Geometry, integrability and field theories

Sylvain Lavau*[1]

[1] *Euler International Mathematical Institute (EIMI) & Steklov Institute, Saint-Petersburg, Russia.*

The idea of the course is to provide notions of topology and geometry to mathematical physicists, as well as present concrete applications of such notions to pure algebraists and geometers. The idea would be not to fall in a too abstract presentation, and to anchor it into examples taken from physics. I propose to start from the basics and grow in complexity to reach higher grounds which are much more intricate.

I propose to advance at a steady pace following a physically informed mathematical path. I would not focus on the logico-deductive process of mathematical proofs but rather on the physical ideas that led to the invention of these notions. Mathematical physicists often find their inspiration in problems and objects set up by theoretical physicists, on which they draw to develop interesting and useful mathematical objects. The latters may be somewhat 'generalizations' of the formers, but they need not encode exactly the physics that inspired them. Theoretical physics is a playground for mathematical physicists who use nice and insightful results to develop fruitful mathematical theories. This course will follow the same line of reasoning: drawing on physical examples to present useful mathematical objects.

A particular attention will be paid to the translation between physics language and mathematical language. In particular, physicists often work in coordinates (this emphasize the local property of their theories) and which facilitates raw computations, whereas mathematicians are more interested in maps, and relationships between objects, spaces, etc. That is why mathematicians favor a coordinate-free approach of geometry. The latter has the advantage of shedding light on geometrical aspects that the picture in coordinates could not provide. During the course, I will try to give an explicit dictionary that shows how one pass from one formulation to the other.

To build this course, I mostly relied on various lectures notes (mostly in the chapter about Poisson Geometry) and on the following sources:

[Baez and Muniain, 1994] Baez, J. and Muniain, J. P. (1994). *Gauge Fields, Knots and Gravity*, volume 4 of *Series on Knots and Everything*. World Scientific, Singapore

[Henneaux and Teitelboim, 1994] Henneaux, M. and Teitelboim, C. (1994). *Quantization of Gauge Systems*. Princeton University Press, Princeton

[Laurent-Gengoux et al., 2013] Laurent-Gengoux, C., Pichereau, A., and Vanhaecke, P. (2013). *Poisson Structures*. Grundlehren Der Mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg

[Lee, 2003] Lee, J. M. (2003). *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer-Verlag, New York

[Lee, 2009] Lee, J. M. (2009). *Manifolds and Differential Geometry*. Number 107 in Graduate Studies in Mathematics. American Mathematical Society, Providence

*lavau@math.univ-lyon1.fr

[Rothe and Rothe, 2010] Rothe, H. J. and Rothe, K. D. (2010). *Classical and Quantum Dynamics of Constrained Hamiltonian Systems*, volume 81 of *World Scientific Lecture Notes in Physics*. World Scientific

Notice that reference [Lee, 2003] is the first (2003) edition. The second (2012) edition has been widely revised and the chapters have been shuffled so that the references to the first edition do not correspond to the same in the second edition.

Notations: we will use Einstein summation convention on sums over space-time coordinates: when an index appears twice (only) in a term, and is such that it appears once as an exponent, and once as a bottom index, then one may get rid of the sum sign, and understand that the sole presence of the repeated indices symbolizes the summation. For exemple, when we write $g_{ij}e^j$ (where g_{ij} symbolizes a metric and e^j is a covector), it mathematically means $\sum_{1\leq j\leq n} g_{ij}e^j$. We would not use this convention for summation other than space-time coordinates which are such that there is one index up and one index down. The kronecker delta will always have one index up and one index down, as in δ^i_j. We also widely use the *rationalized Planck units*, where:

$$c = 4\pi G = \hbar = \varepsilon_0 = k_{\mathrm{B}} = 1$$

Acknowledgments: I would like to thank Nikita Belousov, Casey Blacker, Leonid Ryvkin and Gerard Lavau for their insightful comments and ideas. Their help has been invaluable to improve the quality of the notes and detect the unavoidable mistakes (there are certainly some surviving in the text).

Contents

1 Mathematical background in linear algebra

The idea of the first chapter is to work over finite dimensional vector spaces and define several objects that will be later generalized to manifolds. Let E be a (real) vector space of dimension n. Then it is isomorphic to $\mathbb{R}^n$. A basis of E is a set of n vectors – say $e_1, \ldots, e_n$ – that are linearly independent and that generate the whole vector space.

1.1 The tensor algebra, the symmetric algebra and the exterior algebra

The *tensor algebra* of E – denoted $T(E)$ – is an infinite family of vector spaces $T^0(E), T^1(E), T^2(E), \ldots$ defined recursively as:

$$T^0(E) = \mathbb{R} \quad \text{and, for all } m \geq 0 \quad T^{m+1}(E) = E \otimes T^m E$$

with the convention that $\mathbb{R} \otimes E = E \otimes \mathbb{R} = E$. The symbol $\otimes$ symbolizes a sort of multiplication, not between scalars but between vectors – or more generally tensors, hence the name. More precisely, this tensor product possesses the associativity and distributivity properties of the multiplication operator:

$$x_1 \otimes (x_2 \otimes x_3) = (x_1 \otimes x_2) \otimes x_3 = x_1 \otimes x_2 \otimes x_3$$
$$x_1 \otimes \ldots \otimes (x_i + y) \otimes \ldots \otimes x_m = (x_1 \otimes \ldots \otimes x_i \otimes \ldots \otimes x_m) + (x_1 \otimes \ldots \otimes y \otimes \ldots \otimes x_m)$$

for every $1 \leq i \leq m$ and every vectors $x_1, \ldots, x_m, y \in E$. Notice however that the tensor product $\otimes$ is *not* commutative, contrary to the usual multiplication on scalars. Since we are working on vector spaces, we assume that it is linear in every variable, that is, given any scalar $\lambda \in \mathbb{R}$:

$$\lambda(x_1 \otimes \ldots \otimes x_m) = (\lambda x_1) \otimes \ldots \otimes x_m = x_1 \otimes \ldots \otimes (\lambda x_i) \otimes \ldots \otimes x_m = x_1 \otimes \ldots \otimes (\lambda x_m)$$

Thus, elements of the m-th tensor power of E – denoted $T^m(E)$ or sometimes $E^{\otimes m}$ – are literally products of vectors of E. This has to be contrasted (and not to be confused) with the cartesian product $E \times \ldots \times E$ where multiplication by a scalar satisfies:

$$\lambda(x_1, \ldots, x_m) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_m)$$

and where distributivity over addition is not satisfied:

$$\forall\ 1 \leq i \leq m \qquad (x_1, \ldots, x_i + y, \ldots, x_m) = (x_1, \ldots, x_i, \ldots, x_m) + (0, \ldots, y, \ldots, 0)$$

This comes from the fact that the cartesian product $E \times \ldots \times E$ actually corresponds to the direct sum $E^{\oplus m} = E \oplus \ldots \oplus E$ (m-times). This discussion shows that $T^m(E)$ is of dimension n^m, whereas $E^{\oplus m}$ is of dimension $n \times m$. A basis of $T^m(E)$ is explicitely given by the following tensor products:

$$\left\{ e_{i_1} \otimes \ldots \otimes e_{i_m} \,\middle|\, 1 \leq i_1, \ldots, i_m \leq n \right\} \tag{1.1}$$

Additionally, associativity of the tensor product implies that:

$$T^k(E) \otimes T^l(E) \subset T^{k+l}(E) \tag{1.2}$$

An algebra that is a graded vector space and whose product satisfies a similar condition as Equation (1.2) is called a *graded algebra*:

Definition 1.1. *A* graded vector space *is a family of vector spaces $E = (E_i)_{i\in\mathbb{Z}}$, indexed over $\mathbb{Z}$ (not all E_i need be non-zero). The indices are integers and called* degrees*, and are denoted $|x| = i$ for any homogeneous element $x \in E_i$. We say that E is* non-negatively graded *(resp.* non-positively graded*) if $E = (E_i)_{i\geq 0}$ (resp. $E = (E_i)_{i\leq 0}$).*

A graded algebra *is a graded vector space $A = \bigoplus_{i\in\mathbb{Z}} A_i$ equipped with a $\mathbb{R}$-bilinear operation $\cdot : A \times A \longrightarrow A$ which satisfies:*

$$A_i \cdot A_j \subset A_{i+j}$$

Example 1.2. A vector space E is a graded vector space where every $E_i = 0$ for $i \neq 0$ but $E_0 = E$.

Example 1.3. The tensor algebra is a graded algebra, in which the grading corresponds to the length of the basis elements. This graded algebra is non-negatively graded.

The tensor algebra $T(E)$ contains two particular subspaces[1]: the one formed by linear combinations of fully symmetrized basis elements of $T(E)$ – it is the *symmetric algebra* $S(E)$, and the one formed by linear combinations of fully anti-symmetrized basis elements of $T(E)$ – it is the *exterior algebra* $\bigwedge^\bullet(E)$. Both will be graded algebra, with respect to their product.

Remark 1.4. When we write a bullet $\bullet$ as an index or an exponent we want to emphasize that the space is graded, e.g. $\bigwedge^\bullet(E) = \bigwedge^0(E) \oplus \bigwedge^1(E) \oplus \ldots \oplus \bigwedge^n(E)$.

Both the symmetric algebra and the exterior algebra are actually graded sub-vector spaces of $T(E)$, that is to say: they both decompose as a family of vector spaces $S(E) = \bigoplus_{m=0}^\infty S^m(E)$ and $\bigwedge^\bullet(E) = \bigoplus_{m=0}^n \bigwedge^m(E)$, which are such that $S^m(E), \bigwedge^m(E) \subset T^m(E)$, for every $m \geq 0$. The graded space $S(E)$ is the subspace of $T(E)$ that is invariant under the action of any permutation σ on the labels of the basis vectors. More precisely, for every $m \geq 1$, the space $S^m(E)$ is generated (as a vector subspace of $T^m(E)$) by the following elements:

$$e_{i_1} \odot e_{i_2} \odot \ldots \odot e_{i_m} = \frac{1}{m!} \sum_{\sigma\in S_m} e_{i_{\sigma(1)}} \otimes e_{i_{\sigma(2)}} \otimes \ldots \otimes e_{i_{\sigma(m)}} \tag{1.3}$$

The symmetrized product $\odot$ symbolizes that the tensor $e_{i_1} \odot e_{i_2} \odot \ldots \odot e_{i_m}$ is invariant under the action of any permutation of m elements $\sigma \in S_m$. In particular, invariance under the permutation $(1\ 2)$ reads:

$$e_{i_1} \odot e_{i_2} = e_{i_2} \odot e_{i_1}$$

Hence the symmetric product is *commutative*. Any other combination of permutations leaves the product unchanged. The graded space $S(E)$ equipped with the product $\odot$ is a (commutative) graded algebra because it satisfies a similar condition as Equation (1.2):

$$S^k(E) \odot S^l(E) \subset S^{k+l}(E)$$

Counting the number of ways one can choose m elements (with possible repetitions) among n basis vectors in order to construct the basis elements defined in Equation (1.3), one can check that one obtains all the basis elements by restricting oneself to $1 \leq i_1 \leq i_2 \leq \ldots \leq i_m \leq n$:

$$\Big\{ e_{i_1} \odot \ldots \odot e_{i_m} \ \Big|\ 1 \leq i_1 \leq i_2 \leq \ldots \leq i_m \leq n \Big\} \tag{1.4}$$

Then, the dimension of the space $S^m(E)$ is $\binom{n+m-1}{m}$, thus one can see that it increases with m. The symmetric algebra is thus infinite dimensional, as is the tensor algebra.

[1]Actually the symmetric algebra and the exterior algebra are quotient of the tensor algebra, but there exists a canonical isomorphisms between those and the subspaces of E that we describe.

The exterior algebra, on the other hand, is generated (as a vector space) by elements of $T(E)$ invariant under *signed* permutations. Let us explain what it means. For every $m \geq 1$, the space $\bigwedge^m(E)$, whose elements are called *m-vectors* or *multivectors*, is generated (as a vector subspace of $T^m(E)$) by the following elements:

$$e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_m} = \sum_{\sigma \in S_m} (-1)^\sigma e_{i_{\sigma(1)}} \otimes e_{i_{\sigma(2)}} \otimes \ldots \otimes e_{i_{\sigma(m)}} \tag{1.5}$$

where $(-1)^\sigma$ is the signature of the permutation σ. Using the *Levi-Civita symbol* $\epsilon_{\sigma(1)\ldots\sigma(m)} = (-1)^\sigma \epsilon_{1\ldots m}$, set with the convention that $\epsilon_{1\ldots m} = 1$, one obtains the alternative, more physicists oriented, formula:

$$e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_m} = \sum_{\sigma \in S_m} \epsilon_{\sigma(1)\ldots\sigma(m)}\, e_{i_{\sigma(1)}} \otimes e_{i_{\sigma(2)}} \otimes \ldots \otimes e_{i_{\sigma(m)}}$$

In particular the first few elements are:

$$\begin{aligned}
&\text{for } m = 0 && \bigwedge^0(E) \simeq \mathbb{R} \\
&\text{for } m = 1 && \bigwedge^1(E) \simeq E \\
&\text{for } m = 2 && e_i \wedge e_j = e_i \otimes e_j - e_j \otimes e_i \\
&\text{for } m = 3 && e_i \wedge e_j \wedge e_k = e_i \otimes e_j \otimes e_k + e_j \otimes e_k \otimes e_i + e_k \otimes e_i \otimes e_j \\
& && \qquad - e_i \otimes e_k \otimes e_j - e_k \otimes e_j \otimes e_k - e_j \otimes e_i \otimes e_k
\end{aligned}$$

There exists another convention, which is such that $x \wedge y = \frac{1}{2}(x \otimes y - y \otimes x)$ but this is not convenient for geometrical purposes, but which is the natural product when the exterior algebra is obtained through a quotient of the tensor algebra. These subtleties are discussed at large in Chapter 12 of [Lee, 2003] (Chapter 14 in the 2012 edition).

The *wedge product* $\wedge$ is defined so that the tensor $e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_m}$ is invariant under any *signed* permutation $(-1)^\sigma \sigma$ of m elements. For any permutation $\sigma \in S_m$, the general formula is the following:

$$e_{i_1} \wedge \ldots \wedge e_{i_m} = (-1)^\sigma e_{i_{\sigma(1)}} \wedge \ldots \wedge e_{i_{\sigma(m)}}$$

or, using the Levi-Civita symbol:

$$e_{i_1} \wedge \ldots \wedge e_{i_m} = \epsilon_{\sigma(1)\ldots\sigma(m)} e_{i_{\sigma(1)}} \wedge \ldots \wedge e_{i_{\sigma(m)}}$$

where here, the Einstein summation convention is *not* used! To illustrate these rather abstract formulas, let us pick up the transposition $(1\ 2)$ (of signature -1). Then, invariance of the bivector $e_{i_1} \wedge e_{i_2}$ under the action of the signed permutation $-(1\ 2)$ reads:

$$e_{i_1} \wedge e_{i_2} = -e_{i_2} \wedge e_{i_1} \tag{1.6}$$

The minus sign on the right hand side is the signature of the transposition $(1\ 2)$. Another example σ is the circular permutation $(1\ 2\ 3)$ (of signature $+1$), which is such that e_{i_1} becomes e_{i_2}, e_{i_2} becomes e_{i_3} and e_{i_3} becomes e_{i_1}. This (signed) permutation leaving the trivector $e_{i_1} \wedge e_{i_2} \wedge e_{i_3}$ invariant means that:

$$e_{i_1} \wedge e_{i_2} \wedge e_{i_3} = e_{i_2} \wedge e_{i_3} \wedge e_{i_1} \tag{1.7}$$

More generally, the rule of calculus in the exterior algebra is that, when permuting two elements, a sign appears only when the signature of the chosen transposition is -1. In particular, since it is often difficult to known the signature of a permutation, and since any permutation can be obtained from a sequence of transpositions (permutation of two elements), permuting elements two by two while multiplying by -1 until reaching the image of the desired (signed) permutation is a good technique to obtain the correct sign. Let us illustrate by looking up at the permutation $(1\ 3\ 2) = (1\ 3)(3\ 2)$ (its parity is even so its signature is $+1$). By first using the transposition $(3\ 2)$, and then $(1\ 3)$ one obtains:

$$e_{i_1} \wedge e_{i_2} \wedge e_{i_3} = -e_{i_1} \wedge e_{i_3} \wedge e_{i_2} = e_{i_3} \wedge e_{i_1} \wedge e_{i_2}$$

We can now study the (signed) action of the cycle $(1\ 2\ 3\ 4) = (1\ 2)(2\ 3)(3\ 4)$ (of signature -1) on $e_{i_1} \wedge e_{i_2} \wedge e_{i_3} \wedge e_{i_4}$, which can be obtained through three transpositions:

$$e_{i_1} \wedge e_{i_2} \wedge e_{i_3} \wedge e_{i_4} = -e_{i_1} \wedge e_{i_2} \wedge e_{i_4} \wedge e_{i_3} = e_{i_1} \wedge e_{i_3} \wedge e_{i_4} \wedge e_{i_2} = -e_{i_2} \wedge e_{i_3} \wedge e_{i_4} \wedge e_{i_1} \tag{1.8}$$

Since the permutation $(1\ 2\ 3\ 4)$ is such that e_{i_1} becomes e_{i_2}, e_{i_2} becomes e_{i_3}, e_{i_3} becomes e_{i_4} and e_{i_4} becomes e_{i_1}, one observe that the sign in the right hand side of Equation (1.8) tells us that the parity of $(1\ 2\ 3\ 4)$ is odd. Additionally, we see that the action of the signed permutation $-(1\ 2\ 3\ 4)$ leaves $e_{i_1} \wedge e_{i_2} \wedge e_{i_3} \wedge e_{i_4}$ invariant.

Another efficient way of managing cyclic permutations – instead of decomposing them – is to take the leftmost element, and make it go right through all the terms, so that at each transposition with its neighbor, one adds a minus sign. At each step, we use Equation (1.6) so that we ensure that all expressions are equal. For example the signed action of $(-1)^{k-1}(1\ 2\ \ldots k-1\ k)$ leaves the multivector $e_{i_1} \wedge \ldots \wedge e_{i_m}$ invariant, and that can be shown by making e_{i_1} goes right through the $k-1$ vectors on its right:

$$\begin{aligned} e_{i_1} \wedge e_{i_2} \wedge \ldots \wedge e_{i_k} \wedge \ldots \wedge e_{i_m} &= -e_{i_2} \wedge e_{i_1} \wedge e_{i_3} \wedge \ldots \wedge e_{i_k} \wedge \ldots \wedge e_{i_m} \\ &= e_{i_2} \wedge e_{i_3} \wedge e_{i_1} \wedge \ldots \wedge e_{i_k} \wedge \ldots \wedge e_{i_m} \\ &= (-1)^{k-2} e_{i_2} \wedge \ldots \wedge e_{i_{k-1}} \wedge e_{i_1} \wedge e_{i_k} \wedge \ldots \wedge e_{i_m} \\ &= (-1)^{k-1} e_{i_2} \wedge \ldots \wedge e_{i_k} \wedge e_{i_1} \wedge e_{i_{k+1}} \wedge \ldots \wedge e_{i_m} \end{aligned}$$

The general rule is that for cyclic permutations of the form $(1\ 2\ \ldots k-1\ k)$ the parity is the same as the parity of the integer $k-1$. It is as if the vector e_{i_1} had jumped over $k-1$ elements to get in the right place. This strategy could have been used in Equations (1.6), (1.7) and (1.8), where we obtain that the parity of a transposition is odd, the parity of a circular permutation of three elements is even, whereas the parity of a circular permutation of four elements is odd.

The properties of the wedge product implies in particular that for every $x \in E$, the bivector $x \wedge x$ is zero (in the vector space $\bigwedge^2(E)$). Thus, as soon as the same element of E appears twice in a multivector, then it is automatically zero. For example, let $x_1, \ldots, x_m$ be m linearly independent vectors of E (so in particular $1 \le m \le n$), then:

$$x_1 \wedge x_2 \wedge \ldots \wedge x_m \neq 0$$

In the case that one of the x_i is a linear combination of the others, say $x_i = \sum_{j \neq i} \alpha_j x_j$, then the m-vector is zero, since:

$$x_1 \wedge \ldots \wedge x_m = \sum_{j \neq i} \alpha_j \underbrace{x_1 \wedge \ldots \wedge x_{i-1} \wedge x_j \wedge x_{i+1} \wedge \ldots \wedge x_m}_{=\,0}$$

This has a tremendous consequence: contrary to the symmetric algebra, the exterior algebra is bounded above. A multivector cannot be composed by more than m vectors, for otherwise it vanishes. Hence, contrary to the symmetric algebra, the exterior algebra is of finite dimension.

Due to the fact that the wedge product of two identical elements vanish, one can check that all the basis elements of $\bigwedge^m(E)$ are obtained by restricting oneself to $1 \leq i_1 < i_2 < \ldots < i_m \leq n$, that is to say a basis is formed by the following multivectors:

$$\left\{ e_{i_1} \wedge \ldots \wedge e_{i_m} \,\middle|\, 1 \leq i_1 < i_2 < \ldots < i_m \leq n \right\} \tag{1.9}$$

Then one deduces that the dimension of the vector space $\bigwedge^m(E)$ is $\binom{n}{m}$. One can check that such a dimension is minimal and equal to 1 for $m = 0$ (i.e. when $\bigwedge^0(E) = \mathbb{R}$) and for $m = n$ (i.e. when $\bigwedge^n(E)$ is the one-dimensional vector subspace of $T(E)$ generated by the element $e_1 \wedge e_2 \wedge \ldots \wedge e_n$). The direct sum $\bigwedge^\bullet(E) = \bigoplus_{m=0}^n \bigwedge^m(E)$ is then finite dimensional of total dimension 2^n.

Additionally, the definition of the wedge product has been made so that we have the following property:

$$(e_{i_1} \wedge \ldots \wedge e_{i_p}) \wedge (e_{i_{p+1}} \wedge \ldots \wedge e_{i_m}) = e_{i_1} \wedge \ldots \wedge e_{i_m} \tag{1.10}$$

In particular, the product is associative. This allows us to compute the wedge product of a k-multivectors and l-multi vectors. Notice that the wedge product satisfies Equation (1.10) precisely because of the absence of any scaling factor on the right hand side of Equation (1.5). The wedge product then defines a graded algebra structure on the exterior algebra (hence justifying the name), that is:

$$\bigwedge^k(E) \wedge \bigwedge^l(E) \subset \bigwedge^{k+l}(E)$$

More precisely, for any $\alpha \in \bigwedge^k(E)$ and any $\beta \in \bigwedge^l(E)$, then one has $\alpha \wedge \beta \in \bigwedge^{k+l}(E)$, and it satisfies the following identity:

$$\alpha \wedge \beta = (-1)^{kl} \beta \wedge \alpha$$

We say that the product is *graded commutative*, and the exterior algebra is thus a *(graded) commutative graded algebra*.

Exercise 1.5. The proof is left as an exercise.

Recall that the *dual* of the vector space E is the space denoted E^* of all *linear forms on* E, i.e. all the linear maps $\varphi : E \longrightarrow \mathbb{R}$. While elements of E are called *vectors*, elements of E^* are called *covectors*. Given a basis $e_1, \ldots, e_n$ of E there is a privileged choice of a basis on E^*: the set of linear maps $e^1, \ldots, e^n : E \longrightarrow \mathbb{R}$, that are such that:

$$e^i(e_j) = \delta^i_j \tag{1.11}$$

where here δ^i_j denotes the Kronoecker symbol[2]. Such a choice of basis on E^* can always be made. Notice the localization of the labels i, j: as indices on vectors, as exponents on covectors. This has some importance, and is related to Einstein summation convention: for example, imagine you have a vector $v = v^i e_i$ and a covector $\varphi = \varphi_j e^j$. In particular $\varphi \in E^*$ and can be understood as a linear form $\varphi : E \longrightarrow \mathbb{R}$ which can act on v and define a real number (we assume Einstein summation convention throughout):

$$\varphi(v) = \varphi_j e^j(v^i e_i) = \varphi_j v^i\, e^j(e_i) = \varphi_j v^i \delta^j_i = \varphi_i v^i$$

We passed from the second term to the third by linearity of the dual basis. In the last implicit sum on the right-hand side, we say that the upper and lower indices have been *contracted*. The

[2]Some author use the notation g^i_j instead. Moreover, some authors consider that the real Kronecker symbol is the one with one index up and one index down. In that case, when they write δ_{ij} they mean g_{ij}. We will try to use this convention in the present paper.

result should be an object which does not carry any index, which is precisely the case of the real number $\varphi(v)$.

One can define the tensor algebra of the dual E^* and since we are in finite dimension, $T(E^*) \simeq \big(T(E)\big)^*$. The dual basis of this dual vector space can be obtained from the dual basis $e^1, \dots, e^n$ and the definition of the tensor product. The action of the dual element $e^i \otimes e^j$ on $e_k \otimes e_l$ is given by the following:

$$e^i \otimes e^j (e_k \otimes e_l) = \delta^i_k \delta^j_l$$

This is equal to $+1$ if and only if $k = i$ and $l = j$. From this we deduce that the dual basis to the basis of $T(E)$ (see Equation (1.1)) is made of the following tensor products:

$$\left\{ e^{i_1} \otimes \dots \otimes e^{i_m} \;\middle|\; 1 \leq i_1, \dots, i_m \leq n \right\} \tag{1.12}$$

That is to say:

$$e^{i_1} \otimes \dots \otimes e^{i_m} (e_{j_1} \otimes \dots \otimes e_{j_m}) = \delta^{i_1}_{j_1} \dots \delta^{i_m}_{j_m} \tag{1.13}$$

One can also define a symmetric algebra and an exterior algebra associated to the dual E^*, and we have the following isomorphisms because E is finite dimensional: $S(E^*) \simeq \big(S(E)\big)^*$ and $\bigwedge^\bullet(E^*) \simeq \big(\bigwedge^\bullet(E)\big)^*$. Notice however that the most obvious basis of $S(E^*)$ and $\bigwedge^\bullet(E^*)$ are not the dual basis of (1.4) and (1.9). Indeed, using the definition of the symmetric product (see Equation (1.3)) on the dual basis $e^1, \dots, e^n$ of E^*, one obtains a basis of $S(E^*)$, denoted by vectors of the form $e^{i_1} \odot \dots \odot e^{i_m}$ for $1 \leq i_1 \leq \dots \leq i_m \leq n$. However this basis is not dual to the basis of $S(E)$ given in Equation (1.4), for:

$$e^i \odot e^j (e_k \odot e_l) = \frac{e^i \otimes e^j + e^j \otimes e^i}{2} \frac{e_k \otimes e_l + e_l \otimes e_k}{2} = \frac{(\delta^i_k \delta^j_l + \delta^i_l \delta^j_k)}{2}$$

which is equal to $\frac{1}{2}$ when $k = i$ and $l = j$. Thus the element of $S^2(E^*)$ that would be considered to be dual to $e_i \odot e_j$ is $2\, e^i \odot e^j$. More generally a dual basis to the basis (1.4) is given by:

$$\left\{ m!(e^{i_1} \odot \dots \odot e^{i_m}) \;\middle|\; 1 \leq i_1 \leq i_2 \leq \dots \leq i_m \leq n \right\}$$

This also forms a basis of $S(E^*)$, but one has to remember the factor. The same phenomenon occurs for $\bigwedge^\bullet(E^*)$. Using the definition of the symmetric product (see Equation (1.5)) on the dual basis $e^1, \dots, e^n$ of E^*, one obtains a basis of $\bigwedge^\bullet(E^*)$, denoted by vectors of the form $e^{i_1} \wedge \dots \wedge e^{i_m}$ for $1 < i_1 < \dots < i_m < n$. However this basis is not dual to the basis of $\bigwedge^\bullet(E)$ given in Equation (1.9), for:

$$e^i \wedge e^j (e_k \wedge e_l) = (e^i \otimes e^j - e^j \otimes e^i)(e_k \otimes e_l - e_l \otimes e_k) = 2\,(\delta^i_k \delta^j_l - \delta^i_l \delta^j_k) \tag{1.14}$$

which is equal to 2 when $k = i$ and $l = j$. Thus the element of $\bigwedge^2(E^*)$ that would be considered to be dual to $e_i \wedge e_j$ is $\frac{1}{2} e^i \wedge e^j$. More generally a dual basis to the basis (1.4) is given by:

$$\left\{ \frac{1}{m!}(e^{i_1} \wedge \dots \wedge e^{i_m}) \;\middle|\; 1 \leq i_1 < i_2 < \dots < i_m \leq n \right\}$$

Let us now conclude this subsection by discussing the role of $T(E^*)$, $S(E^*)$ and $\bigwedge^\bullet(E^*)$. The main point is that the tensor algebra of the dual, denoted $T(E^*)$, can be considered to be the vector space of *multi-linear forms on E*. Linear forms on E form the dual space $T^1(E^*) = E^*$. Bilinear forms on E are those functions $B : E \times E \longrightarrow \mathbb{R}$ such that on the one hand it is linear in the first variable $B(\lambda x + \mu y, z) = \lambda B(x, z) + \mu B(y, z)$ (for every $\lambda, \mu \in \mathbb{R}$ and $x, y, z \in E$), and on the other hand a similar identity holds for the second variable. Bilinear forms on E

are precisely the elements of $T^2(E^*) = E^* \otimes E^* \simeq (E \otimes E)^*$. A priori bilinear forms may not be symmetric nor antisymmetric. More generally, a m multilinear form on E is an element $\Theta : E \times E \times \ldots \times E \longrightarrow \mathbb{R}$ which is linear in each of its variables:

$$\Theta(x_1, x_2, \ldots, \lambda\, x_k + \mu\, y, \ldots, x_m) = \lambda\, \Theta(x_1, x_2, \ldots, x_k, \ldots, x_m) + \mu\, \Theta(x_1, x_2, \ldots, y, \ldots, x_m)$$

Although the following result is computational, it is important, and the proof is useful to understand how vectors and covectors interact.

Proposition 1.6. *There is a canonical isomorphism between m multilinear forms on E and the elements of $T^m(E^*)$.*

Proof. Let Θ be a m multilinear form. Then evaluating it on a set of basis vectors $e_{i_1}, \ldots, e_{i_m}$ gives a real number $\Theta(e_{i_1}, \ldots, e_{i_m})$ that we denote by $\Theta_{i_1 \ldots i_m}$. Repeting the process for every combination of m basis vectors of E, one obtains a family of real numbers. Then, the object $\Theta_{i_1 \ldots i_m}\, e^{i_1} \otimes \ldots \otimes e^{i_m}$ (Einstein summation convention implied) is an element of $T^m(E^*)$.

Conversely, let Θ be an element of $T^m(E^*)$, and let us decompose it on the basis (1.12):

$$\Theta = \Theta_{i_1 \ldots i_m}\, e^{i_1} \otimes \ldots \otimes e^{i_m}$$

where $\Theta_{i_1 \ldots i_m} \in \mathbb{R}$ and where the Einstein summation convention has been used. Then, picking up m vectors $x_1 = x_1^{j_1} e_{j_1}, x_2 = x_2^{j_2} e_{j_2}, \ldots, x_m = x_m^{j_m} e_{j_m} \in E$ (Einstein summation convention implied on repeated indices) one can write, using Equation (1.13):

$$\begin{aligned}
\Theta_{i_1 \ldots i_m}\, e^{i_1} \otimes \ldots \otimes e^{i_m}(x_1 \otimes \ldots \otimes x_m) &= \Theta_{i_1 \ldots i_m} x_1^{j_1} \ldots x_m^{j_m} e^{i_1} \otimes \ldots \otimes e^{i_m}(e_{j_1} \otimes \ldots \otimes e_{j_m}) \\
&= \Theta_{i_1 \ldots i_m} x_1^{j_1} \ldots x_m^{j_m} \delta^{i_1}_{j_1} \ldots \delta^{i_m}_{j_m} \\
&= \Theta_{i_1 \ldots i_m} x_1^{i_1} \ldots x_m^{i_m}
\end{aligned}$$

We define this real number to be $\Theta(x_1, \ldots, x_m)$. One can check that the assignment $(x_1, \ldots, x_m) \longmapsto \Theta(x_1, \ldots, x_m)$ is linear in each of its variable. Thus, Θ can be seen as a m multilinear form on E. $\square$

Now recall that – although it is not mathematically totally rigorous – we consider the symmetric algebra and the exterior algebra as subspaces of the tensor algebra. How do they fit in the picture? It turns out that $S^m(E^*)$ is the space of m multilinear forms that are fully symmetric, that is to say, those $\Xi \in T^m(E^*)$ such that, for any choice of permutation $\sigma \in S^m$:

$$\Xi(x_{i_1}, x_{i_2}, \ldots, x_{i_m}) = \Xi(x_{\sigma(i_1)}, x_{\sigma(i_2)}, \ldots, x_{\sigma(i_m)})$$

By Proposition 1.6, the action of a symmetric m multilinear form on a set of m vectors $x_1, \ldots, x_m$ can be written as follows:

$$\Xi(x_1, \ldots, x_m) = \Xi(x_1 \otimes \ldots \otimes x_m) \tag{1.15}$$

when, on the right hand side, we understand that Ξ has been developed on the basis (1.12) of $T^m(E^*)$ and Equation (1.13) is used. On the other hand, the space $\bigwedge^m(E^*)$ is the space of m multilinear forms that are fully anti-symmetric, that is to say, those $\Omega \in T^m(E^*)$ such that, for any choice of permutation $\sigma \in S^m$:

$$\Omega(x_{i_1}, x_{i_2}, \ldots, x_{i_m}) = (-1)^\sigma \Omega(x_{\sigma(i_1)}, x_{\sigma(i_2)}, \ldots, x_{\sigma(i_m)})$$

where $(-1)^\sigma$ is the signature of σ. For example, given a bilinear form $B : E \times E \longrightarrow \mathbb{R}$, the bilinear form A defined by:

$$A(x, y) = B(x, y) - B(y, x)$$

is fully antisymmetric because $A(x, y) = -A(y, x)$. By Proposition 1.6, the action of a fully antisymmetric m multilinear form $\Omega \in \bigwedge^m(E^*)$ on a set of m vectors $x_1, \ldots, x_m$ can be written as follows:

$$\Omega(x_1, \ldots, x_m) = \Omega(x_1 \otimes \ldots \otimes x_m) \tag{1.16}$$

when, on the right hand side, we understand that Ξ has been developed on the basis (1.12) of $T^m(E^*)$ and Equation (1.13) is used. We often call the fully anti-symmetric multilinear forms on E *alternating tensors.*

Exercise 1.7. Given a trilinear form $T \in T^3(E^*)$, check that the trilinear form R defined by:

$$R(x, y, z) = T(x, y, z) + T(y, z, x) + T(z, x, y) - T(x, z, y) - T(z, y, x) - T(y, x, z)$$

is fully antisymmetric.

Last but not least, let us give a formula to compute the value of an alternating tensor, when fed with a bunch of vectors. For every $1 \leq m \leq n$, it is only defined on *decomposable* elements of $\bigwedge^m(E^*)$, i.e. those of the form $\varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_m$, for some $\varphi_i \in E^*$. Here, in particular, the index is not a coordinate index. Pick up such a decomposable element, then one has:

$$\varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_m(x_1, \ldots, x_m) = \det \begin{pmatrix} \varphi_1(x_1) & \varphi_1(x_2) & \ldots & \varphi_1(x_{m-1}) & \varphi_1(x_m) \\ \varphi_2(x_1) & & & & \varphi_2(x_m) \\ \ldots & & \ldots & & \ldots \\ \varphi_{m-1}(x_1) & & & & \varphi_{m-1}(x_m) \\ \varphi_m(x_1) & \varphi_m(x_2) & \ldots & \varphi_m(x_{m-1}) & \varphi_m(x_m) \end{pmatrix} \tag{1.17}$$

for every $x_1, \ldots, x_m \in E$. This formula coincides with Equation (1.16) when $\Omega = \varphi_1 \wedge \ldots \wedge \varphi_m$. Applying this formula to a decomposable alternating 2-tensor $\varphi_1 \wedge \varphi_2$, one has:

$$\varphi_1 \wedge \varphi_2(x_1, x_2) = \varphi_1(x_1)\varphi_2(x_2) - \varphi_1(x_2)\varphi_2(x_1) \tag{1.18}$$

This equation, when $\varphi_1 \wedge \varphi_2 = e^i \wedge e^j$, and when $x_1 = e_k$ and $x_2 = e_l$, gives:

$$e^i \wedge e^j(e_k, e_l) = \delta^i_k \delta^j_l - \delta^i_l \delta^j_k \tag{1.19}$$

which is precisely one half of Equation (1.14). Thus it gives $+1$ when $k = i$ and $l = j$. This shows that looking at elements of $\bigwedge^\bullet(E^*)$ as alternating multilinear forms on E is a well-defined and even legitimate thing to do. See Chapter 15 of the book [Bamberg and Sternberg, 1988] for a good presentation on this topic, and Chapter 12 in [Lee, 2003] (Chapter 14 in the 2012 edition) for a detailed, mathematically oriented one (beware of the notations that are different than here!).

Exercise 1.8. Show that for every alternating tensors $\Omega \in \bigwedge^m(E)$, then its evaluation on identical vectors is zero:

$$\Omega(x_1, \ldots, x, \ldots, x, \ldots, x_m) = 0$$

From this result, deduce that for any choice of vectors $x_1, \ldots, x_m$, if one of the x_i is a linear combination of the others, then $\Omega(x_1, \ldots, x_m) = 0$.

1.2 Scalar product and Hodge star operator

Now suppose that E is additionally equipped with a *pseudo-Riemaniann metric*, that is to say:

Definition 1.9. *A pseudo-Riemaniann metric on a vector space E is a map g from $E \times E$ to $\mathbb{R}$ which is:*

1. *bilinear, e.g. for the first term* $g(\lambda x + \mu y, z) = \lambda g(x,z) + \mu g(y,z)$ *for every* $\lambda, \mu \in \mathbb{R}$ *and* $x, y, z \in E$ *(and the same occurs for the second term)*

2. *symmetric, i.e.* $g(x,y) = g(y,x)$ *for every* $x, y \in E$

3. *non-degenerate, i.e. if* $g(x,y) = 0$ *for every* $y \in E$ *then* $x = 0$

All three items are independent of the choses basis of E. Given the definition of the symmetric algebra, one can see the metric as a bilinear map $g : E \odot E \longrightarrow \mathbb{R}$. One can always define a metric on a vector space since one can check that, given a basis $e_1, \ldots, e_n$, it is sufficient to define g from its action on the basis vectors e_i by:

$$g(e_i, e_i) = 1 \quad \text{and} \quad g(e_i, e_j) = 0 \quad \text{when } i \neq j, \tag{1.20}$$

and then to formally extend it to all of E by assuming it is bilinear. Notice however that there exist alternative choices of scalar product that do not satisfy Equation (1.20), and more generally one writes[3]:

$$g_{ij} = g(e_i, e_j)$$

The metric can then be represented, in a given basis $e_1, \ldots, e_n$, as an $n \times n$ symmetric matrix G, whose components we write g_{ij}. Since the metric is symmetric, i.e. $g_{ij} = g_{ji}$, then there are only $\frac{n(n+1)}{2}$ independent coefficients in the matrix (the diagonal and the upper triangular part, or the diagonal and the lower triangular part). Being symmetric, the matrix can be diagonalized: the number p of positive eigenvalues determines what is called the *signature* of the metric, which is denoted by (p, q), the number q being the number of negative eigenvalues. Notice that another convention uses the reverse order (q, p). Sylvester's law of inertia ensures that the signature of the metric tensor g is invariant under any change of basis. There are no null eigenvalue because the metric is non-degenerate. In particular, the matrix G is invertible. Using these data, one can write the metric explicitly, as a bilinear symmetric form on $E \times E$:

$$g = g_{ij}\, e^i \odot e^j \in S^2(E^*)$$

where we have used the Einstein summation convention. Using the characterization of symmetric multilinear forms as elements of $T(E^*)$ (see Equation (1.15)), an explicit computation then shows that:

$$g(e_k, e_l) = g_{ij}\, e^i \odot e^j (e_k, e_l) = \frac{1}{2} g_{ij}\, (e^i \otimes e^j + e^j \otimes e^i)(e_k \otimes e_l) = \frac{1}{2} g_{ij} (\delta^i_k \delta^j_l + \delta^i_l \delta^j_k) = g_{kl}$$

because $g_{kl} = g_{lk}$. The result is precisely what we should expect.

Remark 1.10. From now on, given a metric g, when we say *orthonormal basis*, we think of a basis $e_1, \ldots, e_n$ satisfying:

$$\begin{aligned} g(e_i, e_i) &= 1 && \text{for every } 1 \leq i \leq p \\ g(e_i, e_i) &= -1 && \text{for every } p+1 \leq i \leq n \\ g(e_j, e_k) &= 0 && \text{for every } j \neq k \end{aligned}$$

where p is the number of negative eigenvalues of g. In other words, we put the eigenvectors with positive eigenvalues (normed to 1) first in order. This is somewhat consistent with some conventions in Minkowski space whose metric's signature we set to $(3, 1) = (- + ++)$: we often consider the time like coordinate to be either the fourth coordinate or the zeroth one. In any case, the first, second and third coordinates are space-like, and correspond to the positive eigenvalue $+1$. Obviously for a pseudo-Riemannian metric there is not negative eigenvalue, as for an Euclidean metric.

[3]This notation is used in general relativity: space-time is a four dimensional manifold and the metric $g_{\mu\nu}$ is a notation for $g(\partial_\mu, \partial_\nu)$.

The metric induces a morphism of vector space $\widetilde{g} : E \longrightarrow E^*$:

$$\begin{aligned} \widetilde{g} :\ & E \longrightarrow E^* \\ & x \longmapsto g(x,\cdot) : y \longmapsto g(x,y) \end{aligned}$$

where we have used the Einstein summation convention. In particular, this definition tells us that acting on a basis vector with $\widetilde{g}$ reads:

$$\widetilde{g}(e_i) = g_{ij}\, e^j$$

Moreover, the nondegeneracy of the metric is equivalent to the injectivity of this map, and hence, of its bijectivity (because E is finite dimensional).

Exercise 1.11. The proof is left as an exercise.

The inverse map is denoted $\widetilde{g}^{-1} : E^* \longrightarrow E$ and it can be used to define a metric g^{-1} on E^* induced from g, which is such that the maps $\widetilde{g}^{-1}$ and $\widetilde{g}$ are isometries[4]. Following the same lines of argument as above, the scalar product g^{-1} gives rise to an isometry $\widetilde{g^{-1}} : E^* \longrightarrow E$, which actually is such that $\widetilde{g^{-1}} = \widetilde{g}^{-1}$. That is to say, we shall have:

$$g^{-1}(\widetilde{g}(x), \widetilde{g}(y)) = g(x,y) \qquad \text{and} \qquad g(\widetilde{g}^{-1}(\varphi), \widetilde{g}^{-1}(\chi)) = g^{-1}(\varphi, \chi)$$

for every $x, y \in E$ and $\varphi, \chi \in E^*$. This is equivalent to the commutativity of the following diagram:

$$\begin{tikzcd}
E \odot E \arrow[dr, bend left, "g"] \arrow[dd, bend right, "\widetilde{g} \odot \widetilde{g}"'] & \\
& \mathbb{R} \\
E^* \odot E^* \arrow[ur, bend right, "g^{-1}"'] \arrow[uu, bend right, "\widetilde{g}^{-1} \odot \widetilde{g}^{-1}"']
\end{tikzcd}$$

There is an $n \times n$ matrix H associated to the metric g^{-1} and the basis $e^1, \ldots, e^n$. We adopt the convention that its components are written g^{ij}, with exponents, so that:

$$g^{-1}(e^i, e^j) = g^{ij}$$

The metric g^{-1} then corresponds to a bilinear symmetric operator on E^* expressed as:

$$g^{-1} = g^{ij}\, e_i \odot e_j \in S^2(E)$$

One can show that the matrix H is the inverse of the matrix G. This implies that the signature of the metric g^{-1} is the same as the one of g.

Exercise 1.12. Using the symmetry of G and the fact that it is invertible, prove that $H = G^{-1}$.

[4] This is a particular form of the Riesz representation theorem in mathematics, which actually applies to infinite dimensional Hilbert spaces.

The maps $\widetilde{g}$ and $\widetilde{g}^{-1}$ satisfy:

$$\widetilde{g}(e_i) = g_{ij}\, e^j \in E^* \qquad \text{and} \qquad \widetilde{g}^{-1}(e^k) = g^{kl}\, e_l \in E$$

where the Einstein summation convention has been used. These equations explain what people mean by saying that the metric raises and lowers the indices. For example, take a vector $x = x^i e_i \in E$, where the $x^i \in \mathbb{R}$ are the coordinates of x with respect to the basis $(e_i)_{1\leq i\leq n}$. In physics in general one is only interested in the coordinates x^i, then lowering the indices i amounts to applying $\widetilde{g}$ to x:

$$\widetilde{g}(x) = x^i g_{ij} e^j = x_j e^j$$

where we have defined $x_j := g_{ij}x^i$, which in the present context are thus the coordinates of $\widetilde{g}(x)$ with respect to the basis $(e^i)_{1\leq i\leq n}$ of E^*. Sometimes, the maps $\widetilde{g}$ and $\widetilde{g}^{-1}$ are called musical isomorphisms, and are denoted $\flat$ (*flat*) and $\sharp$ (*sharp*), respectively:

$$\flat : E \longrightarrow E^* \qquad \text{and} \qquad \sharp : E^* \longrightarrow E$$

They are inverse to one another. This notation is useful because it lightens the notation, by writing $x^\flat$ instead of $\widetilde{g}(x)$, and $\varphi^\sharp$ instead of $\widetilde{g}^{-1}(\varphi)$. Then, while $x = x^i e_i$, we have $x^\flat = x_j e^j$, with $x_j = g_{jk}x^k$. That is why we say that $\flat$ lowers the indices (of the coordinates!) while the map $\sharp$ raises them.

In particular, given a tensor ${A_{i_1\dots i_k}}^{j_1\dots j_l}$ one can raise and lower the indices using the musical isomorphisms, for example:

$$A_{i_1\dots i_{r-1}}{}^{j_0}{}_{i_{r+1}\dots i_k}{}^{j_1\dots j_l} \equiv g^{j_0 i_r} {A_{i_1\dots i_k}}^{j_1\dots j_l} \tag{1.21}$$

where the Einstein summation convention has been used. This has the following mathematical meaning: $A = {A_{i_1\dots i_k}}^{j_1\dots j_l}\, e^{i_1}\otimes\dots\otimes e^{i_k}\otimes e_{j_1}\otimes\dots\otimes e_{j_l}$ is a mixed tensor belonging to $T^k(E^*)\otimes T^l(E)$. The left-hand side of Equation (1.21) is precisely the tensor obtained when one has applied $\sharp = \widetilde{g}^{-1}$ on the r-th covariant leg of A. In other words (with Einstein summation convention):

$$\begin{aligned} &A_{i_1\dots i_{r-1}}{}^{j_0}{}_{i_{r+1}\dots i_k}{}^{j_1\dots j_l}\, e^{i_1}\otimes\dots\otimes e^{i_{r-1}}\otimes e_{j_0}\otimes e^{i_{r+1}}\otimes\dots\otimes e^{i_k}\otimes e_{j_1}\otimes\dots\otimes e_{j_l} \\ &\qquad = \Big(\underbrace{\mathrm{id}_{E^*}\otimes\dots\mathrm{id}_{E^*}}_{r-1 \text{ terms}}\otimes\sharp\otimes\underbrace{\mathrm{id}_{E^*}\otimes\dots\mathrm{id}_{E^*}}_{k-r \text{ terms}}\otimes\underbrace{\mathrm{id}_{E}\otimes\dots\mathrm{id}_{E}}_{l \text{ terms}}\Big)(A) \end{aligned}$$

is an element of $T^{r-1}(E^*)\otimes E\otimes T^{k-r}(E^*)\otimes T^l(E)$.

The metric g can be extended to the exterior algebra $\bigwedge^\bullet(E)$ by using the Gram determinant. For every $1\leq m\leq n$, we will define it first on *decomposable* multivectors, i.e. those elements of $\bigwedge^m(E)$ that are of the form $x_1\wedge\dots\wedge x_m$ for $x_1,\dots,x_m\in E$, and then extend it to all of $\bigwedge^m(E)$ by linearity in each argument. More precisely, let $\alpha,\beta\in\bigwedge^m(E)$ be two decomposable multivectors, so that they can be written as $\alpha = x_1\wedge\dots\wedge x_m$ and $\beta = y_1\wedge\dots\wedge y_m$. Then we define the scalar product of α and β as:

$$\langle\alpha,\beta\rangle = \det\begin{pmatrix} g(x_1,y_1) & g(x_1,y_2) & \dots & g(x_1,y_{m-1}) & g(x_1,y_m) \\ g(x_2,y_1) & & & & g(x_2,y_m) \\ \dots & & \dots & & \dots \\ g(x_{m-1},y_1) & & & & g(x_{m-1},y_m) \\ g(x_m,y_1) & g(x_m,y_2) & \dots & g(x_m,y_{m-1}) & g(x_m,y_m) \end{pmatrix}$$

The determinant on the right hand side ressembles what is called the Gram determinant.

Exercise 1.13. Prove that the so-called scalar product $\langle\alpha,\beta\rangle$ is symmetric and bilinear.

Exercise 1.14. Prove that if $e_1, \dots, e_n$ is an orthonormal basis of E (see Remark 1.10 for a definition), the scalar product satisfies:

$$\langle e_{i_1} \wedge \dots \wedge e_{i_m}, e_{j_1} \wedge \dots \wedge e_{j_m} \rangle = g_{i_1 j_1} g_{i_2 j_2} \dots g_{i_m j_m} \tag{1.22}$$

Then, since any m-multivector can be written as the linear sum of decomposable m-multivectors – such as the basis $(e_{i_1} \wedge \dots \wedge e_{i_m})_{1 \leq i_1 < \dots < i_m \leq n}$ of $\bigwedge^m(E)$ – one can extend the inner product to the whole of $\bigwedge^m(E)$ by enforcing linearity on each argument. For example, let $\alpha = \sum_i \alpha_i$ and $\beta_j = \sum_j \beta_j$ be two m-multivectors written as linear combinations of the decomposable multivectors α_i and β_j; then we set:

$$\langle \alpha, \beta \rangle = \sum_{i,j} \langle \alpha_i, \beta_j \rangle$$

We apply the same idea at every level $1 \leq m \leq n-1$ (for $m = 0$ and $m = n$ the space $\bigwedge^m(E)$ is one-dimensional) so that the scalar product is defined on the entirety of the exterior algebra $\bigwedge^\bullet(E)$. It can be shown that the left hand side does not depend on the decomposition of α and β in terms of decomposable multivectors (see a proof in the Appendix of Chapter 18 of [Bamberg and Sternberg, 1988]). This definition also work on $\bigwedge^\bullet(E^*)$ as well, when one takes g^{-1} instead of g.

Proposition 1.15. *The so-called scalar product $\langle \alpha, \beta \rangle$ is non-degenerate so it indeed bears well its name.*

Proof. One can suppose that $1 \leq m \leq n-1$ and that the basis of E is orthonormal with respect to the metric g. Let α be an m-multivector such that:

$$\langle \alpha, \beta \rangle = 0 \qquad \text{for every } \beta \in \bigwedge^m(E) \tag{1.23}$$

The element α admits the following decomposition on the basis of $\bigwedge^m(E)$:

$$\alpha = \alpha^{i_1 \dots i_m} e_{i_1} \wedge \dots \wedge e_{i_m}$$

where the Einstein summation convention has been used, and where we assumed $i_1 < \dots < i_m$. Apply Equation (1.23) to $e_{i_1} \wedge \dots \wedge e_{i_m}$, for some chosen $i_1 < \dots < i_m$. Then by Equation (1.22), one obtains $0 = \langle \alpha, e_{i_1} \wedge \dots \wedge e_{i_m} \rangle = \alpha^{i_1 \dots i_m}$. Doing this for every basis vector of $\bigwedge^m(E)$, one proves that $\alpha = 0$. ☐

Then, since $\dim(\bigwedge^m(E)) = \dim(\bigwedge^{n-m}(E)) = \binom{n}{m}$, one can use the inner product on the exterior algebra to identify $\bigwedge^m(E)$ and $\bigwedge^{n-m}(E)$, via what is called the *Hodge star operator*. Let $(e_i)_{1 \leq i \leq n}$ be a basis of E and denote by

$$\omega = \frac{1}{\sqrt{|\det(G)|}} e_1 \wedge \dots \wedge e_n \tag{1.24}$$

the standard volume element of E, which also generates the one-dimensional space $\Lambda^n(E)$. The normalization is such that $\langle \omega, \omega \rangle = (-1)^q$, which depends on the number q of negative eigenvalues of the metric.

Exercise 1.16. The proof is left to the reader.

Then, the Hodge star operator is a linear map[5] $\star : \bigwedge^\bullet(E) \longrightarrow \bigwedge^{n-\bullet}(E)$ which is defined on $\bigwedge^m(E)$ (for every $1 \leq m \leq n$) by the following identity:

$$\alpha \wedge (\star\beta) = \langle \alpha, \beta \rangle\, \omega \tag{1.25}$$

for every two m-multivectors $\alpha, \beta \in \bigwedge^m(E)$[6]. Notice that a choice of orientation for E has a direct consequence on the sign of the volume form, and thus on the definition of the Hodge star operator. This indirect definition can be made more explicit by looking at the effect of $\star$ on a basis of $\bigwedge^m(E)$. For any ordered subset $J = \{j_1, \ldots, j_m\}$ of $\{1, \ldots, n\}$ (i.e. such that $1 \leq j_1 < \ldots < j_m \leq n$), let us call $J^c = \{1, \ldots, n\} - J$ the ordered set corresponding to the remaining integer that do not belong to J. We denote the $n - m$ elements of J^c as $j_{m+1}, \ldots, j_n$; they are such that $j_k < j_l$ for $m < k < l$. Then, we denote:

$$e_J = e_{j_1} \wedge \ldots \wedge e_{j_m} \quad \text{and} \quad e_{J^c} = e_{j_{m+1}} \wedge \ldots \wedge e_{j_n}$$

Moreover, let σ_J be the permutation of n elements that sends the ordered set $\{1, \ldots, n\}$ to $\{j_1, \ldots, j_m, j_{m+1}, \ldots, j_n\}$, i.e. it is such that $\sigma_J(k) = j_k$. Thus, under the action of σ_J, the n-form $e_1 \wedge \ldots \wedge e_n$ becomes $e_J \wedge e_{J^c}$. Then, by fixing an ordered set $I = \{i_1, \ldots, i_m\}$ of m elements $1 \leq i_1 < \ldots < i_m \leq n$ and by computing every term of the form $e_J \wedge \star(e_I)$ for every ordered set $J = \{j_1, \ldots, j_m\} \subset \{1, \ldots, n\}$, one obtains the following formula:

$$\star\,(e_I) = \sum_{\substack{J \subset \{1,\ldots,n\} \\ J \text{ ordered}}} (-1)^{\sigma_J} \frac{\langle e_J, e_I \rangle}{\sqrt{|\det(G)|}}\, e_{J^c} \tag{1.26}$$

where $\langle e_J, e_I \rangle$ is a notation for $\langle e_{j_1} \wedge \ldots \wedge e_{j_m}, e_{i_1} \wedge \ldots \wedge e_{i_m} \rangle$, which is a minor of the Gram matrix.

Exercise 1.17. Using Equation (1.25), show that in $\mathbb{R}^3$ with standard basis e_1, e_2, e_3 and with a metric $g = g_{ij}\, e^i \odot e^j$:

$$\star\, e_i = \frac{1}{\sqrt{|\det(G)|}} \Big(g_{1i}\, e_2 \wedge e_3 - g_{2i}\, e_1 \wedge e_3 + g_{3i}\, e_1 \wedge e_2 \Big)$$

$$\star\,(e_i \wedge e_j) = \frac{1}{\sqrt{|\det(G)|}} \left(\det \begin{pmatrix} g_{2i} & g_{2j} \\ g_{3i} & g_{3j} \end{pmatrix} e_1 - \det \begin{pmatrix} g_{1i} & g_{1j} \\ g_{3i} & g_{3j} \end{pmatrix} e_2 + \det \begin{pmatrix} g_{1i} & g_{1j} \\ g_{2i} & g_{2j} \end{pmatrix} e_3 \right)$$

and check that these formulas are indeed those corresponding to Equation (1.26).

Finding an orthonormal basis with respect to the metric (see Remark 1.10 for the definition) is equivalent to diagonalizing the associated matrix G, and rescale the diagonal values so that they become either 1 or -1. The determinant is then the product of the diagonal values g_{ii}, and its absolute value is 1. For any ordered m-index $I = \{i_1, \ldots, i_m\} \subset \{1, \ldots, n\}$, one sets $\lambda_I = g_{i_1 i_1} g_{i_2 i_2} \ldots g_{i_m i_m}$ the product of the m diagonal values. Then, assuming the basis $e_1, \ldots, e_n$ is orthonormal, one has:

$$\star\,(e_I) = (-1)^{\sigma_I} \lambda_I\, e_{I^c} = \pm\, e_{I^c} \tag{1.27}$$

For an explicit formulation using coordinates, see the *nCatLab*.

Let us illustrate Equation (1.26) on several examples. First, the following two identities:

$$\star\,(1_{\mathbb{R}}) = \omega \qquad \text{and} \qquad \star\,\omega = (-1)^q\, 1_{\mathbb{R}}$$

[5] This notation means that for every $1 \leq m \leq n$ the Hodge star operator sends $\bigwedge^m(E)$ to $\bigwedge^{n-m}(E)$.

[6] Notice that in general the Hodge star is usually defined on covectors. In that case the one shall use e^i instead of e_i, and the normalization factor $\sqrt{|\det(G)|}$ instead of $\sqrt{|\det(G^{-1})|}$.

are valid in every case, where $1_{\mathbb{R}}$ is the generator of $\bigwedge^0(E) = \mathbb{R}$, i.e. $1_{\mathbb{R}} = 1$. When $E = \mathbb{R}^2$ with the standard Euclidean metric and standard orientation, $\bigwedge^1(E)$ is two-dimensional as well and can be identified with E, so that the Hodge star operator can be seen as an endomorphism of E and coincides with a rotation by $\frac{\pi}{2}$. This can be checked on any orthonormal basis $\{e_1, e_2\}$ of $E = \bigwedge^1(E)$, since we have:

$$\star\, e_1 = e_2 \qquad \text{and} \qquad \star\, e_2 = -e_1$$

When $E = \mathbb{R}^3$ with the standard Euclidean metric and standard oriention, the two spaces $\bigwedge^1(E) \simeq E$ and $\bigwedge^2(E)$ are both three-dimensional and the Hodge star operator draws a relationship between the wedge product and the cross product:

$$\star(x \times y) = x \wedge y \qquad \text{and} \qquad \star\,(x \wedge y) = x \times y$$

Exercise 1.18. Can you compute the effect of the Hodge star operator on Minkowski space ? The Minkowski metric has signature $(3, 1)$.

Last but not least: the Hodge star operator is not an involution of the exterior algebra, but almost:

$$\star\star\,\alpha = (-1)^{m(n-m)+q}\,\alpha \tag{1.28}$$

for any m-multivector α, and where q is the number of negative eigenvalues in the signature (p, q) of the metric. This implies that, the inverse to the hodge star operator $\star : \bigwedge^m(E) \longrightarrow \bigwedge^{n-m}(E)$, is the operator $\star^{-1} : \bigwedge^{n-m}(E) \longrightarrow \bigwedge^m(E)$ defined by:

$$\star^{-1} = (-1)^{m(n-m)+q}\,\star \tag{1.29}$$

The final identity worth noticing is:

$$\langle \star\,\alpha, \star\,\beta \rangle = (-1)^q \langle \alpha, \beta \rangle \tag{1.30}$$

This equation proves that the Hodge star operator is almost an isometry of the exterior algebra, up to a sign.

Exercise 1.19. Using Exercise 1.17, prove that in $\mathbb{R}^3$ with the given metric g,

$$\star\star\,e_i = (-1)^q\,e_i \quad \text{and} \quad \star\star\,(e_i \wedge e_j) = (-1)^q\,e_i \wedge e_j$$

Exercise 1.20. Using Equation (1.27) and the fact that $(-1)^{\sigma_I^c} = (-1)^{\sigma_I + m(n-m)}$, prove Equation (1.28), that in turns implies Equation (1.30).

Finally notice that usually the Hodge star operator is defined on the exterior algebra of covectors $\bigwedge^\bullet(E^*)$. Pay heed to the differences that this implies: in particular one should use G^{-1} instead of G, and use exponents (resp. indices) in place of indices (resp. exponents).

2 Differential calculus on $\mathbb{R}^n$

This chapter is dedicated to the study of smooth functions, vector fields and differential forms on the euclidean vector space $E = \mathbb{R}^n$. A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ that is infinitely differentiable is called *smooth* and the set of all such functions is denoted $\mathcal{C}^\infty(\mathbb{R}^n)$. It is an infinite dimensional real vector space, and since the product of two smooth functions is always smooth, it is actually an algebra over $\mathbb{R}$. Vector fields are derivations of this algebra, while differential 1-forms are their dual objects.

2.1 Tangent vectors and vector fields on $\mathbb{R}^n$

In this section we generalize the notion of tangent vector to a curve. The idea is the following: assume $n = 3$ and that we represent the trajectory of a physical object in space by a parametrized (differentiable) curve $\gamma : [0,1] \longrightarrow \mathbb{R}^3$. For every $t_0 \in [0,1]$, the velocity vector $\dot\gamma(t_0)$ is often represented as a tangent vector to the curve which has the following properties:

1. it is a 3-dimensional vector based at $\gamma(t_0)$;
2. it is tangent to the curve and points towards the future, i.e. towards the points $\gamma(t_1)$, for small $t_1 > t_0$;
3. its norm is the velocity $\|\dot\gamma(t_0)\|$ of the physical object at time t_0.

The direction and the norm of the tangent vector are somewhat "internal" informations because they can be represented by an abstract vector $X_{\Gamma(t_0)}$ based at the origin of an abstract 3-dimensional space, and which points in the same direction as $\dot\gamma(t_0)$ and has the same norm. The base point at which the velocity vector is defined however is an external information since it depends on the curve γ.

Hence, an abstraction of the velocity vector and of the data contained in the three items above can be equivalently represented by a couple $(\gamma(t_0), X_{\gamma(t_0)})$ of the product space $\mathbb{R}^3 \times \mathbb{R}^3$. The first $\mathbb{R}^3$ is the "position space" (or *configuration space*): it is the space in which the trajectory γ of the physical object takes values. The second $\mathbb{R}^3$ is the "velocity space": for any given point $x \in \mathbb{R}^3$ (of the position space), The curve $\gamma : t \longrightarrow \mathbb{R}^3$ encoding the trajectory of a physical object defines a path in the position space. When t varies in $[0,1]$, the direction and the norm of the tangent vector of γ varies, which in turn defines a path $X : t \longrightarrow X_{\gamma(t)}$ in the velocity space. Thus, the path $t \longrightarrow (\gamma(t), X_{\gamma(t)})$ defines a curve in $\mathbb{R}^3 \times \mathbb{R}^3$ which contains every data on the physical position of the object and its velocity.

Another way of looking at tangent vectors is the following: let $t_0 \in]0,1[$, then the tangent vector $\dot\gamma(t_0)$ acts on any smooth function $f \in \mathcal{C}^\infty(\mathbb{R}^3)$ by:

$$\dot\gamma(t_0)(f) = \frac{d(f \circ \gamma)}{dt}(t_0) \tag{2.1}$$

In particular, if f is a function locally constant at $\gamma(t_0)$, then $\dot\gamma(t_0)(f) = 0$. What is not apparent on this equation is that, although the right hand side involve γ, the left hand side only depends on the velocity vector at the point $\gamma(t_0)$. Any other curve $\eta : [0,1] \longrightarrow \mathbb{R}^3$ such that $\eta(t_0) = \gamma(t_0)$ and such that $\dot\eta(t_0) = \dot\gamma(t_0)$, satisfies $\frac{d(f\circ\gamma)}{dt}(t_0) = \frac{d(f\circ\eta)}{dt}(t_0)$. This implies that we can forget about the dependency on the curve and look at elements of the tangent space as linear maps sending functions to real numbers: for any given point $x \in \mathbb{R}^3$ and any vector $X_x \in \mathbb{R}^3$, pick up a parametrized curve $\gamma : [0,1] \longrightarrow \mathbb{R}^3$ and $t_0 \in]0,1[$ such that $\gamma(t_0) = x$, and that $\dot\gamma(t_0) = X_x$,

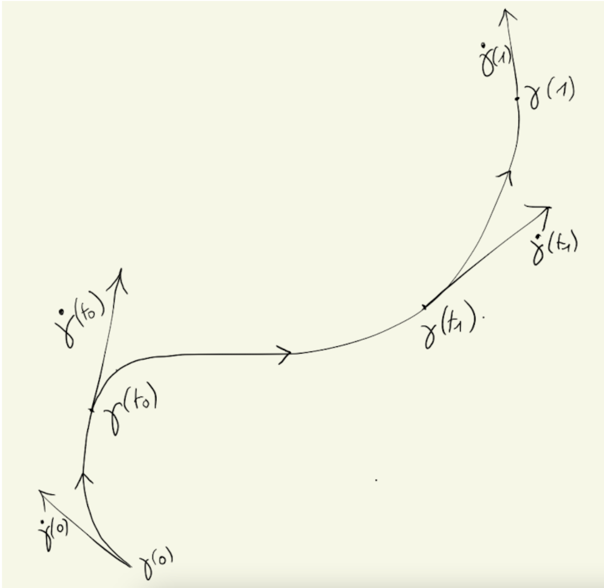

Figure 1: Usually we represent a path and its tangent vectors on the same drawing. The tangent vector $\dot{\gamma}(t)$ is based at the point $\gamma(t)$ but this is not rigorous, mathematically: the norm and the direction of $\dot{\gamma}(t)$ characterizes the tangent vector, and the base point is an external information reminding the reader that the tangent vector is attached to the point $\gamma(t)$.

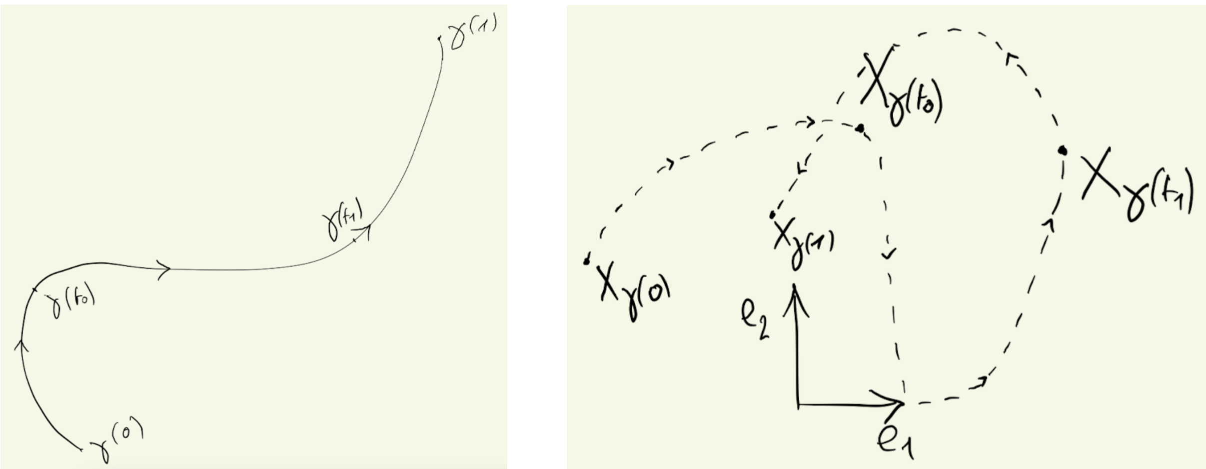

Figure 2: The figure on the left represents the path γ in the "position space" $\mathbb{R}^n$, and the figure on the right is a possible representation of the path $X : t \mapsto X_{\gamma(t)}$ of velocity vectors tangent to the curve γ, in the "velocity space" $\mathbb{R}^n$ (to determine the exact form of this path, one has to compute every $\dot{\gamma}(t)$). For each $t \in [0,1]$, the vector $X_{\gamma(t)}$ has the same norm and the same direction as $\dot{\gamma}(t)$. The path $t \longmapsto (\gamma(t), X_{\gamma(t)})$ in the abstract product space $\mathbb{R}^n \times \mathbb{R}^n$ contains the same data which is represented in Figure 1.

then X_x defines a linear morphism $X_x : \mathcal{C}^\infty(\mathbb{R}^3) \longrightarrow \mathbb{R}$ via Equation (2.1). Due to the properties of the time derivative, one can show that this action satisfies the following properties:

$$X_x(\lambda f + \mu g) = \lambda X_x(f) + \mu X_x(g) \tag{2.2}$$
$$X_x(fg) = X_x(f)g(x) + f(x)X_x(g) \tag{2.3}$$

for any $f, g \in \mathcal{C}^\infty(\mathbb{R}^3)$ and $\lambda, \mu \in \mathbb{R}$. The first equation characterizes the fact that X_x :

$\mathcal{C}^\infty(\mathbb{R}^3) \longrightarrow \mathbb{R}$ is a linear morphism, whereas the second equation implies that X_x acts as what we call *a derivation at* x. Actually, we will see that the action of the vector X_x on a function f can be identified with the directional derivative of f in the direction X_x, evaluated at the point x (see below).

Generalizing this observation to n-dimensional vector spaces gives the following definition: the *tangent space* to $\mathbb{R}^n$ at a given point x is the vector space of linear morphisms that are derivations at x, i.e. all the maps $X_x : \mathcal{C}^\infty(\mathbb{R}^n) \longrightarrow \mathbb{R}$ satisfying Equations (2.2) and (2.3); it is denoted $T_x\mathbb{R}^n$. The following Lemma says that every directional derivative is a derivation at x:

Lemma 2.1. *Let $v \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$. The linear morphism $D_{v,x} : \mathcal{C}^\infty(\mathbb{R}^n) \longrightarrow \mathbb{R}$ defined by:*

$$D_{v,x}(f) = \frac{d}{dt}\bigg|_{t=0} f(x + tv)$$

is a derivation at x, i.e. $D_{v,x} \in T_x\mathbb{R}^n$.

Exercise 2.2. Prove this Lemma, using Equation (2.1).

Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$. We denote by $\frac{\partial}{\partial x^i}\big|_x$ the directional derivative at x associated to the basis vector e_i by Lemma 2.1:

$$\frac{\partial}{\partial x^i}\bigg|_x = D_{e_i,x}$$

The notation is such that the action of $\frac{\partial}{\partial x^i}\big|_x$ precisely coincides with what is expected from such a directional derivative:

$$\frac{\partial}{\partial x^i}\bigg|_x (f) = \frac{\partial f}{\partial x^i}(x)$$

From the definitions of such elements, we deduce the following Lemma:

Lemma 2.3. *The n directional derivatives $\frac{\partial}{\partial x^1}\big|_x, \ldots, \frac{\partial}{\partial x^n}\big|_x$ at the point x are linearly independent, and there is a one-to-one correspondence between vectors v of $\mathbb{R}^n$ and directional derivatives $D_{v,x}$:*

$$v = v^i e_i \qquad \longleftrightarrow \qquad D_{v,x} = v^i \frac{\partial}{\partial x^i}\bigg|_x$$

Proof. This result can be shown as follows: pick up a set of scalars $\lambda_1, \ldots, \lambda_n$ and assume that $\sum_{i=1}^n \lambda_i \frac{\partial}{\partial x^i}\big|_x = 0$. Thus in particular, applying it to the i-th coordinate function $x^i : \mathbb{R}^n \longrightarrow \mathbb{R}$ gives $\lambda_i = 0$. Also one notices that the assignment $v \longmapsto D_{v,x}$ is a linear morphism. This shows that differential derivative decompose as ${}_{v,x} = v^i \frac{\partial}{\partial x^i}\big|_x$. Finally, this result is used to prove that the linear map $v \longmapsto D_{v,x}$ is injective. This proves that there is a one-to-one correspondence between $\mathbb{R}^n$ and the space of directional derivatives at x. □

The following proposition explains why this is also true for derivations at x:

Proposition 2.4. *The n directional derivatives $\frac{\partial}{\partial x^1}\big|_x, \ldots, \frac{\partial}{\partial x^n}\big|_x$ at the point x form a basis of $T_x\mathbb{R}^n$. In particular it means that directional derivatives at x and derivations at x are in one-to-one correspondence, and that $T_x\mathbb{R}^n$ is a n-dimensional vector space.*

Proof. We know by Lemma 2.1 that directional derivatives are derivation at x. This, together with Lemma 2.3, implies that the assignment $v \longmapsto D_{v,x}$ is an injection from $\mathbb{R}^n$ into $T_x\mathbb{R}^n$. We need only show that it is surjective. It can be proven by assigning, to each derivation X_x at x, a vector v so that its i-th coordinate coincides with $X(x^i)$: $v = X_x(x^i)e_i$. Then, showing that $D_{v,x} = X$ is just a matter of using Taylor's series expansion (the Hadamard Lemma). For a detailed proof, see Proposition 3.2 in [Lee, 2003]. □

Thus, any tangent vector X_x at x decomposes in this basis as:

$$X_x = X_x^i \left.\frac{\partial}{\partial x^i}\right|_x \tag{2.4}$$

where the X_x^i are real numbers, and result from applying X_x to the i-th coordinate function $x^i : \mathbb{R}^n \longrightarrow \mathbb{R}$, that is to say:

$$X_x^i = X_x(x^i)$$

The vector $v = X_x^i e_i$ of $\mathbb{R}^n$ which has the same coordinates as X_x then induces a directional derivative $D_{v,x}$ that precisely coincides with X_x:

$$X_x = D_{X_x^i e_i, x}$$

The one-to-one correspondence between directional derivatives at x and derivations at x is summarized by the following sequence of operations:

$$X_x = X_x^i \left.\frac{\partial}{\partial x^i}\right|_x \longrightarrow v = X_x^i e_i \longrightarrow D_{v,x} = v^i \left.\frac{\partial}{\partial x^i}\right|_x = X_x$$

Here, $v^i = X_x^i$ by construction. This sequence also describes the canonical isomorphism between $\mathbb{R}^n$ and $T_x\mathbb{R}^n$. We will often identify $\mathbb{R}^n$ with its image under the canonical bijection $v \mapsto D_{v,x}$, and will either use the notation (x, X_x) or the notation X_x for a tangent vector in $T_x\mathbb{R}^n$, depending on how much emphasis we wish to give to the point x.

Example 2.5. In relativity, if x^μ are coordinates of a point particle in space-time, then the four-velocity, often represented by its coordinate $U^\mu = \frac{dx^\mu}{d\tau}$ (where τ is the proper time), is a tangent vector to the world line of the particle.

So far we have used a geometric perspective (tangent vectors to a curve) to determine algebraic properties that they satisfy: Equation (2.2) and (2.3). Then, we have adopted the other perspective: we started from all the linear morphisms from $\mathcal{C}^\infty(\mathbb{R}^n)$ to $\mathbb{R}^n$ satisfying these equations, and we have shown that they are directional derivatives. Thus we started from algebraic properties to come back to the geometric realm. We will see in the following that this alternance between geometric and algebraic perspectives are central in the discussion. This one-to-one correspondence actually allows us to transform cumbersome geometric considerations into easier algebraic computations, and conversely, to find clear geometric illustrations of algebraic obscure notions. Let us now generalize the notion of tangent vector, to the whole of $\mathbb{R}^n$:

Definition 2.6. *The disjoint union of all tangent spaces:*

$$T\mathbb{R}^n = \bigsqcup_{x \in \mathbb{R}^n} T_x\mathbb{R}^n$$

is called the tangent bundle of $\mathbb{R}^n$*.*

The word ‘bundle’ means that several things of the same kind have been fastened or held together. We call it a *trivial* (vector) bundle because it is homeomorphic[7] to $\mathbb{R}^n \times \mathbb{R}^n$. In the latter cartesian product, we call the space on the left the *base* and the space on the right the *fiber*. The fiber *at* x is the tangent space $T_x\mathbb{R}^n$, which can then be identified with the product $\{x\} \times \mathbb{R}^n$. An element of the tangent bundle is a couple (x, X_x), where x is a point in $\mathbb{R}^n$ and X_x is a tangent vector at x. The projection on the first variable:

$$\begin{aligned} \pi : \quad T\mathbb{R}^n &\longrightarrow \mathbb{R}^n \\ (x, X_x) &\longmapsto x \end{aligned}$$

is surjective, and the pre-image of x through π is the tangent space $T_x\mathbb{R}^n$. This defines a short exact sequence:

$$\begin{tikzcd} 0 \arrow[r] & \mathbb{R}^n \simeq T_x\mathbb{R}^n \arrow[r] & T\mathbb{R}^n \arrow[r, "\pi"] & \mathbb{R}^n \arrow[l, bend left, "\sigma"] \arrow[r] & 0 \end{tikzcd}$$

This sequence *splits*, which means that the map π admits *sections*: continuous maps $\sigma : \mathbb{R}^n \longrightarrow T\mathbb{R}^n$ such that $\pi \circ \sigma = \mathrm{id}_{\mathbb{R}^n}$.

Definition 2.7. *We call* vector fields *over* $\mathbb{R}^n$ *the sections of* π*:*

$$\begin{aligned} X : \quad \mathbb{R}^n &\longrightarrow T\mathbb{R}^n \\ x &\longmapsto (x, X_x) \end{aligned}$$

that are infinitely differentiable (or smooth*) in the second variable (see Scholie 2.8). We denote by* $\mathfrak{X}(\mathbb{R}^n)$ *the* $\mathbb{R}$*-vector space of vector fields on* $\mathbb{R}^n$*.*

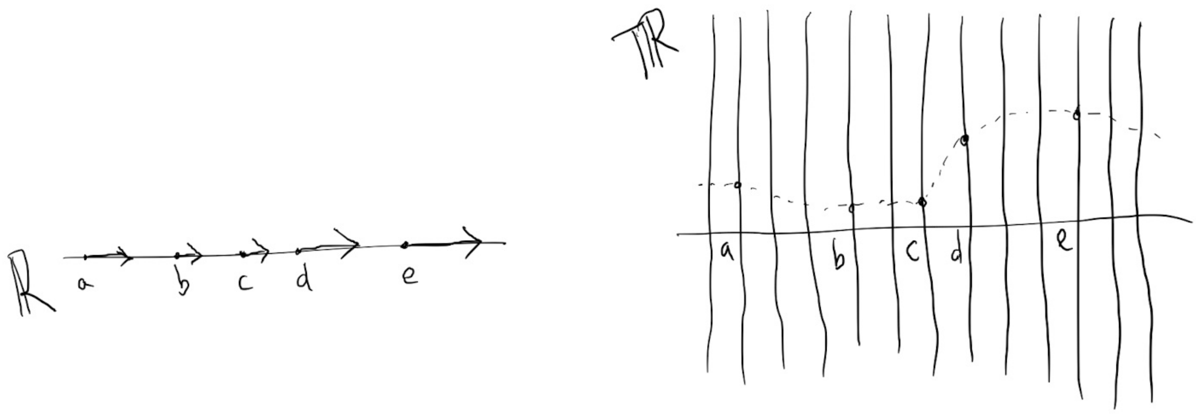

Figure 3: On the left hand side, the ‘geometric’ representation of the tangent vectors to a path in $\mathbb{R}$. On the right hand side, the abstract representation through the tangent bundle of $\mathbb{R}$: over each point x there is a fiber $T_x\mathbb{R} \simeq \mathbb{R}$, and the vector field, tangent to the path at each point, is symbolized by a section (dashed curve) of the vector bundle. The ‘height’ of the section in the fiber over a given point x is equal to the modulus of the tangent vector to the path at x.

By definition, vector fields consist of the assignment to every point x of a tangent vector at x, denoted X_x, which is, additionally, required to vary smoothly over $\mathbb{R}^n$. We will now explain

[7] We will actually see later that it is actually diffeomorphic (the notion of equivalence in the category of smooth manifolds).

what we mean by that. The tangent bundle $T\mathbb{R}^n$ is *trivial*, i.e. it is homeomorphic to the cartesian product $\mathbb{R}^n \times \mathbb{R}^n$. We already know a basis on the base: the vectors $e_1, \ldots, e_n$; let us define a basis on the fiber, denoting it by:

$$\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n} \tag{2.5}$$

This notation is consistent with the notation of the basis vectors of the tangent space $T_x\mathbb{R}^n$ as in Proposition 2.4, because $T_x\mathbb{R}^n \simeq \{x\} \times \mathbb{R}^n$, so that one can makes the straightforward identification:

$$\left.\frac{\partial}{\partial x^i}\right|_x \in T_x\mathbb{R}^n \qquad \longleftrightarrow \qquad \left(x, \frac{\partial}{\partial x^i}\right) \in \{x\} \times \mathbb{R}^n \tag{2.6}$$

Now, given a section X of the tangent bundle, its evaluation at the point x is a tangent vector X_x which can be decomposed on the standard basis of $T_x\mathbb{R}^n$ as in Equation (2.4). Using the one-to-one correspondence (2.6), this gives the following correspondence:

$$X_x \qquad \longleftrightarrow \qquad \left(x, X^i_x \frac{\partial}{\partial x^i}\right) \tag{2.7}$$

Then, for every $1 \leq i \leq n$, this defines an assignment:

$$\begin{aligned} X^i :\ \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto X^i_x \end{aligned}$$

This provides us with the following criterion for smoothness of sections of the tangent bundle:

Scholie 2.8. Smoothness criterion for vector fields *A section $X : \mathbb{R}^n \longrightarrow T\mathbb{R}^n$ being smooth means that the applications:*

$$\begin{aligned} X^i :\ \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto X^i_x \end{aligned}$$

are smooth functions of x (i.e. they are infinitely differentiable).

It turns out that the role of the basis $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ is not only computational, for the sake of the presentation, but has an additional very practical scope. First, the basis elements $\frac{\partial}{\partial x^i}$ can be seen as sections of the tangent bundle, through the following assignment: to every point x, $\frac{\partial}{\partial x^i}$ associates the element $\left.\frac{\partial}{\partial x^i}\right|_x$ in the tangent space $T_x\mathbb{R}^n$. Because the assignment is canonical, we still denote by $\frac{\partial}{\partial x^i}$ these sections, although the reader should remember that, rigorously, they are not the same mathematical objects as the basis vectors (2.5). Second, this set of sections forms a basis of the fiber at each point, by Proposition 2.4.

A set of smooth sections that satisfy these two criteria is called a *frame*. In the present case, the $\frac{\partial}{\partial x^i}$ are in fact *constant* sections, and thus are automatically vector fields on $\mathbb{R}^n$ by Scholie 2.8. They provide a basis for the $\mathcal{C}^\infty$-module of sections, as the following discussion illustrates. Given a vector field X, Scholie 2.8 says that the functions X^i are smooth, so that one can define an additional vector field $X^i \frac{\partial}{\partial x^i}$ on $\mathbb{R}^n$. By equivalence (2.7), we observe that the vector field X and the vector field $X^i \frac{\partial}{\partial x^i}$ coincide at every point x. Thus, one can identity the two vector fields and write:

$$X = X^i \frac{\partial}{\partial x^i} \tag{2.8}$$

It turns out that every vector field can be uniquely decomposed in such a way. This is why we call the functions X^i the *coordinate functions* of the vector field X.

Example 2.9. Examples of vector fields (every coordinate functions are smooth):

$$
\begin{aligned}
X &= y^2 z\frac{\partial}{\partial x} + xe^y\frac{\partial}{\partial y} + 4\frac{\partial}{\partial z} \quad \text{in } \mathbb{R}^3 \\
Y &= 3\,y\sin(t)\frac{\partial}{\partial x} + x^3y^8z^3t^9\frac{\partial}{\partial z} + \arctan(x)\frac{\partial}{\partial t} \quad \text{in } \mathbb{R}^4 \\
Z &= \begin{cases} 0 & \text{when } x \leq 0 \\ e^{-\frac{1}{x}}\frac{\partial}{\partial x} & \text{when } x > 0 \end{cases} \quad \text{in } \mathbb{R} \\
E &= x^i\frac{\partial}{\partial x^i} \quad \text{in } \mathbb{R}^n, \text{ is called the } \textit{Euler vector field}
\end{aligned}
$$

Examples of objects which are *not* vector fields:

$$
1.\quad e^{-\frac{1}{x}}\frac{\partial}{\partial x}, \qquad 2.\quad |x^i|\frac{\partial}{\partial x^i}, \qquad 3.\quad \frac{x}{y-1}\frac{\partial}{\partial z}, \qquad 4. t^{\frac{1}{3}}\frac{\partial}{\partial t}
$$

The first object differs from the vector field Z on the negative semi-axis, and this actually makes a huge difference: although the function $x \longmapsto e^{-\frac{1}{x}}$ is smooth on the right of 0 (its limit is zero), it explodes in the left of 0. This function is not smooth at 0, let alone continuous: that is why the object defined in item 1. is not a vector field. In contrast, to avoid this problem, we have imposed on the object Z to vanish for negative values of x, so that it becomes a well-defined vector field. The object in item 2. is not smooth at zero because the absolute value function is not a smooth function (although it is continuous). The third object is not a vector field because if $x \neq 0$ then the function $y \longmapsto \frac{x}{y-1}$ explodes at 1. The fourth item is not a vector field because it is not differentiable in 0.

Example 2.10. In quantum mechanics (where space time is $\mathbb{R}^4$, say), one can write the wave function ψ in polar form: $\psi = \sqrt{\rho}e^{iS}$, where ρ is a positive smooth function over $\mathbb{R}^4$ and where S is a real-valued smooth function (over $\mathbb{R}^4$). The probability density is $\rho = \psi^\dagger\psi$ and the probability current is denoted $\mathbf{j} = \frac{\rho}{m}\nabla S$, where ∇ (nabla) symbolizes the gradient (with respect to spatial coordinates). In coordinate notations this reads: $\mathbf{j} = \sum_{i=1}^3 \frac{\rho}{m}\frac{\partial S}{\partial x^i}\frac{\partial}{\partial x^i}$; the sum is made over spatial dimensions only. Since the function S is defined all over space-time and is supposedly smooth, for each fixed time t, the probability current is a vector field (on the space $\mathbb{R}^3$). Representing the time with the fourth coordinate x^4, the 4-vector $\rho\frac{\partial}{\partial x^4} + \mathbf{j}$ defines a vector field over space-time $\mathbb{R}^4$. Using the Schrodinger equation, one can show that the divergence of this 4-vector vanishes, which can be interpreted as a *continuity equation for the probability current.*

Example 2.11. Pick up a solution of the heat equation $\frac{\partial u}{\partial t} = \Delta u$ in a homogeneous material of thermal conductivity k (it is a real number then). Assume that this solution is smooth in the four variables (x, y, z, t), or equivalently (x^1, x^2, x^3, x^4). Then, for each time t, the heat flow defined as $\mathbf{q} = -k\nabla u = -\sum_{i=1}^3 k\frac{\partial u}{\partial x^i}\frac{\partial}{\partial x^i}$ is a vector field on $\mathbb{R}^3$ (or at least the part of $\mathbb{R}^3$ where the material is). It indicates at every point of space in which direction the heat flows.

It is important to notice at this point that, when x varies over M, the direction and the norm of the tangent vector X_x varies (it can even vanish at some point !). Hence one sees that a vector field has no "direction" per se, but it is assigned one direction at each point of $\mathbb{R}^n$. Tangent vectors at x were directional derivatives at x; what would be the similar perspective for vector fields?

For every function $f \in \mathcal{C}^\infty(\mathbb{R}^n)$, the assignment $x \longmapsto X_x(f)$ defines a function from $\mathbb{R}^n$ to $\mathbb{R}$. We call this function $X(f)$ and it satisfies, at every point:

$$
X(f)(x) = X_x(f) = X^i_x\frac{\partial f}{\partial x^i}(x)
$$

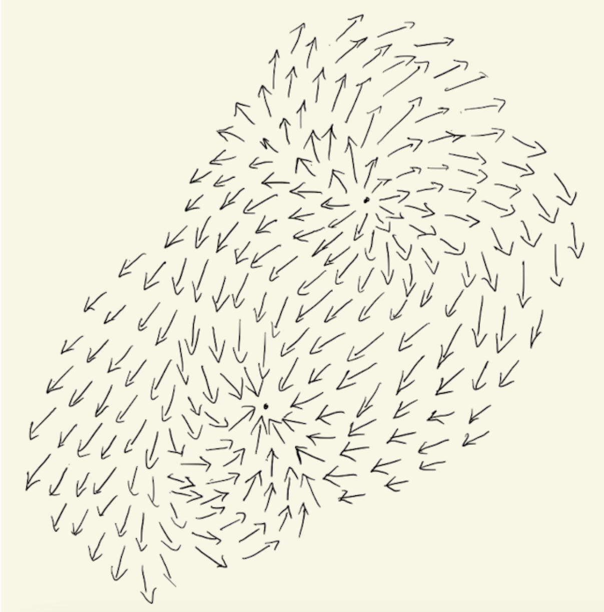

Figure 4: Example of a vector field with two points where it vanishes: one from which the vector field 'flows out', and one where it 'flows in'.

In particular, if f is a constant function, $X(f) = 0$. Because X is smooth, the coordinate functions $X^i : x \longmapsto X^i_x$ are smooth functions of x, as are the derivatives $\frac{\partial f}{\partial x^i}$. Then, the function $X(f)$ is a smooth function. Then the vector field X can be seen as an endomorphism of the (infinite dimensional) vector space $\mathcal{C}^\infty(\mathbb{R}^n)$, also denoted X:

$$\begin{aligned} X: \ \mathcal{C}^\infty(\mathbb{R}^n) &\longrightarrow \mathcal{C}^\infty(\mathbb{R}^n) \\ f &\longmapsto X(f) = X^i \frac{\partial f}{\partial x^i} \end{aligned}$$

This is consistent with the remark that the vector field X can be written as $X^i \frac{\partial}{\partial x^i}$ as explained in Equation (2.8). From this discussion, one sees that the vector field X can be seen as a directional derivative in the direction of X or, said differently, along the *integral curves of* X, i.e. those paths γ in $\mathbb{R}^n$ such that X is always tangent to γ: at each time t, $X_{\gamma(t)} = \dot{\gamma}(t)$. A vector field being a family of tangent vectors indexed over the points of $\mathbb{R}^n$, they inherit the derivation property of tangent vectors Equation (2.3).

A vector field X induces a *derivation* of the algebra of smooth functions $\mathcal{C}^\infty(\mathbb{R}^n)$, i.e. an endomorphism of the (infinite dimensional) vector space $\mathcal{C}^\infty(\mathbb{R}^n)$ that satisfies the following identity:

$$\text{for every } f, g \in \mathcal{C}^\infty(\mathbb{R}^n) \qquad X(fg) = X(f)\cdot g + f \cdot X(g) \tag{2.9}$$

where $\cdot$ symbolizes the multiplication of function in $\mathcal{C}^\infty(\mathbb{R}^n)$. While Equation (2.3) was valid pointwise (because we were working with tangent vectors, defined at a point), Equation (2.9) is valid independently of the point. We denote by $\mathrm{Der}(\mathcal{C}^\infty(\mathbb{R}^n))$ the space of all derivations of $\mathcal{C}^\infty(\mathbb{R}^n)$. Conversely, one can show that any derivation of $\mathcal{C}^\infty(\mathbb{R}^n)$ is a vector field, in the sense of Definition 2.7:

Proposition 2.12. *Vector fields on $\mathbb{R}^n$ are in one-to-one correspondence with derivations of $\mathcal{C}^\infty(\mathbb{R}^n)$:*

$$\mathfrak{X}(\mathbb{R}^n) \simeq \mathrm{Der}(\mathcal{C}^\infty(\mathbb{R}^n))$$

Proof. We have shown that every vector field is a derivation, and we just need to show that a derivation is a vector field. Let $\mathcal{X} \in \mathrm{Der}(\mathcal{C}^\infty(\mathbb{R}^n))$ and define a section of the tangent bundle $X : \mathbb{R}^n \longrightarrow T\mathbb{R}^n$ by:

$$X_x(f) = X(f)(x)$$

This equation makes sense because $\mathcal{X}(f)$ is a smooth function, and the right hand side is its evaluation in x. The object X_x is then a derivation at x, i.e. an element of $T_x\mathbb{R}^n$. One needs only to prove that the assignment $x \longmapsto X_x$ is smooth. This is shown for example in Proposition 4.7 in [Lee, 2003] and in Proposition 8.15 in the 2012 edition. □

This proposition is important because it shows that there is a correspondence between the geometric perspective (vector fields on $\mathbb{R}^n$) and the algebraic perspective (derivations of $\mathcal{C}^\infty(\mathbb{R}^n)$). We have said that passing from one point of view to the other allows to make sense or make things easier. Let us illustrate this strategy by showing that the algebraic perspective is adapted to define a Lie bracket on the space of vector fields $\mathfrak{X}(\mathbb{R}^n)$:

Definition 2.13. *A* Lie algebra *is a (real, possibly infinite dimensional) vector space* $\mathfrak{g}$*, equipped with a bilinear operation* $[\,.\,,.\,]$ *called the* Lie bracket*, which satisfies the following identities:*

$$\begin{array}{ll} \textit{skew-symmetry} & [x,y] = -[y,x] \\ \textit{Jacobi identity} & [x,[y,z]] = [[x,y],z] + [y,[x,z]] \end{array}$$

for every $x, y, z \in \mathfrak{g}$

Remark 2.14. The Jacobi identity is often written under the following, equivalent but more symmetric, form:

$$[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$$

The form presented in Definition 2.13 is useful because it makes clear that "the Lie bracket is a derivation of itself". Here, by *derivation of* $\mathfrak{g}$ we mean any endomorphism $\delta : \mathfrak{g} \longrightarrow \mathfrak{g}$ such that:

$$\delta([x,y]) = [\delta(x), y] + [x, \delta(y)]$$

Then, notice that to every element x of a Lie algebra $(\mathfrak{g}, [\,.\,,.\,])$, we can associate a derivation of $\mathfrak{g}$ via the adjoint action of x on $\mathfrak{g}$:

$$\begin{aligned} \mathrm{ad} : \ \mathfrak{g} &\longrightarrow \mathrm{Der}(\mathfrak{g}) \\ x &\longmapsto \mathrm{ad}_x : y \longmapsto [x,y] \end{aligned}$$

The Jacobi identity ensures that ad_x is a derivation of $\mathfrak{g}$, for every x. The image $\mathrm{ad}(\mathfrak{g}) \subset \mathrm{Der}(\mathfrak{g})$ forms what is called the space of *inner derivation of* $\mathfrak{g}$, sometimes denoted $\mathfrak{inn}(\mathfrak{g})$.

Example 2.15. One can always equip an associative algebra $(A, \cdot)$ with a Lie algebra structure, by setting:

$$[a,b] = a\cdot b - b\cdot a$$

for every $a, b \in A$. In particular, the space of $n \times n$ matrices $\mathcal{M}_n(\mathbb{R})$ (equivalently, the space $\mathrm{End}(E)$ of a n-dimensional vector space) is an associative algebra, so that we can define a Lie bracket on it.

Vector fields are derivations of the (infinite-dimensional) space $\mathcal{C}^\infty(\mathbb{R}^n)$. Derivations are special cases of endomorphisms. However, the composition of two vector fields is *not* a derivation of $\mathcal{C}^\infty(\mathbb{R}^n)$, as the following computation shows:

$$X\big(Y(fg)\big) = X\big(Y(f)\,g + f\,Y(g)\big) = X\big(Y(f)\big)\,g + f\,X\big(Y(g)\big) + \Big(X(f)Y(g) + X(g)Y(f)\Big)$$

The latter parenthesis prevents the composite $X \circ Y$ to be a derivation of $\mathcal{C}^\infty(\mathbb{R}^n)$. Hence, the space of derivations of $\mathcal{C}^\infty(\mathbb{R}^n)$ is not stable under composition. However, inspired by Example 2.15, let us define the following operation on the space of vector fields:

$$[X, Y](f) = X(Y(f)) - Y(X(f)) \tag{2.10}$$

The right-hand side is a smooth function so the left-hand side is a smooth function as well, but one needs to show that the bracket $[X, Y]$ is still a derivation of $\mathcal{C}^\infty(\mathbb{R}^n)$, that is to say: the space $\mathfrak{X}(\mathbb{R}^n) \simeq \mathrm{Der}\big(\mathcal{C}^\infty(\mathbb{R}^n)\big)$ is stable under the action of this bracket. The proof of the following proposition is left as an exercise:

Proposition 2.16. *The $\mathbb{R}$-vector space $\mathfrak{X}(\mathbb{R}^n)$ equipped with this operation is a Lie algebra.*

Exercise 2.17. Prove that Equation (2.10) defines a Lie bracket, i.e. that it is bilinear (with respect to real numbers), skew-symmetric, and that it satisfies the Jacobi identity. By expanding the Lie bracket, prove that the Lie bracket of two vector fields is still a vector field (i.e. a derivation of $\mathcal{C}^\infty(\mathbb{R}^n)$).

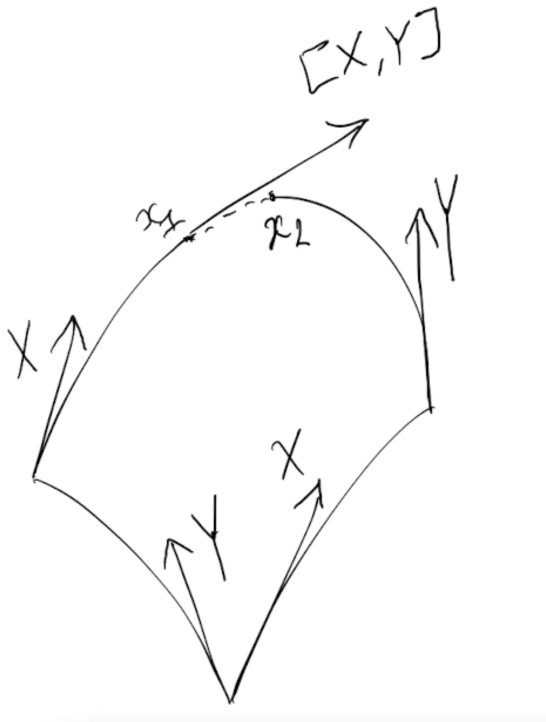

Figure 5: Metaphorical picture of the bracket of vector fields. If one follows for small times first the integral curve of Y, then the integral curve of X, one arrives at the point x_1. Whereas, if one had followed the integral curve of X first, and then that of Y, one arrives at the point x_2. The Lie bracket $[X, Y]$ is the vector field whose integral curve links x_1 to x_2. This discussion can be made rigorous if one makes the time of walking along integral curves tend to 0. See for example page 47 of [Baez and Muniain, 1994], where however the last equation is wrong: the sign should be the opposite, as consequently should be the vector field $[X, Y]$ on figure 10.

A subtle remark has to be made here. We have seen that every vector field can be decomposed on the basis of vectors $\frac{\partial}{\partial x^i}$. However, this basis does *not* form a basis of the $\mathbb{R}$-vector space $\mathfrak{X}(\mathbb{R}^n)$, which is actually infinite dimensional as a real vector space. More precisely:

Scholie 2.18. Algebraic characterization of $\mathfrak{X}(\mathbb{R}^n)$ *The space $\mathfrak{X}(\mathbb{R}^n)$ is, at the same time:*

1. *an infinite dimensional $\mathbb{R}$-vector space;*
2. *a $\mathcal{C}^\infty(\mathbb{R}^n)$-module of finite rank.*

The notion of module over a ring is the generalization of the notion of a vector space over a field. Given a ring $(R, \circ)$, we say that a vector space E is a R-module[8] if there is an action $\cdot$ of R on E which satisfies the following axioms:

$$r \cdot (x+y) = r \cdot x + r \cdot y \qquad (r+s) \cdot x = r \cdot y + r \cdot y$$
$$r \cdot (s \cdot x) = (r \circ s) \cdot x \qquad 1_R \cdot x = x$$

where $r, s \in R$, $x, y \in E$, and where 1_R is the identity of the ring. The reader can check that these axioms are the same axioms that the scalars have to satisfy when acting on a vector space. In our case, the field is $\mathbb{R}$ and the ring is $\mathcal{C}^\infty(\mathbb{R}^n)$. Then, when we say that $\mathfrak{X}(\mathbb{R}^n)$ is a $\mathbb{R}$-vector space we understand that vector fields can be added, and multiplication by real scalars is well-defined. When we say that $\mathfrak{X}(\mathbb{R}^n)$ is a $\mathcal{C}^\infty(\mathbb{R}^n)$-module, we mean that vector fields can be added, and that multiplication by *smooth functions* is well-defined. Notice that, since constant functions can be identified with real scalars, the fact that $\mathfrak{X}(\mathbb{R}^n)$ is a $\mathbb{R}$-vector space is a consequence of the fact that it is a $\mathcal{C}^\infty(\mathbb{R}^n)$-module.

Now, the dimension of a vector space is the minimal number of independent vectors that generate the space (using only multiplication by real scalars and addition). The rank of a module is the maximum number of elements which are linearly independent under the action of the ring. In our case, every vector field X decomposes on the elements $\frac{\partial}{\partial x^i}$ as $X = X^i \frac{\partial}{\partial x^i}$, where the X^i are *smooth functions* (we see the module structure emerging). Moreover, those constant sections are linearly independent over $\mathcal{C}^\infty(\mathbb{R}^n)$ because by definition the identity $X^i \frac{\partial}{\partial x^i} = 0$ implies $X^i = 0$. Thus, $\mathfrak{X}(\mathbb{R}^n)$ is a $\mathcal{C}^\infty(\mathbb{R}^n)$-module of rank n. What is crucial in the present discussion is that the generators of a module need not coincide with a basis of the underlying vector space, because *the multiplication with a ring element generate much different elements than the multiplication with a scalar.* Indeed, one can explicitly compute how the ring of smooth functions $\mathcal{C}^\infty(\mathbb{R}^n)$ acts on a vector field X via multiplication: let $f \in \mathcal{C}^\infty(\mathbb{R}^n)$, then we define the vector field fX to be the unique vector field whose coordinate functions are fX^i, where here we understand the product of two functions. In other words:

$$fX = (f \cdot X^i) \frac{\partial}{\partial x^i}$$

so that $(fX)^i = f \cdot X^i$, where $\cdot$ symbolizes the multiplication of function in $\mathcal{C}^\infty(\mathbb{R}^n)$. Pointwise, this vector field satisfies $(fX)_x = f(x) X_x$. We see how the structure of $\mathcal{C}^\infty(\mathbb{R}^n)$-module only needs n-generators to be defined.

However, $\mathfrak{X}(\mathbb{R}^n)$ is an infinite dimensional vector space. This can be shown by contradiction. Assume there exists a finite number of vector fields $X_1, \dots, X_r$ which form a basis of $\mathfrak{X}(\mathbb{R}^n)$ (as a real vector space), that is: every vector field X would be uniquely written as $X = \sum_{s=1}^r \lambda_s X_s$, where the λ_s are real numbers. Then, given a smooth function f, there exists real scalars $\mu_1, \dots, \mu_r$ such that $fX : \sum_{s=1}^r \mu_s X_s$. On the other hand, multiplying $\sum_{s=1}^r \lambda_s X_s$ by f gives $fX = \sum_{s=1}^r f \lambda_s X_s$. By unicity of the decomposition, $f\lambda_s = \mu_s$ for every $s = 1, \dots, r$, which is impossible most of the time because f need not be constant. The demonstration may be a bit too much abstract. The idea of the proof is that a finite number of elements cannot form a set

[8] Actually, in full generality we only require E to be an abelian group. A vector space is an abelian group with respect to the addition.

of generators for all the vector fields, because multiplication by any function offers much more freedom and variability that can be encoded by a mere finite dimensional vector space.

The Lie algebra structure on $\mathfrak{X}(\mathbb{R}^n)$ is defined on top of the vector space structure. Thus, Scholie 2.18 explains why $\mathfrak{X}(\mathbb{R}^n)$ is a real Lie algebra of infinite dimension, although only a finite number of constant sections $\frac{\partial}{\partial x^i}$ is needed to generate all the vector fields (using the ring multiplication). This additionally explains why the Lie bracket is bilinear with respect to the scalars, but not with respect to the smooth functions. More precisely, since every vector field can be decomposed on the frame $\frac{\partial}{\partial x^i}$, a small computation shows that the Lie bracket of X and Y reads:

$$[X,Y] = \Big(X(Y^i) - Y(X^i)\Big)\frac{\partial}{\partial x^i} \tag{2.11}$$

where we recall that X^i and Y^i are the i-th coordinate functions associated to X and Y, respectively. The Einstein summation convention has been used. Then, for any smooth function $f \in \mathcal{C}^\infty(\mathbb{R}^n)$:

$$[X, fY] = f\,[X,Y] + X(f)\,Y \tag{2.12}$$

where the term on the right hand side has to be understood as the multiplication of the function $X(f)$ with the vector field Y. Equation (2.12) shows that the Lie bracket defined in Equation (2.10) is not linear with respect to the functions, as expected since it should only be linear with respect to real numbers.

Remark 2.19. We conclude this section by introducing an alternative notation for the constant vector fields $\frac{\partial}{\partial x^i}$, that may also be denoted ∂_i:

$$\partial_i \equiv \frac{\partial}{\partial x^i}$$

The position of the index i is indeed at the bottom because one should formally consider the fractional notation $\frac{\partial}{\partial x^i}$ as a fraction of fractions: $\frac{\frac{a}{b}}{\frac{c}{d}}$, where the index i is at the top of the denominator, at the place occupied by the element c. Since the latter fraction can be written as $\frac{ad}{bc}$, and that the element c is at the bottom, this justifies that we place the index i at the bottom of the notation ∂_i. I emphasize that keeping track of the position of indices is central in differential geometry when we work in coordinates. Moreover the above informal reasoning will have some relevance later in the text. Using Equation (2.11), we deduce that the commutator of two vectors of the frame vanishes:

$$[\partial_i, \partial_j] = 0 \tag{2.13}$$

2.2 Cotangent vectors and differential 1-forms on $\mathbb{R}^n$

Now let us turn to the elements dual to tangent vectors and vector fields. Given some point $x \in \mathbb{R}^n$ we call the *cotangent space* and we write $T_x^*\mathbb{R}^n$ the dual of the tangent space at x:

$$T_x^*\mathbb{R}^n = (T_x\mathbb{R}^n)^*$$

Elements of this dual space are called *cotangent vectors* at x. They are linear forms on $T_x\mathbb{R}^n$ and there is a canonical bijection between $(\mathbb{R}^n)^*$ and $T_x^*\mathbb{R}^n$: since a basis of $T_x\mathbb{R}^n$ is given by the vectors $\partial_i|_x = \frac{\partial}{\partial x^i}\Big|_x$, using Equation (1.11) one obtains a dual basis of $T_x^*\mathbb{R}^n$ whose elements are denoted $dx^i|_x$. In particular one has:

$$dx^i|_x(\partial_j|_x) = \delta^i_j \tag{2.14}$$

Thus, for any tangent vector X_x, one has $dx^i|_x(X_x) = X_x^i$. In particular, one observes that it is as if the tangent vector X_x had been fed with the coordinate function $x^i : \mathbb{R}^n \longmapsto \mathbb{R}$ that associates a point to its i-th coordinate:

$$dx^i|_x(X_x) = X_x(x^i)$$

One can now extend the notion of cotangent vectors to the one of covector fields, following what has been said in subsection 2.1.

Definition 2.20. *We define the* cotangent bundle $T^*\mathbb{R}^n$ *to be the union of all cotangent spaces:*

$$T^*\mathbb{R}^n = \bigsqcup_{x\in\mathbb{R}^n} T_x^*\mathbb{R}^n$$

It has several properties: it is a trivial vector bundle over $\mathbb{R}^n$, i.e. it is diffeomorphic to $\mathbb{R}^n \times (\mathbb{R}^n)^*$. As expected from the definition of cotangent spaces, the *fiber* or $T^*\mathbb{R}^n$ is $(\mathbb{R}^n)^*$, the dual space of the fiber of $T\mathbb{R}^n$. Points in $T^*\mathbb{R}^n$ are couples (x, ξ_x), where ξ_x is a notation for cotangent vectors at x; they decompose on the basis of $T_x^*\mathbb{R}^n$ as $\xi_{x,i}dx^i|_x$, where the $\xi_{x,i}$ are the coordinates of ξ_x (they are real numbers). The projection on the first variable, denoted $\tau : T^*\mathbb{R}^n \longrightarrow \mathbb{R}^n$, admits sections:

Definition 2.21. *We call* covector fields – *or* differential 1-forms – *over* $\mathbb{R}^n$ *the sections of* τ*:*

$$\begin{aligned} \xi:\ \mathbb{R}^n &\longrightarrow T^*\mathbb{R}^n \\ x &\longmapsto (x, \xi_x) \end{aligned}$$

that are infinitely differentiable (or smooth*) in the second variable (see Scholie 2.22). We denote by* $\Omega^1(\mathbb{R}^n)$ *the* $\mathbb{R}$*-vector space of covector fields/differential 1-forms on* $\mathbb{R}^n$*.*

Since the cotangent bundle is trivial (i.e. it is diffeomorphic to a cartesian product), one can define a standard basis on its fiber. The fiber $\mathbb{R}^n$ of the tangent bundle is already equipped with a standard basis: the generators $\partial_i = \frac{\partial}{\partial x^i}$. The dual basis would form a basis of the fiber $(\mathbb{R}^n)^*$ of the cotangent bundle; let us denote this basis by:

$$dx^1, \ldots, dx^n$$

and we call it the *dual coframe* to the given frame. This notation is consistent with the notation of the basis vectors of the cotangent spaces. Indeed, since $T_x^*\mathbb{R}^n \simeq \{x\} \times (\mathbb{R}^n)^*$, we can make the following identification:

$$dx^i|_x \in T_x^*\mathbb{R}^n \qquad \longleftrightarrow \qquad (x, dx^i) \in \{x\} \times (\mathbb{R}^n)^* \tag{2.15}$$

Thus, the basis vectors $dx^1, \ldots, dx^n$ can also be seen as *constant* sections of the cotangent bundle: the differential 1-forms dx^i associates, to every point x, the cotangent vector $dx^i|_x$ via the above correspondence. Every cotangent vector ξ_x defined at the point x can be decomposed on the dual basis defined in Equation (2.14) as $\xi_x = \xi_{x,i}dx^i|_x$. Then, because of the injection of $T_x^*\mathbb{R}^n$ into $T^*\mathbb{R}^n$, the one-to-one equivalence defined in Equation (2.15) defines an equivalence:

$$\xi_x \qquad \longleftrightarrow \qquad (x, \xi_{x,i}dx^i)$$

where Einstein summation convention has been used. Then, for every $1 \leq i \leq n$, this defines an assignment:

$$\begin{aligned} \xi_i:\ \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \xi_{i,x} \end{aligned}$$

A priori the coordinate functions ξ_i of a random section ξ are not smooth, unless the section is smooth, i.e. unless it is a covector field. As for vector fields, this actually provides a first criterion for smoothness of covector fields (see Scholie 2.22).

Furthermore, this enables us to understand how sections of $T^*\mathbb{R}^n$ act on vector fields. Recall that we have the following identity, by definition of the dual basis on the fiber of $T^*\mathbb{R}^n$:

$$dx^i(\partial_j) = \delta^i_j$$

By construction, the constant sections dx^i are $\mathcal{C}^\infty(\mathbb{R}^n)$-linear: $dx^i(X^j\partial_j) = X^j dx^i(\partial_j) = X^i$. Then, given a section ξ of $T^*\mathbb{R}^n$ and a vector field X, one has:

$$\xi(X) = \xi_i X^j dx^i(\partial_j) = \xi_i \cdot X^i \tag{2.16}$$

where the Einstein summation convention has been used, and where $\cdot$ symbolizes the multiplication of function in $\mathcal{C}^\infty(\mathbb{R}^n)$. The term on the right of Equation (2.16) is a product of functions, thus the term on the left is a function as well. Evaluating both terms in a point x gives:

$$\xi(X)(x) = \xi_{x,i} X^i_x = \xi_x(X_x)$$

where the term in the middle is a sum of products of real numbers. A priori the function $\xi(X) : x \longmapsto \xi_x(X_x)$ is not smooth, unless ξ is a smooth section of $T^*\mathbb{R}^n$, i.e. unless it is a covector field. This observation provides the second criterion for smoothness of covector fields:

Scholie 2.22. Smoothness criteria for covector fields *A section $\xi : \mathbb{R}^n \longrightarrow T^*\mathbb{R}^n$ being smooth means:*

1. *that the components functions ξ_i:*

$$\begin{aligned} \xi_i :\ \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \xi_{x,i} \end{aligned}$$

   are smooth functions of x (i.e. they are infinitely differentiable);

2. *or that, equivalently*[9]*, for every vector field X, the function:*

$$\begin{aligned} \xi(X) :\ \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \xi_x(X_x) \end{aligned}$$

   is smooth.

Example 2.23. An example of a covector field in $\mathbb{R}^2$:

$$\xi = \big(2xy\cos(x) - x^2y\sin(x)\big)dx + x^2\cos(x)dy$$

We will see in subsection 2.3 that such a differential 1-form is actually the differential of the function $f(x,y) = x^2y\cos(x)$.

Example 2.24. A physically oriented example consists of the connections A_μ. Actually they correspond to a differential 1-form (taking values in a Lie algebra) $A = A_\mu dx^\mu$.

[9]The equivalence of the two criteria is shown by noticing that: 1. the second one is implied by the first one using Equation (2.16) and smoothness of vector fields, and 2. the first one is implied by the second one if one picks up $X = \partial_i$ for every $1 \leq i \leq n$.

Hence there are at least two way at looking at covector fields (= differential 1-forms): one is to see them as smooth sections of the cotangent bundle, and in that case they are smooth if and only if item 1. of Scholie 2.22 is satisfied. Another way of looking at covector fields is to see them as being linear morphisms on the space of vector fields, landing in the smooth functions, that is to say:

$$\Omega^1(\mathbb{R}^n) \simeq \mathrm{Hom}\big(\mathfrak{X}(\mathbb{R}^n), \mathcal{C}^\infty(\mathbb{R}^n)\big) \tag{2.17}$$

The homomorphisms here have to be understood as homomorphisms of $\mathcal{C}^\infty(\mathbb{R}^n)$-modules. More precisely, in that case, a covector field ξ can be seen as a linear morphism:

$$\begin{aligned} \xi:\ \mathfrak{X}(\mathbb{R}^n) &\longrightarrow \mathcal{C}^\infty(\mathbb{R}^n) \\ X &\longmapsto \xi(X) \end{aligned}$$

The fact that this map lands in the smooth functions for every choice of vector field is precisely the content of item 2. of Scholie 2.22. Although the latter perspective is often the most used, the former one is useful to have a glimpse of the geometrical meaning of differential 1-forms.

What is the meaning of covector fields/differential 1-forms? An explanation can be the following: a differential 1-form $\xi : \mathbb{R}^n \longrightarrow T^*\mathbb{R}^n$ defines, at every point x, a linear form $\xi_x : T_x\mathbb{R}^n \longrightarrow \mathbb{R}$ on the tangent space at x. As a map from a n-dimensional space to a 1-dimensional space, the kernel of this linear form is an hyperplane H_x of $T_x\mathbb{R}^n$, that is: a $n-1$-dimensional subspace. This hyperplane separates the n-dimensional space $T_x\mathbb{R}^n$ in two (n-dimensional) open half-spaces. The linear form additionally defines a 'positive' half-space H_x^+ and a 'negative' half-space H_x^-: the former consists of all tangent vectors X_x such that $\xi_x(X_x) > 0$, while the latter consists of all tangent vectors X_x such that $\xi_x(X_x) < 0$. The hyperplane H_x is a separator between these two half-spaces since $\xi_x|_{H_x} = 0$. Since the covector ξ_x is a linear morphism from $T_x\mathbb{R}^n$ to $\mathbb{R}$, its *level sets* are $(n-1)$-dimensional affine subspaces defined as follows:

$$H_{x,t} = \{X_x \,|\, \xi_x(X_x) = t\}$$

for every $t \in \mathbb{R}$. The notation is consistent with the definition of H_x because $H_{x,0} = H_x$. In particular, the positive half-space and the negative half-space are union of level sets:

$$H_x^+ = \bigcup_{t>0} H_{x,t} \qquad \text{and} \qquad H_x^- = \bigcup_{t<0} H_{x,t}$$

The level sets define a partition of $T_x\mathbb{R}^n$ by parallel affine subspaces. The main point here is that a linear form is entirely described from its level sets. Smoothly varying the linear form ξ_x then has the consequence of smoothly changing its level sets and in particular: their inclination and their respective distance. Thus a differential 1-form can be seen as a smooth assignment, to every point x, of a partition of $T_x\mathbb{R}^n$ by parallel affine subspace. Smoothness of this assignment means that the partition (of the fiber $\mathbb{R}^n$) smoothly varies when the base point varies.

These hyperplanes have a geometric significance: let $\xi^\sharp$ be the vector field corresponding to ξ through the musical isomorphism $\sharp$ (where we assume the metric to be the euclidean metric on $\mathbb{R}^n$). Then the hyperplane $H_x = \mathrm{Ker}(\xi_x)$ defines the tangent space to the transversal to the integral curve of $\xi^\sharp$. In other words, the tangent vector $\xi^\sharp_x$ is orthogonal to H_x. The distance between the hyperplanes is an alternative – though equivalent – measure of the length of $\xi^\sharp$: the hyperplanes are closer to one another at the points where $\xi^\sharp$ has a small modulus, and they are more distant to one another at the points where $\xi^\sharp$ has a bigger modulus.

Before moving to the next section, we would go for a quick excursion through the realm of *vector bundles* (over $\mathbb{R}^n$). The idea of a vector bundle is the following: given a k-dimensional vector space, $\mathbb{R}^k$ say, we attach a copy of such a vector space at each and every point of the

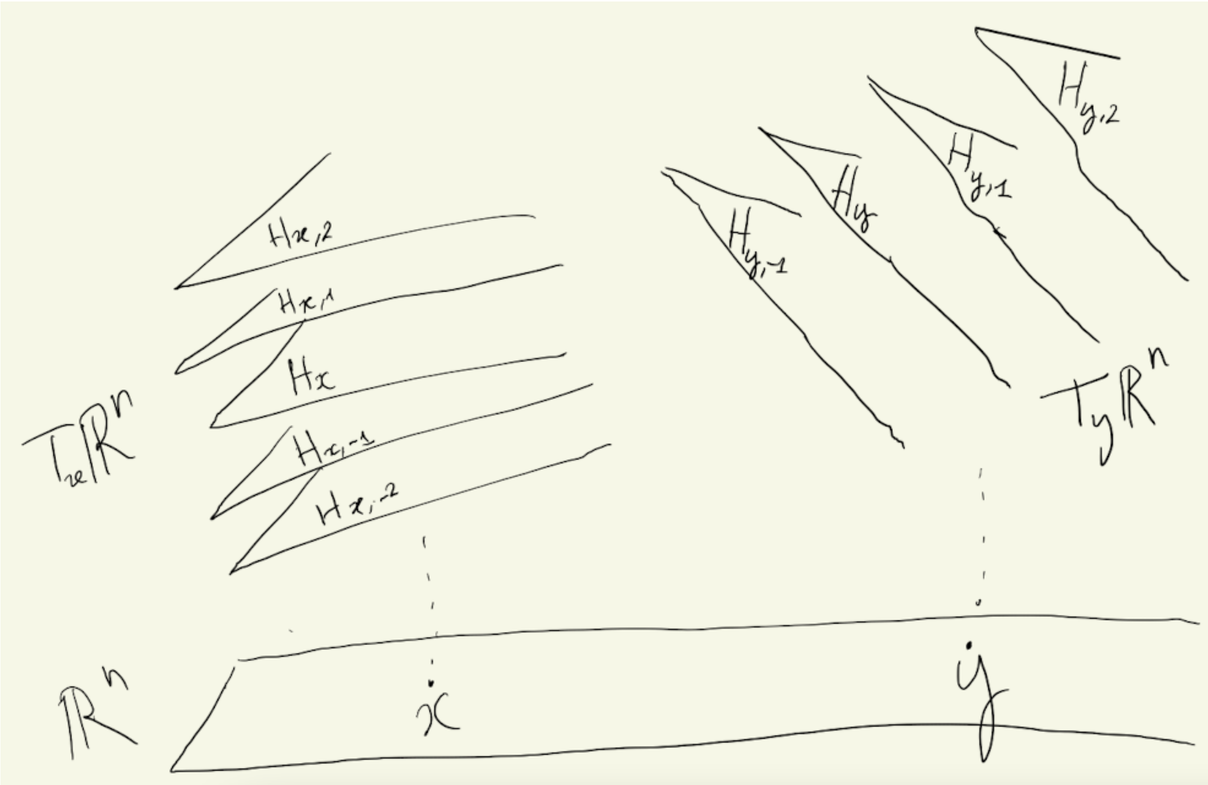

Figure 6: Visual representation of the geometrical meaning of differential forms. When the base point varies smoothly, the partition of the tangent space by $(n-1)$-dimensional affine subspaces varies smoothly: the inclination and the relative distance of the affine hyperspaces is smoothly modified. The hyperspace H_x (resp. H_y) is orthogonal to the tangent vector $\xi^\sharp_x$ (resp. $\xi^\sharp_y$), and can be seen as the tangent space to the transversal to the integral curve of $\xi^\sharp$ at x (resp. y).

space $\mathbb{R}^n$. This form an enormous space denoted E for example, that we require to be sufficiently well-defined (to be clear: it should be a topological space, i.e. a space along with a topology of open sets). The topology on E is chosen so that at least locally, in the neighborhood of any point, say U, E looks like $U \times \mathbb{R}^k$. Trivial vector bundles are precisely those that have this structure globally, i.e. those of the form $\mathbb{R}^n \times \mathbb{R}^k$. It turns out that every vector bundle defined over $\mathbb{R}^n$ has this property. The precise statement is the following:

Definition 2.25. *A* (trivial) vector bundle of rank k *(over $\mathbb{R}^n$) is a topological space E together with a surjective continuous map $\pi : E \longrightarrow \mathbb{R}^n$, satisfying the two following conditions:*

1. *for every $x \in \mathbb{R}^n$, the preimage $\pi^{-1}(x) \subset E$ is a k-dimensional vector space, called the* fiber of E at x *and denoted E_x;*
2. *there exists a homeomorphism $\Phi : E \longrightarrow \mathbb{R}^n \times \mathbb{R}^k$ (called a* trivialization of E*), making the following triangle commutative:*

Figure 7: Wherever it makes sense, the kernel of the differential $d\xi$ defines the tangent space to the transversal to the vector field $\xi^\sharp$.

$$\begin{tikzcd}
E \arrow[rr, "\Phi"] \arrow[dr, "\pi"'] & & \mathbb{R}^n \times \mathbb{R}^k \arrow[dl, "pr_1"] \\
& \mathbb{R}^n &
\end{tikzcd}$$

where $pr_1 : \mathbb{R}^n \times \mathbb{R}^k \longrightarrow \mathbb{R}^n$ is the projection on the first variable, and such that for every $y \in \mathbb{R}^n$, the restriction of Φ to E_y is a linear isomorphism from E_y to $\{y\} \times \mathbb{R}^k \simeq \mathbb{R}^k$.

Remark 2.26. The second item means that for every $u \in E$, one has the following identity:

$$\pi(u) = pr_1 \circ \Phi(u)$$

Notice that, in full generality (i.e. on a smooth manifold), the second item should hold only locally (see e.g. Chapter 5 of [Lee, 2003], Chapter 10 in the 2012 edition). The fact that $\mathbb{R}^n$ is contractible implies that every vector bundle is trivial, and that we wrote this second item from the global perspective.

One should really think of a vector bundle as a bunch of vector spaces stacked together and labeled by points. The set underlying any vector bundle is the disjoint union of its fiber:

$$E = \bigsqcup_{x\in\mathbb{R}^n} E_x$$

There is a natural topology on a disjoint union, that we call the 'disjoint union topology': it is the finest topology that makes the injective functions $\phi_x : E_x \hookrightarrow E$ continuous. More precisely, with respect to this topology, $U \subset E$ is open if and only if $\phi_x^{-1}(U)$ is open in E_x for every $x \in \mathbb{R}^n$. Assuming that every E_x is homeomorphic to $\mathbb{R}^k$ with its standard topology, the disjoint union $E = \bigsqcup_{x\in\mathbb{R}^n} E_x$ equipped with its disjoint union topology is then homeomorphic to the product of topological space $\mathbb{R}^n \times \mathbb{R}^k$, where $\mathbb{R}^n$ has the *discrete topology* and the product has the product topology. Hence, the disjoint union underlying every vector bundle over $\mathbb{R}^n$ is homeomorphic to $\mathbb{R}^n \times \mathbb{R}^k$, with respect to topologies that we do not like though (because we are not interested in working on $\mathbb{R}^n$ with the discrete topology). What additional property does a vector bundle have then, that the mere underlying disjoint union does not have? The answer is that it is equipped with a 'vector bundle topology' – certainly coarser than the disjoint union topology – such that there is an homeomorphism between $E = \bigsqcup_{x\in\mathbb{R}^n} E_x$ equipped with its vector bundle topology and $\mathbb{R}^n \times \mathbb{R}^k$, but here $\mathbb{R}^n$ has its standard topology (which is not discrete!). This is why a vector bundle is much more than its underlying set, the disjoint union of all its fibers.

A *section* of a vector bundle E (over $\mathbb{R}^n$) is a continuous map $\sigma : \mathbb{R}^n \longrightarrow E$ satisfying the following identity:

$$\pi \circ \sigma = \mathrm{id}_{\mathbb{R}^n}$$

In other words, $\sigma(x) \in E_x$ for every x. A section can be symbolically represented as a n-dimensional surface in E, that is projectable onto $\mathbb{R}^n$. Given a section σ, if the map $\Phi \circ \sigma$ is smooth we call σ a *smooth function*. The space of smooth sections of E then consists of smooth functions from $\mathbb{R}^n$ to E and is denoted $\Gamma(E)$ (sometimes, also denoted $\Gamma(\mathbb{R}^n, E)$ or $\mathcal{C}^\infty(\mathbb{R}^n, E)$). As was explained in Scholie 2.18, these spaces are real vector spaces of infinite dimension, and $\mathcal{C}^\infty(\mathbb{R}^n)$-modules of finite rank, k to be precise, as is shown by the following paragraph.

Assume that we have k smooth sections $\sigma_1, \ldots, \sigma_k$ that are fiberwise linearly independent, i.e. for every x, the vectors $\sigma_1(x), \ldots, \sigma_k(x)$ form a basis of E_x. Then, we call such a family a *frame* for E. Since every vector bundle over $\mathbb{R}^n$ is trivial, one can pickup constant orthonormal frames, i.e. for every $1 \leq i \leq n$, the smooth map $\Phi \circ \sigma_i : \mathbb{R}^n \longmapsto \mathbb{R}^n \times \mathbb{R}^k$ is constant, and thus defines a vector $f_i \in \mathbb{R}^k$, so that $f_1, \ldots, f_k$ forms a basis of $\mathbb{R}^k$. Under an intelligent choice of sections, this basis can be made orthonormal. A frame forms a set of generator of the sections of E, with respect to the $\mathcal{C}^\infty(\mathbb{R}^n)$-module structure on $\Gamma(E)$. That is why the rank of this module coincides with the number of vector in the frame, which is the same as the dimension of the fiber: k.

Famous examples of vector bundles are the tangent bundle $T\mathbb{R}^n$ (with fiber $T_x\mathbb{R}^n \simeq \mathbb{R}^n$) and the cotangent bundle $T^*\mathbb{R}^n$ (with fiber $T_x^*\mathbb{R}^n \simeq (\mathbb{R}^n)^*$). Smooth sections of $T\mathbb{R}^n$ are vector fields and smooth sections of $T^*\mathbb{R}^n$ are differential 1-forms:

$$\mathfrak{X}(\mathbb{R}^n) = \Gamma(T\mathbb{R}^n) \qquad \text{and} \qquad \Omega^1(\mathbb{R}^n) = \Gamma(T^*\mathbb{R}^n)$$

A frame for $T\mathbb{R}^n$ is the family of constant vector fields $\partial_1, \ldots, \partial_n$, whereas a frame for $T^*\mathbb{R}^n$ (what we had called a *coframe*) is made of the constant covector fields $dx^1, \ldots, dx^n$. Moreover, drawing on the material presented in subsection 1.1, one can construct the following other vector bundles:

$$\wedge^m T\mathbb{R}_n = \bigsqcup_{x\in\mathbb{R}^n} \wedge^m T_x\mathbb{R}^n \qquad \text{and} \qquad \wedge^m T^*\mathbb{R}_n = \bigsqcup_{x\in\mathbb{R}^n} \wedge^m T_x^*\mathbb{R}^n$$

The notation is transparent: the fiber at a given point x is the m-th exterior power of $T_x\mathbb{R}^n$ (or $T_x^*\mathbb{R}^n$, respectively). These are trivial vector bundles (as is every vector bundle over $\mathbb{R}^n$). Smooth sections of $\wedge^m T\mathbb{R}_n$ are called *m-multivector fields*, whereas smooth sections of $\wedge^m T^*\mathbb{R}_n$ are called *differential m-forms*, and are denoted as follows:

$$\mathfrak{X}^m(\mathbb{R}^n) = \Gamma(\wedge^m T\mathbb{R}^n) \qquad \text{and} \qquad \Omega^m(\mathbb{R}^n) = \Gamma(\wedge^m T^*\mathbb{R}^n)$$

In particular $\mathfrak{X}^0(\mathbb{R}^n) = \Omega^0(\mathbb{R}^n) = \mathcal{C}^\infty(\mathbb{R}^n) \simeq \mathfrak{X}^n(\mathbb{R}^n) \simeq \Omega^n(\mathbb{R}^n)$, and $\mathfrak{X}^1(\mathbb{R}^n) = \mathfrak{X}(\mathbb{R}^n)$. When $m \geq 1$, a frame for $\bigwedge^m T\mathbb{R}^n$ consists of the constant sections $\partial_{i_1} \wedge \ldots \wedge \partial_{i_m}$, whereas a frame for $\bigwedge^m T^*\mathbb{R}^n$ is given by constant sections of the form $dx^{i_1} \wedge \ldots \wedge dx^{i_m}$, for $1 \leq i_1 < \ldots < i_m \leq n$.

Let us now find criteria for smoothness of sections of $\bigwedge^\bullet T^*\mathbb{R}^n$. A (not necessarily smooth but at least continuous) section η of $\bigwedge^m T^*\mathbb{R}^n$ decomposes on this basis as:

$$\eta = \sum_{1 \leq i_1 < \ldots < i_m \leq n} \overline{\eta}_{i_1 \ldots i_m} \, dx^{i_1} \wedge \ldots \wedge dx^{i_m} \tag{2.18}$$

We denote the coordinate functions in the basis $dx^{i_1} \wedge \ldots \wedge dx^{i_m}$ where we assume that $1 \leq i_1 < \ldots < i_m \leq n$ as $\overline{\eta}_{i_1 \ldots i_m}$. However, usually the Einstein summation convention (in which the indices i_k vary from 1 to n and are not ordered) is much more practical. To use it, one needs to do a bit of gymnastics. First define the following functions:

$$\text{for every } 1 \leq i_1 < \ldots < i_m \leq n \qquad \eta_{i_1 \ldots i_m} = \frac{1}{m!} \overline{\eta}_{i_1 \ldots i_m}$$

Then, for every choice of non-ordered indices $i_1, \ldots, i_m \in \{1, ..., n\}$, there is a unique permutation $\sigma \in S_m$ such that $i_{\sigma(1)} < i_{\sigma(2)} < \ldots < i_{\sigma(m)}$. In other words, the permutation σ rearrange the indices so that they come in order. For such a permutation, we define the function $\eta_{i_1 \ldots i_m}$ as follows:

$$\eta_{i_1 \ldots i_m} = (-1)^\sigma \eta_{i_{\sigma(1)} \ldots i_{\sigma(m)}} = \frac{(-1)^\sigma}{m!} \, \overline{\eta}_{i_{\sigma(1)} \ldots i_{\sigma(m)}}$$

Then one can write under Einstein summation convention:

$$\eta = \eta_{i_1 \ldots i_m} \, dx^{i_1} \wedge \ldots \wedge dx^{i_m} \tag{2.19}$$

Exercise 2.27. By using the antisymmetry of the wedge product, prove that Equation (2.19) gives back Equation (2.18).

The section η is at least continuous so the functions $\eta_{i_1 \ldots i_m}$ are continuous functions on $\mathbb{R}^n$ and, as for covector fields (see Scholie 2.22), they are smooth if and only if η is a smooth section, i.e. if and only if $\eta \in \Omega^m(\mathbb{R}^n)$ (η is a differential m-form). Another criterion for smoothness of η is obtained by using Equation (1.17); when fed with m vector fields, η gives the following continuous function:

$$\begin{aligned} \eta(X_1, \ldots, X_m) &= \eta_{i_1 \ldots i_m} \, dx^{i_1} \wedge \ldots \wedge dx^{i_m}(X_1, \ldots, X_m) \\ &= \eta_{i_1 \ldots i_m} \det \begin{pmatrix} X_1^{i_1} & X_2^{i_1} & \ldots & X_{m-1}^{i_1} & X_m^{i_1} \\ X_1^{i_2} & & & & X_m^{i_2} \\ \ldots & & \ldots & & \ldots \\ X_1^{i_{m-1}} & & & & X_m^{i_{m-1}} \\ X_1^{i_m} & X_2^{i_m} & \ldots & X_{m-1}^{i_m} & X_m^{i_m} \end{pmatrix} \end{aligned} \tag{2.20}$$

Since the X_i are vector fields, Scholie 2.8 implies that their coordinate function are infinitely differentiable, which implies that the above determinant, as a product of smooth functions of x, is a smooth function over $\mathbb{R}^n$. Then, it implies that $\eta(X_1, \ldots, X_m)$ is a smooth function if and only if the coordinate functions $\eta_{i_1 \ldots i_m}$, i.e. if and only if η is a differential m-form. The situation can be summarized as follows:

Scholie 2.28. Smoothness criteria for differential m-forms *A section $\eta : \mathbb{R}^n \longrightarrow \wedge^m T^*\mathbb{R}^n$ being smooth means:*

1. *that the coordinate functions $\eta_{i_1 \dots i_m}$ are smooth functions of x;*
2. *or that, equivalently, for every vector fields $X_1, \dots, X_m$, the continuous function $\eta(X_1, \dots, X_m)$ defined in Equation (2.20) is smooth.*

Exercise 2.29. For any given choice of m indices $j_1, \dots, j_m \in \{1, ..., n\}$, show that Equation (2.20) applied to $\partial_{j_1}, \dots, \partial_{j_m}$ gives:

$$\eta(\partial_{j_1}, \dots, \partial_{j_m}) = m!\, \eta_{j_1 \dots j_m} = (-1)^\sigma \overline{\eta}_{j_\sigma(1) \dots j_\sigma(m)}$$

where σ is the unique permutation of m elements such that $j_{\sigma(1)} < j_{\sigma(2)} < \dots < j_{\sigma(m)}$.

The properties of the wedge product on the exterior algebra $\bigwedge^\bullet T^*\mathbb{R}^n$ are transported to the differential forms. So in particular $dx^i \wedge dx^j = -dx^j \wedge dx^i$, and for any $\eta \in \Omega^k(\mathbb{R}^n)$ and $\mu \in \Omega^l(\mathbb{R}^n)$, the object $\eta \wedge \mu$ is a differential $k + l$-form, and:

$$\eta \wedge \mu = (-1)^{kl} \mu \wedge \eta \tag{2.21}$$

This turns the graded vector space $\Omega^\bullet(\mathbb{R}^n)$ into a *(graded) commutative graded algebra.*

2.3 Differential forms on $\mathbb{R}^n$ and the de Rham complex

Scholie 2.22 shows us that an obvious family of covector fields would be those induced by smooth functions. For every $f \in \mathcal{C}^\infty(\mathbb{R}^n)$, let us define the covector field formally denoted df, by the following identity:

$$df(X) = X(f) \tag{2.22}$$

The left hand side is a smooth function by item 2. of Scholie 2.22, as is the right hand side. We call the covector field $df : \mathfrak{X}(\mathbb{R}^n) \longrightarrow \mathcal{C}^\infty(\mathbb{R}^n)$ the *differential of the function f*. Not every covector field is the differential of a function. For example, there is no smooth function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ such that the globally defined covector field $\xi = xdy - ydx$ would be the differential of[10]. Those covector fields that are of the form df for some smooth function f, and thus satisfy Equation (2.22), are called *exact differential 1-forms*. Recall that the basis vectors of the fibre of the cotangent bundle are denoted dx^i; this is not a coincidence, because dx^i is the differential of the coordinate function $x^i : \mathbb{R}^n \longmapsto \mathbb{R}$, and its action on a vector field gives:

$$dx^i(X) = X(x^i) = X^i$$

which is the i-th coordinate function of X.

Exercise 2.30. Show that the covector field defined in Example 2.23 is actually an exact differential 1-form by finding a function f from which it is the differential of.

Let us now compute the coordinates of df for some given f, by applying Equation (2.22) to every generator ∂_i:

$$df(\partial_i) = \frac{\partial f}{\partial x^i}$$

Hence, the covector field df decomposes as follows in the dual coframe:

$$df = \frac{\partial f}{\partial x^i}\, dx^i$$

In other words, the coordinate functions of df coincide with the components of the gradient of f. This is not a coincidence, because we have the following result:

[10] In polar coordinates (r, θ), this covector field reads $r^2 d\theta$, from which we understand that it cannot be written under the form df.

Proposition 2.31. *Given a smooth function $f \in \mathcal{C}^\infty(\mathbb{R}^n)$, there exists a unique vector field on $\mathbb{R}^n$, denoted $\overrightarrow{\mathrm{grad}}(f)$, such that:*

$$df(X) = g(\overrightarrow{\mathrm{grad}}(f), X) \qquad \textit{for every } X \in \mathfrak{X}(\mathbb{R}^n)$$

where g is the standard euclidean metric on the fiber of $T\mathbb{R}^n$.

Since the tangent bundle is *trivial*, it is diffeomorphic to the cartesian product $\mathbb{R}^n \times \mathbb{R}^n$. The metric g appearing in the statement of the proposition is the Euclidean metric defined on the fiber. Thus, on the basis vecors $\partial_1, \dots, \partial_n$, it satisfies $g(\partial_i, \partial_j) = 1$ if $i = j$ and 0 otherwise. Although it is not apparent in the proposition, the metric does not depend on the base point. The metric is bilinear so, for $X = X^i\partial_i$ and $Y = Y^j\partial_j$ two vector fields on $\mathbb{R}^n$, one has:

$$g(X, Y) = g(X^i\partial_i, Y^j\partial_j) = X^iY^i\, g(\partial_i, \partial_j) = \sum_{i=1}^n X^iY^i$$

Notice that we did not use the Einstein summation convention in the rightmost term because the two indices are both exponentiated. It can alternatively be written under this convention as X^iY_i, given that we lowered the second index via the formula $Y_i = g_{ij}Y^j$.

Remark 2.32. Proposition 2.31 is a particular case of a much more general result that states that a pseudo-Riemannian metric on a manifold M defines an isomorphism between TM and T^*M.

Let us now turn to the question of 'dualizing' the Lie bracket, so that we obtain an operator on $T^*\mathbb{R}^n$ that encodes it. Let us first rewrite Equation (2.10) using exact differential 1-forms:

$$df([X, Y]) = X(df(Y)) - Y(df(X)) \tag{2.23}$$

for every $X, Y \in \mathfrak{X}(\mathbb{R}^n)$. Although this equation is satisfied for exact covector fields, it does not mean that it is satisfied for all covector fields:

$$\xi([X, Y]) \stackrel{?}{=} X(\xi(Y)) - Y(\xi(X)) \tag{2.24}$$

We would like to measure 'how far' a given vector field ξ is from satisfying Equation (2.24). This can be done by passing the term on the left-hand side to the right-hand side, so that we can evaluate the difference between $X(\xi(Y)) - Y(\xi(X))$ and $\xi([X, Y])$. To this end, we set (formal notation) $d\xi$ to be the obstruction of a covector field ξ to satisfy Equation (2.24):

$$d\xi(X, Y) = X(\xi(Y)) - Y(\xi(X)) - \xi([X, Y]) \tag{2.25}$$

A covector field satisfies Equation (2.24) if and only if $d\xi = 0$, when evaluated on any two vector fields. We call such covector fields *closed differential 1-forms*. In particular, exact forms are closed.

Notice that since the right-hand side of Equation (2.25) is a smooth function, the object on the left-hand side formally noted $d\xi(X, Y)$ is a smooth function as well. Then, since $d\xi(X, Y) = -d\xi(Y, X)$, $d\xi$ defines a skew-symmetric operator that, when fed with two vector fields X and Y, gives a smooth function $d\xi(X, Y)$ whose evaluation at the point x reads:

$$\begin{aligned} d\xi :\ \ \mathfrak{X}(\mathbb{R}^n) \times \mathfrak{X}(\mathbb{R}^n) &\longrightarrow \mathcal{C}^\infty(\mathbb{R}^n) \\ (X, Y) \quad &\longmapsto d\xi(X, Y) : x \longmapsto X_x(\xi(Y)) - Y_x(\xi(X)) - \xi_x([X, Y]_x) \end{aligned}$$

This is consistent with the definitions of the objects so far, because e.g. $\xi(Y)$ is a smooth function, on which the derivation at x, $X_x : \mathcal{C}^\infty(\mathbb{R}^n) \longrightarrow \mathbb{R}$, acts, hence the term $X_x(\xi(Y))$ is a real number. Although the Lie bracket of two vector fields is not $\mathcal{C}^\infty(\mathbb{R}^n)$ bilinear, one can check that the map $d\xi$ is.

Exercise 2.33. Prove that $d\xi$ is $\mathcal{C}^\infty(\mathbb{R}^n)$ bilinear, i.e. that $d\xi(fX+gY,Z) = f\cdot d\xi(X,Z) + g\cdot d\xi(Y,Z)$ for every $f,g\in\mathcal{C}^\infty(\mathbb{R}^n)$ and $X,Y,Z\in\mathfrak{X}(\mathbb{R}^n)$, and vice versa with respect to the second variable.

Then, it is sufficient to know how $d\xi$ acts on the couples of basis vectors (∂_i,∂_j) to know how it acts on any couple of vector fields. Using Equation (2.13), Equation (2.25) becomes:

$$d\xi(\partial_i,\partial_j) = \frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j} \tag{2.26}$$

The fact that $d\xi(\partial_i,\partial_i) = 0$ is consistent with the fact that $d\xi$ is a skew-symmetric operator. The observation made in Equation (2.26) induces the following result:

Proposition 2.34. *Given a differential 1-form ξ, the skew-symmetric operator $d\xi$ can be seen as a section of the vector bundle $\wedge^2 T^*\mathbb{R}^n$, and reads:*

$$d\xi = \frac{1}{2}\left(\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}\right) dx^i\wedge dx^j \tag{2.27}$$

where the Einstein summation convention (on the two indices i and j!) has been used.

Proof. When one applies the right-hand side of this formula to two vector fields X and Y, one obtains:

$$\begin{aligned}
\frac{1}{2}\left(\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}\right) dx^i\wedge dx^j(X,Y) &= \frac{1}{2}\left(\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}\right)(dx^i\otimes dx^j - dx^j\otimes dx^i)(X,Y)\\
&= \frac{1}{2}\left(\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}\right)(X^i\cdot Y^j - X^j\cdot Y^i)\\
&= \left(X^i\frac{\partial\xi_j}{\partial x^i}\right)\cdot Y^j - \left(Y^i\frac{\partial\xi_j}{\partial x^i}\right)\cdot X^j\\
&= X(\xi_j\cdot Y^j) - Y(\xi_j\cdot X^j) - \xi_j\cdot X(Y^j) + \xi_j\cdot Y(X^j)\\
&= X(\xi(Y)) - Y(\xi(X)) - \xi([X,Y])
\end{aligned}$$

Where the symbol $\cdot$ has been used to symbolize and emphasize the product of two smooth functions. Since indices which are summed over can be relabelled at one's convenance, we have done this between the second line and the third line. The two supplementary terms added on the right in the fourth line compensate the addition of the terms $\xi_j\cdot X(Y^j) - \xi_j\cdot Y(X^j)$ which were necessary to form the terms $X(\xi_j\cdot Y^j) - Y(\xi_j\cdot X^j)$. Passing from the fourth line to the fifth and last line used Equation (2.11). □

Remark 2.35. The right hand side of Equation (2.27) contains redundant terms, since:

$$\left(\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}\right) dx^i\wedge dx^j = -\left(\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}\right) dx^j\wedge dx^i = \left(\frac{\partial\xi_i}{\partial x^j} - \frac{\partial\xi_j}{\partial x^i}\right) dx^j\wedge dx^i$$

The factor $\frac{1}{2}$ precisely compensates such redundancy, so that (2.27) can be rewritten:

$$d\xi = \sum_{1\le i<j\le n}\left(\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}\right) dx^i\wedge dx^j \tag{2.28}$$

Since the bivectors $dx^i\wedge dx^j$ for $i<j$ form a family of generators of $\bigwedge^2 T^*\mathbb{R}^2$, the coordinates functions of $d\xi$ in this basis are the $\frac{\partial\xi_j}{\partial x^i} - \frac{\partial\xi_i}{\partial x^j}$, and *not* one-half of it.

Exercise 2.36. Using Equation (1.19), check that applying Equation (2.27) or (2.28) to the couple (∂_i,∂_j) (beware of the range of the sums!) gives back Equation (2.26).

From Proposition 2.34 we deduce the very important (always true) observation:

Corollary 2.37. *Exact differential 1-forms are closed.*

Proof. We have already seen a proof of such a result by comparing Equations (2.23) and (2.25), but let us use here a more computational approach. Let ξ be an exact differential 1-form. Then there exists f a smooth function on $\mathbb{R}^n$ such that $\xi = df$. In particular it means that $\xi_i = \partial_i f$. Then:

$$d\xi = \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^i \partial x^j} - \frac{\partial^2 f}{\partial x^j \partial x^i}\right) dx^i \wedge dx^j = 0$$

Thus, ξ is a closed form. $\square$

Remark 2.38. We will see later that in the three-dimensional space $\mathbb{R}^3$, Corollary 2.37 is equivalent to the following identity:

$$\overrightarrow{\mathrm{curl}}\Big(\overrightarrow{\mathrm{grad}}(f)\Big) = 0$$

Now that we have an explicit formula for the operator $d\xi$, one may ask: which closed differential 1-forms are also exact? That is to say: which covector fields ξ satisfying Equation (2.24) are actually the differential of a function f, i.e. are such that $\xi = df$? Drawing on Proposition 2.31, this question has an equivalent interpretation in terms of vector fields: which vector field X on $\mathbb{R}^n$ such that $\overrightarrow{\mathrm{curl}}(X) = 0$ (whatever that means in dimension higher than 3) can be written as the gradient of a function f? Indeed, the standard euclidean metric g on the fiber of the tangent bundle defines an isomorphism $\widetilde{g}$ between the fiber of $T\mathbb{R}^n$ and $T^*\mathbb{R}^n$ (see subsection 1.2). The following Lemma is a particular case of Poincaré's Lemma:

Lemma 2.39. (Part of) Poincaré Lemma *Every closed differential 1-form defined on $\mathbb{R}^n$ is an exact form. That is to say, for every $\xi \in \Omega^1(\mathbb{R}^n)$ such that $d\xi = 0$, there exists a smooth function $f \in \mathcal{C}^\infty(\mathbb{R}^n)$ such that $\xi = df$.*

Proof. Let $\xi = \xi_i dx^i \in \Omega^1(\mathbb{R}^n)$, then define the following function:

$$f(x) = \int_0^1 x^i \xi_i(tx) dt$$

This function is smooth because the ξ_i are smooth functions by Scholie 2.22. Differientating f at a given x with respect to the k-th variable, and seeing the function $x \longmapsto \xi_i(tx)$ as the composite function $x \longmapsto tx \longmapsto \xi(tx)$, one obtains:

$$\begin{aligned}
\partial_k f(x) &= \int_0^1 \partial_k\big(x^i \xi_i(tx)\big) dt \\
&= \int_0^1 \delta_k^i \xi_i(tx) dt + \int_0^1 x^i \partial_k\big(\xi_i(tx)\big) dt \\
&= \int_0^1 \xi_k(tx) dt + \int_0^1 x^i t \partial_k \xi_i(tx) dt \\
&= \int_0^1 \xi_k(tx) dt + \int_0^1 t x^i \partial_i \xi_k(tx) dt \\
&= \int_0^1 \frac{d}{dt}\big(t\xi_k(tx)\big) dt \\
&= 1 \cdot \xi_k(x) - 0 \cdot \xi_k(0)
\end{aligned}$$

Here we have used the convention that $\partial_k\big(\xi_i(tx)\big)$ is the derivative in the k-th variable of the function $x \longmapsto \xi_i(tx)$ evaluated at x, whereas $\partial_k \xi_i(tx)$ is the derivative of the function ξ, evaluated at tx. This explains why a factor t appears on the third line. To pass to the fourth line we used Equation (2.26), whose left-hand side is zero because ξ is closed. Thus, we obtain that $\partial_k f = \xi_k$, so that $df = \xi$. $\square$

Remark 2.40. Actually, Poincaré's Lemma is more general: it applies to every differential p-forms, and does not necessarily assume that they are defined globally but only on star-shaped open subsets of $\mathbb{R}^n$.

Remark 2.41. When $\mathbb{R}^n = \mathbb{R}^3$, using the one-to-one correspondence between the fiber of the tangent space and the fiber of the cotangent space, Lemma 2.39 is equivalent to saying that every irrotational vector field X (i.e. such that $\overrightarrow{\mathrm{curl}}(X) = 0$) is conservative (i.e. it is the gradient of a function f).

Let us recall what we have so far: we have shown that for every smooth function, there is a differential 1-form df satisfying Equation (2.22). We have additionally shown that for every differential 1-form ξ, there is a differential 2-form $d\xi$ satisfying Equation (2.25). Additionally, by Corollary 2.37, every exact differential 1-form is closed, and by Lemma 2.39, every closed form is exact. Recalling that $\mathcal{C}^\infty(\mathbb{R}^n) = \Omega^0(\mathbb{R}^n)$, we can summarize the situation by the following sequence of spaces:

$$\Omega^0(\mathbb{R}^n) \xrightarrow{\quad d \quad} \Omega^1(\mathbb{R}^n) \xrightarrow{\quad d \quad} \Omega^2(\mathbb{R}^n) \tag{2.29}$$

Given Equations (2.23) and (2.25), the map d can be understood as the *dual* of the Lie bracket: whereas the Lie bracket is a bilinear map from $\mathfrak{X}^2(\mathbb{R}^n)$ to $\mathfrak{X}^1(\mathbb{R}^n)$, the map $d : \xi \longmapsto d\xi$ is a linear map from $\Omega^1(\mathbb{R}^n)$ to $\Omega^2(\mathbb{R}^n)$.

We have also seen that there is a strong relationship between the map d and the gradient of a function and the curl of a vector field. For example, we have seen that the identity $\overrightarrow{\mathrm{curl}}(\overrightarrow{\mathrm{grad}}(f)) = 0$ is a reformulation of Corollary 2.37. How does the divergence of a vector field enters in the picture? The same question arises for the Laplacian of a function. We are tempted to extend the sequence (2.29) to the right to account for those operators. This is will be the topic of the rest of this subsection. We need first to introduce a few abstract material:

Definition 2.42. *A* chain complex *(of vector spaces) is a graded vector space $E = (E_i)_{i\in\mathbb{Z}}$ equipped with a family of linear morphisms $d = (d_i : E_i \longrightarrow E_{i+1})_{i\in\mathbb{Z}}$:*

$$\ldots \xrightarrow{\quad d_{-3} \quad} E_{-2} \xrightarrow{\quad d_{-2} \quad} E_{-1} \xrightarrow{\quad d_{-1} \quad} E_0 \xrightarrow{\quad d_0 \quad} E_1 \xrightarrow{\quad d_1 \quad} E_2 \xrightarrow{\quad d_2 \quad} \ldots$$

such that $d_{i+1} \circ d_i = 0$. We call the linear operator d the differential *of the chain complex.*

Remark 2.43. In general we do not bother writing all the indices on the maps d_i and we write d instead, being understood that $d\big|_{E_i} = d_i$. In that case $d_{i+1} \circ d_i = 0$ becomes:

$$d^2 = 0$$

Moreover, the graded vector space may be graded above or below, or may be only positively/negatively graded, etc.

Let us now show how the sequence (2.29) can be extended to the right:

$$\Omega^0(\mathbb{R}^n) \xrightarrow{\quad d \quad} \Omega^1(\mathbb{R}^n) \xrightarrow{\quad d \quad} \Omega^2(\mathbb{R}^n) \xrightarrow{\quad d \quad} \ldots \xrightarrow{\quad d \quad} \Omega^{n-1}(\mathbb{R}^n) \xrightarrow{\quad d \quad} \Omega^n(\mathbb{R}^n) \xrightarrow{\quad d \quad} 0 \tag{2.30}$$

where all vector spaces of degree higher than n, on the right, are understood to be null vector spaces. Recall that each $\Omega^m(\mathbb{R}^n)$ is the space of smooth sections of the vector bundle $\bigwedge^m T^*\mathbb{R}^n$. It admits as a set of $\mathcal{C}^\infty(\mathbb{R}^n)$-linearly independent generators the elements:

$$\left\{dx^{i_1} \wedge \ldots \wedge dx^{i_m} \;\middle|\; 1 \leq i_1 < i_2 < \ldots < i_m \leq n\right\}$$

Drawing on what has been said in the discussion following Scholie 2.22 – in particular Equality (2.17) – one can equivalently see $\Omega^m(\mathbb{R}^n)$ as the space of alternating m $\mathcal{C}^\infty(\mathbb{R}^n)$-multi-linear forms on $\mathfrak{X}(\mathbb{R}^n)$ taking values in the smooth functions:

$$\Omega^m(\mathbb{R}^n) \simeq \mathrm{Hom}(\mathfrak{X}^m(\mathbb{R}^n), \mathcal{C}^\infty(\mathbb{R}^n))$$

The homomorphisms here have to be understood as homomorphisms of $\mathcal{C}^\infty(\mathbb{R}^n)$-modules. It means that for any given smooth section $\eta \in \Omega^m(\mathbb{R}^n)$ and any family of vectorfields $X_1, \ldots, X_m$, the element $\eta(X_1, \ldots, X_m)$ is a smooth function. This is a smoothness criterion for differential p-forms.

We have defined the linear morphism $d_0 : \Omega^0(\mathbb{R}^n) \longmapsto \Omega^1(\mathbb{R}^n)$ in Equation (2.22), and we have defined the linear morphism $d_1 : \Omega^1(\mathbb{R}^n) \longmapsto \Omega^2(\mathbb{R}^n)$ in Equation (2.25). In both case we have written df or $d\xi$ but it should be rigorously understood as $d_0 f$ and $d_1\xi$ if one wants to establish a differential whose notation is consistent with Definition 2.42. In the following we will write d instead of d_m because the latter notation is too cumbersome. Generalizing Equation (2.25) to any number of vector field $m \geq 1$, let us define the linear map $d : \Omega^m(\mathbb{R}^n) \longmapsto \Omega^{m+1}(\mathbb{R}^n)$ (should be understood as d_m then) from its action on any section $\eta \in \Omega^m(\mathbb{R}^n)$:

$$\begin{aligned} d\eta(X_1, \ldots, X_m, X_{m+1}) = &\sum_{i=1}^{m+1} (-1)^{i-1} X_i\big(\eta(X_1, \ldots, \widehat{X_i}, \ldots, X_{m+1})\big) \\ &+ \sum_{1 \leq i < j \leq m+1} (-1)^{i+j} \eta([X_i, X_j], X_1, \ldots, \widehat{X_i}, \ldots, \widehat{X_j}, \ldots, X_{m+1}) \end{aligned} \tag{2.31}$$

where the notation $(X_1, \ldots, \widehat{X_i}, \ldots, X_{m+1})$ means that the vector field X_i has been removed from the list of vector fields. In other words, for $2 \leq i \leq m$:

$$(X_1, \ldots, \widehat{X_i}, \ldots, X_{m+1}) = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{m+1})$$

whereas for $i = 0$ we obtain $(X_2, \ldots, X_{m+1})$ and for $i = m + 1$ we obtain $(X_1, \ldots, X_m)$. In a similar fashion we have:

$$([X_i, X_j], X_1, \ldots, \widehat{X_i}, \ldots, \widehat{X_j}, \ldots, X_{m+1}) = ([X_i, X_j], X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{j-1}, X_{j+1}, \ldots, X_{m+1})$$

with similar exceptions for $i = 0$ and $j = m + 1$. First notice that both terms on the right are smooth functions: $\eta(X_1, \ldots, \widehat{X_i}, \ldots, X_{m+1})$ is a smooth function, on which the vector field X_i acts; and one can check that there are only m vector fields in the term $\eta([X_i, X_j], X_1, \ldots, \widehat{X_i}, \ldots, \widehat{X_j}, \ldots, X_{m+1})$, making it a smooth function too. Thus, the right hand side is infinitely differentiable, which make the left-hand side infinitely differentiable.

Exercise 2.44. Check that Equation (2.31) gives back Equation (2.25) when $m = 1$.

Let us now give a formula for $d\eta$ in the basis of generators dx^i. To do this, evaluate Equation (2.31) on m given constant sections taken out of $\partial_1, \ldots, \partial_n$, so that the last term involving the Lie bracket vanishes by Equation (2.13).

Proposition 2.45. *The action of the operator $d : \Omega^\bullet(\mathbb{R}^n) \longrightarrow \Omega^{\bullet+1}(\mathbb{R}^n)$ on a differential m-form $\eta = \eta_{i_1\ldots i_m}\, dx^{i_1} \wedge \ldots \wedge dx^{i_m}$ is given in local coordinates by:*

$$d\eta = \partial_i(\eta_{i_1\ldots i_m})\, dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_m} \tag{2.32}$$

where Einstein summation convention is assumed.

Proof. Write $\eta = \eta_{i_1\ldots i_m}\, dx^{i_1} \wedge \ldots \wedge dx^{i_m}$, where $\eta_{i_1\ldots i_m}$ is a smooth function by Scholie 2.28 and where Einstein summation convention on contracted indices is assumed. Let $\partial_{j_1}, \ldots, \partial_{j_{m+1}}$ play the role of $X_1, \ldots, X_{m+1}$, then Equation (2.31) gives:

$$\begin{aligned}
d\eta(\partial_{j_1}, \ldots, \partial_{j_{m+1}}) &= \sum_{k=1}^{m+1} (-1)^{k-1} \partial_{j_k}\big(\eta(\partial_{j_1}, \ldots, \widehat{\partial_{j_k}}, \ldots, \partial_{j_{m+1}})\big) \\
&\quad + \sum_{1 \le k < l \le m+1} (-1)^{k+l} \eta(\underbrace{[\partial_{j_k}, \partial_{j_l}]}_{=0}, \partial_{j_1}, \ldots, \widehat{\partial_{j_k}}, \ldots, \widehat{\partial_{j_l}}, \ldots, \partial_{j_{m+1}}) \\
&= \sum_{k=1}^{m+1} (-1)^{k-1} \partial_{j_k}\big(m!\, \eta_{i_1\ldots i_m} \delta^{i_1}_{j_1} \ldots \delta^{i_{k-1}}_{j_{k-1}} \delta^{i_{k+1}}_{j_{k+1}} \ldots \delta^{i_{m+1}}_{j_{m+1}}\big) \\
&= m! \sum_{k=1}^{m+1} (-1)^{k-1} \partial_{j_k}\big(\eta_{j_1\ldots j_{k-1} j_{k+1}\ldots j_{m+1}}\big)
\end{aligned} \tag{2.33}$$

where Exercise 2.29 justifies that $m!$ pops out, and where we passed form the second line to the penultimate one by using Equation (2.20). Since the top left hand side of Equation (2.33) is $d\eta$ evaluated on $\partial_{j_1}, \ldots, \partial_{j_{m+1}}$, the same Exercise 2.29 implies that it is equal to $(m+1)!\,(d\eta)_{j_1\ldots j_{m+1}}$. Thus we have:

$$(m+1)!\,(d\eta)_{j_1\ldots j_{m+1}} = m! \sum_{k=1}^{m+1} (-1)^{k-1} \partial_{j_k}\big(\eta_{j_1\ldots j_{k-1} j_{k+1}\ldots j_{m+1}}\big)$$

Multiplying on the left and on the right by $dx^{j_1} \wedge \ldots \wedge dx^{j_{k-1}} \wedge dx^{j_k} \wedge dx^{j_{k+1}} \wedge \ldots \wedge dx^{j_{m+1}}$ and contracting the indices, this implies that $d\eta$ reads:

$$\begin{aligned}
d\eta &= \frac{1}{m+1} \sum_{k=1}^{m+1} (-1)^{k-1} \partial_{j_k}\big(\eta_{j_1\ldots j_{k-1} j_{k+1}\ldots j_{m+1}}\big)\, dx^{j_1} \wedge \ldots \wedge dx^{j_{k-1}} \wedge dx^{j_k} \wedge dx^{j_{k+1}} \wedge \ldots \wedge dx^{j_{m+1}} \\
&= \frac{1}{m+1} \sum_{k=1}^{m+1} \partial_{j_k}\big(\eta_{j_1\ldots j_{k-1} j_{k+1}\ldots j_{m+1}}\big)\, dx^{j_k} \wedge dx^{j_1} \wedge \ldots \wedge dx^{j_{k-1}} \wedge dx^{j_{k+1}} \wedge \ldots \wedge dx^{j_{m+1}} \\
&= \partial_{j_k}\big(\eta_{j_1\ldots j_{k-1} j_{k+1}\ldots j_{m+1}}\big)\, dx^{j_k} \wedge dx^{j_1} \wedge \ldots \wedge dx^{j_{k-1}} \wedge dx^{j_{k+1}} \wedge \ldots \wedge dx^{j_{m+1}}
\end{aligned} \tag{2.34}$$

where we have used Equation (2.21) between the first line and the second line, and where we use Einstein summation convention on repeated indices. But then, they are dummy indices and it does not change anything that we write the m indices $j_1, \ldots, j_{k-1}, j_{k+1}, \ldots, j_{m+1}$ as $i_1, \ldots, i_m$, and j_k as i, at the condition that they appear contracted with themselves in the formula. That is to say, Equation (2.34) can alternatively be written as:

$$d\eta = \partial_i(\eta_{i_1\ldots i_m})\, dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_m}$$

which is the required result. □

Exercise 2.46. Check that Equation (2.32) gives back Equation (2.27) when $m = 1$.

Proposition 2.45 allows us to prove the following proposition in a very elegant way:

Proposition 2.47. *The $\mathcal{C}^\infty(\mathbb{R}^n)$-linear morphism $d : \Omega^\bullet(\mathbb{R}^n) \longrightarrow \Omega^{\bullet+1}(\mathbb{R}^n)$ is a differential, i.e. $d \circ d = 0$.*

Proof. We know already by Corollary 2.37 that $d^2 f = 0$ for any smooth function. Then, let $\eta = \eta_{i_1\ldots i_m}\, dx^{i_1} \wedge \ldots \wedge dx^{i_m} \in \Omega^m(\mathbb{R}^n)$ be a differential m-form, for $m \geq 1$, and apply twice Equation (2.32):

$$\begin{aligned} d^2(\eta) &= d\big(\partial_i(\eta_{i_1\ldots i_m})\, dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_m}\big) \\ &= \partial_j\partial_i(\eta_{i_1\ldots i_m})\, dx^j \wedge dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_m} \end{aligned}$$

But the element $\partial_j\partial_i(\eta_{i_1\ldots i_m})$, symmetric under a permutation $i \leftrightarrow j$, is contracted with an element $dx^j \wedge dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_m}$, which is skew-symmetric under a permutation $i \leftrightarrow j$. Thus their contraction is zero. □

Remark 2.48. There is an alternative proof, much more computational, that relies exclusively on the expression of the differential given in Equation (2.31). This proof ressembles the proof that one could invoke to show the consistency of the Chevalley-Eilenberg differential in the cohomology theory of Lie algebra (recall that the space of vector fields is a Lie algebra!). Doing this alternative proof is a very good training to understand how differential forms interact with vector fields.

Thus, the graded vector space of differential forms $\Omega^\bullet(\mathbb{R}^n) = \big(\Omega^m(\mathbb{R}^n)\big)_{0\leq m\leq n}$, equipped with the differential d is a chain complex. We call it the *de Rham complex* and the differential d is called the *de Rham differential*. It is bounded below and above and it is understood in this complex that for every $m \leq -1$ and every $m \geq n$, $d_m = 0$ (see sequence (2.30)). We conclude this subsection by stating a unicity result that we do not prove, but which is worth knowing:

Proposition 2.49. *The de Rham differential is the unique $\mathcal{C}^\infty(\mathbb{R}^n)$-linear morphism $d : \Omega^\bullet(\mathbb{R}^n) \longrightarrow \Omega^{\bullet+1}(\mathbb{R}^n)$ which satisfies all three following properties:*

1. *on smooth functions (i.e. 0-forms), $df(X) = X(f)$;*
2. *$d \circ d = 0$;*
3. *for every $\eta \in \Omega^k(\mathbb{R}^n)$ and $\mu \in \Omega^l(\mathbb{R}^n)$:*

$$d(\eta \wedge \mu) = (d\eta) \wedge \mu + (-1)^k \eta \wedge (d\mu) \tag{2.35}$$

Proof. This is Theorem 12.14 in [Lee, 2003]. See also the paragraph at the top of page 313 to understand the equivalence between our definition of the de Rham differential and Lee's definition. □

Remark 2.50. Notice that Equation (2.35) implies that d is a *graded derivation* of the commutative graded algebra $(\Omega^\bullet(\mathbb{R}^n), \wedge)$, turning it into a *differential commutative graded algebra*, abbreviated *cdga* (notice the inversion of the letters in the abbreviation).

Example 2.51. *The vector calculus identities.* In three dimensional euclidean space $\mathbb{R}^3$, Proposition (2.47) will translate under an unexpected form. Recall what we said in Remark 2.38: that exact 1-forms are closed translates as the following identity:

$$\overrightarrow{\mathrm{curl}}\Big(\overrightarrow{\mathrm{grad}}(f)\Big) = 0$$

Let us explain this identity from the perspective of differential forms. We saw in Proposition 2.31 that the gradient of a function f is the image through the musical isomorphism $\sharp : T^*\mathbb{R}^n \longrightarrow T\mathbb{R}^n$ of the differential df via the formula:

$$\overrightarrow{\mathrm{grad}}(f) = (df)^\sharp$$

Let us pursue this analogy.

Assume we work in three dimensional euclidean space $\mathbb{R}^3$ with standard coordinates x, y, z (so that, for this discussion, x is a coordinate and not a point). We will not use Einstein summation convention either. Let ξ be a differential 1-form: $\xi = \xi_x dx + \xi_y dy + \xi_z dz$. Then Equation (2.28) tells us that:

$$d\xi = \left(\frac{\partial \xi_y}{\partial x} - \frac{\partial \xi_x}{\partial y}\right) dx \wedge dy + \left(\frac{\partial \xi_z}{\partial y} - \frac{\partial \xi_y}{\partial z}\right) dy \wedge dz + \left(\frac{\partial \xi_x}{\partial z} - \frac{\partial \xi_z}{\partial x}\right) dz \wedge dx$$

We recognize the coordinates of the $\overrightarrow{\mathrm{curl}}$ of the vector field $\xi^\sharp = \xi_x \frac{\partial}{\partial x} + \xi_y \frac{\partial}{\partial y} + \xi_z \frac{\partial}{\partial z}$. Since $d\xi$ is a 2-form, one only needs the Hodge star operator $\star : \Omega^2(\mathbb{R}^3) \longrightarrow \Omega^1(\mathbb{R}^3)$ and the musical isomorphisms to reconstruct the desired relation:

$$(\star\, d\xi)^\sharp = \overrightarrow{\mathrm{curl}}(\xi^\sharp)$$

Equivalently, for every vector field X, one has:

$$\overrightarrow{\mathrm{curl}}(X) = \left(\star\, d(X^\flat)\right)^\sharp$$

Next, pick up a differential 2-form $\eta = \eta_{xy} dx \wedge dy + \eta_{yz} dy \wedge dz + \eta_{zx} dz \wedge d_x$. Let write η_z instead of η_{xy}, η_x instead of η_{yz} and η_y instead of η_{zx}, for a reason that will soon be transparent. The differential of this 2-form is a 3-form, which, under some simple permutations of dx, dy and dz, can be written as:

$$d\eta = \left(\frac{\partial \eta_x}{\partial x} + \frac{\partial \eta_y}{\partial y} + \frac{\partial \eta_z}{\partial z}\right) dx \wedge dy \wedge dz$$

We recognize, in the parenthesis, the divergence of the vector field $(\star\eta)^\sharp = \eta_x \frac{\partial}{\partial x} + \eta_y \frac{\partial}{\partial y} + \eta_z \frac{\partial}{\partial z}$. Then, we have the following identity:

$$\star\, d\eta = \mathrm{div}\left((\star\eta)^\sharp\right) \tag{2.36}$$

The left-hand side is indeed a smooth function because $\star(\Omega^3(\mathbb{R}^3)) = \Omega^0(\mathbb{R}^3) = \mathcal{C}^\infty(\mathbb{R}^3)$. Equivalently, for every vector field X, one has:

$$\mathrm{div}(X) = \star\, d \star (X^\flat) \tag{2.37}$$

Notice that Equation (2.36) is equivalent to Equation (2.37) because in dimension 3, Equation (1.28) tells us that $\star \star \eta = \eta$ for any 2-form η.

Now let us check that the vector calculus identities in $\mathbb{R}^3$ amount to $d^2 = 0$. Let f be a smooth function on $\mathbb{R}^3$, then:

$$\begin{aligned}
\overrightarrow{\mathrm{curl}}\left(\overrightarrow{\mathrm{grad}}(f)\right) &= \overrightarrow{\mathrm{curl}}\left((df)^\sharp\right) \\
&= \left(\star\, d((df^\sharp)^\flat)\right)^\sharp \\
&= \left(\star\, \underbrace{d(df)}_{=0}\right)^\sharp
\end{aligned}$$

To pass from the second line to the third line, we used the fact that $\sharp$ and $\flat$ are inverse to one another. We thus obtain the infamous identity $\overrightarrow{\mathrm{curl}}(\overrightarrow{\mathrm{grad}}(f)) = 0$. Since the Hodge star operator and $\sharp$ are isomorphisms, we conclude that:

$$\overrightarrow{\mathrm{curl}}\left(\overrightarrow{\mathrm{grad}}(f)\right) = 0 \quad \Longleftrightarrow \quad d^2 f = 0$$

Now, turning to the next identity: let X be a vector field on $\mathbb{R}^3$. Then:

$$\begin{aligned}
\operatorname{div}\left(\overrightarrow{\operatorname{curl}}(X)\right) &= \operatorname{div}((\star d(X^\flat))^\sharp) \\
&= \star\, d \star \left(((\star d(X^\flat))^\sharp)^\flat\right) \\
&= \star\, d \star (\star\, d(X^\flat)) \\
&= \star\, \underbrace{d(d(X^\flat))}_{=0}
\end{aligned}$$

Since the Hodge star operator is an isomorphism, we deduce that:

$$\operatorname{div}\left(\overrightarrow{\operatorname{curl}}(X)\right) = 0 \iff d \circ d(X^\flat) = 0$$

Hence, the two most famous vector calculus identities $\overrightarrow{\operatorname{curl}} \circ \overrightarrow{\operatorname{grad}} = 0$ and $\operatorname{div} \circ \overrightarrow{\operatorname{curl}} = 0$ are nothing but Proposition 2.47 applied to $\mathbb{R}^3$. Thus, since on $\mathbb{R}^3$ with the euclidean metric $\star^{-1} = \star$, we have the following commutative diagram:

$$\begin{tikzcd}
 & \mathcal{C}^\infty(\mathbb{R}^3) \arrow[r, "\overrightarrow{\operatorname{grad}}"] \arrow[d, "\operatorname{id}"] & \mathfrak{X}(\mathbb{R}^3) \arrow[r, "\overrightarrow{\operatorname{curl}}"] \arrow[d, "\flat"] & \mathfrak{X}(\mathbb{R}^3) \arrow[r, "\operatorname{div}"] \arrow[d, "\star\flat"] & \mathcal{C}^\infty(\mathbb{R}^3) \arrow[d, "\star"] & \\
0 \arrow[r] & \Omega^0(\mathbb{R}^3) \arrow[r, "d"] & \Omega^1(\mathbb{R}^3) \arrow[r, "d"] & \Omega^2(\mathbb{R}^3) \arrow[r, "d"] & \Omega^3(\mathbb{R}^3) \arrow[r] & 0
\end{tikzcd}$$

2.4 De Rham cohomology and Maxwell equations

Let (E, d) be a chain complex of vector spaces. Then every map $d_i : E_i \longrightarrow E_{i+1}$ has a kernel and an image. We say that an element $x \in E_i$ is *closed* when $d_i x = 0$, whereas it is *exact* when $x = d_{i-1} y$ for some other element $y \in E_{i-1}$. Since $d^2 = 0$, we have the infamous result:

Proposition 2.52. *In a chain complex (E, d), every exact element is closed.*

The converse (that every closed element is exact) is in general not true, and actually those closed elements that are not exact carry important informations on the problem. That is why mathematicians have defined the following central notion in modern mathematics:

Definition 2.53. *Let (E, d) be a chain complex (of vector spaces). We define its* cohomology *as the graded vector vector space $H^\bullet(E) = \left(H^i(E)\right)_{i\in\mathbb{Z}}$, where for each $i \in \mathbb{Z}$, the space $H^i(E)$ is called the i-*th cohomology group *of E and is defined as the quotient:*

$$H^i(E) = \frac{\operatorname{Ker}(d_i)}{\operatorname{Im}(d_{i-1})}$$

We say that that the chain complex is exact *– equivalently, that it is a* resolution *– if $H^i(E) = 0$ for every $i \in \mathbb{Z}$.*

Remark 2.54. While in our context, the cohomology groups $H^i(E)$ are vector spaces, the word ‘group’ is widely used because the notion of cohomology applies to much more general objects than complexes of vector spaces. In any case, a vector space can be seen as an abelian group, with respect to the vector addition.

Elements of $H^i(E)$ are *equivalence classes* of vectors of E_i. For every element $x \in \operatorname{Ker}(d_i) \subset E_i$, we write $[x]$ the corresponding equivalence class in $H^i(E)$. We call $[x]$ the *cohomology class* of x. It has the following meaning: in cohomology, x is identified with every other closed element $x' \in E_i$ that can be written as follows:

$$x' = x + d_{i-1}y$$

for some $y \in E_{i-1}$. In such a case we say that x and x' are *cohomologous* and we write $[x] = [x']$. Therefore, any closed element x whose cohomology class is zero, i.e. such that $[x] = 0 \in H^i(E)$, is exact. To every cohomology class $\theta \in H^i(E)$, there exist an infinite number of representatives, i.e. those closed elements $x \in E_i$ such that $[x] = \theta$, because $x + dy$ would be another valid representative. A priori, there is no better choice of representative, except in certain cases (as we may see later).

The cohomology of the de Rham complex is called the *de Rham cohomology*. We write the cohomology groups of the de Rham complex as $H^i_{dR}(\mathbb{R}^n)$. Lemma 2.39 has shown that closed 1-forms are exact. That is to say, that $H^1_{dR}(\mathbb{R}^n) = 0$. This is actually much more general:

Proposition 2.55. *The de Rham cohomology of $\mathbb{R}^n$ satisfies:*

$$H^i_{dR}(\mathbb{R}^n) \simeq \begin{cases} \mathbb{R} & \text{if } i = 0 \\ 0 & \text{otherwise} \end{cases}$$

Proof. This is a consequence of Poincaré's Lemma, which states that the de Rham cohomology on every star-shaped open set (of a smooth manifold) is trivial (except for the 0-th cohomology group). See Theorem 15.11 in [Lee, 2003]. □

What kind of objects span the 0-th group of de Rham cohomology $H^0_{dR}(\mathbb{R}^n)$? We have the following situation:

$$0 \longrightarrow \Omega^0(\mathbb{R}^n) \xrightarrow{\ d\ } \Omega^1(\mathbb{R}^n) \xrightarrow{\ d\ } \dots$$

Then $H^0_{dR}(\mathbb{R}^n) = \operatorname{Ker}(d_0)$. Since d_0 is the morphism associating, to every function f, its differential, one deduces that $df = 0$ if and only if f is a constant function. Then $H^0_{dR}(\mathbb{R}^n) = \{\text{constant functions on } \mathbb{R}^n\}$, which is indeed a one-dimensional space. Another simple example sits at the other side of the chain complex: we know that $\Omega^n(\mathbb{R}^n)$ is one-dimensional and spanned by the standard volume form $\omega = dx^1 \wedge \dots \wedge dx^n$. Since $d(\Omega^n(\mathbb{R}^n)) = 0$, Proposition (2.55) tells us that there should be a differential $n-1$-form ν such that $\omega = d\nu$. There are several actually: for example $x_1\, dx^2 \wedge \dots \wedge dx^n$ or, more generally, those of the form $(-1)^{k-1} x_k\, dx^1 \wedge \dots \wedge dx^{k-1} \wedge dx^{k+1} \wedge \dots \wedge dx^n$.

Let us now apply all this machinery to Maxwell equations. They are equations that the electric field $\vec{E}$ and the magnetic field $\vec{B}$ should satisfy. Recall what they are (in *three-dimensional space*):

$$\operatorname{div}(\vec{E}) = \rho \tag{2.38}$$

$$\operatorname{div}(\vec{B}) = 0 \tag{2.39}$$

$$\overrightarrow{\operatorname{curl}}(\vec{E}) + \frac{\partial \vec{B}}{\partial t} = 0 \tag{2.40}$$

$$\overrightarrow{\operatorname{curl}}(\vec{B}) - \frac{\partial \vec{E}}{\partial t} = \vec{j} \tag{2.41}$$

We have used the *rationalized Planck units*, where:

$$c = 4\pi G = \hbar = \varepsilon_0 = k_{\mathrm{B}} = 1$$

Although $\vec{E}$ and $\vec{B}$ are usually considered as vector fields, the discussion in Example 2.51 has shown that using the musical isomorphisms allow us to adopt a much more synthesized perspective. However, from the knowledge we have of the differences between the respective behavior of the electric and the magnetic field, we expect that they would not carry the same degree as differential forms. Let us be more specific.

Let $\mathbb{M}^4$ be Minkowski space, i.e. $\mathbb{R}^4$ equipped with a metric $\eta_{\mu\nu}$ of signature $(3,1)$ – the indices ranging from 0 to 3, corresponding to the coordinates t, x, y and z. In other words: $\eta_{00} = -1$, and $\eta_{ii} = +1$ for $1 \leq i \leq 3$. The other components of the metric vanish. The volume form would then be $\omega = dt \wedge dx \wedge dy \wedge dz$. The order is important here because if we had taken t to be the fourth coordinate, then the corresponding volume form $dx \wedge dy \wedge dz \wedge dt$ would be minus ω. This would have repercussions on the definition of the Hodge star operator. The electric and magnetic fields are 1-forms and 2-forms on $\mathbb{R}^4$, respectively:

$$E = E_x\, dx + E_y\, dy + E_z\, dz$$

and

$$B = B_x\, dy \wedge dz + B_y\, dz \wedge dx + B_z\, dx \wedge dy$$

So, in particular, $E = \vec{E}^\flat$ and $B = \star(\vec{B}^\flat \wedge dt)$. We define the *field strength* as the following differential 2-form on $\mathbb{M}^4$:

$$F = B + E \wedge dt$$

In particular, this 2-form decomposes on the canonical frame $dx^\mu \wedge dx^\nu$ of the vector bundle $\bigwedge^2 T^*\mathbb{M}^4$, as:

$$F = \frac{1}{2} F_{\mu\nu}\, dx^\mu \wedge dx^\nu, \qquad \text{where} \qquad F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}$$

The current density $\vec{j}$ can be merged with the charge density ρ into a 4-vector $\vec{J} = \rho\frac{\partial}{\partial t} + \vec{j}$. Using the musical isomorphism $\sharp$, we transform this 4-vector into a differential 1-form J called the *current*:

$$J = -\rho\, dt + j_x\, dx + j_y\, dy + j_z\, dz$$

This allows us to have synthesized Maxwell equations:

Proposition 2.56. Geometric Maxwell equations *Equations* (2.39) *and* (2.40) *are equivalent to the* Bianchi identity*:*

$$dF = 0 \tag{2.42}$$

whereas Equations (2.38) *and* (2.41) *are equivalent to:*

$$\star d \star F = J \tag{2.43}$$

Proof. Equation (2.42) contains two terms:

$$dF = dB + dE \wedge dt \tag{2.44}$$

because $d^2t = 0$. Let us focus on the first term dB, using Proposition (2.45), and deleting the terms containing $dx \wedge dx$, $dy \wedge dy$ or $dz \wedge dz$:

$$\begin{aligned} dB &= \partial_x B_x\, dx \wedge dy \wedge dz + \partial_y B_y\, dy \wedge dz \wedge dx + \partial_z B_z\, dz \wedge dx \wedge dy \\ &\qquad + \partial_t B_x\, dt \wedge dy \wedge dz + \partial_t B_y\, dt \wedge dz \wedge dx + \partial_t B_z\, dt \wedge dx \wedge dy \\ &= \operatorname{div}(\vec{B}) dx \wedge dy \wedge dz + \partial_t B_x\, dy \wedge dz \wedge dt + \partial_t B_y\, dz \wedge dx \wedge dt + \partial_t B_z\, dx \wedge dy \wedge dt \end{aligned}$$

On the other hand, the second term of Equation (2.44) can be written as:

$$\begin{aligned} dE \wedge dt &= \partial_y E_x\, dy \wedge\, dx \wedge dt + \partial_z E_x\, dz \wedge dx \wedge dt \\ &\qquad + \partial_x E_y\, dx \wedge\, dy \wedge dt + \partial_z E_y\, dz \wedge dy \wedge dt \\ &\qquad\qquad + \partial_x E_z\, dx \wedge\, dz \wedge dt + \partial_y E_z\, dy \wedge dz \wedge dt \\ &= (\partial_x E_y - \partial_y E_x) dx \wedge\, dy \wedge dt + (\partial_y E_z - \partial_z E_y) dy \wedge\, dz \wedge dt \\ &\qquad + (\partial_z E_x - \partial_x E_z) dz \wedge\, dx \wedge dt \end{aligned}$$

Writing $dB + dE \wedge dt = 0$, one obtains the following identity:

$$\begin{aligned} 0 &= \operatorname{div}(\vec{B})\, dx \wedge dy \wedge dz + \Big(\partial_t B_x + \partial_y E_z - \partial_z E_y\Big) dy \wedge dz \wedge dt \\ &\qquad + \Big(\partial_t B_y + \partial_z E_x - \partial_x E_z\Big) dz \wedge dx \wedge dt + \Big(\partial_t B_z + \partial_x E_y - \partial_y E_x\Big) dx \wedge dy \wedge dt \end{aligned}$$

Thus, each term in parenthesis is equal to zero, and we obtain Equations (2.39) and (2.40). □

Exercise 2.57. Using the fact that the volume form is $\omega = dt \wedge dx \wedge dy \wedge dz$ in our convention for Minkowski space, show that:

$$(\star F)_{\mu\nu} = \begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & E_z & -E_y \\ -B_y & -E_z & 0 & E_x \\ -B_z & E_y & -E_x & 0 \end{pmatrix}$$

and prove Equation (2.43). Beware of the timelike direction dt that satisfies $\langle dt, dt \rangle = -1$ in Minkowski space.

The Bianchi identity (2.42) implies that the field strength is a closed 2-form. We have seen in Proposition 2.55 that the de Rham cohomology is vanishing, except for zero forms. Then, it means that F is an exact form, i.e. there exists a differential 1-form A such that:

$$F = dA \tag{2.45}$$

Using the musical isomorphism $\sharp$ on the 1-form:

$$A = A_\mu\, dx^\mu = -V\, dt + A_x\, dx + A_y\, dy + A_z\, dz$$

it gives a vector field on $\mathbb{M}^4$ that is written $A^\sharp = V \frac{\partial}{\partial t} + \vec{A}$, where V is the *scalar potential* and $\vec{A}$ is the *vector potential*. To keep this analogy in mind, we often call the differential 1-form A a potential for F. Obviously, in the case where the 2nd group of de Rham cohomology is not zero (this does not happen in $\mathbb{R}^n$ but could happen on other smooth manifolds), it may not be possible to find a vector potential for F. That is why Equation (2.42) is a topological condition. The physical information is contained in Equation (2.43): it is a necessary condition to the existence of a potential A for F.

Exercise 2.58. Check that Equation (2.45), with the potential $A = -V\,dt + A_x\,dx + A_y\,dy + A_z\,dz$ is equivalent to the two equations:

$$\vec{E} = -\overrightarrow{\text{grad}}(V) - \frac{\partial \vec{A}}{\partial t} \tag{2.46}$$
$$\vec{B} = \overrightarrow{\text{curl}}(\vec{A}) \tag{2.47}$$

Where $\overrightarrow{\text{curl}}$ and $\overrightarrow{\text{grad}}$ are considered to be the usual operators in $\mathbb{R}^3$.

Reinjecting Equation (2.45) in Equation (2.43), one obtains the following identity:

$$\star\, d \star dA = J \tag{2.48}$$

We will see later that $\star\, d \star d$ is (in Minkowski space) the d'Alembertian operator $\square = \frac{\partial^2}{\partial t^2} - \Delta$, so that one may show that Equation (2.48) is equivalent to the following two equations:

$$\Delta V + \frac{\partial}{\partial t}\text{div}(\vec{A}) = -\rho \tag{2.49}$$
$$\vec{\square}\vec{A} + \overrightarrow{\text{grad}}\left(\text{div}(\vec{A}) + \frac{\partial V}{\partial t}\right) = \vec{j} \tag{2.50}$$

Under the assumption that $\vec{E}$ and $\vec{B}$ are related to $\vec{A}$ and V through Equations (2.46) and (2.47), Equations (2.49) and (2.50) are equivalent to Equations (2.38) and (2.41). Hence we see that the geometric Maxwell equations are equivalent to the classical Maxwell equations. The Bianchi identity is a topological condition, automatically satisfied in $\mathbb{M}^4$ (but not on every manifold), so that the existence of A depends on the possibility of solving Equation (2.48).

The fact that the choice of potential A is fixed, up to an exact 1-form df – because $d(A+df) = dA = F$ – implies that one can make a specific choice for A that possibly simplifies Equations (2.49) and (2.50). When we make such a choice, we say that we fix the *gauge*. Let us choose the *Lorenz gauge*[11], defined by the condition:

$$\partial_\mu A^\mu = 0 \tag{2.51}$$

The notation A^μ symbolizes the components of the vector field $A^\flat$, so $A^\mu = \eta^{\mu\nu} A_\nu$: $A^0 = V$, and $A^i = A_i$ for $1 \leq i \leq 3$. Then, Equation (2.51) translates as:

$$\frac{\partial V}{\partial t} + \text{div}(\vec{A}) = 0$$

In this gauge, Equations (2.49) and (2.50) become:

$$\square V = \rho$$
$$\vec{\square}\vec{A} = \vec{j}$$

Fixing a gauge allows to obtain differential equations that may be easier to solve. There are several gauges in electromagnetism: the *Coulomb gauge*, where $\text{div}(\vec{A}) = 0$; the *Weyl* – or *temporal* – *gauge*, where $V = 0$. Electromagnetism is one of the simplest gauge theories. Its straightforward generalization is the Yang-Mills theory, whose study is postponed to a later chapter.

[11] From the Danish physicist Ludvig Lorenz, not to be confused with the Dutch physicist Hendrik Lorentz, to whom we attribute the *Lorentz transformations* in the theory of relativity, nor with the American physicist Edward Lorenz, who gave his name to the attractor looking like a butterfly in dynamical systems.

3 Differential calculus on smooth manifolds

In this Chapter, we will introduce the notion of smooth manifold and, relying on the mathematical background of the first chapter, develop the machinery needed to study action functionals and turn to more involved topics. The material presented in Chapter 2 will be central to the present chapter, because we will soon understand that a smooth manifold is *locally* like $\mathbb{R}^n$. It means that at least locally, in a neighborhood of a point, we should think of a smooth manifold as a n-dimensional vector space. The tangent bundle and the cotangent bundle on a smooth manifold, although defined globally, are thus always locally trivial. Differential forms on a manifold are thus always locally exact (because de Rham cohomology on $\mathbb{R}^n$ is almost trivial). Integration of differential forms, though, needs considering the global structure of the manifold. That is why it is often used to probe the topological structure of the manifold, e.g. in topological field theories.

3.1 Smooth manifolds

We emphasize in this presentation the role of functions on manifolds. There is indeed a deep relationship between a manifold, and the algebra of functions on this manifold. One should consider that defining a smooth manifold M from its topology and additional properties satisfied by the open sets is actually equivalent to characterizing what are smooth functions on this manifold M. This point of view illustrates the equivalence between the geometrico-analytic point of view, and the algebraic point of view:

$$\begin{array}{ccc} \text{Geometry} & & \text{Algebra} \\ M & \longleftrightarrow & \mathcal{C}^\infty(M) \end{array}$$

A smooth manifold is a particular case of a topological manifold which, in turn, is defined as follows:

Definition 3.1. *A* topological manifold of dimension n *is a topological space* M *(i.e. a set equipped with a topology of open subsets), that is:*

1. Hausdorff*, i.e. points can be separated by neighborhoods: for every pair of points* $x, y \in M$*, there are disjoint open subsets* $U, V \subset M$ *such that* $x \in U$ *and* $y \in V$*;*
2. second countable*, i.e. there exists a countable basis for the topology of* M*;*
3. locally euclidean*, i.e. every point of* M *has a neighborhood that is homeomorphic to an open subset of* $\mathbb{R}^n$*.*

The first property is a minimal assumption to avoid pathological cases that are not fit to do analysis. The second property means that the topology is generated by a countable family of open sets. This axiom, together with the Hausdorff property, has the following consequence:

Proposition 3.2. *Let* M *be a locally euclidean Hausdorff topological space, then* M *is second-countable if and only if* M *is paracompact – i.e. every open cover of* M *has a locally finite open refinement – and has countably many connected components.*

Proof. See Proposition 2.24 and Exercise 2.15 in [Lee, 2003]. □

This consequence is crucially needed to define partitions of unity, which are central to define integration on smooth manifolds and metrics on a manifold. The last property of Definition 3.1 means, more precisely, that for every point $x \in M$, there exists an open neighborhood U of x and an open subset $\widetilde{U} \subset \mathbb{R}^n$, together with a homeomorphism $\varphi : U \longrightarrow \widetilde{U}$ from U onto its image. We call the pair (U, φ) a *chart* or *coordinate chart* on M. At the cost of translating the image of the map φ in $\mathbb{R}^n$, one can always send x to $0 \in \mathbb{R}^n$. We then say that the chart is *centered* at x; every chart can be made centered at x by substracting the vector $\varphi(x)$. Denoting by $x^1, \ldots, x^n$ the standard coordinates centered at 0 on $\mathbb{R}^n$, we often define by abuse of notation the composite functions $x^i \circ \varphi$ with the same letters x^i. We then call the continuous functions $x^1, \ldots, x^n$ *local coordinates at* x. We define an *atlas* for M to be a collection $\mathscr{A}$ of charts whose domains cover M. Let us now give three pathological examples illustrating why we need the three assumptions in Definition 3.1.

Example 3.3. The 'line with two origins' is second-countable and locally euclidean, but not Hausdorff. It is obtained as the quotient of the union of the two horizontal lines $\{(x, y) \in \mathbb{R}^2 \,|\, y = 1\}$ and $\{(x, y) \in \mathbb{R}^2 \,|\, y = -1\}$ (with their respective subspace topology) under the following relation: $(x, 1) \sim (x, -1)$, whenever $x \neq 0$. Due to this very particular choice of quotient, the two origins cannot be separated by neighborhoods.

Example 3.4. The 'long line' is Hausdorff and locally Euclidean but not second-countable. It consists of segments $[0, 1[$ glued one after the other, but uncountably many times (contrary to the real line). The 'long ray' is the cartesian product $L = \omega_1 \times [0, 1[$ equipped with the order topology that arises from the lexicographical order on L. The long line is obtained by putting together a long ray in each direction (positive and negative).

Example 3.5. The 'figure eight' is Hausdorff and second-countable but not locally Euclidean at the origin.

Example 3.6. Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ be a continuous function. The *graph* of f is the subset of $\mathbb{R}^n \times \mathbb{R}^k$:

$$\Gamma(f) = \big\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^k \,\big|\, y = f(x)\big\}$$

Equipped with the subspace topology, it is a topological manifold. Indeed, denoting the projection on the first factor $pr_1 : (x, y) \longmapsto x$, we set $\varphi = pr_1|_{\Gamma(f)}$. Then φ is a continuous surjective map that has a continuous inverse: $\varphi^{-1}(x) = (x, f(x))$. Then it is a homeomorphism, and $(\Gamma(f), \varphi)$ is a global coordinate chart, turning $\Gamma(f)$ into a topological manifold of dimension n.

The fact that topological manifolds of dimension n are locally homeomorphic to $\mathbb{R}^n$ implies that we may be able to do differential calculus on it. For instance, given a continuous function $f : M \longrightarrow \mathbb{R}$ and a chart (U, φ) on M, one could consider the composition $f \circ \varphi^{-1} : \widetilde{U} \longrightarrow \mathbb{R}$, which is a real-valued function whose domain is an open subset $\widetilde{U}$ of a Euclidean space. Then it would make sense to say that f is *smooth* if and only if for every chart (U, φ) on M, $f \circ \varphi^{-1}$ is infinitely differentiable. However, this definition is not stable when passing from one open set U to another open set V, for the following reason: let (U, φ) and (V, ψ) be two charts whose underlying open sets U and V are overlapping; then, the *transition map* $\varphi \circ \psi^{-1}$ is a homeomorphism from $\psi(U \cap V)$ to $\varphi(U \cap V)$, both open subsets of $\mathbb{R}^n$. However, this map is not necessarily smooth, and this has the following consequence when we write f over the intersection $U \cap V$:

$$f \circ \psi^{-1} = f \circ \varphi^{-1} \circ (\varphi \circ \psi^{-1})$$

Then, even if $f \circ \varphi^{-1}$ and $f \circ \psi^{-1}$ are both differentiable, it does not imply that the function $\varphi \circ \psi^{-1}$ is, which is a bit problematic regarding the derivation rule of composite functions: $\partial_k(g \circ h) = \sum_{i=1}^n \partial_k(h^i)\partial_i(g)$ that one should expect in differential calculus.

To solve this issue, one should restrict the choice of coordinate charts adapted to the topological space M and pick up only a sub-family of those, that are 'compatible', i.e. such that the

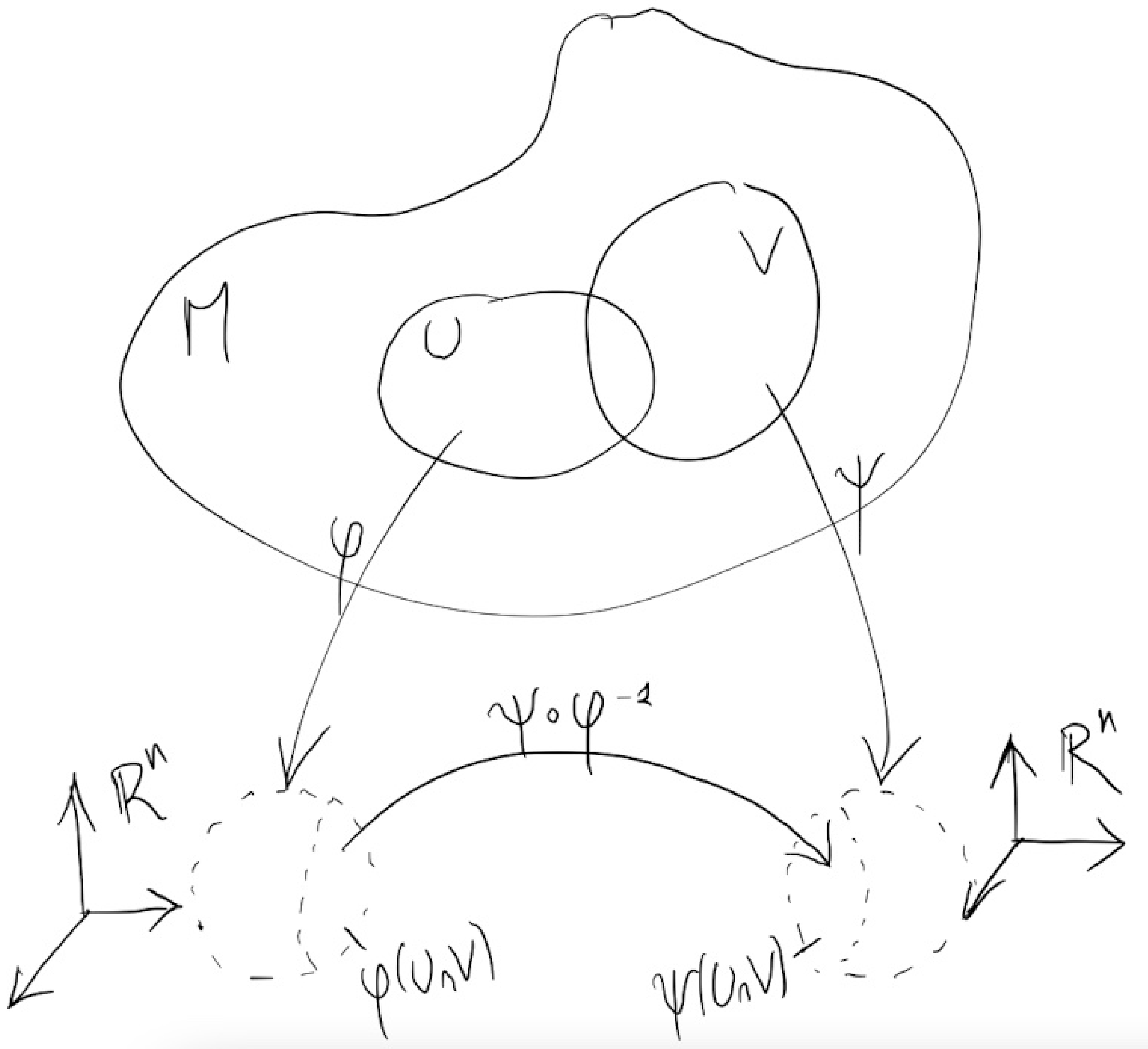

Figure 8: Two overlapping charts (U, φ) and (V, ψ) are smoothly compatible if the map $\psi \circ \varphi^{-1}$: $\varphi(U \cap V) \longrightarrow \psi(U \cap V)$ is a diffeomorphism. A smooth atlas is a collection of smoothly compatible charts covering M.

transition functions between two charts of that family are smooth. More precisely, two charts (U, φ) and (V, ψ) are said to be *smoothly compatible* if either $U \cap V = \emptyset$ or the transition map $\varphi \circ \psi^{-1} : \varphi(U \cap V) \longrightarrow \psi(U \cap V)$ is a diffeomorphism, i.e. a smooth homeomorphism from $\varphi(U \cap V)$ to $\psi(U \cap V)$, whose inverse is smooth as well. An atlas $\mathscr{A}$ is called a *smooth atlas* if any two charts in $\mathscr{A}$ are smoothly compatible with each other. Obviously a given (topological) atlas on M can give rise to several smooth atlases if, for instance, two families of charts covering M are smoothly compatible within the families, but not between them. Given a smooth atlas $\mathscr{A}$ on M, one says that a chart is *smoothly* compatible with the atlas, if this chart is smoothly compatible with every chart comprised in $\mathscr{A}$. The union of all compatible charts to a given smooth atlas $\mathscr{A}$ then defines a smooth atlas that is said *maximal*: it is not contained in any strictly larger smooth atlas. Such a smooth atlas is always very large since it contains every possible choice of smoothly compatible charts on the topological manifold M. Alternatively, one can work with equivalence classes of smooth atlases: two smooth atlases $\mathscr{A}$ and $\mathscr{A}'$ are considered equivalent if every chart of $\mathscr{A}$ is smoothly compatible with $\mathscr{A}'$. That allows working on a manifold with a single smooth atlas, consisting of only a few and practical charts, with the implicit understanding that many other charts and differentiable atlases are equally legitimate. Then, maximal smooth atlases are distinguished representents of their respective equivalence classes of compatible smooth atlases. Lemma 1.10 in [Lee, 2003] provides some understanding of the relationship between maximal smooth atlases and equivalence classes of smooth atlases:

Lemma 3.7. *Let M be a topological manifold of dimension n.*

1. *Every smooth atlas for M is contained in a unique maximal smooth atlas.*

2. *Two smooth atlases are equivalent if and only if their union is a smooth atlas.*

In particular, this shows that there may exist non-equivalent maximal smooth atlases for a given topological manifold M. Then, we can now define the central definition of this subsection:

Definition 3.8. *A* smooth structure *on a topological n dimensional manifold M is a maximal smooth atlas $\mathscr{A}$. A* smooth manifold of dimension n *is a pair $(M, \mathscr{A})$ – often only written M, omitting $\mathscr{A}$ – where M is a topological manifold of dimension n and $\mathscr{A}$ is a smooth structure on M.*

Remark 3.9. The smooth structure is an additional piece of data added to a topological manifold M. Most topological manifolds have uncountably many different smooth structures, but there exist topological manifolds that do not admit any smooth structure.

Example 3.10. The vector space $\mathbb{R}^n$ is a smooth manifold, when equipped with the chart $(\mathbb{R}^n, \mathrm{id}_{\mathbb{R}^n})$: the smooth structure consists of all the charts on $\mathbb{R}^n$ that are compatible with the first one.

Exercise 3.11. Check that the following charts on the 2-sphere are smoothly compatible:

$$\begin{aligned}
U_x^+ &= \{(x,y,z) \in \mathbb{S}^2 \,|\, x > 0\} \qquad && (\text{resp. } U_x^- \text{ for } x < 0)\\
U_y^+ &= \{(x,y,z) \in \mathbb{S}^2 \,|\, y > 0\} \qquad && (\text{resp. } U_y^- \text{ for } y < 0)\\
U_z^+ &= \{(x,y,z) \in \mathbb{S}^2 \,|\, z > 0\} \qquad && (\text{resp. } U_z^- \text{ for } z < 0)
\end{aligned}$$

and thus induce a smooth structure on $\mathbb{S}^2$ (the smooth atlas of every chart compatible with the above three charts). This kind of charts can be generalized to every n-sphere and defines the standard smooth structure on the n-sphere.

It turns out that if the dimension of the topological manifold M is higher than or equal to 1, then it has uncountably many distinct smooth structures (see Problem 1.3 in [Lee, 2003]). Thus we would like a notion of equivalence of smooth structures that mimic the topological equivalence of homeomorphic topological spaces: for this reason we introduce the notion of *diffeomorphism*. Let M, N be smooth manifolds, and let $f : M \longrightarrow N$ be any map (of sets). We say that f is a *smooth map* if for any $x \in M$, there exist smooth charts (U, φ) containing x and (V, ψ) containing $f(x)$ such that $f(U) \subset V$ and the composite map $\psi \circ f \circ \varphi^{-1}$ is smooth in the usual sense (i.e. infinitely differentiable) from $\varphi(U)$ to $\psi(V)$. The smooth map f is a *diffeomorphism* if it is bijective, and its inverse f^{-1} is smooth as well. The coordinate map φ of a smooth chart (U, φ) is a diffeomorphism onto its image $\varphi(U)$. While homeomorphisms define an equivalence relation between topological manifolds, diffeomorphisms define an equivalence relation between smooth manifolds. This relation allows to probe inequivalent smooth structures, for there exist topological manifolds admitting several smooth structures that are not diffeomorphic to one another. Finally, it is always useful to have the local variant of the former notion: $f : M \longrightarrow N$ is called a *local diffeomorphism* if every point $x \in M$ has a neighborhood U such that $f(U)$ is open in N and $f : U \longrightarrow f(U)$ is a diffeomorphism (onto its image).

Example 3.12. The euclidean vector space $\mathbb{R}^n$ has a unique smooth structure (up to diffeomorphism) unless $n = 4$, in which case $\mathbb{R}^4$ admits an uncountable number of non-diffeomorphic smooth structures, and these are called *exotic* $\mathbb{R}^4$. See [Lee, 2003, p. 37] for more details on this deep and exciting topic.

Example 3.13. The situation for the spheres is a bit more complicated. The following table shows how many smooth *types*, i.e. smooth-structures up to diffeomorphism, a n-sphere admits:

Dim.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Types	1	1	1	≥ 1	1	1	28	2	8	6	992	1	3	2	16256	2	16	16	523264	24

The n-spheres whose smooth structure is not diffeomorphic to the standard one are called *exotic sphere*. It is not known yet how many types the 4-sphere possesses.

A smooth map $f : M \longrightarrow \mathbb{R}$ is called a *smooth function* on M. The set of smooth functions on a smooth manifold M is denoted $\mathcal{C}^\infty(M)$. It is a commutative associative unital $\mathbb{R}$-algebra. The definition applies locally as well: any open set U of M inherits a smooth structure by restriction of the atlas to U (this can be seen by applying Lemma 1.23 in [Lee, 2003] to U); then we note $\mathcal{C}^\infty(U)$, or $\Omega^0(U)$, the space of functions on the open set $U \subset M$. Not every function in $\mathcal{C}^\infty(U)$ descend from a function in $\mathcal{C}^\infty(M)$: for example $]0,1[$ is a smooth manifold, whose smooth structure is inherited from its embedding in $\mathbb{R}$, but there are functions on $]0,1[$ that *do not* descend from functions on $\mathbb{R}$, e.g. $f : x \longmapsto \frac{1}{x(1-x)}$. Thus we see that $\mathcal{C}^\infty(U)$ is not a subalgebra of $\mathcal{C}^\infty(M)$. Rather, the assignment $U \longrightarrow \mathcal{C}^\infty(U)$ which associates to any open set a commutative associative unital $\mathbb{R}$-algebra is what is called a *sheaf of (commutative associative unital)* $\mathbb{R}$*-algebras* over M. There is a deep relationship between smooth manifolds and their algebras of functions. As for finite dimensional vector spaces, where the dual space E^* is an alternative characterization of a given vector space E, we expect some sort of duality between a smooth manifold M and its space of smooth functions $\mathcal{C}^\infty(M)$. There exists such a result in operator algebra:

Theorem 3.14. Gel'fand duality *For every arbitrary unital commutative C^*-algebra A there exists a compact Hausdorff topological space X such that A is equivalent to the algebra of complex-valued continuous functions on X: $A \simeq C(X)$. More precisely, there exists an equivalence of categories between the (opposite) category of unital commutative C^*-algebras and the category of compact Hausdorff topological spaces.*

The idea is not to understand this theorem but to see that for any given algebra of a certain type, there exists a geometric space such that this algebra plays the role of a subalgebra of functions – or operators – on this space. We expect this result to hold as well for smooth manifolds, that is to say: to any commutative, associative algebra with unit over $\mathbb{R}$ with some additional property, one can associate a smooth manifold, in the sens of Definition 3.8. Using a metaphor with physics, the algebra of functions would be considered as 'physical observables', and the associated smooth manifold would be what 'could be observed' by using these functions. It is thus meaningful that, given a different choice of observables, then what could be observed would change, and thus the associated manifold. There exists such a correspondence in algebraic geometry, between a choice of a commutative ring R, and its associated set of points that we call the *spectrum* of R: it is the set of prime ideals of R and is denoted $\mathrm{Spec}(R)$. Then, the ring R is considered as playing the role of the ring of functions on $\mathrm{Spec}(R)$. Then a *scheme* is a topological space X admitting a covering by open sets U_i, such that each U_i is the spectrum of a given ring R_i.

We can define a smooth manifold using the same kind of ideas. Let us start from a commutative, associative algebra with unit over $\mathbb{R}$ denoted $\mathscr{C}$ which will play the role of the algebra of smooth functions $\mathcal{C}^\infty(M)$ on the manifold M yet to define. Drawing an analogy from finite dimensional vector spaces, for which the dual of the dual of E is E (this is not true anymore in infinite dimension), we define M – also denoted $|\mathscr{C}|$ – to be the 'dual' of $\mathscr{C}$, i.e. the set of all $\mathbb{R}$-algebra homomorphisms to $\mathbb{R}$:

$$|\mathscr{C}| = \{x : \mathscr{C} \longrightarrow \mathbb{R}, f \longmapsto x(f) \text{ is an } \mathbb{R}\text{-algebra homomorphism}\}$$

To this set we can associate an algebra of 'physical observables' $\widetilde{\mathscr{C}}$, i.e. the $\mathbb{R}$-algebra of objects $\widetilde{f} : |\mathscr{C}| \longrightarrow \mathbb{R}$ associated to some $f \in \mathscr{C}$ via the formula $\widetilde{f}(x) = x(f)$. It turns out that $\mathscr{C}$ is surjective onto $\widetilde{\mathscr{C}}$, but not injective because there may be some non-trivial element $f \in \mathscr{C}$ which satisfies $x(f) = 0$ for every x. Since we want the elements of $\mathscr{C}$ to be in one-to-one correspondence with the physical observables on $M = |\mathscr{C}|$, we require $\mathscr{C}$ to satisfy the additional assumption that:

$$\bigcap_{x \in |\mathscr{C}|} \mathrm{Ker}(x) = 0$$

This condition is equivalent to saying that every element of $\mathscr{C}$ 'observe' at least something – for if $x(f) = 0$ for every $x \in |\mathscr{C}|$, the element f could not be used as a physical observable. Then, under this assumption, one can show that $\mathscr{C}$ becomes canonically isomorphic to the algebra of 'observables' $\widetilde{\mathscr{C}}$. Equipping the set $|\mathscr{C}|$ with the weakest topology for which all such functions are continuous turns $M = |\mathscr{C}|$ into a Hausdorff topological space. Then, $\mathscr{C}$ can be identified through its isomorphism with $\widetilde{\mathscr{C}}$ as a subalgebra of the algebra of continuous functions on M.

At this point, one would expect that the algebra $\mathscr{C}$ represents smooth functions on M. However this claim is still far from reality. A naive postulate would be to additionally require that $\mathscr{C}$ be locally isomorphic to $\mathcal{C}^\infty(\mathbb{R}^n)$ i.e., by assuming that there exists an open cover of M with a family of open sets U_i, $M = \bigcup_i U_i$, such that the restriction of $\mathscr{C}$ to each U_i is isomorphic to $\mathcal{C}^\infty(\mathbb{R}^n)$ as a $\mathbb{R}$-algebra. This would be the algebraic way of saying that the manifold M is locally like $\mathbb{R}^n$. However, this is mathematically not sufficient or does not define a smooth manifold as we understand it. The precise condition that one should require on $\mathscr{C}$ is much more subtle and very close to mathematical notions that are commonly used in algebraic geometry. Since it is not the topic of the current course, I refer to [Nestruev, 2003] for precise statements:

Theorem 3.15. *There is an equivalence of categories between the category of smooth manifolds and the category of complete geometric commutative associative unital $\mathbb{R}$-algebras.*

$$\begin{array}{ccc} \textit{Geometry} & & \textit{Algebra} \\ M & \longrightarrow & \mathcal{C}^\infty(M) \\ |\mathscr{C}| & \longleftarrow & \mathscr{C} \end{array}$$

What should be remembered from this discussion, is that there exists a canonical one-to-one correspondence between smooth manifolds and commutative associative unital $\mathbb{R}$-algebras satisfying additional specific conditions mirroring the topological and differential properties characterizing smooth manifolds. This bijective correspondence will be used frequently when we study gauge theories and constraint surfaces, and ultimately will be the fundamental characterization of graded manifolds in graded geometry.

3.2 Vector bundles, pushforwards, pullbacks

In this section we are interested in smooth maps, and there associated pushforwards and pullbacks. Most notions that we have seen in Chapter 2 will be understood as the 'local' versions of the objects presented in the present section. We have seen that over $\mathbb{R}^n$ a vector bundle is always trivial. This property will only be observed locally for vector bundles over smooth manifolds:

Definition 3.16. *A* vector bundle of rank k *over* M *is a topological space* E *together with a surjective continuous map* $\pi : E \longrightarrow M$*, satisfying the two following conditions:*

1. *for every* $x \in M$*, the preimage* $\pi^{-1}(x) \subset E$ *is a* k*-dimensional vector space, called the* fiber of E at x *and denoted* E_x*;*
2. *for each* $x \in M$*, there exists a neighborhood* U *of* x *in* M *and a homeomorphism* $\Phi_U : \pi^{-1}(U) \to U \times \mathbb{R}^k$ *(called a* local trivialization *of* E *over* U*), making the following triangle commutative:*

$$\begin{tikzcd} \pi^{-1}(U) \arrow[rr, "\Phi_U"] \arrow[dr, "\pi"'] & & U \times \mathbb{R}^k \arrow[dl, "pr_1"] \\ & U & \end{tikzcd}$$

where $pr_1 : U \times \mathbb{R}^k \longrightarrow U$ *is the projection on the first variable; and such that for every* $y \in U$*, the restriction of* Φ_U *to* E_y *is a linear isomorphism from* E_y *to* $\{y\} \times \mathbb{R}^k \simeq \mathbb{R}^k$*.*

The space E *is called the* total space *of the bundle,* M *is called its* base*, and* π *is called its* projection*. If* E *is a smooth manifold,* π *is a smooth map, and the local trivializations can be chosen to be diffeomorphisms, then* E *is said to be a* smooth vector bundle*. If there exists a local trivialization over all of* M *(called a* global trivialization *of* E*), then* E *is said to be a* trivial bundle*. In this case,* E *itself is homeomorphic (resp. diffeomorphic if* E *is smooth) to the product space* $M \times \mathbb{R}^k$*.*

Every point x of M admits a tangent space T_xM, whose definition is straightforward since it does not depend on the neighboring points of x: the tangent space to M at a given point x is the vector space of linear morphisms that are derivations at x, i.e. all the maps $X_x : \mathcal{C}^\infty(M) \longrightarrow \mathbb{R}$ satisfying Equations (2.2) and (2.3). The *tangent bundle* of the smooth manifold M is the disjoint union of the tangent spaces at each point:

$$TM = \bigsqcup_{x \in M} T_xM$$

It can be equipped with a natural topology and a natural smooth structure, making it into a rank n smooth vector bundle over M (see Lemma 4.1 in [Lee, 2003]). Similarly, the *cotangent bundle* it the disjoint union of the cotangent spaces at each point, i.e. the spaces dual to the tangent spaces: $T_x^*M = (T_xM)^*$. It can be showned that it is a smooth vector bundle of rank n (see Proposition 6.5 in [Lee, 2003]).

A *local section* of a vector bundle E over an open set $U \subset M$ is a continuous map $\sigma : U \longrightarrow E$ such that:

$$\pi \circ \sigma = \mathrm{id}_U$$

A *global section* is a local section defined over the whole manifold, i.e. such that $U = M$. When E is a smooth vector bundle and σ is a smooth map, we say it is a *smooth section*. Vector fields and differential 1-forms are smooth sections of the vector bundles TM and T^*M, respectively.

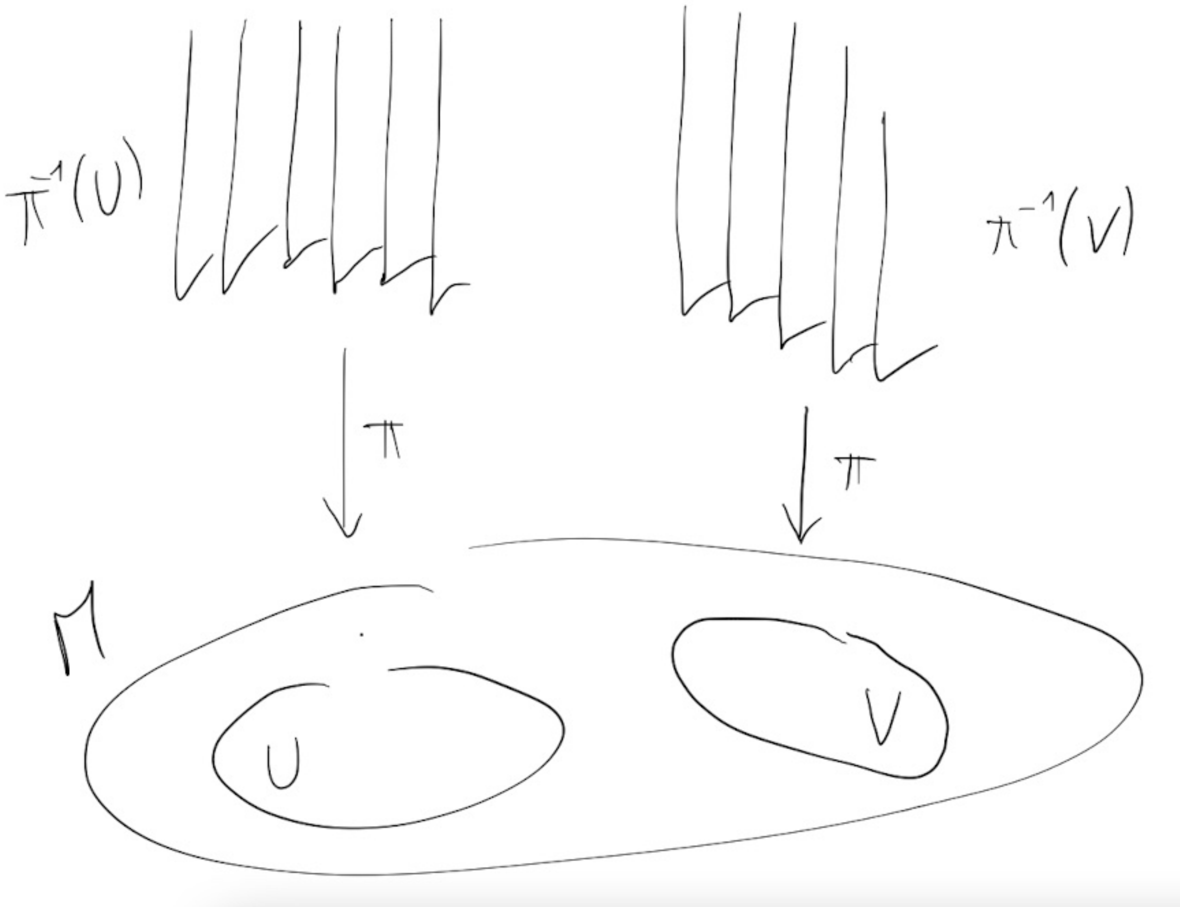

Figure 9: A (smooth) vector bundle is locally trivial, i.e. in the neighborhood of every point, there is an open set U, over which the pre-image $\pi^{-1}(U)$ is homeomorphic (resp. diffeomorphic) to $U \times \mathbb{R}^k$.

Some are defined only locally, while other are defined globally. The space of vector fields on an open set U is noted $\mathfrak{X}(U)$ while the space of differential 1-forms on the same open set is denoted $\Omega^1(U)$. A vector bundle always admits a smooth global section: the *zero section*, that has the particularity that it sends every point $x \in M$ to the zero vector in the fiber E_x. A set of k local sections $\sigma_1, \ldots, \sigma_k$ of E over U is called a *local frame* of E over U if for every $x \in U$, the vectors $\sigma_1(x), \ldots, \sigma_k(x)$ form a basis of the fiber E_x. It is called a *global frame* if $U = M$, and it is called *smooth* if the sections σ_i are smooth sections of the smooth vector bundle E.

Proposition 3.17. *A smooth vector bundle is trivial if and only if it admits a smooth global frame.*

Remark 3.18. Unless explicitly said, in the following we will always assume that vector bundles and their sections are smooth.

The space of smooth local sections of E over U is denoted $\Gamma_U(E)$ or $\Gamma(U, E)$; it is an infinite dimensional $\mathbb{R}$-vector space but a $\mathcal{C}^\infty(U)$-module. If there exists a smooth local frame on U – this occurs U is an open set trivializing E, i.e. satisfying the second item of 3.16 – then one observes that the frame plays the role of independent generators of $\Gamma_U(E)$, with respect to the action of $\mathcal{C}^\infty(U)$. One can always find such a frame in the neighborhood of every point, turning the assignment $U \longrightarrow \Gamma(U, E)$ in what is called a *locally free and finitely generated sheaf* (it is actually what is called a $\mathcal{C}^\infty$*-module*, because $\Gamma(U,E)$ is a $\mathcal{C}^\infty(U)$-module for every U). Pushing the idea further, $\mathcal{C}^\infty(U)$ can be seen as the space of local sections over U of the trivial bundle $M\times\mathbb{R}$. In the same manner that a smooth manifold can be defined by its algebra of functions, a smooth vector bundle can be defined through its space of sections. This fact is a central tenet of the general duality between geometry and algebra. The category of real vector bundles on M is equivalent to the category of locally free and finitely generated sheaves of $\mathcal{C}^\infty$-modules on M. This is the well-known Serre-Swan theorem which, in modern language, can be expressed as:

Theorem 3.19. Serre-Swan *There is an equivalence of categories between smooth vector bundles of finite rank over a smooth manifold M and finitely generated projective (equivalently: locally free) $\mathcal{C}^\infty$-modules over M.*

$$\begin{array}{ccc} \textit{Geometry} & & \textit{Algebra} \\ E & \longleftrightarrow & \Gamma(-,E) \end{array}$$

Proof. It is explained in Chapter 11 of [Nestruev, 2003]. □

Smooth sections of the vector bundle $\bigwedge^m T^*M=\bigsqcup_{x\in M}\bigwedge^m T^*_xM$ are called differential m-forms. They can be either locally defined or globally defined. The de Rham differential acts on these differential forms via Equation (2.31), and it induces the same notion of closeness and exactness of differential forms. For any open set U, the m-th de Rham cohomology group is:

$$H^m_{dR}(U)=\frac{\mathrm{Ker}\big(d:\Omega^m(U)\longrightarrow\Omega^{m+1}(U)\big)}{\mathrm{Im}\big(d:\Omega^{m-1}(U)\longrightarrow\Omega^m(U)\big)}$$

where we understand that $\Omega^{-1}(U)=\Omega^{n+1}(U)=0$. Since a smooth manifold is locally Euclidean, it means that in the neighborhood of every point, the cohomology groups are trivial except at level 0 (see Proposition 2.55), because a small enough open set is homeomorphic to $\mathbb{R}^n$. However, globally, the de Rham cohomology of a smooth manifold has no reason to be trivial. On the contrary, it is often not trivial because it contains information on the topological structure of the manifold, as the following examples show:

Example 3.20. The de Rham cohomology of the n-sphere $\mathbb{S}^n$ satisfies:

$$H^i_{dR}(\mathbb{S}^n)\simeq\begin{cases}\mathbb{R} & \text{if } i=0 \text{ or } i=n\\ 0 & \text{otherwise}\end{cases}$$

Example 3.21. The de Rham cohomology of the n-torus $\mathbb{T}^n$ satisfies:

$$H^i_{dR}(\mathbb{T}^n)\simeq\mathbb{R}^{\binom{n}{i}}$$

In order to define local frames of the tangent and cotangent bundles, one needs to introduce the notion of pushforwards, and pull backs. First, let us define the following important notion:

Definition 3.22. *A* morphism *of vector bundles, or* bundle map*, between smooth vector bundles E (over M) and E' (over N) is a pair (ψ,ϕ) of smooth maps $\psi:M\longrightarrow N$ and $\phi:E\longrightarrow E'$, making the following square commutative:*

$$\begin{tikzcd} E \arrow[r, "\phi"] \arrow[d, "\pi"'] & E' \arrow[d, "\pi'"] \\ M \arrow[r, "\psi"] & N \end{tikzcd}$$

and such that the restriction to the fibers $\phi_p : E_p \longrightarrow E'_{\psi(p)}$ is a linear morphism of vector spaces.

When $N = M$ and $\psi = \mathrm{id}_M$, the above diagram reduces to a triangle:

$$\begin{tikzcd} E \arrow[rr, "\phi"] \arrow[dr, "\pi"'] & & E' \arrow[dl, "\pi'"] \\ & M & \end{tikzcd}$$

Since both E and E' are smooth vector bundles over M, for every smooth section σ the composite $\phi \circ \sigma$ defines a smooth section of E'. Then, fiberwise linearity of ϕ implies that, for every smooth section σ of E, and every function $f \in \mathcal{C}^\infty(M)$, one has:

$$\phi(f(x)\sigma(x)) = f(x)\phi(\sigma(x))$$

Forgetting about the point x, this equation reads: $\phi \circ (f\sigma) = f(\phi \circ \sigma)$. Thus, vector bundle morphisms over the same base manifold are morphisms of the corresponding sheaves of sections that are $\mathcal{C}^\infty$-linear. This is actually an alternative characterization of vector bundle morphisms over a smooth manifold. This is a consequence of the Serre-Swan theorem:

Proposition 3.23. *Let E and E' be smooth vector bundles over a smooth manifold M. Then a map of sheaves $\Phi : \Gamma(-, E) \longrightarrow \Gamma(-, E')$ is linear over $\mathcal{C}^\infty(U)$ for every open set U if and only if there exists a smooth bundle map $\phi : E \longrightarrow E'$ over M such that $\Phi(\sigma) = \phi \circ \sigma$ for all smooth section σ.*

Let us provide an example of such a vector bundle morphism, that will become central in the following parts fo the course:

Definition 3.24. *Let M be a smooth manifold. A* Lie algebroid over M *is a smooth vector bundle A, together with:*

1. *a Lie algebra structure $[\,.\,,.\,]_A : \Gamma(A) \otimes \Gamma(A) \longrightarrow \Gamma(A)$ on the space of sections,*
2. *and a vector bundle morphism $\rho : A \longrightarrow TM$ called the* anchor,

such that the following Leibniz rule holds:

$$[a, fb]_A = f[a, b]_A + \rho(a)(f)\, b \tag{3.1}$$

for every $a, b \in \Gamma(A)$.

A Lie algebroid is a generalization of the tangent bundle, since Equation (3.1) is resembling Equation (2.12). Indeed, the tangent bundle is a particular case of a Lie algebroid, where the anchor is the identity map. Lie algebroids also generalize Lie algebras since a Lie algebra is a Lie algebroid over a point. As Lie algebras are infinitesimal counterparts of Lie groups, Lie algebroids are infinitesimal counterparts of *Lie groupoids*. These objects are widely used in mathematical physics nowadays.

Example 3.25. The space of endomorphisms of $\mathbb{R}^n$ is denoted $\mathrm{End}(\mathbb{R}^n)$ or $\mathfrak{gl}_n(\mathbb{R})$. By Example 2.15, it is a finite dimensional Lie algebra, with respect to the commutator of endomorphisms $[M, N] = M \circ N - N \circ M$. This Lie algebra additionally defines an infinitesimal Lie algebra action on $\mathbb{R}^n$ via the following Lie algebra homomorphism:

$$\begin{aligned}\overline{\rho}:\ \mathfrak{gl}_n(\mathbb{R}) &\longrightarrow \mathfrak{X}(\mathbb{R}^n)\\ M &\longrightarrow X_M : (x, f) \longmapsto \frac{d}{dt}\Big|_{t=0} f(x \cdot \exp(tM))\end{aligned}$$

where, on he right-hand side, the group element acts from the right. On the basis $(E_{i,j})_{1\leq i,j\leq n}$ of $\mathfrak{gl}_n(\mathbb{R})$ this homomorphism then reads at the point x:

$$\overline{\rho}(E_{i,j})_x = X_{E_{i,j},x} = x^i \frac{\partial}{\partial x^j}$$

where the x^i are the coordinates of the point x. These data are sufficient to define a Lie algebroid over $\mathbb{R}^n$ via the following data: $A = \mathbb{R}^n \times \mathfrak{gl}_n(\mathbb{R})$ (it is a trivial vector bundle); $[\,.\,,.\,]_A$ is defined on the constant sections as the bracket on $\mathfrak{gl}_n(\mathbb{R})$ and then it is generalized to every smooth sections by the Leibniz rule (3.1); the anchor map is defined on the frame of constant sections $(E_{i,j})_{1\leq i,j\leq n}$ of A by:

$$\rho(E_{i,j}) = x^i \frac{\partial}{\partial x^j}$$

Then the infinitesimal action of $\mathfrak{gl}_n(\mathbb{R})$ on $\mathbb{R}^n$ straightforwardly translates in the data contained in a Lie algebroid. More generally, the action of a Lie algebra $\mathfrak{g}$ on a manifold M can be encoded in what is called an *action Lie algebroid* $M \times \mathfrak{g}$.

Remark 3.26. One could have defined the infinitesimal action of $\mathfrak{gl}_n(\mathbb{R})$ on $\mathbb{R}^n$ as a left action, but in that case we need to add a minus sign to have a Lie algebra homomorphism:

$$\begin{aligned}\mathfrak{gl}_n(\mathbb{R}) &\longrightarrow \mathfrak{X}(\mathbb{R}^n)\\ M &\longrightarrow X_M : (x, f) \longmapsto -\frac{d}{dt}\Big|_{t=0} f(\exp(tM) \cdot x)\end{aligned} \tag{3.2}$$

and the basis vectors $E_{i,j}$ are sent to $-x^j \frac{\partial}{\partial x^i}$, which is not very practical. If the minus sign had not been present, we would have a Lie algebra *anti-homomorphism* $\mathfrak{g} \longrightarrow \mathfrak{X}(M)$. The choice of a minus sign or, more conveniently, a right action, comes from the following facts (that we summarize very sketchily): to any smooth manifold M, one can associate its set of diffeomorphisms, denoted $\mathrm{Diff}(M)$. This space can be equipped with an infinite dimensional Lie group structure whose local charts are modeled over the infinite vector space $\mathfrak{X}(M)$.

Then, the left invariant vector fields over this Lie group form a Lie algebra $\mathfrak{diff}(M)$, in bijection with the space of vector fields on M. However, the choice of Lie bracket on $\mathfrak{X}(M)$, as defined in Equation (2.10), corresponds to *minus* the Lie bracket on $\mathfrak{diff}(M)$, and we write $\mathfrak{X}(M) \simeq \mathfrak{diff}(M)^{op}$. This can be explained by the fact that the diffeomorphisms act on M from the left, and thus the induced linear map between the Lie algebras of $\mathrm{Diff}(M)$ and $\mathfrak{X}(M)$ is an anti-homomorphism (because the vector field associated to a given diffeomorphism is obtained without involving the minus sign appearing in Equation (3.2)). More generally, a left action of a Lie group G on the manifold M is equivalent to a group homomorphism $G \longrightarrow \mathrm{Diff}(M)$. This group homomorphism induces in turn a Lie algebra homomorphism $\mathfrak{g} \longrightarrow \mathfrak{diff}(M)$, and thus a Lie algebra anti-homomorphism from $\mathfrak{g}$ to $\mathfrak{X}(M)$. However, a right-action of a Lie group G on a manifold M is equivalent to a Lie group homomorphism $G \longrightarrow \mathrm{Diff}(M)^{op}$, where $\mathrm{Diff}(M)^{op}$ is the Lie group modeled on $\mathrm{Diff}(M)$ but with multiplication from the right. Then in that case, the Lie algebra homomorphism $\mathfrak{g} \longrightarrow \mathfrak{diff}(M)^{op}$ is equivalent to a Lie algebra homomorphism $\mathfrak{g} \longrightarrow \mathfrak{X}(M)$. For more details on these questions see Section 3.3 of these lecture notes.

Exercise 3.27. By using the Jacobi identity on $\Gamma(A)$ and Equation (3.1), show that the anchor map is a *Lie algebra homomorphism* from $\Gamma(A)$ to $\mathfrak{X}(M)$. That is to say, it satisfies the following equation:

$$\rho([a,b]_A) = [\rho(a), \rho(b)]$$

for every smooth sections $a, b \in \Gamma(A)$.

Here is another important example of a vector bundle morphism:

Definition 3.28. *Let M, N be smooth manifolds. For every smooth map $F : M \longrightarrow N$ we associate a vector bundle morphism $F_* : TM \longrightarrow TN$ called the* pushforward*, defined on each fiber T_xM as:*

$$F_*(X_x)(f) = X_x(f \circ F)$$

for every $f \in \mathcal{C}^\infty(N)$ and $X_x \in T_xM$.

The pushforward is a vector bundle morphism sending tangent vectors on M to tangent vectors on N:

$$\begin{tikzcd} TM \arrow[r, "F_*"] \arrow[d, "\pi"'] & TN \arrow[d, "\pi'"] \\ M \arrow[r, "F"] & N \end{tikzcd}$$

Given a point $x \in M$ (resp. $y \in N$) and a trivializing neighborhood (U, φ) centered at x (resp. (V, ψ) centered at y), then the matrix of the linear morphism $F_* : T_xM \longrightarrow T_{F(x)}N$ at x is the Jacobian of the smooth map $\psi \circ F \circ \varphi^{-1} : \widetilde{U} \longrightarrow \widetilde{V}$ at $\varphi(x)$, where $\widetilde{U} = \varphi(U)$ is an open subset of $\mathbb{R}^n$ centered at 0 (and respectively for $\widetilde{V}$):

$$F_*\big|_x = \left(\frac{\partial(\psi \circ F \circ \varphi^{-1})^j}{\partial x^i}(\varphi(x)) \right)_{1 \leq i,j \leq n} \tag{3.3}$$

In the above formula, the coordinates x^i in the denominator denote the standard coordinates on $\widetilde{U}$. Notice that the numerator $(\psi \circ F \circ \varphi^{-1})^j$ can alternatively be written $\psi^j \circ F \circ \varphi^{-1}$, where ψ^j is a smooth function on V and denotes the j-th component of ψ with respect to the standard coordinates on $\widetilde{V}$.

The pushforward F_* is then the best linear approximation of F at the point x. The rank of the Jacobian matrix at x characterizes this smooth map at this point and is called the *rank* of F at x. If the rank of F is constant for every point of the smooth manifold M then we say that F has *constant rank* and denote it by $\mathrm{rk}(F)$. We have the following conventions:

1. if $\mathrm{rk}(F) = \dim(M)$ at every point (i.e. F_* is injective everywhere), then F is called an *immersion*;

2. if $\mathrm{rk}(F) = \dim(N)$ at every point (i.e. F_* is surjective everywhere), then F is called a *submersion*.

In both cases, these properties are partly independent from the fact that F being injective, surjective of bijective. For instance, when $\dim(M) = \dim(N)$, F is a (local) diffeomorphism if and only if it is an immersion or a submersion. However, F needs not be a global diffeomorphism: for that it should be either injective or surjective.

Remark 3.29. The pushforward admits several other notations: dF because it is the differential of the map F (so that when $N = \mathbb{R}$, we retrieve the usual differential of functions), TF to symbolize that it is a map between tangent spaces, etc.

Lemma 3.30. *Given two smooth functions $F : M \longrightarrow N$ and $G : N \longrightarrow P$, the pushforward of the composite $G \circ F : M \longrightarrow P$ preserves the order:*

$$(G \circ F)_* = G_* \circ F_*$$

Be aware that although tangent vectors always behave well under pushforwards, it may not be the case for vector fields, i.e. sections of the tangent bundle. This phenomenon actually exists for every vector bundle morphism, so we will study this problem in the general setting. Let E (resp. E') be a smooth vector bundle over the smooth manifold M (resp. N), and let (ψ, ϕ) be a vector bundle morphism from E to E'. Let σ be a smooth section of E, then under the action of ϕ it becomes a map $\sigma^{(\psi,\phi)} : \mathrm{Im}(\psi) \longrightarrow E'$ defined on the subset $\mathrm{Im}(\psi) \subset N$ by:

$$\sigma^{(\psi,\phi)}(\psi(x)) = \phi(\sigma(x))$$

There may be several obstructions to the fact that $\sigma^{(\psi,\phi)}$ forms a smooth section of E'. This can be seen in several situations: 1) local sections should be defined on open sets, but if the smooth map ψ is not open (i.e. if $\psi(U)$ is not necessarily open while U is open) then $\mathrm{Im}(\psi)$ may not be even open in N, so that the map $\sigma^{(\psi,\phi)}$ could not be qualified as a local section of E'; 2) if ψ is not injective then ϕ can send two conflicting informations to the same point of N: take $x, y \in M$ such that $z = \psi(x) = \psi(y)$, but then for any choice of smooth section σ, how would be defined $\sigma^{(\psi,\phi)}(z)$? As $\phi(\sigma(x))$ or as $\phi(\sigma(y))$?

These problems can be explicitly solved if one introduces the notion of *pullback bundle*, that would be introduced as an intermediary bundle between E and E'. Given a smooth map $\psi : M \longrightarrow N$, and a vector bundle E' over N, one defines the *pullback bundle of E' along ψ*, denoted $\psi^!E'$, as the vector bundle over M such that the fiber over the point $x \in M$ is $(\psi^!E')_x = E'_{\psi(x)}$. Thus, as a set, the pullback bundle is the disjoint union $\psi^!E' = \bigsqcup_{x\in M} E'_{\psi(x)}$ and the projection map is denoted $\psi^!\pi' : E'_{\psi(x)} \longmapsto x$. Notice that the fact that we have a disjoint union (and not a mere union) is crucial so that the fibers associated to the pre-image of the same point stay disjoint in $\psi^!E'$. Under this convention, the vector bundle morphism (ψ, ϕ) induces a vector bundle morphism (id_M, ϕ) covering the identity of M:

$$\begin{tikzcd} E \arrow[rr, "\phi"] \arrow[dr, "\pi"'] & & \psi^!E' \arrow[dl, "\psi^!\pi'"] \\ & M & \end{tikzcd}$$

In that context, any local smooth section σ of E induces a local smooth section $\sigma^{(\mathrm{id}_M,\phi)}$ of $\psi^!E'$. Indeed, since $\psi^!E'$ is a vector bundle over M, if σ is defined over an open set U then $\sigma^{(\mathrm{id}_M,\phi)}$ stays defined over the same open set U. Moreover, the possible lack of injectivity of ψ is now solved: the images of two different fibers of E through ϕ are sent to different fibers of $\psi^!E'$ so they cannot be confounded. Thus, even though $\psi(x) = \psi(y)$, the element $\sigma^{(\mathrm{id}_M,\phi)}(x)$ is a vector of the fiber of $\psi^!E'$ over x, while $\sigma^{(\mathrm{id}_M,\phi)}(y) \in (\psi^!E')_y$. To conclude, the vector bundle morphism $(\mathrm{id}_M, \phi) : E \longrightarrow \psi^!E'$ sends smooth sections to smooth sections.

Now, notice that any smooth section τ of E' over some open set $U \subset N$ defines a map $\psi^!\tau : \psi^{-1}(U) \longrightarrow \psi^!E'$ by the following identity:

$$(\psi^!\tau)_x = \tau_{\psi(x)}$$

This assignment is well defined, and it is easy to see that it is additionally smooth, hence $\psi^!\tau$ is a smooth section of $\psi^!E'$ over $\psi^{-1}(U)$. Then, we say that a smooth section $\sigma \in \Gamma(E)$ over some open set $U \subset M$ and a smooth section $\tau \in \Gamma(E')$ over some open set $V \subset N$ containing $\psi(U)$ are (ψ, ϕ)-*related* if we have the following identity over U:

$$\sigma^{(\mathrm{id}_M, \phi)} = \psi^!\tau$$

In that case, we can consider that the image of the smooth section σ through (ψ, ϕ) is τ. Obviously, if ψ is not surjective, σ can be related to many sections τ (that could for instance differ outside $\mathrm{Im}(\psi)$). Moreover, not every smooth section of E is related to a smooth section of E'.

Example 3.31. Let $M = N = \mathbb{R}$ and let $E = E' = \mathbb{R}^2$. Let $\psi(x) = x^2$ and let $\phi(x, y) = (x, xy)$ (the latter is indeed linear on the fibers). Be aware that although $M = N$ and $E = E'$, the fact that ψ is not the identity implies that not all smooth sections of E are (ψ, ϕ) related to smooth sections of E'. Let $\sigma : x \longmapsto (x, \sin(x))$ a smooth section of E; determine what is the section $\sigma^{(\mathrm{id}_M, \phi)} \in \Gamma(\psi^!E')$. Then, find a global smooth section τ of $E!$ which is (ψ, ϕ)-related to σ. Find a smooth section σ' of E such that there exist *no* global smooth section of E' that would be (ψ, ϕ)-related to σ'.

To summarize we have the following situation (this diagram should not be understood as a commutative diagram, but as a metaphor, even though the square on the left is commutative):

$$\begin{tikzcd}
E \arrow[r, "\phi"] \arrow[d, "\pi"'] & \psi^!E' \arrow[d, "\psi^!\pi'"'] & E' \arrow[l, "\psi^!"'] \arrow[d, "\pi'"] \\
M \arrow[r, "\mathrm{id}_M"] & M \arrow[r, "\psi"] & N
\end{tikzcd}$$

Using this construction, we understand that, given a smooth map $F : M \longrightarrow N$, a vector field X on M is F-related to a vector field Y on N, if:

$$F_*X = F^!Y$$

The notion of pullback bundle, as the name indicates, allows to make sense of so-called *pullbacks*:

Definition 3.32. *Let M, N be smooth manifolds. For every smooth map $F : M \longrightarrow N$ we associate the* pullback $F^* : F^!T^*N \longrightarrow T^*M$*, defined on each fiber $T^*_{F(x)}N$ as:*

$$F^*(\xi_{F(x)})(X_x) = \xi_{F(x)}\big(F_*(X_x)\big) \tag{3.4}$$

for every $\xi_{F(x)} \in T^*_{F(x)}N$ and $X_x \in T_xM$.

The pullback is a vector bundle morphism covering the identity of M:

$$\begin{tikzcd}
F^!T^*N \arrow[rr, "F^*"] \arrow[dr, "F^!\pi'"'] & & T^*M \arrow[dl, "\pi"] \\
& M &
\end{tikzcd}$$

Differential 1-forms on N can be pullbacked on M via Equation (3.4) but, contrary to vector fields that do not behave well under pushforwards, differential forms actually behave very well under pullbacks. For every covector field $\xi \in \Omega^1(N)$, the pullback of ξ is the unique section $F^*\xi$ of T^*M defined at x as in Equation (3.4):

$$(F^*\xi)_x = F^*(\xi_{F(x)})$$

Note that there is no ambiguity in the definition of $F^*(\xi)$, contrary to the case of the pushforward of vector fields. This section is smooth because the function $(F^*\xi)(X) : x \longmapsto \xi_{F(x)}(F_*(X_x))$ is a smooth function of x (it can be seen from the fact that $x \longmapsto \xi_{F(x)}$ is a smooth section of $F^!T^*N$, while $x \longmapsto F_*X_x$ is a smooth section of $F^!TN$), thus it satisfies criterion 2. of Scholie 2.22.

Thus, the pullback can be extended to a smooth map $F^* : \Omega^1(N) \longrightarrow \Omega^1(M)$. We can also extend F^* to smooth functions, for if $f \in \mathcal{C}^\infty(N)$, we define, for every $x \in M$:

$$F^*(f)(x) = f(F(x))$$

More generally, for every differential m-form η on N ($m \geq 1$), one defines the pullback of η to M from its action on m vector fields $X_1, \ldots, X_m \in \mathfrak{X}(M)$:

$$F^*(\eta)(X_1, \ldots, X_m) = F^!\eta(F_*X_1, \ldots, F_*X_m)$$

where $F^!\eta \in \Gamma(F^! \wedge^m T^*N)$ is the pullback section of $F^! \wedge^m T^*N$ associated to η, i.e. the smooth map associating to every point $x \in M$ the covector $\eta_{F(x)}$. Using this result, one can extend the pullback as a graded commutative algebra morphism $F^* : \Omega^\bullet(N) \longrightarrow \Omega^\bullet(M)$ from the following identity:

$$F^*(\eta \wedge \mu) = F^*(\eta) \wedge F^*(\mu) \tag{3.5}$$

for any m-form η and p-form μ on N. For a proof of this statement see Lemma 12.10 in [Lee, 2003]. Then, the pullback somehow defines a dual version of a smooth map:

Geometry		Algebra
M	$\longleftrightarrow$	$\mathcal{C}^\infty(M)$
$F : M \longrightarrow N$	$\longleftrightarrow$	$F^* : \Omega^\bullet(N) \longrightarrow \Omega^\bullet(M)$

In this correspondence the pullback is actually characterized by the following algebraic property:

Proposition 3.33. *The pullback $F^* : \Omega^\bullet(N) \longrightarrow \Omega^\bullet(M)$ of the smooth map $F : M \longrightarrow N$ is a morphism of differential graded commutative algebras from $(\Omega^\bullet(N), d_N)$ to $(\Omega^\bullet(M), d_M)$. In particular, it commutes with the respective de Rham differentials d_M on M and d_N on N:*

$$d_M \circ F^* = F^* \circ d_N$$

Proof. The fact that F^* is a morphism of graded commutative algebra is transparent in Equation (3.5). For $m = 0$, let $f \in \mathcal{C}^\infty(N)$ and let X be a vector field on M. Then, one has:

$$F^*(d_N f)(X) = F^! d_N f(F_*X) = F_*X(f) = X(f \circ F) = X(F^*f) = d_M F^*(f)(X)$$

where the third term is an explicitation of the second, as the action of the section $F_*X \in \Gamma(F^!TN)$ on f is understood to be the expected one: $x \longmapsto X_x(f \circ F(x))$. Now let $m \geq 1$, let $\eta \in \Omega^m(N)$ be a differential m-form on N, and let $X_1, \ldots, X_{m+1}$ be m vector fields on

M. One can easily check that on their respective pullback bundles, $F^! d_N \eta = d_N F^! \eta$ and $[F_* X_i, F_* X_j] = F_*[X_i, X_j]$. From this, we deduce:

$$
\begin{aligned}
\big(F^*(d_N\eta)\big)(X_1,\dots,X_{m+1}) &= F^! d_N\eta(F_*X_1,\dots,F_*X_{m+1}) = d_N F^!\eta(F_*X_1,\dots,F_*X_{m+1}) \\
&= \sum_{i=1}^{m+1} (-1)^{i-1} F_*X_i\big(F^!\eta(F_*X_1,\dots,\widehat{F_*X_i},\dots,F_*X_{m+1})\big) \\
&\quad + \sum_{1\le i<j\le m+1} (-1)^{i+j-1} F^!\eta\big(\underbrace{[F_*X_i,F_*X_j]}_{=F_*[X_i,X_j]}, F_*X_1,\dots,\widehat{F_*X_i},\dots,\widehat{F_*X_j},\dots,F_*X_{m+1}\big) \\
&= \sum_{i=1}^{m+1} (-1)^{i-1} X_i\big(F^*(\eta)(X_1,\dots,\widehat{X_i},\dots,X_{m+1})\big) \\
&\quad + \sum_{1\le i<j\le m+1} (-1)^{i+j-1} F^*(\eta)\big([X_i,X_j],X_1,\dots,\widehat{X_i},\dots,\widehat{X_j},\dots,X_{m+1}\big) \\
&= d_M F^*(\eta)
\end{aligned}
$$

□

Since pullback goes the other way around compared to F, contrary to pushforwards, composition of pullbacks is not the pullback of the composite maps:

Lemma 3.34. *Given two smooth functions $F : M \longrightarrow N$ and $G : N \longrightarrow P$, the pullback of the composite $G \circ F : M \longrightarrow P$ reverts the order:*

$$(G \circ F)^* = F^* \circ G^*$$

Remark 3.35. The correspondence between geometry and algebra can be further exploited to describe Lie algebroid morphisms. Without entering into much details, a morphism of Lie algebroids $\phi : A \longrightarrow B$ is a vector bundle morphism that is additionally a Lie algebra morphism on sections, and which is compatible with the anchor map. This complicated condition can be equivalently stated as the following: a Lie algebroid morphism is a morphism of differential commutative graded algebra $\Phi : (\Omega^\bullet(B), d_B) \longrightarrow (\Omega^\bullet(A), d_A)$, where $(\Omega^\bullet(A), d_A)$ is the so-called *Lie algebroid cohomology*. This one-to-one correspondence was originally found by Vaintrob [Vaintrob, 1997], and is still valid for higher Lie algebroids.

Pushforwards and pullbacks allow to define smooth local frames on the tangent and cotangent bundle. Let $x \in M$ and let (U, φ) be a trivializing chart of the tangent bundle (and then, by duality of the fiber, of the cotangent bundle as well) centered at x. We denote by $x^1, \dots, x^n$ the standard coordinates on $\widetilde{U}$ centered at 0 (because $\varphi(x) = 0$), and by abuse of notation they also denote the composite function $x^i \circ \varphi$. Then one can define:

1. a local smooth frame of TM over U from the constant vector fields $\frac{\partial}{\partial x^i}$ on $\widetilde{U} = \varphi(U)$, via the push-forward of $\varphi^{-1} : \widetilde{U} \longrightarrow U$. This time the pushforward is well defined because $\varphi : U \longrightarrow \widetilde{U}$ is a diffeomorphism. For brevity, we denote the induced local smooth frame on U by the same notation $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$ and we call it the *coordinate frame*;

2. a local smooth frame of T^*M over U from the constant covector fields dx^i on $\widetilde{U} = \varphi(U)$ via the pull back of φ. We denote this local smooth frame on U by the same notation $dx^1, \dots, dx^n$ and we call it the *coordinate coframe*.

Both frames are well-defined because they are constant sections. In order to differentiate the frame on $U \subset M$ and the one on $\widetilde{U} \subset \mathbb{R}^n$, we write $\left.\frac{\partial}{\partial x^i}\right|_y$ (resp. $dx^i|_y$) to indicate the former, and $\frac{\partial}{\partial x^i}\Big|_{\varphi(y)}$ (resp. $dx^i|_{\varphi(y)}$) to indicate the latter, for every $y \in U$. A manifold that admits a smooth global frame for the tangent bundle is said *parallelizable*. The only spheres that are parallelizable are $\mathbb{S}^1$, $\mathbb{S}^3$ and $\mathbb{S}^7$.

Example 3.36. For every $1 \leq m \leq n$, the exterior algebra of the cotangent space at each point defines a smooth vector bundle:

$$\wedge^m T^*\mathbb{R}_n = \bigsqcup_{x \in \mathbb{R}^n} \wedge^m T^*_x \mathbb{R}^n$$

Then a smooth local frame consists of the sections $dx^{i_1} \wedge \ldots \wedge dx^{i_m}$ for $1 \leq i_1 < \ldots < i_m \leq n$. It is not a global frame because the coordinates functions x^i are only defined locally.

Now let us understand how the coordinate functions of vector fields and of differential forms transform under a change of local coordinates. Assume that there exists another compatible chart (V, ψ) centered at x so that $\widetilde{V} = \psi(V)$ and $x'^1, \ldots, x'^n$ are the standard coordinates on $\widetilde{V}$ centered at 0. Then, under the change of coordinates $\psi \circ \varphi^{-1} : \widetilde{U} \longrightarrow \widetilde{V}$, the constant sections $\frac{\partial}{\partial x^i}$ transform as:

$$\begin{aligned} \frac{\partial}{\partial x^i}\Big|_y &= (\varphi^{-1})_* \frac{\partial}{\partial x^i}\Big|_{\varphi(y)} && (3.6) \\ &= (\psi^{-1})_* \circ (\psi \circ \varphi^{-1})_* \frac{\partial}{\partial x^i}\Big|_{\varphi(y)} \\ &= (\psi^{-1})_* \left(\frac{\partial(\psi \circ \varphi^{-1})^j}{\partial x^i}(\varphi(y)) \frac{\partial}{\partial x'^j}\Big|_{\psi(y)} \right) \\ &= \frac{\partial(\psi^j \circ \varphi^{-1})}{\partial x^i}(\varphi(y)) \frac{\partial}{\partial x'^j}\Big|_y && (3.7) \end{aligned}$$

for every $y \in U \cap V$. We pass from the first line to the second line by using Lemma 3.30, and from the second line to the third by Equation (3.3). Then the push-forward $(\psi^{-1})_*$ is linear so that we obtain the fourth line.

Remark 3.37. Usually the term $\frac{\partial(\psi^j \circ \varphi^{-1})}{\partial x^i}(\varphi(y))$ is denoted $\frac{\partial x'^j}{\partial x^i}(y)$, because it is transparent and for practical purposes. We will pick up this convention from then on.

A vector field $X \in \mathfrak{X}(U \cap V)$ decomposes as $X^i \frac{\partial}{\partial x^i}$ with respect to the coordinate functions x^i of φ, and $X'^j \frac{\partial}{\partial x'^j}$ with respect to the coordinate functions x'^i of ψ. Then Equations (3.6)-(3.7) show that:

$$X'^j_y = \frac{\partial x'^j}{\partial x^i}(y) X^i_y$$

We observe that under a change of coordinates $x^i \longmapsto x'^i$, the coordinate functions of the vector field transform in the opposite way than the constant sections $\frac{\partial}{\partial x^i}$:

$$\begin{array}{ccc} \frac{\partial}{\partial x^i}\Big|_y & \longrightarrow & \frac{\partial}{\partial x'^i}\Big|_y = \frac{\partial x^j}{\partial x'^i}(y) \frac{\partial}{\partial x^j}\Big|_y \\ X^i_y & \longrightarrow & X'^i_y = \frac{\partial x'^j}{\partial x^i}(y) X^i_y \end{array}$$

The first line has been obtained from Equations (3.6)-(3.7) by inverting the Jacobian matrix $\frac{\partial x'^j}{\partial x^i}$. Since the coordinate functions of vector fields transform in the opposite way than the way in which the canonical frame of the tangent bundle transforms, we say that these coordinates are *contravariant*. Changes of coordinates impact also the way differential forms transform, since for any covector field $\xi = \xi_i \, dx^i = \xi'_j \, dx'^j$, by using Equations (3.6)-(3.7), one has:

$$\xi_{y,i} = \xi_y \left(\frac{\partial}{\partial x^i}\bigg|_y \right) = \frac{\partial x'^j}{\partial x^i}(y)\, \xi_y \left(\frac{\partial}{\partial x'^j}\bigg|_y \right) = \frac{\partial x'^j}{\partial x^i}(y)\, \xi'_{y,j}$$

Here, we consider that the coordinates x^i and x^j are those on $U \cap V$. Thus, we observe that the coordinate functions of differential 1-forms transform in the same way as the constant sections $\frac{\partial}{\partial x^i}$:

$$\begin{array}{ccc} \dfrac{\partial}{\partial x^i}\bigg|_y & \longrightarrow & \dfrac{\partial}{\partial x'^i}\bigg|_y = \dfrac{\partial x^j}{\partial x'^i}(y) \dfrac{\partial}{\partial x^j}\bigg|_y \\ \xi_{y,i} & \longrightarrow & \xi'_{y,i} = \dfrac{\partial x^j}{\partial x'^i}(y)\, \xi_{y,j} \end{array}$$

Since the coordinate functions of differential forms transform in the same way as the way in which the coordinate frame of the tangent bundle transforms, we say that these coordinates are *covariant*. In general the position of the indices indicates when it is a covariant (at the bottom) or a contravariant (at the top) coordinate. The names 'contravariant' and 'covariant' come from the fact that the pushforward functor, assigning to any smooth manifold its tangent bundle and to any smooth function its pushforward, is a covariant functor, while the pullback functor, assigning to any smooth manifold its algebra of functions and to any smooth function between manifolds its pullback, is a contravariant functor.

3.3 Submanifolds in differential geometry

The notion of pullback and pushforward allows to define various kinds of subspaces in a smooth manifold, that can be additionally equipped with a distinguished smooth structure that turn them into *submanifolds*. Since it is a very subtle topic, I strongly advise the reader to refer to [Lee, 2003] and to [Lee, 2009] to get a much more clear understanding of the notions discussed in the present section. There are three main kinds of submanifold objects:

$$\{\text{embedded submanifolds}\} \subset \{\text{weakly embedded submanifolds}\} \subset \{\text{immersed submanifolds}\}$$

An *immersed submanifold* of a smooth manifold M is a subset S, equipped with a smooth structure (i.e. a topology composed of smoothly compatible charts) such that the inclusion $\iota : S \longrightarrow M$ is a smooth map (with respect to the respective smooth structures on S and M) and an immersion. It does not mean that the topology on S is the subspace topology, and in general it will not be! There may exist various non-diffeomorphic smooth structures on the subset S such that the inclusion ι is an immersion. A famous example of an immersed manifold is the figure eight:

Example 3.38. Let $\gamma :]-\pi, \pi[\longrightarrow \mathbb{R}^2$ be the smooth map defined as:

$$\gamma(t) = (\sin(2t), \sin(t))$$

The image of γ, denoted S, is the locus of points (x, y) defined by $x^2 = 4\, y^2(1 - y^2)$. This subset can be equipped with a topology of open sets defined as follows: a subset $U \subset \mathrm{Im}(\gamma)$ is open if and only if $\gamma^{-1}(U)$ is open in the topology of N. This implies in particular that any subset of the form $(\sin(2t), \sin(t))$ for $t \in]-\epsilon, \epsilon[$ is an open set of S. The map γ^{-1} then turn these open sets into smooth open charts, that are smoothly compatible by construction. Then, S is a smooth manifold, but its smooth structure does not descend from the smooth structure on M for the following reason: in the subspace topology, a neighborhood of 0 in S has the shape of a cross, and is not homeomorphic to any region of euclidean space, while in the manifold topology, there exist neighborhoods of the origin that are homeomorphic to an open one-dimensional segment. The inclusion map $\iota : S \longrightarrow M$ being an immersion since $\dot{\gamma}(t)$ never vanishes, the subset S equipped with its smooth manifold structure is an immersed manifold of M.

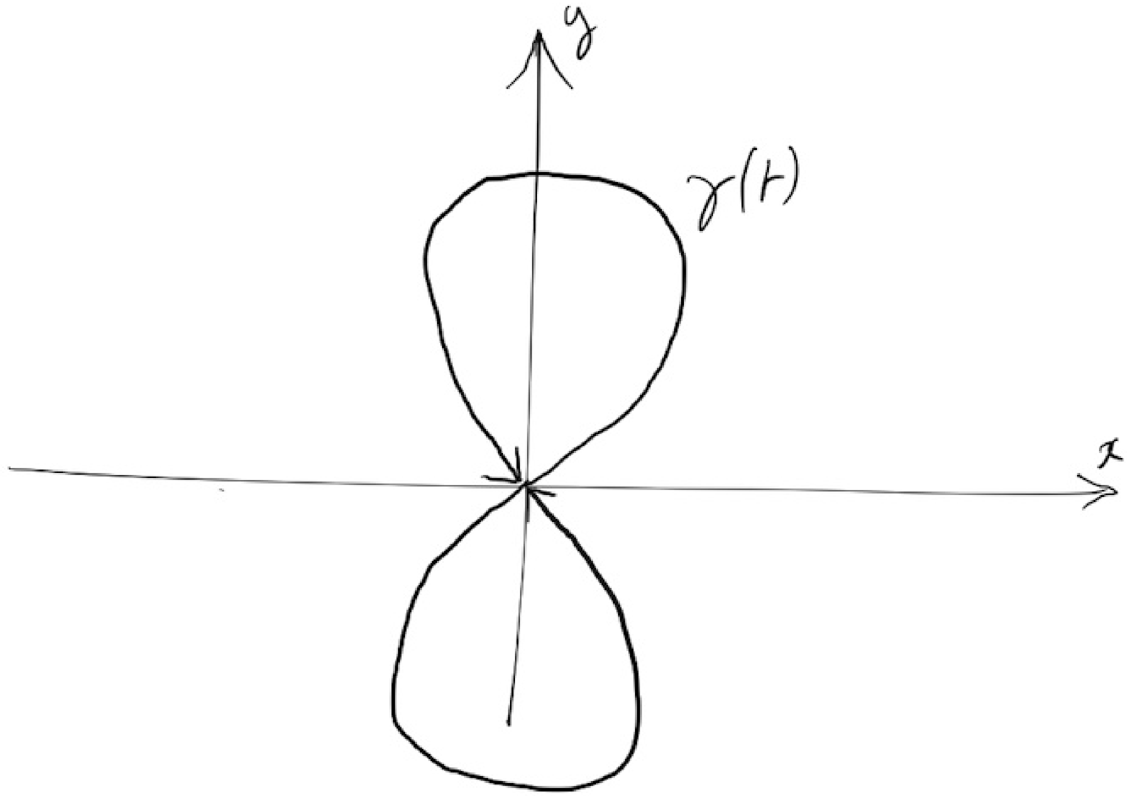

Figure 10: The image of the path γ is a subset of $\mathbb{R}^2$ that has the shape of a 'eight'. In particular it is not simply connected, and at the origin it looks like a crossroad (as a set), although as a topological space, an open neighborhood of the origin is an open set of dimension 1, of the form $\gamma(]-\epsilon,\epsilon[)$.

Example 3.39. Another example consists of any irrational curve on the 2-torus: pick up an irrational number $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ and define S to be the subset of $\mathbb{T}^2$ induced by the slope of slope α (it can be understood visually by using the well-known identification between $\mathbb{T}^2$ and a square). The topology of the irrational curve is so that connected open line segments are open, while the subspace topology does not allow this because the irrational curve is dense in $\mathbb{T}^2$. Thus, the submanifold smooth structure does not descend from the ambient smooth structure so the submanifold is not embedded but merely immersed.

Let us now turn to *embedded* (or *regular*) *submanifolds* of a smooth manifold M: they are subsets S such that the subspace topology of M defines a canonical smooth structure on S (see Theorem 8.2 in [Lee, 2003]). This property is a consequence of the fact that these submanifolds are modeled locally on the standard embedding of $\mathbb{R}^k$ into $\mathbb{R}^n$. More precisely, let $\widetilde{U}$ be an open subset of $\mathbb{R}^n$, then a *k-slice of $\widetilde{U}$* is any subset of the form:

$$\left\{(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n) \in \widetilde{U} \;\middle|\; x^{k+1} = c^{k+1}, \ldots, x^n = c^n\right\}$$

for some constants $c^{k+1}, \ldots, c^n$. Clearly any k-slice is homeomorphic to an open subset of $\mathbb{R}^k$. Let M be a smooth manifold, and let (U, φ) be a smooth chart on M. If S is a subset of U such that $\varphi(S)$ is a k-slice of $\widetilde{U} = \varphi(U)$, then we say simply that S is a k-slice of U. A subset $S \subset M$ is called a *k-dimensional embedded submanifold of M* if for each point $x \in S$, there exists a smooth chart (U, φ) for M such that $x \in U$, and $U \cap S$ is a k-slice of U. Equivalently, S is a k-dimensional embedded manifold in M if every point $x \in S$ is in the domain of a coordinate chart (U, φ) such that:

$$\varphi(U \cap S) = \varphi(U) \cap \{\mathbb{R}^k \times 0\} \tag{3.8}$$

The definition of embedded submanifolds is a local one, so that we can summarize it under the following Lemma:

Lemma 3.40. *Let M be a smooth manifold and let S be a subset of M. Suppose that for some k, every point $x \in S$ has a neighborhood $U \subset M$ such that $U \cap S$ is an embedded k-submanifold of U. Then S is an embedded k-submanifold of M.*

Example 3.41. The figure eight (Figure 10) is not an embedded submanifold because, although any point of the figure eight outside the origin belong to a 1-dimensional slice, there is no open set $U \subset \mathbb{R}^2$ containing the origin such that the intersection of U and the figure eight is an embedded submanifold of U (i.e. a one-dimensional slice of U).

Example 3.42. Let M and N be smooth manifolds of dimensions n and k, respectively, and let $F : M \longrightarrow N$ be a smooth map. Let us call the *graph of F* the following subset of $\mathbb{R}^k \times \mathbb{R}^n$:

$$\mathrm{Gr}(F) = \left\{(y, x) \in \mathbb{R}^k \times \mathbb{R}^n \mid y = F(x)\right\}$$

Indeed, Lemma 8.6 in [Lee, 2003] shows that locally this graph is embedded. Hence, by Lemma 3.40, it is an embedded submanifold.

A nice characterization of embedded submanifolds is obtained through the observation that the slice property carried by embedded submanifolds is equivalent to being locally the level set of a submersion:

Proposition 3.43. Constant rank level set theorem *Let M and N be smooth manifolds, and let $F : M \longrightarrow N$ be a smooth map with constant rank equal to k. Each level set of F is a closed embedded submanifold of codimension k in M. In particular, a subset S of M is an embedded submanifold of M of codimension k if and only if every point $x \in S$ has a neighborhood U in M such that $U \cap S$ is a level set of a submersion $U \longrightarrow \mathbb{R}^k$.*

Proof. See Chapter 8 in [Lee, 2003]. □

A straightforward and very useful relies on the following notions. If $F : M \longrightarrow N$ is a smooth map, a point $x \in M$ is said to be a *regular point of F* if the push-forward $F_* : T_xM \longrightarrow T_{F(x)}N$ is surjective; it is a *critical point* otherwise. A point $y \in N$ is said to be a *regular value of F* if every point of the level set $F^{-1}(y)$ is a regular point, and a *critical value* otherwise. Finally, a level set $F^{-1}(y)$ is called a *regular level set* if y is a regular value; in other words, a regular level set is a level set consisting entirely of regular points. Than, one has:

Theorem 3.44. Regular level set theorem *Every regular level set of a smooth map is a closed embedded submanifold whose codimension is equal to the dimension of the range.*

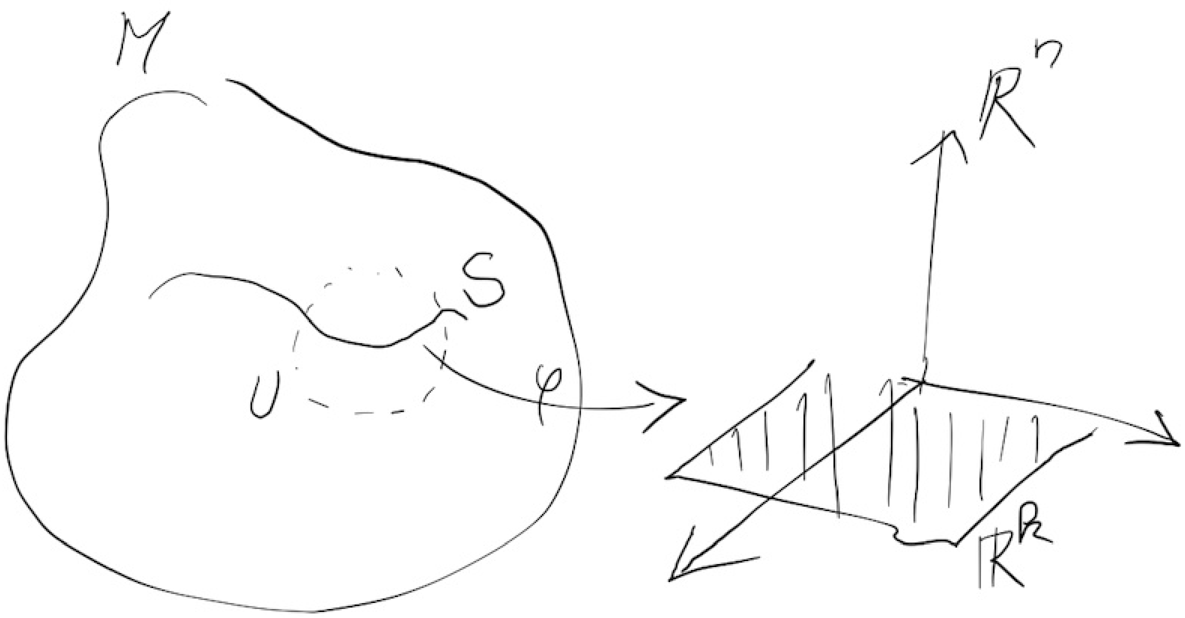

Figure 11: The image of S through the map $\varphi : U \longrightarrow \mathbb{R}^n$
is an open subset of $\mathbb{R}^k \times \{0\}$.

Proof. This is Corollary 8.10 in in [Lee, 2003]. □

Example 3.45. An alternative argument to show that the figure eight is not an embedded submanifold is that the figure eight is the zero level set of the smooth function:

$$\begin{aligned} F:\ \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x,y) &\longmapsto x^2 - 4\,y^2(1-y^2) \end{aligned}$$

This function does not satisfies the latter part of Proposition (3.43) because at $(0,0) \in S$ it is not a submersion.

Example 3.46. Let $\mathbb{R}^n$ be the *configuration space*, with the corresponding coordinates q^i. We call the cotangent bundle $P = T^*\mathbb{R}^n$ the *phase space*, with coordinates q^i and p_i (the latter are linear forms on the fibers). Then a *constraint* is a smooth function $\phi : P \longrightarrow \mathbb{R}$. The 0-level locus of a set of constraints $\phi_1, \dots, \phi_r$ is a subset Σ of P that we call the *constraint surface*. Usually, physicists assume that the constraints satisfy a so-called *regularity condition* that often take the form that for each point $x \in \Sigma$ there exists an open neighborhood U such that only r' constraints $\phi_{i_1}, \dots, \phi_{i_{r'}}$ are functionally independent over U, making $\Sigma \cap U$ an embedded submanifold of U of codimension r' (as a level set of the constant rank smooth map $(\phi_{i_1}, \dots, \phi_{i_{r'}}) : P \longrightarrow \mathbb{R}^{r'}$). Then by Proposition 3.43, the constraint surface is an embedded submanifold of dimension $2n - r'$. For more details, see Chapters 1 and 2 of [Henneaux and Teitelboim, 1994].

Associated to immersed submanifolds and embedded submanifolds, there exist corresponding notions of maps: injective immersions and smooth embeddings. An *injective immersion* between two smooth manifolds S and M is an injective smooth map $F : S \hookrightarrow M$ that is additionally an immersion, i.e. such that the pushforward $F_* : TS \longrightarrow TM$ is injective (we can consider that F_* takes values in TM because F is injective). In particular, an injective smooth map is an immersion if and only if it has constant rank. A *topological embedding* $F : S \hookrightarrow M$ is a continuous map that is a homeomorphism onto its image, where the topology on the image $F(S)$ is the subspace topology induced from the smooth atlas on M. A *smooth embedding* is a topological embedding that is smooth and of constant rank (then it is automatically an immersion). Obviously not every injective immersion is a smooth embedding (not even on its image), however here are two cases where it happens:

1. S is compact
2. F is proper (i.e. $F^{-1}(K)$ is compact if and only if $K \subset M$ is compact)

because in both cases the map $F : S \hookrightarrow M$ is closed (see Proposition 7.4 in [Lee, 2003]). Moreover, immersions locally behave as smooth embeddings, but not globally (hence justifying that the figure eight is the image of an immersion and not an embedding). See Lemma 8.18 in [Lee, 2003] for more details.

Proposition 3.47. *Immersed submanifolds are precisely the images of injective immersions and embedded submanifolds are precisely the images of smooth embeddings.*

Proof. See Chapter 8 in [Lee, 2003]. □

Exercise 3.48. Define the following three open subsets of $\mathbb{R}$:

$$A_- = \Big] -\infty, -\frac{\pi}{2}\Big[, \qquad A_0 = \Big] -\frac{\pi}{2}, +\frac{\pi}{2}\Big[, \qquad A_+ = \Big] +\frac{\pi}{2}, +\infty\Big[$$

Denote by A their disjoint union so, in particular, $A = \mathbb{R} - \{-\frac{\pi}{2}, \frac{\pi}{2}\}$. Let $f : A \longrightarrow \mathbb{R}^2$ be the smooth map defined on each subset as follows:

$$f|_{A_-} = \left(-e^{x-\frac{1}{x+\frac{\pi}{2}}}, e^{x-\frac{1}{x+\frac{\pi}{2}}}\right), \quad f|_{A_0} = (\tan(x), \tan(x)), \quad f|_{A_+} = \left(e^{-x+\frac{1}{x-\frac{\pi}{2}}}, -e^{-x+\frac{1}{x-\frac{\pi}{2}}}\right)$$

Prove that f is an injective immersion (with respect to the standard smooth structures on A and $\mathbb{R}^2$). Draw a conclusion about the image of f, and determine the tangent space of $\mathrm{Im}(f)$ at the point $(0, 0)$.

Now let us study in more details the difference between immersed and embedded submanifolds. Notice that if one had chosen another parametrization for the figure eight in Example 3.38, we would have inherited a totally different topology. Another, alternative smooth map defining the figure eight (as a set) can be chosen to be :

$$\eta(t) = \big(-\sin(2t), \sin(t)\big)$$

and the path corresponding to η would be the symmetric image of that with respect to γ with respect to the vertical axis (see Figure 10). The open sets would not be the same either, because for example the image of $\gamma(]-\epsilon, \epsilon[)$, although a connected open set with respect to the topology induced by γ, would not be open in the topology induced by η, for its preimage would consists of two disjoint intervals, and the origin $t = 0$ (closed point). Then it seems that, although the subset S is uniquely defined as the level sets of the points (x, y) satisfying the equation $x^2 = 4y^2(1 - y^2)$, it admits several – a priori non-equivalent – smooth structures. The above argument shows that the map $\eta^{-1} \circ \gamma :] - \pi, \pi[\longrightarrow] - \pi, \pi[$ is not a smooth map, not even a continuous one. However, if one would have a diffeomorphism ψ from $] - \pi, \pi[$ such that $\eta = \gamma \circ \psi$, we would certainly conclude that the two smooth structures on S can be considered as 'equivalent'. This is not the case, but this equivalence property is worth extending to every immersed submanifolds.

Definition 3.49. *Immersed submanifolds* $N_1 \overset{\varphi_1}{\hookrightarrow} M$ *and* $N_2 \overset{\varphi_2}{\hookrightarrow} M$ *are called* equivalent *when there exists a diffeomorphism* $\psi : N_1 \longrightarrow N_2$ *making the following diagram commutative:*

$$\begin{tikzcd}
N_1 \arrow[dr, "\varphi_1"] \arrow[dd, "\psi"'] & \\
& M \\
N_2 \arrow[ur, "\varphi_2"'] &
\end{tikzcd}$$

This is an equivalence relation on the set of immersed submanifolds of M, and thus each equivalence class has a unique representative $(S, \mathcal{A}, \iota)$ where S is a subset of M with a given smooth structure $\mathcal{A}$ such that the inclusion ι is an immersion. We emphasized the presence of the maximal atlas $\mathcal{A}$ because it will turn out to be central in the discussion. For example, as seen above, there are two possible atlases for the figure eight to be an immersed submanifold, which are not equivalent because the map $\eta \circ \gamma^{-1}$ is not continuous. Thus, the figure eight admits several non-equivalent smooth structures making it an immersed submanifold. More generally now assume that there exist two injective immersions $N_1 \overset{\varphi_1}{\hookrightarrow} M$ and $N_2 \overset{\varphi_2}{\hookrightarrow} M$ whose images coincide $S = \mathrm{Im}(\varphi_1) = \mathrm{Im}(\varphi_2)$. What is the condition on φ_1 and φ_2 for N_1 and N_2 to be equivalent? Obviously, if φ_1 and φ_2 are smooth embeddings (i.e. if S is an embedded submanifold), then a diffeomorphism between N_1 and N_2 satisfying the commutative triangle is $\varphi_2^{-1} \circ \varphi_1$. This solution work for smooth embeddings because they have the following property:

Definition 3.50. *Let* N *and* M *be smooth manifolds. A smooth map* $F : N \longrightarrow M$ *will be called* smoothly universal *if for any smooth manifold* N' *and any smooth map* $H : N' \longrightarrow M$ *such that* $H(N') \subset F(N)$*, there exists a smooth map* $G : N' \longrightarrow N$ *making the following triangle commutative:*

$$\begin{tikzcd}
N \arrow[r, "F"] & M \\
N' \arrow[u, dashed, "G"] \arrow[ur, "H"'] &
\end{tikzcd}$$

Not every injective immersion is smoothly universal as the following discussion shows: given N, N', F and G as in the Definition (and assumming that F is an injective immersion), one naive idea would be to use F^{-1} to lift the smooth map G to H. However, the smoothness of the map H then crucially depends on the smooth structure of N and the properties of the smooth map F or, said differently, if $F(N)$ is an immersed or an embedded manifold. In the latter case, one can always define the map H as the composite $F^{-1} \circ G$, which is a smooth map because F is a diffeomorphism onto its image. However, when $F(N)$ is an immersed submanifold, although well-defined the map $H = F^{-1} \circ G$ needs not be a smooth map with respect to the smooth structure on N. For example, although the two paths γ and η define the same subset S – the figure eight – in $\mathbb{R}^2$, the map $\eta \circ \gamma^{-1}$ is not continuous. This is the content of Scholie 1.31-33 in [Warner, 1983], which contain a nice discussion on this topic. In particular Theorem 1.32 states that the lift G is smooth if and only if it is continuous, thus showing that not having the smoothly universal property has tremendous consequences. This justifies the following definition:

Definition 3.51. *An injective immersion $N \overset{\varphi}{\hookrightarrow} M$ that has the smoothly universal property is called a* weak embedding*. The image of such a map in M is called a* weakly embedded submanifold*.*

We deduce from the above discussion that smooth embeddings are weak embeddings, while injective immersions need not be. Hence the following sequence of inclusions:

$$\{\text{smooth embeddings}\} \subset \{\text{weak embeddings}\} \subset \{\text{injective immersions}\}$$

As for Proposition 3.47, *weakly embedded submanifolds* correspond to the images of weak embeddings (some authors call them *regularly immersed* submanifolds). By construction, they are immersed submanifolds, but need not be embedded submanifolds.

While the underlying set of an immersed submanifold S may admit different smooth structures making the inclusion map $\iota : S \longrightarrow M$ an immersion, weak embeddings carry a universal property making the smooth structure of a weakly embedded manifold unique, up to the equivalence given in Definition 3.49. More precisely, assume that a weakly embedded submanifold S of M is obtained via a weak embedding $\varphi_1 : N_1 \longrightarrow M$ – which is a smooth map with respect to a maximal smooth atlas $\mathcal{A}_1$, and assume moreover that S admits another weak embedding $\varphi_2 : N_2 \longrightarrow M$ with respect to a smooth structure $\mathcal{A}_2$ on N_2. Then, by the smoothly universal property of weak embeddings, both maps $\varphi_1^{-1} \circ \varphi_2 : (N_2, \mathcal{A}_2) \longrightarrow (N_1, \mathcal{A}_1)$ and $\varphi_2^{-1} \circ \varphi_1 : (N_1, \mathcal{A}_1) \longrightarrow (N_2, \mathcal{A}_2)$ are smooth. Being injective and inverse to one another, they define a diffeomorphism between $(N_1, \mathcal{A}_1)$ and $(N_2, \mathcal{A}_2)$, thus showing that the two smooth structures are equivalent in the sense of Definition (3.49):

$$\begin{tikzcd}
(N_1, \mathcal{A}_1) \arrow[dr, "\varphi_1"] \arrow[dd, bend right, "\varphi_2^{-1} \circ \varphi_1"'] & \\
& M \\
(N_2, \mathcal{A}_2) \arrow[ur, "\varphi_2"'] \arrow[uu, bend right, "\varphi_1^{-1} \circ \varphi_2"'] &
\end{tikzcd}$$

Weakly embedded submanifolds can then be considered as those submanifolds that have the right amount of regularity so that they carry only one possible smooth structure making the inclusion map an immersion. We will now give more details on their geometric properties and explain why their smooth structure – although uniquely defined by that of M – is not necessarily induced by the subspace topology. Given a subset S of a smooth manifold M and a point $x \in S$, we denote by $C_x(S)$ the path connected component of x in S, i.e. the set of points that are reachable from x by smooth curves contained entirely in S (here, a smooth curve is a smooth map $\gamma : \mathbb{R} \longrightarrow M$). Then a weakly embedded submanifold is characterized by the following property, that mimick Equation (3.8), but only at the level of path connected components:

Proposition 3.52. *Let M be a smooth manifold and let S be a weakly embedded submanifold of M of dimension k. Then for every $x \in S$, there exists a coordinate chart (U, φ) centered at x such that:*

$$\varphi(C_x(U \cap S)) = \varphi(U) \cap \{\mathbb{R}^k \times 0\} \tag{3.9}$$

Proof. See Propositions 3.19 and 3.20 in [Lee, 2009]. $\square$

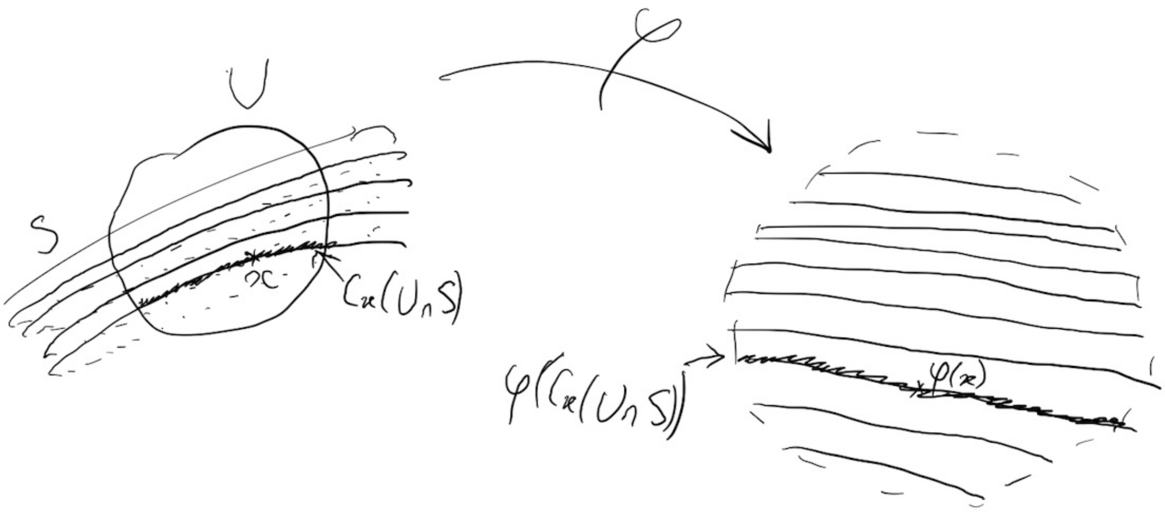

Figure 12: Assume that S is a weakly embedded submanifold of M that is additionally dense in M. Then, although $S \cap U$ consists of an infinite number of disjoint 'plaques', the path component of $x \in S$ in U is connected (by definition). Then its image through $\varphi : U \longrightarrow \mathbb{R}^n$ is an open subset of $\mathbb{R}^k \times \{0\}$.

Once again, we see why the figure eight is not even a weakly embedded submanifold: at the origin, the path-connected component forms a cross shaped set, which does not have the slice property of Equation (3.9). An example of a weakly embedded manifold which is not an embedded manifold is any leaf of the Kronecker foliation of the torus:

Example 3.53. Let $\mathbb{T} = \mathbb{R}^2 \big/ \mathbb{Z}^2$ be the torus and let $X = \frac{\partial}{\partial x} + \alpha \frac{\partial}{\partial y}$ be a vector field on $\mathbb{T}$ such that $\alpha \in \mathbb{R} - \mathbb{Q}$. Then the integral curve of X through any point (x, y) is a dense subset of $\mathbb{T}$, that is a weakly embedded submanifold. Indeed, any open neighborhood U of $(x, y) \in \mathbb{T}$ intersects infinitely many times the leaf through (x, y) (because it is dense). However, the set of points of U which are path-connected to (x, y) satisfy Equation (3.9) when U is taken to be sufficiently small. More generally it has been shown by Štefan in 1974 that leaves of (possibly singular) foliations are weakly embedded submanifolds [Stefan, 1974].

Let us conclude this section by a rather useful result, which is a variation of Proposition 3.52 for immersed submanifolds. Although immersed submanifold do not admit the local structure of embedded or weakly embedded submanifold as a level set of a constant rank smooth map, there exist local distinguished coordinates characterizing open sets of immersed submanifolds:

Proposition 3.54. *Let $N \xhookrightarrow{F} M$ be an immersed submanifold of M and let $x \in N$. Then there exists a connected open neighborhood V of x in N and a coordinate chart (U, φ) centered at $F(x)$ such that:*

$$\varphi(U \cap F(V)) = \varphi(U) \cap (\mathbb{R}^k \times \{0\}) \tag{3.10}$$

Proof. The proof can be found in the discussion on page 131 of [Lee, 2009] and complemented by Proposition 1.35 of [Warner, 1983]. □

Condition (3.10) emphasizes that the image in M of some connected open neighborhood of every point of S is embedded in M. Notice the difference with Lemma 3.40 which characterize embedded submanifolds. Also notice the difference between Equation (3.10) and the one for weakly embedded submanifolds (3.9) and for embedded submanifolds (3.8). We see that in each case the condition is stronger and stronger as we climb the hierarchy of submanifolds:

$$\{\text{embedded submanifolds}\} \subset \{\text{weakly embedded submanifolds}\} \subset \{\text{immersed submanifolds}\}$$

While conditions (3.8) and (3.9) are necessary and sufficient conditions to define embedded and weakly embedded submanifolds (see Propositions 3.19 and 3.20 in [Lee, 2009]), condition (3.10) does not characterize immersed submanifolds, as it is a particular case of the so-called *rank theorem* (see theorem 7.12 in [Lee, 2003]). However if the function F is injective and has constant rank then it implies that it is an immersion and then that the image is an immersed submanifold.

3.4 Distributions and foliations

Submanifolds possess their own tangent bundles, but it is often useful to see them as sub-bundles of the tangent bundle of M. That is why we benefit from the fact that every submanifold – be it immersed, weakly embedded or embedded – is the image of an immersion, to identify the tangent space to a submanifold $S \subset M$ at $x \in S$ with the image of the tangent space T_xS as the image in T_xM of the pushforward of the inclusion map ι_* – or the pushforward of the map $F : N \hookrightarrow M$ defining S:

$$T_{F(x)}S = F_*(T_xN)$$

Then, we often identify the tangent bundle of S (in M) with the subbundle of TM whose base is restricted to S and whose fiber is T_xS at any point $x \in S$. This subbundle satisfies the following nice characterization:

$$T_xS \subset \Big\{X_x \in T_xM \,\Big|\, X_x(f) = 0 \text{ whenever } f \in C^\infty(M) \text{ and } f|_S \equiv 0\Big\}$$

where the inclusion is an equality (at least) when S is an embedded submanifold (see Proposition 8.5 in [Lee, 2003] for a demonstration). However, by Proposition 3.54, one observes that there exists an open neighborhood V of x in S such that:

$$T_xS = \Big\{X_x \in T_xM \,\Big|\, X_x(f) = 0 \text{ whenever } f \in C^\infty(M) \text{ and } f|_{F(V)} \equiv 0\Big\}$$

This equality can be explained by the fact that $F(V)$ is an embedded submanifold of M, and that $T_xS = T_xF(V)$.

Example 3.55. Although the origin in the figure eight is located at the crossroad of two one-dimensional paths, the tangent space at the origin of the figure eight is considered to be one dimensional, since it is the pushforward of $T]-\pi,\pi[$ through γ_*.

Example 3.56. The inclusion may not hold for weakly embedded submanifolds, as the example for the Kronecker foliation shows: since every leaf is dense in $\mathbb{T}$, the only function f that vanish on the leaf passing through x is the zero function, and hence every tangent vector at x satisfies $X_x(f) = 0$. The tangent space to the leaf is one dimensional, hence strictly included into $T_x\mathbb{T}$. However, if one had replaced the condition $f|_S \equiv 0$ by $f|_{C_x(S)} \equiv 0$ – for any smooth function $f \in \mathcal{C}^\infty(U)$, for any arbitrary open neighborhood U of x – then the inclusion would have been an equality for weakly embedded submanifolds, but still not for immersed submanifolds (think of the figure eight).

The tangent bundle of a submanifold $S \subset M$ is a subbundle of the tangent bundle of M, when the base is restricted to S: $TS \subset T_SM$. However, assume now that we have a subbundle of TM defined over the entire manifold M. Then we expect that, under some circumstances, there may exist a family of 'parallel' submanifolds whose tangent bundles are precisely these subbundle.

Definition 3.57. *A* (smooth) distribution *on M is a smooth assignment*[12]*, to every point $x \in M$, of a vector subspace D_x of the tangent space T_xM. We say that the distribution D is* regular *if the function $x \longmapsto \dim(D_x)$ is constant over M – in that case D forms a subbundle of TM – and it is said* singular *(or* generalized*) otherwise. We say that the distribution is* involutive *if the sheaf of smooth sections of D is stable under the Lie bracket of vector fields:*

$$\forall\ X, Y \in \Gamma(D) \qquad [X, Y] \in \Gamma(D)$$

An integral manifold of D *is an immersed submanifold S such that $T_xS = D_x$ for every $x \in S$. A distribution D is said* integrable *if through each point of M passes an integral manifold of D.*

Remark 3.58. Sometimes people define integral manifolds to be those submanifolds that satisfy the following inclusion $T_xS \subset D_x$, and then define a *maximal integral manifold of D* to be an integral manifold that is maximal with respect to inclusion; in particular, which satisfies the equality $T_xS = D_x$. On the other hand, an *invariant manifold of D* would be an immersed submanifold S such that $D_x \subset T_xS$ for every $x \in S$. The name 'invariant' comes from the fact that S is invariant under the action of the flows of sections of D.

Remark 3.59. Notice that the function $x \longmapsto \dim(D_x)$, as a map from a topological space into the integers, is lower semi-continuous, and thus, the rank of the distribution D is locally constant and, in a vicinity of any given any point x, it can only be higher than or equal to that of D_x.

Example 3.60. There exist non-smooth integrable distributions. Let $M = \mathbb{R}^2$ and let D be the distribution defined as follows:

$$D_{(x,y)} = \begin{cases} \{0\} & \text{if } x \neq 0 \\ \langle \partial_y \rangle & \text{if } x = 0 \end{cases}$$

The corresponding integral manifolds are the points (x, y) when $x \neq 0$ and the vertical axis. The distribution is not smooth because there is no way of extending – as a smooth section of D – a non-trivial tangent vector defined at the origin $(0, 0)$ to a small neighborhood because the distribution outside the vertical axis is trivial. Although the distribution D is integrable, we do not consider it forms a *singular foliation* because it does not satisfy the axioms that we will soon present.

An integrable regular (resp. singular) distribution corresponds to what is commonly known as a *regular* (resp. *singular*) *foliation*. We do not want to enter the wide area of foliation theory for now, so we stick to the regular case and to regular distributions. The following definition should certainly be sufficient to understand the basic idea: a *foliation atlas of codimension p* on M (where $0 \leq p \leq n$) is an atlas made of charts call *foliation charts* and that are such that:

1. the image of the domain of any foliation chart (U, φ) through φ decomposes as a product of connected open sets $\varphi(U) = \widetilde{U}' \times \widetilde{U}'' \subset \mathbb{R}^{n-p} \times \mathbb{R}^p$

2. the transition function between two foliation charts (U, φ) and (V, ψ) is of the form:

$$\psi \circ \varphi^{-1}(a, b) = \big(g(a, b), h(b)\big) \in \mathbb{R}^{n-p} \times \mathbb{R}^p \tag{3.11}$$

   where $g : \mathbb{R}^n \longrightarrow \mathbb{R}^{n-p}$ and $h : \mathbb{R}^p \longrightarrow \mathbb{R}^p$ are smooth maps.

Thus the domain U of each foliation chart (U, φ) is partitioned into the connected components of the submanifolds $\varphi^{-1}(\mathbb{R}^{n-p} \times y)$, $y \in \mathbb{R}^p$, called *plaques*. Being the connected components of the level sets of a smooth map of constant rank, plaques are connected embedded submanifolds of M

[12]Here, smooth means that for every tangent vector $X_x \in D_x$, it is always possible to find a locally defined vector field X such that for every point y in a neighborhood of x, $X_y \in D_y$.

of dimension $n-p$. The change-of-charts diffeomorphisms defined in Equation (3.11) preserve the plaques. Then, the union of plaques which overlap amalgamate into an immersed (in fact, weakly embedded) submanifold of M. Those submanifolds which are maximal with respect to inclusion are called *leaves*. More precisely (and we do not have time nor space to detail it here), two points $x, y \in M$ lie on the same leaf if there exists a sequence of foliation charts $U_1, \ldots, U_k$ and a sequence of points $x = x_0, x_1, \ldots, x_k = y$ such that x_{i-1} and x_i lie on the same plaque in U_i. This defines an equivalence relation, so that the leaves of a foliated atlas of codimension p forms a partition of M by disjoint connected immersed submanifolds of dimension $n-p$. This observation justifies the following abstract definition (although one should stick to the idea that a foliation is a partition of M into leaves):

Definition 3.61. *A* foliation of codimension p *on a smooth manifold M is a choice of maximal foliation atlas on M of codimension p.*

Remark 3.62. The immersed submanifolds are actually weakly embedded [Stefan, 1974]. The charts defined in the definition are called *foliated charts* and there exists an atlas for M made of foliated charts, that are additionally compatible to one another, in the sense that the transition maps $\psi \circ \varphi^{-1}$ send slices to slices and preserve their transversal. We call such an atlas a *foliated atlas*. See [Candel and Conlon, 2000] and [Moerdijk and Mrcun, 2003] for details on foliations.

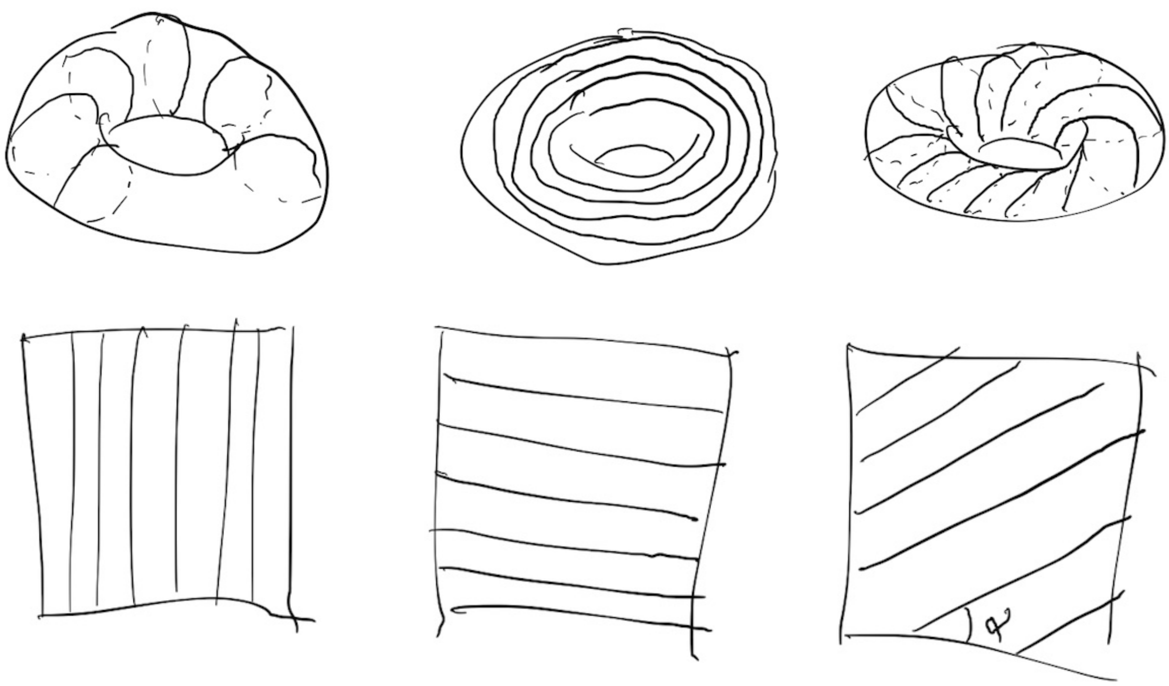

Figure 13: Three different foliations of the torus (and their corresponding equivalent representations on the flat torus): the vertical one, the horizontal one, and the last one being characterized by the slope α. When α is an irrational real number, each leaf is dense in the torus: this is the Kronecker foliation.

The relationship between distributions and foliations is that maximal (with respect to inclusion) connected integral manifolds of an integrable regular distribution form the leaves of a regular foliations:

Proposition 3.63. *Let D be an integrable regular distribution on a smooth manifold M. The collection of all maximal connected integral manifolds of D forms a foliation of M.*

This statement justifies the name 'integrable', since the regular distribution is thus integrable to a regular foliation, such that the leaves are the maximal connected integral manifolds of the distribution. However, this proposition does not tell us under which circumstances a regular distribution D is integrable. First observe that the tangent spaces to the leaves of a regular foliation define an involutive distribution. Thus an integrable distribution is necessarily involutive. The converse is actually also true, and this is the celebrated theorem of Frobenius (although he was not the first to state it):

Theorem 3.64. Frobenius Theorem *A regular distribution D on a smooth manifold is integrable (to a regular foliation) if and only if it is involutive.*

Proof. For more details on this subject, see Chapter 19 in [Lee, 2003], or Chapter 11 in [Lee, 2009], or [Candel and Conlon, 2000] and [Moerdijk and Mrcun, 2003]. □

Example 3.65. Let $\phi_1, \ldots, \phi_r$ be a set of constraints on a phase space $T^*\mathbb{R}^n$, satisfying the regularity condition of Example 3.46: the constraint surface Σ is then a $2n - r'$-dimensional embedded submanifold of $T^*\mathbb{R}^n$. The vector fields $X_i = \{\phi_i, .\}$ generate a distribution on $T^*\mathbb{R}^n$ that is regular of rank r' on the constraint surface. We say that the constraints are *first-class* if the canonical Poisson bracket on $T^*\mathbb{R}^n$ of two such constraints vanishes on the constraint surface, i.e. if we have:

$$\{\phi_i, \phi_j\} = C_{ij}{}^k\, \phi_k$$

where the $C_{ij}{}^k$ are smooth functions on $T^*\mathbb{R}^n$. If otherwise, we say that they are *second-class*. Then, a set of first-class constraints define an involutive, and then integrable, distribution on Σ. The leaves of the induced foliation are immersed (in fact, weakly embedded) submanifolds in Σ (and thus in $T^*\mathbb{R}^n$) of dimension r', and correspond to the gauge equivalent physical configurations.

Exercise 3.66. By using the Jacobi identity satisfied by the Poisson bracket, compute $[X_i, X_j]$ and show that the distribution generated by the X_i is involutive (at least) on Σ.

Now what happens when the distribution is not involutive? It means that there exist (smooth) sections X, Y of D such that their Lie bracket $[X, Y]$ is not a section of D anymore. In particular, there is a point x such that the tangent vector $[X, Y]_x$ does not belong to D_x. Taking the successive brackets of (smooth) sections of D, and evaluating them at the point x thus may generate a subspace at x that is way bigger than D_x. We set $\mathrm{Lie}\big(\Gamma(D)\big)_x$ to be the distribution corresponding to the Lie algebra generated by $\Gamma(D)$ under the successive action of the Lie bracket of vector fields on smooth sections of D:

$$\mathrm{Lie}\big(\Gamma(D)\big)_x = D_x + \mathrm{Span}\Big([X_1, X_2]_x, [[X_1, X_2], X_3]_x, [[[X_1, X_2], X_3], X_4]_x, \ldots \Big|\, X_i \in \Gamma(D)\Big)$$

Notice that it may not be a regular distribution, although interesting things happen when it is:

Definition 3.67. Hormander's condition *Let D be a distribution. We say that D is* bracket generating at x *if:*

$$\mathrm{Lie}\big(\Gamma(D)\big)_x = T_xM$$

We say that D is maximally non-integrable *if D is bracket generating at every point.*

The latter notion comes from the fact that Theorem 3.64 can be reformulated as the statement that a distribution is integrable if and only if $\mathrm{Lie}\big(\Gamma(D)\big)_x = D_x$. Then, obviously, if at some point $\mathrm{Lie}\big(\Gamma(D)\big)_x$ is strictly bigger that D_x, the distribution will not be integrable. Consequently, the situation where $\mathrm{Lie}\big(\Gamma(D)\big)_x = T_xM$ at every point can legitimately be considered as the worst case scenario where D is non-integrable in the worst possible way. However, maximally non-integrable distribution have a nice property: from the fact that if a distribution D is bracket generating at a given point x, every point in a small neighborhood of x can be reached through a so-called 'horizontal' path. A *horizontal path* is a path $\gamma : [0,1] \longrightarrow M$ that is:

1. absolutely continuous on every local coordinate chart, and
2. such that $\dot{\gamma}(t) \in D_{\gamma(t)}$ almost everywhere.

The notion of absolutely continuous paths is often met in the field of control theory under the following form: assume that $X_1, \ldots, X_m$ are smooth sections of D that are defined in a neighborhood of $\gamma([0,1])$, where γ is up to now only a continuous path. Then it is said absolutely continuous if there exist m absolutely continuous functions $u_i \in L^1([0,1])$ such that the following equation holds almost everywhere:

$$\dot{\gamma}(t) = \sum_{i=1}^{m} u_i(t)\, X_{i,\gamma(t)}$$

The functions $u_1, \ldots, u_m$ are called the controls of γ with respect to the vector fields $X_1, \ldots, X_m$. When the distribution is induced by a physical system, and that $D_x \neq T_x M$, we say that the system is *non-holonomic* – joining two points may not be possible if one restricts itself to horizontal paths only – while if $D_x = T_x M$, we say that the system is *holonomic* – one could always join one point of the state space to any other through horizontal paths. Then we have the infamous following result that answer the problem for non-holonomic systems:

Theorem 3.68. Chow-Rashevskii theorem *Let M be a smooth manifold and let D be a smooth distribution that is bracket generating at a given point $x \in M$. Then, there exists a neighborhood of x on which every point can be joined from x by an horizontal path.*

Corollary 3.69. *If D is maximally non-integrable, every two points of the manifold M can be joined through a horizontal path.*

Proof. See Section 3.2 of [Agrachev et al., 2019]. □

Exercise 3.70. Check that the distribution D of rank 2 on $\mathbb{R}^3$ generated by the vector fields $X = \frac{\partial}{\partial x^1} + x^2 \frac{\partial}{\partial x^3}$ and $Y = \frac{\partial}{\partial x^2}$ is maximally non-integrable.

The corollary give some more insight on the denomination *maximally non-integrable*: such a distribution does not have 'leaves' *per se*, and on the contrary, every two points of the manifold can be joined though an absolutely continuous path almost everywhere tangent to the distribution. We will use these notions to explain how Constantin Carathéodory defined a geometric approach to thermodynamics, and how he deduced the existence of a function called the entropy. The following discussion is mainly inspired by Chapter 22 of [Bamberg and Sternberg, 1988].

In thermodynamics, we distinguish between two kinds of physical systems: *closed systems* are those that are spatially bounded and that allow heat transfer with the exterior but no matter transfer of any kind, while *open systems* are those physical systems allowing both heat and matter transfers. Although open systems are those that are found in nature, we will restrict ourselves to closed ones, which are a very practical modelization. To every closed thermodynamical system is associated a thermodynamical state space, consisting of all its equilibrium states. Although we may assume that it is a smooth manifold (possibly with boundary), it turns out that it is often a vector space or a half space. Usually, it admits three types of coordinates: some empirical temperature θ or several, depending on the number of reservoir; some intensive

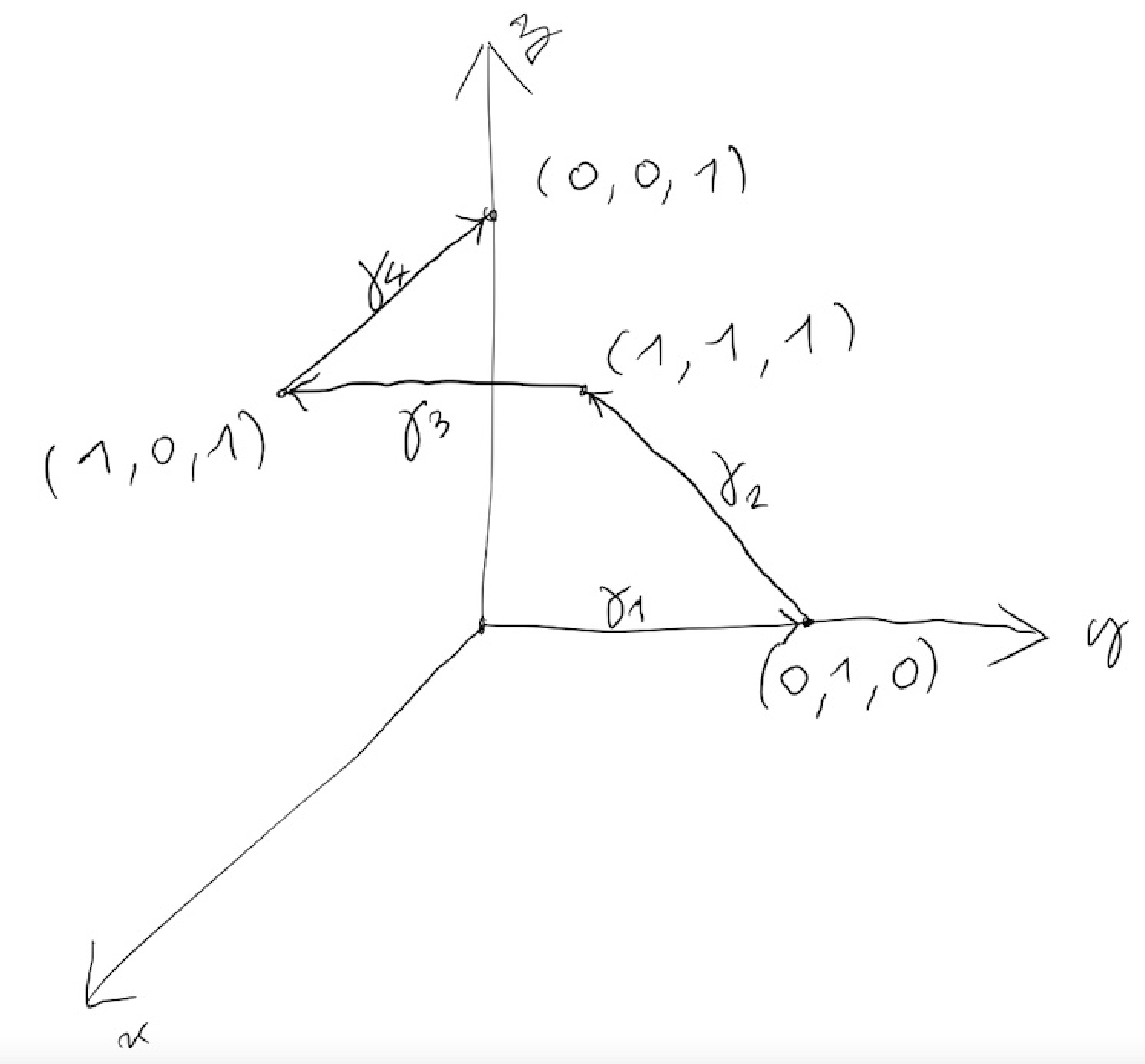

Figure 14: Although the distribution D defined in Exercise 3.70 does not contain the vertical tangent vector $\frac{\partial}{\partial x^3}$, we can however reach the point $(0, 0, 1)$ from $(0, 0, 0)$ through a sequence of paths whose tangent vectors are in D at each point.

variables corresponding to generalized force such as the pression P or a magnetic intensity; and extensive variables measuring variations of volume V or of magnetization, etc. [Zemansky, 1966]. It turns out that the existence of equations of states – such as the one relating the internal energy to the thermodynamic variables, see Scholie 3.71 – implies that the intensive variables can be made dependent on the (then independent) temperatures and on the extensive variables. We will adopt the convention that paths in the state manifold correspond to reversible thermodynamic processes.

There are mainly two kinds of thermodynamic transformations: those in which we apply some *work* W to the physical system, and those in which there is a *heat transfer* Q between the system and the exterior. The corresponding infinitesimal thermodynamic transformations are denoted δW and δQ, respectively. They are differential one-forms which, when integrated over a reversible thermodynamic process represented by a path γ, gives the total amount of work and of heat that has been exchanged:

$$Q_\gamma = \int_0^1 \delta Q(\dot\gamma(t))dt \quad \text{and} \quad W_\gamma = \int_0^1 \delta W(\dot\gamma(t))dt$$

The symbol δ not only symbolizes that the objets δW and δQ correspond to infinitesimal transformations, but also that they are *not* exact one-forms. More precisely, the quantity of work and of heat that is applied to or retrieved from the system depends on the way we apply or retrieve it (it is process dependent). One family of such processes is fundamental in thermodynamics for its usefulness: an *adiabatic process* is a thermodynamic process for which there is no heat transfer with the exterior, i.e. for which $Q_\gamma = 0$. Adiabaticity is a property that is central in Carathéodory's reformulations of the first and the second principle of thermodynamics [Sears, 1966]:

Scholie 3.71. Carathéothodory's first principle of Thermodynamics *For a closed thermodynamic system, in all adiabatic reversible thermodynamic processes between an initial state and a final state, the work does not depend on the path chosen. In particular, this implies that there exists a well-defined function U called the* internal energy *such that its infinitesimal variations satisfies:*

$$dU = \delta Q + \delta W$$

Proof. The proof that the second statement is a consequence of the first can be found in [Sears, 1963]. □

The integration of an exact one form over a path γ joining two points x and y in a simply connected space only depends on the ends points, and not on the path chosen:

$$U(y) - U(x) = \int_\gamma dU = \int_0^1 \dot\gamma(t)(U)dt$$

This makes the internal energy a *state function*, i.e. a function on the state space whose variations only depend on the initial state and the final state of the system – which is not the case for work and heat transfer. There may be different kinds of work δW and one of the most used is the one consisting of increasing or decreasing the volume of a given volume of gas, so that:

$$\delta W = -pdV + \nu_1 d\mu_1 + \ldots$$

The (certainly non-exact) differential one-form $\delta Q = dU - \delta W$ then corresponds to the infinitesimal heat production or absorption. The kernel of the differential one form $\alpha = \delta Q$ defines a distribution $D = \mathrm{Ker}(\alpha)$ such that at every point $x \in M$, $D_x = \mathrm{Ker}(\alpha_x) \subset T_xM$, and that for the sake of the presentation we will assume to be regular. This distribution has rank $n-1$ and then the question is: is it integrable or maximally non-integrable? More precisely, given the equivalence between involutivity and integrability for regular distributions, do we have $\mathrm{Lie}\big(\Gamma(D)\big)_x = D_x$ or, on the contrary, do we have $\mathrm{Lie}\big(\Gamma(D)\big)_x = T_xM$? There exists obviously a middle ground: at some point the distribution may be bracket generating while at others it may not, but we will see that this situation is not met in our context.

If the distribution $D = \mathrm{Ker}(\alpha)$ is maximally non-integrable, it means that every two points of the state space can be joined through a horizontal path, i.e. through a succession of reversible adiabatic transformations. On the contrary, if the distribution D is integrable, then we can deduce some properties of the differential one form δQ. Indeed, one can show that an alternative form of Frobenius theorem states that the graded ideal $\mathcal{I}_\alpha^\bullet = \bigoplus_{1\leq m\leq n} I_\alpha^m$ of the graded commutative algebra $\Omega^\bullet(M)$ generated by α – i.e. $I_\alpha^1 = \mathrm{Span}(\alpha)$ and $I_\alpha^m = \mathrm{Span}(\eta_1 \wedge \ldots \eta_{m-1} \wedge \alpha \,|\, \eta_i \in \Omega^1(M))$ – is actually a *differential graded ideal*, i.e. it is stable under the de Rham differential:

$$d\mathcal{I}_\alpha^\bullet \subset \mathcal{I}_\alpha^\bullet$$

For details on this statement, see for example Theorem 1.3.8 and Exercise 1.3.12 in [Candel and Conlon, 2000]. So, in particular, since $d\alpha \in I_\alpha^2$, there exists a one form η such that, at least locally, $d\alpha = \eta \wedge \alpha$. One can then show that this identity holds if and only if there exist two smooth functions $f, g \in \mathcal{C}^\infty$ such that $\alpha = fdg$. This observation leads to Carathéodory's (partial) reformulation of the second principle of thermodynamics:

Scholie 3.72. Carathéodory's second principle of Thermodynamics, *a.k.a* adiabatic inacessibility *Given a closed system, in every neighborhood of any state x there are states inaccessible from x through adiabatic (reversible) processes. In particular, this implies that there exists two smooth functions – T called the* temperature *and S called the* entropy *– such that the differential form δQ takes the following form:*

$$\delta Q = TdS$$

Proof. If, in the vicinity of each equilibrium state, there are other states which are *not* reachable through adiabatic reversible transformations, then the distribution $D = \mathrm{Ker}(\alpha)$ is *not* maximally non-integrable and, the assumption of non-accessibility holding at every point, we deduce that it is integrable. But then, by the above discussion on integrable distribution of rank $n-1$, we deduce that $\alpha = \delta Q$ can be written as fdg or, for the sake of consistency with traditional notations, $\delta Q = TdS$. The fact that this equality holds globally and not only locally comes from the fact that the thermodynamic state space is often a vector space, on which the cotangent bundle is trivial. □

Remark 3.73. Actually, Carathéodory's principle is not a faithful second principle, because it says nothing about the conditions under which the entropy increases. That is why it is necessary to supplement it with Planck's principle, stating that adiabatic isochoric processes always increase the internal energy of a closed system, hence corresponding to an increase of entropy. The only way that the entropy of a closed system can decrease is when heat is transferred from the system to the exterior. See e.g. [Sears, 1966] and [Zemansky, 1966] for a discussion on the relationships between non-equivalent statements of the second principle.

3.5 Orientation of smooth manifolds and integration of differential forms

Now we have enough material to define integration of differential forms on smooth manifolds. Theoretically one can integrate any differential k-forms, but this relies on advanced mathematics so we would rather only concentrate on integrating differential n-forms. This is consistent with what theoretical physicists mostly do in their everyday life. We would proceed as usual: integration on a manifold M would first be defined locally, because we know how to integrate differential n-forms in $\mathbb{R}^n$, and then using a partition of unity we can sum up all the local contributions to obtain an integral over M. A necessary condition to integrate is to have an *orientable* manifold. In this section we assume that the dimension of manifolds and vector spaces are greater than or equal to 1.

Given a n-dimensional vector space E, we say that two ordered basis $e_1, \ldots, e_n$ and $e'_1, \ldots e'_n$ are *consistently oriented* if the transition matrix from one to the other has positive determinant. This relation is an equivalence relation. Since $\mathbb{R} - \{0\}$ has two disjoint connected components, there are only two equivalence classes of consistently oriented ordered bases: either the determinant of the transition matrix is positive and we stay in the equivalence class, or it is negative, and we change class. We call an *orientation* on E either of those equivalence classes of those consistently oriented ordered bases. There is no absolute choice of orientation on a vector space (except maybe for $\mathbb{R}^n$), there are only relative choices: once we have chosen an ordered basis $e_1, \ldots, e_n$, it is a convention to say that every other consistently oriented ordered basis is *positively oriented.* A basis that is obtained from $e_1, \ldots, e_n$ through a transition matrix with negative determinant is said *negatively oriented.*

Example 3.74. The vector space $\mathbb{R}^n$ admits the following standard basis $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ where 1 is located at the i-th position. We say that the orientation defined by this basis is the standard orientation of $\mathbb{R}^n$.

Lemma 3.75. *Let E be a vector space of dimension $n \geq 1$, and suppose ω is a nonzero element of $\bigwedge^n(E^*)$. The set of ordered bases $e_1, \ldots, e_n$ such that $\omega(e_1, \ldots, e_n)$ has the same sign is an orientation for E.*

Proof. Let $e_1, \ldots, e_n$ and $e'_1, \ldots, e'_n$ be two basis of E and let B the transition matrix from the former to the latter: $e'_i = B_i^j\, e_j$. Then:

$$\omega(e'_1, \ldots, e'_n) = \det(B)\, \omega(e_1, \ldots, e_n)$$

so that $e_1, \ldots, e_n$ and $e'_1, \ldots, e'_n$ are consistently oriented if and only if $\omega(e'_1, \ldots, e'_n)$ and $\omega(e_1, \ldots, e_n)$ have the same sign. □

Thus, choosing an orientation of a vector space E amounts to choosing an element ω of $\bigwedge^n E^*$. One this choice is made, we say that ω is a positively oriented n-covector. For example, if the ordered basis of E is given by $e_1, \ldots, e_n$, the n-covector $\omega = e^1 \wedge \ldots \wedge e^n$ is positively oriented. For any real scalar $\lambda > 0$, $\lambda\, \omega$ is another positively oriented n-covector, while for any real scalar $\mu < 0$, $\mu\, \omega$ is said to be a negatively oriented n-covector. This plays some role in the definition of the Hodge star operator. Indeed, it depends on a choice of orientation of E because the volume form $\omega = e^1 \wedge e^2 \wedge \ldots \wedge e^n$ is given by the choice of an ordered basis $e_1, e_2, \ldots, e_n$. If one had taken the ordered basis $e_2, e_1, e_3, \ldots, e_n$ instead – with reverse orientation, then – the associated volume form positively oriented with respect to the orientation defined by $e_2, e_1, e_3, \ldots, e_n$ would be $\omega' = e^2 \wedge e^1 \wedge e^3 \wedge \ldots \wedge e^n = -\omega$, so that the Hodge star operator $\star'$ associated with ω' would be the opposite to the one associated with ω: $\star' = -\star$.

Since a smooth manifold is locally euclidean, we can define an orientation locally, at the level of the tangent bundle. A *pointwise orientation* on M is the assignment, to every point x, of an orientation of the fiber T_xM. It is always possible to equip a smooth manifold with such a pointwise orientation, but the difficulty comes from having this orientation varying smoothly over the manifold. A local smooth frame $X_1, \ldots, X_n$ of the tangent bundle over some open set U is said *positively oriented* if, for every $x \in U$, the orientation of the basis $X_{1,x}, \ldots, X_{n,x}$ coincides with the orientation of T_xM. A pointwise orientation is said *smooth* if every point of M is in the domain of an oriented local smooth frame. Given two smooth manifolds M and N of the same dimension that admit smooth pointwise orientations, we say that a local diffeomorphism $F : M \longrightarrow N$ is *orientation-preserving* if, for every $x \in M$, F_* takes oriented bases of T_xM to oriented bases of $T_{F(x)}N$, and *orientation-reversing* if it takes (positively) oriented bases of T_xM to negatively oriented bases of $T_{F(x)}N$.

We want to study how the existence of a smooth pointwise orientation translates at the level of charts and transition functions. Let M be a smooth manifold equipped with a (non necessarily smooth) pointwise orientation. Any smooth chart (U, φ) whose coordinate frame $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ is positively (resp. negatively) oriented is called a positively (resp. negatively) *oriented chart.* Any smooth oriented chart (U, φ) on M always induce another chart $(U, \overline{\varphi})$ with reverse orientation. Indeed let $\overline{\varphi}$ be the composite of φ and a reflectional symmetry (which is a smooth map from $\mathbb{R}^n$ to $\mathbb{R}^n$), then $(U, \overline{\varphi})$ has reverse orientation compared to (U, φ). Obviously, there exist choices of pointwise orientations such that some charts are neither positively nor negatively oriented. However we will see that for smooth pointwise orientations, the situation is really nice. We say that two smooth oriented charts (U, φ) and (V, ψ) are *consistently oriented* if the transition map $\psi \circ \varphi^{-1}$ has positive Jacobian determinant, i.e. if it is orientation preserving. An *oriented atlas* on M is a smooth atlas for which all smooth charts are consistently oriented. M is *orientable* if it admits an oriented atlas, and an *orientation of M* is a choice of a maximal oriented atlas. The following proposition shows that the existence of a smooth pointwise orientation on M is equivalent to an orientation on M. Thus, oriented atlases form a subclass of smooth atlases, where the transition functions are not only diffeomorphisms, but also orientation preserving. The relationship between the two notions of orientability is actually very simple:

Proposition 3.76. *Let M be a positive ($n > 1$) smooth manifold equipped with a pointwise orientation. Then it is smooth if and only if it induces an orientation on M.*

Proof. First, notice that a smooth pointwise orientation on M implies that there exists an open cover of positively oriented charts. This can be seen as follows: let $x \in M$ an let $X_1, \ldots, X_n$ be an oriented frame defined on an open neighborhood of x. One can assume that this neighborhood is a smooth chart (U, φ). Then, the induced coordinate frame is either positively oriented, or negatively oriented, but in that latter case the smooth chart $(U, \overline{\varphi})$ is positively oriented. Then we can find an open cover of positively oriented charts.

Assume that the chosen pointwise orientation on M is smooth and pick up such an open cover of oriented charts. Then, by Equation 3.2, on overlapping oriented charts, the transition functions are orientation-preserving. This implies that the open cover of positively oriented charts is an oriented atlas, providing M with an orientation. Conversely, every orientation makes the pointwise orientation smooth because the coordinate frames are oriented frames, and two such frames define the same orientations on the fibers since the transition functions between oriented charts are orientation-preserving by hypothesis. $\square$

Thus, being smoothly pointwise orientable is equivalent to being orientable. If a smooth manifold is orientable, there are essentially two possible choices of orientations. Pick up a tangent space and attribute an orientation to this vector space (here we make a choice between two orientations). Then, by Proposition 3.76, the respective orientations of the other fibers of the tangent bundle will be automatically determined by this first choice. This can be seen from the fact that transitions functions from one oriented chart to another are orientation preserving. Non-orientable manifolds are precisely those manifolds for which there are always at least one transition function that is not orientation preserving, whatever the choice of smooth chart we make. For a zero dimensional manifold, i.e. a point $\{*\}$, an orientation is a choice of map $\{*\} \longmapsto \{\pm 1\}$. We know at least one evident situation where a smooth manifold is orientable:

Proposition 3.77. *Every parallelizable smooth manifold is orientable.*

Example 3.78. Every Lie group is parallelizable, hence is orientable.

Example 3.79. Spheres, planes and tori are orientable.

Example 3.80. The *Mobius bundle* is the vector bundle E over S^1 whose total space is defined as a quotient of $\mathbb{R}^2$ by the following relation:

$$(x, y) \sim (x + n, (-1)^n y)$$

The *Mobius band* is the subset $M \subset E$ of the Mobius bundle that is the image, under the above quotient map, of the set $\{(x, y) \in \mathbb{R}^2 | \, |y| \leq 1\}$. It is a smooth 1-manifold with boundary, which is non orientable.

The most important point with orientations is that we can characterize it through differential forms. A *volume form* on M is a global *nowhere vanishing* smooth section of the vector bundle $\bigwedge^n T^*M$. We usually denote such a section by the letter ω. Over a local smooth chart U, with respect to a coordinate coframe, the volume form decomposes as $\omega = f\, dx^1 \wedge \ldots \wedge dx^n$, for some smooth function $f \in \mathcal{C}^\infty(U)$. The existence of volume forms is tightly connected to that of orientations:

Proposition 3.81. *Let M be a smooth manifold of dimension $n \geq 1$. Any nowhere vanishing differential n-form $\omega \in \Omega^n(M)$ determines a unique orientation of M for which the n-covector $\omega(x) \in \bigwedge^n T^*M$ is positively oriented for every $x \in M$. Conversely, if M is given an orientation, there is a smooth nowhere vanishing differential n-form on M that is positively oriented at each point.*

Proof. Assume that there exists such a volume form ω, so by Lemma 3.75, the evaluation of the volume form ω at a point x induces an orientation of the tangent space T_xM, that is considered to be positively oriented. Now let us check that there exists an oriented smooth atlas for M. Let (U, φ) and (V, ψ) be two intersecting oriented charts. Let us denote by $x^1, \ldots, x^n$ and $x'^1, \ldots, x'^n$ the coordinates respectively associated to the maps φ and ψ. Then, on U the volume forms reads $\omega = f\, dx^1 \wedge \ldots \wedge dx^n$, while on V it reads $\omega = g\, dx'^1 \wedge \ldots \wedge dx'^n$, for two nowhere vanishing functions $f \in \mathcal{C}^\infty(U)$ and $g \in \mathcal{C}^\infty(V)$. Over the intersection $U \cap V$, using the transformation laws found in Equations (3.6)-(3.7), one obtains that:

$$dx'^i\Big|_y = \frac{\partial x'^i}{\partial x^j}(\varphi(y))\, dx^j\Big|_y \tag{3.12}$$

Then, we obtain that $g\, dx'^1 \wedge \ldots \wedge dx'^n = g\, \mathrm{Jac}\big(\psi \circ \varphi^{-1}\big)\, dx^1 \wedge \ldots \wedge dx^n$, where Jac symbolizes the Jacobian determinant. Then, we have:

$$f(y) = g(y)\, \mathrm{Jac}\big(\psi \circ \varphi^{-1}\big)\big(\varphi(y)\big)$$

for every $y \in U \cap V$. The sign of the Jacobian determinant is determined by the sign of the function $\frac{f}{g}$ which is nowhere vanishing over $U \cap V$.

Now, either f and g have the same sign, and then (U, φ) and (V, ψ) are consistently oriented, or they do not have the same sign. However, in that case, one may define another chart $(V, \overline{\psi})$ by changing a sign in the definition of ψ, e.g. $\psi(y) \longmapsto \overline{\psi}(y) = \big(-\psi^1(y), \psi^2(y), \ldots, \psi^n(y)\big)$. This is possible because reflectional symmetries with respect to hyperplanes are diffeomorphisms of $\mathbb{R}^n$. We label the corresponding new coordinates as $\overline{x}^i$, and in particular $\overline{x}^1 = -x'^1$ whereas for $2 \leq i \leq n$, $\overline{x}^i = x'^i$. Then the volume form decomposes in this new coordinate coframe as $\omega = -g\, d\overline{x}^1 \wedge \ldots \wedge d\overline{x}^n$, and then, the new Jacobian determinant reads: $\det\big(\overline{\psi} \circ \varphi^{-1}\big)\big(\varphi(y)\big) = -\frac{f(y)}{g(y)}$ which is now a positive fraction for every $y \in U \cap V$. Thus, the oriented chart $(V, \overline{\psi})$ is consistently oriented with (U, φ). This proves the first statement. For the converse statement – that any orientable smooth manifold admits a volume form – see Proposition 13.4 in [Lee, 2003]. □

Since the vector bundle $\bigwedge^n T^*M$ has rank 1, any other *nowhere vanishing* smooth section $f\,\omega$, where $f \in \mathcal{C}^\infty(M)$, is a volume form as well. Since there are two disjoint connected components in $\mathbb{R} - \{0\}$, there are two equivalence classes of sections of the line bundle $\bigwedge^n T^*M$: those that are positively related to ω, and those that are negatively related to ω. Moreover, those volume forms that are negatively related with ω are still positively related among themselves. Thus, as the proof of Proposition 3.81 shows, picking up any other representent of the equivalence class of ω – i.e. of the form $f\,\omega$ for some strictly positive function – defines the same orientation on

M as ω. Actually, the oriented atlas associated to ω is obtained as the collection of all smooth charts for which the standard volume form on $\mathbb{R}^n$ (induced from the standard oriented basis presented in Example 3.74) pulls back to a positive multiple of ω. That is why some authors define an orientation on M as a choice of an equivalence class of positively related volume forms (see e.g. [Baez and Muniain, 1994, p. 84]):

Corollary 3.82. *There is a one-to-one correspondence between orientations on M and equivalence classes of positively related globally defined volume forms.*

Remark 3.83. There are homological and cohomological characterization of orientability. For example, a smooth manifold is orientable if and only if the first Stiefel-Whitney characteristic class is 0.

Orientability is necessary to define integration on smooth manifolds. Since a smooth manifold is locally euclidean, let us first define integration over $\mathbb{R}^n$, before generalizing to any smooth manifold using pullbacks. A subset of $\mathbb{R}^n$ is a *domain of integration* if its boundary has n-dimensional measure 0. We usually define integration in $\mathbb{R}^n$ defining first the integral of bounded continuous functions on 'rectangles', i.e. products of closed intervals. Then, every continuous function can be locally approximated by such functions, and every domain of integration can be covered by rectangles (given by the closure of open sets inherited from the subspace topology of $\mathbb{R}^n$ on D), so that in the end we can define the integral of bounded continuous functions on any domain of integration. Then, a choice of domain of integration D defines a linear form on the space of bounded continuous functions on D:

$$\begin{aligned} \int_D : \quad \mathcal{C}^0_{\mathrm{b}}(D) &\longrightarrow \mathbb{R} \\ f \quad &\longmapsto \int_D f\,dx^1 \dots dx^n \end{aligned}$$

where the notation $dx^1 \dots dx^n$ is purely abstract and needs not appear. It only reminds the reader that we integrate the function over a subset of $\mathbb{R}^n$. It is sometimes noted $d\mu$ to symbolize the Lebesgue measure. You can find more details on this construction in Appendix A of [Lee, 2003].

This definition straightforwardly generalizes to differential n-forms. Let $U \subset \mathbb{R}^n$ be an open set and let ω be a differential n-form compactly supported on some compact set $K \subset U$:

$$K = \mathrm{supp}(\omega) = \overline{\{x \in M \,|\, \omega(x) \neq 0\}}$$

Lemma 14.1 in [Lee, 2003] shows that there always exists a domain of integration D such that $K \subset D \subset U$. Then, assuming that this differential form can be written as $\omega = f\,dx^1 \wedge \ldots \wedge dx^n$ over its support K, the *integral of ω over U* is given by:

$$\int_U \omega = \int_D f\,dx^1 \dots dx^n$$

The notation here is very convenient: it is as if we had 'erased' the wedges. Notice that the above definition does not depend on the choice of domain of integration $K \subset D \subset U$.

Lemma 3.84. *Suppose U, V are open sets of $\mathbb{R}^n$ and that $F : U \longrightarrow V$ is a diffeomorphism. Let ω be a compactly supported differential n-form on V. Then:*

$$\int_V \omega = \begin{cases} \displaystyle\int_U F^*\omega & \text{if } F \text{ is orientation-preserving} \\ \displaystyle -\int_U F^*\omega & \text{if } F \text{ is orientation-reversing} \end{cases}$$

This lemma provides another formulation of the fact that a differential n-form can be written equivalently in two sets of coordinates $x^1, \dots, x^n$ and $x'^1, \dots, x'^n$, defined over overlapping open sets U and V, and related through a diffeomorphism $F : U \longrightarrow V$, e.g. such that $F(x^1, \dots, x^n) = (x'^1, \dots, x'^n))$. Then if one writes the differential form on V as $\omega = g\, dx'^1 \wedge \dots \wedge dx'^n$, Equation (3.12) implies:

$$\begin{aligned} F^*\omega &= F^*\left(g\, dx'^1 \wedge \dots \wedge dx'^n\right) \\ &= (F^*g)\, F^*(dx'^1) \wedge \dots \wedge F^*(dx'^n) \\ &= g \circ F \cdot \mathrm{Jac}(F)\, dx^1 \wedge \dots \wedge dx^n \end{aligned}$$

where $\cdot$ indicates the multiplication of two smooth functions on U. Thus, we obtain the infamous formula for a change of coordinates under integration:

$$\int_V g(x'^1, \dots, x'^n)\, dx'^1 \dots dx'^n = \int_U g \circ F(x^1, \dots, x^n)\, \mathrm{Jac}(F)(x^1, \dots, x^n)\, dx^1 \dots dx^n$$

We have now enough material to define integration on manifolds. Let M be a n-dimensional oriented smooth manifold. First, let ω be a differential n-form compactly supported in the domain of a single oriented smooth chart (U, φ). Then, we define the *integral of ω over U* as the following objet:

$$\int_U \omega = \int_{\varphi(U)} (\varphi^{-1})^*\omega \tag{3.13}$$

The right-hand side is an integral over an open subset $\widetilde{U} = \varphi(U)$ of $\mathbb{R}^n$. It is well defined because $(\varphi^{-1})^*\omega$ is a compactly supported differential n-forms on this open set. Lemma 3.84 implies that the integral of ω over any other choice of oriented smooth chart (V, ψ) containing its compact support would have given the same result:

$$\int_U \omega = \int_V \omega$$

Now that we have defined integration over compact support, we can extend integration over the whole manifold M by using a notion that is specific to smooth manifolds:

Definition 3.85. *Let $\mathcal{U} = (U_\alpha)_{\alpha \in A}$ be any open cover of M (indexed over some set A). A* partition of unity subordinate to $\mathcal{U}$ *is a collection of continuous functions $(\psi_\alpha : M \longrightarrow \mathbb{R})_{\alpha \in A}$, with the following properties:*

1. $0 \leq \psi_\alpha(x) \leq 1$ *for all $\alpha \in A$ and all $x \in M$;*
2. $\mathrm{supp}(\psi_\alpha) \subset U_\alpha$*;*
3. *for every $x \in M$, there is only a finite number of ψ_α such that $x \in \mathrm{supp}(\psi_\alpha)$;*
4. $\sum_{\alpha \in A} \psi_\alpha(x) = 1$ *for all $x \in M$.*

The third condition is equivalent to saying that the set of supports $\{\mathrm{supp}(\psi_\alpha)\}_{\alpha \in A}$ is locally finite. Because of this condition, the sum in the last item has only finitely many nonzero terms in the neighborhood of each point, so there is no issue of convergence. When the functions ψ_α are smooth, then we say that they form a *smooth* partition of unity. The importance of partition of unity is central in differential geometry, as the following theorem shows:

Theorem 3.86. *Any open cover $\mathcal{U}$ of a smooth manifold admits a smooth partition of unity subordinate to $\mathcal{U}$.*

Proof. The main point is that the fact that a topological manifold is Hausdorff and second-countable implies that it is paracompact (and that it has countably many connected components), which is the crucial property needed to demonstrate the result. However, the proof is long and subtle, so we refer to Chapter 2 of [Lee, 2003]. $\square$

Remark 3.87. While Theorem 3.86 show that smooth partitions of unity subordinate to any open cover of a smooth manifold exist, it is no longer the case for analytic manifolds. Indeed, the proof relies on the existence of smooth bump functions on $[-1, 1]$. Unfortunately, those bump functions are not analytic because of the so-called identity theorem, which is then an obstruction to the existence of analytic partitions of unity subordinate to any open cover of an analytic manifold.

To integrate over a (connected) smooth manifold, one needs an orientation. The latter is needed to integration over the entire manifold in order to ensure that local contributions, as defined by Equation (3.13), do not artefactually cancel one another because of a change in open chart, as shown by the change of sign in Lemma 3.84. Let ω be a compactly supported differential n-form on a connected oriented smooth manifold M. Then there exists a finite open cover $\{(U_i, \varphi_i)\}$ of oriented charts for $\operatorname{supp}(\omega)$, and a partition of unity $\{\psi_i\}$ subordinated to this open cover. We define the integral of ω as follows:

$$\int_M \omega = \sum_i \int_{U_i} \psi_i \, \omega \tag{3.14}$$

For each i, the n-form is compactly supported in U_i, so that the integral on the right-hand side is obtained through Equation (3.13). There are finitely many non-zero integrals on the right because the open cover of $\operatorname{supp}(\omega)$ is finite. It turns out that Equation (3.14) neither depends on the choice of finite cover, nor on the choice of partition of unity (see Lemma 14.5 in [Lee, 2003]). The disconnected case requires to define an orientation on each connected component. In the following proposition are listed several properties of this integral:

Proposition 3.88. *Let M and N be oriented smooth n-dimensional manifolds, and let ω, η be compactly supported differential n-forms on M. Then:*

1. ***Linearity:** for every $a, b \in \mathbb{R}$,*

$$\int_M a\,\omega + b\,\eta = a \int_M \omega + b \int_M \eta$$

2. ***Positivity:** if ω is positively oriented, then $\int_M \omega > 0$;*

3. ***Orientability:** If $F : N \longrightarrow M$ is a diffeomorphism, then:*

$$\int_M \omega = \begin{cases} \displaystyle\int_N F^*\omega & \text{if } F \text{ is orientation-preserving} \\ \displaystyle -\int_N F^*\omega & \text{if } F \text{ is orientation-reversing} \end{cases}$$

Remark 3.89. Equation (3.14) is still valid for non-compactly supported differential forms, but in that case the integral is improper since the sum on the right may not converge. On a compact manifold, the integral is defined for every differential n-form.

We conclude this section by briefly discussing two important results relying on integration of differential forms. We note $\Omega_c^p(M)$ the compactly supported differential p-forms on M. The de Rham differential applies to compactly supported differential forms and induces a cohomology, denoted $H_c^m(M)$. However, this cohomology is different than the de Rham cohomology, as the following observation shows:

Proposition 3.90. *The compactly supported de Rham cohomology of* $\mathbb{R}^n$ *satisfies:*

$$H_c^i(\mathbb{R}^n) \simeq \begin{cases} \mathbb{R} & \text{if } i = n \\ 0 & \text{otherwise} \end{cases}$$

One notices that the m-th compactly supported cohomology group of compact support is isomorphic to the $n-m$-th de Rham cohomology group. Does this extend to general manifolds? For every $0 \leq m \leq n$, notice that the integral defined in Equation (3.14) induces a linear morphism:

$$\begin{aligned} \mathcal{PD}: \ \Omega^m(M) &\longrightarrow \Omega_c^{n-m}(M)^* \\ \eta &\longmapsto \mathcal{PD}(\eta) : \mu \longmapsto \int_M \eta \wedge \mu \end{aligned}$$

Following de Rham, we call $n-m$*-currents* the elements of $\Omega_c^{n-m}(M)^*$; they are related to the notion of distribution. The de Rham differential on compactly supported $n-m$-forms induces a degree -1 differential $d' : \Omega_c^\bullet(M)^* \longrightarrow \Omega_c^{\bullet-1}(M)^*$ defined by $d'\Phi(\mu) = \Phi((-1)^{|\Phi|}d\mu)$. Then, one can show that $\mathcal{PD}$ commutes with the differentials: $\mathcal{PD}(d\eta)(\mu) = d'\mathcal{PD}(\eta)(\mu)$. This result implies that $\mathcal{PD}$ induces a map at the cohomology level, which turns out to be an isomorphism:

Theorem 3.91. Poincaré duality *Let* M *be a smooth orientable manifold, then:*

$$H_{dR}^m(M) \simeq H_c^{n-m}(M)^*$$

Proof. See Exercise 16.6 in [Lee, 2003]. □

We conclude this section by mentioning Stokes' theorem. This result relies on the notion of *manifold with boundary*. We will not enter into the details of this notion, because it would take too much time, but many informations can be found in [Baez and Muniain, 1994] and [Lee, 2003]. The main idea is that a manifold with boundary is locally homeomorphic to the euclidean upper half-plane:

$$\mathbb{H}^n = \{(x^1, \dots, x^n) \in \mathbb{R}^n \,|\, x^n \geq 0\}$$

It means that a chart on a manifold with boundary M is either homeomorphic to an open subset (with respect to the subspace topology) of the interior of $\mathbb{H}^n$, or to an open subset of $\mathbb{H}^n$ which intersects the boundary. The boundary of M is the set ∂M of points of M which are sent to the boundary $\partial\mathbb{H}^n$ of the upper half plane through the coordinate maps. By construction, the boundary of M is a closed embedded submanifold of M (see Exercise 8.5 in [Lee, 2003]). If the manifold M has an orientation, there is a distinguished orientation on its boundary ∂M. In that case, one can define integration of differential n-forms on M and at the same time define integration of differential $n-1$-forms on the boundary ∂M. Stokes' theorem is a statement about the relationship between those two integrals:

Theorem 3.92. Stokes' theorem *Let* M *be a smooth, oriented* n*-dimensional manifold with boundary, and let* ω *be a compactly supported smooth* $n-1$*-form on* M*. Then:*

$$\int_M d\omega = \int_{\partial M} \omega$$

The meaning of the integral on the right hand side and of this theorem are discussed in details in [Baez and Muniain, 1994] and [Lee, 2003].

3.6 Pseudo-Riemannian manifolds and Laplace-de Rham operator

In the last section we used the fact that smooth manifolds admit a tangent bundle, that associates a tangent space to every point, to define an orientation on manifolds. We can use the same strategy to define pseudo-Riemannian metrics on smooth manifolds. First we define a pointwise metric, and we require it to vary smoothly over the manifold.

Definition 3.93. *Let M be an n-dimensional smooth manifold and let $x \in M$. A* metric tensor *on M is a smooth section g of the vector bundle $S^2(T^*M)$ that restricts at every point $x \in M$ to a pseudo-Riemannian metric $g_x : T_xM \times T_xM \longrightarrow \mathbb{R}$. We call a smooth manifold equipped with a metric tensor a* pseudo-Riemannian manifold*; it is said* Riemannian *when the metric tensor is positive definite at every point.*

We know that the tangent space at a point is the best linear approximation of the manifold at that point. The metric tensor at this point is then fed by tangent vectors. However, since g is a smooth tensor, it can be fed by vector fields, and the result defines a smooth function (that is actually a way of characterizing smoothness of g):

$$g(X, Y) = g(Y, X) \in \mathcal{C}^\infty(M) \qquad \text{for every } X, Y \in \mathfrak{X}(M)$$

The smoothness of the metric tensor g is characterized by the fact that the map $x \longmapsto g_x(X_x, Y_x)$ is a smooth map, for every two smooth vector fields X and Y. The metric tensor g can be locally decomposed in a coordinate cotangent basis $dx^1, \dots, dx^n$ over a coordinate chart U as:

$$g = g_{ij}\, dx^i \odot dx^j$$

where $g_{ij} \in \mathcal{C}^\infty(U)$ are smooth functions. So, in particular, with respect to the coordinate tangent frame:

$$g_{ij} = g(\partial_i, \partial_j)$$

Obviously since the metric tensor varies smoothly, its pointwise signature is constant over U (and more generally, over M), and is determined by the eigenvalues of the matrix-valued smooth function $G \in \mathcal{C}^\infty(U, \mathcal{M}_n(\mathbb{R}))$.

Remark 3.94. Let D be a regular smooth distribution on M. Then, assume that we have a smoothly varying metric g_x defined on each subspace D_x and that is smoothly varying, in the sense that for every two smooth sections $X, Y \in \Gamma(D)$, the map $x \longmapsto g_x(X_x, Y_x)$ is a smooth map over M. It is as if the metric was defined in the directions defined by D. We call this 'metric tensor' a *sub-Riemannian metric*, and the smooth manifold M a sub-Riemannian manifold. This metric defines a distance function on the manifold by integrating it over any horizontal path joining the two points:

$$d_\gamma(x, y) = \int_0^1 \sqrt{g_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t))}\,dt$$

The *Carnot-Carathéodory distance* is then the infimum of all such distance, over all the horizontal paths:

$$d_{CC}(x, y) = \inf_{\text{horizontal}\,\gamma} \{d_\gamma(x, y)\}$$

This distance is very useful in sub-Riemannian geometry. For example, Chow-Rashevskii theorem 3.68 can be restated as the following: "the topology induced by the Carnot-Carathéodory metric is equivalent to the intrinsic (locally Euclidean) topology of the manifold".

A pseudo-Rimannian manifold that is additionally an oriented manifold has a distinguished volume form, that we now present. By Corollary 3.82, the fact that M is oriented means that there exist a nowhere vanishing globally defined volume form ω that is positively orientated at every point. The following argument explain that we can chose ω in a certain, adapted form. Since a metric tensor g induces a pointwise metric g^{-1} on the cotangent bundle that varies smoothly, then the following map $\sqrt{|\det(G^{-1})|} = \frac{1}{\sqrt{|\det(G)|}}$ is a smooth function which is well defined and nowhere vanishing. We also sometimes write $\sqrt{|g|}$ instead of $\sqrt{|\det(G)|}$. Since the metric associated to the coordinate cotangent frame $dx^1, \ldots, dx^n$ is g^{-1}, at the price of multiplying ω by a nowhere vanishing positive smooth function, we can always have:

$$\omega = \sqrt{|g|}dx^1 \wedge \ldots \wedge dx^n \tag{3.15}$$

This formula is the counterpart of Equation (1.24) in the context of smooth manifolds, where the exterior algebra $\bigwedge^\bullet(E)$ is the vector bundle $\bigwedge^\bullet T^*M$. Equation (3.15) is the local form of the standard volume element on a pseudo-Riemannian oriented manifold (M, g).

Following the discussion in Section 1.2, the pseudo-Riemannian metric g on M induces a pairing on the fibers of the exterior algebra of the cotangent bundle:

$$\begin{aligned} \langle\,.\,,.\,\rangle:\quad \Omega^m(M)\times\Omega^m(M) &\longrightarrow \mathcal{C}^\infty(\mathbb{R})\\ (\eta,\mu) &\longmapsto \langle\eta,\mu\rangle : x\longrightarrow\langle\eta_x,\mu_x\rangle \end{aligned}$$

for every $0 \leq m \leq n$. It is fiberwisely non-degenerate, but one needs to integrate the function on the right-hand side in order to define an inner product $(\,.\,,.\,)$ on differential forms, via the following formula:

$$(\eta, \mu) = \int_M \langle \eta, \mu\rangle\, \omega \tag{3.16}$$

for every $\eta, \mu \in \Omega^m(M)$, and every $0 \leq m \leq n$, and where ω is the distinguished volume form defined in Equation (3.15). This product may be divergent if the support of one of the arguments does not have compact support. It defines a L^2 norm on those differential forms η that are such that $(\eta, \eta) < +\infty$ (in particular compactly supported differential forms satisfy this condition).

The volume form defined in Equation (3.15) and the fiberwise inner product $\langle\,.\,,.\,\rangle$ also allow to define a Hodge star operator $\star : \Omega^m(M) \longrightarrow \Omega^{n-m}(M)$, as in Equation (1.25). One can then define a $\mathcal{C}^\infty(M)$-linear operator $\delta : \Omega^\bullet(M) \longrightarrow \Omega^{\bullet-1}(M)$ that is 'dual' in some sense to the de Rham differential. On m-forms, it is defined as:

$$\delta = -(-1)^{n(m-1)+q} \star d \star$$

and sends m-forms to $m-1$-forms. By construction, on 0-forms it is zero. Using the definition of the inverse star operator (see Equation (1.29)) $\star^{-1} : \bigwedge^{n-m-1}(M) \longrightarrow \bigwedge^{m-1}(M)$, one deduces that $\delta : \Omega^m(M) \longrightarrow \Omega^{m-1}(M)$ can also be written as:

$$\delta = (-1)^m \star^{-1} d\star \tag{3.17}$$

Then, for every differential m-form η, and any differential $m+1$-form μ, one has the following identity:

$$\star\big(\langle d\eta, \mu\rangle - \langle \eta, \delta\mu\rangle\big) = d(\eta \wedge \star\mu) \tag{3.18}$$

The right-hand side is a n-form, that is why we used a star operator on the scalar in parenthesis on the left-hand side so that it becomes a n-form as well. Thus Equation (3.18) implies that δ is the adjoint of the de Rham differential, with respect to the inner product on differential forms:

$$(\eta, \delta\mu) = (d\eta, \mu)$$

for every $\eta \in \Omega^m(M)$ and $\mu \in \Omega^{m+1}(M)$, where $0 \leq m \leq n-1$.

Exercise 3.95. Prove Equation (3.17).

Exercise 3.96. Prove that the identity $d^2 = 0$ implies that $\delta^2 = 0$.

Definition 3.97. *We call the $\mathcal{C}^\infty(M)$-linear operator $\delta : \Omega^\bullet(M) \longrightarrow \Omega^{\bullet-1}(M)$ the* codifferential*. We define the* Laplace-de Rham operator *as the $\mathcal{C}^\infty(M)$-linear operator $\Delta_{dR} : \Omega^\bullet(M) \longrightarrow \Omega^\bullet(M)$ such that:*

$$\Delta_{dR} = d \circ \delta + \delta \circ d$$

The first term of the Laplace-de Rham operator vanishes on smooth functions, i.e. 0-forms, so that we obtain minus the *Laplace-Beltrami operator*[13]:

$$\Delta_{dR}(f) = -(-1)^q \star d \star df = -\Delta(f)$$

The difference in sign is a convention and descends from the additional sign in δ. The Laplace-de Rham operator is defined to be positive definite, whereas the Laplace-Beltrami operator is usually taken to be negative definite. Since the Laplace-Beltrami operator reads, in coordinates:

$$\Delta(f) = \frac{1}{\sqrt{|g|}}\partial_i\left(\sqrt{|g|}g^{ij}\partial_j(f)\right) \tag{3.19}$$

we deduce that, in Minkowski space-time with signature $(3, 1)$ or, in physics notation, $(-, +, +, +)$, the Laplace-de Rham operator is the d'Alembertian:

$$\Delta_{dR} = \Box = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2} \tag{3.20}$$

Be aware however that, under the convention that the signature of the Minkowski metric is $(1, 3) = (+, -, -, -)$, the right-hand side can be written $\partial^\mu\partial_\mu$. On the contrary, with our convention of signature and using Equation (3.19), the right-hand side of Equation (3.20) reads $-\partial^\mu\partial_\mu$.

The codifferential δ and the Laplace-de Rham operator Δ_{dR} allow to characterize more precisely differential forms and de Rham cohomology. A differential form η that is such that $\delta\eta = 0$ is called *co-closed*, while if there is another differential form μ such that $\eta = \delta\mu$, we say that η is *co-exact*. Differential forms that lie in the kernel of the Laplacian, i.e. those η such that $\Delta_{dR}(\eta) = 0$, are called *harmonic*. We denote by $\mathcal{H}^m(M)$ the space of harmonic differential m-forms, for $0 \leq m \leq n$. Exact, co-exact and harmonic differential forms provide a nice decomposition of the space of differential forms:

Theorem 3.98. Hodge decomposition *Let M be a compact Riemannian manifold, then for every $0 \leq m \leq n$, we have the following decomposition:*

$$\Omega^m(M) = d(\Omega^{m-1}(M)) \oplus \delta(\Omega^{m+1}(M)) \oplus \mathcal{H}^m(M)$$

This direct sum is orthogonal with respect to the inner product defined in Equation (3.16).

This decomposition is very useful to find a distinguished representative of de Rham cohomology classes, because the following corollary proves that each cohomology class has a unique harmonic representative:

Corollary 3.99. *Let M be a compact Riemannian manifold, then for every $0 \leq m \leq n$ we have an isomorphism:*

$$H^m_{dR}(M) \simeq \mathcal{H}^m(M)$$

[13] We cannot justify yet that the Laplace-Beltrami operator $\mathrm{div} \circ \overrightarrow{\mathrm{grad}}$ corresponds $(-1)^q \star d \star d$ but it will be shown later in the course.

Proof. Both proofs of the theorem and of the corollary can be found in Chapter 6 of [Warner, 1983]. □

We conclude this section by the following beautiful remark: integrals and the Hodge star operator allow to write actions in a rather nice way. For example, integrating the Lagrangian density of Maxwell's electromagnetism $F_{\mu\nu}F^{\mu\nu}$ over a pseudo-Riemannian n-dimensional manifold M can be synthesized as (physical notation is on the left):

$$\mathcal{S}_M = \frac{1}{4}\int_M F_{\mu\nu}F^{\mu\nu}\sqrt{|g|}d^n x = \frac{1}{2}\int_M F \wedge \star F$$

One can also write Einstein-Hilbert action (without cosmological constant) as:

$$\mathcal{S}_{EH} = \int_M R\sqrt{|g|}d^n x = \int_M \star R$$

where R is the Ricci scalar. Then, more generally, integrating a Lagrangian density $\mathcal{L}$ over an oriented pseudo-Riemannian smooth manifold M provides the following action:

$$\mathcal{S} = \int_M \star\mathcal{L}$$

Obviously in both cases there is a possible problem of convergence of the integral but we may either work only locally (physical quantities in classical physics do not have non-local properties) so that we can assume that the Lagrangian densities are compactly supported, or we can accept that the integral is not properly defined although while we admit only compactly supported variations of the fields (e.g. δA would be the compactly supported 'variation' of a connection 1-form A), then the induced variation $\delta\mathcal{S}$ would be well-defined (see Section II.4 of [Baez and Muniain, 1994]). This opens the possibility to work on physical theories from a geometric point of views. Gauge theories are precisely theories which benefit from such an approach.

4 Poisson geometry

Poisson geometry draws on the work of mathematicians in the 1960s-1970s striving to formalize Hamiltonian mechanics[14]. Recall that in Hamilton's formulation of classical mechanics, a physical system is characterized by a set of positions q^i and conjugate momenta p_i (where $1 \leq i \leq n$) defining a point in a *phase space* $P = \mathbb{R}^{2n}$, and the evolution of the system is governed by a function $H(q,p)$ called the *Hamiltonian*, so that Hamilton's equations are:

$$\dot{q}^i = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i} \tag{4.1}$$

for every $1 \leq i \leq n$. In this context, the classical Poisson bracket is a skew-symmetric differential operator $\{\,.\,,.\,\} : \mathcal{C}^\infty(P) \times \mathcal{C}^\infty(P) \longrightarrow \mathcal{C}^\infty(P)$ defined as on any two smooth functions $f, g \in \mathcal{C}^\infty(P)$ by:

$$\{f,g\} = \sum_{i=1}^n \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q^i}\frac{\partial f}{\partial p_i} \tag{4.2}$$

Using the Poisson bracket, Hamilton's equations (4.1) become:

$$\dot{q}^i = \{q^i, H\} \qquad \text{and} \qquad \dot{p}_i = \{p_i, H\}$$

for every $1 \leq i \leq n$, where q^i and p_i are considered to be the coordinate functions on P. Then, for every solution $\gamma : t \longmapsto (q^1(t), \ldots, q^n(t), p_1(t), \ldots, p_n(t))$ of the differential equations (4.1), one has:

$$\frac{d(f\circ\gamma)}{dt}(t) = \{f, H\}(\gamma(t))$$

for any smooth function $f \in \mathcal{C}^\infty(P)$. Then, the Hamiltonian defines a vector field $X_H = \{H, .\,\}$ on P, whose integral curves describe the time evolution of the physical system.

The Poisson bracket is central in Hamilton's description of classical mechanics: Poisson had already noticed that the set of functions which are invariant along the integral curves of X_H – the so-called *constants of motion* – is stable under Poisson bracket. Liouville then showed that the existence of a set of n independent constants of motion commuting under the Poisson bracket allows to integrate Hamilton's equations. This result was then later improved by the infamous *action-angle theorem* which, in the situation where the leaves of the constants of motion are compact, provides a distinguished choice of local coordinates which are such that the Hamiltonian takes a very specific and nice form. The Poisson bracket on $\mathbb{R}^{2n}$ can be generalized to smooth manifolds and the aim of this chapter is to show that there are several deep mathematics that are raised by this new notion.

4.1 Poisson manifolds

Keeping in mind the correspondence between algebra and geometry, we first emphasize that Poisson geometry relies on the notion of *Poisson algebra*. Recall that every associative algebra $(A, \cdot)$ gives rise to a Lie bracket:

$$[a,b] = a\cdot b - b\cdot a \tag{4.3}$$

In particular, because of the associativity, a short computation shows that this Lie bracket is a derivation of the associative product:

$$[a, b\cdot c] = [a,b]\cdot c + b\cdot [a,c] \tag{4.4}$$

[14] This section relies on four main sources: [Dufour and Zung, 2005], [Laurent-Gengoux et al., 2013] and [Vaisman, 1994], [Crainic et al., 2021] as well as these lectures notes.

However, the right hand side of Equation (4.3) is trivial when the associative product is commutative, and hence the Lie bracket vanishes. Then, a non-trivial Lie bracket on such a commutative associative algebra should necessarily form exterior, additional data. A Poisson algebra is precisely such an object, where the commutative associative product is compatible with the Lie bracket so that they satisfy Equation (4.4):

Definition 4.1. *A* Poisson algebra *is a $\mathbb{R}$-vector space A equipped with two bilinear products $\cdot$ and $\{\,.\,,\,.\,\}$, such that:*

1. *$(A,\cdot)$ is a commutative associative algebra;*
2. *$(A,\{\,.\,,\,.\,\})$ is a Lie algebra;*
3. *the Lie bracket is a derivation of the associative product:*

$$\{a, b\cdot c\} = \{a,b\}\cdot c + b\cdot\{a,c\}$$

for any elements $a,b,c\in A$.

We call $\{\,.\,,\,.\,\}$ a Poisson bracket*. A* morphism of Poisson algebras *is a map $\phi: A\longrightarrow B$ that is both a morphism of associative algebras and a morphism of Lie algebras.*

Since a Poisson algebra has two main algebraic structures, there are several kinds of ideals and subalgebras: we need to carefully emphasize which product is used in their definitions. We use the denomination *ideal* and *subalgebra* when we refer to these algebraic structure defined with the help of the associative product, and we use the denomination *Lie ideal* and *Lie subalgebra* when we refer to the Poisson bracket. A Poisson ideal (resp. subalgebra) is an ideal (resp. subalgebra) with respect to the associative product *and* to the Poisson bracket. This notions will have a geometric counterpart when we study Poisson manifolds and their submanifolds.

Definition 4.2. *A* Poisson manifold *is a smooth manifold M together with a $\mathbb{R}$-bilinear Lie bracket $\{\,.\,,\,.\,\}$ on the commutative associative algebra of smooth functions $\mathcal{C}^\infty(M)$ which makes it a Poisson algebra. The bracket $\{\,.\,,\,.\,\}$ is called a* Poisson structure on M*. A* Poisson morphism *between two Poisson manifolds M and N is a smooth map $\varphi: M\longrightarrow N$ such that the pullback $\varphi^*:\mathcal{C}^\infty(N)\longrightarrow\mathcal{C}^\infty(M)$ is a morphism of Poisson algebras.*

Example 4.3. The phase space $P=\mathbb{R}^{2n}$, parametrized with the generalized coordinates q^i and their conjugate momenta p_i, together with the Poisson bracket defined in Equation (4.2). In that particular case, the Poisson bracket actually descends from the canonical symplectic structure $\omega=\sum_{i=1}^n dp_i\wedge dq^i$. Then the so-called canonical transformations correspond to Poisson isomorphisms (which in the present case coincide with *symplectomorphisms*).

Example 4.4. To every finite dimensional Lie algebra $(\mathfrak{g},[\,.\,,\,.\,])$ we can associate a *linear Poisson structure* on $\mathfrak{g}^*$. Elements of $\mathfrak{g}$ can then be seen as linear forms on the dual $\mathfrak{g}^*$: indeed, every $x\in\mathfrak{g}$ defines a linear map:

$$\begin{aligned}\overline{x}:\ \mathfrak{g}^* &\longrightarrow \mathbb{R}\\ \xi &\longmapsto \xi(x)\end{aligned}$$

Let $(e_1,\ldots,e_n)$ be a basis of $\mathfrak{g}$. Then by the above assignment they define a system of linear coordinates on $\mathfrak{g}^*$, denoted $\overline{e_1},\ldots,\overline{e_n}$. Every real analytic function on $\mathfrak{g}^*$ can then be expressed in terms of such coordinates functions, and every smooth function on $\mathfrak{g}^*$ can be differentiated with respect to these coordinates. In particular, the commutators $[e_i,e_j]={C_{ij}}^k\,e_k$ define a linear function $\eta_{ij} = \overline{[e_i, e_j]} = C_{ij}{}^k\, \overline{e_k}$ on $\mathfrak{g}^*$. These data allow to define a Poisson structure on the dual space $\mathfrak{g}^*$, called the *linear Poisson structure of* $\mathfrak{g}^*$:

$$\{f, g\} = \sum_{1 \leq i,j \leq n} \eta_{ij} \frac{\partial f}{\partial \overline{e_i}} \frac{\partial g}{\partial \overline{e_j}}$$

for every $f, g \in \mathcal{C}^\infty(\mathfrak{g}^*)$. It is the unique Poisson structure on $\mathfrak{g}^*$ that satisfies the following identity:

$$\{\overline{x}, \overline{y}\} = \overline{[x, y]}$$

for every $x, y \in \mathfrak{g}$.

Example 4.5. Example 4.4 applies for example to $\mathfrak{so}(3)$: let e_1, e_2, e_3 its generators, and $[e_i, e_j] = \sum_{k=1}^3 \epsilon_{ijk} e_k$ be the Lie bracket, where ϵ_{ijk} is the Levi-Civita symbol on three elements. Denoting $X = \overline{e_1}$, $Y = \overline{e_2}$ and $Z = \overline{e_3}$, the corresponding linear Poisson structure on $\mathfrak{so}(3)^*$ satisfies:

$$\{X, Y\} = Z, \qquad \{Y, Z\} = X, \qquad \{Z, X\} = Y.$$

Example 4.6. One can change the former example to the following: instead of $\{X, Y\} = Z$, set $\{X, Y\} = -Z^2 + \frac{1}{4}$, and preserve the other two brackets. This choice defines a non-linear Poisson structure on $\mathfrak{so}(3)^* \simeq \mathbb{R}^3$.

Example 4.7. Example 4.4 extends to Lie algebroids: every Lie algebroid structure on a vector bundle A induces a Poisson manifold structure on A^*. In particular this implies that for every smooth manifold, T^*M is a Poisson manifold. If M is the configuration space associated to a given physical system, with local coordinates $q^1, \ldots, q^n$, then the cotangent bundle T^*M is considered to be the associated phase space, admitting fiberwise local coordinates $p_1, \ldots, p_n$, i.e. $p_k(dq^l) = \delta_k^l$. The Poisson bracket on T^*M is then the canonical one, defined in Equation (4.2).

Let us now deduce some properties of a given Poisson bracket $\{\,.\,,.\,\}$ on a Poisson manifold M. Vector fields on M which are derivations of the Poisson bracket are called *Poisson vector fields*. More precisely, such a vector field X satisfies the following identity:

$$X(\{f, g\}) = \{X(f), g\} + \{f, X(g)\} \tag{4.5}$$

for every $f, g \in \mathcal{C}^\infty(M)$. Given a Poisson bracket, it is not straightforward to deduce which vector fields are Poisson vector fields. However, it turns out that a subclass of those are easily obtained. Recall from Remark 2.14 that to any element x of a Lie algebra one can associate a derivation, called the adjoint action of x, denoted $\mathrm{ad}_x = [x, -]$. This remark applies to Poisson algebras, since they are particular cases of Lie algebras. In particular, let us study how this materializes in $\mathcal{C}^\infty(M)$, when the smooth manifold M is a Poisson manifold. For every $f \in \mathcal{C}^\infty(M)$, we call $X_f = \mathrm{ad}_f = \{f, .\}$ the *Hamiltonian vector field associated to* f. In particular, for any two smooth functions $f, g \in \mathcal{C}^\infty(M)$ we have:

$$dg(X_f) = X_f(g) = \{f, g\} \tag{4.6}$$

Equation (4.6), together with the Jacobi identity imply that the Hamiltonian vector fields have the following nice property:

$$[X_f, X_g] = X_{\{f,g\}} \tag{4.7}$$

In other words the linear map $\mathcal{C}^\infty(M) \longrightarrow \mathfrak{X}(M)$ sending a smooth function to its hamiltonian vector field is a morphism of Lie algebras. Another useful application of Equation (4.6) is in showing that Hamiltonian vector fields are Poisson vector fields because the Jacobi identity for the Poisson bracket can be written as:

$$X_h(\{f, g\}) = \{X_h(f), g\} + \{f, X_h(g)\}$$

for every smooth functions f, g, h. A smooth function of $\mathcal{C}^\infty(M)$ whose hamiltonian vector field is zero is called a *Casimir element*, because it commutes with any other element of the algebra. When M is connected, constant functions on M are always Casimir elements. In Lie theoretic words, the space of Casimir elements corresponds to the center of the Lie algebra $(\mathcal{C}^\infty(M), \{\,.\,,.\,\})$. There may then exists many linearly independent such objects.

Exercise 4.8. Show that the Poisson structure on $\mathbb{R}^3$ as defined in Example 4.6 indeed satisfies Equation (4.7).

Exercise 4.9. Show that the function $C = X^2 + Y^2 + Z^2$ is a Casimir element of the linear Poisson structure on $\mathfrak{so}(3)^*$. Turning to the non-linear structure defined in Example 4.6, check that it admits as a Casimir element (together with constant functions):

$$C = X^2 + Y^2 - \frac{2}{3}Z^3 + \frac{1}{2}Z$$

Given a Poisson bracket, it is not at all evident to deduce which function are Casimir, and which vector fields are Poisson vector fields (up to hamiltonian vector fields). We will give a partial answer to this question using cohomological techniques. The mathematical machinery set up to describe this so called *Poisson cohomology* will eventually provide another, more geometric point of view on Poisson brackets. We first need to generalize the Lie bracket of vector fields to the whole graded algebra $\mathfrak{X}^\bullet(M) = \bigoplus_{i=1}^n \mathfrak{X}^i(M)$. Recall that the space $\mathfrak{X}^i(M)$ represents the sheaf of smooth sections of the vector bundle $\bigwedge^i TM$. Every multivector field is locally decomposable because $\bigwedge^i TM$ admits elements of the form $\partial_{k_1} \wedge \ldots \wedge \partial_{k_i}$ (for $1 \leq k_1 < \ldots < k_i \leq n$) as local frames. Evaluating an element of $\mathfrak{X}^i(M)$ on i smooth functions – which gives back another smooth function – is then done by using Equation (1.17). Moreover, the pair $(\mathfrak{X}^\bullet(M), \wedge)$ is a graded commutative algebra:

$$\mathfrak{X}^i(M) \wedge \mathfrak{X}^j(M) \subset \mathfrak{X}^{i+j}(M)$$

More precisely, for $P \in \mathfrak{X}^i(M)$ and $Q \in \mathfrak{X}^j(M)$ and $i + j$ smooth functions $f_1, \ldots, f_{i+j}$, one has:

$$P \wedge Q(f_1, \ldots, f_{i+j}) = \sum_{\sigma \in Un(i,j)} (-1)^\sigma P(f_{\sigma(1)}, \ldots, f_{\sigma(i)})\, Q(f_{\sigma(i+1)}, \ldots, f_{\sigma(i+j)})$$

where $Un(p, n-p)$ represents the set of $(p, n-p)$-*unshuffles* (other people call it *shuffles*), i.e. those permutations $\sigma \in S_n$ satisfying the following two unshuffling conditions:

$$\sigma(1) < \sigma(2) < \ldots < \sigma(p) \qquad \text{and} \qquad \sigma(p+1) < \sigma(p+2) < \ldots < \sigma(n-1) < \sigma(n)$$

At level 0, i.e. for $\mathfrak{X}^0(M) = \mathcal{C}^\infty(M)$, we understand that wedging with respect to a smooth function f consists in multiplying by this function: $f \wedge P = fP$, for any $P \in \mathfrak{X}^\bullet(M)$.

Example 4.10. Let $P = \partial_x \wedge \partial_y \wedge \partial_z$ be a multivector field on $M = \mathbb{R}^3$. For any three smooth functions f, g, h, we have:

$$P(f, g, h) = \partial_x f \partial_y g \partial_z h - \partial_x f \partial_y h \partial_z g + \circlearrowleft$$

where $\circlearrowleft$ symbolizes circular permutation of the three functions.

While vector fields on a smooth manifold are derivations of smooth functions, multivector fields are multiderivations: $\mathfrak{X}^i(M) \simeq \mathrm{Der}^i(\mathcal{C}^\infty(M))$ (see Lemma 1.2.2 in [Dufour and Zung, 2005]). By multiderivation, we mean the following: for every $P \in \mathfrak{X}^i(M)$ and $f_1, \ldots, f_i, g \in \mathcal{C}^\infty(M)$, we have:

$$P(f_1, \ldots, f_i\, g) = P(f_1, \ldots, f_i)\, g + f_i\, P(f_1, \ldots, g)$$

In particular, since P is fully skew-symmetric with respect to permutations of its variables, the derivation property is true for every slot. Multiderivations can be composed: for $P \in \mathfrak{X}^i(M)$ and $Q \in \mathfrak{X}^j(M)$ two multiderivations (here $1 \leq i, j \leq n$), the composite $P \circ Q$ is not a multiderivation, but a priori no more than a multi-operator on $\mathcal{C}^\infty(M)$ (this was already observed for mere vector fields). More precisely, it acts on $i+j-1$ smooth functions $f_1, \ldots, f_{i+j-1}$ as:

$$P \circ Q(f_1, \ldots, f_{i+j-1}) = \sum_{\sigma \in Un(j,i-1)} (-1)^\sigma P\big(Q(f_{\sigma(1)}, \ldots, f_{\sigma(j)}), f_{\sigma(j+1)}, \ldots, f_{\sigma(i+j-1)}\big) \tag{4.8}$$

while if Q is a smooth function (whatever P is) then we set $P \circ Q = 0$. Equation (4.8) shows that although P has degree i and Q has degree j, the composite $P \circ Q$ does not respect this graduation because it has $i+j-1$ arguments. This is why we decide to create a new grading on $\mathfrak{X}^\bullet(M)$, by shifting the original grading by -1. We denote by $\mathcal{V}^i(M)$ (for $-1 \leq i \leq n-1$) the vector space $\mathfrak{X}^{i+1}(M)$ shifted by a degree -1:

$$\mathcal{V}^i(M) = \mathfrak{X}^{i+1}(M)$$

In other words, we have the following correspondence:

$$\begin{array}{ccccc}
\mathcal{V}^{-1}(M) & \mathcal{V}^{0}(M) & \mathcal{V}^{1}(M) & \ldots & \mathcal{V}^{n-1}(M) \\
\| & \| & \| & & \| \\
\underbrace{\mathfrak{X}^0(M)}_{\mathcal{C}^\infty(M)} & \mathfrak{X}^1(M) & \mathfrak{X}^2(M) & \ldots & \mathfrak{X}^n(M)
\end{array}$$

In particular, smooth functions now belong to $\mathcal{V}^{-1}(M) = \mathfrak{X}^0(M)$, vector fields belong to $\mathcal{V}^0(M) = \mathfrak{X}^1(M)$, and multivector fields of degree i belong to $\mathcal{V}^{i-1}(M)$. We label by $\overline{P}$ the degree (with respect to the new convention, in $\mathcal{V}^\bullet(M)$) of the homogeneous element P. In particular, if $P \in \mathfrak{X}^i(M)$, we have $\overline{P} = i-1$. Given these conventions, we set:

Definition 4.11. *The* Schouten-Nijenhuis bracket *is the $\mathbb{R}$-bilinear graded skew-symmetric bracket on $\mathcal{V}^\bullet(M) = \bigoplus_{i=-1}^{n-1} \mathcal{V}^i(M)$ defined by its action on any two homogeneous multivector fields P, Q of degree ≥ 0:*

$$[P, Q]_{SN} = P \circ Q - (-1)^{\overline{P} \cdot \overline{Q}} Q \circ P \tag{4.9}$$

while, for any function $f \in \mathcal{C}^\infty(M)$:

$$[P, f]_{SN} = P(f, \ldots) \tag{4.10}$$

Remark 4.12. Equation (4.10) means that, if for example in local coordinates $P = P^{i_1 \ldots i_k} \partial_{x^{i_1}} \wedge \ldots \wedge \partial_{x^{i_k}}$ (summation implied), then:

$$[P, f]_{SN} = \sum_{i_1, \ldots, i_k} (-1)^{j+1} P^{i_1 \ldots i_k} \partial_{x^{i_j}}(f) \partial_{x^{i_1}} \wedge \ldots \wedge \widehat{\partial_{x^{i_j}}} \wedge \ldots \wedge \partial_{x^{i_k}}$$

where $\widehat{\partial_{x^{i_j}}}$ means that we omit this term in the wedge product of $k-1$ partial derivatives.

Being *graded skew-symmetric* means that for any two homogeneous multivector fields P, Q, one has:

$$[P, Q]_{SN} = -(-1)^{\overline{P} \cdot \overline{Q}} [Q, P]_{SN} \tag{4.11}$$

This implies in particular that, when one considers P, Q as elements of $\mathfrak{X}^\bullet(M)$ of respective degrees i and j, the Schouten-Nijenhuis bracket reads:

$$[P, Q]_{SN} = -(-1)^{(i-1)(j-1)} [Q, P]_{SN}$$

We see that this definition of the bracket – although equivalent to Equation (4.9) – is not so convenient because of the exponents that do not match the degrees of P and Q. For degree reasons, the bracket of two functions is zero because the sum of their degrees is -2, and the graded vector space $\mathcal{V}^\bullet(M)$ does not possess a vector space of degree -2.

A more explicit formula for Equation (4.9) when P and Q are decomposable multivector fields, may be the following:

$$[X_1\wedge\ldots\wedge X_i, Y_1\wedge\ldots\wedge Y_j]_{SN} = \sum_{\substack{1\leq k\leq i\\ 1\leq l\leq j}} (-1)^{k+l}[X_k, Y_l]\wedge X_1\wedge\ldots\wedge\widehat{X_k}\wedge\ldots\wedge X_i\wedge Y_1\wedge\ldots\wedge\widehat{Y_l}\wedge\ldots\wedge Y_j \tag{4.12}$$

together with, for Equation (4.10):

$$[X_1\wedge\ldots\wedge X_i, f]_{SN} = \sum_{k=1}^{i}(-1)^{k+1}X_i(f)X_1\wedge\ldots\wedge\widehat{X_k}\wedge\ldots\wedge X_i$$

for every vector fields $X_1,\ldots,X_i,Y_1,\ldots,Y_j$, and smooth function $f\in\mathcal{C}^\infty(M)$. The latter expression is convenient because we then have:

$$[X,f]_{SN} = X(f) \quad \text{and} \quad [X\wedge Y, f]_{SN} = X(f)Y - Y(f)X$$

Exercise 4.13. Using Equation (1.17), you may check the identity between formula (4.9) and (4.12) on small decomposable multivector fields, such as $P = X$ and $Q = Y_1\wedge Y_2$.

The Schouten-Nijenhuis bracket has several nice properties. In particular it coincides with the Lie bracket on vector fields when $P,Q\in\mathfrak{X}^1(M)$, since in that case $\overline{P} = \overline{Q} = 0$. It is thus legitimate to wonder whether this bracket generalizes the notion of Lie algebra to that of a *graded Lie algebra* on the graded vector space $\mathcal{V}^\bullet(M)$.

Definition 4.14. *A* graded Lie algebra *is a graded vector space $V = \bigoplus_{i\in\mathbb{Z}} V_i$, equipped with a $\mathbb{R}$-bilinear aperation $[\,.\,,.\,] : V\times V\longrightarrow V$ called a* graded Lie bracket*, and which satisfies the following identities:*

graded skew-symmetry	$[x,y] = -(-1)^{\lvert x\rvert\lvert y\rvert}[y,x]$
graded Jacobi identity	$[x,[y,z]] = [[x,y],z] + (-1)^{\lvert x\rvert\lvert y\rvert}[y,[x,z]]$

for every $x,y,z\in\mathfrak{V}$. A derivation of degree d *of V is an endomorphism $\delta : V^\bullet\longrightarrow V^{\bullet+d}$ such that:*

$$\delta([x,y]) = [\delta(x),y] + (-1)^{\lvert x\rvert d}[x,\delta(y)]$$

We denote $\mathrm{Der}^d(V)$ *the vector space of derivations of degree d of V.*

Remark 4.15. Another, more symmetric form of the graded Jacobi identity exists, but it is not very convenient to use:

$$(-1)^{\lvert x\rvert\lvert z\rvert}[x,[y,z]] + (-1)^{\lvert y\rvert\lvert x\rvert}[y,[z,x]] + (-1)^{\lvert z\rvert\lvert y\rvert}[z,[x,y]] = 0$$

Moreover, the graded Jacobi identity appearing in Definition 4.14 shows that the adjoint action of any element of V is a derivation of V: $\mathrm{ad}_x\in\mathrm{Der}^{\lvert x\rvert}(V)$, for every $x\in V$.

The two conditions satisfied by the graded Lie bracket in Definition 4.14 are slight generalizations of what characterizes a Lie algebra because Lie algebras are graded Lie algebras concentrated in degree 0. The idea with grading is very intuitive: for any two homogeneous elements $x,y\in E$, when we swap x and y to form a new term (either in the bracket or by 'jumping' over), we add a sign $(-1)^{|x||y|}$ in front of the new term hence created, compared to the classical (non-graded) situation. You can see this phenomenon on the right-hand sides of both equations. The same phenomenom happens for differential forms: $\eta\wedge\mu = (-1)^{|\eta||\mu|}\mu\wedge\eta$, making the wedge product *graded commutative* (and not merely commutative). Here, this has interesting consequences: in a graded Lie algebra, we do not necessarily have $[x, x] = 0$ because the graded Lie bracket is not skew-symmetric anymore when $|x|$ is odd because $[x, x] = -(-1)^{1\times1}[x, x]$ hence we cannot conclude on the vanishing of $[x, x]$.

Exercise 4.16. Using the graded Jacobi identity, show that if $x = y$ and $|x|$ is odd, we have:

$$[x, [x, z]] = \frac{1}{2}[[x, x], z]$$

These observations enable us to formulate the following important result:

Proposition 4.17. *The Schouten-Nijenhuis bracket extends the Lie bracket of vector fields to a graded Lie algebra structure on $\mathcal{V}^\bullet(M)$. In particular, the Scouten-Nijenhuis bracket satisfies:*

$$[P, fQ]_{SN} = [P, f]_{SN}Q + f[P, Q]_{SN}$$

for any smooth function f and multivector fields P and Q.

Proof. First of all, one needs to check that the graded vector space $\mathcal{V}^\bullet(M)$ is stable under this bracket. This can be proven on decomposable vector fields, using Equation (4.12). From Equation (4.11), the bracket is obviously graded skew-symmetric. It is just a matter of computation to check with Equation (4.12) that it satisfies the graded Jacobi identity (on decomposable vector fields). See Theorem 1.8.1 in [Dufour and Zung, 2005] for more details. ☐

Remark 4.18. Actually, the Schouten-Nijenhuis bracket is the unique extension of the Lie bracket of vector fields to a graded Lie bracket on the space of alternating multivector fields that makes it into a Gerstenhaber algebra.

The Schouten-Nijenhuis bracket on vector fields allows us to characterize Poisson structures in a more geometric flavored approach. Let $(M, \{\,.\,,.\,\})$ be a Poisson manifold and let $x^1, \ldots, x^n$ be local coordinates on M. Then the Poisson bracket between two functions f, g is locally of the form (see Proposition 1.14 in [Fernandes, 2005]):

$$\{f, g\} = \sum_{1\leq i,j\leq n} \{x^i, x^j\}\frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j} \tag{4.13}$$

This Equation is valid locally in the coordinate neighborhood of any point of a smooth manifold, and it is invariant under change of coordinates $x^i \longmapsto x'^k$. Indeed, using Equations (3.6)-(3.7), we have that (omitting the sum signs):

$$\{x^i, x^j\}\frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j} = \{x^i, x^j\}\frac{\partial x'^k}{\partial x^i}\frac{\partial x'^l}{\partial x^j}\frac{\partial f}{\partial x'^k}\frac{\partial g}{\partial x'^l} = \{x'^k, x'^l\}\frac{\partial f}{\partial x'^k}\frac{\partial g}{\partial x'^l}$$

Then we see from Equation (4.13) that the Poisson bracket can be locally seen as a bivector field $\frac{1}{2}\{x^i, x^j\}\,\partial_i\wedge\partial_j$ which, when evaluated on two smooth functions f, g, give $\{f, g\}$, as the following short calculation (where we have omitted the sum signs) shows:

$$\frac{1}{2}\{x^i, x^j\}\,\partial_i\wedge\partial_j(f, g) = \frac{1}{2}\{x^i, x^j\}\Big(\partial_i f\partial_j g - \partial_i g\partial_j f\Big) = \{x^i, x^j\}\frac{\partial f}{\partial x^i}\frac{\partial g}{\partial x^j}$$

We used Equation (1.18) between the first and the second step. We denote π the unique bivector field whose component in local coordinates is $\pi^{ij} = \{x^i, x^j\}$. Thus, the Poisson bracket uniquely defines a bivector field $\pi \in \mathfrak{X}^2(M)$ via the following identity:

$$\pi(f, g) = \{f, g\} \tag{4.14}$$

Exercise 4.19. By applying the Jacobi identity satisfied by the Poisson bracket to the coordinate functions x^i, x^j, x^k, show that the components of the bivector field π satisfies (this is a local expression):

$$\sum_{s=1}^{n} \pi^{is}\frac{\partial \pi^{jk}}{\partial x^s} + \pi^{js}\frac{\partial \pi^{ki}}{\partial x^s} + \pi^{ks}\frac{\partial \pi^{ij}}{\partial x^s} = 0 \tag{4.15}$$

Obviously, not every bivector field satisfies Equation (4.15). However, those that satisfy it define a Poisson structure on M via Equation (4.14). This translates as the following fundamental fact, due to Lichnerowicz:

Proposition 4.20. *There is a one-to-one correspondence between Poisson structures on a smooth manifold M and bivector fields $\pi \in \mathfrak{X}^2(M)$ such that:*

$$[\pi, \pi]_{SN} = 0 \tag{4.16}$$

Exercise 4.21. Prove that $[\pi, \pi]_{SN} = 0$ is equivalent to Equation (4.15), when evaluated in local coordinates.

Remark 4.22. By the correspondence established by Proposition (4.20), we will now either use the notation $(M, \{\,.\,,.\,\})$ or (M, π) (depending on the context) to denote a Poisson manifold.

Example 4.23. The bivector field associated to the canonical Poisson bracket of Example 4.3 is the following:

$$\pi = \sum_{i=1}^{n} \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i} \tag{4.17}$$

So, in particular, if one relabels the coordinates as $x^i = q^i$ and $x^{n+i} = p_i$ for $1 \leq i \leq n$, then $\pi^{ij}(q,p) = 0$ except when $j = i+n$ or $i+n = j$, and in that case we have $\pi^{i(i+n)} = 1$ and $\pi^{(i+n)n} = -1$, so that $\pi = \frac{1}{2}\pi^{ij}\partial_i \wedge \partial_j$.

Example 4.24. On $\mathbb{R}^3$ one picks the following Poisson bracket:

$$\{f, g\} = x\frac{\partial f}{\partial x}\frac{\partial g}{\partial z} + y\frac{\partial f}{\partial y}\frac{\partial g}{\partial z} - x\frac{\partial g}{\partial x}\frac{\partial f}{\partial z} - y\frac{\partial g}{\partial y}\frac{\partial f}{\partial z}$$

This Poisson bracket corresponds to the following Poisson bivector field:

$$\pi = \left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right) \wedge \frac{\partial}{\partial z}$$

Exercise 4.25. Check that the Poisson bivector defined in Example 4.23 indeed satisfies Equation (4.14), where the Poisson bracket is that of Equation (4.2).

Exercise 4.26. Given two Poisson structures π_0 and π_1 on a smooth manifold M, show that, if $\pi_t = (1-t)\pi_0 + t\pi_1$ is a Poisson structure for some $t \neq 0, 1$, then it is a Poisson structure for all $t \in \mathbb{R}$. We then call the smooth family $(\pi_t)_t$ a *Poisson pencil.*

Seen from $\mathfrak{X}^\bullet(M)$, bivectors fields have degree 2, while seen from $\mathcal{V}^\bullet(M)$, they have degree 1. Bivector fields satisfying Equation (4.16) are called *Poisson bivector fields* (not to be confused with the Poisson vector fields defined in Equation (4.5)). We now show how a Poisson bivector field makes $\mathcal{V}^\bullet(M)$ a chain complex. Let $d_\pi : \mathcal{V}^\bullet(M) \longrightarrow \mathcal{V}^{\bullet+1}(M)$ be the unique $\mathbb{R}$-linear morphism defined on any element $P \in \mathcal{V}^\bullet(M)$ as:

$$d_\pi(P) = [\pi, P]_{SN} \tag{4.18}$$

This operator is well defined, and is indeed of degree 1; it corresponds to the adjoint action of π on the graded Lie algebra $\mathcal{V}^\bullet(M) = \bigoplus_{i=-1}^{n-1} \mathcal{V}^i(M)$. Moreover, the graded Jacobi identity (via Exercise 4.16) together with Equation (4.16) imply that d_π squares to zero:

$$d_\pi^2(P) = [\pi, [\pi, P]_{SN}]_{SN} = \frac{1}{2}[[\pi, \pi]_{SN}, P]_{SN} = 0$$

This operator is often called the *Poisson differential.* These successive facts imply that $(\mathcal{V}^\bullet(M), d_\pi)$ is a chain complex:

$$0 \longrightarrow \underbrace{\mathcal{V}^{-1}(M)}_{\mathcal{C}^\infty(M)} \xrightarrow{d_\pi} \underbrace{\mathcal{V}^{1}(M)}_{\mathfrak{X}(M)} \xrightarrow{d_\pi} \mathcal{V}^1(M) \xrightarrow{d_\pi} \dots \xrightarrow{d_\pi} \mathcal{V}^{n-1}(M) \longrightarrow 0$$

Notice that the above results can equivalently be expressed with respect to the grading on $\mathfrak{X}^\bullet(M)$. The Poisson differential is still defined from Equation (4.18), at the cost of expressing the Schouten-Nijenhuis with respect to the grading of $\mathfrak{X}^\bullet(M)$, via Equation (4.11). Then, we obtain that $(\mathfrak{X}^\bullet(M), d_\pi)$ is a chain complex concentrated in degrees $0, \dots, n$. This is the natural setup to define a cohomology theory:

Definition 4.27. *Let (M, π) be a Poisson manifold. The cohomology of the chain complex $(\mathfrak{X}(M), d_\pi)$ is called the* Poisson cohomology *of (M, π) and is denoted, for $0 \leq i \leq n$:*

$$H^i_\pi(M) = \frac{\mathrm{Ker}(d_\pi : \mathfrak{X}^i(M) \longrightarrow \mathfrak{X}^{i+1}(M))}{\mathrm{Im}(d_\pi : \mathfrak{X}^{i-1}(M) \longrightarrow \mathfrak{X}^i(M))}$$

The map d_π is called the Poisson differential.

Notice that Equation (4.18) can be rewritten in a way that shows a huge similarity with de Rham differential (2.31) (and Chevalley-Eilenberg differential as well):

$$\begin{aligned}(-1)^{m-1}\, d_\pi(P)(f_1, \dots, f_m, f_{m+1}) = &\sum_{i=1}^{m+1} (-1)^{i-1} X_{f_i}(P(f_1, \dots, \widehat{f_i}, \dots, f_{m+1})) \\ &+ \sum_{1\leq i<j\leq m+1} (-1)^{i+j} P(\{f_i, f_j\}, f_1, \dots, \widehat{f_i}, \dots, \widehat{f_j}, \dots, f_{m+1})\end{aligned} \tag{4.19}$$

for every $P \in \mathfrak{X}^m(M)$. The sign $(-1)^m$ could have been got rid of if the Poisson differential had been defined following the alternative, although equivalent, convention: $d_\pi(P) = -[P, \pi]_{SN}$. As can be explicitly be seen in Equation (4.19), the Poisson differential carries information on the Poisson structure on M. The next subsection clarifies the relationship between de Rham cohomology and Poisson cohomology.

Exercise 4.28. Show that the map d_π is a derivation of the Schouten-Nijenhuis bracket.

Let us compute the first few cohomology groups. The 0-th Poisson cohomology is given by:

$$H^0_\pi(M) = \mathrm{Ker}(d_\pi : \mathcal{C}^\infty(M) \longrightarrow \mathfrak{X}(M))$$

By definition of the Schouten-Nijenhuis bracket involving functions, we have $d_\pi(f) = [\pi, f]_{SN} = -\pi(f, -) = -\{f, -\}$. Then, the smooth functions that belong to $H^0_\pi(M)$ consists of those that are such that $\{f, g\} = 0$, i.e. they are Casimir elements of the Lie algebra $(\mathcal{C}^\infty(M), \{\,.\,,\,.\,\})$:

$$H^0_\pi(M) = \text{Casimir elements of } (\mathcal{C}^\infty(M), \{\,.\,,\,.\,\})$$

The dimension of $H^0_\pi(M)$ as a vector space is at least 1, because constant functions on M are Casimir elements (assuming M is connected). Going to the next level, we have:

$$H^1_\pi(M) = \frac{\mathrm{Ker}(d_\pi : \mathfrak{X}(M) \longrightarrow \mathfrak{X}^2(M))}{\mathrm{Im}(d_\pi : \mathcal{C}^\infty(M) \longrightarrow \mathfrak{X}(M))}$$

Elements of $\mathrm{Ker}(d_\pi : \mathfrak{X}(M) \longrightarrow \mathfrak{X}^2(M))$ are characterized by the following property: they are vector fields X such that $[\pi, X]_{SN} = 0$. It corresponds to Equation (4.5), hence such vector fields are Poisson vector fields. The space $\mathrm{Im}(d_\pi : \mathcal{C}^\infty(M) \longrightarrow \mathfrak{X}(M))$ consists of Hamiltonian vector fields on M. These are Poisson vector fields of a particular kind: they are somehow "trivial" in the sense that are the easiest Poisson vector fields to find, for they are automatically given as soon as a Poisson structure is defined. The interesting Poisson vector fields are thus those that are not hamiltonian or, more precisely, the classes of Poisson vector fields up to hamiltonian vector fields, which is precisely what the first cohomology group is:

$$H^1_\pi(M) = \text{classes of non-trivial Poisson vector fields}$$

So, in particular, if $H^1_\pi(M) \neq 0$ there are Poisson vector fields which are not Hamiltonian vector fields. Higher Poisson cohomology groups arise naturally in deformation theory: $H^2_\pi(M)$ may be interpreted as the moduli space of formal infinitesimal deformations of π, while $H^3_\pi(M)$ may be interpreted as the space of obstructions of such deformations [Dufour and Zung, 2005].

Example 4.29. Using the fact that $H^1_{dR}(\mathbb{R}^{2n}) = 0$, we will show in Remark 4.36 that $H^1_\pi(\mathbb{R}^{2n}) = 0$ (where the Poisson structure is the canonical one). It implies that on $\mathbb{R}^{2n}$ equipped with the Poisson bracket defined in Equation (4.2), every Poisson vector field is hamiltonian, i.e. descends from a smooth function. However, contrary to de Rham cohomology which is locally trivial on a smooth manifold, Poisson cohomology needs not be locally trivial on a Poisson manifold because the Poisson structure needs not be non-degenerate.

An alternative view on Poisson vector fields can be made through *Lie derivatives*. First, define the Lie derivative of a vector field Y along the vector field X by the Lie bracket:

$$\mathcal{L}_X(Y) = [X, Y] \tag{4.20}$$

Then, to be consistent with Schouten-Nijenhuis bracket, it implies that on smooth functions, Lie derivatives act as derivations:

$$\mathcal{L}_X(f) = [X, f]_{SN} = X(f) \tag{4.21}$$

More generally, on any multivector field P, the Lie derivative acts as:

$$\mathcal{L}_X(P) = [X, P]_{SN}$$

Then, one notices that the condition (4.5) of X being a Poisson vector field (with respect to the Poisson bivector π) is equivalent to the following equality:

$$\mathcal{L}_X(\pi) = 0$$

Exercise 4.30. Show that the condition that a bivector field B is a Poisson bivector is equivalent to the following identity:

$$\mathcal{L}_{B^\sharp(df)}(B) = 0 \quad \text{for every smooth function } f$$

To conclude this subsection, we show that the Lie derivative can naturally act on differential forms. Using Equations (4.20) and (4.21), one can deduce that the Lie derivative $\mathcal{L}_X$ of a vector field X acts on differential one-forms since it should satisfy a kind of 'derivation property':

$$X(\xi(Y)) = \mathcal{L}_X(\xi)(Y) + \xi(\mathcal{L}_X(Y)) \tag{4.22}$$

for every $\xi \in \Omega^1(M)$ and $Y \in \mathfrak{X}(M)$. Defining the *interior product on differential forms* as the linear operator defined, for every $x \in \mathfrak{X}(M)$, as:

$$\begin{aligned} \iota_X : \Omega^\bullet(M) &\longrightarrow \Omega^{\bullet-1}(M) \\ \eta &\longmapsto \iota_X\eta = \eta(X, \ldots) \end{aligned}$$

we notice that Equation (4.22) is equivalent to writing:

$$\mathcal{L}_X(\xi)(Y) = X(\xi(Y)) - \xi([X,Y]) = Y(\xi(X)) + d\xi(X,Y) = (d\iota_X(\xi) + \iota_X d\xi)(Y)$$

This allows us to find the following characterization:

Definition 4.31. *The Lie derivative of a vector field $X \in \mathfrak{X}(M)$ is the unique derivation of both $\mathfrak{X}^\bullet(M)$ and $\Omega^\bullet(M)$, defined on any multivector field P and differential form η as:*

$$\begin{aligned} \mathcal{L}_X(P) &= [X,P]_{SN} \\ \mathcal{L}_X(\eta) &= d\iota_X(\eta) + \iota_X d\eta \end{aligned}$$

Then, we have a nice result involving Lie derivatives, which is often used in geometry:

Proposition 4.32. *For any two vector fields $X, Y \in \mathfrak{X}(M)$, one has:*

$$[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_{[X,Y]}$$

Proof. This is a direct consequence of the properties of the Schouten-Nijenhuis bracket or the operators d and ι_X. $\square$

4.2 Properties of Poisson bivectors

Let M be a smooth manifold and let $B = \frac{1}{2}b^{ij}\partial_i \wedge \partial_j$ be a (non-necessarily Poisson) bivector field. The local components $(b^{ij})_{1\leq i,j\leq n}$ are smooth functions $b^{ij} : x \longrightarrow b^{ij}(x)$ on U. For simplicity, the latter expression $b^{ij}(x)$ will be denoted b^{ij}_x. Such functions define a smooth function:

$$\begin{aligned} \overline{B} : U &\longrightarrow \mathfrak{gl}_n(\mathbb{R}) \\ x &\longmapsto \begin{pmatrix} b^{11}_x & b^{12}_x & \ldots & b^{1n}_x \\ b^{21}_x & & & b^{2n}_x \\ \ldots & & & \ldots \\ b^{n1}_x & \ldots & \ldots & b^{nn}_x \end{pmatrix} \end{aligned}$$

where $n = \dim(M)$. Because $b^{ij} = -b^{ji}$, this function takes values in skew-symmetric $n \times n$ matrices. We say that the matrix is a *contravariant tensor* because indices appear as exponents and any change of coordinates induce a transformation of the functions b^{ij} following that of the components of vector fields: $b^{kl} = \frac{\partial x^k}{\partial x^i}\frac{\partial x^l}{\partial x^j}b^{ij}$ (summation is implicit). The *rank* of B at a point x is the rank of the corresponding matrix $\overline{B}(x)$. The rank is obviously invariant under coordinate change. We say that a bivector field is *non-degenerate* when it has maximal rank n at every point of M. Since the rank of an anti-symmetric matrix is even, it means that such situation can occur only when M is even dimensional.

To make further sense of $\overline{B}$, let us introduce the notation for the natural pairing between differential two-forms and bivector fields on M. Indeed, the pairing $\langle\, .\, ,.\, \rangle$ between covector and and tangent vectors on M can be extended to decomposable differential 2-forms and bivector fields by the following identity:

$$\langle \xi \wedge \eta, X \wedge Y \rangle = 2\Big(\xi(X)\eta(Y) - \xi(Y)\eta(X)\Big) \tag{4.23}$$

The factor 2 comes from the fact that we have two wedges products on the left-hand side, compared to Equation (1.17) for example. These conventions imply that, for a bivector B, we have:

$$B(f,g) = \frac{1}{2}\langle df \wedge dg, B\rangle \tag{4.24}$$

for any two smooth functions $f, g \in \mathcal{C}^\infty(M)$. Indeed, in local coordinates the bivector field B reads $B = \frac{1}{2}b^{ij}\partial_i \wedge \partial_j$. Then, applying Equation (4.23) to the right hand side of Equation (4.24) gives (summation on repeated indices is implicit):

$$\frac{1}{2}\langle df \wedge dg, B\rangle = \frac{1}{4}b^{ij}\langle df \wedge dg, \partial_i \wedge \partial_j\rangle = \frac{1}{2}b^{ij}(\partial_i(f)\partial_j(g) - \partial_j(f)\partial_i(g)) = \frac{1}{2}b^{ij}\partial_i \wedge \partial_j(f,g) = B(f,g)$$

Remark 4.33. Equation (4.24) straighforwardly applies to $B = \pi$ a Poisson bivector field, although in the litterature the left hand side is often written $\pi(df, dg)$ (not to be confused with $\langle df \wedge dg, B\rangle$ then).

Using the pairing between differential two-forms and bivector fields defined in Equation (4.23), the bivector field B induces a vector bundle morphism:

$$\begin{aligned} B^\sharp : T^*M &\longrightarrow TM \\ (x, \xi_x) &\longmapsto \frac{1}{2}\langle \xi_x \wedge d(-), B_x\rangle \end{aligned}$$

where B_x denotes the evaluation of the bivector field B at x. The term on the right-hand side should be read as follows:

$$\frac{1}{2}\langle \xi_x \wedge d(-), B_x\rangle : f \longmapsto \frac{1}{2}\langle \xi_x \wedge df|_x, B_x\rangle$$

One can check that it is indeed a derivation of smooth functions. More generally, evaluating any differential form η on $B^\sharp(\xi)$ corresponds to the following pairing:

$$\eta(B^\sharp(\xi)) = \frac{1}{2}\langle \xi \wedge \eta, B\rangle \tag{4.25}$$

The definition of $B^\sharp$ has been made so that, when evaluated on exact differential forms (every sufficiently local section of T^*M is exact), it is the unique vector bundle morphism satisfying:

$$B^\sharp(df) = B(f, -)$$

The right hand-side is a vector field on M (or at least an open set U), so the smooth map $B^\sharp$ indeed takes values in the tangent bundle and defines a vector bundle morphism. Since in local coordinates, the right hand side of Equation (4.25) reads $\xi_{i,x}b^{ij}_x\partial_j(f)$ – where the $\xi_{i,x}$ are the components of the covector ξ_x in the basis $dx^1, \ldots, dx^n$ – the rank of the map $B^\sharp$ is the rank of the map $\overline{B} : M \longrightarrow \mathfrak{gl}_n(\mathbb{R})$. The morphism $B^\sharp$ extends as a vector bundle morphism $\wedge^i T^*M \longrightarrow \wedge^i TM$, for $1 \leq i \leq n$, compatible with the wedge product. It means that, setting the action of $B^\sharp$ on $\Omega^0(M) = \mathfrak{X}^0(M) = \mathcal{C}^\infty(M)$ to be the identity map, $B^\sharp$ extends to a morphism of graded commutative algebras $B^\sharp : \Omega^\bullet(M) \longrightarrow \mathfrak{X}^\bullet(M)$:

$$B^\sharp(\eta \wedge \mu) = B^\sharp(\eta) \wedge B^\sharp(\mu)$$

This perspective on bivector fields is quite useful regarding the relationship between de Rham cohomology and Poisson cohomology. Indeed, if $B = \pi$ is a Poisson bivector field, then the Hamiltonian vector field associated to the smooth function f is precisely $X_f = \pi^\sharp(df)$. More generally, the vector bundle morphism $\pi^\sharp : \wedge^i T^*M \longrightarrow \wedge^i TM$ commutes with the respective differentials:

Proposition 4.34. *The graded commutative algebras morphism* $\pi^\sharp : \Omega^\bullet(M) \longrightarrow \mathfrak{X}^\bullet(M)$ *is a chain map:*

$$\pi^\sharp \circ d_{\mathrm{dR}} = d_\pi \circ \pi^\sharp$$

The proof is made by induction on the form degree, and can be found in Proposition 2.1.3 in [Dufour and Zung, 2005]. Then, closed (resp. exact) differential form are sent to closed (resp. exact) multivector fields. The chain map $\pi^\sharp : (\Omega^\bullet(M), d_{\mathrm{dR}}) \longrightarrow (\mathfrak{X}^\bullet(M), d_\pi)$ is an algebra homomorphism, and then induces a homomorphism between cohomology groups that we denote $\pi^\sharp$ as well:

Corollary 4.35. *For any Poisson manifold* (M, π)*, there is a natural homomorphism, called the* Lichnerowicz homomorphism*, between the de Rham cohomology and the Poisson cohomology:*

$$\pi^\sharp : H^\bullet_{\mathrm{dR}}(M) \longrightarrow H^\bullet_\pi(M)$$

Remark 4.36. Then, if π is a non-degenerate bivector field, the Lichnerowicz homomorphism is an isomorphism. This Corollary proves that $H^1_\pi(\mathbb{R}^{2n}) = 0$ so that every Poisson vector field on $\mathbb{R}^{2n}$ (equipped with the standard Poisson bracket) is a Hamiltonian vector field, and that this Poisson structure is 'rigid' in the sense that $H^2_\pi(\mathbb{R}^{2n}) = 0$.

The importance of the vector bundle morphism $\pi^\sharp : T^*M \longrightarrow TM$ is the following: for every $x \in M$ its image in T_xM spans the directions taken by the hamiltonian vector fields at x. As it is, this might be useless, but actually it allows us to understand that hamiltonian vector fields do not necessarily span the entire tangent space, and thus that the transport along these vector fields are constraints in some directions. Hence, for a physical hamiltonian, it means that the Poisson structure on M *constraints* the set of reachable points in the phase space, given an initial point. In particular, if the Poisson bivector is degenerate at a point x, there is no bijection between T^*_xM and T_xM and the integral curves of Hamiltonian vector fields passing through x will not be able to reach every point in the neighborhood of x. This can be explained by the fact that the vector bundle morphism $\pi^\sharp$ defines an integrable distribution, as the following proposition shows:

Proposition 4.37. *Let* (M, π) *be a Poisson manifold. Then* T^*M *is Lie algebroid – called the* cotangent Lie algebroid*, with anchor* $\pi^\sharp : T^*M \longrightarrow TM$ *and with Lie bracket:*

$$\begin{aligned} [\,.\,,.\,]_{T^*M} : \quad \Omega^1(M) \times \Omega^1(M) &\longrightarrow \Omega^1(M) \\ (\alpha, \beta) \quad &\longmapsto [\alpha, \beta]_{T^*M} = \mathcal{L}_{\pi^\sharp(\alpha)}(\beta) - \mathcal{L}_{\pi^\sharp(\beta)}(\alpha) - \frac{1}{2} d(\langle \alpha \wedge \beta, \pi \rangle) \end{aligned}$$

Remark 4.38. Usually, the last term on the last hand side is often written as $d(\pi(\alpha, \beta))$, where the bivector $\pi \in \mathfrak{X}^2(M)$ is here seen as a bilinear form on $\Omega^1(M)$. We chose to use the pairing given by Equation (4.23) for it seems more transparent; see Remark 4.33 for a comparison between the two notations.

Proof. We already know that $\pi^\sharp$ is vector bundle morphism and the bracket $[\,.\,,.\,]_{T^*M}$ is obviously skew-symmetric. Then we only need to show that the bracket satisfies the Jacobi identity and that it is compatible with the anchor map in the sense that they satisfy the Leibniz rule. Since every differential one-form is locally exact, and that the bracket is defined only locally, we will evaluate both the Jacobi identity and the Leibniz rule on exact differential one-forms. Then we can observe the following fact:

Lemma 4.39. *On exact differential one forms, the bracket* $[\,.\,,.\,]_{T^*M}$ *satisfies the following identity:*

$$[df, dg]_{T^*M} = d\{f, g\}$$

for every $f, g \in \mathcal{C}^\infty(M)$.

Exercise 4.40. Prove this lemma by using Proposition (4.31), the definition of $\pi^\sharp$ and the properties of the Lie derivatives given in Proposition 4.31.

Let us now show that $[\,.\,,.\,]_{T^*M}$ satisfies the Jacobi identity on exact differential one-forms; let $f, g, h \in \mathcal{C}^\infty(M)$, then by Lemma 4.39 we obtain:

$$\begin{aligned}
[df, [dg, dh]_{T^*M}]_{T^*M} &= \big[df, d\{g,h\}\big]_{T^*M} \\
&= d\{f, \{g,h\}\} \\
&= d\Big(\{\{f,g\},h\} + \{g,\{f,h\}\}\Big) \\
&= \big[d\{f,g\}, dh\big]_{T^*M} + \big[dg, d\{f,h\}\big]_{T^*M} \\
&= \big[[df,dg]_{T^*M}, dh\big]_{T^*M} + \big[dg, [df,dh]_{T^*M}\big]_{T^*M}
\end{aligned}$$

Notice that the Jacobi identity for $[\,.\,,.\,]_{T^*M}$ is a consequence of the Jacobi identity for the Poisson bracket. Now let us check the Leibniz rule (3.1):

$$\begin{aligned}
[df, gdh]_{T^*M} &= \mathcal{L}_{\pi^\sharp(df)}(gdh) - \mathcal{L}_{\pi^\sharp(gdh)}(df) - \frac{1}{2}d(\langle df \wedge gdh, \pi\rangle) \\
&= \mathcal{L}_{X_f}(gdh) - \mathcal{L}_{gX_h}(df) - \frac{1}{2}d(g\langle df \wedge dh, \pi\rangle) \\
&= \mathcal{L}_{X_f}(g)\, dh + g\,\mathcal{L}_{X_f}(dh) - d(gX_h(f)) - d(g\{f,h\}) \\
&= X_f(g)\, dh + g\,\{f,h\} - d(g\{h,f\}) - d(g\{f,h\}) \\
&= \pi^\sharp(df)(g)\, dh + g\,[df, dh]_{T^*M}
\end{aligned}$$

We used the definition of the Lie derivative as given by Definition 4.31, as well as the definition of Hamiltonian vector fields (see Equation (4.6)). □

The fact that T^*M is a Lie algebroid over M implies that the image of the anchor map $\pi^\sharp$ is a (possibly singular) smooth involutive distribution on M. If the distribution is regular – i.e. has constant rank – Frobenius theorem 3.64 implies that it is integrable to a regular foliation. If the distribution is singular – i.e. if its rank is not constant – then, because it is finitely generated and involutive, it turns out that it also integrates, to what is called a *singular foliation*. The latter notion generalizes the notion of regular foliation in the following way:

Definition 4.41. *A* singular foliation *is a partition $\bigsqcup_\alpha L_\alpha$ of M by disjoint connected weakly embedded submanifolds L_α called* leaves, *such that the induced distribution $D : x \longmapsto T_xL_{\alpha(x)}$ is smooth. Here $\alpha(x)$ denotes the index α such that L_α is the unique leaf passing through x.*

An alternative formulation, closer to that relying on foliated atlases for regular foliations, involves the notion of *distinguished atlas*. We say that M admits a distinguished atlas (with respect to a partition $\mathcal{L} = \bigsqcup_\alpha L_\alpha$ of M into immersed submanifolds, if for every $x \in M$ there exists a chart (U, φ) such that [Stefan, 1974]:

1. $\varphi(U)$ decomposes as a product of connected open sets $\varphi(U) = V \times W \subset \mathbb{R}^p \times \mathbb{R}^{n-p}$;
2. $\varphi(x) = (0,0)$;
3. for any $L \in \mathcal{L}$, $\varphi(L \cap U) = V \times l_L$, with $l_L = \{y \in W \,|\, \varphi^{-1}(0,y) \subset L\}$.

In particular, the last condition implies that the leaves intersecting U have higher than or equal dimension to that passing through x. It is equivalent to requiring that the map $x \longmapsto \dim(L_x)$ (where L_x is the leaf passing through x), going from the topological space M – equipped with

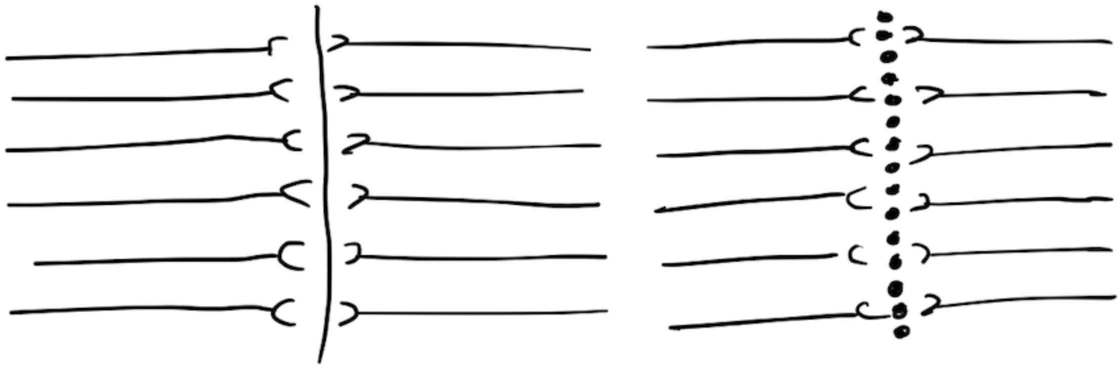

Figure 15: Two examples of partition of $\mathbb{R}^2$: the first one consists of horizontal lines for $x \neq 0$ and the vertical line for $x = 0$, while the second one has points on the vertical axis. The figure on the left is *not* a singular foliation because any tangent vector to the submanifold at $x = 0$ does not satisfy the smoothness condition: any smooth extension of ∂_y in any neighborhood of the origin necessarily contains a vertical part, which is then not tangent to the horizontal leaves outside the vertical axis. On the contrary, the figure on the right is a singular foliation.

the distinguished atlas – to the integer, is continuous. Since the target space has the discrete topology, the map is lower semi-continuous, hence the result. Moreover, it also implies that the map is locally constant on a dense subset of M, which means that the singular leaves are quite rare actually.

Example 4.42. The distribution defined in Example 3.60, although integrable, is *not* a singular foliation because the leaf passing through the origin $(0, 0)$ has higher dimension than its neighbors. Thus it does not admit a distinguished atlas as item 3. fails to be satisfied.

Frobenius' result about integrability can then be generalized to singular distributions thanks to Hermann's theorem:

Theorem 4.43. Hermann Theorem. *A locally finitely generated singular smooth distribution D on a smooth manifold is integrable (to a singular foliation) if and only if it is involutive.*

Remark 4.44. The first assumption, that the distribution is locally finitely generated means the following: for every point $x \in M$, there exist an open neighborhood U of x and a finite number of smooth sections $X_1, \ldots, X_m \in \Gamma(U, D)$ such that, for every open set V such that $\overline{V} \subset U$, the space of smooth sections $\Gamma(V, D)$ is generated as a $\mathcal{C}^\infty(V)$-module by the restrictions of $X_1, \ldots, X_m$ to V. The definition seems complicated but it is made so that the corresponding notion is local.

The idea behind integration of a singular smooth distribution is that the smooth manifold M is foliated by a set of weakly embedded submanifolds called *leaves* such that, given any point x, the tangent space to the leaf through x – denoted L_x – coincides with D_x:

$$T_x L_x = D_x$$

This identity being actually true for every point y of the leaf: $T_y L_x = D_y$. Since the rank of the distribution jumps, the dimension of the leaves will jump as well. A reservoir of examples of integrable distributions come from the following observation:

Proposition 4.45. *The (possibly singular) distribution generated by the anchor of a Lie algebroid is integrable.*

Proof. Let A be a Lie algebroid with anchor map ρ, and set $D_x = \mathrm{Im}(\rho(A_x))$. This is a smooth distribution because each element X_x of D_x admits a preimage $a_x \in A_x$, and it is then sufficient to take any smooth section of A passing through a_x, and to project it to $\mathfrak{X}(M)$ via ρ. The image is a vector field X such that $X_y \in D_y$ for every y is some neighborhood of x. The distribution is locally finitely generated because A is a vector bundle of finite rank, so it admits local frames that induce local generators of $\Gamma(D)$. Finally, it is involutive because $\rho : \Gamma(A) \longrightarrow \mathfrak{X}(M)$ is a homomorphism of Lie algebras. Then, by Hermann Theorem 4.43, the distribution D is integrable to a (possibly singular) foliation. □

Since, for a Poisson manifold (M, π), Proposition 4.37 implies that T^*M is a Lie algebroid, the vector bundle morphism $\pi^\sharp$ defines an integrable generalized distribution $D_\pi = \mathrm{Im}(\pi^\sharp) \subset TM$. The (possibly singular) foliation integrating this distribution is called the *characteristic foliation*. Since the rank of the distribution D at the point x equates that of the image of $\pi^\sharp$ at x and is thus even, the leaves of the foliation induced by a Poisson bivector field will always be even dimensional. We will now explain that they are, in fact, *symplectic manifolds*:

Definition 4.46. *A* symplectic manifold *is a smooth, even dimensional manifold, equipped with a non-degenerate closed two-form ω.*

Remark 4.47. Here, non-degeneracy means that the canonical vector bundle morphism $\omega^\flat = \iota_X(\omega) : TM \longrightarrow T^*M$ induced by ω by contraction with tangent vectors is an isomorphism of vector bundles.

Example 4.48. For every smooth manifold M, the cotangent bundle T^*M is naturally a symplectic manifold: let denote q^i the local coordinate functions on M and p_i the local coordinate functions on the fibers of T^*M, i.e. $p_i(dx^j) = \delta_i^j$. Then the differential 2-form $\omega \in \Omega^2(T^*M)$ defined as $\omega = \sum_{i=1}^n dp_i \wedge dq^i$ is a non-degenerate closed 2-form on T^*M. This result shows that the isomorphism $\omega^\flat : TM \longrightarrow T^*M$ then associates the tangent vector $\frac{\partial}{\partial q^i}$ to dp_i. In Hamiltonian mechanics, the coordinate function p_i is the *conjugate momentum* associated to q^i. Hence, symplectic manifolds represent the canonical setup to do classical mechanics (when it is well-defined). When working in a physical context we may call M the *configuration space* and T^*M the *phase space*. In particular the phase space $\mathbb{R}^{2n}$ presented in Example 4.3 actually corresponds to $T^*\mathbb{R}^n$.

Now let us draw the relationship between symplectic manifolds and Poisson manifolds. At this point, we need not assume that ω is a closed differential form, although we still assume that it is non-degenerate. Then, the vector bundle morphism $\omega^\flat : TM \longrightarrow T^*M$ can be inverted. Its inverse is thus a vector bundle morphism $B^\sharp : T^*M \longrightarrow TM$ satisfying:

$$\omega^\flat \circ B^\sharp(\alpha) = \alpha$$

for every differential one-form $\alpha \in \Omega^1(M)$. We denote X_f the vector field $B^\sharp(df)$ and call it *Hamiltonian vector field of* f (we will soon see that it is not contradictory with the earlier denomination). Then, by construction we have:

$$\omega(X_f, X_g) = \omega^\flat(X_f)(X_g) = df(X_g) = B^\sharp(dg)(f)$$

for any two smooth functions $f, g \in \mathcal{C}^\infty(M)$. Then, by Equations (4.24) and (4.25), there exists a unique non-degenerate bivector field B (hence the notation) such that:

$$\omega(X_f, X_g) = -\frac{1}{2}\langle df \wedge dg, B\rangle = -B(f, g) \tag{4.26}$$

This bivector field is actually not any bivector field:

Proposition 4.49. *ω is a symplectic form if and only if B is a Poisson bivector, that is to say:*

$$d\omega = 0 \quad \Longleftrightarrow \quad [B, B]_{SN} = 0$$

Proof. Let $f, g, h \in \mathcal{C}^\infty(M)$, and we set $X_f = B^\sharp(df)$, $X_g = B^\sharp(dg)$ and $X_h = B^\sharp(dh)$. Then, by Equation (2.31) the de Rham derivative of ω satisfies:

$$d\omega(X_f, X_g, X_h) = X_f(\omega(X_g, X_h)) - \omega([X_f, X_g], X_h) + \circlearrowleft \tag{4.27}$$

where $\circlearrowleft$ symbolizes circular permutation of the three functions. Using successively Equations (4.26), (4.25) and (4.24), one deduces that the first term on the right hand side of Equation (4.27) is $-B(f, B(g, h))$. On the other hand, the second term on the right-hand side of Equation (4.27) can be rewritten as:

$$\begin{aligned} -\omega([X_f, X_g], X_h) &= \omega(X_h, [X_f, X_g]) \\ &= \omega^\flat \circ B^\sharp(dh)([X_f, X_g]) \\ &= [B(f, -), B(g, -)](h) \\ &= B(f, B(g, h)) - B(g, B(f, h)) \end{aligned}$$

Thus, noticing that $-B(g, B(f, h)) = -B(B(h, f), g)$ and writing explicitly the circular permutation, Equation (4.27) can be rewritten:

$$d\omega(X_f, X_g, X_h) = -B(B(f, g), h) - B(B(g, h), f) - B(B(h, f), g)$$

On the right-hand side, one can recognize minus the Schouten-Nijenhuis bracket of B with itself, so that:

$$d\omega(X_f, X_g, X_h) = -[B, B]_{SN}(f, g, h)$$

This prove the claim. $\square$

Thus a symplectic manifold is a Poisson manifold where the Poisson bivector is non-degenerate. Conversely, using Proposition (4.49), one can show the converse statement: any non-degenerate bivector field B on a smooth manifold M gives rise to a non-degenerate differential 2-form ω which is closed – i.e. symplectic – if and only if B is a Poisson bivector. We can summarize these results in the following general statement:

Proposition 4.50. *Let M be an even dimensional smooth manifold. Then there is a one-to-one correspondence between non-degenerate Poisson structures and symplectic structures on M.*

Let M be an even dimensional smooth manifold, equipped with a symplectic form ω, to which correspond a non-degenerate Poisson bivector π. Let us now determinate the relationship between $\omega = \frac{1}{2}\omega_{kl} dx^k \wedge dx^l$ and $\pi = \frac{1}{2}\pi^{ij} \frac{\partial}{\partial x^i} \wedge \frac{\partial}{\partial x^j}$ in local coordinates $x^1, \ldots, x^n$. Evaluating ω on the hamiltonian vector fields $X_{x^i} = \pi^{ij} \frac{\partial}{\partial x^j}$, Equations (4.26) is equivalent to:

$$\omega_{kl} \pi^{ik} \pi^{jl} = -\pi^{ij} \tag{4.28}$$

where summation on repeated indices is implicit. We denote the coefficients of the inverse matrix of $\overline{\pi} = (\pi^{rs})_{rs}$ by π_{rs}, with indices at the bottom to allow contractions, so that $\pi^{rs}\pi_{st} = \delta^r_t$. Then, multiplying both sides of Equation (4.28) with π_{jm} and summing over j we obtain:

$$\pi^{ik} \omega_{km} = \delta^i_m \tag{4.29}$$

Thus, the components ω_{kl} turns out to precisely be π_{kl}, i.e. the coefficients of ω form a matrix that is the inverse matrix of $\overline{\pi}$. Thus, a symplectic form and its associated non-degenerate Poisson bivector somehow represent dual, equivalent pictures.

We have so far shown that when the characteristic foliation of a Poisson manifold consists of one leaf – i.e. when the Poisson bivector is non-degenerate – then the leaf is a symplectic manifold. We want to generalize this result to degenerate Poisson bivectors, by studying the local picture of symplectic manifolds. Recall that the standard Poisson bivector on $\mathbb{R}^{2n}$ is of the form (4.17). It is a non-degenerate Poisson bivector, and the corresponding symplectic form is given in Example 4.48:

$$\pi = \sum_{i=1}^{n} \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i} \qquad \Longleftrightarrow \qquad \omega = \sum_{i=1}^{n} dp_i \wedge dq^i$$

One can check that their respective components π^{ij} and ω_{kl} satisfy Equation (4.29). It turns out that this structure is quite central in symplectic geometry, because every symplectic manifold is locally symplectomorphic to $\mathbb{R}^{2n}$:

Theorem 4.51. Darboux theorem. *Let (M, ω) be a symplectic manifold and let $x \in M$. Then there exists local coordinates (q^i, p_i) centered at x, with respect to which the symplectic form ω is expressed as:*

$$\omega = \sum_{i=1}^{n} dp_i \wedge dq^i$$

In other words, Darboux theorem states that locally, every symplectic manifold locally looks the same. It implies that there are no local invariants in symplectic geometry, contrary to Riemannian geometry for example. The above result occurs when M is symplectic or, equivalently, when it is a non-degenerate Poisson manifold, so that the characteristic foliation consists of one, unique leaf: the total manifold. By Proposition 4.50 we can reformulate Darboux theorem in terms of non-degenerate Poisson structures:

Theorem 4.52. Darboux theorem (Poisson version). *Let (M, π) be a Poisson manifold manifold such that π is non-degenerate, and let $x \in M$. Then there exists local coordinates (q^i, p_i) centered at x, with respect to which the symplectic form ω is expressed as:*

$$\pi = \sum_{i=1}^{n} \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i}$$

Remark 4.53. This theorem sheds light on why the standard Poisson structure on $\mathbb{R}^{2n}$ is 'rigid' in the sense that $H^2_\pi(\mathbb{R}^{2n}) = 0$ and more generally every non-degenerate Poisson bivectors (by Remark 4.36). This is because any small (formal) deformation of such Poisson bivector is still non-degenerate, so they locally still look like the standard structure on $\mathbb{R}^{2n}$. Thus, we cannot 'deform' them.

Now what happens when the Poisson bivector is degenerate, i.e. when its rank does not equate the dimension of the manifold at every point? In that case, the generalized distribution D_π associated to the Poisson bivector π is integrable and its leaves are even dimensional. The following important result generalizing Darboux theorem 4.51 to Poisson manifolds sheds light on what happens locally:

Theorem 4.54. Weinstein splitting theorem. *Let (M, π) be a Poisson manifold of dimension n and let $x \in M$ be an arbitrary point. Denote the rank of the Poisson bivector π at x by*

$2r$, *and let* $s = n - 2r$. *Then, there exists local coordinates* $q^1, ..., q^r, p_1, ..., p_r, z_1, ..., z_s$ *centered at* x, *such that the Poisson bivector reads:*

$$\pi = \sum_{i=1}^{r} \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i} + \sum_{1 \leq k,l \leq s} \phi_{kl} \frac{\partial}{\partial z_k} \wedge \frac{\partial}{\partial z_l} \tag{4.30}$$

where the functions $\phi_{kl} = -\phi_{lk}$ *are smooth functions, which depend on* $z = (z_1, \ldots, z_s)$ *only, and which vanish when* $z = 0$.

Weinstein's theorem is not a result about local coordinates, but a result about the possibility of choosing a special subset of local coordinates satisfying some nice property regarding the Poisson bivector. It is a result of foliation theory that leaves are weakly embedded submanifolds. Then by Proposition 3.52, it always possible to choose, in a vicinity of the point x, coordinates *adapted to the leaf* L_x: the first $2r$ coordinates are local coordinates on L_x, while the last s coordinates represent transversal ones. In particular, the zero locus of the last s coordinates represent the leaf through x in that vicinity, see Figure 16. Weinstein's theorem states that, additionally, a choice of such local coordinates can be made so that, in a vicinity of the point x, the rank of the Poisson bivector field has constant rank $2r$ on the leaf through x, this rank coinciding by definition with the dimension of L_x. This implies in turn that the restriction of the Poisson bivector to L_x is a non-degenerate Poisson bivector $\pi|_{L_x}$ and its form is the standard one, of Theorem 4.52. By Proposition 4.50, this makes L_x a *symplectic manifold.* This fact being true for every point x and thus every leaf of the characteristic foliation, we have finally obtained a full characterization of the latter:

Proposition 4.55. *The leaves of the characteristic foliation of a Poisson manifold are symplectic manifolds.*

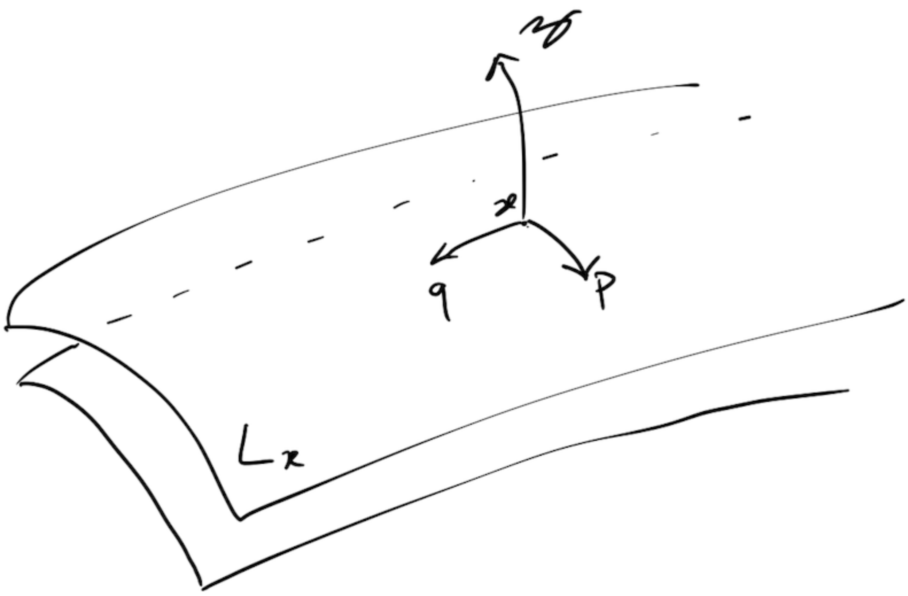

Figure 16: Representation of the local coordinates centered at the point x appearing in Weinstein splitting theorem 4.54. The z-coordinates are transversal to the leaf L_x passing through x.

Example 4.56. The linear Poisson structure defined on the dual of a Lie algebra $\mathfrak{g}$ induces of foliation of $\mathfrak{g}^*$ by symplectic leaves. These actually correspond to the coadjoint orbits of $\mathfrak{g}$ on

$\mathfrak{g}^*$. Polynomial functions on $\mathfrak{g}^*$ (i.e. elements of the universal enveloping algebra of $\mathfrak{g}$) that are constant along these orbits are called Casimir operators. This convention explains why, in Poisson geometry, functions whose hamiltonian vector field is zero are called Casimirs.

Example 4.57. As a particular case of the last example, the Poisson bivector field associated to the Poisson structure of Example 4.5 is the following:

$$\pi = z\partial_x \wedge \partial_y + x\partial_y \wedge \partial_z + y\partial_z \wedge \partial_x$$

where we transformed capital letters into small ones. The symplectic leaves are the concentric spheres or radius r (2-dimensional) and the origin (0-dimensional).

Example 4.58. The Poisson manifold $(\mathbb{R}^3, \pi)$ defined in Example 4.24 induces a distribution D_π generated by the following three hamiltonian vector fields:

$$X_x = x\frac{\partial}{\partial z}, \quad X_y = y\frac{\partial}{\partial z} \quad \text{and} \quad X_z = -x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}$$

This distribution is integrable into a singular foliation: the singular leaves are points of coordinates $(0,0,z)$ because the three vectors fields vanish, while the regular leaves are 2-dimensional vertical planes escaping radially from the vertical axis because then X_z is radial and either X_x, X_y or both are vertical (see Figure 17).

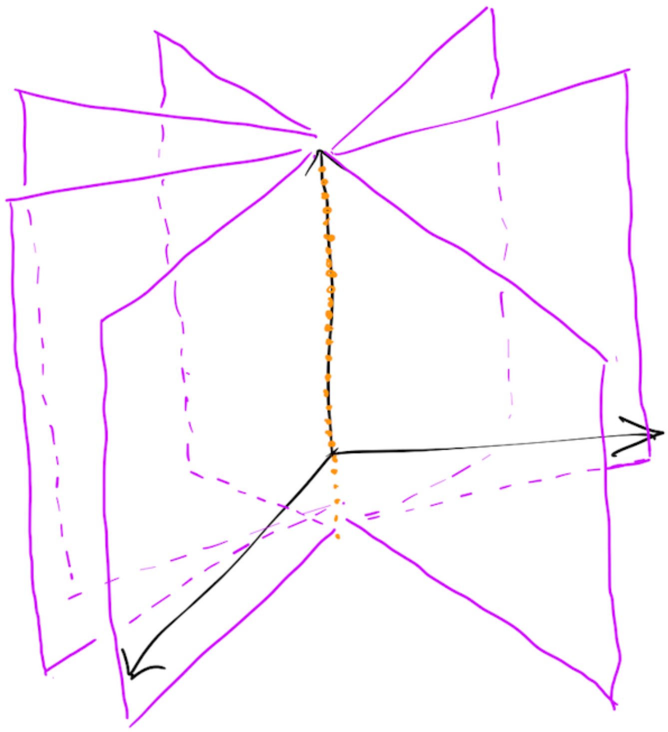

Figure 17: The singular leaves of the characteristic foliation of the Poisson bivector $\pi = (x\,\partial_x + y\,\partial_y) \wedge \partial_z$ are points on the z-axis (in orange, 0-dimensional submanifolds), while the regular leaves are vertical, radial planes (in purple, 2-dimensional submanifolds).

Let us work in the half-space with $x > 0$, and use polar coordinates $(x, y, z) \longmapsto (r, \theta, z)$, where $r = \sqrt{x^2 + y^2} > 0$ and $\theta = \arctan(\frac{y}{x}) \in]-\frac{\pi}{2}, \frac{\pi}{2}[$. Then, the constant vectors $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ become respectively:

$$\frac{\partial}{\partial x} = \cos(\theta)\frac{\partial}{\partial r} - \frac{\sin(\theta)}{r}\frac{\partial}{\partial \theta} \qquad \text{and} \qquad \frac{\partial}{\partial y} = \sin(\theta)\frac{\partial}{\partial r} + \frac{\cos(\theta)}{r}\frac{\partial}{\partial \theta}$$

Then, since $x = r\cos(\theta)$ and $y = r\sin(\theta)$, the Poisson bivector of Example 4.24 becomes:

$$\pi = r\frac{\partial}{\partial r} \wedge \frac{\partial}{\partial z} \tag{4.31}$$

Now, if one performs the following new change of radial coordinate: $r \longmapsto \rho = \ln(r)$, one then obtains $\frac{\partial}{\partial r} \longmapsto \frac{1}{r}\frac{\partial}{\partial \rho}$, so that the Poisson bivector (4.31) becomes:

$$\pi = \frac{\partial}{\partial \rho} \wedge \frac{\partial}{\partial z} \tag{4.32}$$

We have then found the expression of the Poisson bivector π in a set of coordinates (ρ, θ, z) adapted to the situation, although they are *not* those of Weinstein splitting theorem for they are not centered at any point. Expression (4.31) is valid even for $x \leq 0$ (unless $x = y = 0$), because there is no dependence on θ.

Now, to make the connection explicit with Equation (4.30), let (x_0, y_0, z_0) be a point such that $x_0 > 0$, let $\rho_0 = \ln\big(\sqrt{x_0^2 + y_0^2}\big)$ and $\theta_0 = \arctan(\frac{y_0}{x_0})$. Denoting $q = \rho - \rho_0$, $p = z - z_0$ and $\varphi = \theta - \theta_0$, we have a set of (local) coordinates centered at the point (x_0, y_0, z_0), such that (q, p) span the leaf through (x_0, y_0, z_0) – a vertical radial plane – and φ encodes the transversal direction and vanishes on the leaf. Moreover, since we have $\frac{\partial}{\partial \rho} = \frac{\partial}{\partial q}$, $\frac{\partial}{\partial z} = \frac{\partial}{\partial p}$ and $\frac{\partial}{\partial \theta} = \frac{\partial}{\partial \varphi}$, one can write Equation (4.32) as:

$$\pi = \frac{\partial}{\partial q} \wedge \frac{\partial}{\partial p} + \sin(\varphi)\frac{\partial}{\partial \varphi} \wedge \frac{\partial}{\partial \varphi}$$

The last term automatically vanishes because of the antisymmetry of the wedge product. However, we have nonetheless provided a smooth function $\phi : \varphi \longmapsto \sin(\varphi)$ which only depends on φ and vanishes on the level set $\varphi = \theta - \theta_0 = 0$, and managed to write the Poisson bivector as in Equation (4.30). Thus, the set of (local) coordinates (q, p, φ) are those whose existence is claimed by Weinstein splitting theorem. For other regular point with $x \leq 0$, one uses the same argument. For singular points (on the z-axis), the Poisson bivector is zero.

Notice that, since a symplectic manifold is always even dimensional, when the smooth manifold M is odd-dimensional, there will necessarily be zero-dimensional leaves (hence trivial symplectic manifolds). In the case where the rank of π is locally constant at x – i.e. on some open neighborhood U of x – then 1. the foliation induced by the Poisson bivector on U is regular, and 2. there exist s Casimirs such that the symplectic leaves correspond to the level sets of the Casimirs (and then can be taken to be the coordinates z^k). That is why one often call the local coordinates $(q^1, \ldots, q^r, p_1, \ldots, p_r, z^1, \ldots, z^s)$ *Casimir-Darboux coordinates*. Knowing that a singular foliation forms a partition of the ambient manifold M, a corollary of Proposition 4.55 is the following:

Corollary 4.59. *Every point of a Poisson manifold is contained in a unique symplectic leaf.*

We conclude this subsection by the following very beautiful and nice result: one can show that the Poisson bracket can be entirely reconstructed from the data of the symplectic forms on the leaves of the characteristic foliation. One defines a *smooth family of symplectic leaves* on a manifold M as the data of a singular foliation such that each leaf L is a symplectic manifold (L, ω_L), and such that for every $f \in \mathcal{C}^\infty(M)$, the family of tangent vectors $(X_{f,x})_{x \in M}$ defined at each point by $\omega_{L,x}(X_{f,x}, -) = df_x$ (where L is the leaf through x) is a smooth vector field on M. A Poisson manifold obviously induces a smooth family of symplectic leaves, and Vaisman has shown the converse statement [Vaisman, 1994]:

Theorem 4.60. *Let M be a smooth manifold equipped with a smooth family $\mathcal{L}$ of symplectic leaves. Then there exists a unique Poisson structure on M such that the characteristic foliation coincide with the foliation $\mathcal{L}$ (as well as the symplectic structures on the leaves).*

Proof. See Theorem 2.14 in [Vaisman, 1994]. One implication has been proven by the discussion surrounding Weinstein's splitting theorem, the other implication relies on defining $\{f,g\} = X_f(g)$ (since the hamiltonian vector fields are smooth). □

4.3 Submanifolds and reduction in Poisson geometry

The study of submanifolds in Poisson geometry is slightly more intricate than in differential geometry, because one needs to evaluate if the Poisson bracket originally defined on the ambient manifold M descends to the submanifold $S \subset M$. In this section, unless otherwise stated, the word 'submanifold' designates any kind of submanifolds: immersed, weakly embedded or embedded.

Definition 4.61. *A* Poisson submanifold *of a Poisson manifold M is a submanifold $S \overset{\iota}{\hookrightarrow} M$ admitting a Poisson structure such that the inclusion map ι is a Poisson map.*

The immersed or (weakly) embedded submanifold S can always be seen as the image of a injective immersion/weak embedding/smooth embedding $F : N \longrightarrow M$, such that $S = F(N)$. Then one may equivalently consider that the submanifold S is a Poisson submanifold if N is a Poisson manifold and F is a Poisson map. Denoting $\{\,.\,,.\,\}$ (resp. $\{\,.\,,.\,\}_N$) the Poisson bracket on M (resp. on N), this definition implies that the Poisson submanifold $S = F(N)$ is characterized by the fact that:

$$\big\{F^*(f), F^*(g)\big\}_N = F^*\big(\{f,g\}\big) \tag{4.33}$$

for every $f,g \in \mathcal{C}^\infty(M)$. We shall see that in terms of bivector fields, Equation (4.33) can be restated as the fact that the Poisson bivector π defined on M is tangent to S at every point of S: $\pi_x \in \bigwedge^2 T_xS \subset \bigwedge^2 T_xM$ for every $x \in S$. There are additional equivalent characterizations of Poisson submanifolds, both geometric and algebraic:

Proposition 4.62. *Let (M,π) be a Poisson manifold and let S be a submanifold of M. The following are equivalent:*

1. *S is a Poisson submanifold;*
2. *$\pi|_S$ takes values in $\bigwedge^2 TS$;*
3. *$\pi^\sharp(TS^\circ) = 0$;*
4. *$\pi^\sharp(T^*M|_S) \subset TS$;*
5. *all Hamiltonian vector fields are tangent to S.*

Remark 4.63. The notation TS° stands for the *annihilator* of TS. It is the vector bundle over S consisting of all the covectors vanishing on TS. More precisely, for every $x \in S$, one sets:

$$T_xS^\circ = \big\{\xi_x \in T_x^*M \;\big|\; \xi_x(T_xS) = 0\big\}$$

If M is n-dimensional and if S is a r-dimensional submanifold, TS° is a rank $n-r$ vector bundle over S.

Proof. 1. $\iff$ 2. Suppose that the submanifold S is obtained as the image of an injective immersion $F : N \hookrightarrow M$ (weak and smooth embeddings are injective immersions). Then we define $\bigwedge^2 TS$ as the pushforward of the vector bundle $\bigwedge^2 TN$ on M via $F_* \wedge F_*$, and we have $\bigwedge^2 TS \subset \bigwedge^2 TM|_S$.

First, assume that S is a Poisson submanifold of M, i.e. that N admits a Poisson structure $\{\,.\,,.\,\}_N$, and that F is a Poisson map. In full generality, Equation (4.33) can be rewritten in terms of Poisson bivectors as:

$$\big\langle F^*(df \wedge dg), \pi_N \big\rangle_N = F^*\big(\langle df \wedge dg, \pi \rangle_M\big) \tag{4.34}$$

where π (resp. π_N) is the Poisson bivector corresponding to $\{\,.\,,.\,\}$ (resp. $\{\,.\,,.\,\}_N$). On the left hand-side the pairing is taken with respect to TN and T^*N, while on the right-hand side it is taken with respect to TM and T^*M. Equation (4.34) is to be understood as an equality on N or, equivalently, on $S = F(N)$. Restricting Equation (4.34) to S has the following two consequences: one can rewrite the left-hand side as $\langle df \wedge dg, F_* \wedge F_*(\pi_N)\rangle_M$, while dropping F^* on the right-hand side:

$$\big\langle df \wedge dg, F_* \wedge F_*(\pi_N) \big\rangle_M \Big|_S = \langle df \wedge dg, \pi \rangle_M \Big|_S \tag{4.35}$$

where both sides here have to be understood as the restriction to S of the underlying smooth functions. Since the functions f and g are arbitrary, one obtains that, on S, $\pi\big|_S = F_* \wedge F_*(\pi_N)$, which proves item 2 since by definition $\bigwedge^2 TS = F_* \wedge F_*(\bigwedge^2 TN)$.

Conversely, still assumming that $S = F(N)$ is a submanifold of M, then item 2. implies that there exists a bivector field π_N on N such that $\pi\big|_S = F_* \wedge F_*(\pi_N)$. Since F is an injective immersion, it is unique. Moreover, for every open set $U \subset M$ the bivector field π_N satisfies Condition (4.16) on $F^*(\mathcal{C}^\infty(U))$. Let us show that this implies that Condition (4.16) is satisfied in the neighborhood of every point of N. Let $x \in S$ and let $x^1, \ldots, x^n$ be local coordinates adapted to the submanifold S in a neighborhood of x, in the sense of Proposition 3.54. That is to say, there exists a connected coordinate chart $V \subset N$ centered at $y = F^{-1}(x)$ in N and a coordinate chart (U, φ) centered at x such that:

$$\varphi(U \cap F(V)) = \varphi(U) \cap \{\mathbb{R}^k \times 0\}$$

In other words, if the dimension of N is k, we can assume that the first k coordinates $x^1, \ldots, x^k$ of the chart φ are those parametrizing both V and $U \cap F(V) \subset S$, so that the function F becomes $(x^1, \ldots, x^k) \longmapsto (x^1, \ldots, x^k, 0, \ldots, 0)$. Then any function $g \in \mathcal{C}^\infty(V)$ can be written as the pull-back of a smooth function on $\mathcal{C}^\infty(U)$: let $\mu : (x^1, \ldots, x^n) \longmapsto (x^1, \ldots, x^k)$ be the projection along the last $n-k$ coordinates, and let $f = g \circ \mu$. Then $g = f \circ F = F^*(f)$. Since $\mathcal{C}^\infty(V) = F^*(\mathcal{C}^\infty(U))$, then π_N satisfies Condition (4.16) on V. This result being true in the neighborhood of each point of N, we deduce that π_N is a Poisson bivector. Then, since Equation (4.35) holds for arbitrary functions f and g, implying in turn that Equation (4.34) holds, the map $F : N \hookrightarrow M$ is a Poisson map, turning S into a Poisson submanifold of M.

2. $\iff$ 5. Again suppose that S is obtained (at least) as an injective immersion. Let $x \in S$ and let $x^1, \ldots, x^n$ be local coordinates adapted to the submanifold S in a neighborhood of x, as in the proof of the last item. In particular, letting V be a sufficiently small neighborhood of $y = F^{-1}(x)$ as in Proposition 3.54, the first k coordinates $x^1, \ldots, x^k$ parametrize $V \simeq F(V)$, while the last $n-k$ coordinates are transverse to $F(V)$. Then the Poisson bivector π can be decomposed as:

$$\pi = \frac{1}{2}\sum_{i,j=1}^{k} \pi^{ij} \frac{\partial}{\partial x^i} \wedge \frac{\partial}{\partial x^j} + \sum_{i=1}^{k} \sum_{j=k+1}^{n} \pi^{ij} \frac{\partial}{\partial x^i} \wedge \frac{\partial}{\partial x^j} + \frac{1}{2}\sum_{i,j=k+1}^{n} \pi^{ij} \frac{\partial}{\partial x^i} \wedge \frac{\partial}{\partial x^j} \tag{4.36}$$

For any smooth function $f \in \mathcal{C}^\infty(M)$, the association hamiltonian vector field is then:

$$X_f = \sum_{i,j=1}^{k} \pi^{ij} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^j} + \sum_{i=1}^{k} \sum_{j=k+1}^{n} \pi^{ij} \left(\frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^j} - \frac{\partial f}{\partial x^j} \frac{\partial}{\partial x^i} \right) + \sum_{i,j=k+1}^{n} \pi^{ij} \frac{\partial f}{\partial x^i} \frac{\partial}{\partial x^j} \tag{4.37}$$

Only the first term on the right-hand side of Equation (4.36) is a section of $\bigwedge^2 TS$. Thus, if item 2. holds, then the second and third sums are zero on S, proving, using Equation (4.37), that every Hamiltonian vector field is necessarily tangent to S. Conversely, if item 5. holds, the last term of Equation (4.37) necessarily vanishes on S as it is not tangent to S, while in the parenthesis there is a term tangent to S and the other is not. Applying successively Equation (4.37) to the coordinate functions $x^1, \ldots, x^k$ give:

$$X_{x^l} = \sum_{j=1}^{k} \pi^{lj} \frac{\partial}{\partial x^j} + \sum_{j=k+1}^{n} \pi^{lj} \frac{\partial}{\partial x^j}$$

Since these hamiltonian vector fields have to be tangent to S for every $1 \leq l \leq k$, we deduce that on S, we have $\pi^{lj} = 0$ for very $1 \leq l \leq k$ and $k+1 \leq j \leq n$. This proves that the Poisson bivector field π reduces to the first term of Equation (4.36) on the submanifold S, i.e. item 2.

3. $\iff$ 4. Let $\xi \in \Omega^1(M)$ such that $\xi_x \in T_x S^\circ$ for every $x \in S$ and let η be a differential 1-form on M. Then by Equation (4.25) we have, for every $x \in S$:

$$\xi_x(\pi_x^\sharp(\eta_x)) = \frac{1}{2}\langle \eta_x \wedge \xi_x, \pi_x \rangle = -\eta_x(\pi_x^\sharp(\xi_x))$$

This identity being true for every differential one-form η on M and every ξ taking values in TS° on S, this implies that the right-hand side equals 0 if and only if item 3. holds and the left-hand side equals zero if and only if item 4. holds. Then item 3. is equivalent to item 4.

4. $\iff$ 5. The direct implication is straighforward, while for the reverse implication, assume that every Hamiltonian vector field is tangent to S. Since through every point of the cotangent bundle of M passes an exact form, then for every point $x \in S$ and $\xi_x \in T_x^* M$, there exists a smooth function f defined on M such that $df_x = \xi_x$. Then $\pi_x^\sharp(\xi_x) = \pi_x^\sharp(df_x) = X_{f,x}$, which shows that $\pi^\sharp(\xi_x)$ actually takes values in $T_x S$. Since this is true for every point of the cotangent bundle over the submanifold S, one deduces that item 4. holds. □

Example 4.64. Obvious examples of Poisson submanifolds are the symplectic leaves of the characteristic foliation induced by $\pi^\sharp$. They have the property that their Poisson bracket is non-degenerate. More generally, Poisson submanifolds are unions of (open subsets of) symplectic leaves (see Proposition 4.95).

Example 4.65. As seen in Example 4.58, using polar coordinates allow to write the Poisson bivector $\pi = (x\partial_x + y\partial_y) \wedge \partial_z$ on $\mathbb{R}^3$ as $\pi = r\partial_r \wedge \partial_z$. One then straightforwardly sees that the 2-dimensional symplectic leaves (the vertical radial planes) are Poisson submanifolds of $(\mathbb{R}^3, \pi)$.

Exercise 4.66. This exercise is a continuation of Example 4.57, where the symplectic leaves are the concentric spheres in $\mathbb{R}^3$ and aims at showing that they are indeed Poisson submanifolds. The hemi-sphere of radius $r > 0$ located in the positive y half-space admits adapted spherical coordinates (r, θ, φ) on it, where $r > 0$ is the distance from the origin, $\theta \in \,]0, \pi[$ is the angle between the positive z axis and the vector and $\varphi \in \,]0, \pi[$ is the angle between the x axis and the projection of the vector on the Oxy plane. Show that in these spherical coordinates the Poisson bivector of Example 4.57 reads:

$$\pi = \frac{1}{r\sin(\theta)} \partial_\theta \wedge \partial_\varphi$$

and deduce from it that the hemi-sphere of radius $r > 0$ equipped with this Poisson bivector is a Poisson submanifold of $(\mathbb{R}^3, \pi)$ (we know that it should be, as it is (a submanifold of) the level set of the Casimir element $C = x^2 + y^2 + z^2 - r$ of π).

As seen earlier, one can always characterize geometric objects by algebraic ones and vice-versa. This is the goal of the following proposition:

Proposition 4.67. *Let S be a Poisson submanifold of the Poisson manifold M. Then, the multiplicative ideal:*

$$\mathcal{I}_S = \Big\{ f \in \mathcal{C}^\infty(M) \text{ such that } f|_S \equiv 0 \Big\}$$

is a Lie ideal of the Lie algebra $(\mathcal{C}^\infty(M), \{\,.\,,.\,\})$.

Proof. Since every hamiltonian vector field is tangent to S on S, then for any smooth functions $f \in \mathcal{C}^\infty(M)$ and $g \in \mathcal{I}_S$, one has by definition of TS $X_f(g) = 0$ on S, which can be equivalently be written as $\{f, g\}(x) = 0$ for every $x \in S$, that is to say: $\{f, g\} \in \mathcal{I}_S$. This proves that $\mathcal{I}_S$ is a Lie algebra ideal with respect to the Poisson bracket. $\square$

The condition stated in Proposition 4.67 is not sufficient to characterize Poisson manifolds, unless they are embedded. Indeed, for immersed or weakly embedded submanifolds, the tangent space at a point does not necessarily coincide with the set of tangent vectors on M at that point that vanish on I_S (see counter-Example 3.56):

$$T_xS \subset \Big\{ X_x \in T_xM \,\Big|\, X_x(f) = 0 \text{ whenever } f \in \mathcal{I}_S \Big\} \tag{4.38}$$

The fact that $\mathcal{I}_S$ is a Poisson ideal in $\mathcal{C}^\infty(M)$ means that for every $f \in \mathcal{I}_S$ and $g \in \mathcal{C}^\infty(M)$, the smooth function $X_g(f) = \{g, f\}$ vanishes on S, which implies that Hamiltonian vector fields belong to the set on the right hand-side of Equation (4.38). When S is an embedded submanifold, we can conclude that these hamiltonian vector fields are tangent to S, and hence that S is a Poisson submanifold.

Example 4.68. If C is a Casimir function on a Poisson manifold M, then the levels sets of regular values of C are closed embedded submanifolds of M (see Theorem 3.44). Let S be such the level set of such a regular value $\lambda \in \mathbb{R}$, then it is a closed embedded submanifold of M. The ideal of functions vanishing on S is then spanned by the function $x \mapsto C(x) - \lambda$ (see Theorem 1.1 in [Henneaux and Teitelboim, 1994] or pages 95-96 of [Sudarshan and Mukunda, 2015]), and we write $\mathcal{I}_S = \langle C - \lambda \rangle$. This forms a Lie ideal of $(\mathcal{C}^\infty(M), \{\,.\,,.\,\})$ (independently of λ being a regular value or not), as the following argument shows: any smooth function f satisfies $\{f, C-\lambda\} = \{f, C\} = -X_C(f)$, which vanishes on M by definition of C being a Casimir function, hence in particular it vanishes on S, so $\{f, C - \lambda\} \in \mathcal{I}_S$. This also shows that any hamiltonian vector field X_f vanishes on $\mathcal{I}_S$ and, since S is an embedded submanifold because λ is a regular value, inclusion (4.38) becomes an equality. These facts imply that every hamiltonian vector fields are tangent to S, proving that it is a Poisson submanifold of M by item 5. of Proposition 4.62. For example the level sets of the Casimir element of Exercice 4.9 correspond to concentric spheres in $\mathbb{R}^3$, and coincide with the symplectic leaves of $\mathfrak{so}_3(\mathbb{R})$, i.e. its coadjoint orbits.

Poisson submanifolds are actually very rare. As in symplectic geometry, there are weaker notions of submanifolds in Poisson geometry, that possess specific features leading to important applications in mathematical physics: *Poisson-Dirac* submanifolds and *coisotropic* submanifolds. The Poisson bracket of the ambient manifold descends on the former via the so-called *Poisson-Dirac reduction*, while one has to further take a quotient of the latter to define a Poisson bracket: this is the topic of *coisotropic reduction*.

The notion of Poisson-Dirac submanifold relies on the following notion: let S be a – immersed or (weakly) embedded – submanifold, $f \in \mathcal{C}^\infty(S)$ and $x \in S$, then a *local extension of f at x* is the data $(V, U, \widetilde{f})$ of an open neighborhood V of x in S (which then can be embedded into M via Proposition 3.54), an open neighborhood $U \subset M$ of x in M such that $V \subset S \cap U$, and a smooth function $\widetilde{f} \in \mathcal{C}^\infty(U)$, such that $\widetilde{f}$ and f coincide on V: $\widetilde{f}|_V = f|_V$.

Lemma 4.69. *Let S be a – immersed or (weakly) embedded – submanifold of M:*

1. *every smooth function on S can be locally extended;*
2. *if S is closed and embedded, then every smooth functions on S admit global extensions.*

Proof. The proof of the second statement can be found in Lemma 2.27 in [Lee, 2003] and Proposition 1.36 in [Warner, 1983], and heavily rely on the closedness of the submanifold. This statement can be alternatively be described as the following isomorphism: $\mathcal{C}^\infty(S) \simeq \mathcal{C}^\infty(M)\big/\mathcal{I}_S$. Let us now prove the first statement: let $x \in S$ and let V be an open neighborhood of x; Proposition 3.54 tells us that V forms a slice of U, i.e. a closed embedded submanifold of U. In that case, applying the already proven second statement, one can extend $f \in \mathcal{C}^\infty(V)$ to a smooth function $\widetilde{f}$ on U. □

Next, we say that a local extension $(V, U, \widetilde{f})$ of a function f at x is *horizontal* if the Hamiltonian vector field $X_{\widetilde{f}}$ is tangent to S, i.e. if $X_{\widetilde{f},y} \in T_yS$ for every $y \in V$. Although every function on S admits local extensions, it may not be true that it admits horizontal local extensions. Poisson-Dirac submanifolds are precisely those submanifolds in Poisson geometry which have such a property:

Definition 4.70. *A* Poisson-Dirac submanifold *of a Poisson manifold M is a submanifold $S \xhookrightarrow{\iota} M$ which is such that for every point $x \in S$, every smooth function f on S admits an horizontal local extension $(V, U, \widetilde{f})$ at x.*

Remark 4.71. The definition comes from subsection 5.3.2 of [Laurent-Gengoux et al., 2013]. It implies in particular that the pull-back to S of the so-called *Dirac structure* on M corresponding to the Poisson bivector π is a Dirac structure on S. See Section 6 of these lectures notes.

Obviously, every Poisson submanifold is a Poisson-Dirac submanifold since *every* hamiltonian vector field is tangent to a Poisson submanifold. Definition 4.70 however shows that this condition has been profoundly weakened for Poisson-Dirac submanifolds. The main interest of the latter – and the definition has been explicitly chosen to this purpose – is that the Poisson bracket on M descends to S in a unique way to turn S into a Poisson manifold in its own way:

Proposition 4.72. Poisson-Dirac reduction. *Let S be a Poisson-Dirac submanifold of the Poisson manifold $(M, \{\,.\,,.\,\})$. Then there exists a unique Poisson bracket $\{\,.\,,.\,\}_S$ on S such that for every $x \in S$ and every two smooth functions $f, g \in \mathcal{C}^\infty(S)$, one has:*

$$\{f, g\}_S = \{\widetilde{f}, \widetilde{g}\}\big|_V \tag{4.39}$$

for any horizontal local extensions $(V, U, \widetilde{f})$ and $(V, U, \widetilde{g})$ of f and g at x.

Proof. The proof can be found in Proposition 5.24 of [Laurent-Gengoux et al., 2013]. □

Example 4.73. This example is taken from [Fernandes, 2005]: let M be a Poisson manifold and let G be a Lie group properly acting on M via Poisson diffeomorphisms. Then the fixed points set M^G is a Poisson-Dirac submanifold of M.

A Poisson-Dirac submanifold possesses at most one Poisson structure satisfying Equation (4.39), and this Poisson structure is completely determined by the Poisson structure of M. Notice that the Poisson bracket on S is defined from picking up two local extensions whose hamiltonian vector field is tangent to S. It does not work with every local extension, although any other choice of extensions (such that their hamiltonian vector field is tangent to S) gives the same result in Equation (4.39). The fact that not every extension would satisfy Equation (4.39) can be explained from the following observation: contrary to Poisson submanifolds, where the Poisson bivector, restricted to S, takes values in $\bigwedge^2 TS$ (see item 2. of Proposition 4.62), on a Poisson-Dirac submanifold S the Poisson matrix $(\pi^{ij})_{ij}$, with respect to a choice of coordinates adapted to S (see e.g. Proposition 3.54 or Proposition 1.35 in [Warner, 1983]) can be decomposed into blocks and take the form:

$$(\pi^{ij})_{ij} = \begin{pmatrix} A & B \\ -B^t & D \end{pmatrix} \tag{4.40}$$

Then one can show that on S, the anti-diagonal components B and $-B^t$ identically vanish so that the Poisson bivector reduces to two independent terms: $\pi|_S = \pi_1 + \pi_2$, where π_1 corresponds to the A component in the matrix and takes values in $\bigwedge^2 TS$, while π_2 corresponds to the D component. Thus, the Poisson bracket on S by Proposition 4.72 corresponds to π_1, although on S the Poisson bivector $\pi|_S$ contains another component π_2, which only vanishes when evaluated on local extensions whose hamiltonian vector fields are tangent to S. A nice presentation of this issue (in a slightly less general case, however allowing to split TM into a direct sum) is made in the discussion surrounding Lemma 2.15 in these lecture notes. Although Poisson-Dirac submanifolds are very useful for the possibility that Poisson-Dirac reduction offer, Definition 4.70 is a bit obscure so that it is not very clear what does it look like in geometric terms. This is the role of the next proposition:

Proposition 4.74. *Let S be a submanifold of a Poisson manifold M. Then, the following are equivalent:*

1. *S is a Poisson-Dirac submanifold;*
2. *for every $\alpha \in \Omega^1(S)$ there exists open sets $V \subset S$ and $U \subset M$ such that $V \subset S \cap U$, and a differential one-form $\widetilde{\alpha} \in \Omega^1(U)$ such that $\alpha|_V = \iota^*(\widetilde{\alpha}|_U)$ and $\pi^\sharp(\widetilde{\alpha})$ is tangent to S;*
3. *for each $x \in S$, there exist local coordinates on M centered at x such that, if the matrix of π with respect to these coordinates can be written in a block form, then there exists a neighborhood V of x in S such that the matrix is diagonal by block on V;*
4. *$TS \cap \pi^\sharp(TS^\circ) = 0$ and the bivector field π_S induced from π on S via Equation (4.39) is smooth.*

Proof. For item 2. see Lemma 3.29 of these lectures notes, for item 3. see Proposition 5.25 in [Laurent-Gengoux et al., 2013], while for item 4. see subsection 9.2 in [Crainic and Fernandes, 2004]. ☐

This proposition is similar to Proposition 4.62, when S is a Poisson-Dirac submanifold. Since such a submanifold is precisely defined from the behavior of Hamiltonian vector fields of local extensions, the counterpart of item 5. of Proposition 4.62 is item 1. of Proposition 4.74. Item 3. of Proposition 4.62 corresponds to item 2. of Proposition 4.74, via a slight reformulation because not every Poisson-Dirac submanifold S admits a normal bundle N such that $T_xM = T_xS \oplus N_x$ for every $x \in S$ and $\pi|_S$ takes values in $\bigwedge^2 TS \oplus \bigwedge^2 N$. The fact that the Poisson bivector, restricted to S, does not coincide with the component of $\pi|_S$ in $\bigwedge^2 TS$, implies in particular that the inclusion $\iota : S \hookrightarrow M$ is certainly not a Poisson map (except of course if S is a Poisson submanifold). Item 3. is useful to further segregate different kinds of submanifolds within the family of Poisson-Dirac submanifolds. Poisson submanifolds form one extremity of this family, for which $\pi^\sharp(TS^\circ) = 0$. The other extremity is represented by the following condition:

Definition 4.75. *Let S be a Poisson-Dirac submanifold of a Poisson manifold M. We say that S is a* cosymplectic submanifold *(or a* Poisson transversal*) if $TM|_S = TS \oplus \pi^\sharp(TS^\circ)$*

Example 4.76. In Example 4.58, any one-dimensional submanifold of $\mathbb{R}^3 - \{z-\text{axis}\}$ that is transversal to the symplectic leaves is a cosymplectic submanifold of (M, π), on which the Poisson structure is zero (see [Crainic et al., 2021]).

Exercise 4.77. This exercise is taken from [Crainic et al., 2021]. Let (M, π) be a Poisson manifold and let S be a cosymplectic submanifold. Let π_S be the induced Poisson structure obtained by Poisson-Dirac reduction (see Proposition 4.72). Show that if C is a Casimir element of π, then $C|_S$ is a Casimir element of π_S.

For Poisson submanifolds, $\pi^\sharp(TS^\circ)$ has rank zero, while for cosymplectic manifolds, it is of maximal rank $n - \dim(S)$. Subsection 2.2 of these lecture notes give plenty of informations on cosymplectic manifolds. Their main use is that their class contain what are called second class constraint surfaces in Hamiltonian mechanics. To any physical system in hamiltonian mechanics corresponds a configuration manifold Q (with local coordinates the generalized coordinates) and a canonically associated phase space T^*Q (with fiber coordinates the conjugate momenta). The phase space is a symplectic manifold, characterized by the canonical symplectic form $\omega = \sum_i dp_i \wedge dq^i$, dual to a non-degenerate Poisson bivector $\pi = \sum_i \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i}$. A state of the physical system corresponds to a point in the phase space. The equations of motions then govern the evolution of the state of the system and, accordingly, the trajectory of the point to which it is associated.

Sometimes, it may happen that the states that the physical system can occupy are *constrained* (by some physical constraint, such as e.g. the length of the thread of the pendulum). A *constraint* is thus a smooth function on T^*M such that physical states are points of its zero level set. A physical system may admit several (non-necessarily functionally independent) constraints $\phi_1, \ldots, \phi_m$, so that the physical state is contained to the *constraint surface* $\Sigma = \Phi^{-1}(0)$, where $\Phi : T^*Q \longrightarrow \mathbb{R}^r$ is uniquely defined as $\Phi(x) = (\phi_1(x), \ldots, \phi_m(x))$. There are two main kinds of constraints: *first-class constraints* and *second-class constraints*. First class constraints are those constraints whose Poisson bracket with any function vanishing on Σ is zero; we denote them $\varphi_1, \ldots, \varphi_s$. Second-class constraints are those which are not first-class, and are often denoted $\chi_1, \ldots, \chi_r$ (so that $r + s = m$). In particular it means that for any second class constraint χ_k, there exist at least another second class constraint χ_l such that $\{\chi_k, \chi_l\} \neq 0$ on Σ. We define the zero level set of the second-class constraints Σ_0 – it obviously includes Σ. Then Dirac has shown that at least in the neighborhood of the zero level set of the second-class constraints Σ_0, one can define a Poisson bracket on T^*Q (or at least on some tubular neighborhood of Σ_0), because the matrix of functions $C = \big(\{\chi_k, \chi_l\}\big)_{k,l}$ is invertible:

$$\{f, g\}_{Dirac} = \{f, g\} - \{f, \chi_k\}(C^{-1})^{kl}\{\chi_l, g\} \tag{4.41}$$

This bracket, called the *Dirac bracket*, is such that the second-class constraint become Casimirs of this new bracket and that Σ_0 is a symplectic leaf – hence a Poisson submanifold – of $(T^*Q, \{\,.\,,.\,\}_{Dirac}$.

Exercise 4.78. Show that any second class constraint χ_l is a Casimir element of the Dirac bracket.

Let us now explain in geometric terms what is happening. Let M be a symplectic manifold, whose corresponding non-degenerate Poisson bracket is denoted $\{\,.\,,.\,\}$. Let $\Phi : M \longrightarrow \mathbb{R}^r$

(where $r \leq \dim(M)$) be a smooth map and assume that 0 is a *regular value* of Φ. It means that $\Phi_{*,x} : T_xM \longrightarrow T_{\Phi(x)}\mathbb{R}^r$ is surjective for every $x \in \Phi^{-1}(0)$. Since Φ is a smooth map then the map $x \mapsto \mathrm{rk}(\Phi_{*,x})$ is lower semi-continuous, so it means that there exists an open neighborhood U of the origin of $\mathbb{R}^r$ such that Φ_* is surjective on the open set $\Phi^{-1}(U)$. By the regular level set Theorem 3.44, the level sets of every points of U are closed embedded submanifolds of M, which form a regular foliation of $\Phi^{-1}(U)$. We denote $\Sigma_0 = \Phi^{-1}(0)$ the level set of 0. Decomposing the map Φ on the basis of $\mathbb{R}^r$: $\Phi(x) = (\chi_1(x), \chi_2(x), \ldots, \chi_r(x))$ gives r smooth functions $\chi_i \in \mathcal{C}^\infty(M), 1 \leq i \leq r$, called *constraints*. They are said *irreducible* because they are functionally independent, i.e. if there are smooth functions f_i such that $\sum_i f_i\chi_i = 0$ then all the f_i are zero. Then, the level sets of Φ on U are $n-r$ dimensional closed embedded submanifolds. Moreover, the regularity condition can now be stated as follows: $d\chi_1 \wedge \ldots \wedge d\chi_r \in \Gamma(\bigwedge^r T^*M)$ is non vanishing on $\Phi^{-1}(U)$.

Let C be the anti-symmetric matrix of functions whose i, j-th component is:

$$C_{ij} = \{\chi_i, \chi_j\}$$

We further assume that:

$$\det(C) \neq 0 \quad \text{on} \quad \Sigma_0 \tag{4.42}$$

This condition on the smooth functions $(\chi_i)_i$ characterizes *second class constraints*, and Σ_0 is called the second class constraints surface. Since condition (4.42) is an open condition, there exists a tubular neighborhood $V \subset \Phi^{-1}(U)$ of Σ_0 (because Σ_0 is embedded, see Theorem 10.19 in [Lee, 2003]) such that $\det(C) \neq 0$ on the whole of V. Let us show that this condition is central in the properties of Σ_0:

Proposition 4.79. *The second class constraint surface Σ_0 is a cosymplectic submanifold of $(M, \{\,.\,,.\,\})$. In particular it is not a Poisson submanifold.*

Proof. Just for the sake of the exercise, let us first show that Σ_0 is *not* a Poisson submanifold. Since it is embedded in M, by the discussion below Proposition 4.67, the condition for Σ_0 to be a Poisson submanifold is that the multiplicative ideal $\mathcal{I}_{\Sigma_0}$ of smooth functions on M vanishing on Σ_0 is a Lie subalgebra of $(\mathcal{C}^\infty(M), \{\,.\,,.\,\})$. Since Σ_0 is an embedded submanifold, the multiplicative ideal $\mathcal{I}_{\Sigma_0}$ is generated – as a sub-algebra of $\mathcal{C}^\infty(M)$ – by the second class constraints $\chi_1, \ldots, \chi_r$ (see Theorem 1.1 in [Henneaux and Teitelboim, 1994] or pages 95-96 of [Sudarshan and Mukunda, 2015]). Then, if it ever occurred that $\mathcal{I}_{\Sigma_0}$ was a Lie ideal of $\mathcal{C}^\infty(M)$, then it would mean that $\{\mathcal{I}_{\Sigma_0}, \mathcal{I}_{\Sigma_0}\} \subset \mathcal{I}_{\Sigma_0}$. In other words, every Poisson bracket $\{\chi_i, \chi_j\}$ would vanish on Σ_0. But this would contradict condition (4.42). Hence Σ_0 is not a Poisson submanifold.

Now let us proof that on the contrary, Σ_0 is quite far from being a Poisson submanifold, as it is a cosymplectic submanifold. Let us first proof that it is Poisson-Dirac. Since Σ_0 is a closed embedded submanifold, by Lemma 4.69, we know that every smooth function $f \in \mathcal{C}(\Sigma_0)$ admits a global extension, i.e. a smooth function $F \in \mathcal{C}^\infty(M)$ on M such that $F|_S = f$. Actually, we have the following isomorphism: $\mathcal{C}^\infty(\Sigma_0) \simeq \mathcal{C}^\infty(M)\big/\mathcal{I}_{\Sigma_0}$. So, any other choice of function $F + \sum_i \lambda^i\chi_i$ (where the λ^i's are smooth functions on M) is another global extension for f. Let us find such an extension which is horizontal, i.e. whose hamiltonian vector field is tangent to Σ_0. Let us set $\theta_i = \{F, \chi_i\}$ and let us set $\widetilde{f} = F - \theta_k C^{kl}\chi_l$ (summation implied), where for simplicity the C^{kl} denote the coefficients of the inverse matrix C^{-1}. Then one has:

$$X_{\widetilde{f}}(\chi_i) = \{F - \theta_k C^{kl}\chi_l, \chi_i\} = \underbrace{\{F, \chi_i\}}_{=\theta_i} - \underbrace{(\{\theta_k, \chi_i\}C^{kl} + \theta_k\{C^{kl}, \chi_i\})\chi_l}_{\text{vanishes on }\Sigma_0\text{ because of }\chi_l} - \underbrace{\theta_k C^{kl}C^{li}}_{=\theta_k\delta^k_i = \theta_i}$$

which then vanishes on Σ_0. Since the multiplicative ideal $\mathcal{I}_{\Sigma_0}$ is generated by the constraints χ_i, it means that $X_{\widetilde{f}}(\mathcal{I}_{\Sigma_0}) \subset \mathcal{I}_{\Sigma_0}$, i.e. all functions in $X_{\widetilde{f}}(\mathcal{I}_{\Sigma_0})$ vanish on Σ_0. Then, since Σ_0 is embedded, we have the equality (see beginning of subsection 3.4):

$$T_x\Sigma_0 = \left\{ X_x \in T_xM \,\middle|\, X_x(g) = 0 \text{ whenever } g \in \mathcal{I}_{\Sigma_0} \right\} \tag{4.43}$$

Since $X_{\widetilde{f}}$ belongs to the right-hand side, it means that it is tangent to Σ_0. Hence, the smooth function $\widetilde{f} = F - \theta_k C^{kl}\chi_l$ is a horizontal (global) extension of f. This proves that Σ_0 is a Poisson Dirac submanifold of M.

Now, as the constraints χ_i generate $\mathcal{I}_{\Sigma_0}$, and that equality (4.43) holds, we deduce that the differential one-forms $d\chi_i$ form a frame of $T\Sigma_0^\circ$, for they are independent and $d\chi_i(X) = X(\chi_i) = 0$ if and only if the vector field X takes values in $T\Sigma_0$. Then, since the Poisson bivector field π is non-degenerate, it sends the rank r subbundle $T\Sigma_0^\circ$ to a rank r subbundle of $T\Sigma_0$. Since the rank of the vector bundle $T\Sigma_0$ is $n-r$ and that a Poisson-Dirac submanifold satisfies $T\Sigma_0 \cap \pi^\sharp(T\Sigma_0) = 0$, we conclude that $TM|_{\Sigma_0} = T\Sigma_0 \oplus \pi^\sharp(T\Sigma_0)$. In other words, Σ_0 is a cosymplectic submanifold of M. □

Since Σ_0 is a Poisson-Dirac submanifold of $(M, \{\,.\,,.\,\})$, we denote by $\{\,.\,,.\,\}_{\Sigma_0}$ the Poisson bracket inherited by Σ_0 via Poisson-Dirac reduction. More generally using the same arguments as in the proof of Proposition 4.79, one can show that every level sets of the smooth map Φ are cosymplectic submanifolds of $(M, \{\,.\,,.\,\})$ (at least on $\Phi^{-1}(U)$), so they all inherit the a Poisson structure from that on M via Poisson-Dirac reduction. Now, the Dirac bracket defined in Equation (4.41) is another choice of Poisson structure on M (or at least on some tubular neighborhood V of Σ_0), relative to which the second class constraints χ_i are Casimirs. Then, by Example 4.68, we deduce that the level sets of Φ are the symplectic leaves of $(M, \{\,.\,,.\,\}_{Dirac})$ (but *not* of $(M, \{\,.\,,.\,\})$). Then, the second-class constraint surface Σ_0 is a cosymplectic submanifold of $(M, \{\,.\,,.\,\})$ but is a Poisson submanifold of $(M, \{\,.\,,.\,\}_{Dirac})$ (or at least the tubular neighborhood V). What is even more interesting is the following result:

Proposition 4.80. *For simplicity assume that $\{\,.\,,.\,\}_{Dirac}$ is defined on the whole of M. Then the Poisson structure on Σ_0 making it a Poisson submanifold of $(M, \{\,.\,,.\,\}_{Dirac})$ is precisely the Poisson bracket $\{\,.\,,.\,\}_{\Sigma_0}$ inherited from $\{\,.\,,.\,\}$ via Poisson-Dirac reduction.*

Proof. We need to show that for any two smooth functions $f, g \in \mathcal{C}^\infty(\Sigma_0)$, one has on Σ_0:

$$\{\iota^*(f), \iota^*(g)\}_{\Sigma_0} = \{f, g\}_{Dirac}|_{\Sigma_0} \tag{4.44}$$

Since the second class constraints χ_i are Casimirs elements of the Dirac bracket, we have, for every smooth functions $f, g \in \mathcal{C}^\infty(M)$:

$$\{f, g\}_{Dirac}\big|_{\Sigma_0} = \{f - \lambda^i\chi_i, g - \mu^j\chi_j\}_{Dirac}\big|_{\Sigma_0} \tag{4.45}$$

for any family of functions λ^i, μ^j (notice that the equality only holds on Σ_0). In particular, one can make special choices of λ^i and μ^j as in the proof of Proposition 4.79 so that the hamiltonian vector fields of $f - \lambda^i\chi_i$ and $g - \mu^j\chi_j$ are tangent to Σ_0. Then, the very definition of the Dirac bracket implies that we have the following equality:

$$\{f - \lambda^i\chi_i, g - \mu^j\chi_j\}_{Dirac}\big|_{\Sigma_0} = \{f - \lambda^i\chi_i, g - \mu^j\chi_j\}\big|_{\Sigma_0} \tag{4.46}$$

Again, the identity holds only on Σ_0. Now, the fact that f and $f - \lambda^i\chi_i$ coincide on Σ_0 can be written as $\iota^*(f) = \iota^*(f - \lambda^i\chi_i) \in \mathcal{C}^\infty(\Sigma_0)$, where $\iota : \Sigma_0 \longmapsto M$ is the inclusion map. Moreover, since we have that $\mathcal{C}^\infty(\Sigma_0) \simeq \mathcal{C}^\infty(M)\big/\mathcal{I}_{\Sigma_0}$, the smooth function f is a global extension of $\iota^*(f)$, and $f - \lambda^i\chi_i$ is an horizontal one. Then, together with Equations (4.39), (4.45) and (4.46), we deduce that:

$$\begin{aligned}
\{\iota^*(f), \iota^*(g)\}_{\Sigma_0} &= \{\iota^*(f - \lambda^i\chi_i), \iota^*(g - \mu^j\chi_j)\}_{\Sigma_0}\\
&= \{f - \lambda^i\chi_i, g - \mu^j\chi_j\}\big|_{\Sigma_0}\\
&= \{f - \lambda^i\chi_i, g - \mu^j\chi_j\}_{Dirac}\big|_{\Sigma_0}\\
&= \{f, g\}_{Dirac}\big|_{\Sigma_0}
\end{aligned}$$

which is Equation (4.44), as desired. □

Another way of making sense of Proposition 4.80 is the following: for any two smooth functions $f, g \in \mathcal{C}^\infty(\Sigma_0)$, one has:

$$\{f, g\}_{\Sigma_0} = \{\widetilde{f}, \widetilde{g}\}\big|_{\Sigma_0} = \{\widehat{f}, \widehat{g}\}_{Dirac}\big|_{\Sigma_0} \tag{4.47}$$

where on the one hand, $\widetilde{f}, \widetilde{g} \in \mathcal{C}^\infty(M)$ are any horizontal local extensions of f, g (they are required in Poisson-Dirac reduction), and on the other hand $\widehat{f}, \widehat{g} \in \mathcal{C}^\infty(M)$ are *any* local extensions of f, g (since in that case $\iota^*(\widehat{f}) = f$ and $\iota^*(\widehat{g}) = g$, making Equation (4.44) valid). This is the formalized content of Theorem 2.5 in [Henneaux and Teitelboim, 1994]. Let us give a final, alternative formulation: since the second class constraint surface Σ_0 is a cosymplectic submanifold of $(M, \{\,.\,,.\,\})$, the tangent bundle restricted to Σ_0 is a direct sum of the two subbundles $T\Sigma_0$ and $\pi^\sharp(T\Sigma_0^\circ)$, so that one can see the term $\{-, \chi_k\}(C^{-1})^{kl}\{\chi_l, -\}$ in the formula (4.41) defining the Dirac bracket as a bivector field taking values in $\pi^\sharp(T\Sigma_0^\circ)$, which precisely compensates the block D in formula (4.40). Then, the matricial representation of the Dirac bracket with respect to adapted local coordinates around a point $x \in \Sigma_0$ becomes

$$\begin{pmatrix} A & B \\ -B^t & 0 \end{pmatrix} \tag{4.48}$$

Thus, to create the Dirac bracket one has removed the lower right component of the original Poisson bracket represented matricially in formula (4.40). Moreover we can see from the above matrix (Equation (4.48)) that on Σ_0, the bivector associated to $\{\,.\,,.\,\}_{Dirac}$ reduces to A, which takes values in $\bigwedge^2 T\Sigma_0$, as is characteristic for a Poisson submanifold. See subsection 5.1 of [Calvo et al., 2010] for more details on this background story.

Example 4.81. Let $M = T^*\mathbb{R}^2$ be the cotangent bundle of $\mathbb{R}^2$, and let denote the coordinate functions (x, y, p_x, p_y). The canonical (non-degenerate) Poisson bracket on M is then $\pi = \frac{\partial}{\partial x} \wedge \frac{\partial}{\partial p_x} + \frac{\partial}{\partial y} \wedge \frac{\partial}{\partial p_y}$. Let us set $\chi_1 = p_y$ and $\chi_2 = p_x + x - 2y$; these two smooth functions on M make $\Phi = (\chi_1, \chi_2) : M \longrightarrow \mathbb{R}^2$ a submersion. Then, the level set of Φ at 0 is a 2-dimensional plane in M that we denote Σ_0. The Poisson bracket of χ_1 and χ_2 is:

$$\{\chi_1, \chi_2\} = 2$$

So in particular, denoting $C_{ij} = \{\chi_i, \chi_j\}$, one obtains:

$$C = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$$

which is constant on the whole of M. Then $\det(C) \neq 0$, and the inverse matrix is:

$$C^{-1} = \begin{pmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$$

Being defined over the whole of M, the Dirac bracket will be defined on the whole of M:

$$\begin{aligned}\{f,g\}_{Dirac} &= \{f,g\} - \{f,\chi_1\} \times \big(-\frac{1}{2}\big) \times \{\chi_2, g\} - \{f,\chi_2\} \times \frac{1}{2} \times \{\chi_1, g\}\\ &= \{f,g\} + \frac{1}{2}\frac{\partial f}{\partial y}\Big(-\frac{\partial g}{\partial x} + \frac{\partial g}{\partial p_x} - 2\frac{\partial g}{\partial p_y}\Big) - \frac{1}{2}\Big(-\frac{\partial f}{\partial x} + \frac{\partial f}{\partial p_x} - 2\frac{\partial f}{\partial p_y}\Big)\frac{\partial g}{\partial y}\end{aligned}$$

One can check that χ_1 and χ_2 are Casimir elements of the Dirac bracket.

Example 4.82. Another example of a situation where the matrix is invertible on the whole of M is the following: let $M = T^*\mathbb{R}^3$ and let denote the coordinate functions (x, y, z, p_x, p_y, p_z). Let us define the following four linear functions:

$$\chi_1 = x + y, \quad \chi_2 = p_x, \quad \chi_3 = p_y + p_z \quad \text{and} \quad \chi_4 = z - x$$

The level set of $\Phi = (\chi_1, \chi_2, \chi_3, \chi_4) : M \longrightarrow \mathbb{R}^4$ at 0 is a 2-dimensional plane, that we denote Σ_0. This plane is not a Poisson submanifold of M (with respect to its canonical non-degenerate Poisson structure) because the Poisson brackets of the constraints χ_i do not all vanish on Σ_0. Indeed, the matrix C whose i, j-th component is $\{\chi_i, \chi_j\}$ is:

$$C = \begin{pmatrix} 0 & 1 & 1 & 0\\ -1 & 0 & 0 & 1\\ -1 & 0 & 0 & -1\\ 0 & -1 & 1 & 0 \end{pmatrix}$$

It has determinant 4 and is invertible on the whole of M, with inverse matrix:

$$C^{-1} = \frac{1}{2}\begin{pmatrix} 0 & -1 & -1 & 0\\ 1 & 0 & 0 & -1\\ 1 & 0 & 0 & 1\\ 0 & 1 & -1 & 0 \end{pmatrix}$$

The corresponding Dirac bracket is so that the constraints χ_i are Casimirs elements, and the plane Σ_0 is a Poisson submanifold of $(M, \{\,.\,,.\,\}_{Dirac})$.

Example 4.83. Let $M = T\mathbb{R}^2$ and let $\chi_1 = xy - 1$ while $\chi_2 = p_x$. The smooth map $\Phi = (\chi_1, \chi_2) : M \longrightarrow \mathbb{R}^2$ is so that Φ_* is surjective on $\Phi^{-1}(0)$. Then, the preimage $\Sigma_0 = \Phi^{-1}(0)$ is a closed embedded submanifold of M. It has two connected components because the preimage of the first constraint χ_1 has such. The Poisson bracket of the two constraints is:

$$\{\chi_1, \chi_2\} = y$$

Interestingly, this Poisson bracket vanishes on the hyperplane of equation $y = 0$, but this hyperplane does not intersect Σ_0 so the Poisson bracket never vanishes on Σ_0. Hence, the matrix C is:

$$C = \begin{pmatrix} 0 & y\\ -y & 0 \end{pmatrix}$$

Then $\det(C) \neq 0$ on Σ_0 (not on the whole of M) because $y \neq 0$ on the surface, and the inverse matrix is (only defined in a neighborhood of Σ_0):

$$C^{-1} = \begin{pmatrix} 0 & -\frac{1}{y}\\ \frac{1}{y} & 0 \end{pmatrix}$$

The Dirac bracket will then be defined only in a tubular neighborhood of Σ_0:

$$\begin{aligned}\{f,g\}_{Dirac} &= \{f,g\} - \{f,\chi_1\} \times \big(-\frac{1}{y}\big) \times \{\chi_2, g\} - \{f,\chi_2\} \times \frac{1}{y} \times \{\chi_1, g\}\\ &= \{f,g\} + \frac{1}{y}\Big(\frac{\partial f}{\partial p_x}y + \frac{\partial f}{\partial p_y}x\Big)\frac{\partial g}{\partial x} - \frac{1}{y}\frac{\partial f}{\partial x}\Big(\frac{\partial g}{\partial p_x}y + \frac{\partial g}{\partial p_y}x\Big)\end{aligned}$$

The constraints χ_1 and χ_2 are Casimirs of this bracket.

Coming back to our problem in Hamiltonian mechanics, first-class constraints define a submanifold Σ in the embedded cosymplectic submanifold Σ_0. This submanifold is however not a Poisson-Dirac submanifold because first-class constraints satisfy a nullity condition on Σ: $\{\varphi_p, \varphi_q\} = 0$ (so Σ does not satisfy item 4. of 4.74). Rather, the submanifold Σ is what is called a *coisotropic submanifold*. Poisson brackets cannot descend to them, but under some circonstances, to a quotient of them, through a procedure called *Poisson reduction*. The notion of coisotropy is well-known in symplectic geometry, and is attached to submanifolds S whose symplectic orthogonal $TS^{\perp_\omega}$ is a sub-bundle of TS. Since, for a non-degenerate Poisson structure, one has $\pi^\sharp(TS^\circ) = TS^{\perp_\omega}$, the condition that a submanifold is coisotropic is straighforwardly transported to the realm of Poisson geometry:

Definition 4.84. *A* coisotropic submanifold *of a Poisson manifold* (M, π) *is a submanifold* $S \overset{\iota}{\hookrightarrow} M$ *such that* $\pi^\sharp(TS^\circ) \subset TS$.

Example 4.85. Any codimension 1 submanifold S of a Poisson manifold is coisotropic because TS° is 1-dimensional, implying that the right-hand side of Equation (4.25) is zero, implying in turn that $\pi^\sharp(TS^\circ) \subset TS$.

Example 4.86. An interesting example of a coisotropic submanifold is provided by a theorem of A. Weinstein [Weinstein, 1988]: A smooth map $\varphi : (M_1, \pi_1) \longrightarrow (M_2, \pi_2)$ is a Poisson map if and only if its graph $\mathrm{Gr}(\varphi) \subset M_2 \times M_1^-$ is a coisotropic submanifold (where M_1^- is the smooth manifold M_1 equipped with the opposite Poisson structure $-\pi_1$). This statement is the Poisson equivalent of the well-known result in symplectic geometry stating that if M_1, M_2 are symplectic manifolds, then a diffeomorphism $\varphi : (M_1, \omega_1) \longrightarrow (M_2, \omega_2)$ is a symplectomorphism if and only if its graph $\mathrm{Gr}(\varphi) \subset M_2 \times M_1^-$ is a Lagrangian submanifold.

There are two distinguished sub-families of coisotropic subamnifolds: those for which $\pi^\sharp(TS^\circ) = 0$, i.e. Poisson submanifolds, and on the other extreme those for which $\pi^\sharp(TS^\circ) = TS$; they are called *Lagrangian submanifolds* as they correspond to their counterparts in symplectic geometry. Obviously, given the condition established in Definition 4.84 and item 4. of Proposition 4.74, the intersection of the set of coisotropic submanifolds and the set of Poisson-Dirac submanifolds is precisely the set of Poisson submanifolds. As for the other kinds of submanifolds, coisotropic submanifolds have equivalent alternative definitions:

Proposition 4.87. *Let S be a submanifold of a Poisson manifold M. Then, the following are equivalent:*

1. *S is a coisotropic submanifold;*
2. *for every smooth function $f \in \mathcal{C}^\infty(M)$ vanishing on some open set $V \subset S$, the Hamiltonian vector field X_f is tangent to S at every point of V;*
3. *$\langle \bigwedge^2 TS^\circ, \pi \rangle = 0$, where $\langle\, .\, , .\, \rangle$ is the pairing between T^*M and TM.*

Proof. The direction 1. $\Longrightarrow$ 2. is straightforward because $f|_V = 0$ means that $df \in TS^\circ|_V$, so let us turn to the direction 2. $\Longrightarrow$ 1. Let f be such a function vanishing on V and suppose $X_{f,x} \in T_xS = T_xV$ for every point $x \in V$. Let $\xi \in \Gamma(TS^\circ)$ then, one has on V:

$$0 = \xi(X_f) = -df\big(\pi^\sharp(\xi)\big) \tag{4.49}$$

We know for sure that $df \in \Gamma(TS^\circ|_V)$ but the fact that, upon shrinking it, V is an embedded submanifold of M (see Proposition 3.54), implies that $TV^\circ = (TS)^\circ|_V$ is spanned by the pointwise evaluation of exact differential one-forms df for those functions $f \in \mathcal{C}^\infty(M)$ vanishing on V. Since Equation (4.49) holds for every such function, and every $\xi \in \Gamma(TS^\circ)$, one deduces that $\pi^\sharp(\xi)$ is necessarily a tangent vector to S at every point of V. The proof of the equivalence 1. $\Longleftrightarrow$ 3. is straightforward, using Equation (4.25). ☐

Remark 4.88. Notice that in the second item of Proposition 4.87, we did not ask f to vanish on S but on an open set of V precisely because we needed to characterize $TS^\circ|_V = TV^\circ$ as spanned by the exact differential one-forms df. And to do that we needed at least an embedded submanifold, which is true for V but not necessarily for S if immersed.

We have another characterization of coisotropic submanifolds, mimicking Proposition 4.67 for Poisson submanifolds. Indeed, it admits the following counterpart for coisotropic submanifolds:

Proposition 4.89. *Let S be a coisotropic submanifold of the Poisson manifold M. Then, the multiplicative ideal:*

$$\mathcal{I}_S = \Big\{ f \in \mathcal{C}^\infty(M) \text{ such that } f|_S \equiv 0 \Big\}$$

is a Lie subalgebra of the Lie algebra $(\mathcal{C}^\infty(M), \{\,.\,,\,.\,\})$.

The proof of Proposition 4.89 is a straightforward application of Definition 4.84. Notice however that, as for Poisson submanifolds, the converse implication – that the ideal $\mathcal{I}_S$ being a Lie subalgebra of $\mathcal{C}^\infty(M)$ implies that S is a coisotropic submanifold of M – is true only when S is embedded in M.

Example 4.90. Taken from [Crainic et al., 2021]: let $\mathfrak{g}$ be a finite dimensional real Lie algebra and $\mathfrak{g}^*$ be the associated linear Poisson manifold, described in Example 4.4. Let $\xi \in \mathfrak{g}^*$, then the definition of the linear Poisson structure on $\mathfrak{g}^*$ implies that, for any two elements $x, y \in \mathfrak{g}$:

$$\{\overline{x}, \overline{y}\}(\xi) = \overline{[x,y]}(\xi) = \xi([x,y]) \tag{4.50}$$

where $\overline{x}$ is the notation used in Example 4.4 to symbolize the linear form on $\mathfrak{g}^*$ defined as $\overline{x}(\xi) = \xi(x)$. Using Equations (4.24) and (4.25), the left-hand side of Equation (4.50) can be written as:

$$\{\overline{x}, \overline{y}\} = d\overline{y}(\pi^\sharp(d\overline{x})) \tag{4.51}$$

Let V be a subspace of the Lie algebra $\mathfrak{g}$, and let $V^\circ \subset \mathfrak{g}^*$ be its annihilator, that we will denote S in the following. Then, the annihilator of TS is spanned by the elements $d\overline{x}$ for every $x \in V$. This implies that S is a coisotropic submanifold of $\mathfrak{g}^*$ if and only if, for every $x, y \in V$, the right-hand side of Equation (4.51) – evaluated at a point $\xi \in S = V^\circ$ – vanishes, i.e. if and only if the right-hand side of Equation (4.50) vanishes for every $\xi \in V^\circ$. This implies in turn that V° is a coisotropic submanifold of $\mathfrak{g}^*$ if and only if V is a Lie subalgebra of $\mathfrak{g}$. Since $\mathcal{I}_{V^\circ}$ is generated by $V^{\circ\circ} = V$ (because $\mathfrak{g}$ is finite dimensional), we deduce that V° is a coisotropic submanifold of $\mathfrak{g}^*$ if and only if $\mathcal{I}_{V^\circ}$ is a Lie subalgebra of $\mathcal{C}^\infty(\mathfrak{g}^*)$. As a final remark, notice that V° is a Poisson submanifold if and only if V is a Lie ideal, if and only if $\mathcal{I}_{V^\circ}$ is a Lie ideal of $\mathcal{C}^\infty(\mathfrak{g}^*)$.

Contrary to what happens for Poisson-Dirac submanifolds, coisotropic submanifolds are rarely equipped with an induced Poisson bracket. Rather, one may only have a *Poisson reduction* on a quotient of coisotropic submanifolds. Let us first define what is meant by this concept (we use the terminology of subsection 5.2.2 in [Laurent-Gengoux et al., 2013]):

Definition 4.91. *Let $(M, \{\,.\,,\,.\,\})$ be a Poisson manifold and S be a submanifold, N a smooth manifold and $p : S \longrightarrow N$ a surjective submersion:*

$$\begin{tikzcd} S \subset M \arrow[d, two heads, "p"'] \\ N \end{tikzcd}$$

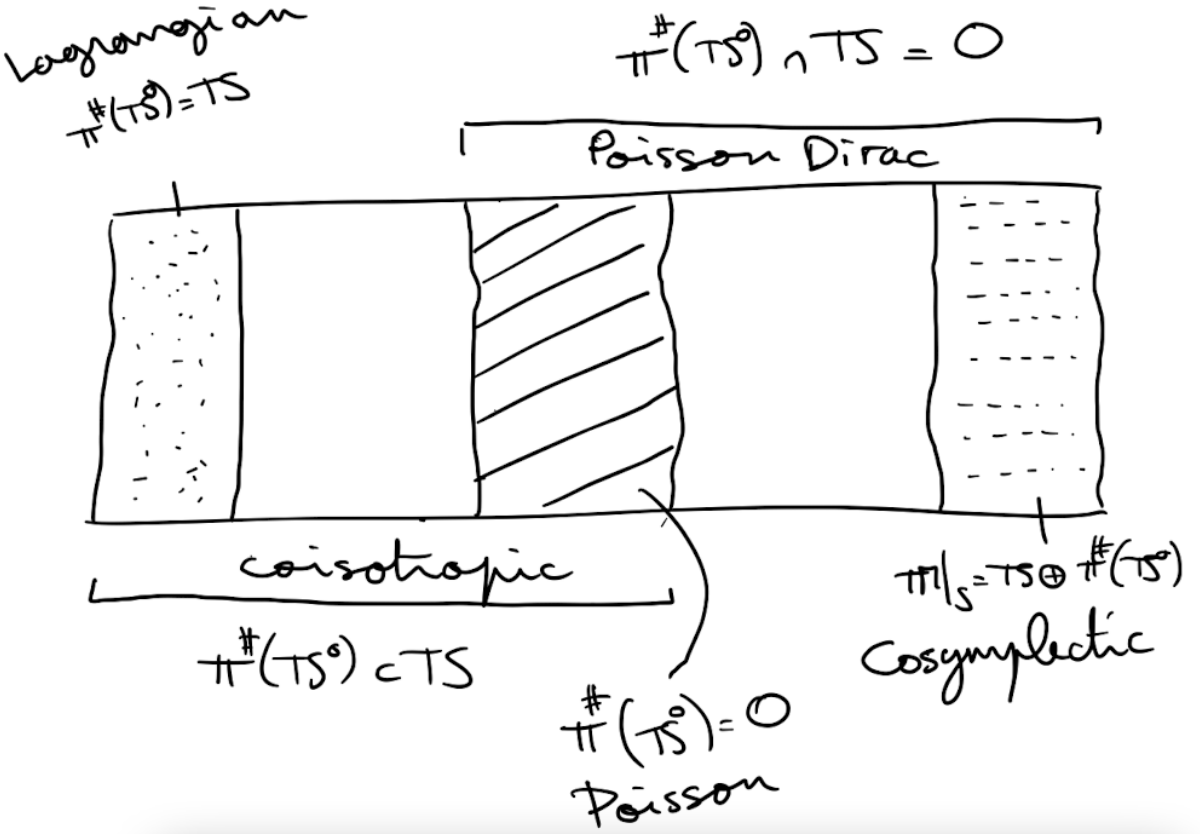

Figure 18: Schematic map of the various families of submanifolds in Poisson geometry. Poisson submanifolds are both coisotropic and Poisson-Dirac submanifolds. Cosymplectic and Lagrangian submanifolds are opposite to Poisson submanifolds in their respective families. See [Zambon, 2011] for additional informations about relationships between various kinds of submanifolds in Poisson geometry.

We say that the triple (M, S, N) *is* Poisson reducible *if there exists a Poisson structure* $\{.,.\}_N$ *on* N *such that, for all open subsets* $V \subset S$ *and* $U \subset M$ *such that* $V \subset U \cap S$*, and for all functions* $f, g \in \mathcal{C}^\infty(p(V))$*, one has:*

$$\{f, g\}_N(p(x)) = \{\widetilde{f}, \widetilde{g}\}(x) \tag{4.52}$$

for every $x \in V$*, and arbitrary local extensions* $\widetilde{f}, \widetilde{g} \in \mathcal{C}^\infty(U)$ *of functions* $f \circ p|_V$ *and* $g \circ p|_V$*.*

Remark 4.92. Poisson reduction is a particular case of what is called Marsden-Ratiu reduction on Poisson manifold, which also generalize Mayer-Marsden-Weinstein reduction on Hamiltonian G-spaces. See these lectures notes, as well as this paper [Falceto and Zambon, 2008].

Example 4.93. If S is a submanifold of M, and f, g are two smooth functions on S, admitting local extensions $\widetilde{f}$ and $\widetilde{g}$, then $\iota^*\widetilde{f} = f$ and $\iota^*\widetilde{g} = g$. Assuming that S is a Poisson submanifold, we set $N = S$, $p = \mathrm{id}_S$, so that Equation (4.33) becomes (4.52). This property being true for every smooth functions f, g, the triple (M, S, S) is Poisson reducible.

It turns out – by Proposition 5.11 of [Laurent-Gengoux et al., 2013] – that a triple (M, S, N) satisfying the conditions of Definition 4.91 is Poisson reducible if and only if:

1. for every function $\widetilde{f} \in \mathcal{C}^\infty(U)$ whose restriction to V is constant of the fibers of p, the hamiltonian vector field $X_{\widetilde{f}}$ is tangent to S at every point of V;

2. for every pair of functions $\widetilde{f}, \widetilde{g} \in \mathcal{C}^\infty(U)$ whose restriction to V is constant of the fibers of p, the restriction of their Poisson bracket to V is constant of the fibers of p.

Then, one can show that in such a case S is a coisotropic submanifold of M. It is thus legitimate to ask under which circumstances a coisotropic S allows a Poisson reduction to some quotient of itself:

Proposition 4.94. *Let S be a coisotropic submanifold of M, and assume that $\pi^\sharp(TS^\circ)$ has constant rank over S (i.e. defines a regular smooth distribution). Then it is integrable in the sense of Frobenius and if the space of leaves N of the corresponding regular foliation is a smooth manifold, the triple (M, S, N) is Poisson reducible.*

Proof. We will show that items 1. and 2. above are satisfied (see also Remark 5.15 in [Laurent-Gengoux et al., 2013]). In the present context, p is the quotient map sending S to the leaf space N, so the fibers of p are the leaves.

First, let $\widetilde{f} \in \mathcal{C}^\infty(U)$ such that it is constant on the leaves, and let ξ a differential one-form taking values in TS° on S. Since S is a coisotropic submanifold, $\pi^\sharp(\xi)$ is a vector field taking values in the regular integrable distribution, that is to say it is tangent to the leaves. Since $\widetilde{f}$ is constant along the leaves, $d\widetilde{f}(\pi^\sharp(\xi)) = \pi^\sharp(\xi)(\widetilde{f}) = 0$ on S. By Equation (4.25), the left-hand side of the former equation is equal to $-\xi(\pi^\sharp(d\widetilde{f}))$. Then it vanishes on S and since ξ takes values in TS°, and that the vanishing of $-\xi(\pi^\sharp(d\widetilde{f}))$ is valid for any such ξ, we deduce that $X_{\widetilde{f}} = \pi^\sharp(d\widetilde{f})$ is tangent to S.

Secondly, let $\widetilde{f}, \widetilde{g} \in \mathcal{C}^\infty(U)$ be two smooth functions which are constant along the leaves and let ξ be a differential one form taking values in TS°. Then, by the first point just proven, $X_{\widetilde{f}}$ and $X_{\widetilde{g}}$ are tangent to S, so is their Lie bracket, and we have:

$$0 = \xi([X_{\widetilde{f}}, X_{\widetilde{g}}]) = \xi(X_{\{\widetilde{f},\widetilde{g}\}}) = \xi(\pi^\sharp(d\{\widetilde{f},\widetilde{g}\})) = -d\{\widetilde{f},\widetilde{g}\}(\pi^\sharp(\xi)) = -\pi^\sharp(\xi)(\{\widetilde{f},\widetilde{g}\})$$

Since by definition, $\pi^\sharp(\xi)$ is a vector field tangent to the leaves, and that $\pi^\sharp(TS^\circ)$ generate all such tangent vector fields, we deduce that the Poisson bracket $\{\widetilde{f},\widetilde{g}\}$ is constant along the leaves, as required. □

This proposition is quite useful to study Hamiltonian under constraints. We have seen earlier that second-class constraints define an embedded cosymplectic submanifold of a Poisson manifold M. On the other hand, first class constraint define an embedded coisotropic submanifold of M (here we assume M to be a symplectic manifold). A quick way to see this is by using the converse of Proposition 4.67, which holds for embedded submanifolds. Assume that we have s constraints $\varphi_1, \ldots, \varphi_s$ which are *irreducible* – i.e. functionally independent – and *regular* – i.e. the differential s-form $d\varphi_1 \wedge \ldots \wedge d\varphi_s$ is nowhere vanishing on the zero level set Σ defined by the smooth map $\Phi = (\varphi_1, \ldots, \varphi_s) : M \longrightarrow \mathbb{R}^s$. This proves that Φ_* is surjective on this level set (actually on a tubular neighborhood), proving in turn that Σ is an embedded submanifold of M.

Being first-class means that $\{\varphi_i, \varphi_j\} = 0$ on Σ for every $1 \leq i, j \leq s$, which is actually equivalent to saying that $\{\varphi_i, f\} = 0$ for every $f \in \mathcal{I}_S$, because every such function is functionally dependent on the constraints since Σ is an embedded submanifold (see Theorem 1.1 in [Henneaux and Teitelboim, 1994] or pages 95-96 of [Sudarshan and Mukunda, 2015]). But this is just the condition that $\mathcal{I}_S$ is a Lie subalgebra of $\mathcal{C}^\infty(M)$. Being embedded, this implies that Σ is a coisotropic submanifold of M. Since the differential one forms $d\varphi_i$ span $T\Sigma^\circ$, the hamiltonian vector fields $X_{\varphi_i} = \pi^\sharp(d\varphi_i)$ span $\pi^\sharp(T\Sigma^\circ)$ and define a regular distribution on Σ (the rank of π

is constant over Σ). By Frobenius theorem this distribution is integrable and the leaf space P is called the *reduced phase space* because its points are the physical states of the system: on the one hand they all satisfy the constraints, and on the other hand we have got rid of the gauge freedom (symbolized by the leaves of the foliation). By Proposition 4.94, if the reduced phase space is a smooth manifold, the Poisson bracket of M descends to P.

However, in most situation, we have a mixed set of constraints, i.e. some of them are first-class and some of them are second-class. Then, the strategy to obtain the physical phase space is first, to perform a Poisson-Dirac reduction on the second-class constraint surface Σ_0, which is then a symplectic embedded submanifold of $(T^*Q, \{\,.\,,.\,\}_{Dirac})$, and second, to consider the first-class constraint surface Σ as a coisotropic submanifold of $(\Sigma_0, \{\,.\,,.\,\}_{\Sigma_0})$ (or equivalently of $(T^*Q, \{\,.\,,.\,\}_{Dirac})$ because the former second-class constraint become first class with respect to the Dirac bracket). By proceeding to a Poisson reduction on Σ, one obtains the physical phase space of the theory. See this chapter for a clear presentation of this approach.

To conclude this section, let us discuss a bit more the relationship between the symplectic leaves of a Poisson manifold and its submanifolds. We know from Theorem 4.60 that there is a one-to-one correspondence between Poisson structures on M and smooth families of symplectic leaves on M. Then a way of defining a Poisson structure on a given submanifold S of M would be to to study the properties of the intersection of S with the symplectic leaves of M:

Proposition 4.95. *Let M be a Poisson manifold and let $S \subset M$ be a submanifold. Then we have the following statements:*

1. *S is a Poisson submanifold if and only if for each symplectic leaf L, the intersection $S \cap L$ is an open set of L;*
2. *S is a Poisson-Dirac submanifold if and only if for each symplectic leaf L, the intersection $S \cap L$ is* clean[15] *and a symplectic submanifold of L, such that these symplectic structure turn the connected components of the intersections $S \cap L$ into a smooth family of symplectic leaves on S, when L ranges over the symplectic leaves of M.*

In both cases, the symplectic leaves induced by the Poisson bivector π_S on S are the connected components of the intersections $S \cap L$, where L ranges over all symplectic leaves of M. Finally, for coisotropic submanifolds, one has the following statement:

3. *assuming that S has clean intersection with all the symplectic leaves of M, S is a coisotropic submanifold if and only if for each symplectic leaf L, the intersection $S \cap L$ is a coisotropic submanifold of L.*

Proof. For Poisson submanifolds, the proof can be found in Proposition 2.12 in [Laurent-Gengoux et al., 2013] or in Proposition 3.26 of these lectures notes. For Poisson-Dirac submanifolds, the proof can be found in Proposition 6 of [Crainic and Fernandes, 2004], and for coisotropic submanifolds it can be found in Proposition 3.29 of the same lectures notes. □

Remark 4.96. The latter statement is more stringent because in general coisotropic submanifolds are far from havin clean intersections with symplectic leaves. See Remark 1. of [Zambon, 2011].

Remark 4.97. There exist plenty of other kinds of submanifolds in Poisson geometry, e.g. *isotropic submanifolds* are those submanifolds S such that $TS \subset \pi^\sharp(TS^\circ)$, *pre-Poisson submanifolds* are those submanifolds S such that the vector bundle $TS + \pi^\sharp(TS^\circ)$ has constant rank, etc.

[15] A *clean intersection* of two submanifolds S and L means that $S \cap L$ is a submanifold satisfying the following condition: $T(S \cap L) = TS \cap TL$. It implies, by the implicit function theorem, that for every $x \in S \cap L$, there exists open neighborhoods $U \subset S$ and $V \subset L$ such that $U \cap V$ is an open neighborhood of x in $S \cap L$. See e.g. this page or the proof of Proposition 5.26 in [Laurent-Gengoux et al., 2013].

4.4 Poisson-sigma model

In Physics, a *sigma model* is a way of encoding an action functional from a smooth map sometimes denoted $\sigma : \Sigma \longrightarrow M$, where Σ and M are smooth manifolds called respectively the *source* and *target* manifolds. Their dimension and the possibly additional structures (such as a pseudo-Riemannian metric or a Poisson structure on M) that these manifolds possess characterize the so-called sigma model. Sigma models are useful for the following reason: the dynamical fields of the physical theory correspond to the composite functions $\sigma^i = x^i \circ \sigma$ on the *target space*. For example the relativistic particle can be seen as a sigma model $X : \mathbb{R} \longrightarrow \mathbb{M}^4$ (where $\mathbb{M}^4$ is Minkowski space), given by the action:

$$S \propto \int_{\mathbb{R}} \eta_{ij}(X(\tau))\dot{X}^i(\tau)\dot{X}^j(\tau)d\tau$$

The trajectory of the particle in space time is parametrized by the proper time τ and is called the *world-line* of the particle. Notice that integration is made over the manifold Σ and not over M, because the independent variables are the coordinates over Σ.

Another example is the Nambu-Goto action for the bosonic relativistic open string is obtained from a sigma model $X : \Sigma \longrightarrow M$, where Σ is a 2-dimensional smooth manifold (with boundaries) called a *world-sheet*, parametrized by a timelike coordinate τ and a spacelike coordinate σ, and M is a pseudo-Riemannian manifold representing spacetime. Then the Nambu-Goto action is:

$$S_{NG} \propto \int_{\Sigma} \sqrt{(g_{\mu\nu}(X)\dot{X}^\mu X'^\nu)^2 - \dot{X}^\mu \dot{X}_\mu X'^\nu X'_\nu}d\tau d\sigma$$

where $\dot{X} = \frac{\partial X}{\partial \tau}$ and $X' = \frac{\partial X}{\partial \sigma}$, and where $g_{\mu\nu}$ is the metric on M.

A gauge theory on a pseudo-Riemannian oriented manifold M may be seen a a particular kind of sigma model: it is characterized by a set of *gauge fields* corresponding to the component of a Lie-algebra valued one-form $A = A_\mu\, dx^\mu = A_\mu^a T_a \otimes dx^\mu \in \Omega^1(M, \mathfrak{g})$, where the T_a form a basis of $\mathfrak{g}$. The Yang-Mills action is then written as:

$$S_{YM} = \frac{1}{2\alpha}\int_M \mathrm{tr}(F \wedge \star F) \tag{4.53}$$

where the F is the field strength associated to A: $F^a = dA^a + \frac{1}{2}[A, A]^a$. Usually, $\mathfrak{g}$ is a semi-simple matrix Lie algebra so that the trace is the usual trace on matrices, however in the more general case, one should think of the trace as symbolizing the Killing form κ on $\mathfrak{g}$[16]. The Lie bracket is that of $\mathfrak{g}$, while the differential form component of A is wedged. More precisely:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + [A_\mu, A_\nu]^a \tag{4.54}$$

Moreover, the notation $F \wedge \star F$ means that the wedge acts with respect to the forms, whereas the Lie algebra components of F is composed with that of $\star F$ (via the adjoint action, say). Much more details can be found in Chapter 3 of Part 2 of [Baez and Muniain, 1994]. It can be seen as a sigma model via the observation that the gauge field is a Lie algebroid morphism $A : TM \longrightarrow \mathfrak{g}$. The source manifold is thus TM while the target manifold is $\mathfrak{g}$.

Exercise 4.98. Show that, decomposing $F = dA + \frac{1}{2}[A, A]$ as $\frac{1}{2}F^a_{\mu\nu}T_a \otimes dx^\mu \wedge dx^\nu$, one indeed finds Equation (4.54).

[16] For finite dimensional semi-simple matrix algebras such as $\mathfrak{sl}_n, \mathfrak{so}_n, \mathfrak{su}_n, \mathfrak{sp}_{2n}$, the Killing form $\kappa(u, v)$ – for any two elements $u, v \in \mathfrak{g}$ – is proportional to $\mathrm{tr}(u \circ v)$, where u, v are in the latter case seen as endomorphisms of $\mathbb{R}^n$ (or $\mathbb{R}^{2n}$ for the symplectic algebra).

The Yang-Mills action can be rewritten by introducing a $n-2$ differential form $X = X^a_\mu T_a \otimes dx^\mu$ taking values in $\mathfrak{g}$:

$$S_{YM} = \int_M \mathrm{tr}\big(X \wedge F + \frac{\alpha}{2} X \wedge \star X\big) \tag{4.55}$$

Then the Euler-Lagrange equation on X is $X = \frac{1}{\alpha} \star F$ (at least when M is a Lorentzian four-dimensional manifold, see Equation (1.28)), so that we retrieve the original action (4.53), upon replacing X by its value. Something particular occur when the manifold M is two dimensional (we will call it Σ), because in that case X is a function and $\star X = X\omega$ where $\omega = \sqrt{|g|}dx^1 \wedge dx^2$ is the normalized volume form on M, as defined in Equation (3.15). In the case where $\mathfrak{g}$ is a finite dimensional semi-simple matrix Lie algebra, the 2-dimensional Yang-Mills action (4.55) becomes (up to some scalar factor):

$$\int_\Sigma \kappa_{ab} X^a (dA^b + \frac{1}{2}[A, A]^b) + \frac{\alpha}{2} \kappa_{ab} X^a X^b \,\omega$$

where $\kappa_{ab} = \mathrm{tr}(\mathrm{ad}_{T_a} \circ \mathrm{ad}_{T_b})$ are the components of the Killing form on $\mathfrak{g}$. Since in this nice situation, the Killing form is a non-degenerate bilinear form on $\mathfrak{g}$, from now on we will use contracted indices instead. Upon integrating by part the term $X^a dA^a$ (assuming, e.g., that the source manifold Σ has no boundary), we obtain:

$$\int_\Sigma A^a \wedge dX_a + \frac{1}{2} X_a [A, A]^a + \frac{\alpha}{2} X_a X^a \,\omega \tag{4.56}$$

Now, observe that $\mathfrak{g}$ is the linear dual of the Poisson vector space $\mathfrak{g}^*$ (see Example 4.4 for more details on linear Poisson structures). In other words, the smooth function X and the differential one-form A take values in $\mathfrak{g}^{**} \simeq \mathfrak{g}$ (because $\mathfrak{g}$ is finite dimensional). This is true for any dimension of the source manifold Σ, but what is characteristic of the 2-dimensional case is that the expression $X_a X^a$ in the last term ressembles the quadratic Casimir element of semi-simple Lie algebras (which usually form the kind of Lie algebras used in gauge theories). More precisely, for X^a a smooth function on M, the element $\sum_{a=1}^{\dim(\mathfrak{g})} X^a X^a T_a \odot T_a$ of the symmetric algebra of $\mathfrak{g}$ can be seen as a polynomial function on $\mathfrak{g}^*$, which actually turns out to be a Casimir element in the sense of Poisson algebras. Thus, we have shown that the 2-dimensional Yang-Mills theory can be reformulated in terms of a sigma model involving a linear Poisson structure (that of $\mathfrak{g}^*$). A natural generalization is then to weaken that condition and allow this theory to be defined on any Poisson manifold:

Definition 4.99. *The* Poisson-sigma model *is a sigma model defined by the following data:*

1. *the source Σ is a 2-dimensional oriented smooth manifold (possibly with boundary);*
2. *the target M is a finite dimensional Poisson manifold, with Poisson bivector π;*
3. *the maps defining the model is a Lie algebroid morphism $(X, A) : T\Sigma \longrightarrow T^*M$;*

and by the following action functional:

$$S_{PSM}(X, A) = \int_\Sigma \langle A, X_* \rangle + \frac{1}{2} \langle A \wedge A, X^! \pi \rangle \tag{4.57}$$

where $\langle\, .\, ,\, .\, \rangle$ denotes the pairing between TM and T^*M, and where $C \in \mathcal{C}^\infty(M)$ is any Casimir function of π.

Let us explain each term in details. The pushforward $X_*: T\Sigma \longrightarrow X^!TM$ can be seen as a one form on Σ taking values in $\Gamma(X^!TM)$. In local coordinates, it can be written as $X_* = dX^i\frac{\partial}{\partial x^i}$ where d is the de Rham differential on Σ and where the x^i are coordinates on M. Then, since A takes values in $\Gamma(X^!T^*M)$, the pairing in the first term is indeed well defined, so that it becomes: $\langle A, X_*\rangle = A_i \wedge dX^i$. In the second term, the notation $X^!\pi$ symbolizes that we evaluate the Poisson bivector π on the image of X in M. In other words, $X^!\pi$ is a section of the pullback vector bundle $X^!\bigwedge^2 TM$. This is justified by the fact that the differential 2-form $A\wedge A$ takes values in $\Gamma(X^!\bigwedge^2 T^*M)$. Then the second term becomes in coordinates: $\frac{1}{2}\langle A\wedge A, X^!\pi\rangle = \frac{1}{2}\pi^{ij}(X)A_i\wedge A_j$ (because π contains $\frac{1}{2}\pi^{ij}$). Then, Equation (4.57), the action functional of the PSM, can be rewritten as:

$$S_{PSM}(X,A) = \int_\Sigma A_i\wedge dX^i + \frac{1}{2}\pi^{ij}(X)A_i\wedge A_j$$

It is usual to add an additional term in the Poisson-sigma model that plays the same role as $\frac{\alpha}{2}X_aX^a\,\omega$ in 2-dimensional Yang-Mills theory. Any choice of Casimir function $C\in\mathcal{C}^\infty(M)$ (relatively to the Poisson bivector π) can be added to the action functional, which then becomes:

$$S_{PSM}(X,A) = \int_\Sigma A_i\wedge dX^i + \frac{1}{2}\pi^{ij}(X)A_i\wedge A_j + \star(C(X))$$

As for the other terms, the Casimir function is evaluated on $\mathrm{Im}(X)\subset M$. The constant $\frac{\alpha}{2}$ that was appearing in Yang-Mills action functional is not apparent in the above formula because it can be absorbed in the Casimir C. Obviously, if $M=\mathfrak{g}^*$ (where $\mathfrak{g}$ is a finite dimensional semi-simple matrix Lie algebra, say), and if the Casimir function is the quadratic Casimir element of $\mathfrak{g}$, then the Poisson-sigma model with Casimir corresponds to the 2-dimensional Yang-Mills action functional (4.56), under the following considerations: 1. the map $X:\Sigma\longrightarrow\mathfrak{g}^*$ is considered to take values in $\mathfrak{g}$ by using the non-degenerate Killing form on $\mathfrak{g}$ which allows to identity $\mathfrak{g}$ and $\mathfrak{g}^*$; 2. the fiber of $T^*\mathfrak{g}^*$ is identified with $\mathfrak{g}$ so that the differential 1-form $A: T\Sigma\longrightarrow T^*\mathfrak{g}^*$ is actually seen as taking values in $\mathfrak{g}$. This can be made explicit by realizing that A is actually a vector bundle morphism $T\Sigma\longrightarrow X^!T^*M$ covering the identity map on Σ, then, evaluating the differential one-form A on a tangent vector at a point x gives an element of the fiber of $T^*_{X(x)}M$, i.e. an element of $\mathfrak{g}$, as required.

If $C(X)=0$ then the Poisson-sigma model becomes a topological field theory, called a *BF-theory*. These are characterized by the following action functional:

$$S_{BF} = \int_\Sigma \mathrm{tr}(B\wedge F)$$

where Σ is a n-dimensional oriented manifold, F is the field strength associated to the gauge potential A (taking values in some Lie algebra, say), while B is a $\mathfrak{g}$ valued differential $n-2$-form. The Euler-Lagrange equations of such topological field theories are:

$$F=0 \qquad\text{and}\qquad d_AB=0$$

where d_A is the covariant derivative associated to the connection A. The solutions of the equations are purely topological: B is a closed 2-form, while the field strength of A vanishes so A does not propagate. Under appropriate assumptions (e.g. Σ is compact without boundary), the Poisson-sigma model is a 2-dimensional BF-theory, since Equation (4.57) can be rewritten as:

$$S_{PSM} = \int_\Sigma X^i\wedge F_i$$

where summation on contracted indices is implicit. Thus, the Poisson-sigma model is a topological field theory that, under the addition of a Casimir function, can encode some physical model such as 2-dimensional Yang-Mills gauge theory.

Exercise 4.100. Check that the action functional of the Poisson-sigma model is invariant under the following gauge transformations:

$$\delta_{(\epsilon,\lambda)}X^i = \lambda_j \pi^{ij} \qquad \text{and} \qquad \delta_{(\epsilon,\lambda)}A_i = d\lambda_i + \frac{\partial \pi^{kl}}{\partial x^i}A_k\epsilon_l$$

where $\epsilon_1, \dots, \epsilon_n$ are smooth functions on Σ and $\lambda = \lambda_i dx^i$ is a differential 1-form on Σ taking values in T^*M. They are obviously nonlinear generalizations of standard gauge transformations.

Another application of the Poisson-sigma model (and actually its original motivation) is to describe 2-dimensional gravity (without matter field). Let Σ be an oriented, 2-dimensional Lorentzian manifold, with metric g (of signature $(1,1)$ then). Let us denote by x^0 and x^1 local coordinates on Σ. Recall that in two dimensions, the symmetries of the Riemann tensor impose that:

$$R_{\mu\nu\alpha\beta} = -\frac{R}{2}\epsilon_{\mu\nu}\epsilon_{\alpha\beta} \tag{4.58}$$

where R is some scalar identified with the Ricci scalar, and $\epsilon_{\mu\nu}$ and $\epsilon_{\alpha\beta}$ are antisymmetric Levi-Civita tensors on two indices, i.e. $\epsilon_{01} = \sqrt{|g|}$ and $\epsilon_{10} = -\sqrt{|g|}$. Due to Equation (4.58) and to the identity $\epsilon_{\mu\nu}\epsilon_{\alpha_\beta} = g_{\mu\beta}g_{\nu\alpha} - g_{\mu\alpha}g_{\nu\beta}$, the Einstein tensor:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu}$$

identically vanish on Σ. This is problematic since the vacuum Einstein field equation is $G_{\mu\nu} = 0$. The fact that it is automatically satisfied in 2-dimensional gravity shows that 2-dimensional gravity without matter does not yield propagating gravitational modes. That is why physicists usually allow the Einstein-Hilbert Lagrangian to take more intricate forms in 2-dimensions. One of particular importance is a the $f(R)$*-gravity*, in which the Ricci scalar is replaced by a well-behaved function:

$$\int_\Sigma \star(f(R)) = -\int_\Sigma \sqrt{|g|}f(R)dx^0dx^1$$

Then one may show that under rather common assumptions, this action can be rewritten in terms of an auxiliary field ϕ called the *dilaton* and another well-behaved function $V(\phi)$:

$$\int_\Sigma \sqrt{|g|}(\phi R - V(\phi))dx^0dx^1 \tag{4.59}$$

See for example Section 7 of [Schmidt, 1999] for an explicit treatment of this replacement.

Exercise 4.101. Show that, for $f(R) = R^2$, we have the usual Gaussian integral:

$$\frac{1}{2}\int_\Sigma \sqrt{|g|}R^2dx^0dx^1 = \int_\Sigma \sqrt{|g|}(\phi R - \frac{1}{2}\phi^2)dx^0dx^1$$

Let us now rewrite the $f(R)$-lagrangian using zweibein and a spin connection, á la Palatini (see Chapter 3 Part III of [Baez and Muniain, 1994] for a treatment of Palatini formalism in n dimensions). The idea (in two dimensions) is the following: the metric g is locally diagonalizable, and even better, by a diagonal matrix of the form:

$$g \sim \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Then, in the neighborhood of every point, there exist two locally defined vector fields e_+ and e_- defining a frame of $T\Sigma$, such that in the local coordinate defined by this frame g takes the above diagonal form. We call the pair (e_+, e_-) a *zweibein* – the 2-dimensional analogs of tetrads in 4-dimensions and of vielbein in n dimensions. In particular, noting e^a for the differential one-form on Σ dual to e_a, where $a = \pm$, we have $e^+ \wedge e^- = \sqrt{|g|}dx^0dx^1$. While the Einstein-Hilbert Lagrangian in n dimensions is invariant under diffeomorphisms, its reformulation in terms of vielbein is only invariant under the gauge group $SO(n-1,1)$ (encoding every possible Lorentz rotations of the orthonormal frame). In two dimensions, this group is one dimensional, hence abelian. The gauge invariance under the Lorentz group is encoded by a connection ω called the *spin connection*. It is a differential 1-form on Σ taking values in $\mathfrak{so}(1,1)$ (or $\mathfrak{so}(1,1)$ when working on a n-dimensional space-time), satisfying the following compatibility condition:

$$De^a \equiv de^a + \omega^a_b e^b = 0 \tag{4.60}$$

This condition implies that ω is uniquely expressed in terms of the zwiebein and it dual. Since the gauge group $SO(1,1)$ is abelian, the curvature of the spin connection reduces to $d\omega$, so that we have:

$$R\sqrt{|g|}dx^0dx^1 = -2d\omega$$

To implement the constraint (4.60) in the $f(R)$-lagrangian, one has to introduce two Lagrange multiplicators X_+, X_-, so that the action (4.59) can be rewritten as:

$$\int_\Sigma \phi\, d\omega + X_a De^a + \frac{1}{2}V(\phi)\, e^+ \wedge e^- = \int_\Sigma \underbrace{\omega \wedge d\phi + e^a \wedge dX_a}_{A_i\, dX^i} + \underbrace{X_a\, \omega^a_b e^b + \frac{1}{2}V(\phi)\, e^+ \wedge e^-}_{\frac{1}{2}\pi^{ij}(X)A_i \wedge A_j}$$

The expression on the right-hand side corresponds to a Poisson-sigma model, where the Poisson manifold M is $\mathbb{R}^3$, where the scalar function $X : \Sigma \longrightarrow M$ is the triplet (X^+, X^-, ϕ) and where the differential one-form $A \in \Omega^1(\Sigma, X^! T^* M)$ is the triplet (e^+, e^-, ω). This shows that the action functional of $f(R)$ 2-dimensional gravity (without matter) can be expressed as a Poisson-sigma model. There are additional applications of this model to other topological field theories.

5 A geometric perspective on Bergmann-Dirac formalism

From the 1930s physicists have tried to find a way of ‘quantizing’ existing classical physical theories in order to find out what would a quantum field theory look like [Rosenfeld, 1932]. Classical mechanics appeared as a limit of non-relativistic quantum mechanics, formulated in terms of a Hamiltonian and of position and momenta operators. Physicists then were hoping to develop the canonical formalism associated with Hamiltonian mechanics to relativistic field theories. In particular, a possible goal was to obtain quantum electrodynamics by quantizing Maxwell electromagnetism, and some quantum theory of gravity by quantizing general relativity. Unfortunately, both of those theories possess inner symmetries (gauge symmetries and coordinate invariance) which prevent to straightforwardly obtain the Hamiltonian from the Lagrangian, as is usually possible in classical mechanics. Indeed, it has been shown that if a Lagrangian is covariant under a set of symmetries – i.e. if its expression stays invariant – then the Legendre transform form the the Lagrangian to the Hamiltonian cannot be performed. On the contrary, one has to add several *constraints* in the hamiltonian picture to account for the non-invertibility of the Legendre transform. Existence of constraints characterize physical theories with internal symmetries such as gauge symmetries.

Although another alternative path was followed in the 1940s to provide the first central example of a quantum field theory – namely: quantum electrodynamics – Peter Bergmann and Paul Dirac proposed in the late 40s-early 50s an alternative approach to quantization [Bergmann, 1949, Dirac, 1950]. This canonical quantization procedure relied on obtaining first the Hamiltonian corresponding to the given Lagrangian characterizing the action principle, and then quantize the Hamiltonian as well as the various position and momenta operators, together with the several *constraints* emerging from the procedure. More generally, any smooth function f of the canonical coordinates should be sent to an operator via a quantization map $\mathcal{Q}$: $\mathcal{Q}(f) = F$, having natural properties such that $\mathcal{Q}(1) = \mathrm{Id}$ – the identity operator. In this latter step, Dirac requires that the Lie bracket of operators and the Poisson brackets of observables (smooth functions on the phase space) obey the following compatibility condition:

$$\mathcal{Q}_{\{f,g\}} = \frac{1}{i\hbar}[\mathcal{Q}_f, \mathcal{Q}_g] \tag{5.1}$$

Although the procedure seems perfectly viable on the paper, and that the first part of the procedure is well-known, there is actually no unique way of quantizing a classical theory. Indeed, one usually promotes the position q^k and conjugate momenta p_k coordinates to operators Q^k, P_k on a Hilbert space, and require that their Lie Bracket is proportional to $i\hbar$, but there may exist alternative choice of coordinates that would thus give other quantized operator. Moreover, when one has a product of conjugates coordinates – such as $qp = pq$, say – there is no standard way of assigning an operator because the operators associated to p_k and q^k do not commute. There exists a convention specified by Weyl, which comes close to achieve this task, but a no-go theorem by Groenewold proves that there is no quantization scheme such that Equation (5.1) is satisfied at any polynomial order. This is why the canonical quantization proposed by Dirac is then usually performed only for unambiguous classical theories for which the Hamiltonian has nice properties.

Another huge problem in Dirac’s quantization procedure is the treatment of constraints. As we will see, the quantization scheme $\mathcal{Q}$ obviously sends every constraint ϕ^a to an operator, but it does not say what convention one should impose on the action of $\Phi^a = \mathcal{Q}(\phi^a)$ on the vectors of the Hilbert space. Additional procedures have been developed to handle this problem which arise as soon as one wants to quantize a gauge theory: the *BV formalism* on the one hand (in the Lagrangian picture) and the *BFV/BRST formalism* (in the Hamiltonian picture)[17]. This

[17] BV stands for Batalin-Vilkovisky and BFV stands for Batalin-Fradkin-Vilkovisky.

section is devoted to study all those quantization procedures, and we will mostly rely on the following texts: [Gitman and Tyutin, 1990], [Henneaux and Teitelboim, 1994], and the incredibly pedagogical [Matschull, 1996] and [Rothe and Rothe, 2010]. Other useful resources [Bergmann, 1949,Bergmann and Goldberg, 1955,Dirac, 1964,Sudarshan and Mukunda, 2015] are of historical interest and may also be pedagogical on particular aspects of the topic. Eventually, there are two alternative quantization scheme provided by mathematicians: *geometric quantization* and *deformation quantization*. While the former tries to provide a Hilbert space and an quantization transformation $\mathcal{Q}$, the second tries to deform the algebra of observables – i.e. smooth functions on the phase space – so that we obtain a non-commutative associative algebra resembling the operator algebra physicists look for. None of them give a definite answer to quantization because they have their own, respective, issues. The problem of quantization is thus a very intricate one, and is still under investigation.

5.1 Lagrangian and Hamiltonian formalism from a geometric point of view

We begin the review of Dirac's canonical formalism with a non-relativistic physical model, to later turn to relativistic field theory. Let us start with a given *configuration space* represented by a n-dimensional oriented smooth manifold Q (possibly with boundary). In this section the points of Q are denoted q – instead of x. The local coordinate functions on Q are denoted by q^i – instead of x^i – and can express the position of several particles, the length of a spring, the charge of a capacitor etc. That is why they are called *generalized coordinates*. Let us now fix a trivializing chart U of both TQ and T^*Q, admitting local coordinates q^i on the base U. The tangent bundle TQ over U admits fiberwise coordinate functions $v^i : TQ \longmapsto \mathbb{R}$ that are a mere rewriting of the constant covector fields on Q denoted dq^i. In particular for every tangent vector $X \in T_qQ$, $v^i(X) = v^i(X^j \frac{\partial}{\partial q^j}) = X^i$. That is why we will often denote tangent vectors at q as $v \in T_qQ$, so that by abuse of notation, we would identify the components of v in the basis $\frac{\partial}{\partial q^i}$ with v^i.

The cotangent bundle T^*Q over U also admits fiberwise coordinate functions denoted p_i and defined as expected: $p_i(\xi) = p_i(\xi_j dq^j) = \xi_i$ for any covector field ξ. For this reason, the coordinates p_i can be identified to the locally defined constant vector fields $\frac{\partial}{\partial q^i}$. In particular we set $p_i(v^j) = \delta_i^j$ so that the p_i are the dual coordinates to the v^i, explaining why the former are called *conjugate momenta*. This also justifies that we call T^*Q the *phase space* – sometimes denoted P – since it contains the configurations as well as the momenta of the configuration space Q. We will often denote covector fields as the letter p, so that a point in the cotangent bundle T^*Q would be denoted (q, p). By abuse of notation, we identify the components of p (resp. v) in the basis dq^i (resp. $\frac{\partial}{\partial q^i}$) with p_i (resp. v^i). Since the tangent and cotangent bundles need not be trivial vector bundles, both v^i and p_i are only defined locally on Q. More precisely, the coordinates q^i are local coordinates on the trivializing neighborhood U of q, which in turn implies that the coordinate v^i and p^i are fiberwise linear coordinates globally defined on the fiber.

Example 5.1. The cotangent bundle represent the natural setup to do Hamiltonian mechanics. Let us illustrate this property by analyzing the pendulum (of mass m and length L) from a Poisson/symplectic geometry perspective. The physical system is parametrized by the angle θ so that we set the space of all possible angles – i.e. the *configuration space* – to be the circle S^1. The conjugate momentum to the generalized coordinate $q = \theta$ is denoted p so that it is interpreted as the fiberwise linear coordinate on the phase space T^*S^1. The symplectic structure on this cotangent bundle is the standard one, i.e. $\omega = dp \wedge dq$, where $q = \theta$. The corresponding non-degenerate Poisson structure on T^*S^1 is thus given by:

$$\{f,h\} = \frac{\partial f}{\partial q}\frac{\partial h}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial h}{\partial q}$$

for every two smooth functions $f,h \in \mathcal{C}^\infty(T^*S^1)$.

Let us define the following smooth function on T^*S^1:

$$H = \frac{p^2}{2mL} + mgL(1-\cos(\theta)) \tag{5.2}$$

where g is a constant positive parameter that may be fixed at 9,8 if one wants to reproduce the gravitational force equivalent. We call this function (5.2) the "Hamiltonian of the system" and compute its hamiltonian vector field $X_H \in \mathfrak{X}(T^*S^1)$:

$$X_H = \{H,-\} = -\frac{p}{mL}\frac{\partial}{\partial\theta} + mgL\sin(\theta)\frac{\partial}{\partial p} \tag{5.3}$$

An *integral curve* of the vector field $-X_H$ is a smooth path $\gamma : \mathbb{R} \longrightarrow T^*S^1, t \longmapsto (\theta(t), p(t))$ which is such that the tangent vector $\dot\gamma(t) = \dot\theta(t)\frac{\partial}{\partial\theta} + \dot p(t)\frac{\partial}{\partial p}$ at the point $\gamma(t) = (\theta(t), p(t))$ is equal to $-X_H|_{(\theta(t),p(t))}$. Alternatively, it corresponds to the level set of the smooth function H. By isolating the two components $\dot\theta(t)$ and $\dot p(t)$ forming $\dot\gamma(t)$ at the point $\gamma(t) = (\theta(t), p(t))$ and equating them to that of $-X_H$ at the same point, one has, for every t:

$$\dot\theta(t) = -X_H(\theta) = \{\theta, H\} = \frac{\partial H}{\partial p}$$
$$\dot p(t) = -X_H(p) = \{p, H\} = -\frac{\partial H}{\partial\theta}$$

Thus, the integral curves of the vector field $-X_H$ are precisely those path $\gamma : \mathbb{R} \longrightarrow T^*S^1$ whose components $\theta(t)$ and $p(t)$ satisfy the Hamilton equations of motion. This implies that such integral curves are the physical solutions of the Hamilton equations which means that, starting from a point (q_0, p_0) on the phase space T^*S^1, the physical motion of the pendulum obliges to follow the integral curve of the vector field $-X_H$ passing through (q_0, p_0). More abstractly, we say that $-X_H$ points towards the flow of physical time[18]. Drawing such integral curves using the expression (5.3) gives the well-known phase portrait, Fig. 19.

We have thus seen in Example 5.1 that the mathematics developed in Poisson geometry is well-adapted to describe physical systems in the Hamiltonian formalism. However, this was only possible because every point of the phase space could be used as an initial condition. Sometimes in physics, it may happen that not every point of the phase space can be chosen to be a set of initial conditions. In that case one cannot straightforwardly apply Hamiltonian formalism to the model, and a more refined formalism is required: *constrained Hamiltonian formalism.* We will spend the rest of this section on this topic.

Definition 5.2. *A* Lagrangian *is a fiberwise convex smooth function $L \in \mathcal{C}^\infty(TQ)$ on the tangent bundle of Q. By* fiberwise convex, *we mean that, for every $q \in Q$, the function $L(q,-) : T_qQ \longrightarrow \mathbb{R}$ is a smooth convex function, i.e. it is such that its* Hessian *symmetric matrix (written in local coordinates):*

$$\mathscr{H}_{ij}(q,v) = \frac{\partial^2 L(q,v)}{\partial v^i \partial v^j} \tag{5.4}$$

has non negative determinant for every $v \in T_qQ$.

[18] The minus sign comes down to the choice of defining the hamiltonian vector field of a smooth function f as $\{f,-\}$ and not as $\{-,f\}$, although the latter convention exists.

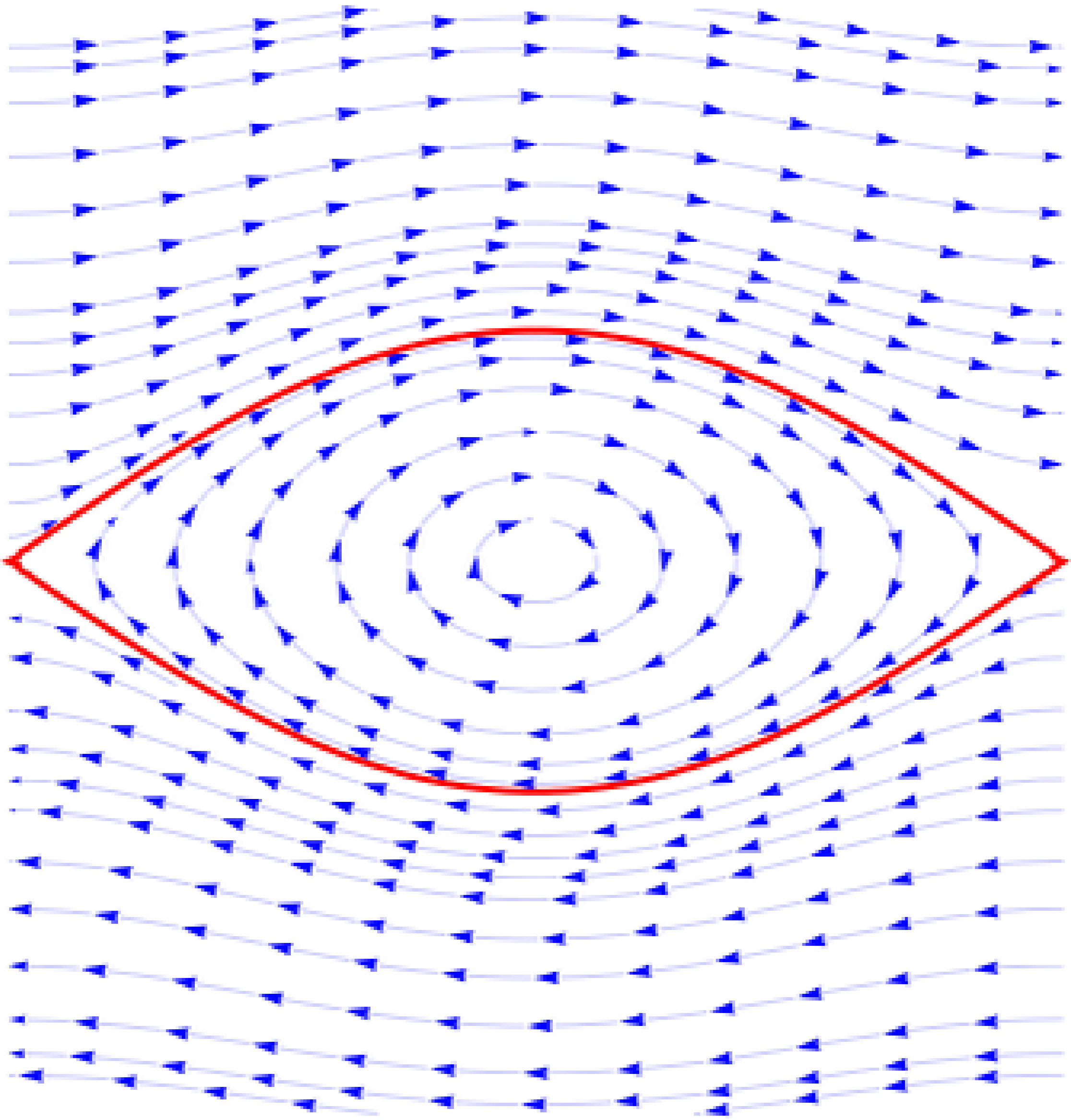

Figure 19: Phase portrait of the pendulum build from a purely symplectic/Poisson geometry perspective. The horizontal axis represents the angular coordinate $q = \theta$ between $-\pi$ and π, while the vertical axis represents the conjugate momenta p. The arrow heads represent the direction of $-X_H$ (flow of physical time) and the lines its corresponding integral curves. The *separatrix* is actually made of four submanifolds: 2 points (singular leaves) at $\theta = \pm\pi$ and $p = 0$, while the upper (resp. lower) red line is directed toward the right (resp. the left) but never reaches π (resp. $-\pi$). There is an additional singular leaf at $(0, 0)$. Hence this phase portrait is indeed a singular foliation, integrating the distribution generated by $-X_H$. Picture taken from Wolfram Alpha.

Recall that, here, we consider that $v \in T_qQ$ and we identify the coordinate functions $v^i : T_qQ \longrightarrow \mathbb{R}$ with the components of v in the basis $\frac{\partial}{\partial q^i}$. To any smooth path $\gamma : \mathbb{R} \longrightarrow M$, one can associate a tangent vector at the point $\gamma(t)$, which we denote $\dot{\gamma}(t) \in T_{\gamma(t)}M$ (see subsection 2.1). One can then evaluate the Lagrangian function along this path: $t \longmapsto L(\gamma(t), \dot{\gamma}(t))$. A priori, one can always pick up any kind of path on Q, but physicists have a recipe to determine which kind of path would correspond to the time evolution of the physical system whose state is encoded by the generalized coordinates q. Indeed, such a path γ should satisfy some differential equations called the *Euler-Lagrange equations*, under appropriate boundary conditions. Any other choice of path would be considered as non-physical. They proceed as follows: the (non-relativistic) physical model is characterized by a so-called *action*, which depends exclusively on the choice of path γ:

$$S(\gamma) = \int_{\mathbb{R}} L(\gamma(t), \dot{\gamma}(t))\, dt$$

Often the path admits well-defined boundary conditions so the integral converges. To stick with physicists' notation, we will now write the time dependency of the Lagrangian with respect to the chosen path as $L(q, \dot{q})$ instead of $L(\gamma(t), \dot{\gamma}(t))$, where $\dot{q}$ denotes the time derivative of the generalized coordinate $q = \gamma(t)$ at time t, which geometrically corresponds to the vector $\dot{\gamma}(t)$ tangent to the curve γ at time t.

Assuming that smooth path γ corresponding to physical evolution are extrema of the action – i.e. stationary points, one requires that an infinitesimal variation of the action with respect to an infinitesimal change of path would vanish if the original path is a physical path. More precisely, assume that γ_0 is a smooth path in Q corresponding to a physical evolution of the system, then $S(\gamma_0)$ should be an extremum of the function S, and thus the infinitesimal variations of S around γ_0 should be zero:

$$0 = \delta S = \int_{\mathbb{R}} \delta L\, dt$$

where the variation should be understood to be taken at γ_0 (stationary point of the action). Computing the variation of L with respect to infinitesimal change of path – i.e. with respect to coordinates q and v – and with respect to the boundary conditions gives the following identity :

$$\delta S = -\sum_{i=1}^{n} \int_{\mathbb{R}} E_i(q, \dot{q}, \ddot{q}) \delta q^i\, dt$$

where the $E_i(q, \dot{q}, \ddot{q})$ are defined as:

$$E_i(q, \dot{q}, \ddot{q}) = \frac{d}{dt}\frac{\partial L(q, \dot{q})}{\partial v^i} - \frac{\partial L(q, \dot{q})}{\partial q^i} \tag{5.5}$$

for every $1 \leq i \leq n$. Hence, a smooth path γ_0 corresponding to a physical evolution of the system (given appropriate initial state and boundary conditions), being a stationary point of the action, should make Equation (5.5) vanish when $(q, \dot{q}) = (\gamma_0(t), \dot{\gamma_0}(t))$. In other words, a path corresponding to a physical evolution of the system should necessarily satisfy the infamous *Euler-Lagrange equations*:

$$\frac{d}{dt}\frac{\partial L}{\partial v^i} - \frac{\partial L}{\partial q^i} = 0 \tag{5.6}$$

for every $1 \leq i \leq n$. Conversely, we will consider that solutions of these equations – i.e. smooth paths $\gamma : \mathbb{R} \longrightarrow M$ such that $(\gamma(t), \dot{\gamma}(t))$ are solutions of the Euler-Lagrange equations – are precisely the paths characterizing physical evolution of the system.

Now, since we assume that the Lagrangian does not have explicit time dependence, expanding the time derivative in the Euler-Lagrange equations (5.6) gives the following:

$$\mathscr{H}_{ij}(q, \dot{q})\, \ddot{q}^j = \frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial v^i \partial q^j}\dot{q}^j \tag{5.7}$$

One then sees that the accelerations are uniquely solvable in terms of the positions and the velocities only if the Hessian matrix $\mathscr{H}_{ij}(q, \dot{q})$ is invertible, i.e. if $\det(\mathscr{H}_{ij}(q, \dot{q})) \neq 0$. When this is the case, i.e when one can write the Equations (5.7) as $\ddot{q}^i =$ something depending only on q and $\dot{q}$ then the theory of ordinary differential equations says that, given a set of initial conditions, there is a unique solution of this Cauchy problem (at least) in a small neighborhood of these initial conditions. In other words, it means that the time evolution of the physical system – the evolution of the couple $(q, \dot{q})$ – is guaranteed to depend *only* on the initial conditions. However, when the Hessian matrix has vanishing determinant, the left hand side of Equation (5.7) vanishes so that the Cauchy problem does not admit a unique solution. Such a situation happens when the Lagrangian admits local symmetries.

Definition 5.3. *A (local) symmetry of the physical system is a (local) diffeomorphism of the configuration space Q – i.e. it is a transformation of the generalized coordinates – such that the form of the Lagrangian is left unchanged (up to a total derivative), so that the action is invariant.*

It has indeed been shown in the late 1940s that a Lagrangian admitting local symmetries has a vanishing Hessian (see Appendix A of [Rothe and Rothe, 2010] which is a modern reformulation of [Bergmann, 1949]). In that case, as is explained in Chapter 2 of [Rothe and Rothe, 2010], one has to dig into the constraints that the Lagrangian imposes on the system by carefully studying the null eigenvectors of the Hessian matrix. This opens the treatment of the quantization of gauge theories via the Batalin-Vilkovisky formalism. Notice however that in Dirac's canonical quantization procedure, one quantize the theory from the Hamiltonian perspective because in quantum mechanics the Hamiltonian has a central role. Let us give a bit more details on how hamiltonian mechanics enter the picture.

Definition 5.4. *Let $L : TQ \longrightarrow \mathbb{R}$ be a Lagrangian (assumed to be a convex function) and define the* canonical hamiltonian *to be the following function on the extended tangent bundle $\mathbb{T}Q = TQ \oplus T^*Q$:*

$$H_c(q, v, p) = \langle p, v\rangle_q - L(q, v) \tag{5.8}$$

where $\langle p, v\rangle_q$ denotes the pairing between T^*_qM and T_qM.

This function is called the *canonical hamiltonian* because it corresponds to the usual definition of the hamiltonian for unconstrained systems. Let U be a trivializing chart of both TQ and T^*Q and let q^i, v^i and p_i the corresponding local coordinates on the base, and on the fibers of $TQ|_U$ and $T^*Q|_U$, respectively. Since $\langle p, v\rangle_q = \sum_{i=1}^n p_i v^i$, by differentiating the canonical hamiltonian with respect to p_i one obtains:

$$v_i = \frac{\partial H_c}{\partial p^i}$$

Let us compute the derivative of H_c with respect to v^i:

$$\frac{\partial H_c}{\partial v^i} = p_i - \frac{\partial L}{\partial v^i} \tag{5.9}$$

Then, the points of $\mathbb{T}_qQ$ for which $\frac{\partial H}{\partial v^i} = 0$ are those such that:

$$p_i = \frac{\partial L}{\partial v^i} \tag{5.10}$$

Now, notice that the Euler-Lagrange equations (5.6) are second-order differential equations, which can be reformulated as two sets of first-order differential equations:

$$v^i = \dot{q}^i \qquad \text{and} \qquad \mathscr{H}_{ij}(q, v)\,\dot{v}^j = \frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial v^i \partial q^j} v^j$$

where we have assumed that some path $\gamma : \mathbb{R} \longrightarrow TQ$ defines a solution, so that $(q(t), \dot{q}(t)) = (\gamma(t), \dot{\gamma}(t))$. These equations are equivalent to the following set of equations, called the *implicit Euler-Lagrange equations*:

$$v^i = \dot{q}^i, \qquad p_i = \frac{\partial L}{\partial v^i} \qquad \text{and} \qquad \dot{p}_i = \frac{\partial L}{\partial q^i} \tag{5.11}$$

The equivalence can indeed be straightforwardly calculated, and the latter equations can be obtained as the variation of the following action, where the p_i have the role of Lagrange multipliers in what is called the *Hamilton-Pontryagin action*:

$$S = \int \left(L(q,v) + p_i(\dot{q}^i - v^i)\right) dt \tag{5.12}$$

The set of Equations (5.11) can then be recasted using the canonical Hamiltonian:

$$\dot{q}^i = \frac{\partial H_c}{\partial p_i}, \qquad \frac{\partial H_c}{\partial v^i} = 0 \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial H_c}{\partial q^i} \tag{5.13}$$

These equations descend from the variation of the following action:

$$S = \int \left(p_i \dot{q}^i - H_c(q,v,p)\right) dt$$

which is actually a rewriting of Equation (5.12). Then, we see how Hamiltonian can be a very efficient way of recasting Euler-Lagrange equations (5.6) into first-order differential equations.

In classical mechanics, the Hamiltonian is the *Legendre transform* of the Lagrangian. Usually the Legendre transform of a convex function $x \longmapsto f(x)$ – with domain of definition I – is a smooth function $p \longmapsto f^*(p)$ defined via evaluating the supremum of the concave function $x \mapsto px - f(x)$ over I, for each p such that this supremum is finite. Denoting I^* the subset of $\mathbb{R}$ whose elements $p \in I^*$ are such that $\sup\limits_{x\in I}\big(px - f(x)\big) < +\infty$, one sets:

$$f^*(p) = \sup_{x\in I}\big(px - f(x)\big) \tag{5.14}$$

Under the assumption that the derivative of f is invertible there is an explicit formula for f^*:

$$f^*(p) = px - f(x)\Big|_{x=(f')^{-1}(p)} \tag{5.15}$$

where here one really should understand x and p as real numbers so it makes sense to have $(f')^{-1}(p)$. Equation (5.15) is the kind of formula one usually uses in thermodynamics, where Helmholtz free energy A and Gibbs free energy G are obtained by performing Legendre transforms (up to a sign) of the internal energy U and enthalpy H, respectively. There, we usually do not explicitly check that the derivative of U and H with respect to the entropy is invertible although it is implicitly used when we do the Legendre transform using Formula (5.15) instead of Formula (5.14).

In our context, we precisely chose the Lagrangian to be convex so that we can take its Legendre transform. We will slightly extend the meaning of the latter by considering that it is a map from the tangent bundle to the cotangent bundle, thus providing an explanation for the formula $p_i = \frac{\partial L}{\partial v^i}$. The Lagrangian is supposed to be a convex function, i.e. its Hessian $\mathscr{H}$ has non-negative determinant. The Legendre transform is then performed with respect to the coordinates v^i. In geometric terms, the Legendre transform between the Lagrangian and the

Hamiltonian corresponds to performing a Legendre transform of the function $L(q,-) \in \mathcal{C}^\infty(T_qQ)$ for every q. In other words, at a fixed point q, the function $H_0(q,p)$ is defined as:

$$H_0(q,p) = \sup_{v \in T_qM} \big(H_c(q,v,p)\big) \tag{5.16}$$

when such supremum exists. The Hamiltonian is a smooth function on (a subset of) T^*Q, hence it depends only on the generalized coordinates q^i and on the conjugate momenta p_i. As for the rest of the section, we will use the Legendre transform from a more geometrical point of view. We will adopt an 'in-between' perspective where we mostly work in local coordinates over a trivializing chart $U \subset M$ to treat hamiltonian constraints (as physicists do), and at the same time we will adopt from time to time a global coordinate-free perspective to address issues that will inevitably arise along the way (as mathematicians do). We will mostly rely on the following resources: on the mathematical side, the Legendre transform had been investigated by Tulczyjew [Tulczyjew, 1977] and constrained hamiltonians by Marsden and Yoshimura [Yoshimura and Marsden, 2007] (see also most of references therein), while on the physical side there exist well established sources ont constrained hamiltonian systems [Gitman and Tyutin, 1990], [Henneaux and Teitelboim, 1994], [Rothe and Rothe, 2010], see also these notes.

To provide a geometric flavour to this discussion, let us then generalize the Legendre transform to the tangent and cotangent bundles:

Definition 5.5. *The* Legendre transform *is a base point preserving smooth map from TQ to T^*Q (but not necessarily a vector bundle morphism) given by:*

$$\begin{aligned} \mathscr{L}:\ TQ &\longrightarrow T^*Q \\ (q,v) &\longmapsto \left(q, p : w \mapsto \frac{d}{ds}\Big|_{s=0} L(q, v+sw)\right) \end{aligned}$$

On the right-hand side, the element w is a tangent vector at q. Thus, the element p – image of v via $\mathscr{L}$ – is a linear form on T_qQ, sending w to $\frac{d}{ds}\big|_{s=0} L(q,v+sw)$. This definition does not depend on the local coordinates, but the function $\mathscr{L}$ can be decomposed on the local frame dq^i as:

$$\mathscr{L}(q,v) = \sum_{i=1}^n \mathscr{L}_i(q,v)dq^i = \sum_{i=1}^n \frac{\partial L}{\partial v^i} dq^i$$

so that $\mathscr{L}_i = \frac{\partial L}{\partial v^i} \in \mathcal{C}^\infty(TQ)$ symbolize the components of the function and we indeed obtain again that $p_i\big(\mathscr{L}(q,v)\big) = \frac{\partial L}{\partial v^i}(q,v)$. As a base point preserving smooth map from TQ to T^*Q, the Legendre transform gives rise to a submanifold $N_{\mathscr{L}}$ of the total extended tangent bundle $\mathbb{T}Q = TQ \oplus T^*Q$, defined as:

$$N_{\mathscr{L}} = \big\{(q,v,p)\,|\,(q,p) = \mathscr{L}(q,v)\big\} \subset \mathbb{T}Q$$

This submanifold is the disjoint union over the points $q \in Q$ of the graphs of the smooth maps $\mathscr{L}(q,-) : TQ \longrightarrow T^*Q$, i.e. $N_{\mathscr{L}} \cap \mathbb{T}_qQ = \mathrm{Gr}\big(\mathscr{L}(q,-)\big)$. Seeing the Legendre transform from this geometrical viewpoint allows to retrieve the usual definition:

Lemma 5.6. *The submanifold $N_{\mathscr{L}}$ is the set of points $(q,v,p) \in \mathbb{T}Q$ such that v is a critical point of the smooth function $x \longmapsto \langle p,x\rangle_q - L(q,x)$.*

Proof. Let $(q,v,p) \in \mathbb{T}Q = TQ \oplus T^*Q$, then $p_i\big(\mathscr{L}(q,v)\big) = \frac{\partial L}{\partial v^i}(q,v)$ if and only if v satisfies $\frac{\partial(\langle p,v\rangle_q - L(q,v))}{\partial v^i} = 0$, i.e. if and only if v is a critical point of $x \longmapsto \langle p,x\rangle_q - L(q,x)$. $\square$

By Lemma 5.6, the restriction of the function H_c to $N_{\mathscr{L}}$ does not depend on v because for any given choice of pair (q, p), any critical point v of $x \mapsto \langle p, x\rangle_q - L(q, x)$ gives the same critical value (because the supremum is unique). So, in particular:

$$\frac{\partial H_c}{\partial v^i}\bigg|_{N_{\mathscr{L}}} = 0$$

This equation implies that the canonical hamiltonian induces a smooth function defined on the image of the Legendre transform $\mathrm{Im}(\mathscr{L}) \subset T^*Q$:

$$H_0(q, p) = H_c(q, v, p) \text{ for any triple } (q, v, p) \in N_{\mathscr{L}}$$

This latter equation can be summarized as:

$$H_0 = H_c\big|_{N_{\mathscr{L}}} \tag{5.17}$$

The notation is not innocent since Lemma 5.6 tells us that that H_0 is precisely the smooth function H_0 defined in Equation (5.16). The Hamiltonian H_0 is not defined on the entirety of the cotangent bundle, except if the function $\mathscr{L}$ is invertible. In that latter case:

$$H_0(q, p) = \langle p, \mathscr{L}^{-1}(q, p)\rangle_q - L\big(q, \mathscr{L}^{-1}(q, p)\big)$$

When it is not invertible, it is still possible to have an explicit expression for H_0 in terms of q and p but this requires to introduce local sections of the Legendre transform, see Equation (5.21). The condition for $\mathscr{L}$ to be invertible goes down to the non-vanishing of the determinant of its Jacobian matrix $\mathscr{J}(q, v) = (\frac{\partial \mathscr{L}_i}{\partial v^j})_{i,j}$. But this amounts to the non-vanishing of the determinant of the Hessian of the Lagrangian, for:

$$\frac{\partial \mathscr{L}_i}{\partial v^j} = \frac{\partial^2 L}{\partial v^i \partial v^j}$$

In other words, $\mathscr{J} = \mathscr{H}$ and, in light of the discussion following Equation (5.7), one concludes that *when the Lagrangian admits local symmetries, the Legendre transform is not invertible.*

Example 5.7. Let us use Example 1 of [Rothe and Rothe, 2010], p. 8. The configuration manifold is $Q = \mathbb{R}^2$, and the Lagrangian is:

$$L(q, v) = \frac{1}{2}v_x^2 + v_x y + \frac{1}{2}(x - y)^2$$

where x, y are the standard coordinates on $\mathbb{R}^2$ and v_x, v_y are those on the tangent space. Fix $q = (x, y) \in Q$, then the Hessian of L is computed using Equation (5.4):

$$\mathscr{H}_{ij}(q, v) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

This is obviously a singular matrix, which means that L is a singular Lagrangian, i.e. it admits an infinitesimal symmetry, given by the following transformations:

$$\delta x = \epsilon_x(t) \quad \text{and } \delta y = \epsilon_y(t) \quad \text{such that} \quad \epsilon_y = \epsilon_x - \dot{\epsilon_x}$$

Since the Hessian is singular, we expect by the above discussion that the Legendre transform is not bijective. Indeed, applying the definition of the Legendre transform, one has:

$$\mathscr{L}_x(q, v) = \mathscr{L}(q, v)(\partial_x) = v_x + y \quad \text{and} \quad \mathscr{L}_y(q, v) = \mathscr{L}(q, v)(\partial_y) = 0$$

Then, we obtain that:

$$\mathrm{Im}(\mathscr{L}) = \big\{(q, p) \text{ such that there exists } v \in T_qQ \text{ satisfying } p = (v_x + y)dx\big\} \subset T^*Q$$

One can straightforwardly check that the Jacobian of the Legendre transform coincides with the Hessian of the Lagrangian.

Example 5.8. Let us use Example 2 of [Rothe and Rothe, 2010], p. 8, first studied in [Henneaux and Teitelboim, 1994]. The configuration space is $Q = \mathbb{R}^3$, and the Lagrangian is:

$$L(q,v) = \frac{1}{2}(v_y - e^x)^2 + \frac{1}{2}(v_z - y)^2$$

At a given point $q = (x,y,z)$, the Hessian matrix is given by:

$$\mathscr{H}_{ij}(q,v) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The matrix is singular, so is the Lagrangian, which means that it admits a local symmetry. Indeed, it is given by:

$$\delta x = e^{-x}\frac{d^2}{dt^2}\alpha(t), \quad \delta y = \frac{d}{dt}\alpha(t) \quad \text{and} \quad \delta z = \alpha(t)$$

for any smooth function of the time $\alpha(t)$. The Legendre transform should be singular as well. It is given by:

$$\mathscr{L}_x(q,v) = 0, \quad \mathscr{L}_y(q,v) = v_y - e^x \quad \text{and} \quad \mathscr{L}_z(q,v) = v_z - y$$

Then, we obtain that:

$$\mathrm{Im}(\mathscr{L}) = \big\{(q,p) \text{ such that there exists } v \in T_qQ \text{ satisfying } p = (v_y - e^x)dy + (v_z - y)dz\big\} \subset T^*Q$$

One can straightforwardly check that the Jacobian of the Legendre transform coincides with the Hessian of the Lagrangian.

5.2 Hamiltonian under constraints

Let us now dwelve into the case where $\mathscr{L}$ is possibly not invertible, by assuming however that the rank of the smooth function $\mathscr{L}(q,-) : T_qQ \longrightarrow T^*_qQ$ is constant over Q, and we denote this rank $R_{\mathscr{L}}$, for some $1 \leq R_{\mathscr{L}} \leq n$. Let fix $q \in Q$ and let U be a trivializing chart of TQ (and hence of T^*Q as well). Let $v \in T_qQ$, then there exists a reindexing of the coordinates q^i (and thus of the coordinates v^i and p_i) such that:

1. the first $R_{\mathscr{L}}$ coordinates are labelled with a latin index from the beginning of the alphabet $1 \leq a \leq R_{\mathscr{L}}$, while the last $n - R_{\mathscr{L}}$ coordinates are labelled with a greek index from the beginning of the alphabet $R_{\mathscr{L}} + 1 \leq \alpha \leq n$, and
2. the minor $\big(\frac{\partial \mathscr{L}_a}{\partial v^b}\big)_{1 \leq a,b \leq R_{\mathscr{L}}}$ of the Jacobian matrix $\mathscr{J}$ is non-singular at (q,v)[19].

In other words, the $R_{\mathscr{L}}$ functions $\mathscr{L}_a \in \mathcal{C}^\infty(TQ|_U)$ are functionally independent in some open neighborhood $V \subset TQ$ of the point (q,v). Then, the remaining $n - R_{\mathscr{L}}$ functions $\mathscr{L}_\alpha$ are functionally dependent on the former: for each $R_{\mathscr{L}} + 1 \leq \alpha \leq n$ and each base point q, there exists a functional relationship which smoothly depend on q:

$$\mathscr{L}_\alpha = \psi_\alpha(q, \mathscr{L}_a) \tag{5.18}$$

where we understand that each functional ψ_α depends on potentially all the $\mathscr{L}_a$. This argument is an adaptation of the proof of the Rank theorem in [Lee, 2003], see in particular Equation

[19] Since the matrix $\mathscr{J} = \mathscr{H}$ is symmetric, it is always possible to isolate such a minor.

(7.9). Notice that the relabelling of coordinates utterly depends on the chosen tangent vector $(q, v) \in T_qQ$ for the functions $\mathscr{L}_i$ may vary a lot over $TQ|_U$. Hence, the functional dependency (5.18) is in theory only defined locally, in the neighborhood of a given tangent vector, while at another point, we may have another reindexing and correspondingly another dependency. Moreover, the choice of a different minor in $\mathscr{J} = \mathscr{H}$ gives different independent coordinates and thus different functions ψ. However, the number of independent functions would always stay equal to $R_{\mathscr{L}}$.

For every $q \in Q$ let us set $\Gamma_q = \mathrm{Im}(\mathscr{L}(q, -))$ and $\Gamma = \mathrm{Im}(\mathscr{L}) = \bigcup_{q \in Q} \Gamma_q$; it is a subset of T^*Q and we will now study its property. Any covector (q, p) lying in the subspace Γ_q satisfies:

$$p_i = \mathscr{L}_i = \frac{\partial L}{\partial v^i}(q, v) \tag{5.19}$$

for some $(q, v) \in T_qQ$ and a local choice of coordinates q^i, v^i, p_i. Fix a covector $(\widetilde{q}, \widetilde{p}) \in \Gamma$ and a preimage $(\widetilde{q}, \widetilde{v})$ through the Legendre transform. Then from the discussion leading to Equation (5.18), there exists an open neighborhood $V \subset TQ$ of $(\widetilde{q}, \widetilde{v})$ and a reindexing of the coordinates q^i (and thus of the coordinates p_i as well) in two sets such that the coordinates of any covector $(q, p) \in \mathscr{L}(V)$ satisfy:

$$p_a = \mathscr{L}_a \quad \text{and} \quad p_\alpha = \psi_\alpha(q, p_a) \tag{5.20}$$

This is a mere rewriting of Equation (5.19), where we have replaced the terms $\mathscr{L}_i$ by p_i since they coincide on Γ. Moreover, we have used Equation (5.18) to p_α in terms of the p_a. This latter set of equations is a priori only valid on $\mathscr{L}(V)$. However, since on the open set V the functions $\mathscr{L}_a$ are independent and coincide with the p_a on $\mathscr{L}(V)$, one can see the ψ_α as functions of $p_a$and locally extend them outside $\mathscr{L}(V)$ by replacing $\mathscr{L}_a$ by p_a in their argument. See Equation (7.9) in the proof of the Rank theorem in [Lee, 2003] to understand the dependency of ψ_α in terms of independent functions. Let W be such a small neighborhood of $(\widetilde{q}, \widetilde{p})$ on which we formally extend these functions $\psi_\alpha \in \mathcal{C}^\infty(W)$ (it needs not contain the whole of $\mathscr{L}(V)$). Then one can define the following set of smooth functions on W:

$$\phi_\alpha(q, p) := p_\alpha - \psi_\alpha(q, p_a) \qquad \text{for every } R_{\mathscr{L}} + 1 \leq \alpha \leq n$$

called *primary constraints*. In particular these functions only depend on the generalized coordinates and on (part of) the conjugate momenta. The adjective *primary* denotes a further distinction between additional constraints that we will discuss next. The functions ϕ_α actually emerge naturally in the proof of the Rank theorem in [Lee, 2003]. Notice that the choice of a different minor in $\mathscr{J} = \mathscr{H}$ gives different independent coordinates and thus different primary constraints.

Remark 5.9. For reasons that will soon become clear, the triple (W, p_a, ϕ_α) is called a *constrained chart adapted to* $(\widetilde{q}, \widetilde{p})$ (often we will omit to mention the dependency of these data on the original choice of point $(\widetilde{q}, \widetilde{p})$). Since the definition of such charts depend on the choice of preimage of $(\widetilde{q}, \widetilde{p})$, every point of Γ might admit as many adapted constrained charts as it possesses preimages.

The choice of coordinates on Q has been made precisely so that the functions $\mathscr{L}_a$ form a set of independent functions on V and that they span the same subspace of W as the first p_a coordinates (see the rank theorem [Lee, 2003]). Moreover, since each primary constraint ϕ_α involves linearly a different p_α, they form another independent set of functions, and since they altogether form an independent set of functions on W, it turns the constrained chart (W, p_a, ϕ_α) into a coordinate chart of T^*Q. Then, since the vanishing of the primary constraints is equivalent to the second set of equations (5.20), we conclude that the primary constraint characterize the set $W \cap \mathscr{L}(V)$. Indeed, since the primary constraint are functionally independent on W – we say that they are *irreducible* – we conclude that the smooth map $\Phi = (\phi_1, \ldots, \phi_{n-R_{\mathscr{L}}}) : W \longrightarrow \mathbb{R}^{n-R_{\mathscr{L}}}$

has constant rank. Then, since Φ is surjective (one is free to chose any value for the p_α, whatever value for p_a has been chosen), it implies that it is a submersion (Theorem 7.14 in [Lee, 2003]). Being the zero level set of a submersion, the set $W \cap \mathscr{L}(V)$ is a *closed embedded submanifold* of $W \subset T^*Q$ (Corollary 8.9 in [Lee, 2003]). However, it does not imply that $W \cap \Gamma$ is an embedded submanifold of W, for Γ might be an immersed submanifold of T^*Q and have self intersection corresponding to the image through $\mathscr{L}$ of subset of TQ located far from V. More precisely the primary constraints depend primarily on the choice of preimage of $(\widetilde{q}, \widetilde{p})$. Although the matrix $\mathscr{J}(\widetilde{q}, v)$ has rank $R_{\mathscr{L}}$ for every $v \in T_{\widetilde{q}}M$, another choice of preimage $(\widetilde{q}, \widetilde{v})$ and of open set $V' \subset TQ|_U$ may imply another form of dependency from the components $\mathscr{L}_i$. That is to say: another reindexing of the coordinates q^i, as well as another dependency between the corresponding $\mathscr{L}_\alpha$, leading to a redifinition of the ψ_α and hence of the primary constraints defined on another neighborhood W' of $(\widetilde{q}, \widetilde{p})$. The vanishing of these new constraints would this turn make the set $W' \cap \mathscr{L}(V')$ – not necessarily coinciding with $W \cap \mathscr{L}(V)$ – a closed embedded submanifold. That would certainly not prevent Γ to be an immersed submanifold, with possible intersections. To avoid such annoying cases, physicists usually assume that the functions ϕ_α satisfy a so-called *regularity* condition (see alternative formulations on p. 7 of [Henneaux and Teitelboim, 1994]):

Scholie 5.10. Regularity condition on primary constraints. *For every covector $(\widetilde{q}, \widetilde{p}) \in \Gamma$, and any constrained chart (W, p_a, ϕ_α) adapted to $(\widetilde{q}, \widetilde{p})$, the subset $W \cap \Gamma$ is assumed to coincide with the zero level set of the* primary constraints *ϕ_α.*

By Lemma 3.40, the regularity condition presented in Scholie 5.10 implies that Γ is an embedded submanifold of T^*Q. For every point $(\widetilde{q}, \widetilde{p}) \in \Gamma$, and any constrained chart (W, p_a, ϕ_α), the coordinates (p_a, ϕ_α) form a set of local coordinates on W adapted to Γ. More precisely, the coordinates (q^i, p_a) form a local coordinate chart for Γ (because every point on $W \cap \Gamma$ can be retrieved from these data in a unique and smooth way using the smooth functions ψ_α), while the constraints ϕ_α are coordinate transverse to Γ.

Definition 5.11. *We call the embedded submanifold $\Gamma = \mathrm{Im}(\mathscr{L})$ (also denoted $\Gamma^{(1)}$) the* primary constraint surface*.*

The fact that Γ is an embedded submanifold of T^*Q implies in particular that the Legendre transform $\mathscr{L}$ is a submersion. It then admits local sections: for any point $(\widetilde{q}, \widetilde{p}) \in \Gamma$ and adapted constrained chart (W, p_a, ϕ_α), there exists a smooth injective map $\nu : \Gamma \longrightarrow TQ$ such that $\mathscr{L}(q, \nu(q,p)) = (q,p)$ for any $(q,p) \in \Gamma$[20]. This map does actually depend only on $R_{\mathscr{L}}$ momenta, that we can chose to be the p_a, i.e. $\nu(q,p) = \nu(q^i, p_a)$. This allows to find an explicit expression of H_0 in terms of q and p only, and make sense of Equation (5.17) even when the Legendre transform is not invertible. The submanifold $\nu(\Gamma) \oplus \Gamma \subset \mathbb{T}Q$ is by construction a submanifold of $N_{\mathscr{L}}$. Since H_c does not depends on v over $N_{\mathscr{L}}$ (and hence, of the section ν), we deduce that the 'restriction' of the canonical hamiltonian H_c to $\nu(\Gamma) \oplus \Gamma$ gives an explicit formulation of the smooth function H_0, as defined sloppily in (5.17). In local coordinates in an adapted constrained chart (W, p_a, ϕ_α), we indeed have:

$$H_0(q, p_a) = p_a \nu^a(q, p_a) + \psi_\alpha(q, p_a) \nu^\alpha(q, p_a) - L(q, \nu(q, p_a)) \tag{5.21}$$

The hamiltonian does only depend on the first $R_{\mathscr{L}}$ coordinates because ν does, but this implies in turn that H_0 is defined only on $\Gamma = \mathrm{Im}(\mathscr{L})$.

[20] To define a section of $\mathscr{L}$ one needs only Γ to be a weakly embedded submanifold of T^*Q, because in that case one can show using Definition 3.51 that the Legendre transform $\mathscr{L}$ defines a smooth map onto Γ, which is a necessary condition for $\mathscr{L}$ to be a submersion. Being an immersed submanifold would certainly not be sufficient.

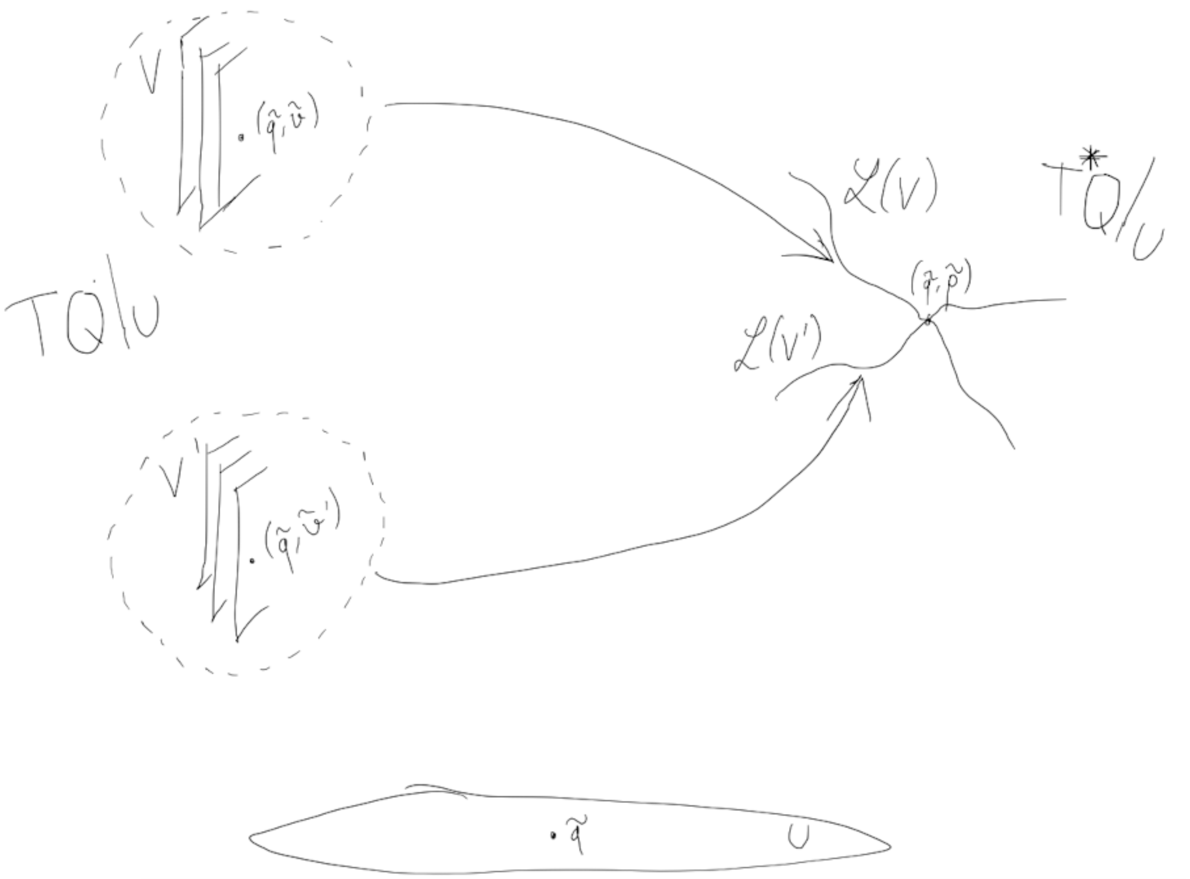

Figure 20: This is a situation we do not want: that different choices of preimages of $(\widetilde{q}, \widetilde{p})$ have neighborhoods V, V' whose image through the Legendre transform $\mathscr{L}$ do not coincide in the vicinity of $(\widetilde{q}, \widetilde{p})$. That is why we ask for the regularity condition, so that the primary constraint surface is an embedded submanifold.

$$
\begin{tikzcd}
 & TQ & \\
\Gamma \arrow[ur, hook, "\nu"] \arrow[dr, hook] \arrow[rr, dashed, bend right=60, "H_0"'] & \oplus \arrow[r, "H_c"] & \mathbb{R} \\
 & T^*Q &
\end{tikzcd}
$$

Notice that we replaced the $n - R_{\mathscr{L}}$ conjugate momenta p_α by ψ_α because they are thus defined on the constraint surface Γ. Hence the hamiltonian H_0 does not depend on the $n - R_{\mathscr{L}}$ conjugate momenta p_α. Moreover, because of Equation (5.9), H_0 does not depend on the choice of section ν. A careful discussion about this independence can be found in Proposition 1 (section 3.3) of [Rothe and Rothe, 2010]. Moreover, although the local expression of H_0 depends on the original choice of splitting between independent momenta p_a and dependent momenta p_α on Γ (and then ultimately on the choice of invertible minor of the matrix $\mathscr{J} = \mathscr{H}$), any other choice would give a function H_0' that would coincide with H_0 on Γ. Some physicists emphasize that this discussion is purely local (see e.g. page 24 in [Gitman and Tyutin, 1990]) while other assume that local coordinates are actually global coordinates (i.e. they work on a vector space), so that Equation (5.21) is valid globally (see e.g. page 10 in [Henneaux and Teitelboim, 1994]). Under this assumption, the hamiltonian H_0 is a smooth function on the primary constraint surface, i.e. $H_0 \in \mathcal{C}^\infty(\Gamma)$.

Extending H_0 out of the constraint surface is actually necessary to proceed to Hamiltonian treatment of constrained systems. Indeed, there is no Poisson bracket on Γ so one cannot formally write Hamilton's equations with H_0 in their classical form. One would need to *replace* H_0 by a smooth function on the phase space, so that Hamiltonian treatment of the system could be done. Moreover it would solve the practical issue raised by the fact that Equation (5.21) is only defined locally. Thus, a smooth function $H \in \mathcal{C}^\infty(T^*Q)$ which coincides with H_0 on Γ, i.e. such that $H|_\Gamma = H_0$, would then be a global smooth extension of H_0 to the whole of phase space, and would be a potential candidate to perform Hamiltonian analysis on T^*Q. If Γ is not closed, it may not exist (see Lemma 4.69) but as physicists, we will assume such global extension exists (this is the case in particular if we assume that the primary constraints are finite and defined globally). The choice of the map H is physically not relevant because physics only occurs on the constraint surface. It turns out however that the primary constraints should be explicitly taken into account so that one prefers to use the following function:

Definition 5.12. *Assume that there is a finite number of globally defined constraints ϕ_α defining a closed embedded submanifold $\Gamma \subset T^*Q$, and let u^α be yet unspecified smooth functions on T^*Q (that physicists sometimes identify with* velocities*). Let H be a smooth function which coincides with H_0 on the primary constraint surface Γ. Then we define the* total Hamiltonian *to be the smooth function:*

$$H_T = H + u^\alpha \phi_\alpha \tag{5.22}$$

Remark 5.13. By definition, $H_T|_\Gamma = H_0$, but we will see that the presence of the constraints are necessary for the consistency of the dynamics. For more informations on this function see e.g. the Corollary on page 31 of [Rothe and Rothe, 2010], or a similar but less general discussion on page 16 of [Dirac, 1964], or a more obscure but quite interesting approach in section 2.1 of [Gitman and Tyutin, 1990].

Example 5.14. Using the Lagrangian of Example 5.7, one observes that the Legendre transform has rank 1 so the dimension of $\Gamma = \mathrm{Im}(\mathscr{L})$ is 3 (because Q has dimension 2, to which we add 1 for the rank of $\mathscr{L}$). The only dependent function $\mathscr{L}_\alpha$ is $\mathscr{L}_y$ and it vanishes. Thus, the only primary constraint is $\phi = p_y$, and the primary constraint surface is characterized by the vanishing of this constraint, i.e. $\Gamma = p_y^{-1}(0)$. The canonical Hamiltonian H is given by:

$$H_c(q,v,p) = p_x v_x + p_y v_y - \left(\frac{1}{2}v_x^2 + v_x y + \frac{1}{2}(x-y)^2\right)$$

On the constraint surface $\Gamma = \mathrm{Im}(\mathscr{L})$, we know that $p_y = \mathscr{L}_y = 0$ and $p_x = \mathscr{L}_x = v_x + y$. By construction, these relations are also valid on $N_{\mathscr{L}}$. Thus, evaluating H_c on the latter gives:

$$H_0(q,p) = p_x(p_x - y) - \left(\frac{1}{2}(p_x-y)^2 + (p_x-y)y + \frac{1}{2}(x-y)^2\right) = \frac{1}{2}p_x^2 - \frac{1}{2}x^2 + xy - yp_x \tag{5.23}$$

Alternatively, one would obtain this expression plugging in Equation (5.21) the following section of Γ to TQ is $\nu_x = p_x - y$ and $\nu_y = 0$. While H_0 is supposedly defined only over Γ, one can straightforwardly extend it to the whole phase space $T^*\mathbb{R}^2$ as a function H, and then define the total hamiltonian as:

$$H_T = H + up_y \tag{5.24}$$

where $u \in \mathcal{C}^\infty(T^*\mathbb{R}^2)$ is still an unfixed smooth function acting as a parameter.

Example 5.15. Using the Lagrangian of Example 5.8, one observes that the Legendre transform has rank 2, so that the constraint surface Γ is a 5-dimensional submanifold of $T^*\mathbb{R}^3$. The only primary constraint is $p_x = 0$ so that $\Gamma = p_x^{-1}(0)$. On this submanifold, we have moreover $p_y = \mathscr{L}_y = v_y - e^x$ and $p_z = \mathscr{L}_z = v_z - y$. These identities – together with $p_x = 0$ – are also valid on the submanifold $N_{\mathscr{L}} \subset \mathbb{T}Q$. Thus, evaluating the canonical Hamiltonian H_c on $N_{\mathscr{L}}$ gives:

$$H_0 = p_y(p_y + e^x) + p_z(p_z + y) - \left(\frac{1}{2}p_y^2 + \frac{1}{2}p_z^2\right) = \frac{1}{2}p_y^2 + \frac{1}{2}p_z^2 + p_y e^x + y p_z \tag{5.25}$$

While this function is supposedly defined only over Γ, one can straightforwardly extend it to the whole phase space $T^*\mathbb{R}^2$ as a function H, so that the total Hamiltonian is:

$$H_T = H + u p_x$$

where $u \in \mathcal{C}^\infty(T^*\mathbb{R}^2)$ is still an unfixed smooth function acting as a parameter.

To justify the use of H_T, let us differentiate H_0 with respect to the canonical variables q and p. A detailed discussion about this can be found in Proposition 2 (section 3.3) of [Rothe and Rothe, 2010]. First, deriving Equation (5.21) with respect to p_a and noticing that $p_i = \frac{\partial L}{\partial v^i}$ on Γ, one obtains that the terms $p_a \frac{\partial \nu^a}{\partial p_i} + \psi_\alpha \frac{\partial \nu^\alpha}{\partial p_i}$ cancels out with $\frac{\partial L}{\partial v^j}\frac{\partial \nu^j}{\partial p_i}$ so that we obtain:

$$\frac{\partial H_0}{\partial p_a} = \nu^a + \frac{\partial \psi_\alpha}{\partial p_a}\nu^\alpha \tag{5.26}$$

We see that there is no contribution of the derivatives of ν with respect to p_a. Notice however that this observation is valid only on the primary constraint surface Γ, and thus so is Equation (5.26). By definition of ϕ_α, Equation (5.26) can be straightforwardly rewritten:

$$\frac{\partial H_0}{\partial p_a} = \nu^a - \frac{\partial \phi_\alpha}{\partial p_a}\nu^\alpha \tag{5.27}$$

Unfortunately the set of Equations (5.27) does not include the derivative with respect to the p_α since H_0 does not depend on them. However, relying on this fact and that $\frac{\partial \phi_\beta}{\partial p_\alpha} = \delta_\beta^\alpha$ on Γ, one may add a set of additional tautological equations:

$$\frac{\partial H_0}{\partial p_\alpha} = \nu^\alpha - \frac{\partial \phi_\beta}{\partial p_\alpha}\nu^\beta \tag{5.28}$$

Hence we notice that a priori Equations (5.27) and (5.28) do not involve time whatsoever.

Next, differentiating Equation (5.21) with respect to q^i and noticing that $p_i = \frac{\partial L}{\partial v^i}$ on Γ, one obtains that the terms $p_a \frac{\partial \nu^a}{\partial q^i} + \psi_\alpha \frac{\partial \nu^\alpha}{\partial q^i}$ cancels out with $\frac{\partial L}{\partial v^j}\frac{\partial \nu^j}{\partial q^i}$, so that we obtain:

$$\frac{\partial H_0}{\partial q^i} = -\frac{\partial \phi_\alpha}{\partial q^i}\nu^\alpha - \frac{\partial L}{\partial q^i} \tag{5.29}$$

Notice that we had replaced ψ_α by $-\phi_\alpha$ since by construction their derivative with respect to q^i coincide. Now, assume that we restrict our study to a smooth curve $\gamma : \mathbb{R} \longrightarrow M$ so that $q = \gamma(t)$ and the vector field corresponding to the velocity at time t is tangent to the curve at every time t and lives in the image of the section ν, i.e. $\dot{q}(t) = \dot{\gamma}(t) = \nu(q(t), p(t))$. The image through $\mathscr{L}$ of the path $t \longmapsto \big(q(t), \dot{q}(t)\big)$ defines a path in the phase space $t \longmapsto \big(q(t), p(t)\big)$. Then, one may add $\frac{d}{dt}\big(\frac{\partial L}{\partial v^i}\big)$ to Equation (5.29) and substract $\dot{p}_i$ (since they compensate one another on Γ by Equation (5.10)), to obtain:

$$\frac{\partial H_0}{\partial q^i} = -\dot{p}_i - \frac{\partial \phi_\alpha}{\partial q^i}\nu^\alpha(t) + E_i\big(q(t), \nu(t), \dot{\nu}(t)\big) \tag{5.30}$$

where $E_i(q(t), \nu(t), \dot{\nu}(t))$ is the smooth function defined in Formula (5.5), which vanishes precisely when the path is a solution of the Euler-Lagrange equations (5.6). Now, assume that the path γ is a solution of the Euler-Lagrange equations (5.6), and that we have $\nu(t) = \nu(q(t), p(t)) = \dot{\gamma}(t) = \dot{q}(t)$. Then, the image of such a path through the Legendre transform $\mathscr{L}$ defines a path $t \longrightarrow (q(t), p(t))$ staying in the primary constraint surface Γ, and whose time derivative satisfy the infamous *Hamilton equations of motion* satisfied by q^i and p_i:

$$\dot{q}^i = \frac{\partial H_0}{\partial p_i} + \frac{\partial \phi_\alpha}{\partial p_i}\nu^\alpha \tag{5.31}$$

$$\dot{p}_i = -\frac{\partial H_0}{\partial q^i} - \frac{\partial \phi_\alpha}{\partial q^i}\nu^\alpha \tag{5.32}$$

We obtained these equations by gathering Equations (5.27), (5.28) with Equations (5.30) and reordering the terms. Notice that, due to the constraints, they do not precisely respect the usual form of Hamilton's equations of motions. We will soon see how one can recast these in this form.

Recall that, although the first Hamilton equations of motion (5.31) are mere consequences of the Legendre transform (and are valid without assuming that ν is of the form $\dot{\gamma}(t)$), the second ones (5.32) are satisfied if the Euler-Lagrange equations (5.6) are satisfied (this is a consequence, and not an equivalence). Moreover, in both case we see that, for points of T^*Q to be considered as potential candidates for physical states of the system – or equivalently, for paths to be considered physical trajectories in the phase space – they at least need to live on Γ, where the Hamiltonian is defined. It does not mean however that every point of the primary constraint surface Γ is an admissible physical state – and we will see that in general they do not. Finally, notice that Equations (5.31) and (5.32) can be recasted in a system of Equations which ressembles more Hamilton equations of motions, at the cost of enforcing the constraint equations:

$$\begin{cases} \dot{q}^i &= \frac{\partial}{\partial p_i}\Big(H_0 + \phi_\alpha \nu^\alpha\Big) \\ \dot{p}_i &= -\frac{\partial}{\partial q^i}\Big(H_0 + \phi_\alpha \nu^\alpha\Big) \\ \phi_\alpha = 0 \end{cases}$$

where here ϕ_α is evaluated on the smooth path $(q(t), p(t))$. This set of equations is consistent with the set of equations (5.13): indeed, if one adds $\phi_\alpha \nu^\alpha$ to Equation (5.21), one obtains Equation (5.8) for $v = \nu(q, p)$. Then Equations $\frac{\partial H_c}{\partial v^i} = 0$ imply that $H_c(q, \nu(q, p), p)$ does not depend on the section ν, or equivalently, that we are working on $N_{\mathscr{L}}$, which is alternatively said by imposing $\phi_\alpha = 0$ and $p_a = \mathscr{L}_a$.

Unfortunately, since H_0 is a priori not defined outside the primary constraint surface Γ, we cannot write the above set of equations with the help of the Poisson bracket. For this a function defined all over the phase space would be necessary. However, the presence of the terms $H_0 + \phi_\alpha \nu^\alpha$ reminds us of the discussion surrounding Equation (5.22) where we said that replacing H_0 by any smooth function $H \in \mathcal{C}^\infty(T^*Q)$ such that $H|_\Gamma = H_0$ would lead to the same physics and, more importantly, would open the use of the Poisson bracket on the phase space. Indeed, the corollary of Proposition 3 in [Rothe and Rothe, 2010] shows that for any smooth function $H \in \mathcal{C}^\infty(T^*Q)$ such that $H|_\Gamma = H_0$, there exists smooth functions $u^\alpha \in \mathcal{C}^\infty(T^*Q)$ such that the Hamiltonian equations of motions can be recasted as:

$$\begin{cases} \dot{q}^i &= \frac{\partial}{\partial p_i}\Big(H + \phi_\alpha u^\alpha\Big) \\ \dot{p}_i &= -\frac{\partial}{\partial q^i}\Big(H + \phi_\alpha u^\alpha\Big) \\ \phi_\alpha = 0 \end{cases}$$

The justification comes from the fact that, since H coincides with H_0 on the constraint surface Γ, it may be written (at least locally) as $H_0 + \phi_\alpha \lambda^\alpha$ (see the proof of Proposition 5.17), and the smooth functions λ^α are so that on the constraint surface, one has $\lambda^\alpha = \frac{\partial H}{\partial p_\alpha}$. It then implies that $u^\alpha = \nu^\alpha - \lambda^\alpha$. The latter hamiltonian equations of motions are quite convenient because they are defined outside of Γ, if not on the whole phase space (when the constraints are so defined).

The fact that the primary constraints appear explicitly in the above set of equations also justifies that the correct Hamiltonian is not H, but the total Hamiltonian $H_T = H + u^\alpha \phi_\alpha$, as postulated in Equation (5.22). Indeed, denoting $\{\,.\,,.\,\}$ the canonical Poisson bracket on the cotangent bundle, associated to the canonical symplectic form on T^*Q, one can then recast Hamilton's equations of motions as:

$$\begin{cases} \dot{q}^i &= +\frac{\partial H_T}{\partial p_i} = \{q^i, H_T\} \\ \dot{p}_i &= -\frac{\partial H_T}{\partial q^i} = \{p_i, H_T\} \\ \phi_\alpha = 0 \end{cases} \tag{5.33}$$

These equations are a consequence of the extended Euler-Lagrange equations (5.13) and thus, of the original ones as well. Thus, although $H_T|_\Gamma = H_0$, the presence of the primary constraints in its definition are of utter importance. We will see later that we can find a set of equations extending (5.33) which is equivalent to the Euler-Lagrange equations. The discussion appearing in section 2.1 of [Gitman and Tyutin, 1990] is quite interesting althoug a bit obscure, because it justifies that although the splitting into independent conjugate momenta p_a and dependent ones p_α on Γ is not unique (one could have chosen another set of independent coordinates p_a), the hamiltonian H_0 is uniquely defined and the total hamiltonian forms a class of function 'equivalent' to that of H_0. See also [Rothe and Rothe, 2010] for additional food for thoughts.

5.3 The Bergmann-Dirac algorithm

The importance of the primary constraint surface in the Hamiltonian formalism of singular Lagrangian theories can be best shown after introducing some adapted notation:

Definition 5.16. *We say that two functions $f, g \in \mathcal{C}^\infty(T^*Q)$ are* weakly equivalent *if they coincide on Γ, and we note:*

$$f \approx g$$

For clarity, we say that they are strongly equivalent *if they coincide on the whole of phase space T^*Q.*

Being weakly equivalent is an equivalence relation, and this notion will be thoroughly used in the text. Since Γ is an embedded submanifold of Q defined as a level set of a set of smooth functions, it turns out that any smooth function vanishing on Γ is functionally locally dependent on the primary constraints:

Proposition 5.17. *Let $f \in \mathcal{C}^\infty(T^*Q)$ be a smooth function that is weakly equivalent to the zero function: $f \approx 0$. Then, for every point $(\widetilde{q}, \widetilde{p}) \in \Gamma$ and any choice of adapted constrained chart (W, p_a, ϕ_α) there exist $f_\alpha \in \mathcal{C}^\infty(T^*Q)$ such that $f = \sum_\alpha f_\alpha \phi_\alpha$ on W.*

Proof. The proof is given in Appendix of Chapter 1 of [Henneaux and Teitelboim, 1994], or alternatively in section 3.3 of [Rothe and Rothe, 2010], where it was adopted from [Sudarshan and Mukunda, 2015]. □

Remark 5.18. Notice that the statement of Proposition 5.17 is a local one, while in the references cited for the proof, the statement is a global one. The discrepancy comes from the fact that in physics textbooks, the constraints are defined globally over the phase space T^M. This is an extra assumption that is *not* a consequence of the Legendre transform. On the contrary, we have shown that using the Rank theorem, only local statement can be made on the form of the constraints. Then, while physicists usually think of constraints as a finite set of globally defined constraints, mathematicians should definitely think of them as a locally free and finitely generated subsheaf of the sheaf of smooth functions $\mathcal{C}^\infty(M)$. One may even weaken this assumption by asking it to be only locally finitely generated.

Using the notation of Definition 5.16, one can recast equations (5.33) as:

$$\dot{q}^i \approx \{q^i, H_T\} \tag{5.34}$$
$$\dot{p}_i \approx \{p_i, H_T\} \tag{5.35}$$

The Poisson bracket has to be evaluated on Γ *after* it has been computed – i.e. we compute $\{q^i, H_0\}$ but $\{q^i, H_T\}$ and then we apply $\phi_\alpha = 0$. The total Hamiltonian thus defines the flow of time when we restrict ourselves to the primary constraint surface Γ. Equations (5.34) and (5.35) imply in turn that the total Hamiltonian computes the dynamics of any smooth function which is evaluated on any physical path sitting in Γ. More precisely, let $f \in \mathcal{C}^\infty(T^*Q)$ be any smooth function, and let $(\widetilde{q}, \widetilde{p})$ be any point of Γ. Then for small times t, and under the assumption that the undefined parameters u^α are fixed, there is a unique path $t \mapsto (q(t), p(t))$ such that:

$$\begin{cases} q(0) = \widetilde{q} \\ p(0) = \widetilde{p} \end{cases} \qquad \text{and} \qquad \begin{cases} \dot{q}^i(0) = \{q^i, H_T\}(\widetilde{q}, \widetilde{p}) \\ \dot{p}_i(0) = \{p_i, H_T\}(\widetilde{q}, \widetilde{p}) \end{cases} \tag{5.36}$$

which is contained in Γ, i.e. such that $\phi_\alpha(q(t), p(t)) = 0$ for all times t. We then define the following real numbers:

$$\dot{f}(q(t), p(t)) = \frac{\partial f}{\partial q^i}(q(t), p(t))\dot{q}^i(t) + \frac{\partial f}{\partial p_i}(q(t), p(t))\dot{p}_i(t)$$

By unicity of the Cauchy problem the value of $\dot{f}$ only depends on the point, and not on the path. The right-hand side is not only a smooth function of the time t, but also of the base point $(\widetilde{q}, \widetilde{p})$. Then, we can define a smooth assignment:

$$\begin{aligned} \dot{f}: \quad \Gamma &\longrightarrow \mathbb{R} \\ (\widetilde{q}, \widetilde{p}) &\longmapsto \frac{\partial f}{\partial q^i}(\widetilde{q}, \widetilde{p})\dot{q}^i(0) + \frac{\partial f}{\partial p_i}(\widetilde{q}, \widetilde{p})\dot{p}_i(0) \end{aligned} \tag{5.37}$$

where $\dot{q}^i(0)$ and $\dot{p}_i(0)$ are uniquely defined by Conditions (5.36). Since the primary constraint surface is an embedded submanifold of T^*Q, the assignment (5.37) admits at least locally a smooth extension, and two such extensions coincide on Γ. Then we have the following important result about dynamics:

Lemma 5.19. *Let $f \in \mathcal{C}^\infty(T^*Q)$ be any smooth function, and let $\dot{f}$ be any smooth extension of the associated smooth assignment (5.37); then:*

$$\dot{f} \approx \{f, H_T\}$$

Since every physical solution of the Hamilton equations (5.33) should be contained in the contraint surface Γ, it means that if one evaluates the primary constraints ϕ_α on any such physical path $t \mapsto (q(t), p(t))$, one has $\dot{\phi}_\alpha(q(t), p(t)) = 0$ because $\phi_\alpha(q(t), p(t)) = 0$ for all t. Using Lemma 5.19, this necessary condition reads:

$$\{\phi_\alpha, H_T\} \approx 0 \tag{5.38}$$

We call this equation the *persistence* of the primary constraints ϕ_α. Developing the total hamiltonian, computing the bracket, and eventually evaluating the result on Γ gives:

$$\{\phi_\alpha, H\} + u^\beta\{\phi_\alpha, \phi_\beta\} \approx 0 \tag{5.39}$$

Recall that H is *any* smooth function on T^*Q that coincides with H_0 on the constraint surface: $H|_\Gamma = H_0$. Then there are several possible issues for each index $\alpha = R_{\mathscr{L}} + 1, \dots, n$:

1. either the equation reduces to $0 = 0$ in that case nothing new is known;
2. either the Poisson bracket of ϕ_α with some of the primary constraints does not vanish weakly and then the equation involves the u's and thus impose a relationship between some of them;
3. or the Poisson bracket of ϕ_α with *all* the primary constraints does vanish weakly, i.e. $\forall\,\beta\ \{\phi_\alpha, \phi_\beta\} \approx 0$ for this particular α, then the equation reduces to $\{\phi_\alpha, H\} \approx 0$. This equation is independent of the primary constraints, otherwise it would be of the first kind. This defines a new *second-stage* constraint[21] $\phi^{(2)}_\alpha = \{\phi_\alpha, H\} \in \mathcal{C}^\infty(T^*Q)$.

Remark 5.20. The freedom of choice in the parameters u^α is thus utterly tied to the existence of constraints whose Poisson bracket with other constraints vanishes on Γ. As we will see later these are called *first-class constraints* and are the markers of gauge symmetries (i.e. arbitrary transformations of the coordinates through time), through what is called the *Dirac conjecture*.

The vanishing of this second-stage constraint $\phi^{(2)}_\alpha$ (for this particular α) on the smooth path $t \mapsto (q(t), p(t))$ is then interpreted as a necessary condition for the equations $\dot{\phi}_\alpha \approx 0$ to be satisfied. Notice that one cannot just replace the identities

$$\phi_\alpha(q(t), p(t)) = 0 \text{ for every } t \tag{5.40}$$

by the vanishing of the second-stage constraints (together with some initial condition $\phi_\alpha(q(0), p(0)) = 0$) because we originally used Equations (5.40) to define the second-stage constraints. From this, we deduce that physical solutions of Hamilton equations (5.33) should in fact be contained in the intersection of the contraint surface Γ and of the zero level set of all the second-stage constraints $\phi^{(2)}_\alpha$ (for every α for which they exist). Thus, let us define $\Gamma^{(2)}$ to be the subamnifold of T^*Q corresponding to the zero level set of the primary (equivalently, first-stage) and second-stage constraints:

$$\Gamma^{(2)} = \Gamma \cap \bigcap_\alpha \big(\phi^{(2)}_\alpha\big)^{-1}(0)$$

This definition is consistent because if $\phi^{(2)}_\alpha = 0$ – i.e. there is no second-stage constraint associated to ϕ_α – then $\big(\phi^{(2)}_\alpha\big)^{-1}(0) = T^*Q$. Moreover, until the next step of the algorithm, the weak equivalence sign is interpreted to be defined relatively to the submanifold $\Gamma^{(2)}$ defined by *all* the constraints generated up to this point: all the primary (equivalently, first-stage) and second-stage constraints, and not only the primary ones.

Thus, given that the vanishing of the second-stage constraints are necessary conditions for Hamilton equations (5.33) to stand, we deduce that any solution of the Hamilton equations lives in $\Gamma^{(2)}$. The persistence of the primary constraints being conditioned to the persistence of the second-stage constraints, one then should necessarily have $\dot{\phi}^{(2)}_\alpha(q(t), p(t)) = 0$ for any physical path $t \mapsto (q(t), p(t))$ (in addition to the condition that $\phi_\alpha(q(t), p(t)) = 0$ for any such path). In other words, one needs to impose that $\dot{\phi}^{(2)}_\alpha \approx 0$ (on $\Gamma^{(2)}$ then), which translates as:

$$\{\phi^{(2)}_\alpha, H\} + u^\beta\{\phi^{(2)}_\alpha, \phi_\beta\} \approx 0 \tag{5.41}$$

[21] The terminology "second-stage", "third-stage", etc. is taken from [Gitman and Tyutin, 1990].

This equation then falls into either one of the three kinds of the above cases (see text before Remark 5.20). In particular, if Equations (5.41) are not trivial, they will either provide a new relationship between the undefined parameters u's, or a set of *third-stage* constraints $\phi_\alpha^{(3)} = \{\phi_\alpha^{(2)}, H\}$. We impose the weak equivalence in Equation (5.41) relatively $\Gamma^{(2)}$ because there may happen that the third-stage constraints could be redundant with the primary or second-stage constraints. Putting the latter to zero would then enforce the former to be automatically zero as well, and we could then avoid any redundancy.

Persistence of the second-stage constraints requires the third-stage constraints to vanish over any physical path $t \mapsto (q(t), p(t))$. We then define $\Gamma^{(3)}$ to be the surface defined by all the constraints found up to this point: primary (first-stage), second-stage, and third-stage constraints:

$$\Gamma^{(3)} = \Gamma^{(2)} \cap \bigcap_\alpha (\phi_\alpha^{(3)})^{-1}(0)$$

Any physical path satisfying Hamilton equations (5.33) should then belong to this third-stage constraint surface. As for the second step, until the next step of the algorithm, the weak equivalence sign is now interpreted to be defined relatively to the submanifold $\Gamma^{(2)}$ defined by *all* the constraints generated up to this point: all the primary, second-stage *and* third-stage constraints. And then, the algorithm goes on with $\phi_\alpha^{(3)}$ whose time derivative should vanish on $\Gamma^{(3)}$ as a necessary condition for $\phi_\alpha^{(2)}$ and thus ϕ_α to stay invariant through time. The vanishing of $\phi_\alpha^{(3)}$ translates as:

$$\{\phi_\alpha^{(3)}, H\} + u^\beta \{\phi_\alpha^{(3)}, \phi_\beta\} \approx 0$$

where the weak equivalence sign is now understood to be computed with respect to $\Gamma^{(3)}$. If these equations are not trivial, we may find four-stage constraints, and then fifth-stage constraints and so on, but the algorithm terminates because the dimension of the phase space T^*Q is finite. We end up, for each $\alpha = R_{\mathscr{L}} + 1, \ldots, n$, with a sequences of k-th stage constraints $\phi_\alpha^{(k)}$ (the $\phi_\alpha^{(k)}$ are considered to be smooth functions, at least on some local neighborhood W of a fixed point $(\widetilde{q}, \widetilde{p})$ on Γ), and the sequence terminates, for each α, at some integer $k_\alpha \geq 1$. In other words, $\phi_\alpha^{(k_\alpha)} \neq 0$ while $\phi_\alpha^{(k_\alpha)} = 0$ (as smooth functions defined over W or T^*Q). Then, for every $1 \leq k \leq k_\alpha - 1$, one has:

$$\phi_\alpha^{(k+1)} = \{\phi_\alpha^{(k)}, H\}$$

See section 3.4 in [Rothe and Rothe, 2010] for more details on this iterative algorithm.

Example 5.21. We have seen in Example 5.14 that the Hamiltonian H_0 could be straightforwardly extended to the whole phase space as a function H. Since there is only one primary constraint $\phi = p_y$, the second term in Equation (5.39) vanishes and persistence of the primary constraint then reads:

$$\{p_y, H\} \approx 0 \tag{5.42}$$

The parameter u is thus left undetermined, and using Equation (5.23) one obtains that Equation (5.42) is equivalent to:

$$x - p_x \approx 0$$

This is a necessary condition so that $\dot{\phi} \approx 0$ for all times. Then it is promoted to a second-stage constraint $\phi^{(2)} = x - p_x$. Persistence of this constraint does not give rise to any new constraint, and the algorithm stops there.

Example 5.22. Let us proceed in the same way for Example 5.15. There was only one primary constraint $\phi = p_x$. Persistence of this constraint gives the following condition:

$$\{p_x, H\} \approx 0$$

where H is the straightforward extension of the function H_0 defined in Equation (5.25). This gives the following condition:

$$p_y e^x \approx 0$$

which in turn implies that we have a second-stage constraint $\phi^{(2)} = p_y$ (the dependence on x does not appear because $e^x \neq 0$). Persistence of this second stage constraint reads:

$$\{p_y, H\} \approx 0$$

which in turn implies that $p_z \approx 0$. This necessary condition for the persistence of $\phi^{(2)}$ – and then of ϕ altogether – gives rise to the following third-stage constraint $\phi^{(3)} = p_z$. Persistence of this function does not provide any new constraint so the algorithm stops here.

All k-th stage constraints are called *secondary constraints*, in order to emphasize that they come after imposing some condition on the primary constraints. The subset of T^*Q consisting (at least locally) of the zero level set of the set Ω of all the constraints (primary *and* secondary) is called the *(secondary) constraint surface* and is denoted Σ:

$$\Sigma = \bigcap_{k \geq 1} \Gamma^{(k)}$$

This surface is independent on the choice of primary (and then secondary) constraints that is originally made (that was already implicit in the discussion following Remark 5.9). As for the primary constraint surface $\Gamma^{(1)} = \Gamma$, the secondary constraint surface Σ is assumed to satisfy a regularity condition similar to that of Scholie 5.10.

Scholie 5.23. Regularity condition on secondary constraints. *The secondary constraint surface Σ is an embedded submanifold of T^*Q.*

Definition 5.16 and Lemma 5.19 then become modified so that they are now defined with respect to the secondary constraint surface Σ, and not Γ anymore. From now on, we will then use the notation $\approx$ of Definition 5.16 to indicate equivalence of function on the constraint surface Σ. Any physical path – i.e. a solution of Hamilton equations (5.33) – should then be sitting in Σ. This is a consequence of the fact that secondary constraints, which are hidden in the persistence conditions of the primary constraints, are actually needed to draw an equivalence with Euler-Lagrange equations:

Proposition 5.24. *The Euler-Lagrange equations (5.6) are equivalent to the following set of Hamilton equations:*

$$\begin{cases} \dot{q}^i &= \{q^i, H_T\} \\ \dot{p}_i &= \{p_i, H_T\} \\ \phi_\alpha^{(k)} &= 0 \quad \text{for all } k \geq 1 \text{ such that } \phi_\alpha^{(k)} \text{ exists} \end{cases}$$

where we choose the notation $\{\phi_\alpha, \phi_\alpha^{(2)}, \phi_\alpha^{(3)}, \ldots, \phi_\alpha^{(k)}, \ldots\}_{\alpha,k}$ to denote the set of constraints, both primary and secondary.

Proof. See page 29 of [Rothe and Rothe, 2010] and the subsequent section. □

Remark 5.25. In particular it implies that the image through the Legendre transform of a solution $t \mapsto (q(t), \dot{q}(t))$ of the Euler-Lagrange equations sits in the secondary constraint surface Σ.

At this stage the primary and secondary constraints are certainly not all functionally independent. Notice moreover that the difference between primary and secondary is not so clear because the k-th stage secondary constraint $\phi_\alpha^{(k+1)} = \{\phi_\alpha^{(k)}, H\}$ (when it exists) often involves primary constraints in its expression, which then vanish when evaluated over Γ. Moreover, even the choice of primary constraint is not unique, since the choice of a minor in the Hessian matrix of the Lagrangian determines the primary constraints. Another choice of minor would have lead to another set of primary constraints $\phi'^{(1)}_\alpha = \phi'_\alpha$, equivalent to the original set of primary constraints $\phi^{(1)}_\alpha = \phi_\alpha$ through a linear transformation, because they determine the same primary constraint surface Γ. The secondary constraints associated to the ϕ'_α then would have been weakly equivalent to the set of secondary constraints associated to the ϕ_α:

$$\phi'^{(k)}_\alpha \approx \sum_{l\geq 1} \Lambda^\beta_{(l)\alpha} \phi^{(l)}_\beta \tag{5.43}$$

where summation over repeated indices is implicit and the $\Lambda_{(l)}$ are a family of matrices. Thus, secondary constraints are often mixed with primary constraints. However, some author value primary constraints as carrying noticeable information: see e.g. page 39 and page 72 of [Gitman and Tyutin, 1990], or subsection 3.3.2 in [Rothe and Rothe, 2010]. In the latter reference, it is postulated that the parameters u^α appearing in the total Hamiltonian may be considered as the projections of the velocities on the zero eigenspace of the Hessian of the Lagrangian. The primary constraints in this context simply state that these velocities stay finite.

Example 5.26. Let us modify Example 5.7 so that its Lagrangian is obtained as a limit $\alpha \longrightarrow 0$ of the following (non-singular) Lagrangian:

$$L(q,v) = \frac{1}{2}v_x^2 + v_x y + \frac{\alpha}{2}v_y^2 + \frac{1}{2}(x-y)^2$$

As much as $\alpha \neq 0$ the Hessian of L is non-singular:

$$\mathscr{H}_{ij}(q,v) = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}$$

The Legendre transform is then bijective and given by:

$$\mathscr{L}_x(q,v) = \mathscr{L}(q,v)(\partial_x) = v_x + y \quad \text{and} \quad \mathscr{L}_y(q,v) = \mathscr{L}(q,v)(\partial_y) = \alpha v_y$$

In particular, the relationship between velocities and momenta is given by $v_x = p_x - y$ and $v_y = \frac{p_y}{\alpha}$. Thus, evaluating the canonical Hamiltonian H_c on $N_{\mathscr{L}}$ as in Example 5.14, gives H_0:

$$\begin{aligned} H_0(q,p;\alpha) &= p_x(p_x - y) + p_y\frac{p_y}{\alpha} - \left(\frac{1}{2}(p_x-y)^2 + (p_x-y)y + \frac{1}{2\alpha}p_y^2 + \frac{1}{2}(x-y)^2\right) \\ &= \frac{1}{2}p_x^2 + \frac{1}{2\alpha}p_y^2 - \frac{1}{2}x^2 + xy - yp_x \end{aligned} \tag{5.44}$$

Let us rewrite the second term $\frac{1}{2\alpha}p_y^2$ as $\frac{1}{2\alpha}\psi^2$ where $\psi = \alpha v_y$, because this is how the y-velocity and the y-momentum are related to one another via the Legendre transform. This rewriting emphasizes that, although the denominator makes the fraction $\frac{1}{2\alpha}$ diverge when $\alpha \to 0$, at the same time the numerator ψ^2 will converge to 0 twice quicker, making the overall term to vanish. Then, we have $H_0(q,p;\alpha) \underset{\alpha\to 0}{\longrightarrow} H_0(q,p)$, where $H_0(q,p)$ is the Hamiltonian defined in Equation (5.23). One can recast Equation (5.44) as:

$$H_0(q,p;\alpha) = H_0(q,p) + \frac{1}{2}v_y p_y$$

Under this form, one is reminded of the general form of the total Hamiltonian of Equation (5.24). Then one shows that the parameter u is indeed related to the velocities (they may be interpreted as coordinates on the preimage of p_x, as is explained on page 10 of [Henneaux and Teitelboim, 1994]) while the primary constraint emerge naturally from the formalism.

5.4 First-class and second-class constraints, gauge transformations

Thus, the splitting of the set of constraints Ω into primary and secondary constraints is not really relevant. A better distinction is that of *first-class* and *second-class* constraints, originally proposed by Dirac and which has deep relationship with gauge transformations and the *Dirac conjecture*.

Definition 5.27. *A smooth function $f \in \mathcal{C}^\infty(T^*Q)$ is said to be* first-class *if its Poisson bracket with* every *constraint vanishes weakly on Σ. It is said* second-class *otherwise.*

Definition 5.27 tells us that second-class functions satisfy the following non-triviality condition: for any second-class function f there is a point x on the constraint surface Σ and a constraint ϕ_{α_0} such that $\{f, \phi_{\alpha_0}\}(x) \neq 0$. This implies that the bracket $\{f, \phi_{\alpha_0}\}$ is actually non zero in a small neighborhood $U \subset T^*Q$ of x. A priori this is a local property because there is no reason preventing the brackets $\{f, \phi_\alpha\}$ to be weakly equivalent to 0 outside this neighborhood, i.e. they totally could vanish everywhere else on $\Sigma \cap (T^*Q \backslash U)$. However as is usual in Bergmann-Dirac formalism, physicists often consider that such property is global, i.e. that second-class functions will satisfy the non-triviality condition at every point of Σ (in fact at every point of the second-class constraint manifold Σ_0, see subsection 5.5). On the other hand, although first-class functions need not be vanishing on the constraint surface Σ, their Poisson bracket with any constraint ϕ_α is strongly equivalent to a linear combination of the constraints (see Proposition 5.17): $\{f, \phi_\alpha\} = \sum_\beta F_{\alpha\beta}\phi_\beta$. This observation straightforwardly implies the following nice property:

Proposition 5.28. *The Poisson bracket of two first-class functions is first class.*

Proof. Let $f, g \in \mathcal{C}^\infty(T^*Q)$ be first-class functions, so we have:

$$\{f, \phi_\alpha\} = \sum_\beta F_{\alpha\beta}\phi_\beta \quad \text{and} \quad \{g, \phi_\alpha\} = \sum_\beta G_{\alpha\beta}\phi_\beta$$

To evaluate if the Poisson bracket of f and g is first-class we compute:

$$\begin{aligned}\{\{f,g\}, \phi_\alpha\} &= \{\{f, \phi_\alpha\}, g\} + \{f, \{g, \phi_\alpha\}\} \\ &= \{F_{\alpha\beta}\phi_\beta, g\} + \{f, G_{\alpha\beta}\phi_\beta\} \\ &= F_{\alpha\beta}\{\phi_\beta, g\} + \{F_{\alpha\beta}, g\}\phi_\beta + G_{\alpha\beta}\{f, \phi_\beta\} + \{f, G_{\alpha\beta}\}\phi_\beta\end{aligned}$$

which identically vanish on Σ because on the one hand $\phi_\alpha \approx 0$ and on the other hand the functions f and g are first-class. □

Example 5.29. If one considers that the weak equivalence is now defined with respect to the secondary constraint surface – as it should be, the total Hamiltonian H_T is a first class function. Indeed, it satisfies by construction, the persistence equation $\{\phi_\alpha, H_T\} \approx 0$ is satisfied for both primary and secondary constraints.

The characterization made in Definition 5.27 enables us to divide the set of constraints Ω into first-class and second-class constraints. Usually first-class constraints are denoted φ_j and second class constraints are denoted χ_l. The bracket of any first class constraint φ_j with any other constraint vanishes on Σ, while the remaining constraints are second class. As emphasized earlier, physicists often consider that second-class constraints are such that the non-triviality condition is satisfied globally over Σ. In particular, a second-class constraint can have a non-weakly vanishing bracket only if the other argument is another second class constraint (for it the latter was a first class constraint the bracket would vanish on Σ). Notice however that any there is some latitude in the choice of first-class and second-class constraint, since e.g. one cannot make the difference between $\{\chi_l, .\}$ and $\{\chi_l + a_l^j \varphi_j, .\}$, when evaluated on the constraint surface. Conversely, adding a linear combination of squares of second class constraint to a first-class constraint defines another first class constraint: $\varphi_j \mapsto \varphi_j + b_j^{kl}\chi_k\chi_l$, see section 1.3.1 in [Henneaux and Teitelboim, 1994]. Then, reinterpreting Proposition 5.28 in light of these ambiguities, saying that the Poisson bracket of two first-class constraints is again a first-class function – hence a first-class constraint – amounts to saying that it is linear in first class constraints and square in second class. Moreover, the result of the Poisson brackets between first class constraints with second class constraints must be linear in first-class constraints and linear in second-class constraints.

Let D be the matrix made of the Poisson brackets of constraints $\big(\{\phi_\alpha^{(k)}, \phi_\beta^{(l)}\}\big)$. We assume from now on that the rank of this matrix is (at least locally) constant on the constraint surface Σ, and that is rank is r. The rank is necessary an even integer because D is a skew-symmetric matrix. Then, at the cost of redefining the constraints via linear combinations, it is possible to split the set of constraints into a set of r independent second-class constraints, the remaining constraints being first-class. The way it is often done is to first define an independent set of constraints generating Ω, and then use linear combinations of constraints as in Equation (5.43) in order to obtain another set of independent constraints, for which the number of second-class constraints is minimal. By construction, this set of constraints satisfies the property that no linear combination of second-class constraints is first class.

Then, reorganize the lines and columns of the matrix D (in the basis defined from the new set of constraints) to put the first class constraints first. By definition, a bracket involving a first-class constraint vanishes weakly, so the line and the column associated to this constraint should vanish on Σ. Then the algorithm results in a matrix of the form, when evaluated on Σ:

$$\begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix}$$

where C is a square matrix such that $C_{kl} = \{\chi_k, \chi_l\}$, and whose determinant does not vanish. Indeed, if it vanished, then there would be a linear combination of columns which would vanish, implying in turn the existence of a linear combination of second-class constraints being a first-class constraint, which is not possible given the assumptions. Then, we deduce that C is a $r \times r$ matrix of maximal rang and thus, that this number r corresponds to the number of the independent second-class constraints χ_k (see also Chapter 2 of [Dirac, 1964] for an explanation of this process).

The role of first-class and second-class constraints has an important consequence on the treatment of Hamilton's equations of motion. The total Hamiltonian (5.22) contains primary constraints, which are split into first-class and second-class constraints. Depending on this characteristic, the persistence condition (5.38) then admits two totally different possibilities, most visible on Equation (5.39):

1. if ϕ_α is second-class, then Equation (5.39) leads to a new condition on the parameters u^β;

2. if ϕ_α is first-class, then Equation (5.39) leads to a new (second-stage) constraint.

From this observation we conclude that the secondary constraints arise from the first-class primary constraints. More generally, assume that every secondary constraint has been found and that the set of constraints has been split into independent first-class and second-class constraints, where the latter set is the smallest possible. For this specific computation, we will label the primary constraint with an upper index (1), as the second-stage, third-stage, and k-stage constraints usually carry a (k) index. Then both primary and secondary constraints should satisfy (by construction) the persistence equation (5.39) (where the $\approx$ sign is evaluated with respect to the whole set of constraints). Then, by definition of first-class and second-class constraints, we have:

$$\{\varphi_i, H\} \approx 0 \tag{5.45}$$

$$\{\chi_k, H\} + u^l\{\chi_k, \chi_l^{(1)}\} \approx 0 \tag{5.46}$$

Since we assumed that the number of independent second-class constraints is minimal, the matrix $C_{mn} = \{\chi_m, \chi_n\}$ is invertible and one can then compute explicitly the parameters u^l as functions of the canonical variables:

$$u^l = (C^{-1})^{kl}\{\chi_k, H\} \tag{5.47}$$

where the sum is made over all k, and where l is a label associated to a primary second-class constraint only. Notice that we have defined the parameters u^l by a strong equation, rather than as a weak equality. The classical equations of motion are insensitive to this, but it will turn out to be relevant for defining the corresponding quantum theory. Moreover, under the assumption that we have a minimal set of second-class constraints, the solution u^l of Equation (5.46) are unique for the following reason: indeed, the only freedom in the choice of u^l would come from the solutions of the homogeneous equation $v^l\{\chi_k, \chi_l^{(1)}\} \approx 0$, but if v^l is such an arbitrary solution, then it would mean that $v^l\chi_l^{(1)}$ is a first class constraint because it commutes with all second class constraints, which is impossible given the original assumption. See subsection 3.5 in [Rothe and Rothe, 2010] for additional informations and subsections 1.1.7 and 1.1.10 of [Henneaux and Teitelboim, 1994] for a treatment of this problem form the other way around (one first solve Equations (5.46) and then realize that the free parameters v^l can be absorbed into the coefficients associated with the primary first-class constraints).

Then, we see that the division between first-class and second-class constraints such that the latter are minimal in number allow us to identify which 'velocities' u^α can be fully determined. We see on Equation (5.45) that those parameters associated to the primary first-class constraints cannot be determined by Hamilton equations. They will remain undetermined so they are then free functions of time and will play the role of gauge parameters of the system. On the contrary, as discussed above the 'velocities' associated to the primary second-class constraints are fully determined. This observation leads us to the following conclusion

Definition 5.30. *Let H be the smooth function which coincides with H_0 on the primary constraint surface Γ used into Definition 5.12, and let u^l be the uniquely defined parameters associated to the primary second-class constraints $\chi_l^{(1)}$ in Equation (5.47). Then we define the* first-class Hamiltonian *as:*

$$H' = H + u^l\chi_l^{(1)}$$

Contrary to the original definition of the total Hamiltonian, the first class hamiltonian provides a splitting of H_T into two first-class functions:

$$H_T = H' + v^i\varphi_i^{(1)} \tag{5.48}$$

Authors often consider that it is permissible to add any linear combination of primary first-class constraints to H' (this would correspond to a rewriting of the free parameters v^i), but for clarity we would not consider this option here. The main characteristic of the first-class Hamiltonian is that it is first-class:

Lemma 5.31. *The first-class hamiltonian is a first-class function.*

Proof. The so-called first-class hamiltonian is indeed first-class: for any constraint ϕ_α (be it primary or secondary), we have:

$$\{H', \phi_\alpha\} = \{H_T, \phi_\alpha\} + \{v^i, \phi_\alpha\}\varphi_i^{(1)} - v^i\{\varphi_i^{(1)}, \phi_\alpha\}$$

The first term vanishes on-shell – i.e. on the constraint surface Σ – by Example 5.29. The second term is proportional to a constraint while the third term vanishes on Σ because $\varphi_i^{(1)}$ is a first-class constraint. We see that the first-class hamiltonian is not uniquely defined because any other choice of free parameters v^i still gives a first-class function. □

We will now study the relationship between first-class constraints and gauge transformations, and what does the latter mean. Ideally, the physical state of a system at any time t should be determined by a unique point $(q(t), p(t))$, if the path $t \mapsto (q(t), p(t))$ satisfies the Hamilton equations of motion. However, it may well happen that at each time t, the state of the system can be specified by various, equivalent points of the phase space. In other words, although the state of the system at time t is uniquely defined once given a point $(q(t), p(t))$, the converse is not true, i.e. there is more than one set of values of the canonical variables representing the same physical state. Ideally, we would expect the equations of motions to fully determine the time evolution of physical states. However, we have seen that some parameters in the total Hamiltonian – those v^i associated with the primary first-class constraints $\varphi_i^{(1)}$ – are still unspecified. This implies that, given a physical state at time t_1, determined by a point $(q(t_1), p(t_1))$, the solution of the equations of motion corresponds to a path $t \mapsto (q(t), p(t))$ in phase space which depends on the value of the afore mentioned free parameters v^i, until a terminal state at time t_2. Although different such parameters induce different endpoints, we consider that a physical observable shall not depend on such arbitrary variation – *because* they are arbitrary. In other words, any ambiguity in the canonical variables at any time should be a physically irrelevant ambiguity. We are then led to define the following more general notion:

Definition 5.32. *A* gauge theory *is a physical theory in which the general solution to the equations of motion contain arbitrary functions of (space)-time. In that case an initial state gives rise to different time evolutions, and we call* gauge transformations *the infinitesimal transformations sending one such time evolution to another.*

We will now explore in what sense the primary first-class constraints generate (at least part of) the gauge transformations. Let f be a smooth function on T^*Q and let us compute two different time evolutions of f depending of two different sets of arbitrary parameters attached to the primary first-class constraints. Let $t_2 = t_1 + \delta t$, then by Lemma 5.19 and Equation (5.48) we have:

$$\begin{aligned} f(t_2) &= f(t_1) + \dot{f}(t_1)\delta t + \mathcal{O}((\delta t)^2) \\ &\approx f(t_1) + \{f, H_T\}(t_1)\delta t + \mathcal{O}((\delta t)^2) \\ &\approx f(t_1) + \{f, H'\}(t_1)\delta t + v^i\{f, \varphi_i^{(1)}\}(t_1)\delta t + \mathcal{O}((\delta t)^2) \end{aligned} \tag{5.49}$$

where the v^i are a set of unspecified smooth functions associated to the primary first-class constraints $\varphi_i^{(1)}$, possibly depending on time t. Now, if one takes up another set of parameters v'^i, the time evolution of f is now:

$$f(t_2) \approx f(t_1) + \{f, H'\}(t_1)\delta t + v'^i\{f, \varphi_i^{(1)}\}(t_1)\delta t + \mathcal{O}((\delta t)^2) \tag{5.50}$$

Notice that here, although we used the same notation, the value of $f(t_2)$ differ in both expressions (5.49) and (5.50) because the point $(q(t_2), p(t_2))$ is not the same in both cases. Notice that by construction, the first-class Hamiltonian H' is however the same in (5.49) and (5.50). Then they cancel out when we compute the difference between the two expressions of $f(t_2)$ and we have:

$$\delta f \approx -\delta\epsilon^i \{f, \varphi_i^{(1)}\}(t_1)\delta t + \mathcal{O}((\delta t)^2)$$

where $\delta\epsilon^i = (v'^i - v^i)\delta t$. Thus, the primary first-class constraints $\varphi_i^{(1)}$ are generators of local transformations with infinitesimal parameters $(v'^i - v^i)\delta t$. More precisely, denoting $X_{\varphi_i^{(1)}}$ the hamiltonian vector field associated to $\varphi_i^{(1)}$, one has:

$$\delta f \approx \delta\epsilon^i X_{\varphi_i^{(1)}}(f) + \mathcal{O}((\delta t)^2) \tag{5.51}$$

Since the parameters v^i and v'^i – and hence ϵ^i – are arbitrary, these transformations can be considered as legitimate gauge transformations. In other words, the primary first-class constraints generate (a subset of) gauge transformations.

We will now show that the primary first-class constraint cannot be the only set of functions generating gauge transformations. Indeed, starting from the former discussion, assume that between $t_2 = t_1 + \delta t$ and $t_3 = t_2 + \delta t$ we decide to apply $H' + v'^i\varphi_i^{(1)}$ to $f(t_2)$ as defined in (5.49) and $H' + v^i\varphi_i^{(1)}$ to $f(t_2)$ as defined in (5.50). Then we compare $f(t_3)$ obtained via the first path, and $f(t_3)$ obtained via the second path. Since we expect that any ambiguity in the canonical variables at any time should be a physically irrelevant ambiguity, we deduce that the difference between the two values of f at time t_3 is the result of a gauge transformation.

Theorem 5.33. *Beyond the primary first-class constraints, the set of generators of gauge transformations contains the following smooth functions:*

1. *the Poisson bracket of any two primary first-class constraints;*
2. *the Poisson bracket of any primary first-class constraint and the first-class Hamiltonian.*

Proof. We first apply $H' + v'^i\varphi_i^{(1)}$ to $f(t_2)$ as defined in (5.49) and we obtain:

$$\begin{aligned}
f(t_3) &= f(t_2) + \{f, H'\}(t_2)\delta t + v'^j\{f, \varphi_j^{(1)}\}(t_2)\delta t + \mathcal{O}((\delta t)^2) \\
&\approx f(t_1) + \{f, H'\}(t_1)\delta t + v^i\{f, \varphi_i^{(1)}\}(t_1)\delta t \\
&\qquad + \{f + \{f, H'\}\delta t + v^i\{f, \varphi_i^{(1)}\}\delta t, H'\}(t_1)\delta t \\
&\qquad\qquad + v'^j\{f + \{f, H'\}\delta t + v^i\{f, \varphi_i^{(1)}\}\delta t, \varphi_j^{(1)}\}(t_1)\delta t + \mathcal{O}((\delta t)^3) \\
&\approx f(t_1) + 2\{f, H'\}\delta t + v^i\{f, \varphi_i^{(1)}\}\delta t + v'^j\{f, \varphi_j^{(1)}\}\delta t + \{\{f, H'\}, H'\}(\delta t)^2 \\
&\qquad + \{f, \varphi_i^{(1)}\}\{v^i, H'\}(\delta t)^2 + v^i\{\{f, \varphi_i^{(1)}\}, H'\}(\delta t)^2 + v'^j\{\{f, H'\}, \varphi_j^{(1)}\}(\delta t)^2 \\
&\qquad\qquad + v'^j v^i\{\{f, \varphi_i^{(1)}\}, \varphi_j^{(1)}\}(\delta t)^2 + v'^j\{f, \varphi_i^{(1)}\}\{v^i, \varphi_j^{(1)}\}(\delta t)^2 + \mathcal{O}((\delta t)^3)
\end{aligned} \tag{5.52}$$

where the evaluation at time t_1 is implicit for each term. Then we apply $H' + v^i\varphi_i^{(1)}$ to $f(t_2)$ as defined in (5.50), and we obtain:

$$\begin{aligned}
f(t_3) &\approx f(t_1) + 2\{f, H'\}\delta t + v'^i\{f, \varphi_i^{(1)}\}\delta t + v^j\{f, \varphi_j^{(1)}\}\delta t + \{\{f, H'\}, H'\}(\delta t)^2 \\
&\qquad + \{f, \varphi_i^{(1)}\}\{v'^i, H'\}(\delta t)^2 + v'^i\{\{f, \varphi_i^{(1)}\}, H'\}(\delta t)^2 + v^j\{\{f, H'\}, \varphi_j^{(1)}\}(\delta t)^2 \\
&\qquad\qquad + v^j v'^i\{\{f, \varphi_i^{(1)}\}, \varphi_j^{(1)}\}(\delta t)^2 + v^j\{f, \varphi_i^{(1)}\}\{v'^i, \varphi_j^{(1)}\}(\delta t)^2 + \mathcal{O}((\delta t)^3)
\end{aligned} \tag{5.53}$$

where again the evaluation at time t_1 is implicit. Computing the difference between Equations (5.52) and (5.53) one obtains, after reordering the terms and noticing that their respective first line cancel each other:

$$
\begin{aligned}
\delta f \approx \Big(&\underline{\{f,\varphi_i^{(1)}\}\{v'^i,H'\}}\overline{-\{f,\varphi_i^{(1)}\}\{v^i,H'\}}\Big)(\delta t)^2 \\
&\underline{+\,v'^i\Big(\{\{f,\varphi_i^{(1)}\},H'\}-\{\{f,H'\},\varphi_i^{(1)}\}\Big)}(\delta t)^2 \\
&\overline{+\,v^j\Big(\{\{f,H'\},\varphi_j^{(1)}\}-\{\{f,\varphi_j^{(1)}\},H'\}\Big)}(\delta t)^2 \\
&+v^jv'^i\Big(\{\{f,\varphi_i^{(1)}\},\varphi_j^{(1)}\}-\{\{f,\varphi_j^{(1)}\},\varphi_i^{(1)}\}\Big)(\delta t)^2 \\
&+v^j\{f,\varphi_i^{(1)}\}\{v'^i,\varphi_j^{(1)}\}(\delta t)^2-v'^j\{f,\varphi_i^{(1)}\}\{v^i,\varphi_j^{(1)}\}+\mathcal{O}((\delta t)^3)
\end{aligned}
$$

Let us show that the two underlined terms combine to give the term $\{f,\{v'^i\varphi_i^{(1)},H'\}\}$ *on the constraint surface*. Then it is straightforward to antisymmetrize the computation and deduce that the two overlined terms give the term $-\{f,\{v^j\varphi_j^{(1)},H'\}\}$, again on Σ. Indeed, since the constraints vanish on Σ, we can rewrite the first term while the regarding the second we use the Jacobi identity for the Poisson bracket:

$$
\{f,\varphi_i^{(1)}\}\{v'^i,H'\}+v'^i\Big(\{\{f,\varphi_i^{(1)}\},H'\}-\{\{f,H'\},\varphi_i^{(1)}\}\Big)\approx\{f,\varphi_i^{(1)}\{v'^i,H'\}\}+v'^i\{f,\{\varphi_i^{(1)},H'\}\}
$$

The last term is weakly equivalent to the following one: $\{f,v'^i\{\varphi_i^{(1)},H'\}\}$, because H' being first-class, its bracket with any linear combination of constraints such as $\{\varphi_i^{(1)},H'\}$ vanishes on Σ. Thus, the sum $\{f,\varphi_i^{(1)}\{v'^i,H'\}\}+v'^i\{f,\{\varphi_i^{(1)},H'\}\}$ is weakly equivalent to $\{f,\{v'^i\varphi_i^{(1)},H'\}\}$ as desired.

Now, let us compute the sum $v^jv'^i\{\{f,\varphi_i^{(1)}\},\varphi_j^{(1)}\}+v^j\{f,\varphi_i^{(1)}\}\{v'^i,\varphi_j^{(1)}\}$, when it is restricted to the constraint surface. Start by factorizing out v^j and make v'^i enter the bracket so that the sum is weakly equivalent to:

$$
v^jv'^i\{\{f,\varphi_i^{(1)}\},\varphi_j^{(1)}\}+v^j\{f,\varphi_i^{(1)}\}\{v'^i,\varphi_j^{(1)}\}\approx v^j\{\{f,\varphi_i^{(1)}\}v'^i,\varphi_j^{(1)}\}
$$

Then the bracket $\{\{f,\varphi_i^{(1)}\}v'^i,\varphi_j^{(1)}\}$ is weakly equivalent to $\{\{f,\varphi_i^{(1)}v'^i\},\varphi_j^{(1)}\}$ because:

$$
\begin{aligned}
\{\{f,\varphi_i^{(1)}v'^i\},\varphi_j^{(1)}\}&=\{\{f,\varphi_i^{(1)}\}v'^i,\varphi_j^{(1)}\}+\{\varphi_i^{(1)}\{f,v'^i\},\varphi_j^{(1)}\}\\
&=\{\{f,\varphi_i^{(1)}\}v'^i,\varphi_j^{(1)}\}+\{\varphi_i^{(1)},\varphi_j^{(1)}\}\{f,v'^i\}+\varphi_i^{(1)}\{\{f,v'^i\},\varphi_j^{(1)}\}
\end{aligned}
$$

Both last terms vanish on the constraint surface; the middle one because the Poisson bracket of two first class constraints vanishes on Σ by definition. Then, it turns out that the term $v^j\{\{f,\varphi_i^{(1)}\}v'^i,\varphi_j^{(1)}\}$ is weakly equivalent to $v^j\{\{f,\varphi_i^{(1)}v'^i\},\varphi_j^{(1)}\}$, which in turn is weakly equivalent to $\{\{f,\varphi_i^{(1)}v'^i\},\varphi_j^{(1)}v^j\}$. By antisymmetry, we straightforwardly deduce that the sum $-v^jv'^i\{\{f,\varphi_j^{(1)}\},\varphi_i^{(1)}\}-v'^i\{f,\varphi_j^{(1)}\}\{v^j,\varphi_i^{(1)}\}$ is weakly equivalent to $-\{\{f,\varphi_j^{(1)}v^j\},\varphi_i^{(1)}v'^i\}$. By the Jacobi identity, the two terms combine and give:

$$
\{\{f,\varphi_i^{(1)}v'^i\},\varphi_j^{(1)}v^j\}-\{\{f,\varphi_j^{(1)}v^j\},\varphi_i^{(1)}v'^i\}=\{f,\{\varphi_i^{(1)}v'^i,\varphi_j^{(1)}v^j\}\}
$$

Gathering all the simplifications we obtained, we have the following weak equivalence, which characterize the gauge transformation applied to f:

$$
\delta f\approx\{f,\{v'^i\varphi_i^{(1)},H'\}\}(\delta t)^2-\{f,\{v^j\varphi_j^{(1)},H'\}\}(\delta t)^2+\{f,\{v'^i\varphi_i^{(1)},v^j\varphi_j^{(1)}\}\}(\delta t)^2+\mathcal{O}((\delta t)^3)
$$

where the evaluation at time t_1 is implicit. Then we see that the gauge transformation δf is generated by the three terms $\{v'^i\varphi_i^{(1)}, H'\}, \{v^j\varphi_j^{(1)}, H'\}$ and $\{v'^i\varphi_i^{(1)}, v^j\varphi_j^{(1)}\}$. Since the parameters v'^i and v^j are arbitrary, we have the result. □

Theorem 5.33 shows us that primary first-class constraints are not the only functions acting as generators of gauge transformations. The Poisson brackets $\{v'^i\varphi_i^{(1)}, H'\}, \{v^j\varphi_j^{(1)}, H'\}$ and $\{v'^i\varphi_i^{(1)}, v^j\varphi_j^{(1)}\}$ should indeed generate gauge transformations as well. Since by definition of the Bergmann-Dirac algorithm the bracket $\{\varphi_i^{(1)}, H'\} \approx \{\varphi_i^{(1)}, H\}$ is a second-stage constraint, we deduce that the two first brackets involve secondary constraints. Moreover, the third bracket $\{v'^i\varphi_i^{(1)}, v^j\varphi_j^{(1)}\}$ vanishes on Σ by definition of first-class constraints, hence it is strongly equivalent to a linear combination of primary and secondary constraints. Eventually, by Proposition 5.28 we know that the brackets $\{v'^i\varphi_i^{(1)}, H'\}, \{v^j\varphi_j^{(1)}, H'\}$ and $\{v'^i\varphi_i^{(1)}, v^j\varphi_j^{(1)}\}$ are first-class functions. Together with the above arguments, it implies that these brackets are not only strongly equivalent to linear combinations of primary and secondary constraints, but these constraints are all *first-class*.

Although it is in general not possible to infer from these observations alone that *every* secondary first-class constraint generates a gauge transformation, we will usually assume that it is the case (this is the Dirac conjecture, Scholie 5.34). Indeed, we have seen that the distinction between primary and secondary constraint is contingent because it heavily relies on the original choice of coordinates when we perform the Legendre transform. On the contrary, first-class and second-class constraints is a fundamental distinction, brought up by the Poisson structure of T^*Q. Additionally, first-class constraints form a Lie subalgebra of $\mathcal{C}^\infty(T^*Q)$ so they form an ideal candidate for generators of gauge transformations, but one has then to consider *all of them*. Moreover, we will see later that quantization methods put the first-class constraints on the same footings; there is no known quantization scheme if one does only consider part of them as gauge generators. For more details and further discussion, see the fruitful subsection 1.2.1 of [Henneaux and Teitelboim, 1994]. These observations led Dirac to formulate the following assumption:

Scholie 5.34. Dirac conjecture. *The generators of the gauge transformations are the first-class constraints, both primary* and *secondary.*

The status of the Dirac conjecture is debated. Its name first – a conjecture – would in general implicitly say that it has not yet been proven. However, fifty years of discussion have shown that the well-grounded character of this statement seems to mostly depend on its interpretation. For example, Henneaux and Teitelboim use the following Lagrangian defined on $\mathbb{R}^2$ (subsection 1.2.2 of [Henneaux and Teitelboim, 1994]):

$$L = \frac{1}{2}e^y v_x^2$$

The first constraint is a primary first-class constraint $\varphi^{(1)} = p_y$, and it induces a unique secondary first-class constraint $\varphi^{(2)} = \frac{1}{2}e^{-y}p_x^2$ which turns out to coincide with the Hamiltonian. There are no other constraints. Then, if one considers that the true secondary first-class constraint is $\widetilde{\varphi}^{(2)} = p_x$ – as Henneaux and Teitelboim did – one observes that it does not generate gauge transformations. These authors chose to pass from $\varphi^{(2)}$ to the mathematically equivalent constraint $\widetilde{\varphi}^{(2)}$ because they considered as 'true' constraints those that can serve as coordinate transverse to the secondary constraint surface.

However, Rothe and Rothe have shown (subsection 6.4 of [Rothe and Rothe, 2010]) that if one sticks to the secondary first-class constraint $\varphi^{(2)}$ then it generates a gauge transformation. They say that the ambiguity in the Dirac conjecture comes from an ambiguity in the interpretation of what is a 'true' first-class constraint, and that the validity of Dirac conjecture depends crucially on the chosen form for the constraints. The replacement of constraints by a formally equivalent set of constraints in fact may obliterate the full symmetry of the total action and will lose some important physical informations. Hence, this example shows that mathematically equivalent constraints may not be physically equivalent. Discussions about the Dirac conjecture have been vivid in the 1980s and the literature on the topic is rich [Gotay, 1983, Stefano, 1983, Cabo, 1986, Gràcia and Pons, 1988, Cabo and Louis-Martinez, 1990]. See also subsection 3.3.2 of [Henneaux and Teitelboim, 1994] for a proof of Dirac conjecture under mild assumptions. From now, we will stick to the modern view that the conjecture holds.

Now that we have determined all the generators of gauge transformations, we soon realize that the total Hamiltonian H_T does not contain every first-class constraints, and thus cannot generate all the gauge transformations. Thus we are led to adding the remaining first-class constraints to H_T to obtain a proper, more general Hamiltonian:

Definition 5.35. *Assume that there are p first-class constraints in total and let the v^i be arbitrary smooth parameters on the canonical coordinates (possibly depending on time also). Then we define the* extended Hamiltonian *as the following smooth function:*

$$H_E = H' + v^i \varphi_i$$

where H' is the first-class Hamiltonian.

Thus, the extended Hamiltonian contains the primary second-class constraints (hidden into H') and all the first-class constraints. When Hamilton equations involve the extended Hamiltonian, all the gauge transformations are allowed to be performed. However, a physical observable should not be depend on such gauge transformations. Hence the choice of Hamiltonian one picks up in the equations of motion – be it H', H_T or H_E – will not have any consequence on the smooth functions that are physically relevant, but will impact any other smooth function. Notice however that, while the total Hamiltonian was directly obtained from the Lagrangian formalism and would give back the Euler-Lagrange equations (see Proposition 5.24), the extended Hamiltonian is a new feature of the Hamiltonian formalism that does not have a Lagrangian counterpart (see subsection 1.2.3 in [Henneaux and Teitelboim, 1994]). The extended Hamiltonian allows to set all the first-class constraints on the same footing, which is a useful to apply canonical quantization. The difference between the total and the extended Hamiltonians are further explored in sections 3.2 and 3.3 of [Henneaux and Teitelboim, 1994], as well as section 5.4 of [Rothe and Rothe, 2010].

5.5 The geometry of the constraint surface

Let us now address the geometrical meaning of first-class and second-class constraints. Fix once and for all some functionally independent first-class and second-class constraints, and assume that the latter are minimal in number. Then the zero-level set of the second-class constraints is called the *second-class constraint manifold* and forms a *cosymplectic* submanifold Σ_0 of $(M, \{\,.\,,.\,\})$ (see subsection 4.3). It is assumed to be an embedded submanifold, and that the rank of the matrix D has constant rank over it (and not only on Σ). We then know that in a tubular neighborhood W of Σ_0 (or at least locally) one can define a Poisson bracket – called the *Dirac bracket*, see Equation (4.41) – so that Σ_0 is a symplectic leaf of $(W, \{\,.\,,.\,\}_{Dirac})$. In particular, we have shown that the Poisson bracket on Σ_0 induced by the Poisson-Dirac reduction coincides with the Dirac bracket (see Equation (4.47)). So the second-class manifold is a symplectic manifold and can be taken to be a replacement of the original phase space.

Then the first-class constraint define a submanifold of Σ_0, which turns out to be the constraint surface Σ. For simplicity in the following we will often assume that the constraints are defined globally over the entire phase space T^*Q, but remember that in full generality the results are only defined on a tubular neighborhood $W \subset T^*Q$ of Σ, or at least locally around each point:

Lemma 5.36. *The constraint surface Σ is a* coisotropic *submanifold of the second-class constraint manifold $(\Sigma_0, \{\,.\,,.\,\}_{Dirac})$. When the Dirac bracket is defined globally over T^*Q, then the constraint surface Σ is a* coisotropic *submanifold of $(T^*Q, \{\,.\,,.\,\}_{Dirac})$.*

Proof. Let us show the second point directly. Assuming that the constraints – both first-class *and* second-class – are globally defined, we have that the secondary constraint surface Σ is a closed embedded submanifold of T^*Q. By Proposition 4.89 it is sufficient to show that the ideal $\mathcal{I} = \mathrm{Span}(\varphi_j, \chi_k)$ of vanishing functions on Σ generated in $\mathcal{C}^\infty(T^*Q)$ by the first class *and* the second class constraints is a Lie subalgebra of $(\mathcal{C}^\infty(T^*Q), \{\,.\,,.\,\}_{Dirac})$. The definition of the Dirac bracket, Equation (4.41), has the following consequences:

1. the second-class constraints χ_l are Casimir elements of the Dirac bracket, so the Dirac bracket with any of them vanish, so in particular on Σ;

2. the first-class constraints φ_i are such that $\{\varphi_i, \mathcal{I}\}$ vanish on Σ by Definition 5.27, which implies that on Σ we have:

$$\{\varphi_i, .\,\}_{Dirac} = \{\varphi_i, .\,\} - \underbrace{\{\varphi_i, \chi_k\}}_{=0} C^{kl}\{\chi_l, .\,\} = \{\varphi_i, .\,\} \tag{5.54}$$

   This implies in turn that $\{\varphi_i, \varphi_j\}_{Dirac} = \{\varphi_i, \varphi_j\}$ on Σ which, by Definition 5.27, vanish on Σ.

These observations show that we have that the smooth functions belonging to the set $\{\mathcal{I}, \mathcal{I}\}_{Dirac}$ vanish on the secondary constraint surface Σ, i.e. they belong to $\mathcal{I}$. This proves that $\mathcal{I}$ is a subalgebra of $(\mathcal{C}^\infty(T^*Q), \{\,.\,,.\,\}_{Dirac})$. □

Remark 5.37. There exists an alternative way of getting rid of the second class constraints: instead of using the Poisson-Dirac reduction, one extends the phase space so that the second class constraints become first class. This is called the Fadeev-Jackiw reduction and is described in section 1.4.3 of [Henneaux and Teitelboim, 1994], in Section 4.4 of [Rothe and Rothe, 2010], and in the references there-in. Extending the phase space in such a way seems to correspond to solving the problem of coisotropic embedding of presymplectic manifolds into a bigger symplectic manifold [Gotay, 1982].

The proof of Lemma 5.36 has a very interesting consequence: Equation (5.54) shows that on the (secondary) constraint surface Σ, the hamiltonian vector fields of a first-class constraint φ_i – either computed with respect to the original Poisson bracket $\{\,.\,,.\,\}$ or the Dirac bracket – coincide. Indeed, on Σ, the properties of first class functions imply that we have $\{\varphi_i, .\,\} = \{\varphi_i, .\,\}_{Dirac}$ (see Equation (5.54)). There is then no ambiguity of talking about hamiltonian vector fields of first class constraints, when we restrict ourself to the constraint surface Σ. However for second-class constraints, it is another story because they form Casimir elements of the Dirac bracket. We then have the following important, geometric observation, with physical ramifications:

Proposition 5.38. *The hamiltonian vector fields X_{χ_l} associated to the second class constraints $\{\chi_l\}$ – and computed with respect to the original Poisson bracket $\{\,.\,,.\,\}$ – are nowhere tangent to the second-class constraint manifold Σ_0 (and hence to Σ).*

The hamiltonian vector fields X_{φ_i} associated to the first-class constraints $\{\varphi_i\}$ are everywhere tangent to the secondary constraint surface Σ and moreover induce a regular foliation on Σ.

Proof. For the first point, we will assume, as physicists do, that the non-triviality condition of second-class constraints is satisfied globally over the second-class constraint manifold Σ_0. The hamiltonian vector fields associated to the second class constraints are nowhere tangent to the second-class constraint manifold Σ_0 because for any second-class constraint χ_k and any point of Σ_0, there is at least one bracket $\{\chi_k, \chi_l\} = X_{\chi_k}(\chi_l)$ with another second-class constraint χ_l which does not vanish at this point. So the action of the Hamiltonian vector field X_{χ_k} on the ideal of vanishing functions on Σ_0 never lands in this ideal, so X_{χ_k} is not tangent to Σ_0, and hence to $\Sigma \subset \Sigma_0$.

On the contrary, by Definition of first-class functions 5.27, $\{\varphi_i, \mathcal{I}\} \subset \mathcal{I}$, so the hamiltonian vector fields X_{φ_i} are tangent to Σ. Now let us show the last item: if the set of first class-constraints is not irreducible, the regularity condition 5.23 implies that there are at least k independent first-class constraints which generate all the others. This is a local condition because even if the constraints are defined over the whole of T^*Q, their generators may change. So, locally, the set of first-class constraints is generated by k constraints. Let D be the smooth distribution generated by the hamiltonian vector fields X_{φ_i}. It has constant, finite rank k.

Now let us show that it is involutive. Since, moreover, Σ is a coisotropic submanifold of $(T^*Q, \{\,.\,,.\,\}_{Dirac})$, the multiplicative ideal $\mathcal{I}$ of vanishing functions on Σ is a subalgebra of $(\mathcal{C}^\infty(T^*Q), \{\,.\,,.\,\}_{Dirac})$, i.e. there exist smooth functions C_{ij}^k on T^*Q such that $\{\varphi_i, \varphi_j\}_{Dirac} = C_{ij}^k \varphi_k$. Then, by Equation (4.7) and Equation (4.41) we have:

$$[X_{\varphi_i}, X_{\varphi_j}] = X_{\{\varphi_i,\varphi_j\}} = X_{\{\varphi_i,\varphi_j\}_{Dirac}} + X_{\{\varphi_i,\chi_k\}C^{kl}\{\chi_l,\varphi_j\}} \tag{5.55}$$

The second term on the right-hand side reads:

$$X_{\{\varphi_i,\chi_k\}C^{kl}\{\chi_l,\varphi_j\}} = C^{kl}\{\chi_l, \varphi_j\}X_{\{\varphi_i,\chi_k\}} + \{\varphi_i, \chi_k\}X_{C^{kl}\{\chi_l,\varphi_j\}} \approx 0$$

It indeed vanishes on Σ because $\{\varphi_i, \chi_k\}$ and $\{\chi_l, \varphi_j\}$ vanish on Σ by definition of first-class functions. The first term on the right-hand side of Equation (5.55) reads:

$$X_{\{\varphi_i,\varphi_j\}_{Dirac}} = X_{C_{ij}^k\varphi_k} = \varphi_k X_{C_{ij}^k} + C_{ij}^k X_{\varphi_k} \approx C_{ij}^k X_{\varphi_k}$$

We then conclude that Equation (5.55) can be written on Σ as follows:

$$[X_{\varphi_i}, X_{\varphi_j}] \approx C_{ij}^k X_{\varphi_k}$$

Then, the regular distribution D is involutive on Σ (and a priori only on Σ). We say that the algebra of vector fields generated by the X_{φ_i} *close on-shell*. By Frobenius theorem, it is integrable and induces a regular foliation (on Σ). □

One can then proceed to Poisson reduction, at the condition that the leaf space corresponding to the foliation of Proposition 5.38 is a smooth manifold. The leaf space is then called the *reduced phase space* and is denoted Σ_{ph}, because it corresponds to the true non-gauge equivalent physical states of the system. When the leaf space if a smooth manifold, Proposition 4.94 applies and the Dirac bracket descends from Σ to Σ_{ph} by Poisson reduction, so that the resulting Poisson bracket is non-degenerate on Σ_{ph}. On the reduced phase space the equations of motions are the usual Hamilton's ones (see Appendix 2.A in [Henneaux and Teitelboim, 1994] to obtain more informations on the symplectic structure on Σ_{ph}). The smooth functions on Σ_{ph} are the physical observables, and thus we should characterize the space $\mathcal{C}^\infty(\Sigma_{ph})$ explicitly before quantizing the theory. Although it would seem desirable to work on the reduced phase space, it turns out that one often loses desirable features of the physical model such as Lorentz manifest invariance or, in the case of field theory, polynomiality of fields and locality in space. Moreover, it is often impossible to reformulate the theory in terms of gauge invariant quantities only and then to subsequently quantize it from the reduced phase space picture. It is thus often a better choice to carry along the dynamical variables and keep track of the first-class constraints, without using Poisson reduction, and then quantize the theory (see subsection 2.2.3 of [Henneaux and Teitelboim, 1994]). This is the object of BRST formalism, which provides an algebraic formulation of gauge invariant functions, i.e. physical observables.

Assume now that the constraints are globally defined, so that Σ is a closed embedded submanifold of T^*Q. The constraint surface is characterized by the ideal $\mathcal{I} = \mathrm{Span}(\varphi_j, \chi^k)$ of vanishing functions on Σ generated in $\mathcal{C}^\infty(T^*Q)$ by the first class *and* the second class constraints. Then, by Lemma 4.69 we have:

$$\mathcal{C}^\infty(\Sigma) \simeq \mathcal{C}^\infty(T^*Q)\Big/\mathcal{I}$$

By Lemma 5.36, Σ is a coisotropic submanifold of $(T^*Q, \{\,.\,,.\,\}_{Dirac})$ (or possibly only on a tubular neighborhood W of Σ). Indeed under the Dirac bracket the second class constraints behave as zero thus $\mathcal{I}$ is a Lie subalgebra of $\mathcal{C}^\infty(T^*Q)$, i.e. $\{\mathcal{I}, \mathcal{I}\}_{Dirac} \subset \mathcal{I}$. Let us now make sense of the Poisson reduction to Σ_{ph} in light of the knowledge we have on gauge transformations.

One may consider the set of gauge transformations as a family of infinitesimal transformations on $\mathcal{C}^\infty(T^*Q)$. Each gauge transformation is proportional to a (set of) parameters ϵ^i (which in subsection 5.4 corresponds to the difference $v'^i - v^i$ for example), where a priori i ranged from 1 to p, the number of first-class constraints, both primary and secondary (see Dirac conjecture, Scholie 5.34). We can then consider the family of parameters ϵ^i as the respective components of a smooth section ϵ of the trivial vector bundle $E = \mathbb{R}^p \times T^*Q$. We then denote the corresponding gauge transformation as $\delta_\epsilon : \mathcal{C}^\infty(T^*Q) \longrightarrow \mathcal{C}^\infty(T^*Q)$, sending any smooth function f to the following one:

$$\delta_\epsilon(f) = \epsilon^i X_{\varphi_i}(f) \tag{5.56}$$

where the X_{φ_i} are the Hamiltonian vector fields associated to the first-class constraints φ_i (we know that on Σ they do not depend if we picked up the Poisson bracket or the Dirac bracket to define them). Consequently, the vector fields are independent if the primary first-class constraints are irreducible, i.e. if the constraints are themselves independent. Moreover notice that the dependence in δt – which is explicit in Equation (5.51) – has been suppressed in Equation (5.56) because its role was only to emphasize the infinitesimal character of the transformations.

Thus, we have defined a $\mathcal{C}^\infty(T^*Q)$-linear map $\delta : \Gamma(E) \longrightarrow \mathfrak{X}(T^*Q)$, sending a section $\epsilon \in \Gamma(E)$ to the corresponding gauge generator. While the space $\mathfrak{X}(T^*Q)$ is a Lie algebra (of infinite dimension), it is a priori not the case for $\Gamma(E)$. This property would be nonetheless very natural, so that in the best case E would be a Lie algebroid over T^*Q such that δ would be its anchor map. However this situation often never happen. In order to distinguish which situations are met in our physical theories, we introduce the following nomenclature:

Definition 5.39. *The infinite dimensional space of gauge transformations $\mathscr{G} = \mathrm{Im}(\delta) \subset \mathfrak{X}(T^*Q)$ is abusively called the* algebra of gauge transformations. *We say that the algebra of gauge transformations is* closed *if $\mathscr{G}$ is a Lie subalgebra of the Lie algebra of vector fields $(\mathfrak{X}(T^*Q), [\,.\,,.\,])$. It is said* open *otherwise. We say that the algebra of gauge transformations is* closed on-shell *if:*

$$[\mathrm{Im}(\delta), \mathrm{Im}(\delta)]\big|_\Sigma \subset \mathrm{Im}(\delta)\big|_\Sigma$$

The last denomination can be equivalently stated as follows: the algebra is closed *on-shell* if for every smooth sections $\epsilon, \eta \in \Gamma(E)$, the Lie bracket of their image via δ – which is a genuine vector field on T^*Q – turns out to be weakly equivalent to a gauge transformation in the following sense:

$$[\delta_\epsilon, \delta_\eta](f) \approx \delta_\rho(f)$$

for any smooth function $f \in \mathcal{C}^\infty(T^*Q)$ and some $\rho \in \Gamma(E)$. In that case, if the map δ is injective – so that $\mathscr{G}$ and $\Gamma(E)$ are isomorphic – then one may induce a Lie bracket on $\Gamma(E)$ via the identification $[\epsilon, \eta] = \rho$. The algebra of gauge transformation is rarely closed, but it turns out that:

Lemma 5.40. *The algebra of gauge transformations is closed on-shell.*

Proof. Indeed, Proposition 5.38 tells us that the Hamiltonian vector fields associated to the first-class constraints define an involutive distribution D on Σ. Since by construction (see Equation (5.56)) $\delta_\epsilon|_\Sigma$ and $\delta_\eta|_\Sigma$ take values in D, we deduce that the Lie bracket $[\delta_\epsilon, \delta_\eta]|_\Sigma$ takes values in D, i.e. it can be decomposed on the basis of vectors X_{φ_i}, with coordinates ρ^i. This proves that $[\delta_\epsilon, \delta_\eta]|_\Sigma = \delta_\rho|_\Sigma$ for $\rho = \rho^i e_i$, where $e_1, \dots, e_p$ is the canonical (global) frame of E. $\square$

Since the tangent spaces to the leaves of the foliation integrating the distribution D on Σ are generated by the hamiltonian vector fields X_{φ_j} associated to the first-class constraint φ_j, we deduce that the leaves of the regular foliation characterize gauge equivalent physical states. In other words, two points on the same leaf – this is a geometric equivalence relation – are 'gauge equivalent' in the sense that any physical observable $O \in \mathcal{C}^\infty(\Sigma)$ should take the same value in these two points. Physical observables are said to be the *gauge invariant functions on* Σ and, in the geometric picture, they correspond to those functions being constant *along* each leaf of the foliation. Then they are invariant with respect to the vector fields tangent to the leaves:

$$X_{\varphi_i}(O) = 0$$

for every first-class constraints φ_i, both primary *and* secondary.

The gauge invariant functions on Σ are then constant along the the leaves of the foliation induced by the vector fields X_{φ_i}. As we assume that the leaf space Σ_{ph} is a smooth manifold, we deduce that the gauge invariant functions on the constraint surface Σ pass to the quotient $\Sigma \twoheadrightarrow \Sigma_{ph}$ and define smooth functions on Σ_{ph}. Conversely, any smooth function on Σ_{ph} – the putative physical observables – induce a smooth function on Σ which is constant along the leaves of the foliation, i.e. gauge invariant. Then we conclude that, as expected, there is a one to one correspondence between physical observables (on Σ_{ph}) and gauge invariant smooth functions (on Σ). Moreover, the properties of Poisson reduction, and in particular the fact that Proposition 4.94 applies to the current situation, shows that taking the Dirac bracket of any (global) extensions of two gauge invariant functions on Σ is equivalent to taking the reduced Poisson bracket of the corresponding physical observables on Σ_{ph}.

Definition 5.41. *By abuse of denomination, we will call* gauge-invariant function *any smooth function $f \in \mathcal{C}^\infty(T^*Q)$ (or possibly only on a tubular neighborhood W of Σ) such that:*

$$X_{\varphi_j}(f) \approx 0 \tag{5.57}$$

for every first class constraint φ_j.

Any such gauge invariant function, when restricted to the constraint surface Σ, is a proper gauge invariant function $O = f|_\Sigma \in \mathcal{C}^\infty(\Sigma)$, and any such latter function admits a (possibly global) extension satisfying the assumption Definition 5.41. We would now characterize the space of physical observables by using such extended notion of gauge invariant function. It is indeed easier to work on T^*Q as we do not work on a quotient, as would be the case if one worked with $\mathcal{C}^\infty(\Sigma)$. However, notice that while there was a one-to-one correspondence between physical observables – i.e. smooth functions on Σ_{ph} – and gauge invariant functions on Σ, there are much more gauge-invariant functions $f \in \mathcal{C}^\infty(T^*Q)$ as characterized in Definition 5.41. The latter are however quite useful to characterize the space of function $\mathcal{C}^\infty(\Sigma_{ph})$ because they are defined over the whole phase space T^*Q, which has a more regular smooth structure than the constraint surface Σ.

Condition (5.57) equivalently means that $X_f(\mathcal{I}) \approx 0$, where here the Hamiltonian vector field is computed with respect to the Dirac bracket. Since these constraints span the ideal $\mathcal{I}$ of functions vanishing on Σ and since Σ is a (closed) embedded submanifold of T^*Q, it implies that the hamiltonian vector field of f (with respect to the Dirac bracket) is tangent to Σ (see beginning of subsection 3.4). Hence, the smooth functions $f \in \mathcal{C}^\infty(T^*Q)$ inducing gauge invariant functions on Σ or, equivalently, physical observables, are precisely those smooth functions $f \in \mathcal{C}^\infty(T^*Q)$ such that $\{f, \mathcal{I}\}_{Dirac} \approx 0$. The maximal subalgebra of $\mathcal{C}^\infty(T^*Q)$ generated by such functions and containing the ideal $\mathcal{I}$ is denoted $\mathcal{N}$. Then, physical observables can be identified with the quotient of the latter by the former:

$$\mathcal{C}^\infty(\Sigma_{ph}) \simeq \mathcal{N}\big/\mathcal{I}$$

The reduced phase space Σ_{ph} corresponds to the classical unconstrained picture in Hamiltonian mechanics; it is then open to quantization. However it is difficult to quantize directly the theory on the reduced phase space. On the contrary, Dirac tried to quantize the theory by quantizing the canonical coordinates on the total phase space T^*Q – which is a well known procedure – together with the constraints. This is necessary to keep track of which function of q and p are physical observables or not. We will see that canonical quantization has several obstructions and *no-go theorems*. That is why other strategies, such as the BRST quantization scheme, tried to make sense of the dynamics on the reduced phase space using an algebraic picture. It avoids the complications due to the quotienting and provides a tractable model of $\mathcal{C}^\infty(\Sigma_{ph})$.

5.6 More about generating functions and the Legendre transform

In this section we will provide some material non-related to Bergmann-Dirac formalism but which however is relevant for those interested into the geometrization of classical mechanics. The tangent bundle of the cotangent bundle $T(T^*Q)$ is a $4n$-dimensional manifold and a rank $2n$ vector bundle over T^*Q. Over a local trivializing open set $U \subset M$, the cotangent bundle T^*Q can be trivialized as the product of U (with coordinate functions q^i) with the fiber $\mathbb{R}^n$ (with coordinate functions p_i), so that $T(T^*Q)$ is locally isomorphic to $TU \times T\mathbb{R}^n$. The same holds for the cotangent bundle of the cotangent bundle $T^*(T^*Q)$. A differential form on T^*Q is a section of $\wedge^\bullet T^*(T^*Q)$; since T^*Q can be locally seen as the product of a trivializing open set U and the fiber $\mathbb{R}^n$, one understands that a differential form on T^*Q is locally generated by products of covector fields dq^i on Q with covector fields dp_i on the fiber. Here the de Rham differential is the de Rham differential on T^*Q, so that dq^i should indeed be seen as a constant section of $T^*(T^*Q)$, although its action is trivial on the fiber. With these conventions in hand, we observe that the cotangent bundle T^*Q is a symplectic manifold: there exists a closed non-degenerate 2-form $\omega \in \Omega^2(T^*Q)$ called the *Poincaré 2-form*. This two form is canonical (see below) and can be written in local coordinates q^i, p_i as: $\omega = \sum_{i=1}^n dp_i \wedge dq^i$. The Poincaré 2-form has the particularity of being exact and this represents a defining feature: it is the de Rham derivative $\omega = d\theta$ of the so-called *Liouville-Poincaré* – or *tautological* – *one-form*. This differential form is the unique globally defined one-form whose expression in local coordinates is $\theta = \sum_{i=1}^{n} p_i dq^i$. It admits a coordinate-free definition that we will now provide.

Definition 5.42. *Let $\pi_Q : T^*Q \longrightarrow M$ be the projection associated to the bundle structure on the cotangent bundle of Q. Then, the* tautological one form *is the unique differential one-form on T^*Q defined pointwise as follow:*

$$\text{for every } (q,p) \in T^*Q \qquad\qquad \theta_{(q,p)} = p \circ (\pi_M)_*|_{(q,p)}$$

where $(\pi_Q)_*|_{(q,p)} : T_{(q,p)}(T^*Q) \longrightarrow T_qQ$ is the push-forward of π_M at the point (q,p).

The tautological one-form is the unique one-form that 'cancels' pullback: any differential one-form $\sigma \in \Omega^1(M)$ can be seen as a smooth section $\sigma : M \longrightarrow T^*Q$. Then, the push-forward of σ is a vector bundle map $\sigma_* : TQ \longrightarrow T(T^*Q)$, as well as the pull-back $\sigma^* : T^*(T^*Q) \longrightarrow T^*Q$. Then, pulling back a differential one form on T^*Q via σ gives a differential one form on Q. Then, the tautological one-form is the unique one-form on T^*Q such that:

$$\sigma^*(\theta) = \sigma$$

Another, alternative perspective is the following: the tautological one-form is the only differential one-form θ on T^*Q such that:

$$\theta = \pi_Q^* \circ \pi_{T^*Q}(\theta)$$

where $\pi_Q^* : T^*Q \longrightarrow T^*(T^*Q)$ is the pull-back of π_M while $\pi_{T^*Q} : T^*(T^*Q) \longrightarrow T^*Q$ is the projection associated to the bundle structure on the cotangent bundle of T^*Q. The smooth manifold $T^*(T^*Q)$ is called a *double vector bundle* and the tautological one-form is used to characterize the Legendre transform between TQ and T^*Q [Tulczyjew, 1977, Yoshimura and Marsden, 2007].

$$\begin{tikzcd}
\theta \in T^*(T^*Q) \arrow[rr, "\mathrm{id}"] \arrow[dr, "\pi_{T^*Q}"'] & & \theta \in T^*(T^*Q) \\
& T^*Q \arrow[ur, "\pi_M^*"'] &
\end{tikzcd}$$

References

[Agrachev et al., 2019] Agrachev, A., Barilari, D., and Boscain, U. (2019). *A Comprehensive Introduction to Sub-Riemannian Geometry*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[Baez and Muniain, 1994] Baez, J. and Muniain, J. P. (1994). *Gauge Fields, Knots and Gravity*, volume 4 of *Series on Knots and Everything*. World Scientific, Singapore.

[Bamberg and Sternberg, 1988] Bamberg, P. and Sternberg, S. (1988). *A Course in Mathematics for Students of Physics*, volume 2. Cambridge University Press, Cambridge.

[Bergmann, 1949] Bergmann, P. G. (1949). Non-Linear Field Theories. *Phys. Rev.*, 75(4):680–685.

[Bergmann and Goldberg, 1955] Bergmann, P. G. and Goldberg, I. (1955). Dirac Bracket Transformations in Phase Space. *Phys. Rev.*, 98(2):531–538.

[Cabo, 1986] Cabo, A. (1986). On Dirac's conjecture for systems having only first-class constraints. *J. Phys. A: Math. Gen.*, 19(5):629–638.

[Cabo and Louis-Martinez, 1990] Cabo, A. and Louis-Martinez, D. (1990). On Dirac's conjecture for Hamiltonian systems with first- and second-class constraints. *Phys. Rev. D*, 42(8):2726–2735.

[Calvo et al., 2010] Calvo, I., Falceto, F., and Zambon, M. (2010). Deformation of Dirac structures along isotropic subbundles. *Reports on Mathematical Physics*, 65(2):259–269.

[Candel and Conlon, 2000] Candel, A. and Conlon, L. (2000). *Foliations I*. Number 23 in Graduate Studies in Mathematics. American Mathematical Society, Providence.

[Crainic and Fernandes, 2004] Crainic, M. and Fernandes, R. L. (2004). Integrability of Poisson Brackets. *Journal of Differential Geometry*, 66(1):71–137.

[Crainic et al., 2021] Crainic, M., Fernandes, R. L., and Marcut, I. (2021). *Lectures on Poisson Geometry*. Number 217 in Graduate Studies in Mathematics. American Mathematical Society, Providence.

[Dirac, 1950] Dirac, P. a. M. (1950). Generalized Hamiltonian Dynamics. *Canadian Journal of Mathematics*, 2:129–148.

[Dirac, 1964] Dirac, P. A. M. (1964). *Lectures on Quantum Mechanics*. Number 2 in Monographs Series. Belfer Graduate School of Science, Yeshiva University, New York.

[Dufour and Zung, 2005] Dufour, J.-P. and Zung, N. T. (2005). *Poisson Structures and Their Normal Forms*. Number 242 in Progress in Mathematics. Birkhäuser, Basel.

[Falceto and Zambon, 2008] Falceto, F. and Zambon, M. (2008). An Extension of the Marsden–Ratiu Reduction for Poisson Manifolds. *Lett Math Phys*, 85(2):203–219.

[Fernandes, 2005] Fernandes, R. L. (2005). A Note on Proper Poisson Actions. *arXiv:math/0503147*.

[Gitman and Tyutin, 1990] Gitman, D. M. and Tyutin, I. V. (1990). *Quantization of Fields with Constraints*. Springer Series in Nuclear and Particle Physics. Springer-Verlag, Berlin Heidelberg.

[Gotay, 1982] Gotay, M. J. (1982). On coisotropic imbeddings of presymplectic manifolds. *Proc. Amer. Math. Soc.*, 84(1):111–114.

[Gotay, 1983] Gotay, M. J. (1983). On the validity of Dirac's conjecture regarding first-class secondary constraints. *J. Phys. A: Math. Gen.*, 16(5):L141–L145.

[Gràcia and Pons, 1988] Gràcia, X. and Pons, J. M. (1988). Gauge generators, Dirac's conjecture, and degrees of freedom for constrained systems. *Annals of Physics*, 187(2):355–368.

[Henneaux and Teitelboim, 1994] Henneaux, M. and Teitelboim, C. (1994). *Quantization of Gauge Systems.* Princeton University Press, Princeton.

[Laurent-Gengoux et al., 2013] Laurent-Gengoux, C., Pichereau, A., and Vanhaecke, P. (2013). *Poisson Structures.* Grundlehren Der Mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg.

[Lee, 2003] Lee, J. M. (2003). *Introduction to Smooth Manifolds.* Graduate Texts in Mathematics. Springer-Verlag, New York.

[Lee, 2009] Lee, J. M. (2009). *Manifolds and Differential Geometry.* Number 107 in Graduate Studies in Mathematics. American Mathematical Society, Providence.

[Matschull, 1996] Matschull, H.-J. (1996). Dirac's Canonical Quantization Programme. *arXiv:quant-ph/9606031.*

[Moerdijk and Mrcun, 2003] Moerdijk, I. and Mrcun, J. (2003). *Introduction to Foliations and Lie Groupoids.* Number 91 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[Nestruev, 2003] Nestruev, J. (2003). *Smooth Manifolds and Observables.* Graduate Texts in Mathematics. Springer-Verlag, New York.

[Rosenfeld, 1932] Rosenfeld, L. (1932). La théorie quantique des champs. *Annales de l'institut Henri Poincaré*, 2(1):25–91.

[Rothe and Rothe, 2010] Rothe, H. J. and Rothe, K. D. (2010). *Classical and Quantum Dynamics of Constrained Hamiltonian Systems*, volume 81 of *World Scientific Lecture Notes in Physics.* World Scientific.

[Schmidt, 1999] Schmidt, H.-J. (1999). The Classical Solutions of Two-dimensional Gravity. *General Relativity and Gravitation*, 31(8):1187–1210.

[Sears, 1963] Sears, F. W. (1963). A Simplified Simplification of Carathéodory's Treatment of Thermodynamics. *American Journal of Physics*, 31(10):747–752.

[Sears, 1966] Sears, F. W. (1966). Modified Form of Carathéodory's Second Axiom. *American Journal of Physics*, 34(8):665–666.

[Stefan, 1974] Stefan, P. (1974). Accessible Sets, Orbits, and Foliations with Singularities. *Proc. London Math. Soc.*, s3-29(4):699–713.

[Stefano, 1983] Stefano, R. D. (1983). Modification of Dirac's method of Hamiltonian analysis for constrained systems. *Phys. Rev. D*, 27(8):1752–1765.

[Sudarshan and Mukunda, 2015] Sudarshan, E. C. G. and Mukunda, N. ([1974] 2015). *Classical Dynamics: A Modern Perspective.* World Scientific.

[Tulczyjew, 1977] Tulczyjew, W. M. (1977). The Legendre transformation. *Annales de l'I.H.P. Physique théorique*, 27(1):101–114.

[Vaintrob, 1997] Vaintrob, A. Y. (1997). Lie algebroids and homological vector fields. *Russ. Math. Surv.*, 52(2):428.

[Vaisman, 1994] Vaisman, I. (1994). *Lectures on the Geometry of Poisson Manifolds*. Progress in Mathematics. Birkhäuser Basel.

[Warner, 1983] Warner, F. W. (1983). *Foundations of Differentiable Manifolds and Lie Groups*. Graduate Texts in Mathematics. Springer-Verlag, New York.

[Weinstein, 1988] Weinstein, A. (1988). Coisotropic calculus and Poisson groupoids. *Journal of the Mathematical Society of Japan*, 40(4):705–727.

[Yoshimura and Marsden, 2007] Yoshimura, H. and Marsden, J. E. (2007). Dirac Structures and the Legendre Transformation for Implicit Lagrangian and Hamiltonian Systems. In Allgüwer, F., Fleming, P., Kokotovic, P., Kurzhanski, A., Kwakernaak, H., Rantzer, A., Tsitsiklis, J., Bullo, F., and Fujimoto, K., editors, *Lagrangian and Hamiltonian Methods for Nonlinear Control 2006*, Lecture Notes in Control and Information Sciences, pages 233–247, Berlin, Heidelberg. Springer.

[Zambon, 2011] Zambon, M. (2011). Submanifolds in Poisson geometry: A survey. In Ebeling, W., Hulek, K., and Smoczyk, K., editors, *Complex and Differential Geometry*, Springer Proceedings in Mathematics, pages 403–420, Berlin, Heidelberg. Springer.

[Zemansky, 1966] Zemansky, M. W. (1966). Kelvin and Caratheodory—A Reconciliation. *American Journal of Physics*, 34(10):914–920.