SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Hybrid Feedback Control on Manifolds

av

Lars Lidvall

2026 - No M7

Hybrid Feedback Control on Manifolds

Lars Lidvall

Självständigt arbete i matematik 30 högskolepoäng, avancerad nivå

Handledare: Boris Shapiro

2026

Abstract

In this report we introduce the concept of hybrid feedback control and apply it to smooth manifolds. The problem of obtaining robust global asymptotic stability is desirable for applications, but it is not solvable with standard control techniques for not contractable manifolds. Compact manifolds are not contractable, but the smooth compact connected manifolds SO(2) and SO(3) appear in many autonomous vehicles problems. We will solve the problem of designing a robust global asymptotically stable controller for smooth compact connected manifolds in the framework of hybrid feedback control. This result will be applied to products of SO(2) and SO(3).

Sammanfattning

I denna rapport introducerar vi konceptet hybrid återkopplingsreglering och tillämpar det på glatta mångfalder. Att uppnå robust global asymptotisk stabilitet är önskevärt för tillämpningar, men det är omöjligt att uppnå med vanlig kontroll för icke-kontrakterbara mångfalder. Kompakta mångfalder är inte kontrakterbara, men de glatta kompakta sammanhängande mångfalderna SO(2) och SO(3) förekommer i många problem som har med autonoma fordon att göra. Vi kommer att formulera en robust globalt asymptotiskt stabil regulator för glatta kompakta sammanhängande mångfalder inom ramverket för hybrid återkopplingsreglering. Detta resultat kommer att tillämpas på produkter av SO(2) och SO(3).

Acknowledgements

I would like to thank my supervisor Boris for his support, as well as my family and friends that have made my life better outside the development of this thesis. Especially SLAM, Betty, Adam, and Tara.

Contents

Chapter 1

Introduction

Control theory is concerned with steering dynamical systems toward desired states. A fundamental goal is to achieve *global asymptotic stability*, which is to ensure that a system converges to a target state from any initial condition, and remains remains close to the target if once reached close enough. For systems defined on Euclidean space, this is a well-understood theory. Indeed, you can construct a system that converges to an given point x^* by considering a function that has x^* as its global minimum, and has no other critical points, for example $\|x - x^*\|^2$, and then flow towards the minimum using gradient flow.

In applications, you might consider the orientation of a rigid body in two or three dimensions. These form the spaces $\mathrm{SO}(2) \cong \mathbb{S}^1$ and $\mathrm{SO}(3) \cong \mathbb{R}P^3$, which are compact smooth manifolds. The same argument does not work in this case, since on a compact manifold, any continuous function must both have a minimum and a maximum. Therefore there must be at least two critical points, where gradient flow becomes stuck if you start at the maximum. For $\mathrm{SO}(3)$ there must be at least 4 critical points, consisting of one global minimum, one global maximum, and two saddle points. Even if you do not start at a critical point, now gradient flow can flow your state towards a saddle point if you are unlucky, and then become stuck.

This thesis addresses this obstruction using *hybrid feedback control*, which combines continuous flow dynamics with discrete jumps. The central result, following the framework of Montgomery and Sanfelice [MS24], is that for any compact connected smooth manifold and any target point, one can design a hybrid controller achieving robust global asymptotic stability. The idea is to use a Morse function which has the target point as its global minimum, then flow along gradient flow and measure if the state is close to a critical point that is not the global minimum. If the measurement reads that the state is close to such a critical point, the flow is switched from gradient flow to a *breeze vector field*, which pushes the state away far enough that gradient flow will not flow towards that point again.

The report is organized as follows. Chapter 2 develops the necessary background for topology and smooth manifolds. Chapter 3 covers Morse theory, including the Morse lemma, the Morse number, and perfect Morse functions. Chapter 4 introduces hybrid dynamical systems and hybrid feedback control. Chapter 5 proves the main theorem and applies it to products of $\mathrm{SO}(2)$ and $\mathrm{SO}(3)$.

Chapter 2

Topological Spaces and Smooth Manifolds

2.1 Topological spaces

2.1.1 Open sets

Definition 2.1.1 (Topological space)**.** A *topological space* is a pair $(X, \mathcal{O})$ where X is an arbitrary set and $\mathcal{O}$ is a collection of subsets of X fulfilling the following criteria.

- The basic subsets $\varnothing$ and X are elements of $\mathcal{O}$. That is, $\varnothing, X \in \mathcal{O}$.
- If you are given an arbitrary collection of elements of $\mathcal{O}$, then the union must be an element of $\mathcal{O}$. That is, given sets $U_i \in \mathcal{O}$, $i \in I$ then $\bigcup_{i\in I} U_i \in \mathcal{O}$.
- If you are given *finitely* many elements of $\mathcal{O}$, then the intersection must also be an element of $\mathcal{O}$. That is, given $U_1, \ldots, U_m \in \mathcal{O}$ then $\bigcap_{i=1}^{m} U_i \in \mathcal{O}$.

Elements of $\mathcal{O}$ are called *open* sets of $(X, \mathcal{O})$, and $\mathcal{O}$ is called a *topology* on X.

Any subset of a topological space inherits a "natural" topology, called the *subspace topology.*

Definition 2.1.2 (Subspace topology)**.** Given a topological space $(X, \mathcal{O})$ and a subset $S \subseteq X$, the collection $\mathcal{O}_{\text{subset}} = \{U \cap S : U \in \mathcal{O}\}$ is called the subspace topology on S.

Lemma 2.1.1. *Given a topological space $(X, \mathcal{O})$ and a subset $S \subseteq X$ then $(S, \mathcal{O}_{subset})$ is a topological space.*

Remark. The topological space $(S, \mathcal{O}_{\text{subset}})$ may then be called a subspace of $(X, \mathcal{O})$.

Proof. Because $X, \varnothing \in \mathcal{O}$ then $S = X \cap S$ and $\varnothing = \varnothing \cap S$ are elements of $\mathcal{O}_{\text{subset}}$. Moreover, given $U_i \in \mathcal{O}_{\text{subset}}$ we know that $U_i = V_i \cap S$ for some $V_i \in \mathcal{O}$, and therefore $\bigcup_i U_i = \bigcup_i (V_i \cap S) = (\bigcup_i V_i) \cap S$. Since each $V_i \in \mathcal{O}$, we have $\bigcup_i V_i \in \mathcal{O}$, and therefore $(\bigcup_i V_i) \cap S \in \mathcal{O}_{\text{subset}}$. Similarly $\cap_i U_i = \cap_i (V_i \cap S) = (\cap_i V_i) \cap S$, so $\cap_i U_i \in \mathcal{O}_{\text{subset}}$. □

Example 1 (Two point space)**.** $X = \{1, 2\}$ can be equipped with several topological structures. For example the *trivial space* $(X, \mathcal{O}_{\text{trivial}})$ where $\mathcal{O}_{\text{trivial}} = \{\varnothing, \{1, 2\}\}$ contains as few open sets as possible; the *discrete space* $(X, \mathcal{O}_{\text{discrete}})$ where $\mathcal{O}_{\text{discrete}} = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}$ contains as many open sets as possible; and the *particular point space* $(X, \mathcal{O}_{\text{particular}})$ where $\mathcal{O}_{\text{particular}} = \{\varnothing, \{1\}, \{1, 2\}\}$ is something in between (here 1 is the "particular point").

One can check that the three collections $\mathcal{O}_{\text{trivial}}$, $\mathcal{O}_{\text{discrete}}$, and $\mathcal{O}_{\text{particular}}$ indeed fulfill the three criteria of Definition 2.1.1 for $X = \{1, 2\}$, and they give different topologies. It is important clarify which topology one is talking about. But in practice there is typically a "canonical" or "natural" topology that a set can be endowed with, and then instead of denoting the topological space by $(X, \mathcal{O}_{\text{natural}})$ one usually just writes the set X where the topology clear from the context. The following example is of such a flavor, but we will be explicit for clarity.

Example 2 (Euclidean space $\mathbb{E}^m$)**.** $X = \mathbb{R}^m$ with $\mathcal{O} = \mathcal{O}_{\text{Euclidean}}$ being the Euclidean topology on $\mathbb{R}^m$. This makes the m-dimensional Euclidean space $\mathbb{E}^m = (\mathbb{R}^m, \mathcal{O}_{\text{Euclidean}})$, where $\mathcal{O}_{\text{Euclidean}}$ is *generated* by (open) Euclidean balls $B_r(c) = \{x \in \mathbb{R}^m : d(c, x) < r\}$. Here $c \in \mathbb{R}^m$, $r \in \mathbb{R}_{>0}$ and d is the Euclidean distance function $d(c, x) = \|c - x\| = \sqrt{(c_1 - x_1)^2 + \cdots + (c_m - x_m)^2}$.

Remark. "Generated" means that we begin with this collection of sets and use them to make a topology by including every union and finite intersection of these sets in $\mathcal{O}$. A better term is that the Euclidean balls form a *basis* for the Euclidean topology.

Definition 2.1.3 (Basis)**.** Given a topological space $(X, \mathcal{O})$, a collection $\mathcal{B} \subseteq \mathcal{O}$ of open sets forms a *basis* for the topology $\mathcal{O}$ if and only if every open set is a union of open sets in $\mathcal{B}$. That is, for every $U \in \mathcal{O}$ there exists open sets $B_i \in \mathcal{B}$, $i \in I$ such that $U = \bigcup_{i \in I} B_i$

Remark. Observe that every topological space has a basis, since we can always take $\mathcal{B} = \mathcal{O}$. What is interesting is to see how small a basis $\mathcal{B}$ can be made.

Example 3 (Euclidean bases)**.** $\mathbb{E}^m = (\mathbb{R}^m, \mathcal{O}_{\text{Euclidean}})$ has the basis $\mathcal{B} = \{B_r(c) : c \in \mathbb{R}^m,\ r \in \mathbb{R}_{>0}\}$, but we can make the basis smaller by only considering "rational balls" (rational centers and radii) $\mathcal{B}_{\text{rational}} = \{B_r(c) : c \in \mathbb{Q}^m,\ r \in \mathbb{Q}_{>0}\}$. This is indeed a basis because $\mathbb{Q}^m$ is dense in $\mathbb{R}^m$, so for any open set $U \in \mathcal{O}_{\text{Euclidean}}$, we may write U as the union of all rational balls in U. This basis $\mathcal{B}_{\text{rational}}$ is smaller than $\mathcal{B}$ in the significant way that $\mathcal{B}$ contains uncountably many open sets, but $\mathcal{B}_{\text{rational}}$ contains only countably many.

Definition 2.1.4 (Product space)**.** Let $(X_1, \mathcal{O}_1), \ldots, (X_n, \mathcal{O}_n)$ be topological spaces. Then $(X_1 \times \cdots \times X_n, \mathcal{O})$ is a topological space with basis $\mathcal{B} = \{U_1 \times \cdots \times U_n : U_i \in \mathcal{O}_i\}$.

Example 4 (Euclidean space as a product)**.** $\mathbb{E}^m$ is the product space of m copies of $\mathbb{E}^1$.

2.1.2 Properties of topological spaces

We now introduce several properties that a topological space can have. These properties will later be used to define manifolds and to establish results about functions on them.

The smallest possible size of a basis for a topological space becomes a way of quantifying how many open sets the topology contains. This motivates the following definition.

Definition 2.1.5 (Second countable)**.** A topological space $(X, \mathcal{O})$ is called *second countable* if and only if it admits a basis which only contains at most countably many open sets.

Lemma 2.1.2. *Euclidean space $\mathbb{E}^m$ is a second countable topological space.*

Proof. See Example 3. □

However, there are also topologies in which there are in some sense too few open sets. It turns out that having enough open sets such that any two points can be separated into two non-overlapping open sets is useful, and this property is called Hausdorff.

Definition 2.1.6 (Hausdorff property)**.** A topological space $(X, \mathcal{O})$ is called *Hausdorff* if and only if for any two distinct points, there exists two non-overlapping open sets containing one of the points each. That is, for every $x_1, x_2, \in X$, $x_1 \neq x_2$ there exists $U_1, U_2 \in \mathcal{O}$ such that $x_1 \in U_1$, $x_2 \in U_2$ and $U_1 \cap U_2 = \varnothing$.

Remark. One can say that in Hausdorff spaces points can always be separated into different open sets.

Nonexample 5. The space $(X, \mathcal{O}_{\text{particular}}) = (\{1, 2\}, \{\varnothing, \{1\}, \{1, 2\}\})$ is not Hausdorff, because its points 1 and 2, cannot be put into two different open sets (that do not overlap). Neither is the space $(X, \mathcal{O}_{\text{trivial}}) = (\{1, 2\}, \{\varnothing, \{1, 2\}\})$. Indeed, these spaces have "too few" open sets.

Lemma 2.1.3. *Euclidean space $\mathbb{E}^m$ is a Hausdorff topological space.*

Proof. Consider two different points $p, q \in \mathbb{R}^m$, $p \neq q$ and let the Euclidean distance between them be $r = d(p, q) > 0$. They can be separated into different open sets by covering them with balls that have a radius of half this distance. That is $p \in B_{r/2}(p)$ and $q \in B_{r/2}(q)$ and $B_{r/2}(p) \cap B_{r/2}(q) = \varnothing$. Therefore Euclidean space $\mathbb{E}^m$ is Hausdorff. □

Another property of a topological space is whether it is "in one piece". A space that can be split into two separate parts, each being open, would in some sense consist of two independent spaces. We call spaces that cannot be split in this way *connected*.

Definition 2.1.7 (Connected)**.** A topological space $(X, \mathcal{O})$ is called *connected* if and only if it can *not* be written as a union $X = U_1 \cup U_2$ of two nonempty disjoint open sets $U_1, U_2 \in \mathcal{O}$, $U_1, U_2 \neq \varnothing$, $U_1 \cap U_2 = \varnothing$.

Example 6. Euclidean space $\mathbb{E}^m$ is connected. The discrete space $(\{1, 2\}, \mathcal{O}_{\text{discrete}})$ is not connected, since $\{1, 2\} = \{1\} \cup \{2\}$ with $\{1\} \cup \{2\} \in \mathcal{O}_{\text{discrete}}$ and clearly $\{1\} \cap \{2\} = \varnothing$.

Definition 2.1.8 (Compact)**.** A subset $K \subseteq X$ of a topological space $(X, \mathcal{O})$ is called *compact* if and only if for every cover of K with open sets $K \subseteq \bigcup_{i \in I} U_i$, with $U_i \in \mathcal{O}$, we can always choose a finite number of these open sets $J \subseteq I$, with J finite, and still cover it, that is, $K \subseteq \bigcup_{j \in J} U_j$.

We say that $(X, \mathcal{O})$ is compact if and only if X is compact (as a subset of itself $X \subseteq X$).

Remark. In short, one usually says "every open cover has as finite subcover".

Sometimes its easier to phrase facts about topology with the complement of open sets. These sets are called *closed*.

Definition 2.1.9 (Closed)**.** Let $(X, \mathcal{O})$ be a topological space. A subset $E \subseteq X$ is *closed* if $X \setminus E \in \mathcal{O}$.

Lemma 2.1.4 (Closed subset of a compact space)**.** *Let $(X, \mathcal{O})$ be compact, and let $E \subseteq K$ be closed. Then E is compact.*

Proof. Let $E \subseteq \bigcup_{i \in I} U_i$ where U_i are open. Since E is closed $X \setminus E$ is open, and hence $X = (X \setminus E) \cup \bigcup_{i \in I} U_i$ is an open cover of X. Since X is compact there exists a finite subcover $X = (X \setminus E) \cup \bigcup_{j \in J} U_j$ or possibly $X = \bigcup_{j \in J} U_j$. In either case, since $E \cap (X \setminus E) = \varnothing$, $E \subseteq X$ means that $E \subseteq \bigcup_{j \in J} U_j$. This is a finite subcover of the original cover $E \subseteq \bigcup_{i \in I} U_i$, and hence E is compact. □

For subspaces of Euclidean space, there is a useful theorem.

Lemma 2.1.5 (Heine-Borel)**.** *A subspace of Euclidean space is compact if and only if it is closed and bounded.*

Remark. A subspace S of $\mathbb{E}^m$ is bounded if there exists a ball $B_r(c)$ such that $S \subset B_r(c)$.

Proof. See Theorem 2.41 of [Rud76]. □

Definition 2.1.10 (Closure)**.** Let S be a subspace of $(X, \mathcal{O})$. The *closure* $\overline{S}$ of S is the set of points $p \in X$ such that every open set $U \in \mathcal{O}$ with $p \in U$ satisfies $U \cap S \neq \varnothing$.

Remark. The closure $\overline{S}$ is the smallest closed set in X containing S in the sense that if T is a closed set in X with $S \subseteq T$, then $\overline{S} \subseteq T$.

Lemma 2.1.6. *In a Hausdorff space $(X, \mathcal{O})$, if $F \subseteq X$ is a finite set of points then F is closed.*

Proof. Say that $F = \{p_1, \ldots, p_n\}$. Firstly, $X \setminus \{p_i\}$ is open because for every $q \in X \setminus \{p_i\}$ by the Hausdorff property there exists an open set U_q such that $q \in U_q$ and $p_i \notin U_q$. Hence $X \setminus \{p_i\} = \bigcup_{q \in X \setminus \{p_i\}} U_q$ is a union of open sets, and hence open. Furthermore $X \setminus F = \bigcap_{i=1}^{n} X \setminus \{p_i\}$, which is a finite intersection of open sets and is hence open. □

2.2 Continuous functions and homeomorphisms

2.2.1 Continuous functions

Considering Example 1, we note that one can also have the topology $\mathcal{O}_{\text{particular 2}} = \{\varnothing, \{2\}, \{1, 2\}\}$, which is similar to the topology $\mathcal{O}_{\text{particular}}$. Maybe we should consider them the same. To be able to say that two topological spaces should be treated as the same, we need a map between them preserving the topological structure.

We will formalize this idea by considering functions $f : X_1 \to X_2$ between two topological spaces $(X_1, \mathcal{O}_1)$ and $(X_2, \mathcal{O}_2)$. Under such a function, every part of X_1 will be mapped into X_2, and we will say that the topological structure of $(X_1, \mathcal{O}_1)$ is preserved by f in $(X_2, \mathcal{O}_2)$ if the function f is *continuous*. In this sense, two topological spaces will be treated as identical if there exists a function between the spaces which is continuous both ways (meaning a continuous function which has a continuous inverse). Such functions will be called *homeomorphisms*.

Definition 2.2.1 (Continuous function)**.** Given two topological spaces $(X_1, \mathcal{O}_1)$, $(X_2, \mathcal{O}_2)$, a function $f : X_1 \to X_2$ is called a *continuous* function from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$ if and only if the preimage of each open set in X_2 is open in X_1. In other words, for every $U_2 \in \mathcal{O}_2$ we have $f^{-1}(U_2) \in \mathcal{O}_1$.

Remark. One could guess that a continuous function is a function which sends open sets to open sets (such functions are called "open"), but this definition would not work well with the definition of a topology. This is because preimages preserve arbitrary unions and finite intersections, whereas images do not in general.

There are also further reasons, like generalizing real analysis, where it turns out that in $\mathbb{E}^m$ this definition of continuity is equivalent to the standard definition of continuity as in Lemma 2.2.5.

Example 7 (Constant functions are continuous)**.** For any two (nonempty) topological spaces $(X_1, \mathcal{O}_1)$, $(X_2, \mathcal{O}_2)$ there always exists continuous functions, since we may take a point $q \in X_2$ and consider the constant function $c_q : X_1 \to X_2,\ x \mapsto q$ for each $x \in X_1$.

Such a function is continuous because for any open set $V \in \mathcal{O}_2$, if $q \in V$ then $c_q^{-1}(V) = X_1$ and if $q \notin V$ then $c_q^{-1}(V) = \varnothing$, and as we know $X_1, \varnothing \in \mathcal{O}_1$.

Remark. Here we assumed that $X_2 \neq \varnothing$, which is fine since we do not care about the case $X_2 = \varnothing$. (If you are curious, given that $X_2 = \varnothing$ and $X_1 \neq \varnothing$ then there exists no function from X_1 to X_2, and if $X_1 = \varnothing$ then there exists exactly one function from X_1 to X_2 called the empty function $\varnothing$ which is automatically continuous, since its preimage is constantly equal to $\varnothing$.)

Lemma 2.2.1 (Composition of continuous functions is continuous)**.** *Given two continuous functions f from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$ and g from $(X_2, \mathcal{O}_2)$ to $(X_3, \mathcal{O}_3)$, then the composition $g \circ f$ is a continuous function from $(X_1, \mathcal{O}_1)$ to $(X_3, \mathcal{O}_3)$.*

Proof. Let $W \in \mathcal{O}_3$ be an arbitrary open set. Then $(g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W))$, where g being continuous means that $g^{-1}(W) \in \mathcal{O}_2$, and therefore, f being continuous means that $f^{-1}(g^{-1}(W)) \in \mathcal{O}_1$. Hence $g \circ f$ is continuous. □

Lemma 2.2.2 (Restriction of continuous function is continuous)**.** *Let f be a continuous function from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$ and $S \subseteq X_1$. Then the restriction $f|_S : S \to f(S),\ x \mapsto f(x)$ is continuous between the subspaces $S \subseteq X_1$ and $f(S) \subseteq X_2$.*

Proof. Let V be open in $f(S)$. By the subspace topology, there exists $W \in \mathcal{O}_2$ such that $V = W \cap f(S)$, and $f|_S^{-1}(V) = f|_S^{-1}(W) \cap f|_S^{-1}(f(S)) = f|_S^{-1}(W) \cap S = f^{-1}(W) \cap S$, since $f|_S^{-1}(W) \subseteq S$. Because f is continuous $f^{-1}(W) \in \mathcal{O}_1$, so by the subspace topology $f^{-1}(W) \cap S = f|_S^{-1}(V)$ is open in S. □

Lemma 2.2.3 (Continuous sends compact to compact)**.** *Let f be a continuous function from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$ and let $K \subseteq X_1$ be compact. Then $f(K) \subseteq X_2$ is compact.*

Proof. Let $f(K) \subseteq \bigcup_{i \in I} U_i$, $U_i \in \mathcal{O}_2$. Then $K \subseteq f^{-1}(f(K)) \subseteq \bigcup_{i \in I} f^{-1}(U_i)$, since the preimage distributes over unions, and $f^{-1}(U_i) \in \mathcal{O}_1$ because f is continuous. By compactness of K, there exists a finite subcover $K \subseteq \bigcup_{j \in J} f^{-1}(U_j)$, and then

$$f(K) \subseteq f\left(\bigcup_{j \in J} f^{-1}(U_j)\right) = \bigcup_{j \in J} f(f^{-1}(U_j)) \subseteq \bigcup_{j \in J} U_j \tag{2.1}$$

which is a finite subcover of $f(K) \subseteq \bigcup_{i \in I} U_i$. Therefore $f(K)$ is compact. □

Lemma 2.2.4 (A continuous function attains is maximum and minimum)**.** *Let $(X, \mathcal{O})$ be compact, and let f be a continuous function from $(X, \mathcal{O})$ to $\mathbb{E}^1$. Then there exists (not necessarily unique) $x_{min}, x_{max} \in X$ such that $f(x_{min}) \leq f(x)$ and $f(x_{max}) \geq f(x)$ for every $x \in X$.*

Proof. Since f is continuous and $(X, \mathcal{O})$ is compact, $f(X) \subset \mathbb{R}$ is compact. By Heine-Borel, $f(X)$ is closed and bounded. Every closed and bounded subset of $\mathbb{E}^1$ is a finite union of closed intervals $f(X) = \bigcup_{i=1}^{N} [a_i, b_i]$, $a_1 \leq b_1 < a_2 \leq b_2 < \cdots < a_N \leq b_N$. Hence $a_1 \leq f(x) \leq b_N$ and since $a_1, b_N \in f(X)$ that means that there exists $x_{\min} \in f^{-1}(\{a_1\})$ and $x_{\max} \in f^{-1}(\{b_N\})$, which then satisfy $f(x_{\min}) = a_1 \leq f(x)$ and $f(x_{\max}) = b_N \geq f(x)$ for every $x \in X$. □

Lemma 2.2.5 (Continuity in Euclidean space)**.** *Assume that we have a function $f : A \to B$ between two Euclidean subspaces $A \subseteq \mathbb{E}^m$ and $B \subseteq \mathbb{E}^n$. For $x = (x_1, \ldots, x_m) \in A$ this is then a function of the form $(x_1, \ldots, x_m) \mapsto (f_1(x_1, \ldots, x_m), \ldots, f_n(x_1, \ldots, x_m))$ with $(f_1(x), \ldots, f_n(x)) = y \in B$, viewing the component functions $f_1, \ldots, f_n$ as functions $A \to \mathbb{R}$.*

This function f is continuous if and only if each of the component functions f_i fulfill the criterion

$$\lim_{x \to a} f_i(x) = f_i(a)$$

for each $a \in A$.

Proof. See Theorem 4.10(a) of [Rud76]. □

2.2.2 Homeomorphisms

Definition 2.2.2 (Homeomorphism)**.** A continuous function from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$ is said to be a homeomorphism if and only if it is invertible and the inverse is also a continuous function from $(X_2, \mathcal{O}_2)$ to $(X_1, \mathcal{O}_1)$. That is, f is a homeomorphism if and only if f is continuous, f^{-1} exists, and f^{-1} is continuous.

Remark. Hence, if f is a homeomorphism from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$, then f^{-1} is a homeomorphism from $(X_2, \mathcal{O}_2)$ to $(X_1, \mathcal{O}_1)$

Remark. It turns out that having a homeomorphism from X_1 to X_2 (here keeping the topologies implicit) vaguely means that X_1 can be transformed to X_2 without tearing X_1 apart, and X_2 can be transformed to X_1 without tearing X_2 apart. In other words, f being a homeomorphism from X_1 to X_2 is equivalent to it transforming X_1 to X_2 without tearing apart or putting different parts of X_1 *adjacent/next to each other* (reverse of tearing apart) (and its inverse f^{-1} does the same, transforming X_2 to X_1).

Lemma 2.2.6 (Composition of homeomorphisms is a homeomorphism)**.** *Given two homeomorphisms f from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$ and g from $(X_2, \mathcal{O}_2)$ to $(X_3, \mathcal{O}_3)$, then the composition $g \circ f$ is a homeomorphism from $(X_1, \mathcal{O}_1)$ to $(X_3, \mathcal{O}_3)$.*

Lemma 2.2.7 (Restriction of homeomorphism is homeomorphism)**.** *If f is a homeomorphism from $(X_1, \mathcal{O}_1)$ to $(X_2, \mathcal{O}_2)$ and $S \subseteq X_1$ then $f|_S : S \to f(S)$ is a homeomorphism between the subspaces $S \subseteq X_1$ and $f(S) \subseteq X_2$.*

Proof. By Lemma 2.2.2, $f|_S$ is continuous, and since f is a bijection so is $f|_S$, meaning that $f|_S^{-1}$ exists. What remains to be shown is that $g = f|_S^{-1} : f(S) \to S$ is continuous. Let V be open in S, meaning that there exists $W \in \mathcal{O}_1$ such that $V = W \cap S$, and so $g^{-1}(W \cap S) = g^{-1}(W) \cap g^{-1}(S) = f_S(W) \cap f|_S(S) = f_S(W) \cap f(S) = f(W) \cap f(S)$, where $f(W) \in \mathcal{O}_2$ because f^{-1} is continuous and $f(W) = (f^{-1})^{-1}(W)$. By the subspace topology $f(W) \cap f(S) = g^{-1}(V)$ is open in $f(S)$. □

Nonexample 8 ($I \to \mathbb{S}^1$)**.** Consider the subsets $I = [0, 1[\subset \mathbb{R}^1$ and $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \subset \mathbb{R}^2$ as topological subspaces of $\mathbb{E}^1$ and $\mathbb{E}^2$ respectively. The function $p : I \to \mathbb{S}^1$ given by $p(t) = (\cos(2\pi t), \sin(2\pi t))$ is continuous and bijective, but it is not a homeomorphism.

We see that it is not a homeomorphism because for the inverse, the limit $\lim_{(x,y)\to(1,0)} p^{-1}(x, y)$ does not exist, since approaching $(1, 0)$ from different directions along $\mathbb{S}^1$ yields incompatible limits for p^{-1}.

Nonexample 9 (Figure eight)**.** Considering the second remark after Definition 2.2.2, for low-dimensional Euclidean space $m \leq 3$ where we can rely on our intuition, that is, we can imagine that most examples of continuous bijections that are not homeomorphisms will be similar to the previous nonexample. Another nonexample is the figure eight $F \subset \mathbb{R}^2$ as a subspace of $\mathbb{E}^2$, given by $q :]0, 2\pi[\to F,\ t \mapsto (\sin(t), \sin(2t))$, which is a continuous bijection, but not a homeomorphism because $\lim_{(x,y)\to(0,0)} q^{-1}(x, y)$ does not exist.

Example 10. The function $f : \{1, 2\} \to \{1, 2\},\ 1 \mapsto 2,\ 2 \mapsto 1$ is continuous from $(X, \mathcal{O}_{\text{particular}}) = (\{1, 2\}, \{\varnothing, \{1\}, \{1, 2\}\})$ to $(X, \mathcal{O}_{\text{particular 2}}) = (\{1, 2\}, \{\varnothing, \{2\}, \{1, 2\}\})$ because $f^{-1}(\varnothing) = \varnothing$, $f^{-1}(\{2\}) = \{1\}$, $f^{-1}(\{1, 2\}) = \{1, 2\}$ and $\varnothing, \{1\}, \{1, 2\} \in \mathcal{O}_{\text{particular}}$.

We note that the inverse function in this case coincides with f itself, but checking that the inverse is continuous takes into account the topology, which is checking that the preimage of the sets in $\mathcal{O}_{\text{particular}}$ are in $\mathcal{O}_{\text{particular 2}}$. That is, if we denote the inverse function by g (which we call g instead of f^{-1} in this case to avoid confusing with the preimage), then we note that $g^{-1}(\varnothing) = \varnothing$, $g^{-1}(\{1\}) = \{2\}$, $g^{-1}(\{1, 2\}) = \{1, 2\}$ and $\varnothing, \{2\}, \{1, 2\} \in \mathcal{O}_{\text{particular 2}}$, and so the inverse of f exists and is also continuous.

Therefore the topological spaces $(X, \mathcal{O}_{\text{particular}})$ and $(X, \mathcal{O}_{\text{particular 2}})$ are homeomorphic.

Example 11 ($\mathbb{E}^1$ is homeomorphic to any open interval)**.** Consider $\mathbb{E}^1$ and an open interval $]a,b[\subset\mathbb{R}$ as a subspace. Note first that $]a,b[$ is homeomorphic to $]0,1[$ using the function $x\mapsto\frac{x-a}{b-a}$ which is homeomorphic to $]-1,1[$ using $x\mapsto 2x-1$, which is homeomorphic to $]-\pi/2,\pi/2[$ using $x\mapsto\frac{\pi}{2}x$ which is homeomorphic to $\mathbb{E}^1$ using $x\mapsto\tan(x)$. By composing the homeomorphisms we obtain a homeomorphism between the arbitrary open interval $]a,b[$ and $\mathbb{E}^1$.

Lemma 2.2.8. *$\mathbb{E}^m$ is homeomorphic to any open ball in $\mathbb{E}^m$.*

Proof. Firstly we have a homeomorphism $B_r(c)\to B_1(0)$ by shifting the origin and scaling down the radius $p\mapsto\frac{1}{r}(p-c)$. Then $B_1(0)\mapsto\mathbb{R}^m$ by keeping the origin fixed and scaling the distance to the origin from $]0,1[$ to $]0,\infty[$. Say that a point $p\in B_1(0)$ has distance to the origin $d=\|p\|\neq 0$, then $]0,1[\to]0,\pi/2[$ using $d\mapsto\frac{\pi}{2}d$ and $]0,\pi/2[\to]0,\infty[$ using tan. To make sure that the new distance to the origin is given by the composition $d_{\text{new}}=\tan\left(\frac{\pi}{2}\|p\|\right)$, we first scale the distance to the origin to 1 by $p\mapsto p/\|p\|$, and then scale this distance to be d_{new} by multiplying by d_{new}. Therefore, we have a homeomorphism $p\mapsto\tan\left(\frac{\pi}{2}\|p\|\right)\frac{p}{\|p\|}$ if we exclude the origin $B_1(0)\setminus\{0\}\to\mathbb{R}^m\setminus\{0\}$, which extends to a homeomorphism $B_1(0)\to\mathbb{R}^m$ by sending the origin to itself. In total, we obtain a homeomorphism

$$\begin{aligned} h_{r,c}:B_r(c)\to\mathbb{R}^m,\quad & p\mapsto\tan\left(\frac{\pi}{2r}\|p-c\|\right)\frac{p-c}{\|p-c\|},\quad p\neq c\\ & c\mapsto 0. \end{aligned}\tag{2.2}$$

□

2.3 Manifolds

2.3.1 Locally Euclidean spaces and coordinate charts

Definition 2.3.1 (Locally Euclidean space)**.** A topological space $(X,\mathcal{O})$ is called *locally Euclidean* of dimension m if and only if every point of X is contained in some open set homeomorphic to $\mathbb{E}^m$.

Lemma 2.3.1 (Locally Euclidean equivalent definition)**.** *A topological space $(X,\mathcal{O})$ is* locally Euclidean *of dimension m if and only if every point of X is contained in some open set which is homeomorphic to an* arbitrary open subset *of $\mathbb{E}^m$.*

Proof. Because $\mathbb{E}^m$ is an open subset of itself, what we need to show is that for a given point $p\in X$, if there exists a homeomorphism $\varphi:U\to V$ with $p\in U\in\mathcal{O}$ where V is an arbitrary open set in $\mathbb{E}^m$ then there exists a homeomorphism $\widehat{\varphi}:\widehat{U}\to\mathbb{R}^m$ with $p\in\widehat{U}\in\mathcal{O}$.

Say that such a homeomorphism φ is given. Because V is open in $\mathbb{E}^m$ it contains some open ball $B_r(c)\subset V$ around every point $c\in V$. Take $c=\varphi(p)$. Because φ is a homeomorphism it is continuous, meaning that $\widehat{U}=\varphi^{-1}(B_r(\varphi(p)))$ is open in U, where we note that $p\in\widehat{U}$. Then $\widehat{U}\in\mathcal{O}$ since $\widehat{U}$ being open in U (with the subspace topology) means that $\widehat{U}=W\cap U$ for some $W\in\mathcal{O}$, and so $U\in\mathcal{O}$ means that their finite intersection is open $\widehat{U}=W\cap U\in\mathcal{O}$. Finally $\varphi|_{\widehat{U}}:\widehat{U}\to B_r(\varphi(p))$ is a homeomorphism by Lemma 2.2.7, and so $\widehat{\varphi}=h_{r,\varphi(p)}\circ\varphi|_{\widehat{U}}:\widehat{U}\to\mathbb{R}^m$ is a homeomorphism as desired. □

Definition 2.3.2 (Coordinate chart)**.** A *coordinate chart* (U,φ) of a locally Euclidean topological space $(X,\mathcal{O})$ of dimension m is a pair (U,φ) consisting of an open set $U\in\mathcal{O}$ homeomorphic to an open subset V of $\mathbb{E}^m$ and a homeomorphism $\varphi:U\to V$.

Remark. Because the target of φ is $V \subseteq \mathbb{R}^m$ with component functions $\varphi = (\varphi_1, \ldots, \varphi_m)$, which given a point $p \in U$ sends it to the coordinates $(\varphi_1(p), \ldots, \varphi_m(p)) \in \mathbb{R}^m$. Hence the pair (U, φ) can be seen as a "coordinate" chart.

Remark. Following the proof of Lemma 2.3.1, for any $p \in X$ we can always a coordinate chart of the form $\varphi : U \to \mathbb{R}^m$ with $p \in U$.

Lemma 2.3.2 (Restriction of coordinate chart is coordinate chart)**.** *If (U, φ) is a coordinate chart and $V \subseteq U$ is open, then $(V, \varphi|_V)$ is again a coordinate chart.*

Proof. By Lemma 2.2.7 $\varphi|_V : V \to \varphi(V)$ is a homeomorphism, and $\varphi(V)$ is open in $\mathbb{E}^m$ since φ is a homeomorphism. □

2.3.2 Manifolds

Definition 2.3.3 (Manifold)**.** A *manifold* (of dimension m) is a topological space $(M, \mathcal{O})$ that is second countable, Hausdorff, and locally Euclidean (of dimension m).

Example 12 (Euclidean space $\mathbb{E}^m$)**.** Euclidean space $\mathbb{E}^m$ is a manifold, because we have shown it is second countable, Hausdorff, and it is locally Euclidean since for any $p \in \mathbb{R}^m$ we have the homeomorphism $\mathrm{id}_{\mathbb{R}^m} : \mathbb{R}^m \to \mathbb{R}^m,\ (x_1, \ldots, x_m) \mapsto (x_1, \ldots, x_m)$.

Lemma 2.3.3 (Subsets of second countable, Hausdorff)**.** *If a topological space $(X, \mathcal{O})$ is second countable, then any subspace $S \subseteq X$ is also second countable. If $(X, \mathcal{O})$ is Hausdorff, then any subspace $S \subseteq X$ is Hausdorff.*

Proof. This follows immediately from the definitions of subspace, second countable, and Hausdorff space. □

Remark. As a result of this lemma, given a manifold $(M, \mathcal{O})$, a subspace S is also a manifold if and only if S is locally Euclidean.

Nonexample 13 (Figure eight)**.** The figure eight in nonexample 9 is not a manifold because it is not locally Euclidean at $p = (0, 0) \in F$. This is because every open set containing p has four branches, which is not homeomorphic to an open interval (which has two branches).

Example 14 (Circle)**.** The circle $\mathbb{S}^1$ considered as a subspace of $\mathbb{E}^2$ *is* a manifold, of dimension 1. Another property that $\mathbb{S}^1$ has is that it is *compact*, while Euclidean space is not.

Lemma 2.3.4 (Homeomorphisms preserve topological properties)**.** *If two topological spaces $(X_1, \mathcal{O}_1)$ and $(X_2, \mathcal{O}_2)$ are homeomorphic, then $(X_1, \mathcal{O}_1)$ is*

- *Hausdorff*
- *second countable*
- *locally Euclidean (of dimension m)*
- *a manifold (of dimension m)*
- *compact*
- *connected*

if and only if $(X_2, \mathcal{O}_2)$ is also.

Proof. See p.105 of [Mun00]. □

Lemma 2.3.5 (Properties of product space)**.** $(X_1, \mathcal{O}_1), \ldots, (X_n, \mathcal{O}_n)$ *has property P (where P is one of the properties in Lemma 2.3.4), if and only if the product space* $X_1 \times \cdots \times X_n$ *has property P.*

(This is true except that we need to account for the dimension, where $\dim(X_1 \times \cdots \times X_n) = \dim(X_1) + \cdots + \dim(X_n)$ *if P is locally Euclidean or a manifold.)*

Proof. See [Lee11, Proposition 3.31, 3.35] and [Mun00, Theorem 26.7, 23.6]. □

We will see how compact manifolds appear in applications. In particular the sets of orientations of a rigid body in 2D and 3D space (given a reference frame), denoted by SO(2) and SO(3).

Example 15 (SO(2), the set of 2D orientations)**.** Given a rigid body in 2D space, we can give it a reference frame in $\mathbb{R}^2$, where the orientation of this rigid body (bijectively) corresponds to a rotation of this reference frame. Hence we may view SO(2) as the set of 2D rotations, given by

$$\mathrm{SO}(2) = \left\{R \in \mathbb{R}^{2\times 2} \ : \ R^{\mathsf{T}}R = I_2, \ \det(R) = 1\right\} = \left\{ \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \ : \ \theta \in [0, 2\pi[\right\}$$

where each element represents the action of counterclockwise rotating by some angle. Identifying $\mathbb{R}^{2\times 2}$ with $\mathbb{R}^4$ by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto (a, b, c, d)$$

we may give $\mathbb{R}^{2\times 2}$ the topology of Euclidean space $\mathbb{E}^4$, and then SO(2) can be seen as a subspace of $\mathbb{E}^4$ in this way.

Now we see that SO(2) is homeomorphic to the circle $\mathbb{S}^1$ by

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \mapsto (\cos(\theta), \sin(\theta)) \tag{2.3}$$

and they will hence be viewed as the same topological manifolds. In particular, SO(2) is compact.

Since orientations of a rigid body in 2D space can be identified with $\mathbb{S}^1$, one might now believe that orientations of a rigid body in 3D space can be identified with $\mathbb{S}^2$. However, this is *not* the case, since given an object in 3D space, a point on $\mathbb{S}^2$ only represents a direction, where to obtain an orientation we would need to specify how much the object should rotate about that axis.

Hence, one might think that SO(3) can be represented by $\mathbb{S}^2 \times \mathbb{S}^1$, but this is *not* the case either, since, for example, rotating by the angle $\theta \in \mathbb{S}^1$ about the axis $x \in \mathbb{S}^2$ corresponds to the same orientation as rotating by the angle $-\theta$ about the axis $-x$, but indeed (x, θ) and $(-x, -\theta)$ are different points in $\mathbb{S}^2 \times \mathbb{S}^1$. Moreover, rotating by $\theta = 0$ about any axis would result in no change of orientation, and hence the points $(x, 0)$ would need to all be the same. For these reasons, SO(3) must be a different set.

Example 16 (SO(3), the set of 3D orientations)**.** Similarly to SO(2), given a rigid body in 3D space, we can associate to it a reference frame in $\mathbb{R}^3$, where the orientation of this body bijectively corresponds to a rotation of this reference frame, meaning that SO(3) can be viewed as the set of 3D rotations

$$\mathrm{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \ \ R^{\mathsf{T}}R = I, \ \det(R) = 1\}.$$

Now we will again think of this as being a subspace of Euclidean space. In particular view $\mathbb{R}^{3\times 3}$ as $\mathbb{R}^9$ endowed with the Euclidean topology $\mathbb{E}^9$, and then see SO(3) as a subspace of this space. (Note that the order in which we map the elements of $\mathbb{R}^{3\times 3}$ to $\mathbb{R}^9$ does not matter in a topological sense, since different such maps would just result in the components of the $\mathbb{R}^9$ elements being permuted, and a permutation will be a homeomorphism and therefore result in homeomorphic spaces, which we view as the same topologically.)

Remark. It turns out that SO(3) is homeomorphic to $\mathbb{RP}^3$, that is, the 3-dimensional *real projective space*, which is a compact manifold, and hence SO(3) is compact as well.

In addition to being compact manifolds, SO(2) and SO(3) are also *smooth*, which we will discuss next.

2.4 Smooth manifolds

2.4.1 Smooth structure

To be able to discuss differentiability on manifolds, we begin by reducing the problem to Euclidean space, where we understand the meaning of differentiability.

Definition 2.4.1 (Smooth compatibility of change of coordinates)**.** Two coordinate charts (U_1, φ_1) and (U_2, φ_2) are said to be *smoothly compatible* if and only if the *change of coordinates* functions $\varphi_2 \circ \varphi_1^{-1}$ and $\varphi_1 \circ \varphi_2^{-1}$ are infinitely differentiable.

Remark. Note that a change of coordinates function $\varphi_2 \circ \varphi_1^{-1}$ is defined on $\varphi_1(U_1 \cap U_2) \subseteq \mathbb{R}^m$, and its target is $\varphi_2(U_1 \cap U_2) \subseteq \mathbb{R}^m$, so it is a function whose differentiability we can discuss.

Remark. Note further that if $U_1 \cap U_2 = \varnothing$ then $\varphi_2 \circ \varphi_1^{-1} : \varnothing \to \varnothing$ is the "empty function", which we say is infinitely differentiable by definition.

Definition 2.4.2 (Smooth atlas)**.** A *smooth atlas* $\mathcal{A}$ of a manifold $(M, \mathcal{O})$ is a collection of coordinate charts $\mathcal{A} = \{(U_i, \varphi_i) : i \in I\}$ such that any two are smoothly compatible and the open sets cover M. That is, for $i, j \in I$ we have $\varphi_j \circ \varphi_i^{-1}$ is infinitely differentiable, and $\bigcup_{i\in I} U_i = M$.

Definition 2.4.3 (Equivalence of smooth atlases)**.** Two smooth atlases $\mathcal{A}_1$ and $\mathcal{A}_2$ of the same manifold M are said to be equivalent if and only if $\mathcal{A}_1 \cup \mathcal{A}_2$ also is a smooth atlas of M.

Definition 2.4.4 (Smooth structure)**.** A *smooth structure* $[\mathcal{A}]$ for a manifold $(M, \mathcal{O})$ is an equivalence class of smooth atlases on M under the above equivalence relation.

Lemma 2.4.1 (Restriction of coordinate chart is a coordinate chart)**.** *Let $(M, \mathcal{O})$ be a manifold, $[\mathcal{A}]$ a smooth structure for this manifold, and $(U, \varphi) \in \mathcal{A}$. If $V \subseteq U$ and $V \in \mathcal{O}$ then $(V, \varphi|_V) \in \mathcal{A}' \in [\mathcal{A}]$.*

Proof. By Lemma 2.3.2, $(V, \varphi|_V)$ is a coordinate chart. We will now show that $\mathcal{A}' = \mathcal{A} \cup \{(V, \varphi|_V)\}$ is a smooth atlas equivalent to $\mathcal{A}$. Beginning with smooth compatibility, let $(W, \psi) \in \mathcal{A}$ and consider the change of coordinates $\psi \circ (\varphi|_V)^{-1} : \varphi|_V(V \cap W) \to \psi(V \cap W)$. Now $\varphi|_V(V \cap W) = \varphi(V \cap W)$, and since $V \cap W \subseteq U \cap W$, we have $\varphi(V \cap W) \subseteq \varphi(U \cap W)$, and because $(U, \varphi), (W, \psi) \in \mathcal{A}$ then $\psi \circ \varphi^{-1}$ is smooth. On $\varphi(V \cap W)$ we have $(\varphi|_V)^{-1} = \varphi^{-1}$, so $\psi \circ (\varphi|_v)^{-1} = (\psi \circ \varphi^{-1})|_{\varphi(V\cap W)}$ which is the restriction of a smooth function to an open subset, and is hence smooth. Similarly $\varphi|_V \circ \psi^{-1} = (\varphi \circ \psi^{-1})|_{\psi(V\cap W)}$ which is smooth, showing that $(V, \varphi|_V)$ is smoothly compatible with every coordinate chart in $\mathcal{A}$, meaning that $\mathcal{A}'$ is a smooth atlas. Further, this atlas is equivalent to $\mathcal{A}$ since $\mathcal{A} \cup \mathcal{A}' = \mathcal{A}'$ which is a smooth atlas, as shown. □

Lemma 2.4.2. *Given a manifold $(M, \mathcal{O})$ and a smooth structure $[\mathcal{A}]$ for this manifold, the open sets $U \in \mathcal{O}$ are those that map to open sets in Euclidean space when applying a coordinate chart. That is, $\mathcal{O} = \{V : \varphi(U \cap V) \in \mathcal{O}_{Euclidean},\ (U, \varphi) \in \mathcal{A}',\ \mathcal{A}' \in [\mathcal{A}]\}$.*

Remark. Hence, the topology can be reconstructed from the smooth structure.

Proof. See Lemma 1.35 of [Lee13]. □

Definition 2.4.5 (Smooth manifold)**.** A smooth manifold (of dimension m) is a pair $(M, [\mathcal{A}])$ such that the topology induced by $[\mathcal{A}]$ makes M into a manifold (of dimension m), and $[\mathcal{A}]$ into a smooth structure for that manifold.

Example 17 (Countable discrete set)**.** If we define $\mathbb{R}^0 = \{0\}$, so $\mathbb{E}^0 = (\{0\}, \{0\})$, then any countable discrete set D has the smooth atlas $\mathcal{A} = \{(\{p\}, 0_p) : p \in D\}$ for $p \in D$, where $0_p : \{p\} \to \mathbb{R}^0,\ p \mapsto 0$. This makes $(D, [\mathcal{A}])$ into a smooth manifold with the discrete topology $\mathcal{O} = \mathcal{P}(D)$ (every subset of D is open).

Example 18 (Open subsets are smooth manifolds)**.** Let $(M, [\mathcal{A}])$ be a smooth manifold and $O \subseteq M$ be an arbitrary open subset. Then $(O, [\mathcal{A}|_{\mathcal{O}}])$ is a smooth manifold where if $\mathcal{A} = \{(U_i, \varphi_i) : i \in I\}$, we define $\mathcal{A}|_O := \{(U_i \cap O, \varphi_i|_{U_i \cap O})\}$.

Remark. We note that indeed $\varphi_i|_{U_i \cap O} : U_i \cap O \to \varphi_i(U_i \cap O)$ is a homeomorphism, because $\varphi_i : U_i \to \mathbb{R}^m$ is, and so, since $U_i \cap O$ is open in U_i by the definition of subspace topology, the image of $U_i \cap O$ under φ_i becomes an open set in $\mathbb{E}^m$ which is sufficient by lemma 2.3.1.

Example 19 (Product of smooth manifolds is smooth)**.** Let $(M, [\mathcal{A}])$ and $(N, [\mathcal{B}])$ be smooth manifolds and let $\mathcal{A} = \{(U_i, \varphi_i) : i \in I\}$ and $\mathcal{B} = \{(V_j, \psi_j) : j \in J\}$. The *product manifold* $(M \times N, [\mathcal{A}_{M \times N}])$ is then a smooth manifold where $\mathcal{A}_{M \times N} = \{(U_i \times V_j, \varphi_i \times \psi_j) : (i, j) \in I \times J\}$ and $\varphi_i \times \psi_j(p, q) := (\varphi_i(p), \psi_j(q))$.

Lemma 2.4.3 (Global coordinate chart implies smoothness)**.** *Given a manifold $(M, \mathcal{O})$, if there exists a homeomorphism $\varphi : M \to \mathbb{R}^m$ from this manifold to the Euclidean space $\mathbb{E}^m$, including (M, φ) as a coordinate chart automatically makes $(M, [\{(M, \varphi)\}])$ into a smooth manifold.*

Proof. This follows from the fact that $\mathcal{A} = \{(M, \varphi)\}$ is a smooth atlas, because it covers the manifold M and if two coordinate charts are taken from $\mathcal{A}$ they are necessarily the same (M, φ), and hence there is only one change of coordinates function

$$\varphi \circ \varphi^{-1} = \mathrm{id}_{\mathbb{R}^m} : (p_1, \ldots, p_m) \mapsto (p_1, \ldots, p_m) \tag{2.4}$$

which is infinitely differentiable because

$$\begin{aligned} \frac{\partial\, \mathrm{id}_{\mathbb{R}^m}}{\partial x_i}(p) &= \frac{\partial(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m)}{\partial x_i}(p) \\ &= (0, \ldots, 0, 1, 0, \ldots 0)(p) \\ &= (0, \ldots, 0, 1, 0, \ldots 0) \\ &= e_i \end{aligned} \tag{2.5}$$

which is constant, so

$$\frac{\partial^2\, \mathrm{id}_{\mathbb{R}^m}}{\partial x_j \partial x_i}(p) = \frac{\partial e_i}{\partial x_j}(p) = 0(p) = 0. \tag{2.6}$$

Any higher order derivative would then also be 0. □

Corollary 2.4.4 (Euclidean space $\mathbb{E}^m$)**.** *The pair $\mathbb{E}^m = (\mathbb{R}^m, [\mathcal{A}])$ with $\mathcal{A} = \{(\mathbb{R}^m, \mathrm{id}_{\mathbb{R}^m})\}$ is a smooth manifold, where $\mathrm{id}_{\mathbb{R}^m} : \mathbb{R}^m \to \mathbb{R}^m,\ (p_1, \ldots, p_m) \mapsto (p_1, \ldots, p_m)$.*

Remark. The function $\mathrm{id}_{\mathbb{R}^m}$ has component functions $\mathrm{id}_{\mathbb{R}^m} = (x_1, \ldots, x_m)$ where these are defined by $x_i : \mathbb{R}^m \to \mathbb{R}^1,\ (p_1, \ldots, p_m) \mapsto p_i$.

Remark. The smooth structure obtained from the atlas $\mathcal{A} = \{(\mathbb{R}^m, \mathrm{id}_{\mathbb{R}^m})\}$ is called the canonical smooth structure on $\mathbb{R}^m$. The induced topology from this smooth structure is the Euclidean topology, and for $m \neq 4$ it is the only smooth structure that induces the Euclidean topology (up to diffeomorphism). For $m = 4$ there are also uncountably many other (not diffeomorphic) smooth structures that induce the Euclidean topology. They are called exotic $\mathbb{R}^4$. [Sco05]

Lemma 2.4.5 (SO(n))**.** *The set of nD rotations* $\mathrm{SO}(n) = \{R \in \mathbb{R}^{n\times n} : R^\mathsf{T}R = I,\ \det(R) = 1\}$ *has a canonical smooth structure, making it a compact smooth manifold of dimension* $m = n(n-1)/2$.

Proof. See Example 7.28 of [Lee13]. □

2.4.2 Smooth functions and diffeomorphisms

Now that we have smooth manifolds, we wish to study functions between them. For functions between Euclidean spaces, we already understand what it means to be smooth, that is, infinitely differentiable. For a function $f : M \to N$ between smooth manifolds, the idea is to reduce to the Euclidean case by expressing f in coordinate charts.

Definition 2.4.6 (Smooth function)**.** A *smooth function* f from a smooth manifold $(M, [\mathcal{A}])$ to a smooth manifold $(N, [\mathcal{B}])$ is a function $f : M \to N$ that is smooth in Euclidean space. That is, for every $p \in M$ there exists a coordinate chart $(U, \varphi) \in \mathcal{A}' \in [\mathcal{A}]$ with $p \in U$ and a coordinate chart $(V, \psi) \in \mathcal{B}' \in [\mathcal{B}]$ with $f(p) \in V$ such that $f(U) \subseteq V$ and the expression of f in Euclidean space in this coordinate chart $\psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(V)$ is infinitely differentiable.

Remark. To make the composition $\psi \circ f \circ \varphi^{-1}$ make sense, we are here actually considering the restriction $\widehat{f} : U \to V$.

Remark. It turns out that if we do not impose that $f(U) \subseteq V$, then this is not really a problem, since if $\psi \circ f \circ \varphi^{-1} : \varphi(U \cap f^{-1}(V)) \to \psi(V)$ is infinitely differentiable then (U, φ) can be restricted to a the smaller coordinate chart (U', φ') where $U' = U \cap f^{-1}(V)$. where indeed $f(U') \subseteq V$. However, to make this a coordinate chart U' needs to be open, which is true if f is continuous, which we will show next. Hence, after Lemma 2.4.6, we can without loss of generality assume that $f(U) \subseteq V$.

Lemma 2.4.6 (Smooth functions are continuous)**.** *A smooth function f from $(M, [\mathcal{A}])$ to $(N, [\mathcal{B}])$ is a continuous function.*

Proof. Let $W \subseteq N$ be open. We wish to show that $f^{-1}(W) \subseteq M$ is open. It suffices to show that for every $p \in f^{-1}(W)$ there exists an open set $O_p \subseteq M$ with $p \in O_p \subseteq f^{-1}(W)$, since then $f^{-1}(W) = \bigcup_{p\in f^{-1}(W)} O_p$, which is a union of open sets and hence open.

Let $p \in f^{-1}(W)$, so $f(p) \in W$. Since f is smooth, by definition there exist coordinate charts $(U, \varphi) \in \mathcal{A}' \in [\mathcal{A}]$ with $p \in U$ and $(V, \psi) \in \mathcal{B}' \in [\mathcal{B}]$ with $f(p) \in V$ such that $\psi \circ \widehat{f} \circ \varphi^{-1} : \varphi(U) \to \psi(V)$ is infinitely differentiable. In particular, by Lemma 2.2.5, $\psi \circ \widehat{f} \circ \varphi^{-1}$ is continuous. Consider the open set $V \cap W \subseteq N$, which is nonempty since $f(p) \in V \cap W$. Since $\psi : V \to \psi(V)$ is a homeomorphism, $\psi(V \cap W)$ is open in $\psi(V)$, and therefore open in $\mathbb{E}^n$, where $n = \dim(N)$. By continuity of $\psi \circ \widehat{f} \circ \varphi^{-1}$, we have that $(\psi \circ \widehat{f} \circ \varphi^{-1})^{-1}(\psi(V \cap W)) \subseteq \varphi(U)$ is open in $\varphi(U)$ and hence open in $\mathbb{E}^m$ where $m = \dim(M)$, and this is equal to

$$\begin{aligned}(\psi \circ \widehat{f} \circ \varphi^{-1})^{-1}(\psi(V \cap W)) &= \varphi \circ \widehat{f}^{-1} \circ \psi^{-1}(\psi(V \cap W)) \\ &= \varphi \circ \widehat{f}^{-1}(V \cap W) \\ &= \varphi(f^{-1}(V \cap W) \cap U).\end{aligned} \tag{2.7}$$

Since $\varphi : U \to \varphi(U)$ is a homeomorphism, $O_p = f^{-1}(V \cap W) \cap U$ is open in U, and therefore open in M. Finally, $p \in O_p$ because $p \in U$ and $f(p) \in V \cap W$, and $O_p \subseteq f^{-1}(W)$ because $f^{-1}(V \cap W) \subseteq f^{-1}(W)$. □

Definition 2.4.7 (Diffeomorphism)**.** A *diffeomorphism* is a smooth function that is invertible and the inverse is also a smooth function.

Remark. Two smooth manifolds are considered the same if there exists a diffeomorphism between them, and they are then called diffeomorphic. For example, SO(2) is diffeomorphic to $\mathbb{S}^1$ by the function in (2.3).

Lemma 2.4.7 (Coordinate charts are diffeomorphisms)**.** *Given a smooth manifold $(M, [\mathcal{A}])$, any coordinate chart (U, φ) becomes a diffeomorphism $\varphi : U \to \varphi(U)$ between the manifolds $(U, [\mathcal{A}|_U])$ and the subspace $\varphi(U)$ of $\mathbb{E}^m$.*

Proof. Firstly, since $\varphi(U) \subseteq \mathbb{R}^m$, we have that $(\varphi(U), \mathrm{id}_{\varphi(U)})$ is a coordinate chart for $\varphi(U)$. Choosing the coordinates (U, φ) on U gives the expression for φ in Euclidean space

$$\mathrm{id}_{\varphi(U)} \circ \varphi \circ \varphi^{-1} = \mathrm{id}_{\varphi(U)} : (p_1, \ldots, p_m) \mapsto (p_1, \ldots, p_m), \tag{2.8}$$

which we have shown is infinitely differentiable. Similarly, the expression for $\varphi^{-1} : \varphi(U) \to U$ in Euclidean space also becomes

$$\varphi \circ (\varphi^{-1}) \circ \mathrm{id}^{-1}_{\varphi(U)} = \mathrm{id}^{-1}_{\varphi(U)} = \mathrm{id}_{\varphi(U)}, \tag{2.9}$$

which is infinitely differentiable, and so φ is a diffeomorphism. □

Lemma 2.4.8 (Smoothness is independent of coordinates)**.** *If $f : M \to N$ is smooth, then for any coordinate chart (U_2, φ_2) on M and (V_2, ψ_2) on N, the expression for f in the Euclidean space in these coordinate charts $\psi_2 \circ f \circ \varphi_2^{-1}$ is infinitely differentiable.*

Proof. This follows from smooth compatibility of coordinate charts, because for $p \in M$, say that $f : M \to N$ is smooth in the coordinates (U, φ), (V, ψ), with $p \in U$, $f(p) \in V$ then for any other coordinate chart (U_2, φ_2), (V_2, ψ_2) with $p \in U_2$, $f(p) \in V_2$, we obtain that

$$\psi_2 \circ f \circ \varphi_2^{-1} = \psi_2 \circ (\psi^{-1} \circ \psi) \circ f \circ (\varphi^{-1} \circ \varphi) \circ \varphi_2^{-1} = (\psi_2 \circ \psi^{-1}) \circ (\psi \circ f \circ \varphi^{-1}) \circ (\varphi \circ \varphi_2^{-1}), \tag{2.10}$$

which is the composition of three smooth functions, and is therefore smooth by the standard chain rule. □

2.5 Tangent space and derivative

2.5.1 Tangent vectors

Definition 2.5.1 (Smooth curves)**.** A *smooth curve* c on a smooth manifold M is a smooth function $c : I \to M$, where $0 \in I = (a, b)$ is an open interval in $\mathbb{E}^1$ containing zero.

Remark. We will typically use the variable t in the interval I and think of it as time, and $t = 0 \in I$ will be called initial moment.

We can now construct tangent vectors on manifolds by considering velocity vectors of these curves. As usual we move the problem to the Euclidean space in order to have objects we can actually work with. It can happen that two different curves have the same initial velocity, and in this case we want to identify the curves to obtain a unique correspondence to a vector in Euclidean space.

Definition 2.5.2 (Equivalence of curves)**.** Two curves c_1, c_2 with the same initial point $c_1(0) = c_2(0) = p \in M$ will be considered equivalent if and only if they also have the same initial velocity vector in Euclidean space. That is, if there exists a coordinate chart (U, φ) with $p \in U$ such that

$$(\varphi \circ c_1)'(0) = (\varphi \circ c_2)'(0),$$

where $'$ denotes taking the derivative.

Remark. Because curves have Euclidean space as their domain, we may use standard coordinates as a coordinate chart, meaning that the expression of c_1 in Euclidean space becomes $\varphi \circ c_1 \circ \mathrm{id}^{-1} = \varphi \circ c_1$, and similarly for c_2.

Lemma 2.5.1. *The relation in Definition 2.5.2 is well-defined, in the sense that it is independent of the coordinate chart.*

Proof. It is well-defined because if it holds in one coordinate chart (U_1, φ_1) with $p \in U_1$, then for any other coordinate chart (U_2, φ_2) with $p \in U_2$, the chain rule gives that

$$\begin{aligned}
(\varphi_2 \circ c_1)'(0) &= (\varphi_2 \circ \varphi_1^{-1} \circ \varphi_1 \circ c_1)'(0) \\
&= (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(c_1(0))} \cdot (\varphi_1 \circ c_1)'(0) \\
&= (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \cdot (\varphi_1 \circ c_1)'(0) \\
&= (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \cdot (\varphi_1 \circ c_2)'(0) \\
&= (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(c_2(0))} \cdot (\varphi_1 \circ c_2)'(0) \\
&= (\varphi_2 \circ \varphi_1^{-1} \circ \varphi_1 \circ c_2)'(0) \\
&= (\varphi_2 \circ c_2)'(0).
\end{aligned} \tag{2.11}$$

Thus the equality holds in this coordinate chart as well. □

Remark. Because $\varphi_2 \circ \varphi_1^{-1}$ is a function from a subset of $\mathbb{R}^m$ to $\mathbb{R}^m$, its derivative is the Jacobian matrix $(\varphi_2 \circ \varphi_1^{-1})'$, where $(\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)}$ is evaluating each entry of the matrix at the point $\varphi_1(p) \in \mathbb{R}^m$. Moreover $(\varphi_1 \circ c_2)'(0)$ is a vector, and so $\cdot$ is the multiplication of a matrix by a vector.

Lemma 2.5.2. *The relation in Definition 2.5.2 is an equivalence relation.*

Proof. Denote this relation by $\sim$. It is obviously reflexive, that is, $c \sim c$, because $(\varphi \circ c)'(0) = (\varphi \circ c)'(0)$ trivially.

It is also trivially symmetric $c_1 \sim c_2 \implies c_2 \sim c_1$ because $(\varphi \circ c_1)'(0) = (\varphi \circ c_2)'(0)$ obviously means that $(\varphi \circ c_2)'(0) = (\varphi \circ c_1)'(0)$.

Transitivity $c_1 \sim c_2,\ c_2 \sim c_3 \implies c_1 \sim c_3$ follows from the independence of coordinate charts, because if $c_1 \sim c_2$ by $(\varphi_1 \circ c_1)'(0) = (\varphi_1 \circ c_2)'(0)$ and $c_2 \sim c_3$ by $(\varphi_2 \circ c_2)'(0) = (\varphi_2 \circ c_3)'(0)$, then we know by the independence that also $(\varphi_1 \circ c_2)'(0) = (\varphi_1 \circ c_3)'(0)$ and then we immediately obtain $(\varphi_1 \circ c_1)'(0) = (\varphi_1 \circ c_2)'(0) = (\varphi_1 \circ c_3)'(0)$ and therefore $c_1 \sim c_3$. □

Now we have motivated the use of the word "equivalent" in Definition 2.5.2. Every two possible curves on M with an initial point $p \in M$ that could correspond to the same vector in $\mathbb{R}^m$ are equivalent. Therefore each equivalence class of curves corresponds to a unique vector in $\mathbb{R}^m$. This equivalence class will be called a *tangent vector* at the point $p \in M$. The *tangent space* at that point will be the set of all such collections. For this set we will define a vector space structure, where the point on the manifold will be the origin of the vector space.

Definition 2.5.3 (Tangent vector)**.** A *tangent vector* $v = [c]$ at a point $p \in M$ of a smooth manifold is the equivalence class of curves on M with $c(0) = p$.

Definition 2.5.4 (Tangent space)**.** The *tangent space* T_pM at a point $p \in M$ (or with origin $p \in M$) of a smooth manifold M is the set of all tangent vectors at that point.

The vector in $\mathbb{R}^m$ to which the tangent vector $[c]$ at $p \in M$ corresponds depends on the coordinate chart (U, φ). The vector will be $(\varphi \circ c)'(0)$, so we may consider a function $[c] \mapsto (\varphi \circ c)'(0)$, which corresponds to the derivative, or Jacobian, of φ evaluated at p.

Definition 2.5.5 (Derivative of a coordinate chart)**.** The *derivative* $T_p\varphi$ of a coordinate chart (U, φ) at a point $p \in U$ is the function

$$T_p\varphi : T_pM \to \mathbb{R}^m, \quad [c] \mapsto (\varphi \circ c)'(0).$$

Lemma 2.5.3. *$T_p\varphi : T_pM \to \mathbb{R}^m$ is well-defined and is a bijection.*

Proof. Because tangent vectors $[c]$ were constructed in such a way that they correspond to a unique vector in $\mathbb{R}^m$ by the function $T_p\varphi$, this means that $T_p\varphi$ is well-defined and injective.

It is also surjective because given a vector $w \in \mathbb{R}^m$ we may consider a curve of the form $c(t) = \varphi^{-1}(\varphi(p) + tw)$, taking $t \in I \subseteq \mathbb{R}$ in a small enough interval I such that $\mathrm{im}(c) \subseteq U$. Then we see that

$$T_p\varphi([c]) = (\varphi(\varphi^{-1}(\varphi(p) + tw)))'|_{t=0} = (\varphi(p) + tw)'|_{t=0} = w|_{t=0} = w. \tag{2.12}$$

Hence $T_p\varphi$ is surjective, and therefore a bijection. □

Remark. Notice that the derivative of any coordinate chart function φ is bijective, in the sense that it relies on φ being invertible and every change of coordinates function $\varphi \circ \psi^{-1}$ being differentiable, as seen in the proofs above. We will see that the derivative of a function (which we define below) does not necessarily have a bijective derivative.

Now because $\mathbb{R}^m$ is a vector space and given a coordinate chart we have a bijection to T_pM, we can turn T_pM into a vector space as well by moving our elements of T_pM to $\mathbb{R}^m$, doing our vector operations there, and then moving back to T_pM. But we will show that the vector space structure is independent of the coordinate chart. To do this we first need a formula for changing derivative, that is, the chain rule.

Lemma 2.5.4 (Chain rule)**.** $T_p\varphi_2 = (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \circ T_p\varphi_1$.

Proof. This result follows from the usual chain rule as in (2.11), since for any $v = [c] \in T_pM$

$$\begin{aligned} T_p\varphi_2([c]) &= (\varphi_2 \circ c)'(0) \\ &= (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \cdot (\varphi_1 \circ c)'(0) \\ &= (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \cdot T_p\varphi_1([c]). \end{aligned} \tag{2.13}$$

Since the equality holds for every input $[c] \in T_pM$, it holds for functions, where the matrix-vector multiplication $(\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \cdot T_p\varphi_1([c])$ can be rewritten as the composition $(\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \circ T_p\varphi_1$ evaluated at $[c]$, viewing $(\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)}$ as a function $\mathbb{R}^m \to \mathbb{R}^m$, $w \mapsto (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \cdot w$. □

Remark. Because $T_p\varphi_1, T_p\varphi_2$ are bijections, they are invertible. Thus we obtain the following formula

$$\begin{aligned} T_p\varphi_2 = (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \circ T_p\varphi_1 &\implies (T_p\varphi_2) \circ (T_p\varphi_1)^{-1} = (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)} \\ &\implies (T_p\varphi_1)^{-1} = (T_p\varphi_2)^{-1} \circ (\varphi_2 \circ \varphi_1^{-1})'|_{\varphi_1(p)}. \end{aligned} \tag{2.14}$$

Proposition 2.5.5 (Tangent space is a vector space)**.** *The tangent space T_pM at a point $p \in M$ (of a smooth manifold M of dimension m) is real vector space of dimension m with the vector addition and multiplication by a scalar given by*

$$\begin{aligned} v + w &:= (T_p\varphi)^{-1}(T_p\varphi(v) + T_p\varphi(w)) \\ rv &:= (T_p\varphi)^{-1}(rT_p\varphi(v)). \end{aligned}$$

Here $v, w \in T_pM$, $r \in \mathbb{R}$ and (U, φ) with $p \in U$ is an arbitrary coordinate chart on M.

Proof. First we show that the operations are independent of the coordinate chart. Say that (U_1, φ_1) and (U_2, φ_2) are two coordinate charts with $p \in U_1$ and $p \in U_2$. Then $(T_p\varphi_1)^{-1}(T_p\varphi_1(v) + T_p\varphi_1(w)) = (T_p\varphi_2)^{-1}(T_p\varphi_2(v) + T_p\varphi_2(w))$ because by Lemma 2.5.4, and using the fact that matrix-vector multiplication is linear we get

$$\begin{aligned} (T_p\varphi_1)^{-1}(T_p\varphi_1(v) + T_p\varphi_1(w)) &= (T_p\varphi_1)^{-1}\Big((\varphi_1 \circ \varphi_2^{-1})'|_{\varphi_2(p)} \ \cdot \ (T_p\varphi_2(v) + T_p\varphi_2(w))\Big) \\ &= (T_p\varphi_1)^{-1} \circ (\varphi_1 \circ \varphi_2^{-1})'|_{\varphi_2(p)}(T_p\varphi_2(v) + T_p\varphi_2(w)) \\ &= (T_p\varphi_2)^{-1}(T_p\varphi_2(v) + T_p\varphi_2(w)). \end{aligned} \tag{2.15}$$

Further $(T_p\varphi_1)^{-1}(rT_p\varphi_1(v)) = (T_p\varphi_2)^{-1}(rT_p\varphi_2(v))$ because in $\mathbb{R}^m$ we have $rw = (rI) \cdot w$ with I being the identity matrix, and matrices of the form rI are those that commute with every other matrix, so

$$\begin{aligned} (T_p\varphi_1)^{-1}(rT_p\varphi_1(v)) &= (T_p\varphi_1)^{-1}(r(\varphi_1 \circ \varphi_2^{-1})'|_{\varphi_2(p)} \ \cdot \ T_p\varphi_2(v)) \\ &= (T_p\varphi_1)^{-1}((rI)(\varphi_1 \circ \varphi_2^{-1})'|_{\varphi_2(p)} \ \cdot \ T_p\varphi_2(v)) \\ &= (T_p\varphi_1)^{-1}((\varphi_1 \circ \varphi_2^{-1})'|_{\varphi_2(p)}(rI) \ \cdot \ T_p\varphi_2(v)) \\ &= (T_p\varphi_1)^{-1} \circ (\varphi_1 \circ \varphi_2^{-1})'|_{\varphi_2(p)}((rI) \ \cdot \ T_p\varphi_2(v)) \\ &= (T_p\varphi_2)^{-1}((rI) \ \cdot \ T_p\varphi_2(v)) \\ &= (T_p\varphi_2)^{-1}(rT_p\varphi_2(v)). \end{aligned} \tag{2.16}$$

This means that the definitions of $v + w$ and rv are independent of the coordinate chart and are hence well-defined.

Now we need to show that all the axioms for a vector space are fulfilled for T_pM with these operations. To do that, we note that the derivative of a coordinate chart will be a linear function $T_pM \to \mathbb{R}^m$ by definition, in the sense that

$$\begin{aligned} v + w := (T_p\varphi)^{-1}(T_p\varphi(v) + T_p\varphi(w)) &\iff T_p\varphi(v + w) = T_p\varphi(v) + T_p\varphi(w) \\ rv := (T_p\varphi)^{-1}(rT_p\varphi(v)) &\iff T_p\varphi(rv) = rT_p\varphi(v). \end{aligned} \tag{2.17}$$

Since the axioms of a vector space holds for $\mathbb{R}^m$, the linearity property together with the fact that $T_p\varphi$ is a bijection means that the axioms for a vector space hold for T_pM as well, and that it will be an isomorphism of vector spaces (since the inverse of a linear function automatically is linear). Hence T_pM is a vector space of dimension m over $\mathbb{R}$. □

Corollary 2.5.6. *$T_p\varphi : T_pM \to \mathbb{R}^m$ is an isomorphism (linear bijection) of vector spaces.*

Proposition 2.5.7 (Euclidean space as a vector space)**.** *There is a canonical isomorphism of vector spaces $T_p\mathbb{E}^m \to \mathbb{R}^m$.*

Proof. Because $\mathbb{E}^m$ has a canonical coordinate chart $(\mathbb{R}^m, \mathrm{id})$ this gives a canonical isomorphism of vector spaces $T_p \,\mathrm{id}_{\mathbb{R}^m} : T_p\mathbb{E}^m \to \mathbb{R}^m, \ [c] \mapsto c'(0)$. □

2.5.2 Coordinate vectors and bases

Remark. One can view $\mathbb{R}^m$ as the set of vectors in $\mathbb{R}^m$ with origin $0 = (0, \dots, 0)$, which is the set of tangent vectors of $\mathbb{E}^m$ at 0. That is $T_p\mathbb{E}^m \cong \mathbb{R}^m \cong T_0\mathbb{E}^m$. In this sense, the isomorphism $T_p \operatorname{id}_{\mathbb{R}^m}$ moves the origin from p to 0, and it is canonical because 0 is the canonical point of $\mathbb{R}^m$. This is how we will view $\mathbb{R}^m$ further, and identify these spaces.

Because $\mathbb{R}^m$ has a canonical basis $e_1 = (1, 0, \dots, 0, 0)$, $\dots$, $e_m = (0, 0, \dots, 0, 1)$, given a coordinate chart (U, φ) we can obtain a canonical basis in these coordinates for the tangent space T_pM at a point $p \in U$. We can do this by seeing to which vectors in T_pM that $e_1, \dots, e_m$ correspond to by applying the inverse of the derivative $(T_p\varphi)^{-1} : \mathbb{R}^m \to T_pM$. These vectors will be called the coordinate vectors of the given coordinate chart.

Definition 2.5.6 (Canonical basis of $\mathbb{R}^m$)**.** The canonical vector space basis of $\mathbb{R}^m$ consists of the vectors $e_1 = (1, 0, \dots, 0, 0)$, $\dots$, $e_m = (0, 0, \dots, 0, 1)$.

Definition 2.5.7 (Coordinate vectors)**.** The coordinate vectors $\frac{\partial}{\partial\varphi_1}|_p$, $\dots$, $\frac{\partial}{\partial\varphi_m}|_p$ of a coordinate chart (U, φ) at a point $p \in U$ are defined by

$$\frac{\partial}{\partial\varphi_i}|_p := (T_p\varphi)^{-1}(e_i).$$

Remark. In the Euclidean space $\mathbb{E}^m$, the canonical coordinate chart $(\mathbb{R}^m, \mathrm{id})$, with $\mathrm{id} = (x_1, \dots, x_m)$ gives

$$\frac{\partial}{\partial x_i}|_p := (T_p \operatorname{id}_{\mathbb{R}^m})^{-1}(e_i).$$

So we have a canonical identification of $\frac{\partial}{\partial x_i}|_p$ with e_i, for each point $p \in \mathbb{R}^m$.

Remark. If $m = 1$ then there is just one basis vector for $\mathbb{R}^1$, $e_1 = 1$, and we can denote the coordinate vectors by

$$\frac{d}{d\varphi}|_p := (T_p\varphi)^{-1}(1) \quad \text{and} \quad \frac{d}{dx}|_p := (T_p \operatorname{id}_{\mathbb{R}^1})^{-1}(1).$$

Remark. For now it may seem strange that the vectors are denoted by what looks like partial derivative operators, but we will soon see that there is a natural correspondence between vectors and first order partial derivatives, motivating our notation.

Lemma 2.5.8 (Coordinate basis)**.** *For any coordinate chart (U, φ) with $p \in U$ the coordinate vectors $\frac{\partial}{\partial\varphi_1}|_p$, $\dots$, $\frac{\partial}{\partial\varphi_m}|_p$ form a basis for the vector space T_pM. This will be called the* coordinate basis *of the tangent space for this coordinate chart.*

Remark. Lemma 2.5.8 means that given a coordinate chart (U, φ), any vector $V_p \in T_pM$ has unique coordinates $v_{\varphi_i}(p) \in \mathbb{R}$ such that

$$V_p = \sum_{i=1}^{m} v_{\varphi_i}(p) \frac{\partial}{\partial\varphi_i}|_p.$$

In other words, if $\mathcal{B} = \left(\frac{\partial}{\partial\varphi_1}|_p, \dots, \frac{\partial}{\partial\varphi_m}|_p\right)$ is the chosen basis for T_pM, then the vector V_p has coordinates

$$[V_p]_{\mathcal{B}} = \begin{bmatrix} v_{\varphi_1}(p) \\ \vdots \\ v_{\varphi_m}(p) \end{bmatrix}$$

in this basis.

Proof. Since $T_p\varphi : T_pM \to \mathbb{R}^m$ is an isomorphism of vector spaces, and $(e_1, \dots, e_m)$ is a basis for $\mathbb{R}^m$, the coordinate basis is a basis for T_pM. □

2.5.3 Derivative

Definition 2.5.8 (Derivative at a point)**.** The *derivative* D_pf (also called the Jacobian, or the tangent map) of a smooth function $f : M \to N$ at a point $p \in M$ is the function

$$D_pf : T_pM \to T_{f(p)}N, \quad [c] \mapsto [f \circ c]$$

sending the vector at $p \in M$ corresponding to the curve c on M to the vector at $f(p) \in N$ corresponding to the curve $f \circ c$ on N.

Lemma 2.5.9. *The function $D_pf : T_pM \to T_{f(p)}N$ is well-defined and linear.*

Proof. Indeed, let (U, φ) be an arbitrary coordinate chart on M with $p \in U$ and (V, ψ) be an arbitrary coordinate chart on N with $f(p) \in V$. Then

$$\begin{aligned} D_pf([c]) &= [f \circ c] \\ &= (T_{f(p)}\psi)^{-1} \circ T_{f(p)}\psi([f \circ c]) \\ &= (T_{f(p)}\psi)^{-1}((\psi \circ f \circ c)'(0)) \\ &= (T_{f(p)}\psi)^{-1}((\psi \circ f \circ \varphi^{-1} \circ \varphi \circ c)'(0)) \\ &= (T_{f(p)}\psi)^{-1}((\psi \circ f \circ \varphi^{-1})'|_{\varphi(p)} \cdot (\varphi \circ c)'(0)) \\ &= (T_{f(p)}\psi)^{-1}((\psi \circ f \circ \varphi^{-1})'|_{\varphi(p)} \cdot T_p\varphi([c])) \\ &= ((T_{f(p)}\psi)^{-1} \circ (\psi \circ f \circ \varphi^{-1})'|_{\varphi(p)} \circ T_p\varphi)([c]). \end{aligned} \tag{2.18}$$

Here we observe that $D_pf = (T_{f(p)}\psi)^{-1} \circ (\psi \circ f \circ \varphi^{-1})'|_{\varphi(p)} \circ T_p\varphi$ is the composition of three well-defined linear functions, and therefore D_pf is also linear and well-defined. □

Lemma 2.5.10 (Chain rule)**.** *Given two smooth functions $f : M \to N$, $g : N \to L$, one has*

$$D_p(g \circ f) = (D_{f(p)}g) \circ D_pf.$$

Proof. This follows from the next computation, using the fact that composition of functions is associative.

$$D_p(g \circ f)([c]) = [g \circ f \circ c] = D_{f(p)}g([f \circ c]) = D_{f(p)}g(D_pf[c]) = ((D_{f(p)}g) \circ D_pf)([c]). \tag{2.19}$$

□

Lemma 2.5.11 (Derivative of diffeomorphism is isomorphism)**.** *Let $f : M \to N$ be a diffeomorphism. Then $D_pf : T_pM \to T_{f(p)}N$ is an isomorphism, with inverse $D_{f(p)}f^{-1} : T_{f(p)}N \to T_pM$.*

Proof. By the chain rule

$$D_{f(p)}f^{-1} \circ D_pf = D_p \operatorname{id}_M = \operatorname{id}_{T_pM}, \quad D_pf \circ D_{f(p)}f^{-1} = D_{f(p)} \operatorname{id}_N = \operatorname{id}_{T_{f(p)}N}, \tag{2.20}$$

so D_pf has the inverse $D_{f(p)}f^{-1}$, and since $D_{f(p)}f^{-1}$ is linear, D_pf is an isomorphism. □

Proposition 2.5.12 (Derivative of coordinate chart is the derivative of a smooth function)**.** *Under the identification of Proposition 2.5.7, one has*

$$T_p\varphi \cong D_p\varphi.$$

Proof. By Lemma 2.4.7, $\varphi : M \to \mathbb{R}^m$ is a smooth function between M and the Euclidean space $\mathbb{E}^m$. Therefore we can compute

$$(T_{\varphi(p)} \operatorname{id}) \circ D_p\varphi([c]) = T_{\varphi(p)} \operatorname{id}([\varphi \circ c]) = (\varphi \circ c)'(0) = T_p\varphi([c]). \tag{2.21}$$

□

Proposition 2.5.13 (Derivative is the Jacobian)**.** *For a smooth function $f : \mathbb{R}^m \to \mathbb{R}^n$,*

$$D_p f \cong f'(p)$$

under the canonical identification of Proposition 2.5.7.

Proof. Under the canonical coordinate chart $(\mathbb{R}^m, \mathrm{id}_{\mathbb{R}^m})$ and $(\mathbb{R}^n, \mathrm{id}_{\mathbb{R}^n})$ we obtain from (2.18) that $D_p f = (T_{f(p)} \mathrm{id}_{\mathbb{R}^n})^{-1} \circ f'|_p \circ T_p \mathrm{id}_{\mathbb{R}^m}$. That is, for $v \in \mathbb{R}^n$

$$T_{f(p)} \mathrm{id}_{\mathbb{R}^n} \circ D_p f \circ (T_p \mathrm{id}_{\mathbb{R}^m})^{-1}(v) = f'(p) \cdot v. \tag{2.22}$$

□

Lemma 2.5.14 (Formula for the derivative)**.** *Let $f : M \to N$ be a smooth function and (U, φ) with $p \in U$ and (V, ψ) with $f(p) \in V$ be two coordinate charts.*

The matrix representation of the linear function $D_p f : T_p M \to T_{f(p)} N$ in the respective coordinate bases is given by the Jacobian $(\psi \circ f \circ \varphi^{-1})'|_{\varphi(p)}$. That is

$$D_p f \left(\frac{\partial}{\partial \varphi_i}|_p \right) = \sum_{j=1}^{n} \frac{\partial(\psi_j \circ f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)) \; \frac{\partial}{\partial \psi_j}|_{f(p)}.$$

Remark. The jth component function of $\psi \circ f \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ is $\psi_j \circ f \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}$, where ψ_j is the jth component function of $\psi : V \to \mathbb{R}^n$. Moreover $\frac{\partial(\psi_j \circ f \circ \varphi^{-1})}{\partial x_i} : \mathbb{R}^m \to \mathbb{R}$ is the standard partial derivative with respect to the ith variable.

Remark. Another formulation of Lemma 2.5.14 is as follows. Denote these coordinate bases by $\mathcal{B} = \left(\frac{\partial}{\partial \varphi_1}|_p, \ldots, \frac{\partial}{\partial \varphi_m}|_p \right)$ for the vector space $T_p M$ and $\mathcal{C} = \left(\frac{\partial}{\partial \psi_1}|_{f(p)}, \ldots, \frac{\partial}{\partial \psi_n}|_{f(p)} \right)$ for $T_{f(p)} N$. Then the matrix for $D_p f$ is given by

$$[D_p f]^{\mathcal{C}}_{\mathcal{B}} = (\psi \circ f \circ \varphi^{-1})'(\varphi(p)) = \begin{bmatrix} \frac{\partial(\psi_1 \circ f \circ \varphi^{-1})}{\partial x_1}(\varphi(p)) & \ldots & \frac{\partial(\psi_1 \circ f \circ \varphi^{-1})}{\partial x_m}(\varphi(p)) \\ \vdots & \ddots & \vdots \\ \frac{\partial(\psi_n \circ f \circ \varphi^{-1})}{\partial x_1}(\varphi(p)) & \ldots & \frac{\partial(\psi_n \circ f \circ \varphi^{-1})}{\partial x_m}(\varphi(p)) \end{bmatrix}.$$

Proof. From (2.18), using the fact that $(T_{f(p)}\psi)^{-1}$ is linear we obtain

$$\begin{aligned}
D_p f \left(\frac{\partial}{\partial \varphi_i}|_p \right) &= (T_{f(p)}\psi)^{-1}((\psi \circ f \circ \varphi^{-1})'|_{\varphi(p)} \cdot T_p \varphi \left(\frac{\partial}{\partial \varphi_i}|_p \right)) \\
&= (T_{f(p)}\psi)^{-1}((\psi \circ f \circ \varphi^{-1})'|_{\varphi(p)} \cdot e_i) \\
&= (T_{f(p)}\psi)^{-1} \begin{bmatrix} \frac{\partial(\psi_1 \circ f \circ \varphi^{-1})}{\partial x_1}(\varphi(p)) & \ldots & \frac{\partial(\psi_1 \circ f \circ \varphi^{-1})}{\partial x_m}(\varphi(p)) \\ \vdots & \ddots & \vdots \\ \frac{\partial(\psi_n \circ f \circ \varphi^{-1})}{\partial x_1}(\varphi(p)) & \ldots & \frac{\partial(\psi_n \circ f \circ \varphi^{-1})}{\partial x_m}(\varphi(p)) \end{bmatrix} e_i \\
&= (T_{f(p)}\psi)^{-1} \begin{bmatrix} \frac{\partial(\psi_1 \circ f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)) \\ \vdots \\ \frac{\partial(\psi_n \circ f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)) \end{bmatrix} \\
&= (T_{f(p)}\psi)^{-1} \left(\sum_{j=1}^{n} \frac{\partial(\psi_j \circ f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)) \; e_j \right) \\
&= \sum_{j=1}^{n} \frac{\partial(\psi_j \circ f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)) \; (T_{f(p)}\psi)^{-1}(e_j) \\
&= \sum_{j=1}^{n} \frac{\partial(\psi_j \circ f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)) \; \frac{\partial}{\partial \psi_j}|_{f(p)}.
\end{aligned} \tag{2.23}$$

□

2.6 Vector fields

2.6.1 Definitions and local expressions

We note that the information concerning a tangent vector is two-fold. The first part is which point it is tangent to the manifold (which tangent space it belongs to), and the second part is what vector we consider. We collect the information about tangent vectors on a manifold into its so-called tangent bundle.

Definition 2.6.1 (Tangent bundle)**.** The *tangent bundle* TM of a manifold M is the set

$$TM = \{(p, v) : p \in M, v \in T_pM\} = \bigcup_{p\in M} \{p\} \times T_pM = \bigsqcup_{p\in M} T_pM.$$

Remark. The tangent bundle has a natural topology and smooth structure inherited from the manifold, making the tangent bundle itself into a smooth manifold. The dimension of the tangent bundle is twice that of the manifold, accounting both for the dimension of the base and vectors at each point of the tangent bundle.

Remark. The tangent bundle of a product manifold $M \times N$ is

$$T(M \times N) = \bigsqcup_{(p,q)\in M\times N} T_{(p,q)}(M \times N) \cong \bigsqcup_{(p,q)\in M\times N} T_pM \oplus T_qN =: TM \oplus TN$$

where $T_pM \oplus T_qN = T_pM \times T_qN$ as a set, and the $\oplus$ is to emphasize that the vector space structure is defined component-wise, so $\dim(T_pM \oplus T_qN) = \dim(T_pM) + \dim(T_qN)$.

The derivative can naturally be extended to a function between tangent bundles as follows.

Definition 2.6.2 (Derivative)**.** The *derivative* Df of a smooth function $f : M \to N$ is the function

$$Df : TM \to TN, \quad (p, v) \mapsto (f(p), D_pf(v)).$$

Remark. For smooth functions $c : I \to M$ (where I is any open subset of $\mathbb{E}^1$), there is a canonical way for the derivative to associate a vector in $T_{c(t)}M$ for each $t \in I$. Firstly the canonical identification $T_tI \cong \mathbb{R}$ is made so that $TI \cong I \times \mathbb{R}$ and we obtain a function

$$I \times \mathbb{R} \to TM, \quad (t, r) \mapsto (c(t), D_tc \cdot r),$$

which we identify with the function

$$\dot{c} : I \to TM, \quad t \mapsto (c(t), D_tc),$$

where the derivative D_tc has been considered as the vector $D_tc((D_t \operatorname{id}_I)^{-1}(1)) \in T_{c(t)}M$.

Definition 2.6.3 (Vector fields)**.** A *vector field* V is a section of the tangent bundle, meaning that it is a function of the form

$$V : M \to TM, \quad p \mapsto V(p) = (p, V_p),$$

where V_p is some vector $V_p \in T_pM$.

Definition 2.6.4 (Coordinate vector fields)**.** Given a coordinate chart (U, φ) we obtain vector fields

$$\frac{\partial}{\partial\varphi_i} : U \to TU, \quad p \mapsto \left(p, \frac{\partial}{\partial\varphi_i}|_p\right)$$

which will be called the *coordinate vector fields* of (U, φ).

Definition 2.6.5 (Local expression for vector fields)**.** Given a vector field $V : M \to TM$, a coordinate chart (U, φ) and a point $p \in U$, we can express the vector $V_p \in T_pM$ as

$$V_p = \sum_{i=1}^{m} v_{\varphi_i}(p) \frac{\partial}{\partial \varphi_i}|_p$$

or, equivalently,

$$V|_U = \sum_{i=1}^{m} v_{\varphi_i} \frac{\partial}{\partial \varphi_i}.$$

This $V|_U$ will be called the local expression for V given the coordinate chart (U, φ).

Remark. The coordinates v_{φ_i} for V will possibly be different for each point, so they will be functions

$$v_{\varphi_i} : U \to \mathbb{R}.$$

Definition 2.6.6 (Smooth vector fields)**.** A smooth vector field V is a vector field $V : M \to TM$ which locally has coordinates given by smooth functions. That is, for every $p \in M$ there exists a coordinate chart (U, φ) with $p \in U$ such that if $V|_U = \sum_{i=1}^{m} v_{\varphi_i} \frac{\partial}{\partial \varphi_i}$ then $v_{\varphi_i} : U \to \mathbb{R}$ are smooth functions.

Example 20 (Ordinary differential equation (ODE) of order one)**.** Every vector field $V : M \to TM$ encodes an *ODE* of order one, which is an expression of the form

$$D_t c = V_{c(t)}$$

or, equivalently,

$$\dot{c} = V \circ c$$

where c an unknown smooth function $I \to M$.

Lemma 2.6.1 (Complete flows on compact manifolds)**.** *If M is a compact smooth manifold and $V : M \to TM$ is a smooth vector field, then for any $p \in M$ there exists a unique smooth solution $c : \mathbb{R} \to M$ to $\dot{c} = V \circ c$ with $c(0) = p$.*

Proof. See Theorem 9.12(a) and Corollary 9.17 of [Lee13]. □

2.6.2 Directional derivatives

Definition 2.6.7 (Directional derivative)**.** The *directional derivative* $V(f)$ of a smooth function $f : M \to \mathbb{R}$ in the direction of the vector field $V : M \to TM$ is the function

$$V(f) : M \to \mathbb{R}, \quad p \mapsto (D_p f(V_p)).$$

Remark. Here $D_p f(V_p)$ has been identified with $(T_{f(p)} \mathrm{id}_{\mathbb{R}})(D_p f(V_p))$, which means that $V(f)$ is the vector component of the composition $Df \circ V$.

Lemma 2.6.2 (Local expression for directional derivative)**.** *Given a smooth function $f : M \to \mathbb{R}$ and a vector field $V : M \to TM$ the local expression for the function $V(f)$ in a given coordinate chart (U, φ) is given by*

$$V(f)|_U : U \to \mathbb{R}, \quad p \mapsto \sum_{i=1}^{m} v_{\varphi_i}(p) \frac{\partial (f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)).$$

Remark. Note that $f \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}$, and that $\frac{\partial(f\circ\varphi^{-1})}{\partial x_i}$ is the standard partial derivative with respect to the ith variable.

Proof. This follows from a computation, using the expression in Lemma 2.5.14 and applying the linearity of the derivative. For $p \in U$

$$\begin{aligned}
(V(f))(p) &= (T_{f(p)}\,\mathrm{id}_{\mathbb{R}})(D_p f(V_p)) \\
&= (T_{f(p)}\,\mathrm{id}_{\mathbb{R}})(D_p f\left(\sum_{i=1}^{m} v_{\varphi_i}(p)\frac{\partial}{\partial\varphi_i}|_p\right)) \\
&= \sum_{i=1}^{m} v_{\varphi_i}(p)\ (T_{f(p)}\,\mathrm{id}_{\mathbb{R}})(D_p f\left(\frac{\partial}{\partial\varphi_i}|_p\right)) \\
&= \sum_{i=1}^{m} v_{\varphi_i}(p)\ (T_{f(p)}\,\mathrm{id}_{\mathbb{R}})\left(\frac{\partial(\mathrm{id}_{\mathbb{R}}\circ f\circ\varphi^{-1})}{\partial x_i}(\varphi(p))\ \frac{d}{dx}|_{f(p)}\right) \\
&= \sum_{i=1}^{m} v_{\varphi_i}(p)\frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \cdot T_{f(p)}\,\mathrm{id}_{\mathbb{R}}\left(\frac{d}{dx}|_{f(p)}\right) \\
&= \sum_{i=1}^{m} v_{\varphi_i}(p)\frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \cdot 1 \\
&= \sum_{i=1}^{m} v_{\varphi_i}(p)\frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)).
\end{aligned} \tag{2.24}$$

□

Remark. In particular, considering the restriction $f|_U : U \to \mathbb{R}$ we obtain that

$$\left(\frac{\partial}{\partial\varphi_i}(f|_U)\right)(p) = \frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)),$$

where this smooth function will naturally be denoted by

$$\frac{\partial f}{\partial\varphi_i} : U \to \mathbb{R}, \quad p \mapsto \frac{\partial f}{\partial\varphi_i}(p) := \frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)).$$

Therefore we obtain the simpler formula

$$V(f)|_U : U \to \mathbb{R}, \quad p \mapsto \sum_{i=1}^{m} v_{\varphi_i}(p)\frac{\partial f}{\partial\varphi_i}(p).$$

Corollary 2.6.3 (Directional derivative is smooth)**.** *If $f : M \to \mathbb{R}$ is a smooth function and $V : M \to TM$ is a smooth vector field, then the directional derivative $V(f) : M \to \mathbb{R}$ is a smooth function.*

Proof. Using the coordinate chart (U, φ) with $p \in U$ on M and the standard coordinate chart $(\mathbb{R}^1, \mathrm{id}_{\mathbb{R}^1})$ on $\mathbb{E}^1$ means that the corresponding expression for $V(f)$ in Euclidean space becomes

$$\mathrm{id}_{\mathbb{R}^1} \circ V(f) \circ \varphi^{-1} = V(f) \circ \varphi^{-1} = \sum_{i=1}^{m} v_{\varphi_i} \circ \varphi^{-1} \cdot \frac{\partial(f\circ\varphi^{-1})}{\partial x_i}, \tag{2.25}$$

where the smoothness of $f : M \to \mathbb{R}$ means that $\frac{\partial(f\circ\varphi^{-1})}{\partial x_i}$ is smooth. Therefore $V(f) \circ \varphi^{-1}$ is smooth if $v_{\varphi_i} \circ \varphi^{-1}$ is, which is true if v_{φ_i} is smooth, by Lemma 2.4.8. □

Definition 2.6.8 (Vector fields as derivations)**.** By Corollary 2.6.3, smooth vector fields $V \in \Gamma(TM)$ can be seen as functions

$$V : C^\infty(M) \to C^\infty(M) \quad f \mapsto V(f).$$

Here $\Gamma(TM)$ denotes the set of smooth vector fields $V : M \to TM$ and $C^\infty(M)$ denotes the set of smooth functions $f : M \to \mathbb{R}$.

Corollary 2.6.4 (Vector fields are first order linear differential operators)**.** *For a smooth function $f : \mathbb{R}^m \to \mathbb{R}$ in Euclidean space, the vector field V is the first order linear partial derivative operator given by*

$$V = g_1 \frac{\partial}{\partial x_1} + \cdots + g_m \frac{\partial}{\partial x_m} \mapsto V(f) = g_1 \frac{\partial f}{\partial x_1} + \cdots + g_m \frac{\partial f}{\partial x_m}.$$

Proof. This follows from Lemma 2.6.2, using the standard coordinate chart $(\mathbb{R}^m, \mathrm{id}_{\mathbb{R}^m})$, where then $V(f)|_{\mathbb{R}^m} = V(f)$ and $f \circ \mathrm{id}_{\mathbb{R}^m}^{-1} = f$, and the coordinates v_{φ_i} are smooth functions $g_i : \mathbb{R}^n \to \mathbb{R}$. □

Lemma 2.6.5 (Sum and product of smooth functions)**.** *Given two smooth functions $f, g : M \to \mathbb{R}$, their sum and their product*

$$\begin{aligned} f + g : M \to \mathbb{R}, \quad p \mapsto f(p) + g(p) \\ f \cdot g : M \to \mathbb{R}, \quad p \mapsto f(p) \cdot g(p) \end{aligned}$$

are smooth functions.

Proof. Let $p \in M$ be an arbitrary point, and let (U, φ) be an arbitrary coordinate chart with $p \in U$. Then in Euclidean space $f \circ \varphi^{-1}$ and $g \circ \varphi^{-1}$ are infinitely differentiable by smoothness of f and g, and so by the linearity of the derivative, we know that $(f+g) \circ \varphi^{-1} = f \circ \varphi^{-1} + g \circ \varphi^{-1}$ can be differentiated infinitely many times

$$(f \circ \varphi^{-1} + g \circ \varphi^{-1})' = (f \circ \varphi^{-1})' + (g \circ \varphi^{-1})' \tag{2.26}$$

and similarly $(f \cdot g) \circ \varphi^{-1} = (g \circ \varphi^{-1}) \cdot (f \circ \varphi^{-1})$ can be differentiated infinitely many times. Indeed, by the product rule

$$((g \circ \varphi^{-1}) \cdot (f \circ \varphi^{-1}))' = (g \circ \varphi^{-1})' \cdot (f \circ \varphi^{-1}) + (g \circ \varphi^{-1}) \cdot (f \circ \varphi^{-1})'. \tag{2.27}$$

Hence $f \cdot g$ and $f + g$ can be differentiated infinitely many times in Euclidean space, and are hence smooth functions. □

Lemma 2.6.6 (Linearity of vector fields)**.** *Vector fields $V : C^\infty(M) \to C^\infty(M)$ are $\mathbb{R}$-linear, that is*

$$V(f + g) = V(f) + V(g) \quad \textit{and} \quad V(rf) = rV(f)$$

for smooth functions $f, g : M \to \mathbb{R}$ and real numbers $r \in \mathbb{R}$.

Proof. This follows from the linearity of partial derivatives in Euclidean space. We can see this

by letting $p \in M$ be arbitrary and (U, φ) be an arbitrary chart with $p \in U$, meaning that

$$\begin{aligned} V(rf+g)(p) &= \sum_{i=1}^{m} v_{\varphi_i}(p) \frac{\partial((rf+g)\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \\ &= \sum_{i=1}^{m} v_{\varphi_i}(p) \frac{\partial(r\cdot f\circ\varphi^{-1} + g\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \\ &= \sum_{i=1}^{m} v_{\varphi_i}(p) \left(r\cdot \frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) + \frac{\partial(g\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \right) \\ &= r\cdot \sum_{i=1}^{m} v_{\varphi_i}(p) \cdot \frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) + \sum_{j=1}^{m} v_{\varphi_i}(p) \frac{\partial(g\circ\varphi^{-1})}{\partial x_j}(\varphi(p)) \\ &= rV(f)(p) + V(g)(p), \end{aligned} \tag{2.28}$$

where setting $r = 1$ and $g \equiv 0$ respectively give $V(f+g) = V(f) + V(g)$ and $V(rf) = rV(f)$, since the above equality was for an arbitrary $p \in M$. □

Lemma 2.6.7 (Product rule)**.** *Given two smooth functions $f, g : M \to \mathbb{R}$, the product*

$$f \cdot g : M \to \mathbb{R}, \quad p \mapsto f(p)\cdot g(p)$$

obeys the product rule when applying the directional derivative in the direction of a vector field $V : M \to TM$, being

$$V(f\cdot g) = V(f)\cdot g + f\cdot V(g).$$

Proof. This follows from the product rule in Euclidean space by the following computation. For any $p \in M$, let (U, φ) be an arbitrary coordinate chart with $p \in U$ and observe that

$$\begin{aligned} &V(f\cdot g)(p) \\ &= \sum_{i=1}^{m} v_{\varphi_i}(p) \cdot \frac{\partial((f\cdot g)\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \\ &= \sum_{i=1}^{m} v_{\varphi_i}(p) \cdot \frac{\partial((f\circ\varphi^{-1})\cdot(g\circ\varphi^{-1}))}{\partial x_i}(\varphi(p)) \\ &= \sum_{i=1}^{m} v_{\varphi_i}(p) \cdot \left(\frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p))\cdot(g\circ\varphi^{-1})(\varphi(p)) + (f\circ\varphi^{-1})(\varphi(p))\cdot\frac{\partial(g\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \right) \\ &= \sum_{i=1}^{m} v_{\varphi_i}(p) \cdot \left(\frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p))\cdot g(p) + f(p)\cdot\frac{\partial(g\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \right) \\ &= \sum_{i=1}^{m} v_{\varphi_i}(p) \cdot \frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p))\cdot g(p) + f(p)\cdot v_{\varphi_i}(p)\cdot\frac{\partial(g\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \\ &= \left(\sum_{i=1}^{m} v_{\varphi_i}(p) \cdot \frac{\partial(f\circ\varphi^{-1})}{\partial x_i}(\varphi(p)) \right)\cdot g(p) + f(p)\cdot\left(\sum_{j=1}^{m} v_{\varphi_i}(p)\cdot\frac{\partial(g\circ\varphi^{-1})}{\partial x_j}(\varphi(p)) \right) \\ &= V(f)(p)\cdot g(p) + f(p)\cdot V(g)(p). \end{aligned} \tag{2.29}$$

Because this holds for any $p \in M$, the result follows. □

2.6.3 Composition of vector fields

Because vector fields are functions $V, W : C^\infty(M) \to C^\infty(M)$ this means that we can compose two vector fields V, W in order to obtain a new function $W \circ V : C^\infty(M) \to C^\infty(M)$.

Lemma 2.6.8 (Local expression for the composition of vector fields)**.** *Given a smooth function $f : M \to \mathbb{R}$ and two smooth vector fields $V, W : C^\infty(M) \to C^\infty(M)$, in a given coordinate chart (U, φ) the local expression for the function $(W \circ V)(f)$ is given by*

$$(W \circ V)(f)|_U = \sum_{i,j=1}^{m} w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j} \frac{\partial f}{\partial \varphi_i} + \sum_{k,l=1}^{m} v_{\varphi_k} w_{\varphi_l} \frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l} : U \to \mathbb{R}.$$

Remark. Here

$$\frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l} := \frac{\partial^2}{\partial \varphi_k \partial \varphi_l}(f|_U) := \left(\frac{\partial}{\partial \varphi_k} \circ \frac{\partial}{\partial \varphi_l} \right)(f|_U) = \frac{\partial}{\partial \varphi_k} \left(\frac{\partial f}{\partial \varphi_l} \right) : U \to \mathbb{R}, \tag{2.30}$$

where then

$$\begin{aligned} \frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l}(p) &= \frac{\partial}{\partial \varphi_k} \left(\frac{\partial f}{\partial \varphi_l} \right)(p) \\ &= \frac{\partial}{\partial \varphi_k} \left(\frac{\partial (f \circ \varphi^{-1})}{\partial x_l} \circ \varphi \right)(p) \\ &= \left(\frac{\partial}{\partial x_k} \left(\left(\frac{\partial (f \circ \varphi^{-1})}{\partial x_l} \circ \varphi \right) \circ \varphi^{-1} \right) \right)(\varphi(p)) \\ &= \left(\frac{\partial}{\partial x_k} \left(\frac{\partial (f \circ \varphi^{-1})}{\partial x_l} \right) \right)(\varphi(p)) \\ &= \frac{\partial^2 (f \circ \varphi^{-1})}{\partial x_k \partial x_l}(\varphi(p)). \end{aligned} \tag{2.31}$$

Because $f \circ \varphi^{-1} : \mathbb{R}^m \to \mathbb{R}$ is smooth we obtain

$$\frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l}(p) = \frac{\partial^2 (f \circ \varphi^{-1})}{\partial x_k \partial x_l}(\varphi(p)) = \frac{\partial^2 (f \circ \varphi^{-1})}{\partial x_l \partial x_k}(\varphi(p)) = \frac{\partial^2 f}{\partial \varphi_l \partial \varphi_k}(p). \tag{2.32}$$

Proof. Using linearity and the product rule for vector fields, we obtain

$$\begin{aligned} (W \circ V)(f)|_U &= W|_U (V|_U (f|_U)) \\ &= W|_U \left(\sum_{i=1}^{m} v_{\varphi_i} \cdot \frac{\partial f}{\partial \varphi_i} \right) \\ &= \sum_{i=1}^{m} W|_U \left(v_{\varphi_i} \cdot \frac{\partial f}{\partial \varphi_i} \right) \\ &= \sum_{i=1}^{m} W|_U (v_{\varphi_i}) \cdot \frac{\partial f}{\partial \varphi_i} + v_{\varphi_i} \cdot W|_U \left(\frac{\partial f}{\partial \varphi_i} \right) \\ &= \sum_{i=1}^{m} W|_U (v_{\varphi_i}) \cdot \frac{\partial f}{\partial \varphi_i} + \sum_{k=1}^{m} v_{\varphi_k} \cdot W|_U \left(\frac{\partial f}{\partial \varphi_k} \right), \end{aligned} \tag{2.33}$$

where

$$W|_U (v_{\varphi_i}) = \sum_{j=1}^{m} w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j} \quad \text{and} \quad W|_U \left(\frac{\partial f}{\partial \varphi_k} \right) = \sum_{l=1}^{m} w_{\varphi_l} \frac{\partial}{\partial \varphi_l} \left(\frac{\partial f}{\partial \varphi_k} \right) = \sum_{l=1}^{m} w_{\varphi_l} \frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l} \tag{2.34}$$

giving us the desired result. □

Remark. The standard way of computing partial derivative operators, is applying them from left-to-right, distributing over addition and applying the product rule when necessary, where a partial derivative applied to another becomes a higher order partial derivative where the order of differentiation does not matter because it will be applied to a smooth function. In other words,

$$\begin{aligned}(W \circ V)|_U &= \left(\sum_{j=1}^m w_{\varphi_j} \frac{\partial}{\partial \varphi_j}\right) \circ \left(\sum_{i=1}^m v_{\varphi_i} \frac{\partial}{\partial \varphi_i}\right) \\ &= \sum_{i,j=1}^m w_{\varphi_j} \frac{\partial}{\partial \varphi_j} \circ \left(v_{\varphi_i} \frac{\partial}{\partial \varphi_i}\right) \\ &= \sum_{i,j=1}^m w_{\varphi_j} \left(\frac{\partial v_{\varphi_i}}{\partial \varphi_j} \frac{\partial}{\partial \varphi_i} + v_{\varphi_i} \frac{\partial^2}{\partial \varphi_j \partial \varphi_i}\right) \\ &= \sum_{i,j=1}^m w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j} \frac{\partial}{\partial \varphi_i} + v_{\varphi_i} w_{\varphi_j} \frac{\partial^2}{\partial \varphi_j \partial \varphi_i} \\ &= \sum_{i,j=1}^m w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j} \frac{\partial}{\partial \varphi_i} + \sum_{k,l=1}^m v_{\varphi_k} w_{\varphi_l} \frac{\partial^2}{\partial \varphi_l \partial \varphi_k} \\ &= \sum_{i,j=1}^m w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j} \frac{\partial}{\partial \varphi_i} + \sum_{k,l=1}^m v_{\varphi_k} w_{\varphi_l} \frac{\partial^2}{\partial \varphi_k \partial \varphi_l}.\end{aligned} \tag{2.35}$$

This recipe can be used to obtain formulas for higher order partial derivatives $V_3 \circ V_2 \circ V_1$, etc.. We emphasize that higher order partial derivatives do not correspond to a single vector field. It is only first order partial derivative operators that correspond to vector fields.

Observe that the second order term is symmetric

$$\sum_{k,l=1}^m v_{\varphi_k} w_{\varphi_l} \frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l} = \sum_{k,l=1}^m w_{\varphi_k} v_{\varphi_l} \frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l} \tag{2.36}$$

but the first order term is not (necessarily)

$$\sum_{i,j=1}^m w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j} \frac{\partial f}{\partial \varphi_i} \neq \sum_{i,j=1}^m v_{\varphi_j} \frac{\partial w_{\varphi_i}}{\partial \varphi_j} \frac{\partial f}{\partial \varphi_i}. \tag{2.37}$$

This implies that in the expression for $(V \circ W)(f)|_U - (W \circ V)(f)|_U$, the second order term would disappear, and only a term of the first order would remain, being

$$\begin{aligned}(V \circ W)(f)|_U - (W \circ V)(f)|_U &= \sum_{i,j=1}^m v_{\varphi_j} \frac{\partial w_{\varphi_i}}{\partial \varphi_j} \frac{\partial f}{\partial \varphi_i} - w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j} \frac{\partial f}{\partial \varphi_i} \\ &= \sum_{i=1}^m \left(\sum_{j=1}^m v_{\varphi_j} \frac{\partial w_{\varphi_i}}{\partial \varphi_j} - w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j}\right) \frac{\partial f}{\partial \varphi_i}\end{aligned} \tag{2.38}$$

Because first order partial derivatives correspond to vector fields, this means that we obtain a *vector field* $[V, W]|_U = (V \circ W)|_U - (W \circ V)|_U$ (at least locally). That is

$$[V, W]|_U = \sum_{i=1}^m \left(\sum_{j=1}^m v_{\varphi_j} \frac{\partial w_{\varphi_i}}{\partial \varphi_j} - w_{\varphi_j} \frac{\partial v_{\varphi_i}}{\partial \varphi_j}\right) \frac{\partial}{\partial \varphi_i}. \tag{2.39}$$

It turns out that this is a vector field on all of M which is called the *Lie bracket* of V and W.

2.7 Dual spaces, tensors, and tensor fields

2.7.1 Dual spaces and 1-forms

Definition 2.7.1 (Dual space and dual vectors)**.** Given a vector space T with real numbers as scalars, the *dual space* of T is the vector space T^*, consisting of the linear functions $F : T \to \mathbb{R}$, with addition and scalar multiplication defined pointwise as

$$F_1 + F_2 : T \to \mathbb{R},\ v \mapsto F_1(v) + F_2(v) \quad \text{and} \quad rF_1 : T \to \mathbb{R},\ v \mapsto rF_1(v).$$

The elements of a dual vector space will be called *dual vectors*.

Lemma 2.7.1 (Dual basis)**.** *If T is a vector space of dimension m, then T^* is also a vector space of dimension m. In particular, if $(v_1, \dots, v_m)$ is a basis for T, then $(F_{v_1}, \dots, F_{v_m})$ is a basis for T^*, where these are functions defined by*

$$F_{v_i}(v_i) = 1, \quad F_{v_i}(v_j) = 0 \quad (\text{for } j \neq i),$$

and extending by linearity. $(F_{v_1}, \dots, F_{v_m})$ is known as the dual basis of $(v_1, \dots, v_m)$.

Remark. Here extending by linearity means that if $v = c_1v_1 + \cdots + c_{i-1}v_{i-1} + c_iv_i + c_{i+1}v_{i+1} + \cdots + c_mv_m$, then from just the linearity of F_{v_i} we obtain that

$$F_{v_i}(v) = c_1F_{v_i}(v_1) + \cdots + c_{i-1}F_{v_i}(v_{i-1}) + c_iF_{v_i}(v_i) + c_{i+1}F_{v_i}(v_{i+1}) + \cdots + c_mF_{v_i}(v_m) = c_i \quad (2.40)$$

is well-defined.

Proof. See Theorem 3.11 of [Rom08]. □

Definition 2.7.2 (Dual tangent bundle)**.** Given a smooth manifold M, the *dual tangent bundle* is the set

$$T^*M := \{(p, t) : p \in M, t \in T_p^*M\},$$

where here $T_p^*M := (T_pM)^*$.

Definition 2.7.3 (1-forms)**.** A *1-form* F is a function is a section of the dual tangent bundle, meaning that is a function of the form

$$F : M \to T^*M, \quad p \mapsto F(p) = (p, F_p),$$

where F_p is some linear function $F_p \in T_p^*M$.

Definition 2.7.4 (Coordinate 1-forms)**.** Given a coordinate chart (U, φ), the *coordinate 1-forms* $d\varphi_i$ are the functions

$$d\varphi_i : U \to T^*U, \quad p \mapsto (p, d\varphi_i|_p)$$

where $(d\varphi_1|_p, \dots, d\varphi_m|_p)$ is the dual basis of $\left(\frac{\partial}{\partial\varphi_1}|_p, \dots, \frac{\partial}{\partial\varphi_m}|_p\right)$.

Definition 2.7.5 (Smooth 1-forms)**.** A *smooth 1-form* F is a 1-form, where for any $p \in M$ there exists a coordinate chart (U, φ) with $p \in U$ such that if

$$F_p = \sum_{i=1}^{m} f_{\varphi_i}(p)\ d\varphi_i|_p$$

then $f_{\varphi_i} : U \to \mathbb{R}$ are smooth functions.

Definition 2.7.6 (Pairing of a 1-form and a vector field)**.** Given a smooth 1-form $F : M \to T^*M$ and a smooth vector field $V : M \to TM$, these can be paired into a smooth function

$$F(V) : M \to \mathbb{R}, \quad p \mapsto F_p(V_p)$$

In this sense smooth 1-forms can be seen as functions $F : \Gamma(TM) \to C^\infty(M),\ V \mapsto F(V)$ and smooth vector fields can be seen as functions $V : \Gamma(T^*M) \to C^\infty(M),\ F \mapsto F(V)$.

Lemma 2.7.2 (Local expression of pairing)**.** *The pairing of a smooth 1-form $F : M \to T^*M$ and a smooth vector field $V : M \to TM$ is a smooth function, and if (U, φ) is a coordinate chart, then the local expression for the pairing $F(V)$ is*

$$F(V)|_U : U \to \mathbb{R} \quad p \mapsto \sum_{i=1}^{m} f_{\varphi_i}(p) v_{\varphi_i}(p).$$

Proof. Let (U, φ) be a coordinate chart, and

$$F|_U = \sum_{i=1}^{m} f_{\varphi_i} d\varphi_i, \quad V|_U = \sum_{j=1}^{m} v_{\varphi_j} \frac{\partial}{\partial \varphi_j} \tag{2.41}$$

where the linearity of 1-forms (for every $p \in U$) means that

$$d\varphi_i \left(\sum_{j=1}^{m} v_{\varphi_j} \frac{\partial}{\partial \varphi_j} \right) = v_{\varphi_i} \tag{2.42}$$

and so

$$F(V)|_U = F|_U(V|_U) = \sum_{i=1}^{m} f_{\varphi_i} v_{\varphi_i} \tag{2.43}$$

as desired. □

2.7.2 Tensors

We will generalize the concept of dual vectors and 1-forms, and these objects will be called (a, b)-tensors and (a, b)-tensor fields.

Definition 2.7.7 ((a, b)-tensors)**.** An (a, b)-tensor t on a vector space T is a function

$$t : (T^*)^a \times T^b \to \mathbb{R},$$

taking a dual vectors and b vectors as input and outputting a real number, with the requirement that it should be *multilinear*.

Remark. We note that t takes $a + b$ total inputs, where multilinearity means that if all but one number of inputs $(a + b) - 1$ are specified, then the resulting function is linear.

Example 21 (Dual vectors)**.** Dual vectors F are $(0, 1)$-tensors, since they are linear functions $F : T \to \mathbb{R}$.

Example 22 (Scalar product)**.** The scalar product $\cdot : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R},\ (v, w) \mapsto v \bullet w$ is a $(0, 2)$-tensor on $\mathbb{R}^m$, because it is linear in the first argument $(rv) \bullet w = r(v \bullet w)$, and $(v_1 + v_2) \bullet w = v_1 \bullet w + v_2 \bullet w$, and similarly in the second argument.

Example 23 (Determinant)**.** The determinant of a matrix with variable columns $(v_1, \ldots, v_m) \mapsto \det \begin{bmatrix} v_1 & \ldots & v_m \end{bmatrix}$ is a $(0, m)$-tensor on $\mathbb{R}^m$.

Example 24 (Vector)**.** Given any vector $v \in T$, it can be seen as the $(1,0)$-tensor $v : T^* \to \mathbb{R},\ F \mapsto F(v)$, which we will define to be linear by setting $(rF)(v) := rF(v)$ and $(F_1+F_2)(v) := F_1(v) + F_2(v)$.

Definition 2.7.8 (Sum and product)**.** Given two (a,b)-tensors $s,t : (T^*)^a \times T^b \to \mathbb{R}$, we can define their sum and their product to be the sum and product

$$s+t : (T^*)^a \times T^b \to \mathbb{R}, \quad (F_1,.,F_a,v_1,.,v_b) \mapsto s(F_1,.,F_a,v_1,.,v_b) + t(F_1,.,F_a,v_1,.,v_b)$$

and

$$s \cdot t : (T^*)^a \times T^b \to \mathbb{R}, \quad (F_1,.,F_a,v_1,.,v_b) \mapsto s(F_1,.,F_a,v_1,.,v_b) \cdot t(F_1,.,F_a,v_1,.,v_b)$$

as real numbers $\mathbb{R}$.

Given an (a,b)-tensor s and (c,d)-tensor t, we note that s takes a dual vectors and b vectors as input, and t takes c dual vectors and d vectors as input. In total, s and t take $a+c$ dual vectors and $b+d$ vectors as input. Hence we can imagine combining s and t into a new $(a+c,b+d)$-tensor $s \otimes t$, by simply taking the first a dual vectors and the first b vectors, inputting this in s to get a real number r_1, then taking the remaining c dual vectors and d vectors, inputting this in t to get another real number r_2, and then combining r_1 and r_2 into a new real number such that this function $s \otimes t$ becomes multilinear. It turns out that simply multiplying these numbers $r_1 \cdot r_2$ does the trick. The resulting function $s \otimes t$ is known as the tensor product of s and t.

Definition 2.7.9 (Tensor product)**.** Given an (a,b)-tensor s and a (c,d)-tensor t on a vector space T, the *tensor product* $s \otimes t$ of s and t is the $(a+c,b+d)$-tensor given by

$$s \otimes t : (T^*)^{a+c} \times T^{b+d}, \quad (F_1,.,F_a,G_1,.G_c,v_1,.v_b,w_1,.,w_d) \mapsto s(F_1,.,F_a,v_1,.v_b) \cdot t(G_1,.G_c,w_1,.,w_d).$$

Example 25 (Scalar product)**.** It turns out that we can construct the scalar product on $\mathbb{R}^m$ using the tensor product. Given vectors in the standard basis $\mathcal{B} = (e_1,\ldots,e_m)$

$$[V]_{\mathcal{B}} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}, \quad [W]_{\mathcal{B}} = \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} \tag{2.44}$$

their scalar product is given by

$$V \bullet W = v_1 \cdot w_1 + \ldots v_m \cdot w_m. \tag{2.45}$$

Now denote the dual of the standard basis be $(F_{e_1},\ldots,F_{e_m})$. Then $F_{e_i}(V) = v_i$ and $F_{e_i}(W) = w_i$, so

$$V \bullet W = F_{e_1}(V) \cdot F_{e_1}(W) + \cdots + F_{e_m}(V) \cdot F_{e_m}(W). \tag{2.46}$$

But then we see that

$$F_{e_i} \otimes F_{e_i} : (V,W) \mapsto F_{e_i}(V) \cdot F_{e_i}(W) \tag{2.47}$$

and so $- \bullet - : (\mathbb{R}^m)^2 \to \mathbb{R}, \quad (V,W) \mapsto V \bullet W$ is the $(0,2)$-tensor

$$- \bullet - = \sum_{i=1}^{m} F_{e_i} \otimes F_{e_i}. \tag{2.48}$$

Definition 2.7.10 (Set of all (a,b)-tensors)**.** Given a vector space T, the set of (a,b)-tensors on T will be denoted by $T^{(a,b)}$, which becomes a vector space with the addition of (a,b)-tensors as defined above, and scalar multiplication defined by $rt : (T^*)^a \times T^b \to \mathbb{R}, \quad (F_1,.,F_a,v_1,.,v_b) \mapsto rt(F_1,.,F_a,v_1,.,v_b)$.

Remark. Observe that $T = T^{(1,0)}$, $T^* = T^{(0,1)}$, and we may define $T^{(0,0)} := \mathbb{R}$.

Lemma 2.7.3 (Basis of $T^{(a,b)}$)**.** *If T is a m-dimensional vector space, then $T^{(a,b)}$ is a m^{a+b}-dimensional vector space. Moreover, if $(v_1, \dots, v_m)$ is a basis for T and $(F_{v_1}, \dots, F_{v_m})$ its dual basis, then the m^{a+b} different tensors*

$$v_{i_1} \otimes \cdots \otimes v_{i_a} \otimes F_{j_1} \otimes \cdots \otimes F_{j_b}$$

where $i_1, \dots, i_a, j_1, \dots, j_b \in \{1, \dots, m\}$ form a basis for $T^{(a,b)}$.

Proof. See Corollary 12.12 of [Lee13]. □

2.7.3 Tensor fields

Definition 2.7.11 (Tensor bundle)**.** If M is a smooth manifold, we define its (a, b)-tensor bundle to be the set

$$T^{(a,b)}M := \{(p, t) : p \in M, t \in T_p^{(a,b)}M\}$$

where $T_p^{(a,b)}M := (T_pM)^{(a,b)}$.

Definition 2.7.12 ((a, b)-tensor fields)**.** An (a, b) tensor field F on a smooth manifold M is a function

$$F : M \to T^{(a,b)}M, \quad p \mapsto (p, t_p).$$

The sum and the product of (a, b)-tensor fields $F, G : M \to T^{(a,b)}M$ are defined by

$$F + G : M \to T^{(a,b)}M, \quad p \mapsto (p, F_p + G_p)$$
$$F \cdot G : M \to T^{(a,b)}M, \quad p \mapsto (p, F_p \cdot G_p)$$

and tensor product of an (a, b)-tensor field $F : M \to T^{(a,b)}M$ and a (c, d)-tensor field $G : M \to T^{(c,d)}M$ is defined by

$$F \otimes G : M \to T^{(a+c,b+d)}M, \quad p \mapsto (p, F_p \otimes G_p).$$

Definition 2.7.13 (Coordinate (a, b)-tensor fields)**.** Given a coordinate chart, (U, φ) the m^{a+b} different tensor fields

$$\frac{\partial}{\partial \varphi_{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial \varphi_{i_a}} \otimes d\varphi_{j_1} \cdots \otimes d\varphi_{j_b} : U \to T^{(a,b)}U$$

are known as the coordinate (a, b)-tensor fields for the coordinate chart (U, φ).

Lemma 2.7.4 (Local expression)**.** *Given any (a, b)-tensor field $F : M \to T^{(a,b)}M$, for a given coordinate chart (U, φ) it can locally be expressed as*

$$F|_U = \sum_{i_1,\dots,i_a,j_1,\dots,j_m=1}^{m} f_{\varphi_{i_1},\dots,\varphi_{i_m},\varphi_{j_1},\dots,\varphi_{j_m}} \frac{\partial}{\partial \varphi_{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial \varphi_{i_a}} \otimes d\varphi_{j_1} \cdots \otimes d\varphi_{j_b},$$

where $f_{\varphi_{i_1},\dots,\varphi_{i_m},\varphi_{j_1},\dots,\varphi_{j_m}} : U \to \mathbb{R}$.

Proof. Proof follows from the (a, b)-tensors $\frac{\partial}{\partial \varphi_{i_1}}|_p \otimes \cdots \otimes \frac{\partial}{\partial \varphi_{i_a}}|_p \otimes d\varphi_{j_1}|_p \cdots \otimes d\varphi_{j_b}|_p$ constituting a basis for $T_p^{(a,b)}M$ for every $p \in U$. □

Definition 2.7.14 (Smooth tensor fields)**.** A *smooth* (a,b)-tensor field $F : M \to T^{(a,b)}M$ on M is an (a,b)-tensor field on M such that for each point $p \in M$, there exists a coordinate chart (U,φ) with $p \in U$ such that the local expression $F|_U$ in this coordinate chart has smooth functions $f_{\varphi_{i_1},\dots,\varphi_{i_m},\varphi_{j_1},\dots,\varphi_{j_m}} : U \to \mathbb{R}$ as coordinates.

Remark. Let $\Gamma(T^{(a,b)}M)$ denote the set of smooth (a,b)-tensor fields $M \to T^{(a,b)}M$ on M. Then $\Gamma(T^{(0,0)}M)$ is the set of smooth functions, $\Gamma(T^{(1,0)}M)$ is the set of smooth vector fields, $\Gamma(T^{(0,1)}M)$ is the set of smooth 1-forms, and $\Gamma(T^{(0,2)}M)$ is the set of smooth *bilinear forms*, on M, which will appear in the next subsection.

Lemma 2.7.5 (Evaluation of tensor fields on 1-forms and vector fields)**.** *Given an (a,b)-tensor field G, a 1-forms $F_1,\dots,F_a$ and b vector fields $V_1,\dots,V_b$, their* evaluation *defines a function*

$$G(F_1,.,F_a,V_1,.V_b) : M \to \mathbb{R}, \quad p \mapsto G_p((F_1)_p,.,(F_a)_p,(V_1)_p,.,(V_b)_p)$$

by pairing For a given coordinate chart (U,φ) this is locally given by

$$G(F_1,.,F_a,V_1,.,V_b)|_U = \sum_{i_1,\dots,i_a,j_1,\dots,j_m=1}^{m} g_{\varphi_{i_1},\dots,\varphi_{i_m},\varphi_{j_1},\dots,\varphi_{j_m}} \cdot (f_1)_{\varphi_{i_1}} \cdot \cdot (f_a)_{\varphi_{i_a}} \cdot \cdot (v_1)_{\varphi_{j_1}} \cdot \cdot (v_b)_{\varphi_{j_m}}.$$

Hence $G(F_1,.,F_a,V_1,.,V_b)$ is a smooth function if G, F_1, …, F_a, V_1, …, V_b are smooth tensor fields.

Proof. This follows from the fact that tensor fields are multilinear at each point $p \in M$, and proceeding similarly to the proof of Lemma 2.7.2. □

Remark. Because tensor fields F are such that they are multilinear for each p and that the evaluation of a smooth tensor field with smooth 1-forms and smooth vector fields yields a smooth function, we obtain that smooth (a,b)-tensor fields can be viewed as functions

$$F : \Gamma(T^*M)^a \times \Gamma(TM)^b \to C^\infty(M)$$

that are $C^\infty(M)$-linear. We call an arbitrary function $F : \Gamma(T^*M)^a \times \Gamma(TM)^b \to C^\infty(M)$ that is $C^\infty(M)$-linear "tensorial", and in fact, any such function is necessarily a smooth (a,b)-tensor field $F \in \Gamma(T^{(a,b)}M)$ [Lee13, Lemma 12.24].

2.8 Riemannian geometry

2.8.1 Riemannian metrics

Definition 2.8.1 (Bilinear form)**.** A *bilinear form* is a $(0,2)$-tensor.

Remark. Let B be a bilinear form on the vector space T. If $\mathcal{B} = (V_1,\dots,V_m)$ is a vector space basis for T, we can gather the information about B in this basis by calculating $B(V_i,V_j) \in \mathbb{R}$, which we can put in a matrix

$$[B]_{\mathcal{B}} = \begin{bmatrix} B(V_1,V_1) & \dots & B(V_1,V_m) \\ \vdots & \ddots & \vdots \\ B(V_m,V_1) & \dots & B(V_m,V_m) \end{bmatrix}.$$

Moreover, if $R = \sum_{i=1}^m r_i V_i$ and $S = \sum_{i=1}^m s_i V_i$, then

$$B(R,S) = [R]_{\mathcal{B}}^{\mathsf{T}}[B]_{\mathcal{B}}[S]_{\mathcal{B}} = \begin{bmatrix} r_1 & \dots & r_m \end{bmatrix} \begin{bmatrix} B(V_1,V_1) & \dots & B(V_1,V_m) \\ \vdots & \ddots & \vdots \\ B(V_m,V_1) & \dots & B(V_m,V_m) \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix}. \tag{2.49}$$

Definition 2.8.2 (Types of bilinear forms)**.** A bilinear form B is called

- Symmetric if and only if $B(V,W) = B(W,V)$
- Non-degenerate if and only if $B(V,W) = 0$ for every $W \in T$ implies that necessarily $V = 0$
- Positive-definite if and only if $B(V,V) > 0$ for every $V \neq 0$.

Remark. Note that if B is positive definite, then it is also non-degenerate.

Definition 2.8.3 (Inner product)**.** An *inner product* G on a vector space is a symmetric and positive-definite bilinear form.

Definition 2.8.4 (Riemannian metric)**.** A *Riemannian metric* G on a manifold M is a smooth $(0,2)$-tensor field where for each $p \in M$, G_p is an inner product on T_pM.

Remark. Given a point $p \in M$ and a coordinate chart (U,φ) with $p \in U$, in the coordinate basis $\mathcal{B}_\varphi|_p = \left(\frac{\partial}{\partial\varphi_1}|_p, \ldots, \frac{\partial}{\partial\varphi_m}|_p\right)$ we can express the matrix representation of G_p in this basis as

$$[G_p]_{\mathcal{B}_\varphi|_p} = \begin{bmatrix} g_{\varphi_1,\varphi_1}(p) & \cdots & g_{\varphi_1,\varphi_m}(p) \\ \vdots & \ddots & \vdots \\ g_{\varphi_1,\varphi_m}(p) & \cdots & g_{\varphi_m,\varphi_m}(p) \end{bmatrix}, \tag{2.50}$$

which is symmetric because G_p is symmetric, and it is invertible because G_p is non-degenerate. The elements of the inverse of this matrix

$$([G_p]_{\mathcal{B}_\varphi|_p})^{-1} = \begin{bmatrix} g^{\varphi_1,\varphi_1}(p) & \cdots & g^{\varphi_1,\varphi_m}(p) \\ \vdots & \ddots & \vdots \\ g^{\varphi_1,\varphi_m}(p) & \cdots & g^{\varphi_m,\varphi_m}(p) \end{bmatrix}, \tag{2.51}$$

will be denoted by upper-case indices, where this matrix also is symmetric because G_p is symmetric. Now we obtain formulas relating these functions g_{φ_i,φ_j} and g^{φ_k,φ_k} using that $([G_p]_{\mathcal{B}_\varphi|_p})([G_p]_{\mathcal{B}_\varphi|_p})^{-1} = I$. We get

$$\sum_{j=1}^{m} g_{\varphi_i,\varphi_j}(p) \cdot g^{\varphi_j,\varphi_k}(p) = \begin{cases} 1, & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}. \tag{2.52}$$

We may drop writing the p and consider $[G]_{\mathcal{B}_\varphi} : U \to \mathbb{R}^{m^2},\ p \mapsto [G_p]_{\mathcal{B}_\varphi|_p}$, and similarly with $([G]_{\mathcal{B}_\varphi})^{-1} : U \to \mathbb{R}^{m^2},\ p \mapsto ([G_p]_{\mathcal{B}_\varphi|_p})^{-1}$.

Definition 2.8.5 (Riemannian manifold)**.** A Riemannian manifold $(M, [\mathcal{A}], G)$ is a smooth manifold $(M, [\mathcal{A}])$ together with a Riemannian metric G on it.

Remark. We denote Riemannian manifolds by (M,G) or simply M, if the smooth structure and Riemannian metric are implied by context.

Definition 2.8.6 (Riemannian norm)**.** Let (M,G) be a Riemannian manifold. The *norm* of a tangent vector $V_p \in T_pM$ is defined to be $\|V_p\|_G = \sqrt{G_p(V_p,V_p)}$.

Definition 2.8.7 (Riemannian distance)**.** Let (M,G) be a Riemannian manifold. The *distance* between two points $p,q \in M$ is given by

$$d_G(p,q) = \inf_c \int_0^1 \|\dot{c}(t)\|_G dt$$

Example 26 (Euclidean space as a Riemannian manifold)**.** Euclidean space $\mathbb{E}^m$ has a canonical Riemannian metric $< -, - >$ which is defined by

$$< -, - >:= \sum_{i=1}^{m} dx_i \otimes dx_i.$$

This means that in the standard coordinates, the matrix representation for $< -, - >_p$ is given by the identity matrix I for every $p \in M$. This corresponds to the standard scalar product $\bullet$ on $\mathbb{R}^m$.

It turns out that every smooth manifold can be equipped with a Riemannian structure. The idea for proving this is to define the metric locally by coordinate $(0,2)$-tensors and put it together to a globally defined metric by using a *partition of unity*.

Definition 2.8.8 (Partition of unity)**.** Let M be a smooth manifold and $\mathcal{C} = \{U_i : i \in I\}$ an open cover of M. A *partition of unity* subordinate to $\mathcal{C}$ is a family of smooth functions $\{\chi_i : M \to \mathbb{R} \ : \ i \in I\}$ such that

- $\mathrm{im}(\chi_i) \subseteq [0,1]$
- $\mathrm{supp}(\chi_i) = \overline{\{p \in M : \chi_i(p) \neq 0\}} \subset U_i$
- For every $p \in M$ there exists an open set V with $p \in V$ such that $V \cap \mathrm{supp}(\chi_i) \neq \varnothing$ for only finitely many i.
- $\sum_{i \in I} \chi_i(p) = 1$ for every $p \in M$.

Lemma 2.8.1 (Partition of unity always exists)**.** *Let M be a smooth manifold and let $\mathcal{C}$ be an open cover of M. Then there exists a partition of unity subordinate to $\mathcal{C}$.*

Proof. See Theorem 2.23 of [Lee13]. □

Lemma 2.8.2 (Smooth manifolds can be equipped with a Riemannian structure)**.** *Let $(M, [\mathcal{A}])$ be a smooth manifold. Then there exists a Riemannian metric G on M.*

Proof. Since $\mathcal{A} = \{(U_i, \varphi_i) : i \in I\}$ is an atlas, it constitutes an open cover $\cup_i U_i = M$, and therefore by Lemma 2.8.1 there exists a partition of unity $\{\chi_i : i \in I\}$ subordinate to $\{U_i : i \in I\}$. If we denote $\varphi_i = ((\varphi_i)_1, \dots (\varphi_i)_m)$, then for each $i \in I$, we can consider the smooth tensor field on U_i defined by

$$G_i = \sum_{j=1}^{m} d(\varphi_i)_j \otimes d(\varphi_i)_j.$$

Because χ_i is smooth and $\chi_i(p) = 0$ for $p \notin U_i$ then

$$\chi_i(p)(G_i)_p = \begin{cases} \chi_i(p) \sum_{j=1}^{m} d(\varphi_i)_j|_p \otimes d(\varphi_i)_j|_p, & p \in U_i \\ 0, & \text{otherwise} \end{cases}$$

(where here 0 means the $(0,2)$-tensor which always evaluates to $0 \in \mathbb{R}$) means that $\chi_i G_i$ is a smooth tensor field on all of M. Let now

$$G = \sum_{i \in I} \chi_i G_i.$$

The claim is that G is a Riemannian metric on M. Since each G_i is symmetric this means that G is symmetric, and since each G_i is positive-definite and $\chi_i(p) \geq 0$ this means that $G_p(v,v) \geq 0$. Moreover, by positive-definiteness each $(G_i)_p(v,v) > 0$ for $v \neq 0$, and so $G_p(v,v) = 0$ for $v \neq 0$ only if $\chi_i(p) = 0$ for every $i \in I$. However $\sum_{i \in I} \chi_i(p) = 1$ prevents this, and so G_p is positive definite for every $p \in M$. □

Lemma 2.8.3 (Product metric)**.** *Let $(M_1, G_1), \dots (M_n, G_n)$ be Riemannian manifolds. Then (M, G) is a Riemannian manifold, where $M = M_1 \times \cdots \times M_n$ and we let for each $p = (p_1, \dots, p_n) \in M$*

$$\begin{aligned}(G)_p : (T_{p_1}M_1 \times T_{p_1}M_1) \times \cdots \times (T_{p_n}M_n \times T_{p_n}M_n) &\to \mathbb{R}, \\ (V_1, W_1, \dots, V_n, W_n) &\mapsto (G_1)_{p_1}(V_1, W_1) + \cdots + (G_n)_{p_n}(V_n, W_n).\end{aligned}$$

This G is called the product metric.

Proof. Since each G_i is positive definite, the sum G is zero if and only if each G_i is zero, meaning that G is positive definite. In the coordinate basis $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_n)$ given by the coordinate chart $(U_1 \times \cdots \times U_n, (\varphi_1, \dots, \varphi_n))$ we get the block-diagonal matrix

$$[(G)_p]_{\mathcal{B}} = \begin{bmatrix} [(G_1)_{p_1}]_{\mathcal{B}_1} & & \\ & \ddots & \\ & & [(G_n)_{p_n}]_{\mathcal{B}_n} \end{bmatrix}, \tag{2.53}$$

which then only consists of smooth functions, since each G_i is Riemannian. Therefore G is also Riemannian. □

2.8.2 Gradient

Definition 2.8.9 (Gradient)**.** The *gradient* $\operatorname{grad}_G f$ of a smooth function $f : M \to \mathbb{R}$ defined on a Riemannian manifold (M, G) is a smooth vector field

$$\operatorname{grad}_G f : M \to TM$$

defined by the equation

$$G(\operatorname{grad}_G f, V) = V(f). \tag{2.54}$$

Remark. The gradient is well-defined because G is non-degenerate.

Lemma 2.8.4 (Local expression for gradient)**.** *Given a coordinate chart (U, φ), the local expression for the vector field $\operatorname{grad}_G f$ is given by*

$$\operatorname{grad}_G f|_U = \sum_{i,j=1}^{m} \left(g^{\varphi_i, \varphi_j} \frac{\partial f}{\partial \varphi_j} \right) \frac{\partial}{\partial \varphi_i}.$$

Proof. Given the coordinate chart (U, φ), let

$$\operatorname{grad}_G f|_U = \sum_{i=1}^{m} y_{\varphi_i} \frac{\partial}{\partial \varphi_i}$$

where the functions $y_{\varphi_i} : U \to \mathbb{R}$ are to be determined. Now let $V|_U = \frac{\partial}{\partial \varphi_j}$, so that

$$\begin{aligned} G(\operatorname{grad}_G f, V)|_U &= G|_U(\operatorname{grad}_G f|_U, V|_U) \\ &= \begin{bmatrix} y_{\varphi_1} & \dots & y_{\varphi_m} \end{bmatrix} \begin{bmatrix} g_{\varphi_1,\varphi_1} & \cdots & g_{\varphi_1,\varphi_m} \\ \vdots & \ddots & \vdots \\ g_{\varphi_1,\varphi_m} & \cdots & g_{\varphi_m,\varphi_m} \end{bmatrix} e_j \\ &= \begin{bmatrix} y_{\varphi_1} & \dots & y_{\varphi_m} \end{bmatrix} \begin{bmatrix} g_{\varphi_1,\varphi_j} \\ \vdots \\ g_{\varphi_m,\varphi_j} \end{bmatrix} \\ &= y_{\varphi_1} g_{\varphi_1,\varphi_j} + \cdots + y_{\varphi_m} g_{\varphi_m,\varphi_j} \end{aligned} \tag{2.55}$$

and

$$V(f)|_U = \frac{\partial f}{\partial \varphi_j}. \tag{2.56}$$

Therefore, for each $j = 1, \dots, m$ we obtain

$$\begin{cases} y_{\varphi_1} g_{\varphi_1,\varphi_1} + \cdots + y_{\varphi_m} g_{\varphi_1,\varphi_m} & = \frac{\partial f}{\partial \varphi_1} \\ & \vdots \\ y_{\varphi_1} g_{\varphi_m,\varphi_1} + \cdots + y_{\varphi_m} g_{\varphi_m,\varphi_m} & = \frac{\partial f}{\partial \varphi_m} \end{cases} \tag{2.57}$$

which is a linear system of equations

$$[G]_{\mathcal{B}_\varphi} [\operatorname{grad}_G f]_{\mathcal{B}_\varphi} = \begin{bmatrix} \frac{\partial f}{\partial \varphi_1} \\ \vdots \\ \frac{\partial f}{\partial \varphi_m} \end{bmatrix}. \tag{2.58}$$

This has a unique solution because G is non-degenerate, given by

$$\begin{aligned} [\operatorname{grad}_G f]_{\mathcal{B}_\varphi} &= ([G]_{\mathcal{B}_\varphi})^{-1} \begin{bmatrix} \frac{\partial f}{\partial \varphi_1} \\ \vdots \\ \frac{\partial f}{\partial \varphi_m} \end{bmatrix} \\ &= \begin{bmatrix} g^{\varphi_1,\varphi_1} & \dots & g^{\varphi_1,\varphi_m} \\ \vdots & \ddots & \vdots \\ g^{\varphi_1,\varphi_m} & \dots & g^{\varphi_m,\varphi_m} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \varphi_1} \\ \vdots \\ \frac{\partial f}{\partial \varphi_m} \end{bmatrix} \\ &= \begin{bmatrix} g^{\varphi_1,\varphi_1} \frac{\partial f}{\partial \varphi_1} + \cdots + g^{\varphi_1,\varphi_m} \frac{\partial f}{\partial \varphi_m} \\ \vdots \\ g^{\varphi_m,\varphi_1} \frac{\partial f}{\partial \varphi_1} + \cdots + g^{\varphi_m,\varphi_m} \frac{\partial f}{\partial \varphi_m} \end{bmatrix} \end{aligned} \tag{2.59}$$

which is the result we wanted to prove. □

Lemma 2.8.5 (Gradient is smooth)**.** *For any Riemannian manifold (M, G) and any smooth function $f : M \to \mathbb{R}$, the gradient vector field $\operatorname{grad}_G f : M \to TM$ is a smooth vector field.*

Proof. To prove this we need to show that the coordinates of $\operatorname{grad}_G f$ in a given coordinate chart consist of smooth functions. Given a coordinate chart (U, φ), this means that we need to show that

$$\sum_{j=1}^{m} g^{\varphi_i,\varphi_j} \frac{\partial f}{\partial \varphi_j} : U \to \mathbb{R} \tag{2.60}$$

is smooth. Since the functions $\frac{\partial f}{\partial \varphi_j}$ are smooth, it suffices to show that the functions g^{φ_i,φ_j} are smooth. This follows from those functions being entries of the matrix

$$([G]_{\mathcal{B}_\varphi})^{-1} = \frac{1}{\det([G]_{\mathcal{B}_\varphi})} \cdot \operatorname{adj}([G]_{\mathcal{B}_\varphi}), \tag{2.61}$$

where both $\det([G]_{\mathcal{B}_\varphi})$ and every entry of $\operatorname{adj}([G]_{\mathcal{B}_\varphi})$ is a sum of products of entries from $[G]_{\mathcal{B}_\varphi}$ which are smooth functions g_{φ_k,φ_l}. □

Chapter 3

Morse Theory

3.1 Critical points and gradient flow

Definition 3.1.1 (Gradient flow)**.** For a Riemannian manifold (M, G), every smooth function $f : M \to \mathbb{R}$ encodes a (negative) *gradient flow*, which is the first order ODE corresponding to the smooth vector field $\operatorname{grad}_G f$, that is,

$$\dot{c} = -(\operatorname{grad}_G f) \circ c,$$

where $c : I \to M$ is a smooth function.

Remark. For the definition of $\dot{c}$, see the remark to Definition 2.6.2.

Definition 3.1.2 (Critical point)**.** For a smooth function $f : M \to \mathbb{R}$, a *critical point* $p \in M$ of f is a point which satisfies $D_p f = 0 : T_p M \to \mathbb{R},\ v \mapsto 0$.

Lemma 3.1.1 (Zeros of gradient)**.** *Let (M, G) be a Riemannian manifold and $f : M \to \mathbb{R}$ a smooth function. A point $p \in M$ is a critical point of f if and only if $(\operatorname{grad}_G f)_p = 0$.*

Remark. The Riemannian structure is irrelevant in the sense that if $(\operatorname{grad}_G f)_p = 0$ for one metric, then it holds for any metric.

Proof. ($\Longrightarrow$) Let $p \in M$ be a critical point of f. Then $D_p f = 0$, and so, by the definition of gradient $G_p((\operatorname{grad}_G f)_p, V_p) = D_p f(V_p) = 0$ for all $V_p \in T_p M$. By the definition of Riemannian metric, G_p is non-degenerate, and so this can only hold if $(\operatorname{grad}_G f)_p = 0$.

($\Longleftarrow$) If $(\operatorname{grad}_G f)_p = 0$, then $D_p f(V_p) = G_p((\operatorname{grad}_G f)_p, V_p) = G_p(0, V_p) = 0$, showing that p is a critical point of f. □

Lemma 3.1.2 (Gradient flow strictly decreases f)**.** *Let (M, G) be a Riemannian manifold, $f : M \to \mathbb{R}$ a smooth function, and $c : I \to M$ a solution to $\dot{c} = -(\operatorname{grad}_G f) \circ c$. Then $(f \circ c)'(t) = -G_{c(t)}((\operatorname{grad}_G f)_{c(t)}, (\operatorname{grad}_G f)_{c(t)}) \leq 0$, with equality if and only if $c(t)$ is a critical point of f.*

Proof. By the chain rule (Lemma 2.5.10) $(f \circ c)'(t) = (D_{c(t)} f)(D_t c)$ (where $D_t c$ is canonically considered as the vector $D_t c((D_t \operatorname{id}_I)^{-1} 1) \in T_{c(t)} M$ as in the remark to Definition 2.6.2). Since $\dot{c} = -(\operatorname{grad}_G f) \circ c$, equivalently $D_t c = -(\operatorname{grad}_G f)_{c(t)}$, and so

$$(f \circ c)'(t) = (D_{c(t)} f)(-(\operatorname{grad}_G f)_{c(t)}) = -(D_{c(t)} f)((\operatorname{grad}_G f)_{c(t)}) \tag{3.1}$$

by linearity. By the definition of the gradient $D_{c(t)} f(v) = G_{c(t)}((\operatorname{grad}_G f)_{c(t)}, v)$ for any $v \in T_{c(t)} M$. Applying this with $v = (\operatorname{grad}_G f)_{c(t)}$ gives

$$(f \circ c)'(t) = -G_{c(t)}((\operatorname{grad}_G f)_{c(t)}, (\operatorname{grad}_G f)_{c(t)}). \tag{3.2}$$

Since $G_{c(t)}$ is positive definite by the definition of Riemannian metric $G_{c(t)}(v,v) \geq 0$ for all $v \in T_{c(t)}M$ with equality if and only if $v = 0$. Therefore $(f \circ c)'(t) \leq 0$, with equality if and only if $(\operatorname{grad}_G f)_{c(t)} = 0$, which by Lemma 3.1.1 holds if and only if $c(t)$ is a critical point of f. □

Lemma 3.1.3 (Convergence of gradient flow)**.** *Let (M,G) be a compact Riemannian manifold, $f : M \to \mathbb{R}$ a smooth function, and $c : \mathbb{R} \to M$ a solution to $\dot{c} = -(\operatorname{grad}_G f) \circ c$. If for some $t_0 \in \mathbb{R}$ the set $\{q \in M \ : \ f(q) \leq f(c(t_0))\}$ contains exactly one critical point p^*, then $\lim_{t\to\infty} c(t) = p^*$.*

Proof. Since f is smooth, it is also continuous by Lemma 2.4.6, and so any sublevel set $\{q \in M : f(q) \leq S\} = f^{-1}(]-\infty, S])$ is a closed subset of M. Since M is compact the sublevel set is also compact. By Proposition 3.7 of [HM94] this means that c converges to a connected subset $C \subseteq P$ of the set of critical points P of f. By Lemma 3.1.2, $f \circ c$ is non-increasing, so $f(c(t)) \leq f(c(t_0))$ for every $t \geq t_0$, and hence $C \subseteq \{q \in M \ : \ f(q) \leq f(c(t_0))\}$. Since also $C \subseteq P$, this means that $C \subseteq \{q \in M \ : \ f(q) \leq f(c(t_0))\} \cap P = \{p^*\}$ by assumption. Hence c converges to p^*. □

3.2 The Hessian and Morse functions

Definition 3.2.1 (Hessian)**.** Given a smooth manifold M and a smooth function $f : M \to \mathbb{R}$, if $p \in M$ is a critical point for f, then we define the *Hessian* of f at $p \in M$ to be the bilinear form

$$\operatorname{Hess}(f)|_p : T_pM \times T_pM \to \mathbb{R}, \quad (V_p, W_p) \mapsto (V \circ W)(f)(p)$$

where V is any vector field equal to V_p at p and W is any vector field equal to W_p at p.

Lemma 3.2.1 (Hessian is well-defined and symmetric)**.** *Given the vectors V_p and W_p, the Hessian $\operatorname{Hess}(f)|_p(V_p, W_p)$ at a point $p \in M$ is well-defined, in the sense that it does not depend on the choice of vector fields. Moreover, it is bilinear and symmetric $\operatorname{Hess}(f)|_p(V_p, W_p) = \operatorname{Hess}(f)|_p(W_p, V_p)$.*

Proof. Let $p \in M$ be a given critical point of f, and let (U, φ) be an arbitrary coordinate chart with $p \in U$. Because p is a critical point of f, we have that

$$\frac{\partial f}{\partial \varphi_i}(p) = \left(\frac{\partial}{\partial \varphi_i}(f)\right)(p) = D_p f\left(\frac{\partial}{\partial \varphi_i}|_p\right) = 0. \tag{3.3}$$

Further if V and W are arbitrary vector fields, then by Lemma 2.6.8

$$\begin{aligned}
(V \circ W)(f)(p) &= \sum_{i,j=1}^{m} v_{\varphi_j}(p)\frac{\partial w_{\varphi_i}}{\partial \varphi_j}(p)\frac{\partial f}{\partial \varphi_i}(p) + \sum_{k,l=1}^{m} v_{\varphi_k}(p) w_{\varphi_l}(p)\frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l}(p) \\
&= \sum_{k,l=1}^{m} v_{\varphi_k}(p) w_{\varphi_l}(p)\frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l}(p) \\
&= \sum_{k,l=1}^{m} d\varphi_k|_p(V_p) d\varphi_l|_p(W_p)\frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l}(p) \\
&= \sum_{k,l=1}^{m} d\varphi_l|_p(W_p) d\varphi_k|_p(V_p)\frac{\partial^2 f}{\partial \varphi_k \partial \varphi_l}(p) \\
&= \sum_{k,l=1}^{m} d\varphi_l|_p(W_p) d\varphi_k|_p(V_p)\frac{\partial^2 f}{\partial \varphi_l \partial \varphi_k}(p) \\
&= (W \circ V)(f)(p).
\end{aligned} \tag{3.4}$$

We see that $\text{Hess}(f)|_p$ only depends on V_p and W_p and is symmetric. It is also bilinear because we can write it as the following $(0,2)$-tensor on T_pM

$$\text{Hess}(f)|_p = \sum_{k,l=1}^{m} \frac{\partial^2 f}{\partial\varphi_k \partial\varphi_l}(p)\ d\varphi_k|_p \otimes d\varphi_l|_p. \tag{3.5}$$

□

Corollary 3.2.2 (Pointwise expression for Hessian)**.** *Let $p \in M$ be a critical point of $f : M \to \mathbb{R}$. If (U, φ) is a coordinate chart with $p \in U$ then $\text{Hess}(f)|_p$ is the symmetric bilinear form given by*

$$\text{Hess}(f)|_p = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial\varphi_i \partial\varphi_j}(p)\ d\varphi_i|_p \otimes d\varphi_j|_p$$

on the tangent space T_pM.

Remark. We then see that the matrix representation of $\text{Hess}(f)|_p$ in the coordinate chart (U, φ) is given by

$$[\text{Hess}(f)|_p]_{\mathcal{B}_{\varphi}|_p} = \begin{bmatrix} \frac{\partial^2 f}{\partial\varphi_1\partial\varphi_1}(p) & \cdots & \frac{\partial^2 f}{\partial\varphi_1\partial\varphi_m}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial\varphi_m\partial\varphi_1}(p) & \cdots & \frac{\partial^2 f}{\partial\varphi_m\partial\varphi_m}(p) \end{bmatrix}.$$

Remark. Observe that $\text{Hess}(f)$ under this definition is not a $(0,2)$-tensor field unless $Df \equiv 0$. Indeed, assume that there exists a point $p \in M$ where $D_pf \neq 0$. Then for one of the basis vectors $\frac{\partial}{\partial\varphi_1}|_p, \ldots, \frac{\partial}{\partial\varphi_m}|_p$, say $\frac{\partial}{\partial\varphi_\alpha}|_p$, we would have $\frac{\partial f}{\partial\varphi_\alpha}(p) = D_pf\left(\frac{\partial}{\partial\varphi_\alpha}|_p\right) \neq 0$, and so the terms $v_{\varphi_j}(p)\frac{\partial w_{\varphi_\alpha}}{\partial\varphi_j}(p)\frac{\partial f}{\partial\varphi_\alpha}(p)$ do not only depend on the values of $v_{\varphi_j}(p) \in \mathbb{R}$, $w_{\varphi_\alpha}(p) \in \mathbb{R}$, but it depends on all values of $w_{\varphi_\alpha} : U \to \mathbb{R}$ in some open subset of U containing p in order to calculate the directional derivative $\frac{\partial w_{\varphi_\alpha}}{\partial\varphi_j}$. Therefore it is not a tensor field.

Remark. $\text{Hess}(f)$ can be made into a $(0,2)$-tensor field if we define it to be

$$\text{Hess}^{\nabla}(f)(V, W) = (V \circ W)(f) - Df(\nabla_V W)$$

where ∇ is a "connection". (See Example 4.22 of [Lee18].) This would eliminate the problematic terms $v_{\varphi_j}(p)\frac{\partial w_{\varphi_\alpha}}{\partial\varphi_j}(p)\frac{\partial f}{\partial\varphi_\alpha}(p)$ to make it a tensor field.

If M is a Riemannian manifold, then there exists a canonical connection called the "Levi-Civita connection", also known as the "covariant derivative". If ∇ is the Levi-Civita connection on the Riemannian manifold (M, G) then

$$\text{Hess}^{\nabla}(f)(V, W) = G(\nabla_V(\text{grad}_G f), W)$$

and locally, given a coordinate chart (U, φ) then

$$\text{Hess}^{\nabla}(f)|_U = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial\varphi_i\partial\varphi_j}\ d\varphi_i \otimes d\varphi_j - \sum_{i,j,k=1}^{m} \Gamma^{\varphi_k}_{\varphi_i,\varphi_j}\ d\varphi_i \otimes d\varphi_j,$$

where $\Gamma^{\varphi_k}_{\varphi_i,\varphi_j} : U \to \mathbb{R}$ are the coordinates of the vector field

$$\nabla_{\frac{\partial}{\partial\varphi_i}}\left(\frac{\partial}{\partial\varphi_i}\right) = \sum_{k=1}^{m} \Gamma^{\varphi_k}_{\varphi_i,\varphi_j}\ \frac{\partial}{\partial\varphi_k} : U \to TU.$$

They are called the "Christoffel symbols" for the coordinate chart (U, φ).

Lemma 3.2.3 (Hessian of composition)**.** *Let $g : M \to \mathbb{R}$ and $\psi : N \to M$ be smooth, and let $\psi(Q)$ be a critical point of g. Then*

$$[\mathrm{Hess}(g \circ \psi)|_Q] = [D_Q\psi]^\mathsf{T}[\mathrm{Hess}(g)|_{\psi(Q)}][D_Q\psi]$$

expressed in appropriate coordinate bases.

Proof. By the chain rule $D_Q(g \circ \psi) = (D_{\psi(Q)}g) \circ (D_Q\psi) = 0$ since $\psi(Q)$ is a critical point of g, and therefore Q is a critical point of $g \circ \psi$. Let now (U, φ) and (V, ξ) be coordinate charts around $Q \in U$ and $\psi(Q) \in V$ respectively, and let $\mathcal{B}$ be the coordinate basis of (U, φ) for T_QN and $\mathcal{C}$ be the coordinate basis of (V, ξ) for $T_{\psi(Q)}M$.

By the definition of the Hessian we obtain that

$$\mathrm{Hess}(g \circ \psi)|_Q(V_Q, W_Q) = \sum_{i,j} d\varphi_i|_Q(V_Q)\, d\varphi_j|_Q(W_Q)\, \frac{\partial^2(g \circ \psi)}{\partial\varphi_i\partial\varphi_j}(Q). \tag{3.6}$$

Applying the chain rule twice, and using that $\frac{\partial g}{\partial \xi_k}(\psi(Q)) = 0$ for every k since $\psi(Q)$ is a critical point gives

$$\begin{aligned}\frac{\partial^2(g \circ \psi)}{\partial\varphi_i\partial\varphi_j}(Q) &= \sum_{k,l} \frac{\partial(\xi_k \circ \psi)}{\partial\varphi_i}(Q)\, \frac{\partial(\xi_l \circ \psi)}{\partial\varphi_j}(Q)\, \frac{\partial^2 g}{\partial\xi_k\partial\xi_l}(\psi(Q)) \\ &= \sum_{k,l} \left([D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}\right)_{ki} \left([D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}\right)_{lj} \frac{\partial^2 g}{\partial\xi_k\partial\xi_l}(\psi(Q)).\end{aligned} \tag{3.7}$$

Therefore

$$\begin{aligned}&\mathrm{Hess}(g\circ\psi)|_Q(V_Q, W_Q) \\ &= \sum_{k,l} \left(\sum_i \left([D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}\right)_{ki} d\varphi_i|_Q(V_Q)\right) \left(\sum_j \left([D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}\right)_{lj} d\varphi_j|_Q(V_Q)\right) \frac{\partial^2 g}{\partial\xi_k\partial\xi_l}(\psi(Q)) \\ &= \sum_{k,l} \left([D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}[V_Q]_{\mathcal{B}}\right)_k \left([\mathrm{Hess}(g)|_{\psi(Q)}]_{\mathcal{C}}\right)_{kl} \left([D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}[W_Q]_{\mathcal{B}}\right)_l \\ &= [V_Q]^\mathsf{T}_{\mathcal{B}} \left([D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}\right)^\mathsf{T} [\mathrm{Hess}(g)|_{\psi(Q)}]_{\mathcal{C}}[D_Q\psi]^{\mathcal{C}}_{\mathcal{B}}[W_Q]_{\mathcal{B}}.\end{aligned} \tag{3.8}$$

Since $\mathrm{Hess}(g \circ \psi)|_Q(V_Q, W_Q) = [V_Q]^\mathsf{T}_{\mathcal{B}}[\mathrm{Hess}(g \circ \psi)|_Q]_{\mathcal{B}}[W_Q]_{\mathcal{B}}$, we are done. □

Definition 3.2.2 (Non-degenerate critical point)**.** Given a smooth function $f : M \to \mathbb{R}$, a critical point $p \in M$ is called *non-degenerate* if and only if the Hessian of f at p is a non-degenerate bilinear form.

Definition 3.2.3 (Morse function)**.** A smooth function $f : M \to \mathbb{R}$ is called a *Morse function* if all of its critical points are non-degenerate.

Lemma 3.2.4 (Sard-Morse)**.** *Every smooth manifold has at least one Morse function.*

Proof. See Theorem 1.2 of Chapter 6 of [Hir76]. □

Remark. "Almost all" smooth functions $f : M \to \mathbb{R}$ are Morse, in the sense that the Morse functions form an open and "dense" subset of $C^\infty(M)$ equipped with the "Whitney C^2" topology.

3.3 The Morse lemma and its consequences

Firstly we will establish the *Morse lemma*, for which we first need to define the index of a bilinear form.

Definition 3.3.1 (Index)**.** The *index* k of a bilinear form $B : V \times V \to \mathbb{R}$ is the largest possible dimension $0 \leq k \leq m$ of a subspace $W \subseteq V$ such that the restriction $B|_W : W \times W \to \mathbb{R}$ is *negative-definite*.

Remark. Here $B|_W$ being negative-definite means that $-B|_W$ is positive-definite, or in other words, if $w \neq 0$ then $B(w, w) < 0$ for every $w \in W$.

Remark. The index of B is also equal to the number of negative eigenvalues of B, counted with multiplicity.

Lemma 3.3.1 (Morse lemma)**.** *If $p \in M$ is a non-degenerate critical point of $f : M \to \mathbb{R}$, then there exists a coordinate chart $(U_{f,p}, \varphi_{f,p})$, with $p \in U_{f,p}$ and $\varphi_{f,p} = (x_1, \dots x_c, y_1, \dots, y_k)$ such that $\varphi_{f,p}(p) = 0$ and*

$$f(q) = f(p) + (x_1(q))^2 + \cdots + (x_c(q))^2 - (y_1(q))^2 - \cdots - (y_k(q))^2$$

for every $q \in U_{f,p}$, where k is the index of $\mathrm{Hess}(f)|_p$.

Proof. See Lemma 2.2 of [Mil63]. □

Definition 3.3.2 (Morse coordinate chart)**.** A coordinate chart fulfilling the Morse lemma (for f) is called a Morse coordinate chart (for f).

Definition 3.3.3 (Index of non-degenerate critical point)**.** If the smooth function $f : M \to \mathbb{R}$ is clear by context, the *index* of p is defined to be the index of $\mathrm{Hess}(f)|_p$.

Definition 3.3.4 (Local minimum)**.** Let $f : M \to \mathbb{R}$ be a smooth function. A point $p \in M$ is called a *local minimum* of f if and only if there exists an open set $U \subseteq M$ with $p \in U$ such that $f(p) \leq f(q)$ for every $q \in U$. If moreover $f(p) < f(q)$ for every $q \in U$ with $q \neq p$, then p is called a *strict local minimum*.

Remark. A local minimum is necessarily a critical point, since if $D_p f \neq 0$ then there exists a curve c with $c(0) = p$ and $(f \circ c)'(0) < 0$, meaning f decreases along c and so p cannot be a local minimum.

Lemma 3.3.2 (Local minimum of Morse function)**.** *Let $f : M \to \mathbb{R}$ be a Morse function. Then p is a local minimum of f if and only if p has index $k = 0$. Moreover, if p is a local minimum of f, it is a strict local minimum.*

Proof. ($\implies$) Let p be a local minimum of f in the open set $W \subseteq M$. As noted, this means that it is a critical point of f. By the Morse lemma there exists a coordinate chart (U, φ) with $p \in U$, $\varphi = (x_1, \dots, x_c, y_1, \dots, y_k)$ such that $\varphi(p) = 0$ and $f(q) = f(p) + (x_1(q))^2 + \cdots + (x_c(q))^2 - (y_1(q))^2 - \cdots - (y_k(q))^2$.

By way of contradiction, let $k \geq 1$. Since $p \in W$ then $W \cap U \neq \varnothing$ and $W \cap U \subseteq U$, so the equation given by the Morse lemma holds in $W \cap U$, which is an open set because U and W are open. Since $\varphi : U \to V$, where V is open in Euclidean space, $\varphi(p) = 0$ means that there exists a small ball $B_r(0) \subseteq V$, $r > 0$, and so $(0, \dots, 0, \varepsilon, 0, \dots, 0) \in V$, for every $0 \leq \varepsilon < r$, and hence $q_\varepsilon = \varphi^{-1}(0, \dots, 0, \varepsilon, 0, \dots, 0) \in U$ for every $0 \leq \varepsilon < r$. Since $W \cap U$ is an open set containing p, then $q_\varepsilon \in U \cap W$ for every $0 \leq \varepsilon < r'$, where $0 < r' \leq r$ is some number. Hence there exists a $q_\varepsilon \in U \cap W$ for some $\varepsilon > 0$, where by the Morse lemma

$f(q_\varepsilon) = f(p) + 0^2 + \cdots + 0^2 - \varepsilon^2 - 0^2 - \cdots - 0^2 = f(p) - \varepsilon^2 < f(p)$, contradicting p being a local minimum of f in $W \supseteq U \cap W$.

($\Longleftarrow$) Let p have index 0. By the Morse lemma $f(p) = f(q) - (x_1(q))^2 - \cdots - (x_m(q))^2 = f(q) - \|\varphi(q)\|^2$, where $m = \dim(M)$. Since $\varphi(p) = 0$ and φ is bijective, then $\varphi(q) = 0 \iff q = p$, and hence $\|\varphi(q)\|^2 > 0 \iff q \neq p$, showing that $f(p) < f(q)$ for $p \neq q$. □

Definition 3.3.5 (Critical values)**.** Let $f : M \to \mathbb{R}$ be a smooth function. If $p \in M$ is a critical point of f, then $f(p) \in \mathbb{R}$ is called a *critical value* of f.

Lemma 3.3.3 (Morse function with unique minimum)**.** *If M is a connected smooth manifold and $p^* \in M$, then there exists a Morse function $f : M \to \mathbb{R}$ having p^* as its unique local minimum with all critical values distinct.*

Proof. See Chapter 5 of [Wal16]. □

Corollary 3.3.4 (Translating and rescaling)**.** *If $f : M \to \mathbb{R}$ is a Morse function having $p^* \in M$ as a unique local minimum and all critical values distinct, by replacing f with $K(f - f(p^*))$ for a sufficiently large $K > 0$, we may additionally assume that $f(p^*) = 0$, $f(p_i) \geq 1$ for each critical point that is not a local minimum p_i, and $|f(p_i) - f(p_j)| \geq 1$ for $i \neq j$.*

Lemma 3.3.5 (Morse function for a product)**.** *Let $f_1 : M_1 \to \mathbb{R}, \ldots f_n : M_n \to \mathbb{R}$ be smooth functions. Then*

$$f : M_1 \times \cdots \times M_n \to \mathbb{R}, \quad (q_1, \ldots, q_n) \mapsto f_1(q_1) + \cdots + f_n(q_n)$$

is a Morse function on $M_1 \times \cdots \times M_n$ if and only if each $f_i : M_i \to \mathbb{R}$ is a Morse function on M_i for $1 \leq i \leq n$.

Moreover $(p_1, \ldots, p_n)$ is a critical point of f if and only if $p_1, \ldots, p_n$ are critical points of $f_1, \ldots, f_n$ respectively, and the index of $(p_1, \ldots, p_n)$ is the sum of the indices of $p_1, \ldots, p_n$. In particular, $(p_1, \ldots, p_n)$ is the unique local minimum of f if and only if $p_1, \ldots, p_n$ are the unique local minima of $f_1, \ldots, f_n$.

Proof. Let $p = (p_1, \ldots, p_n) \in M_1 \times \cdots \times M_n$, and let (U_i, φ_i) be a coordinate chart on M_i around p_i for each $1 \leq i \leq n$. Then $(U_1 \times \cdots \times U_n, (\varphi_1, \ldots, \varphi_n))$ is a coordinate chart on $M_1 \times \cdots \times M_n$ around p and

$$f \circ (\varphi_1, \ldots, \varphi_n)^{-1}(x) = f_1 \circ \varphi_1^{-1}(x_1) + \cdots + f_n \circ \varphi_n^{-1}(x_n), \tag{3.9}$$

where $x = (x_1, \ldots, x_n) \in \varphi_1(U_1) \times \cdots \times \varphi_n(U_n)$, by the definition of f.

Since p is a critical point if and only if $D_p f = 0$, it is a critical point if and only if every partial derivative of $f \circ (\varphi_1, \ldots, \varphi_n)^{-1}$ is 0 at $x = (\varphi_1(p_1), \ldots, \varphi_n(p_n))$. Since the terms in (3.9) are independent, this happens if and only if every partial derivative of every term is equal to 0, which is equivalent to p_i being a critical point of f_i for every $1 \leq i \leq n$.

Let $\mathcal{B}$ be the coordinate basis $(U_1 \times \cdots \times U_n, (\varphi_1, \ldots, \varphi_n))$ and $\mathcal{B}_i$ be the coordinate basis of (U_i, φ_i). At a critical point p, the matrix representation of $\mathrm{Hess}(f)|_p$ in the basis $\mathcal{B}$ is the block-diagonal matrix

$$[\mathrm{Hess}(f)|_p]_{\mathcal{B}} = \begin{bmatrix} [\mathrm{Hess}(f_1)|_{p_1}]_{\mathcal{B}_1} & & \\ & \ddots & \\ & & [\mathrm{Hess}(f_n)|_{p_n}]_{\mathcal{B}_n} \end{bmatrix} \tag{3.10}$$

because the mixed partial derivatives $\frac{\partial^2 f}{\partial (\varphi_i)_k \partial (\varphi_j)_l} = 0$ for $i \neq j$, since no term of (3.9) depends on both φ_i and φ_j.

The characteristic polynomial of a block-diagonal matrix is the product of the characteristic polynomial of each block, and so $\mathrm{Hess}(f)|_p$ is non-degenerate (determinant 0) if and only if each $\mathrm{Hess}(f_i)|_{p_i}$ is non-degenerate, meaning that f is a Morse function if and only if each f_i is a Morse function. Moreover, this also means that the eigenvalues of $\mathrm{Hess}(f)|_p$ is the union (counted with multiplicity) of the eigenvalues of each $\mathrm{Hess}(f_i)|_{p_i}$.

Since the index is the number of negative eigenvalues, this means that the index of p will be the sum of the indices of the p_i. By Lemma 3.3.2, p is a local minimum if and only if its index is 0, which then happens if and only if the index of each p_i is 0, that is, if and only if p_i is a local minimum of f_i for each $1 \leq i \leq n$. This is therefore the unique minimum if and only if each p_i is the unique local minimum of f_i for every $1 \leq i \leq n$. □

Lemma 3.3.6 (Morse coordinate chart on product)**.** *For $1 \leq i \leq n$, let $f_i : M_i \to \mathbb{R}$ be a Morse function, and (U_i, φ_i) a Morse coordinate chart for f_i centered at the critical point $p_i \in U_i$. Then $(U, \varphi) = (U_1 \times \cdots \times U_n, \varphi)$ is a Morse coordinate chart for*

$$f : M_1 \times \cdots \times M_n \to \mathbb{R}, \quad (q_1, \ldots, q_n) \mapsto f_1(q_1) + \cdots + f_n(q_n)$$

centered at the critical point $p = (p_1, \ldots, p_n) \in U$, where

$$\widetilde{\varphi} : U_1 \times \cdots \times U_n \to \varphi_1(U_1) \times \cdots \times \varphi_n(U_n), \quad (q_1, \ldots, q_n) \mapsto (\varphi_1(q_1), \ldots, \varphi_n(q_n))$$

and $\varphi = \rho \circ \widetilde{\varphi}$, where ρ is a permutation of the coordinates.

Proof. By the definition of the product topology, $U_1 \times \cdots \times U_n$ is open, and since $\varphi_i(U_i)$ is an open set in $\mathbb{E}^{\dim(M_i)}$, then $\varphi_1(U_1) \times \cdots \times \varphi_n(U_n)$ is an open set in $\mathbb{E}^{\dim(M_1)+\cdots+\dim(M_n)} = \mathbb{E}^{\dim(M_1 \times \cdots \times M_n)}$. Since each φ_i is a diffeomorphism, so is $\widetilde{\varphi}$ with inverse $\widetilde{\varphi}^{-1}(a_1, \ldots, a_n) = (\varphi_1^{-1}(a_1), \ldots, \varphi_n^{-1}(a_n))$. Therefore $(U, \widetilde{\varphi})$ is a chart, and given that $\varphi_i(p_i) = 0$, then $\widetilde{\varphi}(p) = 0$, and if we let $\varphi_i = (x_{i,1}, \ldots, x_{i,c_i}, y_{i,1}, \ldots, y_{i,k_i})$ then

$$\varphi = (x_{1,1}, \ldots, x_{1,c_1}, y_{1,1}, \ldots, y_{1,k_1}, \ldots, x_{n,1}, \ldots, x_{n,c_n}, y_{n,1}, \ldots, y_{n,k_n}) \tag{3.11}$$

so we choose the permutation ρ so that

$$\varphi = \rho \circ \widetilde{\varphi} = (x_{1,1}, \ldots, x_{n,c_n}, y_{1,1}, \ldots, y_{n,k_n}) = (\alpha_1, \ldots, \alpha_c, \beta_1, \ldots, \beta_k) \tag{3.12}$$

where $c = c_1 + \cdots + c_n$ and $k = k_1 + \cdots + k_n$. Since ρ is just a permutation, (U, φ) is a chart, and since (U_i, φ_i) is a Morse chart for f_i

$$f \circ \varphi^{-1} = \sum_{i=1}^{n} f_i(p_i) + \sum_{i=1}^{n} \sum_{j=1}^{c_i} x_{i,j}^2 - \sum_{i=1}^{n} \sum_{j=1}^{k_i} y_{i,j} = f(p) + \alpha_1^2 + \cdots + \alpha_c^2 - \beta_1^2 - \cdots - \beta_k^2 \tag{3.13}$$

which fulfills the Morse lemma, since the index of p is equal to the sum of the indices $\mathrm{index}(p) = \sum_{i=1}^{n} \mathrm{index}(p_i) = k_1 + \cdots + k_n = k$. □

3.4 Counting critical points

3.4.1 Isolation and finiteness

Lemma 3.4.1 (Non-degenerate critical points are isolated)**.** *If $p \in M$ is a non-degenerate critical point of $f : M \to \mathbb{R}$ then there exists an open set $U \subseteq M$ such that $p \in U$ and no other critical point of f is in U.*

Proof. By the Morse lemma there exists a coordinate chart (U, φ) with $p \in U$ and $\varphi = (x_1, \dots, x_c, y_1, \dots, y_k)$ such that $\varphi(p) = 0$ and $f|_U = f(p) + x_1^2 + \cdots + x_c^2 - y_1^2 - \cdots - y_k^2$. A point $q \in U$ is a critical point of f if and only if $D_q f = 0$, and by Lemma 2.5.14 we can calculate $D_q f$ in U by

$$\begin{aligned} D_q f\left(\frac{\partial}{\partial x_i}|_q\right) = \frac{\partial(f \circ \varphi^{-1})}{\partial_i}(\varphi(q)) = \frac{\partial}{\partial_i}(f(p) + \mathrm{id}_1^2 + \cdots + \mathrm{id}_c^2 - \mathrm{id}_{c+1}^2 - \cdots - \mathrm{id}_m^2)|_{\varphi(q)} \\ = 2\,\mathrm{id}_i\,|_{\varphi(q)} = 2x_i(q), \end{aligned} \tag{3.14}$$

where $\mathrm{id} = (\mathrm{id}_1, \dots, \mathrm{id}_m) : \mathrm{im}(\varphi) \to \mathrm{im}(\varphi)$ is the identity on $\mathrm{im}(\varphi) \subseteq \mathbb{R}^m$ and similarly

$$\begin{aligned} D_q f\left(\frac{\partial}{\partial y_j}|_q\right) = \frac{\partial(f \circ \varphi^{-1})}{\partial_{c+j}}(\varphi(q)) = \frac{\partial}{\partial_{c+j}}(f(p) + \mathrm{id}_1^2 + \cdots + \mathrm{id}_c^2 - \mathrm{id}_{c+1}^2 - \cdots - \mathrm{id}_m^2)|_{\varphi(q)} \\ = -2\,\mathrm{id}_{c+j}\,|_{\varphi(q)} = -2y_j(q). \end{aligned} \tag{3.15}$$

Therefore $D_q f = 0$ if and only if $2x_i(q) = 0$ and $-2y_j(q) = 0$ for every $1 \le i \le c$ and $1 \le j \le k$, that is $x_i(q) = 0$ and $y_j(q) = 0$, meaning that $\varphi(q) = 0$. However, $\varphi(p) = 0$ so bijectivity of φ means that $p = q$, and so the only critical point in U is the non-degenerate critical point p. □

Corollary 3.4.2 (Finite critical points on compact manifolds)**.** *Let M be a compact smooth manifold and $f : M \to \mathbb{R}$ be a Morse function. Then f has finitely many critical points.*

Remark. If M is any smooth manifold (not necessarily compact), the set of critical points of a Morse function is at most countable, since by Lemma 3.4.1 the critical points are isolated, and an isolated subset of a second countable space is countable.

Proof. Let C denote the set of critical points of f. Since f is a Morse function, by definition every critical is non-degenerate, so by Lemma 3.4.1, for each $p \in C$ there exists an open set $U_p \subseteq M$ with $p \in U_p$ and $C \cap U_p = \{p\}$. Moreover, for each $q \in M \setminus C$ we have $D_q f \neq 0$, meaning that in some coordinate chart (U, φ) with $q \in U$, some partial derivative $\frac{\partial(f \circ \varphi^{-1})}{\partial x_i}(\varphi(q)) \neq 0$. Since this partial derivative is a continuous function, it remains nonzero in some open set $V_q \subseteq \mathrm{im}(\varphi)$ with $\varphi(q) \in V_q$. Therefore, by continuity of φ, $\varphi^{-1}(V_q) = W_q \subseteq M$ is an open set, and it fulfills $q \in W_q$ and $C \cap W_q = \varnothing$. The collection

$$\{U_p : p \in C\} \cup \{W_q : q \in M \setminus C\}$$

is an open cover of M, and by compactness there exists a finite subcover. Since each W_q contains no point of C, every point of C must be covered by one of the finitely many U_p in the subcover. Since each U_p contains exactly one point of C, the set C is finite. □

3.4.2 Betti numbers and Morse inequalities

Definition 3.4.1 (Betti numbers)**.** Given a compact manifold M and a field $\mathbb{F}$, the kth *Betti number* $b_k(M, \mathbb{F})$, for $k \ge 0$, is defined by $b_k(M, \mathbb{F}) = \dim_{\mathbb{F}}(H_k(M, \mathbb{F}))$, where $H_k(M, \mathbb{F})$ is the kth *homology group* of M with coefficients in $\mathbb{F}$.

Remark. We will not define homology in this text, but in short, $b_k(M, \mathbb{F})$ counts the number of independent "k-dimensional cycles that are not boundaries" (k-dimensional holes) with linear dependence taken over $\mathbb{F}$ (detectable over $\mathbb{F}$). The Betti numbers are topological invariants, meaning that homeomorphic manifolds have the same Betti numbers.

Remark. The value of $b_k(M, \mathbb{F})$ does not always depend on $\mathbb{F}$. For example $b_0(M, \mathbb{F}) \ge 1$ is the number of connected components of $M \neq \varnothing$, and $b_k(M, \mathbb{F}) = 0$ for $k > \dim(M)$, are both independent of $\mathbb{F}$. However, the value can depend on $\mathbb{F}$, for example $b_1(\mathrm{SO}(3), \mathbb{R}) = 0$ and $b_1(\mathrm{SO}(3), \mathbb{Z}_2) = 1$, but it only depends on the "characteristic" of $\mathbb{F}$.

Lemma 3.4.3 (Strong Morse inequalities)**.** *Let M be a compact smooth manifold of dimension m, $f : M \to \mathbb{R}$ a Morse function, $c_{f,k}$ the number of critical points of f of index k, and $\mathbb{F}$ any field. Then for each $0 \leq j \leq m$*

$$\sum_{k=0}^{j}(-1)^{j-k}b_k(M,\mathbb{F}) \leq \sum_{k=0}^{j}(-1)^{j-k}c_{f,k}$$

with equality when $j = m$.

Proof. See Section 5 of Part I of [Mil63]. □

Remark. In the case when $j = m$, then

$$\chi(M) = \sum_{k=0}^{m}(-1)^{m-k}b_k(M,\mathbb{F}) = \sum_{k=0}^{m}(-1)^{m-k}c_{f,k}$$

where $\chi(M)$ is the *Euler characteristic* of M [Mil63, §5].

Corollary 3.4.4 (Weak Morse inequalities)**.** *Under the same assumptions as Lemma 3.4.3, then $c_{f,k} \geq b_k(M,\mathbb{F})$ for each $0 \leq k \leq m$.*

Proof. To avoid confusion with naming of index, let us prove that $c_{f,j} \geq b_j(M,\mathbb{F})$ for each $0 \leq j \leq m$. For $j = 0$, the strong Morse inequality gives $c_{f,0} \geq b_0(M,\mathbb{F})$ directly. For $1 \leq j \leq m$ adding the strong Morse inequality for j and for $j-1$ gives $c_{f,j} \geq b_j(M,\mathbb{F})$ (since the terms for a given k appear with opposite signs in the inequalities for j and $j-1$, and hence cancel when added). □

3.4.3 Morse number and perfect Morse functions

Definition 3.4.2 (Morse number)**.** The *Morse number* $\mathrm{Morse}(M)$ of a compact smooth manifold M of dimension m is the minimum total number of critical points over all Morse functions on M. That is, $\mathrm{Morse}(M) = \min\{\sum_{k=0}^{m} c_{f,k} \ : \ f : M \to \mathbb{R} \text{ is Morse}\}$.

Remark. By Lemma 3.4.2 this is a finite number and hence well-defined.

Corollary 3.4.5 (Lower bound on Morse number)**.** *For any compact smooth manifold M of dimension m*

$$\max_{\mathbb{F}} \sum_{k=0}^{m} b_k(M,\mathbb{F}) \leq \mathrm{Morse}(M) = \min_{f} \sum_{k=0}^{m} c_{f,k}.$$

Proof. This follow from the Weak Morse inequalities, summing over k and then making the inequality as sharp as possible by taking the maximum over the lower bound and the minimum over the upper bound. □

Lemma 3.4.6 (Morse functions on compact manifolds cannot only have minima)**.** *Let M be a compact smooth manifold with dimension $m \geq 1$. Then $\mathrm{Morse}(M) \geq 2$, and any Morse function $f : M \to \mathbb{R}$ has at least one critical point that is not a local minimum.*

Proof. Since f is continuous by Lemma 2.4.6 and M is compact, then f attains its maximum at a point $p \in M$. A maximum is in particular a local maximum, which must be a critical point (define and argue analogously to Definition 3.3.4). Arguing analogously to Lemma 3.3.2, p has index $m \geq 1$, and p is not a local minimum since by Lemma 3.3.2, local minima have index 0. Hence f has at least one point that is not a local minimum, and since this holds for any Morse function $\min_f c_{f,0} \geq 1$ and $\min_f c_{f,m} \geq 1$, meaning that $\mathrm{Morse}(M) \geq 2$. □

Definition 3.4.3 (Perfect Morse function)**.** A Morse function $f : M \to \mathbb{R}$ on a compact smooth manifold M of dimension m is called *perfect* over $\mathbb{F}$ if and only if $c_{f,k} = b_k(M, \mathbb{F})$ for every $0 \leq k \leq m$.

Lemma 3.4.7 (Perfect Morse functions achieve the Morse number)**.** *Let $f : M \to \mathbb{R}$ be a Morse function on the compact smooth manifold M of dimension m. Then f is perfect over $\mathbb{F}$ if and only if*

$$\sum_{k=0}^{m} c_{f,k} = \text{Morse}(M) = \sum_{k=0}^{m} b_k(M, \mathbb{F}),$$

meaning that equality holds in Corollary 3.4.5 and both f and $\mathbb{F}$ achieve this equality. Moreover, in this case $\mathbb{F}$ maximizes each term individually $\max_{\mathbb{F}'} b_k(M, \mathbb{F}') = b_k(M, \mathbb{F}) = c_{f,k}$.

Proof. ($\implies$) Suppose f is perfect over $\mathbb{F}$, so $c_{f,k} = b_k(M, \mathbb{F})$ for every k. By Corollary 3.4.5,

$$\sum_{k=0}^{m} b_k(M, \mathbb{F}) \leq \text{Morse}(M) \leq \sum_{k=0}^{m} c_{f,k} = \sum_{k=0}^{m} b_k(M, \mathbb{F})$$

so $\text{Morse}(M) = \sum_{k=0}^{m} b_k(M, \mathbb{F})$. If there were a field $\mathbb{F}'$ with $\sum_{k=0}^{m} b_k(M, \mathbb{F}') > \sum_{k=0}^{m} b_k(M, \mathbb{F})$ then Corollary 3.4.5 would give $\text{Morse}(M) \geq \sum_{k=0}^{m} b_k(M, \mathbb{F}') > \text{Morse}(M)$, a contradiction. Hence $\mathbb{F}$ maximizes $\sum_{k=0}^{m} b_k(M, \mathbb{F})$ and achieves the upper bound in Corollary 3.4.5. Moreover, for any field $\mathbb{F}'$, the weak Morse inequalities give $b_k(M, \mathbb{F}) \leq c_{f,k} = b_k(M, \mathbb{F})$ for each k, so $\mathbb{F}$ maximizes each term individually.

($\impliedby$) Suppose f realizes the minimum $\min_f \sum_{k=0}^{m} c_{f,k} = \text{Morse}(M) = \sum_{k=0}^{m} b_k(M, \mathbb{F})$. That is $\sum_{k=0}^{m} c_{f,k} = \sum_{k=0}^{m} b_k(M, \mathbb{F})$ The weak Morse inequalities give $c_{f,k} \geq b_k(M, \mathbb{F})$ for each k, and equality of the sums then forces $c_{f,k} = b_k(M, \mathbb{F})$ for every k. Hence f is perfect over $\mathbb{F}$. □

Remark. Because of this lemma, if M is a manifold that admits a perfect Morse function, we can say that $f : M \to \mathbb{R}$ is *perfect* if and only if $\sum_{k=0}^{m} c_{f,k} = \text{Morse}(M)$.

Lemma 3.4.8 (Betti numbers of $\text{SO}(n)$)**.** $\max_{\mathbb{F}} b_k(\text{SO}(n), \mathbb{F}) = b_k(\text{SO}(n), \mathbb{Z}/2)$.

Proof. In [Hat02], see Proposition 3D.3, Corollary 3A.4 (with $A = \mathbb{Z}$, $G = \mathbb{F}$), and Proposition 3A.5(3,5). □

Lemma 3.4.9 ($\text{SO}(n)$ admits perfect Morse functions)**.** $\text{Morse}(\text{SO}(n)) = 2^{n-1}$*, and if* $D = \text{diag}(d_1, \ldots, d_n)$ *with* $0 \leq d_1 < \cdots < d_n$ *then*

$$f : \text{SO}(n) \to \mathbb{R}, \quad R \mapsto \text{tr}(DR) = d_1 R_{11} + \cdots + d_n R_{nn}$$

is a perfect Morse function of $\text{SO}(n)$ *having I as its unique local (and global) maximum.*

Remark. $\text{tr}(DR) =< D^{\mathsf{T}}, R >=< D, R >$ where $< -, - >$ is the Frobenius inner product, which in general is an inner product on the vector space of rectangular matrices with any specified number of rows and columns [RC12, Section 5.2].

Remark. The critical points are the 2^{n-1} matrices of the form

$$\begin{bmatrix} \pm 1 & 0 & \cdots & 0 \\ 0 & \pm 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \pm 1 \end{bmatrix} \tag{3.16}$$

which have determinant 1. The number 2^{n-1} can be calculated by choosing the first $n-1$ signs arbitrarily, and then choosing the last sign such that the determinant is equal to 1.

Proof. Firstly f is a perfect Morse function by Theorem 3 of [Sol16], and its critical points are as in (3.16) by Lemma 2 of [Sol16]. Since there are 2^{n-1} critical points, $\mathrm{Morse}(\mathrm{SO}(n)) = 2^{n-1}$. This is attained by $\mathbb{F} = \mathbb{Z}/2$ by Lemma 3.4.8. Secondly, f is continuous by Lemma 2.4.6, and since $\mathrm{SO}(n)$ is compact f attains its maximum, which necessarily at a critical point, arguing as in the proof of Lemma 3.4.6. Since $0 \leq d_1 < \cdots < d_n$, the critical point where f attains the largest value is at $R = I$, therefore I is the unique global maximum.

Since it is a global maximum, it is necessarily a local maximum, and uniqueness follows from the following. Let P be a critical point of f as in 3.16, and let $1 \leq i_1 < \cdots < i_t \leq n$, be the indices $i \in \{i_1, \ldots, i_t\}$ where $P_{ii} = 1$. Then, by the first formula on p.319 of [Sol16], the index of P is $(i_1 - 1) + \cdots + (i_t - 1)$, where P is a local maximum if and only if its index is $\dim(\mathrm{SO}(n)) = n(n-1)/2$ by the Morse lemma. Every P which is not I has at least one entry which is -1, so the index of P is strictly less than the index of I. Since I is a local maximum, this means that P is not a local maximum, and hence I is the unique local maximum. As a sanity check we compute

$$\mathrm{index}(I) = (1-1)+(2-1)+\cdots+(n-1) = 1+\cdots+(n-1) = n(n-1)/2 = \dim(\mathrm{SO}(n)). \quad (3.17)$$

□

Corollary 3.4.10 (Sum of Betti numbers)**.** $\sum_{k=0}^{m} b_k(\mathrm{SO}(n), \mathbb{Z}/2) = 2^{n-1}$*, where* $m = \dim(\mathrm{SO}(n)) = n(n-1)/2$.

Lemma 3.4.11 (Rotation and shift)**.** *The function*

$$f : \mathrm{SO}(n) \to \mathbb{R}, \quad R \mapsto \mathrm{tr}(D) - \mathrm{tr}(D(P^*)^{\mathsf{T}} R)$$

is a perfect Morse function of $\mathrm{SO}(n)$ *having* $P^* \in \mathrm{SO}(n)$ *as its unique local (and global) minimum with* $f(P^*) = 0$*,* $f(P_i) \geq 1$ *for each critical point* P_i*,* $1 \leq i \leq 2^{n-1} - 1$ *which is not a local minimum, and* $|f(P_i) - f(P_j)| \geq 1$ *for* $i \neq j$*, where* $D = \mathrm{diag}(d_1, \ldots, d_n)$ *and* $d_i = 2^{i-1}$.

Remark. $P_i = P^*\mathcal{E}$, where $\mathcal{E} \neq I$ is of the form (3.16), $\det(\mathcal{E}) = 1$, and $f(P^*\mathcal{E}) = \sum_{j\,:\,\mathcal{E}_{jj}=-1} 2^i$.

Proof. Since $0 \leq d_1 = 1 < \cdots < d_n = 2^{n-1}$, by Lemma 3.4.9, $g : R \mapsto \mathrm{tr}(DR)$ is a perfect Morse function with I as its unique local and global maximum, and therefore $R \mapsto \mathrm{tr}(D(P^*)^{-1}R) = \mathrm{tr}(D(P^*)^{\mathsf{T}}R)$ has P^* as its unique local and global maximum. We make this the unique local and global minimum by multiplying with -1, that is $R \mapsto -\mathrm{tr}(D(P^*)^{\mathsf{T}}R)$. Since $P^* \mapsto -\mathrm{tr}(D(P^*)^{\mathsf{T}}P^*) = -\mathrm{tr}(D)$, we make P^* map to 0 by adding $\mathrm{tr}(D)$. That is $f : R \mapsto \mathrm{tr}(D) - \mathrm{tr}(D(P^*)^{\mathsf{T}}R)$ has P^* as its unique local and global minimum with $f(P^*) = 0$.

Now, f is a perfect Morse function with critical points P^*P, where P is as in (3.16) because of the following. Firstly, the function $\psi : \mathrm{SO}(n) \to \mathrm{SO}(n)$, $R \mapsto (P^*)^{\mathsf{T}}R$ is a diffeomorphism, and by the chain rule

$$D_Q(g \circ \psi) = (D_{\psi(Q)}g) \circ D_Q\psi = 0 \iff (D_{\psi(Q)}g) = 0 \quad (3.18)$$

since ψ being a diffeomorphism means that $D_Q\psi$ is an isomorphism. Therefore, Q is a critical point of $g \circ \psi$ if and only if $\psi(Q) = (P^*)^{\mathsf{T}}Q = P$ is a critical point of g, that is, $Q = P^*P$.

Secondly, by Lemma 3.2.3, $[\mathrm{Hess}(g \circ \psi)|_Q] = [D_Q\psi]^{\mathsf{T}}[\mathrm{Hess}(g)|_{\psi(Q)}][D_Q\psi]$, and since $D_Q\psi$ is an isomorphism, it is invertible. Therefore, by Sylvester's law of inertia [RC12, Theorem 4.5.8], the index of Q is the same as the index of $\psi(Q)$. Therefore $g \circ \psi$ is perfect with the same Betti numbers as g.

Thirdly, multiplying by -1 does not change the set of critical points, and it negates the Hessian at each critical point, meaning that if a critical point of $g \circ \psi$ had index k, then the same critical point of $-g \circ \psi$ will have index $m - k$, where $m = \dim(\mathrm{SO}(n))$. Therefore

$$\mathrm{Morse}(\mathrm{SO}(n)) = \sum_{k=0}^{m} c_{g\circ\psi,k} = \sum_{k=0}^{m} c_{-g\circ\psi,m-k} = \sum_{k=0}^{m} c_{-g\circ\psi,k}, \quad (3.19)$$

meaning that $-g \circ \psi$ is also perfect.

Finally, adding the constant $\operatorname{tr}(D)$ does not affect the critical points or the Hessian, meaning that $f = \operatorname{tr}(D) - g \circ \psi$ is perfect with critical points P^*P.

Since P^* is the unique critical point which is as local minimum, and f has 2^{n-1} critical points by Lemma 3.4.9, there are $N = 2^{n-1} - 1$ critical points which are not local minima.

What remains to be proven is that $f(P_i) \geq 1$ and $|f(P_i) - f(P_j)| \geq 1$ for $i \neq j$. for the critical points P_i which are not local minima. Now $f(P^*P) = \operatorname{tr}(D) - g(P)$, and by Lemma 3.4.9, P is of the form (3.16), so by the second formula on p.319 of [Sol16]

$$g(P) = 2 \sum_{i\,:\,P_{ii}=1} d_i - \sum_{i=1}^{n} d_i = 2 \sum_{i\,:\,P_{ii}=1} 2^{i-1} - \operatorname{tr}(D) = \sum_{i\,:\,P_{ii}=1} 2^i - \operatorname{tr}(D), \tag{3.20}$$

and since $\operatorname{tr}(D) = \sum_{i=1}^{n} 2^{i-1}$

$$f(P^*P) = 2\operatorname{tr}(D) - \sum_{i\,:\,P_{ii}=1} 2^i = \sum_{i=1}^{n} 2^i - \sum_{i\,:\,P_{ii}=1}^{n} 2^i = \sum_{i\,:\,P_{ii}=-1} 2^i \tag{3.21}$$

because $P_{ii} \in \{1, -1\}$. For each $P \neq I$, the value $f(P^*P) = \sum_{i\,:\,P_{ii}=-1} 2^i$ represents a unique positive binary integer, and therefore all critical values are distinct, they are at least 1, and at least 1 apart. □

3.5 Perfection on products

Lemma 3.5.1 (Number of critical points)**.** *Let $f_i : M_i \to \mathbb{R}$ be a Morse function with critical point set C_i, $1 \leq i \leq n$. Then*

$$f : M_1 \times \cdots \times M_n \to \mathbb{R}, \quad (q_1, \ldots, q_n) \mapsto f_1(q_1) + \cdots + f_n(q_n)$$

has the critical point set $C = C_1 \times \cdots \times C_n$. Moreover, if C^k is the set of critical points of f with index k, and C_i^k the set of critical points of f_i with index k, then

$$C^k = \bigcup_{k_1 + \cdots + k_n = k} C_1^{k_1} \times \cdots \times C_n^{k_n}.$$

Proof. By Lemma 3.3.5 $p \in C$ if and only if $p = (p_1, \ldots, p_n)$, where $p_i \in C_i$ for $1 \leq i \leq n$. Therefore $C = C_1 \times \cdots \times C_n$. Also by Lemma 3.3.5, the index of p is the sum of the indices of the p_i, so p has index k if and only if the index of p_i is k_i and $k_1 + \cdots + k_n = k$. The result follows. □

Lemma 3.5.2 (Perfect Morse function on product)**.** *Let $f_1 : M_1 \to \mathbb{R}, \ldots, f_n : M_n \to \mathbb{R}$ be perfect Morse functions over the same field $\mathbb{F}$. Then*

$$f : M_1 \times \cdots \times M_n \to \mathbb{R}, \quad (q_1, \ldots, q_n) \mapsto f_1(q_1) + \cdots + f_n(q_n)$$

is a perfect Morse function over $\mathbb{F}$.

Proof. By Lemma 3.3.5, f is a Morse function, and implicitly we are assuming that the M_i are compact to define perfection, where then $M_1 \times \ldots M_n$ is compact by Lemma 2.3.5, so we can define perfection on it.

By Lemma 3.5.1

$$c_{f,k} = \sum_{k_1 + \cdots + k_n = k} c_{f_1,k_1} \ldots c_{f_n,k_n} = \sum_{k_1 + \cdots + k_n = k} b_{k_1}(M_1, \mathbb{F}) \ldots b_{k_n}(M_n, \mathbb{F}) \tag{3.22}$$

since each f_i is perfect over $\mathbb{F}$. By the *Künneth formula over a field* [Hat02, Corollary 3B.7]

$$b_k(M_1 \times B_2) = \sum_{j=0}^{k} b_j(M_1, \mathbb{F}) b_{k-j}(B_2, \mathbb{F}) \tag{3.23}$$

which we can apply again for $B_2 = M_2 \times B_3$, and so on, until we arrive at

$$b_k(M_1 \times \cdots \times M_n, \mathbb{F}) = \sum_{k_1+\cdots+k_n=k} b_{k_1}(M_1, \mathbb{F}) \dots b_{k_n}(M_n, \mathbb{F}). \tag{3.24}$$

By (3.22), this means that $c_{f,k} = b_k(M_1 \times \cdots \times M_n, \mathbb{F})$, so f is perfect. □

Theorem 1 (Perfect prepared Morse function on product)**.** *The function*

$$f : \mathrm{SO}(n_1)^{b_1} \times \cdots \times \mathrm{SO}(n_a)^{b_a}, \quad (\mathbf{R}_{ij})_{1\le i\le a,\ 1\le j\le b_a} \mapsto \sum_{i=1}^{a}\sum_{j=1}^{b_i} \mathrm{tr}(\mathbf{D}_{ij}) - \mathrm{tr}(\mathbf{D}_{ij}(\mathbf{P}_{ij}^*)^\mathsf{T}\mathbf{R}_{ij})$$

is a perfect Morse function, having $P^* = (\mathbf{P}_{ij}^*)_{1\le i\le a,\ 1\le j\le b_a}$ as its unique local (and global) minimum, with $f(P^*) = 0$, $f(P_i) \ge 1$ for each critical point P_i, $1 \le i \le 2^{\sum_{i=1}^{a}(n_i-1)b_i} - 1$, which is not a local minimum, and $|f(P_i) - f(P_j)| \ge 1$ for $i \ne j$, where $\mathbf{D}_{11} = \mathrm{diag}(1, \dots, 2^{n_1-1})$, $\mathbf{D}_{12} = \mathrm{diag}(2^{n_1}, \dots, 2^{n_1+n_1-1})$, (continuing lexicographically), $\mathbf{D}_{ij} = \mathrm{diag}(2^{s_{ij}}, \dots, 2^{s_{ij}+n_i-1})$, $s_{ij} = (j-1)n_i + \sum_{k=1}^{i-1} n_k b_k$.

Proof. By Lemma 3.4.9 and Lemma 3.5.2, f is a perfect Morse function. By Lemma 3.3.5, P^* is the unique local minimum, and therefore the unique global minimum by compactness. Inserting P^* in the expression for f makes each term equal $\mathrm{tr}(\mathbf{D}_{ij}) - \mathrm{tr}(\mathbf{D}_{ij}(\mathbf{P}_{ij}^*)^\mathsf{T}\mathbf{P}_{ij}^*) = \mathrm{tr}(\mathbf{D}_{ij}) - \mathrm{tr}(\mathbf{D}_{ij}) = 0$, and therefore $f(P^*) = 0$. By Lemma 3.5.1 and Lemma 3.4.11, f has $N + 1 = \left(2^{n_1-1}\right)^{b_1} \dots \left(2^{n_a-1}\right)^{b_a} = 2^{\sum_{i=1}^{a}(n_i-1)b_i}$ critical points, and hence N critical points which are not local minima.

What remains to be proven is that $f(P_i) \ge 1$ and $|f(P_i) - f(P_j)| \ge 1$ for $i \ne j$, which we will prove similarly to the proof of Lemma 3.4.11, by showing that the critical values are distinct positive binary integers. By Lemma 3.3.5 and Lemma 3.4.11, the critical points of f are of the form $(\mathbf{P}_{ij}^*\mathbf{P}_{ij}^{L_{ij}})_{1\le i\le a,\ 1\le j\le b_a}$, where each $\mathbf{P}_{ij}^{L_{ij}} \in \mathrm{SO}(n_i)$, $1 \le L_{ij} \le 2^{n_i-1}$ is one of the 2^{n_i-1} critical points of the form (3.16) ($\mathbf{P}_{ij}^{L_{ij}} = \mathrm{diag}(\varepsilon_1, \dots, \varepsilon_{n_i})$, $\varepsilon_k \in \{1, -1\}$, $\det(\mathbf{P}_{ij}^{L_{ij}}) = 1$). Applying the same calculation as in the proof of Lemma 3.4.11 to each term $f_{ij} : \mathbf{R}_{ij} \mapsto \mathrm{tr}(\mathbf{D}_{ij}) - \mathrm{tr}(\mathbf{D}_{ij}(\mathbf{P}_{ij}^*\mathbf{R}_{ij}))$ with $\mathbf{D}_{ij} = \mathrm{diag}(2^{s_{ij}}, \dots, 2^{s_{ij}+n_i-1})$ (so that $d_k = 2^{s_{ij}+k-1}$), we obtain

$$f_{ij}(\mathbf{P}_{ij}^*\mathbf{P}_{ij}^{L_{ij}}) = \sum_{k\,:\,\left(\mathbf{P}_{ij}^{L_{ij}}\right)_{kk}=-1} 2^{s_{ij}+k} \tag{3.25}$$

and therefore

$$f((\mathbf{P}_{ij}^*\mathbf{P}_{ij}^{L_{ij}})_{1\le i\le a, 1\le j\le b_a}) = \sum_{i=1}^{a}\sum_{j=1}^{b_i} \sum_{k\,:\,\left(\mathbf{P}_{ij}^{L_{ij}}\right)_{kk}=-1} 2^{s_{ij}+k}. \tag{3.26}$$

The claim is now that the exponents $s_{ij} + k$, for $1 \le i \le a$, $1 \le j \le b_i$, $1 \le k \le n_i$, are all distinct. Indeed, ordering the set $S = \{(i, j, k) : 1 \le i \le a,\ 1 \le j \le b_i, 1 \le k \le n_i\}$ lexicographically, yields that the lexicographical position of (i, j, k) is equal to

$$\begin{aligned} &|\{(i', j', k') \in S : i' < i\}| + |\{(i, j', k') \in S : j' < j\}| + |\{(i, j, k') \in S : k' < k\}| + 1 \\ &\qquad = \left(\sum_{l=1}^{i-1} n_l b_l\right) + n_i(j-1) + (k-1) + 1 \\ &\qquad = s_{ij} + k \end{aligned} \tag{3.27}$$

and so there is a bijection $S \cong \{s_{ij} + k : (i,j,k) \in S\} = \{1, \ldots, |S|\}$, meaning that the exponents are all distinct.

Now, to every critical point $(\mathbf{P}_{ij}^* \mathbf{P}_{ij}^{L_{ij}})_{1 \le i \le a, 1 \le j \le b_a}$ we can associate a unique subset of S

$$T = \left\{ (i,j,k) \in S : \left(\mathbf{P}_{ij}^{L_{ij}} \right)_{kk} = -1 \right\} \subset S, \tag{3.28}$$

which is unique because $\mathbf{P}_{ij}^{L_{ij}}$ is uniquely determined by its diagonal entries (since it is a diagonal matrix with values in $\{1, -1\}$). Hence

$$f((\mathbf{P}_{ij}^* \mathbf{P}_{ij}^{L_{ij}})_{1 \le i \le a, 1 \le j \le b_a}) = \sum_{(i,j,k) \in T} 2^{s_{ij}+k} \tag{3.29}$$

is a unique sum of powers of 2, so the critical values are unique non-negative numbers, expressed in binary. Since P^* is uniquely associated to $T = \varnothing$, every other critical value must be a positive integer. Since the values are unique, they must also be at least 1 apart. □

Chapter 4

Hybrid Feedback Control

Some systems do not have continuous dynamics, but rather exhibit a combination of continuous and discrete changes. We will call these *hybrid dynamical systems*. In control applications, steering typically produces a continuous change, whereas switching or button presses may induce discrete state changes. We will introduce control of hybrid dynamical systems by letting our state be governed by a *hybrid plant* system and our controller be governed by a *hybrid controller* system for which the total dynamics will be governed by the *interconnection* of the plant and controller systems. Given a hybrid plant the goal of hybrid feedback control is to find a hybrid controller such that their interconnection has desirable properties.

4.1 Hybrid Dynamical Systems

4.1.1 Motivating example and hybrid time

One system that naturally follows hybrid dynamics is that of a bouncing ball. We will go through this example and see that we need to introduce a richer notion of time that records not only the elapsed continuous time but also the number of discrete changes that have occurred.

Example 27 (Bouncing ball 1)**.** Consider the dynamics of a ball falling straight down until it hits the ground, at which point the ball bounces up and loses some energy.

Say that it has height x over the ground and velocity v while in the air, meaning that the state of the system is given by $c : I \to \mathbb{R}^2,\ t \mapsto c(t) = (x(t), v(t))$. While the ball is falling it follows the continuous dynamics

$$\dot{c} = \begin{bmatrix} \dot{x} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} v \\ -g \end{bmatrix} \tag{4.1}$$

where $g \approx 9.8 m/s^2$ is the acceleration due to gravity. The ball hits the ground when $x(t) = 0$, and then the direction of motion will be flipped and the ball will lose some energy. So when $x(t) = 0$ the velocity $v(t)$ will change to $-rv(t)$, where $0 < r < 1$ is the "coefficient of restitution". This is a discrete change of the system, and the whole dynamical system can be written as

$$\begin{cases} \dot{c} = \begin{bmatrix} v \\ -g \end{bmatrix}, & c \in \mathbb{R}^2 \\ c = (x, v) \mapsto (x, -rv), & \text{if } c \in \{0\} \times \mathbb{R}. \end{cases} \tag{4.2}$$

Before proceeding we note that there is a slight issue with the discrete change of the system. This is because the discrete change does not move the state out of the condition for a discrete change, so it is not well-defined how many discrete changes should occur at this time. To fix this we introduce the concept of "hybrid time", for which a solution c to the system depends

not just on the continuous time in the interval I, and we will declare that the system (4.2) is a rule that a solution must follow, where the number of discrete changes is encoded in the domain of definition.

The idea will be to consider a continuous time $t \in \mathbb{R}$ and a discrete time $j \in \mathbb{N} = \{0, 1, 2, \dots\}$, encoding the number of discrete changes, where a solution to a hybrid system will first be an interval of continuous time $I_0 = [t_0, t_1] \subseteq \mathbb{R}$ where the solution follows the continuous dynamics of the system for $t \in]t_0, t_1[$ and no discrete change has yet happened, so $j = 0$. Then at $t = t_1$ the first discrete change happens, so $j = 1$, and then there is some interval of continuous time $I_1 = [t_1, t_2]$ where the solutions follows the continuous dynamics of the system for $t \in]t_1, t_2[$, until another discrete change happens at $t = t_2$, and so on.

Each point in time will then be encoded by a pair (t, j), and we note that there is a natural chronology outlined here. Time (t, j) comes after a time $(\widehat{t}, \widehat{j})$ if the continuous time is later $t > \widehat{t}$ or if the continuous time is the same $t = \widehat{t}$ but the number of discrete changes is larger, that is, $j > \widehat{j}$. This gives an ordering, because if the continuous time is less $\widehat{t} < t$, then the number of total discrete changes at that time is definitely less than or equal $\widehat{j} \leq j$. We note that we can describe this ordering by saying that $(t, j) \geq (\widehat{t}, \widehat{j})$ if and only if $t + j \geq \widehat{t} + \widehat{j}$.

Definition 4.1.1 (Hybrid time domain)**.** A *hybrid time domain* is a set $E \subset \mathbb{R} \times \mathbb{N}$ of pairs (t, j), where for each $j \leq N$ there exists an interval $I_j = [t_j, t_{j+1}] \subseteq \mathbb{R}$ such that

$$E = \bigcup_{j=0}^{N} I_j \times \{j\}.$$

This is also a totally ordered set with $(t, j) \geq (\widehat{t}, \widehat{j})$ if and only if $t+j \geq \widehat{t}+\widehat{j}$, for $(t, j), (\widehat{t}, \widehat{j}) \in E$.

Remark. Here we use the convention $0 \leq N \leq \infty$, and $-\infty \leq t_j \leq t_{j+1} \leq \infty$, so we get a sequence $-\infty \leq t_0 \leq t_1 \leq t_2 \leq \cdots \leq \infty$ in $\mathbb{R} \cup \{-\infty, \infty\}$, where we consider each interval $[t_j, t_{j+1}]$ to be a subset of $\mathbb{R}$, so $[-\infty, b] =]-\infty, b]$, and $[a, \infty] = [a, \infty[$ and $[-\infty, \infty] =]-\infty, \infty[= \mathbb{R}$. We note that this also means that $E \subset \mathbb{R} \times \mathbb{N}$, and so the relation $t + j \geq \widehat{t} + \widehat{j}$ makes sense. Moreover, at most one of the values $t_0, t_1, t_2 \dots$ is equal to $-\infty$, and at most one of the values is equal to ∞, which is true because the values t_j, t_{j+1} arise from intervals $I_j \subseteq \mathbb{R}$.

Example 28. One situation we will encounter is the following. We have a finite number of discrete changes N and these happen at N distinct times $t_1 < t_2 < \cdots < t_N$. The starting time t_0 is typically equal to 0 and the ending time t_{N+1} is typically equal to infinity. That is $t_0 = 0 < t_1 < t_2 < \cdots < t_N < t_{N+1} = \infty$, and so

$$E = ([0, t_1] \times \{0\}) \cup ([t_1, t_2] \times \{1\}) \cup \dots \cup ([t_{N-1}, t_N] \times \{N-1\}) \cup ([t_N, \infty[\times \{N\}). \quad (4.3)$$

If we wish to encode two discrete changes happening at the same time, say at time t_α, then this would mean that $t_\alpha = t_{\alpha+1}$, so we would then have discrete changes at times $t_1 < t_2 < \dots t_{\alpha-1} < t_\alpha = t_{\alpha+1} < t_{\alpha+2} < \cdots < t_N$, meaning that $I_\alpha = \{t_\alpha\}$, and then

$$([t_\alpha, t_{\alpha+1}] \times \{\alpha\}) \cup ([t_{\alpha+1}, t_{\alpha+2}] \times \{\alpha+1\}) = (\{t_\alpha\} \times \{\alpha\}) \cup ([t_\alpha, t_{\alpha+2}] \times \{\alpha+1\}). \quad (4.4)$$

Continuing like this we can represent the situation of doing any finite number of discrete changes at arbitrary times.

Example 29. A non-typical but also possible situation is when we have an infinite number of discrete changes $N = \infty$ happening at a single continuous time. If this is the case, then we note that this can only happen at one continuous time, and that this will be at a final finite continuous time $t_n = t_{n+1} = t_{n+2} = \cdots < \infty$. This is because all continuous times after t_n

must be equal, since otherwise there could not be an infinite number of discrete changes. In this case

$$E = ([t_0, t_1] \times \{0\}) \cup \ldots \cup ([t_{n-1}, t_n] \times \{n-1\}) \cup (\{t_n\} \times \{n, n+1, n+2, \ldots\}). \quad (4.5)$$

Example 30 (Bouncing ball 2)**.** Continuing the example with the bouncing ball, but with hybrid time, we now consider functions $c : E \to \mathbb{R}^2$, $(t, j) \mapsto (x(t, j), v(t, j))$, where E is a hybrid time domain and let $\dot{c}(t, j) = \frac{dc}{dt}(t, j)$, and $c^+(t, j) = c(t, j + 1)$. The hybrid dynamical system will now be written

$$\begin{cases} \dot{c} = \begin{bmatrix} v \\ -g \end{bmatrix}, & c \in \mathbb{R}^2 \\ c^+ = (x, v)^+ = (x, -rv), & \text{if } c \in \{0\} \times \mathbb{R}, \end{cases} \quad (4.6)$$

and the function c will be considered a solution to this system, if given

$$E = [t_0, t_1] \times \{0\} \quad \cup \quad [t_1, t_2] \times \{1\} \cup \ldots \quad (4.7)$$

that indeed for

$$(t, j) \in]t_0, t_1[\times\{0\} \cup]t_1, t_2[\times\{1\} \cup \ldots \quad (4.8)$$

that we have

$$\dot{c}(t, j) = \begin{bmatrix} v(t, j) \\ -g \end{bmatrix} \quad (4.9)$$

and for

$$(t, j) \in \{(t_1, 0), (t_2, 1), \ldots\} \quad (4.10)$$

that $c(t, j) \in \{0\} \times \mathbb{R}$ and

$$c^+(t, j) = c(t, j + 1) = (0, v(t, j + 1)) = (0, -rv(t, j)). \quad (4.11)$$

4.1.2 Defining and generalizing hybrid dynamical systems

Now we can finally define hybrid dynamical systems. We start with the usual definition for subsets of $\mathbb{R}^n$ and then appropriately generalize them to smooth manifolds. Prior to the definition, we first note that later when we define feedback we need some more flexibility for the discrete changes, or for the *jump function* as it will be called. It turns out that we need the possibility for at least two different discrete changes (see the discussion in subsection 4.2.3), which is not possible if we have a function, which would give a unique discrete change. Therefore we require that the so-called jump function to be set-valued.

Without restriction we can also allow the *flow function*, corresponding to the continuous dynamics, to be set valued. This is not necessary for feedback, but makes our definition for a hybrid dynamical system agree with that of a hybrid inclusion system in [GST12], and makes it easier to define robustness.

Definition 4.1.2 (Set-valued function)**.** The notation $f : A \rightrightarrows B$ means a that $f : A \to \mathcal{P}(B)$, that is $f(a)$ is a collection of elements of B for every $a \in A$. If $f(a)$ is a singleton set for every $a \in A$, then we can view f as a normal function $f : A \to B$. An important case to keep in mind is if $f(a) = \varnothing$ is an empty collection.

Definition 4.1.3 (Hybrid dynamical system)**.** A classical *hybrid dynamical system* is a 4-tuple $\mathcal{H} = (C, F, D, G)$ where $C \subseteq \mathbb{R}^m$ is a set, $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is a function, $D \subseteq \mathbb{R}^m$ is a set, and $G : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is a function.

Remark. Definition 4.1.3 corresponds to a blueprint for (possibly many different kinds of) solutions to a dynamical system of the form

$$\mathcal{H}: \begin{cases} \dot{c} \in F(c), & c \in C \\ c^+ \in G(c), & c \in D, \end{cases}$$

where $c : E \to C$ is some unspecified *solution*. Here the hybrid time is implicit, but for clarity it could be written as follows.

$$\mathcal{H}: \begin{cases} \dot{c}(t,j) \in F(c(t,j)), & \text{if } c(t,j) \in C \\ c^+(t,j) \in G(c(t,j)), & \text{if } c(t,j) \in D. \end{cases}$$

Remark. From now on we will call the continuous change of a hybrid system the *flow* and the discrete change the *jump*, where we will call C the *flow set*, F the *flow function*, D the *jump set* and G the *jump function*. A point of the system $c(t,j)$ will be called a *state*, and $c : E \to \mathbb{R}^m$ will be called the *state function*.

Generalization to manifolds will be done in the following way.

Definition 4.1.4 (Hybrid dynamical system on manifold)**.** A general *hybrid dynamical system* is a 4-tuple $\mathcal{H} = (C, F, D, G)$ where $C \subseteq M$ is a set, M is a smooth manifold, $F : M \rightrightarrows TM$ is such that if $F(p) \subseteq \{p\} \times T_pM$, $D \subseteq M$ is a set, and $G : M \rightrightarrows M$ is a function.

Remark. M will be called the state space C, F, D and G will be called the same as if it were a classical hybrid dynamical system, and when mentioning a hybrid dynamical system we will refer to a general hybrid dynamical system.

Remark. We choose this as our generalization of hybrid dynamical systems, since firstly, if $M = \mathbb{E}^m$ then this definition agrees with Definition 4.1.3, see Definition 2.11 of [San21] (in the context of closed-loop systems which we will get to in the next section), which agrees with Definition 2.2 of [GST12].

Secondly, if the flow function is single-valued $F : M \to TM$ so it becomes a vector field and $D = \varnothing$ so there is no discrete dynamics, then this definition agrees with standard definitions for continuous dynamical systems on manifolds, see the definition of integral curve in [Lee13] (which agrees with standard definitions for autonomous first-order ODEs if $M = \mathbb{E}^m$) and in Definition 1.14 of [AS04], if we further assume that F is a smooth vector field.

Thirdly, if the jump function is single-valued $G : M \to M$ and $C = \varnothing$ so there is no continuous dynamics, then our definition agrees with that of discrete dynamical systems (on manifolds).

4.1.3 Solutions to hybrid dynamical systems

Definition 4.1.5 (Flow time set and jump time set)**.** Given a hybrid time domain

$$E = \bigcup_{j=0}^{N} [t_j, t_{j+1}] \times \{j\},$$

we define the *flow time set* of E to be

$$\text{FlowTime}(E) := \bigcup_{j=0}^{N}]t_j, t_{j+1}[\times \{j\}$$

and the *jump time set* of E to be

$$\text{JumpTime}(E) := \{(t_{j+1}, j) : 0 \leq j < N\}.$$

Remark. We note that if $N < \infty$, $t_{N+1} < \infty$ and $t_0 > -\infty$ then we have the following partition of E

$$E = \{(t_0, 0)\} \cup \text{FlowTime}(E) \cup \text{JumpTime}(E) \cup \{(t_{N+1}, N)\}. \tag{4.12}$$

If $t_0 > -\infty$ and either $N = \infty$ or $t_{N+1} = \infty$ then we instead have this partition

$$E = \{(t_0, 0)\} \cup \text{FlowTime}(E) \cup \text{JumpTime}(E). \tag{4.13}$$

Finally, if $t_0 = -\infty$ and either $N = \infty$ or $t_{N+1} = \infty$ then this is the partition of E

$$E = \text{FlowTime}(E) \cup \text{JumpTime}(E). \tag{4.14}$$

Definition 4.1.6 (Solution to hybrid dynamical system)**.** A *solution* $c : E \to C$ to a hybrid dynamical system (C, F, D, G) is a function

$$c : E \to C,$$

where E is a hybrid time domain, and for each $(t, j) \in \text{FlowTime}(E)$

$$c|_{\text{FlowTime}(E)}(-, j) :]t_j, t_{j+1}[\to C, \quad t \mapsto c(t, j)$$

is a smooth function such that

$$c(t, j) \in C \quad \text{and} \quad \dot{c}(t, j) \in F(c(t, j)), \quad \text{for } (t, j) \in \text{FlowTime}(E)$$

and

$$c(t, j) \in D \quad \text{and} \quad c^+(t, j) = G(c(t, j)), \quad \text{for } (t, j) \in \text{JumpTime}(E).$$

Remark. Here we have defined

$$\dot{c}(t, j) := \dot{(c|_{\text{FlowTime}(E)}}(-, j))(t)$$

and

$$c^+(t, j) := c(t, j + 1).$$

(Note the derivative dot in the right-hand side.)

Remark. We now allow solutions to have an arbitrary amount of jumps made at a given time t_j, but alternatively we could enforce that solutions just do one jump each time the state enters the jump set. This is what makes sense in the example with the bouncing ball. If we do this, we have a clearer correspondence between a given hybrid dynamical system and a (unique) solution to this system, given that we have an initial point c_0 for the state $c(t_0, 0) = c_0 \in M$, similar to standard dynamical systems. However, for our analysis, there is no need to enforce this.

4.2 Hybrid Control Systems

Now that we have the framework of hybrid dynamical systems, we are ready to introduce the notion of control on these types of systems. Similarly to when defining hybrid dynamical systems, we will rely on the definition in [GST12] and [San21] such that if the system exhibits no jumps, the following *hybrid plant* will correspond to a dynamical system as in [AS04] or [Lee13].

The idea will be to introduce a *control parameter* to a hybrid dynamical system, turning it into a *hybrid plant*, which is a generalization of hybrid dynamical system, where *input* and *measurement* are allowed. Then this control parameter will be governed by a *hybrid controller*, which is another generalization of a hybrid dynamical system, where *input* and *output* are allowed. The *interconnection* of the hybrid plant and hybrid controller will be taking the input of the hybrid plant to be the output of the hybrid controller and the input of the hybrid controller will be the measurement (or *feedback*) of the hybrid plant. This process is called *closing the loop*, and hopefully results in a hybrid dynamical system, now called a *closed-loop hybrid system*.

4.2.1 Hybrid plant

Control theory originates from applications, where we model a real life machine, or *plant*, as a dynamical system, where the state can be influenced, or *controlled*, by some input $u \in U$, which we call a *control parameter*. Now we wish to do this with hybrid dynamical systems, where we will mathematically model plants in the following way.

Instead of a flow function $F : M \to TM$ we will allow F to depend on a control parameter $F : M \times U \to TM$. To distinguish this from a flow function of a hybrid dynamical system, we will typically denote this flow function by $F_P : M_P \to TM$, where $M_P = M \times U$ and P stands for plant. If the control parameter is specified, then we want the plant dynamics to correspond to a hybrid dynamical system, so we wish that $F_P(-, u) : M \to TM$ is a (smooth) vector field.

Now, for the flow set, we will consider some subset $C_P \subseteq M \times U$, which means that for a given point $p \in M$ we might only allow a subset of the control parameters u and vice-versa.

Moreover, for the jump function, now denoted G_P, we will also allow the jump set D_P to be an arbitrary subset of $M \times U$, so it could only include a subset of the control parameters.

Finally we wish to be able to measure the state $c(t, j)$, and do computations with the measurement $h(c(t, j))$. The computations we wish to do require the measurement to be in a normed (real) vector space, which we for simplicity choose to be $\mathbb{R}^l$, where l corresponds to the dimensionality of what we measure. Since we can not measure more than the whole state $c(t, j) \in M$, we can without loss of generality assume that $l \leq \dim(M)$.

Definition 4.2.1 (Hybrid plant)**.** A *hybrid plant* is a 5-tuple $\mathcal{H}_P = (C_P, F_P, D_P, G_P, h)$, where $C_P \subseteq M \times U$ and M is a smooth manifold, U is a set, $F_P : M \times U \to TM$ is a function such that $F_P(-, u) : M \to TM$ is a vector field for each $u \in U$, $D_P \subseteq M \times U$ is a subset, $G_P : D_P \to M$ is a function, and $h : M \to \mathbb{R}^l$ is a function where l is arbitrary. This system can be written as

$$\mathcal{H}_P : \begin{cases} \dot{z} = F_P(z, u), & (z, u) \in C_P \\ z^+ = G_P(z, u), & (z, u) \in D_P \\ y = h(z) \end{cases}$$

where $z : E \to M$, and $u \in U$ is left unspecified.

Remark. C_P is called the *flow set*, M the *state space*, U the *input set* or *control value set*, F_P the *flow function*, D_P the *jump set*, G_P the *jump function*, and h the *measurement function*.

4.2.2 Hybrid controller

Now we need a scheme for changing the value of the control parameter $u \in U$. The idea is that we will construct another generalization of a hybrid dynamical system, called a *hybrid controller*, usually denoted by $\mathcal{H}_K$ (with K standing for controller, since the letter C is already occupied). This will have an output $\zeta \in U$ that can be used as the input $u = \zeta$ to a given hybrid plant.

However, our output should depend on the state of the hybrid plant, since if we do not know what the state of the hybrid plant is, how would we know how to control it to a desirable state? To address this issue, we will allow hybrid controllers to have an input v which is the measurement of a hybrid plant $y \in \mathbb{R}^l$. For simplicity we will let $v = y$, so $v \in \mathbb{R}^l$.

The rule for what the output ζ should be should then depend on the input v. Since we want more flexibility, we will also allow the hybrid controller to have its own dynamics. We will let the dynamics of the controller be similar to that of a hybrid dynamical system, so that if we set $u = \zeta$ and $v = y$ this would result in a new *interconnected* system. If we are lucky it could be described as a hybrid dynamical system, meaning that we stay within the same framework.

To distinguish the state space of the hybrid plant from the state space of the hybrid controller we will typically denote the state space of the hybrid controller by N, where then naturally $C_K \subseteq N \times \mathbb{R}^l$, since we have input $v \in \mathbb{R}^l$. We will have a flow function F_K, jump set G_K and jump function G_K similar to a hybrid plant. But instead of a measurement function, we will have an output function κ. As discussed, for a given hybrid plant, this function should depend on both the state of the hybrid controller and the state of the measurement of the hybrid plant. Since the input of the hybrid controller will be the measurement of the hybrid plant, we more generally have that $\kappa : N \times \mathbb{R}^l \to U$, where without a hybrid plant given, U is just a set.

Definition 4.2.2 (Hybrid controller)**.** A *hybrid controller* is a 5-tuple $\mathcal{H}_K = (C_K, F_K, D_K, G_K, \kappa)$ such that $C_K \subseteq N \times \mathbb{R}^l$ where N is a smooth manifold and l is arbitrary, $F_K : N \times \mathbb{R}^l \to TN$ is a function such that $F_K(-, v) : N \to TN$ is a vector field for each $v \in \mathbb{R}^l$, $D_K \subseteq N \times \mathbb{R}^l$ is a subset, $G_K : N \times \mathbb{R}^l \to N$ is a function, and $\kappa : N \times \mathbb{R}^l \to U$ is a function where U is a set. This system can be written as

$$\mathcal{H}_K : \begin{cases} \dot{\eta} = F_K(\eta, v), & (\eta, v) \in C_K \\ \eta^+ = G_K(\eta, v), & (\eta, v) \in D_K \\ \zeta = \kappa(\eta, v) \end{cases}$$

where $\eta : E' \to N$, and $v \in \mathbb{R}^l$ is left unspecified.

Remark. C_K is called the *flow set*, $\mathbb{R}^l$ the *input set*, N the *state space*, F_K the *flow function*, D_K the *jump set*, G_K the *jump function*, U the *output set*, and κ the *output function*.

A more specific situation we will consider is when we wish for our control parameter $u \in U$ to come from a discrete set $U = Q$, where each value $u = q \in Q$ corresponds to some mode of the flow function F_P. For simplicity we will consider just two modes $q \in \{0, 1\}$, where the hybrid controller $\mathcal{H}_K$ governing this value q, which we will call a *logic variable*, will be quite simple, having no flow $F_K = 0$, and the jump function switching between the two modes. That is, if $q = 0$ then $G_K(q, v) = 1$ and if $q = 1$ then $G_K(q, v) = 0$. This can be accomplished by $G_K(q, v) = 1 - q$.

Definition 4.2.3 (Logic variables)**.** A *logic variable* q is the state of a hybrid controller of the form

$$\mathcal{H}_K : \begin{cases} \dot{q} = 0, & (q, v) \in C_K \\ q^+ = 1 - q, & (q, v) \in D_K \\ \zeta = \kappa(q, v). \end{cases}$$

That is $\mathcal{H}_K = (Q \times \mathbb{R}^l, 0, D_K, 1 - \mathrm{id}_1, \kappa)$, where $Q = \{0, 1\}$ and $\mathrm{id}_1 : Q \times \mathbb{R}^l \to Q,\ (q, v) \mapsto q$.

4.2.3 Hybrid feedback and closed-loop hybrid systems

We will now consider the *interconnection* of a hybrid plant $\mathcal{H}_P$ and a hybrid controller H_K. This is setting the input $u \in U$ of the hybrid plant to be equal to the output $\zeta(i, j) = \kappa(\eta(i, j), v) \in U$ of the hybrid controller, as well as setting the input $v \in \mathbb{R}^l$ of the hybrid controller to be equal to the measurement $y(i, j) = h(z(i, j)) \in \mathbb{R}^l$ of the hybrid plant. Therefore $u = \kappa(\eta, h(z))$ and we can rewrite the hybrid plant $\mathcal{H}_P = (C_P, F_P, D_P, G_P, h)$ as

$$\mathcal{H}_P^{\sim} : \begin{cases} \dot{z} = F_P(z, \kappa(\eta, h(z))), & (z, \kappa(\eta, h(z))) \in C_P \\ z^+ = G_P(z, \kappa(\eta, h(z))), & (z, \kappa(\eta, h(z))) \in D_P \\ v = h(z). \end{cases} \tag{4.15}$$

Since $v = h(z(i,j))$ we may rewrite the hybrid controller $\mathcal{H}_K = (C_K, F_K, D_K, G_K, \kappa)$ as

$$\mathcal{H}_K^{\sim} : \begin{cases} \dot{\eta} = F_K(\eta, h(z)), & (\eta, h(z)) \in C_K \\ \eta^+ = G_K(\eta, h(z)), & (\eta, h(z)) \in D_K \\ u = \kappa(\eta, h(z)). \end{cases} \tag{4.16}$$

However, we see that these $\mathcal{H}_P^{\sim}$ and $\mathcal{H}_K^{\sim}$ do not constitute a hybrid plant and hybrid controller as we have defined them. If we define a new state $x = (z, \eta)$ with $z = x_1$ and $\eta = x_2$ we obtain a new system, but the dynamics are not fully determined in all cases. The cases where both components of $x = (z, \eta)$ flow or where exactly one component jumps are clear, but the other cases remain ambiguous. The cases to consider are as follows.

$$\begin{cases} \dot{x} = \begin{bmatrix} F_P(x_1, \kappa(x_2, h(x_1))) \\ * \end{bmatrix}, & (x_1, \kappa(x_2, h(x_1))) \in C_P \text{ and } (x_2, h(x_1)) \notin C_K \\ \dot{x} = \begin{bmatrix} * \\ F_K(x_2, h(x_1)) \end{bmatrix}, & (x_1, \kappa(x_2, h(x_1))) \notin C_P \text{ and } (x_2, h(x_1)) \in C_K \\ \dot{x} = \begin{bmatrix} F_P(x_1, \kappa(x_2, h(x_1))) \\ F_K(x_2, h(x_1)) \end{bmatrix}, & (x_1, \kappa(x_2, h(x_1))) \in C_P \text{ and } (x_2, h(x_1)) \in C_K \\ x^+ = (G_P(x_1, \kappa(x_2, h(x_1))), x_2), & (x_1, \kappa(x_2, h(x_1))) \in D_P \text{ and } (x_2, h(x_1)) \notin D_K \\ x^+ = (x_1, G_K(x_2, h(x_1))), & (x_1, \kappa(x_2, h(x_1))) \notin D_P \text{ and } (x_2, h(x_1)) \in D_K \\ x^+ = *, & (x_1, \kappa(x_2, h(x_1))) \in D_P \text{ and } (x_2, h(x_1)) \in D_K \end{cases} \tag{4.17}$$

where "$*$" has been put where it is more ambiguous what should happen. In the first two cases we could reasonably let $* = 0$, since this would impose the dynamics of keeping that variable fixed. However, no matter what $*$ would be in these cases, we would have some new dynamics that do not immediately come from the dynamics of $\mathcal{H}_P$ or $\mathcal{H}_K$ which could cause problems even if $\mathcal{H}_P$ and $\mathcal{H}_K$ are well-behaved systems. Therefore we disregard the first two cases entirely.

In the last case, either one jump $(G_P(...), x_2)$ or $(x_1, G_K(...))$ could be made, or possibly both $(G_P(...), G_K(...))$. It turns out that allowing either variable to jump is no problem, but the simultaneous jump does not entirely make sense. This is because $G_P(x_1, \kappa(x_2, h(x_1)))$ depends on the value of x_2 before this jump is made, and $G_K(x_2, h(x_1))$ depends on the value of x_1 before this jump is made. Hence allowing the simultaneous jump $(G_P(...), G_K(...))$ would mean that x_2 should be the value before the jump is made to evaluate $G_P(...)$ correctly, but at the same time x_2 should be equal to $G_K(...)$ since the second variable jumps at the same time. This does not make sense (and similarly for x_1), and therefore in the last case, either jump is allowed, but not both simultaneously. This motivates the following definition.

Definition 4.2.4 (Interconnection)**.** Given a hybrid plant $\mathcal{H}_P = (C_P, F_P, D_P, G_P, h)$ with state space M and a hybrid controller $\mathcal{H}_K = (C_K, F_K, D_K, G_K, \kappa)$ with state space N, the *interconnection* of $\mathcal{H}_P$ and $\mathcal{H}_K$ is the hybrid dynamical system $\mathcal{H} = (C, F, D, G)$ with state space $M \times N$, where

$$D_1 = \{(x_1, x_2) \in M \times N \ : \ (x_1, \kappa(x_2, h(x_1))) \in D_P\}$$
$$D_2 = \{(x_1, x_2) \in M \times N \ : \ (x_2, h(x_1)) \in D_K\}$$

and

$$C = \{(x_1, x_2) \in M \times N \ : \ (x_1, \kappa(x_2, h(x_1))) \in C_P \text{ and } (x_2, h(x_1)) \in C_K\}$$
$$F : M \times N \to TM \oplus TN, \quad (x_1, x_2) \mapsto \begin{bmatrix} F_P(x_1, \kappa(x_2, h(x_1))) \\ F_K(x_2, h(x_1)) \end{bmatrix}$$
$$D = D_1 \cup D_2$$
$$G : M \times N \rightrightarrows M \times N,$$
$$(x_1, x_2) \mapsto \begin{cases} \{\ (G_P(x_1, \kappa(x_2, h(x_1))), x_2)\ \}, & \text{if } x \in D_1 \setminus D_2 \\ \{\ (x_1, G_K(x_2, h(x_1)))\ \}, & \text{if } x \in D_2 \setminus D_1 \\ \{\ (G_P(x_1, \kappa(x_2, h(x_1))), x_2),\ (x_1, G_K(x_2, h(x_1)))\ \}, & \text{if } x \in D_1 \cap D_2. \end{cases}$$

Remark. The first two cases for G are singletons, and the third case is a set with two elements. There is no need to specify $G(x)$ for $x \notin D$, since jumps are only exhibited for $x \in D$, so the value of $G(x)$ in this case is irrelevant.

Remark. This constitutes a hybrid dynamical system since M and N are smooth manifolds, so is $M \times N$, by Example 19. Moreover $T(M \times N) \cong TM \oplus TN$ canonically.

Definition 4.2.5 (Closed-loop system)**.** The interconnection of a hybrid plant $\mathcal{H}_P$ and hybrid controller $\mathcal{H}_K$ is also called the *closed-loop* system of $\mathcal{H}_P$ and $\mathcal{H}_K$.

Remark. The reason for this name can be seen in the figure below.

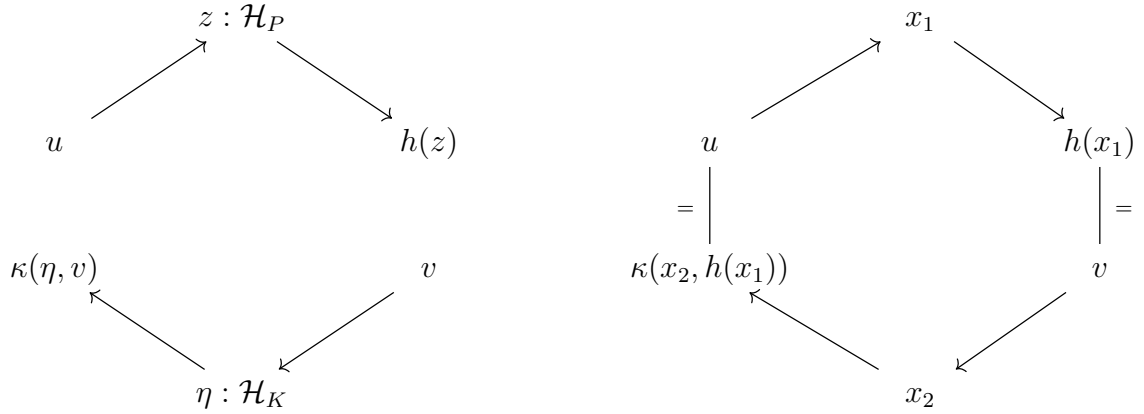

Figure 4.1: "Closing the loop" of the systems $\mathcal{H}_P$ and $\mathcal{H}_K$ (left) to create the *closed-loop* system (right).

Definition 4.2.6 (Hybrid feedback control)**.** Given a hybrid plant, *Hybrid feedback control* is the process of designing a hybrid controller such that their interconnection becomes a system that has desirable properties.

Remark. One desirable property is that the interconnection is a hybrid dynamical system. Other desirable properties could be *stability* of a specified point and *robustness* to disturbances, which we will cover in the next section.

Example 31 (Switching between vector fields)**.** Consider a hybrid plant where no jumps are exhibited ($D_P = \varnothing$) and the flow function is single-valued. This can be written as

$$\mathcal{H}_P : \begin{cases} \dot{z} = F_P(z, u), & (z, u) \in C_P \\ y = h(z) \end{cases} \tag{4.18}$$

corresponding to a standard control system, like in [AS04]. If $C_P \subset M \times \{0,1\}$ we may consider a particularly simple system $F_P(z,q) = (1-q)F_0(z) + qF_1(z)$, where $F_0, F_1 : M \to TM$ are vector fields. The idea is that if $u = q$ is a logic variable, then this can be seen as a system where we are switching between two different vector fields.

Say that when $(q, h(z)) \in D_K \subset \{0,1\} \times \mathbb{R}^l$ is when we wish to switch which vector field is used for the flow for z. This means that we have a controller

$$\mathcal{H}_K : \begin{cases} q^+ = 1 - q, & (q,v) \in D_K \\ \zeta = \kappa(q,v) \end{cases} \tag{4.19}$$

that governs the logic variable (which by definition has no flow $F_K = 0$). Moreover we will let $\kappa(q,v) = q$, so that the interconnection of $\mathcal{H}_P$ and $\mathcal{H}_K$, being $v = y = h(z)$, $u = \zeta = \kappa(q,v) = \kappa(q,h(z)) = q$ will be given by

$$\mathcal{H} : \begin{cases} \dot{z} = (1-q)F_0(z) + qF_1(z) & (z,q) \in C \\ q^+ = 1 - q & (z,q) \in D \end{cases} \tag{4.20}$$

where $C = \{(z,q) \in M \times \{0,1\} \ : \ (z, \kappa(q,h(z))) \in C_P\} = C_P$ (since $\kappa(q,h(z)) = q$) and $D = \{(z,q) \in M \times \{0,1\} : (q,h(z)) \in D_K\}$. Now we can write C_P and D_K as $C_P = C_0 \times \{0\} \cup C_1 \times \{1\}$ and $D_K = \{0\} \times D_0 \cup \{1\} \times D_1$, so

$$\mathcal{H} : \begin{cases} \dot{z} = (1-q)F_0(z) + qF_1(z) & z \in C_q \\ q^+ = 1 - q & h(z) \in D_q \end{cases} \tag{4.21}$$

where $C_0, C_1 \subseteq M$ and $D_0, D_1 \subseteq \mathbb{R}^l$ are arbitrary subsets to be specified. If we want a unique solution to this system (given an initial position) to be possible, we need $C_q \cap h^{-1}(D_q) = \varnothing$, and if we want any initial position $(z_0, q_0) \in M \times \{0,1\}$ to be possible, we need $C_q \cup h^{-1}(D_q) = M$. Therefore, if we specify C_1 and D_0 then $h^{-1}(D_1) = M \setminus C_1$ and $C_0 = M \setminus h^{-1}(D_0)$ necessarily. Without loss of generality we can let $D_1 = h(M \setminus C_1)$, since letting D_1 be a larger subset of $\mathbb{R}^l$ (outside the image of h) would never be measured.

Hence, if we want a unique solution and any initial position to be possible, it suffices to specify $C_1 \subset M$ and $D_0 \subset \mathbb{R}^l$, which may both be chosen arbitrarily. Equivalently to choosing D_0, one can choose the set $h^{-1}(D_0) \subset M$, that is, the set of $z \in M$ such that $h(z) \in D_0$, and then let $D_0 = h(h^{-1}(D_0))$ without loss of generality.

4.3 Desired properties of closed-loop systems

One of the main goals of control theory is to steer a dynamical system towards a particular point, keep the state near that point, and moreover to have the state converge to that point. If this can be done given any initial condition, this is known as *global asymptotic stability*.

Typically we steer a system using feedback control, but in real world applications, the measurement of the state used for feedback will be imprecise and other errors may well accrue. It may happen that under these assumptions, global asymptotic stability is lost. Therefore, another goal is to ensure that global asymptotic stability is maintained (at least to some neighborhood), even if there is some (bounded) error in the measurement function. This is known as *robust* feedback.

4.3.1 Stability and asymptotics

For applications we often care about the stability of a system. For example keeping an airplane using auto-pilot oriented correctly even if there is turbulence, or keeping a self-driving car on

the road even if there are changes in terrain. In these examples, the orientation of the airplane $p^* \in \mathrm{SO}(3)$, and the desired speed and direction of steering $p^* \in \mathbb{R} \times \mathrm{SO}(2)$ become particular points of the corresponding system that we wish to steer the system towards and then stay near that point.

In our framework, we would model each of the examples as a hybrid plant, with flow, jump and measurement function corresponding to the physical dynamics of the systems. Each control value would correspond to a way that we are allowed to alter the state, being the orientation of the airplane, and the speed and direction of steering of the self-driving car, and the measurement function would correspond to what sensors we have, like a gyroscope, speedometer, and so on.

One goal is then to come up with a hybrid controller such that the interconnection $\mathcal{H}$ would only have such solutions where if we are close to the desired state p^*, then we will stay close. This is known as the point p^* being *stable* for $\mathcal{H}$. We will mathematically capture this in the following way.

Say that a solution to $\mathcal{H}$ is $c : E \to C$. We can think of c being near p^* at a time (t, j) if $c(t, j) \in U$, where $U \subseteq C$ is some open set with $p^* \in U$ that we think consist of points that are near enough p^* for our applications. We can then say that we have some stability if there is some notion of c being close enough to p^* such that after c becomes close enough to p^*, it will stay close to p^* for all time. We will say that c is close enough to p^* if there exists some open subset $\widehat{U} \subseteq U$ such that if $c(\widehat{t}, \widehat{j}) \in \widehat{U}$ then for all future time $(t, j) \geq (\widehat{t}, \widehat{j})$ we stay near p^*, that is $c(t, j) \in U$. We will say that $p^* \in C$ is *stable* for $\mathcal{H}$ if this holds for every open set U we may choose.

Definition 4.3.1 (Stable point)**.** A point $p^* \in C$ is *stable* for a hybrid dynamical system $\mathcal{H}$ with state space C if and only if for each open set $U \subseteq C$ with $p^* \in U$ there exists an open set $\widehat{U} \subseteq U$ such that for every solution $c : E \to C$ of $\mathcal{H}$, if $c(\widehat{t}, \widehat{j}) \in \widehat{U}$ then $c(t, j) \in U$ for every $(t, j) \in E$ with $(t, j) \geq (\widehat{t}, \widehat{j})$

Remark. For simplicity we have defined stability for points of hybrid dynamical systems, but without much effort this can be defined for subsets of hybrid inclusion systems or more general hybrid systems.

Now we wish to capture the notion of a solution converging to a particular point p^*. That is, something like $\lim_{(t,j)\to\infty} c(t, j) = p^*$. But what do we mean with this? Firstly, we note that if $c : E \to M$ is a solution to some hybrid dynamical system, the hybrid time domain E matters a lot. Clearly to consider something like $\lim_{(t,j)\to\infty} c(t, j)$ we need $(t, j) \in E$, and we also have an order on E, where we have that $(t, j) > (\widehat{t}, \widehat{j})$ if and only if $t + j > \widehat{t} + \widehat{j}$, where $t, \widehat{t} \in \mathbb{R}$ and $j, \widehat{j} \in \mathbb{N}$, so $t + j, \widehat{t} + \widehat{j} \in \mathbb{R}$. Hence we can import the concept of a of limit with real numbers $\lim_{r\to\infty} f(r)$ into our setup. That is, to calculate something like $\lim_{(t,j)\to\infty} c(t, j)$, it requires the existence of a sequence $(t_k, j_k) \in E$ such that $t_k + j_k \to \infty$ as $k \to \infty$. Such a sequence does not necessarily exist, unless the total number of jumps $N = \infty$ or the final time $t_{N+1} = \infty$. If this is the case, we will call E *positively unbounded* and c *positively complete*.

Definition 4.3.2 (Complete solution)**.** A solution $c : E \to M$ of a hybrid dynamical system is called *positively complete* if and only if E is *positively unbounded* if and only if $N = \infty$ or if $t_{N+1} = \infty$.

Remark. Here N and t_{N+1} are as in the definition of hybrid time domain 4.1.1.

To finally calculate the limit, we note that the values $c(t, j) \in M$ lie in the manifold M, so we will use the standard notion of limit for topological spaces together with the standard notion of $r \to \infty$ for real numbers.

Definition 4.3.3 (Ultimately in a set)**.** Let $c : E \to M$ be a solution to a hybrid dynamical system and $U \subseteq L$. We say that c is *ultimately in* U if there exists a number T_U such that $c(t, j) \in U$ for every $(t, j) \in E$ with $t + j > T_U$.

Definition 4.3.4 (Limit of solution)**.** Let $p^* \in M$ and let $c : E \to M$ be a positively complete solution to a hybrid dynamical system with M as its state space. We say that c *converges* to p^* if c is ultimately in U for every open set $U \subseteq M$ with $p^* \in U$. If this is the case we may write

$$\lim_{(t,j)\to\infty} c(t,j) = p^*,$$

and say that p^* is the *limit* of c.

Now say that $c : E \to M$ is a given solution to a hybrid dynamical system $\mathcal{H}$. Clearly, keeping only a part $\widehat{E} \subseteq E$ of the hybrid time domain (where $\widehat{E}$ is itself a hybrid time domain) results in another solution $\widehat{c} = c|_{\widehat{E}} : \widehat{E} \to M$ to the same hybrid dynamical system $\mathcal{H}$. Therefore, when studying asymptotic behavior (where we want as large of a hybrid time domain as possible) we can without loss of generality study *maximal* solutions, which are those solutions $c_{\max} : E_{\max} \to C$ which is not the restriction of any other solution. That is, there does *not* exist a solution $c : E \to M$ with $E_{\max} \subset E$.

Definition 4.3.5 (Maximal solution)**.** A *maximal* solution $c : E \to M$ to a hybrid dynamical system $\mathcal{H}$ is a solution to $\mathcal{H}$ where there does *not* exist another solution $c_{\text{larger}} : E_{\text{larger}} \to C$ such that $E \subset E_{\text{larger}}$.

Now we are ready for the main definitions of this section. Firstly is the idea of the *positively invariant set* of a point $p^* \in C$ for a hybrid dynamical system. This will be the set of points $c_0 \in C$ such that if a solution passes through that point $c(\widehat{t},\widehat{j}) = c_0$ it necessarily converges to p^*. Without loss of generality we can always assume that a given solution is maximal, but to be able to talk about the limit of a solution we need the solution to be positively complete. This leads to the following definition.

Definition 4.3.6 (Positively invariant set)**.** The *positively invariant set* $W^+_{\mathcal{H}}(p^*)$ of a point $p^* \in M$ for a hybrid dynamical system $\mathcal{H}$ with state space M is the set of points $c_0 \in M$ such that any maximal solution $c : E \to M$ with $c(\widehat{t},\widehat{j}) = c_0$ for some $(\widehat{t},\widehat{j}) \in E$ is positively complete and converges to p^*.

Remark. We could also define the *negatively invariant* $W^-_{\mathcal{H}}(p^*)$ to be the set of points $c_0 \in C$ such that any maximal solution $c : E \to C$ with $c(\widehat{t},\widehat{j}) = c_0$ for some $(\widehat{t},\widehat{j}) \in E$ is *negatively complete* and *diverges from* p^*, where *negatively complete* means that the starting time is $t_0 = -\infty$, and that c diverges from p^* means that $\lim_{(t,j)\to-\infty} c(t,j) = p^*$, defined correspondingly.

Now we will define what it means to be an *attractive* point for a hybrid dynamical system. This is a point p^* which has a large positively invariant set $W^+_{\mathcal{H}}(p^*)$, where by large we mean that it contains an open set $U \subseteq W^+_{\mathcal{H}}(p^*)$ with p^* in it.

Definition 4.3.7 (Attractive point)**.** A point $p^* \in M$ is *attractive* for a hybrid dynamical system $\mathcal{H}$ with state space M if there exists an open set $U \subseteq M$ with $p^* \in U$ such that $U \subseteq W^+_{\mathcal{H}}(p^*)$.

Definition 4.3.8 (Basin of attraction)**.** If p^* is an attractive point for $\mathcal{H}$, then we will call its positively invariant set $W^+_{\mathcal{H}}(p^*)$ the *basin of attraction* of p^*.

Definition 4.3.9 (Globally attractive point)**.** A point $p^* \in M$ is *globally attractive* for a hybrid dynamical system $\mathcal{H}$ with state space M if p^* is an attractive point for $\mathcal{H}$ and its basin of attraction is the whole state space $W^+_{\mathcal{H}}(p^*) = M$.

Definition 4.3.10 (Asymptotically stable point)**.** A point $p^* \in M$ is *asymptotically stable* for a hybrid dynamical system $\mathcal{H}$ with state space M if it is both stable and attractive for $\mathcal{H}$.

In some sense, this is what we want for a specified point p^*, where this means that if we are close enough to p^* then we will stay close to p^* and moreover we will converge to p^*. However ideally we want this to be true no matter how far away from p^* we are, which will be known as *global asymptotic stability* of this point p^*.

Definition 4.3.11 (Globally asymptotically stable point)**.** A point $p^* \in M$ is *globally asymptotically stable* for a hybrid dynamical system $\mathcal{H}$ with state space M if it is both stable and globally attractive for $\mathcal{H}$.

Given a hybrid plant $\mathcal{H}_P$ with state space M and a desired state $z^* \in M$, a goal is then to find a hybrid controller $\mathcal{H}_K$ with state space N and a desired state $\eta^* \in N$ such that the interconnection $\mathcal{H}$ of the hybrid plant and hybrid controller has $p^* = (z^*, \eta^*)$ as a globally asymptotically stable point.

4.3.2 Robustness

This report follows the article [MS24], which do not develop a well-defined framework for robustness, and it is therefore not fully rigorously proven in the article. It turns out that there are some complications to generalizing robustness to manifolds. We will discuss this here, and simply cite [MS24] for the result about robustness. The idea is as follows.

Suppose we have found a hybrid controller such that the interconnection $\mathcal{H} = (C, F, D, G)$ globally asymptotically stabilizes a desired point $p^* \in M \times N$. In practice, $\mathcal{H}$ will not be realized exactly, since the measurement function $h : M \to \mathbb{R}^l$ is subject to sensor noise, and the flow function F is subject to actuator noise, environmental disturbances, and modelling error. A *robust* controller is one where if the system is perturbed, it retains the desired properties, or at least an approximation to them. A way to model this could be to consider perturbed functions $h + e_h : M \to \mathbb{R}^l$ instead of h and $F + e_F : N \times M \to TM \oplus TN$, however then the errors e_h and e_F could not be time dependent if we still want to end up with a well-defined hybrid dynamical system. This does not mirror reality well, and hence we require a more general setup.

A solution is to use a definition inspired by Definition 6.27 of [GST12], which defines a *perturbed* hybrid system $\mathcal{H}^\varepsilon = (C^\varepsilon, F^\varepsilon, D^\varepsilon, G^\varepsilon)$, which will capture the perturbations mentioned, and is still a hybrid dynamical system. The perturbed hybrid system is based upon inflating C, F, D and G by an ε ball, so to inflate F we need a norm on the tangent space, which can obtain by equipping $M \times N$ with a Riemannian metric. The canonical choice would be to equip M and N with a Riemannian metric each $\mathcal{G}^M$, $\mathcal{G}^N$, and let $\mathcal{G} = \mathcal{G}^M + \mathcal{G}^N$. Here M and N are smooth manifolds, so they can be equipped with Riemannian metrics by Lemma 2.8.2. To inflate C, D and G we need to have a distance on $M \times N$, which we obtain from the Riemannian metric $\mathcal{G}$.

In [GST12] the inflation is both more general, and larger in the case of F and G than we have mentioned here. Specifically, instead of $\varepsilon \geq 0$ we would have a function $\rho : L \to \mathbb{R}_{\geq 0}$ and the domain of F and G is first inflated by $\rho(x)$ and then the target of this inflated function is also inflated by $\rho(x)$. For F, the convex hull and closure is also applied before the final inflation. Without the closure of the convex hull we would have $G^\rho(x) = G(\{x\}^{\rho(x)} \cap D)^{\rho(x)}$, and $F^\rho(x) = \bigcup_{(y,v) \in F(\{x\}^{\rho(x)} \cap C)} \{y\} \times \{v\}^{\rho(x)}$. However, $\overline{\text{co}} F(\{x\}^{\rho(x)} \cap C)$ would take some care to define, since $F(\{x\}^{\rho(x)} \cap C) \subseteq TL$. There is a natural topology on TL, so taking the closure is no problem, however for the convex hull we would need to be able to add vectors with different base points. This can be done by introducing parallel transport, to transport the vectors to the same base point, where they can be added.

Chapter 5

Applying Hybrid Feedback Control to Compact Manifolds

5.1 Designing a hybrid controller

The following section is based upon the article [MS24], where the scheme described has been expressed with the rigorous definitions that we developed in the previous chapter.

5.1.1 Main theorem

Theorem 2 (Robust globally asymptotic controller)**.** *Let M be a compact connected smooth manifold, and let $p^* \in M$. There exists smooth vector fields $F_0, F_1 : M \to TM$, a measurement function $h : M \to \mathbb{R}^2$, and a hybrid controller $\mathcal{H}_K$ with state space $N = \{0, 1\}$, such that the interconnection of the hybrid plant*

$$\mathcal{H}_P : \begin{cases} \dot{z} = (1-q)F_0(z) + qF_1(z), & (z,q) \in C_P \subset M \times \{0,1\} \\ y = h(z) \end{cases}$$

with $\mathcal{H}_K$ has $(p^*, 0)$ as a globally asymptotically stable point. Moreover, the global asymptotic stability is robust.

The idea to prove the theorem is as follows. We will let $f : M \to \mathbb{R}$ be a Morse function with p^* as its unique local and global minimum, where we will decrease f towards its global minimum using gradient flow $\dot{z} = (-\operatorname{grad}_G f) \circ z$. However, gradient flow might instead converge to another critical point $p_i \neq p^*$ of f (a saddle point). We can prevent this by first measuring if we are close to a critical point by measuring $\| \operatorname{grad}_G f(z) \|$, which is close to 0 if and only if z is close to a critical point of f. In this case we switch z to instead follow a *breeze vector field* Y which will move z away from p_i. It is now a problem if we switch away from gradient flow when z becomes close to p^*, but we can ensure that z is not close to p^* before switching by also measuring $f(z)$. To make this scheme robust we need some margin for when we switch between gradient flow and the breeze vector field.

5.1.2 Proof step 1: Steady breeze

Definition 5.1.1 (Breeze vector field)**.** Let $f : M \to \mathbb{R}$ be a Morse function, and denote its critical points that are not local minima by p_i, $i \in I$. A smooth vector field $Y : M \to TM$ is called a *breeze vector field* for f if there exists Morse coordinate charts (U_i, φ_i) for f with $p_i \in U_i$ and $\varphi_i = (x_{i,1}, \ldots, x_{i,l_i}, y_{i,1}, \ldots, y_{i,k_i})$ such that $Y|_{U_i} = \frac{\partial}{\partial y_{i,1}}$ for every $i \in I$.

Lemma 5.1.1 (Existence of breeze vector fields)**.** *For any Morse function $f : M \to \mathbb{R}$ on a compact smooth manifold M, there exists a breeze vector field.*

Proof. By Lemma 3.4.2, compactness of M implies that f has finitely many critical points, and hence finitely many that are not local minima. Denote these by $p_1, \dots, p_N$.

For each p_i, by the Morse lemma there exists a Morse coordinate chart (W_i, φ_i), with $p_i \in W_i$ and $\varphi_i = (x_{i,1}, \dots, x_{i,l_i}, y_{i,1}, \dots, y_{i,k_i})$ such that $f|_{W_i} = f(p_i) + \|x_i\|^2 - \|y_i\|^2$, where $k_i \geq 0$ is the index of p_i and $x_i = (x_{i,1}, \dots, x_{i,l_i})$, $y_i = (y_{i,1}, \dots, y_{i,k_i})$. Since p_i is not a local minimum $k_i \geq 1$ by Lemma 3.3.2, so $y_{i,1}$ exists. Now we define the smooth vector field $Y_i = \frac{\partial}{\partial y_{i,1}}$ on W_i. By Lemma 2.4.1 we can restrict our coordinate chart to arbitrary open subsets of W_i, which we note are still Morse coordinate charts, and by Lemma 3.4.1, the $p_i \notin W_j$ for $i \neq j$, so we can shrink the W_i to make them be pairwise disjoint. Explicitly, since M is Hausdorff for each $i \neq j$ there exists open sets $V_{i,j}$ and $V_{j,i}$ with $p_i \in V_{i,j}$ and $V_{i,j} \cap V_{j,i} = \varnothing$, so we may replace W_i with $W_i' = W_i \cap \bigcap_{j \neq i} V_{i,j}$, which is open because its a finite intersection of open sets. Now $p_i \in W_i'$ and the W_i' are disjoint since for $i \neq j$ we have $W_i' \subseteq V_{i,j}$ and $W_j' \subseteq V_{j,i}$, so $V_{i,j} \cap V_{j,i} = \varnothing$ means $W_i' \cap W_j' = \varnothing$.

By Lemma 2.1.6, $M \setminus \{p_1, \dots, p_N\}$ is open, so $\{W_1, \dots, W_N, M \setminus \{p_1, \dots, p_N\}\}$ is an open cover of M. By Lemma 2.8.1, there exists a partition of unity $\{\chi_1, \dots, \chi_N, \chi_{N+1}\}$ subordinate to this cover. Hence we may define

$$Y = \sum_{i=1}^{N} \chi_i Y_i \tag{5.1}$$

as a well-defined smooth vector field on M. The claim is now that there exists open sets $U_i \subseteq W_i$ with $p_i \in U_i$ such that $Y|_{U_i} = Y_i|_{U_i}$. Indeed, since the W_j are pairwise disjoint and $\operatorname{supp}(\chi_j) \subset W_j$, we have $\chi_j|_{W_i} = 0$ for $j \neq i$. Moreover $\operatorname{supp}(\chi_{N+1}) \subset M \setminus \{p_1, \dots, p_N\}$, and since the support is closed, $M \setminus \operatorname{supp}(\chi_{N+1})$ is open. Because $p_i \notin \operatorname{supp}(\chi_{N+1})$, $p_i \in U_i = W_i \cap (M \setminus \operatorname{supp}(\chi_{N+1}))$, which is an open set because it is an intersection of two open sets, and $\chi_{N+1}|_{U_i} = 0$. Therefore, $\chi_j|_{U_i} = 0$ for $i \neq j$, and by the definition of partition of unity $\sum_{i=1}^{N+1} \chi_i = 1$, so $\sum_{j=1}^{N+1} \chi_j|_{U_i} = \chi_i|_{U_i} = 1$. This gives

$$Y|_{U_i} = \sum_{j=1}^{N} \chi_j|_{U_i} Y_j|_{U_i} = Y_i|_{U_i} = \frac{\partial}{\partial y_{i,1}}\bigg|_{U_i} = \frac{\partial}{\partial (y_{i,1}|_{U_i})}, \tag{5.2}$$

where again by Lemma 2.4.1 $(U_i, \varphi_i|_{U_i})$ is a coordinate chart. It is a Morse coordinate chart, with $y_{i,1}|_{U_i}$ being in the same coordinate as $y_{i,1}$, since the local expression in this coordinate chart is the same as with φ_i, just restricted. □

Lemma 5.1.2 (Breeze coordinate charts)**.** *Let M be a compact smooth manifold, $f : M \to \mathbb{R}$ a Morse function, and denote the critical points of f that are not local minima by p_i, $1 \leq i \leq N$, with index $k_i \geq 1$. Then there exists a constant $A > 0$ such that for each p_i there exists Morse coordinate charts $(V_{i,A}, \varphi_i)$ such that*

$$V_{i,A} = \varphi_i^{-1}\{(\alpha, \beta) \in \mathbb{R}^{l_i + k_i} : |\beta_1| < \sqrt{2A},\ \|\alpha\|^2 + \beta_2^2 + \cdots + \beta_{k_i}^2 < A\}$$

(here $l_i + k_i = \dim(M)$) their closures $\overline{V}_{i,A}$ are pairwise disjoint, and no local minimum of f is within $\overline{V}_{i,A}$ for any i.

Proof. Firstly, $1 \leq i \leq N$ by Lemma 3.4.2 and Lemma 3.4.6 since M is compact. Moreover, since f is Morse, the critical points of f are isolated by Lemma 3.4.1 (there exists disjoint open $W_i \ni p_i$), so the Morse coordinate charts (U_i, ψ_i) given by the Morse lemma can be shrunk to be disjoint $(U_i \cap W_i, \psi_i|_{U_i \cap W_i})$ by Lemma 2.4.1. Moreover, by Lemma 3.4.1, then no local minimum of f is within any W_i.

Because ψ_i is a homeomorphism, by Lemma 2.2.7 any restriction of ψ_i is a homeomorphism onto its image, so since $\psi_i(p_i) = 0$ and $0 \in E_{i,A} = \{(\alpha, \beta) \in \mathbb{R}^{l_i+k_i} : |\beta_1| \leq \sqrt{2A},\ \|\alpha\|^2 + \beta_2^2 + \cdots + \beta_{k_i}^2 \leq A\}$ for any $A > 0$ and $E_{i,A}$ is inside any open subset containing 0 for A small enough, then for A small enough $\psi_i^{-1}(E_{i,A}) = \overline{V}_{i,A} \subset U_i \cap W_i$. Therefore $V_{i,A} \subset U_i \cap W_i$, which is open since ψ_i is continuous, which then makes $(V_{i,A}, \psi_i|_{V_{i,A}})$ Morse coordinate charts, and the $\overline{V}_{i,A}$ are pairwise disjoint and do not contain any local minimum of f since the $U_i \cap W_i$ are pairwise disjoint and do not contain any local minimum of f. □

Definition 5.1.2 (Breeze sets)**.** Charts $(V_{i,A}, \varphi_i)$ fulfilling the conclusion of Lemma 5.1.2 will be called *breeze coordinate charts* (and $V_{i,A}$ *breeze sets*) for f with parameter A.

Lemma 5.1.3 (Bound in breeze sets)**.** *Let $z \in V_{i,A}$. Then $f(z) > f(p_i) - 3A$.*

Proof. Since $(V_{i,A}, \varphi_i)$ are Morse coordinate charts, if we let $\varphi_i(z) = (\alpha, \beta)$ the Morse lemma says that $f(z) = f(p_i) + \|\alpha\|^2 - \|\beta\|^2 \geq f(p_i) - \|\beta\|^2 = f(p_i) - \beta_1^2 - (\beta_2^2 + \cdots + \beta_{k_i}^2) > f(p_i) - \sqrt{2A}^2 - A = f(p_i) - 3A$, by the definition of $V_{i,A}$. □

Lemma 5.1.4 (Steady breeze)**.** *Let M be a compact smooth manifold, $f : M \to \mathbb{R}$ a Morse function, and Y a breeze vector field for f with Morse coordinate charts that are breeze coordinate charts $(V_{i,A}, \varphi_i)$. Every solution c_{z_0} of $\dot{c}_{z_0} = Y \circ c_{z_0}$ with $c_{z_0}(0) = z_0 \in V_{i,A}$ first exits $V_{i,A}$ at $q_{z_0,i} = c_{z_0}(T_{z_0,i}) \in \overline{V}_{i,A}$ in finite time $T_{z_0,i} = \inf\{t > 0 : c_{z_0}(t) \notin V_{i,A}\}$ with*

$$f(c_{z_0}(T_{z_0,i})) < f(p_i) - A.$$

Proof. Since Y is a breeze vector field for f, then $p_i \in V_{i,A}$, $\varphi_i = (x_{i,1}, \ldots, x_{i,l_i}, y_{i,1}, \ldots, y_{i,k_i})$, and $Y|_{V_{i,A}} = \frac{\partial}{\partial y_{i,1}}$. Because M is compact, by Lemma 2.6.1 there exists a unique solution $c_{z_0} : \mathbb{R} \to M$ to $\dot{c}_{z_0} = Y \circ c_{z_0}$ with initial point $c_{z_0}(0) = z_0$. Now, $Y|_{V_{i,A}} = \frac{\partial}{\partial y_{i,1}}$ means that for $c_{z_0}(t) \in V_{i,A}$

$$c_{z_0}(t) = \varphi_i^{-1}(a_1, \ldots, a_{l_i}, b_1 + t, b_2, \ldots, b_{k_i}) \tag{5.3}$$

if we let $\varphi_i(z_0) = (a_1, \ldots, a_{l_i}, b_1, \ldots, b_{k_i}) \in \mathbb{R}^m$. By Equation 5.3 and the definition of breeze coordinate chart, we see that for $z_0 \in V_{i,A}$ then c_{z_0} will exit $V_{i,A}$ at $\beta_1 = \sqrt{2A}$ or $\beta_1 = -\sqrt{2A}$, and so for $t > 0$, c_{z_0} exits $V_{i,A}$ when $b_1 + t = \sqrt{2A}$, that is $T_{z_0,i} = \sqrt{2A} - b_1 < 2\sqrt{2A} < \infty$.

At the exit point $q_{z_0,i} = c_{z_0}(T_{z_0,i})$, and $y_{i,1}(q_{z_0,i}) = \sqrt{2A}$, and since the other coordinates have been constant they still satisfy

$$x_{i,1}(q_{z_0,i})^2 + \cdots + x_{i,l_i}(q_{z_0,i})^2 + y_{i,2}(q_{z_0,i})^2 + \cdots + y_{i,k_i}(q_{z_0,i})^2 < A. \tag{5.4}$$

Therefore, by the Morse lemma

$$\begin{aligned} f(q_{z_0,i}) &= f(p_i) + x_{i,1}(q_{z_0,i})^2 + \cdots + x_{i,l_i}(q_{z_0,i})^2 - 2A - y_{i,2}(q_{z_0,i})^2 - \cdots - y_{i,k_i}(q_{z_0,i})^2 \\ &\leq f(p_i) + x_{i,1}(q_{z_0,i})^2 + \cdots + x_{i,l_i}(q_{z_0,i})^2 - 2A \\ &= f(p_i) + x_{i,1}(q_{z_0,i})^2 + \cdots + x_{i,l_i}(q_{z_0,i})^2 - 2A + y_{i,2}(q_{z_0,i})^2 + \cdots + y_{i,k_i}(q_{z_0,i})^2 \\ &< f(p_i) + A - 2A \\ &= f(p_i) - A. \end{aligned} \tag{5.5}$$

□

5.1.3 Proof step 2: Detecting breeze sets with margin for robustness

The goal now is to construct sets $B_i \subset V_{i,A}$ that we can detect using our measurement $h(z) = (\|(\mathrm{grad}_G f)_z\|_G, f(z))$. Since $(\mathrm{grad}_G f)_z = 0$ if and only if z is a critical point of f by Lemma

3.1.1, then $\{z \in M : \|(\operatorname{grad}_G f)_z\|_G < r\}$ is a union of open sets containing the critical points of f that can be made arbitrarily small, depending on r. If we make $f(p^*)$ be the smallest critical value and sure that the critical values of f have some positive distance to the critical value $f(p^*)$, then we can use the measurement $f(z)$ to determine if we are close to p^* or not.

It turns out that we can always find a Morse function f that has p^* as its global minimum, and all critical values are arbitrarily far apart, for example at least 1 apart. This also makes us be able to measure exactly which critical point we are closest to, not just if we are closest to p^* or not.

Definition 5.1.3 (Prepared Morse function)**.** Let M be a compact smooth manifold, and $f : M \to \mathbb{R}$ a Morse function with a unique local minimum p^* and all critical values distinct. Let the critical points that are not p^* be denoted by $p_1, \dots, p_N$. We will call f *prepared* if additionally

- $f(p^*) = 0$ and $f(q) > 0$ for $q \neq p^*$
- $c_i = f(p_i) \geq 1$ for each $1 \leq i \leq N$
- $|c_i - c_j| \geq 1$ for $i \neq j$.

Lemma 5.1.5 (Existence of prepared Morse functions)**.** *Every compact connected smooth manifold admits a prepared Morse function.*

Proof. By Corollary 3.3.4, connectedness implies that there exists a Morse function $f : M \to \mathbb{R}$ with a unique local minimum p^* such that $f(p^*) = 0$, and if $\{p_i : i \in I\}$ are the other critical points, then $f(p_i) = c_i \geq 1$ and $|c_i - c_j| \geq 1$ for $i \neq j$. Moreover, the number N is finite by Corollary 3.4.2, and p^* is a global minimum because M is compact, so f attains its global minimum at some point, which in particular is a local minimum, and since p^* is the unique local minimum it must be the global minimum. □

Lemma 5.1.6 (Detecting breeze sets)**.** *Let (M, G) be a compact Riemannian manifold, $f : M \to \mathbb{R}$ a prepared Morse function, and $V_{i,A}$, $1 \leq i \leq N$ breeze sets for f. Then there there exists a constant $r_A > 0$ which can be taken small enough such that the following holds.*

$$\Delta_{r_A} = \{z \in M \ : \ \|(\operatorname{grad}_G f)_z\|_G \leq r_A, \ f(z) \geq 1/2\} \quad \subseteq \quad \bigcup_{i=1}^{N} V_{i,A}.$$

Proof. Let $O_A = \bigcup_{i=1}^N V_{i,A}$. To show that for some $r > 0$ we have $\Delta_r \subseteq O_A$, we will show the equivalent statement $M \setminus O \subseteq M \setminus \Delta_r = \{z \in M \ : \ \|(\operatorname{grad}_G f)_z\|_G > r \text{ or } f(z) < 1/2\}$.

To show this, let $z \in M \setminus O_A$. We begin by showing the weaker claim that $\|(\operatorname{grad}_G f)_z\|_G > 0$ or $f(z) < 1/2$. Indeed, suppose $\|(\operatorname{grad}_G f)_z\|_G = 0$ and $f(z) \geq 1/2$. By Lemma 3.1.1, z is a critical point of f, so $z \in \{p^*, p_1, \dots, p_N\}$. Since f is prepared $f(p^*) = 0$, so $f(z) \geq 1/2$ means that $z = p_i$ for some i. But then $z \in V_{i,A}$, contradicting $z \in M \setminus O_A$.

We will now use compactness of M to show that the same constant $r > 0$ can be taken for every $z \in M \setminus O_A$, such that $f(z) < 1/2$ or $\|(\operatorname{grad}_G f)_z\|_G > r$. To do this, we need to construct an open cover. In the case $f(z) < 1/2$ we associate to z the number $r_z = 1$ (it can be any positive number) and the set $W_z = \{w \in M : f(w) < 1/2\}$ which clearly contains z, and is open because f is continuous. If however $f(z) \geq 1/2$ but $\|(\operatorname{grad}_G f)_z\|_G > 0$, then we let $r_z = \|(\operatorname{grad}_G f)_z\|_G/2 > 0$ and let $W_z = \{w \in M : \|(\operatorname{grad}_G f)_w\|_G > r_z\}$, which similarly contains z. Since $\operatorname{grad}_G f$ is smooth by Lemma 2.8.5, and G is smooth by definition of Riemannian metric, then $\sigma(w) = G_w((\operatorname{grad}_G f)_w, (\operatorname{grad}_G f)_w) = \|(\operatorname{grad}_G f)_w\|_G^2$ is smooth and hence continuous by Lemma 2.4.6. Therefore $W_z = \sigma^{-1}]r_z^2, \infty[$ is open in this case too.

In either case, for every $w \in W_z$, either $\|(\operatorname{grad}_G f)_w\|_G > r_z$ or $f(w) < 1/2$, so $w \in M \setminus \Delta_{r_z}$, and hence $w \in M \setminus \Delta_r$ for every $r \leq r_z$. The collection $\{W_z : z \in M \setminus O_A\} \cup \{O_A\}$ is an open cover of M, so by compactness of M, there is a finite subcover, which may or may not include O_A. Since $O \neq M$, at least one of the W_z must be included in the finite subcover. Let $\{W_{z_1}, \ldots, W_{z_k}\}$ be the sets in the finite subcover which is not O, so that $M \setminus O \subseteq W_{z_1} \cup \cdots \cup W_{z_k}$. Now let $r_A = \min\{r_{z_1}, \ldots, r_{z_k}\} > 0$.

Finally, if $w \in M \setminus O$, then $w \in W_{z_j}$ for some $1 \leq j \leq k$, so either $\|(\operatorname{grad}_G f)_w\|_G > r_{z_j} \geq r_A$ or $f(w) < 1/2$, meaning $w \in M \setminus \Delta_{r_A}$. Hence $\Delta_{r_A} \subseteq O_A$. □

Definition 5.1.4 (Switch sets)**.** Given the setup of Lemma 5.1.6, define the *switch sets* of f to be

$$B_{i,A} = \Delta_{r_A} \cap V_{i,A} = \{z \in V_{i,A} \ : \ \|(\operatorname{grad}_G f)_z\|_G \leq r_A, \ f(z) \geq 1/2\},$$

for $1 \leq i \leq N$, and $\Delta_{r_A} = \bigcup_{i=1}^N B_{i,A}$ with the $B_{i;A}$ pairwise disjoint.

5.1.4 Proof step 3: Defining the hybrid systems

Now we fix (M, G) to be a compact connected Riemannian manifold, $f : M \to \mathbb{R}$ to be a prepared Morse function, and $0 < A \leq 1/2$.

($0 < A \leq 1/2$ since $|c_i - c_j| \geq a$ and we need $2A < a$. and $c_i \geq 1 + c_0$, $c_0 = 0$ and we need some number $c_0 < R < c_i$, and $b \geq R$, so we take $b \geq 1/2$)

The hybrid systems will be as in Example 31, with $F_0 = -\operatorname{grad}_G f$, $F_1 = Y$, measurement function $h : M \to \mathbb{R}^2$, $z \mapsto (\|(\operatorname{grad}_G f)_z\|_G, f(z))$, and $P_1 = \bigcup_{i=1}^N V_{i,A}$. We will let $\widetilde{D}_0 = h^{-1}(D_0) = \bigcup_{i=1}^N B_{i,A}$, choosing $D_0 = \{(a, b) \in \mathbb{R}^2 : a \leq r_A \text{ and } b \geq 1/2\}$, $C_0 = M \setminus h^{-1}(D_0)$ and $D_1 = h(M \setminus P_1)$, $\widetilde{D}_1 = M \setminus P_1$, so

$$C_P = \left(M \setminus \bigcup_{i=1}^N B_{i,A}\right) \times \{0\} \cup \left(\bigcup_{i=1}^N V_{i,A}\right) \times \{1\} \quad = \quad C_0 \times \{0\} \cup P_1 \times \{1\} \tag{5.6}$$

and

$$H_P : \begin{cases} \dot{z} = (1-q)F_0(z) + qF_1(z), & (z, q) \in C_P \\ y = (\|(\operatorname{grad}_G f)_z\|_G, \ f(z)). \end{cases} \tag{5.7}$$

Similarly

$$D_K = \{0\} \times D_0 \cup \{1\} \times D_1 \tag{5.8}$$

and

$$H_K : \begin{cases} q^+ = 1 - q, & (q, y) \in D_K \\ \zeta = q, \end{cases} \tag{5.9}$$

so, as in Example 31, $C = C_P$ and

$$D = \left(\bigcup_{i=1}^N B_{i,A}\right) \times \{0\} \cup \left(M \setminus \bigcup_{i=1}^N V_{i,A}\right) \times \{1\} \quad = \quad \widetilde{D}_0 \times \{0\} \cup \widetilde{D}_1 \times \{1\}. \tag{5.10}$$

Therefore

$$\mathcal{H} : \begin{cases} \dot{z} = (1-q)F_0(z) + qF_1(z) & (z, q) \in C \\ q^+ = 1 - q & (z, q) \in D \end{cases} \tag{5.11}$$

or informally

$$\mathcal{H} : \begin{cases} \dot{z} = (1-q)F_0(z) + qF_1(z) & z \in C_q \\ q^+ = 1 - q & h(z) \in D_q. \end{cases} \tag{5.12}$$

Lemma 5.1.7 (Hysteresis)**.** *$\widetilde{D}_0 \subset P_1$ and $\widetilde{D}_1 \subset C_0$.*

Proof. By Definition 5.1.4 and Lemma 5.1.6 we have $h^{-1}(D_0) \subset P_1$, so by taking complements we obtain $M \setminus P_1 \subset M \setminus h^{-1}(D_0)$, that is $h^{-1}(D_1) \subset C_0$. □

5.1.5 Proof step 4: Desired properties

Lemma 5.1.8 (Finite jumps)**.** *Every maximal solution to $\mathcal{H}$ has at most $2N+1$ jumps.*

Proof. Let $(z, q) : E \to M \times \{0, 1\}$ be a maximal solution to $\mathcal{H}$. Since hybrid dynamical systems are time-invariant (in the continuous time variable), we can assume that $(0, 0) \in E$. We begin with the case $q(0, 0) = 0$.

If a jump occurs, the first jump will then be when $(z, q) \in \widetilde{D}_0 \times \{0\}$, where now $q = 1$ and $z \in V_{i,A}$ for some i by the hysteresis lemma, meaning that z will flow along Y. By the steady breeze lemma, z exits $V_{i,A}$ in finite time, and then $f(u) < c_i - A$ at the exit point $u \in \overline{V}_{i,A} \setminus \bigcup_{j=1}^{N} V_{j,A} \subseteq \widetilde{D}_1$. Since $(u, q) \in \widetilde{D}_1 \times \{1\}$ we get an immediate jump back to $q = 0$, where by the hysteresis lemma $\widetilde{D}_1 \subset C_0$, so $(u, q) \in C_0 \times \{0\}$, with z now flowing along $-\operatorname{grad}_G f$.

By Lemma 3.1.3, either z will converge to p^* or eventually $z \in \widetilde{D}_0$, for which $(z, q) \in \widetilde{D}_0 \times \{0\}$ again, meaning that now $q = 1$ and $z \in V_{j,A}$ for some j by the hysteresis lemma. By Lemma 5.1.3 $f(z) > c_j - 3A$, but Since $f(u) < c_i - A$ and f has only decreased by Lemma 3.1.2, then $c_j - 3A < f(z) < c_i - A$, meaning that $c_j < c_i + 2A$. By way of contradiction, assume that $c_j > c_i$, so since f is prepared we have that $c_j \geq c_i + 1$, but then $A \leq 1/2$ contradicts this, since $c_i + 1 \leq c_j < c_i + 2A = c_i + 1$ is impossible. Therefore $z \in V_{j,A}$ for some $c_j < c_i$.

Since there is a finite number of critical values (that are not the global minimum) $1 \leq i \leq N$, and the argument above showed that two jumps will trigger every time $(z, q) \in \widetilde{D}_0 \times \{0\}$, this means that at most $2N$ jumps will occur.

It remains to handle the case $q(0, 0) = 1$. If $z(0, 0) \in P_1$, then $z(0, 0) \in V_{i,A}$ for some i, where as in the analysis above, the steady breeze lemma applies and the state exits at a point $u \in \widetilde{D}_1$, where now a jump occurs $q = 0$ and $u \in V_0$ by the hysteresis lemma. As in the argument above, if another jump happens $z \in \widetilde{D}_0$ then $z \in V_{j,A}$ for some $c_j < c_i$. Hence in there are still only at most $2N$ jumps.

If $z(0, 0) \in \widetilde{D}_1$, an immediate jump occurs $q(0, 1) = 0$, after which we are in the $q = 0$ situation analyzed above, contributing at most $2N$ more jumps, for a total of $2N+1$ jumps. □

Lemma 5.1.9 (Positive completeness)**.** *Every maximal solution to $\mathcal{H}$ is positively complete.*

Proof. Since there is a finite number of jumps by Lemma 5.1.8, continuous time part of the flow time set will be a finite union of intervals, for which the flow exists for all time in each interval because M is compact, by Lemma 2.6.1. Hence the last interval must stretch to infinity, meaning that the solution is positively complete. □

Lemma 5.1.10 (Global attractivity)**.** *Every maximal solution of $\mathcal{H}$ converges to $(p^*, 0)$.*

Proof. Let $(z, q) : E \to M \times \{0, 1\}$ be a maximal solution with last jump at (t^*, j^*). As seen in the proof of Lemma 5.1.8, necessarily $q(t^*, j^*) = 0$, and since this was the last jump $q(t, j^*) = 0$ and $z(t, j^*) \in C_0$ for every $t \geq t^*$. Let now $j = j^*$ and $t \geq t^*$. Firstly $z \in C_0$ means that $z \notin \widetilde{D}_0$, so $\|(\operatorname{grad}_G f)_z\|_G > r_\rho$ or $f(z) < 1/2$. By way of contradiction, assume that $\|(\operatorname{grad}_G f)_z\| > r_\rho$ holds for all $t \geq t^*$ then by Lemma 3.1.1 there is some bounded from below distance to the critical points of f, where then Lemma 3.1.2 means that $f(z(-, j^*))'(t) \leq -\delta$ for some $\delta > 0$, since $f(z(-, j^*))$ is smooth by Definition 4.1.6 and hence continuous by Lemma 2.4.6, and hence in finite time $f(z) < 0$, contradicting that 0 is the global minimum value of f. Hence, for some

$T \geq t^*$ we have $f(z) < 1/2$. Since $c_i \geq 1$, the set critical points in the set $\{p : f(p) < 1/2\}$ contains only p^*, so by Lemma 3.1.3, z converges to p^*, and hence (z, q) converges to $(p^*, 0)$. □

Lemma 5.1.11 (Stability)**.** *The point $(p^*, 0)$ is stable for $\mathcal{H}$.*

Proof. Let $U \subseteq M \times \{0, 1\}$ be an open set with $(p^*, 0) \in U$ and choose $0 < \varepsilon < 1/2$ small enough such that $\widehat{U} = \{p : f(p) < \varepsilon\} \times \{0\} \subseteq U$, which we can do since $f(p^*) = 0$, $f(p) \geq 0$, and f is continuous. Now let $c = (z, q) : E \to M \times \{0, 1\}$ be a solution to $\mathcal{H}$ with some $(\widehat{t}, \widehat{j}) \in E$ such that $c(\widehat{t}, \widehat{j}) \in \widehat{U}$. This means that $q(\widehat{t}, \widehat{j}) = 0$ and $f(z(\widehat{t}, \widehat{j})) < 1/2$, so $z(\widehat{t}, \widehat{j}) \notin \widetilde{D}_0$, that is $z(\widehat{t}, \widehat{j}) \in C_0$. For $(t, j) \geq (\widehat{t}, \widehat{j})$ we have that z flows along gradient flow, so by Lemma 3.1.2, $f(z) < \varepsilon$ and so $z \in C_0$, meaning that there will be no more jump, so $q = 0$, and therefore $c = (z, q) \in \widehat{U} \subseteq U$. □

Lemma 5.1.12 (Global asymptotic stability)**.** *The point $(p^*, 0)$ is globally asymptotically stable for $\mathcal{H}$.*

Proof. This is simply Lemma 5.1.10 and Lemma 5.1.11. □

Lemma 5.1.13 (Robustness)**.** *The global asymptotic stability of $(p^*, 0)$ for the interconnection $\mathcal{H}$ is robust.*

Proof. By the hysteresis lemma there is some positive distance between the boundary of $B_{i,A} \subset V_{i,A}$ and the boundary of $V_{i,A}$. If we now let C_i be the minimum value of $\| \operatorname{grad}_G f \|$, then by the definition of $B_{i,A}$ we have $C_i > r_A$. Therefore, we have a margin of robustness $\frac{1}{2} \min\{r_A, C_i - r_A\}$ for each $1 \leq i \leq N$, where an error of this size or smaller when measuring $\| \operatorname{grad}_G f \|$ still ensures that if we measure that we are inside $\bigcup_{i=1}^N B_{i,A}$ then we are inside $\bigcup_{i=1}^N V_{i,A}$ (hysteresis robust). Taking the minimum over all i gives us a margin of robustness for interconnection. Since N is finite we have a positive margin, and therefore the global asymptotic stability is robust. □

5.1.6 Having as few jumps as possible

By Lemma 3.4.6, $N \geq 1$, so for any Morse function, there is at least one jump which might be necessary. In applications, it could be that every jump (switching of vector fields) has some inherent cost, or that every jump could be a source for failure. Since the number of jumps is bounded by $2N + 1$ (or $2N$ if $q = 0$ initially) by Lemma 5.1.8, an idea could be to make N as small as possible, by choosing a fitting Morse function.

The smallest number of critical points $N+1$ a Morse function can have on a compact smooth manifold M is called the Morse number of M. For applications, we might therefore want to choose a prepared Morse function which achieves the Morse number, which can be done for $\mathrm{SO}(n)$ by Lemma 3.4.11 (and products of different $\mathrm{SO}(n_i)$ by Theorem 1).

Another way to lower the bound on the number of jumps is to ensure that the gradient flow is *Morse-Smale* [MS24, §6.3.6].

5.2 Application to Products of $\mathrm{SO}(2)$ and $\mathrm{SO}(3)$

5.2.1 Riemannian structure of $\mathrm{SO}(n)$

Lemma 5.2.1. *The tangent space $T_R\mathrm{SO}(n) = \{R\Omega : \Omega^\mathsf{T} = -\Omega\} = R \cdot \mathfrak{o}(n)$, where $\mathfrak{o}(n)$ is the set of skew-symmetric matrices.*

Proof. Let $\gamma :]-\varepsilon, \varepsilon[\to \mathrm{SO}(n)$ be a smooth curve with $\gamma(0) = R$. Since $\gamma(t) \in \mathrm{SO}(n)$, we have $\gamma(t)^\mathsf{T}\gamma(t) = I$. Then (37) of the matrix cookbook [PP12] implies that taking the derivative

yields $\gamma'(t)^\mathsf{T}\gamma(t) + \gamma(t)^\mathsf{T}\gamma'(t) = 0$. Now let $\gamma'(0) = A \in T_R\mathrm{SO}(n)$, so setting $t = 0$ means that $A^\mathsf{T}R + R^\mathsf{T}A = 0$. Let now $\Omega = R^\mathsf{T}A$, where then $\Omega^\mathsf{T} + \Omega = 0$ and $A = R\Omega$. Hence $T_R\mathrm{SO}(n) \subseteq \{R\Omega : \Omega^\mathsf{T} = -\Omega\}$.

For the other inclusion, let $\Omega \in \mathfrak{o}(n)$ and $R \in \mathrm{SO}(n)$. Then $\gamma(t) = Re^{t\Omega}$ satisfies $\gamma(0) = R$, and

$$\gamma(t)^\mathsf{T}\gamma(t) = R^\mathsf{T}(e^{t\Omega})^\mathsf{T}e^{t\Omega}R = R^\mathsf{T}e^{t\Omega^\mathsf{T}}e^{t\Omega}R = R^\mathsf{T}e^{-t\Omega}e^{t\Omega}R = R^\mathsf{T}R = I \tag{5.13}$$

by [Hal15, Proposition 2.3] and $\det(\gamma(t)) = \det(e^{t\Omega}) = e^{\mathrm{tr}(t\Omega)} = e^{0_n} = I$ by [Hal15, Proposition 2.3, Theorem 2.12], where $\Omega^\mathsf{T} = -\Omega$ forces $\mathrm{tr}(\Omega) = 0$. Therefore $\gamma(t) \in \mathrm{SO}(n)$, and $\gamma'(t) = R\Omega e^{t\Omega}$ by [Hal15, Proposition 2.4], and so $\gamma'(0) = R\Omega$, meaning that $\{R\Omega : \Omega^\mathsf{T} = -\Omega\} \subseteq T_R\mathrm{SO}(n)$, completing the proof. □

Lemma 5.2.2 (Frobenius inner product)**.** $(\mathrm{SO}(n), <-,->)$ *is a Riemannian manifold, where* $<-,->\in T^{(0,2)}\mathrm{SO}(n)$ *is the Frobenius inner product*

$$<-,->_R\colon T_R\mathrm{SO}(n) \times T_R\mathrm{SO}(n) \to \mathbb{R}, (A, B) \mapsto <A, B>_R = \mathrm{tr}(A^\mathsf{T}B).$$

Proof. For each $R \in \mathrm{SO}(2)$, $<-,->_R$ is an inner product by [RC12, Section 5.2]. Moreover $<-,->$ is a smooth, since for any chart (U, φ) we have

$$g_{\varphi_i,\varphi_j}(R) = \left\langle \frac{\partial}{\partial\varphi_i}|_R, \frac{\partial}{\partial\varphi_i}|_R \right\rangle_R = \mathrm{tr}\left(\frac{\partial}{\partial\varphi_i}|_R^\mathsf{T}\frac{\partial}{\partial\varphi_j}|_R\right) \tag{5.14}$$

where $R \mapsto \frac{\partial}{\partial\varphi_k}|_R$ is a matrix with entries that are smooth functions, and tr is a polynomial in the entries, meaning that $R \mapsto g_{\varphi_i,\varphi_j}(R)$ is smooth. □

From now on we will let $\mathrm{SO}(n)$ be equipped with the Frobenius inner product, and use it to compute the gradient of our Morse function.

Lemma 5.2.3 (Gradient of Morse function)**.** *The gradient of* $f : \mathrm{SO}(n) \to \mathbb{R}$, $R \mapsto -\mathrm{tr}(QR)$ *with respect to the Frobenius inner product is*

$$(\mathrm{grad}\, f)_R = R\,\mathrm{skew}(-R^\mathsf{T}Q)$$

where skew : $A \mapsto \frac{A-A^\mathsf{T}}{2}$ *is the* skew-symmetrization*. (Here Q is any $n \times n$ matrix.)*

Proof. We begin by computing D_Rf. Let γ be a curve in $\mathrm{SO}(n)$ with $\gamma(0) = R$ and $\gamma'(0) = V \in T_R\mathrm{SO}(n)$. Then

$$D_Rf(V) = (f \circ \gamma)'(0) = \mathrm{tr}(Q\gamma(-)')(0) = -\mathrm{tr}(Q\gamma'(0)) = -\mathrm{tr}(QV) \tag{5.15}$$

since $\frac{d}{dt}\mathrm{tr}(A(t)) = \frac{d}{dt}\sum_i A_{ii}(t) = \sum_i \frac{d}{dt}A_{ii}(t)$. The defining property of the gradient is that $<(\mathrm{grad}\, f)_R, V>_R = D_Rf(V)$, that is, if we let $X = (\mathrm{grad}\, f)_R \in T_R\mathrm{SO}(n)$, then

$$\mathrm{tr}(X^\mathsf{T}V) = -\mathrm{tr}(QV) \tag{5.16}$$

but $X \neq -Q^\mathsf{T}$ in general, since we need $X \in T_R\mathrm{SO}(n) = R \cdot \mathfrak{o}(n)$. The idea will be to project $-Q^\mathsf{T} \in \mathbb{R}^{n\times n}$ down to $R \cdot \mathfrak{o}(n)$.

We show this in a few steps. Firstly $\mathbb{R}^{n\times n}$ splits as an orthogonal direct sum with respect to the Frobenius inner product $\mathbb{R}^{n\times n} = \mathrm{Sym}(n) \oplus \mathfrak{o}(n)$, where $\mathrm{Sym}(n)$ denotes the symmetric $n \times n$ matrices. These are orthogonal, because for any $\Omega \in \mathfrak{o}(n), S \in \mathrm{Sym}(n)$, we have that

$$<\Omega, S> = \mathrm{tr}(\Omega^\mathsf{T}S) = -\mathrm{tr}(\Omega S) = -\mathrm{tr}(S\Omega) = -\mathrm{tr}(S^\mathsf{T}\Omega) = -<S, \Omega> = -<\Omega, S> \tag{5.17}$$

so $< \Omega, S >= 0$, and they make up all of $\mathbb{R}^{n\times n}$, since for any $A \in \mathbb{R}^{n\times n}$ we have $A = \frac{A-A^\mathsf{T}}{2} + \frac{A+A^\mathsf{T}}{2}$, where $\frac{A-A^\mathsf{T}}{2} \in \mathfrak{o}(n)$ and $\frac{A+A^\mathsf{T}}{2} \in \mathrm{Sym}(n)$. Therefore, $\mathrm{skew}(A) = \frac{A-A^\mathsf{T}}{2}$ is the orthogonal projection onto $\mathfrak{o}(n)$, meaning that for any $A \in \mathbb{R}^{n\times n}$ and $\Omega \in \mathfrak{o}(n)$ we have

$$< A, \Omega >= \left\langle \mathrm{skew}(A) + \frac{A+A^\mathsf{T}}{2}, \Omega \right\rangle =< \mathrm{skew}(A), \Omega > + \left\langle \frac{A+A^\mathsf{T}}{2}, \Omega \right\rangle =< \mathrm{skew}(A), \Omega > . \tag{5.18}$$

Hence, for any $R \in \mathrm{SO}(n)$ we have that $\mathbb{R}^{n\times n} = R\cdot\mathfrak{o}(n) \oplus R\cdot\mathrm{Sym}(n)$, since $A = R\frac{R^\mathsf{T}A-(R^\mathsf{T}A)^\mathsf{T}}{2} + R\frac{R^\mathsf{T}A+(R^\mathsf{T}A)^\mathsf{T}}{2}$, and $< R\Omega, RS >= \mathrm{tr}(\Omega^\mathsf{T}R^\mathsf{T}RS) = \mathrm{tr}(\Omega^\mathsf{T}S) =< \Omega, S >= 0$, meaning that $R\mathrm{skew}(R^\mathsf{T}A)$ is the orthogonal projection onto $R\cdot\mathfrak{o}(n)$, and therefore

$$< A, R\Omega >=< R\mathrm{skew}(R^\mathsf{T}A), R\Omega > . \tag{5.19}$$

Finally, since $V \in T_R\mathrm{SO}(n) = R\cdot\mathfrak{o}(n)$ we have that

$$-\mathrm{tr}(QV) =< -Q, V >=< R\mathrm{skew}(-R^\mathsf{T}Q), V > \tag{5.20}$$

and therefore $< X, V >_R =< R\mathrm{skew}(-R^\mathsf{T}Q), V >_R$ for every $V \in T_R\mathrm{SO}(n)$, where we can now conclude that

$$(\mathrm{grad}\, f)_R = X = R\mathrm{skew}(-R^\mathsf{T}Q). \tag{5.21}$$

□

Corollary 5.2.4 (Gradient of prepared Morse)**.** *The gradient of $f : \mathrm{SO}(n) \to \mathbb{R}$, $R \mapsto \mathrm{tr}(D) - \mathrm{tr}(D(P^*)^\mathsf{T}R)$ with respect to the Frobenius inner product is*

$$(\mathrm{grad}\, f)_R = R\mathrm{skew}(P^*DR),$$

where D is symmetric.

Proof. Letting $Q = D(P^*)^\mathsf{T}$, and noting that $\mathrm{tr}(D)$ when added to $-\mathrm{tr}(QR)$ would disappear in (5.15), and not affect any other computation, we obtain that

$$(\mathrm{grad}\, f)_R = R\mathrm{skew}(-R^\mathsf{T}D(P^*)^\mathsf{T}) = R\mathrm{skew}(-(P^*DR)^\mathsf{T}) = (\mathrm{grad}\, f)_R = R\mathrm{skew}(P^*DR) \tag{5.22}$$

since D symmetric, and $\mathrm{skew}(-A^\mathsf{T}) = \mathrm{skew}(A)$. □

Definition 5.2.1 (Standard basis of $\mathfrak{o}(n)$)**.** For $1 \le a < b \le n$, define the matrix $E_{ab} \in \mathfrak{o}(n)$ by

$$E_{ab} = e_a e_b^\mathsf{T} - e_b e_a^\mathsf{T},$$

where $e_a \in \mathbb{R}^n$ denotes the ath standard basis vector.

Remark. Here $(E_{ab})_{ab} = 1$, $(E_{ab})_{ba} = -1$ and all other entries are 0. These are clearly linearly independent, and since $1 \le a < b \le n$, there are $n(n-1)/2$ of them. The dimension $\dim(\mathfrak{o}(n)) = \dim(T_I\mathrm{SO}(n)) = \dim(\mathrm{SO}(n)) = n(n-1)/2$, so this forms a basis, meaning that every $\Omega \in \mathfrak{o}(n)$ can be written uniquely as $\Omega = \sum_{a<b}\omega_{ab}E_{ab}$ for coordinates $\omega_{ab} \in \mathbb{R}$.

Remark. We order the matrices E_{ab} lexicographically, meaning that we let

$$(E_{12}, E_{13}, \ldots, E_{1n}, E_{23}, \ldots, E_{2n}, \ldots, E_{(n-1)n})$$

be the order of the basis. For instance $\mathfrak{o}(2)$ has the basis

$$(E_{12}) = \left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right)$$

and $\mathfrak{o}(3)$ has the basis

$$(E_{12}, E_{13}, E_{23}) = \left(\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}\right).$$

5.2.2 Morse charts for SO(2) and SO(3)

To apply the hybrid controller of Theorem 2, we need explicit Morse coordinate charts for the critical points P_i of the prepared Morse function f. Since $\mathrm{SO}(n)$ is a "Lie group", the standard choice of chart is based upon the matrix exponential

$$\mathbb{R}^{\frac{n(n-1)}{2}} \cong T_I\mathrm{SO}(n) = \mathfrak{o}(n) \to \mathrm{SO}(n), \quad \Omega \mapsto P_i e^{\Omega} \tag{5.23}$$

which is a "local diffeomorphism", meaning that there exists open sets $0_n \in V_i \subseteq \mathfrak{o}(n)$ and $P_i \in U_i \subseteq \mathrm{SO}(n)$ such that $V_i \to U_i$, $\Omega \mapsto P_i e^{\Omega}$ is a diffeomorphism. Therefore, the inverse of this function will be a chart of $\mathrm{SO}(n)$, since $\mathfrak{o}(n) \cong \mathbb{R}^{\dim(\mathrm{SO}(n))}$.

To do computations with this chart we need to evaluate the matrix exponential, which is an infinite series. This is quite complicated for most n, but in the cases $n = 2$ and $n = 3$ there are simpler formulas. The case $n = 2$ coincides with the famous Euler's formula $e^{i\theta} = \cos(\theta) + i\sin(\theta)$, and the case $n = 3$ has *Rodrigues' formula* (also discovered by Euler [CG89]).

Lemma 5.2.5 (Euler's formula)**.** *For* $\Omega = \omega E_{12} \in \mathfrak{o}(2)$

$$e^{\Omega} = \cos(\omega)I + \sin(\omega)E_{12}.$$

Proof. We first note that

$$E_{12}^2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}^2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = -I. \tag{5.24}$$

Therefore

$$\begin{aligned} e^{\omega E_{12}} = \sum_{k=0}^{\infty} \frac{\omega^k}{k!} E_{12}^k &= \sum_{k=0}^{\infty} \frac{\omega^{2k}}{(2k)!} E_{12}^{2k} + \sum_{k=0}^{\infty} \frac{\omega^{2k+1}}{(2k+1)!} E_{12}^{2k+1} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{\omega^{2k}}{(2k)!} I + \sum_{k=0}^{\infty} (-1)^k \frac{\omega^{2k+1}}{(2k+1)!} E_{12} \\ &= \cos(\omega)I + \sin(\omega)E_{12}. \end{aligned} \tag{5.25}$$

□

Definition 5.2.2 (θ function)**.** Let

$$\begin{aligned} &\theta : \mathfrak{o}(2) \to \mathbb{R}, \quad \Omega = \omega E_{12} \mapsto |\omega| \\ &\theta : \mathfrak{o}(3) \to \mathbb{R}, \quad \Omega = \omega_{12}E_{12} + \omega_{13}E_{13} + \omega_{23}E_{23} \mapsto \sqrt{\omega_{12}^2 + \omega_{13}^2 + \omega_{23}^2} \end{aligned}$$

and if Ω is clear from context, we may write θ instead of $\theta(\Omega)$.

Remark. Note that $\theta = 0 \iff \Omega = 0$, and under the isomorphism $\mathfrak{o}(3) \cong \mathbb{R}^3$, then $\theta(\Omega) = \|\widetilde{\Omega}\|$.

Lemma 5.2.6 (Rodrigues' formula)**.** *For* $\Omega \in \mathfrak{o}(3)$

$$e^{\Omega} = I + \frac{\sin(\theta)}{\theta}\Omega + \frac{1 - \cos(\theta)}{\theta^2}\Omega^2.$$

Remark. Here $\sin(\theta)/\theta$ and $1 - \cos(\theta)/\theta^2$ are viewed as their power series expansions, making them smooth (analytic) functions, where for $\theta = 0$, then $\sin(\theta)/\theta = 1$ and $(1-\cos(\theta))/\theta^2 = 1/2$. Commonly one writes $\mathrm{sinc}(t) = \sin(t)/t$, $t \neq 0$, $\mathrm{sinc}(0) = 1$.

Proof. The idea is to use Cayley-Hamilton [RC12, Theorem 2.4.3.2]. Since

$$\det(\Omega - \lambda I) = \det \begin{bmatrix} -\lambda & \omega_{12} & \omega_{13} \\ -\omega_{12} & -\lambda & \omega_{23} \\ -\omega_{13} & -\omega_{23} & -\lambda \end{bmatrix} = -\lambda(\lambda^2 + \theta^2) \tag{5.26}$$

then $-\Omega(\Omega^2 + \theta^2 I) = 0$, that is $\Omega^3 = \theta^2\Omega$. Using this and splitting the even an odd terms in the power series for e^Ω yields the result. □

Remark. It turns out that $\pm i\theta$ are the non-zero eigenvalues of Ω for $n = 2, 3$, and this trend continues, but for $n \geq 4$ there are more eigenvalues that we need to obtain in order to find a nice expression for e^Ω using Cayley-Hamilton. In fact $\lfloor n/2 \rfloor$ eigenvalue pairs. See [GX03]. This makes the expressions much more complicated very quickly, and since there is no closed formula in general for polynomials of degree 5 or higher, even if we express the eigenvalues as roots of a polynomial in θ^2, for $n \geq 10$ there is no general closed form formula. Therefore it is very fitting that $n = 2, 3$ are most needed for applications.

Since the goal is to use the inverse of $\Omega \mapsto P_i e^\Omega$ as a chart, if we want to use this chart for applications we need a closed formula. We will call the inverse the *matrix logarithm* $\log_{P_i}$, which we will develop in a few steps. To invert $P_i e^\Omega \mapsto \Omega$, the idea will be to squeeze out as much information about e^Ω as possible, using functions that are familiar to us.

Lemma 5.2.7 (Trace identities)**.** *For $\Omega = \omega E_{12} \in \mathfrak{o}(2)$, then*

$$\operatorname{tr}(e^\Omega) = 2\cos(\omega)$$

and for $\Omega \in \mathfrak{o}(3)$

$$\operatorname{tr}(e^\Omega) = 1 + 2\cos(\theta).$$

Proof. The proof follows from Euler's formula and Rodrigues' formula. Since $\operatorname{tr}(\Omega) = 0$ in both cases, for $n = 2$

$$\operatorname{tr}(e^\Omega) = \cos(\omega)\operatorname{tr}(I) + \sin(\omega)\operatorname{tr}(E_{12}) = 2\cos(\omega) \tag{5.27}$$

and for $n = 3$

$$\operatorname{tr}(e^\Omega) = \operatorname{tr}(I) + \frac{\sin(\theta)}{\theta}\operatorname{tr}(\Omega) + \frac{1 - \cos(\theta)}{\theta^2}\operatorname{tr}(\Omega^2) = 3 + \frac{1 - \cos(\theta)}{\theta^2}\operatorname{tr}(\Omega^2). \tag{5.28}$$

Now

$$\Omega^2 = \begin{bmatrix} 0 & \omega_{12} & \omega_{13} \\ -\omega_{12} & 0 & \omega_{23} \\ -\omega_{13} & -\omega_{23} & 0 \end{bmatrix}^2 = \begin{bmatrix} -\omega_{12}^2 - \omega_{13}^2 & -\omega_{13}\omega_{23} & \omega_{12}\omega_{23} \\ -\omega_{13}\omega_{23} & -\omega_{12}^2 - \omega_{23}^2 & -\omega_{12}\omega_{13} \\ \omega_{12}\omega_{23} & -\omega_{12}\omega_{13} & -\omega_{13}^2 - \omega_{23}^2 \end{bmatrix} \tag{5.29}$$

and therefore

$$\operatorname{tr}(\Omega^2) = -2\omega_{12}^2 - 2\omega_{13}^2 - 2\omega_{23}^2 = -\theta^2, \tag{5.30}$$

meaning that

$$\operatorname{tr}(e^\Omega) = 3 - 2 + 2\cos(\theta) = 1 + 2\cos(\theta). \tag{5.31}$$

□

For $n = 2$ we now have a formula $\cos(\omega) = \operatorname{tr}(e^\Omega)/2$, and hence if we choose $|\omega| < \pi$, then we obtain the matrix logarithm $\omega = \arccos(\operatorname{tr}(e^\Omega)/2)$, $\Omega = \omega E_{12}$. Looking at Rodrigues' formula, an idea to extract Ω would be to apply a function which preserves Ω but kills I and Ω^2. Since Ω is skew-symmetric, and we have observed that Ω^2 is symmetric, the function skew will do the trick.

Lemma 5.2.8 (Skew identity)**.** *For* $\Omega \in \mathfrak{o}(3)$

$$\operatorname{skew}(e^{\Omega}) = \operatorname{sinc}(\theta)\,\Omega.$$

Proof. By (5.29), Ω^2 is symmetric, so

$$\begin{aligned}\operatorname{skew}(e^{\Omega}) = \frac{1}{2}(e^{\Omega} - (e^{\Omega})^{\mathsf{T}}) &= \frac{1}{2}\left(I - I + \frac{\sin(\theta)}{\theta}(\Omega - \Omega^{\mathsf{T}}) + \frac{1-\cos(\theta)}{\theta^2}(\Omega^2 - (\Omega^2)^{\mathsf{T}})\right) \\ &= \frac{1}{2}\left(\frac{\sin(\theta)}{\theta}2\Omega + \frac{1-\cos(\theta)}{\theta^2}(\Omega^2 - \Omega^2)\right) \\ &= \frac{\sin(\theta)}{\theta}\Omega \\ &= \operatorname{sinc}(\theta)\Omega.\end{aligned} \tag{5.32}$$

□

Now we can extract $\Omega = \operatorname{skew}(e^{\Omega})/\operatorname{sinc}(\theta)$ if we can obtain θ. By the trace identity, for $n = 3$ we have $\cos(\theta) = (\operatorname{tr}(e^{\Omega}) - 1)/2$, so if $\theta(\Omega) < \pi$ then $\theta = \arccos((\operatorname{tr}(e^{\Omega}) - 1)/2)$, and we have obtained the matrix logarithm.

Lemma 5.2.9 (Matrix logarithm)**.** *Let* $n \in \{2, 3\}$ *and define* $V = \{\Omega : \theta(\Omega) < \pi\}$, $U = e^V = \{e^{\Omega} : \theta(\Omega) < \pi\}$ *accordingly. Then* $\log : U \to V$

$$R \mapsto \arccos(\operatorname{tr}(R)/2), \qquad (n = 2)$$

$$R \mapsto \frac{\operatorname{skew}(R)}{\operatorname{sinc}(\arccos((\operatorname{tr}(R) - 1)/2))}, \qquad (n = 3)$$

is a diffeomorphism, with smooth inverse $\exp : V \to U$, $\Omega \mapsto e^{\Omega}$.

Proof. By Lemma 5.2.7 and Lemma 5.2.8, $\exp^{-1} = \log$, and log is smooth because its a composition of smooth functions. The set $U = \theta^{-1}(]-\infty, \pi[)$ is open because θ is continuous, and therefore $V = \log^{-1}(U)$ is open. Finally exp is smooth by [Hal15, Proposition 2.4], meaning that log is a diffeomorphism. □

Corollary 5.2.10 (Exponential chart)**.** *Let* $n \in \{2, 3\}$ *and* $P \in \mathrm{SO}(n)$. *Then*

$$\log_P : P \cdot U \to V, \quad R \mapsto \log(P^{\mathsf{T}}R)$$

is a chart around $P \in P \cdot U \subset \mathrm{SO}(n)$.

Proof. By Lemma 5.2.9, log is a diffeomorphism with domain U, and therefore $R \mapsto \log(P^{\mathsf{T}}R)$ is a diffeomorphism, with domain $P \cdot U = Pe^V = \{Pe^{\Omega} : \theta(\Omega) < \pi\}$. Since $e^0 = I$ and $\theta(0) = 0 < \pi$, then $I \in U$ and therefore $P \in P \cdot U$. □

Definition 5.2.3 (Setup)**.** Fix n and let $D = \operatorname{diag}(d_1, \ldots, d_n)$ with $0 \le d_1 < \cdots < d_n$, and for $1 \le i \le 2^{n-1}$, define the 2^{n-1} matrices $\mathcal{E}_i = \operatorname{diag}(\varepsilon_1^{(i)}, \ldots, \varepsilon_n^{(i)})$, $\varepsilon_l^{(i)} \in \{1, -1\}$, $\det(\mathcal{E}_i) = 1$. Moreover, let $P^* \in \mathrm{SO}(n)$ and define

$$f : \mathrm{SO}(n) \to \mathbb{R}, \quad R \mapsto \operatorname{tr}(D) - \operatorname{tr}(D(P^*)^{\mathsf{T}}R).$$

By Lemma 3.4.11, f has critical points $P_i = P^*\mathcal{E}_i$, and for each critical point we define

$$\lambda_{jk}^{(i)} = \varepsilon_j^{(i)} d_j + \varepsilon_k^{(i)} d_k.$$

Note that $\lambda_{jk}^{(i)} \neq 0$ since the d_l are distinct.

Since $\log_{P_i}$ is a chart around P_i, given f, we now wish to find a function z_i such that $\varphi_i = z_i \circ \log_{P_i}$ is a Morse chart for f, that is $f(R) = f(P_i) + ((\varphi_i)_1(R))^2 + \cdots + ((\varphi_i)_{c_i}(R))^2 - ((\varphi_i)_{c_i+1}(R))^2 - \cdots - ((\varphi_i)_m(R))^2$. Since R is in the domain of $\log_{P_i}$, we can write $R = P_i e^{\Omega}$, and then

$$f(R) = f(P_i) + ((z_i)_1(\Omega))^2 + \cdots + ((z_i)_{c_i}(\Omega))^2 - ((z)_{c_i+1}(\Omega))^2 - \cdots - ((z)_m(\Omega))^2 = f(P_i) + g_i(\Omega) \tag{5.33}$$

so we wish to study $g_i(\Omega) = f(R) - f(P_i)$ and use it to find z_i.

Lemma 5.2.11 (Morse form function)**.** *Define as in Definition 5.2.3. Then*

$$g_i : \mathfrak{o}(n) \to \mathbb{R}, \quad \Omega \mapsto f(P_i e^{\Omega}) - f(P_i)$$

fulfills $f(R) = f(P_i) + g_i(\Omega)$ for $R = P_i e^{\Omega}$, and if we let $\widetilde{D}_i = D\mathcal{E}_i = \operatorname{diag}(d_1\varepsilon_1^{(i)}, \ldots, d_n\varepsilon_n^{(i)})$, then

$$g_i(\Omega) = \operatorname{tr}(\widetilde{D}_i) - \operatorname{tr}(\widetilde{D}_i e^{\Omega}).$$

Proof. Since $P_i = P^*\mathcal{E}_i$ and $P^* \in \mathrm{SO}(n)$ we have $(P^*)^{\mathsf{T}}P^* = I$ and

$$\begin{aligned} g_i(\Omega) &= (\operatorname{tr}(D) - \operatorname{tr}(D(P^*)^{\mathsf{T}}P^*\mathcal{E}_i e^{\Omega})) - (\operatorname{tr}(D) - \operatorname{tr}(D(P^*)^{\mathsf{T}}P^*\mathcal{E}_i)) \\ &= \operatorname{tr}(D\mathcal{E}_i) - \operatorname{tr}(D\mathcal{E}_i e^{\Omega}) \\ &= \operatorname{tr}(\widetilde{D}_i) - \operatorname{tr}(\widetilde{D}_i e^{\Omega}). \end{aligned} \tag{5.34}$$

□

Lemma 5.2.12 (Morse charts for $n = 2$)**.** *Let $n = 2$ and define as in Definition 5.2.3. Then for each critical point $P_i \in \{P^*, -P^*\}$ let $U_i = P_i \cdot U = \{P_i e^{\omega E_{12}} : E_{12} \in \mathfrak{o}(2), |\omega| < \pi\} \subset \mathrm{SO}(2)$, $V = \{\omega E_{12} \in \mathfrak{o}(2) : |\omega| < \pi\} \cong \{\omega : |\omega| < \pi\} \subset \mathbb{R}$ and*

$$z_i : V \to z_i(V), \quad \omega \mapsto \sqrt{2|\lambda_{12}^{(i)}|}\sin(\omega/2).$$

Then (U_i, φ_i) is a Morse chart for f at P_i where $\varphi_i = z_i \circ \log_{P_i}$, so $\varphi_i(P_i e^{\omega E_{12}}) = z_i(\omega)$ and

$$f(R) = f(P_i) + \operatorname{sgn}(\lambda_{12}^{(i)})(\varphi_i(R))^2.$$

Proof. By Lemma 5.2.11, the idea to find z_i is to study $g_i(\Omega) = \operatorname{tr}(\widetilde{D}_i) - \operatorname{tr}(\widetilde{D}_i e^{\Omega})$. By Euler's formula

$$\widetilde{D}_i e^{\Omega} = \widetilde{D}_i e^{\omega E_{12}} = \begin{bmatrix} d_1\varepsilon_1^{(i)} & 0 \\ 0 & d_2\varepsilon_2^{(i)} \end{bmatrix} \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix} = \begin{bmatrix} d_1\varepsilon_1^{(i)}\cos(\omega) & d_1\varepsilon_1^{(i)}\sin(\omega) \\ -d_2\varepsilon_2^{(i)}\sin(\omega) & d_2\varepsilon_2^{(i)}\cos(\omega) \end{bmatrix}, \tag{5.35}$$

so

$$g_i(\Omega) = (d_1\varepsilon_1^{(i)} + d_2\varepsilon_2^{(i)}) - (d_1\varepsilon_1^{(i)} + d_2\varepsilon_2^{(i)})\cos(\omega) = \lambda_{12}^{(i)}(1 - \cos(\omega)), \tag{5.36}$$

where $\lambda_{12}^{(i)} \neq 0$ since $d_2 > d_1 \geq 0$. Now we want to write this on the form $\pm z^2$, where z is smooth. Double angle formula for cos gives

$$\begin{aligned} g_i(\Omega) = \lambda_{12}^{(i)}(1 - \cos(\omega)) = \lambda_{12}^{(i)}\, 2\sin^2(\omega/2) &= \frac{|\lambda_{12}^{(i)}|}{|\lambda_{12}^{(i)}|}\lambda_{12}^{(i)}\, 2\sin^2(\omega/2) \\ &= \operatorname{sgn}(\lambda_{12}^{(i)})\, 2|\lambda_{12}^{(i)}|\sin^2(\omega/2) \\ &= \operatorname{sgn}(\lambda_{12}^{(i)}) \left(\sqrt{2|\lambda_{12}^{(i)}|}\sin(\omega/2)\right)^2 \\ &= \operatorname{sgn}(\lambda_{12}^{(i)})(z_i(\Omega))^2 \\ &= \operatorname{sgn}(\lambda_{12}^{(i)})(\varphi_i(P_i e^{\Omega}))^2. \end{aligned} \tag{5.37}$$

Therefore, by the definition of g_i, $f(R) = f(P_i) + \operatorname{sgn}(\lambda_{12}^{(i)})(\varphi_i(R))^2$. What remains to show is that $\varphi_i = z_i \circ \log_{P_i}$ is a diffeomorphism. By Lemma 5.2.9, $\log_{P_i}$ is a diffeomorphism, and z_i is a diffeomorphism because it has the smooth inverse $z_i(r) = 2\arcsin(r/\sqrt{2|\lambda_{12}^{(i)}|})$. Therefore φ_i is a composition of diffeomorphisms, and hence a diffeomorphism. □

Lemma 5.2.13 (Morse charts for $n = 3$)**.** *Let $n = 3$ and define as in Definition 5.2.3. Then for each critical point $P_i \in \{P^*, P^*\mathcal{E}_2, P^*\mathcal{E}_3, P^*\mathcal{E}_4\}$ let $U_i = P_i \cdot U = \{P_i e^{\Omega} : \Omega \in \mathfrak{o}(3),\ \theta(\Omega) < \pi\} \subset \mathrm{SO}(3)$, $V = \{\Omega \in \mathfrak{o}(3) : \theta(\Omega) < \pi\} \cong \{\widetilde{\Omega} = (\omega_{12}, \omega_{13}, \omega_{23}) : \|\widetilde{\Omega}\| < \pi\} \subset \mathbb{R}^3$ and*

$$z_i : V \to z_i(V), \quad \Omega \mapsto \left(\sqrt{|\lambda_{jk}^{(i)}|/2}\ \omega_{jk}\ \operatorname{sinc}(\theta/2)\right)_{jk\in\{12,13,23\}}.$$

Then (U_i, φ_i) is a Morse chart for f at P_i where $\varphi_i = z_i \circ \log_{P_i}$, so $\varphi_i(P_i e^{\Omega}) = z_i(\Omega)$ and

$$f(R) = f(P_i) + \operatorname{sgn}(\lambda_{12}^{(i)})\left((\varphi_i(R))_{12}\right)^2 + \operatorname{sgn}(\lambda_{13}^{(i)})\left((\varphi_i(R))_{13}\right)^2 + \operatorname{sgn}(\lambda_{23}^{(i)})\left((\varphi_i(R))_{23}\right)^2.$$

Remark. The signs in the expression in the Morse chart is not necessarily in any standard order. Therefore we can permute the coordinates $\rho_i : \mathbb{R}^3 \to \mathbb{R}^3$ and take the $\rho_i \circ \varphi_i$ as the chart instead.

Proof. We follow the same structure as the proof of the case $n = 2$. By Rodrigues' formula

$$\widetilde{D}_i e^{\Omega} = \widetilde{D}_i + \frac{\sin(\theta)}{\theta}\widetilde{D}_i\Omega + \frac{1-\cos(\theta)}{\theta^2}\widetilde{D}_i\Omega^2 \tag{5.38}$$

where $\operatorname{tr}(\widetilde{D}_i\Omega) =< \widetilde{D}_i, \Omega >= 0$ by (5.17) since $\widetilde{D}_i$ is diagonal (and hence symmetric). Therefore

$$\begin{aligned} g_i(\Omega) = \operatorname{tr}(\widetilde{D}_i) - \operatorname{tr}(\widetilde{D}_i e^{\Omega}) &= \operatorname{tr}(\widetilde{D}_i) - \operatorname{tr}(\widetilde{D}_i) - \frac{\sin(\theta)}{\theta}\operatorname{tr}(\widetilde{D}_i\Omega) - \frac{1-\cos(\theta)}{\theta^2}\operatorname{tr}(\widetilde{D}_i\Omega^2) \\ &= -\frac{1-\cos(\theta)}{\theta^2}\operatorname{tr}(\widetilde{D}_i\Omega^2). \end{aligned} \tag{5.39}$$

Since multiplying with $\widetilde{D}_i$ just multiplies the rows of Ω^2 by the corresponding entries of $\widetilde{D}_i$, $\operatorname{tr}(\widetilde{D}_i\Omega^2) = \sum_{j=1}^{3} d_j\varepsilon_j^{(i)}\Omega_{jj}^2$, so (5.29) gives that

$$\begin{aligned} \operatorname{tr}(\widetilde{D}_i\Omega^2) &= d_1\varepsilon_1^{(i)}(\omega_{12}^2 + \omega_{13}^2) + d_2\varepsilon_2^{(i)}(\omega_{12}^2 + \omega_{23}^2) + d_3\varepsilon_3^{(i)}(\omega_{13}^2 + \omega_{23}^2) \\ &= \lambda_{12}^{(i)}\omega_{12}^2 + \lambda_{13}^{(i)}\omega_{13}^2 + \lambda_{23}^{(i)}\omega_{23}^2. \end{aligned} \tag{5.40}$$

This is quadratic expression in the entries ω_{jk}, so we just need $(1-\cos(\theta))/\theta^2$ to also be expressed as a square. Double angle formula for cos gives

$$\frac{1-\cos(\theta)}{\theta^2} = \frac{1}{\theta^2}(1-\cos(\theta)) = \frac{2}{\theta^2}\sin^2(\theta/2) = \frac{1}{2}\frac{\sin^2(\theta/2)}{(\theta/2)^2} = \frac{1}{2}\operatorname{sinc}^2(\theta/2). \tag{5.41}$$

Hence

$$g_i(\Omega) = \sum_{jk\in\{12,13,23\}} \frac{1}{2}\lambda_{jk}^{(i)}\,\omega_{jk}^2\,\operatorname{sinc}^2(\theta/2) = \sum_{jk\in\{12,13,23\}} \operatorname{sgn}(\lambda_{jk}^{(i)})\left(\sqrt{|\lambda_{jk}^{(i)}|/2}\ \omega_{jk}\ \operatorname{sinc}(\theta/2)\right)^2, \tag{5.42}$$

where by the definition of g_i, $f(R) = f(P_i) + \sum_{jk\in\{12,13,23\}} \operatorname{sgn}(\lambda_{jk}^{(i)})(\varphi_i(R))^2$, which is what we wanted to show. It remains to prove that $\varphi_i = z_i \circ \log_{P_i}$ is a diffeomorphism. Again, since $\log_{P_i}$ is a diffeomorphism, we just need to show that $z_i : V \to \operatorname{im}(z_i)$ is a diffeomorphism. We do this by finding the inverse explicitly, and showing that is is smooth.

Let $\Omega \in V = \{\Omega \in \mathfrak{o}(3) : \theta(\Omega) < \pi\}$, and $z_i(\Omega) = r = (r_{12}, r_{13}, r_{23}) \in \mathrm{im}(z_i) \subseteq \mathbb{R}^3$. We wish to obtain an expression for $\Omega \cong (\omega_{12}, \omega_{13}, \omega_{23})$. Now, by the definition z_i, if we let $c_{jk} = \sqrt{|\lambda^{(i)}_{jk}|/2}$, then $r_{jk} = c_{jk}\,\omega_{jk}\,\mathrm{sinc}(\theta/2)$. Therefore $\omega_{jk} = r_{jk}/(c_{jk}\,\mathrm{sinc}(\theta/2))$ and

$$\theta^2 = \omega_{12}^2 + \omega_{13}^2 + \omega_{23}^2 = \left(\left(\frac{r_{12}}{c_{12}}\right)^2 + \left(\frac{r_{13}}{c_{13}}\right)^2 + \left(\frac{r_{23}}{c_{23}}\right)^2\right)\frac{1}{\mathrm{sinc}^2(\theta/2)}. \tag{5.43}$$

If we let $C = \mathrm{diag}(c_{12}, c_{13}, c_{23})$, then we can write this as $\mathrm{sinc}^2(\theta/2)\theta^2 = \|C^{-1}r\|^2$, but $\mathrm{sinc}(\theta/2) = 2\sin(\theta/2)/\theta$, so

$$4\frac{\sin^2(\theta/2)}{\theta^2}\theta^2 = 4\sin^2(\theta/2) = \|C^{-1}r\|^2 \tag{5.44}$$

and therefore $\sin(\theta/2) = \|C^{-1}r\|/2$, where $\Omega \in V$ means that $\theta/2 \in [0, \pi/2[$ and so $\theta = 2\arcsin(\|C^{-1}r\|/2)$. Hence

$$\omega_{jk} = \frac{r_{jk}}{c_{jk}}\frac{1}{\mathrm{sinc}(\theta/2)} = \frac{r_{jk}}{c_{jk}}\frac{\theta/2}{\sin(\theta/2)} = \frac{r_{jk}}{c_{jk}}\frac{\arcsin(\|C^{-1}r\|/2)}{\|C^{-1}r\|/2} = \frac{r_{jk}}{c_{jk}}\mathrm{asinc}(\|C^{-1}r\|/2) \tag{5.45}$$

if we define $\mathrm{asinc} :]-1, 1[\to \mathbb{R}$, $t \mapsto \arcsin(t)/t$, $0 \mapsto 1$ as in (5.47). This is a smooth function in r, since $\mathrm{asinc}(t)$ only has even powers in its power series expansion, and $t^2 = (\|C^{-1}r\|/2)^2 = \|C^{-1}r\|^2/4$ is smooth in r. This means that

$$z_i^{-1} : \mathrm{im}(z_i) \to V, \quad (r_{jk}) \mapsto (\omega_{jk}) = \left(\frac{r_{jk}}{\sqrt{|\lambda^{(i)}_{jk}|/2}}\mathrm{asinc}\left(\frac{1}{\sqrt{2}}\sqrt{\frac{r_{12}^2}{|\lambda^{(i)}_{12}|} + \frac{r_{13}^2}{|\lambda^{(i)}_{13}|} + \frac{r_{23}^2}{|\lambda^{(i)}_{23}|}}\right)\right) \tag{5.46}$$

where $jk \in \{12, 13, 23\}$, since $c_{jk}^2 = |\lambda^{(i)}_{jk}|/2$, and $(1/2)\sqrt{\frac{1}{2}} = \frac{1}{\sqrt{2}}$. Since z_i^{-1} is continuous and V is open, so is $\mathrm{im}(z_i)$, so smoothness of z_i^{-1} means that z_i is a diffeomorphism.

Therefore $\varphi_i = z_i \circ \log_{P_i}$ is a diffeomorphism, and hence a chart. □

Definition 5.2.4 (asinc)**.** The function $\mathrm{asinc} :]-1, 1[\to \mathbb{R}$, $t \mapsto \arcsin(t)/t$, $0 \mapsto 1$ is smooth (analytic) with power series expansion

$$\mathrm{asinc}(t) = \frac{\arcsin(t)}{t} = \frac{1}{t}\left(t + \frac{1}{2}\frac{t^3}{3} + \frac{1}{2}\frac{3}{4}\frac{t^5}{5} + \dots\right) = 1 + \frac{1}{2}\frac{t^2}{3} + \frac{1}{2}\frac{3}{4}\frac{t^4}{5} + \dots. \tag{5.47}$$

5.2.3 Breeze on SO(2) and SO(3)

Lemma 5.2.14 (Breeze set on SO(2))**.** *Let $f : \mathrm{SO}(2) \to \mathbb{R}$, $R \mapsto \mathrm{tr}(D) - \mathrm{tr}(D(P^*)^{\mathsf{T}}R)$, where $D = \mathrm{diag}(1, 2)$. Then f has the critical points P^* and $-P^*$ where P^* is the unique local and global minimum of f. Moreover for $0 < A \leq 3$, (V_A, φ) is a breeze chart for f, where*

$$V_A = \left\{R : |\omega| < 2\arcsin(\sqrt{A/3})\right\}$$

and

$$\varphi(R) = \sqrt{6}\sin(\omega/2)$$

setting $\omega = \log_{-P^*}(R)$.

(Under the isomorphism $\mathrm{SO}(2) \cong \mathbb{S}^1 \cong \mathbb{R}/2\pi$, if $P^* \cong (\cos(\omega^*), \sin(\omega^*)) \cong \omega^*$ then $-P^* \cong \omega^* + \pi$ and $R \cong \omega + (\omega^* + \pi)$.)

Proof. By Lemma 3.4.11, f has the properties described. By Lemma 5.2.12, there is a Morse chart (U_2, φ_2) around $P_2 = -P^* = P^*(-I)$, such that

$$U_2 = \{R : |\omega| < \pi\},$$

$\varepsilon_l^{(2)} = -1$ and $\lambda_{12}^{(2)} = d_1\varepsilon_1^{(2)} + d_2\varepsilon_2^{(2)} = -1 - 2 = -3$, meaning that $\varphi_2(R) = \sqrt{6}\sin(\omega/2)$ and since $f(-P^*) = 2\operatorname{tr}(D) = 6$

$$f(R) = 6 - (\varphi(R)^2. \tag{5.48}$$

Therefore, by Definition 5.1.2,

$$V_A = \{R : |\varphi_2(R)| < \sqrt{2A}\} = \{R : |\sqrt{6}\sin(\omega/2)| < \sqrt{2A}\} = \left\{R : |\sin(\omega/2)| < \sqrt{A/3}\right\}$$

so, since $|\omega| < \pi$, setting $0 < \sqrt{A/3} \leq 1$, that is $0 < A \leq 3$ yields

$$V_A = \left\{R : |\omega| < 2\arcsin(\sqrt{A/3})\right\} \subseteq U_2 \tag{5.49}$$

since $0 < 2\arcsin(t) \leq \pi$ for $0 < t \leq 1$. □

Lemma 5.2.15 (Breeze vector field on SO(2))**.** *Let f and (V_A, φ) be as in Lemma 5.2.14. Then there exists a breeze vector field $Y : \mathrm{SO}(2) \to T\mathrm{SO}(2)$ for f with breeze set V_A, where for $R \in V_A$*

$$Y_R = \frac{2}{\sqrt{f(R)}} RE_{12}$$

Proof. We want $Y_R = \frac{\partial}{\partial\varphi|_R} = (D_R\varphi)^{-1}(1) = D_{\varphi(R)}\varphi^{-1}(1) = D_{\varphi(R)}\varphi^{-1}$. Now, since $\varphi^{-1} = (z \circ \log_{-P^*})^{-1} = -P^*e^{(-)} \circ z^{-1} = -P^*e^{z^{-1}}$, where $z(\omega E_{12}) = \sqrt{6}\sin(\omega/2)$, so $z^{-1}(r) = 2\arcsin(r/\sqrt{6})E_{12}$, and

$$D_\Omega(-P^*e^{(-)}) = -P^*e^\Omega \tag{5.50}$$

and

$$D_r z^{-1} = \frac{2}{\sqrt{6}} \frac{1}{\sqrt{1 - r^2/6}} E_{12} = \frac{2}{\sqrt{6 - r^2}} E_{12} \tag{5.51}$$

so by the chain rule

$$D_r\varphi^{-1} = D_{z^{-1}(r)}(-P^*e^{(-)}) \circ D_r z^{-1} = \frac{2}{\sqrt{6 - r^2}} \left(-P^*e^{z^{-1}(r)}E_{12}\right) \tag{5.52}$$

and therefore, when $r = \varphi(R)$ we have $z^{-1}(r) = \Omega$ (where $R = -P^*e^\Omega$), and $6 - \varphi(R)^2 = f(R)$ by (5.48), and therefore

$$Y_R = D_{\varphi(R)}\varphi^{-1} = \frac{2}{\sqrt{6 - \varphi(R)^2}} \left(-P^*e^\Omega E_{12}\right) = \frac{2}{\sqrt{f(R)}} RE_{12}. \tag{5.53}$$

By Lemma 5.1.1, we assemble this to a global vector field via a partition of unity. (However, the value of Y is not relevant outside V_A.) □

Lemma 5.2.16 (Breeze set on SO(3))**.** *Let $f : \mathrm{SO}(3) \to \mathbb{R}$, $R \mapsto \operatorname{tr}(D) - \operatorname{tr}(D(P^*)^\mathsf{T}R)$, where $D = \operatorname{diag}(1, 2, 4)$. Then f has the critical points P^*, $P_1 = P^*\mathcal{E}_1$, $P_2 = P^*\mathcal{E}_2$, and $P_3 = P^*\mathcal{E}_3$ where P^* is the unique local and global minimum of f and*

$$\mathcal{E}_1 = \operatorname{diag}(-1, -1, 1), \quad \mathcal{E}_2 = \operatorname{diag}(-1, 1, -1), \quad \mathcal{E}_3 = \operatorname{diag}(1, -1, -1).$$

Moreover $(V_{i,A}, \varphi_i)$, $1 \le i \le 3$ *are breeze charts for* f *if* $0 < A < 2/3$, *where*

$$V_{1,A} = \psi_1^{-1}\rho\left\{(c,b,a) : |a| < \sqrt{2A},\ c^2+b^2 < A\right\},$$
$$V_{2,A} = \psi_2^{-1}\left\{(a,b,c) : |b| < \sqrt{2A},\ a^2+c^2 < A\right\},$$
$$V_{3,A} = \psi_3^{-1}\left\{(a,b,c) : |c| < \sqrt{2A},\ a^2+b^2 < A\right\},$$

where $\rho(c,b,a) = (a,b,c)$, $\varphi_1 = \rho\circ\psi_1|_{V_{1,A}}$ *and* $\varphi_i = \psi_i|_{V_i}$ *for* $i = 2,3$ *where the* ψ_i *are the Morse charts of Lemma 5.2.13.*

Proof. By Lemma 3.4.11, f has the desired properties. Now we find a coordinate that contributes negatively in the expression for f in each Morse chart (U_i, ψ_i) given by Lemma 5.2.13. By the proof of Lemma 3.4.11, the index of P_1, P_2, and P_3 are 1, 2 and 3 respectively so each coordinate contributes negatively around P_3 and, $\lambda_{12}^{(1)} = -3 < 0$ for P_1, and $\lambda_{12}^{(2)} = 1 > 0$, so $\lambda_{13}^{(2)} < 0$ for P_2. Therefore we can permute the coordinates of ψ_1 so that for P_1 the 12 coordinate is last, to have an order of the coordinates as in the Morse lemma. Specifically we can let

$$V_{1,A} = \psi_1^{-1}\rho\left\{(c,b,a) : |a| < \sqrt{2A},\ c^2+b^2 < A\right\} \tag{5.54}$$

where $\rho(c,b,a) = (a,b,c)$ and

$$\begin{aligned} V_{2,A} &= \psi_2^{-1}\left\{(a,b,c) : |b| < \sqrt{2A},\ a^2+c^2 < A\right\} \\ V_{3,A} &= \psi_3^{-1}\left\{(a,b,c) : |c| < \sqrt{2A},\ a^2+b^2 < A\right\}. \end{aligned} \tag{5.55}$$

Since $\rho^{-1} = \rho$, this means that we have breeze charts $(V_{1,A}, \rho\circ\psi_1|_{V_{1,A}})$, and $(V_{i,A}, \psi_i|_{V_{i,A}})$ $i = 2,3$.

Finally we need to make sure that A is small enough such that $V_{i,A} \subseteq U_i = \{P_i e^{\Omega} : \theta(\Omega) < \pi\}$. By Lemma 5.2.13, the Morse chart ψ_i satisfies $\|C_i^{-1}\psi_i(R)\|^2 = 4\sin^2(\theta/2)$, where $C_i = \mathrm{diag}(c_{jk}^{(i)})$ with $c_{jk}^{(i)} = \sqrt{|\lambda_{jk}^{(i)}|/2}$, so $R \in U_i$ if and only if $\|C_i^{-1}\psi_i(R)\|^2 < 4$. For $R \in V_{i,A}$ the breeze set conditions give $\sum_{jk}(\psi_i(R)_{jk})^2 < 3A$, and therefore

$$\|C_i^{-1}\psi_i(R)\|^2 = \sum_{jk}\frac{(\psi_i(R)_{jk})^2}{(c_{jk}^{(i)})^2} \le \frac{\sum_{jk}(\psi_i(R)_{jk})^2}{\min_{jk}(c_{jk}^{(i)})^2} < \frac{3A}{\min_{jk}(c_{jk}^{(i)})^2}. \tag{5.56}$$

This is less than 4 whenever $A < \frac{4}{3}\min_{jk}(c_{jk}^{(i)})^2$. Computing from the λ-values, $\min_{jk}(c_{jk}^{(1)})^2 = 1$, $\min_{jk}(c_{jk}^{(2)})^2 = \frac{1}{2}$, and $\min_{jk}(c_{jk}^{(3)})^2 = \frac{1}{2}$, so the condition $A < \frac{2}{3}$ guarantees $V_{i,A} \subseteq U_i$ for each i. □

Lemma 5.2.17 (Breeze vector field on SO(3))**.** *Let* f *and* $(V_{i,A}, \varphi_i)$, $1 \le i \le 3$, *be as in Lemma 5.2.16. Then there exists a breeze vector field* $Y : \mathrm{SO}(3) \to T\mathrm{SO}(3)$ *for* f *with breeze charts* $(V_{i,A}, \varphi_i)$, *where for* $R = P_i e^{\Omega} \in V_{i,A}$

$$Y_R = \left.\frac{\partial}{\partial y_{i,1}}\right|_R = D_{\varphi_i(R)}\varphi_i^{-1}(e_{y_{i,1}}), \tag{5.57}$$

where $e_{y_{i,1}} \in \mathbb{R}^3$ *is the standard basis vector in the* $y_{i,1}$*-coordinate direction of* φ_i.

Proof. Since $\varphi_i^{-1} = \psi_i^{-1}\circ\rho_i^{-1}$ and ρ_i is linear (so $D\rho_i^{-1} = \rho_i^{-1}$ as a linear map), the chain rule gives

$$D_{\varphi_i(R)}\varphi_i^{-1}(e_{y_{i,1}}) = D_{\psi_i(R)}\psi_i^{-1}\big(\rho_i^{-1}(e_{y_{i,1}})\big) = D_{\psi_i(R)}\psi_i^{-1}(e_{(jk)^*}), \tag{5.58}$$

where $e_{(jk)^*} = \rho_i^{-1}(e_{y_{i,1}})$ is the standard basis vector in the breeze coordinate direction of ψ_i. Since $\rho_2 = \rho_3 = \mathrm{id}$, for $i = 2, 3$ we have $e_{(jk)^*} = e_{y_{i,1}}$, giving $e_{r_{13}}$ for P_2 and $e_{r_{23}}$ for P_3. For $i = 1$, $\rho_1 = \rho$ swaps the first and third coordinates, so $e_{y_{1,1}} = e_3$ and $\rho^{-1}(e_3) = e_1 = e_{r_{12}}$.

To evaluate $D_{\psi_i(R)}\psi_i^{-1}(e_{(jk)^*})$, let $s = \psi_i(R)$ and $\Omega_0 = z_i^{-1}(s)$, so that $R = P_i e^{\Omega_0}$. Consider the curve

$$c(t) = \psi_i^{-1}(s + te_{(jk)^*}) = P_i e^{\Omega(t)}, \qquad \Omega(t) = z_i^{-1}(s + te_{(jk)^*}), \tag{5.59}$$

so that $c(0) = R$ and $\dot{\Omega}(0) = (Dz_i^{-1})_s(e_{(jk)^*})$. The formula [Bla+09, Lemma 2] for the derivative of a matrix exponential gives

$$\left(e^{\Omega(-)}\right)'(0) = e^{\Omega_0} \int_0^1 e^{-u\Omega_0}\, \Omega'(0)\, e^{u\Omega_0}\, du, \tag{5.60}$$

so we obtain

$$Y_R = D_{\psi_i(R)}\psi_i^{-1}(e_{(jk)^*}) = P_i \left(e^{\Omega(-)}\right)'(0) = R \int_0^1 e^{-u\Omega_0}\, H\, e^{u\Omega_0}\, du, \tag{5.61}$$

where $H = (Dz_i^{-1})_{\psi_i(R)}(e_{(jk)^*}) \in \mathfrak{o}(3)$. We can calculate H by computing the derivative of z_i^{-1}, as seen in (5.46). This is smooth on $V_{i,A}$ since z_i^{-1} is smooth on $\psi_i(U_i)$ and $V_{i,A} \subseteq U_i$ for $A < 2/3$ by Lemma 5.2.16. Again, the global extension to all of SO(3) can be done as in the proof of Lemma 5.1.1, using a partition of unity subordinate to $\{V_{1,A}, V_{2,A}, V_{3,A}, \mathrm{SO}(3) \setminus \{P_1, P_2, P_3\}\}$. □

5.2.4 Main theorem applied to products of SO(2) and SO(3)

For this subsection, let

$$M = \mathrm{SO}(2)^{b_1} \times \mathrm{SO}(2)^{b_2} = M_1 \times \cdots \times M_L \tag{5.62}$$

where $b_1, b_2 \geq 0$, $L = b_1 + b_2 \geq 1$ and fix a target point $P^* = (P_1^*, \ldots, P_L^*) \in M$. By Theorem 1

$$f\big((R_l)_{l=1}^L\big) = \sum_{l=1}^L f_l(R_l), \qquad f_l(R_l) = \mathrm{tr}(D_l) - \mathrm{tr}\big(D_l (P_l^*)^{\mathsf{T}} R_l\big) \tag{5.63}$$

is a prepared perfect Morse function on M with unique local and global minimum, where $D_l = \mathrm{diag}(2^{s_l}, 2^{s_l+1})$ for SO(2)-factors and $D_l = \mathrm{diag}(2^{s_l}, 2^{s_l+1}, 2^{s_l+2})$ for SO(3)-factors, where s_l is given by Theorem 1. Again by Theorem 1, the total number of critical points of f that are not the local minimum is

$$N = 2^{b_1+2b_2} - 1. \tag{5.64}$$

Finally we equip M with the (Frobenius) product metric (see Lemma 2.8.3)

$$G_{(P_1,\ldots,P_L)} = (G_1)_{P_1} + \cdots + (G_L)_{P_L} \tag{5.65}$$

where we choose each $(G_l)_{P_l} = <>_{P_l}$ to be the Frobenius inner product

Lemma 5.2.18 (Gradient of product Morse function)**.** *The Riemannian gradient of f with respect to G is*

$$(\mathrm{grad}_G f)_{(R_l)} = \big(R_l\, \mathrm{skew}(P_l^* D_l R_l)\big)_{l=1}^L.$$

Proof. Since $G_{(R_l)}(u, v) = \sum_l G_{l,R_l}(u_l, v_l)$ for $u = (u_l), v = (v_l) \in T_{(R_l)}M = \bigoplus_l T_{R_l}M_l$, the defining property of the gradient gives

$$G_{(R_l)}\big((\mathrm{grad}_G f)_{(R_l)}, v\big) = D_{(R_l)}f(v) = \sum_{l=1}^L D_{R_l} f_l(v_l) = \sum_{l=1}^L G_{l,R_l}\big((\mathrm{grad}_{G_l} f_l)_{R_l}, v_l\big).$$

Comparing factor by factor yields $(\mathrm{grad}_G f)_{(R_l)} = ((\mathrm{grad}_{G_l} f_l)_{R_l})_l$, and Corollary 5.2.4 applied to each factor gives the result. □

The Morse coordinate chart at any critical point $P = (P_l)_l$ of f is obtained from Lemma 3.3.6 by applying the individual Morse charts from Lemma 5.2.12 and Lemma 5.2.13 to each factor, then permuting coordinates to place all positive directions before all negative ones. We denote the resulting product chart by

$$\varphi_P = \rho_P \circ (\varphi_{P_1}, \dots, \varphi_{P_l}), \tag{5.66}$$

where ρ_P is the permutation placing all positive coordinates first.

Lemma 5.2.19 (Breeze chart on products of SO(2) and SO(3))**.** *For each non-minimum critical point $P = (P_l)_l$ of f, let $l_0 = l_0(P) = \min\{l : P_l \neq P_l^*\}$ denote the index of the first non-minimum factor. Then, for $A > 0$ sufficiently small, $(V_{P,A}, \varphi_P)$ is a breeze chart at P, where φ_P is the product chart* (5.66) *and*

$$V_{P,A} = V_{P_{l_0},\, A/(2L)} \;\times \prod_{l \neq l_0} \varphi_{P_l}^{-1}\left\{ q \in \mathbb{R}^{\dim M_l} : \|q\|^2 < \frac{A}{2L} \right\}, \tag{5.67}$$

using the individual breeze set $V_{P_{L_0},A/(2L)}$ from Lemma 5.2.14 (if $M_{l_0} = \mathrm{SO}(2)$) or Lemma 5.2.16 (if $M_{l_0} = \mathrm{SO}(3)$). The first unstable coordinate of φ_P is $y_{P,1} = y_{l_0,1}$, the first unstable coordinate of the individual chart $\varphi_{P_{L_0}}$.

Proof. The chart φ_P is a Morse chart at P by Lemma 3.3.6, with index $k = \sum_l k_l \geq 1$. The product chart reorders coordinates as (α, β) with $\beta_1 = y_{l_0,1}$ as claimed.

We verify the breeze set conditions for $(R_l) \in V_{P,A}$. For factor l_0, the individual breeze set with parameter $A/(2L)$ gives $|y_{l_0,1}| < \sqrt{2 \cdot A/(2L)} = \sqrt{A/L}$ and $\|\alpha^{(l_0)}\|^2 + \sum_{j\geq 2}(y_{l_0,j})^2 < A/(2L)$. For each $l \neq l_0$, the ball condition gives $\|\varphi_{P_l}(R_l)\|^2 < A/(2L)$. Since $L \geq 1$, we have $\sqrt{A/L} \leq \sqrt{2A}$, and therefore

$$|\beta_1| = |y_{l_0,1}| < \sqrt{A/L} \leq \sqrt{2A},$$

and

$$\|\alpha\|^2 + \sum_{j\geq 2} \beta_j^2 \;\leq\; \Big(\|\alpha^{(l_0)}\|^2 + \sum_{j\geq 2} y_{l_0,j}^2\Big) + \sum_{l\neq l_0} \|\varphi_{P_l}(R_l)\|^2 \;<\; L \cdot \frac{A}{2L} = \frac{A}{2} < A,$$

so both conditions of Definition 5.1.2 are satisfied. Pairwise disjointness of the closures $\overline{V}_{P,A}$ for A small enough follows from Lemma 5.1.2 applied to the compact manifold M with Morse function f. The absence of P^* from any $\overline{V}_{P,A}$ holds because $P^* \notin \overline{V}_{P_{L_0},A/(2L)}$ (by Lemma 3.4.1, the Morse chart neighbourhoods of distinct critical points are disjoint) and $\overline{V}_{P,A} \subseteq \overline{V}_{P_{L_0},A/(2L)} \times \prod_{l\neq l_0} M_l$. □

Lemma 5.2.20 (Breeze vector field on products of SO(2) and SO(3))**.** *There exists a smooth breeze vector field $Y : M \to TM$ for f with breeze coordinate charts $(V_{P,A}, \varphi_P)$ from Lemma 5.2.19. For each non-minimum critical point P with $l_0 = l_0(P)$, the restriction to $V_{P,A}$ satisfies $Y|_{V_{P,A}} = \partial/\partial y_{P,1}$ and is given concretely by*

$$Y_{(R_l)} = (0, \dots, 0,\; Y_{l_0}(R_{l_0}),\; 0, \dots, 0) \;\in \bigoplus_{l=1}^{L} T_{R_l} M_l, \tag{5.68}$$

where Y_{l_0} is the individual breeze vector field of factor l_0 from Lemma 5.2.15 or 5.2.17.

Proof. The formula (5.68) defines a smooth vector field on each $V_{P,A}$ since Y_{l_0} is smooth on $V_{P_{L_0},A/(2L)} \supseteq \pi_{l_0}(V_{P,A})$, the projection to the l_0 factor. To verify that $Y|_{V_{P,A}} = \partial/\partial y_{P,1}$, we note that flowing along $(0, \dots, Y_{l_0}, \dots, 0)$ keeps all factors $l \neq l_0$ unchanged, so their chart

coordinates $\varphi_{P_l}(R_l)$ are constant, and the l_0 factor flows exactly along Y_{l_0}. Since $Y_{l_0}|_{V_{P_{L_0},A/(2L)}} = \partial/\partial y_{l_0,1} = \partial/\partial y_{P,1}$ by Lemma 5.2.15 or 5.2.17, this gives $Y|_{V_{P,A}} = \partial/\partial y_{P,1}$ as required. The global extension to all of M follows from Lemma 5.1.1, using a partition of unity subordinate to $\{V_{P,A} : P \text{ non-minimum}\} \cup \{M \setminus \{P_1, \dots, P_N\}\}$. □

Theorem 3 (Global asymptotic stabilization of $\mathrm{SO}(2)^{b_1} \times \mathrm{SO}(3)^{b_2}$)**.** *Let $M = \mathrm{SO}(2)^{b_1} \times \mathrm{SO}(3)^{b_2}$, $P^* = (P_l^*)_l \in M$, and f, G, Y be as above. The hybrid system $\mathcal{H}$ from Theorem 2, with*

- $F_0 = -\operatorname{grad}_G f$, *explicitly* $(F_0)_{(R_l)} = \big(-R_l \operatorname{skew}(P_l^* D_l R_l)\big)_l$ *by Lemma 5.2.18,*
- $F_1 = Y$, *the product breeze vector field of Lemma 5.2.20,*
- $h\big((R_l)\big) = \big(\|(\operatorname{grad}_G f)_{(R_l)}\|_G,\ f\big((R_l)\big)\big)$,

has $(P^*, 0) \in M \times \{0, 1\}$ as a globally asymptotically stable point, and the global asymptotic stability is robust. Every maximal solution executes at most

$$2N + 1 = 2^{b_1+2b_2+1} - 1 \tag{5.69}$$

jumps.

Proof. M is compact and connected since each $\mathrm{SO}(n)$ is compact and connected, and products preserve both properties (Lemma 2.3.5). By Theorem 1, f is a prepared Morse function with unique local minimum P^*, $f(P^*) = 0$, $N = 2^{b_1+2b_2} - 1$ where the other critical points $P_1, \dots, P_N$ satisfy $f(P_i) \geq 1$ and $|f(P_i) - f(P_j)| \geq 1$ for $i \neq j$. The vector field $F_0 = -\operatorname{grad}_G f$ is smooth by Lemma 5.2.18. The charts $(V_{P,A}, \varphi_P)$ from Lemma 5.2.19 are breeze coordinate charts in the sense of Definition 5.1.2. The breeze vector field Y from Lemma 5.2.20 satisfies $Y|_{V_{P,A}} = \partial/\partial y_{P,1}$. All hypotheses of Theorem 2 are therefore satisfied, giving robust global asymptotic stability of $(P^*, 0)$. The jump bound follows from Lemma 5.1.8 with N as in (5.64), giving $2N + 1 = 2(2^{b_1+2b_2} - 1) + 1 = 2^{b_1+2b_2+1} - 1$. □

Remark. The Morse function f of Theorem 1, and hence minimizes N to the Morse number $\mathrm{Morse}(M) = 2^{b_1+2b_2}$ by Lemma 3.5.2 and Theorem 1. Any other prepared Morse function on M would have at least as many non-minimum critical points.

References

[AS04] Andrei A. Agrachev and Yuri L. Sachkov. *Control Theory from the Geometric Viewpoint*. Springer, 2004.

[Bla+09] S. Blanes et al. "The Magnus expansion and some of its applications". In: *Physics Reports* 470.5-6 (2009). Preprint available as arXiv:0810.5488v1 [math-ph], pp. 151–238.

[CG89] Hui Cheng and K. C. Gupta. "An Historical Note on Finite Rotations". In: *Journal of Applied Mechanics* 56.1 (Mar. 1989), pp. 139–145. DOI: 10.1115/1.3176034. URL: https://iel.ucdavis.edu/publication/1989/j_ASMEAM.pdf.

[GST12] Rafal Goebel, Ricardo G. Sanfelice, and Andrew R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 2012.

[GX03] Jean Gallier and Dianna Xu. "Computing Exponentials of Skew-Symmetric Matrices and Logarithms of Orthogonal Matrices". In: *International Journal of Robotics and Automation* 18.1 (2003), pp. 10–20. URL: https://cs.brynmawr.edu/~dxu/206-2550-2.pdf.

[Hal15] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. 2nd ed. Springer, 2015.

[Hat02] Allen Hatcher. *Algebraic Topology*. Cambridge: Cambridge University Press, 2002.

[Hir76] Morris W. Hirsch. *Differential Topology*. Springer, 1976.

[HM94] Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. Springer, 1994.

[Lee11] John M. Lee. *Introduction to Topological Manifolds*. 2nd ed. Springer, 2011.

[Lee13] John M. Lee. *Introduction to Smooth Manifolds*. 2nd ed. Springer, 2013.

[Lee18] John M. Lee. *Introduction to Riemannian Manifolds*. 2nd ed. Springer, 2018.

[Mil63] John Milnor. *Morse Theory*. Princeton University Press, 1963.

[MS24] Richard Montgomery and Ricardo G. Sanfelice. "Hybrid Control, Morse Theory and Ivan Kupka". In: *Ivan Kupka Legacy: A Tour Through Controlled Dynamics*. Ed. by Bernard Bonnard et al. American Institute of Mathematical Sciences, 2024. Chap. 6, pp. 135–152.

[Mun00] James R. Munkres. *Topology*. 2nd ed. Prentice Hall, 2000.

[PP12] Kaare Brandt Petersen and Michael Syskind Pedersen. *The Matrix Cookbook*. Version 20121115. Nov. 2012. URL: http://www2.imm.dtu.dk/pubdb/p.php?3274.

[RC12] Horn R.A. and Johnson C.R. *Matrix Analysis*. 2nd ed. Cambridge University Press, 2012.

[Rom08] Steven Roman. *Advanced Linear Algebra*. 3rd ed. Springer, 2008.

[Rud76] Walter Rudin. *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill, 1976.

[San21] Ricardo G. Sanfelice. *Hybrid Feedback Control*. Princeton University Press, 2021.

[Sco05] Alexandru Scorpan. *The Wild World of 4-Manifolds*. American Mathematical Society, 2005.

[Sol16] Mehmet Solgun. "Perfect Morse Function on $SO(n)$". In: *European Journal of Pure and Applied Mathematics* (2016).

[Wal16] C. T. C. Wall. *Differential Topology*. Cambridge University Press, 2016.