dh(p)$ is surjective), then, adding a basis $\{v_1, \ldots, v_k\}$ of T_pM to $\{\nabla h_1(p), \ldots, \nabla h_{n-k}(p)\}$ produces a basis of $\mathbb{R}^n$. To produce $\mathcal{O}_p$, we use $\{\pm v_1, v_2, \ldots, v_k\}$, with the sign such that

$$\{\nabla h_1(p), \ldots, \nabla h_{n-k}(p), \pm v_1, \ldots, v_k\} \sim \{e_1, \ldots, e_n\},$$

that is, has the same orientation as the canonical basis.

[11] Every manifold is locally the zero set of a function $h : U \to \mathbb{R}^{n-k}$. If the same function can also be used globally, then M is orientable.

For what we said before, it's clear that a basis $\{\pm v_1, \dots, v_k\}$ of T_pM satisfying these conditions always exists. Suppose now there are two bases $\{v_1, \dots, v_k\}$ and $\{w_1, \dots, w_k\}$ of T_pM such that $\{\nabla h_1(p), \dots, \nabla h_{n-k}(p), v_1, \dots, v_k\} \sim \{\nabla h_1(p), \dots, \nabla h_{n-k}(p), w_1, \dots, w_k\}$. Then also $\{v_1, \dots, v_k\} \sim \{w_1, \dots, w_k\}$, which can be proved with analysing the change-of-basis matrix from one basis to the other. This shows that, for every p, we have a well-defined orientation $\mathcal{O}_p$ of T_pM.

Fix now a parametrization $\psi : V \to M$ with connected domain $V \subseteq \mathbb{R}^k$. Our aim is to prove that the family $\{\mathcal{O}_{\psi(x)}\}_{x \in V}$ is constant, i.e. $(D\psi(x)^{-1})_*(\mathcal{O}_{\psi(x)})$ does not depend on x. Pick a base point $a \in V$ and choose a corresponding basis $\{v_1, \dots, v_k\} \in \mathcal{O}_{\psi(a)}$ as earlier. If we can exhibit a smooth family of vectors $w_i(x) \in T_{\psi(x)}M$ such that $w_i(a) = v_i$ and

$$\det[\nabla h_1(\psi(x)), \dots, \nabla h_{n-k}(\psi(x)), w_1(x), \dots, w_k(x)] > 0, \quad \forall x \in V, \tag{5.1}$$

then, by definition of $\mathcal{O}_{\psi(x)}$, the list $\{w_1(x), \dots, w_k(x)\}$ belongs to $\mathcal{O}_{\psi(x)}$ for every x. Consequently the orientations $(D\psi(x)^{-1})_*(\mathcal{O}_{\psi(x)})$ all coincide with the orientation of $\mathbb{R}^k$, hence are constant on V. So it only remains to construct such $w_i(x)$.

Set $z_i := D\psi(a)^{-1}[v_i] \in \mathbb{R}^k$ for $i = 1, \dots, k$. Because ψ is regular, $D\psi(a)$ has full rank, and is a linear isomorphism, so it maps the basis $\{v_1, \dots, v_k\}$ into $\{z_1, \dots, z_k\}$, which must be then a basis of $\mathbb{R}^k$. For every $x \in V$ define

$$w_i(x) := D\psi(x)[z_i] = D\psi(x)[D\psi(a)^{-1}[v_i]].$$

The map $x \mapsto w_i(x)$ is C^1 since $D\psi$ is, and $w_i(a) = v_i$. Moreover, with such definition of $w_i(x)$, the determinant in (5.1) is positive for every $x \in V$, since it is satisfied at the base point $x = a$, varies continuously in x and it cannot vanish.
Because $\{w_1(x), \dots, w_k(x)\} \subset \mathcal{O}_{\psi(x)}$ for every x, and $(D\psi(x)^{-1})_*(\mathcal{O}_{\psi(x)})$ is the orientation of $\mathbb{R}^k$ determined once for all by $\{z_1, \dots, z_k\}$, the family $(D\psi(x)^{-1})_*(\mathcal{O}_{\psi(x)})$ is indeed constant on V. □

5.2 Integrating a differential k-form

Henceforth, M will denote a k-dimensional oriented manifold and α will denote a C^0 differential form. Before starting, let us mention two measure theoretic existence results which will be useful for the definition of the integral. There will be two definitions, one for each of the following properties. We can always find:

- A countable family of parametrizations $\psi_i : V_i \to M$ and Borel subsets $E_i \subseteq V_i$ such that M is the disjoint union of the images $\psi_i(E_i)$;
- A countable family of parametrizations $\psi_i : V_i \to M$ such that M is the disjoint union of the images $\psi_i(V_i)$, plus a negligible set N, such that $\mathcal{H}^k(N) = 0$.

If M is oriented, we can require that each parametrization is oriented. If M is compact, we can even require a finite number of parametrizations and Borel sets, with connected domains.

Definition 5.7 (Integral of a Differential k-form). In the previous setting, given a differential k-form α and a family of *oriented* parametrizations ψ_i, we let

$$\int_M \alpha := \sum_i \int_{E_i} \psi_i^* \alpha \quad \text{or} \quad \int_M \alpha := \sum_i \int_{W_i} \phi_i^* \alpha.$$

Remark 5.8. Note the difference: in the first case we are taking Borel subsets and we cover M completely. In the second case, we do not take Borel subsets, and we cover M up to a negligible set. But the integral in the two cases coincides.[12]

Example 5.9.

ADD...

5.3 Connection Between Intrinsic and Extrinsic Approaches

Eventually, we clarify the geometric intuition behind differential forms and their relation to the extrinsic perspective. In particular, we embrace the physicists' interpretation of differential forms as representing infinitesimal "volume patches" (or "area patches") to build an intuitive understanding of integration on manifolds.

[12] The fact that the choice of a family of parametrizations is not important, as long as they preserve the orientation, parallels the fact that line integrals in complex analysis are invariant under reparametrizations.

The connection between the two approaches is given by a differential k-form ω with, conventionally, unit volume. If the manifold is oriented and endowed with an inner product, we can construct such a form canonically:

Definition 5.10 (Volume Form). Let $(M, \mathcal{O}_p)$ be an oriented k-dimensional manifold. At each point $p \in M$, let $\{v_1, \ldots, v_k\}$ be an orthonormal basis of T_pM compatible with the orientation $\mathcal{O}_p$. The *volume form* of M (of class C^1) is differential k-form given by

$$\omega(p) := v_1^* \wedge \cdots \wedge v_k^*,$$

and it is independent on the choice of orthonormal basis.

Intuitively, $\omega(p)$ acts as a "unit volume patch" centered at p. When evaluated on vectors $v_1, \ldots, v_k \in T_pM$, it yields

$$\omega(p)[v_1, \ldots, v_k] = \det \begin{pmatrix} v_1^*(v_1) & \ldots & v_1^*(v_k) \\ \vdots & & \vdots \\ v_k^*(v_1) & \ldots & v_k^*(v_k) \end{pmatrix},$$

which measures the *oriented volume* of the parallelepiped formed by the v_i's.
Geometrically, v_i^* represents an infinitesimal increment dv_i in the direction of v_i, and the wedge product $\wedge$ combines directions to form higher-dimensional oriented volume elements: for instance, $v_i^* \wedge v_j^*$ is an infinitesimal oriented area element (parallelogram) in the plane spanned by v_i and v_j. The antisymmetry of the wedge product naturally encodes orientation. Thus, $\omega(p)$ represents an infinitesimal k-volume element whose geometric shape (cube, parallelepiped, cylinder...) depends on the chosen coordinates, as ω provides a coordinate-free description.
In order to express this unitary measure in Hausdorff sense, we define an inner product on the space $\Lambda^k T_pM \simeq \mathbb{R}$ by requiring that

$$\langle \omega(p), \omega(p) \rangle = \|\omega(p)\|^2 = 1,$$

which expresses the unitary nature of the volume form. In this setting, integration can be intrinsically described using ω and the k-dimensional Hausdorff measure $\mathcal{H}^k$:

$$\boxed{\int_M f \, d\mathcal{H}^k = \int_M f \, \omega} \qquad \text{and} \qquad \boxed{\int_M \alpha = \int_M \langle \alpha, \omega \rangle \, d\mathcal{H}^k},$$

where f is a scalar function and $\alpha \in \Omega^k(M)$ is a k-form.
IMPROVE The first identity expresses that integrating a function with respect to the Hausdorff measure is equivalent to integrating it against the unit volume form ω. The second can be understood with the following (intuitive) reasoning:

$$\int_M \alpha \overset{1}{=} \int_V \varphi^* \alpha \overset{2}{=} \int_V \langle \alpha, \omega \rangle \, dx_1 \ldots dx_k = \int_M \langle \alpha, \omega \rangle \, d\mathcal{H}^k,$$

where 1 writes M in local coordinates and 2 interprets the integration as a sum of the evaluations of multilinear forms on infinitesimal parallelepipeds, specifically, on the tangent vectors $d\varphi(\partial/\partial x_i)$.

Remark 5.11. In the extrinsic setting, it is clear that the integral may diverge if the manifold M is non-compact or if the differential form α does not have compact support. Therefore, we should require

$$\int_M \|\alpha(x)\| \, d\mathcal{H}^k(x) < +\infty,$$

where $\|\cdot\|$ is the norm induced on k-forms by the inner product on $\Lambda^k T_pM$.

5.4 Stokes' Theorem

All the machinery developed so far, differentiable manifolds, differential forms, pullbacks, and the exterior derivative, culminates in a single, unifying result: the general form of Stokes' Theorem. This theorem stands as the central result of differential geometry in its purely smooth setting, before entering the realms of topology or Riemannian geometry. In order to introduce it, we are only left with the notion of manifolds with boundary.
Intuitively, $M = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ is not a manifold. However, we say that M is a *manifold with boundary*, $\partial M = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is the *boundary* of M, and $M' = M \setminus \partial M$ is a manifold in the usual sense.

Definition 5.12 (Manifold with Boundary). Given $k \geq 1$, a k-dimensional manifold with boundary is a subset $M \subseteq \mathbb{R}^n$ such that, for any $p \in M$, there exists a C^1 local diffeomorphism $\Phi : U \to \mathbb{R}^n$ with its image from an open set $U \subseteq \mathbb{R}^n$ containing p such that either

$$M \cap U = \{x \in U : \Phi(x) \in \mathbb{R}^k \times \{0\}\}, \tag{5.2}$$

or[13]

$$M \cap U = \{x \in U : \Phi(x) \in (-\infty, 0] \times \mathbb{R}^{k-1} \times \{0\}\}. \tag{5.3}$$

Definition 5.13 (Boundary of a Manifold). Given a k-dimensional manifold M with boundary, we call boundary $\partial M \subseteq M$ the set of points p such that M at p resembles a k-dimensional closed half-plane up to a diffeomorphism, i.e., satisfying (5.3).

Remark 5.14. Morally, a manifold M with boundary is a manifold M' plus some points ∂M that, added to M', are such that $M' \cup \partial M$ resembles a closed half-space.

Proposition 5.15. *The set ∂M is a $(k-1)$-dimensional manifold.*

If M is oriented, then ∂M inherits a *canonical orientation*, called *boundary orientation*. Given Φ as above, if $\det(D\Phi(p)) > 0$, i.e. $D\Phi(p) : T_pM \to \mathbb{R}^k \times \{0\}$ preserves the orientation, then we simply take the orientation of $T_p(\partial M)$ given by the basis

$$\{D\Phi(p)^{-1}[e_2], \dots, D\Phi(p)^{-1}[e_k]\};$$

if $\det(D\Phi(p)) < 0$, we take the opposite. One checks easily that this construction does not depend on the choice of chart Φ.
In practice, one may use the *outer-normal-first* rule: choose any vector $\nu \in T_pM$ pointing "outward" of M and use it to complete a basis $\{v_2, \dots, v_k\}$ of $T_p(\partial M)$ to an oriented basis $\{\nu, v_2, \dots, v_k\}$ of T_pM. Then $\{v_2, \dots, v_k\}$ induces the boundary orientation if $\{\nu, v_2, \dots, v_k\}$ has the orientation of M.

Example 5.16. The closed unit ball $\overline{B}_1(0) \subset \mathbb{R}^n$ is a compact, oriented manifold with boundary, endowed with the standard orientation of $\mathbb{R}^n$. Its boundary is the sphere $S^{n-1} = \partial\overline{B}_1(0)$. At each point $x \in S^{n-1}$, the outward-pointing normal in $T_x\overline{B}_1(0)$ is exactly the radial vector x. Hence, if we choose any basis $\{v_2, \dots, v_n\}$ of the tangent space T_xS^{n-1}, then the basis

$$\{\, x,\ v_2, \dots, v_n\}$$

of $T_x\overline{B}_1(0)$ agrees with the standard orientation of $\mathbb{R}^n$. By the outer-normal-first convention, this induces the canonical boundary orientation on S^{n-1}.

Remark 5.17. Given a manifold with boundary M, then $M \setminus \partial M$ is an ordinary manifold without boundary, where integration is defined as usual. To integrate on the boundary, we should define parametrizations for M in a suitable way to take into account the boundary points; they are obtained by restricting C^1 maps to the intersection of V with a closed halfspace. Integration on M is then defined by pulling back forms to these half-space parametrizations and integrating as usual.

Theorem 5.18 (Stokes' Theorem). *Let M be a compact, oriented, k-dimensional manifold ($k \geq 1$) with boundary, possibly $\partial M = \emptyset$. Given a differential $(k-1)$-form α of class C^1, we have*

$$\int_M d\alpha = \oint_{\partial M} \alpha.$$

Next, we list some trivial observations:

i) If $\partial M = \emptyset$, so that M is a compact, oriented manifold in the usual sense, then $\int_M d\alpha = 0$.

ii) If $k = 1$ and $M = [a, b]$, Stokes' theorem becomes the Fundamental Theorem of Calculus.

iii) With some work, the divergence theorem could be obtained as a special case of Stokes.

iv) If we apply Stokes twice on $d(d\beta)$, with β being a differential $(k-2)$-form, we get

$$\int_M d(d\beta) = \int_{\partial M} d\beta = \int_{\partial(\partial M)} \beta = 0,$$

since the boundary ∂M is always a manifold in the usual sense (without boundary). In some sense, the fact that $\partial \circ \partial = 0$ is the dual notion of $d \circ d = 0$ (we already know that $d(d\beta) = 0$.

Example 5.19.

ADD...

[13] It can be shown that those conditions cannot both occur, as the choice of the diffeomorphism is irrelevant.

6 Constrained Optimization

In this final section, we study optimization problems under constraints, introducing the principle of Lagrange multipliers and the Karush–Kuhn–Tucker (KKT) conditions. As an application, we provide a variational proof of the spectral theorem for symmetric matrices.

6.1 Lagrange Multipliers

Theorem 6.1 (Fermat). *Given a C^1 function $f : M \to \mathbb{R}$ on a manifold M, if x_0 is an optimum point for f, then x_0 is a critical point, i.e., $Df(x_0) = 0$ on $T_{x_0}M$*

Proof. Without loss of generality, assume that x_0 is a minimum point: indeed, if x_0 is a maximum point, then we can replace f with $-f$. Take a C^1 parametrization $\gamma : [0, \varepsilon] \to M$ with $\gamma(0) = x_0$. Then $f \circ \gamma$ is C^1 and

$$f \circ \gamma(t) \geq f(x_0) = f \circ \gamma(0) \quad \text{and} \quad Df(x_0)[\gamma'(0)] = (f \circ \gamma)'(0) = 0 \implies Df(x_0)[v] \geq 0 \quad \forall v \in T_{x_0}M.$$

Replacing v with $-v$ we also have $Df(x_0) \leq 0$, concluding that $Df(x_0) = 0$. □

Theorem 6.2 (Lagrange Multipliers). *If x_0 is a constrained minimum (or maximum) point for f, among points $x \in U \subseteq \mathbb{R}^n$ subject to the constraint $h_1(x) = \cdots = h_\ell(x) = 0$, then there exist $\mu_1, \ldots, \mu_\ell \in \mathbb{R}$, called Lagrange multipliers, such that*

$$\nabla f(x_0) + \mu_1 \nabla h_1(x_0) + \cdots + \mu_\ell \nabla h_\ell(x_0) = 0,$$

provided that the vectors $\nabla h_1(x_0), \ldots, \nabla h_\ell(x_0)$ are linearly independent (non degeneracy condition).

Proof.
The thesis is equivalent to

$$\nabla f(x_0) \in \operatorname{span}\{\nabla h_i(x_0) \mid i = 1, \ldots, \ell\}.$$

Consider the function

$$h := (h_1, \ldots, h_\ell) : U \to \mathbb{R}^\ell,$$

where $U \subseteq \mathbb{R}^n$. Its differential $Dh(x_0)$ is an $\ell \times n$ matrix of rank ℓ, since its rows are $\nabla h_i(x_0)^T$, which are linearly independent. Therefore, at x_0 there exists an $\ell \times \ell$ minor of $Dh(x)$ with nonzero determinant. Since this determinant varies continuously in x, we deduce the same property in a neighbourhood $U' \subseteq U$ of x. Thus, $Dh(x)$ has full rank in U'. Define

$$M := \{x \in U' : h(x) = 0\}.$$

We see that M is a $n - \ell$ dimensional manifold with tangent space

$$T_{x_0}M = \ker Dh(x_0) = \{v : \langle \nabla h_i(x_0), v\rangle = 0 \quad \text{for all } i = 1, \ldots, \ell\} = \left(\operatorname{span}\{\nabla h_i(x_0) | i = 1, \ldots, \ell\}\right)^\perp.$$

Therefore, we only need to show that $\nabla f(x_0) \perp T_{x_0}M$, which in turn is equivalent to $\langle \nabla f(x_0), v\rangle = 0$ for every $v \in T_{x_0}M$. We thus compute

$$\langle \nabla f(x_0), v\rangle = Df(x_0)[v] = D\tilde{f}(x_0)[v] = 0,$$

where $\tilde{f} : M \to \mathbb{R}$ is the restriction of f on M, to which we apply Fermat's Theorem. □

6.2 Application: Spectral Theorem

In this section we present a digression on spectral theory. Our goal is to prove the spectral theorem for real symmetric matrices using only the principle of Lagrange multipliers, rather than relying on the fundamental theorem of algebra to guarantee the existence of eigenvalues.

Theorem 6.3 (Spectral Theorem). *Given a linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ given by $f(x) = Ax$ for a symmetric matrix A, there exists an orthonormal basis $\{v_1, \ldots, v_n\}$ consisting of eigenvectors of A.*

Proof.

We begin by showing that a real eigenvalue for A exists. Consider the constrained optimization problem

$$\min \underbrace{x^T A x}_{f(x)} \qquad \text{sub } x \in S^{n-1} = \big\{x \in \mathbb{R}^n : \underbrace{|x|^2 - 1}_{h(x)} = 0\big\}.$$

Since f is C^1 and S^{n-1} is compact, a minimum x_0 exists. Since $\nabla h(x) = 2x \neq 0$ for every $x \in S^{n-1}$, we can apply the Lagrange multipliers principle, giving the existence of $\lambda \in \mathbb{R}$ such that

$$\nabla f(x_0) + \lambda \nabla h(x_0) = 0.$$

To find $\nabla f(x_0)$, let us compute $Df(x_0)[v]$:

$$\begin{aligned}
Df(x_0)[v] &= \lim_{\varepsilon\to 0} \frac{f(x_0+\varepsilon v) - f(x_0)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{(x_0+\varepsilon v)^T A (x_0+\varepsilon v) - x_0^T A x_0}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \frac{x_0^T A x_0 + \varepsilon (x_0^T A v + v^T A x_0) + \varepsilon^2 v^T A v - x_0^T A x_0}{\varepsilon} \\
&= x_0^T A v + v^T A x_0 = v^T A^T x_0 + v^T A x_0 = 2 v^T A x_0 = \langle v, 2Ax_0\rangle,
\end{aligned}$$

where in the last passage we used the symmetry of A. Since this is valid for every $v \in \mathbb{R}^n$, we deduce that $\nabla f(x_0) = 2Ax_0$. hence, we obtain

$$2Ax_0 = -\lambda 2 x_0.$$

Since $x_0 \neq 0$ as $x_0 \in S^{n-1}$, we found an eigenvector $x_0 \neq 0$ and a corresponding eigenvalue $-\lambda \in \mathbb{R}$ for A.

Next, we show, by induction on n, the existence of an orthonormal basis of eigenvectors of A. For $n = 1$ there is nothing to prove, since $[\lambda]$ is already diagonalized.
Assuming the statement true for $k < n$, let us prove it for $k = n$. Since we just proved that an eigenvector v_1 exists, we can normalize it to $e_1 = v_1/\|v_1\|$ and then complete $\{e_1\}$ to an orthonormal basis $\{e_1, \dots, e_n\}$ with Gram Schmidt. Let $P := [e_1 \mid e_2 \mid \cdots \mid e_n]$. Then

$$P^{-1} A\, P = \begin{pmatrix} \lambda_1 & * \\ * & C \end{pmatrix},$$

where C is an $(n-1)\times(n-1)$ matrix and λ_1 is the eigenvalue corresponding to v_1. Since A is symmetric and symmetric matrices are closed under orthogonal change of basis, then $P^{-1}AP$ is symmetric and C must also be symmetric. By inductive hypothesis, we know there exists an orthonormal basis of $\mathbb{R}^{n-1}$ made of eigenvectors of C. Let us call this $\{u_2, \dots, u_n\}$. Then define

$$e_1, \quad v_i = \sum_{j=2}^{n} (u_i)_j\, w_j \quad (i = 2, \dots, n).$$

Then $\{e_1, v_2, \dots, v_n\}$ is an orthonormal basis of eigenvectors of A, with real eigenvalues. □

Actually, there is a second formulation of the spectral theorem. First, let us recall a fundamental result on bilinear forms:

Definition 6.4 (Scalar Product). Given a vector space V (finite or infinite dimensional) on $\mathbb{R}$, a scalar product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is an operator satisfying the following properties:

i) *Symmetry*: $\langle x, y\rangle = \langle y, x\rangle$;

ii) *Linearity:* $\langle \lambda x + \mu y, z\rangle = \lambda\langle x, z\rangle + \mu\langle y, z\rangle$. Combining this property with symmetry we deduce that the scalar product is a bilinear form.

iii) *Positive-Definiteness*: $\langle x, x\rangle \geq 0$, with the equality if and only if $x = 0$.

Theorem 6.5 (Bilinear Form). *Given a scalar product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$, where $\dim V = n$, there exists a unique symmetric and positive definite matrix $A \in \mathbb{R}^{n\times n}$ with respect to the chosen basis, such that, for every $x, y \in V$,*

$$\langle x, y\rangle = x^T A y$$

Proof.
Taking a basis $\{e_1, \dots, e_n\}$ of V, write $x = \sum_{i=1}^n x_i e_i$ and $y = \sum_{j=1}^n y_j e_j$ and then use bilinearity:

$$\langle x, y \rangle = \left\langle \sum_{i=1}^n x_i e_i, \sum_{i=1}^n y_j e_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n x_i y_j \langle e_i, e_j \rangle = x^T A y, \quad \text{where} \quad A_{ij} = \langle e_i, e_j \rangle.$$

Using symmetry and positive-definiteness we also conclude that A is symmetric and positive definite. □

Theorem 6.6 (Spectral theorem (bilinear forms)). *Any nondegenerate symmetric bilinear form can be diagonalized in an orthonormal basis for a fixed positive-definite inner product.*
More precisely, let V be a finite-dimensional real vector space equipped with two symmetric bilinear forms

$$\langle x, y \rangle_1, \quad \langle x, y \rangle_2,$$

where $\langle \cdot, \cdot \rangle_1$ is positive-definite (i.e., it is a scalar product) and $\langle \cdot, \cdot \rangle_2$ is nondegenerate. Then there exists a basis $\{v_1, \dots, v_n\}$ of V which is orthonormal with respect to $\langle \cdot, \cdot \rangle_1$ and orthogonal with respect to $\langle \cdot, \cdot \rangle_2$.

Proof.
We now sketch the equivalence of the two formulations. First, choose a basis of V that is orthonormal for $\langle \cdot, \cdot \rangle_1$. In that basis, applying 6.5, there exists a unique symmetric matrix $A \in \mathbb{R}^{n \times n}$

$$\langle x, y \rangle_1 = x^T y, \quad \langle x, y \rangle_2 = x^T A\, y, \quad \forall x, y \in \mathbb{R}^n.$$

Since A is real and symmetric, the standard Spectral Theorem gives an orthonormal basis $\{v_1, \dots, v_n\}$ (with respect to the standard inner product $x^T y$) consisting of eigenvectors of A. Hence $\{v_i\}$ is orthonormal for $\langle \cdot, \cdot \rangle_1$. We now check orthonormality with respect to $\langle \cdot, \cdot \rangle_2$. For every i, j,

$$\langle v_i, v_j \rangle_2 = v_i^T\, A\, v_j = v_i^T (\lambda_j\, v_j) = \lambda_j\, (v_i^T v_j) = \begin{cases} \lambda_j & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

Therefore, we now only need to divide each v_i by $\operatorname{sgn}(\lambda_i)\sqrt{|\lambda_i|}$ to normalize the basis. In turn, this proof shows that $\langle \cdot, \cdot \rangle_2$ can be represented by a diagonal matrix with respect to this new basis. □

As a corollary of the spectral theorem, it turns out that any linear map is obtained by stretching each axis by a nonnegative vector, up to pre and post composing with linear isometries.

Corollary 6.7 (Singular Value Decomposition). *Given a matrix $A \in \mathbb{R}^{m \times n}$, we can find an orthogonal matrix $P \in \mathbb{R}^{m \times m}$ and an orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ such that $A = PDQ$, where D has nonnegative entries and has the form $\begin{pmatrix} D' \\ 0 \end{pmatrix}$ or $\begin{pmatrix} D' & 0 \end{pmatrix}$, where D' is an $m \times m$ or $n \times n$ diagonal matrix, depending on whether $m \geq n$ or $n < m$.*

ADD...

Proof.

□

6.3 Fenchel-Legendre-Moreau Transform

In this section we expand convex optimization with some additional material that goes beyond the scope of differential geometry, but turns out to be pivotal in many fields of applied mathematics.

Definition 6.8 (Lower semicontinuity). A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous at x_0 if

$$\liminf_{x \to x_0} f(x) \geq f(x_0)$$

or, equivalently, if for every $\alpha \in \mathbb{R}$, the set

$$\{x \in \mathbb{R}^n : f(x) > \alpha\}$$

is open.

Example 6.9. Consider the following functions:

$$f(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0 \end{cases} \qquad g(x) = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

f is lower semicontinuous since continuity is broken only due to an upward shift; meanwhile, g is not lower semicontinuous since continuity is broken due to a downward shift, hence g is upper semicontinuous.

We state, without proof, the following useful characterization:

Proposition 6.10. *A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous if and only if there exists a family $\{\ell_i\}_{i\in I}$ of affine functions $\ell_i : \mathbb{R}^n \to \mathbb{R}$ such that*

$$f(x) = \sup_{i\in I}\{\ell_i(x)\} = \sup_{i\in I}\{\xi_i^T x + a_i\}.$$

On convex and lower semicontinuous functions it holds the well known Jensen's inequality:

Theorem 6.11 (Jensen). *Given a probability space $\{\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}^n$ with $\mathbb{E}[|X]] < +\infty$, we have*

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

for any convex and lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

Proof.
Exploiting Proposition 6.10 and the linearity of the expectation, we have

$$f(\mathbb{E}[X]) = \sup_{i\in I}\left[\xi_i^T \mathbb{E}[X] + a_i\right] = \sup_{i\in I}\left[\mathbb{E}[\xi_i^T X + a_i]\right] \leq \sup_{i\in I}\left[\mathbb{E}[f(X)]]\right] = \mathbb{E}[f(X)],$$

where the inequality holds since $\xi_i^T X + a_i \leq f(X)$. $\square$

Remark 6.12. Given a_i as defined in the previous proposition, we can "optimize" its value by selecting, for every i, the highest value such that the inequality $\xi^T x + a \leq f(x)$ still holds for all $x \in \mathbb{R}^n$. Such value is precisely

$$\alpha_\xi := \inf_{x\in\mathbb{R}^n}[f(x) - \xi^T x],$$

where we removed the dependence from i for ease of notation. Note how

$$\alpha_\xi = \inf_{x\in\mathbb{R}^n}[e^x - \xi^T x] = -\infty \quad \forall \xi < 0.$$

Therefore, $\alpha_\xi \in [-\infty, +\infty)$.

Proposition 6.13. *A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous if and only if*

$$f(x) = \sup_{\xi\in\mathbb{R}^n}[\xi^T x + \alpha_\xi] = \sup_{\xi\in\mathbb{R}^n}\left[\xi^T x + \inf_{x\in\mathbb{R}^n}[f(x) - \xi^T x]\right]$$

Proof.
If $f(x) = +\infty$, then $\alpha_\xi = +\infty$ and $\xi^T x + \alpha_\xi = +\infty$, so the claim holds. Since $f(x) \geq \xi^T x + \alpha_\xi$ for all $\xi \in \mathbb{R}^n$ and all $x \in \mathbb{R}^n$, we clearly have

$$f(x) \geq \sup_{\xi\in\mathbb{R}^n}[\xi^T x + \alpha_\xi] \quad \forall x \in \mathbb{R}^n.$$

On the other hand, α_ξ is defined to maximize $\xi_i^T x + \alpha_i$, so that

$$f(x) = \sup_{i\in I}\{\ell_i(x)\} \leq \sup_{i\in I}\{\xi_i^T x + \alpha_{\xi_i}\} \leq \sup_{\xi\in\mathbb{R}^n}\{\xi^T x + \alpha_\xi\}.$$

Combining these two inequalities we deduce the equality $f(x) = \sup_{\xi\in\mathbb{R}^n}\{\xi^T x + \alpha_\xi\}$. $\square$

Definition 6.14 (Legendre transform). Given $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, its Fenchel-Legendre-Moreau transform $f^\star : \mathbb{R}^n \to [-\infty, +\infty]$ is given by

$$f^\star(\xi) := -\alpha_\xi := \sup_{x\in\mathbb{R}^n}[\xi^T x - f(x)].$$

Remark 6.15. The following formulas hold:

$$\boxed{f^\star(\xi) = \sup_{x\in\mathbb{R}^n} [\xi^T x - f(x)]} \qquad \boxed{f(x) = \sup_{\xi\in\mathbb{R}^n} [\xi^T x - f^\star(\xi)]} \qquad \boxed{f(x) + f^\star(\xi) \geq \xi^T x}$$

In particular, the second formula implies that $\boxed{f = f^{\star\star}}$, switching the roles of x and ξ.

Remark 6.16. Given $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ proper, convex, lower semicontinuous, if both f and $f^\star$ are differentiable, then $(f')^{-1} = (f^\star)'$.

Remark 6.17. We can easily show that if $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, lower semicontinuous and proper, then $f^\star > -\infty$ and $f^\star$ is convex, lower semicontinuous and proper. As a consequence, calling $\mathcal{C}$ the class of proper, convex, lower semicontinuous functions $\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the operation

$$f \mapsto f^\star$$

maps the class $\mathcal{C}$ to itself.

6.4 Karush-Kuhn-Tucker conditions

In this section we will analyze the Karush-Kuhn-Tucker conditions.
We start by expressing the theorem for C^1 functions.

Theorem 6.18 (KKT conditions). *Given C^1 functions $f, g_1, \ldots, g_k, h_1, \ldots, h_\ell : U \to \mathbb{R}$, if x_0 is a constrained minimum point of f, among points $x \in U$ subject to*

$$g_1(x) \leq 0, \ldots, g_k(x) \leq 0, h_1(x) = \cdots = h_\ell(x) = 0,$$

assuming that $\nabla g_1(x_0), \ldots, \nabla g_k(x_0), \nabla h_1(x_0), \ldots, \nabla h_\ell(x_0)$ are linearly independent, there exist $\lambda_1, \ldots, \lambda_k \geq 0$ and $\mu_1, \ldots, \mu_\ell \in \mathbb{R}$ such that

$$\nabla f(x_0) + \sum_{i=1}^k \lambda_i \nabla g_i(x_0) + \sum_{j=1}^\ell \mu_j \nabla h_j(x_0) = 0,$$

as well as $\lambda_i g_i(x_0) = 0$ for all $i = 1, \ldots, k$.

ADD...

Proof.

□

Remark 6.19. Of course, for a constrained maximum point, we can simply replace f with $-f$, and in such case $\lambda_1, \ldots, \lambda_k \leq 0$.

Next, we state the theorem for convex functions, not necessarily differentiable.

Definition 6.20 (Subdifferential). Given a convex function $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$, the subdifferential of f at x_0 is the set

$$\partial f(x_0) := \{w \in \mathbb{R}^n : f(x) \geq f(x_0) + w^T(x - x_0) \quad \forall x \in C\}.$$

It follows from the definition that $0 \in \partial f(x_0)$ if and only if x_0 is a global minimum point.

Theorem 6.21 (KKT conditions). *Given a convex function $f : \mathbb{R}^n \to \mathbb{R}$, convex functions $g_1, \ldots, g_k : \mathbb{R}^n \to \mathbb{R}$ and affine functions $h_1, \ldots, h_\ell : \mathbb{R}^n \to \mathbb{R}$, define*

$$U = \{x \in \mathbb{R}^n : g_i(x) \leq 0 \quad \text{and} \quad h_j(x) = 0 \quad \forall i = 1, \ldots, k, \forall j = 1, \ldots, \ell\}.$$

Assume also the existence of a point $x^{int} \in U$ such that $g_i(x^{int}) < 0$ for all $i = 1, \ldots, k$ (Slater's condition). Then x_0 is a constrained minimizer for f, i.e.,

$$x_0 \in \arg\min_{x\in U} f(x),$$

if and only if $x_0 \in U$ and there exist $\lambda_1, \ldots, \lambda_k \geq 0$ and $\mu_1, \ldots, \mu_\ell \in \mathbb{R}$ such that

$$0 \in \partial f(x_0) + \sum_{i=1}^k \lambda_i \partial g_i(x_0) + \sum_{j=1}^\ell \mu_j \partial h_j(x_0),$$

as well as $\lambda_i g_i(x_0) = 0$ for all $i = 1, \ldots, k$.

Proof.

Sufficiency: suppose $x_0 \in U$ and there exist $\lambda_1, \dots, \lambda_k \geq 0$ and $\mu_1, \dots, \mu_\ell \in \mathbb{R}$ satisfying stationarity and complementary slackness. Define

$$\tilde{f}(x) := f(x) + \sum_{i=1}^{k} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x).$$

Since we have $h_j(x_0) = 0$ for all j, as well as complementary slackness, then

$$\tilde{f}(x_0) = f(x_0).$$

Moreover, we know that $0 \in \partial \tilde{f}(x_0)$: this means that x_0 is a global minimum point for $\tilde{f}$, giving

$$f(x_0) = \tilde{f}(x_0) \leq \tilde{f}(x) = f(x) + \sum_{i=1}^{k} \lambda_i g_i(x) \leq f(x) \quad \forall x \in \mathbb{R}^n,$$

since $h_j(x) = 0$ and $\lambda_i g_i(x) \leq 0$.[a]

Necessity: suppose $x_0 \in U$ is a constrained minimizer for f. Define the Lagrangian $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^k_+ \times \mathbb{R}^\ell \to \mathbb{R}$:

$$\mathscr{L}(x, \lambda, \mu) := f(x) + \sum_{i=1}^{k} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x).$$

Step 1: we will first prove that

$$f(x_0) = \inf_{x \in \mathbb{R}^n} \sup_{\lambda \geq 0, \mu \in \mathbb{R}^\ell} \mathscr{L}(x, \lambda, \mu) \tag{6.1}$$

Indeed,

$$\sup_{\lambda \geq 0, \mu \in \mathbb{R}^\ell} \mathscr{L}(x, \lambda, \mu) = f(x) + \sup_{\lambda \geq 0} \left\{ \sum_{i=1}^{k} \lambda_i g_i(x) \right\} + \sup_{\mu \in \mathbb{R}^\ell} \sum_{j=1}^{\ell} \mu_j h_j(x).$$

The first term is a constant, the second is 0 whenever $g_i(x) \leq 0$ for every i but $+\infty$ once any $g_i(x) > 0$, and the third term is zero whenever $h_j(x) = 0$, but $+\infty$ once any $h_j(x) \neq 0$. Therefore,

$$\sup_{\lambda \geq 0, \mu \in \mathbb{R}^\ell} \mathscr{L}(x, \lambda, \mu) = \begin{cases} f(x) & \text{if } g_i(x) \leq 0 \text{ for every } i \text{ and } h_j(x) = 0 \text{ for every } j, \\ +\infty & \text{otherwise} \end{cases}.$$

Since x_0 minimizes f on the constraint, we deduce (6.1).

Step 2: assume **strong convex duality**, or **strong lagrangian duality**:

$$\inf_{x \in \mathbb{R}^n} \sup_{\lambda \geq 0, \mu \in \mathbb{R}^\ell} \mathscr{L}(x, \lambda, \mu) = \sup_{\lambda \geq 0, \mu \in \mathbb{R}^\ell} \inf_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda, \mu) \tag{6.2}$$

Moreover, assume that LHS is achieved by x_0 (which we already know) and RHS is achieved by $(\lambda^\star, \mu^\star) \in \mathbb{R}^k_+ \times \mathbb{R}^\ell$, meaning

$$f(x_0) = \inf_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \geq 0 \\ \mu \in \mathbb{R}^\ell}} \mathscr{L}(x, \lambda, \mu) = \sup_{\substack{\lambda \geq 0 \\ \mu \in \mathbb{R}^\ell}} \mathscr{L}(x_0, \lambda, \mu) \quad \text{and} \quad \sup_{\substack{\lambda \geq 0 \\ \mu \in \mathbb{R}^\ell}} \inf_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda, \mu) = \inf_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda^\star, \mu^\star).$$

Therefore,

$$f(x_0) = \inf_{x \in \mathbb{R}^n} \sup_{\substack{\lambda \geq 0 \\ \mu \in \mathbb{R}^\ell}} \mathscr{L}(x_0, \lambda, \mu) = \sup_{\substack{\lambda \geq 0 \\ \mu \in \mathbb{R}^\ell}} \mathscr{L}(x_0, \lambda, \mu) = \mathscr{L}(x_0, \lambda^\star, \mu^\star) = \inf_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda^\star, \mu^\star) \leq \mathscr{L}(x, \lambda^\star, \mu^\star) \quad \forall x \in \mathbb{R}^n$$

This implies that x_0 is a global minimum point for

$$\tilde{f}(x) := f(x) + \sum_{i=1}^{k} \lambda_i^\star g_i(x) + \sum_{j=1}^{\ell} \mu_j^\star h_j(x),$$

and hence $0 \in \partial \tilde{f}(x_0)$. As we will see, proving strong convex duality will yield optimal $(\lambda^\star, \mu^\star)$ as a byproduct, concluding the proof.

Step 3: Let us prove **strong convex duality**, or **strong lagrangian duality**

$$\inf_{x\in\mathbb{R}^n} \sup_{\lambda\geq 0, \mu\in\mathbb{R}^\ell} \mathscr{L}(x,\lambda,\mu) = \sup_{\lambda\geq 0, \mu\in\mathbb{R}^\ell} \inf_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda,\mu) \tag{6.3}$$

and show that the supremum in RHS is achieved by $(\lambda^\star, \mu^\star) \in \mathbb{R}^k_+ \times \mathbb{R}^\ell$.

For that purpose, we introduce some perturbation in the optimization problem: for any $r \in \mathbb{R}^k$ and $s \in \mathbb{R}^\ell$, define

$$U(r,s) = \{x \in \mathbb{R}^n : g_i(x) \leq r_i \quad \text{and} \quad h_j(x) = s_j \quad \forall i = 1,\dots,k, \forall j = 1,\dots,\ell\}.$$

We define the value function $v : \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R} \cup \{\pm\infty\}$ as

$$v(r,s) := \inf_{x\in\mathbb{R}^n} \{f(x) | x \in U(r,s)\} = \inf_{x\in U(r,s)} f(x).$$

Thanks to the convexity of f, g and h, it can be shown that v is convex.
Its Fenchel conjugate is

$$\begin{aligned}
v^\star(\zeta,\xi) &= \sup_{r,s}[\zeta^T r + \xi^T s - v(r,s)] = \sup_{r,s}[\zeta^T r + \xi^T s + \sup_{x\in U(r,s)} -f(x)] \\
&= \sup_{x\in\mathbb{R}^n} \sup_{\substack{r_i \geq g_i(x) \\ s_j = h_j(x)}} \Big[-f(x) + \zeta^T r + \xi^T s \Big] = \sup_{x\in\mathbb{R}^n} \Big[-f(x) + \xi^T h(x) + \sup_{r_i\geq g_i(x)} \zeta^T r \Big] \\
&= \begin{cases} \sup_{x\in\mathbb{R}^n} \Big[-f(x) + \sum_{j=1}^\ell \xi_j h_j(x) + \sum_{i=1}^k \zeta_i g_i(x) \Big] & \text{if } \zeta \leq 0 \\ +\infty & \text{otherwise} \end{cases} \\
&\overset{\substack{\lambda=-\zeta \\ \mu=-\xi}}{=} \begin{cases} \sup_{x\in\mathbb{R}^n} -\Big[f(x) + \sum_{i=1}^k \lambda_i g_i(x) + \sum_{j=1}^\ell \mu_j h_j(x) \Big] & \text{if } \lambda \geq 0 \\ +\infty & \text{otherwise} \end{cases} \\
&= \begin{cases} \sup_{x\in\mathbb{R}^n} -\mathscr{L}(x,\lambda,\mu) & \text{if } \lambda \geq 0 \\ +\infty & \text{otherwise} \end{cases} = \begin{cases} -\inf_x \mathscr{L}(x,\lambda,\mu)\,, & \lambda \geq 0, \\ +\infty\,, & \text{otherwise.} \end{cases}
\end{aligned}$$

Let $\bar{x} \in \mathbb{R}^n$ satisfy the Slater condition $g_i(\bar{x}) < 0$ $(i = 1,\dots,k)$ and $h_j(\bar{x}) = 0$ $(j = 1,\dots,\ell)$. Choose $\varepsilon > 0$ such that $g_i(\bar{x}) + \varepsilon < 0$ for every i. Then for every (r,s) with $\|r\|_\infty < \varepsilon$ and $\|s\|_\infty < \varepsilon$ we still have $\bar{x} \in U(r,s)$, hence $v(r,s) < +\infty$. Therefore

$$(0,0) \in \operatorname{int}(\operatorname{dom} v).$$

Because v is proper, convex and finite on a neighbourhood of $(0,0)$, it is continuous (hence lower-semicontinuous) at that point. By the Fenchel–Moreau theorem for closed convex functions we obtain the biconjugate identity

$$v(0,0) = v^{**}(0,0).$$

Evaluating at $(0,0)$ yields

$$\inf_{x\in\mathbb{R}^n} \sup_{\lambda\geq 0, \mu\in\mathbb{R}^\ell} \mathscr{L}(x,\lambda,\mu) = f(x_0) = v(0,0) = v^{\star\star}(0,0) = \sup_{(\lambda,\mu)} [-v^\star(\lambda,\mu)] = \sup_{\substack{\lambda\geq 0 \\ \mu\in\mathbb{R}^\ell}} [\inf_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda,\mu)],$$

concluding the proof. $\square$

[a]Note that Slater's condition is not needed for sufficiency. Moreover, note that sufficiency does not hold in the non-convex setting.