GENERIC CONTROLLABILITY OF EQUIVARIANT SYSTEMS AND APPLICATIONS TO PARTICLE SYSTEMS AND NEURAL NETWORKS

ANDREI AGRACHEV AND CYRIL LETROUIT

ABSTRACT. There exist many examples of systems which have some symmetries, and which one may monitor with symmetry-preserving controls. Since symmetries are preserved along the evolution, full controllability is not possible, and controllability has to be considered inside sets of states with same symmetries. We prove that generic systems with symmetries are controllable in this sense.

This result has several applications, for instance: (i) generic controllability of particle systems when the kernel of interaction between particles plays the role of a mean-field control; (ii) generic controllability for families of vector fields on manifolds with boundary; (iii) universal interpolation for neural networks architectures with "generic" self-attention-type layers - a type of layers ubiquitous in recent neural networks architectures, e.g., in the Transformers architecture.

The tools we develop could help address various other questions of control of equivariant systems.

CONTENTS

1. INTRODUCTION AND MAIN RESULTS

1.1. **A motivating example.** Given n points at positions $x_1, \ldots, x_n \in \mathbb{R}^d$ (a point cloud), their empirical measure is the probability measure $\mu = \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j}$. We are interested in time-dependent evolutions in the set of point clouds, and more precisely in controlling these evolutions, which amounts to operating control in the set of empirical measures. To monitor these evolutions, we control the interactions of the points, assumed to be of the form

$$\forall i \in \{1, \ldots, n\}, \quad x_i(0) = x_i^0 \in \mathbb{R}^d \text{ and } \frac{d}{dt}x_i(t) = K_t\Big(x_i(t), \frac{1}{n}\sum_{j=1}^{n}\delta_{x_j(t)}\Big) \tag{1}$$

where $(K_t)_{t \in \mathbb{R}}$ is our control, and for any time $t \in \mathbb{R}$, K_t belongs to a given time-independent family $\mathcal{K}$. This is a particular type of *mean-field* control.

Said with words, the influence felt by $x_i(t)$ and generated by the points $x_j(t)$, $j \neq i$, is given by the time-dependent kernel K_t which we control. Each kernel $K \in \mathcal{K}$ depends both on $x_i(t)$ and all other positions $x_j(t)$, but not on their labels: K_t writes $K_t(x_i(t), \mu(t))$ where $\mu(t) = \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j(t)}$. This form of monitoring of particle systems evolution through control of the kernel arises for instance in neural networks architectures, as we will see later.

If τ denotes the permutation $\tau = (ij)$ acting on $(\mathbb{R}^d)^n$ by permuting the i-th and j-th particles, then this action commutes with the evolution (1). This observation implies the important property that the evolution (1) induces an evolution in the set of empirical measures; only

positions of the points matter, not their labels. Moreover, (1) and the induced evolution of empirical measures necessarily preserve the mass of individual points, since

$$x_i(0) = x_j(0) \Rightarrow \forall t \in \mathbb{R}, \ x_i(t) = x_j(t) \tag{2}$$

(here and in the sequel, we assume that the evolution (1) is globally well-posed).

Given a fixed time $T > 0$ and initial and final empirical measures μ^0, μ^1, our goal is to choose $K_t \in \mathcal{K}$ for each $t \in [0, T]$ in a way that the solution to (1) with $\mu(0) = \mu^0$ satisfies $\mu(T) = \mu^1$. To achieve this goal, the following mass-preservation constraint must hold, due to (2): to each point in the support of μ^0 must correspond a point in the support of μ^1 with the same mass, and this correspondance must be one-to-one. Equivalently,

$$\exists x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}^d, \text{ such that } \mu^0 = \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j}, \ \mu^1 = \frac{1}{n} \sum_{j=1}^{n} \delta_{y_j},$$

$$\text{and for any } i, j \in \{1, \ldots, n\}, \ x_i = x_j \text{ if and only if } y_i = y_j. \tag{3}$$

Our aim in this paper is to prove that if the constraint (3) is satisfied, this control problem is generically feasible. More precisely, if $\mathcal{K}$ contains at least two elements, and that they are "generic" enough, the condition (3) is the only constraint that initial and final data must satisfy to achieve our goal of sending μ^0 to μ^1 through an evolution of the form (1).

It turns out that this problem may be rephrased abstractly as a control problem in a manifold M endowed with the action of a compact Lie group G. In the above example, $M = (\mathbb{R}^d)^n$ and $G = \mathfrak{S}_n$ is the symmetric group which acts on M by permuting copies of $\mathbb{R}^d$ (see Example 1.1). As already observed, permutations commute with the evolution (1). In the general case, we consider only evolutions in M following vector fields which are equivariant under the action of G (see definition below). Consequently, motion in M is constrained to remain inside some strata. In the above example, each stratum gathers points clouds with the same repartition of mass, i.e., two point clouds belong to the same stratum if and only if their empirical measures μ^0 and μ^1 satisfy (3). In the general case, strata are given by connected components of sets of points in M whose isotropy groups (=stabilizers) are in the same conjugacy class.

Solving this abstract control problem requires to dive into the subtle geometry of equivariant dynamical systems. As a result, it offers a wide range of applications, sometimes far from our original example of particle systems. Also, we believe that the tools we develop to solve this problem could be useful to address other questions related to control of equivariant systems.

Section 1.2 is devoted to the formulation of our main results, and Section 1.3 sketches some of its applications, which are developed in more detail in Section 2.

1.2. **Main results.** Let M be a real analytic[1] manifold and let G be a compact Lie group acting analytically on M. M is thus called an analytic G-manifold, and

$$M_G := M/G = \{Gq \mid q \in M\}$$

is the orbit space, each set Gq being an orbit, or a G-orbit. The set of C^∞ vector fields on M is denoted by $\mathrm{Vec}(M)$. In the sequel, for $s \in \mathbb{N}$, we use the notation $[s] = \{1, \ldots, s\}$.

The isotropy group at $q \in M$ is

$$G_q = \{g \in G \mid gq = q\}.$$

Notice that if q, q' belong to the same G-orbit, then G_q and $G_{q'}$ are conjugate. The action of G on M induces a natural stratification

$$M_G = \bigsqcup_{i \in \mathcal{I}_G} S_G^i \tag{4}$$

where for each $i \in \mathcal{I}_G$, S_G^i is a connected component of the set of all orbits whose representatives have isotropy groups conjugate to[2] some given subgroup H_i of G.

---

[1] in this paper, all real analytic manifolds are assumed to be paracompact and second countable.

[2] this does not depend on the representative

While reading the rest of this section, it might be helpful to keep in mind the example of Section 1.1, summarized here.

Example 1.1. *In Section 1.1, $G = \mathfrak{S}_n$ acts on $M = (\mathbb{R}^d)^n$ through*

$$\sigma \cdot (x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)}).$$

Then M_G may be identified with the set of n-points empirical measures, i.e., probability measures of the form $\mu = \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j}$. Two elements of M_G belong to the same stratum in the decomposition (4) if the corresponding empirical measures, denoted for instance by μ^0 and μ^1, satisfy (3). The number of strata is finite.

The stratification (4) is sometimes called "stratification by isotropy types". As recalled in Section 3.3, $\mathcal{I}_G$ is countable and each stratum S_G^i is a smooth submanifold.

Given $g \in G$ we define the diffeomorphism P_g on M by $P_g q = gq$. The pushforward $Y = (P_g)_* X$ of a vector field X through P_g is given by $Y(q) = (dP_g)_{P_g^{-1}(q)}(X(P_g^{-1}(q)))$ for any $q \in M$. Equivariant vector fields are those which are compatible with the action of G: $X \in \mathrm{Vec}(M)$ is equivariant under the action of G if for any $g \in G$ there holds

$$(P_g)_* X = X. \tag{5}$$

Any equivariant vector field induces a vector field on M_G. We denote by $\mathrm{Vec}^G(M_G)$ the set of vector fields induced on M_G by C^∞ equivariant vector fields on M. This set is equipped with the (induced) C^∞ topology on compact sets, whose definition is recalled in Section 3.1.

We are interested in controllability properties in the quotient set M_G, with vector fields in $\mathrm{Vec}^G(M_G)$. We will see in Section 3 that this is tightly related, but not equivalent, with controllability in M with equivariant vector fields (controllability in M_G is slightly weaker). We prove in Lemma 3.9 that any element of $\mathrm{Vec}^G(M_G)$ is tangent to strata defined in (4). Therefore, any integral curve of $\mathrm{Vec}^G(M_G)$ is contained in a single stratum, which implies that it is not possible to connect points $q, q' \in M_G$ by integral curves of $\mathrm{Vec}^G(M_G)$ if q, q' do not belong to the same stratum. However, it might be possible to connect them if the two points q, q' belong to the same stratum. This observation motivates the following definition:

Definition 1.2 (Controllability in strata). *Let $X_1, \ldots, X_k \in \mathrm{Vec}^G(M_G)$. We say that controllability holds in strata if for any $q, q' \in M_G$ belonging to the same stratum in (4), there exist $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that*

$$q' = e^{t_1 X_{i_1}} \circ \ldots \circ e^{t_m X_{i_m}} q$$

where e^{tX} denotes the flow at time t of the vector field X on M_G.

Our first main result is the following:

Theorem 1.3. *There exists for any $k \geq 2$ a set of k-uples $(X_1, \ldots, X_k) \in (\mathrm{Vec}^G(M_G))^k$ which is residual in $(\mathrm{Vec}^G(M_G))^k$ and for which controllability holds in strata.*

This roughly means that controllability in strata holds for "almost any" k-uples of vector fields in $\mathrm{Vec}^G(M_G)$. Since $\mathrm{Vec}^G(M_G)$ is infinite-dimensional, residual sets are an appropriate framework to state "almost-sure" properties.

We actually prove a slightly stronger statement than Theorem 1.3; namely, we prove the controllability in the leaves of the foliation generated by the equivariant fields on $M \setminus M'$, where M' is defined in Section 3. The precise statement is given in Theorem 3.3.

In view of applications[3], it is natural to extend our result to the simultaneous control of N points in M_G. Simultaneous control (also called "ensemble control") means that with a single control that is shared by all N points in M_G, we seek to drive the N initial points to their N respective targets (see [AS20]). In Example 1.1, this means driving N empirical measures to N

---

[3]The main application we have in mind here is to self-attention layers of neural networks, see Section 2.4. Neural networks with self-attention layers are designed to map billions of sequences to billions of target sequences. Self-attention layers are implemented for instance in the Transformers architecture [V+17], whose success in machine learning calls for mathematical explanations.

other empirical measures, evolving each of them independently (i.e. the empirical measures do not interact with each other), but with the same interaction kernel K_t which may depend on time.

Definition 1.4 (Simultaneous controllability in strata)**.** *Let $X_1, \ldots, X_k \in \mathrm{Vec}^G(M_G)$ for some $k \geq 2$. We say that simultaneous controllability in strata (of dimension ≥ 2) holds if for any $N \in \mathbb{N}$, any $q_1, \ldots, q_N, q'_1, \ldots, q'_N \in M_G$ satisfying:*

  (i) *for any distinct $i, j \in [N]$, $q_i \neq q_j$ and $q'_i \neq q'_j$*
  (ii) *for any $j \in [N]$, q_j and q'_j belong to the same stratum in (4), and the dimension of this stratum is ≥ 2*

the following conclusion holds: there exist $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that

$$\forall j \in [N], \qquad q'_j = e^{t_1 X_{i_1}} \circ \ldots \circ e^{t_m X_{i_m}} q_j.$$

We prove the following result, which is a generalization to the equivariant framework of [AS20, Theorem 1]:

Theorem 1.5. *For any $k \geq 2$, there exists a set of k-uples of equivariant C^∞-vector fields $(X_1, \ldots, X_k)$ which is residual in $(\mathrm{Vec}^G(M_G))^k$, and for which simultaneous controllability in strata holds.*

[AS20, Theorem 1] can be recovered by taking G reduced to the identity. Our proof of Theorem 1.5 (and of Theorem 1.3) is constructive, whereas the proof of [AS20, Theorem 1] relied on the multijet transversality theorem as a black-box.

1.3. **Applications.** Theorems 1.3 and 1.5 have various applications, which are developed in Section 2.

  (i) *Control on manifolds with boundary* (Section 2.1). Indeed, any manifold with boundary may be written as the quotient of a manifold without boundary by a reflection. This application of Theorem 1.3 is technically the simplest where G is non-trivial, since G is just $\mathbb{Z}/2\mathbb{Z}$.
  (ii) *Control of the spectrum of symmetric (or Hermitian) matrices* (Section 2.2). We apply Theorem 1.3 to the case where M is the set of symmetric matrices, and G is the orthogonal group, acting by conjugation on M. Each element A of M/G may be identified with the spectrum of any of its representatives (=symmetric matrices), i.e., the collection of eigenvalues seen up to permutations. The stratum to which A belongs depends on the cardinality of each packet of coincident eigenvalues of A.
  (iii) *Control of particle systems* (Section 2.3). This covers the example presented in Section 1.1, and its generalizations. In this case the particles live in some manifold W, so $M = W^n$, and $G = \mathfrak{S}_n$ (the particles are indistinguishable). The orbit space M_G is identified with the set of n-points empirical measures on W.
  (iv) *Universal interpolation for generic self-attention layers in neural networks* (Section 2.4). This can be framed as a particular case of the previous application. In this case each element of M_G represents for instance a sentence, each element of W is a word embedding ("a token"), and the time-evolution corresponds to evolution across layers. The goal explained in Section 1.1 of sending the initial empirical measure to the target one typically represents a translation task, realized sentence by sentence (and not word by word).
  (v) *Control of quantum systems with symmetries*, notably symmetric Ising spin networks of n spin $\frac{1}{2}$ particles (Section 2.5). These networks of n states evolve according to Hamiltonians which are invariant under permutations of the spins. This is a particular case of application (iii), in the case where W is the unit sphere of $\mathbb{C}^2$, in which spins live.

1.4. **Open questions.** Here are a few open questions which we believe of particular interest:

  (1) generalize Theorem 1.3 to the case where G is not compact. Natural examples are the groups of translations and homotheties (centered at the origin) in Euclidean spaces.

(2) generalize Theorem 1.3 (or the simpler theorem by Lobry [Lob72], see Section 1.5) to the case where M has infinite dimension. This would possibly have applications to control of measures, of diffeomorphisms, and control of the spectrum of general self-adjoint or Hermitian operators of infinite dimension.

(3) in the present paper, we only study the controllability problem, but the *optimal control* problem is also certainly worth studying, for instance for an ensemble of N points on M_G. This is natural in view of application to neural networks architectures. See [Sca23] in the case where there is no group acting on M.

Also, let us mention here that our assumption of analyticity on M and G (which is satisfied in all natural examples) is technical. Although we do not know how to avoid it, we do not believe that this assumption is fundamental.

1.5. **Bibliography.** The idea of proving controllability for "generic dynamics" as in Theorem 1.3 is not new: in [Lob72], Lobry proved that for a generic family of $k \geq 2$ vector fields on a connected manifold M without boundary, any couple of points of M may be connected by an integral path of the family. An outcome of our approach is an extension of Lobry's result to manifolds with boundary, see Corollary 2.1. Lobry's paper has been extended in [AS20] to the case of ensembles of points on M evolving according to a shared open loop control; the points are not interacting with each other but they are driven by the same control. This has direct applications to universal interpolation for so-called "neural ODEs". Theorem 1.5 in the present work generalizes [AS20, Theorem 1] to the equivariant framework, which is relevant among other applications to neural networks equipped with self-attention layers (see Section 2.4). Compared to [Lob72] and [AS20], our proofs are constructive and do not use (multi-)jet transversality. Also, the fact that M is endowed with a group action is not a mere additional technicality: our proofs require a detailed understanding of the structure of orbits and strata of G-manifolds, partly based on the so-called slice theorem.

It is important to mention that there already exists a vast literature on equivariant dynamical systems, see e.g. the book [Fie07] for a detailed account. Controllability and observability of equivariant dynamical systems, which are basically systems with symmetries, are well-developed subjects, see for instance [BMR08], [BMR09], and [MGH22] for a review. To our knowledge, generic controllability has never been studied in this framework, and the applications which we propose are original. We believe our fine analysis of equivariant dynamics (e.g., Lemma 3.9) is of independent interest and could be useful to address other questions related to control of equivariant systems.

More references on applications of our results are given in Section 2.

1.6. **Organization of the paper.** The applications sketched in Section 1.3 are developed in Section 2.

We prove Theorem 1.3 in Section 3. For this, we do not use equivariant transversality theory[4]. Using this theory might seem to be a natural lead since the papers [Lob72], [AS20] rely on transversality theory, but it turns out to be very cumbersome since equivariant transversality theory has many subtleties. Instead, our proofs rely on averaging techniques and codimension computations, and, as already said, require a detailed understanding of the structure of orbits and strata of G-manifolds, partly based on the so-called slice theorem.

In Section 4, we prove Theorem 1.5, which is a generalization of the proof of Theorem 1.3.

---

[4]see [Fie07] for an exhaustive treatment of this theory.

2. APPLICATIONS

This section develops the applications which have been sketched in Section 1.3.

2.1. **Manifolds with boundary.** Let $\widetilde{M}$ be a real-analytic manifold with smooth boundary $\partial \widetilde{M} \neq \emptyset$. We denote by $\mathcal{U}$ the set of C^∞-vector fields on $\widetilde{M}$, defined up to the boundary $\partial \widetilde{M}$, and which are tangent to $\partial \widetilde{M}$. This set is endowed with the C^∞ topology on compact sets. Corollary 2.1 asserts that for a generic k-uple ($k \geq 2$) of elements of $\mathcal{U}$, any two points belonging either to the same connected component of $\partial \widetilde{M}$ or to the same connected component of the interior of $\widetilde{M}$ may be connected by an integral path of the vector fields. This result is a generalization to manifolds with boundary of a result due to Lobry [Lob72] (which does not require the analyticity assumption).

Corollary 2.1 (Generic control on manifolds with boundary). *Let $\widetilde{M}$ be a real-analytic and connected manifold with smooth boundary $\partial \widetilde{M}$. Then for any integer $k \geq 2$ there exists a residual set of k-uples $(X_1, \ldots, X_k) \in \mathcal{U}^k$ for which the following property holds. For any q, q' belonging either to the same connected component of $\partial \widetilde{M}$ or to the same connected component of the interior of $\widetilde{M}$, there exist $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that $q' = e^{t_1 X_{i_1}} \circ \ldots \circ e^{t_m X_{i_m}} q$.*

Let us explain how Corollary 2.1 follows from Theorem 1.3. To any compact manifold $\widetilde{M}$ with smooth boundary is naturally associated a G-manifold M, with $G = \mathbb{Z}_2$ (here $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$), constructed as follows. First, the double M of $\widetilde{M}$ is formed by gluing together two copies of $\widetilde{M}$ along their common boundary. There is a natural action of $\mathbb{Z}_2$ by reflection on the manifold M fixing the common boundary, and sending each point of the first copy of $\widetilde{M}$ to the same point in the second copy of $\widetilde{M}$, and vice versa. Then, the quotient space $M/\mathbb{Z}_2$ can be identified with $\widetilde{M}$.

The strata of $\widetilde{M} \simeq M/\mathbb{Z}_2$ are the connected components of the boundary and of the interior of $\widetilde{M}$. As an illustration, if $\widetilde{M}$ is a disk, then M is a 2-dimensional sphere, and the set of fixed points through the mirror action is an equator of M. Equivariant vector fields on M are tangent to this equator. Then Theorem 1.3 applied to $\widetilde{M}$ yields Corollary 2.1. Of course, there exists an "ensemble version" of Corollary 2.1, which follows from Theorem 1.5.

2.2. **Spectrum of matrices.** Our results also have applications to control of symmetric and Hermitian matrices. Let us consider the natural action of the orthogonal group $G = \mathcal{O}_n$ on the space of symmetric matrices $\mathcal{S}_n$: any $P \in \mathcal{O}_n$ acts by $\mathcal{S}_n \ni S \mapsto PSP^\top$. By diagonalization of symmetric matrices, we identify an element of $\mathcal{S}_n/\mathcal{O}_n$ with the empirical measure of the eigenvalues of any of its representatives in $\mathcal{S}_n$, i.e.,

$$\mathcal{S}_n/\mathcal{O}_n \simeq \mathbb{R}^n/\mathfrak{S}_n \simeq \mathcal{M}_n(\mathbb{R})$$

where $\mathfrak{S}_n$ denotes the symmetric group over n elements, acting on $\mathbb{R}^n$ by permuting the coordinates, and $\mathcal{M}_n(\mathbb{R})$ is the set of empirical measures over n points in $\mathbb{R}$, that is, the set of probability measures on $\mathbb{R}$ of the form $\mu = \frac{1}{n} \sum_{j=1}^n \delta_{x_j}$.

Given $S \in \mathcal{S}_n/\mathcal{O}_n$, we denote by $\lambda_1 < \ldots < \lambda_{n(S)}$ the *distinct* eigenvalues of any representative, and by $m_1, \ldots, m_{n(S)}$ their respective multiplicities. The tuple $(m_1, \ldots, m_{n(S)})$ is called the ordered multiplicities of S. Then $S, S' \in \mathcal{S}_n/\mathcal{O}_n$ belong to the same stratum if and only if their ordered multiplicities coincide, i.e., $n(S) = n(S')$ and $m_i(S) = m_i(S')$ for any $1 \leq i \leq n(S) = n(S')$. As a side remark, recall that it is known since Von Neumann and Wigner [VW29] that in $\mathcal{S}_n$, the set of matrices with two coincident eigenvalues has codimension 2.

We denote by $\mathcal{W}$ the set of vector fields on $\mathcal{S}_n/\mathcal{O}_n \simeq \mathcal{M}_n(\mathbb{R})$ induced by $\mathcal{O}_n$-equivariant C^∞ vector fields on $\mathcal{S}_n$. Along each integral curve in $\mathcal{S}_n/\mathcal{O}_n$ of the family $\mathcal{W}$, ordered multiplicities are preserved. Our result may be phrased as follows: for generic k-uples ($k \geq 2$) of elements of $\mathcal{W}$, it is possible to transfer by appropriate composition of the flows of these vector fields any empirical measure of eigenvalues to any other empirical measure of eigenvalues with the same ordered multiplicities.

Corollary 2.2 (Generic control of the spectrum of symmetric matrices)**.** *For any integer $k \geq 2$ there exists a residual set of k-uples $(X_1, \ldots, X_k) \in \mathcal{W}^k$, for which the following property holds. For any $S, S' \in \mathcal{S}_n/\mathcal{O}_n \simeq \mathcal{M}_n(\mathbb{R})$ whose ordered multiplicities coincide, there exist $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that $S' = e^{t_1 X_{i_1}} \circ \ldots \circ e^{t_m X_{i_m}} S$.*

This statement is a consequence of Theorem 1.3, and there exists an "ensemble version" which follows from Theorem 1.5. Also, an analogous statement holds for the natural action of the unitary group on the space of $n \times n$ Hermitian matrices.

2.3. **Particle systems.**

We develop now the application of our results to (mean-field) control of particle systems, making rigorous Section 1.1. If $n \in \mathbb{N}$ and W is a manifold, the set $\mathcal{M}_n(W)$ of empirical measures over n points in W, that is, the set of measures of the form $\mu = \frac{1}{n} \sum_{j=1}^n \delta_{x_j}$, also naturally carries a G-manifold structure. It is isomorphic to the quotient of W^n by the action of the symmetric group $\mathfrak{S}_n$ given by

$$\sigma : (x_1, \ldots, x_n) \mapsto (x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \tag{6}$$

for $\sigma \in \mathfrak{S}_n$. The isotropy group G_q as $q = (x_1, \ldots, x_n)$ is not reduced to the identity if and only if at least two of the x_i's are equal.

In the sequel we assume that W is real-analytic, connected and of dimension ≥ 2 (the case where $\dim(W) = 1$ is actually treated in Section 2.2). Two points $q, q' \in W^n$, $q = (x_1, \ldots, x_n)$, $q' = (x'_1, \ldots, x'_n)$ have conjugate isotropy groups if and only if there exists $h : W \to W$ and $\sigma \in \mathfrak{S}_n$ such that $h(x_\ell) = x'_{\sigma(\ell)}$ for any $\ell \in [n]$, in other words if and only if the numbers of pairs/triples/quadruples/... of points among $x_1, \ldots, x_n$ which are equal coincide with the same numbers computed for $x'_1, \ldots, x'_n$. This condition is actually necessary and sufficient for q and q' to belong to the same stratum: since W is connected and of dimension ≥ 2, it is easy to construct a smooth path from $\frac{1}{n} \sum_{j=1}^n \delta_{x_j}$ to $\frac{1}{n} \sum_{j=1}^n \delta_{x'_j}$.

It is possible to give an analytic characterization of $\mathfrak{S}_n$-equivariant vector fields on W^n. For this we denote by $\mathcal{M}_n^\bullet(W)$ the set of couples (x, μ) where $\mu \in \mathcal{M}_n(W)$ and $x \in \operatorname{supp}(\mu)$. Then equivariant vector fields are in one-to-one correspondance with functions $f : \mathcal{M}_n^\bullet(W) \to T_\bullet W$ where this notation means that $f(x, \mu) \in T_x W$ for $x \in \operatorname{supp}(\mu)$. Let us describe this one-to-one correspondance. If $f : \mathcal{M}_n^\bullet(W) \to T_\bullet W$, then for $x = (x_1, \ldots, x_n) \in W^n$, we define

$$V(x) = (f(x_1, \mu), \ldots, f(x_n, \mu)) \in T_x W^n, \qquad \text{where } \mu = \frac{1}{n} \sum_{j=1}^n \delta_{x_j}.$$

It is immediate to verify that V is an equivariant vector field on W^n. Conversely, if $V = (V_1, \ldots, V_n)$ is an equivariant vector field, for $(x, \mu) \in \mathcal{M}_n^\bullet(W)$ we set

$$f(y, \mu) = V_1(y, x_2, \ldots, x_n) \in T_y W \tag{7}$$

where we have written $\mu = \frac{1}{n}(\delta_y + \sum_{j=2}^n \delta_{x_j})$. The right-hand side in (7) does not depend on the order in which we put $x_2, \ldots, x_n$ since V is equivariant. Therefore (7) yields a well-defined $f : \mathcal{M}_n^\bullet \to T_\bullet W$.

We denote by $\mathcal{V}$ the set of functions $f : \mathcal{M}_n^\bullet(W) \to T_\bullet W$ such that the equivariant vector field on W^n associated with f is C^∞ and generates a globally defined flow. The set $\mathcal{V}$ is endowed with the C^∞ topology on compact sets.

According to the above characterization of equivariant vector fields on W^n, Theorem 1.3 reads as follows in this context:

Corollary 2.3 (Generic control of particle systems)**.** *For any integer $k \geq 2$ there exists a residual set of tuples $(f_1, \ldots, f_k) \in \mathcal{V}^k$ for which the following property holds. For any $\mu^0, \mu^1 \in \mathcal{M}_n(W)$ in the same stratum, written as*

$$\mu^0 = \frac{1}{n} \sum_{r=1}^n \delta_{x_r^0} \qquad and \qquad \mu^1 = \frac{1}{n} \sum_{r=1}^n \delta_{x_r^1} \tag{8}$$

there exist $m \in \mathbb{N}$, $0 = t_0 < t_1 < \ldots < t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that the unique solution to the system of coupled ODEs

$$\forall \ell \in [n], \ \forall j \in [m], \ \forall t \in [t_{j-1}, t_j[, \qquad \frac{d}{dt} x_\ell(t) = f_{i_j}(x_\ell(t), \mu(t)), \qquad \mu(t) = \frac{1}{n} \sum_{r=1}^{n} \delta_{x_r(t)} \quad (9)$$

with initial values $(x_1(0), \ldots, x_n(0)) = (x_1^0, \ldots, x_n^0)$ satisfies

$$\frac{1}{n} \sum_{r=1}^{n} \delta_{x_r(t_m)} = \mu^1. \tag{10}$$

With the stronger Theorem 3.3, it is even possible to impose that $x_r(t_m) = x_r^1$ for any $r \in [n]$ instead of the weaker condition (10) (but for this it is necessary that the points x_r^0, x_r^1 in the writing (8) are numbered in a way that $x_r^0 = x_{r'}^0$ if and only if $x_r^1 = x_{r'}^1$, which is possible since μ^0 and μ^1 belong to the same stratum).

There are other natural group actions on W^n for which Theorem 1.3 has natural corollaries, for instance the action by rotations and/or reflections when $W = \mathbb{R}^d$. We leave the precise statements to the reader. We also mention that the same G-manifold structure has been used to model the geometry of chords of music instruments, see e.g. [Tym06].

2.4. Universal interpolation with generic self-attention layers.

The application of Theorem 1.3 (or Theorem 1.5) to particle systems, developed in the previous section, is relevant in machine learning. One of our motivations for this paper is actually to understand the possibilities of approximation offered by a relatively new neural network architecture introduced in [V+17], called Transformers, which play nowadays a central role in the inner workings of large language models (the last letter in "Chat-GPT" stands for Transformers): more precisely, we would like to study which classes of functions these neural networks architectures are able to approximate. If this class is large, it suggests that the architecture is able to handle many different types of data and problems.

Approximation and interpolation properties of some neural networks (see [CLLS23] for precise definitions) are known to be equivalent to controllability properties of some non-linear systems of ODEs in discrete or continuous time. In the past 5 years, tools from geometric control like Lie bracketing have therefore been used to study the controllability properties of so-called ResNets (standing for "residual neural networks") and their continuous-time version called neural ODEs, see e.g. [CLT20], [TG22], [AS20].

However, Transformers are not of the same nature as ResNets and neural ODEs. The main difference is that they incorporate self-attention layers, which may be seen from the mathematical point of view as interacting particle systems or evolutions in the set of (empirical) measures (see [VBT20], [SABP22], [GLPR23a], [GLPR23b]). The results of the present paper give insights on the approximation/interpolation properties of (generalized) self-attention layers, if one forgets about the rest of the architecture of Tranformers which, in addition to self-attention layers, usually incorporate also normalization layers and multi-layer perceptrons.

Our definition of self-attention layers is much broader than the specific self-attention layers used in practice: we call "generalized self-attention layer" any vector field on $(\mathbb{R}^d)^n$ which may be written for some $f : \mathcal{M}_n^\bullet(\mathbb{R}^d) \to T_\bullet \mathbb{R}^d$ as

$$(x_1, \ldots, x_n) \mapsto (f(x_1, \mu), \ldots, f(x_n, \mu)), \qquad \text{where} \quad \mu = \frac{1}{n} \sum_{r=1}^{n} \delta_{x_r}. \tag{11}$$

In other words, generalized self-attention layers are nothing else than the infinitesimal-time version of a permutation-equivariant sequence-to-sequence maps. In discrete time, they would take the form of a discrete system of coupled ODEs

$$x_i((m+1)\Delta t) = x_i(m\Delta t) + (\Delta t) f(x_i(m\Delta t), \mu(m\Delta t)), \qquad \mu(m\Delta t) = \frac{1}{n} \sum_{r=1}^{n} \delta_{x_r(m\Delta t)} \tag{12}$$

for some fixed $\Delta t > 0$, and any $m \in \mathbb{N}$. In the terminology of neural networks, the term $x_i(m\Delta t)$ in the right-hand side is called a skip-connection.

Our result, which is a genericity result, works for "almost all generalized self-attention layers", but does not say anything about the interpolation properties of the specific self-attention layers used in practice, which correspond to the choice of functions f of the form

$$f\left(x, \frac{1}{n}\sum_{j=1}^{n}\delta_{x_j}\right) = \sum_{h=1}^{H}\frac{\sum_{j=1}^{n}e^{\langle Q_h x, K_h x_j\rangle}V_h x_j}{\sum_{j=1}^{n}e^{\langle Q_h x, K_h x_j\rangle}} \tag{13}$$

for some $d \times d$ matrices Q_h, K_h, V_h (see below for literature on this problem).

The study of the interpolation properties of generalized self-attention layers boils down to a problem of controllability of interacting particle system of the form presented in Corollary 2.3 (or Corollary 2.4 below). Our goal is not to control an interacting particle system with fixed interaction kernel by acting on a subset of particles, which is a classical question; instead, in our problem, the controls are directly given by a family of interaction kernels.

One particular feature of generalized self-attention layers (or equivalently interacting particle systems) is that they are equivariant with respect to permutation of particles. Therefore, approximation/interpolation properties are considered in the class of permutation-equivariant maps. Recall that equivariant neural networks are of particular interest because they maintain their performance even when the input data undergoes certain transformations, such as rotations, translations, or scaling, and they have a reduced number of parameters compared to non-equivariant counterparts, as they exploit the inherent symmetries in the data.

In the context of particle systems, Theorem 1.5 takes the form of Corollary 2.4 below. Interpreted as a universal interpolation result, it says that if we take $k \geq 2$ sufficiently generic maps (=layers) of the form (11), then it is possible for any given initial set of N point clouds and any given final set of N (target) point clouds, to compose these k maps[5], possibly many times and in some appropriate order, in a way to move each of the N point clouds to its target. In other words,

universal interpolation is a generic property of k-uples of generalized self-attention layers.

Corollary 2.4. *For any integer $k \geq 2$ there exists a residual set of tuples $(f_1, \ldots, f_k) \in \mathcal{V}^k$ for which the following property holds. Let $N \in \mathbb{N}$ and $\mu_1^0, \ldots, \mu_N^0, \mu_1^1, \ldots, \mu_N^1$ be distinct elements of $\mathcal{M}_n(W)$ such that for any $h \in [N]$, μ_h^0 and μ_h^1 are in the same stratum. For any $h \in [N]$ we write*

$$\mu_h^0 = \frac{1}{n}\sum_{r=1}^{n}\delta_{x_{h,r}^0} \qquad and \qquad \mu_h^1 = \frac{1}{n}\sum_{r=1}^{n}\delta_{x_{h,r}^1}.$$

Then there exist $m \in \mathbb{N}$, $0 = t_0 < t_1 < \ldots < t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that for any $h \in [N]$, the unique solution to the system of coupled ODEs

$$\forall \ell \in [n], \forall j \in [m], \forall t \in [t_{j-1}, t_j[, \quad \frac{d}{dt}x_{h,\ell}(t) = f_{i_j}(x_{h,\ell}(t), \mu_h(t)), \quad \mu_h(t) = \frac{1}{n}\sum_{r=1}^{n}\delta_{x_{h,r}(t)} \tag{14}$$

with initial condition $(x_{h,1}(0), \ldots, x_{h,n}(0)) = (x_{h,1}^0, \ldots, x_{h,n}^0)$ satisfies

$$\frac{1}{n}\sum_{r=1}^{n}\delta_{x_{h,r}(t_m)} = \mu_h^1.$$

Compared to previous results in the literature [YBRRK19], [ADTK23], our result does not need to incorporate multi-layer perceptrons to achieve universal interpolation. Besides, it is not restricted to the Euclidean setting and works for data on general manifolds W, thus it may be considered as a statement in geometric deep learning (see [BBLSV17], and [G+23] for geometric deep learning with equivariant neural networks). However, its main drawback is that it does not rely on self-attention layers used in practice, namely (13), but on "generic" self-attention

---

[5]At the level of the discretized equation (12), this would mean alternating between different f for different values of m.

layers, of the general form (11). Also, it deals with universal interpolation instead of universal approximation.

It would be relevant to prove an analogue of Corollary 2.3 in restricted classes of particle systems. We have in mind the following type of statement: for generic k-uples of functions chosen in a class of evolutions $\mathcal{F}$ smaller than $\mathcal{V}_p$, the same conclusion as in Corollary 2.3 holds. For instance, does it hold when $\mathcal{F}$ is the family of evolutions driven by pairwise interactions? I.e., each f_i in (9) is of the form

$$f_i\Big(x, \frac{1}{n}\sum_{j=1}^n \delta_{x_j}\Big) = \sum_{j=1}^n K_i(x, x_j)$$

for some kernel K_i.

Another family $\mathcal{F}$ with meaningful applications is given by the attention dynamics in Transformers without layer normalization, i.e., each f_i would be of the form (13) (see [SABP22], [GLPR23a], [GLPR23b]) for some $H \in \mathbb{N}$, and some $d \times d$ matrices Q_h, K_h, V_h.

Finally, let us mention that there are other equivariant neural networks architectures for which universal approximation theorems in the class of equivariant maps have already been proved, see for instance [KP19] for Graph Neural Networks. Our results might also be applicable to this setting.

2.5. **Control of quantum systems with symmetries.** Theorem 1.3 also applies to quantum systems controlled through Hamiltonians which display some symmetries. Let us develop one example. The papers [CZDP17], [AD18] study symmetric Ising spin network where n spin $\frac{1}{2}$ particles (i.e., vectors in the unit sphere of $\mathbb{C}^2$) interact through permutation-invariant Hamiltonians. What we call here a permutation-invariant Hamiltonian is a Hamiltonian which is invariant under permutation of the spins, for instance

$$H_{zz} = \sum_{1 \leq k < m \leq n} \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \underbrace{\sigma_z}_{k^{\text{th}} \text{ place}} \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \underbrace{\sigma_z}_{m^{\text{th}} \text{ place}} \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1}$$

or else

$$H_{xyz} = \sum_{i,j,k \text{ distinct}} \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \underbrace{\sigma_x}_{i^{\text{th}} \text{ place}} \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \underbrace{\sigma_y}_{j^{\text{th}} \text{ place}} \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1} \otimes \underbrace{\sigma_z}_{k^{\text{th}} \text{ place}} \otimes \mathbf{1} \otimes \ldots \otimes \mathbf{1}$$

where

$$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y := \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \qquad \sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are the Pauli matrices, and $\mathbf{1}$ is the 2×2 identity matrix. We will say that two networks $q_1 = (q_1^{(1)}, \ldots, q_1^{(n)})$ and $q_2 = (q_2^{(1)}, \ldots, q_2^{(n)})$ of n labelled spin $\frac{1}{2}$ particles (each $q_i^{(j)}$ belongs to the unit sphere of $\mathbb{C}^2$) have same symmetries if $q_1^{(i)} = q_1^{(j)} \Leftrightarrow q_2^{(i)} = q_2^{(j)}$, for any $i, j \in [n]$. Our results (precisely, Theorem 3.3) imply that if we are given at least two generic permutation-invariant Hamiltonians, and two networks q_1, q_2 of n spin $\frac{1}{2}$ particles having same symmetries, then there exists a path from q_1 to q_2 given by appropriate compositions of the unitary evolutions associated to the Hamiltonians. In other words, subspace controllability (see [Dal23, Section 1]) generically holds.

The focus of [AD18] is different but related: the authors give explicit examples of permutation-invariant Hamiltonians for which they are able to prove controllability and to design control laws.

More generally, concerning the Lie bracket approach to control of quantum systems, we also mention the paper [Llo95], in which it is shown that generically, a quantum logic gate with two or more inputs is computationally universal, i.e., copies of the gate can be "wired together" to effect any desired logic circuit, and to perform any desired unitary transformation on a set of quantum variables. This result exactly corresponds to Lobry's theorem [Lob72], applied in the particular context of unitary transformations: in other words, it asserts that if one can apply some Hamiltonians (at least two) repeatedly to a few variables at a time one can in general effect any desired unitary time evolution on an arbitrarily large number of variables.

3. Proof of Theorem 1.3

This section is devoted to the proof of Theorem 1.3. We actually prove a slightly stronger statement, given in Theorem 3.3. In the whole paper, we work under the assumptions that M is a real analytic manifold and G is a compact Lie group acting analytically on M.

3.1. **A stronger statement.** This subsection is devoted to introducing notation and stating Theorem 3.3.

Orbits, i.e., sets of the form Gx, are closed submanifold (see e.g. [Fie07, Corollary 3.1.17]). We denote by M' the set of points $x \in M$ at which all equivariant vector fields are tangent to the orbit Gx. This set M' is closed and G-invariant.

Example 3.1. *If G acts transitively on M, then $M' = M$. Even if G is finite, the set M' may be non-empty: for instance, if M is the unit circle in $\mathbb{R}^2$ and G acts by permuting coordinates, then M' has cardinal 2.*

Although the above example shows that M' is not empty in general, there holds $M' = \emptyset$ in all applications listed in Section 2 and in Example 1.1. When $M' \neq \emptyset$, controllability is not necessarily possible inside connected components of M', as shown in Example 3.5, therefore we need to adapt the definition of controllability, see Definition 3.2.

We denote by $\mathrm{Vec}^G(M)$ the set of equivariant C^∞ vector fields on M, equipped with the C^∞ topology on compact sets defined as follows. For any chart (ϕ, U) of M, any compact set $K \subset U$, any $X \in \mathrm{Vec}^G(M)$, $k \in \mathbb{N}$ and $\varepsilon > 0$, we consider the set $\mathcal{U}(X, \phi, K, \varepsilon, k)$ of vector fields $Y \in \mathrm{Vec}^G(M)$ such that

$$\sup_{x \in \phi(K), 0 \leq |\alpha| \leq k} \|D^\alpha(\phi_* X)(x) - D^\alpha(\phi_* Y)(x)\| < \varepsilon.$$

The C^∞ topology on compact sets is defined as the set of all such sets $\mathcal{U}(X, \phi, K, \varepsilon, k)$. Since no confusion is possible, the quotient topology on $\mathrm{Vec}^G(M_G)$ is also called C^∞ topology on compact sets.

The set $\mathrm{Vec}^G(M)$ is an involutive distribution, therefore it is integrable according to Frobenius' theorem. We foliate M with maximal connected integral manifolds of this distribution. An *equivariant leaf* is any of the leaves of this foliation. As a consequence, it is possible to connect any two points in an equivariant leaf by following equivariant vector fields.

Definition 3.2 (Controllability in equivariant leaves)**.** *Let $X_1, \ldots, X_k \in \mathrm{Vec}^G(M)$. We say that controllability holds in equivariant leaves if for any $q, q' \in M \setminus M'$ belonging to the same equivariant leaf, there exist $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that*

$$q' = e^{t_1 X_{i_1}} \circ \ldots \circ e^{t_m X_{i_m}} q$$

where e^{tX} denotes the flow at time t of the vector field X on M.

Theorem 3.3. *There exists for any $k \geq 2$ a set of k-uples $(X_1, \ldots, X_k) \in (\mathrm{Vec}^G(M))^k$ which is residual in $(\mathrm{Vec}^G(M))^k$ and for which controllability in equivariant leaves holds.*

We refer the reader to Example 3.14 for an illustration of the concept of controllability in equivariant leaves.

Recall that the Lie algebra generated by a family $\mathcal{F} \subset \mathrm{Vec}^G(M)$ is the smallest sub-algebra of $\mathrm{Vec}^G(M)$ containing $\mathcal{F}$, namely

$$\mathrm{Lie}(\mathcal{F}) := \mathrm{Span} \left\{ [Y_1, \ldots, [Y_{j-1}, Y_j]], Y_i \in \mathcal{F}, j \in \mathbb{N} \right\}.$$

Theorem 3.3 is a corollary of the following result (see Section 3.5 for a proof of this implication).

Theorem 3.4. *For any $k \geq 2$, there exists a set of k-uples $(X_1, \ldots, X_k) \in (\mathrm{Vec}^G(M))^k$ which is residual in $(\mathrm{Vec}^G(M))^k$ and for which*

$$\forall q \in M \setminus M', \quad \mathrm{Lie}_q(X_1, \ldots, X_k) = \mathrm{Vec}^G(M)_{|q}. \tag{15}$$

We end this section with an example showing that controllability does not necessarily hold inside connected components of M' even if they form equivariant leaves, which explains why the set M' is excluded in Definition 3.2.

Example 3.5. *We consider $M = \mathbb{R}^d/\mathbb{Z}^d$ for some $d > 2$, and G the set of all translations in M. The action is transitive. Then equivariant vector fields are constant vector fields, automatically tangent to the unique orbit, therefore $M' = M$. However, for any $k \leq d - 1$ equivariant vector fields $X_1, \ldots, X_k$, controllability cannot hold inside M'. Indeed, since $X_1, \ldots, X_k$ are constant, the integral curves of the distribution spanned by $X_1, \ldots, X_k$ are strict subsets of M.*

Remark 3.6. *We will see in Corollary 3.10 that M'/G is a set of isolated points in M_G, and that each point in M'/G is a stratum in (4). Therefore controllability in the sense of Definition 1.2 is automatically satisfied in M'/G, although all elements of $\mathrm{Vec}^G(M_G)$ vanish in M'/G by definition of M'. This is why the set M'/G does not need to be excluded in the statement of Theorem 1.3. On the other side, Example 3.5 shows that controllability does not hold inside M', which illustrates why M' needs to be removed in the statement of Theorem 3.3 (actually it is removed in Definition 3.2).*

3.2. **Averaging.** We gather in this section a definition and a proposition that are required for the rest of the proof.

First we define the averaging operator, which turns any vector field on M to an equivariant one by averaging over G-orbits. Recall that equivariant vector fields are vector fields on M, and that they induce vector fields on M_G.

Definition 3.7 (Averaging). *Given $X \in \mathrm{Vec}(M)$, we set*

$$X^{eq} = \int_G (P_g)_* X \ d\mu_G(g) \in \mathrm{Vec}^G(M) \tag{16}$$

where μ_G is the normalized Haar measure on G.

Averaging commutes with Lie brackets, as shown in the following proposition in which X is required to be equivariant.

Proposition 3.8 (Averaging and bracketing commute). *If $X \in \mathrm{Vec}^G(M)$, then for any $Y \in \mathrm{Vec}(M)$,*

$$[X, Y^{eq}] = [X, Y]^{eq} \tag{17}$$

Proof. Since X is equivariant, there holds $(P_g)_* X = X$ for any $g \in G$, hence

$$[X, Y^{eq}] = \int_G [X, (P_g)_* Y] \ d\mu_G(g) = \int_G [(P_g)_* X, (P_g)_* Y] \ d\mu_G(g) = [X, Y]^{eq}$$

where for the last equality we used $(P_g)_*[X, Y] = [(P_g)_* X, (P_g)_* Y]$. $\square$

3.3. **Density of vector fields transverse to G-orbits.** We introduce a stratification of M, whose link with the stratification (4) is clarified below. The action of G on M induces a natural stratification

$$M = \bigsqcup_{i \in \mathcal{I}} S^i \tag{18}$$

where for each $i \in \mathcal{I}$, S^i is a connected component of the set of points of M whose isotropy group is conjugate to some given subgroup H_i of G. The slice theorem recalled in the proof of Lemma 3.9 has several consequences. Firstly, according to [Fie07, Proposition 3.7.2], each S^i is a smooth submanifold. Secondly, $\mathcal{I}$ is countable: this is e.g. a consequence of the proof of [Fie07, Proposition 3.7.4]. Moreover, as a consequence of Lemma 3.9 below, any equivariant leaf (see Section 3.1) is contained in one of the strata in (18). The image of any stratum in (18) under quotient by G is a stratum in (4). Conversely, the preimage of any stratum in (4) is the union of a finite number of strata in (18).

Recall that M is assumed real-analytic. For any invariant subset $U \subset M$, we denote by $\mathcal{X}^\omega_{\mathrm{eq}}(U)$ (resp. $\mathcal{X}^\omega_{\mathrm{eq}}(U)$) the set of real-analytic (resp. smooth) equivariant vector fields on U

and by $C_G^\infty(U)$ the set of smooth G-invariant real-valued functions on U. We consider the $C_G^\infty(U)$-module

$$\mathcal{A}_U = \left\{ \sum_{j \in \mathcal{J}} a_j Y_j \mid \mathcal{J} \text{ finite}, \ \forall j \in \mathcal{J}, a_j \in C_G^\infty(U), \ Y_j \in \mathcal{X}_{\text{eq}}^\omega(U) \right\}.$$

Then we set

$$\mathcal{A} = \{ X \in \mathcal{X}_{\text{eq}}^\infty(M) \mid \forall x, \exists U_x \neq \emptyset \text{ invariant open neighborhood of } x \text{ s.t. } X_{|U_x} \in \mathcal{A}_{U_x} \}$$

For $x \in M$, we denote by A_x the operator $A_x : \mathcal{A} \to T_x M$,

$$A_x : X \mapsto X^{eq}(x) \tag{19}$$

(see (16)). Its image is denoted by $\text{Im}(A_x)$. We recall that G-orbits, i.e., sets of the form Gx for some $x \in M$, are closed submanifolds (see e.g. [Fie07, Corollary 3.1.17]).

Lemma 3.9. *Let S be one of the strata in (18) and let $x \in S$. Then*

$$\text{Im}(A_x) \subset T_x S \subset \text{Im}(A_x) + T_x(Gx). \tag{20}$$

Lemma 3.9 is fundamental in the sequel, and it may be tested for instance on Example 3.14 for which the inclusion of $\text{Im}(A_x)$ in $T_x S$ is strict.

Proof of Lemma 3.9. Our proof of (20) is based on the slice theorem (Theorem 3.5.2 in [Fie07]) whose statement is the following: for every $x \in M$, it is possible to choose a smooth family of pairwise-disjoint slices

$$\mathcal{S} = \{ S_y \mid y \in Gx \}$$

satisfying the following properties:

- For $y \in Gx$, $S_y \subset M$ is a G_y-invariant embedded disk of M of dimension $\dim(M) - \dim(Gx)$ which is transverse to Gx.
- For $y \in Gx$ and $g \in G$, $g \cdot S_y = S_{g \cdot y}$. In particular, $g \cdot S_y = S_y$ for $g \in G_y$.
- For $y \in Gx$, S_y is G-equivariantly diffeomorphic via $\exp_y$ to the representation given by the linear action of G_y on $(T_y Gx)^\perp$.
- For $z \in S_y$, G_z is a subgroup of G_y (see Lemma 3.7.1(a) in [Fie07]).
- $GS_x = \bigcup_{y \in Gx} S_y$ is an open G-invariant neighbourhood of Gx which is G-equivariantly diffeomorphic to the twisted product $G \times_{G_x} S_x$.

These slices are actually defined as follows. We first define a G-invariant metric ν on M by taking any Riemannian metric on M, then pushing it forward by the G-action and finally averaging the result with respect to the Haar measure. For $y \in Gx$ we denote by $N_y(\varepsilon)$ the set of $v \in (T_y Gx)^\perp$ such that $\nu_y(v) < \varepsilon$. Then the slices S_y for $y \in Gx$ are given by $S_y = \exp_y(N_y(\varepsilon))$ for some ε sufficiently small and depending only on x. The key point is that the exponential map $\exp : TM \to M$ is G-equivariant, giving the linear action of G_y on $(T_y Gx)^\perp$. It is easily checked that if $z = \exp_y(v)$ for some $v \in N_y(\varepsilon)$, then

$$G_z = \{ g \in G_y \mid g_*(v) = v \}. \tag{21}$$

Therefore, the slice theorem follows from this construction.

We first prove that

$$\text{Im}(A_x) = T_x E_x \tag{22}$$

where E_x is the set of points with same isotropy group as x. Notice that E_x is locally diffeomorphic near x to the subspace of $(T_x Gx)^\perp$ given by vectors which are invariant under the linear action of G_x (by (21)), hence it is a submanifold. Also, this observation shows that $\text{Im}(A_x) \subset T_x E_x$ since equivariant vector fields evaluated at x are invariant under G_x.

We then prove $T_x E_x \subset \text{Im}(A_x)$. Let $v \in T_x E_x$, then $h_* v = v$ for any $h \in G_x$. We define a vector field on S_x as follows, using the third point of the slice theorem: we consider the preimage u of v through the diffeomorphism $\exp_x$, which is a vector at the origin in $(T_x Gx)^\perp$ which verifies $g_*(u) = u$ for any $g \in G_x$, we extend u to a constant vector field on $(T_x Gx)^\perp$, and then we push it forward to S_x through $\exp_x$. This vector field is well-defined and equivariant

because if $g_2 z = g_1 z$ for some $z \in S_x$, then $g_2 = g_1 h$ for some $h \in G_x$ according to the second point of the slice theorem. We extend this vector field to $G S_x$ by pushforward by G, obtaining an equivariant and analytic (because the action is analytic) vector field. We then use a smooth G-invariant cut-off[6] equal to 1 on $G S_x$ to obtain an element $X \in \mathcal{A}$, equal to v on $G S_x$. The involution $g \mapsto g^{-1}$ preserves the normalized Haar measure μ_G on G because compact Lie groups are unimodular (i.e. the left-invariant measure is also right-invariant). Hence

$$X^{eq}(x) = \int_G ((P_g)_* X)(x)\, d\mu_G(g) = \int_G ((P_{g^{-1}})_* X)(x)\, d\mu_G(g) = \int_G (P_{g^{-1}})_* (P_g)_* v\, d\mu_G(g) = v$$

therefore $v \in \mathrm{Im}(A_x)$ which concludes the proof of (22).

To prove (20) it is now sufficient to prove $T_x E_x \subset T_x S \subset T_x E_x + T_x(Gx)$. The inclusion $T_x E_x \subset T_x S$ follows from the definition of strata. Let us prove that $T_x S \subset T_x E_x + T_x(Gx)$. For this, we use the fourth point of the slice theorem. Since all elements of S have isotropy groups conjugated to G_x, we deduce $S_x \cap S \subset E_x$. Hence

$$T_x S \subset T_x(S_x \cap S) + T_x(Gx) \subset T_x E_x + T_x(Gx),$$

which concludes the proof of (20). □

The following corollary may be deduced from the above proof.

Corollary 3.10. *M'/G is a set of isolated points, and each point of M'/G is a stratum in (4).*

Proof. Let $x \in M'$, and denote by S the stratum containing x. If the smooth submanifold Gx has dimension 0, then according to Lemma 3.9 there holds $T_x S = \{0\}$, therefore Gx is a stratum in (4), reduced to a point. If Gx has dimension ≥ 1, since $Gx \subset S$ and $x \in M'$, Lemma 3.9 implies that $T_x S = T_x(Gx)$. The slice theorem shows that if $z \in S_x \setminus \{x\}$ (where S_x denotes the slice at x, see proof of Lemma 3.9), then $G_z \subsetneq G_x$. Therefore $Gx \in M'/G$ is a stratum in (4).

As recalled at the beginning of Section 3.3, the number of strata is locally finite. Therefore, M'/G is a set of isolated points. □

We say that $X \in \mathrm{Vec}(M)$ is transverse to the G-orbit at x if $X(x) \notin T_x(Gx)$. Lemma 3.9 is useful to prove the following result.

Lemma 3.11. *Assume $k \geq 2$. Then there exists a residual set of k-uples $(X_1, \ldots, X_k) \in \mathcal{A}^k$ such that for any $x \in M \setminus M'$, at least one of the vectors $X_1(x), \ldots, X_k(x)$ is transverse to the G-orbit at x.*

Proof. In the proof of Lemma 3.9 we recalled the slice theorem. We also recall from [Fie07, Proposition 3.7.4] that if M is a compact G-manifold or a G-representation, then the number of isotropy types (i.e. different isotropy subgroups, up to conjugation) for the G-action is finite. It follows from the proof of [Fie07, Proposition 3.7.4] that even if M is not assumed compact, for any compact set $K \subset M$, the number of strata of M in the sense of the stratification (18) (in particular strata are connected sets) which have non-empty intersection with K is finite. This property is called Property **P** in the sequel.

We exhaust the open set $M \setminus M'$ by increasing compact sets M_j, $j \in \mathbb{N}$ (this is possible thanks to Whitney's embedding theorem) assumed to be G-invariant:

$$M = \bigcup_{j \in \mathbb{N}} M_j, \qquad \forall j \in \mathbb{N}, \, M_j \subset M_{j+1}, \qquad G M_j = M_j.$$

In the sequel $j \in \mathbb{N}$ is fixed. Let S be a stratum intersecting M_j. The set S/G is a smooth manifold, and we denote its dimension by $\ell(S)$. The dimension of $T_x(Gx)$ does not depend on $x \in S$ (because G_x, $G_{x'}$ are conjugate for $x, x' \in S$) and there holds $\ell(S) = \dim(T_x S / T_x(Gx))$ for any $x \in S$.

We consider for $x \in M$ the linear map $h_x : \mathcal{A}^k \to (T_x M)^k$

$$h_x : (X_1, \ldots, X_k) \mapsto (A_x X_1 \bmod(T_x(Gx)), \ldots, A_x X_k \bmod(T_x(Gx))).$$

---

[6]this is the main reason why we introduced the module $\mathcal{A}$: to be able to make cut-offs of analytic fields.

According to Lemma 3.9, for $x \in S$, the application h_x has rank $k\ell(S)$. The domain of h_x is infinite-dimensional, but to compute codimensions, we restrict h_x to a finite-dimensional space, while preserving its range: for any $x \in M_j$, we choose a finite dimensional subspace of $\mathcal{A}^k$ such that for any y in some open neighborhood of x the restriction of h_y to this subspace has same range as h_y. Covering the compact set M_j with a finite number of such open neighborhoods, we end-up with a finite dimensional subspace $F_j \subset \mathcal{A}$ such that the restriction $\tilde{h}_x := h_{x|F_j^k}$ has also rank $k\ell(S)$ for any $x \in M_j \cap S$. Its kernel $\tilde{h}_x^{-1}(0) \subset F_j^k$ has codimension $k\ell(S)$ for $x \in M_j \cap S$. The union $\bigcup_{x \in M_j \cap S} \tilde{h}_x^{-1}(0)$ is a subset of F_j^k of codimension $\geq (k-1)\ell(S)$. This quantity is ≥ 1 as soon as $\ell(S) \geq 1$ (since $k \geq 2$).

The above reasoning implies that for any $j \in \mathbb{N}$, for any stratum S having non-empty intersection with M_j, $\bigcup_{x \in M_j \cap S} h_x^{-1}(0)$ is a subset of codimension $\geq (k-1)\ell(S)$. Taking the union over the locally finite number (according to Property **P** above) of strata for which $\ell(S \cap M_j) \neq 0$, we obtain that the codimension of $\bigcup_{x \in M_j} h_x^{-1}(0)$ is ≥ 1. Therefore there exists a dense set of k-uples $(X_1, \ldots, X_k) \in \mathcal{A}^k$ on M such that for any $x \in M_j$ there exists $i \in \{1, \ldots, k\}$ having the property that $X_i(x)$ is transverse to the G-orbit at x. This set of k-uples is open since M_j is compact. Taking the intersection of these sets over $j \in \mathbb{N}$, we obtain Lemma 3.11. $\square$

Remark 3.12. *Assume G is finite. Since any orbit is discrete, a vector field X is transverse to the G-orbit at x if and only if $X(x) \neq 0$. Therefore, Lemma 3.11 means that for an open dense set of k-uples $(X_1, \ldots, X_k)$ of elements of $\mathcal{A}^k$, for any $x \in M \setminus M'$ at least one of the vectors $X_1(x), \ldots, X_k(x)$ is $\neq 0$.*

3.4. **Proof of Theorem 3.4.** Recall the notation

$$\mathrm{ad}_X^0 Y = Y, \qquad \mathrm{ad}_X^k Y = [X, \mathrm{ad}_X^{k-1} Y],$$

for any $k \geq 1$ and any vector fields X, Y. Our proof of Theorem 3.4 is based on the following lemma.

Lemma 3.13. *Let $\mathcal{K} \subset M$ be a G-invariant compact set. Let $X, Y \in \mathcal{A}$ and let $\mathcal{O} \subset \mathcal{K}$ be a tubular open set of the form*

$$\mathcal{O} = \bigsqcup_{t \in]-T, T[} e^{tX} \Sigma \tag{23}$$

where $\Sigma \subset M$ is a G-invariant hypersurface, X is transverse to Σ, and $T > 0$ is small enough so that (23) defines tubular coordinates in $\mathcal{O}$. There exists $p(\mathcal{K}) \in \mathbb{N}$ (depending only on $\mathcal{K}$) such that for any $p \geq p(\mathcal{K})$, any $\varepsilon > 0$ and any neighborhood of the closure $\overline{\mathcal{O}}$, there exists $Z \in \mathcal{A}$ supported in this neighborhood and with $\|Z\|_{C^p} \leq \varepsilon$ such that

$$\forall q \in \mathcal{O}, \quad \mathrm{Span}\left(\mathrm{ad}_X^0(Y + Z)(q), \ldots, \mathrm{ad}_X^p(Y + Z)(q)\right) = \mathrm{Vec}^G(\mathcal{O})_{|q}.$$

We postpone the proof of Lemma 3.13 to the end of this section, and first explain how to finish the proof of Theorem 3.15. In the C^∞ topology on compact sets (defined in Section 3.1), k-uples of vector fields satisfying (15) form a countable intersection of open sets. Therefore, we only need to prove their density.

Let us fix $(X_1, \ldots, X_k) \in (\mathrm{Vec}^G(M))^k$. By density of $\mathcal{A}$ in $\mathrm{Vec}^G(M)$ together with Lemma 3.11, we may assume that $(X_1, \ldots, X_k) \in \mathcal{A}^k$ satisfy the conclusion of Lemma 3.11, i.e., for any $x \in M \setminus M'$, at least one of the vectors $X_1(x), \ldots, X_k(x)$ is transverse to the G-orbit at x. Thanks to Lemma 3.11, we cover $M \setminus M'$ with G-invariant open sets $(\mathcal{O}_j)_{j \in \mathcal{J}}$ which are tubular neighborhoods of the form (23) for some hypersurface $\Sigma := \Sigma_j$ transverse to the G-orbits in $\mathcal{O}_j$, some $T := T_j \in \mathbb{R}^+$ and some $X := X_{i_j}$ where $i_j \in \{1, \ldots, k\}$ for any $j \in \mathcal{J}$. We may assume that this covering is locally finite, i.e., for any compact subset of $M \setminus M'$, the number of elements of this covering which intersect this compact set is finite, therefore $\mathcal{J} = \mathbb{N}$ or $\mathcal{J} = \{1, \ldots, J\}$ for some $J \in \mathbb{N}$. We also fix an increasing sequence $(\mathcal{K}_\ell)_{\ell \in \mathbb{N}}$ of compact sets such that $M = \bigcup_{\ell \in \mathbb{N}} \mathcal{K}_\ell$.

We modify the vector fields $X_1, \ldots, X_k$ inductively, for $j = 1, 2, \ldots$ ($j \in \mathcal{J}$). At the end of step $j \in \mathcal{J}$, we have the vector fields $X_1^{(j)}, \ldots, X_k^{(j)}$ and we ensure that:

(1) $X_1^{(j)}, \ldots, X_k^{(j)}$ satisfy

$$\forall q \in \bigcup_{j' \leq j} \mathcal{O}_{j'}, \qquad \mathrm{Lie}_q(X_1, \ldots, X_k) = \mathrm{Vec}^G(M)_{|q}.$$

(2) For any $j' \in \mathcal{J}$, $X_{i_{j'}}^{(j)}$ is transverse to $\Sigma_{j'}$.

Fix $j \in \mathcal{J}$ and assume that step $j - 1$ has been done. We pick $\alpha_j \in [k] \setminus \{i_j\}$ (arbitrarily). We modify $X_{\alpha_j}^{(j-1)}$ in a neighborhood of $\mathcal{O}_j$. The modification only affects X_{α_j}: it is of the form

$$X_{\alpha_j}^{(j)} = X_{\alpha_j}^{(j-1)} + \varphi_j Z_j \tag{24}$$
$$X_i^{(j)} = X_i^{(j-1)} \text{ for } i \neq \alpha_j.$$

In particular $X_{i_j}^{(j)} = X_{i_j}^{(j-1)}$. Let us explain the construction of φ_j and Z_j.

Let us fix an index ℓ such that $\mathcal{O}_j \subset \mathcal{K}_\ell$. We choose Z_j thanks to Lemma 3.13 such that for any $q \in \mathcal{O}_j$ the vectors

$$ad_{X_{i_j}^{(j)}}^k X_{\alpha_j}^{(j)}(q) = ad_{X_{i_j}^{(j-1)}}^k (X_{\alpha_j}^{(j-1)} + Z_j)(q), \qquad k = 0, \ldots, p$$

span $\mathrm{Vec}^G(M)_{|q}$. The regularity index p above only depends only on $\mathcal{K}_\ell$ but not on $\mathcal{O}_j$ according to Lemma 3.13. We also take φ_j a $C^\infty(M)$ and G-invariant cutoff function supported near $\mathcal{O}_j$, with value 1 in $\mathcal{O}_j$, and 0 outside a small neighborhood of $\mathcal{O}_j$. Moreover, we require the following properties:

- $\|Z_j\|_{C^p} \leq \varepsilon 2^{-j}$.
- (1) and (2) are satisfied.

The second bullet is guaranteed by taking $\|Z_j\|_{C^p}$ sufficiently small and the support of φ_j to be a sufficiently small neighborhood of the closure $\overline{\mathcal{O}}_j$. Here we use the fact that the covering of $M \setminus M'$ is locally finite, hence the transversality condition (2), which is an open condition, is perturbed only for a finite number of $j' \in \mathcal{J}$, and therefore remains true if the perturbation is sufficiently small.

Once j has run over $\mathcal{J}$, and at each step a perturbation of the form (24) has been added, we obtain modified vector fields which we denote by $X_1', \ldots, X_k' \in \mathrm{Vec}^G(M)$. Convergence of the series of modifications is guaranteed by the first bullet above, with $\|X_i' - X_i\|_{C^{p_\ell}(\mathcal{K}_\ell)} \leq \varepsilon$ for any $i \in [k]$ and for some p_ℓ depending only on $\mathcal{K}_\ell$. Moreover, the vector fields satisfy (15) thanks to (1). This concludes the proof of Theorem 3.15.

We finally prove Lemma 3.13.

Proof of Lemma 3.13. Any module generated by real-analytic vector fields is locally finitely generated, due to the Nötherian property of the ring of germs of real-analytic functions (see [Fri67, Theorem I.9]). Therefore, the module $\mathcal{A}_\mathcal{K}$ is locally finitely generated: there exist $m \in \mathbb{N}$ and analytic vector fields $f_1, \ldots, f_m$ on $\mathcal{K}$ such that

$$\mathcal{A}_\mathcal{K} = \left\{ \sum_{i=1}^m a_i f_i \mid \forall i \in [m], \ a_i \in C_G^\infty(\mathcal{K}) \right\}. \tag{25}$$

Of course, the analytic vector fields $f_1, \ldots, f_m$, when restricted to $\mathcal{O}$, also generate $\mathcal{A}_\mathcal{O}$. But it is important for us that the number m of vector fields depends only on $\mathcal{K}$ (not on $\mathcal{O} \subset \mathcal{K}$). Besides, $f_1, \ldots, f_m \in \mathcal{X}_{eq}^\omega(\mathcal{K})$ since all vector fields in $\mathcal{A}_\mathcal{K}$ are equivariant.

Since $\mathrm{ad}_X f_i \in \mathcal{A}_\mathcal{O}$ (according to (17)) and $\mathcal{A}_\mathcal{O}$ is finitely generated by the f_j,

$$\mathrm{ad}_X f_i = \sum_{j=1}^m a_{ij} f_j$$

for some $a_{ij} \in C_G^\infty(\mathcal{O})$, $1 \leq i, j \leq m$. We set

$$f_i^t := (e^{tX})_* f_i \tag{26}$$

and observe that

$$\frac{d}{dt}f_i^t = -\sum_{j=1}^{m}(a_{ij} \circ e^{-tX})f_j^t, \qquad f_i^0 = f_i. \tag{27}$$

Since the equation (27) is linear and the f_j verify (25), the f_j^t, $j = 1, \ldots, m$ verify

$$\mathcal{A}_{\mathcal{O}} = \left\{ \sum_{i=1}^{m} a_i f_i^t \mid \forall i \in [m], \ a_i \in C_G^\infty(\mathcal{O}) \right\} \tag{28}$$

for any $t \in]-T, T[$. For some sufficiently large $p \in \mathbb{N}$ and some matrix $\alpha \in \mathcal{M}_{m \times (p+1)}$, both to be chosen later, we consider for $i = 0, \ldots, p$

$$g_i = \sum_{j=1}^{m} \alpha_{ij} f_j. \tag{29}$$

Then $g_i^t := (e^{tX})_* g_i$ verify

$$g_i^t = \sum_{j=1}^{m} \alpha_{ij} f_j^t. \tag{30}$$

In the sequel, each point of $\mathcal{O}$ is written as $e^{tX}q_0$ where $t \in]-T, T[$ and $q_0 \in \Sigma$, thanks to (23). For such g_i's, we consider

$$Z(e^{tX}q_0) = (e^{tX})_* \left(\sum_{i=0}^{p} \frac{t^i}{i!} g_i(q_0) \right). \tag{31}$$

We want to compute $\mathrm{ad}_X^\gamma Z$ for $\gamma = 0, \ldots, p$. For this, we notice

$$[(e^{\varepsilon X})_* Z](e^{tX}q_0) = (e^{\varepsilon X})_* (e^{(t-\varepsilon)X})_* \left(\sum_{i=0}^{p} \frac{(t-\varepsilon)^i}{i!} g_i(q_0) \right) = \sum_{i=0}^{p} \frac{(t-\varepsilon)^i}{i!} g_i^t(e^{tX}q_0)$$

where for the second equality we used $g_i^t = (e^{tX})_* g_i$. We differentiate γ times with respect to ε at 0: we obtain

$$[\mathrm{ad}_X^\gamma Z](e^{tX}q_0) = (-1)^\gamma \frac{d^\gamma}{d\varepsilon^\gamma}_{|\varepsilon=0} \sum_{i=0}^{p} \frac{(t-\varepsilon)^i}{i!} g_i^t(e^{tX}q_0) = \sum_{i=\gamma}^{p} \frac{t^{i-\gamma}}{(i-\gamma)!} g_i^t(e^{tX}q_0).$$

Therefore writing $\mathrm{ad}_X^\gamma Y = \sum_{j=1}^{m} \beta_{\gamma j} f_j$ where $\beta = (\beta_{kj})$ is a $(p+1) \times m$ matrix, we obtain

$$\mathrm{ad}_X^\gamma (Y + Z)(e^{tX}q_0) = \sum_{j=1}^{m} \underbrace{\left(\beta_{\gamma j}(e^{tX}q_0) + \sum_{i=\gamma}^{p} \frac{t^{i-\gamma}}{(i-\gamma)!} \alpha_{ij} \right)}_{:=\delta_{\gamma j}} f_j^t(e^{tX}q_0) \tag{32}$$

thanks to (30). The goal is to choose the constant-coefficients matrix $\alpha = (\alpha_{ij})$ (with $0 \le \gamma \le p$ and $1 \le j \le m$) in a way that $\delta = (\delta_{\gamma j}) \in \mathcal{M}_{(p+1) \times m}$ defined in (32) has rank m at any point in $\mathcal{O}$. We notice that $\eta(t) \in \mathcal{M}_{(p+1) \times (p+1)}$ defined by its coefficients

$$\eta_{\gamma i}(t) = \mathbf{1}_{i \ge \gamma} \frac{t^{i-\gamma}}{(i-\gamma)!} \tag{33}$$

is a triangular matrix with non-zero diagonal coefficients, hence it is invertible. Thus, $\delta = \beta + \eta\alpha$ has rank m if and only if $\eta^{-1}\beta + \alpha$ has rank m. When $x = e^{tX}q_0$ varies in $\mathcal{O}$, $\eta(t)^{-1}\beta(e^{tX}q_0)$ describes a submanifold of $\mathcal{M}_{(p+1) \times m}$ of dimension $\le n = \dim(M)$. Hence, if $p + 1 \ge n + m$, then for α in a codimension $p + 2 - m - n \ge 1$ submanifold, δ has rank m at any point in the neighborhood.

Fix such α and take Z according to (29), (31). As a consequence of (28), (32) and the fact that δ has rank m, we have at any point $q \in \mathcal{O}$

$$\mathrm{Span}\left(\mathrm{ad}_X^0(Y+Z)(q), \ldots, \mathrm{ad}_X^p(Y+Z)(q) \right) = \mathrm{Span}(f_1^t(q), \ldots, f_m^t(q)) = \mathrm{Vec}^G(\mathcal{O})_{|q}.$$

Finally, recall that m depends only on $\mathcal{K}$, therefore $p(\mathcal{K}) := n + m - 1$ depends only on $\mathcal{K}$, which concludes the proof. $\qquad\square$

3.5. **Proof of Theorem 3.3 and an example.** Each equivariant leaf is by definition a submanifold of M. Let $X_1, \ldots, X_k$ be a tuple satisfying the conclusion of Theorem 3.4. We may apply Chow-Rashevskii's theorem [AS04, Theorems 5.1 and 5.2] in any equivariant leaf L. We obtain that for any $q, q' \in L$, there exist $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{R}$ and $i_1, \ldots, i_m \in [k]$ (not necessarily distinct) such that $q' = e^{t_1 X_{i_1}} \circ \ldots \circ e^{t_m X_{i_m}} q$. In other words, controllability in equivariant leaves holds, which concludes the proof of Theorem 3.3.

The following example illustrates the fact that controllability does not necessarily generically hold in strata of M given in (18). This is because strata are possibly larger sets than equivariant leaves.

Example 3.14. *Let $m, \ell \geq 1$ and take $M = \mathbb{S}^m \times \mathbb{R}^\ell$ equipped with the action of the orthogonal group $G = O(m+1)$ given by $O \cdot (a, b) = (Oa, b)$ for $(a, b) \in \mathbb{S}^m \times \mathbb{R}^\ell$. The equivariant vector fields are all vector fields that are tangent to $\mathbb{R}^\ell$ and do not depend on the point on the sphere. There is a single stratum for (18), equal to M, and controllability cannot hold in the whole stratum. Equivariant leaves are sets of the form $L_a = \{(a, x) \mid x \in \mathbb{R}^\ell\}$ for $a \in \mathbb{S}^m$, and controllability in equivariant leaves is possible (and generic according to Theorem 3.3). Also, controllability in $M/G = \mathbb{R}^\ell$ holds since C^∞ equivariant vector fields induce on $M/G \simeq \mathbb{R}^\ell$ all smooth fields.*

3.6. **Proof of Theorem 1.3.** Recall that the image of any stratum in (18) under quotient by G is a stratum in (4). Combining Theorem 3.4 with Lemma 3.9 and Corollary 3.10, we obtain:

Theorem 3.15. *For any $k \geq 2$, there exists a set of k-uples $(X_1, \ldots, X_k) \in (\mathrm{Vec}^G(M_G))^k$ which is residual in $(\mathrm{Vec}^G(M))^k$ and for which*

$$\forall q \in M_G, \quad \mathrm{Lie}_q(X_1, \ldots, X_k) = T_q S_G^i. \tag{34}$$

Here $i = i(q)$ denotes the index of the stratum containing q.

Theorem 1.3 follows from Theorem 3.15 and Chow-Rashevskii's theorem.

4. PROOF OF THEOREM 1.5

This section is devoted to the proof of Theorem 1.5. Since an intersection of residual sets is still a residual set, it is sufficient to prove the result for fixed N, i.e., simultaneous controllability in strata for any N points with N fixed. Let us fix $N \in \mathbb{N}$. For a vector field $X \in \mathrm{Vec}(M)$, consider its N-fold, defined on the product M^N as

$$X^N(x_1, \ldots, x_N) = (X(x_1), \ldots, X(x_N)).$$

For $X, Y \in \mathrm{Vec}(M)$ and $N \geq 1$ we define the Lie bracket of the N-folds X^N, Y^N on M^N "componentwise": $[X^N, Y^N] = [X, Y]^N$ where $[X, Y]$ is the Lie bracket of X, Y on M. The same holds for the iterated Lie brackets.

We set

$$M^{(N)} = \{(q_1, \ldots, q_N) \in (M \setminus M')^N \mid \forall i \neq j, \ Gq_i \cap Gq_j = \emptyset\}.$$

Given equivariant vector fields $X_1, \ldots, X_k$ on M, we say that their N-folds $X_1^N, \ldots, X_k^N$ form a bracket-generating system in equivariant leaves on $U \subset M^{(N)}$ if

$$\forall (q_1, \ldots, q_N) \in U, \quad \mathrm{Lie}_{(q_1, \ldots, q_N)}(X_1^N, \ldots, X_k^N) = \mathrm{Vec}^G(M)_{|q_1} \times \ldots \times Vec^G(M)_{|q_N}. \tag{35}$$

This equality is written for points in $U \subset M^{(N)}$ because it cannot hold in $M^N \setminus M^{(N)}$ due to equivariance. Notice that for $N > 1$, (35) for $U = M^{(N)}$ is strictly stronger than the property (15) (which corresponds to the case $N = 1$).

In this section we prove the following statement:

Theorem 4.1. *For any $N \geq 1$, there is a residual set of k-uples of vector fields $(X_1, \ldots, X_k)$ in $(\mathrm{Vec}^G(M))^k$, such that for any $(X_1, \ldots, X_k)$ from this set the N-folds $(X_1^N, \ldots, X_k^N)$ form a bracket generating system in equivariant leaves on $M^{(N)}$.*

Recall that the image of any stratum in (18) under quotient by G is a stratum in (4). Combining Theorem 4.1 with Lemma 3.9 and Corollary 3.10, we obtain the following result (where the definition of N-fold of elements of $\mathrm{Vec}^G(M_G)$ is deduced from N-fold of elements of $\mathrm{Vec}^G(M)$).

Theorem 4.2. *For any $k \geq 2$, there exists a set of k-uples $(X_1, \ldots, X_k) \in (\mathrm{Vec}^G(M_G))^k$ which is residual in $(\mathrm{Vec}^G(M_G))^k$ and for which for any distinct $q_1, \ldots, q_N \in M_G$,*

$$\mathrm{Lie}_{(q_1,\ldots,q_N)}(X_1^N, \ldots, X_k^N) = T_{q_1}S_G^{i(q_1)} \times \ldots \times T_{q_N}S_G^{i(q_N)}. \tag{36}$$

Here $i(q)$ denotes the index of the stratum in (4) containing $q \in M_G$.

Theorem 1.5 is a direct consequence of Theorem 4.2 combined with the Chow-Rashevskii theorem, as in [AS20, Proposition 3.1]. The condition in Definition 1.4 that the stratum has dimension ≥ 2 is necessary because in dimension 1, ordering of points is preserved (and indeed, $q_1, \ldots, q_N$ are assumed distinct in (36), therefore they cannot cross).

The rest of this section is devoted to the proof of Theorem 4.1. Since the proof consists in a slight modification of the proof of Theorem 3.4, we only provide the key ideas and highlight the modifications compared to the proof of Theorem 3.4.

We write $M^{(N)}$ as a union of compact sets $M_s^{(N)}$, $s \in \mathbb{N}$, which are invariant under the action of G^N given by

$$(g_1, \ldots, g_N) \cdot (q_1, \ldots, q_N) = (g_1 q_1, \ldots, g_N q_N).$$

It is sufficient to prove Theorem 4.1 in $M_s^{(N)}$ for fixed $s \in \mathbb{N}$, instead of $M^{(N)}$, since taking intersection over s of the sets of vector fields will yield a countable intersection of residual sets, which is still a residual set. Therefore, we fix $s \in \mathbb{N}$ and we prove that there exists $p_s \in \mathbb{N}$ such that

> the set of k-uples of vector fields $(X_1, \ldots, X_k)$ in $(\mathrm{Vec}^G(M))^k$, such that the length $\leq p_s$ brackets of the N-folds $(X_1^N, \ldots, X_k^N)$ generate in $M_s^{(N)}$ all equivariant leaves is open and dense in $(\mathrm{Vec}^G(M))^k$.

Moreover, for any compact set $\mathcal{K} \subset M$, the regularity index p_s in the above statement may be taken the same for all s such that $M_s^{(N)} \subset \mathcal{K}$.

Openness in the above statement is immediate since $M_s^{(N)}$ is compact. Therefore, we only need to prove the density.

Let us fix $(X_1, \ldots, X_k) \in (\mathrm{Vec}^G(M))^k$. By density of $\mathcal{A}$ in $\mathrm{Vec}^G(M)$ together with Lemma 3.11, we may assume that $(X_1, \ldots, X_k) \in \mathcal{A}^k$ satisfy the conclusion of Lemma 3.11, i.e.,

$$\forall x \in M \setminus M', \exists i \in [k] \text{ such that } X_i(x) \text{ is transverse to the } G\text{-orbit at } x. \tag{37}$$

For each $(q_1, \ldots, q_N) \in M_s^{(N)}$ we do the following. Thanks to (37) we fix a linear combination $a_1 X_1 + \ldots + a_k X_k$ which is transverse to the G-orbit at q_i for any $i \in [N]$. Then we consider a product of tubular neighborhoods

$$\mathcal{O}_1 \times \ldots \times \mathcal{O}_N \subset M^{(N)} \tag{38}$$

where $q_\ell \in \mathcal{O}_\ell$, each $\mathcal{O}_\ell$ is of the form (23), and the $\mathcal{O}_\ell$'s have empty intersection, which is possible since $q_1, \ldots, q_N$ are pairwise distinct. We assume that the neighborhood (38) is sufficiently small so that the linear combination $a_1 X_1 + \ldots + a_k X_k$ is transverse in $\mathcal{O}_1 \cup \ldots \cup \mathcal{O}_N$ to the G-orbits.

Doing this for any $(q_1, \ldots, q_N) \in M_s^{(N)}$, we have obtained an open covering of the compact set $M_s^{(N)}$, from which we select a finite sub-covering $\widetilde{V}_1, \ldots, \widetilde{V}_{j_0}$. For fixed $j \in [j_0]$, we have by definition

$$\widetilde{V}_j = \mathcal{O}_1^{(j)} \times \ldots \times \mathcal{O}_N^{(j)} \tag{39}$$

where each $\mathcal{O}_\ell^{(j)}$ is of the form (23) (with $t \in]-T_\ell^{(j)}, T_\ell^{(j)}[$ and $\Sigma_\ell^{(j)}$ denotes the hypersurface), and $\mathcal{O}_\ell^{(j)}, \mathcal{O}_{\ell'}^{(j)}$ are separated ("at positive distance") for any $\ell \neq \ell'$. For any $j \in [j_0]$, there exist

$a_1^{(j)}, \ldots, a_k^{(j)} \in \mathbb{R}$ such that $a_1^{(j)} X_1 + \ldots + a_k^{(j)} X_k$ is transverse to the G-orbits in

$$V_j = \bigsqcup_{\ell=1}^N \mathcal{O}_\ell^{(j)}. \tag{40}$$

(Notice that the V_j's do not necessarily have empty intersection.)

We perturb $(X_1, \ldots, X_k)$ in $V_1, \ldots, V_{j_0}$ successively; at step j we perturb the vector fields in V_j, and the vector fields which we obtain at the end of this step are denoted by $(X_1^{(j)}, \ldots, X_k^{(j)})$.

We need to modify the construction made in the proof of Lemma 3.13. Fix a step $j \in [j_0]$. We set $X = a_1^{(j)} X_1 + \ldots + a_k^{(j)} X_k$ and

$$Y = \begin{cases} X_1 \text{ if } (a_1^{(j)}, \ldots, a_k^{(j)}) \neq (\lambda, 0, \ldots, 0) \text{ for any } \lambda \in \mathbb{R} \\ X_2 \text{ otherwise.} \end{cases}$$

We construct below a perturbation Z of Y, supported near V_j. We use the same idea of Taylor expansion as in (31), just taking larger p in order to generate all directions in the tangent space at any $(e^{t_1 X} q_1, \ldots, e^{t_N X} q_N)$. We introduce the set $\mathcal{C}$ of parameters $(\mathbf{t}, \mathbf{q})$ such that $\mathbf{t} = (t_1, \ldots, t_N)$ with $t_\ell \in]-T_\ell^{(j)}, T_\ell^{(j)}[$, and $\mathbf{q} = (q_1, \ldots, q_N) \in \Sigma_1^{(j)} \times \ldots \times \Sigma_N^{(j)}$.

The vector field Z is constructed in the following way. We pick analytic vector fields $f_1, \ldots, f_m$ as in the proof of Lemma 3.13, in order for (25) to hold. We follow (26) to (30) in all $\mathcal{O} = \mathcal{O}_\ell^{(j)}$, $i = 1, \ldots, N$, and define Z through (31): for each $\ell \in [N]$ there is a matrix $\alpha = \alpha^{(\ell)}$, which defines Z in $\mathcal{O}_\ell^{(j)}$. The vector field Z is thus supported near V_j. An analogous formula to (32) holds:

$$\operatorname{ad}_X^\gamma (Y + Z)(e^{t_\ell X} q_\ell) = \sum_{r=1}^m \underbrace{\left(\beta_{\gamma r}^{(\ell)}(e^{t_\ell X} q_\ell) + \sum_{i=\gamma}^p \frac{t^{i-\gamma}}{(i-\gamma)!} \alpha_{ir}^{(\ell)} \right)}_{:= \delta_{\gamma r}^{(\ell)}} f_r^t(e^{t_\ell X} q_\ell) \tag{41}$$

The goal is to choose the constant-coefficients matrices $\alpha^{(\ell)} = (\alpha_{ir}^{(\ell)})$ (with $0 \leq \gamma \leq p$ and $r \in [m]$, $\ell \in [N]$) in a way that the $N(p+1) \times Nm$ block matrix

$$\delta' = \begin{pmatrix} \delta^{(1)} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \delta^{(2)} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \delta^{(N)} \end{pmatrix}$$

(whose N^2 blocks are of size $(p+1) \times m$) has rank Nm for any $(\mathbf{t}, \mathbf{q}) \in \mathcal{C}$. We denote by η' the $N(p+1) \times N(p+1)$ matrix defined by N^2 blocks of size $(p+1) \times (p+1)$, all equal to 0 except the on-diagonal ones taken equal to the matrix η introduced in (33). We define similarly β' and α' as $N(p+1) \times Nm$ block matrices having blocks equal to 0 except ℓ^{th} diagonal block equal respectively to $\beta^{(\ell)}$ and $\alpha^{(\ell)}$ which appear in (41). We obtain $\delta' = \beta' + \eta' \alpha'$.

Since η' is invertible, δ' has rank Nm if and only if $\eta'^{-1}\beta' + \alpha'$ has rank Nm. When $(\mathbf{t}, \mathbf{q})$ vary in $\mathcal{C}$, $\eta'^{-1}\beta'$ describes a submanifold of $\mathcal{M}_{(p+1)N \times mN}$ of dimension $\leq nN$, where $n = \dim(M)$. Hence, if $p + 1 \geq (m + n)N$, then for α in a codimension $p + 2 - (m + n)N \geq 1$ submanifold, δ' has rank mN at any point in the neighborhood.

For α' and Z taken in this way, we obtain that the Lie algebra generated by the (restrictions of the) N-folds vector fields $X_1^N, \ldots, X_k^N$ equals $\operatorname{Vec}^G(\mathcal{O}_1) \times \ldots \times \operatorname{Vec}^G(\mathcal{O}_N)$ - this is a direct consequence of the fact that δ' has rank Nm together with (28).

Besides, we take the perturbations α' and Z sufficiently small so that the condition that $a_1^{(j')} X_1 + \ldots + a_k^{(j')} X_k$ is transverse to the G-orbits in $V_{j'}$ for any $j' \in [j_0]$ is preserved.

By definition, if we set $X_1' = X_{1,j_0}$ and $X_2' = X_{2,j_0}$, then their N-folds $X_1'^{(N)}$ and $X_2'^{(N)}$ form a bracket-generating system in equivariant leaves in all V_j ($j \in [j_0]$), hence in $M_s^{(N)}$. This concludes the proof of Theorem 4.1.

References

[AS04] A. Agrachev and Y. Sachkov. *Control theory from the geometric viewpoint*, Vol. 2. Springer Science & Business Media, 2004.

[AS20] A. Agrachev and A. Sarychev, Control in the spaces of ensembles of points, *SIAM J. Control Optim.*, 58.3, 1579-1596, 2020.

[ADTK23] S. Alberti, N. Dern, L. Thesing and G. Kutyniok, Universal Approximation for Efficient Transformers, *In: Topological, Algebraic and Geometric Learning Workshops, PMLR*, 72-86, 2023.

[AD18] F. Albertini, D. D'Alessandro, Controllability of symmetric spin networks, *J. Math. Phys.*, 59.5, 2018.

[BMR08] S. Bonnabel, P. Martin and P. Rouchon, Symmetry-preserving observers, *IEEE Trans. Autom. Control.*, 53.11, 2514-2526, 2008.

[BMR09] S. Bonnabel, P. Martin and P. Rouchon, Non-linear symmetry-preserving observers on Lie groups, *IEEE Trans. Autom. Control.*, 54.7, 1709-1713, 2009.

[BBLSV17] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, Geometric deep learning: going beyond euclidean data, *IEEE Signal Processing Magazine*, 34.4, 18-42, 2017.

[CZDP17] J. Chen, H. Zhou, C. Duan and X. Peng, Preparing Greenberger-Horne-Zeilinger and W states on a long-range Ising spin model by global controls, *Physical Review A*, 95.3, 032340, 2017.

[CLLS23] J. Cheng, Q. Li, T. Lin and Z. Shen, Interpolation, approximation and controllability of deep neural networks, *arXiv preprint arXiv:2309.06015*, 2023.

[CLT20] C. Cuchiero, M. Larsson and J. Teichmann, Deep neural networks, generic universal interpolation, and controlled ODEs, *SIAM J. Math. Data Sci.*, 2.3, 901-919, 2020.

[Dal23] D. D'Alessandro, Subspace controllability and Clebsch-Gordan decomposition of symmetric quantum networks, *arXiv preprint arXiv:2307.12908*, 2023.

[Fie07] M. Field, *Dynamics and symmetry*, Vol. 3. World Scientific, 2007.

[Fri67] J. Frisch, Points de platitude d'un morphisme d'espaces analytiques complexes, *Inventiones mathematicae*, 4, 118-138, 1967.

[G+23] J. E. Gerken, J. Aronsson, O. Carlsson, H. Linander, F. Ohlsson, C. Petersson and D. Persson, Geometric deep learning and equivariant neural networks, *Artificial Intelligence Review*, 56.12, 14605-14662, 2023.

[GLPR23a] B. Geshkovski, C. Letrouit, Y. Polyanskiy and P. Rigollet, The emergence of clusters in self-attention dynamics, *Advances in Neural Information Processing Systems*, 36, 2023.

[GLPR23b] B. Geshkovski, C. Letrouit, Y. Polyanskiy and P. Rigollet, A mathematical perspective on Transformers, *arXiv preprint arXiv:2312.10794*, 2023.

[KP19] N. Keriven and G. Peyré, Universal invariant and equivariant graph neural networks, *Advances in Neural Information Processing Systems*, 32, 2019.

[Llo95] S. Lloyd, Almost any quantum logic gate is universal, *Physical review letters*, 75(2), 346, 1995.

[Lob72] C. Lobry, Une propriété générique des couples de champs de vecteurs, *Czechoslov. Math. J.*, 22.2, 230-237,1972.

[LPM15] M. T. Luong, H. Pham & C. D. Manning, Effective approaches to attention-based neural machine translation, *In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 1412-1421, 2015.

[MGH22] R. Mahony, P. van Goor, and T. Hamel, Observer design for nonlinear systems with equivariance, *Annual Review of Control, Robotics, and Autonomous Systems*, 5, 221-252, 2022.

[Mar70] J. Martinet, Sur les singularités des formes différentielles, *Ann. de l'Institut Fourier*, 95-178, 1970.

[SABP22] M. E. Sander, P. Ablin, M. Blondel and G. Peyré, Sinkformers: Transformers with doubly stochastic attention, *International Conference on Artificial Intelligence and Statistics*, PMLR. 3515-3530 2022.

[Sca23] A. Scagliotti, Optimal control of ensembles of dynamical systems, *ESAIM: Control, Optimisation and Calculus of Variations*, 29(22), 2023.

[TG22] P. Tabuada, and B. Gharesifard, Universal approximation power of deep residual neural networks through the lens of control, *IEEE Trans. Autom. Control.*, 2022.

[Tym06] D. Tymoczko, The geometry of musical chords, *Science* 313.5783, 72-74, 2006.

[V+17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, 30, 2017.

[VW29] J. von Neumann and E. P. Wigner, Über das Verhalten von Eigenwerten bei adiabatischen Prozessen, *Physikalische Zeitschrift* 30, 467-470, 1929.

[VBT20] J. Vuckovic, A. Baratin and R. Tachet des Combes, A mathematical theory of attention, *arXiv preprint arXiv:2007.02876* (2020).

[YBRRK19] C. Yun, S. Bhojanapalli, A.S. Rawat, S. Reddi and S. Kumar. Are Transformers universal approximators of sequence-to-sequence functions? *International Conference on Learning Representations*, 2019.

Andrei Agrachev. SISSA, via Bonomea 265, 34136 Trieste, Italy
Email address: agrachevaa@gmail.com

Cyril Letrouit. Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, Université Paris-Saclay, Bâtiment 307, 91405 Orsay Cedex & CNRS UMR 8628
Email address: cyril.letrouit@universite-paris-saclay.fr