Journal of Geometric Mechanics
12(2), 165–308, 2020
DOI: 10.3934/jgm.2020013

Nonholonomic and constrained variational mechanics

Andrew D. Lewis*

2018/12/24
Last updated: 2025/04/16

Abstract

Equations governing mechanical systems with nonholonomic constraints can be developed in two ways: (1) using the physical principles of Newtonian mechanics; (2) using a constrained variational principle. Generally, the two sets of resulting equations are not equivalent. While mechanics arises from the first of these methods, sub-Riemannian geometry is a special case of the second. Thus both sets of equations are of independent interest.

The equations in both cases are carefully derived using a novel Sobolev analysis where infinite-dimensional Hilbert manifolds are replaced with infinite-dimensional Hilbert spaces for the purposes of analysis. A useful representation of these equations is given using the so-called constrained connection derived from the system's Riemannian metric, and the constraint distribution and its orthogonal complement. In the special case of sub-Riemannian geometry, some observations are made about the affine connection formulation of the equations for extremals.

Using the affine connection formulation of the equations, the physical and variational equations are compared and conditions are given that characterise when all physical solutions arise as extremals in the variational formulation. The characterisation is complete in the real analytic case, while in the smooth case a locally constant rank assumption must be made. The main construction is that of the largest affine subbundle variety of a subbundle that is invariant under the flow of an affine vector field on the total space of a vector bundle.

Keywords. Nonholonomic mechanics, calculus of variations, affine differential geometry, Sobolev spaces of mappings, linear and affine vector fields on vector bundles, invariant subbundles, sub-Riemannian geometry

AMS Subject Classifications (2020). 34C45, 46E35, 49K05, 53B05, 53C05, 53C17, 58A30, 70F25, 70G45

Contents

*Professor, Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada
Email: andrew.lewis@queensu.ca, URL: http://www.mast.queensu.ca/~andrew/
Research supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada

1. Introduction

For mechanical systems not subject to nonholonomic constraints—sometimes called "holonomic mechanical systems"—it is well-known that the physical motions are the extremals for a problem in the calculus of variations involving a physically meaningful Lagrangian. It is well-known that this property of physical motions breaks down when the system is subject to nonholonomic constraints. There *is* a natural problem in the calculus of variations—a problem with constraints—that one can associate to mechanical systems with nonholonomic constraints; it is just that the extremals from the calculus of variations problem are not generally physical motions. In the case when the conservative forces are absent, this calculus of variations problem is, however, equivalent to the determination of extremals in sub-Riemannian geometry. This, therefore, gives rise to two natural sets of governing equations associated with the data that describe a mechanical system subject to nonholonomic constraints, one physical and one variational, and both interesting in their own right.

There has been a literature devoted to the comparison of the two sorts of equations describing constrained motion, and the modern take on this seems to originate with papers of Kozlov [1992] and Kharlomov [1992]. Other work includes [Borisov, Mamaev, and Bizyaev 2017, Cardin and Favretti 1996, Favretti 1998, Gràcia, Marin-Solano, and Muñoz-Lecanda 2003, Kupka and Oliva 2001, Lewis and Murray 1995, Vershik and Gershkovich 1990, Zampieri 2000]. Our interest is in providing a characterisation of those physical motions that also arise as extremals for the constrained variational problem. Some work has been done on this problem [e.g., Cortés, de León, Martín de Diego, and Martínez 2002, Crampin and Mestdag 2010, Favretti 1998, Fernandez and Bloch 2008, Jóźwikowski and Respondek 2019, Rumiantsev 1978, Terra 2018]; we refer to the introduction of [Jóźwikowski and Respondek 2019] for a nice review of the literature on this topic. Our approach and conclusions differ from what presently exists in the literature. While much (but not all) of the existing literature considers general Lagrangians, we work exclusively with kinetic energy minus potential energy Lagrangians. This allows us to take advantage of the geometric structure of such systems. Also, most of the work in the literature derives certain sufficient conditions, sometimes involving additional system structure, that allows one to conclude that all physical trajectories are also constrained variational trajectories. While the work of Cortés, de León, Martín de Diego, and Martínez [2002] in principle offers a complete resolution to the problem of when all physical trajectories are also constrained variational trajectories, this resolution comes in the form of an iterative "algorithm" which requires certain regularity conditions and which offers very little insight as to just when the algorithm yields an affirmative answer. Also, Cortés, de León, Martín de Diego, and Martínez do not consider singular trajectories. By contrast, we are able here to offer a complete resolution to the comparison problem for the most important class of Lagrangians in the real analytic case,[1] while our results for the smooth case give sufficient conditions and require assumptions of the locally constant rank of some subbundles. Moreover, our results are of an insightful nature in multiple ways, connecting the detailed geometry of the interaction of the constraint distribution and the Riemannian metric defining the kinetic energy.

[1] Of course, the real analytic case has the most relevance to physics.

1.1. Contribution of paper. We restrict ourselves to "kinetic energy minus potential energy" Lagrangians, and characterise in Section 7 the comparison of solutions of the two constrained problems using the interaction of the Levi-Civita affine connection and the distribution. The end result of our detailed constructions is an affine vector field that describes the evolution of the adjoint variable (i.e., the Lagrange multiplier) for the constrained variational problem, and it is this vector field that allows us to nicely characterise cases where nonholonomic trajectories are also constrained variational trajectories. The most interesting of our results can be seen as analogous to the following question from linear algebra:

> *Let* V *be a finite-dimensional* $\mathbb{R}$*-vector space, let* $\mathsf{U} \subseteq \mathsf{V}$ *be a subspace, let* $A \in \mathrm{End}(\mathsf{V})$*, and let* $b \in \mathsf{V}$*. Determine all solutions to the problem*
>
> $$\begin{aligned} \dot{x}(t) &= A(x(t)) + b, \\ x(t) &\in \mathsf{U}. \end{aligned}$$

Of course, there are some technicalities that distinguish our problem from this simple one, but this simple problem is useful to keep in mind.

Another objective of our presentation is to develop, in a simple context, a methodology for doing Sobolev-type nonlinear analysis on manifolds; this is given in Section 3. For the setting of this paper, the analysis involves the space of curves on a manifold. A typical technique for doing this type of analysis is to develop the structure of an infinite-dimensional Hilbert manifold for the space of curves [e.g., Klingenberg 1995, Kupka and Oliva 2001, Terra and Kobayashi 2004a, Terra and Kobayashi 2004b]. This type of analysis has the benefit that, once one has at hand the manifold structure, all of the standard tools of differential geometry are made available. The drawback of the methodology is that the infinite-dimensional manifold structure can be difficult to work with. The approach we develop in this paper is that, given finite-dimensional manifolds M and N, one can replace a single mapping $\Phi\colon \mathsf{M} \to \mathsf{N}$ with the family of functions $f \circ \Phi\colon \mathsf{M} \to \mathbb{R}$, one for each smooth function $f\colon \mathsf{N} \to \mathbb{R}$. By taking this point of view, one works, not with the space of mappings which does not have a vector space structure, but with the space of functions which does have a vector space structure. Indeed, we are able to do all of the analysis we need in the paper while working explicitly only with the space the space $\mathrm{H}^1([t_0, t_1]; \mathbb{R})$ of absolutely continuous functions on the interval $[t_0, t_1]$ that are square integrable with square integrable derivative. This is a point of view that has been explored in a variety of ways in a variety of settings. For example, the replacement of the nonlinear manifold structure with the linear structure of its space of functions is a device reminiscent of algebraic geometry, and gives rise to a sort of "algebraic analysis" that is explored for smooth differential geometry, for example, in the book of Nestruev [2003]. Agrachev and Gamkrelidze [1978] use this idea of function evaluations as the basis for their "chronological calculus" used to study flows of vector fields. These ideas are further explored by Jafarpour and Lewis [2014], and indeed this latter work, combined with our modest undertakings here, can be used as a basis for a comprehensive methodology for Sobolev-type analysis on manifolds. Some explorations along these lines have been carried out by Convent and Van Schaftingen [2016a, 2016b, 2019]. In this paper, we make use of these ideas to characterise spaces of curves that satisfy a linear velocity constraint and/or endpoint constraints. We reproduce in our Section 5.1

the results of Kupka and Oliva [2001, §5], while only using elementary methods (the proofs themselves are not necessarily trivial, mind).

Another novel feature of our presentation is the development in Section 4.1 of some results for the invariance of subsets, not generally submanifolds, of a manifold under the flow of a vector field. Of special interest is the situation where the subset is a (not necessarily locally constant rank) subbundle of a vector bundle and where the vector field has some interesting structure relative to the vector bundle structure, e.g., linear or affine. We give a useful infinitesimal characterisation for the invariance of such subbundles under such vector fields in Sections 4.2, 4.3, 4.4, and 4.5. Using these constructions, in Section 4.6 we are able to build the "largest invariant affine subbundle variety contained in a subbundle." This construction plays a crucial rôle in our comparison results of Section 7.

1.2. An outline of the paper. In Section 2 we overview some constructions and notation concerning vector bundles, subbundles and affine subbundles, connections in vector bundles, vector fields on the total space of a vector bundle, Riemannian geometry, and the geometry of subbundles of the tangent bundle. In Section 3 we develop our methodology for the nonlinear analysis that we will use to deduce the two sets of equations of motion which we will ultimately compare. Results concerning subbundles and affine subbundles invariant under linear and affine vector fields in vector bundles are developed in Section 4. The equations governing nonholonomic mechanics and constrained variational mechanics are developed in Section 5. The equations we produce are those derived by Kupka and Oliva [2001], but we do this a little more comprehensively than do Kupka and Oliva, filling in some gaps in their written arguments, correcting some confusing typography, and banishing the use of coordinates. We also cast the equations in a new way using constructions from Section 2.11 involving affine connections adapted to distributions. It is these new representations of the governing equations that makes possible a systematic and comprehensive comparison of nonholonomic mechanics and constrained variational mechanics. In Section 6 we make some connections between constrained variational mechanics and sub-Riemannian geometry. The affine connection formalism we use here provides some new tools for problems in sub-Riemannian geometry, where the Hamiltonian approach mainly prevails in the current literature. In Section 7 we present the main new results of the paper, which are this comprehensive comparison of nonholonomic mechanics with constrained variational mechanics. We point out how we can encompass existing results, in cases when this is easily done.

1.3. Background and notation. We use standard set theoretic terminology, with the possible exception that we use "$\subseteq$" to denote the inclusion of a set in another, and use "$\subset$" to denote strict inclusion. By id_X we denote the identity map on a set X. If $f\colon X \to Y$ is a map of sets and if $A \subseteq X$, we denote by $f|A$ the restriction of f to A. For sets $X_1, \dots, X_k$, we denote by

$$\mathrm{pr}_j\colon X_1 \times \cdots \times X_k \to X_j, \qquad j \in \{1, \dots, k\},$$

the projections. By $\mathbb{Z}$ we denote the set of integers, while $\mathbb{Z}_{>0}$ and $\mathbb{Z}_{\geq 0}$ denote the sets of positive and nonnegative integers. By $\mathbb{R}$ we denote the set of real numbers, while $\mathbb{R}_{>0}$ denotes the set of positive real numbers.

By $\mathbb{R}^{m\times n}$ we denote the set of $m \times n$ matrices with real entries. The rank of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ we denote by $\mathrm{rank}(\boldsymbol{A})$. The $n \times n$ identity matrix we denote by $\boldsymbol{I}_n$.

An affine subspace of a $\mathbb{R}$-vector space V is a subset A such that $sv_1 + (1-s)v_2 \in \mathsf{A}$ for every $v_1, v_2 \in \mathsf{A}$ and $s \in \mathbb{R}$. We denote by $L(\mathsf{A})$ the linear part of A defined by

$$L(\mathsf{A}) = \{v - v_0 \mid v \in \mathsf{A}\}$$

for some $v_0 \in \mathsf{A}$. We refer to [Berger 1987, Chapter 2] for background on affine spaces. For a $\mathbb{R}$-vector space V and for $S \subseteq \mathsf{V}$, we denote by $\operatorname{span}_{\mathbb{R}}(S)$ the smallest subspace of V containing S and by $\operatorname{aff}_{\mathbb{R}}(S)$ the smallest affine subspace of V containing S.

For $\mathbb{R}$-vector spaces U and V, $\mathrm{L}(\mathsf{U};\mathsf{V})$ denotes the set of linear mappings from U to V. By $\mathsf{V}^* = \mathrm{L}(\mathsf{V};\mathbb{R})$ we denote the algebraic dual of V. For $A \in \mathrm{L}(\mathsf{U};\mathsf{V})$, we denote by $A^* \in \mathrm{L}(\mathsf{V}^*;\mathsf{U}^*)$ the algebraic dual. The pairing of $\alpha \in \mathsf{V}^*$ with $v \in \mathsf{V}$ will be denoted by one of

$$\alpha(v),\ \ \alpha \cdot v,\ \ \langle \alpha; v\rangle,$$

whichever seems most aesthetically pleasing in the moment. By $\mathrm{T}^r_s(\mathsf{V})$ we denote the r-contravariant and s-covariant tensors on V, i.e.,

$$\mathrm{T}^r_s(\mathsf{V}) = \underbrace{\mathsf{V} \otimes \cdots \otimes \mathsf{V}}_{r \text{ times}} \otimes \underbrace{\mathsf{V}^* \otimes \dots \otimes \mathsf{V}^*}_{s \text{ times}}.$$

We denote by $\operatorname{End}(\mathsf{V})$ the endomorphisms of V, i.e., $\operatorname{End}(\mathsf{V}) = \mathrm{L}(\mathsf{V};\mathsf{V})$. By $\bigwedge^k(\mathsf{V}^*)$ and $\mathrm{S}^k(\mathsf{V}^*)$ we denote the k-fold alternating and symmetric tensors on V, respectively.

If A is a $(0,2)$-tensor and B is a $(2,0)$-tensor on a finite-dimensional $\mathbb{R}$-vector space V, we denote by

$$A^\flat\colon \mathsf{V} \to \mathsf{V}^*,\ \ B^\sharp\colon \mathsf{V}^* \to \mathsf{V}$$

the mappings defined by

$$\langle A^\flat(u); v\rangle = A(v,u),\ \ \langle \alpha; B^\sharp(\beta)\rangle = B(\alpha,\beta), \qquad u, v \in \mathsf{V},\ \alpha, \beta \in \mathsf{V}^*.$$

If $(\mathsf{V}, \mathbb{G})$ is a $\mathbb{R}$-inner product space and if $S \subseteq \mathsf{V}$, we denote by $S^\perp$ the subspace orthogonal to S.

For a topological space $\mathcal{X}$ and for $A \subseteq \mathcal{X}$, we denote by $\operatorname{int}(A)$, $\operatorname{cl}(A)$, and $\operatorname{bd}(A)$ the interior, closure, and boundary of A, respectively.

By $\mathsf{B}(r, \boldsymbol{x}) \subseteq \mathbb{R}^n$ we denote the open ball of radius r and centre $\boldsymbol{x}$.

For Banach spaces E and F, an open set $\mathcal{U} \subseteq \mathsf{E}$, and a mapping $\Phi\colon \mathcal{U} \to \mathsf{F}$, we denote by $\boldsymbol{D}\Phi(u)\colon \mathsf{E} \to \mathsf{F}$ the Fréchet derivative of Φ at u, when this exists.

We shall be concerned with functions of two variables, $(s,t) \mapsto f(s,t)$. For such functions, we have the partial derivatives

$$\begin{aligned}
\partial_1 f(s,t) &= \lim_{h\to 0} \frac{f(s+h,t) - f(s,t)}{h},\\
\partial_2 f(s,t) &= \lim_{h\to 0} \frac{f(s,t+h) - f(s,t)}{h},\\
\partial_1\partial_2 f(s,t) &= \lim_{h\to 0} \frac{\partial_2 f(s+h,t) - \partial_2 f(s,t)}{h},\\
\partial_2\partial_1 f(s,t) &= \lim_{h\to 0} \frac{\partial_1 f(s,t+h) - \partial_1 f(s,t)}{h},
\end{aligned}$$

defined when the limits exist.

For an interval $I \subseteq \mathbb{R}$, $A \subseteq \mathbb{R}$, and for $p \in [1, \infty)$, we denote by $\mathrm{L}^p(I; A)$ the set of measurable A-valued functions f on I for which

$$\int_I |f(t)|^p \, \mathrm{d}t < \infty.$$

The norm on $\mathrm{L}^p(I; \mathbb{R})$ we denote by

$$\|f\|_{\mathrm{L}^p} = \left(\int_I |f(t)|^p \, \mathrm{d}t \right)^{1/p}.$$

For $s \in \mathbb{Z}_{\geq 0}$, by $\mathrm{H}^s(I; \mathbb{R})$ we denote the set of measurable functions whose first s distributional derivatives are in $\mathrm{L}^2([t_0, t_1]; \mathbb{R})$. We denote the norm on $\mathrm{H}^s(I; \mathbb{R})$ by

$$\|f\|_{\mathrm{H}^s} = \sum_{a=0}^{s} \|f^{(a)}\|_{\mathrm{L}^2},$$

$f^{(a)}$ being the ath derivative of f. Of course, $\mathrm{H}^0([t_0, t_1]; \mathbb{R}) = \mathrm{L}^2([t_0, t_1]; \mathbb{R})$.

Our geometric notation mainly follows [Abraham, Marsden, and Ratiu 1988]. Manifolds will be assumed to be smooth, Hausdorff, and paracompact. We shall at some crucial points require real analyticity of the manifolds and geometric objects we use. To cover the smooth and real analytic cases, we shall allow the regularity classes $r \in \{\infty, \omega\}$, $r = \infty$ being the smooth case and $r = \omega$ being the real analytic case. For a manifold M, the tangent bundle is denoted by $\pi_{\mathsf{TM}} \colon \mathsf{TM} \to \mathsf{M}$ and the cotangent bundle is denoted by $\pi_{\mathsf{T}^*\mathsf{M}} \colon \mathsf{T}^*\mathsf{M} \to \mathsf{M}$. The set of C^r-mappings from a manifold M to a manifold N is denoted by $\mathrm{C}^r(\mathsf{M}; \mathsf{N})$. We abbreviate $\mathrm{C}^r(\mathsf{M}) = \mathrm{C}^r(\mathsf{M}; \mathbb{R})$.

For a C^r-vector bundle $\pi \colon \mathsf{E} \to \mathsf{M}$, we denote the fibre at $x \in \mathsf{M}$ by E_x. We will sometimes denote the zero vector in E_x by 0_x. The set of C^r-sections of $\pi \colon \mathsf{E} \to \mathsf{M}$ we denote by $\Gamma^r(\mathsf{E})$. If $S \subseteq \mathsf{M}$, we denote by $\mathsf{E}|S$ the restriction of E to S, i.e., $\mathsf{E}|S = \pi^{-1}(S)$. The trivial bundle $\mathbb{R}^k \times \mathsf{M}$ is denoted by $\mathbb{R}^k_{\mathsf{M}}$. If $\pi \colon \mathsf{E} \to \mathsf{M}$ is a C^r-vector bundle and if $\Phi \in \mathrm{C}^r(\mathsf{N}; \mathsf{M})$ is a C^r-mapping of manifolds, then $\Phi^*\pi \colon \Phi^*\mathsf{E} \to \mathsf{N}$ is the pull-back vector bundle, with

$$\Phi^*\mathsf{E} = \{(e, y) \in \mathsf{E} \times \mathsf{N} \mid \pi(e) = \Phi(y)\}$$

and $\Phi^*\pi(e, y) = y$.

The bundle of k-jets of local sections of a C^r-vector bundle $\pi \colon \mathsf{E} \to \mathsf{M}$ we denote by $\mathsf{J}^k\mathsf{E}$.

The derivative of $\Phi \in \mathrm{C}^\infty(\mathsf{M}; \mathsf{N})$ is denoted by $T\Phi \colon \mathsf{TM} \to \mathsf{TN}$. We denote $T_x\Phi = T\Phi|\mathsf{T}_x\mathsf{M}$. If $f \in \mathrm{C}^r(\mathsf{M})$, then $\mathrm{d}f \in \Gamma^r(\mathsf{T}^*\mathsf{M})$ is defined by

$$\langle \mathrm{d}f(x); v_x \rangle = T_x f(v_x), \qquad v_x \in \mathsf{TM}.$$

If $I \subseteq \mathbb{R}$ is an interval and if $\gamma \colon I \to \mathsf{M}$ is differentiable at $t \in I$, then we denote $\gamma'(t) = T_t\gamma(1)$. For a diffeomorphism $\Phi \colon \mathsf{M} \to \mathsf{N}$, for tensor fields A on M and B on N, Φ_*A is the push-forward of A by Φ and Φ^*B is the pull-back of B by Φ.

We shall completely eschew any use of local coordinates, but for certain technical results we shall properly embed manifolds in some Euclidean space $\mathbb{R}^N$, assuming in such instances that manifolds are second countable, e.g., connected. The existence of such embeddings is proved by Whitney [1936] in the smooth case and by Grauert [1958] in the real analytic case. The principal manner in which we shall use these embeddings is according to the following result.

1.1 Lemma: (Global generators for vector fields and one-forms) *Let $r \in \{\infty, \omega\}$ and let M be a second countable C^r-manifold. Then the following statements hold:*

(i) there exist $X_1, \dots, X_N \in \Gamma^r(\mathsf{TM})$ such that, if $X \in \Gamma^r(\mathsf{TM})$, then

$$X = f^1 X_1 + \dots + f^N X_N$$

for some $f^1, \dots, f^N \in \mathrm{C}^r(\mathsf{M})$;

(ii) there exist $g^1, \dots, g^N \in \mathrm{C}^r(\mathsf{M})$ such that, if $\beta \in \Gamma^r(\mathsf{T}^*\mathsf{M})$, then

$$\beta = f^1 \mathrm{d}g^1 + \dots + f^N \mathrm{d}g^N$$

for some $f^1, \dots, f^N \in \mathrm{C}^r(\mathsf{M})$.

Proof: We assume that we have a proper embedding $\iota \colon \mathsf{M} \to \mathbb{R}^N$. If $x \in \mathsf{M}$, we shall simply write $\iota(x) = x$. We denote by $\widehat{X}_1, \dots, \widehat{X}_N$ the coordinate vector fields on $\mathbb{R}^N$ and by $\hat{g}^1, \dots, \hat{g}^N$ the coordinate functions.

(i) Denote by $X_1, \dots, X_N \in \Gamma^r(\mathsf{TM})$ the vector fields on M obtained by requiring that $X_j(x)$ be the orthogonal projection of $\widehat{X}_j(x)$ (with respect to the Euclidean metric) onto $\mathsf{T}_x\mathsf{M}$. If $x \in \mathsf{M}$ and if $v \in \mathsf{T}_x\mathsf{M}$, then $v \in \mathsf{T}_x\mathbb{R}^N$ and so there are unique $v^1, \dots, v^N \in \mathbb{R}$ such that

$$v = v^1 \widehat{X}_1(x) + \dots + v^N \widehat{X}_N(x).$$

We then have, by orthogonal projection,

$$v = v^1 X_1(x) + \dots + v^N X_N(x),$$

and the result follows by performing the previous constructions for $v = X(x)$ for every $x \in \mathsf{M}$.

(ii) For $x \in \mathsf{M}$ and $\alpha \in \mathsf{T}_x^*\mathsf{M}$, let $\alpha = T_x^*\iota(\hat{\alpha})$ for some $\hat{\alpha} \in \mathsf{T}_x^*\mathbb{R}^N$. We can make $\hat{\alpha}$ unique by requiring that $\hat{\alpha}$ is orthogonal (with respect to the Euclidean inner product on $\mathsf{T}_x^*\mathbb{R}^N$) to the annihilator of $\mathsf{T}_x\mathsf{M}$. We can then write

$$\hat{\alpha} = \alpha_1 \mathrm{d}\hat{g}^1(x) + \dots + \alpha_N \mathrm{d}\hat{g}^N(x)$$

for unique $\alpha_1, \dots, \alpha_N \in \mathbb{R}$. We then have

$$\alpha = \alpha_1 T_x^*\iota(\mathrm{d}\hat{g}^1(x)) + \dots + \alpha_N T_x^*\iota(\mathrm{d}\hat{g}^N(x)).$$

If $g^j = \iota^*\hat{g}^j$, $j \in \{1, \dots, N\}$, are the restrictions of the coordinate functions to M, then this gives

$$\alpha = \alpha_1 \mathrm{d}g^1(x) + \dots + \alpha_N \mathrm{d}g^N(x)$$

using the commuting of pull-back and differential [Abraham, Marsden, and Ratiu 1988, Theorem 7.4.4]. The result follows by performing the previous constructions for $\alpha = \beta(x)$ for every $x \in \mathsf{M}$. ■

The flow of a vector field $X \in \Gamma^r(\mathsf{TM})$ is denoted by Φ_t^X, so that the solution to the initial value problem

$$\xi'(t) = X \circ \xi(t), \quad \xi(0) = x,$$

is $t \mapsto \Phi^X_t(x)$. Many of our constructions and results do not require vector fields to be complete, but a crucial component of our analysis requires completeness of a certain vector field. If $X \in \Gamma^r(\mathsf{TM})$ and if $f \in \mathrm{C}^r(\mathsf{M})$, by $\mathscr{L}_X f \in \mathrm{C}^r(\mathsf{M})$ we denote the Lie derivative of f with respect to X. The Lie bracket of vector fields $X, Y \in \Gamma^r(\mathsf{TM})$ is the vector field $[X,Y]$ defined by

$$\mathscr{L}_{[X,Y]} f = \mathscr{L}_X \mathscr{L}_Y f - \mathscr{L}_Y \mathscr{L}_X f.$$

Let $(\mathsf{M}, \mathbb{G})$ be a Riemannian manifold. By $\|\cdot\|_{\mathbb{G}}$ we denote the fibre norm defined by $\mathbb{G}$. We denote by $\overset{\mathbb{G}}{\nabla}$ the Levi-Civita affine connection. We denote by exp the Riemannian exponential, which we regard as a mapping from a neighbourhood of the zero section in TM into M. If $f \in \mathrm{C}^\infty(\mathsf{M})$, we denote $\operatorname{grad} f = \mathbb{G}^\sharp \circ \mathrm{d}f$.

For a C^r-manifold M, we denote by $\mathscr{C}^r_{\mathsf{M}}$ the sheaf of C^r-functions over M. The stalk of this sheaf at $x \in \mathsf{M}$ is denoted by $\mathscr{C}^r_{x,\mathsf{M}}$. For a C^r-vector bundle $\pi\colon \mathsf{E} \to \mathsf{M}$, we denote by $\mathscr{G}^r_{\mathsf{E}}$ the sheaf of C^r-sections of E, thought of as a $\mathscr{C}^r_{\mathsf{M}}$-module. By $\mathscr{G}^r_{x,\mathsf{E}}$ we denote the stalk of $\mathscr{G}^r_{\mathsf{E}}$ at $x \in \mathsf{M}$.

We shall on occasion require the following lemma which is elementary in the smooth case, but is less elementary in the real analytic case.

1.2 Lemma: (Globally defined sections with a prescribed jet at a point) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $k \in \mathbb{Z}_{\geq 0}$, and let $\Xi \in \mathsf{J}^k\mathsf{E}_x$. Then there exists $\xi \in \Gamma^r(\mathsf{E})$ such that $j_k\xi(x) = \Xi$.*

Proof: Let $\mathscr{E}^k_x$ be the sheaf of sections of E whose k-jets vanish at x. Thus

$$\mathscr{E}^k_x(\mathcal{U}) = \begin{cases} \{\xi \in \Gamma^r(\mathsf{E}|\mathcal{U}) \mid j_k\xi(x) = 0\}, & x \in \mathcal{U}, \\ \Gamma^r(\mathsf{E}|\mathcal{U}), & x \notin \mathcal{U}. \end{cases}$$

Consider the short exact sequence of sheaves

$$0 \longrightarrow \mathscr{E}^k_x \longrightarrow \mathscr{G}^r_{\mathsf{E}} \longrightarrow \mathscr{G}^r_{\mathsf{E}}/\mathscr{E}^k_x \longrightarrow 0$$

One readily verifies that

$$\mathscr{G}^r_{x,\mathsf{E}}/\mathscr{E}^k_{x,y} \simeq \begin{cases} \mathsf{J}^k\mathsf{E}_x, & x = y, \\ 0, & x \neq y. \end{cases}$$

This short exact sequence of sheaves gives rise to the long exact sequence for global sections

$$0 \longrightarrow \mathrm{H}^0(\mathscr{E}^k_x) \longrightarrow \mathrm{H}^0(\mathscr{G}^r_{\mathsf{E}}) \longrightarrow \mathrm{H}^0(\mathscr{G}^r_{\mathsf{E}}/\mathscr{E}^k_x) \longrightarrow \mathrm{H}^1(\mathscr{E}^k_x) \longrightarrow \cdots$$

We claim that $H^1(\mathscr{E}^k_x) = 0$. We consider the smooth and real analytic cases separately.

1. In the smooth case, [Wells Jr. 2008, Proposition 3.11] (along with [Wells Jr. 2008, Examples 3.4(d, e)] and [Wells Jr. 2008, Proposition 3.5]), immediately gives the vanishing of $\mathrm{H}^p(\mathscr{E}^k_x)$ for $p \in \mathbb{Z}_{>0}$.
2. In the real analytic case, first, by Oka's Theorem,[2] $\mathscr{G}^\omega_{\mathsf{E}}$ is coherent. By a standard argument using Hadamard's Lemma (cf. Lemma 1 from the proof of Proposition 4.3),

[2]For a proof of Oka's Theorem in the holomorphic case, see [Grauert and Remmert 1984, Theorem 2.5.2]. The same proof works in the real analytic case.

one shows that $\mathscr{E}_x^k$ is locally finitely generated (by monomials, in coordinates). Thus $\mathscr{E}_x^k$ is coherent in the real analytic case by [Grauert and Remmert 1984, Example 2, pg 235]. Now, by Cartan's Theorem B in the real analytic case [Cartan 1957, Proposition 6], the sheaf cohomology of $\mathscr{E}_x^k$ vanishes in positive degree, particularly in degree 1.

As $H^1(\mathscr{E}_x^k) = 0$, we have the surjectivity of the mapping

$$\mathrm{H}^0(\mathscr{G}_{\mathsf{E}}^r) = \Gamma^r(\mathsf{E}) \ni \xi \mapsto j_k\xi(x) \in \mathrm{H}^0(\mathscr{G}_{\mathsf{E}}^r/\mathscr{E}_x^k) \simeq \mathsf{J}^k\mathsf{E}_x,$$

which is what is to be proved. ■

List of symbols. For convenient reference we list the commonly used, but not necessarily commonplace, notation that we use, along with its place of definition.

X^*	: dual of linear vector field X, 18
∂_1, ∂_2	: partial derivatives for functions of two real variables, 7
$\hat{\Phi}$	: function associated to Sobolev space-valued function, 56
$\widehat{(A,b)}$	: linear map on $\mathsf{V} \oplus \mathbb{R}$ determined by $A \in \mathrm{End}(\mathsf{V})$ and $b \in \mathsf{V}$, 85
$\overset{\mathrm{D}}{\nabla}$	: constrained connection, 41
$\overset{\mathrm{G}}{\nabla}$	: Levi-Civita connection for Riemannian metric $\mathbb{G}$, 10
$A^\flat$	: endomorphism associated with $(0,2)$-tensor A, 7
$B^\sharp$	: endomorphism associated with $(2,0)$-tensor A, 7
Δ_0	: subspace associated with subspace $\Delta \subseteq \mathsf{V}^* \oplus \mathbb{R}$, 36
Δ_1	: subspace associated with subspace $\Delta \subseteq \mathsf{V}^* \oplus \mathbb{R}$, 36
$\delta\Phi$	: variational derivative of Φ, 64
$\delta\sigma$	: variation field associated with variation σ, 58
$\Lambda(\mathsf{F})$	: annihilator subbundle of subbundle F, 27
$\nu\sigma$	: velocity field associated with variation σ, 58
Φ_t^X	: flow for vector field X, 9
$\mathscr{C}_{\mathsf{M}}^r$	: sheaf of C^r-functions on M, 10
$\mathscr{G}_{\mathsf{E}}^r$	: sheaf of C^r-sections of E, 10
$\mathscr{I}_{\mathsf{S}}$	: ideal sheaf of a variety S, 69
A_{D}	: linear part of adjoint equation, 148
$\mathsf{A}(\Delta)$	: affine bundle associated with defining subbundle Δ, 37
$\mathrm{aff}_{\mathbb{R}}(S)$	: affine span of S, 7
$\mathrm{Aff}^r(\mathsf{E})$	: affine functions on vector bundle E, 21
A_Υ	: linear part of adjoint equation along curve, 146
b_{D}	: constant part of adjoint equation, 148
b_Υ	: constant part of adjoint equation along curve, 146
$\boldsymbol{D}f$	: Fréchet derivative of f, 7
$\langle\cdot,\cdot\rangle_{\mathrm{D}}$	: Dirichlet semi-inner product, 54
A^{e}	: vertical evaluation of endomorphism A, 15
$(\mathscr{G}_{\mathsf{E}^*}^r(\mathcal{U}))^{\mathrm{e}}$	: vertical evaluation sheaf, 72

λ^{e}	:	vertical evaluation of section λ, 21
ev_f	:	evaluation mapping, 51
F_{D}	:	Frobenius curvature of distribution D, 41
$\hat{F}_{\mathsf{D}}$	:	pull-back of F_{D}, 146
G_{D}	:	geodesic curvature for distribution D, 41
f^{h}	:	horizontal lift of function f, 21
hlft	:	horizontal lift mapping, 14
hor	:	horizontal projection, 14
$\mathrm{H}^s(I;\mathbb{R})$	:	sth-order Sobolev space of functions, 8
$\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{E})$	:	sth-order Sobolev space of sections along curve γ, 50
$\mathrm{H}^s([t_0,t_1];\mathsf{M})$	:	sth-order Sobolev space of curves, 49
$\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})$	:	sth-order Sobolev space of curves with derivatives in D, 50
X^{h}	:	horizontal lift of vector field X, 15
$\boldsymbol{I}_n$	:	$n\times n$ identity matrix, 6
K_Γ	:	connector, 14
$L(\mathsf{A})$	:	linear part of affine space A, 7
$\mathrm{Lin}^r(\mathsf{E})$	:	linear functions on vector bundle E, 21
pr_j	:	projection from $X_1\times\cdots\times X_k$ onto the jth factor, 6
$\mathbb{R}^k_{\mathsf{M}}$	:	trivial vector bundle, 8
$\mathbb{R}^{m\times n}$	:	set of matrices with m rows and n columns, 6
S_{D}	:	second fundamental form of distribution D, 41
$\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$	:	tangent space to $\mathrm{H}^1([t_0,t_1];\mathsf{M})$, 62
$\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$	:	tangent space to $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$, 113
$T\Phi$	:	derivative of mapping Φ, 8
ver	:	vertical projection, 14
vlft	:	vertical lift mapping, 14
ξ^{v}	:	vertical lift of section ξ, 15
$\widehat{X}$	:	linear vector field on $\mathsf{E}\oplus\mathbb{R}_{\mathsf{M}}$ associated to affine vector field X on E, 85
$X^{\mathrm{nh}}_{\mathsf{D}}$	:	vector field describing nonholonomic dynamics, 147
$X^{\mathrm{reg}}_{\mathsf{D}}$	:	regular adjoint vector field, 148
$X^{\mathrm{sing}}_{\mathsf{D}}$	:	singular adjoint vector field, 148

Acknowledgements. Discussions with Connor Boyd and Bahman Gharesifard were helpful in the development of some ideas in the paper. The author thanks Mike Roth for helping to clarify the notions of invariance presented in Section 4.

2. Geometric preliminaries

In this section we develop the tools we need to state our main results. Some of the constructions we present are made for review and to present the notation we use. However, some of the developments are nonstandard.

2.1. Connections on vector bundles. We shall require a few particular constructions concerning connections, and in this section we overview the required material. One objective is to make perfectly clear the meaning of the covariant derivative of a locally absolutely continuous section along a locally absolutely continuous curve.

We let $r \in \{\infty, \omega\}$ and consider a C^r-vector bundle $\pi\colon \mathsf{E} \to \mathsf{M}$. We note that the bundle $\mathsf{J}^1\mathsf{E}$ of 1-jets of local sections of E is an affine bundle modelled on $\pi^*\mathsf{T}^*\mathsf{M} \otimes \pi^*\mathsf{E}$. A ***$\mathbf{C}^r$-connection*** in E is a C^r-section of $j_1\pi$, $\Gamma\colon \mathsf{E} \to \mathsf{J}^1\mathsf{E}$. The connection Γ is ***linear*** if Γ is a morphism of vector bundles according to the diagram

$$\begin{tikzcd} \mathsf{E} \arrow[r, "\Gamma"] \arrow[d, "\pi"'] & \mathsf{J}^1\mathsf{E} \arrow[d, "\pi\circ j_1\pi"] \\ \mathsf{M} \arrow[r, equal] & \mathsf{M} \end{tikzcd}$$

The relationship of this to covariant differentiation is accomplished as follows. Let $X \in \Gamma^r(\mathsf{TM})$ and $\xi \in \Gamma^r(\mathsf{E})$. Then $j_1\xi - \Gamma \circ \xi$ is a C^r-section of $\pi \circ j_1\pi$ that covers ξ:

$$\begin{tikzcd} \mathsf{J}^1\mathsf{E} \arrow[r, "j_1\pi"] & \mathsf{E} \\ \mathsf{M} \arrow[u, "j_1\xi-\Gamma\circ\xi"] \arrow[ur, "\xi"'] & \end{tikzcd}$$

Thus

$$j_1\xi(x) - \Gamma \circ \xi(x) \in \mathsf{J}^1_{\xi(x)}\mathsf{E} \simeq \mathsf{T}^*_x\mathsf{M} \otimes \mathsf{E}_x.$$

We then can define the section $\nabla_X\xi \in \Gamma^r(\mathsf{E})$ by

$$\nabla_X\xi(x) = (j_1\xi(x) - \Gamma \circ \xi(x))(X(x)).$$

Let us define the horizontal/vertical decomposition of a vector bundle associated with a connection Γ in $\pi\colon \mathsf{E} \to \mathsf{M}$. For $x \in \mathsf{M}$, $e \in \mathsf{E}_x$, and $e_1 \in \mathsf{J}^1_e\mathsf{E}$, let $\xi \in \Gamma^r(\mathsf{E})$ be such that $\xi(x) = e$ and $j_1\xi(x) = e_1$. There then exists a unique and well-defined linear mapping $L_{e_1} \in \mathsf{L}(\mathsf{T}_x\mathsf{M}; \mathsf{T}_e\mathsf{E})$ satisfying $L_{e_1}(v) = T_x\xi(v)$. We then have a C^r-vector bundle morphism P^H_Γ of TE (as a vector bundle over E) defined by

$$P^H_\Gamma(u_e) = L_{\Gamma(e)} \circ T_e\pi(u_e), \qquad u_e \in \mathsf{T}_e\mathsf{E}.$$

One can verify that (1) $\ker(P^H_\Gamma) = \ker(T\pi)$ and (2) $\mathsf{TE} = \ker(P^H_\Gamma) \oplus \operatorname{image}(P^H_\Gamma)$. We then denote the horizontal subbundle by $\mathsf{HE} = \operatorname{image}(P^H_\Gamma)$ and the vertical subbundle by $\mathsf{VE} = \ker(P^H_\Gamma)$. We denote by $P^V_\Gamma = \mathrm{id}_{\mathsf{TE}} - P^H_\Gamma$ the projection onto VE. Just by definition, P^V_Γ is a C^r-vector bundle mapping according to the diagram

$$\begin{tikzcd} \mathsf{TE} \arrow[r, "P^V_\Gamma"] \arrow[d, "\pi_{\mathsf{TE}}"'] & \mathsf{VE} \arrow[d, "\pi_{\mathsf{TE}}|\mathsf{VE}"] \\ \mathsf{E} \arrow[r, equal] & \mathsf{E} \end{tikzcd}$$

If Γ is additionally linear, P_Γ^V is a C^r-vector bundle mapping according to the diagram

$$\begin{tikzcd} \mathsf{TE} \arrow[r, "P_\Gamma^V"] \arrow[d, "T\pi"'] & \mathsf{VE} \arrow[d, "\nu_{\mathsf{E}}"] \\ \mathsf{TM} \arrow[r, "\pi_{\mathsf{TM}}"'] & \mathsf{M} \end{tikzcd}$$

where $\nu_{\mathsf{E}} = \pi \circ (\pi_{\mathsf{TE}}|\mathsf{VE})$.

We shall sometimes use hor in place of P_Γ^H and ver in place of P_Γ^V.

We can refine this further via the horizontal and vertical lift isomorphisms

$$\mathrm{hlft}\colon \pi^*\mathsf{TM} \to \mathsf{HE}$$

and

$$\mathrm{vlft}\colon \mathsf{E} \oplus \mathsf{E} \to \mathsf{VE}.$$

These are defined by requiring that $\mathrm{hlft}(v_x, e_x)$ be the unique horizontal vector satisfying $T_{e_x}\pi(\mathrm{hlft}(v_x, e_x)) = v_x$ and by

$$\mathrm{vlft}(f_x, e_x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (e_x + tf_x).$$

Note that vlft is canonical while hlft depends on the connection.

We then define the ***connector*** associated with Γ to be the mapping

$$K_\Gamma = \mathrm{pr}_1 \circ \mathrm{vlft}^{-1} \circ P_\Gamma^V \colon \mathsf{TE} \to \mathsf{E}, \tag{2.1}$$

where $\mathrm{pr}_1\colon \mathsf{E} \oplus \mathsf{E} \to \mathsf{E}$ is the projection onto the first factor. This is a C^r-vector bundle mapping according to the diagram

$$\begin{tikzcd} \mathsf{TE} \arrow[r, "K_\Gamma"] \arrow[d, "\pi_{\mathsf{TE}}"'] & \mathsf{E} \arrow[d, "\pi"] \\ \mathsf{E} \arrow[r, "\pi"'] & \mathsf{M} \end{tikzcd}$$

When Γ is additionally linear, then the connector is additionally a C^r-vector bundle mapping according to the diagram

$$\begin{tikzcd} \mathsf{TE} \arrow[r, "K_\Gamma"] \arrow[d, "T\pi"'] & \mathsf{E} \arrow[d, "\pi"] \\ \mathsf{TM} \arrow[r, "\pi_{\mathsf{TM}}"'] & \mathsf{M} \end{tikzcd}$$

This allows us to characterise covariant differentiation by the formula

$$\nabla_X \xi(x) = K_\Gamma \circ T\xi \circ X(x). \tag{2.2}$$

A crucial observation obtained from the preceding formula is that $\nabla_X \xi(x)$ depends only on the derivative of ξ in the direction of X. This allows us to differentiate locally absolutely continuous sections along locally absolutely continuous curves. Consider a continuous curve $\gamma\colon I \to \mathsf{M}$ defined on an interval $I \subseteq \mathbb{R}$ and a continuous section $\eta\colon I \to \mathsf{E}$ over γ, i.e.,

$$\begin{tikzcd} I \arrow[r, "\eta"] \arrow[dr, "\gamma"'] & \mathsf{E} \arrow[d, "\pi"] \\ & \mathsf{M} \end{tikzcd}$$

Suppose that both γ and η are differentiable at $t \in I$. Let $X \in \Gamma^r(\mathsf{TM})$ be such that $X \circ \gamma(t) = \gamma'(t)$, by Lemma 1.2. Similarly, let $\xi \in \Gamma^r(\mathsf{E})$ be such that $T_{\gamma(t)}\xi(\gamma'(t)) = \eta'(t)$. We then define

$$\nabla_{\gamma'(t)}\eta(t) \triangleq \nabla_X \xi(\gamma(t)).$$

If γ and η are locally absolutely continuous, then we can define a section $\nabla_{\gamma'}\eta$ over γ by

$$\nabla_{\gamma'}\eta(t) = \nabla_{\gamma'(t)}\eta(t), \qquad \text{a.e. } t \in I.$$

Moreover, (2.2) implies that

$$P^H_\Gamma(\eta'(t)) = \operatorname{hlft}(\gamma'(t), \eta(t)), \quad P^V_\Gamma(\eta'(t)) = \operatorname{vlft}(\nabla_{\gamma'(t)}\eta(t), \eta(t)), \qquad \text{a.e. } t \in I. \tag{2.3}$$

2.2. Vector fields on the total space of a vector bundle.

An essential rôle is played in our main results by certain vector fields defined on the total space of a vector bundle and the dual of a vector bundle. Let us define the types of vector fields that will arise on the total space of a vector bundle.

2.1 Definition: (Vector fields on the total space of a vector bundle) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, and let $X_0 \in \Gamma^r(\mathsf{TM})$.

(i) A vector field $X \in \Gamma^r(\mathsf{TE})$ is a ***linear*** (resp. ***affine***) vector field over X_0 if

(a) it projects to X_0, i.e., the diagram

$$\begin{tikzcd} \mathsf{E} \arrow[r, "X"] \arrow[d, "\pi"'] & \mathsf{TE} \arrow[d, "T\pi"] \\ \mathsf{M} \arrow[r, "X_0"'] & \mathsf{TM} \end{tikzcd}$$

commutes and

(b) it is a C^r-vector bundle morphism (resp. affine bundle morphism) of the preceding diagram.

(ii) If $\xi \in \Gamma^r(\mathsf{E})$, the ***vertical lift*** of ξ is the vector field $\xi^{\mathrm{v}} \in \Gamma^r(\mathsf{TE})$ defined by

$$\xi^{\mathrm{v}}(e) = \operatorname{vlft}(\xi \circ \pi(e), e).$$

(iii) For $A \in \Gamma^r(\operatorname{End}(\mathsf{E}))$, the ***vertical evaluation*** of A is the vector field $A^{\mathrm{e}} \in \Gamma^r(\mathsf{TE})$ defined by

$$A^{\mathrm{e}}(e) = \operatorname{vlft}(A(e), e).$$

Additionally assume that ∇ is a C^r-linear connection in E.

(iv) The ***horizontal lift*** of X_0 is the vector field $X_0^{\mathrm{h}} \in \Gamma^r(\mathsf{TE})$ defined by

$$X_0^{\mathrm{h}}(e) = \operatorname{hlft}(X_0 \circ \pi(e), e).$$ •

The following lemma assembles all of the above ingredients.

2.2 Lemma: (Linear and affine vector fields, and linear connections) *Let $r \in \{\infty, \omega\}$, let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, and let $X_0 \in \Gamma^r(\mathsf{TM})$. Then the following statements hold:*

(i) if $X^{\mathrm{lin}} \in \Gamma^r(\mathsf{TE})$ is a linear vector field over X_0, then there exists $A_{X^{\mathrm{lin}}} \in \Gamma^r(\mathrm{End}(\mathsf{E}))$ such that

$$X^{\mathrm{lin}} = X_0^{\mathrm{h}} + A^{\mathrm{e}}_{X^{\mathrm{lin}}};$$

(ii) if $X^{\mathrm{aff}} \in \Gamma^r(\mathsf{TE})$ is an affine vector field over X_0, then there exists $A_{X^{\mathrm{aff}}} \in \Gamma^r(\mathrm{End}(\mathsf{E}))$ and $b_{X^{\mathrm{aff}}} \in \Gamma^r(\mathsf{E})$ such that

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}}_{X^{\mathrm{aff}}} + b^{\mathrm{v}}_{X^{\mathrm{aff}}}.$$

Proof: (i) Since ∇ is a linear connection, the vertical projection ver, and therefore also the horizontal projection hor, are vector bundle mappings with respect to the following diagram:

$$\begin{tikzcd}
\mathsf{TE} \arrow[r, "\mathrm{ver},\mathrm{hor}"] \arrow[d, "\mathsf{T}\pi"'] & \mathsf{TE} \arrow[d, "T\pi"] \\
\mathsf{TM} \arrow[r, "\mathrm{id}_{\mathsf{TM}}"'] & \mathsf{TM}
\end{tikzcd}$$

[Kolář, Michor, and Slovák 1993, §11.10]. Therefore, since X^{lin} is a linear vector field over X_0, we have a vector bundle mapping $\mathrm{hor}(X^{\mathrm{lin}})$ determined by the following diagram:

$$\begin{tikzcd}
\mathsf{E} \arrow[r, "\mathrm{hor}(X^{\mathrm{lin}})"] \arrow[d, "\pi"'] & \mathsf{TE} \arrow[d, "T\pi"] \\
\mathsf{M} \arrow[r, "X_0"'] & \mathsf{TM}
\end{tikzcd}$$

Thus we conclude two things: (1) $\mathrm{hor}(X^{\mathrm{lin}}) = X_0^{\mathrm{h}}$ since both $\mathrm{hor}(X^{\mathrm{lin}})$ and X_0^{h} are horizontal vector fields projecting to X_0; (2) X_0^{h} is a linear vector field over X_0. Thus $X^{\mathrm{lin}} - X_0^{\mathrm{h}}$ is a linear vector field over the zero vector field. This shows that $X^{\mathrm{lin}} - X_0^{\mathrm{h}}$ is vertical and so we have that $\mathrm{ver}(X^{\mathrm{lin}})$ is a linear vector field. Thus we have the following vector bundle mapping:

$$\begin{tikzcd}
\mathsf{E} \arrow[rr, "\mathrm{vlft}^{-1}\circ\mathrm{ver}(X^{\mathrm{lin}})"] \arrow[dr, "\pi"'] & & \mathsf{E} \oplus \mathsf{E} \arrow[dl, "\pi\oplus\pi"] \\
& \mathsf{M} &
\end{tikzcd}$$

Since $\mathrm{ver}(X^{\mathrm{lin}})$ is a vector field on E, we have

$$\mathrm{vlft}^{-1} \circ \mathrm{ver}(X^{\mathrm{lin}})(e_x) = (A_{X^{\mathrm{lin}}}(e_x), e_x)$$

for $A_{X^{\mathrm{lin}}} \in \mathrm{End}(\mathsf{E})_x$, and this gives the assertion.

(ii) This follows from the observation that an affine map between vector spaces has the form of the sum of a linear map and a constant map. ■

2.3. Flows of vector fields on the total space of a vector bundle. It will be useful to have at hand characterisations of the flows of the various vector fields considered above.

2.3 Lemma: (Flows of vector fields on the total space of a vector bundle) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $X \in \Gamma^r(\mathsf{TE})$ be a linear vector field over X_0, let $\xi \in \Gamma^r(\mathsf{E})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Then the following statements hold:*

(i) if $x \in \mathsf{M}$ and $t \in \mathbb{R}$ are such that $\Phi_t^{X_0}(x)$ is defined, then

(a) $\Phi_t^X(e)$ is defined for every $e \in \mathsf{E}_x$,

(b) $\Phi_t^X(e) \in \mathsf{E}_{\Phi_t^{X_0}(x)}$ for every $e \in \mathsf{E}_x$, and

(c) $\Phi_t^X|\mathsf{E}_x\colon \mathsf{E}_x \to \mathsf{E}_{\Phi_t^{X_0}(x)}$ is an isomorphism of $\mathbb{R}$-vector spaces;

(ii) $\Phi_t^{\xi^{\mathrm{v}}}(e) = e + t\xi(\pi(e))$ for every $(t, e) \in \mathbb{R} \times \mathsf{E}$;

(iii) $\Phi_t^{A^{\mathrm{e}}}(e) = \mathrm{e}^{A(\pi(e))t}(e)$ for every $(t, e) \in \mathbb{R} \times \mathsf{E}$;

(iv) $\Phi_t^{X_0^{\mathrm{h}}}(e)$ is the parallel transport of e along the curve $t \mapsto \Phi_t^{X_0}(\pi(e))$.

Proof: (i) This is shown, for example, in [Kolář, Michor, and Slovák 1993, §47.9].

(ii) This follows from the definition of vertical lift.

(iii) We note that, if $t \mapsto \Upsilon(t)$ is an integral curve for A^{e}, then, since A^{e} is vertical, $T_{\Upsilon(t)}\pi(\Upsilon'(t)) = 0$ and

$$\mathrm{ver}(\Upsilon'(t)) = \mathrm{vlft}(A(\Upsilon(t)), \Upsilon(t)).$$

This is a linear differential equation in E_x, where $x = \pi \circ \Upsilon(t)$ for all t, and so we have $\Upsilon(t) = \mathrm{e}^{A(x)t}(\Upsilon(0))$.

(iv) This is the content of [Kobayashi and Nomizu 1963, §II.3]. ■

The following characterisation of integral curves of affine vector fields will allow us to connect the formulation in this section to formulations that will arise in Section 7.

2.4 Lemma: (Covariant derivative characterisation of integral curves of affine vector fields) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$, let ∇ be a C^r-linear connection in E, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and let $b \in \Gamma^r(\mathsf{E})$. For a curve $\Upsilon\colon I \to \mathsf{E}$, the following are equivalent:*

(i) Υ is an integral curve of $X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$;

(ii) $\nabla_{\gamma'}\Upsilon = A \circ \Upsilon + b \circ \gamma$, where $\gamma = \pi \circ \Upsilon$.

Proof: Let $\gamma = \pi \circ \Upsilon$ so that Υ is to be thought of as a section of E along γ. Then Υ is an integral curve of $X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$ if and only if

$$\Upsilon' = X_0^{\mathrm{h}} \circ \Upsilon + A^{\mathrm{e}} \circ \Upsilon + b^{\mathrm{v}} \circ \Upsilon.$$

Taking the vertical part of this equation and using the second of the equations in (2.3) gives the lemma. ■

The following adaptation of the variation of constants formula for linear ordinary differential equations will also be useful. In the statement and proof of the lemma we use some notation and results that we will introduce and prove, respectively, below.

2.5 Proposition: (Variation of constants formula for the flow of an affine vector field) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $X^{\mathrm{lin}} \in \Gamma^r(\mathsf{TE})$ be a C^r-linear vector field over X_0, and let $b \in \Gamma^r(\mathsf{E})$. Define $X^{\mathrm{aff}} = X^{\mathrm{lin}} + b^{\mathrm{v}}$. Let $(t, x) \in \mathbb{R} \times \mathsf{M}$ be such that $\Phi^{X_0}_t(x)$ is defined. Then, for $F \in \mathrm{Lin}^r(\mathsf{E})$ (see (2.4) for notation),*

$$F \circ \Phi^{X^{\mathrm{aff}}}_t(e) = F \circ \Phi^{X^{\mathrm{lin}}}_t(e) + \int_0^t F \circ \Phi^{X^{\mathrm{lin}}}_{t-\tau}(b \circ \Phi^{X_0}_\tau(\pi(e)))\,\mathrm{d}\tau, \qquad e \in \mathsf{E}.$$

Proof: We first make a calculation:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\bigg(F \circ \Phi^{X^{\mathrm{lin}}}_t(e) + &\int_0^t F \circ \Phi^{X^{\mathrm{lin}}}_{t-\tau}(b \circ \Phi^{X_0}_\tau(\pi(e)))\,\mathrm{d}\tau\bigg)\\
&= \langle \mathrm{d}F(\Phi^{X^{\mathrm{lin}}}_t(e)); X^{\mathrm{lin}}(\Phi^{X^{\mathrm{lin}}}_t(e))\rangle + F \circ b \circ \Phi^{X_0}_t(\pi(e))\\
&\quad + \int_0^t \langle \mathrm{d}F(\Phi^{X^{\mathrm{lin}}}_{t-\tau}(b \circ \Phi^{X_0}_\tau(\pi(e)))); X^{\mathrm{lin}}(\Phi^{X^{\mathrm{lin}}}_{t-\tau}(b \circ \Phi^{X_0}_\tau(\pi(e))))\rangle\,\mathrm{d}\tau\\
&= \langle \mathrm{d}F; X^{\mathrm{lin}}\rangle(\Phi^{X^{\mathrm{lin}}}_t(e)) + \int_0^t \langle \mathrm{d}F; X^{\mathrm{lin}}\rangle \circ \Phi^{X^{\mathrm{lin}}}_{t-\tau}(b \circ \Phi^{X_0}_t(\pi(e)))\,\mathrm{d}\tau\\
&\quad + F \circ b(\Phi^{X_0}_t(\pi(e))).\end{aligned}$$

By Lemma 2.10 below, we have

$$\langle \mathrm{d}F; X^{\mathrm{aff}}\rangle = \langle \mathrm{d}F; X^{\mathrm{lin}}\rangle + (F \circ b)^{\mathrm{h}}$$

(noting that, since $F \in \mathrm{Lin}^r(\mathsf{E})$, $F = \lambda^{\mathrm{e}}$ for some $\lambda \in \Gamma^r(\mathsf{E}^*)$). We have

$$(F \circ b)^{\mathrm{h}}(\Phi^{X^{\mathrm{aff}}}_t(e)) = F \circ b(\Phi^{X_0}_t(\pi(e)))$$

since X^{aff} projects to X_0. Thus our initial calculation shows that the derivative of the linear function F as a function of time is the derivative of F with respect to X^{lin} plus the derivative of F with respect to b^{v}. Since the right-hand side of the asserted expression evaluates to $F(e)$ at $t = 0$, we conclude that this right-hand side gives the evolution of F along integral curves of X^{aff}, as claimed. ■

2.4. The dual of a linear vector field. If $\pi\colon \mathsf{E} \to \mathsf{M}$ is a vector bundle, then E^* is the set of vector bundle maps from E to the trivial vector bundle $\mathbb{R}_{\mathsf{M}}$. This is the ***dual vector bundle*** for E. We denote the canonical projection by $\pi^*\colon \mathsf{E}^* \to \mathsf{M}$, acknowledging the possible confusion of the projection π^* with the pull-back by the projection π. Let $\mathcal{U} \subseteq \mathsf{M}$ be an open subset and let $\Phi\colon \mathsf{E}|\mathcal{U} \to \mathsf{E}$ have the property that it is a vector bundle isomorphism onto its image over the map $\Phi_0\colon \mathcal{U} \to \mathsf{M}$ which is a diffeomorphism onto its image. The ***dual*** of Φ is the map $\Phi^*\colon (\Phi(\mathsf{E}|\mathcal{U}))^* \to (\mathsf{E}|\mathcal{U})^*$ defined by $\Phi^*|\mathsf{E}^*_{\Phi_0(x)} = (\Phi|\mathsf{E}_x)^*$.

Associated with a linear vector field on E is its dual, determined according to the following.

2.6 Definition: (Dual of a linear vector field) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $X \in \Gamma^r(\mathsf{TE})$ be a linear vector field over X_0. The ***dual vector field*** of X is the vector field X^* on E^* defined by

$$X^*(\alpha_x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\Phi^X_{-t})^*(\alpha_x) \qquad \bullet$$

We can prove some fundamental properties of the dual of a linear vector field. To do so, it is convenient to introduce some notation. For a vector bundle $\pi\colon \mathsf{E} \to \mathsf{M}$ with dual bundle $\pi^*\colon \mathsf{E}^* \to \mathsf{M}$ and for a linear vector field X on E, we denote by $X \times X^*$ the vector field on $\mathsf{E} \times \mathsf{E}^*$ defined by $X \times X^*(e, \alpha) = (X(e), X^*(\alpha))$. We also note that the Whitney sum $\mathsf{E} \oplus \mathsf{E}^*$ is the submanifold of $\mathsf{E} \times \mathsf{E}^*$ given by

$$\mathsf{E} \oplus \mathsf{E}^* = \{(v, \alpha) \in \mathsf{E} \times \mathsf{E}^* \mid \ \pi(v) = \pi^*(\alpha)\}.$$

With this notation we have the following result.

2.7 Lemma: (Properties of the dual vector field) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $X \in \Gamma^r(\mathsf{TE})$ be a linear vector field over X_0. Then the dual vector field X^* has the following properties:*

(i) X^* is a linear vector field over X_0;

(ii) the vector field $X \times X^*$ is tangent to $\mathsf{E} \oplus \mathsf{E}^* \subseteq \mathsf{E} \times \mathsf{E}^*$;

(iii) if $X \oplus X^*$ denotes the restriction of $X \times X^*$ to $\mathsf{E} \oplus \mathsf{E}^*$ and if $f_{\mathsf{E}}\colon \mathsf{E} \oplus \mathsf{E}^* \to \mathbb{R}$ denotes the function $f_{\mathsf{E}}(e \oplus \alpha) = \alpha(e)$, then $\mathscr{L}_{X \oplus X^*} f_{\mathsf{E}} = 0$.

Moreover, if $Y \in \Gamma^r(E^*)$ is a linear vector field over X_0, then

(iv) the vector field $X \times Y$ is tangent to $\mathsf{E} \oplus \mathsf{E}^* \subseteq \mathsf{E} \times \mathsf{E}^*$ and,

(v) if $X \oplus Y$ denotes the restriction of $X \times Y$ to $\mathsf{E} \oplus \mathsf{E}^*$, then $\mathscr{L}_{X \oplus Y} f_{\mathsf{E}} = 0$ only if $Y = X^*$.

Proof: (i) Note that $\pi^* \circ (\Phi^X_{-t})^*(\alpha) = \Phi^{X_0}_t \circ \pi^*(\alpha)$. Therefore,

$$\begin{aligned} T_\alpha \pi^*(X^*(\alpha)) &= T_\alpha \pi^* \left(\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\Phi^X_{-t})^*(\alpha) \right) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \pi^* \circ (\Phi^X_{-t})^*(\alpha) \\ &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \Phi^{X_0}_t \circ \pi^*(\alpha) = X_0(\pi^*(\alpha)). \end{aligned}$$

Thus X^* projects to X_0. Since the flow of X^*, by definition, consists of local isomorphisms of E^*, it follows from [Kolář, Michor, and Slovák 1993, §47.9] that X^* is a linear vector field.

(ii) The submanifold $\mathsf{E} \oplus \mathsf{E}^*$ is the preimage of the diagonal submanifold

$$\Delta = \{(x, y) \in \mathsf{M} \times \mathsf{M} \mid \ x = y\}$$

under the projection $\pi \times \pi^*\colon \mathsf{E} \times \mathsf{E}^* \to \mathsf{M} \times \mathsf{M}$. Since $\pi \times \pi^*$ is a surjective submersion, it is transversal to Δ [Abraham, Marsden, and Ratiu 1988, Definition 3.5.10]. Therefore, by [Abraham, Marsden, and Ratiu 1988, Theorem 3.5.12],

$$(e, \alpha) \in (\pi \times \pi^*)^{-1}(x, x) \quad \Longrightarrow \quad T_{(e,\alpha)}(\mathsf{E} \oplus \mathsf{E}^*) = T_{(e,\alpha)}(\pi \times \pi^*)^{-1}(\mathsf{T}_{(x,x)}\Delta).$$

For $(e,\alpha)\in\mathsf{E}\oplus\mathsf{E}^*$ we have

$$T_e\pi(X(e))=T_\alpha\pi^*(X^*(\alpha))=X_0(\pi(e))$$

and so

$$(T_e\pi(X(e)),T_\alpha\pi^*(X^*(\alpha)))\in\mathsf{T}_{(\pi(e),\pi^*(\alpha))}\Delta.$$

Now we compute

$$T_{(e,\alpha)}(\pi\times\pi^*)((X\times X^*)(e,\alpha))=(T_e\pi(X(e)),T_\alpha\pi^*(X^*(\alpha)))\in\mathsf{T}_{(\pi(e),\pi^*(\alpha))}\Delta,$$

giving the result.

(iii) We have

$$\begin{aligned}f_{\mathsf{E}}(\Phi_t^{X\oplus X^*}(e,\alpha))&=f_{\mathsf{E}}(\Phi_t^X(e),\Phi_t^{X^*}(\alpha))=\langle\Phi_t^{X^*}(\alpha);\Phi_t^X(e)\rangle\\&=\langle\alpha;(\Phi_t^{X^*})^*\circ\Phi_t^X(e)\rangle=\langle\alpha;\Phi_{-t}^X\circ\Phi_t^X(e)\rangle=\alpha(e),\end{aligned}$$

from which the result follows by [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.10].

(iv) In the proof of part (ii), we only used the fact that $X=X^*$ are linear vector fields over X_0 in the proof. Thus the proof applies to the linear vector field Y over the same vector field X_0 as X.

(v) For every $(v,\alpha)\in\mathsf{E}_x\oplus\mathsf{E}_x^*$ we have

$$\begin{aligned}\alpha(v)&=f_{\mathsf{E}}(\Phi_t^{X\oplus Y}(v,\alpha))=f_{\mathsf{E}}(\Phi_t^X(v),\Phi_t^Y(\alpha))\\&=\langle\Phi_t^Y(\alpha);\Phi_t^X(v)\rangle=\langle\alpha;(\Phi_t^Y)^*\circ\Phi_t^X(v)\rangle.\end{aligned}$$

We conclude, therefore, that $(\Phi_t^Y)^*\circ\Phi_t^X=\mathrm{id}_{\mathsf{E}}$ and so $\Phi_t^Y=(\Phi_{-t}^X)^*$, as desired. ■

Let us determine the dual of a linear vector field represented in the decomposition of Lemma 2.2.

2.8 Lemma: (Duals of decomposed linear vector fields) *Let $r\in\{\infty,\omega\}$, let $\pi\colon\mathsf{E}\to\mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, and let $X_0\in\Gamma^r(\mathsf{TM})$. For $A\in\Gamma^r(\mathrm{End}(\mathsf{E}))$, the dual of the linear vector field $X=X_0^{\mathrm{h}}+A^{\mathrm{e}}$ is $X^*=X_0^{\mathrm{h},*}-(A^*)^{\mathrm{e}}$. Moreover, $X_0^{\mathrm{h},*}$ is the horizontal lift of X_0 corresponding to the dual linear connection in E^*.*

Proof: We have

$$f_{\mathsf{E}}(\Phi_t^{-A^{\mathrm{e}}\oplus(A^*)^{\mathrm{e}}}(e,\alpha))=\langle\Phi_t^{-(A^*)^{\mathrm{e}}}(\alpha);\Phi_t^{A^{\mathrm{e}}}(e)\rangle=\langle\mathrm{e}^{-A^*(\pi^*(\alpha))t}(\alpha);\mathrm{e}^{A(\pi(e))t}(e)\rangle.$$

Thus

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}f_{\mathsf{E}}(\Phi_t^{A^{\mathrm{e}}\oplus(A^*)^{\mathrm{e}}}(e,\alpha))=-\langle A^*(\alpha);e\rangle+\langle\alpha;A(e)\rangle=0.$$

By [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.10], we have $\mathscr{L}_{-A^{\mathrm{e}}\oplus(A^*)^{\mathrm{e}}}f_{\mathsf{E}}=0$. By Lemma 2.7(iii) this gives

$$\mathscr{L}_{(X_0^{\mathrm{h}}+A^{\mathrm{e}})\oplus(X_0^{\mathrm{h},*}-(A^*)^{\mathrm{e}})}f_{\mathsf{E}}=\mathscr{L}_{X_0^{\mathrm{h}}\oplus X_0^{\mathrm{h},*}}f_{\mathsf{E}}+\mathscr{L}_{A^{\mathrm{e}}\oplus(-(A^*)^{\mathrm{e}})}f_{\mathsf{E}}=0.$$

The first assertion in the result follows from Lemma 2.7(v).

For the final assertion, we make three observations from which the assertion follows:

1. the flows of the horizontal lifts X_0^{h} and $X_0^{\mathrm{h},*}$ are given by parallel translation Lemma 2.3(iv);
2. the flow of $X_0^{\mathrm{h},*}$ is the dual of the inverse flow of X_0^{h} by definition;
3. the parallel translation by the dual of a linear connection is the dual of the inverse of parallel translation of the linear connection (by definition of the dual of a linear connection). ■

2.5. Functions on the total space of a vector bundle. In this section we introduce some special classes of functions on vector bundles, and indicate how to differentiate these with respect to the special kinds of vector fields we introduced in the preceding sections.

The functions we consider are the following.

2.9 Definition: (Functions on the total space of a vector bundle) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle.

(i) The ***horizontal lift*** of $f \in \mathrm{C}^r(\mathsf{M})$ is $f^{\mathrm{h}} \in \mathrm{C}^r(\mathsf{E})$ defined by $f^{\mathrm{h}} = \pi^* f$.

(ii) The ***vertical evaluation*** of $\lambda \in \Gamma^r(\mathsf{E}^*)$ is $\lambda^{\mathrm{e}} \in \mathrm{C}^r(\mathsf{E})$ defined by $\lambda^{\mathrm{e}}(e) = \langle \lambda \circ \pi(e); e\rangle$. •

Associated with these notions we introduce the notation

$$\begin{aligned} \mathrm{Lin}^r(\mathsf{E}) &= \{F \in \mathrm{C}^r(\mathsf{E}) \mid F|\mathsf{E}_x \text{ is linear for each } x \in \mathsf{M}\}, \\ \mathrm{Aff}^r(\mathsf{E}) &= \{F \in \mathrm{C}^r(\mathsf{E}) \mid F|\mathsf{E}_x \text{ is affine for each } x \in \mathsf{M}\}. \end{aligned} \tag{2.4}$$

Clearly we have an isomorphism (of $\mathrm{C}^r(\mathsf{M})$-modules) $\lambda \mapsto \lambda^{\mathrm{e}}$ of $\mathrm{Lin}^r(\mathsf{E})$ with $\Gamma^r(\mathsf{E}^*)$. Given $F \in \mathrm{Aff}^\infty(\mathsf{E})$, there exists a unique $f \in \mathrm{C}^\infty(\mathsf{M})$ determined by

$$\begin{tikzcd} \mathsf{E} \arrow[r, "F"] \arrow[d, "\pi"'] & \mathbb{R} \\ \mathsf{M} \arrow[ur, "f"'] & \end{tikzcd}$$

Thus we also have an isomorphism (of $\mathrm{C}^r(\mathsf{M})$-modules) $\lambda \oplus f \mapsto \lambda^{\mathrm{e}} + f^{\mathrm{h}}$ of $\Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}})$ with $\mathrm{Aff}^r(\mathsf{E})$. We shall use the notation

$$(\lambda, f)^{\mathrm{e}} = \lambda^{\mathrm{e}} + f^{\mathrm{h}}. \tag{2.5}$$

These isomorphisms will arise frequently in our presentation in multiple ways.

Let us now see how to differentiate the special classes of affine functions just introduced with respect to the special classes of vector fields considered in Section 2.2.

2.10 Lemma: (Differentiating functions on vector bundles with respect to vector fields on vector bundles) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, and let ∇ be a C^r-connection in E. Let $f \in \mathrm{C}^r(\mathsf{M})$, let $\lambda \in \Gamma^r(\mathsf{E}^*)$, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $\xi \in \Gamma^r(\mathsf{E})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Then the following statements hold:*

(i) $\mathscr{L}_{X_0^{\mathrm{h}}} f^{\mathrm{h}} = (\mathscr{L}_{X_0} f)^{\mathrm{h}}$;

(ii) $\mathscr{L}_{\xi^{\mathrm{v}}} f^{\mathrm{h}} = 0$;

(iii) $\mathscr{L}_{A^{\mathrm{e}}} f^{\mathrm{h}} = 0$;

(iv) $\mathscr{L}_{X_0^{\mathrm{h}}}\lambda^{\mathrm{e}} = (\nabla_{X_0}\lambda)^{\mathrm{e}}$;

(v) $\mathscr{L}_{\xi^{\mathrm{v}}}\lambda^{\mathrm{e}} = \langle\lambda;\xi\rangle^{\mathrm{h}}$;

(vi) $\mathscr{L}_{A^{\mathrm{e}}}\lambda^{\mathrm{e}} = (A^*\lambda)^{\mathrm{e}}$.

Proof: (i) We have

$$\langle \mathrm{d}(\pi^* f)(e); X_0^{\mathrm{h}}(e)\rangle = \langle \mathrm{d}f\circ\pi(e); T_e\pi(X_0(e))\rangle = \langle \mathrm{d}f\circ\pi(e); X_0\circ\pi(e)\rangle,$$

which is the desired result.

(ii) Since f^{h} is constant on fibres of π and ξ^{v} is tangent to fibres, we have

$$f^{\mathrm{h}}(e + t\xi\circ\pi(e)) = f(e).$$

Differentiating with respect to t at $t = 0$ gives the result.

(iii) Again, f^{h} is constant on fibres and A^{e} is tangent to fibres. Thus we have

$$f^{\mathrm{h}}(\Phi_t^{A^{\mathrm{e}}}(e)) = f(e).$$

Differentiation with respect to t at $t = 0$ gives the result.

(iv) Let $e \in \mathsf{E}$ and let $t \mapsto \gamma(t)$ be the integral curve for X_0 satisfying $\gamma(0) = \pi(e)$ and let $t \mapsto \Upsilon(t)$ be the integral curve for X_0^{h} satisfying $\Upsilon(0) = e$. Then, by Lemma 2.3(iv), $t \mapsto \Upsilon(t)$ is the parallel translation of e along γ and, as such, we have $\nabla_{\gamma'(t)}\Upsilon(t) = 0$. Then

$$\begin{aligned}
\mathscr{L}_{X_0^{\mathrm{h}}}\lambda^{\mathrm{e}}(e) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \langle\lambda\circ\gamma(t);\Upsilon(t)\rangle \\
&= \langle\nabla_{\gamma'(t)}\lambda(t);\Upsilon(t)\rangle\big|_{t=0} + \langle\lambda\circ\gamma(t);\nabla_{\gamma'(t)}\Upsilon(t)\rangle\big|_{t=0} \\
&= \langle\nabla_{X_0}\lambda\circ\pi(e); e\rangle,
\end{aligned}$$

as claimed.

(v) Here, using Lemma 2.3(ii), we compute

$$\begin{aligned}
\mathscr{L}_{\xi^{\mathrm{v}}}\lambda(e) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \langle\lambda(e + t\xi\circ\pi(e)); e + t\xi\circ\pi(e)\rangle \\
&= \langle\lambda\circ\pi(e);\xi\circ\pi(e)\rangle = \langle\lambda;\xi\rangle^{\mathrm{h}}(e),
\end{aligned}$$

so completing the proof.

(vi) By Lemma 2.3(iii), we have

$$\begin{aligned}
\mathscr{L}_{A^{\mathrm{e}}}\lambda^{\mathrm{e}}(e) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \langle\lambda\circ\pi(e); \mathrm{e}^{A(\pi(e))t}(e)\rangle = \langle\lambda\circ\pi(e); A(\pi(e))(e)\rangle \\
&= \langle A^*(\lambda\circ\pi(e)); e\rangle = (A^*\lambda)^{\mathrm{e}}(e),
\end{aligned}$$

as desired. ■

2.6. The symplectic structure of the tangent bundle of a Riemannian manifold. In Section 6 we shall relate constrained variational mechanics to sub-Riemannian geometry, and in doing so it will be convenient to have at hand some nice formulae for the pull-back of the canonical symplectic structure of the cotangent bundle to the tangent bundle by the metric-canonical diffeomorphism. We follow [Paternain 1999, §1.3.2, §1.4].

Let $r \in \{\infty, \omega\}$. Let us first intrinsically describe the canonical symplectic structure of the cotangent bundle of a C^r-manifold M. We begin by describing a canonical one-form on the cotangent bundle. We define $\theta_0 \in \Gamma^r(\mathsf{T}^*\mathsf{T}^*\mathsf{M})$ by

$$\langle \theta_0(\alpha_x); X_{\alpha_x} \rangle = \langle \alpha_x; T_{\alpha_x}\pi_{\mathsf{T}^*\mathsf{M}}(X_{\alpha_x}) \rangle, \qquad \alpha_x \in \mathsf{T}^*\mathsf{M},\ X_{\alpha_x} \in \mathsf{T}_{\alpha_x}\mathsf{T}^*\mathsf{M}.$$

Let us name the one-form θ_0 and define the canonical symplectic two-form.

2.11 Definition: (Liouville one-form, symplectic two-form) For $r \in \{\infty, \omega\}$ and for a C^r-manifold M,

(i) the one-form θ_0 is the ***Liouville one-form*** and

(ii) the two-form $\omega_0 = -\mathrm{d}\theta_0$ is the ***canonical symplectic two-form***

on $\mathsf{T}^*\mathsf{M}$. •

Now let $\mathbb{G}$ be a C^r-Riemannian metric on M and consider the vector bundle isomorphism $\mathbb{G}^\flat\colon \mathsf{TM} \to \mathsf{T}^*\mathsf{M}$. The following lemma describes the pull-back of ω_0 to TM. We denote by $K_{\mathbb{G}}$ the connector associated with the Levi-Civita connection for $\mathbb{G}$ as in (2.1).

2.12 Lemma: (The canonical symplectic form on the tangent bundle of a Riemannian manifold) *Let $r \in \{\infty, \omega\}$ and let $(\mathsf{M}, \mathbb{G})$ be a C^r-Riemannian manifold. Then*

(i) $(\mathbb{G}^\flat)^*\theta_0(X_{v_x}) = \mathbb{G}(v_x, T_{v_x}\pi_{\mathsf{TM}}(X_{v_x}))$ *and*

(ii) $(\mathbb{G}^\flat)^*\omega_0(X_{v_x}, Y_{v_x}) = \mathbb{G}(T_{v_x}\pi_{\mathsf{TM}}(X_{v_x}), K_{\mathbb{G}}(Y_{v_x})) - \mathbb{G}(K_{\mathbb{G}}(X_{v_x}), T_{v_x}\pi_{\mathsf{TM}}(Y_{v_x}))$

for $v_x \in \mathsf{TM}$, $X_{v_x}, Y_{v_x} \in \mathsf{T}_{v_x}\mathsf{TM}$.

Proof: (i) We calculate

$$\begin{aligned}(\mathbb{G}^\flat)^*\theta_0(X_{v_x}) &= \theta_0(T_{v_x}\mathbb{G}^\flat(X_{v_x})) = \langle \mathbb{G}^\flat(v_x); T_{\mathbb{G}^\flat(v_x)}\pi_{\mathsf{T}^*\mathsf{M}}(T_{v_x}\mathbb{G}^\flat(X_{v_x}))\rangle \\ &= \mathbb{G}(v_x, T_{\mathbb{G}^\flat(v_x)}(\pi_{\mathsf{T}^*\mathsf{M}} \circ \mathbb{G}^\flat)(X_{v_x})) = \mathbb{G}(v_x, T_{v_x}\pi_{\mathsf{TM}}(X_{v_x})),\end{aligned}$$

as asserted.

(ii) This part of the lemma will follow if we can show that

$$\mathrm{d}((\mathbb{G}^\flat)^*\theta_0)(X_{v_x}, Y_{v_x}) = \mathbb{G}(T_{v_x}\pi_{\mathsf{TM}}(Y_{v_x}), K_{\mathbb{G}}(X_{v_x})) - \mathbb{G}(K_{\mathbb{G}}(Y_{v_x}), T_{v_x}\pi_{\mathsf{TM}}(X_{v_x})), \\ v_x \in \mathsf{TM},\ X_{v_x}, Y_{v_x} \in \mathsf{T}_{v_x}\mathsf{TM},$$

by virtue of the fact that $\mathrm{d}((\mathbb{G}^\flat)^*\theta_0) = (\mathbb{G}^\flat)^*\mathrm{d}\theta_0$ [Abraham, Marsden, and Ratiu 1988, Theorem 7.4.4].

Let $v_x \in \mathsf{TM}$ and let $X_{v_x}, Y_{v_x} \in \mathsf{T}_{v_x}\mathsf{TM}$. Let $X_0, Y_0, X_1, Y_1 \in \Gamma^r(\mathsf{TM})$ be such that

$$X_{v_x} = \mathrm{hlft}(X_0(x), v_x) + \mathrm{vlft}(X_1(x), v_x),\quad Y_{v_x} = \mathrm{hlft}(Y_0(x), v_x) + \mathrm{vlft}(Y_1(x), v_x).$$

Note that $K_{\mathbb{G}}(X_{v_x}) = X_1(x)$ and $K_{\mathbb{G}}(Y_{v_x}) = Y_1(x)$. Denote $X = X_0^{\mathrm{h}} + X_1^{\mathrm{v}}$ and $Y = Y_0^{\mathrm{h}} + Y_1^{\mathrm{v}}$. Let us also define, for $Z \in \Gamma^\infty(\mathsf{TM})$,

$$\begin{aligned}\phi_Z\colon \mathsf{TM} &\to \mathbb{R} \\ v &\mapsto \mathbb{G}(v, Z(\pi_{\mathsf{TM}}(v))).\end{aligned}$$

Let us see how to differentiate functions such as this.

1 Sublemma: *For* $X, Z \in \Gamma^\infty(\mathsf{TM})$ *we have*

(i) $\mathscr{L}_{X^{\mathrm{h}}}\phi_Z = \phi_{\overset{\mathbb{G}}{\nabla}_X Z}$ *and*

(ii) $\mathscr{L}_{X^{\mathrm{v}}}\phi_Z = \mathbb{G}(X, Z)$.

Proof: (i) Let $v \in \mathsf{TM}$ and let $\gamma\colon [0, T] \to \mathsf{M}$ be an integral curve of X through $\pi_{\mathsf{TM}}(v)$. Let $\Upsilon\colon [0, T] \to \mathsf{TM}$ be the vector field along γ defined by parallel translating v. Then, using Lemma 2.3(iv),

$$\begin{aligned}\mathscr{L}_{X^{\mathrm{h}}}\phi_Z(v) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \phi_Z(\Phi_t^{X^{\mathrm{h}}}(v)) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathbb{G}(\Upsilon(t), Z \circ \gamma(t))\\ &= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\Upsilon(t), Z \circ \gamma(t)) + \mathbb{G}(\Upsilon(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)} Z \circ \gamma(t))\\ &= \mathbb{G}(v, \overset{\mathbb{G}}{\nabla}_X Z(\pi_{\mathsf{TM}}(v))),\end{aligned}$$

as desired.

(ii) We compute, using Lemma 2.3(ii),

$$\mathscr{L}_{X^{\mathrm{v}}}\phi_Z(v) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathbb{G}(v + tX(\pi_{\mathsf{TM}}(v)), Z(\pi_{\mathsf{TM}}(v))) = \mathbb{G}(X, Z)(\pi_{\mathsf{TM}}(v)),$$

as claimed. ▼

With these preliminaries and using [Abraham, Marsden, and Ratiu 1988, Proposition 7.4.11], we calculate

$$\begin{aligned}\mathrm{d}((\mathbb{G}^\flat)^*\theta_0)(X, Y) &= \mathscr{L}_X\langle(\mathbb{G}^\flat)^*\theta_0; Y\rangle - \mathscr{L}_Y\langle(\mathbb{G}^\flat)^*\theta_0; X\rangle - \langle(\mathbb{G}^\flat)^*\theta_0; [X, Y]\rangle\\ &= \mathscr{L}_X\phi_{Y_0} - \mathscr{L}_Y\phi_{X_0} - \phi_{[X_0, Y_0]}\\ &= \phi_{\overset{\mathbb{G}}{\nabla}_{X_0} Y_0} - \phi_{\overset{\mathbb{G}}{\nabla}_{Y_0} X_0} + \mathbb{G}(X_1, Y_0) - \mathbb{G}(Y_1, X_0) - \phi_{[X_0, Y_0]}\\ &= \mathbb{G}(K_{\mathbb{G}} \circ X, Y_0) - \mathbb{G}(K_{\mathbb{G}} \circ Y, X_0),\end{aligned}$$

which is the desired assertion. ■

Paternain [1999] makes use of the Riemannian metric of Sasaki [1958] on TM to prove the preceding lemma, but, as we see, this is not necessary.

We shall denote

$$\theta_{\mathbb{G}} = (\mathbb{G}^\flat)^*\theta_0, \quad \omega_{\mathbb{G}} = (\mathbb{G}^\flat)^*\omega_0. \tag{2.6}$$

2.7. Varieties. In Section 4 we shall consider vector fields that leave subsets of manifolds and vector bundles invariant. It will be essential for our results to have the desired generality that we allow for these subsets to be more general than submanifolds and subbundles. In this section and the next three we present the sorts of objects we shall work with when discussing invariance.

First we consider subsets of manifolds we work with, generalising the notion of a submanifold.

2.13 Definition: (C^r-variety) Let $r \in \{\infty, \omega\}$ and let M be a C^r-manifold. A subset $\mathsf{S} \subseteq \mathsf{M}$ is a ***C^r-variety*** if, for any $x \in \mathsf{M}$, there exists a neighbourhood $\mathcal{U}$ of x and $f^1, \dots, f^k \in \mathrm{C}^r(\mathcal{U})$ such that

$$\mathsf{S} \cap \mathcal{U} = \bigcap_{j=1}^{k} (f^j)^{-1}(0). \qquad \bullet$$

In words, a C^r-variety is a subset that is locally the intersection of the level set of finitely many functions of class C^r. Note that, in the case of $r = \infty$, the notion of a C^∞-variety is equivalent to that of a closed set. The following lemma which proves this is well-known, but we could not find a reference for it.

2.14 Lemma: (C^∞-varieties are precisely closed sets) *If $\mathcal{U}$ is an open subset of a smooth manifold M, then there exists $f \in \mathrm{C}^\infty(\mathsf{M})$ such that $f(x) \in \mathbb{R}_{>0}$ for all $x \in \mathcal{U}$ and $f(x) = 0$ for all $x \in \mathsf{M} \setminus \mathcal{U}$.*

Proof: We shall construct f as the limit of a sequence of smooth functions converging in the weak C^∞-topology. We equip M with a Riemannian metric $\mathbb{G}$. Let $g \in \mathrm{C}^\infty(\mathsf{M})$. If $K \subseteq \mathsf{M}$ is compact and if $k \in \mathbb{Z}_{\geq 0}$, we define

$$\|g\|_{k,K} = \sup\{\|\overset{\mathbb{G}}{\nabla}{}^j g(x)\|_{\mathbb{G}} \mid x \in K,\ j \in \{0, 1, \dots, k\}\},$$

where $\|\cdot\|_{\mathbb{G}}$ indicates the norm induced on tensors by the norm associated with the Riemannian metric. One readily sees that the family of seminorms $\|\cdot\|_{k,K}$, $k \in \mathbb{Z}_{\geq 0}$, $K \subseteq \mathsf{M}$ compact, defines a locally convex topology agreeing with other definitions of the weak topology. Thus, if a sequence $(g^j)_{j \in \mathbb{Z}_{>0}}$ satisfies

$$\lim_{j \to \infty} \|g - g^j\|_{k,K} = 0, \qquad k \in \mathbb{Z}_{\geq 0},\ K \subseteq \mathsf{M} \text{ compact},$$

then g is infinitely differentiable [Michor 1980, §4.3].

We suppose that M is connected since, if it is not, we can construct f for each connected component, which suffices to give f on M. Since M is paracompact, connectedness allows us to conclude that M is second countable [Abraham, Marsden, and Ratiu 1988, Proposition 5.5.11]. Using Lemma 2.76 of [Aliprantis and Border 2006], we let $(K_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of compact subsets of $\mathcal{U}$ such that $K_j \subseteq \mathrm{int}(K_{j+1})$ for $j \in \mathbb{Z}_{>0}$ and such that $\cup_{j \in \mathbb{Z}_{>0}} K_j = \mathcal{U}$. For $j \in \mathbb{Z}_{>0}$, let $g^j \colon \mathsf{M} \to [0, 1]$ be a smooth function such that $g^j(x) = 1$ for $x \in K_j$ and $g^j(x) = 0$ for $x \in \mathsf{M} \setminus K_{j+1}$; see [Abraham, Marsden, and Ratiu 1988, Proposition 5.5.8]. Let us define $\alpha_j = \|g^j\|_{j,K_{j+1}}$ and take $\epsilon_j \in \mathbb{R}_{>0}$ to satisfy $\epsilon_j < (\alpha_j 2^j)^{-1}$. We define f by

$$f(x) = \sum_{j=1}^{\infty} \epsilon_j g^j(x),$$

and claim that f as defined satisfies the conclusions of the lemma.

First of all, since each of the functions g^j takes values in $[0, 1]$, we have

$$|f(x)| \leq \sum_{j=1}^{\infty} |\epsilon_j g^j(x)| \leq \sum_{j=1}^{\infty} \epsilon_j \|g^j\|_{0,K_{j+1}} \leq \sum_{j=1}^{\infty} \epsilon_j \|g^j\|_{j,K_{j+1}} \leq \sum_{j=1}^{\infty} \frac{1}{2^j} \leq 1,$$

and so f is well-defined and continuous by the Weierstrass M-test. If $x \in \mathcal{U}$, then there exists $N \in \mathbb{Z}_{>0}$ such that $x \in K_N$. Thus $g^N(x) = 1$ and so $f(x) \in \mathbb{R}_{>0}$. If $x \in \mathsf{M} \setminus \mathcal{U}$ then $g^j(x) = 0$ for all $j \in \mathbb{Z}_{>0}$ and so $f(x) = 0$. All that remains to show is that f is infinitely differentiable.

Let $x \in \mathsf{M}$, let $m \in \mathbb{Z}_{>0}$, and let $j \in \mathbb{Z}_{\geq 0}$ be such that $j \leq m$. If $x \notin K_{m+1}$ then g^m is zero in a neighbourhood of x, and so $\|\overset{\mathbb{G}}{\nabla}{}^j g^m(x)\|_{\mathbb{G}} = 0$. If $x \in K_{m+1}$ then

$$\begin{aligned}\|\overset{\mathbb{G}}{\nabla}{}^j g^m(x)\|_{\mathbb{G}} \leq\ & \sup\{\|\overset{\mathbb{G}}{\nabla}{}^j g^m(x')\|_{\mathbb{G}} \mid\ x' \in K_{m+1}\} \\ & \leq \sup\{\|\overset{\mathbb{G}}{\nabla}{}^j g^m(x')\|_{\mathbb{G}} \mid\ x' \in K_{m+1},\ j \in \{0, 1, \dots, m\}\} = \alpha_m.\end{aligned}$$

Thus, whenever $j \leq m$ we have $\|\overset{\mathbb{G}}{\nabla}{}^j g^m(x)\|_{\mathbb{G}} \leq \alpha_m$ for every $x \in \mathsf{N}$.

Let us define $f^m \in \mathrm{C}^\infty(\mathsf{M})$ by

$$f^m(x) = \sum_{j=1}^m \epsilon_j g^j(x).$$

Let $K \subseteq \mathsf{M}$ be compact, let $r \in \mathbb{Z}_{\geq 0}$, and let $\epsilon \in \mathbb{R}_{>0}$. Take $N \in \mathbb{Z}_{>0}$ sufficiently large that

$$\sum_{m=m_1+1}^{m_2} \frac{1}{2^m} < \epsilon,$$

for $m_1, m_2 \geq N$ with $m_1 < m_2$, this being possible by convergence of $\sum_{j=1}^\infty \frac{1}{2^j}$. Then, for $m_1, m_2 \geq N$,

$$\begin{aligned}\|f^{m_1} - f^{m_2}\|_{k,K} &= \sup\{\|\overset{\mathbb{G}}{\nabla}{}^j f^{m_1}(x) - \overset{\mathbb{G}}{\nabla}{}^j f^{m_2}(x)\|_{\mathbb{G}} \mid\ x \in K,\ j \in \{0, 1, \dots, k\}\} \\ &= \sup\left\{\left\|\sum_{m=m_1+1}^{m_2} \epsilon_m \overset{\mathbb{G}}{\nabla}{}^j g^m(x)\right\|_{\mathbb{G}} \;\middle|\; x \in K,\ j \in \{0, 1, \dots, k\}\right\} \\ &\leq \sup\left\{\sum_{m=m_1+1}^{m_2} \epsilon_m \|\overset{\mathbb{G}}{\nabla}{}^j g^m(x)\|_{\mathbb{G}} \;\middle|\; x \in K,\ j \in \{0, 1, \dots, k\}\right\} \\ &\leq \sum_{m_1+1}^{m_2} \frac{1}{2^m} < \epsilon.\end{aligned}$$

Thus, for every $k \in \mathbb{Z}_{\geq 0}$ and $K \subseteq \mathsf{M}$ compact, $(f^m)_{m \in \mathbb{Z}_{>0}}$ is a Cauchy sequence in the seminorm $\|\cdot\|_{k,K}$. Completeness of the weak C^∞-topology implies that the sequence $(f^m)_{m \in \mathbb{Z}_{>0}}$ converges to a function that is infinitely differentiable. ■

The lemma implies that C^∞-varieties are too general to expect to be able to say much about them. Indeed, in the smooth case we shall restrict ourselves to the consideration of submanifolds. However, in the real analytic case, we consider C^ω-varieties that are not submanifolds.

2.8. Generalised and cogeneralised subbundles. We shall encounter subsets of vector bundles that, like subbundles, are comprised of a union of fibres that are subspaces of fibre of the vector bundle, but, unlike subbundles, the dimension of these fibres is not locally constant. To study these sorts of objects in a systematic way, there needs to be some regularity assumptions made. In this section we present two natural forms of such regularity, both of which we shall use, and give some properties of these.

First let us make some initial definitions.

2.15 Definition: (Generalised subbundle, cogeneralised subbundle) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, and let $\mathsf{F} \subseteq \mathsf{E}$ be such that, for each $x \in \mathsf{M}$, $\mathsf{F}_x \triangleq \mathsf{E}_x \cap \mathsf{F}$ is a subspace. Denote $\Lambda(\mathsf{F}) \subseteq \mathsf{E}^*$ by asking that $\Lambda(\mathsf{F})_x \triangleq \mathsf{E}^*_x \cap \Lambda(\mathsf{F})$ be the annihilator of F_x.

(i) The subset F is a ***C^r-generalised subbundle*** if, for each $x \in \mathsf{M}$, there exists a neighbourhood $\mathcal{U}_x$ of x and C^r-sections $(\xi_i)_{i \in I_x}$ of $\mathsf{E}|\mathcal{U}_x$ such that

$$\mathsf{F}_y = \operatorname{span}_{\mathbb{R}}(\xi_i(y) \mid \ i \in I_x), \qquad y \in \mathcal{U}_x.$$

We call the sections $(\xi_i)_{i \in I_x}$ ***local generators*** for F on $\mathcal{U}_x$.

(ii) The subset F is a ***C^r-cogeneralised subbundle*** if $\Lambda(\mathsf{F})$ is a C^r-generalised subbundle.

(iii) If F is a C^r-generalised or a C^r-cogeneralised subbundle, $x \in \mathsf{M}$ is a ***regular point*** of F if there is a neighbourhood $\mathcal{U}$ of x such that $\mathsf{F}|\mathcal{U}$ has constant rank. If $x \in \mathsf{M}$ is not a regular point for F, then it is a ***singular point*** for F. •

We shall adapt some usual notation for vector bundles to generalised or cogeneralised subbundles.

2.16 Definition: (Constructions with generalised or cogeneralised subbundles) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, and let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-generalised or a C^r-cogeneralised subbundle.

(i) If $S \subseteq \mathsf{M}$, the ***restriction*** of F to S is $\mathsf{F}|S = \pi^{-1}(S) \cap \mathsf{F}$.

(ii) By $\Gamma^r(\mathsf{F})$ we denote the C^r-sections of E taking values in F:

$$\Gamma^r(\mathsf{F}) = \{\xi \in \Gamma^r(\mathsf{E}) \mid \ \xi(x) \in \mathsf{F}_x, \ x \in \mathsf{M}\}.$$

(iii) By $\mathscr{G}^r_{\mathsf{F}}$ we denote the sheaf of C^r-sections of E taking values in F: for $\mathcal{U} \subseteq \mathsf{M}$ open,

$$\mathscr{G}^r_{\mathsf{F}}(\mathcal{U}) = \{\xi \in \mathscr{G}^r_{\mathsf{E}}(\mathcal{U}) \mid \ \xi(x) \in \mathsf{F}_x, \ x \in \mathcal{U}\}.$$ •

The following result will be essential in our discussion of invariant subbundles.

2.17 Lemma: (Cogeneralised subbundles are closed) *Let $r \in \{\infty, \omega\}$ and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. If $\mathsf{F} \subseteq \mathsf{E}$ is a C^r-cogeneralised subbundle, then F is closed.*

Proof: We consider the smooth and real analytic cases separately.

First, in the smooth case, we assume that M is connected, since, if it is not, the argument we give can be applied to each connected component. Note that Sussmann [2008] proves that a smooth generalised subbundle has a finite set of global generators. That is, if $\mathsf{G} \subseteq \mathsf{E}$ is a generalised subbundle, then there are $\xi_1, \dots, \xi_k \in \Gamma^\infty(\mathsf{E})$ such that

$$\mathsf{G}_x = \operatorname{span}_{\mathbb{R}}(\xi_1(x), \dots, \xi_k(x)), \qquad x \in \mathsf{M}.$$

In this case, we see that G is the image of the vector bundle map

$$\begin{aligned}\Phi\colon \mathbb{R}^k_{\mathsf{M}} &\to \mathsf{E}\\ ((a_1,\dots,a_k),x) &\mapsto a_1\xi_1(x)+\dots+a_k\xi_k(x).\end{aligned}$$

Now, since F is a smooth cogeneralised subbundle, $\Lambda(\mathsf{F})$ is a smooth generalised subbundle. Thus $\Lambda(\mathsf{F}) = \operatorname{image}(\Phi)$ for a vector bundle mapping Φ as above. Since $\Lambda(\operatorname{image}(\Phi)) = \ker(\Phi^*)$, we have

$$\mathsf{F} \simeq \Lambda(\Lambda(\mathsf{F})) = \ker(\Phi^*).$$

Thus F is the preimage of the zero section of a smooth vector bundle under a smooth vector bundle map. Thus F is closed since the zero section is closed.

In the real analytic case, one can show, with a great deal of work, that, for a real analytic generalised subbundle G and for $x \in \mathsf{M}$, there exists a neighbourhood $\mathcal{U}$ of x and sections $\xi_1,\dots,\xi_k \in \Gamma^\omega(\mathsf{E})$ such that

$$\mathsf{G}_x = \operatorname{span}_{\mathbb{R}}(\xi_1(x),\dots,\xi_k(x)), \qquad x \in \mathcal{U},$$

[Lewis 2012, Theorem 5.2]. Using this result and the same arguments as in the smooth case above, it follows that, for each $x \in \mathsf{M}$, there is a neighbourhood $\mathcal{U}$ of x such that $\mathsf{F}|\mathcal{U}$ is a closed subset of $\mathsf{E}|\mathcal{U}$. To show that F is then closed, let $(e_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in F that converges to $e \in \mathsf{E}$. There is then a neighbourhood $\mathcal{U}$ of $\pi(e)$ such that $e_j \in \pi^{-1}(\mathcal{U})$ for j sufficiently large. Since $\mathsf{F}|\mathcal{U}$ is closed (possibly after shrinking $\mathcal{U}$), it follows that $e \in \mathsf{F}$, and so F is closed. ■

For generalised subbundles, the proof of the preceding lemma immediately gives the following result.

2.18 Corollary: (The fibres of a generalised subbundle are generated by global sections) *Let $r \in \{\infty,\omega\}$ and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. If $\mathsf{F} \subseteq \mathsf{E}$ is a C^r-generalised subbundle, then*

$$\mathsf{F}_x = \{\xi(x) \mid \xi \in \Gamma^r(\mathsf{F})\}, \qquad x \in \mathsf{M}.$$

More specifically, for each $x \in \mathsf{M}$, there exists a neighbourhood $\mathcal{U}$ of x and $\xi_1,\dots,\xi_k \in \Gamma^r(\mathsf{F})$ such that

$$\mathsf{F}_y = \operatorname{span}_{\mathbb{R}}(\xi_1(y),\dots,\xi_k(y)), \qquad y \in \mathcal{U}.$$

For cogeneralised subbundles, the proof of the lemma gives the following.

2.19 Corollary: (Cogeneralised subbundles are varieties) *Let $r \in \{\infty,\omega\}$ and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. If $\mathsf{F} \subseteq \mathsf{E}$ is a C^r-cogeneralised subbundle, then it is a C^r-variety.*

One could, therefore, say that cogeneralised subbundles are "linear varieties." We shall expand on this idea below when we discuss affine subbundle varieties.

An important feature of the definitions is the following characterisation of the set of regular points for a generalised or cogeneralised subbundle.

2.20 Lemma: (Regular points for generalised and cogeneralised subbundles) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$, and let F be a C^r-generalised or a C^r-cogeneralised subbundle of E. Then there exists an open dense subset $\mathcal{U} \subseteq \mathsf{M}$ such that $\mathsf{F}|\mathcal{U}$ is a C^r-subbundle of $\mathsf{E}|\mathcal{U}$.*

Proof: We suppose that M is connected so that E has constant fibre dimension, say m. The lemma will follow in the general case by applying the proof here to each connected component of M.

First suppose that F is a C^r-generalised subbundle. For $j \in \{0, 1, \dots, m+1\}$, denote

$$\mathcal{V}_j = \{x \in \mathsf{M} \mid \ \dim_{\mathbb{R}}(\mathsf{F}_x) \geq j\};$$

note that $\mathcal{V}_{m+1} = \varnothing$. Denote $\mathcal{U}_j = \operatorname{int}(\mathcal{V}_j \setminus \mathcal{V}_{j+1})$, $j \in \{0, 1, \dots, m\}$, and $\mathcal{U} = \cup_{j=0}^m \mathcal{U}_j$. We will show that $\mathcal{U}$ satisfies the conclusions of the lemma.

First we claim that $\mathcal{V}_j$ is open for each $j \in \{0, 1, \dots, m+1\}$. Let $x \in \mathsf{M}$ and suppose that $\dim_{\mathbb{R}}(\mathsf{F}_x) \geq j$. Let $\mathcal{U}_x$ be a neighbourhood of x and let $(\xi_i)_{i \in I_x}$ be local generators for F on $\mathcal{U}_x$. Then there exist $i_1, \dots, i_j \in I_x$ so that $\xi_{i_1}(x), \dots, \xi_{i_j}(x)$ are linearly independent. By continuity, $\xi_{i_1}(y), \dots, \xi_{i_j}(y)$ are linearly independent for y in some neighbourhood of x. Thus $\mathcal{V}_j$ is open.

It is clear that $\mathcal{V}_{j+1} \subseteq \mathcal{V}_j$.

Let us show that $\mathcal{V}_j \setminus \operatorname{bd}(\mathcal{V}_{j+1})$ is open. Let $x \in \mathcal{V}_j \setminus \operatorname{bd}(\mathcal{V}_{j+1})$. Then either $x \in \mathcal{V}_{j+1}$ or $x \in \mathcal{V}_j \setminus \operatorname{cl}(\mathcal{V}_{j+1})$. If $x \in \mathcal{V}_{j+1}$, then there is a neighbourhood of x in $\mathcal{V}_{j+1}$, just by openness of $\mathcal{V}_{j+1}$. If $x \in \mathcal{V}_j \setminus \operatorname{cl}(\mathcal{V}_{j+1})$, then there is a neighbourhood of x in $\mathcal{V}_j \setminus \operatorname{cl}(\mathcal{V}_{j+1})$, again just by openness of $\mathcal{V}_j \setminus \operatorname{cl}(\mathcal{V}_{j+1})$. Thus $\mathcal{V}_j \setminus \operatorname{bd}(\mathcal{V}_{j+1})$ is open in $\mathcal{V}_j$.

Let us show that $\mathcal{V}_j \setminus \operatorname{bd}(\mathcal{V}_{j+1})$ is dense in $\mathcal{V}_j$. Let $x \in \mathcal{V}_j$ and let $\mathcal{N}$ be a neighbourhood of x. We have three mutually exclusive cases.

1. If $x \notin \operatorname{cl}(\mathcal{V}_{j+1})$, then there is a neighbourhood of x in $\mathcal{V}_j \setminus \operatorname{cl}(\mathcal{V}_{j+1})$ which, therefore, necessarily intersects $\mathcal{N}$. Thus $x \in \operatorname{cl}(\mathcal{V}_j \setminus \operatorname{bd}(\mathcal{V}_{j+1}))$.
2. If $x \in \mathcal{V}_{j+1}$, then there is a neighbourhood of x in $\mathcal{V}_{j+1}$. Thus $x \in \operatorname{cl}(\mathcal{V}_j \setminus \operatorname{bd}(\mathcal{V}_{j+1}))$.
3. Finally, if $x \in \operatorname{bd}(\mathcal{V}_{j+1})$, obviously $x \in \operatorname{cl}(\mathcal{V}_j \setminus \operatorname{bd}(\mathcal{V}_{j+1}))$.

Note that, if $\dim_{\mathbb{R}}(\mathsf{F}_x) = j$, then $x \in \mathcal{C}_j \triangleq \mathcal{V}_j \setminus \mathcal{V}_{j+1}$, $j \in \{0, 1, \dots, k\}$. We have $\mathcal{U}_j = \operatorname{int}(\mathcal{C}_j)$. Since $\mathsf{F}|\mathcal{U}_j$ has rank j, $\mathsf{F}|\mathcal{U}_j$ is a C^r-subbundle of $\mathsf{E}|\mathcal{U}_j$ since, from any local generators for F in a neighbourhood of $x \in \mathcal{U}_j$, we can find j of them that are a local basis for sections. Thus $\mathsf{F}|\mathcal{U}$ is also a C^r-subbundle of $\mathsf{E}|\mathcal{U}$.

It remains to show that $\mathcal{U}$ is open and dense in M. Being a union of the open sets $\mathcal{U}_0, \mathcal{U}_1, \dots, \mathcal{U}_m$, it is certainly open. Now let $x \in \mathsf{M} \setminus \mathcal{U}$. Then $x \in \mathcal{C}_j$ for some $j \in \{0, 1, \dots, m\}$. This means that $x \in \operatorname{bd}(\mathcal{V}_{j+1})$. Thus any neighbourhood of x intersects at least one of $\mathcal{V}_j$ or $\mathcal{V}_{j+1}$, whence it intersects at least one of $\mathcal{U}_j$ or $\mathcal{U}_{j+1}$.

This gives the lemma when F is a C^r-generalised subbundle. If F is a C^r-cogeneralised subbundle, then $\Lambda(\mathsf{F})$ is a C^r-generalised subbundle. Thus, as we just showed, there is an open dense subset $\mathcal{U} \subseteq \mathsf{M}$ such that $\Lambda(\mathsf{F})|\mathcal{U}$ is a C^r-subbundle of $\mathsf{E}^*|\mathcal{U}$. Thus, for $x \in \mathcal{U}$, there is a neighbourhood $\mathcal{V} \subseteq \mathcal{U}$ of x and $k \in \{0, 1, \dots, m\}$ such that $\dim_{\mathbb{R}}(\Lambda(\mathsf{F})_y) = k$ for $y \in \mathcal{V}$. Therefore, $\dim_{\mathbb{R}}(\mathsf{F}_y) = m - k$ for $y \in \mathcal{V}$, and so $\mathsf{F}|\mathcal{U}$ has locally constant rank, and so is a C^r-subbundle of $\mathsf{E}|\mathcal{U}$. ■

The following result gives an instance of generalised and cogeneralised subbundles.

2.21 Lemma: (The kernel and image of a vector bundle map are generalised and cogeneralised subbundles) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ and $\theta\colon \mathsf{F} \to \mathsf{M}$ be C^r-vector bundles, and let $\Phi\colon \mathsf{E} \to \mathsf{F}$ be a C^r-vector bundle mapping. Then the following statements hold:*

(i) $\operatorname{image}(\Phi)$ *is a C^r-generalised subbundle;*

(ii) $\ker(\Phi)$ *is a C^r-cogeneralised subbundle.*

Proof: (i) Let $\mathcal{U} \subseteq \mathsf{M}$ be an open set for which there exists a basis $\xi_1, \dots, \xi_k \in \Gamma^r(\mathsf{E}|\mathcal{U})$ of sections of E over $\mathcal{U}$. Then $\Phi \circ \xi_1, \dots, \Phi \circ \xi_k$ are local generators for $\operatorname{image}(\Phi|(\mathsf{E}|\mathcal{U}))$.

(ii) We note that $\ker(\Phi) = \Lambda(\operatorname{image}(\Phi^*))$. ■

2.9. Generalised and cogeneralised affine subbundles. By virtue of Lemma 2.21, one can think of the subbundles of Section 2.8 as being either sets of linear *equations* in vector bundles (in the case of generalised subbundles) or the sets of *solutions* of linear equations (in the cogeneralised case). In this section we extend this to sets of affine equations. In the next section we will consider sets of solutions to affine equations.

Our first definition is the following.

2.22 Definition: (Generalised affine subbundle) Let $r \in \{\infty, \omega\}$ and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. A subset $\mathsf{B} \subseteq \mathsf{E}$ is a ***C^r-generalised affine subbundle*** if, for each $x \in \mathsf{M}$, there exists a neighbourhood $\mathcal{U}_x$ of x and C^r-sections $(\xi_i)_{i \in I_x}$ of $\mathsf{E}|\mathcal{U}_x$ such that

$$\mathsf{B} \cap \mathsf{E}_y = \operatorname{aff}_{\mathbb{R}}(\xi_i(y)|\ i \in I_x), \qquad y \in \mathcal{U}_x.$$

We call the sections $(\xi_i)_{i \in I_x}$ ***local generators*** for B on $\mathcal{U}_x$. We denote $\mathsf{B}_x = \mathsf{B} \cap \mathsf{E}_x$, $x \in \mathsf{M}$. •

The following characterisation of generalised affine subbundles is one we shall frequently use.

2.23 Lemma: (Characterisation of generalised affine subbundles) *Let $r \in \{\infty, \omega\}$ and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. For a subset $\mathsf{B} \subseteq \mathsf{E}$, the following statements are equivalent:*

(i) B *is a C^r-generalised affine subbundle;*

(ii) there exists $\xi_0 \in \Gamma^r(\mathsf{E})$ and a C^r-generalised subbundle $\mathsf{F} \subseteq \mathsf{E}$ such that

$$\mathsf{B} \cap \mathsf{E}_x = \xi_0(x) + \mathsf{F}_x, \qquad x \in \mathsf{M}.$$

Proof: We first prove a few simple linear algebraic facts. In the following, we shall take as our definition of an affine subspace of a vector space that by which an affine subspace is such that it contains the bi-infinite line passing through any two points.

1 Sublemma: *Let V be a $\mathbb{R}$-vector space. The following statements hold:*

(i) a subset $\mathsf{A} \subseteq \mathsf{V}$ is an affine subspace if and only if there exists $v_0 \in \mathsf{A}$ and a subspace $\mathsf{U} \subseteq \mathsf{V}$ such that $\mathsf{A} = v_0 + \mathsf{U}$;

(ii) if, for an affine subspace $\mathsf{A} \subseteq \mathsf{V}$, we have

$$\mathsf{A} = v_0 + \mathsf{U} = v_0' + \mathsf{U}'$$

for $v_0, v_0' \in \mathsf{A}$ and for subspaces $\mathsf{U}, \mathsf{U}' \subseteq \mathsf{V}$, we have $\mathsf{U} = \mathsf{U}'$ and $\pi_{\mathsf{U}}(v_0) = \pi_{\mathsf{U}}(v_0')$, where $\pi_{\mathsf{U}}\colon \mathsf{V} \to \mathsf{V}/\mathsf{U}$ is the canonical projection.

Proof: (i) Let $v_0 \in \mathsf{A}$ and define $\mathsf{U} = \{v - v_0 \mid \ v \in \mathsf{A}\}$. The result will be proved if we prove that U is a subspace. Let $v - v_0 \in U$ for some $v \in \mathsf{A}$ and let $a \in \mathbb{R}$. Then

$$a(v - v_0) = av + (1-a)v_0 - v_0,$$

and so $a(v - v_0) \in \mathsf{U}$ since $av + (1-a)v_0 \in \mathsf{A}$. For $v_1 - v_0, v_2 - v_0 \in \mathsf{U}$ with $v_1, v_2 \in \mathsf{A}$ we have

$$(v_1 - v_0) + (v_2 - v_0) = (v_1 + v_2 - v_0) - v_0.$$

Thus we will have $(v_1 - v_0) + (v_2 - v_0) \in \mathsf{U}$ if we can show that $v_1 + v_2 - v_0 \in \mathsf{A}$. However, we have

$$\begin{aligned} & v_1 - v_0, v_2 - v_0 \in \mathsf{U}, \\ \implies \quad & 2(v_1 - v_0), 2(v_2 - v_0) \in \mathsf{U}, \\ \implies \quad & 2(v_1 - v_0) + v_0, 2(v_2 - v_0) + v_0 \in \mathsf{A}, \\ \implies \quad & \tfrac{1}{2}(2(v_1 - v_0) + v_0) + \tfrac{1}{2}(2(v_2 - v_0) + v_0) \in \mathsf{A}, \end{aligned}$$

which gives the result after we notice that

$$\tfrac{1}{2}(2(v_1 - v_0) + v_0) + \tfrac{1}{2}(2(v_2 - v_0) + v_0) = v_1 + v_2 - v_0.$$

(ii) The equality

$$\{v_0 + u \mid \ u \in \mathsf{U}\} = \{v_0' + u' \mid \ u' \in \mathsf{U}'\}$$

implies that $v_0 = v_0' + u'$ for some $u' \in \mathsf{U}'$. Thus $v_0 - v_0' \in \mathsf{U}'$. In similar manner, $v_0' - v_0 \in \mathsf{U}$. Now let $u' \in \mathsf{U}'$. Thus

$$v_0' + u' = v_0 + u \qquad \implies \qquad u' = u + v_0 - v_0'$$

for some $u \in \mathsf{U}$, and so $u' \in \mathsf{U}$. Thus $\mathsf{U}' \subseteq \mathsf{U}$. As the opposite inclusion is established similarly, we have $\mathsf{U} = \mathsf{U}'$. We also have

$$\{v_0 + u \mid \ u \in \mathsf{U}\} = \{v_0' + (v_0 - v_0') + u \mid \ u \in \mathsf{U}\} = \{v_0' + u \mid \ u \in \mathsf{U}\},$$

as desired. In particular, $v_0 + 0 = v_0' + u$ for some $u \in \mathsf{U}$ and so $v_0 - v_0' \in \mathsf{U}$, as desired. ▼

Now we proceed with the proof.

Suppose that B is a C^r-generalised affine subbundle. Let $\mathscr{U} = (\mathcal{U}_a)_{a \in A}$ be an open cover for M such that, for each $a \in A$, we have local generators $(\xi_{ai})_{i \in I_a}$ for B on $\mathcal{U}_a$. For $a \in A$, fix $i_0 \in I_a$ and denote $\xi_{a0} = \xi_{ai_0}$. As in the first part of the sublemma, for $x \in \mathcal{U}_a$, we have $\mathsf{B}_x = \xi_{a0}(x) + \mathsf{F}_x$, where

$$\mathsf{F}_x = \operatorname{span}_{\mathbb{R}}(\xi_{ai}(x) - \xi_{a0}(x) \mid \ \ i \in I_a).$$

By the second part of the sublemma, the subspace F_x is well-defined, independently of the choice of $a \in A$ for which $x \in \mathcal{U}_a$. Note that this then defines a C^r-generalised subbundle F. If $\mathcal{U}_a \cap \mathcal{U}_b \neq \varnothing$, then the second part of the sublemma gives $\xi_{a0}(x) - \xi_{b0}(x) \in \mathsf{F}_x$ for $x \in \mathcal{U}_a \cap \mathcal{U}_b$. Said otherwise,

$$\xi_{a0}|\mathcal{U}_a \cap \mathcal{U}_b - \xi_{b0}|\mathcal{U}_a \cap \mathcal{U}_b \in \mathscr{G}^r_{\mathsf{F}}(\mathcal{U}_a \cap \mathcal{U}_b). \tag{2.7}$$

Said yet otherwise, if $\pi_{\mathsf{F}}\colon \mathscr{G}^r_{\mathsf{E}} \to \mathscr{G}^r_{\mathsf{E}}/\mathscr{G}^r_{\mathsf{F}}$ is the projection onto the quotient sheaf, then

$$\pi_{\mathsf{F}}(\xi_{a0})|\mathcal{U}_a \cap \mathcal{U}_b = \pi_{\mathsf{F}}(\xi_{b0})|\mathcal{U}_a \cap \mathcal{U}_b.$$

Since $\mathscr{G}^r_{\mathsf{E}}/\mathscr{G}^r_{\mathsf{F}}$ is a sheaf, there exists $\sigma \in (\mathscr{G}^r_{\mathsf{E}}/\mathscr{G}^r_{\mathsf{F}})(\mathsf{M})$ such that

$$\sigma|\mathcal{U}_a = \pi_{\mathsf{F}}(\xi_{a0}), \qquad a \in A.$$

We will show that $\sigma = \pi_{\mathsf{F}}(\xi_0)$, where $\xi_0 \in \Gamma^r(\mathsf{E})$ is such that

$$\pi_{\mathsf{F}}(\xi_0)|\mathcal{U}_a = \pi_{\mathsf{F}}(\xi_{a0}), \qquad a \in A,$$

and, by the second part of the sublemma, this will establish that $\mathsf{B}_x = \xi_0 + \mathsf{F}_x$.

We will use constructions from sheaf and Čech cohomology, and we refer to [Ramanan 2005, §4.5] for the background notions.

To do this, we first claim that, for any open set $\mathcal{U} \subseteq \mathsf{M}$, the sheaf $\mathscr{G}^r_{\mathsf{F}}|\mathcal{U}$ is acyclic. In the smooth case, this follows from [Wells Jr. 2008, Proposition 3.11] (along with [Wells Jr. 2008, Examples 3.4(d, e)] and [Wells Jr. 2008, Proposition 3.5]). In the real analytic case, we note that $\mathscr{G}^\omega_{\mathsf{F}}$ is coherent by [Lewis 2012, Corollary 4.11]. Thus $\mathscr{G}^\omega_{\mathsf{F}}|\mathcal{U}$ is acyclic in the real analytic case by Cartan's Theorem B [Cartan 1957, Proposition 6].

It follows, therefore, by Leray's Theorem [Ramanan 2005, Theorem 5.3] that the Čech cohomology $\check{\mathrm{H}}^1(\mathscr{U};\mathscr{G}^r_{\mathsf{F}})$ vanishes.

The 0-cochain $(\xi_{a0})_{a\in A} \in \check{\mathrm{C}}^0(\mathscr{U};\mathscr{G}^r_{\mathsf{E}})$ satisfies (2.7), and so, keeping in mind that $\ker(\pi_{\mathsf{F}}) = \mathscr{G}^r_{\mathsf{F}}$, this means that $(\pi_{\mathsf{F}}(\xi_{a0}))_{a\in A} \in \check{\mathrm{Z}}^1(\mathscr{U};\ker(\pi_{\mathsf{F}}))$, the Čech 1-cocycles of the kernel sheaf relative to the open cover. By the vanishing of $\check{\mathrm{H}}^1(\mathscr{U};\mathscr{G}^r_{\mathsf{F}})$, we thus have a 1-coboundary $(\eta_a)_{a\in A} \in \check{\mathrm{C}}^0(\mathscr{U},\ker(\pi_{\mathsf{F}}))$ such that

$$\eta_b|\mathcal{U}_a \cap \mathcal{U}_b - \eta_a|\mathcal{U}_a \cap \mathcal{U}_b = \xi_{a0}|\mathcal{U}_a \cap \mathcal{U}_b - \xi_{b0}|\mathcal{U}_a \cap \mathcal{U}_b.$$

Let $\zeta_a \in \mathscr{G}^r_{\mathsf{E}}(\mathcal{U}_a)$ be given by $\zeta_a = \xi_{a0} + \eta_a$ and note that

$$\zeta_a|\mathcal{U}_a \cap \mathcal{U}_b = (\xi_{a0} + \eta_a)|\mathcal{U}_a \cap \mathcal{U}_b = (\xi_{b0} + \eta_b)|\mathcal{U}_a \cap \mathcal{U}_b = \zeta_b|\mathcal{U}_a \cap \mathcal{U}_b.$$

Since $\mathscr{G}^r_{\mathsf{F}}$ is a sheaf, there exists $\xi_0 \in \mathscr{G}^r_{\mathsf{E}}(\mathsf{M})$ such that $\xi_0|\mathcal{U}_a = \zeta_a$, $a \in A$. Moreover,

$$\pi_{\mathsf{F}}(\xi_0)|\mathcal{U}_a = \pi_{\mathsf{F}}(\zeta_a) = \pi_{\mathsf{F}}(\xi_{a0} + \eta_a) = \pi_{\mathsf{F}}(\xi_{a0}),$$

which gives this implication of the lemma.

For the other, suppose that we have $\xi_0 \in \Gamma^r(\mathsf{E})$ and a C^r-generalised subbundle $\mathsf{F} \subseteq \mathsf{E}$ such that

$$\mathsf{B} \cap \mathsf{E}_x = \xi_0(x) + \mathsf{F}_x, \qquad x \in \mathsf{M}.$$

Then, by the first part of the sublemma, for each $x \in \mathsf{M}$, there exists a neighbourhood $\mathcal{U}_x$ of x such that

$$\mathsf{B} \cap \mathsf{E}_x = \operatorname{aff}_{\mathbb{R}}(\xi_0(x) + \xi_i(x) \mid \ i \in I_x),$$

where $(\xi_i)_{i\in I_x}$ are local generators for F on $\mathcal{U}_x$. Thus $(\xi_0|\mathcal{U}_x + \xi_i)_{i\in I}$ are local generators for an affine subbundle which equals B. ■

The generalised subbundle F is called the ***linear part*** of the generalised affine subbundle B and is denoted by $L(\mathsf{B})$.

Based on the lemma, we make the following definition.

2.24 Definition: (Cogeneralised affine subbundle) Let $r \in \{\infty, \omega\}$ and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. A subset $\mathsf{B} \subseteq \mathsf{E}$ is a ***C^r-cogeneralised affine subbundle*** if there exists $\xi_0 \in \Gamma^r(\mathsf{E})$ and a C^r-cogeneralised subbundle $\mathsf{F} \subseteq \mathsf{E}$ such that

$$\mathsf{B}_x \triangleq \mathsf{B} \cap \mathsf{E}_x = \xi_0(x) + \mathsf{F}_x, \qquad x \in \mathsf{M}. \qquad \bullet$$

Many properties of generalised or cogeneralised affine subbundles are immediately deduced from the corresponding properties of their linear part, which is a generalised or cogeneralised subbundle. We shall freely use such properties.

We will require the analogue of $\Lambda(\mathsf{F})$ for a generalised or cogeneralised subbundle. Let us first indicate this just on the level of linear algebra.

2.25 Lemma: (Affine functions whose zeros prescribe an affine subspace) *Let V be a $\mathbb{R}$-vector space and let $\mathsf{B} \subseteq \mathsf{V}$ be an affine subspace. Let $u_0 \in \mathsf{V}$ be such that $\mathsf{B} = u_0 + L(\mathsf{B})$. Then*

$$\mathsf{B} = \{v \in \mathsf{V} \mid \ \alpha(v) - \langle \alpha; u_0 \rangle = 0,\ \alpha \in \Lambda(L(\mathsf{B}))\}.$$

Proof: We have

$$\begin{aligned}\mathsf{B} &= \{u_0 + u \mid \ u \in L(\mathsf{B})\} = \{v \in \mathsf{V} \mid \ v - u_0 \in L(\mathsf{B})\} \\ &= \{v \in \mathsf{V} \mid \ \alpha(v - u_0) = 0,\ \alpha \in \Lambda(L(\mathsf{B}))\},\end{aligned}$$

as claimed. ■

Based on the lemma, for a C^r-generalised or a C^r-cogeneralised affine subbundle B of a C^r-vector bundle $\pi\colon \mathsf{E} \to \mathsf{M}$, define $\Lambda_{\mathrm{aff}}(\mathsf{B}) \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ by

$$\Lambda_{\mathrm{aff}}(\mathsf{B})_x = \{(\alpha, -\langle \alpha; u \rangle) \mid \ \alpha \in \Lambda(L(\mathsf{B}))_x\}, \tag{2.8}$$

where $u \in \mathsf{E}_x$ is such that $\mathsf{B}_x = u + L(\mathsf{B})_x$. We regard $(\alpha, -\langle \alpha; u \rangle) \in \Lambda_{\mathrm{aff}}(\mathsf{B})_x$ to be an affine function F_α on E_x by

$$F_\alpha(v) = \langle \alpha; v \rangle - \langle \alpha; u \rangle.$$

If $\lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))$, then we let $F_\lambda\colon \mathsf{E} \to \mathbb{R}$ be defined by $F_\lambda|\mathsf{E}_x = F_{\lambda(x)}$. Note that

$$F_\lambda = \lambda^{\mathrm{e}} - \langle \lambda; \xi_0 \rangle^{\mathrm{h}} \tag{2.9}$$

if $\mathsf{B} = \xi_0 + L(\mathsf{B})$.

We then have the following analogue of Corollary 2.18.

2.26 Lemma: (Affine functions defining cogeneralised affine subbundles) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, and let $\mathsf{B} \subseteq \mathsf{E}$ be a C^r-cogeneralised affine subbundle given by $\xi_0 + L(\mathsf{B})$ for $\xi_0 \in \Gamma^r(\mathsf{E})$. Then, for each $x \in \mathsf{M}$,*

$$\mathsf{B}_x = \{v \in \mathsf{E}_x \mid \ F_\lambda(v) = 0,\ \lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))\}.$$

Proof: The hypotheses imply that $\Lambda(L(\mathsf{B}))$ is a C^r-generalised subbundle of E^*. Let $x \in \mathsf{M}$. By Corollary 2.18,

$$\Lambda(L(\mathsf{B}))_x = \{\lambda(x) \mid \ \lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))\}.$$

Therefore,

$$\Lambda_{\mathrm{aff}}(\mathsf{B})_x = \{(\lambda(x), -\langle \lambda(x); \xi_0(x) \rangle) \mid \ \lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))\}$$

and so, by Lemma 2.25,

$$\begin{aligned}\mathsf{B}_x &= \{v \in \mathsf{E}_x \mid \ \lambda(x)(v) - \langle\lambda(x);\xi_0(x)\rangle,\ \lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))\} \\ &= \{v \in \mathsf{E}_x \mid \ F_\lambda(v) = 0,\ \lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))\},\end{aligned}$$

as claimed. ■

There is an important point that should be made here concerning the rôle played by the notion of a cogeneralised affine subbundle.

2.27 Remark: (On cogeneralised affine subbundles I) We note that a cogeneralised affine subbundle is not quite the natural idea of something dual to a generalised affine subbundle. Indeed, while for generalised and cogeneralised subbundles, one has the equation/solution duality, if one thinks of generalised affine subbundles as being affine equations, then cogeneralised affine subbundles *do not* play the rôle of the corresponding solutions. We shall have more to say on this in Remark 2.34 below. •

The preceding remark notwithstanding, we shall discuss cogeneralised subbundles in some detail, since in Section 4 we shall present a theory for invariant subbundles. Since this is a subject that seems to not have received much consideration in the literature, it seems worthwhile to be comprehensive about this.

2.10. Affine subbundle varieties. With the closing Remark 2.27 of the preceding section in mind, let us turn to a discussion of what *should* be regarded as the object naturally dual to a generalised affine subbundle, and to a sort of object which features prominently in our results of Section 4 and of the application of these results in Section 7 to the problem of comparison of trajectories to nonholonomic and constrained variational systems.

To set the groundwork, let us consider a little linear algebra. In particular, we wish to recast the classical notion of a system of linear inhomogeneous algebraic equations. Thus let V be a finite-dimensional $\mathbb{R}$-vector space, and let $A \in \mathrm{End}(\mathsf{V})$ and $b \in \mathsf{V}$. We denote by

$$\mathrm{Sol}(A, b) = \{v \in \mathsf{V} \mid \ A(v) + b = 0\}$$

the set of solutions to the corresponding system of inhomogeneous linear equations. Note that

$$\begin{aligned}\mathrm{Sol}(A, b) &= \{v \in \mathsf{V} \mid \ \langle\lambda; A(v)\rangle + \langle\lambda; b\rangle = 0,\ \lambda \in \mathsf{V}^*\} \\ &= \{v \in \mathsf{V} \mid \ \langle A^*(\lambda); v\rangle + \langle\lambda; b\rangle = 0,\ \lambda \in \mathsf{V}^*\},\end{aligned}$$

leading us to define the subspace

$$\mathrm{Sol}^*(A, b) = \{(A^*(\lambda), \langle\lambda; b\rangle) \in \mathsf{V}^* \oplus \mathbb{R} \mid \ \lambda \in \mathsf{V}^*\}$$

of $\mathsf{V}^* \oplus \mathbb{R}$. Note that this subspace is distinguished by having positive codimension. The following lemma indicates the importance of this condition, and as well characterises the conditions for existence of solutions using subspaces of $\mathsf{V}^* \oplus \mathbb{R}$.

2.28 Lemma: (Systems of linear equations and subspaces of $\mathsf{V}^* \oplus \mathbb{R}$) *For a finite-dimensional $\mathbb{R}$-vector space V, the following statements concerning a subspace $\Delta \subseteq \mathsf{V}^* \oplus \mathbb{R}$ are equivalent:*

(i) Δ has positive codimension;

(ii) there exists $A \in \mathrm{End}(\mathsf{V})$ and $b \in \mathsf{V}$ such that $\Delta = \mathrm{Sol}^*(A, b)$.

Moreover, if $\Delta \subseteq \mathsf{V}^* \oplus \mathbb{R}$ is a subspace of positive codimension and if $A \in \mathrm{End}(\mathsf{V})$ and $b \in \mathsf{V}$ are such that $\Delta = \mathrm{Sol}^*(A, b)$, then

(iii) $\mathrm{Sol}(A, b) = \{v \in \mathsf{V} \mid (v, 1) \in \Lambda(\Delta)\}$,

(iv) $\mathrm{Sol}(A, b) \neq \varnothing$ if and only $(0, 1) \notin \Delta$.

Finally,

(v) if $A \in \mathrm{End}(\mathsf{V})$ and if $b \in \mathsf{V}$, then there exists a unique subspace $\Delta \subseteq \mathsf{V} \oplus \mathbb{R}$ such that

$$\mathrm{Sol}(A, b) = \{v \in \mathsf{V} \mid (v, 1) \in \Lambda(\Delta)\}.$$

Proof: The equivalence of (i) and (ii) is easily established. Indeed, as $\dim(\Delta) \leq \dim(\mathsf{V}^*)$, there exists a surjective linear map $L \in \mathrm{L}(\mathsf{V}^*; \Delta)$. As $\Delta \subseteq \mathsf{V}^* \oplus \mathbb{R}$, this means that we can write $L(\lambda, a) = (A^*(\lambda), \langle \lambda; b \rangle)$ for some $A \in \mathrm{End}(\mathsf{V})$ and $b \in \mathsf{V}$.

For the parts (iii) and (iv) of the lemma, suppose that Δ, A, and b are as posited.

(iii) We compute

$$\mathrm{Sol}(A, b) = \{v \in \mathsf{V} \mid \langle A^*(\lambda); v \rangle + \langle \lambda; b \rangle = 0\} = \{v \in \mathsf{V} \mid (v, 1) \in \Lambda(\Delta)\}.$$

(iv) Suppose that $(0, 1) \in \Delta$ and that $\mathrm{Sol}(A, b) \neq \varnothing$. Then, for $v \in \mathrm{Sol}(A, b)$ we have

$$1 = \langle (0, 1); (v, 1) \rangle = 0,$$

which contradiction ensures that $\mathrm{Sol}(A, b) = \varnothing$.

For the converse, suppose that $(0, 1) \notin \Delta$. Since Δ is a subspace, this implies that $(0, a) \notin \Delta$ for any $a \neq 0$. Since $\Delta = \mathrm{Sol}^*(A, b)$, the definition of $\mathrm{Sol}^*(A, b)$ means that

$$\lambda \in \ker(A^*) \implies \langle \lambda; b \rangle = 0.$$

Thus b annihilates $\ker(A^*)$ and so $b \in \mathrm{image}(A)$. Thus $\mathrm{Sol}(A, b) \neq \varnothing$.

Finally, for part (v), we note that the existence assertion follows by taking $\Delta = \mathrm{Sol}^*(A, b)$. For uniqueness, let us make a trivial general observation. If $S \subseteq \mathsf{V} \oplus \mathbb{R}$, then

$$\{(\lambda, a) \in \mathsf{V}^* \oplus \mathbb{R} \mid \langle \lambda; v \rangle + ab = 0,\ (v, b) \in S\}$$

is simply the annihilator of S, and is a subspace uniquely prescribed by S. The conclusion here follows by taking

$$S = \{(v, 1) \mid v \in \mathrm{Sol}(A, b)\}. \qquad \blacksquare$$

We note that the subspace Δ is uniquely defined by the set of solutions $\mathrm{Sol}(A, b)$, while this set of solutions does not uniquely define A and b. However, one can recover the important ingredients of A and b from Δ. Let us see how to do this. First note that $\mathrm{Sol}(A, b)$ is determined by $\ker(A)$ (since $\mathrm{Sol}(A, b)$ is an affine space with linear part equal to $\ker(A)$)

and by $b + \operatorname{image}(A) \in \mathsf{V}/\operatorname{image}(A)$ (since vectors $b, b' \in \mathsf{V}$ satisfy $\operatorname{Sol}(A, b) = \operatorname{Sol}(A, b')$ if and only if $b - b' \in \operatorname{image}(A)$). Let

$$\pi_0 \colon \mathsf{V}^* \oplus \mathbb{R} \to (\mathsf{V}^* \oplus \mathbb{R})/(\{0\} \oplus \mathbb{R}) \simeq \mathsf{V}^*.$$

Given a subspace $\Delta \subseteq \mathsf{V}^* \oplus \mathbb{R}$, denote

$$\Delta_1 = \pi_0(\Delta) \subseteq \mathsf{V}^*. \tag{2.10}$$

Also define

$$\Delta_0 = \Delta \cap (\{0\} \oplus \mathbb{R}). \tag{2.11}$$

With this notation, we have the following result.

2.29 Lemma: (Recovering A and b from Δ) *Let V be a finite-dimensional $\mathbb{R}$-vector space, let $\Delta \subseteq \mathsf{V}^* \oplus \mathbb{R}$ be a subspace of positive codimension, and let $A \in \operatorname{End}(\mathsf{V})$ and $b \in \mathsf{V}$ be such that $\Delta = \operatorname{Sol}^*(A, b)$. Then*

(i) $\operatorname{Sol}(A, b) \neq \varnothing$ *if and only if* $\Delta_0 = \{(0, 0)\}$*;*

(ii) $\operatorname{image}(A^*) = \Delta_1$*;*[3]

(iii) $\ker(A^*) = \{\lambda \in \mathsf{V}^* \mid (0, \langle\lambda; b\rangle) \in \Delta_0\}$*.*[4]

Proof: (i) Note that $\Delta_0 = \{(0, 0)\}$ if and only if $(0, 1) \notin \Delta_0$, just because Δ_0 is either a zero- or one-dimensional subspace. Thus this part of the result follows from Lemma 2.28(iv).

(ii) We have

$$\begin{aligned}
\Delta_1 &= \{\lambda \in \mathsf{V}^* \mid (\lambda, a) \in \Delta \text{ for some } a \in \mathbb{R}\} \\
&= \{\lambda \in \mathsf{V}^* \mid (\lambda, a) \in \mathsf{V}^* \oplus \mathbb{R},\ \lambda = A^*(\mu),\ a = \langle\mu; b\rangle \text{ for some } \mu \in \mathsf{V}^*\} \\
&= \operatorname{image}(A^*),
\end{aligned}$$

as claimed.

(iii) We have

$$\begin{aligned}
\ker(A^*) &= \{\lambda \in \mathsf{V}^* \mid A^*(\lambda) = 0\} \\
&= \{\lambda \in \mathsf{V}^* \mid (A^*(\lambda), \langle\lambda; b\rangle) \in \Delta_0\} \\
&= \{\lambda \in \mathsf{V}^* \mid (0, \langle\lambda; b\rangle) \in \Delta_0\},
\end{aligned}$$

as desired. ■

With the above considerations in mind, and as indicated by (2.5), we identity the set of affine functions on a vector bundle $\pi \colon \mathsf{E} \to \mathsf{M}$ with sections of the vector bundle $\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$. We shall notationally distinguish these things, however, by $\operatorname{Aff}^r(\mathsf{E})$ and $\Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}})$ in order to attempt to clarify the ways in which we will think of what is effectively the same thing.

With the preceding as motivation, we make the following definition.

---

[3] Note that specifying $\operatorname{image}(A^*)$ is equivalent to specifying $\ker(A)$.

[4] Note that specifying $\ker(A^*)$ is equivalent to specifying $\operatorname{image}(A)$. Also, the condition here specifies $b + \operatorname{image}(A) \in \mathsf{V}/\operatorname{image}(A)$ by prescribing its annihilator in $(\mathsf{V}/\operatorname{image}(A))^* \simeq (\operatorname{image}(A))^*$.

2.30 Definition: (Defining subbundle, affine subbundle variety) Let $r \in \{\infty, \omega\}$ and let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle.

(i) A ***C^r-defining subbundle*** is a C^r-generalised subbundle Δ of $\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ for which Δ_x has positive codimension in $\mathsf{E}^*_x \oplus \mathbb{R}$ for each $x \in \mathsf{M}$.

(ii) A subset $\mathsf{A} \subseteq \mathsf{E}$ is a ***C^r-affine subbundle variety*** if there exists a C^r-defining subbundle $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ for which

$$\mathsf{A} = \{e \in \mathsf{E} \mid \langle \lambda; e\rangle + a = 0,\ (\lambda, a) \in \Delta_{\pi(e)}\}. \tag{2.12}$$

We shall write $\mathsf{A} = \mathsf{A}(\Delta)$ when we wish to prescribe the defining subbundle giving rise to the affine subbundle variety A.

(iii) The set

$$\mathsf{S}(\mathsf{A}) = \{x \in \mathsf{M} \mid \mathsf{A} \cap \mathsf{E}_x \neq \varnothing\}$$

is the ***base variety*** for A.

For $x \in \mathsf{S}(\mathsf{A})$, we denote $\mathsf{A}_x = \mathsf{A} \cap \mathsf{E}_x$. •

For $x \in \mathsf{S}(\mathsf{A})$, we note that A_x is an affine subspace of E_x. Thus, by Lemma 2.28(v), there is a unique subspace $\Delta_x \subseteq \mathsf{E}^*_x \oplus \mathbb{R}$ such that

$$\mathsf{A}_x = \{e \in \mathsf{E}_x \mid (e, 1) \in \Lambda(\Delta_x)\}. \tag{2.13}$$

This shows that, given an affine subbundle variety A, any defining subbundle Δ for which $\mathsf{A} = \mathsf{A}(\Delta)$ is uniquely determined for $x \in \mathsf{S}(\mathsf{A})$. It is *not* the case, however, that the defining subbundle is uniquely determined at points not in $\mathsf{S}(\mathsf{A})$, of course.

Note that it is often most practical to talk of defining subbundles in the absence of the associated affine subbundle varieties since the former is always a perfectly well-defined subbundle, while the latter may be the empty set.

Similarly to Corollary 2.19 for cogeneralised subbundles, we have the following result.

2.31 Corollary: (Affine subbundle varieties are varieties) *Let $r \in \{\infty, \omega\}$ and let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. If $\mathsf{A} \subseteq \mathsf{E}$ is a C^r-affine subbundle variety, then it is a C^r-variety.*

Proof: Let $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ be a defining subbundle for A. Let $e \in \mathsf{A}$ and let $x = \pi(e)$. By Lemma 2.23 and the proof of Lemma 2.17, there is a neighbourhood $\mathcal{U}$ of x and globally sections $\lambda^1 \oplus f^1, \dots, \lambda^k \oplus f^k \in \Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}})$ that generate $\Delta|\mathcal{U}$. If we define $F^j = (\lambda^j)^{\mathrm{e}} + (f^j)^{\mathrm{h}} \in \mathrm{Aff}^r(\mathsf{E})$, $j \in \{1, \dots, k\}$, we have

$$\mathsf{A} \cap \pi^{-1}(\mathcal{U}) = \bigcap_{j=1}^{k} (F^j)^{-1}(0),$$

showing that A is indeed locally the intersection of the zeros of finitely many C^r-functions. ■

Another consequence of all of this is the following.

2.32 Corollary: (The base variety of an affine subbundle variety is a variety) *Let $r \in \{\infty, \omega\}$ and let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle. If $\mathsf{A} \subseteq \mathsf{E}$ is a C^r-affine subbundle variety, then $\mathsf{S}(\mathsf{A})$ is a C^r-variety if it is nonempty.*

Proof: Let $x \in \mathsf{S}(\mathsf{A})$. As in the proof of Corollary 2.31, let $\mathcal{U}$ be a neighbourhood of x and let $F^1, \dots, F^k \in \mathrm{Aff}^r(\mathsf{E})$ be such that

$$\mathsf{A} \cap \pi^{-1}(\mathcal{U}) = \bigcap_{j=1}^{k} (F^j)^{-1}(0).$$

Write $F^j = (\lambda^j)^{\mathrm{e}} + (f^j)^{\mathrm{h}}$ for $\lambda^1, \dots, \lambda^k \in \Gamma^r(\mathsf{E}^*)$ and $f^1, \dots, f^k \in \mathrm{C}^r(\mathsf{M})$. The conditions for $y \in \mathcal{U}$ to be in $\mathsf{S}(\mathsf{A})$ are expressed by the conditions that there exists $v \in \mathsf{E}_y$ for which

$$\langle \lambda^j(y); v \rangle + f^j(y) = 0, \qquad j \in \{1, \dots, k\}.$$

We can assume that E is trivialised over $\mathcal{U}$ via local sections $\xi_1, \dots, \xi_m$. We denote $\lambda^j_l = \langle \lambda^j; \xi_l \rangle$, $j \in \{1, \dots, k\}$, $l \in \{1, \dots, m\}$. We write

$$v = \sum_{l=1}^{m} v^l \xi_l(y).$$

The defining conditions for $y \in \mathcal{U}$ to be in $\mathsf{S}(\mathsf{A})$ are then that there exists $v^1, \dots, v^m \in \mathbb{R}$ satisfying

$$\sum_{l=1}^{m} \lambda^j_l(y) v^l + f^j(y) = 0, \qquad j \in \{1, \dots, k\}. \tag{2.14}$$

If we define matrices

$$\boldsymbol{A}_0(y) = \begin{bmatrix} \lambda^1_1(y) & \cdots & \lambda^1_m(y) \\ \vdots & \ddots & \vdots \\ \lambda^k_1(y) & \cdots & \lambda^k_m(y) \end{bmatrix}, \quad \boldsymbol{A}_1(y) = \begin{bmatrix} \lambda^1_1(y) & \cdots & \lambda^1_m(y) & f^1(y) \\ \vdots & \ddots & \vdots & \vdots \\ \lambda^k_m(y) & \cdots & \lambda^k_m(y) & f^k(y) \end{bmatrix},$$

the condition for the existence of $v^1, \dots, v^s \in \mathbb{R}$ satisfying (2.14) is equivalent to the ranks of $\boldsymbol{A}_0(y)$ and $\boldsymbol{A}_1(y)$ being equal. For $m \in \mathbb{Z}_{\geq 0}$, let

$$\mathsf{S}_0(m) = \{y \in \mathcal{U} \mid \ \mathrm{rank}(\boldsymbol{A}_0(y)) = m\}, \quad \mathsf{S}_1(m) = \{y \in \mathcal{U} \mid \ \mathrm{rank}(\boldsymbol{A}_1(y)) = m\}.$$

These two subsets are submanifolds, as the following general lemma proves.

1 Lemma: *Let $r \in \{\infty, \omega\}$, let M be C^r-manifold, let $m, n \in \mathbb{Z}_{>0}$, let $k \in \mathbb{Z}_{\geq 0}$ satisfy $k \leq \max\{m, n\}$, and let $\boldsymbol{A} \colon \mathsf{M} \to \mathbb{R}^{m \times n}$ be a C^r-function. Then*

$$\{x \in \mathsf{M} \mid \ \mathrm{rank}(\boldsymbol{A}(x)) = k\}$$

is a C^r-variety whenever it is nonempty.

Proof: Let

$$\mathsf{S}(\boldsymbol{A}, \geq k) = \{x \in \mathsf{M} \mid \ \boldsymbol{A}(x) \geq k\},$$

and note that $\mathsf{S}(\boldsymbol{A}, \geq k)$ is an open subset of M, Indeed, since the condition for membership in $\mathsf{S}(\boldsymbol{A}, \geq k)$ is that there be a $k \times k$ subdeterminant of $\boldsymbol{A}$ which is nonzero, continuity of determinants gives the desired openness. Now, if $x \in \mathsf{S}(\boldsymbol{A}, \geq k)$, then we can permute rows and columns of $\boldsymbol{A}(y)$ to arrive at a matrix $\boldsymbol{B}(y)$ of the form

$$\boldsymbol{B}(y) = \begin{bmatrix} \boldsymbol{B}_{11}(y) & \boldsymbol{B}_{12}(y) \\ \boldsymbol{B}_{21}(y) & \boldsymbol{B}_{22}(y) \end{bmatrix},$$

where $\boldsymbol{B}_{11} \in \mathbb{R}^{k \times k}$ satisfies $\operatorname{rank}(\boldsymbol{B}_{11}(x)) = k$. Since $\boldsymbol{B}(y)$ is obtained from $\boldsymbol{A}(y)$ by mere swapping of rows and columns, we have $\operatorname{rank}(\boldsymbol{B}(y)) = \operatorname{rank}(\boldsymbol{A}(y))$ for all $y \in \mathsf{M}$. By continuity, there is a neighbourhood $\mathcal{U}$ of x such that $\operatorname{rank}(\boldsymbol{B}_{11}(y)) = k$ for $y \in \mathcal{U}$. Now consider the matrix

$$\boldsymbol{P}(y) = \begin{bmatrix} \boldsymbol{I}_k & -\boldsymbol{B}_{11}^{-1}(y)\boldsymbol{B}_{12}(y) \\ \boldsymbol{0} & \boldsymbol{I}_{m-k} \end{bmatrix},$$

which is invertible and is a C^r-function of $y \in \mathcal{U}$. We directly compute

$$\boldsymbol{B}(y)\boldsymbol{P}(y) = \begin{bmatrix} \boldsymbol{B}_{11}(y) & \boldsymbol{0} \\ \boldsymbol{B}_{21}(y) & \boldsymbol{B}_{22}(y) - \boldsymbol{B}_{21}(y)\boldsymbol{B}_{11}^{-1}(y)\boldsymbol{B}_{12}(y) \end{bmatrix}.$$

Thus

$$\begin{aligned} \{y \in \mathcal{U} \mid \ \operatorname{rank}(\boldsymbol{A}(y)) = k\} &= \{y \in \mathcal{U} \mid \ \operatorname{rank}(\boldsymbol{B}(y)) = k\} \\ &= \{y \in \mathcal{U} \mid \ \boldsymbol{B}_{22}(y) - \boldsymbol{B}_{21}(y)\boldsymbol{B}_{11}^{-1}(y)\boldsymbol{B}_{12}(y) = \boldsymbol{0}\}. \end{aligned}$$

This last set is the intersection of the zeros of finitely many C^r-functions (namely the functions defined by the entries of the matrix that is required to be zero), and this gives the result. ▼

Now we note that

$$\mathsf{S}(\mathsf{A}) \cap \mathcal{U} = \bigcup_{m \in \mathbb{Z}_{\geq 0}} (\mathsf{S}_0(m) \cap \mathsf{S}_1(m)). \tag{2.15}$$

Motivated by this, we have the following lemma which is well-known, but for which we give a proof since typically proofs of these simple facts are embedded in a more complicated setting.

2 Lemma: *Finite intersections and unions of C^r-varieties are C^r-varieties.*

Proof: It suffices to consider the intersection and union of two C^r-varieties. Let $\mathsf{S}, \mathsf{T} \subseteq \mathsf{M}$ be C^r-varieties. Let $x \in \mathsf{S} \cap \mathsf{T}$ and let $\mathcal{U}$ be a neighbourhood of x such that

$$\mathsf{S} \cap \mathcal{U} = \bigcap_{i=1}^{k} (f^i)^{-1}(0), \quad \mathsf{T} \cap \mathcal{U} = \bigcap_{j=1}^{l} (g^j)^{-1}(0),$$

for $f^1, \dots, f^k, g^1, \dots, g^l \in \mathrm{C}^r(\mathcal{U})$. Then

$$(\mathsf{S} \cap \mathsf{T}) \cap \mathcal{U} = \left(\bigcap_{i=1}^{k} (f^i)^{-1}(0) \right) \cap \left(\bigcap_{j=1}^{l} (g^j)^{-1}(0) \right),$$

showing that $\mathsf{S}\cap\mathsf{T}$ is a C^r-variety. Next let $x\in\mathsf{S}\cup\mathsf{T}$ and let $\mathcal{U}$ be a neighbourhood of x such that

$$\mathsf{S}\cap\mathcal{U}=\bigcap_{i=1}^{k}(f^i)^{-1}(0),\quad \mathsf{T}\cap\mathcal{U}=\bigcap_{j=1}^{l}(g^j)^{-1}(0),$$

for $f^1,\dots,f^k,g^1,\dots,g^l\in\mathrm{C}^r(\mathcal{U})$. We claim that

$$(\mathsf{S}\cup\mathsf{T})\cap\mathcal{U}=\bigcap_{\substack{i\in\{1,\dots,k\},\\ j\in\{1,\dots,l\}}}(f^ig^j)^{-1}(0).$$

Indeed, if $y\in(\mathsf{S}\cup\mathsf{T})\cap\mathcal{U}$, then $f^i(y)=0$, $i\in\{1,\dots,l\}$, or $g^j(y)=0$, $j\in\{1,\dots,k\}$. Thus

$$y\in\bigcap_{\substack{i\in\{1,\dots,k\},\\ j\in\{1,\dots,l\}}}(f^ig^j)^{-1}(0).$$

Conversely, suppose that $y\in\mathcal{U}-(\mathsf{S}\cup\mathsf{T})$. Then there exists $i\in\{1,\dots,k\}$ and $j\in\{1,\dots,l\}$ such that $f^i(y)\neq 0$ and $g^j(y)\neq 0$. Thus $f^i(y)g^j(y)\neq 0$ and so

$$y\notin\bigcap_{\substack{i\in\{1,\dots,k\},\\ j\in\{1,\dots,l\}}}(f^ig^j)^{-1}(0).$$

which gives the result. ▼

The corollary follows immediately from (2.15) and the lemma. ■

The picture one should have in one's mind concerning an affine subbundle variety $\mathsf{A}\subseteq\mathsf{E}$ is encoded in the following diagram:

$$\begin{tikzcd}
\mathsf{A} \arrow[r, "\text{affine on fibres}"] \arrow[d, "\pi|\mathsf{A}"'] & \mathsf{E} \arrow[d, "\pi"] \\
\mathsf{S}(\mathsf{A}) \arrow[r] & \mathsf{M}
\end{tikzcd}$$

The column on the left is comprised of C^r-varieties while the column on the right is comprised of C^r-manifolds. Thus one should think of A as being a "singular affine bundle" over the "singular manifold" $\mathsf{S}(\mathsf{A})$.

Let us introduce some terminology that we will find useful.

2.33 Definition: (Total, partial, null defining subbundle) Let $r\in\{\infty,\omega\}$ and let $\pi\colon\mathsf{E}\to\mathsf{M}$ be a C^r-vector bundle. A C^r-defining subbundle Δ is

(i) ***total*** if $\mathsf{S}(\mathsf{A}(\Delta))=\mathsf{M}$,

(ii) ***partial*** if $\mathsf{S}(\mathsf{A}(\Delta))\neq\varnothing$ and $\mathsf{S}(\mathsf{A})\subset\mathsf{M}$, and

(iii) ***null*** if $\mathsf{S}(\mathsf{A}(\Delta))=\varnothing$. •

Let us further punctuate the issues raised in Remark 2.27 regarding the rôle played by cogeneralised affine subbundles.

2.34 Remark: (On cogeneralised affine subbundles II) In the case that a C^r-defining subbundle Δ is total, one might be inclined to say that the corresponding affine subbundle variety $\mathsf{A}(\Delta)$ should be the same thing as a C^r-cogeneralised affine subbundle. This is not true, since it is generally not the case that one can find a C^r-section ξ_0 of $\mathsf{A}(\Delta)$. Indeed, it may not even be the case that one can find a continuous section of $\mathsf{A}(\Delta)$. Such matters are discussed by Fefferman and Kollár [2013]. All that one can say with any generality is that, if the fibres of $\mathsf{A}(\Delta)$ have locally constant rank, then it is indeed a C^r-cogeneralised subbundle, in fact a C^r-subbundle. While true, this fact turns a blind eye to the interesting lack of correspondence between total defining subbundles and cogeneralised affine subbundles. •

2.11. Distributions on Riemannian manifolds. As essential rôle in our main results in Section 7 is played by certain constructions involving the interaction of distributions and Riemannian metrics. Our presentation here is derived in part from developments of Lewis [1998].

Let $r \in \{\infty, \omega\}$, let M be a C^r-manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle. We shall sometimes call D a ***distribution*** on M. We shall need some particular constructions concerning the interaction of distributions and Riemannian geometry. Thus we additionally introduce a C^r-Riemannian metric $\mathbb{G}$ and denote by $\mathsf{D}^\perp$ the $\mathbb{G}$-orthogonal complement to D. We denote by

$$P_{\mathsf{D}}, P_{\mathsf{D}^\perp} \colon \mathsf{TM} \to \mathsf{TM}$$

the $\mathbb{G}$-orthogonal projections onto D and $\mathsf{D}^\perp$, respectively. Let us define a few objects that can be built from this data.

2.35 Definition: (Constrained connection, second fundamental form, Frobenius curvature, geodesic curvature) Let $r \in \{\infty, \omega\}$, let $(\mathsf{M}, \mathbb{G})$ be a C^r-Riemannian manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle.

(i) The ***constrained connection*** for D is the affine connection $\overset{\mathsf{D}}{\nabla}$ on M defined by

$$\overset{\mathsf{D}}{\nabla}_X Y = \overset{\mathbb{G}}{\nabla}_X Y + (\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y), \qquad X, Y \in \Gamma^r(\mathsf{TM}).$$

(ii) The ***second fundamental form*** for D is the tensor field

$$S_{\mathsf{D}} \in \Gamma^r(\mathsf{D}^\perp \otimes \mathsf{T}^*\mathsf{M} \otimes \mathsf{D}^*)$$

defined by $S_{\mathsf{D}}(X, Y) = -(\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y)$ for $X \in \Gamma^r(\mathsf{TM})$ and $Y \in \Gamma^r(\mathsf{D})$.

(iii) The ***Frobenius curvature*** for D is the tensor field

$$F_{\mathsf{D}} \in \Gamma^r(\mathsf{D}^\perp \otimes \textstyle\bigwedge^2(\mathsf{D}^*))$$

defined by $F_{\mathsf{D}}(X, Y) = S_{\mathsf{D}}(X, Y) - S_{\mathsf{D}}(Y, X)$ for $X, Y \in \Gamma^r(\mathsf{D})$.

(iv) The ***geodesic curvature*** for D is the tensor field

$$G_{\mathsf{D}} \in \Gamma^r(\mathsf{D}^\perp \otimes \mathrm{S}^2(\mathsf{D}^*))$$

defined by $G_{\mathsf{D}}(X, Y) = S_{\mathsf{D}}(X, Y) + S_{\mathsf{D}}(Y, X)$ for $X, Y \in \Gamma^r(\mathsf{D})$. •

These definitions contain some implicit assertions that must be proved. We prove these, along with a few other facts, in the following lemma. In the statement of the lemma, we make use of the operation

$$\langle X : Y\rangle = \overset{\mathbb{G}}{\nabla}_X Y + \overset{\mathbb{G}}{\nabla}_Y X,$$

which is called the ***symmetric product*** by Lewis [1998].

2.36 Lemma: (Constructions for distributions on Riemannian manifolds) *Let $r \in \{\infty, \omega\}$, let $(\mathsf{M}, \mathbb{G})$ be a C^r-Riemannian manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle. Then the following statements hold:*

(i) $S_{\mathsf{D}}(X, Y) = P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y)$ *for* $X \in \Gamma^r(\mathsf{TM})$ *and* $Y \in \Gamma^r(\mathsf{D})$*;*

(ii) $\overset{\mathsf{D}}{\nabla}_X Y = P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_X Y)$ *for* $X \in \Gamma^r(\mathsf{TM})$ *and* $Y \in \Gamma^r(\mathsf{D})$*;*

(iii) $P_{\mathsf{D}}^\perp((\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}}^\perp)(Y)) = P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y) = 0$ *for* $X \in \Gamma^r(\mathsf{TM})$ *and* $Y \in \Gamma^r(\mathsf{D}^\perp)$*;*

(iv) $F_{\mathsf{D}}(X, Y) = P_{\mathsf{D}^\perp}([X, Y])$ *for* $X, Y \in \Gamma^r(\mathsf{D})$*;*

(v) $G_{\mathsf{D}}(X, Y) = P_{\mathsf{D}^\perp}(\langle X : Y\rangle)$ *for* $X, Y \in \Gamma^r(\mathsf{D})$*;*

(vi) $S_{\mathsf{D}} = \frac{1}{2}(G_{\mathsf{D}} + F_{\mathsf{D}})$*.*

Proof: (i) First, if $Y \in \Gamma^r(\mathsf{D})$, then

$$\begin{aligned} & P_{\mathsf{D}^\perp}(Y) = 0, \\ \Longrightarrow \quad & (\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y) + P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y) = 0, \\ \Longrightarrow \quad & P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y) + P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y) = 0, \quad P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y) = 0, \end{aligned}$$

since

$$P_{\mathsf{D}^\perp} \circ P_{\mathsf{D}^\perp} = P_{\mathsf{D}^\perp}, \quad P_{\mathsf{D}} \circ P_{\mathsf{D}^\perp} = 0.$$

Thus $(\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y) \in \Gamma^r(\mathsf{D}^\perp)$ and, consequently,

$$P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y) = -(\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y),$$

as claimed.

(ii) Using the computations from the first part of the proof,

$$\overset{\mathsf{D}}{\nabla}_X Y = \overset{\mathbb{G}}{\nabla}_X Y + (\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y) = \overset{\mathbb{G}}{\nabla}_X Y - P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y) = P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_X Y),$$

for $X \in \Gamma^r(\mathsf{TM})$ and $Y \in \Gamma^r(\mathsf{D})$.

(iii) For $Y \in \Gamma^r(\mathsf{D}^\perp)$ and $X \in \Gamma^r(\mathsf{TM})$ we have

$$\begin{aligned} & P_{\mathsf{D}^\perp}(Y) = Y \\ \Longrightarrow\ & (\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y) + P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y) = \overset{\mathbb{G}}{\nabla}_X Y \\ \Longrightarrow\ & P_{\mathsf{D}^\perp}((\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y)) + P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y) = P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y), \end{aligned}$$

and the conclusion follows from this.

(iv) For $X, Y \in \Gamma^r(\mathsf{D})$ we have

$$F_{\mathsf{D}}(X,Y) = S_{\mathsf{D}}(X,Y) - S_{\mathsf{D}}(Y,X) = -(\overset{\mathbb{G}}{\nabla}_X P_{\mathsf{D}^\perp})(Y) + (\overset{\mathbb{G}}{\nabla}_Y P_{\mathsf{D}^\perp})(X)$$
$$= P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_X Y - \overset{\mathbb{G}}{\nabla}_Y X) = P_{\mathsf{D}}([X,Y]),$$

using part (i) and the fact that $\overset{\mathbb{G}}{\nabla}$ is torsion-free.

Part (v) follows similarly to part (iv) and part (vi) is immediate from the definitions. ■

The statement of part (iii) is a bit of an outlier since it essentially involves evaluating S_{D} on an argument taking values in $\mathsf{D}^\perp$. However, S_{D} is defined only for arguments taking values in D. Thus the statement is "improper," in some sense. However, we shall use this conclusion in the proof of Proposition 6.7, so we state it here.

For $Y \in \Gamma^r(\mathsf{D})$, denote by $S_{\mathsf{D}}(Y) \in \Gamma^r(\mathsf{D}^\perp \otimes \mathsf{T}^*\mathsf{M})$ the tensor field defined by $S_{\mathsf{D}}(Y)(X) = S_{\mathsf{D}}(X,Y)$, $X \in \Gamma^r(\mathsf{TM})$. In a similar manner, for $Y \in \Gamma^r(\mathsf{D})$, we denote by $F_{\mathsf{D}}(Y), G_{\mathsf{D}}(Y) \in \Gamma^r(\mathsf{D}^\perp \otimes \mathsf{D}^*)$ the tensor fields defined by

$$F_{\mathsf{D}}(Y)(X) = F_{\mathsf{D}}(X,Y), \quad G_{\mathsf{D}}(Y)(X) = G_{\mathsf{D}}(X,Y),$$

respectively, for $X \in \Gamma^r(\mathsf{D})$. The $\mathbb{G}$-transposes of $S_{\mathsf{D}}(Y)$, $F_{\mathsf{D}}(Y)$, and $G_{\mathsf{D}}(Y)$ we denote by $S^*_{\mathsf{D}}(Y)$, $F^*_{\mathsf{D}}(Y)$, and $G^*_{\mathsf{D}}(Y)$ so that

$$\mathbb{G}(S^*_{\mathsf{D}}(Y)(\alpha), X) = \mathbb{G}(\alpha, S_{\mathsf{D}}(X,Y)), \qquad X, Y \in \Gamma^r(\mathsf{TM}),\ \alpha \in \Gamma^r(\mathsf{D}^\perp),$$

and

$$\mathbb{G}(F^*_{\mathsf{D}}(Y)(\alpha), X) = \mathbb{G}(\alpha, F_{\mathsf{D}}(X,Y)), \quad \mathbb{G}(G^*_{\mathsf{D}}(Y)(\alpha), X) = \mathbb{G}(\alpha, G_{\mathsf{D}}(X,Y))$$

for $X \in \Gamma^r(\mathsf{D})$, $\alpha \in \Gamma^r(\mathsf{D}^\perp)$.

We will also be interested in representations of these tensors with the orders of the arguments flipped. To this end, for $\alpha \in \Gamma^r(\mathsf{D}^\perp)$, define $F^\star_{\mathsf{D}}(\alpha), G^\star_{\mathsf{D}}(\alpha) \in \Gamma^r(\mathsf{D} \otimes \mathsf{D}^*)$ by

$$\mathbb{G}(X, F^\star_{\mathsf{D}}(\alpha)(Y)) = \mathbb{G}(\alpha, F_{\mathsf{D}}(X,Y)), \quad \mathbb{G}(X, G^\star_{\mathsf{D}}(\alpha)(Y)) = \mathbb{G}(\alpha, G_{\mathsf{D}}(X,Y)),$$

for $X, Y \in \Gamma^r(\mathsf{D})$ and $\alpha \in \Gamma^r(\mathsf{D}^\perp)$. Thus we have

$$F^\star_{\mathsf{D}}(\alpha)(Y) = F^*_{\mathsf{D}}(Y)(\alpha), \quad G^\star_{\mathsf{D}}(\alpha)(Y) = G^*_{\mathsf{D}}(Y)(\alpha), \quad Y \in \Gamma^r(\mathsf{D}),\ \alpha \in \Gamma^r(\mathsf{D}^\perp),$$

Let us make some observations about the preceding constructions.

2.37 Remarks: (Constructions for distributions on Riemannian manifolds)

1. The constrained connection $\overset{\mathsf{D}}{\nabla}$ is a C^r-affine connection on M that restricts to a C^r-linear connection in D. Of course, there are many affine connections on M that agree with $\overset{\mathsf{D}}{\nabla}$ when restricted to D, though the one we give is arguably the most natural as it arises merely from the $\mathbb{G}$-orthogonal decomposition of $\overset{\mathbb{G}}{\nabla}$. The constrained connection seems to have originated in the work of Synge [1928], but the development we give is that of Lewis [1998].
2. The definition of the second fundamental form is a natural adaptation of the theory of Riemannian geometry for submanifolds [e.g., Lee 2018, Chapter 8].

3. It is clear that the Frobenius curvature of D vanishes if and only if D is integrable (noting that D is integrable if and only if it is involutive as it is a subbundle of TM [Abraham, Marsden, and Ratiu 1988, Theorem 4.3.3]).
4. An important difference between the theory for distributions as we present here and the theory for submanifolds concerns geodesic invariance. For submanifolds, one has the equivalence of the conditions (a) geodesic invariance of a submanifold (i.e., geodesics with initial conditions tangent to the submanifold remain in the submanifold), (b) the connection restricts to the submanifold, and (c) vanishing of the second fundamental form. For distributions, the second and third of these conditions are equivalent (as is clear), but they do not imply geodesic invariance (i.e., geodesics with initial conditions in the distribution have subsequence tangent vectors also in the distribution). In fact, Lewis [1998] shows that a distribution D is geodesically invariant for an affine connection ∇ if and only if
$$\langle X : Y\rangle \in \Gamma^r(\mathsf{D}), \qquad X, Y \in \Gamma^r(\mathsf{D}).$$[5]
From this we see that a distribution D is geodesically invariant if and only if its geodesic curvature vanishes. •

2.12. Characteristic subbundle of a distribution. In our study of the equivalence of the two equations for constrained motion, we shall encounter a geometric condition whose meaning will be helpful to understand. A basic concept in this understanding is the following.

2.38 Definition: (Vector fields and flows leaving distributions invariant) Let $r \in \{\infty, \omega\}$, let M be a C^r-manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle. Let $X \in \Gamma^r(\mathsf{TM})$.

(i) The distribution D is ***invariant*** under X if $[X, Y] \in \Gamma^r(\mathsf{D})$ for $Y \in \Gamma^r(\mathsf{D})$.

(ii) The distribution D is ***flow-invariant*** under X if, for each $x \in \mathsf{M}$, there exist a neighbourhood $\mathcal{U} \subseteq \mathsf{M}$ of x and $T \in \mathbb{R}_{>0}$ such that, for each $t \in [-T, T]$, $\Phi^X_t|\mathcal{U}$ is defined and
$$(\Phi^X_t)^*(Y|\Phi^X_t(\mathcal{U})) \in \Gamma^r(\mathsf{D}|\mathcal{U}), \qquad Y \in \Gamma^r(\mathsf{D}).$$ •

For a distribution whose rank is not locally constant—sometimes called a "generalised distribution" to distinguish from the nice locally constant rank case—the relationship between infinitesimal invariance and invariance is that they agree when $r = \omega$ and they agree when $r = \infty$ under a finite generation hypothesis. Let us give here a full proof of the correspondence between these notions in the locally constant rank case.

2.39 Proposition: (Invariance and infinitesimal invariance of distributions under vector fields) *Let $r \in \{\infty, \omega\}$, let M be a C^r-manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle. Then, for $X \in \Gamma^r(\mathsf{TM})$, D is invariant under X if and only if it is flow-invariant under X.*

Proof: Suppose that D is invariant under X. Let $x \in \mathsf{M}$ and let $\mathcal{V}$ be a neighbourhood of x with the following properties:

1. there exists $T \in \mathbb{R}_{>0}$ such that $\Phi^X_t(y)$ is defined for $t \in [-T, T]$ and $y \in \mathcal{V}$;
2. there exists $Y_1, \dots, Y_m \in \Gamma^r(\mathsf{D})$ such that $(Y_1(y), \dots, Y_m(y))$ is a basis for D_y for $y \in \mathcal{V}$.

[5] The proof by Lewis is done in the smooth case, but the proof works also in the real analytic case.

The existence of a neighbourhood $\mathcal{V}$ having the first property follows from the semicontinuity properties of the maximal interval of existence for integral curves of a vector field [Abraham, Marsden, and Ratiu 1988, Proposition 4.1.24]. The existence of a neighbourhood $\mathcal{V}$ having the second property follows in the smooth case since D is a subbundle, and using cutoff functions. In the real analytic case, the existence of such a neighbourhood $\mathcal{V}$ can be inferred by Cartan's Theorem A [Cartan 1957, Proposition 6], cf. Corollary 2.18. In any case, if $Y \in \Gamma^r(\mathsf{D})$, then

$$Y|\mathcal{V} = \eta^1(Y_1|\mathcal{V}) + \cdots + \eta^m(Y_m|\mathcal{V})$$

for some $\eta^1, \dots, \eta^m \in \mathrm{C}^r(\mathcal{V})$. Now let $\mathcal{U}$ be a neighbourhood of x such that $\Phi_t^X(y) \in \mathcal{V}$ for every $t \in [-T, T]$ and $y \in \mathcal{U}$. The existence of such a neighbourhood $\mathcal{U}$ follows by continuity of the flow. Thus we can write

$$[X, Y_j]|\mathcal{V} = \sum_{k=1}^m f_j^k(Y_j|\mathcal{V}), \qquad j \in \{1, \dots, m\},$$

for $f_j^k \in \mathrm{C}^r(\mathcal{V})$.

Define

$$v_j(t, y) = (\Phi_t^X)^* Y_j(y), \qquad j \in \{1, \dots, m\},\ t \in [-T, T],\ y \in \mathcal{U}.$$

Note that $t \mapsto v_j(t, y)$ is a curve in $\mathsf{T}_y\mathsf{M}$. By [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.19] we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} v_j(t, y) &= \frac{\mathrm{d}}{\mathrm{d}t}(\Phi_t^X)^* Y_j(y) = (\Phi_t^X)^*[X, Y_j](y) = (\Phi_t^X)^* \left(\sum_{k=1}^m f_j^k Y_k\right)(y) \\ &= \sum_{k=1}^m (\Phi_t^X)^* f_j^k(y)(\Phi_t^X)^* Y_k(y) = \sum_{k=1}^m (\Phi_t^X)^* f_j^k(y) v_k(t, y).\end{aligned}$$

Define $\boldsymbol{A}_y(t) \in \mathbb{R}^{m \times m}$ by

$$A_{y,j}^k(t) = (\Phi_t^X)^* f_j^k(y), \qquad j, k \in \{1, \dots, m\},$$

and let $\boldsymbol{\Psi}_y \colon \mathbb{R} \to \mathbb{R}^{m \times m}$ be the solution to the matrix initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Psi}_y(t) = \boldsymbol{A}_y(t)\boldsymbol{\Psi}_y(t), \qquad \boldsymbol{\Psi}_y(0) = \boldsymbol{I}_m.$$

We claim that

$$v_j(t, y) = \sum_{j=1}^m \Psi_{y,j}^k(t) Y_k(y).$$

Indeed,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{k=1}^m \Psi_{y,j}^k(t) Y_k(y)\right) &= \sum_{k=1}^m \frac{\mathrm{d}}{\mathrm{d}t}\Psi_{y,j}^k(t) Y_k(y) = \sum_{k,l=1}^m A_{y,j}^l(t)\Psi_{y,l}^k(t) Y_k(y) \\ &= \sum_{l=1}^m (\Phi_t^X)^* f_j^l(y)\left(\sum_{k=1}^m \Psi_{y,l}^k(t) Y_k(y)\right).\end{aligned}$$

Moreover,

$$\sum_{k=1}^{m} \Psi^k_{y,j}(0) Y_k(y) = Y_j(y), \qquad v_j(0,y') = Y_j(y').$$

Thus

$$t \mapsto v_j(t,y) \text{ and } t \mapsto \sum_{k=1}^{m} \Psi^k_{y,j}(t) Y_k(y)$$

satisfy the same differential equation with the same initial condition. Thus they are equal. This gives

$$(\Phi^X_t)^* Y_j(y) = \sum_{k=1}^{m} \Psi^k_{y,j}(t) Y_k(y)$$

for every $t \in [-T,T]$ and $y \in \mathcal{U}$.

Now let $Y \in \Gamma^r(\mathsf{D})$ and write

$$Y|\mathcal{V} = \eta^1(Y_1|\mathcal{V}) + \cdots + \eta^m(Y_m|\mathcal{V}).$$

for $\eta^j \in \mathrm{C}^r(\mathcal{V})$, $j \in \{1,\dots,m\}$. Therefore, for $y \in \mathcal{U}$ and $t \in [-T,T]$,

$$(\Phi^X_t)^* Y(y) = \sum_{j=1}^{k} \eta^j(y) (\Phi^X_t)^* Y_j(y) = \sum_{j=1}^{k} \eta^j(y) \sum_{i=1}^{k} \Psi^i_{y,j}(t) Y_i(y),$$

and so $(\Phi^X_t)^*(Y|\Phi^X_t(\mathcal{U})) \in \Gamma^r(\mathsf{D}|\mathcal{U})$, giving this part of the proposition.

Now suppose that D is flow-invariant under X. Let $x \in \mathsf{M}$ and let $T \in \mathbb{R}_{>0}$ be such that $\Phi^X_t(x)$ is defined for $T \in [-T,T]$. Let $Y \in \Gamma^r(\mathsf{D})$. By hypothesis we have

$$(\Phi^X_t)^* Y(x) \in \mathsf{D}_x, \qquad t \in [-T,T].$$

Therefore, by [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.19], we have

$$[X,Y](x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\Phi^X_t)^* Y(x) \in \mathsf{D}_x$$

since D_x is closed in $\mathsf{T}_x\mathsf{M}$. ■

With these notions of invariance, we can make the following definitions.

2.40 Definition: (Characteristic vector field, characteristic distribution) Let $r \in \{\infty, \omega\}$, let M be a C^r-manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle.

(i) A vector field $X \in \Gamma^r(\mathsf{TM})$ is a ***characteristic vector field*** for D if $X \in \Gamma^r(\mathsf{D})$ and if D is invariant under X.

(ii) The ***characteristic distribution*** of D is the C^r-generalised subbundle $C(\mathsf{D}) \subseteq \mathsf{TM}$ defined by

$$C(\mathsf{D})_x = \{X(x) \mid X \text{ a characteristic vector field}\}. \qquad \bullet$$

The following characterisation of characteristic vector fields and of the characteristic distribution will be useful for us.

2.41 Lemma: (Characterisation of characteristic vector fields and the characteristic distribution) *Let $r \in \{\infty, \omega\}$, let $(\mathsf{M}, \mathbb{G})$ be a C^r-Riemannian manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle. Then the following statements hold:*

(i) $X \in \Gamma^r(\mathsf{D})$ is a characteristic vector field for D if and only if $\ker(F^*_{\mathsf{D}}(X(x))) = \mathsf{D}^{\perp}_x$ for every $x \in \mathsf{M}$;

(ii) $C(\mathsf{D})_x = \{v_x \in \mathsf{D}_x \mid \ \ker(F^*_{\mathsf{D}}(v_x)) = \mathsf{D}^{\perp}_x\}$.

Proof: Since the second assertion follows immediately from the first, we just prove the first.

We have that $X \in \Gamma^r(\mathsf{D})$ is a characteristic vector field if and only if

$$\begin{aligned}
& [X,Y](x) \in \mathsf{D}_x, && Y \in \Gamma^r(\mathsf{D}),\ x \in \mathsf{M},\\
\iff & F_{\mathsf{D}}(X(x), Y(x)) = 0, && Y \in \Gamma^r(\mathsf{D}),\ x \in \mathsf{M},\\
\iff & \mathbb{G}(\alpha(x), F_{\mathsf{D}}(X(x), Y(x))) = 0, && \alpha \in \Gamma^r(\mathsf{D}^{\perp}),\ Y \in \Gamma^r(\mathsf{D}),\ x \in \mathsf{M},\\
\iff & \mathbb{G}(Y(x), F^*_{\mathsf{D}}(X(x))(\alpha(x))) = 0, && \alpha \in \Gamma^r(\mathsf{D}^{\perp}),\ Y \in \Gamma^r(\mathsf{D}),\ x \in \mathsf{M},\\
\iff & F^*_{\mathsf{D}}(X(x))(\alpha(x)) = 0, && \alpha \in \Gamma^r(\mathsf{D}^{\perp}),\ x \in \mathsf{M},\\
\iff & \ker(F^*_{\mathsf{D}}(X(x))) = \mathsf{D}^{\perp}_x, && x \in \mathsf{M},
\end{aligned}$$

as claimed. ■

We shall turn the preceding constructions on their head a little, since it is this altered form in which we shall be interested. We are interested in an understanding of $\ker(F^*_{\mathsf{D}}(u))$ for $u \in \mathsf{D}_x$.

2.42 Definition: (Characteristic, cocharacteristic and non-cocharacteristic vectors) Let $r \in \{\infty, \omega\}$, let $(\mathsf{M}, \mathbb{G})$ be a C^r-Riemannian manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be C^r-subbundle.

(i) A vector $u \in \mathsf{D}_x$ is a ***characteristic vector*** if $F^*_{\mathsf{D}}(u) = 0$.

(ii) For $u \in \mathsf{D}_x$, a vector $\alpha \in \mathsf{D}^{\perp}_x$ is ***cocharacteristic*** for u if $F^*_{\mathsf{D}}(u)(\alpha) = 0$.

(iii) For $u \in \mathsf{D}_x$, a vector $\alpha \in \mathsf{D}^{\perp}_x$ is ***non-cocharacteristic*** for u if $F^*_{\mathsf{D}}(u)(\alpha) \neq 0$. •

The point is that, if $u \in \mathsf{D}_x$ is a characteristic vector, then every vector in $\mathsf{D}^{\perp}_x$ is cocharacteristic for u, whereas, if u is not a characteristic vector, then there are non-cocharacteristic vectors for u.

3. Sobolev spaces of curves on a manifold

In this section we develop a framework for performing geometric analysis with the space of curves on a Riemannian manifold. This is typically carried out by exploiting the structure of an infinite-dimensional Hilbert manifold possessed by the space of such curves [e.g., Klingenberg 1995, §2.3]. Rather than working with infinite-dimensional nonlinear geometry, we reduce the problem to infinite-dimensional linear analysis by working with function evaluations. In this section we also put our framework to use to describe some special classes of curves that will be useful for our analysis in Section 5 for deriving the governing equations for nonholonomic mechanics and constrained variational mechanics.

Since we have a lot of constructions, definitions, and results in this section, let us provide a roadmap to help the reader understand how the story will unfold.

1. In Section 3.1 we introduce the various classes of curves we use in the paper. The basic player is the space $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$ of absolutely continuous curves on M that are square integrable with square integrable derivative. We characterise curves γ in this space by characterising $f \circ \gamma$ for $f \in \mathrm{C}^\infty(\mathsf{M})$, and it is this idea that characterises our approach, in general.

   We consider, specially, curves in $\mathrm{H}^1([t_0, t_1]; \mathsf{E})$, where E is the total space of a vector bundle $\pi \colon \mathsf{E} \to \mathsf{M}$. Here curves are characterised by composition, not with general smooth functions, but with smooth fibre-affine functions, cf. Definition 2.9. This is how the particular structure of the vector bundle is accounted for in our framework. As part of this development of curves in the total space of a vector bundle, we consider curves that are to be thought of as sections over a curve in $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$. These vector spaces of sections over a curve will be important for us in a multitude of ways.

   Some of the classes of curves considered in this section have mechanical significance, such as curves with fixed endpoints and curves with tangent vectors in a distribution. These will be studied in greater detail in Section 5.1.

2. The topology of $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$ is described in Section 3.2. Our definition of this space as a topological space relies only on the functions $f \circ \gamma$, and so gives a description of the topology that involves only the Hilbert space $\mathrm{H}^1([t_0, t_1]; \mathbb{R})$. We prove, using this description of the topology, that various of the subsets of curves and subsets of sections along curves that we introduce are, in fact, closed subsets. This ensures that their relative topology is comparatively friendly. We postpone to Section 5.1 a discussion of the differentiable structure of some of these spaces of curves.

3. A key ingredient in providing a differentiable structure to spaces of curves is to have at hand a notion of calculus in our framework. In Section 3.3 we develop some calculus in $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$ by considering the calculus of curves in the space of curves $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$. These differentiable curves give us access to subsequent definitions for tangent vectors, etc. Consistent with our approach, we do this by reducing definitions to those involving only $\mathrm{H}^1([t_0, t_1]; \mathbb{R})$, where only standard differential calculus in Banach spaces suffices. We further simplify this approach by reducing questions of differentiability and derivatives of curves to questions involving elementary calculus of $\mathbb{R}$-valued functions of two variables. These simple methods for determining differentiability and derivatives are the basis for the applicability of the methods we introduce.

4. Some tools in the calculus of variations are developed in Section 3.4. Specifically we

define the notions of variation and infinitesimal variation of a curve $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$. We make full use of our simplified calculus developed in Section 3.3. These notions of variation and infinitesimal variation get us started towards defining the notion of a tangent vector in our approach.

5. Indeed, this notion of tangent vector, and the associated notion of tangent spaces, are described in Section 3.5. Using these notions we extend our calculus from curves with values in $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$ to mappings from $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$ to a manifold. Again, our methods enable one to reduce questions of differentiability to differentiability of $\mathbb{R}$-valued functions of two real variables.
6. In Section 3.6 we consider mappings between spaces of curves, and we extend our calculus to such mappings. Once again, in our approach we are able to reduce the questions of differentiability to that of $\mathbb{R}$-valued functions of two variables.
7. In Section 3.8 we define a technical device, the weak covariant derivative for distributional sections. This will come up in the proof of Lemma 1 from the proof of Theorem 5.22.

The preceding outline of what we do in this chapter to develop tools for nonlinear Sobolev-type analysis is a beginning of what ought to be possible. One should be able to develop a more comprehensive set of tools applicable to perform this analysis in far more general settings than we use here, while always reducing the analysis to that of scalar-valued functions. Some tools for working with the higher-order derivatives required for such analysis are presented by Jafarpour and Lewis [2014]. Ideas very much inline with what we describe here are given in the series of papers by Convent and Van Schaftingen [2016a, 2016b, 2019].

The constructions and results in this section do not depend on the regularity of manifolds, metrics, and connections, and for this reason we work in the smooth category in this section.

3.1. Curves and sections along curves. Let M be a smooth manifold. We first consider classes of curves on M. Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$ and denote, for $s \in \mathbb{Z}_{\geq 0}$,

$$\mathrm{H}^s([t_0, t_1]; \mathsf{M}) = \{\gamma\colon [t_0, t_1] \to \mathsf{M} \mid f \circ \gamma \in \mathrm{H}^s([t_0, t_1]; \mathbb{R}),\ f \in \mathrm{C}^\infty(\mathsf{M})\}. \tag{3.1}$$

Another way of introducing these classes of curve is to ask that, for each $t \in [t_0, t_1]$, there exists a chart $(\mathcal{U}, \phi)$ about $\gamma(t)$ such that the components of $\phi \circ \gamma$ are members of the usual Sobolev spaces $\mathrm{H}^s(I_t; \mathbb{R})$ for some interval I_t about t. Let us outline how this definition is equivalent to the one we give.

1. The coordinate definition is independent of coordinate chart because the classical Sobolev spaces are invariant under uniformly smooth changes of coordinate [Adams and Fournier 2003, Theorem 3.41]. The assumption of all derivatives being uniformly continuous will always hold if the domain is compact. This suffices in our case since our domain $[t_0, t_1]$ is compact, and so the image of a curve can be covered with finitely many relatively compact coordinate charts.
2. One may assume that coordinate functions are restrictions of globally defined smooth functions to the chart domain by use of bump functions. Equivalently, one can see this by using the Whitney Embedding Theorem [Whitney 1936]. This latter approach has

the benefit of applying in any regularity category where embeddings in Euclidean space exist, e.g., the real analytic category [Grauert 1958].

3. A consequence of the preceding is that our definition of $\mathrm{H}^s([t_0,t_1];\mathsf{M})$ agrees with the standard one in the case of $\mathsf{M}=\mathbb{R}$.

Given $x_0,x_1\in\mathsf{M}$, denote

$$\mathrm{H}^s([t_0,t_1];\mathsf{M};x_0)=\{\gamma\in\mathrm{H}^s([t_0,t_1];\mathsf{M})\mid\ \gamma(t_0)=x_0\}$$

and

$$\mathrm{H}^s([t_0,t_1];\mathsf{M};x_0,x_1)=\{\gamma\in\mathrm{H}^s([t_0,t_1];\mathsf{M})\mid\ \gamma(t_0)=x_0,\ \gamma(t_1)=x_1\}.$$

Now suppose that we additionally have a smooth subbundle $\mathsf{D}\subseteq\mathsf{TM}$ and that $s\in\mathbb{Z}_{>0}$. Here we denote

$$\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})=\{\gamma\in\mathrm{H}^s([t_0,t_1];\mathsf{M})\mid\ \gamma'(t)\in\mathsf{D}_{\gamma(t)}\text{ a.e. }t\in[t_0,t_1]\}$$

and

$$\begin{aligned}\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D};x_0)&=\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})\cap\mathrm{H}^s([t_0,t_1];\mathsf{M};x_0),\\ \mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)&=\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})\cap\mathrm{H}^s([t_0,t_1];\mathsf{M};x_0,x_1).\end{aligned}$$

Next we consider a vector bundle $\pi\colon\mathsf{E}\to\mathsf{M}$. We recall from (2.4) the notions of affine and linear functions on E. We use these functions to characterise curves in the total space of a vector bundle.

3.1 Lemma: (Sobolev spaces of curves in the total space of a vector bundle) *If $\pi\colon\mathsf{E}\to\mathsf{M}$ is a smooth vector bundle and if $s\in\mathbb{Z}_{\geq 0}$, then*

$$\mathrm{H}^s([t_0,t_1];\mathsf{E})=\{\xi\colon[t_0,t_1]\to\mathsf{E}\mid\ F\circ\xi\in\mathrm{H}^s([t_0,t_1];\mathbb{R}),\ F\in\mathrm{Aff}^\infty(\mathsf{E})\}.$$

Proof: This is a consequence of the fact that, about any point $e\in\mathsf{E}$, one can choose a coordinate chart comprised of globally defined affine functions (e.g., the coordinates defined by a vector bundle chart). ■

Now fix $\gamma\colon[t_0,t_1]\to\mathsf{M}$ and denote

$$\gamma^*\mathsf{E}=\{(t,e)\in[t_0,t_1]\times\mathsf{E}\mid\ \gamma(t)=\pi(e)\}.$$

We wish to think about sections of E along γ. For $s\in\mathbb{Z}_{\geq 0}$, we have

$$\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{E})=\{\xi\colon[t_0,t_1]\to\mathsf{E}\mid\ \pi\circ\xi=\gamma,\ F\circ\xi\in\mathrm{H}^s([t_0,t_1];\mathbb{R}),\ F\in\mathrm{Aff}^\infty(\mathsf{E})\}.$$

One can verify that

$$\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{E})=\{\xi\in\mathrm{H}^s([t_0,t_1];\mathsf{E})\mid\ \pi\circ\xi=\gamma\};$$

that is, to characterise sections along a curve, it suffices to work with smooth affine functions, and not general smooth functions, just as in Lemma 3.1. In the case that $s=0$, we replace the symbol "H^0" with "L^2," this being sensible since the spaces are vector spaces.

The curve γ automatically inherits the regularity of the section.

3.2 Lemma: (Regularity of curves covered by regular sections) *Let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a smooth vector bundle and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Let $s \in \mathbb{Z}_{\geq 0}$. If $\gamma\colon [t_0, t_1] \to \mathsf{M}$ and if $\xi \in \mathrm{H}^s([t_0, t_1]; \gamma^*\mathsf{E})$, then $\gamma \in \mathrm{H}^s([t_0, t_1]; \mathsf{M})$.*

Proof: Since $\pi^* f \in \mathrm{Aff}^\infty(\mathsf{E})$ for every $f \in \mathrm{C}^\infty(\mathsf{M})$, it follows that

$$f \circ \gamma = \pi^* f \circ \xi \in \mathrm{H}^s([t_0, t_1]; \mathbb{R}),$$

as claimed. ■

Similarly as was done for curves, for $\gamma \in \mathrm{H}^s([t_0, t_1]; \mathsf{M}; x_0)$, denote

$$\mathrm{H}^s([t_0, t_1]; \gamma^*\mathsf{E}; x_0) = \{\xi \in \mathrm{H}^s([t_0, t_1]; \gamma^*\mathsf{E}) \mid \xi(t_0) = 0\}$$

and, for $\gamma \in \mathrm{H}^s([t_0, t_1]; \mathsf{M}; x_0, x_1)$, denote

$$\mathrm{H}^s([t_0, t_1]; \gamma^*\mathsf{E}; x_0, x_1) = \{\xi \in \mathrm{H}^s([t_0, t_1]; \gamma^*\mathsf{E}) \mid \xi(t_0) = 0,\ \xi(t_1) = 0\}.$$

3.2. Topology on spaces of curves and sections along curves. Part of what we do in this work is develop a means of rigorously working with the spaces $\mathrm{H}^s([t_0, t_1]; \mathsf{M})$, which are not vector spaces, without needing to introduce infinite-dimensional manifolds, which is the standard methodology one uses in these cases. The approach we describe here uses evaluation by functions to replace the nonlinear space $\mathrm{H}^s([t_0, t_1]; \mathsf{M})$ with linear spaces $\mathrm{H}^s([t_0, t_1]; \mathbb{R})$, indexed by smooth functions.

To begin, let M be a smooth manifold, let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$, and let $s \in \mathbb{Z}_{\geq 0}$. Given $f \in \mathrm{C}^\infty(\mathsf{M})$, we have a mapping

$$\begin{aligned} \mathrm{ev}_f\colon \mathrm{H}^s([t_0, t_1]; \mathsf{M}) &\to \mathrm{H}^s([t_0, t_1]; \mathbb{R}) \\ \gamma &\mapsto f \circ \gamma. \end{aligned}$$

The use of these maps is obviously suggested by our very definition of the space $\mathrm{H}^s([t_0, t_1]; \mathsf{M})$. Here we use these maps to render $\mathrm{H}^s([t_0, t_1]; \mathsf{M})$ a topological space with the family of semimetrics

$$\rho^s_{a,f}\colon \mathrm{H}^s([t_0, t_1]; \mathsf{M})^2 \to \mathbb{R}_{\geq 0}, \qquad a \in \{0, 1, \dots, s\},\ f \in \mathrm{C}^\infty(\mathsf{M}),$$

defined by

$$\rho^s_{a,f}(\gamma_1, \gamma_2) = \left(\int_{t_0}^{t_1} |(f \circ \gamma_1)^{(a)}(t) - (f \circ \gamma_2)^{(a)}(t)|^2 \,\mathrm{d}t\right)^{1/2}.$$

The resulting topology is then easily verified to be the initial topology for the mappings ev_f, $f \in \mathrm{C}^\infty(\mathsf{M})$. Let us make some comments on this topology.

1. If M is connected, M can be embedded in $\mathbb{R}^N$ for some $N \in \mathbb{Z}_{>0}$, and so there are finitely many functions $f^1, \dots, f^N \in \mathrm{C}^\infty(\mathsf{M})$ (see Lemma 1.1(ii)) so that the topology of $\mathrm{H}^s([t_0, t_1]; \mathsf{M})$ is determined by the semimetrics ρ^s_{a,f^j}, $a \in \{0, 1, \dots, s\}$, $j \in \{1, \dots, N\}$. This implies that the topology can be described by its convergent sequences, and we shall frequently make use of this fact. This observation then immediately carries over to the case when M is not connected by applying it to each connected component.
2. The completeness of $\mathrm{H}^s([t_0, t_1]; \mathbb{R})$ implies the completeness of $\mathrm{H}^s([t_0, t_1]; \mathsf{M})$.

3. By the Sobolev Embedding Theorem [Adams and Fournier 2003, Theorem 4.12], we have a continuous embedding

$$\mathrm{H}^{s+1}([t_0,t_1];\mathbb{R}) \to \mathrm{C}^s([t_0,t_1];\mathbb{R}).$$

Since the semimetrics

$$\rho^{\infty}_{a,f}(\gamma_1,\gamma_2) = \sup\{|(f\circ\gamma_1)^{(a)}(t) - (f\circ\gamma_2)^{(a)}(t)| \mid t\in[t_0,t_1]\}, \qquad a\in\{0,1,\dots,s\},\ f\in\mathrm{C}^\infty(\mathsf{M}),$$

define the usual topology on $\mathrm{C}^s([t_0,t_1];\mathsf{M})$, we infer that we have a continuous embedding

$$\mathrm{H}^{s+1}([t_0,t_1];\mathsf{M}) \to \mathrm{C}^s([t_0,t_1];\mathsf{M}). \tag{3.2}$$

Let us verify that the subsets of $\mathrm{H}^s([t_0,t_1];\mathsf{M})$ specified in the preceding section are closed.

3.3 Lemma: (Closed subsets of curves) *Let* M *be a smooth manifold, let* $t_0,t_1\in\mathbb{R}$ *satisfy* $t_0<t_1$*, let* $x_0,x_1\in\mathsf{M}$*, and let* $\mathsf{D}\subseteq\mathsf{TM}$ *be a smooth subbundle. Then, for* $s\in\mathbb{Z}_{>0}$*, the following are closed subsets of* $\mathrm{H}^s([t_0,t_1];\mathsf{M})$*:*

(i) $\mathrm{H}^s([t_0,t_1];\mathsf{M};x_0)$;

(ii) $\mathrm{H}^s([t_0,t_1];\mathsf{M};x_0,x_1)$;

(iii) $\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})$;

(iv) $\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$;

(v) $\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$.

Proof: (i) It suffices to show that the map

$$\begin{aligned}\mathrm{ev}_{t_0}\colon \mathrm{H}^s([t_0,t_1];\mathsf{M}) &\to \mathsf{M}\\ \gamma &\mapsto \gamma(t_0)\end{aligned}$$

is continuous. To verify this, let $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathrm{H}^s([t_0,t_1];\mathsf{M})$ converging to $\gamma\in\mathrm{H}^s([t_0,t_1];\mathsf{M})$. Thus

$$\lim_{j\to\infty}\rho^s_{a,f}(\gamma,\gamma_j)=0, \qquad a\in\{0,1,\dots,s\},\ f\in\mathrm{C}^\infty(\mathsf{M}),$$

and then by (3.2),

$$\lim_{j\to\infty}\rho^\infty_{a,f}(\gamma,\gamma_j)=0, \qquad a\in\{0,1,\dots,s-1\}, f\in\mathrm{C}^\infty(\mathsf{M}).$$

Since $s>0$, it follows that

$$\lim_{j\to\infty}|f\circ\gamma(t_0)-f\circ\gamma_j(t_0)|=0, \qquad f\in\mathrm{C}^\infty(\mathsf{M}).$$

This, in turn, implies that $(\gamma_j(t_0))_{j\in\mathbb{Z}_{>0}}$ converges to $\gamma(t_0)$, giving continuity of ev_{t_0}.

(ii) Here we can show, similarly to the preceding part of the proof, that the mapping

$$\begin{aligned}\mathrm{ev}_{(t_0,t_1)}\colon \mathrm{H}^s([t_0,t_1];\mathsf{M}) &\to \mathsf{M}\times\mathsf{M}\\ \gamma &\mapsto (\gamma(t_0),\gamma(t_1))\end{aligned}$$

is continuous.

(iii) We shall first show that the mapping

$$\begin{aligned}\hat{P}_{\mathsf{D}^\perp}\colon \mathrm{H}^s([t_0,t_1];\mathsf{M}) &\to \mathrm{H}^0([t_0,t_1];\mathsf{TM})\\ \gamma &\mapsto P_{\mathsf{D}^\perp}\circ\gamma'\end{aligned}$$

is well-defined and continuous.

Let $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathrm{H}^s([t_0,t_1];\mathsf{M})$ converging to γ. We claim that $(\gamma_j')_{j\in\mathbb{Z}_{>0}}$ converges to γ' in $\mathrm{H}^0([t_0,t_1];\mathsf{TM})$. To show this, we must show that $(F\circ\gamma_j')_{j\in\mathbb{Z}_{>0}}$ converges to $F\circ\gamma'$ for every $F\in\mathrm{Aff}^\infty(\mathsf{TM})$. Since this is immediately true for the affine functions π_{TM}^*f, $f\in\mathrm{C}^\infty(\mathsf{M})$, it suffices to show this for linear functions F. Since $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ converges to γ in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$, it follows that, for any $f\in\mathrm{C}^\infty(\mathsf{M})$,

$$\begin{aligned}0 &= \lim_{j\to\infty}\rho^s_{1,f}(\gamma_j,\gamma)\\ &= \lim_{j\to\infty}\left(\int_{t_0}^{t_1}|(f\circ\gamma_j)'(t)-(f\circ\gamma)'(t)|^2\,\mathrm{d}t\right)^{1/2}\\ &= \lim_{j\to\infty}\left(\int_{t_0}^{t_1}|\langle\mathrm{d}f(\gamma(t));\gamma_j'(t)\rangle-\langle\mathrm{d}f(\gamma(t));\gamma'(t)\rangle|^2\,\mathrm{d}t\right)^{1/2}.\end{aligned}$$

By Lemma 1.1(ii), there exists $f^1,\dots,f^N\in\mathrm{C}^\infty(\mathsf{M})$ such that, if $F\in\mathrm{Lin}^\infty(\mathsf{TM})$, then

$$F=\phi_F^1\mathrm{d}f^1+\dots+\phi_F^N\mathrm{d}f^N$$

for some $\phi_F^1,\cdots\phi_F^N\in\mathrm{C}^\infty(\mathsf{M})$. Since $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ converges uniformly to γ by (3.2), there exists a compact $K\subseteq\mathsf{M}$ so that $\mathrm{image}(\gamma_j)\subseteq K$, $j\in\mathbb{Z}_{>0}$, and $\mathrm{image}(\gamma)\subseteq K$. Let $F\in\mathrm{Lin}^\infty(\mathsf{TM})$ and denote

$$M=\sup\{|\phi_F^l(x)|\mid x\in K,\ l\in\{1,\dots,N\}\}.$$

Now we calculate

$$\begin{aligned}\rho^0_{0,F}(\gamma_j',\gamma') &= \left(\int_{t_0}^{t_1}|F\circ\gamma_j'(t)-F\circ\gamma'(t)|^2\,\mathrm{d}t\right)^{1/2}\\ &\le \sum_{l=1}^N\left(\int_{t_0}^{t_1}|\phi_F^l(\gamma_j(t))\langle\mathrm{d}f^l(\gamma_j(t));\gamma_j'(t)\rangle-\phi_F^l(\gamma(t))\langle\mathrm{d}f^l(\gamma(t));\gamma'(t)\rangle|^2\,\mathrm{d}t\right)^{1/2}\\ &\le M\sum_{l=1}^N\rho^s_{1,f^l}(\gamma_j,\gamma).\end{aligned}$$

Thus

$$\lim_{j\to\infty}\rho^0_{0,F}(\gamma_j',\gamma')=0,$$

giving continuity of the mapping $\gamma\mapsto\gamma'$ from $\mathrm{H}^s([t_0,t_1];\mathsf{M})$ to $\mathrm{H}^0([t_0,t_1];\mathsf{TM})$.

Now we show continuity of the mapping $\xi\mapsto P_{\mathsf{D}^\perp}\circ\xi$ from $\mathrm{H}^0([t_0,t_1];\mathsf{TM})$ to itself. As above, it suffices to show that, if $(\xi_j)_{j\in\mathbb{Z}_{>0}}$ converges to ξ, then $(F\circ P_{\mathsf{D}^\perp}\circ\xi_j)_{j\in\mathbb{Z}_{>0}}$

converges to $F \circ P_{\mathsf{D}^\perp} \circ \xi$ for every $F \in \mathrm{Lin}^\infty(\mathsf{TM})$. However, this follows immediately since $F \circ P_{\mathsf{D}^\perp} \in \mathrm{Lin}^\infty(\mathsf{TM})$.

Finally, we show that $\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})$ is closed in $\mathrm{H}^s([t_0,t_1];\mathsf{M})$. Let $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})$ converging to $\gamma \in \mathrm{H}^s([t_0,t_1];\mathsf{M})$. Note that $P_{\mathsf{D}^\perp} \circ \gamma'_j = 0$, $j \in \mathbb{Z}_{>0}$. Therefore, by the just demonstrated continuity of $\hat{P}_{\mathsf{D}^\perp}$,

$$P_{\mathsf{D}^\perp} \circ \gamma' = P_{\mathsf{D}^\perp}\left(\lim_{j\to\infty} \gamma'_j\right) = \lim_{j\to\infty} P_{\mathsf{D}^\perp} \circ \gamma'_j = 0,$$

and we conclude that $\gamma' \in \mathrm{H}^s([t_0,t_1];\mathsf{M};\mathsf{D})$.

Parts (iv) and (v) follow from the preceding parts of the proof since the intersection of closed sets is closed. ■

We shall require these subsets to have more regularity than being merely closed. However, we shall have to wait until we have some calculus at hand before we can make sense of such additional regularity.

We will make a few more purely topological constructions before we start doing calculus. We shall topologise the spaces $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{E})$ of sections along a curve $\gamma \in \mathrm{H}^s([t_0,t_1];\mathsf{M})$. This topology can be defined by the family of seminorms

$$\|\xi\|_{a,F} = \left(\int_{t_0}^{t_1} |(F\circ\xi_1)^{(a)}(t)|^2\,\mathrm{d}t\right)^{1/2}, \qquad a \in \{0,1,\dots,s\},\ F \in \mathrm{Lin}^\infty(\mathsf{E}).$$

Again, these definitions are suggested by our very definition of these spaces, along with the fact that, since we are considering sections over one fixed curve γ in M, the semimetrics ρ^s_{a,π^*f} will always evaluate to zero; that is, we need only consider linear functions when defining the topology. If $\mathbb{G}$ is a fibre metric on E, this topology can equivalently, and more easily, be defined by a single inner product:

$$\langle \xi_1,\xi_2\rangle_{\mathrm{H}^s} = \sum_{a=0}^{s} \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}{}^a_{\gamma'(t)}\xi_1(t), \overset{\mathbb{G}}{\nabla}{}^a_{\gamma'(t)}\xi_2(t))\,\mathrm{d}t.$$

We leave to the reader the quite simple exercise of showing that the two topologies agree.

We shall also use the following semi-inner product for $\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{E})$:

$$\langle \xi_1,\xi_2\rangle_{\mathrm{D}} = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\xi_1(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\xi_2(t))\,\mathrm{d}t, \tag{3.3}$$

called the ***Dirichlet semi-inner product***. If $\|\cdot\|_{\mathrm{D}}$ denotes the corresponding seminorm, note that $\|\xi\|_{\mathrm{D}} = 0$ if and only if ξ is constant. Therefore, as a consequence, $\langle\cdot,\cdot\rangle_{\mathrm{D}}$ is an inner product on $\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{E};x_0,x_1)$.

Let us verify that the subsets of $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM})$ specified in the preceding section are closed.

3.4 Lemma: (Closed subsets of sections along a curve) *Let M be a smooth manifold, let $t_0,t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$, let $x_0,x_1 \in \mathsf{M}$, let $\mathsf{D} \subseteq \mathsf{TM}$ be a smooth subbundle, and let $\gamma \in \mathrm{H}^s([t_0,t_1];\mathsf{M})$. Then, for $s \in \mathbb{Z}_{>0}$, the following are closed subsets of $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM})$:*

(i) $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM};x_0)$*;*

(ii) $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM};x_0,x_1)$*;*
(iii) $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{D})$*;*
(iv) $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{D};x_0)$*;*
(v) $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{D};x_0,x_1)$*.*

Proof: (i) It is enough to show that the map

$$\begin{aligned}\mathrm{ev}_{t_0}\colon \mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM}) &\to \mathsf{TM}\\ \xi &\mapsto \xi(t_0)\end{aligned}$$

is continuous. Let $(\xi_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM})$ converging to ξ. We then have that $(F\circ\xi_j)_{j\in\mathbb{Z}_{>0}}$ converges to $F\circ\xi$ for $F\in\mathrm{Lin}^\infty(\mathsf{TM})$. Therefore, since $(F\circ\xi_j)_{j\in\mathbb{Z}_{>0}}$ converges uniformly by (3.2), $(F\circ\xi_j(t_0))_{j\in\mathbb{Z}_{>0}}$ converges to $F\circ\xi(t_0)$. As this must hold for every $F\in\mathrm{Lin}^\infty(\mathsf{TM})$, we conclude that $(\xi(t_0))_{j\in\mathbb{Z}_{>0}}$ converges to $\xi(t_0)$, giving the desired continuity.

(ii) Here we can show that the map

$$\begin{aligned}\mathrm{ev}_{(t_0,t_1)}\colon \mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM}) &\to \mathsf{TM}\times\mathsf{TM}\\ \xi &\mapsto (\xi(t_0),\xi(t_1))\end{aligned}$$

is continuous, rather as in the first part of the proof, and this gives the desired conclusion.

(iii) We shall first show that the mapping

$$\begin{aligned}\hat{P}_{\mathsf{D}^\perp}\colon \mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM}) &\to \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{TM})\\ \xi &\mapsto P_{\mathsf{D}^\perp}\circ\xi\end{aligned}$$

is continuous. Thus let $(\xi_j)_{j\in\mathbb{Z}_{>0}}$ converge to ξ in $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM})$ so that

$$\lim_{j\to\infty}\left(\int_{t_0}^{t_1}|(F\circ(\xi_j-\xi))^{(a)}(t)|^2\,dt\right)^{1/2}=0,\qquad a\in\{0,\dots,s\},\ F\in\mathrm{Lin}^\infty(\mathsf{TM}).$$

In particular, this holds if we replace "F" with "$F\circ P_{\mathsf{D}^\perp}$", and this then gives

$$\lim_{j\to\infty}\left(\int_{t_0}^{t_1}|F\circ\hat{P}_{\mathsf{D}^\perp}(\xi_j-\xi)(t)|^2\,dt\right)^{1/2}=0,\qquad a\in\{0,\dots,s\},\ F\in\mathrm{Lin}^\infty(\mathsf{TM}),$$

which establishes the desired continuity.

Now we suppose that $(\xi_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM};\mathsf{D})$ converging to $\xi\in\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM})$. Then

$$\hat{P}_{\mathsf{D}^\perp}(\gamma_j)=0\quad\Longrightarrow\quad\hat{P}_{\mathsf{D}^\perp}(\gamma)=0,$$

and so $\gamma\in\mathrm{H}^s([t_0,t_1];\gamma^*\mathsf{TM};\mathsf{D})$, as desired.

Parts (iv) and (v) follow from the preceding parts of the proof since the intersection of closed sets is closed. ■

3.3. Calculus on spaces of curves I. We shall need to do calculus involving mappings whose domain and/or codomain is one of our spaces of curves. We build up this calculus piece by piece, starting in this section with differentiability of curves in the space of curves.

First we define a suitable version of differentiability for mappings with values in $\mathrm{H}^s([t_0,t_1];\mathsf{M})$.

3.5 Definition: (Differentiability for mappings with values in $\mathrm{H}^s([t_0,t_1];\mathsf{M})$) Let M and N be smooth manifolds, let $t_0,t_1\in\mathbb{R}$ satisfy $t_0<t_1$, and let $s,k\in\mathbb{Z}_{\geq 0}$. A mapping $\Phi\colon\mathsf{N}\to\mathrm{H}^s([t_0,t_1];\mathsf{M})$ is of ***class*** $\mathbf{C}^{\boldsymbol{k}}$ if $\mathrm{ev}_f\circ\Phi\colon\mathsf{N}\to\mathrm{H}^s([t_0,t_1];\mathbb{R})$ is of class C^k for every $f\in\mathrm{C}^\infty(\mathsf{M})$. •

This definition of differentiability is natural, given our definition of the Sobolev spaces of curves $\mathrm{H}^s([t_0,t_1];\mathsf{M})$. However, to prove that the definition is useful requires some analysis. To do this, let us introduce some notation. Given $\Phi\colon\mathsf{N}\to\mathrm{H}^s([t_0,t_1];\mathsf{M})$, we have a map

$$\begin{aligned}\hat{\Phi}\colon\mathsf{N}\times[t_0,t_1]&\to\mathsf{M}\\(y,t)&\mapsto\Phi(y)(t).\end{aligned}$$

For $t\in[t_0,t_1]$, we define

$$\begin{aligned}\hat{\Phi}^t\colon\mathsf{N}&\to\mathsf{M}\\y&\mapsto\hat{\Phi}(y,t)\end{aligned}$$

and, for $y\in\mathsf{N}$, we define

$$\begin{aligned}\hat{\Phi}_y\colon[t_0,t_1]&\to\mathsf{M}\\t&\mapsto\hat{\Phi}(y,t).\end{aligned}$$

Let us determine the properties of $\hat{\Phi}$ that characterise differentiability. It is possible to do this in general, however, we are solely interested here in C^1-mappings defined on an interval in $\mathbb{R}$. We are also primarily interested in working with $\mathrm{H}^s([t_0,t_1];\mathsf{M})$ when $s=1$. Therefore, we focus our analysis on this case. The workings of the general situation can easily be deduced from what we do by adding some notation.

Our first result is the following.

3.6 Lemma: (Curves of class $\mathbf{C}^1$ in $\mathbf{H}^1([t_0,t_1];\mathbb{R})$) *Let $J\subseteq\mathbb{R}$ be an interval, let $\sigma\colon J\to\mathrm{H}^1([t_0,t_1];\mathbb{R})$ be of class C^1 (in the usual sense of a mapping between open subsets of Banach spaces), and let $\boldsymbol{D}\sigma\colon J\to\mathrm{H}^1([t_0,t_1];\mathbb{R})$ be the derivative. Then the following statements hold:*

(i) $\hat{\sigma}_s$ is absolutely continuous for every $s\in J$;

(ii) $\hat{\sigma}^t\in\mathrm{C}^1(J;\mathbb{R})$ for each $t\in[t_0,t_1]$;

(iii) $\partial_1\hat{\sigma}(s,t)=\widehat{\boldsymbol{D}\sigma}(s,t)$ for $(s,t)\in J\times[t_0,t_1]$;

(iv) $\partial_1\hat{\sigma}\colon J\times[t_0,t_1]\to\mathbb{R}$ is continuous;

(v) the mixed partial derivatives $\partial_1\partial_2\hat{\sigma}$ and $\partial_2\partial_1\hat{\sigma}$ exist almost everywhere and agree almost everywhere on $J\times[t_0,t_1]$.

Proof: (i) This follows just because σ takes values in $\mathrm{H}^1([t_0,t_1];\mathbb{R})$.

We shall prove parts (ii)-(iv) together. Let $[s_0,s_1]\subseteq J$ be a compact subinterval and note that $\sigma|[s_0,s_1]$ is uniformly continuous. Thus, for $\epsilon\in\mathbb{R}_{>0}$, there exists $\delta\in\mathbb{R}_{>0}$ such that

$$\|\boldsymbol{D}\sigma(s)-\boldsymbol{D}\sigma(s')\|_{\mathrm{H}^1}<\epsilon,\qquad s,s'\in[s_0,s_1],\ |s-s'|<\delta.$$

For $s\in[s_0,s_1]$, let $h\in\mathbb{R}$ be such that $|h|\in(0,\delta)$ and such that $s+h\in[s_0,s_1]$. By the Mean Value Theorem [Abraham, Marsden, and Ratiu 1988, Proposition 2.4.8], there exists a between s and $s+h$ such that

$$\frac{\sigma(s+h)-\sigma(s)}{h}=\boldsymbol{D}\sigma(a).$$

Therefore,
$$\left\|\frac{\sigma(s+h)-\sigma(s)}{h}-\boldsymbol{D}\sigma(s)\right\|_{\mathrm{H}^1}=\|\boldsymbol{D}\sigma(a)-\boldsymbol{D}\sigma(s)\|_{\mathrm{H}^1}<\epsilon.$$
By (3.2), there exists $M\in\mathbb{R}_{>0}$ such that
$$\left|\frac{\sigma(s+h)(t)-\sigma(s)(t)}{h}-\boldsymbol{D}\sigma(s)(t)\right|<M\epsilon,\qquad t\in[t_0,t_1].$$
This all shows that, for $\epsilon\in\mathbb{R}_{>0}$, there exists $\delta\in\mathbb{R}_{>0}$ such that
$$\left|\frac{\hat{\sigma}(s+h,t)-\hat{\sigma}(s,t)}{h}-\widehat{\boldsymbol{D}\sigma}(s,t)\right|<\epsilon$$
for $t\in[t_0,t_1]$, $s\in[s_0,s_1]$, and $|h|<\delta$. This means, in particular, that

1. $\hat{\sigma}^t$ is differentiable at each $s\in[s_0,s_1]$.
2. $\partial_1\hat{\sigma}=\widehat{\boldsymbol{D}\sigma}$ is continuous on $[s_0,s_1]\times[t_0,t_1]$.

As this hold for any compact subinterval $[s_0,s_1]\subseteq J$, we obtain the conclusions (ii)–(iv) of the lemma.

(v) Let $[s_0,s_1]\subseteq J$ be a compact subinterval. Since
$$\boldsymbol{D}\hat{\sigma}_s=(\widehat{\boldsymbol{D}\sigma})_s\in\mathrm{H}^1([t_0,t_1];\mathbb{R}),\qquad s\in[s_0,s_1],$$
it follows that $\partial_2\partial_1\hat{\sigma}(s,t)$ exists for almost every $(s,t)\in[s_0,s_1]\times[t_0,t_1]$. Since $\boldsymbol{D}\sigma$ is continuous, it is bounded on $[s_0,s_1]$, and so there exists $M\in\mathbb{R}_{>0}$ such that
$$\|\boldsymbol{D}\sigma(s)\|_{\mathrm{H}^1}\le M,\qquad s\in[s_0,s_1].$$
This implies that, in particular,
$$\int_{t_0}^{t_1}|\partial_2\partial_1\hat{\sigma}(s,t)|^2\,\mathrm{d}t\le M,\qquad s\in[s_0,s_1].$$
By continuity of the inclusion
$$\mathrm{L}^2([t_0,t_1];\mathbb{R})\subseteq\mathrm{L}^1([t_0,t_1];\mathbb{R}),$$
this means that, possibly by suitably modifying M, we have
$$\int_{t_0}^{t_1}|\partial_2\partial_1\hat{\sigma}(s,t)|\,\mathrm{d}t\le M,\qquad s\in[s_0,s_1].$$
This then gives
$$\partial_2\partial_1\hat{\sigma}\in\mathrm{L}^1([s_0,s_1]\times[t_0,t_1];\mathbb{R}).$$
By [Minguzzi 2015, Theorem 7], we conclude that $\partial_1\partial_2\hat{\sigma}(s,t)$ exists and that
$$\partial_2\partial_1\hat{\sigma}(s,t)=\partial_1\partial_2\hat{\sigma}(s,t)$$
for almost every $(s,t)\in[s_0,s_1]\times[t_0,t_1]$. Since this is true for every compact subinterval $[s_0,s_1]\subseteq J$, this part of the lemma follows. ■

The lemma enables an understanding of the differentiability of a curve in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$.

3.7 Lemma: (Curves of class C^1 in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$) *Let $J \subseteq \mathbb{R}$ be an interval and let $\sigma\colon J \to \mathrm{H}^1([t_0,t_1];\mathsf{M})$ be of class C^1 (in the sense of Definition 3.5). Let $f \in \mathrm{C}^\infty(\mathsf{M})$. Then*

(i) $f \circ \hat{\sigma}_s$ is absolutely continuous for every $s \in J$,

(ii) $f \circ \hat{\sigma}^t \in \mathrm{C}^1(J;\mathbb{R})$ for every $t \in [t_0,t_1]$,

(iii) $\partial_1(f\circ\hat{\sigma})(s,t) = \widehat{\boldsymbol{D}(f\circ\sigma)}(s)(t)$ for $(s,t) \in J \times [t_0,t_1]$, and

(iv) $\partial_1(f\circ\hat{\sigma})\colon J \times [t_0,t_1] \to \mathbb{R}$ is continuous.

(v) the mixed partial derivatives $\partial_1\partial_2(f\circ\hat{\sigma})$ and $\partial_2\partial_1(f\circ\hat{\sigma})$ exist almost everywhere and agree almost everywhere on $J \times [t_0,t_1]$.

Proof: This follows immediately from the definition of σ being class C^1 and from Lemma 3.6. ■

As with our definition of $\mathrm{H}^s([t_0,t_1];\mathsf{M})$, we should verify that our definition of derivative agrees with standard coordinate versions. This follows along the lines of our brief discussion following (3.1), and noting that (1) the mapping between Sobolev spaces induced by changes of coordinate is continuous and (2) the definition of derivative for Banach space-valued functions is independent of equivalent norms.[6]

Let us now investigate what one can say about continuously differentiable curves in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ by working with the above definitions and constructions using post-composition with a smooth function. We note that

$$\partial_2(f\circ\widehat{\sigma})(s,t) = (f\circ\hat{\sigma}_s)'(t) = \langle \mathrm{d}f(\hat{\sigma}(s,t)); (\hat{\sigma}_s)'(t)\rangle.$$

From this expression, we note that the knowledge of $\partial_2(f\circ\widehat{\sigma})(s,t)$ for every $f \in \mathrm{C}^\infty(\mathsf{M})$ uniquely determines $(\hat{\sigma}_s)'(t) \in \mathsf{T}_{\hat{\sigma}(s,t)}\mathsf{M}$. Let us denote $\nu\sigma(s)(t) = (\hat{\sigma}_s)'(t)$. In like manner, we denote $\delta\sigma(s)(t) = (\hat{\sigma}^t)'(s)$. We shall use all manner of notation associated with these constructions. It is our intention that the notation appear natural, even if it is a bit cumbersome. With apologies out of the way, we have the following notation:

$$\begin{aligned}\nu\sigma(s)(t) &= \nu\hat{\sigma}(s,t) = \nu\hat{\sigma}_s(t) = \nu\hat{\sigma}^t(s),\\ \delta\sigma(s)(t) &= \delta\hat{\sigma}(s,t) = \delta\hat{\sigma}_s(t) = \delta\hat{\sigma}^t(s).\end{aligned}$$

In Figure 1 we illustrate how one should envision these quantities.

We next indicate where $\nu\sigma$ and $\delta\sigma$ takes their values.

3.8 Lemma: (The derivatives of a C^1-curve in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$) *Let M be a smooth manifold, let $J \subseteq \mathbb{R}$ be an interval, and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. For a continuously differentiable mapping $\sigma\colon J \to \mathrm{H}^1([t_0,t_1];\mathsf{M})$, we have*

(i) $\nu\sigma(s) \in \mathrm{H}^0([t_0,t_1];\mathsf{TM})$ and

(ii) $\delta\sigma(s) \in \mathrm{H}^1([t_0,t_1];\mathsf{TM})$

for all $s \in J$.

[6] As with our remarks following (3.1), there are matters of compactness of the domain of the change of coordinate that must be considered, but these are not problematic since our curves are defined on a domain $[t_0,t_1]$ that is compact.

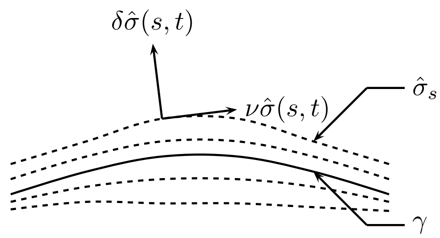

Figure 1. A depiction of $\nu\hat{\sigma}$ and $\delta\hat{\sigma}$. Note that $\nu\hat{\sigma}_s$ is the tangent vector field for $\hat{\sigma}_s$ and $\delta\hat{\sigma}^t$ is the tangent vector field for $\hat{\sigma}^t$.

Proof: Throughout the proof, we fix $s \in J$.

(i) We must show that $F \circ \nu\hat{\sigma}_s \in \mathrm{L}^2([t_0,t_1];\mathbb{R})$ for every $F \in \mathrm{Aff}^\infty(\mathsf{TM})$. First let $f \in \mathrm{C}^\infty(\mathsf{M})$. Then

$$\pi_{\mathsf{TM}}^* f \circ \nu\hat{\sigma}_s = f \circ \hat{\sigma}_s \in \mathrm{H}^1([t_0,t_1];\mathbb{R}) \subseteq \mathrm{L}^2([t_0,t_1];\mathbb{R}).$$

Next let $F \in \mathrm{Lin}^\infty(\mathsf{TM})$. By Lemma 1.1(ii), write

$$F = \phi^1 \mathrm{d}f^1 + \cdots + \phi^N \mathrm{d}f^N$$

for $\phi^l, f^l \in \mathrm{C}^\infty(\mathsf{M})$, $l \in \{1,\dots,N\}$. Since $f^l \circ \sigma$ takes values in $\mathrm{H}^1([t_0,t_1];\mathbb{R})$, we have

$$\partial_2(f^l \circ \hat{\sigma}_s) \in \mathrm{L}^2([t_0,t_1];\mathbb{R}), \qquad l \in \{1,\dots,N\}. \tag{3.4}$$

Now calculate

$$F \circ \nu\hat{\sigma}_s(t) = \sum_{l=1}^N \phi^l(\hat{\sigma}(s,t)\langle \mathrm{d}f^l(\hat{\sigma}(s,t)); \nu\hat{\sigma}_s(t)\rangle = \sum_{l=1}^N \phi^l(\hat{\sigma}(s,t))\partial_2(f^l \circ \hat{\sigma}_s)(t),$$

giving the desired conclusion by virtue of (3.4).

(ii) We must show that $F \circ \delta\hat{\sigma}_s \in \mathrm{H}^1([t_0,t_1];\mathbb{R})$ for every $F \in \mathrm{Aff}^\infty(\mathsf{TM})$. First let $f \in \mathrm{C}^\infty(\mathsf{M})$. Then

$$\pi_{\mathsf{TM}}^* f \circ \delta\hat{\sigma}_s = f \circ \hat{\sigma}_s \in \mathrm{H}^1([t_0,t_1];\mathbb{R}).$$

Next let $F \in \mathrm{Lin}^\infty(\mathsf{TM})$. Write

$$F = \phi^1 \mathrm{d}f^1 + \cdots + \phi^N \mathrm{d}f^N,$$

as above. Then

$$F \circ \delta\hat{\sigma}_s(t) = \sum_{l=1}^N \phi^l(\hat{\sigma}(s,t))\langle \mathrm{d}f^l(\hat{\sigma}(s,t)); \delta\hat{\sigma}_s(t)\rangle = \sum_{l=1}^N \phi^l(\hat{\sigma}(s,t))\boldsymbol{D}(f^l \circ \sigma)(s)(t).$$

Continuous differentiability of $f^l \circ \sigma$ means that

$$\boldsymbol{D}(f^l \circ \sigma)(s) \in \mathrm{H}^1([t_0,t_1];\mathbb{R}), \qquad l \in \{1,\dots,N\},$$

and this gives the lemma. ∎

The lemma then makes sense of the following definition.

3.9 Definition: (Derivative for mappings into spaces of curves) Let M be a smooth manifold, let $J \subseteq \mathbb{R}$ be an interval, and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. For a continuously differentiable mapping $\sigma\colon J \to \mathrm{H}^1([t_0,t_1];\mathsf{M})$, the ***derivative*** of σ is the mapping

$$\delta\sigma\colon J \to \mathrm{H}^1([t_0,t_1];\mathsf{TM})$$

defined by

$$\langle \mathrm{d}f(\hat{\sigma}(s,t)); \delta\sigma(s)(t)\rangle = \partial_1(f\circ\widehat{\sigma})(s,t), \qquad f \in \mathrm{C}^\infty(\mathsf{M}). \qquad \bullet$$

3.10 Remark: (The story for $\mathbf{H}^s([t_0,t_1];\mathsf{M})$)

1. We shall encounter differentiable curves with values in $\mathrm{H}^0([t_0,t_1];\mathsf{M})$. The adaptation from $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ to $\mathrm{H}^0([t_0,t_1];\mathsf{M})$ is, of course, easy: one merely throws away conditions and conclusions from our development above that involve "$\frac{\mathrm{d}}{\mathrm{d}t}$." We shall use this adaptation without further discussion when we require it.
2. It is also possible to extend the above development to higher-order Sobolev spaces of curves, $\mathrm{H}^s([t_0,t_1];\mathsf{M})$, $s \geq 2$. To do so without using coordinates requires working carefully with jet bundles. Jafarpour and Lewis [2014] give some constructions along these lines that are useful. •

3.4. Variations and infinitesimal variations. Having the definition and some properties for the derivative of a curve in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$, we may use these constructions to develop some of the usual players in the calculus of variations, such as variations of curves. Indeed, our first definitions give standard constructions from the calculus of variations in our setting.

3.11 Definition: (Variation, infinitesimal variation) Let M be a smooth manifold, let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$, and let $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$.

(i) A ***variation*** of γ is a mapping $\sigma\colon (-a,a) \to \mathrm{H}^1([t_0,t_1];\mathsf{M})$, where

  (a) $a \in \mathbb{R}_{>0}$,

  (b) $\sigma(0) = \gamma$, and

  (c) σ is continuously differentiable.

(ii) An ***infinitesimal variation*** of γ is an element of $\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{TM})$. •

One would like to think of an infinitesimal variation as being the "derivative" of a variation, and the following result makes this clear.

3.12 Lemma: (Variations and infinitesimal variations) *Let M be a smooth manifold, let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$, and let $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$. Then the following statements hold:*

(i) if $\sigma\colon (-a,a) \to \mathrm{H}^1([t_0,t_1];\mathsf{M})$ is a variation of γ, then $\delta\sigma(0)$ is an infinitesimal variation of γ;

(ii) if $\delta \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{TM})$, then there exists a variation σ of γ such that $\delta = \delta\sigma(0)$.

Proof: (i) This follows from Lemma 3.8.

(ii) Let $\mathbb{G}$ be a smooth Riemannian metric on M. Since δ is absolutely continuous, it is continuous. Thus it is bounded on the compact domain $[t_0,t_1]$, i.e.,

$$\sup\{\|\delta(t)\|_{\mathbb{G}} \mid t \in [t_0,t_1]\} < \infty.$$

Therefore, by a compactness argument, there exists $a \in \mathbb{R}_{>0}$ such that $\exp(s\delta(t))$ is defined for $s \in (-a, a)$ and $t \in [t_0, t_1]$. Define

$$\begin{aligned} \hat{\sigma} \colon (-a, a) \times [t_0, t_1] &\to \mathsf{M} \\ (s, t) &\mapsto \exp(s\delta(t)). \end{aligned}$$

We claim that, if σ is defined as usual, i.e., $\sigma(s)(t) = \hat{\sigma}(s, t)$, then it is a variation with the desired properties. Let $f \in \mathrm{C}^\infty(\mathsf{M})$. We must show that

1. $f \circ \hat{\sigma}_s \in \mathrm{H}^1([t_0, t_1]; \mathbb{R})$ for every $s \in (-a, a)$,
2. $s \mapsto f \circ \hat{\sigma}_s$ is continuous, and
3. $\delta\sigma(0) = \delta$.

Let $s \in (-a, a)$. Since $f \circ \hat{\sigma}_s$ is continuous, we have $f \circ \hat{\sigma}_s \in \mathrm{L}^2([t_0, t_1]; \mathbb{R})$. We also have

$$\partial_2(f \circ \hat{\sigma})(s, t) = s\langle \mathrm{d}f(\hat{\sigma}(s, t)); T\exp(\delta'(t))\rangle. \tag{3.5}$$

Since $\delta \in \mathrm{H}^1([t_0, t_1]; \mathsf{TM})$, $\delta' \in \mathrm{H}^0([t_0, t_1]; \mathsf{TTM})$, cf. Lemma 3.8(i). Therefore, since $T\exp$ is smooth, we conclude that $\partial_2(f \circ \hat{\sigma}_s) \in \mathrm{H}^0([t_0, t_1]; \mathsf{TM})$. From this we conclude that $f \circ \hat{\sigma}_s \in \mathrm{H}^1([t_0, t_1]; \mathbb{R})$, giving the first of our three desired conclusions.

To obtain continuity of $s \mapsto f \circ \hat{\sigma}_s$, we must show that one can make both of the integrals

$$\int_{t_0}^{t_1} |f \circ \hat{\sigma}(s_1, t) - f \circ \hat{\sigma}(s_2, t)|^2 \, dt$$

and

$$\int_{t_0}^{t_1} |\partial_2(f \circ \hat{\sigma})(s_1, t) - \partial_2(f \circ \hat{\sigma})(s_2, t)|^2 \, dt$$

small by making s_1 and s_2 close.

For the first, let $\epsilon \in \mathbb{R}_{>0}$. Continuity of $(s, t) \mapsto f \circ \hat{\sigma}(s, t)$ gives the existence of $\delta \in \mathbb{R}_{>0}$ such that, if $|s_1 - s_2| < \delta$, then

$$\int_{t_0}^{t_1} |f \circ \hat{\sigma}(s_1, t) - f \circ \hat{\sigma}(s_2, t)|^2 \, dt < \epsilon.$$

For the second integral, again let $\epsilon \in \mathbb{R}_{>0}$. By continuity of

$$s \mapsto \mathrm{d}f(\hat{\sigma}(s, t)),$$

smoothness of $T\exp$, and since $\delta' \in \mathrm{H}^0([t_0, t_1]; \mathsf{TTM})$, there exists $\delta_1 \in \mathbb{R}_{>0}$ such that, if $|s_1 - s_2| < \delta_1$,

$$\int_{t_0}^{t_1} |\langle \mathrm{d}f(\hat{\sigma}(s_1, t)) - \mathrm{d}f(\hat{\sigma}(s_2, t)); T\exp(\delta'(t))\rangle|^2 \, \mathrm{d}t < \frac{\epsilon}{2a^2}.$$

For the same reasons, there exists $M \in \mathbb{R}_{>0}$ such that

$$\int_{t_0}^{t_1} |\langle \mathrm{d}f(\hat{\sigma}(s_2, t)); T\exp(\delta'(t))\rangle|^2 \leq M.$$

We then let $\delta_2 \in \mathbb{R}_{>0}$ be such that, if $|s_1 - s_2| < \delta_2$, then $|s_1 - s_2|^2 < \frac{\epsilon}{2M}$. By (3.5), we then have

$$\begin{aligned}\int_{t_0}^{t_1} &|\partial_2(f\circ\hat{\sigma})(s_1,t) - \partial_2(f\circ\hat{\sigma})(s_2,t)|^2\,dt \\ &= \int_{t_0}^{t_1} |s_1\langle \mathrm{d}f(\hat{\sigma}(s_1,t)); T\exp(\delta'(t))\rangle - s_2\langle \mathrm{d}f(\hat{\sigma}(s_2,t)); T\exp(\delta'(t))\rangle|^2\,\mathrm{d}t \\ &\leq |s_1|^2 \int_{t_0}^{t_1} |\langle \mathrm{d}f(\hat{\sigma}(s_1,t)) - \mathrm{d}f(\hat{\sigma}(s_2,t)); T\exp(\delta'(t))\rangle|^2\,\mathrm{d}t \\ &\quad + |s_1 - s_2|^2 \int_{t_0}^{t_1} |\langle \mathrm{d}f(\hat{\sigma}(s_2,t)); T\exp(\delta'(t))\rangle|^2\,\mathrm{d}t < \epsilon.\end{aligned}$$

The preceding arguments give the second of our three desired conclusions.

Finally, we compute

$$\begin{aligned}\partial_1(f\circ\hat{\sigma}(0,t)) &= \langle \mathrm{d}f(\hat{\sigma}(0,t)); (\hat{\sigma}^t)'(0)\rangle \\ &= \langle \mathrm{d}f(\hat{\sigma}(0,t)); \delta(t)\rangle,\end{aligned}$$

since $s \mapsto \exp(s\delta(t))$ is the geodesic with initial velocity $\delta(t)$. Thus the definition of $\delta\sigma$ gives $\delta\sigma(0) = \delta(t)$, giving the final of the three desired conclusions. ■

We can now define tangent vectors and the tangent spaces for $\mathrm{H}^1([t_0,t_1];\mathsf{M})$. We have not established—and will not establish—a manifold structure for $\mathrm{H}^1([t_0,t_1];\mathsf{M})$, and so we cannot really think of what we define as being a tangent space in the strictest sense of the word. However, our definitions obviously so closely represent the usual notions that we do not feel guilty when we eliminate the quotation marks around geometric names for objects that do not have their usual strict geometric meaning.

With this round of apologies out of the way, we proceed with definitions.

3.13 Definition: (Tangent vector, tangent space) Let M be a smooth manifold and let $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$. Let $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$.

(i) A ***tangent vector*** at γ is an element of $\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{TM})$.

(ii) The union of all tangent vectors at γ is the ***tangent space*** to $\mathrm{H}^1([t_0,t_1];\mathsf{M})$, and we denote this by $\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$. •

Note that Lemma 3.12(ii) ensures that, if $\xi \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$, then there is a continuously differentiable curve $\sigma\colon J \to \mathrm{H}^1([t_0,t_1];\mathsf{M})$ satisfying $\delta\sigma(0) = \xi$. This is in agreement with what we know about tangent vectors.

We close this section by proving an important result about swapping the order of covariant derivatives along $\hat{\sigma}_s$ and $\hat{\sigma}^t$. This can be seen as the geometric consequence of the equality of mixed partial derivatives in Lemma 3.6(v). Typically this sort of lemma is proved by using the Lie bracket of the vector fields defining the "time" and "variation" parameters. However, this is problematic, in general, since these vector fields are not regular enough to allow this. However, our use of smooth function evaluations allows us to give a geometric proof without needing Lie brackets of things that do not have Lie brackets.

3.14 Lemma: (Swapping covariant derivatives for C^1-curves in $H^1([t_0, t_1]; \mathsf{M})$) *Let $(\mathsf{M}, \mathbb{G})$ be a smooth Riemannian manifold, let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$, let $J \subseteq \mathbb{R}$ be an open interval. Let $\sigma: J \to H^1([t_0, t_1]; \mathsf{M})$ be a C^1-curve. Then, for each $s \in J$,*

$$\overset{\mathbb{G}}{\nabla}_{\delta\hat{\sigma}_s(t)} \nu\hat{\sigma}^t(s) = \overset{\mathbb{G}}{\nabla}_{\nu\hat{\sigma}_s(t)} \delta\hat{\sigma}_s(t)$$

for almost every $t \in [t_0, t_1]$.

Proof: We compute

$$\partial_1(f \circ \hat{\sigma})(s,t) = \langle \mathrm{d}f(\hat{\sigma}(s,t)); \delta\hat{\sigma}(s,t)\rangle$$

and then

$$\partial_2\partial_1(f \circ \hat{\sigma})(s,t) = \langle \overset{\mathbb{G}}{\nabla}_{\hat{\sigma}'_s(t)} \mathrm{d}f(\hat{\sigma}(s,t)); \delta\hat{\sigma}(s,t)\rangle + \langle \mathrm{d}f(\hat{\sigma}(s,t)); \overset{\mathbb{G}}{\nabla}_{\hat{\sigma}'_s(t)} \delta\hat{\sigma}_s(t)\rangle. \tag{3.6}$$

Similarly,

$$\partial_1\partial_2(f \circ \hat{\sigma})(s,t) = \langle \overset{\mathbb{G}}{\nabla}_{(\hat{\sigma}^t)'(s)} \mathrm{d}f(\hat{\sigma}(s,t)); \nu\hat{\sigma}(s,t)\rangle + \langle \mathrm{d}f(\hat{\sigma}(s,t)); \overset{\mathbb{G}}{\nabla}_{(\hat{\sigma}^t)'(s)} \nu\hat{\sigma}^t(s)\rangle. \tag{3.7}$$

We now employ an elementary geometric sublemma.

1 Sublemma: *If $(\mathsf{M}, \mathbb{G})$ is a smooth Riemannian manifold and if α is a smooth one-form, then, for $x \in \mathsf{M}$,*

$$\mathrm{d}\alpha(u,v) = \langle \overset{\mathbb{G}}{\nabla}_u \alpha; v\rangle - \langle \overset{\mathbb{G}}{\nabla}_v \alpha; u\rangle, \qquad u, v \in \mathsf{T}_x\mathsf{M}.$$

Proof: Let $U, V \in \Gamma^\infty(\mathsf{TM})$ be such that $U(x) = u$ and $V(x) = v$. Then

$$\mathrm{d}\alpha(U,V) = \mathscr{L}_U\langle \alpha; V\rangle - \mathscr{L}_V\langle \alpha; U\rangle - \langle \alpha; [U,V]\rangle$$

by [Abraham, Marsden, and Ratiu 1988, Proposition 7.4.11]. We also have

$$\mathscr{L}_U\langle \alpha; V\rangle = \langle \overset{\mathbb{G}}{\nabla}_U \alpha; V\rangle + \langle \alpha; \overset{\mathbb{G}}{\nabla}_U V\rangle$$

and

$$\mathscr{L}_V\langle \alpha; U\rangle = \langle \overset{\mathbb{G}}{\nabla}_V \alpha; U\rangle + \langle \alpha; \overset{\mathbb{G}}{\nabla}_V U\rangle.$$

Combining the above formulae, and using the fact that $\overset{\mathbb{G}}{\nabla}$ is torsion-free, gives the sublemma. ▼

The sublemma gives

$$\langle \overset{\mathbb{G}}{\nabla}_{\hat{\sigma}'_s(t)} \mathrm{d}f(\hat{\sigma}(s,t)); \delta\hat{\sigma}(s,t)\rangle - \langle \overset{\mathbb{G}}{\nabla}_{(\hat{\sigma}^t)'(s)} \mathrm{d}f(\hat{\sigma}(s,t)); \nu\hat{\sigma}(s,t)\rangle = \mathrm{dd}f(\nu\hat{\sigma}(s,t), \delta\hat{\sigma}(s,t)) = 0,$$

after noting that

$$\hat{\sigma}'_s(t) = \nu\hat{\sigma}(s,t), \quad (\hat{\sigma}^t)'(s) = \delta\hat{\sigma}(s,t).$$

Combining this with (3.6) and (3.7), and also recalling Lemma 3.6(v), gives

$$\langle \mathrm{d}f(\hat{\sigma}(s,t)); \overset{\mathbb{G}}{\nabla}_{\hat{\sigma}'_s(t)} \delta\hat{\sigma}_s(t) - \overset{\mathbb{G}}{\nabla}_{(\hat{\sigma}^t)'(s)} \nu\hat{\sigma}^t(s)\rangle = 0.$$

As this holds for every $f \in C^\infty(\mathsf{M})$, the lemma follows. ■

3.5. Calculus on spaces of curves II. We now extend our calculus from Section 3.3 from curves with values in spaces of curves to mappings from spaces of curves to manifolds. We do this by making use of the notion of tangent space from the preceding section.

We begin with the following definition.

3.15 Definition: (Variational differentiability for mappings with domain $\mathrm{H}^1([t_0,t_1];\mathsf{M})$) Let M and N be smooth manifolds, and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Let $\Phi\colon \mathrm{H}^1([t_0,t_1];\mathsf{M}) \to \mathsf{N}$ and let $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$. The mapping Φ is ***variationally differentiable*** in the direction $\xi \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ if the mapping $g \circ \Phi \circ \sigma$ is differentiable at 0 for every $g \in \mathrm{C}^\infty(\mathsf{N})$, where σ is a variation of γ with $\delta\sigma(0) = \xi$. •

The rôle of the post-composition with the smooth function g seems less necessary—indeed, possibly seems intrusive—in the present framework. However, in the next section we shall need a more sophisticated form of differentiability, and in this setting the post-composition by a smooth function will allow for simpler definitions, and will be consistent with our definition above.

Note that, for $g \in \mathrm{C}^\infty(\mathsf{N})$ and for a variation σ of $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$, we have

$$\boldsymbol{D}(g \circ \Phi \circ \sigma)(0) = \langle \mathrm{d}g(\Phi(\gamma)); (\Phi \circ \sigma)'(0)\rangle.$$

Thus a knowledge of $\boldsymbol{D}g \circ \Phi \circ \sigma(0)$ for every $g \in \mathrm{C}^\infty(\mathsf{N})$ determines $(\Phi \circ \sigma)'(0)$. We denote

$$\delta\Phi(\gamma;\xi) = (\Phi \circ \sigma)'(0) \in T_{\Phi(\gamma)}\mathsf{N},$$

where $\xi = \delta\sigma(0)$. We call $\delta\Phi(\gamma;\xi)$ the ***variational derivative*** of Φ at γ in the direction of $\xi \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$.

With this notation, we can further refine our notion of differentiability.

3.16 Definition: (Differentiability for mappings with domain $\mathrm{H}^1([t_0,t_1];\mathsf{M})$) Let M and N be smooth manifolds, and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Let $\Phi\colon \mathrm{H}^1([t_0,t_1];\mathsf{M}) \to \mathsf{N}$ and let $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$. The mapping Φ is ***differentiable*** at γ if

(i) for every $\xi \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$, Φ is variationally differentiable at γ in the direction ξ and

(ii) there exists a linear map

$$T_\gamma\Phi\colon \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M}) \to \mathsf{T}_{\Phi(\gamma)}\mathsf{N}$$

such that $T_\gamma\Phi(\xi) = \delta\Phi(\gamma;\xi)$ for all $\xi \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$.

We call $T_\gamma\Phi$ the ***derivative*** of Φ at γ. •

We comment that this notion of derivative *does not* yet agree with the usual notion of Fréchet derivative, as for the latter one needs continuity of $\mathsf{T}_\gamma\Phi$ with respect to γ, cf. [Abraham, Marsden, and Ratiu 1988, Corollary 2.4.10]. While this is something we could impose and verify in all cases where we use the derivative, we pull up short of this since the technicalities would take us far afield.

3.6. Calculus on spaces of curves III. Now we extend our calculus to work for mappings with domain and codomain both being spaces of curves. Following Remark 3.10–1, we can extend the analysis to higher-order Sobolev spaces of curves, but we shall here use the lower-order space $\mathrm{H}^0([t_0,t_1];\mathsf{M})$, as this is the case we shall predominantly use.

3.17 Definition: (Variational differentiability for mappings with domain $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ and codomain $\mathrm{H}^0([t_0,t_1];\mathsf{N})$) Let M and N be smooth manifolds, and let $t_0,t_1\in\mathbb{R}$ satisfy $t_0<t_1$. Let $\Phi\colon \mathrm{H}^1([t_0,t_1];\mathsf{M})\to\mathrm{H}^0([t_0,t_1];\mathsf{N})$ and let $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M})$. The mapping Φ is ***variationally differentiable*** in the direction $\xi\in\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ if the mapping $\mathrm{ev}_g\circ\Phi\circ\sigma$ is differentiable at 0 for every $g\in\mathrm{C}^\infty(\mathsf{N})$, where σ is a variation of γ with $\delta\sigma(0)=\xi$. •

We note that

$$\mathrm{ev}_g\circ\Phi\circ\sigma(s)(t)=g(\Phi\circ\sigma(s)(t)).$$

We adopt our usual notation and write

$$\widehat{\Phi\circ\sigma}(s,t)=\Phi\circ\sigma(s)(t).$$

Then, according to our constructions preceding Lemma 3.8, we define

$$\delta(\Phi\circ\sigma)\colon J\to\mathrm{H}^0([t_0,t_1];\mathsf{TN})$$

by requiring that

$$\partial_1(g\circ\widehat{\Phi\circ\sigma})(s,t)=\langle \mathrm{d}g(\widehat{\Phi\circ\sigma}(s,t));\delta(\Phi\circ\sigma)(s)(t)\rangle.$$

If $\delta\sigma(0)=\xi$, then we denote

$$\delta\Phi(\gamma;\xi)=\delta(\Phi\circ\sigma)(0)\in\mathrm{H}^0([t_0,t_1];\mathsf{TN}),$$

which we call the ***variational derivative*** of Φ at γ in the direction of $\xi=\delta\sigma(0)$.

With this notation, we make the following definitions.

3.18 Definition: (Differentiability for mappings with domain $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ and codomain $\mathrm{H}^0([t_0,t_1];\mathsf{N})$) Let M and N be smooth manifolds, let $t_0,t_1\in\mathbb{R}$ satisfy $t_0<t_1$, and let $r,s\in\mathbb{Z}_{\geq0}$. Let $\Phi\colon\mathrm{H}^1([t_0,t_1];\mathsf{M})\to\mathrm{H}^0([t_0,t_1];\mathsf{N})$ and let $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M})$. The mapping Φ is ***differentiable*** at γ if

(i) for every $\xi\in\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$, Φ is variationally differentiable at γ in the direction ξ and

(ii) there exists a linear map

$$T_\gamma\Phi\colon\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})\to\mathsf{T}_{\Phi(\gamma)}\mathrm{H}^0([t_0,t_1];\mathsf{N})$$

such that $T_\gamma\Phi(\xi)=\delta\Phi(\gamma;\xi)$ for all $\xi\in\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$.

We call $T_\gamma\Phi$ the ***derivative*** of Φ at γ. •

3.7. The functor $\mathrm{H}^1([t_0,t_1];\bullet)$. The assignment $\mathsf{M}\mapsto\mathrm{H}^1([t_0,t_1];\mathsf{M})$ assigns a topological space to a smooth manifold. We see that morphisms in the category of smooth manifolds induce morphisms in the category of topological spaces, and this allows us to think of the assignment as a functor. In this section we explore some attributes of this functor.

We begin by establishing the natural way of assigning a mapping of curves to a mapping of manifolds.

3.19 Lemma: (The mapping of curves associated with a mapping of manifolds) *Let M and N be smooth manifolds, let $\Phi \in \mathrm{C}^\infty(\mathsf{M};\mathsf{N})$, and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Then the mapping*

$$\begin{aligned} \mathrm{H}^1([t_0,t_1];\Phi) \colon \mathrm{H}^1([t_0,t_1];\mathsf{M}) &\to \mathrm{H}^1([t_0,t_1];\mathsf{N}) \\ \gamma &\mapsto \Phi \circ \gamma \end{aligned}$$

is continuous.

Proof: First of all, we clearly have $\mathrm{H}^1([t_0,t_1];\Phi)(\gamma) \in \mathrm{H}^1([t_0,t_1];\mathsf{N})$ if $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$ since $g \circ \Phi \in \mathrm{C}^\infty(\mathsf{M})$ for $g \in \mathrm{C}^\infty(\mathsf{N})$. Thus we need only establish the continuity of $\mathrm{H}^1([t_0,t_1];\Phi)$. Let $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ converging to $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$. Then $(f \circ \gamma_j)_{j\in\mathbb{Z}_{>0}}$ converges to $f \circ \gamma$ for every $f \in \mathrm{C}^\infty(\mathsf{M})$. In particular, $(g \circ \Phi \circ \gamma_j)_{j\in\mathbb{Z}_{>0}}$ converges to $g \circ \Phi \circ \gamma$ for every $g \in \mathrm{C}^\infty(\mathsf{N})$. Thus $(\Phi \circ \gamma_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\Phi \circ \gamma$ and this gives the desired continuity. ■

If is clear that $\mathrm{H}^1([t_0,t_1];\mathrm{id}_\mathsf{M})$ is the identity on $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ and that, if $\Phi \in \mathrm{C}^\infty(\mathsf{M};\mathsf{N})$ and $\Psi \in \mathrm{C}^\infty(\mathsf{N};\mathsf{P})$, then

$$\mathrm{H}^1([t_0,t_1];\Psi \circ \Phi) = \mathrm{H}^1([t_0,t_1];\Psi) \circ \mathrm{H}^1([t_0,t_1];\Phi).$$

This then, indeed, defines a (covariant) functor $\mathrm{H}^1([t_0,t_1];\bullet)$ from the category of smooth manifolds to the category of topological spaces.

Let us also prove that the morphism $\mathrm{H}^1([t_0,t_1];\Phi)$ is differentiable, with differentiability as in Definition 3.18. This will require the reader to adapt Definition 3.18 from mappings into H^0 to mappings into H^1. This is easily done and gives the following result.

3.20 Lemma: (Differentiability of induced mappings in spaces of curves) *Let M and N be smooth manifolds, let $\Phi \in \mathrm{C}^\infty(\mathsf{M};\mathsf{N})$, and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Then the mapping $\mathrm{H}^1([t_0,t_1];\Phi)$ is differentiable at each $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$. Moreover,*

$$T_\gamma \mathrm{H}^1([t_0,t_1];\Phi)(\xi)(t) = T_{\gamma(t)}\Phi(\xi(t)), \qquad \xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M}).$$

Proof: Let $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$ and let $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M})$. By Lemma 3.12, let $\sigma \colon J \to \mathrm{H}^1([t_0,t_1];\mathsf{M})$ be a variation of γ such that $\delta\sigma(0) = \xi$. Let $g \in \mathrm{C}^\infty(\mathsf{M})$. To show that $\mathrm{H}^1([t_0,t_1];\Phi)$ is variationally differentiable at γ in the direction ξ, we must show that, if we define

$$\hat{\alpha}(s,t) = g \circ \Phi \circ \hat{\sigma}(s,t),$$

then the mapping $\alpha \colon J \to \mathrm{H}^1([t_0,t_1];\mathbb{R})$ is continuously differentiable. This, however, is immediate by the definition of continuously differentiability of σ and since $g \circ \Phi \in \mathrm{C}^\infty(\mathsf{M})$. Moreover, we have

$$\delta\alpha(0)(t) = \langle \mathrm{d}g(\Phi \circ \hat{\sigma}(0,t)); T_{\sigma(0,t)}\Phi(\delta\sigma(0)(t))\rangle,$$

and so we conclude that

$$T_\gamma \mathrm{H}^1([t_0,t_1];\Phi)(\xi) = T_{\gamma(t)}\Phi(\xi(t)),$$

giving differentiability of $\mathrm{H}^1([t_0,t_1];\Phi)$ at γ since the expression on the right is linear in ξ. ■

3.8. Weak covariant derivatives along curves. Let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a smooth vector bundle and let ∇ be a linear connection in E. For $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$, we have the operator

$$\nabla_{\gamma'}\colon \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{E}) \to \mathrm{L}^2([t_0, t_1]; \gamma^*\mathsf{E}).$$

We wish to extend this to an operator on "distributional sections" acting on "test sections" defined along γ. The difficulty in doing this geometrically is that the curve γ is generally not smooth, so the notion of a smooth compactly supported section along γ is problematic. Thus we here develop the framework for doing this.

First we give the definition of the space of test sections. For a smooth vector bundle $\pi\colon \mathsf{E} \to \mathsf{M}$, we denote by

$$\mathscr{D}(\mathsf{E}) = \{\Xi \in \Gamma^\infty(\mathsf{E}) \mid \ \mathrm{supp}(\Xi) \text{ compact}\}$$

the sections of E with compact support. We topologise this space in the usual way, by requiring that a sequence $(\Xi)_{j\in\mathbb{Z}_{>0}}$ in $\mathscr{D}(\mathsf{E})$ converges to zero if there exists a compact $K \subseteq \mathsf{M}$ such that $\mathrm{supp}(\Xi_j) \subseteq K$, $j \in \mathbb{Z}_{>0}$, and if $(\Xi_j)_{j\in\mathbb{Z}_{>0}}$ and all of its jets converge uniformly to zero.

3.21 Definition: (Test sections along a curve) For a smooth vector bundle $\pi\colon \mathsf{E} \to \mathsf{M}$ and for $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$, the ***test sections*** of E along γ consists of the sections

$$\begin{aligned} \gamma^*\Xi\colon [t_0, t_1] &\to \gamma^*\mathsf{E} \\ t &\mapsto (\Xi \circ \gamma(t), \gamma(t)) \end{aligned}$$

of $\gamma^*\mathsf{E}$ where

(i) $\Xi \in \mathscr{D}(\mathsf{E})$ and

(ii) $\mathrm{supp}(\Xi) \cap \gamma([t_0, t_1]) \subseteq \gamma((t_0, t_1))$.

By $\gamma^*\mathscr{D}(\mathsf{E})$ we denote the set of test sections of E over γ. We topologise $\gamma^*\mathscr{D}(\mathsf{E})$ by requiring that a sequence $(\gamma^*\Xi_j)_{j\in\mathbb{Z}_{>0}}$ converges to zero if

(i) there exists a compact $K \subseteq \mathsf{M}$ with $K \cap \gamma([t_0, t_1]) \subseteq \gamma((t_0, t_1))$ such that $\mathrm{supp}(\Xi_j) \subseteq K$, $j \in \mathbb{Z}_{>0}$, and

(ii) $(\Xi_j)_{j\in\mathbb{Z}_{>0}}$ and all of its jets converge uniformly to zero. •

Now we can define distributional sections of a vector bundle.

3.22 Definition: (Distributional sections of a vector bundle) Let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a smooth vector bundle and let $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$. A ***distributional section*** of E over γ is a continuous mapping from $\gamma^*\mathscr{D}(\mathsf{E}^*)$ to $\mathbb{R}$. We denote the set of distributional sections of E over γ by $\gamma^*\mathscr{D}'(\mathsf{E})$. We use the weak topology for $\gamma^*\mathscr{D}'(\mathsf{E})$; explicitly, we define the topology by requiring that a sequence $(\theta_j)_{j\in\mathbb{Z}_{>0}}$ in $\gamma^*\mathscr{D}'(\mathsf{E})$ converge to zero if $(\theta_j(\gamma^*\lambda))_{j\in\mathbb{Z}_{>0}}$ converge to zero for every $\gamma^*\lambda \in \gamma^*\mathscr{D}(\mathsf{E}^*)$. •

Note that, if $\xi \in \mathrm{L}^1([t_0, t_1]; \gamma^*\mathsf{E})$, then we define $\theta_\xi \in \gamma^*\mathscr{D}'(\mathsf{E})$ by

$$\langle \theta_\xi; \gamma^*\lambda \rangle = \int_{t_0}^{t_1} \langle \lambda \circ \gamma(t); \xi(t) \rangle \, \mathrm{d}t, \qquad \gamma^*\lambda \in \gamma^*\mathscr{D}(\mathsf{E}^*).$$

One can verify easily that the mapping $\xi \mapsto \theta_\xi$ is continuous.

Now we define the covariant derivative of a distributional section along a curve. Thus we suppose that E possesses a vector bundle connection ∇. We note that, if $\xi \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{E})$ and if $\gamma^*\lambda \in \gamma^*\mathscr{D}(\mathsf{E}^*)$, then we have

$$\begin{aligned}\langle \nabla_{\gamma'}\xi; \gamma^*\lambda\rangle &= \int_{t_0}^{t_1} \langle \lambda \circ \gamma(t); \nabla_{\gamma'(t)}\xi(t)\rangle \\ &= \int_{t_0}^{t_1} \frac{\mathrm{d}}{\mathrm{d}t}\langle \lambda \circ \gamma(t); \xi(t)\rangle \,\mathrm{d}t - \int_{t_0}^{t_1} \langle \nabla_{\gamma'(t)}\lambda \circ \gamma(t); \xi(t)\rangle \,\mathrm{d}t \\ &= -\int_{t_0}^{t_1} \langle \nabla_{\gamma'(t)}\lambda \circ \gamma(t); \xi(t)\rangle \,\mathrm{d}t,\end{aligned}$$

using the fact that $\lambda \circ \gamma(t_0) = 0$ and $\lambda \circ \gamma(t_1) = 0$. Motivated by this, for $\theta \in \gamma^*\mathscr{D}'(\mathsf{E})$, we define its ***covariant derivative*** along γ by

$$\langle \nabla_{\gamma'}\theta; \gamma^*\lambda\rangle = -\langle \theta; \nabla_{\gamma'}\gamma^*\lambda\rangle, \qquad \gamma^*\lambda \in \gamma^*\mathscr{D}(\mathsf{E}^*).$$

Showing that $\nabla_{\gamma'}\theta \in \gamma^*\mathscr{D}'(\mathsf{E})$ amounts to showing that $(\nabla_{\gamma'}\gamma^*\lambda_j)_{j\in\mathbb{Z}_{>0}}$ converges to zero if $(\gamma^*\lambda_j)_{j\in\mathbb{Z}_{>0}}$ converges to zero. This, however, is obvious by definition of convergence to zero in $\gamma^*\mathscr{D}(\mathsf{E}^*)$.

4. Invariant generalised or cogeneralised subbundles and affine subbundles, and affine subbundle varieties

In Section 7.1 we will lift the equation for constrained variational mechanics from the base space of a vector bundle to the total space of the vector bundle. The resulting equation is an affine equation in the vector bundle. Moreover, in Section 7.2 we shall see that it is interesting to consider when these affine equations leave invariant a cogeneralised subbundle obtained as the kernel of a certain vector bundle mapping. In this section we consider this setup in a general framework. A great deal of the required complication arises from the fact that we need to consider generalised and cogeneralised subbundles that have nonconstant rank. Moreover, as we are unaware of existing results of this nature, we are a little more comprehensive in our approach than is required by our subsequent use of these results.

In this section we carefully distinguish between smooth and real analytic regularity, as this plays an essential rôle in our results.

4.1. Varieties invariant under vector fields. Before we specialise to generalised and cogeneralised subbundles, to generalised and cogeneralised affine bundles, and to affine subbundle varieties that are invariant under linear and affine vector fields, it is illustrative to first introduce our notions of invariance in a more general setting. To this end, we consider vector fields that leave invariant some subset of a manifold, allowing the case where the subset may not be a submanifold. We do require, however, that the subset have *some* structure, namely that of a variety as given in Section 2.7.

For a general subset $S \subseteq \mathsf{M}$, we shall make the following construction.

4.1 Definition: (Ideal sheaf of a subset) Let $r \in \{\infty, \omega\}$ and let M be a C^r-manifold. The ***ideal sheaf*** of $S \subseteq \mathsf{M}$ is the subsheaf (of $\mathscr{C}^r_{\mathsf{M}}$-modules) $\mathscr{I}_S \subseteq \mathscr{C}^r_{\mathsf{M}}$ defined by

$$\mathscr{I}_S(\mathcal{U}) = \begin{cases} \{f \in \mathrm{C}^r(\mathcal{U}) \mid \ f|(S \cap \mathcal{U}) = 0\}, & S \cap \mathcal{U} \neq \varnothing, \\ \mathrm{C}^r(\mathcal{U}), & S \cap \mathcal{U} = \varnothing. \end{cases}$$ •

With this terminology, we can make a definition of what we mean for a vector field to leave a subset invariant. To make the definition, we note that, if $X \in \Gamma^r(\mathsf{TM})$, we have a sheaf morphism (in the category of $\mathbb{R}$-vector spaces)

$$\mathscr{L}_X : \mathscr{C}^r_{\mathsf{M}} \to \mathscr{C}^r_{\mathsf{M}}$$

defined by, for $f \in \mathscr{C}^r_{\mathsf{M}}(\mathcal{U})$,

$$\mathscr{L}_X f(x) = \langle \mathrm{d}f(x); X(x) \rangle, \qquad x \in \mathcal{U}.$$

This is, of course, simply the sheaf version of the ordinary Lie derivative of functions with respect to a vector field; thus we are guilt-free in using the same notation.

4.2 Definition: (Invariant subsets for $\mathbf{C}^r$-vector field) Let $r \in \{\infty, \omega\}$ and let M be a C^r-manifold. Let $X \in \Gamma^r(\mathsf{TM})$ and let $S \subseteq \mathsf{M}$.

(i) The subset S is ***invariant*** under X if $\mathscr{L}_X(\mathscr{I}_S) \subseteq \mathscr{I}_S$.

(ii) The subset S is ***flow-invariant*** under X if $\Phi^X_t(x) \in S$ for every $(t, x) \in \mathbb{R} \times S$ for which $\Phi^X_t(x)$ is defined. •

While the preceding definitions are made for arbitrary subsets, they are really only useful for C^r-varieties as it is only for varieties that one can usefully connect the notions of invariance and flow-invariance.

4.3 Proposition: (Relationship between invariance and flow-invariance) *Let $r \in \{\infty, \omega\}$ and let M be a C^r-manifold. For $X \in \Gamma^r(\mathsf{TM})$ the following statements hold:*

(i) a subset $S \subseteq \mathsf{M}$ is invariant under X if it is flow-invariant under X;

(ii) if $r = \omega$, then a C^r-variety $\mathsf{S} \subseteq \mathsf{M}$ is flow-invariant under X if it is invariant under X;

(iii) if $r = \infty$ and if $\mathsf{S} \subseteq \mathsf{M}$ is a C^∞-submanifold, then S is flow-invariant under X if it is invariant under X.

Proof: (i) Let $\mathcal{U} \subseteq \mathsf{M}$ be open, let $f \in \mathscr{I}_S(\mathcal{U})$, and let $x \in \mathcal{U}$. Let $T \in \mathbb{R}_{>0}$ be such that $\Phi^X_t(x)$ exists and is in $\mathcal{U}$ for $t \in (-T, T)$. Since S is flow-invariant under X and since $f \in \mathscr{I}_S(\mathcal{U})$, $f \circ \Phi^X_t(x) = 0$ for every $t \in (-T, T)$. Thus, by [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.10],

$$\mathscr{L}_X f(x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f \circ \Phi^X_t(x) = 0,$$

and so $\mathscr{L}_X f \in \mathscr{I}_S(\mathcal{U})$.

(ii) Let $x \in \mathsf{M}$ and let $\mathcal{U}$ be a neighbourhood of x. Let $T \in \mathbb{R}_{>0}$ be such that $\Phi^X_t(x)$ exists and is in $\mathcal{U}$ for $t \in (-T, T)$. By [Sontag 1998, Proposition C.3.12], the mapping $t \mapsto \Phi^X_t(x)$ is real analytic. Thus, for $f \in \mathrm{C}^\omega(\mathcal{U})$, $t \mapsto f \circ \Phi^X_t(x)$ is real analytic. Moreover, by an elementary induction,

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}\right|_{t=0} f \circ \Phi^X_t(x) = \underbrace{\mathscr{L}_X \cdots \mathscr{L}_X f}_{k \text{ times}}(x).$$

Again by induction, if $f \in \mathscr{I}_{\mathsf{S}}(\mathcal{U})$, then

$$\underbrace{\mathscr{L}_X \cdots \mathscr{L}_X f}_{k \text{ times}} \in \mathscr{I}_{\mathsf{S}}(\mathcal{U}), \qquad k \in \mathbb{Z}_{\geq 0}.$$

Therefore, for $f \in \mathscr{I}_{\mathsf{S}}(\mathcal{U})$, we have, possibly after shrinking T,

$$f \circ \Phi^X_t(x) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \underbrace{\mathscr{L}_X \cdots \mathscr{L}_X f}_{k \text{ times}}(x), \qquad t \in (-T, T).$$

Thus,

$$x \in \mathsf{S} \implies \Phi^X_t(x) \in \mathsf{S}, \qquad t \in (-T, T).$$

Let us show that the above arguments show that S is flow-invariant under X under the current hypotheses. Suppose that it is not. Then there exists $x \in \mathsf{S}$ and $t \in \mathbb{R}$ such that $\Phi^X_t(x)$ is defined, but $\Phi^X_t(x) \not\in \mathsf{S}$. We can assume for concreteness that $t \in \mathbb{R}_{>0}$. Let

$$t_* = \sup\{t \in \mathbb{R}_{\geq 0} \mid \Phi^X_t(x) \in \mathsf{S}\}.$$

Since S is closed and since $t \mapsto \Phi_t^X(x)$ is continuous, $\Phi_{t_*}^X(x) \in \mathsf{S}$. The argument above then gives $T \in \mathbb{R}_{>0}$ such that, for $t \in (-T, T)$,

$$\Phi_t^X \circ \Phi_{t_*}^X(x) \in \mathsf{S} \quad \Longrightarrow \quad \Phi_{t+t_*}^X(x) \in \mathsf{S},$$

in contradiction with the definition of t_*.

(iii) Let $x \in \mathsf{S}$, and let $\mathcal{U}$ be a neighbourhood of x in M and $f^1, \dots, f^k \in \mathrm{C}^r(\mathcal{U})$ be such that

$$\mathsf{S} \cap \mathcal{U} = \bigcap_{j=1}^k (f^j)^{-1}(0).$$

Since S is a submanifold, we can assume that $\mathrm{d}f^1(x), \dots, \mathrm{d}f^k(x)$ are linearly independent. By shrinking $\mathcal{U}$ we can ensure that $\mathrm{d}f^1, \dots, \mathrm{d}f^k$ are linearly independent on $\mathcal{U}$. We now make use of a lemma.

1 Lemma: *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be a neighbourhood of $\mathbf{0}$, let $\mathsf{S} \subseteq \mathbb{R}^n$ be the subspace*

$$\mathsf{S} = \{(x^1, \dots, x^n) \in \mathbb{R}^n \mid x^1 = \dots = x^k = 0\},$$

and let $f \in \mathrm{C}^\infty(\mathcal{U})$ satisfy $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathsf{S} \cap \mathcal{U}$. Let $\mathrm{pr}_{\mathsf{S}} \colon \mathbb{R}^n \to \mathsf{S}$ be the natural projection onto the first k-components. Then there exists a neighbourhood $\mathcal{V} \subseteq \mathcal{U}$ of $\mathbf{0}$ and $g^1, \dots, g^k \in \mathrm{C}^\infty(\mathcal{V})$ such that

$$f(\boldsymbol{x}) = \sum_{j=1}^k g^j(\boldsymbol{x}) x^j, \qquad \boldsymbol{x} \in \mathcal{V}.$$

Proof: The hypothesis is that f vanishes on $\mathsf{S} \cap \mathcal{U}$. Let $\mathcal{W} \subseteq \mathsf{S}$ be a neighbourhood of $\mathbf{0}$ and let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}(\epsilon, \boldsymbol{y}) \subseteq \mathcal{U}$ for all $\boldsymbol{x} \in \mathcal{W}$, possibly after shrinking $\mathcal{W}$. Let

$$\mathcal{V} = \cup_{\boldsymbol{x} \in \mathcal{W}} \mathsf{B}(\epsilon, \boldsymbol{x}).$$

Let $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{V}$ (with $\boldsymbol{x}_1 \in \mathsf{S}$) and define

$$\begin{aligned} \gamma_{\boldsymbol{x}} \colon [0,1] &\to \mathbb{R} \\ t &\mapsto f(\boldsymbol{x}_1, t\boldsymbol{x}_2). \end{aligned}$$

We calculate

$$\begin{aligned} f(\boldsymbol{x}) = f(\boldsymbol{x}_1, \boldsymbol{x}_2) &= f(\boldsymbol{x}_1, \boldsymbol{x}_2) - f(\boldsymbol{x}_1, \mathbf{0}) \\ &= \gamma_{\boldsymbol{x}}(1) - \gamma_{\boldsymbol{x}}(0) = \int_0^1 \gamma'_{\boldsymbol{x}}(t)\,\mathrm{d}t \\ &= \sum_{j=1}^k x^j \int_0^1 \frac{\partial f}{\partial x^j}((\boldsymbol{x}_1, t\boldsymbol{x}_2))\,\mathrm{d}t = \sum_{j=1}^k x^j g^j(\boldsymbol{x}), \end{aligned}$$

where

$$g^j(\boldsymbol{x}) = \int_0^1 \frac{\partial f}{\partial x^j}(\boldsymbol{x}_1, t\boldsymbol{x}_2)\,\mathrm{d}t, \qquad j \in \{1, \dots, k\}.$$

By standard theorems on interchanging derivatives and integrals [Jost 2005, Theorem 16.11], we can conclude that $g^1, \dots, g^k$ are smooth since f is smooth. ▼

The lemma implies that $f^1,\dots,f^k$ generate $\mathscr{I}_{\mathsf{S}}(\mathcal{U})$, possibly after shrinking $\mathcal{U}$. Then, by hypothesis,

$$\mathscr{L}_X f^j = g^j_1 f^1 + \cdots + g^j_k f^k$$

for some $g^j_l \in \mathrm{C}^\infty(\mathcal{U})$, $j,l\in\{1,\dots,k\}$. Therefore,

$$\langle \mathrm{d}f^j(x); X(x)\rangle = 0, \qquad j\in\{1,\dots,k\},\ x\in \mathsf{S}\cap\mathcal{U}.$$

We conclude from this that $X|\mathcal{U}$ is tangent to $\mathsf{S}\cap\mathcal{U}$. As this holds in some neighbourhood of any point in S, we conclude that X is tangent to S. From this we conclude that S is flow-invariant under X.[7] ■

4.2. Generalised and cogeneralised subbundles invariant under linear vector fields. We wish to consider linear vector fields on the total space of a vector bundle $\pi\colon \mathsf{E}\to\mathsf{M}$ that leave invariant a generalised or cogeneralised subbundle $\mathsf{F}\subseteq\mathsf{E}$, allowing the case where F may not be a subbundle. The following elementary lemma will be frequently called upon. We remind the reader of the notions of vertical evaluation introduced in Definition 2.1 and of the annihilator of a generalised subbundle introduced in Definition 2.15.

4.4 Lemma: (Vertical evaluations and the ideal sheaf of a cogeneralised subbundle) *Let $r\in\{\infty,\omega\}$, let $\pi\colon\mathsf{E}\to\mathsf{M}$ be a C^r-vector bundle, and let $\mathsf{F}\subseteq\mathsf{E}$ be a C^r-cogeneralised subbundle. Define a sheaf morphism (of $\mathscr{C}^r_{\mathsf{M}}$-modules)*

$$\cdot^{\mathrm{e}}\colon \mathscr{G}^r_{\mathsf{E}^*}\to\mathscr{C}^r_{\mathsf{E}}$$

by

$$(\mathscr{G}^r_{\mathsf{E}^*}(\mathcal{U}))^{\mathrm{e}} = \{\lambda^{\mathrm{e}} \mid \lambda\in\mathscr{G}^r_{\mathsf{E}^*}(\mathcal{U})\}$$

for $\mathcal{U}\subseteq\mathsf{M}$ open. Then $(\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}\subseteq\mathscr{I}_{\mathsf{F}}$ and, moreover, for $e\in\mathsf{F}$, there is a neighbourhood $\mathcal{V}\subseteq\mathsf{E}$ of e such that

$$\mathsf{F}\cap\mathcal{V} = \{e'\in\mathcal{V} \mid \lambda^{\mathrm{e}}(e')=0,\ \lambda\in\mathscr{G}^r_{\Lambda(\mathsf{F})}(\pi(\mathcal{V}))\}.$$

Proof: The only not completely obvious assertion is the final one, but this follows from Corollary 2.18. ■

The idea of the lemma is that, to carve out the cogeneralised subbundle F, it suffices to use vertical evaluations of sections of $\Lambda(\mathsf{F})$. We note that, as a consequence of this, C^r-cogeneralised subbundles are C^r-varieties (stated as Corollary 2.19). However, C^r-generalised subbundles are not generally C^r-varieties, e.g., they are not generally closed. Nonetheless, we will give useful theories of invariance and flow-invariance for both generalised and cogeneralised subbundles.

As a first step towards this, we now introduce the notions of invariance in which we shall be interested. To do so we first note that, as a consequence of Lemma 2.2(i) and parts (iv) and (vi) of Lemma 2.10, we have

$$\mathscr{L}_X((\mathscr{G}^r_{\mathsf{E}^*})^{\mathrm{e}}) \subseteq (\mathscr{G}^r_{\mathsf{E}^*})^{\mathrm{e}}, \tag{4.1}$$

whenever X is a linear vector field on $\pi\colon\mathsf{E}\to\mathsf{M}$.

[7] We assume the well-known and "obvious" fact that, if a vector field is tangent to a submanifold, then the submanifold is flow-invariant under the vector field.

4.5 Definition: (Generalised or cogeneralised subbundles invariant under linear vector fields) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $X \in \Gamma^r(\mathsf{TE})$ be a linear vector field over X_0.

(i) A C^r-cogeneralised subbundle $\mathsf{F} \subseteq \mathsf{E}$ is ***invariant*** under X if $\mathscr{L}_X((\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}) \subseteq (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}$.

(ii) A C^r-generalised subbundle $\mathsf{F} \subseteq \mathsf{E}$ is ***invariant*** under X if $\Lambda(\mathsf{F})$ is invariant under X^*.

(iii) A C^r-generalised or a C^r-cogeneralised subbundle $\mathsf{F} \subseteq \mathsf{E}$ is ***flow-invariant*** under X if $\Phi^X_t(e) \in \mathsf{F}$ for every $(t, e) \in \mathbb{R} \times \mathsf{F}$ for which $\Phi^{X_0}_t(\pi(e))$ is defined. •

4.6 Remark: (Notions of invariance for subbundles) There is an issue that must be addressed here. Note that, if $\mathsf{F} \subseteq \mathsf{E}$ is a C^r-cogeneralised subbundle, by virtue of being a C^r-variety of E (by Corollary 2.19) it has an ideal sheaf $\mathscr{I}_{\mathsf{F}}$. One can then ask whether the notion of invariance of F under a linear vector field agrees with that of Definition 4.2. This question boils down to whether $(\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}$ generates $\mathscr{I}_{\mathsf{F}}$ as a $\mathscr{C}^r_{\mathsf{E}}$-module. This is certainly true when F is a subbundle, but we were not able to prove this when F is not regular. However, our approach and Lemma 4.4 obviates the need to know this, and gives the results that we want. Nonetheless, this does leave open an interesting question. •

Let us state a more or less obvious result.

4.7 Lemma: (Correspondence between flow-invariance of F and $\Lambda(\mathsf{F})$) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-generalised or a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $X \in \Gamma^r(\mathsf{TE})$ be a linear vector field over X_0. Then F is flow-invariant under X if and only if $\Lambda(\mathsf{F})$ is flow-invariant under X^*.*

Proof: Suppose that F is invariant under X, let $\alpha \in \Lambda(\mathsf{F})$, and let $x = \pi^*(\alpha)$. Let $t \in \mathbb{R}$ be such that $\Phi^{X_0}_t(x)$ exists and compute, for $e \in \mathsf{F}_{\Phi^X_t(\pi^*(x))}$,

$$\langle \Phi^{X^*}_t(\alpha); e\rangle = \langle \alpha; \Phi^X_{-t}(e)\rangle = 0,$$

since $\Phi^X_{-t}(e) \in \mathsf{F}_x$. Thus $\Phi^{X^*}_t(\alpha) \in \Lambda(\mathsf{F})_x$.

The proof of the other implication is carried out similarly. ■

Let us explore the relationship between subbundles that are invariant and those that are flow-invariant.

4.8 Proposition: (Relationship between invariance and flow-invariance under linear vector fields) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-generalised or a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$, and let $X^{\mathrm{lin}} \in \Gamma^r(\mathsf{TE})$ be a linear vector field over X_0. Consider the following statements:*

(i) F is flow-invariant under X^{lin};

(ii) F is invariant under X^{lin}.

Then (i) $\Longrightarrow$ (ii) and, if either (1) $r = \omega$ or (2) $r = \infty$ and F is a subbundle, then (ii) $\Longrightarrow$ (i).

Proof: (i) $\Longrightarrow$ (ii) First we suppose that F is a C^r-cogeneralised subbundle. Let $\mathcal{U} \subseteq \mathsf{M}$ be open, let $\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})$, and let $e \in \mathsf{F}|\mathcal{U}$. Let $T \in \mathbb{R}_{>0}$ be such that $\Phi^{X^{\mathrm{lin}}}_t(e)$ exists and is in $\pi^{-1}(\mathcal{U})$ for $t \in (-T, T)$. Since F is flow-invariant under X^{lin} and since $\lambda^{\mathrm{e}} \in (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}(\mathcal{U})$,

$\lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{lin}}}(e) = 0$ for every $t \in (-T, T)$. Thus, by [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.10],

$$\mathscr{L}_{X^{\mathrm{lin}}} \lambda^{\mathrm{e}}(e) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{lin}}}(e) = 0,$$

and so $\mathscr{L}_{X^{\mathrm{lin}}} \lambda^{\mathrm{e}} \in (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}(\mathcal{U})$ by Lemmata 2.2(i), and 2.10(iv) and (vi).

Now let F be a C^r-generalised subbundle that is flow-invariant under X^{lin}. By Lemma 4.7 and the first part of the lemma, $\Lambda(\mathsf{F})$ is invariant under $X^{\mathrm{lin},*}$. By definition, this is the same thing as F being invariant under X^{lin}.

(ii) $\Longrightarrow$ (i) Let us first suppose that F is a C^ω-cogeneralised subbundle. Let $e \in \mathsf{E}$ and let $\mathcal{U}$ be a neighbourhood of $\pi(e)$. Let $T \in \mathbb{R}_{>0}$ be such that $\Phi_t^{X^{\mathrm{lin}}}(e)$ exists and is in $\pi^{-1}(\mathcal{U})$ for $t \in (-T, T)$. By [Sontag 1998, Proposition C.3.12], the mapping $t \mapsto \Phi_t^{X^{\mathrm{lin}}}(e)$ is real analytic. Thus, for $\lambda \in \Gamma^\omega(\mathsf{E}|\mathcal{U})$, $t \mapsto \lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{lin}}}(e)$ is real analytic. Moreover, by an elementary induction on $k \in \mathbb{Z}_{>0}$,

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}\right|_{t=0} \lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{lin}}}(e) = \underbrace{\mathscr{L}_{X^{\mathrm{lin}}} \cdots \mathscr{L}_{X^{\mathrm{lin}}}}_{k \text{ times}} \lambda^{\mathrm{e}}(x).$$

Again by induction and by the current hypotheses, if $\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})$, then

$$\underbrace{\mathscr{L}_{X^{\mathrm{lin}}} \cdots \mathscr{L}_{X^{\mathrm{lin}}} \lambda^{\mathrm{e}}}_{k \text{ times}} \in (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}(\mathcal{U}), \qquad k \in \mathbb{Z}_{\geq 0},$$

by virtue of (4.1). Therefore, for $\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})$, we have, possibly after shrinking T,

$$\lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{lin}}}(e) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \underbrace{\mathscr{L}_{X^{\mathrm{lin}}} \cdots \mathscr{L}_{X^{\mathrm{lin}}} \lambda^{\mathrm{e}}}_{k \text{ times}}(e) = 0, \qquad t \in (-T, T).$$

By Corollary 2.18, we conclude that, if $e \in \mathsf{F}$, then $\Phi_t^{X^{\mathrm{lin}}}(e) \in \mathsf{F}$.

Let us show that the above arguments show that F is flow-invariant under X^{lin} under the current hypotheses. Suppose that it is not. Then there exists $e \in \mathsf{F}$ and $t \in \mathbb{R}$ such that $\Phi_t^{X^{\mathrm{lin}}}(e)$ is defined, but $\Phi_t^{X^{\mathrm{lin}}}(e) \notin \mathsf{F}$. We can assume for concreteness that $t \in \mathbb{R}_{>0}$. Let

$$t_* = \sup\{t \in \mathbb{R}_{\geq 0} \mid \ \Phi_t^{X^{\mathrm{lin}}}(e) \in \mathsf{F}\}.$$

Since F is closed by Lemma 2.17 and since $t \mapsto \Phi_t^{X^{\mathrm{lin}}}(e)$ is continuous, $\Phi_{t_*}^{X^{\mathrm{lin}}}(e) \in \mathsf{F}$. The argument above then gives $T \in \mathbb{R}_{>0}$ such that, for $t \in (-T, T)$,

$$\Phi_t^{X^{\mathrm{lin}}} \circ \Phi_{t_*}^{X^{\mathrm{lin}}}(e) \in \mathsf{F} \quad \Longrightarrow \quad \Phi_{t+t_*}^{X^{\mathrm{lin}}}(e) \in \mathsf{F},$$

in contradiction with the definition of t_*.

The preceding gives this part of the proposition when F is a C^ω-cogeneralised subbundle. Next suppose that F is a C^ω-generalised subbundle invariant under X^{lin}. Then, by definition, $\Lambda(\mathsf{F})$ is a C^ω-cogeneralised subbundle that is invariant under $X^{\mathrm{lin},*}$. By the first half of this part of the proof, $\Lambda(\mathsf{F})$ is flow-invariant under $X^{\mathrm{lin},*}$. By Lemma 4.7, it follows that F is flow-invariant under X^{lin}.

Finally, suppose that F is a smooth subbundle. Let $\mathcal{U} \subseteq \mathsf{M}$ be open and let $\xi_1, \dots, \xi_m \in \Gamma^\infty(\mathsf{E})$ be a basis of local sections for which $\xi_1, \dots, \xi_k$ are a local basis for F, this being possible since F is a subbundle. Let $\lambda^1, \dots, \lambda^m \in \Gamma^\infty(\mathsf{E}^*)$ be the dual basis, i.e.,

$$\langle \lambda^a(x); \xi_b(x) \rangle = \begin{cases} 1, & a = b, \\ 0, & a \neq b, \end{cases} \qquad x \in \mathcal{U}.$$

Note that $\xi_1, \dots, \xi_k$ generate $\mathscr{G}^\infty_{\mathsf{F}}(\mathcal{U})$ and that $\lambda^{k+1}, \dots, \lambda^m$ generate $\mathscr{G}^\infty_{\Lambda(\mathsf{F})}(\mathcal{U})$. More germanely,

$$\mathsf{F}|\mathcal{U} = \cap_{a=k+1}^m ((\lambda^a)^{\mathrm{e}})^{-1}(0)$$

and $\mathrm{d}(\lambda^{k+1})^{\mathrm{e}}(e), \dots, \mathrm{d}(\lambda^m)^{\mathrm{e}}(e)$ are linearly independent for $e \in \mathsf{E}|\mathcal{U}$. Now we have, by (4.1),

$$\mathscr{L}_{X^{\mathrm{lin}}}(\lambda^a)^{\mathrm{e}} = f^a_{k+1}(\lambda^{k+1})^{\mathrm{e}} + \dots + f^a_m(\lambda^m)^{\mathrm{e}}, \qquad a \in \{k+1, \dots, m\},$$

for $f^a_{k+1}, \dots, f^a_m \in \mathrm{C}^\infty(\mathcal{U})$. Therefore,

$$\langle \mathrm{d}(\lambda^a)^{\mathrm{e}}(e); X^{\mathrm{lin}}(e) \rangle = 0, \qquad e \in \mathsf{F}|\mathcal{U},\ a \in \{k+1, \dots, m\}.$$

We conclude that X^{lin} is tangent to F and so F is flow-invariant. ■

In the next section we shall give conditions for invariance of generalised and cogeneralised subbundles under linear vector fields as a special case of such conditions for affine vector fields.

4.3. Generalised and cogeneralised subbundles invariant under affine vector fields. In this section we extend the analysis of the preceding section to generalised and cogeneralised subbundles invariant under affine vector fields. Thus the situation we consider is as follows. Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $X^{\mathrm{aff}} \in \Gamma^r(\mathsf{TE})$ be an affine vector field over X_0. If we suppose that E is equipped with a C^r-linear connection, then we can write

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$$

for $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$ and $b \in \Gamma^r(\mathsf{E})$, as in Lemma 2.2. Our conditions in this section then extend those from the previous section where $b = 0$.

We first have the following result concerning invariance of cogeneralised subbundles under affine vector fields.

4.9 Proposition: (Cogeneralised subbundles invariant under an affine vector field) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and let $b \in \Gamma^r(\mathsf{E})$. Consider the following statements:*

(i) F is flow-invariant under the affine vector field $X^{\mathrm{aff}} \triangleq X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$;

(ii) the following conditions hold:

(a) $b(x) \in \mathsf{F}_x$ for $x \in \mathsf{M}$;

(b) $A(\mathsf{F}_x) \subseteq \mathsf{F}_x$ for $x \in \mathsf{M}$;

(c) $\nabla_{X_0}(\mathscr{G}^r_{\Lambda(\mathsf{F})}) \subseteq \mathscr{G}^r_{\Lambda(\mathsf{F})}$.

Then (i) $\Longrightarrow$ (ii) and, if either (1) $r=\omega$ or (2) $r=\infty$ and F is a subbundle, then (ii) $\Longrightarrow$ (i).

Proof: (i) $\Longrightarrow$ (ii) We shall prove each of the three conditions in sequence.

(ii)(a) Let $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$. Then, $\lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{aff}}}(e) = 0$ for all $(t,e) \in \mathbb{R} \times \mathsf{F}$ for which the expression is defined. Therefore, by [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.10],

$$\mathscr{L}_{X^{\mathrm{aff}}}\lambda^{\mathrm{e}}(e) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{aff}}}(e) = 0,$$

and so $\mathscr{L}_{X^{\mathrm{aff}}}\lambda^{\mathrm{e}}(e) = 0$ for $e \in \mathsf{F}$. By Lemma 2.10, we have

$$\mathscr{L}_{X^{\mathrm{aff}}}\lambda^{\mathrm{e}} = (\nabla_{X_0}\lambda)^{\mathrm{e}} + (A^*\lambda)^{\mathrm{e}} + \langle \lambda; b\rangle^{\mathrm{h}}$$

for $\lambda \in \Gamma^r(\mathsf{E}^*)$. Therefore, since, for $x \in \mathsf{M}$, $0_x \in \mathsf{F}_x$, we have

$$0 = (\nabla_{X_0}\lambda)^{\mathrm{e}}(0_x) + (A^*\lambda)^{\mathrm{e}}(0_x) + \langle \lambda; b\rangle^{\mathrm{h}}(0_x) = \langle \lambda(x); b(x)\rangle.$$

Since

$$\{\lambda(x) \mid\ \lambda \in \Gamma^r(\Lambda(\mathsf{F}))\} = \Lambda(\mathsf{F})_x$$

by Corollary 2.18, we conclude that $b(x) \in \mathsf{F}_x$ for every $x \in \mathsf{M}$.

(ii)(b) We assume that $b(x) \in \mathsf{F}_x$ for every $x \in \mathsf{M}$ since we have just shown that this follows from our current hypotheses. Let $\mathcal{U} \subseteq \mathsf{M}$ be open and let $\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})$. Then, by Lemma 2.10(v), we have

$$\mathscr{L}_{b^{\mathrm{v}}}\lambda^{\mathrm{e}} = \langle \lambda; b|\mathcal{U}\rangle^{\mathrm{h}},$$

and so, for any $\mathcal{V} \subseteq \mathsf{E}$ open, we have

$$\mathscr{L}_{b^{\mathrm{v}}}(\{\lambda^{\mathrm{e}}|\mathcal{V} \mid\ \lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\pi(\mathcal{V}))\}) \subseteq \mathscr{I}_{\mathsf{F}}(\mathcal{V}).$$

Thus we have

$$\mathscr{L}_{b^{\mathrm{v}}}((\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}) \subseteq \mathscr{I}_{\mathsf{F}}. \tag{4.2}$$

We now employ a lemma.

1 Lemma: *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{E}' \subseteq \mathsf{E}$ be a C^r-subbundle, let ∇ be a C^r-linear connection in E, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. If E' is invariant under $X_0^{\mathrm{h}} + A^{\mathrm{e}}$, then $A(\mathsf{E}'_x) \subseteq \mathsf{E}'_x$ for every $x \in \mathsf{M}$.*

Proof: Since E' is a submanifold, E' is invariant under $X_0^{\mathrm{h}} + A^{\mathrm{e}}$ if and only if it is flow-invariant under $X_0^{\mathrm{h}} + A^{\mathrm{e}}$ if and only if

$$X_0^{\mathrm{h}}(e') + A^{\mathrm{e}}(e') \in \mathsf{TE}', \qquad e' \in \mathsf{E}'.$$

Thus we have

$$X_0^{\mathrm{h}}(e') + A^{\mathrm{e}}(e') \in T_{e'}\mathsf{E}' \quad\Longrightarrow\quad \mathrm{ver}(X_0^{\mathrm{h}}(e') + A^{\mathrm{e}}(e')) \in \mathrm{ver}(T_{e'}\mathsf{E}') \quad\Longrightarrow\quad A^{\mathrm{e}}(e') \in \mathsf{V}_{e'}\mathsf{E}'.$$

Since $\mathsf{V}_{e'}\mathsf{E}' \simeq \mathsf{E}'_{\pi(e')}$, this gives $A^{\mathrm{e}}(\mathsf{E}'_x) \subseteq \mathsf{E}'_x$ for every $x \in \mathsf{M}$. The result follows by definition of A^{e}. ▼

Now, for $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, we have

$$\mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}+b^{\mathrm{v}}}\lambda^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\mathsf{M}), \quad \mathscr{L}_{b^{\mathrm{v}}}\lambda^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\mathsf{M}),$$

the first by hypothesis and by Proposition 4.3(i), and the second by (4.2). Therefore,

$$\mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}}\lambda^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\mathsf{M}) \cap (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}(\mathsf{M}) = (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}(\mathsf{M}). \tag{4.3}$$

Now let $\mathcal{U} \subseteq \mathsf{M}$ be an open dense set for which $\mathsf{F}|\mathcal{U}$ is a subbundle, as in Lemma 2.20. Then, by the preceding lemma and (4.3), $\mathsf{F}|\mathcal{U}$ is invariant under $A|\mathcal{U}$. Now let $x \in \mathsf{M}$ and let $e \in \mathsf{F}_x$. Let $(e_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{F}|\mathcal{U}$ converging to e. Then $A(e_j) \in \mathsf{F}|\mathcal{U}$. By Lemma 2.17 and continuity of A we have

$$A(e) = \lim_{j\to\infty} A(e_j) \in \mathsf{F}_x.$$

(ii)(c) Now we can assume that $b(x) \in \mathsf{F}_x$ and $A(\mathsf{F}_x) \subseteq \mathsf{F}_x$ for every $x \in \mathsf{M}$. As in (4.2), $\mathscr{L}_{b^{\mathrm{v}}}((\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}) \subseteq \mathscr{I}_{\mathsf{F}}$. Let $\mathcal{U} \subseteq \mathsf{M}$ be open and let $\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})$. Then, for $e \in \pi^{-1}(\mathcal{U})$, we have

$$\mathscr{L}_{A^{\mathrm{e}}}\lambda^{\mathrm{e}}(e) = (A^*\lambda)^{\mathrm{e}}(e) = \langle A^*\lambda(\pi(e)); e\rangle = \langle \lambda(\pi(e)); A(e)\rangle.$$

Therefore, for $\mathcal{V} \subseteq \mathsf{E}$ open,

$$\mathscr{L}_{A^{\mathrm{e}}}(\{\lambda^{\mathrm{e}}|\mathcal{V} \mid \lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\pi(\mathcal{V}))\}) \subseteq \mathscr{I}_{\mathsf{F}}(\mathcal{V}).$$

Thus we have

$$\mathscr{L}_{A^{\mathrm{e}}}((\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}) \subseteq \mathscr{I}_{\mathsf{F}} \cap (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}} = (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}. \tag{4.4}$$

Now, for $\mathcal{U} \subseteq \mathsf{M}$ open and for $\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})$, we have $\lambda^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\pi^{-1}(\mathcal{U}))$. Therefore, with the current hypotheses, we have

$$\mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}+b^{\mathrm{v}}}\lambda^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\pi^{-1}(\mathcal{U})), \;\; \mathscr{L}_{A^{\mathrm{e}}}\lambda^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\pi^{-1}(\mathcal{U})), \;\; \mathscr{L}_{b^{\mathrm{v}}}\lambda^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\pi^{-1}(\mathcal{U})),$$

the first by hypothesis and by Proposition 4.3(i), the second by (4.4), and the third by (4.2). Therefore, by Lemma 2.10, we have

$$\mathscr{L}_{X_0^{\mathrm{h}}}\lambda^{\mathrm{e}} = (\nabla_{X_0}\lambda)^{\mathrm{e}} \in \mathscr{I}_{\mathsf{F}}(\pi^{-1}(\mathcal{U})) \cap \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})^{\mathrm{e}}.$$

Therefore, $\nabla_{X_0}\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\mathcal{U})$, as desired.

(ii) $\Longrightarrow$ (i) Let $\mathcal{V} \subseteq \mathsf{E}$ be open. Let $\lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\pi(\mathcal{V}))$. Using Lemma 2.10, we have

$$\mathscr{L}_{b^{\mathrm{v}}}\lambda^{\mathrm{e}} = \langle \lambda; b\rangle^{\mathrm{h}} = 0.$$

We point out that this Lie derivative not only vanishes on F, but it is everywhere zero. Therefore,

$$\mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}+b^{\mathrm{v}}}((\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}) = \mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}}((\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}),$$

simply by Lemma 2.10 and since

$$(\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}} \subseteq (\mathscr{G}^r_{\mathsf{E}^*})^{\mathrm{e}}.$$

Now suppose that $\mathsf{F} \cap \mathcal{V} = \varnothing$. In this case we immediately have

$$\mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}+b^{\mathrm{v}}}((\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}) \subseteq (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}.$$

Next suppose that $\mathsf{F} \cap \mathcal{U} \neq \varnothing$. In a similar manner,

$$\mathscr{L}_{A^{\mathrm{e}}}\lambda^{\mathrm{e}}(e) = \langle \lambda(\pi(e)); A(e)\rangle = 0, \qquad e \in \mathsf{F} \cap \mathcal{V}.$$

Finally,

$$\mathscr{L}_{X_0^{\mathrm{h}}}\lambda^{\mathrm{e}}(e) = (\nabla_{X_0}\lambda)^{\mathrm{e}}(e) = \langle \nabla_{X_0}\lambda(\pi(e')); e\rangle = 0, \qquad e \in \mathsf{F} \cap \mathcal{V}.$$

Thus

$$\mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}+b^{\mathrm{v}}}(\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}} = \mathscr{L}_{X_0^{\mathrm{h}}+A^{\mathrm{e}}}(\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}} \subseteq (\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}.$$

Using this fact, the proof of this part of the proposition can be carried out just as are the corresponding parts of Proposition 4.8. ■

The analogous result for generalised subbundles is the following.

4.10 Proposition: (Generalised subbundles invariant under an affine vector field) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-generalised subbundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and let $b \in \Gamma^r(\mathsf{E})$. Consider the following statements:*

(i) F is flow-invariant under the affine vector field $X^{\mathrm{aff}} \triangleq X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$;

(ii) the following conditions hold:

(a) $b(x) \in \mathsf{F}_x$ for $x \in \mathsf{M}$;

(b) $A(\mathsf{F}_x) \subseteq \mathsf{F}_x$ for $x \in \mathsf{M}$;

(c) $\nabla_{X_0}(\mathscr{G}^r_{\mathsf{F}}) \subseteq \mathscr{G}^r_{\mathsf{F}}$.

Then (i) $\Longrightarrow$ (ii) and, if either (1) $r = \omega$ or (2) $r = \infty$ and F is a subbundle, then (ii) $\Longrightarrow$ (i).

Proof: Let us denote $X^{\mathrm{lin}} = X_0^{\mathrm{h}} + A^{\mathrm{e}}$. For $\lambda \in \Gamma^r(\mathsf{E}^*)$, define

$$X_\lambda^{\mathrm{aff},*} = X_0^{\mathrm{h},*} - A^* + \lambda^{\mathrm{v}}.$$

By Lemma 2.8, $X_\lambda^{\mathrm{aff},*} = X^{\mathrm{lin},*} + \lambda^{\mathrm{v}}$.

Let us prove a lemma.

1 Lemma: *With the preceding notation, the following statements hold:*

(i) if F is flow-invariant under X^{aff}, then $\Lambda(\mathsf{F})$ is flow-invariant under $X_\lambda^{\mathrm{aff},*}$ for every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$;

(ii) if $\Lambda(\mathsf{F})$ is flow-invariant under $X^{\mathrm{lin},*}$ and if $b \in \Gamma^r(\mathsf{F})$, then F is flow-invariant under X^{aff}.

Proof: (i) Let $\alpha \in \Lambda(\mathsf{F})$ and let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0} \circ \pi^*(\alpha)$ is defined. Let $e \in \mathsf{F}_{\Phi_t^{X_0} \circ \pi^*(\alpha)}$. Let $\Theta \in \mathrm{Lin}^r(\mathsf{E}^*)$ be such that $\Theta \circ \pi(e) = e$. This is possible by Lemma 1.2. Then we use

Proposition 2.5 to compute, for $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$,

$$\begin{aligned}
\big\langle \Phi_t^{X_\lambda^{\mathrm{aff},*}}(\alpha); e\big\rangle &= \Theta \circ \Phi_t^{X_\lambda^{\mathrm{aff},*}}(\alpha) \\
&= \Theta \circ \Phi_t^{X^{\mathrm{lin},*}}(\alpha) + \int_0^t \Theta \circ \Phi_{t-\tau}^{X^{\mathrm{lin},*}}(\lambda \circ \Phi_\tau^{X_0}(\pi^*(\alpha)))\,\mathrm{d}\tau \\
&= \langle \Phi_t^{X^{\mathrm{lin},*}}(\alpha); e\rangle + \int_0^t \langle \Phi_{t-\tau}^{X^{\mathrm{lin},*}}(\lambda \circ \Phi_\tau^{X_0}(\pi^*(\alpha))); e\rangle\,\mathrm{d}\tau \\
&= \langle \alpha; \Phi_{-t}^{X^{\mathrm{lin}}}(e)\rangle + \int_0^t \langle \lambda \circ \Phi_\tau^{X_0}(\pi^*(\alpha)); \Phi_{\tau-t}^{X^{\mathrm{lin}}}(e)\rangle\,\mathrm{d}\tau = 0.
\end{aligned}$$

As this must hold for every $e \in \mathsf{F}_{\Phi_t^{X_0}(\pi^*(\alpha))}$, we conclude that $\Phi_t^{X_\lambda^{\mathrm{aff},*}}(\alpha) \in \Lambda(\mathsf{F})$.

(ii) Let $e \in \mathsf{F}$ and let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0} \circ \pi(e)$ is defined. Let $\alpha \in \Lambda(\mathsf{F})_{\Phi_t^{X_0}(\pi(e))}$. Let $F \in \mathrm{Lin}^r(\mathsf{E})$ be such that $F \circ \pi^*(\alpha) = \alpha$. Then compute, by Proposition 2.5,

$$\begin{aligned}
\langle \alpha; \Phi_t^{X^{\mathrm{aff}}}(e)\rangle &= F \circ \Phi_t^{X^{\mathrm{aff}}}(e) \\
&= F \circ \Phi_t^{X^{\mathrm{lin}}}(e) + \int_0^t F \circ \Phi_{t-\tau}^{X^{\mathrm{lin}}}(b \circ \Phi_\tau^{X_0}(\pi(e)))\,\mathrm{d}\tau \\
&= \langle \alpha; \Phi_t^{X^{\mathrm{lin}}}(e)\rangle + \int_0^t \langle \alpha; \Phi_{t-\tau}^{X^{\mathrm{lin}}}(b \circ \Phi_\tau^{X_0}(\pi(e)))\rangle\,\mathrm{d}\tau \\
&= \langle \Phi_{-t}^{X^{\mathrm{lin},*}}(\alpha); e\rangle + \int_0^t \langle \Phi_{\tau-t}^{X^{\mathrm{lin},*}}(\alpha); b \circ \Phi_\tau^{X_0}(\pi(e))\rangle\,\mathrm{d}\tau = 0.
\end{aligned}$$

As this must hold for every $\alpha \in \Lambda(\mathsf{F})_{\Phi_t^{X_0}(\pi(e))}$, we conclude that $\Phi_t^{X^{\mathrm{aff}}}(e) \in \mathsf{F}$. ▼

(i) $\Longrightarrow$ (ii) Since F is a C^r-generalised subbundle, $\Lambda(\mathsf{F})$ is a cogeneralised subbundle.

Since F is flow-invariant under X^{aff}, part (i) of the lemma gives that $\Lambda(\mathsf{F})$ is flow-invariant under $X_\lambda^{\mathrm{aff},*}$. Since $\Lambda(\mathsf{F})$ is a C^r-cogeneralised subbundle, Proposition 4.9 implies that

1. $\lambda(x) \in \Lambda(\mathsf{F})_x$ for every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$ and $x \in \mathsf{M}$ (this is redundant and so harmless),
2. $A^*(\Lambda(\mathsf{F})_x) \subseteq \Lambda(\mathsf{F})_x$ for $x \in \mathsf{M}$ (which implies that $A(\mathsf{F}_x) \subseteq \mathsf{F}_x$ for every $x \in \mathsf{M}$), and
3. $\nabla_{X_0}(\mathscr{G}_{\mathsf{F}}^r) \subseteq \mathscr{G}_{\mathsf{F}}^r$.

Since F in flow-invariant under X^{aff} and $\Lambda(\mathsf{F})$ is flow-invariant under $X_\lambda^{\mathrm{aff},*}$ for every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, we have

$$f_{\mathsf{E}} \circ \Phi_t^{X^{\mathrm{aff}} \oplus X_\lambda^{\mathrm{aff},*}}(e, \alpha) = \langle \Phi_t^{X_\lambda^{\mathrm{aff},*}}(\alpha); \Phi_t^{X^{\mathrm{aff}}}(e)\rangle = 0$$

for every $(e, \alpha) \in \mathsf{F} \oplus \Lambda(\mathsf{F})$ and every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$. Therefore, by [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.10],

$$\mathscr{L}_{X^{\mathrm{aff}} \oplus X_\lambda^{\mathrm{aff},*}} f_{\mathsf{E}}(e, \alpha) = 0$$

for every $(e, \alpha) \in \mathsf{F} \oplus \Lambda(\mathsf{F})$ and every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$. Thus

$$0 = \mathscr{L}_{X^{\mathrm{lin}} \oplus X^{\mathrm{lin},*}} f_{\mathsf{E}}(e, \alpha) + \mathscr{L}_{b^{\mathrm{v}} \oplus \lambda^{\mathrm{v}}} f_{\mathsf{E}}(e, \alpha) = \mathscr{L}_{b^{\mathrm{v}} \oplus \lambda^{\mathrm{v}}} f_{\mathsf{E}}(e, \alpha)$$

for every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, by Lemma 2.7(iii). By the Leibniz Rule, we have

$$\mathscr{L}_{b^{\mathrm{v}} \oplus \lambda^{\mathrm{v}}} f_{\mathsf{E}}(e, \alpha) = \langle \alpha; b \circ \pi(e) \rangle + \langle \lambda \circ \pi^*(\alpha); e \rangle$$

for every $(e, \alpha) \in \mathsf{E} \oplus \mathsf{E}^*$. Taking $(e, \alpha) \in \mathsf{F} \oplus \Lambda(\mathsf{F})$, this gives

$$\langle \alpha; b \circ \pi(e) \rangle = 0, \qquad \alpha \in \Lambda(\mathsf{F})_{\pi^*(\alpha)},$$

and so $b \in \Gamma^r(\mathsf{F})$. Combining this with our conclusions 2 and 3 above, we get this part of the proposition.

(ii) $\implies$ (i) Our current hypotheses give the following:

1. $\lambda(x) \in \Lambda(\mathsf{F})_x$ for every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$ and $x \in \mathsf{M}$ (obviously);
2. $A^*(\Lambda(\mathsf{F})_x) \subseteq \Lambda(\mathsf{F})_x$ for $x \in \mathsf{M}$;
3. $\nabla_{X_0}(\mathscr{G}^r_{\mathsf{F}}) \subseteq \mathscr{G}^r_{\mathsf{F}}$.

Therefore, by Proposition 4.9, we conclude that the C^r-cogeneralised subbundle $\Lambda(\mathsf{F})$ is flow-invariant under $X^{\mathrm{aff},*}_{\lambda}$ for every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$. In particular, if we take $\lambda = 0$ and apply part (ii) of the lemma above, we conclude that F is flow-invariant under X^{aff}. ■

4.4. Generalised and cogeneralised affine subbundles invariant under affine vector fields. Our next collection of subbundle invariance results concerns the invariance of affine subbundles invariant under the flow of affine vector fields. As with the preceding section, we separately consider the cases of generalised and cogeneralised affine subbundles.

First we give an affine analogue of Lemma 4.4, recalling the notation of (2.5).

4.11 Lemma: (Vertical evaluations and the ideal sheaf of a cogeneralised affine subbundle) *Let $r \in \{\infty, \omega\}$, let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, and let $\mathsf{B} \subseteq \mathsf{E}$ be a C^r-cogeneralised affine subbundle given by $\mathsf{B} = \xi_0 + L(\mathsf{B})$. Define a sheaf morphism (of $\mathscr{C}^r_{\mathsf{M}}$-modules)*

$$\cdot^{\mathrm{e}} \colon \mathscr{G}^r_{\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}} \to \mathscr{C}^r_{\mathsf{E}}$$

by

$$(\mathscr{G}^r_{\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}}(\mathcal{U}))^{\mathrm{e}} = \{\lambda^{\mathrm{e}} \oplus f^{\mathrm{h}} \mid (\lambda, f) \in \mathscr{G}^r_{\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}}(\mathcal{U})\}$$

for $\mathcal{U} \subseteq \mathsf{M}$ open. Then $(\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}} \subseteq \mathscr{I}_{\mathsf{B}}$ and, moreover, for $e \in \mathsf{B}$, there is a neighbourhood $\mathcal{V} \subseteq \mathsf{E}$ of e such that

$$\mathsf{B} \cap \mathcal{V} = \{e' \in \mathcal{V} \mid (\lambda \oplus f)^{\mathrm{e}}(e') = 0,\ \lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\pi(\mathcal{V})),\ f = -\langle \lambda; \xi_0 \rangle\}.$$

Proof: Given Lemma 2.25, the only not completely obvious assertion is the final one, but this follows from Lemma 2.26. ■

Let us begin with a characterisation of an affine subbundle that is invariant under an affine vector field.

4.12 Lemma: (Characterisations of (co)generalised affine subbundles invariant under affine vector fields) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{B} \subseteq \mathsf{E}$ be a C^r-generalised or a C^r-cogeneralised affine subbundle given by $\mathsf{B} = \xi_0 + L(\mathsf{B})$ for $\xi_0 \in \Gamma^r(\mathsf{E})$, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $X^{\mathrm{aff}} \in \Gamma^r(\mathsf{TE})$ be an affine vector field over X_0. Write $X^{\mathrm{aff}} = X^{\mathrm{lin}} + b^{\mathrm{v}}$ for a linear vector field X^{lin} and for $b \in \Gamma^r(\mathsf{E})$. Then the following statements are equivalent:*

(i) B *is flow-invariant under* X^{aff}*;*

(ii) $L(\mathsf{B})$ *is flow-invariant under* X^{lin} *and* $\Phi_t^{X^{\mathrm{aff}}} \circ \xi_0(x) \in \mathsf{B}$ *for every* $(t,x) \in \mathbb{R} \times \mathsf{M}$ *for which* $\Phi_t^{X_0}(x)$ *is defined;*

Proof: We note that, by Proposition 2.5,

$$\Phi_t^{X^{\mathrm{aff}}}\colon \mathsf{E}_x \to \mathsf{E}_{\Phi_t^{X_0}(x)}$$

is an affine mapping with linear part equal to $\Phi_t^{X^{\mathrm{lin}}}$. Therefore, for $e \in \mathsf{E}$,

$$\Phi_t^{X^{\mathrm{lin}}}(e - \xi_0(\pi(e))) = \Phi_t^{X^{\mathrm{aff}}}(e) - \Phi_t^{X^{\mathrm{aff}}}(\xi_0(\pi(e))).$$

We shall use this formula in our proof.

(i) $\Longrightarrow$ (ii) Let $u \in L(\mathsf{B})$ and let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0}(\pi(u))$ is defined. We then have

$$\begin{aligned}\Phi_t^{X^{\mathrm{lin}}}(u) &= \Phi_t^{X^{\mathrm{lin}}}(u + \xi_0(\pi(u)) - \xi_0(\pi(u)))\\ &= \Phi_t^{X^{\mathrm{aff}}}(u + \xi_0(\pi(u))) - \Phi_t^{X^{\mathrm{aff}}}(\xi_0(\pi(u))) \in L(\mathsf{B})\end{aligned}$$

since B is flow-invariant under X^{aff} and $u + \xi_0(\pi(u)), \xi_0(\pi(u)) \in \mathsf{B}$. As a part of this argument, we have used the fact that $\Phi_t^{X^{\mathrm{aff}}} \circ \xi_0(x) \in \mathsf{B}$, just since B is flow-invariant under X^{aff}.

(ii) $\Longrightarrow$ (i) Let $e \in \mathsf{B}$ and let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0}(\pi(e))$ is defined. Then we have

$$\Phi_t^{X^{\mathrm{aff}}}(e) = \Phi_t^{X^{\mathrm{lin}}}(e - \xi_0(\pi(e))) + \Phi_t^{X^{\mathrm{aff}}}(\xi_0(\pi(e))) \in \mathsf{B}$$

since $e - \xi_0(\pi(e)) \in L(\mathsf{B})$ and $\Phi_t^{X^{\mathrm{aff}}}(\xi_0(e)) \in \mathsf{B}$. ■

First we consider the case of a cogeneralised affine subbundle.

4.13 Proposition: (Cogeneralised affine subbundles invariant under an affine vector field) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{B} \subseteq \mathsf{E}$ be a C^r-cogeneralised affine subbundle given by $\mathsf{B} = \xi_0 + L(\mathsf{B})$ for $\xi_0 \in \Gamma^r(\mathsf{E})$, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and let $b \in \Gamma^r(\mathsf{E})$. Consider the following statements:*

(i) B *is flow-invariant under the affine vector field* $X^{\mathrm{aff}} \triangleq X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$*;*

(ii) the following conditions hold:

(a) $A(L(\mathsf{B})_x) \subseteq L(\mathsf{B})_x$ *for* $x \in \mathsf{M}$*;*

(b) $\nabla_{X_0}(\mathscr{G}^r_{\Lambda(L(\mathsf{B}))}) \subseteq \mathscr{G}^r_{\Lambda(L(\mathsf{B}))}$*;*

(c) $(\nabla_{X_0}\xi_0 - A \circ \xi_0 - b)(x) \in L(\mathsf{B})_x$ *for every* $x \in \mathsf{M}$*.*

Then (i) $\implies$ (ii) and, if either (1) $r = \omega$ or (2) $r = \infty$ and $L(\mathsf{B})$ is a subbundle, then (ii) $\implies$ (i).

Proof: (i) $\implies$ (ii) Since B is flow-invariant under X^{aff}, $L(\mathsf{B})$ is flow-invariant under X^{lin}. Thus, by Proposition 4.9 (applied to linear vector fields), parts (ii)(a) and (ii)(b) hold.

Now let $\lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))$ and recall from (2.9) the notation F_λ. By Proposition 4.3 and Lemma 2.26, $\mathscr{L}_{X^{\mathrm{aff}}} F_\lambda \in \mathscr{I}_{\mathsf{B}}$. Using the decompositions

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}, \quad F_\lambda = \lambda^{\mathrm{e}} - \langle \lambda; \xi_0 \rangle^{\mathrm{h}}$$

and Lemma 2.10, we compute

$$\begin{aligned} \mathscr{L}_{X^{\mathrm{aff}}} F_\lambda &= (\nabla_{X_0}\lambda)^{\mathrm{e}} + (A^*\lambda)^{\mathrm{e}} + \langle \lambda; b\rangle^{\mathrm{h}} - (\mathscr{L}_{X_0}\langle \lambda; \xi_0\rangle)^{\mathrm{h}} \\ &= F_{\nabla_{X_0}\lambda} + (A^*\lambda)^{\mathrm{e}} + \langle \lambda; b\rangle^{\mathrm{h}} - \langle \lambda; \nabla_{X_0}\xi_0\rangle^{\mathrm{h}}. \end{aligned}$$

Let us write

$$(A^*\lambda)^{\mathrm{e}} = (A^*\lambda)^{\mathrm{e}} - \langle A^*\lambda; \xi_0\rangle^{\mathrm{h}} + \langle A^*\lambda; \xi_0\rangle^{\mathrm{h}} = F_{A^*\lambda} + \langle \lambda; A \circ \xi_0\rangle^{\mathrm{h}}.$$

By (ii)(b), $F_{\nabla_{X_0}\lambda} \in \mathscr{I}_{\mathsf{B}}$. By (ii)(a), $F_{A^*\lambda} \in \mathscr{I}_{\mathsf{B}}$. Therefore,

$$\mathscr{L}_{X^{\mathrm{aff}}} F_\lambda \in \mathscr{I}_{\mathsf{B}} \quad \implies \quad \langle \lambda; A \circ \xi_0 + b - \nabla_{X_0}\xi_0\rangle^{\mathrm{h}} \in \mathscr{I}_{\mathsf{B}}.$$

Therefore, for every $e \in \mathsf{B}$, we have

$$\langle \lambda; A \circ \xi_0 + b - \nabla_{X_0}\xi_0\rangle^{\mathrm{h}}(e) = 0.$$

Since horizontal lifts of functions from M to E are constant on fibres, we can conclude that $\langle \lambda; A \circ \xi_0 + b - \nabla_{X_0}\xi_0\rangle^{\mathrm{h}} = 0$. This, however, implies that

$$A \circ \xi_0(x) + b(x) - \nabla_{X_0}\xi_0(x) \in L(\mathsf{B})_x, \qquad x \in \mathsf{M},$$

by Corollary 2.18. Thus (ii)(c) holds as well.

(ii) $\implies$ (i) First let us consider the case when $r = \omega$. By our computations above and our given hypotheses, for $\lambda \in \Gamma^r(\Lambda(L(\mathsf{B})))$, we have

$$\mathscr{L}_{X^{\mathrm{aff}}} F_\lambda = F_{(\nabla_{X_0} + A^*)\lambda}. \tag{4.5}$$

Thus $\mathscr{L}_{X^{\mathrm{aff}}} F_\lambda \in \mathscr{I}_{\mathsf{B}}$ by our current hypotheses and by Lemma 2.26. By induction,

$$\underbrace{\mathscr{L}_{X^{\mathrm{aff}}} \cdots \mathscr{L}_{X^{\mathrm{aff}}} F_\lambda}_{k \text{ times}} = F_{(\nabla_{X_0} + A^*)^k\lambda} \in \mathscr{I}_{\mathsf{B}}.$$

By the argument from the proof of Proposition 4.3(ii), this gives

$$F_\lambda \circ \Phi_t^{X^{\mathrm{aff}}}(e) = 0$$

for every $(t, e) \in \mathbb{R} \times \mathsf{B}$ for which $\Phi_t^{X_0}(\pi(e))$ is defined. By Lemma 2.26 we conclude that $\Phi_t^{X^{\mathrm{aff}}}(e) \in \mathsf{B}$ for every $(t, e) \in \mathbb{R} \times \mathsf{B}$ for which $\Phi_t^{X_0}(\pi(e))$ is defined, i.e., B is flow-invariant under X^{aff}.

Now we consider the case when $r = \infty$ and $L(\mathsf{B})$ is a subbundle. Let $\mathcal{U} \subseteq \mathsf{M}$ be open and let $\xi_1, \dots, \xi_m \in \Gamma^\infty(\mathsf{E})$ be a basis of local sections for which $\xi_1, \dots, \xi_k$ are a local basis for $L(\mathsf{B})$, this being possible since $L(\mathsf{B})$ is a subbundle. Let $\lambda^1, \dots, \lambda^m \in \Gamma^\infty(\mathsf{E}^*)$ be the dual basis, i.e.,

$$\langle \lambda^a(x); \xi_b(x) \rangle = \begin{cases} 1, & a = b, \\ 0, & a \neq b, \end{cases} \qquad x \in \mathcal{U}.$$

Note that $\xi_1, \dots, \xi_k$ generate $\mathscr{G}^\infty_{L(\mathsf{B})}(\mathcal{U})$ and that $\lambda^{k+1}, \dots, \lambda^m$ generate $\mathscr{G}^\infty_{\Lambda(L(\mathsf{B}))}(\mathcal{U})$. More germanely,

$$L(\mathsf{B})|\mathcal{U} = \cap_{a=k+1}^m ((\lambda^a)^{\mathrm{e}})^{-1}(0)$$

and $\mathrm{d}(\lambda^{k+1})^{\mathrm{e}}(e), \dots, \mathrm{d}(\lambda^m)^{\mathrm{e}}(e)$ are linearly independent for $e \in \mathsf{E}|\mathcal{U}$. Now we have, under our current hypotheses

$$\nabla_{X_0} \lambda^a = \sum_{b=k+1}^m f^a_b \lambda^b, \quad A^* \lambda^a = \sum_{b=k+1}^m g^a_b \lambda^b, \qquad a \in \{k+1, \dots, m\},$$

for some $f^a_b, g^a_b \in \mathrm{C}^\infty(\mathcal{U})$, $a, b \in \{k+1, \dots, m\}$. Now, using (4.5), we have

$$\mathscr{L}_{X^{\mathrm{aff}}} F_{\lambda^a} = \sum_{b=k+1}^m (f^a_b + g^a_b) F_{\lambda^b}, \qquad a \in \{k+1, \dots, m\}.$$

Therefore, by Lemma 2.26, we conclude that X^{aff} is tangent to B and so B is flow-invariant. ■

For generalised affine subbundles, the invariance result is the following.

4.14 Proposition: (Generalised affine subbundles invariant under an affine vector field) *Let $r \in \{\infty, \omega\}$, let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{B} \subseteq \mathsf{E}$ be a C^r-generalised affine subbundle given by $\mathsf{B} = \xi_0 + L(\mathsf{B})$ for $\xi_0 \in \Gamma^r(\mathsf{E})$, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and let $b \in \Gamma^r(\mathsf{E})$. Consider the following statements:*

(i) B is flow-invariant under the affine vector field $X^{\mathrm{aff}} \triangleq X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$;

(ii) the following conditions hold:

(a) $A(L(\mathsf{B})_x) \subseteq L(\mathsf{B})_x$ for $x \in \mathsf{M}$;

(b) $\nabla_{X_0}(\mathscr{G}^r_{L(\mathsf{B})}) \subseteq \mathscr{G}^r_{L(\mathsf{B})}$;

(c) $(\nabla_{X_0}\xi_0 - A \circ \xi_0 - b)(x) \in L(\mathsf{B})_x$ for every $x \in \mathsf{M}$.

Then (i) $\implies$ (ii) and, if either (1) $r = \omega$ or (2) $r = \infty$ and $L(\mathsf{B})$ is a subbundle, then (ii) $\implies$ (i).

Proof: (i) $\implies$ (ii) From Lemma 4.12 we conclude that, under the current hypotheses, $L(\mathsf{B})$ is flow-invariant under X^{lin}. By Proposition 4.10, parts (ii)(a) and (ii)(b) hold, and we shall use the fact that these conditions hold in the rest of this part of the proof.

By Lemma 2.20, let $\mathcal{U} \subseteq \mathsf{M}$ be an open and dense subset such that $L(\mathsf{B})|\mathcal{U}$ is a subbundle. Let $\alpha \in \Lambda(L(\mathsf{B}))|\mathcal{U}$ and note that, by Proposition 4.13, we have

$$\langle \alpha; (\nabla_{X_0}\xi_0 - A \circ \xi_0 - b)(\pi^*(\alpha)) \rangle = 0.$$

If $\alpha \in \Lambda(L(\mathsf{B}))$, let $(\alpha_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\Lambda(L(\mathsf{B}))|\mathcal{U}$ converging to α, this being possible by Lemma 2.17. Then

$$\langle\alpha; (\nabla_{X_0}\xi_0 - A\circ\xi_0 - b)(\pi^*(\alpha))\rangle = \lim_{j\to\infty}\langle\alpha_j; (\nabla_{X_0}\xi_0 - A\circ\xi_0 - b)(\pi^*(\alpha_j))\rangle = 0.$$

Thus we also have (ii)(c).

(ii) $\Longrightarrow$ (i) By Proposition 4.10 we have that $L(\mathsf{B})$ is flow-invariant under X^{lin}.

Let $x \in \mathsf{M}$ and let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0}(x)$ is defined. Note that

$$\Phi_t^{X^{\mathrm{aff}}}\circ\xi_0(x) \in \mathsf{B}_{\Phi_t^{X_0}(x)} \iff \Phi_t^{X^{\mathrm{aff}}}\circ\xi_0(x) - \xi_0\circ\Phi_t^{X_0}(x) \in L(\mathsf{B})_{\Phi_t^{X_0}(x)}.$$

We compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\Big(\Phi_t^{X^{\mathrm{aff}}}\circ\xi_0(x) - \xi_0\circ\Phi_t^{X_0}(x)\Big) &= X_0^{\mathrm{h}}\circ\xi_0(x) + A^{\mathrm{e}}\circ\xi_0(x) + b^{\mathrm{v}}\circ\xi_0(x) - T_x\xi_0(X_0)\\ &= X_0^{\mathrm{h}}\circ\xi_0(x) + A^{\mathrm{e}}\circ\xi_0(x) + b^{\mathrm{v}}\circ\xi_0(x)\\ &\quad - \mathrm{hlft}(X_0(x),\xi_0(x)) - \mathrm{vlft}(\nabla_{X_0}\xi_0(x),\xi_0(x))\\ &= \mathrm{vlft}(A\circ\xi_0(x) + b(x) - \nabla_{X_0}\xi_0(x),\xi_0(x)),\end{aligned}$$

using (2.2). Thus

$$t \mapsto \Phi_t^{X^{\mathrm{aff}}}\circ\xi_0(x) - \xi_0\circ\Phi_t^{X_0}(x)$$

is a vertical curve and so is tangent to $L(\mathsf{B})$ if and only if it is tangent to the fibres of $L(\mathsf{B})$. Since the tangent vector to this vertical curve in the fibre E_x is

$$A\circ\xi_0(x) + b(x) - \nabla_{X_0}\xi_0(x) \in L(\mathsf{B})_x,$$

it follows that $\Phi_t^{X^{\mathrm{aff}}}\circ\xi_0(x) \in \mathsf{B}$ for every $(t,x) \in \mathbb{R}\times\mathsf{B}$ for which $\Phi_t^{X_0}(x)$ is defined. This part of the result now follows from Lemma 4.12. ∎

4.5. Affine subbundle varieties and defining subbundles invariant under affine vector fields. Now we turn to characterising invariance of affine subbundle varieties and their defining subbundles, as defined in Section 2.9. To do this, consistent with Lemmata 4.4 and 4.11, we first characterise the ideal sheaf we use.

4.15 Lemma: (The ideal sheaf of an affine subbundle variety) *Let $r \in \{\infty,\omega\}$, let $\pi\colon \mathsf{E}\to\mathsf{M}$ be a C^r-vector bundle, and let $\mathsf{A}\subseteq\mathsf{E}$ be a nonempty affine subbundle variety with defining subbundle $\Delta\subseteq\mathsf{E}^*\oplus\mathbb{R}_{\mathsf{M}}$. Then, for $e\in\mathsf{A}$, there is a neighbourhood $\mathcal{V}\subseteq\mathsf{E}$ of e such that*

$$\mathsf{A}\cap\mathcal{V} = \{e'\in\mathcal{V} \mid (\lambda,f)^{\mathrm{e}}(e') = 0, (\lambda,f)\in\mathscr{G}_\Delta^r(\pi(\mathcal{V}))\}.$$

Proof: Since Δ is a C^r-generalised subbundle, this follows from Corollary 2.18. ∎

Since the definition of an affine subbundle variety is made only relative to a defining subbundle, one expects that there should be a connection between the invariance properties of an affine subbundle variety and that of a defining subbundle. In order to talk about invariance properties of defining subbundles, we need to ascertain the relevant vector field on $\mathsf{E}^*\oplus\mathbb{R}_{\mathsf{M}}$ with respect to which we discuss invariance. Let us set this up.

We let $r \in \{\infty, \omega\}$, and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle with ∇ a C^r-linear connection in E. We consider an affine vector field

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}},$$

for $X_0 \in \Gamma^r(\mathsf{TM})$, $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and $b \in \Gamma^r(\mathsf{E})$. Define a connection $\widehat{\nabla}$ on $\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}$ by

$$\widehat{\nabla}_X(\xi, f) = (\nabla_X \xi, \mathscr{L}_X f), \qquad X \in \Gamma^r(\mathsf{TM}),\ (\xi, f) \in \Gamma^r(\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}).$$

This is the connection obtained as a direct sum of ∇ and the canonical flat connection on $\mathbb{R}_{\mathsf{M}}$. We then define a linear vector field on the vector bundle $\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}$ by

$$\widehat{X}^{\mathrm{aff}} = X_0^{\mathrm{h}} + \widehat{(A, b)}^{\mathrm{e}},$$

where the horizontal lift in the first term on the right is that associated with the connection $\widehat{\nabla}$ and, for the second term on the right, $\widehat{(A, b)}$ is the section of $\Gamma^r(\mathrm{End}(\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}))$ defined by

$$\widehat{(A, b)}(\xi, f) = (A \circ \xi + f b, 0), \qquad (\xi, f) \in \Gamma^r(\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}).$$

Note that

$$\widehat{(A, b)}^*(\lambda, g) = (A^* \circ \lambda, \langle \lambda; b \rangle), \qquad (\lambda, g) \in \Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}),$$

The next lemma explains this definition of the linear vector field $\widehat{X}^{\mathrm{aff}}$, noting that there is a correspondence between $\mathrm{Lin}^r(\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}})$ with $\mathrm{Aff}^r(\mathsf{E})$. Let us be explicit about this. Let $(\lambda, g) \in \Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}})$. This defines a linear function $\widehat{F}_{(\lambda, g)}$ on $\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}$ by

$$\widehat{F}_{(\lambda, g)}(e, a) = \langle \lambda(\pi(e)); e \rangle + a g(\pi(e))$$

and an affine function $F_{(\lambda, g)}$ on E in the usual way:

$$F_{(\lambda, g)}(e) = \langle \lambda(\pi(e)); e \rangle + g(\pi(e)).$$

Note that

$$F_{(\lambda, g)}(e) = \widehat{F}_{(\lambda, g)}(e, 1),$$

consistent with Lemma 2.28(iii).

With this in mind, we have the following lemma.

4.16 Lemma: (Linear vector fields on $\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}$ correspond to affine vector fields on E) *We let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle with ∇ a C^r-linear connection in E. If*

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}},$$

is an affine vector field on E for $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$ and $b \in \Gamma^r(\mathsf{E})$, then the linear vector field $\widehat{X}^{\mathrm{aff}}$ on $\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}$ satisfies

$$\mathscr{L}_{\widehat{X}^{\mathrm{aff}}} \widehat{F}_{(\lambda, g)}(e, 1) = \mathscr{L}_{X^{\mathrm{aff}}} F_{(\lambda, g)}(e), \qquad e \in \mathsf{E},\ (\lambda, g) \in \Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}).$$

Proof: Note that

$$F_{(\lambda,g)} = \lambda^{\mathrm{e}} + g^{\mathrm{h}}, \quad \widehat{F}_{(\lambda,g)} = (\lambda, g)^{\mathrm{e}}.$$

Using Lemma 2.10, we calculate

$$\mathscr{L}_{X^{\mathrm{aff}}} F_{(\lambda,g)} = (\nabla_{X_0}\lambda)^{\mathrm{e}} + (\mathscr{L}_{X_0} g)^{\mathrm{h}} + (A^*\lambda)^{\mathrm{e}} + \langle \lambda; b\rangle^{\mathrm{h}}$$

and

$$\begin{aligned}\mathscr{L}_{\widehat{X}^{\mathrm{aff}}} \widehat{F}_{(\lambda,g)} &= (\widehat{\nabla}_{X_0}(\lambda, g))^{\mathrm{e}} + (\widehat{(A,b)}^*(\lambda,g))^{\mathrm{e}} \\ &= (\nabla_{X_0}\lambda, \mathscr{L}_{X_0} g)^{\mathrm{e}} + (A^*\lambda, \langle\lambda; b\rangle)^{\mathrm{e}}.\end{aligned}$$

Thus the lemma holds by making the identification $\mathrm{Lin}^r(\mathsf{E}\oplus\mathbb{R}_{\mathsf{M}}) \simeq \mathrm{Aff}^r(\mathsf{E})$ indicated before the statement of the lemma. ■

Now, since there is not a unique correspondence between defining subbundles and their associated affine subbundle varieties (many defining subbundles might give rise to the same affine subbundle variety), the way we shall characterise the invariance properties of an affine subbundle variety is as according to the following definition.

4.17 Definition: (Affine subbundle varieties and defining subbundles invariant under affine vector fields) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, and let $X^{\mathrm{aff}} \in \Gamma^r(\mathsf{TE})$ be an affine vector field over X_0.

(i) A defining subbundle $\Delta \subseteq \mathsf{E}^*\oplus\mathbb{R}_{\mathsf{M}}$ is ***invariant*** under X^{aff} if $\mathscr{L}_{\widehat{X}^{\mathrm{aff},*}}((\mathscr{G}^r_\Delta)^{\mathrm{e}}) \subseteq (\mathscr{G}^r_\Delta)^{\mathrm{e}}$.

(ii) A nonempty C^r-affine subbundle variety $\mathsf{A} \subseteq \mathsf{E}$ is ***invariant*** under X^{aff} if there exists a defining subbundle Δ such that $\mathsf{A} = \mathsf{A}(\Delta)$ and such that Δ is invariant under X^{aff}.

(iii) A defining subbundle $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ is ***flow-invariant*** under X^{aff} if $\Phi_t^{\widehat{X}^{\mathrm{aff},*}}(\beta, a) \in \Delta$ for every $(t, (\beta, a)) \in \mathbb{R}\times\Delta$ for which $\Phi_t^{X_0}(\pi^*(\beta))$ is defined.

(iv) A nonempty C^r-affine subbundle variety $\mathsf{A} \subseteq \mathsf{E}$ is ***flow-invariant*** under X^{aff} if $\Phi_t^{X^{\mathrm{aff}}}(e) \in \mathsf{A}$ for every $(t, e) \in \mathbb{R}\times\mathsf{A}$ for which $\Phi_t^{X_0}(\pi(e))$ is defined. •

A remark similar to Remark 4.6 can be made here concerning the possible conflicting notions of invariance. While the issues raised by this are interesting, we sidestep them in our approach by virtue of Lemma 4.15.

Let us prove a basic result regarding the relationship of the flow-invariance of an affine subbundle variety versus that of a defining subbundle.

4.18 Lemma: (Correspondence between flow-invariance of $\boldsymbol{\Delta}$ and $\mathsf{A}(\boldsymbol{\Delta})$) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}$ be a C^r-defining subbundle, and let A be a C^r affine subbundle variety. Let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, let $b \in \Gamma^r(\mathsf{E})$, and consider the affine vector field*

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}.$$

Then the following statements hold:

(i) A is flow-invariant under X^{aff} if and only if

$$\{(e, 1) \mid e \in \mathsf{A}\}$$

is flow-invariant under $\widehat{X}^{\mathrm{aff}}$;

(ii) if Δ is flow-invariant under X^{aff}, then $\mathsf{A}(\Delta)$ is flow-invariant under X^{aff};

(iii) if A is flow-invariant under X^{aff}, then

(a) $\mathsf{S}(\mathsf{A})$ is flow-invariant and

(b) for any defining subbundle Δ for A, $\Delta \cap (\pi^* \times \mathrm{pr}_1)^{-1}(\mathsf{S}(\mathsf{A}))$ is flow-invariant under $\widehat{X}^{\mathrm{aff},*}$.

Proof: (i) The definition of $\widehat{X}^{\mathrm{aff}}$ shows that $t \mapsto (\Upsilon(t), \alpha(t))$ is an integral curve for $\widehat{X}^{\mathrm{aff}}$ if and only if

$$\begin{aligned}
&\Upsilon'(t) = X_0^{\mathrm{h}} \circ \Upsilon(t) + A^{\mathrm{e}} \circ \Upsilon(t) + \alpha(t) b^{\mathrm{v}} \circ \Upsilon(t), \\
&\alpha'(t) = 0.
\end{aligned}$$

Thus we see that $t \mapsto \Upsilon(t)$ is an integral curve for X^{aff} if and only if $t \mapsto (\Upsilon(t), 1)$ is an integral curve for $\widehat{X}^{\mathrm{aff}}$, giving this part of the result.

(ii) By part (i) it is sufficient to show that, under the given hypothesis, the set

$$\{(e,1) \in \mathsf{E} \times \mathbb{R} \mid e \in \mathsf{A}(\Delta)\}$$

is flow-invariant under $\widehat{X}^{\mathrm{aff}}$. By (2.13) we note that

$$(e,1) \in \Lambda(\Delta_x) \iff e \in \mathsf{A}(\Delta)_x \iff (e,1) \in \{(e',1) \in \mathsf{E}_x \oplus \mathbb{R} \mid e' \in \mathsf{A}(\Delta)_x\}. \tag{4.6}$$

Now let $e \in \mathsf{A}(\Delta)$, let $x = \pi(e)$, and let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0}(x)$ is defined. Let $(\lambda, a) \in \Delta_{\Phi_t^{X_0}(x)}$ and compute,

$$\langle (\lambda, a); \Phi_t^{\widehat{X}^{\mathrm{aff}}}(e,1)\rangle = \langle \Phi_{-t}^{\widehat{X}^{\mathrm{aff},*}}(\lambda, a); (e,1)\rangle = 0$$

since $\Phi_{-t}^{\widehat{X}^{\mathrm{aff},*}}(\lambda, a) \in \Delta_x$ and since

$$(e,1) \in \{(e',1) \in \mathsf{E}_x \times \mathbb{R} \mid e' \in \mathsf{A}(\Delta)_x\} = \{(e',1) \in \mathsf{E}_x \oplus \mathbb{R} \mid (e',1) \in \Lambda(\Delta_x)\},$$

using (4.6). Using the characterisation of the integral curves of $\widehat{X}^{\mathrm{aff}}$ from the proof of part (i), we have

$$\Phi_t^{\widehat{X}^{\mathrm{aff}}}(e,1) = (\Phi_t^{X^{\mathrm{aff}}}(e), 1).$$

Thus

$$\begin{aligned}
\Phi_t^{\widehat{X}^{\mathrm{aff}}}(e,1) &\in \{(e',1) \in \mathsf{E}_{\Phi_t^{X_0}(x)} \oplus \mathbb{R} \mid (e',1) \in \Lambda(\Delta_{\Phi_t^{X_0}(x)})\} \\
&= \{(e',1) \in \mathsf{E}_{\Phi_t^{X_0}(x)} \oplus \mathbb{R} \mid e' \in \mathsf{A}_{\Phi_t^{X_0}(x)}\}
\end{aligned}$$

by (4.6). This shows that, if Δ is flow-invariant under $\widehat{X}^{\mathrm{aff},*}$, then the set

$$\{(e,1) \in \mathsf{E} \oplus \mathbb{R} \mid e \in \mathsf{A}(\Delta)\}$$

is flow-invariant under $\widehat{X}^{\mathrm{aff}}$, as desired.

(iii)(a) Let $x \in \mathsf{S}(\mathsf{A})$ and let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0}(x)$ exists. Let $e \in \mathsf{A}_x$. Then $e' \triangleq \Phi_t^{X^{\mathrm{aff}}}(e) \in \mathsf{A}$. Therefore,

$$x' = \pi(e') = \Phi_t^{X_0}(e) \in \mathsf{S}(\mathsf{A}),$$

as desired.

(iii)(b) By part (i), the set

$$\{(e,1) \in \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}} \mid e \in \mathsf{A}_x\}$$

is flow-invariant under $\widehat{X}^{\mathrm{aff}}$. Let $x \in \mathsf{S}(\mathsf{A})$ and let $(\lambda, a) \in \Delta_x$. Let $t \in \mathbb{R}$ be such that $\Phi_t^{X_0}(x)$ is defined and let $e \in \mathsf{A}_{\Phi_t^{X_0}(x)}$. We compute

$$\langle \Phi_t^{\widehat{X}^{\mathrm{aff},*}}(\lambda, a); (e,1)\rangle = \langle (\lambda, a); \Phi_{-t}^{\widehat{X}^{\mathrm{aff}}}(e,1)\rangle = 0$$

since

$$\Phi_{-t}^{\widehat{X}^{\mathrm{aff}}}(e,1) \in \{(e',1) \in \mathsf{E}_x \times \mathbb{R} \mid (e',1) \in \mathsf{A}_x\} = \{(e',1) \in \mathsf{E}_x, \mathbb{R} \mid (e',1) \in \Lambda(\Delta_x)\},$$

using (4.6). Thus

$$\Phi_t^{\widehat{X}^{\mathrm{aff},*}}(\lambda, a) \in \Lambda(\{(e',1) \in \mathsf{E}_{\Phi_t^{X_0}(x)} \oplus \mathbb{R} \mid e' \in \mathsf{A}_{\Phi_t^{X_0}(x)}\}) = \Lambda(\Lambda(\Delta_x)) = \Delta_x,$$

again using (4.6). ■

Note that, since defining subbundles are generalised subbundles, their characterisation is made by reference to Proposition 4.10, specialising to the case of the proposition when the vector field is linear. To do this, it is convenient to define, for a defining subbundle $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ and for $x \in \mathsf{M}$, subspaces $\Delta_{1,x}$ and $\Delta_{0,x}$ of E^* and $\mathsf{E}_x^* \oplus \mathbb{R}$, respectively, as in (2.10) and (2.11). Note that Lemma 2.29(ii) associates $\Delta_{1,x}$ with the linear part of the affine subspace fibre $\mathsf{A}(\Delta)_x$ for $x \in \mathsf{S}(\mathsf{A}(\Delta))$.

We then arrive at the following result concerning flow-invariant defining subbundles.

4.19 Proposition: (Defining subbundles invariant under an affine vector field) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ be a C^r-defining subbundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and let $b \in \Gamma^r(\mathsf{E})$. Consider the affine vector field*

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$$

and the following statements:

(i) Δ is flow-invariant under X^{aff};

(ii) the following conditions hold:

(a) $\widehat{(A,b)}^*(\Delta_x) \subseteq \Delta_x$ for $x \in \mathsf{M}$;

(b) $\nabla_{X_0}(\mathscr{G}^r_{\Delta_1}) \subseteq \mathscr{G}^r_{\Delta_1}$.

Then (i) $\Longrightarrow$ (ii) and, if either (1) $r = \omega$ or (2) $r = \infty$ and Δ is a subbundle, then (ii) $\Longrightarrow$ (i).

Proof: In the proof, let us abbreviate subbundles of $\mathsf{E}^* \oplus \mathbb{R}$ by

$$\Lambda_0 = \{0\} \oplus \mathbb{R}_{\mathsf{M}}, \quad \Lambda_1 = \mathsf{V}^* \oplus \{0\}.$$

We shall make the identification $\Lambda_1 \simeq (\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}})/\Lambda_0$. Note that

$$\Delta_0 = \Delta \cap \Lambda_0, \quad \Delta_1 = \Delta \cap \Lambda_1.$$

(i) $\implies$ (ii) By definition, Δ is (flow-)invariant under the affine vector field X^{aff} if Δ is (flow-)invariant under the linear vector field $\widehat{X}^{\mathrm{aff},*}$. By Proposition 4.10 (specialised to linear vector fields), since Δ is flow-invariant under $\widehat{X}^{\mathrm{aff},*}$ we conclude that

1. $\widehat{(A,b)}^*(\Delta_x) \subseteq \Delta_x$ for $x \in \mathsf{M}$ and
2. $\widehat{\nabla}_{X_0}(\mathscr{G}^r_\Delta) \subseteq \mathscr{G}^r_\Delta$.

Condition 1 is exactly part (ii)(a).

Note that, for an open set $\mathcal{U} \subseteq \mathsf{M}$,

$$\widehat{\nabla}_{X_0}(\lambda, g) = (\nabla_{X_0}\lambda, \mathscr{L}_{X_0}g), \qquad (\lambda, g) \in \mathscr{G}^r_{\mathsf{E}^*\oplus\mathbb{R}}(\mathcal{U}).$$

Thus

$$\widehat{\nabla}_{X_0}|\mathscr{G}^r_{\Lambda_1} = \nabla_{X_0}, \quad \widehat{\nabla}_{X_0}|\mathscr{G}^r_{\Lambda_0} = \mathscr{L}_{X_0}.$$

We have

$$\Delta_{0,x} = \begin{cases} \Lambda_{0,x}, & x \notin \mathsf{S}(\mathsf{A}(\Delta)), \\ \{(0,0)\}, & x \in \mathsf{S}(\mathsf{A}(\Delta)). \end{cases}$$

By Lemma 2.29(i), we have $\mathscr{G}^r_{\Delta_0} \simeq \mathscr{I}_{\mathsf{S}(\mathsf{A}(\Delta))}$. By Lemma 4.18(ii), $\mathsf{A}(\Delta)$ is flow-invariant under X^{aff}. By Lemma 4.18(iii)(a), $\mathsf{S}(\mathsf{A}(\Delta))$ is flow-invariant under X_0. By Proposition 4.3, $\mathscr{G}^r_{\Delta_0}$ is invariant under $\widehat{\nabla}_{X_0}$. Thus $\widehat{\nabla}_{X_0}$ descends to a sheaf morphism on the quotient

$$\mathscr{G}^r_{\mathsf{E}^*\oplus\mathsf{M}}/\mathscr{G}^r_{\Lambda_0} \simeq \mathscr{G}^r_{\Lambda_1}.$$

Moreover, under this identification, the morphism on the quotient sheaf is ∇_{X_0}. Thus condition 2 above gives

$$\nabla_{X_0}(\mathscr{G}^r_{\Delta_1}) \subseteq \mathscr{G}^r_{\Delta_1},$$

noting that $\mathscr{G}^r_{\Delta_1}$ is the image of $\mathscr{G}^r_\Delta$ under the quotient. This gives part (ii)(b).

(ii) $\implies$ (i) Our observations from the preceding part of the proof show that Δ is invariant under X^{aff} if and only if the conditions of parts (ii)(a) and (ii)(b) hold. Thus this part of the proposition follows from the corresponding part of Proposition 4.10 (specialised to linear vector fields). ■

Note that the preceding result says nothing about whether the defining subbundle is total, partial, or null. The following result, all of whose conclusions follow from already proven results, summarises how one should approach the matter of determining the properties of a defining subbundle that is flow-invariant under X^{aff}. This should be regarded as providing a list of constraints that can be enforced to determine whether one can find a partial defining subbundle that is flow-invariant under X^{aff}.

4.20 Proposition: (Total, partial, and null defining subbundles invariant under an affine vector field) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ be a C^r-defining subbundle, let $X_0 \in \Gamma^r(\mathsf{TM})$, let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$, and let $b \in \Gamma^r(\mathsf{E})$. Consider the affine vector field*

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$$

and assume that Δ is flow-invariant under X^{aff}. Then

(i) *$A^*(\Delta_{1,x}) \subseteq \Delta_{1,x}$ for $x \in \mathsf{M}$.*

Also

(ii) *$\mathsf{S}(\mathsf{A}(\Delta)) = \{x \in \mathsf{M} \mid \Delta_{0,x} = \{(0,0)\}\}$ and so Δ is partial if and only if $\mathsf{S}(\mathsf{A}(\Delta)) \neq \varnothing$ and is total if and only if $\mathsf{S}(\mathsf{A}(\Delta)) = \mathsf{M}$.*

Moreover, if Δ is partial or total, then

(iii) *$b(x) \in \Lambda(\Delta_{1,x})$ for $x \in \mathsf{S}(\mathsf{A}(\Delta))$;*

(iv) *$\mathscr{L}_{X_0}(\mathscr{I}_{\mathsf{S}(\mathsf{A}(\Delta))}) \subseteq \mathscr{I}_{\mathsf{S}(\mathsf{A}(\Delta))}$.*

Proof: We adopt the notation from the proof of Proposition 4.19.

(i) Recall that

$$\widehat{(A,b)}^*(\lambda, a) = (A^*(\lambda), \langle \lambda; b\rangle), \qquad (\lambda, a) \in \Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}),$$

and so Λ_0 is invariant under $\widehat{(A,b)}^*$. Thus $\widehat{(A,b)}^*$ descends to an endomorphism on the quotient

$$(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}})/\Lambda_0 \simeq \Lambda_1.$$

Under the identification, this endomorphism is A^*. This gives this part of the result.

(ii) This follows from Lemma 2.29(i).

(iii) By Lemma 2.29(i), for $x \in \mathsf{S}(\mathsf{A}(\Delta))$ we have $\Delta_{0,x} = \{(0,0)\}$. In this case, for $x \in \mathsf{S}(\mathsf{A}(\Delta))$, we have

$$\widehat{(A,b)}^*(\Delta_x) \subseteq \Delta_x \quad \iff \quad \begin{array}{l} \langle \lambda; b\rangle = 0, \qquad \lambda \in \Delta_{1,x}, \\ A^*(\Delta_{1,x}) \subseteq \Delta_{1,x}. \end{array}$$

This gives this part of the result, in particular.

(iv) By Lemma 4.18(ii), $\mathsf{A}(\Delta)$ is flow-invariant under X^{aff}. By Lemma 4.18(iii)(a), $\mathsf{S}(\mathsf{A}(\Delta))$ is flow-invariant under X_0. By Proposition 4.3 this part of the result follows. ■

The preceding results enable us to identify flow-invariant defining subbundles. Having identified one of these, by Lemma 4.18 one automatically gets a flow-invariant affine subbundle variety, at least when the defining subbundle is not null. What the result does *not* do is answer the question of whether, given a flow-invariant affine subbundle variety, one can find a corresponding flow-invariant defining subbundle. Fortunately, we are not required to answer this question since, as part of our constructions of the next section, we will be naturally led first to a flow-invariant defining subbundle.

4.6. Invariant affine subbundle varieties contained in subbundles. We now investigate our discussion from the preceding sections from a different angle. The question in which we are interested is the following.

4.21 Question: (Integral curves of affine vector fields that leave invariant a cogeneralised subbundle) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$, let $b \in \Gamma^r(\mathsf{E})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}.$$

With this data, the basic question we consider is:

Are there integral curves of X^{aff} that leave F invariant?

The question has a few different components to it. First of all, it is an existential question. As well, assuming that the existential question has been answered in the affirmative, one can then ask about the character of *all* integral curves of X^{aff} that leave F invariant. •

We address the above question by first understanding the structure of *all* integral curves of an affine vector field that leave invariant a cogeneralised subbundle.

We start by considering the case of linear vector fields.

4.22 Theorem: (The largest invariant cogeneralised subbundle of a cogeneralised subbundle) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote*

$$X^{\mathrm{lin}} = X_0^{\mathrm{h}} + A^{\mathrm{e}}.$$

Let

$$\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}}) = \{e \in \mathsf{F} \mid \Phi_t^{X^{\mathrm{lin}}}(e) \in \mathsf{F} \text{ for all } t \in \mathbb{R} \text{ such that } \Phi_t^{X_0}(\pi(e)) \text{ is defined}\}.$$

Then the following statements hold:

(i) $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}}) \subseteq \mathsf{F}$*;*

(ii) $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ *is flow-invariant under* X^{lin}*;*

(iii) there exists a subsheaf $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})$ *of* $\mathscr{G}^r_{\Lambda(\mathsf{F})}$ *such that*

$$\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})_x = \{e \in \mathsf{E}_x \mid F(e) = 0,\ F = \lambda^{\mathrm{e}},\ [\lambda]_x \in \mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})_x\};$$

(iv) $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ *is a C^r-cogeneralised subbundle if*

(a) X_0 *is complete or*

(b) $r = \infty$*;*

(v) if $\mathsf{L} \subseteq \mathsf{F}$ *is a C^r-cogeneralised subbundle that is flow-invariant under X^{lin}, then* $\mathsf{L} \subseteq \mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$.

Proof: Just as we shall show in the proof of the more general Theorem 4.23 below, we have

1. $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}}) \subseteq \mathsf{F}$;
2. $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ is flow-invariant under X^{lin};
3. if $\mathsf{L} \subseteq \mathsf{F}$ is a C^r-cogeneralised subbundle that is flow-invariant under X^{lin}, then $\mathsf{L} \subseteq \mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$.

Let us deduce an alternative characterisation of $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$. Let

$$D(X_0) = \{(t,x) \in \mathbb{R} \times \mathsf{M} \mid \Phi_t^{X_0}(x) \text{ is defined}\}$$

and let $\Phi_{X_0}\colon D(X_0) \to \mathsf{M}$ be the flow, i.e., $\Phi_{X_0}(t,x) = \Phi_t^{X_0}(x)$. By Lemma 2.3(i),

$$D(X^{\mathrm{lin}}) \triangleq \{(t,e) \in \mathbb{R} \times \mathsf{E} \mid \Phi_t^{X^{\mathrm{lin}}}(e) \text{ is defined}\} = (\mathrm{id}_{\mathbb{R}} \times \pi)^{-1}(D(X_0)).$$

Denote by $\Phi_{X^{\mathrm{lin}}}\colon D(X^{\mathrm{lin}}) \to \mathsf{E}$ the flow, i.e., $\Phi_{X^{\mathrm{lin}}}(t,e) = \Phi_t^{X^{\mathrm{lin}}}(e)$. If $\lambda \in \Gamma^r(\mathsf{E}^*)$, then there is the associated section $\Phi_{X_0}^*\lambda$ of the pull-back bundle $\Phi_{X_0}^*\pi\colon \Phi_{X_0}^*\mathsf{E}^* \to D(X_0)$ defined by

$$\Phi_{X_0}^*\lambda(t,x) = (\lambda \circ \Phi_t^{X_0}(x), (t,x)).$$

Associated with this is the function $\Phi_{X^{\mathrm{lin}}}^*\lambda^{\mathrm{e}} \in \mathrm{Lin}^r(\Phi_{X_0}^*\mathsf{E})$ defined by

$$\Phi_{X^{\mathrm{lin}}}^*\lambda^{\mathrm{e}}(e,(t,x)) = \lambda^{\mathrm{e}} \circ \Phi_{X^{\mathrm{lin}}}(t,e) = \langle \lambda \circ \Phi_t^{X_0}(x); \Phi_t^{X^{\mathrm{lin}}}(e)\rangle.$$

We claim that

$$\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})_x = \{e \in \mathsf{E}_x \mid \Phi_{X^{\mathrm{lin}}}^*\lambda^{\mathrm{e}}(e,(t,x)) = 0,\ \lambda \in \Gamma^r(\Lambda(\mathsf{F})),\ (t,x) \in D(X_0)\}. \tag{4.7}$$

First let $e \in \mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})_x$, let $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, and let $t \in \mathbb{R}$ be such that $(t,x) \in D(X_0)$. Then, since $e \in \mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ and $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, it follows that

$$\Phi_{X^{\mathrm{lin}}}^*\lambda^{\mathrm{e}}(e,(t,x)) = \langle \lambda \circ \Phi_t^{X_0}(x); \Phi_t^{X^{\mathrm{lin}}}(e)\rangle = 0.$$

Now let $e \in \mathsf{E}_x$ be such that $\Phi_{X^{\mathrm{lin}}}^*\lambda^{\mathrm{e}}(e,(t,x)) = 0$ for all $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$ and t such that $(t,x) \in D(X_0)$. Then we have

$$\lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{lin}}}(e) = \langle \lambda \circ \Phi_t^{X_0}(x); \Phi_t^{X^{\mathrm{lin}}}(e)\rangle = \Phi_{X^{\mathrm{lin}}}^*\lambda^{\mathrm{e}}(e,(t,x)) = 0.$$

By Corollary 2.18 we conclude that $\Phi_t^{X^{\mathrm{lin}}}(e) \in \mathsf{F}$ and so we have verified our claim (4.7).

As a consequence of this,

$$\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})_x = \bigcap \{e \in \mathsf{E}_x \mid \langle \lambda \circ \Phi_t^{X_0}(x); \Phi_t^{X^{\mathrm{lin}}}(e)\rangle = 0,\ \lambda \in \Gamma^r(\Lambda(\mathsf{F})),\ (t,x) \in D(X_0)\}.$$

Since

$$\{e \in \mathsf{E}_x \mid \langle \lambda \circ \Phi_t^{X_0}(x); \Phi_t^{X^{\mathrm{lin}}}(e)\rangle = 0\}$$

is a subspace for every $\lambda \in \Gamma^r(\mathsf{E}^*)$ and $t \in \mathbb{R}$ such that $(t,x) \in D(X_0)$, we see that $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})_x$ is a subspace of E_x. We now turn our attention to the regularity of $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ by defining a subsheaf $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})$ with the property asserted in (iii).

Define a subsheaf $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})$ of $(\mathscr{G}^r_{\Lambda(\mathsf{F})})^{\mathrm{e}}$ by

$$\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})(\mathcal{U}) = \{(\Phi_t^{X^{\mathrm{lin}}})^*(\lambda^{\mathrm{e}}|\Phi_{-t}^{X_0}(\mathcal{U})) \mid \lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\Phi_{-t}^{X_0}(\mathcal{U})),\ (t,x) \in D(X_0),\ x \in \mathcal{U}\}. \tag{4.8}$$

Then, by virtue of (4.7), for $\mathcal{U} \subseteq \mathsf{M}$ open,

$$\pi^{-1}(\mathcal{U}) \cap \mathsf{L}(\mathsf{F}, X^{\mathrm{lin}}) = \{e \in \pi^{-1}(\mathcal{U}) \mid F(e) = 0,\ F \in \mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})(\mathcal{U})\}, \tag{4.9}$$

i.e., $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ is the zero set for the local sections of $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})$. This establishes (iii).

Next we prove part (iv), considering two cases.

First we consider the case of $r = \infty$. In this case, we have that, in the terminology of [Lewis 2012], $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})$ is patchy. Therefore, by virtue of Proposition 3.23 of [Lewis 2012] and the definition of patchy sheaves, we conclude that, for each $x \in \mathsf{M}$, there is a neighbourhood $\mathcal{U}$ of x and local sections $(\lambda_i)_{i \in I_x}$ from $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})(\mathcal{U})$ generating $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})(\mathcal{U})$ as a $\mathscr{C}^\infty_{\mathsf{M}}(\mathcal{U})$-module. This shows that $\Lambda(\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}}))$ is a C^∞-generalised subbundle, and so $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ is a C^∞-cogeneralised subbundle by (4.9).

Now we consider the case when $r = \omega$ and X_0 is complete. In this case, for $\mathcal{U} \subseteq \mathsf{M}$ open, the sections

$$\lambda^{\mathrm{e}} \circ \Phi_t^{X^{\mathrm{lin}}}(\Phi_{-t}^{X_0}(\mathcal{U})), \qquad \lambda \in \Gamma^\omega(\Lambda(\mathsf{F})),\ t \in \mathbb{R},$$

are generators for $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})(\mathcal{U})$ as a $\mathscr{C}^\omega_{\mathsf{M}}(\mathcal{U})$-module. Thus, for each $x \in \mathsf{M}$, there is a neighbourhood $\mathcal{U}$ of x such that $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})(\mathcal{U})$ is generated, as a $\mathscr{C}^\omega_{\mathsf{M}}(\mathcal{U})$-module, by some family of sections of E^* restricted to $\mathcal{U}$. Thus $\Lambda(\mathsf{F})$ is a C^ω-generalised subbundle, and so $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ is a C^ω-cogeneralised subbundle by (4.9). ∎

In the proof we constructed a subsheaf $\mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})$ of $\mathscr{G}^r_{\mathsf{E}^*}$ by (4.8). We define a subset $\Lambda(\mathsf{F}, X^{\mathrm{lin}})$ of E^* by

$$\Lambda(\mathsf{F}, X^{\mathrm{lin}})_x \triangleq \Lambda(\mathsf{F}, X^{\mathrm{lin}}) \cap \mathsf{E}^*_x = \{\lambda(x) \mid [\lambda]_x \in \mathscr{L}^*(\mathsf{F}, X^{\mathrm{lin}})_x\}.$$

Under the technical hypotheses of part (iv), $\Lambda(\mathsf{F}, X^{\mathrm{lin}})$ is a C^r-generalised subbundle and the associated C^r-cogeneralised subbundle is $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$. We should think (1) of $\Lambda(\mathsf{F}, X^{\mathrm{lin}})$ as being the smallest subbundle of E^* that annihilates F and is flow-invariant under $X^{\mathrm{lin},*}$, and (2) of $\mathsf{L}(\mathsf{F}, X^{\mathrm{lin}})$ as being the largest subbundle of E that is contained in F and is flow-invariant under $X^{\mathrm{lin},*}$.

Now we consider the general case of affine vector fields.

4.23 Theorem: (The largest invariant affine subbundle variety contained in a cogeneralised subbundle) *Let $r \in \{\infty, \omega\}$, let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$, let $b \in \Gamma^r(\mathsf{E})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote*

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}.$$

Let

$$\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}}) = \{e \in \mathsf{F} \mid \Phi_t^{X^{\mathrm{aff}}}(e) \in \mathsf{F} \text{ for all } t \in \mathbb{R} \text{ such that } \Phi_t^{X_0}(\pi(e)) \text{ is defined}\}$$

and let

$$\mathsf{S}(\mathsf{F}, X^{\mathrm{aff}}) = \{x \in \mathsf{M} \mid \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})_x \triangleq \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}}) \cap \mathsf{E}_x \neq \varnothing\}.$$

Then the following statements hold:

(i) $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}}) \subseteq \mathsf{F}$*;*

(ii) $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ *is flow-invariant under* X^{aff}*;*

(iii) $\mathsf{S}(\mathsf{F}, X^{\mathrm{aff}})$ *is flow-invariant under* X_0*;*

(iv) there exists a subsheaf $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})$ *of* $\mathscr{G}^r_{\mathsf{E}^*\oplus\mathbb{R}_\mathsf{M}}$ *such that, for* $x \in \mathsf{S}(\mathsf{F}, X^{\mathrm{aff}})$*,*

$$\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})_x = \{e \in \mathsf{E}_x \mid\ F(e) = 0,\ F = \lambda^{\mathrm{e}} + f^{\mathrm{h}},\ [(\lambda, f)]_x \in \mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})_x\};$$

(v) $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ *is a* C^r*-affine subbundle variety of* E *if*

(a) X_0 *is complete or*

(b) $r = \infty$*;*

(vi) if $\mathsf{B} \subseteq \mathsf{F}$ *is a* C^r*-affine subbundle variety that is flow-invariant under* X^{aff}*, then* $\mathsf{B} \subseteq \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$.

Proof: The conclusion (i) follows by definition of $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$.

For (ii), let $e \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ and let $t \in \mathbb{R}$ be such that $\Phi^{X_0}_t(\pi(e))$ exists. Then $e' \triangleq \Phi^{X^{\mathrm{aff}}}_t(e) \in \mathsf{F}$ by definition of $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$. To show that $e' \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$, let $t' \in \mathbb{R}$ be such that $\Phi^{X_0}_{t'}(\pi(e'))$ is defined. Then

$$\Phi^{X^{\mathrm{aff}}}_{t'}(e') = \Phi^{X^{\mathrm{aff}}}_{t+t'}(e) \in \mathsf{F}$$

since $e \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$, giving $e' \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$, as desired.

For (iii), let $x \in \mathsf{S}(\mathsf{F}, X^{\mathrm{aff}})$ and let $t \in \mathbb{R}$ be such that $\Phi^{X_0}_t(x)$ exists. Let $e \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}}) \cap \mathsf{E}_x$. Then we showed above that $e' \triangleq \Phi^{X^{\mathrm{aff}}}_t(e) \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$. Therefore,

$$x' = \pi(e') = \Phi^{X_0}_t(e) \in \mathsf{S}(\mathsf{F}, X^{\mathrm{aff}}),$$

as desired.

To show (vi), let B be as stated and let $e \in \mathsf{B}$. Then, because B is flow-invariant under X^{aff} and since $\mathsf{B} \subseteq \mathsf{F}$, $\Phi^{X^{\mathrm{aff}}}_t(e) \in \mathsf{F}$ for all $t \in \mathbb{R}$ such that $\Phi^{X_0}_t(\pi(e))$ is defined. That is to say, $e \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$.

It thus remains to prove (iv) and (v).

First let us devise an alternative characterisation of $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$. Let

$$D(X_0) = \{(t, x) \in \mathbb{R} \times \mathsf{M} \mid\ \Phi^{X_0}_t(x) \text{ is defined}\}$$

and let $\Phi_{X_0} \colon D(X_0) \to \mathsf{M}$ be the flow, i.e., $\Phi_{X_0}(t, x) = \Phi^{X_0}_t(x)$. Denote

$$D(X^{\mathrm{aff}}) = \{(t, e) \in \mathbb{R} \times \mathsf{E} \mid\ \Phi^{X^{\mathrm{aff}}}_t(e) \text{ is defined}\}$$

and we have $D(X^{\mathrm{aff}}) = (\mathrm{id}_{\mathbb{R}} \times \pi)^{-1}(D(X_0))$ by Lemma 2.3(i) and Proposition 2.5. We have the mapping

$$\begin{aligned}\Phi_{X^{\mathrm{aff}}} \colon D(X^{\mathrm{aff}}) &\to \mathsf{E}\\ (t, x) &\mapsto \Phi^{X^{\mathrm{aff}}}_t(e).\end{aligned}$$

For $\lambda \in \Gamma^r(\mathsf{E}^*)$ and $f \in \mathrm{C}^r(\mathsf{M})$, there is the affine function $\lambda^{\mathrm{e}} + f^{\mathrm{h}} \in \mathrm{Aff}^r(\mathsf{E})$ given by (2.5), and so the associated affine function $\Phi^*_{X^{\mathrm{aff}}}(\lambda^{\mathrm{e}} + f^{\mathrm{h}}) \in \mathrm{Aff}^r(\Phi^*_{X_0}\mathsf{E})$ defined by

$$\begin{aligned}\Phi^*_{X^{\mathrm{aff}}}(\lambda^{\mathrm{e}} + f^{\mathrm{h}})(e, (t, x)) &= (\lambda^{\mathrm{e}} + f^{\mathrm{h}}) \circ \Phi_{X^{\mathrm{aff}}}(t, e)\\ &= \langle \lambda \circ \Phi^{X_0}_t(x); \Phi^{X^{\mathrm{aff}}}_t(e)\rangle + f \circ \Phi^{X_0}_t(x).\end{aligned}$$

We claim that, for $x \in \mathsf{S}(\mathsf{F}, X^{\mathrm{aff}})$,

$$\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})_x = \{e \in \mathsf{E}_x \mid \Phi^*_{X^{\mathrm{aff}}}\lambda^{\mathrm{e}}(e,(t,x)) = 0,\ \lambda \in \Gamma^r(\Lambda(\mathsf{F})),\ (t,x) \in D(X_0)\}. \tag{4.10}$$

First let $e \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})_x$, let $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, and let $t \in \mathbb{R}$ be such that $(t,x) \in D(X_0)$. Then, since $e \in \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ and $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, it follows that

$$\Phi^*_{X^{\mathrm{aff}}}\lambda^{\mathrm{e}}(e,(t,x)) = \langle \lambda \circ \Phi^{X_0}_t(x); \Phi^{X^{\mathrm{aff}}}_t(e)\rangle = 0.$$

Now let $e \in \mathsf{E}_x$ be such that $\Phi^*_{X^{\mathrm{aff}}}\lambda^{\mathrm{e}}(e,(t,x)) = 0$ for all $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$ and t such that $(t,x) \in D(X_0)$. Then we have

$$\lambda^{\mathrm{e}} \circ \Phi^{X^{\mathrm{aff}}}_t(e) = \langle \lambda \circ \Phi^{X_0}_t(x); \Phi^{X^{\mathrm{aff}}}_t(e)\rangle = \Phi^*_{X^{\mathrm{aff}}}\lambda^{\mathrm{e}}(e,(t,x)) = 0.$$

By Corollary 2.18 we conclude that $\Phi^{X^{\mathrm{aff}}}_t(e) \in \mathsf{F}$ and so we have verified our claim (4.10).

Note that, because the flow of X^{aff} is not linear but affine, we have that $(\Phi^{X^{\mathrm{aff}}}_t)^*\lambda^{\mathrm{e}}$ is an affine function for $\lambda \in \Gamma^r(\mathsf{E}^*)$. Thus we can regard $(\Phi^{X^{\mathrm{aff}}}_t)^*\lambda^{\mathrm{e}}$ as a section of $\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ via (2.5). With this in mind, define a subsheaf $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})$ of $\mathscr{G}^r_{\mathsf{E}^*\oplus\mathbb{R}_{\mathsf{M}}}$ by

$$\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})(\mathcal{U}) = \{(\Phi^{X^{\mathrm{aff}}}_t)^*(\lambda^{\mathrm{e}}|\Phi^{X_0}_{-t}(\mathcal{U})) \mid \lambda \in \mathscr{G}^r_{\Lambda(\mathsf{F})}(\Phi^{X_0}_{-t}(\mathcal{U})),\ (t,x) \in D(X_0),\ x \in \mathcal{U}\}. \tag{4.11}$$

Then, by virtue of (4.10), for $\mathcal{U} \subseteq \mathsf{M}$ open,

$$\pi^{-1}(\mathcal{U}) \cap \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}}) = \{e \in \pi^{-1}(\mathcal{U}) \mid F(e) = 0,\ F \in \mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})(\mathcal{U})\}, \tag{4.12}$$

i.e., $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ is the zero set for the local sections of $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})$. This proves (iv).

We consider the two cases for part (v).

The first case we consider is that of $r = \infty$. In this case, we have that, in the terminology of [Lewis 2012], $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})$ is patchy. Therefore, by virtue of Proposition 3.23 of [Lewis 2012] and the definition of patchy sheaves, we conclude that, for each $x \in \mathsf{M}$, there is a neighbourhood $\mathcal{U}$ of x and local sections $((\lambda_i, f_i))_{i\in I_x}$ from $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})(\mathcal{U})$ generating $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})(\mathcal{U})$ as a $\mathscr{C}^\infty_{\mathsf{M}}(\mathcal{U})$-module. This shows that $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ has a defining bundle that is a C^∞-generalised subbundle, and so $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ is a C^∞-affine subbundle variety by (4.12).

Next we suppose that $r = \omega$ and that X_0 is complete. In this case, for $\mathcal{U} \subseteq \mathsf{M}$ open, the sections

$$\lambda^{\mathrm{e}} \circ \Phi^{X^{\mathrm{aff}}}_t(\Phi^{X_0}_{-t}(\mathcal{U})), \qquad \lambda \in \Gamma^\omega(\Lambda(\mathsf{F})),\ t \in \mathbb{R},$$

are generators for $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})(\mathcal{U})$ as a $\mathscr{C}^\omega_{\mathsf{M}}(\mathcal{U})$-module. Thus, for each $x \in \mathsf{M}$, there is a neighbourhood $\mathcal{U}$ of x such that $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})(\mathcal{U})$ is generated, as a $\mathscr{C}^\omega_{\mathsf{M}}(\mathcal{U})$-module, by some family of sections of $\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ restricted to $\mathcal{U}$. Thus $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ has a defining bundle that is a C^ω-generalised subbundle, and so $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ is a C^ω-affine subbundle variety by (4.12). ■

In the proof we constructed a subsheaf $\mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})$ of $\mathscr{G}^r_{\mathsf{E}^*\oplus\mathbb{R}_{\mathsf{M}}}$ by (4.11). We define a subset $\Delta(\mathsf{F}, X^{\mathrm{aff}})$ of $\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ by

$$\Delta(\mathsf{F}, X^{\mathrm{aff}})_x \triangleq \Delta(\mathsf{F}, X^{\mathrm{aff}}) \cap \mathsf{E}^*_x \oplus \mathbb{R} = \{(\lambda(x), g(x)) \mid [(\lambda, g)]_x \in \mathscr{A}^*(\mathsf{F}, X^{\mathrm{aff}})_x\}.$$

Under the technical hypotheses of part (v), $\Delta(\mathsf{F}, X^{\mathrm{aff}})$ is a C^r-defining subbundle and the associated C^r-affine subbundle variety is $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$. The definition of $\Delta(\mathsf{F}, X^{\mathrm{aff}})$ and

Lemma 4.16 ensure that $\Delta(\mathsf{F}, X^{\mathrm{aff}})$ is invariant under $\widehat{X}^{\mathrm{aff},*}$. We should thus think (1) of $\Delta(\mathsf{F}, X^{\mathrm{aff}})$ as being the smallest subbundle of $\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ whose associated affine functions annihilate F and that is flow-invariant under $\widehat{X}^{\mathrm{aff},*}$, and (2) of $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ as being the largest affine subbundle variety contained in F that is flow-invariant under X^{aff}.

Let us see how to use Theorem 4.23 to answer Question 4.21, leaving aside the technicalities of when certain sheaves are sheaves of sections of generalised subbundles. We do this in the general case of affine vector fields, with linear vector fields being an easier special case.

1. Determine the smallest generalised subbundle $\Delta(\mathsf{F}, X^{\mathrm{aff}})$ consisting of affine functions that vanish on F and which is invariant under the flow of $\widehat{X}^{\mathrm{aff},*}(\mathsf{F}, X^{\mathrm{aff}})$. Explicitly,
$$\Delta(\mathsf{F}, X^{\mathrm{aff}}) = \{\Phi_t^{\widehat{X}^{\mathrm{aff},*}}(\lambda, 0) \mid \lambda \in \Lambda(\mathsf{F})_x,\ (t, x) \in D(X_0)\}.$$
2. Define the corresponding affine subbundle variety $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$.
3. We then have
$$\{e \in \mathsf{E} \mid \Phi_t^{X^{\mathrm{aff}}} \in \mathsf{F},\ (t, \pi(e)) \in D(X_0)\} = \mathsf{A}(\mathsf{F}, X^{\mathrm{aff}}).$$
That is to say, $\mathsf{A}(\mathsf{F}, X^{\mathrm{aff}})$ consists of all initial conditions through which integral curves of X^{aff} remain in F.

Of course, the preceding "algorithm" is not practical, relying as it does on knowing the flow of the affine vector field X^{aff}. The following two results give the corresponding associated infinitesimal conditions.

We begin with the case of invariance under linear vector fields.

4.24 Theorem: (Cogeneralised subbundles invariant under a linear vector field and contained in a cogeneralised subbundle) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let L be a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote*
$$X^{\mathrm{lin}} = X_0^{\mathrm{h}} + A^{\mathrm{e}}$$
and suppose that L is flow-invariant under X^{lin}. Consider the following statements:

(i) $\mathsf{L} \subseteq \mathsf{F}$;

(ii) the following conditions hold:

  (a) $A(\mathsf{L}_x) \subseteq \mathsf{F}_x$ *for* $x \in \mathsf{M}$;

  (b) $\nabla_{X_0}(\mathscr{G}^r_{\Lambda(\mathsf{F})}) \subseteq \mathscr{G}^r_{\Lambda(\mathsf{L})}$.

Then (i) $\Longrightarrow$ (ii) and, if either (1) $r = \omega$ or (2) $r = \infty$ and F is a subbundle, then (ii) $\Longrightarrow$ (i).

Proof: (i) $\Longrightarrow$ (ii) Since L is flow-invariant under X^{lin}, by Proposition 4.9, we have
$$A(\mathsf{L}_x) \subseteq \mathsf{L}_x \subseteq \mathsf{F}_x, \qquad x \in \mathsf{M},$$
and
$$\mathscr{G}^r_{\Lambda(\mathsf{F})} \subseteq \mathscr{G}^r_{\Lambda(\mathsf{L})} \quad \Longrightarrow \quad \nabla_{X_0}(\mathscr{G}^r_{\Lambda(\mathsf{F})}) \subseteq \nabla_{X_0}(\mathscr{G}^r_{\Lambda(\mathsf{L})}) \subseteq \mathscr{G}^r_{\Lambda(\mathsf{L})},$$

as desired.

(ii) $\implies$ (i) For $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$ and $e \in \mathsf{L}$, by Lemma 2.10 we have

$$\mathscr{L}_{X^{\mathrm{lin}}}\lambda^{\mathrm{e}}(e) = \langle \nabla_{X_0}\lambda(\pi(e)); e\rangle + \langle \lambda(\pi(e)); A(e)\rangle = 0.$$

Since this holds for every $\lambda \in \Gamma^r(\Lambda(\mathsf{F}))$, from Proposition 4.9 we conclude that all integral curves of X^{lin} with initial conditions in L remain in F. Since L is flow-invariant under X^{lin}, we conclude that $\mathsf{L} \subseteq \mathsf{F}$. ■

Next we consider the case of affine vector fields. Here we wish to obtain conditions on a defining subbundle Δ that ensure that its corresponding affine subbundle variety $\mathsf{A}(\Delta)$ remains in a given subbundle F. However, because $\mathsf{A}(\Delta)$ may be empty, we would like instead to make the problem into one that always has a solution, and then leave the matter of checking whether $\mathsf{A}(\Delta)$ is nonempty to something one can do afterwards.

We first note that, by Lemma 2.28(iii),

$$\widehat{\Lambda(\Delta)} \triangleq \{(e,1) \in \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}} \mid e \in \mathsf{A}(\Delta)\} = \Lambda(\Delta) \cap (\mathsf{E} \times \{1\}).$$

By Lemma 4.18, if $\mathsf{A}(\Delta) \subseteq \mathsf{E}$ is flow-invariant under the affine vector field X^{aff}, then $\widehat{\Lambda(\Delta)}$ is flow-invariant under $\widehat{X}^{\mathrm{aff}}$. Clearly

$$\mathsf{A}(\Delta) \subseteq \mathsf{F} \iff \{(e,1) \in \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}} \mid e \in \mathsf{A}(\Delta)\} \subseteq \widehat{\mathsf{F}} \triangleq \{(e,1) \in \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}} \mid e \in \mathsf{F}\}.$$

Therefore, one seeks a defining bundle $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ that is flow-invariant under X^{aff} (meaning, by definition, that it is flow-invariant under $\widehat{X}^{\mathrm{aff},*}$) and satisfies

$$\widehat{\Lambda(\Delta)} \triangleq \Lambda(\Delta) \cap (\mathsf{E} \times \{1\}) \subseteq \widehat{\mathsf{F}}. \tag{4.13}$$

The following lemma examines the linear algebra of this condition, recalling from (2.10) the definition of $\Delta_1 \simeq \mathrm{pr}_1(\Delta)$.

4.25 Lemma: (Algebraic properties of invariant defining subbundles annihilating a cogeneralised subbundle) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let Δ be a C^r-defining subbundle for which $\widehat{\Lambda(\Delta)}_x \neq \varnothing$ for every $x \in \mathsf{M}$, let $X_0 \in \Gamma^r(\mathsf{M})$, let $b \in \Gamma^r(\mathsf{E})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote*

$$X^{\mathrm{lin}} = X_0^{\mathrm{h}} + A^{\mathrm{e}}, \quad X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$$

and suppose that Δ is flow-invariant under X^{aff}. The following statements hold:

(i) $\widehat{\mathsf{F}} = \{(e,a) \in \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}} \mid F_{(\lambda,g)}(e,a) = 0,\ \lambda \in \Gamma^r(\Lambda(\mathsf{F})),\ g \in \mathrm{C}^r(\mathsf{M})\}$, *where*

$$F_{(\lambda,g)} = (\lambda, g)^{\mathrm{e}} - \hat{\pi}^* g, \qquad \lambda \in \Gamma^r(\mathsf{E}^*),\ g \in \mathrm{C}^r(\mathsf{M}),$$

and where $\hat{\pi}\colon \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}} \to \mathsf{M}$ is the vector bundle projection (note that this definition of $F_{(\lambda,g)}$ differs from those for both $\widehat{F}_{(\lambda,g)}$ and $F_{(\lambda,g)}$ given preceding the statement of Lemma 4.16, but agrees with that in (2.9));

(ii) $L(\widehat{\Lambda(\Delta)}) = \Lambda(\Delta) \cap (\mathsf{E} \oplus 0) = \Lambda(\Delta_1) \oplus 0$;

(iii) $\Lambda(L(\widehat{\Lambda(\Delta)})) = \Delta + (0 \oplus \mathbb{R}_{\mathsf{M}}) = \Delta_1 \oplus \mathbb{R}_{\mathsf{M}}$*;*

(iv) the condition (4.13) holds if and only if (a) $\widehat{\Lambda(\Delta)} \cap \widehat{\mathsf{F}} \neq \varnothing$ *and (b)* $\Lambda(\Delta_1) \subseteq \mathsf{F}$*;*

(v) $\Lambda(\Delta_1)$ *is invariant under* X^{lin}*.*

Proof: (i) We note that

$$\widehat{\mathsf{F}} = (0,1) + \mathsf{F} \oplus 0 \subseteq \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}.$$

Let us verify that, as asserted, $F_{(\lambda,g)}$ agrees with the formula of (2.9) applied to our current setting. For $(\lambda, g) \in \Gamma^r(\mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}})$, the formula (2.9) gives

$$\begin{aligned} F_{(\lambda,g)}(e,a) &= \langle(\lambda(\pi(e)), g(\pi(e))); (e,a)\rangle - \langle(\lambda(\pi(e)), g(\pi(e))); (0,1)\rangle \\ &= \langle\lambda(\pi(e)); e\rangle + g(\pi(e))a - g(\pi(e)), \end{aligned}$$

whence $F_{(\lambda,g)} = (\lambda, g)^{\mathrm{e}} - \hat{\pi}^* g$, as asserted. This part of the result then follows from Lemma 2.26.

(ii) We have

$$\widehat{\Lambda(\Delta)} \subseteq \mathsf{E} \times \{1\}$$

and

$$L(\mathsf{E} \times \{1\}) = \mathsf{E} \oplus 0.$$

Therefore,

$$L(\widehat{\Lambda(\Delta)}) \subseteq \mathsf{E} \oplus 0.$$

Now suppose that $(e, 0) \in L(\widehat{\Lambda(\Delta)})$. Then

$$(e,0) + (e',1) = (e+e',1) \in \widehat{\Lambda(\Delta)}, \qquad (e',1) \in \widehat{\Lambda(\Delta)}.$$

Therefore, since $\widehat{\Lambda(\Delta)} \subseteq \Lambda(\Delta)$ and since $\Lambda(\Delta)_x$ is a subspace for each $x \in \mathsf{M}$,

$$(e,0) = (e+e',1) - (e',1) \in \Lambda(\Delta), \qquad (e',1) \in \widehat{\Lambda(\Delta)}.$$

Thus $L(\widehat{\Lambda(\Delta)}) \subseteq \Lambda(\Delta) \cap (\mathsf{E} \oplus 0)$. Conversely, suppose that $(e, 0) \in \Lambda(\Delta)$. Then, for every $(e', 1) \in \widehat{\Lambda(\Delta)}$,

$$(e',1) + (e,0) = (e'+e,1) \in \Lambda(\Delta),$$

again since $\widehat{\Lambda(\Delta)} \subseteq \Delta(\Lambda)$. This means that $(e', 1) + (e, 0) \in \widehat{\Lambda(\Delta)}$ for every $(e', 1) \in \widehat{\Lambda(\Delta)}$, whence $(e, 0) \in L(\widehat{\Lambda(\Delta)})$. Thus we have $\Lambda(\Delta) \cap (\mathsf{E} \oplus 0) \subseteq L(\widehat{\Lambda(\Delta)})$.

Now let us show that $\Lambda(\Delta) \cap (\mathsf{E} \oplus 0) = \Lambda(\Delta_1) \oplus 0$. To do so, let us denote by

$$\mathrm{pr}_1 \colon \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}} \to \mathsf{E}^*$$

the projection and by

$$i_1 \colon \mathsf{E} \to \mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}$$

the inclusion. Note that $\Delta_1 = \mathrm{pr}_1(\Delta)$ and that $i_1^* = \mathrm{pr}_1$.

Let $e \in \Lambda(\Delta_1)$. Then $(e, 0) \in \mathsf{E} \oplus 0$, obviously. Also, if $(\lambda, a) \in \Delta$, then

$$\langle(\lambda,a); (e,0)\rangle = \langle(\lambda,a); i_1(e)\rangle = \langle\mathrm{pr}_1(\lambda,a); e\rangle = 0,$$

and so $(e,0) \in \Lambda(\Delta)$. Thus $\Lambda(\Delta_1) \oplus 0 \subseteq \Lambda(\Delta) \cap (\mathsf{E} \oplus 0)$.

Next let $(e,0) \in \Lambda(\Delta) \cap (\mathsf{E} \oplus 0)$. Let $(\lambda, a) \in \Delta$ so that $\lambda = \mathrm{pr}_1(\lambda, a) \in \Delta_1$. Then

$$\langle \lambda; e\rangle = \langle (\lambda,0);(e,0)\rangle = \langle \mathrm{pr}_1(\lambda,a); e\rangle = \langle (\lambda,a); i_1(e)\rangle = \langle (\lambda,a);(e,0)\rangle = 0,$$

and so $\Lambda(\Delta) \cap (\mathsf{E} \oplus 0) \subseteq \Lambda(\Delta_1) \oplus 0$.

(iii) We have

$$\begin{aligned} \Lambda(L(\widehat{\Lambda(\Delta)})) &= \Lambda(\Lambda(\Delta) \cap (\mathsf{E} \oplus 0)) \\ &= \Lambda(\Lambda(\Delta)) + \Lambda(\mathsf{E} \oplus 0) \\ &= \Delta + (0 \oplus \mathbb{R}_{\mathsf{M}}). \end{aligned}$$

By (ii) we also have $\Lambda(L(\widehat{\Lambda(\Delta)})) = \Delta_1 \oplus \mathbb{R}_{\mathsf{M}}$.

(iv) Suppose that the condition (4.13) holds. Clearly $\widehat{\Lambda(\Delta)} \cap \widehat{\mathsf{F}} \neq \varnothing$. Since $\widehat{\Lambda(\Delta)} \subseteq \widehat{\mathsf{F}}$ and since $L(\widehat{\mathsf{F}}) = \mathsf{F} \oplus 0$, it follows that $L(\widehat{\Lambda(\Delta)}) \subseteq \mathsf{F} \oplus 0$. By part (iv), we have $\Lambda(\Delta_1) \subseteq \mathsf{F}$.

Now suppose that $\widehat{\Lambda(\Delta)} \cap \widehat{\mathsf{F}} \neq \varnothing$ and that $\Lambda(\Delta_1) \subseteq \mathsf{F}$. The arguments from the preceding part of the proof show that $L(\widehat{\Lambda(\Delta)}) \subseteq L(\widehat{\mathsf{F}})$. If $e \in \widehat{\Lambda(\Delta)} \cap \widehat{\mathsf{F}}$, this means that

$$e + e' \in \widehat{F}, \qquad e' \in L(\widehat{\Lambda(\Delta)}) \subseteq L(\widehat{\mathsf{F}}),$$

which gives $\widehat{\Lambda(\Delta)} \subseteq \widehat{\mathsf{F}}$, as asserted.

(v) Note that $\Lambda(\Delta)$ is flow-invariant under $\widehat{X}^{\mathrm{aff}}$ by Lemma 4.7, since Δ is flow-invariant under $\widehat{X}^{\mathrm{aff},*}$. Since $\mathsf{E} \times \{1\}$ is flow-invariant under $\widehat{X}^{\mathrm{aff}}$ (by Lemma 4.18(ii), taking $\Delta = \{0\}$ and so $\mathsf{A}(\Delta) = \mathsf{E}$), we conclude that the cogeneralised affine subbundle $\widehat{\Lambda(\Delta)}$ is flow-invariant under $\widehat{X}^{\mathrm{aff}}$, being the intersection of two flow-invariant sets. By Lemma 4.12 and since $\widehat{X}^{\mathrm{aff}}$ is a linear vector field, $L(\widehat{\Lambda(\Delta)})$ is flow-invariant under $\widehat{X}^{\mathrm{aff}}$. Recall that

$$\widehat{X}^{\mathrm{aff}} = X_0^{\mathrm{h}} + \widehat{(A,b)}^{\mathrm{e}},$$

the horizontal lift being that of the connection $\widehat{\nabla}$ on $\mathsf{E} \oplus \mathbb{R}_{\mathsf{M}}$ induced by the connection ∇ on E and the flat connection on $\mathbb{R}_{\mathsf{M}}$, and

$$\widehat{(A,b)}(e,a) = (A(e) + ab, 0), \qquad (e,a) \in \mathsf{E}_x \oplus \mathbb{R}.$$

Thus, by Proposition 4.9 (specialised to linear vector fields), we have

1. $\widehat{(A,b)}(L(\widehat{\Lambda(\Delta)})) \subseteq L(\widehat{\Lambda(\Delta)})$,
2. $\widehat{\nabla}_{X_0}(\mathscr{G}^r_{\Lambda(L(\widehat{\Lambda(\Delta)}))}) \subseteq \mathscr{G}^r_{\Lambda(L(\widehat{\Lambda(\Delta)}))}$.

By part (ii), we have

$$L(\widehat{\Lambda(\Delta)}) = \Lambda(\Delta_1) \oplus 0,$$

whence

$$\Lambda(L(\widehat{\Lambda(\Delta)})) = \Delta_1 \oplus \mathbb{R}_{\mathsf{M}}, \tag{4.14}$$

Note, then, that condition 1 is equivalent to $A(\Lambda(\Delta_1)) \subseteq \Lambda(\Delta_1)$ since $\widehat{(A,b)}(e,0) = (A(e),0)$ for $e \in \mathsf{E}$. Now let $\mathcal{U} \subseteq \mathsf{M}$ be open, and let $(\lambda, g) \in \mathscr{G}^r_{\Lambda(L(\widehat{\Lambda(\Delta)}))}(\mathcal{U})$ and note that

$$\widehat{\nabla}_{X_0}(\lambda, g) = (\nabla_{X_0}\lambda, \mathscr{L}_{X_0} g) \in \mathscr{G}^r_{\Lambda(L(\widehat{\Lambda(\Delta)}))}(\mathcal{U}).$$

Projecting this relation onto the first factor and taking note of (4.14), we deduce that

$$\nabla_{X_0}\lambda \in \mathscr{G}^r_{\Delta_1}(\mathcal{U}), \qquad \lambda \in \mathscr{G}^r_{\Delta_1}(\mathcal{U}).$$

Conversely, if λ satisfies the preceding relation and if $g \in \mathscr{C}^r_{\mathsf{M}}(\mathcal{U})$, then

$$\widehat{\nabla}_{X_0}(\lambda, g) = (\nabla_{X_0}\lambda, \mathscr{L}_{X_0}g) \in \mathscr{G}^r_{\Lambda(L(\widehat{\Lambda(\Delta)}))}(\mathcal{U}).$$

Thus the condition 2 above is equivalent to $\nabla_{X_0}(\mathscr{G}^r_{\Delta_1}) \subseteq \mathscr{G}^r_{\Delta_1}$. In summary, we have

1. $\widehat{(A,b)}(\Lambda(\Delta_1)) \subseteq \Lambda(\Delta_1)$ and
2. $\widehat{\nabla}_{X_0}(\mathscr{G}^r_{\Delta_1}) \subseteq \mathscr{G}^r_{\Delta_1}$.

Since Δ_1 is a generalised subbundle (being the image under a C^r-vector bundle mapping of a generalised subbundle), Proposition 4.10(ii) gives precisely these two conditions for invariance of $\Lambda(\Delta_1)$, as asserted. ■

The following result concerns the criterion $\Lambda(\Delta)_1 \subseteq \mathsf{F}$ from part (iv) of the lemma, making use of the flow-invariance property of part (v) of the lemma.

4.26 Theorem: (Linear parts of defining subbundles invariant under an affine vector field and annihilating a cogeneralised subbundle) *Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let Δ be a C^r-defining subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$, let $b \in \Gamma^r(\mathsf{E})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote*

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}$$

and suppose that Δ is flow-invariant under X^{aff}. Consider the following statements:

(i) $\Lambda(\Delta_1) \subseteq \mathsf{F}$;

(ii) the following conditions hold:

  (a) $A(\Lambda(\Delta_{1,x})) \subseteq \mathsf{F}_x$ *for* $x \in \mathsf{M}$;

  (b) $\nabla_{X_0}(\mathscr{G}^r_{\Lambda(\mathsf{F})}) \subseteq \mathscr{G}^r_{\Delta_1}$.

Then (i) $\Longrightarrow$ (ii) and, if either (1) $r = \omega$ or (2) $r = \infty$ and F is a subbundle, then (ii) $\Longrightarrow$ (i).

Proof: This follows firstly from Lemma 4.25(v), and then Theorem 4.24. ■

One can combine the previous results with Proposition 4.13 to obtain the following procedure for finding invariant affine subbundles contained in a given subbundle.

We first consider the linear case.

4.27 Remark: (Finding invariant cogeneralised subbundles contained in a cogeneralised subbundle) Let $r \in \{\infty, \omega\}$, let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$ be complete, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote

$$X^{\mathrm{lin}} = X_0^{\mathrm{h}} + A^{\mathrm{e}}.$$

Find a flow-invariant cogeneralised subbundle $\mathsf{L} \subseteq \mathsf{F}$ satisfying the following algebraic/differential conditions:

1. $A(\mathsf{L}) \subseteq \mathsf{F}$;
2. $\nabla_{X_0}(\mathscr{G}^r_{\Lambda(\mathsf{F})}) \subseteq \mathscr{G}^r_{\Lambda(\mathsf{L})}$.

We shall say that L satisfying these conditions is ***(X^{lin}, F)-admissible***. The resulting cogeneralised subbundle L is then flow-invariant under X^{lin} and is contained in F. •

In the affine case, we have the following.

4.28 Remark: (Finding invariant affine subbundle varieties contained in a cogeneralised subbundle) Let $r \in \{\infty, \omega\}$, let $\pi \colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, let ∇ be a C^r-linear connection in E, let $\mathsf{F} \subseteq \mathsf{E}$ be a C^r-cogeneralised subbundle, let $X_0 \in \Gamma^r(\mathsf{M})$ be complete, let $b \in \Gamma^r(\mathsf{E})$, and let $A \in \Gamma^r(\mathrm{End}(\mathsf{E}))$. Denote

$$X^{\mathrm{aff}} = X_0^{\mathrm{h}} + A^{\mathrm{e}} + b^{\mathrm{v}}.$$

Find a flow-invariant defining subbundle $\Delta \subseteq \mathsf{E}^* \oplus \mathbb{R}_{\mathsf{M}}$ satisfying the following algebraic/differential conditions:

1. $A(\Lambda(\Delta_{1,x})) \subseteq \mathsf{F}_x$ for $x \in \mathsf{M}$;
2. $\nabla_{X_0}(\mathscr{G}^r_{\Lambda(\mathsf{F})}) \subseteq \mathscr{G}^r_{\Delta_1}$.

We shall say that Δ satisfying these conditions is ***(X^{aff}, F)-linearly admissible***. Having found such a Δ, check the following:

3. the set $\mathsf{S}(\mathsf{A}(\Delta)) = \{x \in \mathsf{M} \mid (0,1) \notin \Delta_x\}$ is nonempty.

If an $(X^{\mathrm{aff}}, \mathsf{F})$-linearly admissible defining subbundle Δ satisfies this latter condition, we shall say that it is ***(X^{aff}, F)-admissible***. The resulting affine subbundle variety $\mathsf{A}(\Delta)$ is then flow-invariant under X^{aff} and is contained in F. •

The methodology outlined in the preceding constructions involve some interesting partial differential equations with algebraic constraints. With some effort, it might be possible to apply the integrability theory for partial differential equations of, e.g., Goldschmidt [1967a, 1967b] to arrive at the obstructions to solving these equations. An application of the resulting conditions to the setup of Section 7 would doubtless lead to some interesting answers to the central questions of this paper.

5. Nonholonomic and constrained variational mechanics

In this section we derive the two sets of equations whose correspondences we study. The equations we produce here are derived by Kupka and Oliva [2001], and we fill in some of the missing steps in their proofs. Additionally, we provide intrinsic proofs for some steps that are carried out using coordinates by Kupka and Oliva. For the most part, however, our derivations are intended to be an illustration of the methodology of Section 3 for working with spaces of curves.

We begin in Section 5.1 by characterising the tangent spaces to various classes of curves, using our constructions from Section 3.6. In Section 5.2 we introduce the energy functions we consider in the paper and indicate how to differentiate these, using the calculus from Section 3.5. The equations of nonholonomic mechanics are derived in Section 5.3, and we reiterate here that it is these equations of Section 5.3 that correspond in physics to the Newton–Euler equations. This is developed in a general and geometric setting by Lewis [2017]. By contrast, the constrained variational equations developed in Section 5.4 do not generally produce equations that correspond to the physical equations of motion; instead, this setting does reproduce the equations for extremals in sub-Riemannian geometry, as we explore in Section 6.

For subsequent brevity, let us make a definition encompassing the data in which we shall be interested in this section.

5.1 Definition: (Constrained simple mechanical system) Let $r \in \{\infty, \omega\}$, A $\mathbf{C}^r$-***constrained simple mechanical system*** is a quadruple $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ where

(i) M is a C^r-manifold (the ***configuration manifold***),

(ii) $\mathbb{G}$ is a C^r-Riemannian metric on M (the ***kinetic energy metric***),

(iii) V is a C^r-function (the ***potential energy function***), and

(iv) $\mathsf{D} \subseteq \mathsf{TM}$ is a C^r-subbundle (the ***constraint distribution***). •

5.2 Remark: (The subbundle assumption for the velocity constraints) In some physical systems, the velocity constraints do not describe a subbundle. In the physical systems of which we are aware, the failure of the velocity constraints to be a subbundle is a result of constraint forces aligning, and so the annihilator of the velocity constraints drops rank in such configurations. As a consequence, when the velocity constraints fail to be a subbundle, they are instead a *co*generalised subbundle (not a generalised subbundle). Moreover, we are not aware of any means of determining the equations of motion when the velocity constraints are not a subbundle. It seems like there is something in the Laws of Nature that is yet to be understood for nonholonomic mechanical systems.

The upshot of the above discussion is the following two points:

1. the assumption that the velocity constraints describe a subbundle is made *with* loss of physical generality;
2. we are not aware of any way of overcoming this loss of generality. •

5.1. Submanifolds of curves and their tangent spaces.

In this section we consider how some of the subsets of curves from Section 3.1 can be thought of as submanifolds of $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$. We shall also describe the tangent spaces to these submanifolds. To do this, we shall use reasoning closely resembling the usual Implicit Function Theorem arguments, using our notions of derivative and tangent space from Sections 3.5 and 3.6.

5.1.1. Curves with endpoint constraints. We first consider the subsets

$$\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0), \mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1) \subseteq \mathrm{H}^1([t_0,t_1];\mathsf{M}).$$

We recall from the proof of Lemma 3.3 that we had defined the two evaluation maps

$$\begin{aligned} \mathrm{ev}_{t_0} \colon \mathrm{H}^1([t_0,t_1];\mathsf{M}) &\to \mathsf{M} \\ \gamma &\mapsto \gamma(t_0) \end{aligned}$$

and

$$\begin{aligned} \mathrm{ev}_{(t_0,t_1)} \colon \mathrm{H}^1([t_0,t_1];\mathsf{M}) &\to \mathsf{M}\times\mathsf{M} \\ \gamma &\mapsto (\gamma(t_0),\gamma(t_1)), \end{aligned}$$

and had shown that they are continuous. Here we consider their differentiability and their derivatives.

5.3 Lemma: (Regular points for the evaluation maps) *Let M be a smooth manifold, let $t_1, t_1 \in \mathbb{R}$ with $t_0 < t_1$, and let $x_0, x_1 \in \mathsf{M}$. Then, for $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$, the following statements hold:*

(i) the mappings ev_{t_0} and $\mathrm{ev}_{(t_0,t_1)}$ are differentiable at γ with derivatives

$$T_\gamma \mathrm{ev}_{t_0}(\xi) = \xi(t_0), \quad T_\gamma \mathrm{ev}_{(t_0,t_1)}(\xi) = (\xi(t_0),\xi(t_1));$$

(ii) the derivatives

$$T_\gamma \mathrm{ev}_{t_0} \colon \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M}) \to \mathsf{T}_{\gamma(t_0)}\mathsf{M}$$

and

$$T_\gamma \mathrm{ev}_{(t_0,t_1)} \colon \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M}) \to \mathsf{T}_{\gamma(t_0)}\mathsf{M} \oplus \mathsf{T}_{\gamma(t_1)}\mathsf{M}$$

are surjective.

Proof: (i) For $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M})$, let σ be a variation of γ for which $\xi = \delta\sigma(0)$. Then $\mathrm{ev}_{t_0} \circ \sigma(s) = \hat{\sigma}(s,t_0)$ and so $\delta\mathrm{ev}_{t_0}(\gamma;\xi) = \xi(t_0)$. The map $\xi \mapsto \xi(t_0)$ is clearly linear, and this gives the differentiability of ev_{t_0}. Similarly, one ascertains that $\delta\mathrm{ev}_{(t_0,t_1)}(\gamma;\xi) = (\xi(t_0),\xi(t_1))$, and linearity of the map $\xi \mapsto (\xi(t_0),\xi(t_1))$ then gives the differentiability of $\mathrm{ev}_{(t_0,t_1)}$.

(ii) If we can show that $T_\gamma\mathrm{ev}_{(t_0,t_1)}$ is surjective, then $T_\gamma\mathrm{ev}_{t_0}$ is also surjective. So let $(v_0,v_1) \in \mathsf{T}_{\gamma(t_0)}\mathsf{M} \oplus \mathsf{T}_{\gamma(t_1)}\mathsf{M}$. Let $X \in \Gamma^\infty(\mathsf{TM})$ be such that $X(\gamma(t_0)) = v_0$ and $X(\gamma(t_1)) = v_1$. Then define $\xi \colon [t_0,t_1] \to \mathsf{TM}$ by $\xi = X \circ \gamma$. Since X is smooth and $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$, we ascertain that

$$\int_{t_0}^{t_1} |F\circ\xi(t)|^2 \, \mathrm{d}t = \int_{t_0}^{t_1} |F\circ X\circ\gamma(t)|^2 \, \mathrm{d}t < \infty$$

and

$$\int_{t_0}^{t_1} |(F\circ\xi)'(t)|^2 \, \mathrm{d}t = \int_{t_0}^{t_1} |(F\circ X\circ\gamma)'(t)|^2 \, \mathrm{d}t < \infty$$

for every $F \in \mathrm{Aff}^\infty(\mathsf{TM})$, showing that $\xi \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$. This gives surjectivity of $T_\gamma\mathrm{ev}_{(t_0,t_1)}$, as desired. ■

Now we note that

$$\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0) = \mathrm{ev}_{t_0}^{-1}(x_0)$$

and

$$\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1) = \mathrm{ev}_{(t_0,t_1)}^{-1}(x_0,x_1).$$

Thus, by the previous lemma and by appealing to an Implicit Function Theorem rationale, it makes sense to say that $\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0)$ and $\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ are submanifolds of $\mathrm{H}^1([t_0,t_1];\mathsf{M})$. Moreover, by the same rationale, we should say that the tangent space to $\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0)$ (resp. $\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$) at γ is $\ker(T_\gamma \mathrm{ev}_{t_0})$ (resp. $\ker(T_\gamma \mathrm{ev}_{(t_0,t_1)})$).

Whether or not one agrees with our calling things "tangent spaces" or "submanifolds," the following essential punchline remains valid and is all that we require in our subsequent development: given $x_0,x_1\in\mathsf{M}$, a curve $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$, and an infinitesimal variation $\delta\in\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$, there exists a variation σ of γ satisfying $\delta\sigma(0)=\delta$.

5.1.2. Curves with derivatives in a distribution. Now we consider the addition to the data of the subbundle $\mathsf{D}\subseteq\mathsf{TM}$. We first wish to prove that $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$ is a submanifold of $\mathrm{H}^1([t_0,t_1];\mathsf{M})$, and we do this in a manner similar to that in the preceding section, with an argument reminiscent of the Implicit Function Theorem, using the calculus from Section 3.6.

We recall from the proof of Lemma 3.3 the mapping

$$\begin{aligned}\hat{P}_{\mathsf{D}^\perp}\colon \mathrm{H}^1([t_0,t_1];\mathsf{M}) &\to \mathrm{H}^0([t_0,t_1];\mathsf{D}^\perp)\\ \gamma &\mapsto P_{\mathsf{D}^\perp}\circ\gamma',\end{aligned}$$

where we evidently have introduced a Riemannian metric $\mathbb{G}$ on M. Let us consider the differentiability properties of this map.

5.4 Lemma: (Regular points for the projection onto a distribution) *Let* $(\mathsf{M},\mathbb{G})$ *be a smooth Riemannian manifold, let* $\mathsf{D}\subseteq\mathsf{TM}$ *be a smooth subbundle, and let* $t_0,t_1\in\mathbb{R}$ *satisfy* $t_0<t_1$*. Then the following statements hold:*

(i) for $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M})$*, the mapping* $\hat{P}_{\mathsf{D}^\perp}$ *is differentiable at* γ *with derivative*

$$T_\gamma\hat{P}_{\mathsf{D}^\perp}(\xi) = P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi) + (\overset{\mathbb{G}}{\nabla}_{\xi}P_{\mathsf{D}^\perp})(\gamma');$$

(ii) for $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$*, the mapping*

$$T_\gamma\hat{P}_{\mathsf{D}^\perp}\colon \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})\to\mathsf{T}_\gamma\mathrm{H}^0([t_0,t_1];\mathsf{D}^\perp)$$

is surjective.

Proof: (i) Let $\sigma\colon J\to\mathrm{H}^1([t_0,t_1];\mathsf{M})$ be a variation of γ. Note that

$$\widehat{\hat{P}_{\mathsf{D}^\perp}\circ\sigma}(s,t) = \hat{P}_{\mathsf{D}^\perp}\circ\sigma(s)(t) = P_{\mathsf{D}^\perp}\circ\nu\hat{\sigma}(s,t).$$

Given this, to show that $\hat{P}_{\mathsf{D}^\perp}$ is continuous, it will suffice to show that, for $F\in\mathrm{Aff}^\infty(\mathsf{TM})$, the function

$$(s,t)\mapsto F\circ P_{\mathsf{D}^\perp}\circ\nu\hat{\sigma}(s,t)$$

defines a differentiable curve in $\mathrm{H}^0([t_0,t_1];\mathsf{TM})$, keeping in mind that, because we are working with H^0 and not H^1, we can ignore the conditions for differentiability that depend on t-derivatives, cf. Remark 3.10–1. First let $f\in\mathrm{C}^\infty(\mathsf{M})$ and compute

$$\partial_1(\pi_{\mathsf{TM}}^*f\circ P_{\mathsf{D}^\perp}\circ\nu\hat{\sigma})(s,t)=\langle \mathrm{d}f(\hat{\sigma}(s,t));(\overset{\mathrm{G}}{\nabla}_{(\hat{\sigma}^t)'(s)}P_{\mathsf{D}^\perp})(\nu\hat{\sigma}^t(s))+P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{(\hat{\sigma}^t)'(s)}\nu\hat{\sigma}^t(s))\rangle. \tag{5.1}$$

As we argued in the proof of Lemma 3.3(iii) (making use of Lemma 1.1(i)), knowledge of an estimate for the right-hand side of this previous expression gives a corresponding estimate for $F\in\mathrm{Lin}^\infty(\mathsf{TM})$. Thus it suffices to show that, for any $f\in\mathrm{C}^\infty(\mathsf{M})$ and any variation σ of γ, if we denote

$$\hat{\alpha}(s,t)=\pi_{\mathsf{TM}}^*f\circ P_{\mathsf{D}^\perp}\circ\nu\hat{\sigma}(s,t)$$

and

$$\hat{\beta}(s,t)=\langle \mathrm{d}f(\hat{\sigma}(s,t));(\overset{\mathrm{G}}{\nabla}_{(\hat{\sigma}^t)'(s)}P_{\mathsf{D}^\perp})(\nu\hat{\sigma}^t(s))+P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{(\hat{\sigma}^t)'(s)}\nu\hat{\sigma}^t(s))\rangle,$$

then the associated mappings

$$\alpha,\beta\colon J\to\mathrm{L}^2([t_0,t_1];\mathbb{R})$$

are well-defined and continuous.

Let us first consider α. Here we have $\alpha(s)=f\circ\sigma(s)$, and so this immediately give well-definedness and continuity of α, since σ takes values in

$$\mathrm{H}^0([t_0,t_1];\mathsf{M})\subseteq\mathrm{H}^1([t_0,t_1];\mathsf{M})$$

and since, by definition of the topology on $\mathrm{H}^1([t_0,t_1];\mathsf{M})$,

$$\mathrm{ev}_f\colon\mathrm{H}^0([t_0,t_1];\mathsf{M})\to\mathrm{L}^2([t_0,t_1];\mathbb{R})$$

is continuous.

Now we turn to β. We first prove a technical sublemma.

1 Sublemma: *There exist $X_1,\dots,X_N\in\Gamma^\infty(\mathsf{TM})$ such that*

$$\nu\hat{\sigma}(s,t)=\hat{\nu}^1(s,t)X_1(\hat{\sigma}(s,t))+\cdots+\hat{\nu}^N(s,t)X_N(\hat{\sigma}(s,t))$$

and

$$\delta\hat{\sigma}(s,t)=\hat{\delta}^1(s,t)X_1(\hat{\sigma}(s,t))+\cdots+\hat{\delta}^N(s,t)X_N(\hat{\sigma}(s,t))$$

for continuously differentiable functions $\nu^1,\dots,\nu^N\colon J\to\mathrm{L}^2([t_0,t_1];\mathbb{R})$ and continuous functions $\delta^1,\dots,\delta^N\colon J\to\mathrm{H}^1([t_0,t_1];\mathbb{R})$.

Proof: By Lemma 1.1(i), let $X_1,\dots,X_N\in\Gamma^\infty(\mathsf{TM})$ be global generators for $\Gamma^\infty(\mathsf{TM})$ as a $\mathrm{C}^\infty(\mathsf{M})$-module. As in the proof of Lemma 1.1(i), $X_1,\dots,X_N$ are the orthogonal projections of the coordinate vector fields $\widehat{X}_1,\dots,\widehat{X}_N$ on $\mathbb{R}^N$, where we have an embedding of M in $\mathbb{R}^N$. Let $\Sigma\colon J\to\mathrm{H}^0([t_0,t_1];\mathbb{R}^N_{\mathsf{M}})$ be continuously differentiable and such that $\pi\circ\Sigma=\sigma$, where $\pi\colon\mathbb{R}^N_{\mathsf{M}}\to\mathsf{M}$ is the projection. There are then unique

$$\hat{\Sigma}_1,\dots,\hat{\Sigma}_N\colon J\times[t_0,t_1]\to\mathbb{R}$$

such that
$$\hat{\Sigma}(s,t)=\hat{\Sigma}^1(s,t)\widehat{X}_1(\hat{\sigma}(s,t))+\cdots+\hat{\Sigma}^N(s,t)\widehat{X}_N(\hat{\sigma}(s,t)).$$
It is clear that, if $F^1,\dots,F^N\in\mathrm{Lin}^\infty(\mathbb{R}^N_{\mathsf{M}})$ comprise the standard dual basis, then
$$F^l\circ\hat{\Sigma}(s,t)=\hat{\Sigma}^l(s,t),\qquad l\in\{1,\dots,N\}.$$
Thus we conclude that
$$\Sigma^1,\dots,\Sigma^N\colon J\to\mathrm{L}^2([t_0,t_1];\mathbb{R})$$
are continuously differentiable. This construction applies, in particular, to $\Sigma=\nu\sigma$, in which case we get the assertion of the sublemma by projecting
$$\nu\hat{\sigma}(s,t)=\hat{\Sigma}^1(s,t)\widehat{X}_1(\hat{\sigma}(s,t))+\cdots+\hat{\Sigma}^N(s,t)\widehat{X}_N(\hat{\sigma}(s,t))$$
onto TM.

Moreover, the same construction applies if Σ takes values in $\mathrm{H}^1([t_0,t_1];\mathbb{R}^N_{\mathsf{M}})$, in which case we can apply the result to $\Sigma=\delta\sigma$, noting in this case that $\delta\sigma$ is not continuously differentiable, but continuous. ▼

Given the sublemma, we see that we can write
$$\hat{\beta}(s,t)=\sum_{l,m=1}^N\beta^0_{lm}(\hat{\sigma}(s,t))\hat{\delta}^l(s,t)\hat{\nu}^m(s,t)+\sum_{l,m=1}^N\beta^1_{lm}(\hat{\sigma}(s,t))\hat{\delta}^l(s,t)\partial_1\hat{\nu}^m(s,t)$$
for $\beta^0_{lm},\beta^1_{lm}\in\mathrm{C}^\infty(\mathsf{M})$, $l,m\in\{1,\dots,N\}$. For fixed s, $\hat{\sigma}_s$ and $\hat{\delta}^l_s$, $l\in\{1,\dots,N\}$, are continuous and so bounded on $[t_0,t_1]$. Since $\hat{\nu}^l_s,\partial_1\hat{\nu}^l_s\in\mathrm{L}^2([t_0,t_1];\mathbb{R})$, $l\in\{1,\dots,N\}$, we conclude that $\hat{\beta}_s\in\mathrm{L}^2([t_0,t_1];\mathbb{R})$ for each $s\in J$, and so β is well-defined.

To show that β is continuous, let $s_0\in J$ and let $(s_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in J converging to s_0. Let $K\subseteq J$ be a compact subinterval such that $s_j\in K$, $j\in\mathbb{Z}_{\geq 0}$.

Since σ is continuous, $\sigma(K)\subseteq\mathrm{H}^1([t_0,t_1];\mathsf{M})$ is compact. By definition of the topology for $\mathrm{H}^1([t_0,t_1];\mathsf{M})$,
$$\mathrm{ev}_{\beta^a_{lm}}\circ\sigma(K)\subseteq\mathrm{H}^1([t_0,t_1];\mathbb{R}),\qquad a\in\{0,1\},\ l,m\in\{1,\dots,N\},$$
are compact. Therefore, by (3.2), these subsets are also compact in $\mathrm{C}^0([t_0,t_1];\mathbb{R})$. Thus, for $a\in\{0,1\}$ and $l,m\in\{1,\dots,N\}$,
$$\sup\{\|\mathrm{ev}_{\beta^a_{lm}}\circ\sigma(s)\|_\infty\mid s\in K\}=\sup\{|\hat{\beta}^a_{lm}(s,t)|\mid (s,t)\in K\times[t_0,t_1]\}\leq M_1\tag{5.2}$$
for some $M_1\in\mathbb{R}_{>0}$. By similar reasoning, since $\delta^l\colon J\to\mathrm{H}^1([t_0,t_1];\mathbb{R})$, $l\in\{1,\dots,N\}$, are continuous, there exists $M_2\in\mathbb{R}_{>0}$ such that
$$\sup\{|\hat{\delta}^l(s,t)|\mid (s,t)\in K\times[t_0,t_1]\}\leq M_2,\qquad l\in\{1,\dots,N\}.\tag{5.3}$$

Note that the sequence $(\sigma(s_j))_{j\in\mathbb{Z}_{>0}}$ converges in $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ to $\sigma(s_0)$. By definition of the topology on $\mathrm{H}^1([t_0,t_1];\mathsf{M})$, the sequences $(\mathrm{ev}_{\beta^a_{lm}}\circ\sigma(s_j))_{j\in\mathbb{Z}_{>0}}$, $a\in\{0,1\}$, $l,m\in\{1,\dots,N\}$, converge in $\mathrm{H}^1([t_0,t_1];\mathbb{R})$ to $\mathrm{ev}_{\beta^a_{lm}}\circ\sigma(s_0)$. By (3.2), these sequences converge uniformly. Similarly, since $(\delta^l(s_j))_{j\in\mathbb{Z}_{>0}}$ converges in $\mathrm{H}^1([t_0,t_1];\mathbb{R})$, it converges uniformly to $\delta^l(s_0)$. Thus $(\beta^a_{lm}(s_j)\delta^l(s_j))_{j\in\mathbb{Z}_{>0}}$ converges uniformly to $\beta^a_{lm}(s_0)\delta^l(s_0)$. Therefore, for $\epsilon\in\mathbb{R}_{>0}$, there exists $n_1\in\mathbb{Z}_{>0}$ such that

$$|\beta^a_{lm}(\hat{\sigma}(s_0,t))\hat{\delta}^l(s_0,t)-\beta^a_{lm}(\hat{\sigma}(s_j,t))\hat{\delta}^l(s_0,t)|\leq\frac{\epsilon}{4N^2M_2},\\ t\in[t_0,t_1],\ a\in\{0,1\},\ l,m\in\{1,\dots,N\},\ j\geq n_1. \tag{5.4}$$

The sequences $(\nu\sigma(s_j))_{j\in\mathbb{Z}_{>0}}$ and $(\delta_1\nu\sigma(s_j))_{j\in\mathbb{Z}_{>0}}$ converge to $\nu\sigma(s_0)$ and $\delta_1\nu\sigma(s_0)$ in $\mathrm{H}^0([t_0,t_1];\mathsf{TM})$ and $\mathrm{H}^0([t_0,t_1];\mathsf{TTM})$, respectively. By the definition of the topology on $\mathrm{H}^0([t_0,t_1];\mathsf{TM})$ and $\mathrm{H}^0([t_0,t_1];\mathsf{TTM})$, the sequences $(\nu^l(s_j))_{j\in\mathbb{Z}_{>0}}$ and $(\partial_1\nu^l(s_j))_{j\in\mathbb{Z}_{>0}}$, $l\in\{1,\dots,N\}$, converge to $\nu^l(s_0)$ and $\partial_1\nu^l(s_0)$ in $\mathrm{L}^2([t_0,t_1];\mathbb{R})$. Therefore the sequences are also bounded, and so, for one thing, there exists $M_3\in\mathbb{R}_{>0}$ such that

$$\|\nu^l(s_j)\|_{\mathrm{L}^2},\|\partial_1\nu^l(s_j)\|_{\mathrm{L}^2}\leq M_3,\qquad l\in\{1,\dots,N\},\ j\in\mathbb{Z}_{>0}, \tag{5.5}$$

and, for another thing, for $\epsilon\in\mathbb{R}_{>0}$, there exists $n_2\in\mathbb{R}_{>0}$ such that

$$\|\nu^l(s_0)-\nu^l(s_j)\|_{\mathrm{L}^2},\|\partial_1\nu^l(s_0)-\partial_1\nu^l(s_j)\|_{\mathrm{L}^2}<\frac{\epsilon}{4N^2M_1M_2},\qquad l\in\{1,\dots,N\},\ j\geq n_2. \tag{5.6}$$

Thus

$$\begin{aligned}
|\hat{\beta}(s_0,t)-&\hat{\beta}(s_j,t)|\\
&\leq\sum_{l,m=1}^N|\beta^0_{lm}(\hat{\sigma}(s_0,t))\hat{\delta}^l(s_0,t)\hat{\nu}^m(s_0,t)-\beta^0_{lm}(\hat{\sigma}(s_j,t))\hat{\delta}^l(s_j,t)\hat{\nu}^m(s_j,t)|\\
&\quad+\sum_{l,m=1}^N|\beta^1_{lm}(\hat{\sigma}(s_0,t))\partial_1\hat{\delta}^l(s_0,t)\hat{\nu}^m(s_0,t)-\beta^1_{lm}(\hat{\sigma}(s_j,t))\partial_1\hat{\delta}^l(s_j,t)\hat{\nu}^m(s_j,t)|\\
&\leq\sum_{l,m=1}^N|\beta^0_{lm}(\hat{\sigma}(s_0,t))\hat{\delta}^l(s_0,t)(\hat{\nu}^m(s_0,t)-\hat{\nu}^m(s_j,t))|\\
&\quad+\sum_{l,m=1}^N|(\beta^0_{lm}(\hat{\sigma}(s_0,t))\hat{\delta}^l(s_0,t)-\beta^0_{lm}(\hat{\sigma}(s_j,t))\hat{\delta}^l(s_j,t))\hat{\nu}^m(s_j,t)|\\
&\quad+\sum_{l,m=1}^N|\beta^1_{lm}(\hat{\sigma}(s_0,t))\hat{\delta}^l(s_0,t)(\partial_1\hat{\nu}^m(s_0,t)-\partial_1\hat{\nu}^m(s_j,t))|\\
&\quad+\sum_{l,m=1}^N|(\beta^1_{lm}(\hat{\sigma}(s_0,t))\hat{\delta}^l(s_0,t)-\beta^1_{lm}(\hat{\sigma}(s_j,t))\hat{\delta}^l(s_j,t))\partial_1\hat{\nu}^m(s_j,t)|.
\end{aligned}$$

Using the estimates (5.2)–(5.6), we have

$$\|\beta(s_0)-\beta(s_j)\|_{\mathrm{L}^2}<\epsilon,\qquad j\geq\max\{n_1,n_2\},$$

showing that $(\beta(s_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\beta(s_0)$, as desired.

Finally, evaluating (5.1) at $s=0$ and using Lemma 3.14 gives

$$\delta(\pi^*f\circ P_{\mathsf{D}^\perp}\circ\nu\sigma)(0,t)=\langle\mathrm{d}f(\gamma(t));P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\xi(t))+(\overset{\mathrm{G}}{\nabla}_{\xi(t)}P_{\mathsf{D}^\perp})(\gamma'(t))\rangle.$$

As this hold for every $f \in \mathrm{C}^\infty(\mathsf{M})$, we conclude that

$$T_\gamma \hat{P}_{\mathsf{D}^\perp}(\xi) = P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\xi) + (\overset{\mathrm{G}}{\nabla}_{\xi} P_{\mathsf{D}^\perp})(\gamma'),$$

as claimed.

(ii) Let $\eta \in \mathsf{T}_{\hat{P}_{\mathsf{D}^\perp}(\gamma')}\mathrm{H}^0([t_0,t_1];\mathsf{D}^\perp)$. By Lemma 2.36(i), since $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$, we have

$$(\overset{\mathrm{G}}{\nabla}_{\xi} P_{\mathsf{D}^\perp})(\gamma') = -S_{\mathsf{D}}(\xi,\gamma') \in \mathrm{H}^0([t_0,t_1];\mathsf{D}^\perp).$$

By Lemma 2.36(ii), if $\xi \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$, we have

$$P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\xi) = \overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\xi \in \mathrm{L}^2([t_0,t_1],\gamma^*\mathsf{D}^\perp).$$

Thus the linear differential equation

$$\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\xi - S_{\mathsf{D}}(\xi,\gamma') = \eta$$

for sections ξ of $\mathsf{D}^\perp$ along γ has solutions for initial conditions in $\mathsf{D}^\perp$, and this gives the desired surjectivity of $T_\gamma \hat{P}_{\mathsf{D}^\perp}$. ■

Let us denote

$$\mathrm{Z}^0([t_0,t_1];\mathsf{D}^\perp) = \{\xi \in \mathrm{H}^0([t_0,t_1];\mathsf{D}^\perp) \mid \ \mathrm{image}(\xi) \subseteq \mathsf{M} \subseteq \mathsf{D}^\perp\}.$$

Note that

$$\mathrm{Z}^0([t_0,t_1];\mathsf{D}^\perp) = \mathrm{H}^1([t_0,t_1];\zeta)(\mathrm{H}^1([t_0,t_1];\mathsf{M})),$$

where $\zeta\colon \mathsf{M} \to \mathsf{D}^\perp$ is the zero section and where we are making reference to $\mathrm{H}^1([t_0,t_1];\zeta)$ as a functor as in Section 3.7. Since ζ is an injective immersion, by Lemma 3.20 we can assert that $\mathrm{Z}^0([t_0,t_1];\mathsf{D}^\perp)$ is a submanifold of $\mathrm{H}^0([t_0,t_1];\mathsf{D}^\perp)$. Since we clearly have

$$\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D}) = \hat{P}_{\mathsf{D}^\perp}^{-1}(\mathrm{Z}^0([t_0,t_1];\mathsf{D}^\perp)),$$

and since the preceding lemma shows that points in $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$ are regular points for $\hat{P}_{\mathsf{D}^\perp}$, we can assert that $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$ is a submanifold of $\mathrm{H}^1([t_0,t_1];\mathsf{M})$ with tangent space at $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$ given by

$$\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D}) = \{\xi \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M}) \mid \ P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\xi) - S_{\mathsf{D}}(\xi,\gamma') = 0\}.$$

5.1.3. Curves with derivatives in a distribution and with endpoint constraints. Now let us turn to the matter of whether $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$ and $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$ are submanifolds of $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$. First we consider the former.

5.5 Lemma: (Regular points for the evaluation map for curves with values in a distribution) *Let* $(\mathsf{M}, \mathbb{G})$ *be a smooth Riemannian manifold, let* $\mathsf{D} \subseteq \mathsf{TM}$ *be a smooth subbundle, and let* $t_0, t_1 \in \mathbb{R}$ *satisfy* $t_0 < t_1$*. Then, for* $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$*, the following statements hold:*

(i) the mapping

$$\begin{aligned} \mathrm{ev}_{t_0} \colon \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}) &\to \mathsf{M} \\ \gamma &\mapsto \gamma(t_0) \end{aligned}$$

is differentiable at γ *with derivative* $T_\gamma \mathrm{ev}_{t_0}(\xi) = \xi(t_0)$*;*

(ii) the derivative

$$T_\gamma \mathrm{ev}_{t_0} \colon \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}) \to \mathsf{T}_{\gamma(t_0)}\mathsf{M}$$

is surjective.

Proof: (i) For $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$, let σ be a variation of γ for which $\xi = \delta\sigma(0)$. Then $\mathrm{ev}_{t_0} \circ \sigma(s) = \hat{\sigma}(s, t_0)$ and so $\delta \mathrm{ev}_{t_0}(\gamma; \xi) = \xi(t_0)$. This gives differentiability of ev_{t_0} at γ.

(ii) Let $v_0 \in \mathsf{T}_{\gamma(t_0)}\mathsf{M}$ and let $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$ satisfy the initial value problem

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi - S_{\mathsf{D}}(\xi, \gamma') = 0, \qquad \xi(t_0) = v_0.$$

We claim that $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$. Indeed, since $S_{\mathsf{D}}(\xi, \gamma') \in \mathrm{H}^0([t_0, t_1]; \mathsf{D}^\perp)$, we have

$$P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi) - S_{\mathsf{D}}(\xi, \gamma') = 0,$$

giving the desired conclusion. ■

The lemma allows us to assert that $\mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0)$ is a submanifold of $\mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$ with tangent space at $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0)$ given by

$$\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0) = \{\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}) \mid P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi) - S_{\mathsf{D}}(\xi, \gamma') = 0,\ \xi(t_0) = 0\}.$$

Let us now consider curves in the distribution fixing the endpoint at t_1 as well. Here we must consider whether the mapping

$$\begin{aligned} \mathrm{ev}_{t_1} \colon \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0) &\to \mathsf{M} \\ \gamma &\mapsto \gamma(t_1) \end{aligned}$$

is differentiable at $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0)$ and whether the derivative is surjective. In this case, one verifies, just as in the proof of Lemma 5.5, that ev_{t_1} is differentiable. However, it is *not* generally the case that $T_\gamma \mathrm{ev}_{t_1}$ is surjective. Because of this, we introduce the following terminology.

5.6 Definition: (D-regular curve, D-singular curve) Let $(\mathsf{M}, \mathbb{G})$ be a smooth Riemannian manifold, let $\mathsf{D} \subseteq \mathsf{TM}$ be a smooth subbundle, and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. A curve $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$ is

(i) ***D-regular*** if

$$T_\gamma \mathrm{ev}_{t_1} \colon T_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; \gamma(t_0)) \to \mathsf{T}_{\gamma(t_1)}\mathsf{M}$$

is surjective and is

(ii) ***D-singular*** if it is not D-regular. •

Let us consider some singular curves to show that the above classification has content. A simple situation where singular curves arise is given in the following result.

5.7 Proposition: (The tangent space of $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$ when D is integrable) *Let* $(\mathsf{M},\mathbb{G})$ *be a Riemannian manifold and let* $\mathsf{D}\subseteq\mathsf{TM}$ *be a smooth subbundle. Then the following two statements are equivalent:*

(i) D *is integrable;*

(ii) for every $x_0\in\mathsf{M}$*, for every* $t_0,t_1\in\mathbb{R}$ *with* $t_0<t_1$*, and for every* $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$*, we have* $\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)=\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0)$.

Moreover,

(iii) all curves $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$ *are* D*-singular.*

Proof: (i) $\implies$ (ii) Let $x_0\in\mathsf{M}$, let $t_0,t_1\in\mathbb{R}$ satisfy $t_0<t_1$, let $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$, and let $\xi\in\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$. Suppose that $\xi\notin\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0)$. Then there exists $\tau\in(t_0,t_1]$ such that $\xi(\tau)\notin\mathsf{D}_{\gamma(\tau)}$. Let $\Lambda(\mathsf{D},x_0)$ be the leaf of the foliation associated with D through x_0 [Abraham, Marsden, and Ratiu 1988, Theorem 4.4.7]. Since γ is absolutely continuous with derivative almost everywhere in D, $\gamma(t)\in\Lambda(\mathsf{D},x_0)$ for all $t\in[t_0,t_1]$. By Lemma 3.12, let σ be a variation of γ for which $\xi=\delta\sigma$. By Lemma 3.7(ii), the curve $s\mapsto\hat\sigma(s,\tau)$ is a differentiable curve whose derivative at $s=0$ is not in $\mathsf{D}_{\gamma(t)}$. Thus, for small s_0, $\hat\sigma(s_0,\tau)\notin\Lambda(\mathsf{D},x_0)$. Since $\hat\sigma(s_0,t_0)=x_0\in\Lambda(\mathsf{D},x_0)$ and since the absolutely continuous curve $t\mapsto\hat\sigma(s_0,t)$ has tangent vector in D for almost every $t\in[t_0,t_1]$, it follows that $\hat\sigma(s_0,\tau)\in\Lambda(\mathsf{D},x_0)$. This contradiction means that we must have $\xi\in\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0)$.

Now let $\xi\in\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0)$. Define a variation of γ by

$$\begin{aligned}\hat\sigma\colon(-a,a)\times[t_0,t_1]&\to\mathsf{M}\\(s,t)&\mapsto\exp(s\xi(t)),\end{aligned}$$

where exp is the exponential map associated with the constrained connection $\overset{\mathsf{D}}{\nabla}$, cf. the proof of Lemma 3.12. Note that $s\mapsto\hat\sigma(s,t)$ is a geodesic for $\overset{\mathsf{D}}{\nabla}$ with initial condition $\xi(t)\in\mathsf{D}_{\gamma(t)}$. Since D is geodesically invariant under $\overset{\mathsf{D}}{\nabla}$, cf. Remark 2.37–4, this curve must lie in $\Lambda(\mathsf{D},x_0)$. Thus $\hat\sigma(s,t)\in\Lambda(\mathsf{D},x_0)$ for all $(s,t)\in(-a,a)\times[t_0,t_1]$. Thus σ takes values in $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$. Thus $\xi=\delta\sigma\in\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$.

(ii) $\implies$ (i) Let $x\in\mathsf{M}$, let $u,v\in\mathsf{D}_x$, let γ be the geodesic of $\overset{\mathsf{D}}{\nabla}$ satisfying $\gamma'(0)=v$, and let $\xi\in\mathsf{T}_\gamma\mathrm{H}^1([0,1];\mathsf{M})$ satisfy the initial value problem

$$\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\xi(t)-S_{\mathsf{D}}(\xi(t),\gamma'(t))=0,\qquad\xi(0)=0.$$

Then

$$F_{\mathsf{D}}(u,v)=P_{\mathsf{D}}^{\perp}(\overset{\mathbb{G}}{\nabla}_\xi\gamma'(0)-\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi(0))=-(P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi-S_{\mathsf{D}}(\xi,\gamma))(0)=0,$$

giving integrability of D.

(iii) Let $\gamma\in\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$ and let $\sigma\colon(-a,a)\to\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$ be a variation of γ satisfying the distribution and endpoint constraints. As we saw above, we have $\hat\sigma(s,t)\in\Lambda(\mathsf{D},x_0)$ for all $(s,t)\in(-a,a)\times[t_0,t_1]$. By Lemma 3.12, if $\xi\in\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$, we have $\xi=\delta\sigma(0)$ for a variation σ of γ as just described. Thus

$$T_\gamma\mathrm{ev}_{t_1}(\xi)(\xi)=\partial_1\hat\sigma(0,t_1)\in\mathsf{D}_{\gamma(t_1)}.$$

This precludes $T_\gamma\mathrm{ev}_{t_1}$, restricted to $\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0)$ from being surjective. Thus γ is a D-singular curve. ■

However, even when the sections of D satisfy the bracket generating condition, it might still be the case that $T_\gamma \mathrm{ev}_{t_1}$ is not surjective.

5.8 Example: (A curve whose right endpoint mapping does not have surjective derivative [Liu and Sussmann 1994, §2.3]) We take $\mathsf{M} = \mathbb{R}^3$ and let D be the subbundle

$$\mathsf{D}_{(x,y,z)} = \mathrm{span}_{\mathbb{R}}\left(\frac{\partial}{\partial x}, (1-x)\frac{\partial}{\partial y} + x^2\frac{\partial}{\partial z}\right).$$

One can readily verify that this distribution is bracket generating. We consider the curve $\gamma\colon [0,1] \to \mathbb{R}^3$ defined by $\gamma(t) = (0,t,0)$. Thus we take $t_0 = 0$, $t_1 = 1$, $x_0 = (0,0,0)$ and $x_1 = (0,1,0)$. To describe the tangent space of $\mathsf{T}_\gamma \mathrm{H}^1([0,1];\mathbb{R}^3;\mathsf{D})$, we let $\mathbb{G}$ be the Euclidean metric for $\mathbb{R}^3$ and compute the matrix representative of $P_{\mathsf{D}^\perp}$ to be

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{x^4}{x^4+(x-1)^2} & \frac{x^2(x-1)}{x^4+(x-1)^2} \\ 0 & \frac{x^2(x-1)}{x^4+(x-1)^2} & \frac{(x-1)^2}{x^4+(x-1)^2} \end{bmatrix}.$$

Let $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M})$ and write

$$\xi(t) = \xi^x(t)\frac{\partial}{\partial x} + \xi^y(t)\frac{\partial}{\partial y} + \xi^z(t)\frac{\partial}{\partial z}.$$

Note that the vector field $Y = \frac{\partial}{\partial y}$ has the property that $Y(\gamma(t)) = \gamma'(t)$ for $t \in [0,1]$. Therefore, we compute

$$S_{\mathsf{D}}(\xi(t), \gamma'(t)) = P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_\xi Y)(t) = 0.$$

We also compute

$$P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\xi(t)) = \dot{\xi}^z(t).$$

Therefore,

$$\mathsf{T}_\gamma \mathrm{H}^1([0,1];\mathbb{R}^3;\mathsf{D}) = \{\xi \in \mathsf{T}_\gamma \mathrm{H}^1([0,1];\mathbb{R}^3) \mid \dot{\xi}^z(t) = 0\}.$$

We then see that

$$\begin{aligned} &\mathsf{T}_\gamma \mathrm{H}^1([0,1];\mathbb{R}^3;\mathsf{D},(0,0,0)) \\ &\qquad = \{\xi \in \mathsf{T}_\gamma \mathrm{H}^1([0,1];\mathbb{R}^3) \mid \xi(0) = 0,\ \dot{\xi}^z(t) = 0\} \\ &\qquad = \{\xi \in \mathsf{T}_\gamma \mathrm{H}^1([0,1];\mathbb{R}^3) \mid \xi^x(0) = \xi^y(0) = 0,\ \xi^z(t) = 0,\ t \in [0,1]\}. \end{aligned}$$

It follows that

$$\mathrm{image}(T_\gamma \mathrm{ev}_{t_1}) = \{(u,v,w) \in \mathsf{T}_{(0,1,0)}\mathbb{R}^3 \mid w = 0\} \subset \mathsf{T}_{(0,1,0)}\mathbb{R}^3,$$

and so γ is not a regular point for ev_{t_1}. •

Finally, we can give a differential equation that characterises D-singular curves.

5.9 Proposition: (A characterisation of D-singular curves) *Let* $(\mathsf{M}, \mathbb{G})$ *be a smooth Riemannian manifold, let* $\mathsf{D} \subseteq \mathsf{TM}$ *be a smooth subbundle, and let* $t_0, t_1 \in \mathbb{R}$ *satisfy* $t_0 < t_1$*. For* $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$*, the following statements are equivalent:*

(i) γ *is a* D*-singular curve;*

(ii) there exists a nowhere zero $\lambda \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}^\perp)$ *such that*

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda + S^*_{\mathsf{D}}(\gamma')(\lambda) = 0.$$

Proof: Let us make some preliminary computations first.

For $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; \gamma(t_0))$ we have

$$P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi) - S_{\mathsf{D}}(\gamma')(\xi) = 0.$$

If we write $\xi = \xi^\parallel + \xi^\perp$ where $\xi^\parallel = P_{\mathsf{D}} \circ \xi$ and $\xi^\perp = P_{\mathsf{D}^\perp} \circ \xi$, then we have

$$\begin{aligned} P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi^\parallel) - S_{\mathsf{D}}(\gamma')(\xi^\parallel) + P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi^\perp) - S_{\mathsf{D}}(\gamma')(\xi^\perp) \\ = \overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\xi^\perp - S_{\mathsf{D}}(\gamma')(\xi^\perp) - G_{\mathsf{D}}(\gamma', \xi^\parallel) = 0. \end{aligned}$$

Thus we see that $\xi^\parallel$ can be freely chosen, with $\xi^\parallel(t_0) = 0$, and then $\xi^\perp$ is obtained as the solution to the initial value problem

$$\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\xi^\perp - S_{\mathsf{D}}(\gamma')(\xi^\perp) = G_{\mathsf{D}}(\gamma', \xi^\parallel), \qquad \xi^\perp(t_0) = 0.$$

Let $\lambda \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}^\perp)$. Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{G}(\xi(t), \lambda(t)) = \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\xi(t), \lambda(t)) + \mathbb{G}(\xi(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\lambda(t))$$

and since $\xi(t_0) = 0$,

$$\mathbb{G}(\xi(t_1), \lambda(t_1)) = \int_{t_0}^{t_1} \Big(\mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\xi(t), \lambda(t)) + \mathbb{G}(\xi(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\lambda(t))\Big)\, \mathrm{d}t.$$

Now we have

$$\mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi, \lambda) = \mathbb{G}(P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi), \lambda) = \mathbb{G}(S_{\mathsf{D}}(\gamma', \xi), \lambda) = \mathbb{G}(\xi, S^*_{\mathsf{D}}(\gamma')(\lambda)),$$

which gives

$$\mathbb{G}(\xi(t_1), \lambda(t_1)) = \int_{t_0}^{t_1} \Big(\mathbb{G}(\xi(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\lambda(t) + S^*_{\mathsf{D}}(\gamma'(t))(\lambda(t)))\Big)\, \mathrm{d}t. \tag{5.7}$$

Now we proceed with the two implications of the proof.

(i) $\Longrightarrow$ (ii) Since $T_\gamma \mathrm{ev}_{t_1}$ is not surjective, there exists $v_1 \in \mathsf{T}_{\gamma(t_1)}\mathsf{M}$ such that $\mathbb{G}(v_1, \xi(t_1)) = 0$ for all $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; \gamma(t_0))$. Let us write $v_1 = v_1^\parallel + v_1^\perp$ with $v_1^\parallel \in \mathsf{D}_{\gamma(t_1)}$ and $v_1^\perp \in \mathsf{D}^\perp_{\gamma(t_1)}$. Let $\lambda \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}^\perp)$ satisfy the final value problem

$$\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda + P_{\mathsf{D}^\perp} \circ S^*_{\mathsf{D}}(\gamma')(\lambda) = 0, \quad \lambda(t_1) = v_1^\perp.$$

Since $\overset{\mathsf{D}^\perp}{\nabla}$ leaves $\mathsf{D}^\perp$ invariant by Lemma 2.36(ii), we see that λ is indeed a section of $\mathsf{D}^\perp$ over γ. Now, for $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};\gamma(t_0))$, we have

$$0 = \mathbb{G}(\xi(t_1), v_1) = \mathbb{G}(\xi^\perp(t_1), \lambda(t_1)) + \mathbb{G}(\xi^\parallel(t_1), v_1^\parallel). \tag{5.8}$$

From (5.7), and noting the definition of λ, we have

$$\begin{aligned}
\mathbb{G}(\xi^\perp(t_1), \lambda(t_1)) &= \mathbb{G}(\xi(t_1), \lambda(t_1)) \\
&= \int_{t_0}^{t_1} \left(\mathbb{G}(\xi(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\lambda(t) + S^*_{\mathsf{D}}(\gamma'(t))(\lambda(t))) \right) \mathrm{d}t \\
&= \int_{t_0}^{t_1} \left(\mathbb{G}(\xi^\parallel(t), P_{\mathsf{D}}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\lambda(t)) + P_{\mathsf{D}}(S^*_{\mathsf{D}}(\gamma'(t))(\lambda(t)))) \right) \mathrm{d}t.
\end{aligned}$$

By (5.8), we can then write

$$0 = \mathbb{G}(\xi^\parallel(t_1), v_1^\parallel) + \int_{t_0}^{t_1} \left(\mathbb{G}(\xi^\parallel(t), P_{\mathsf{D}}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\lambda(t)) + P_{\mathsf{D}} \circ S^*_{\mathsf{D}}(\gamma'(t))(\lambda(t))) \right) \mathrm{d}t.$$

As we saw above, $\xi^\parallel$ can be freely chosen, to satisfy $\xi^\parallel(t_0) = 0$. Choosing $\xi^\parallel$ to be pointwise orthogonal to

$$P_{\mathsf{D}}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\lambda) + P_{\mathsf{D}} \circ S^*_{\mathsf{D}}(\gamma')(\lambda),$$

it follows that $v_1^\parallel = 0$. Again since we can freely choose $\xi^\parallel$ (up to its vanishing at t_0), it then follows that

$$P_{\mathsf{D}}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\lambda) + P_{\mathsf{D}} \circ S^*_{\mathsf{D}}(\gamma')(\lambda) = 0.$$

Combining this with the definition of λ gives

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\lambda + S^*_{\mathsf{D}}(\gamma')(\lambda) = 0.$$

As long as we take $v_1^\perp \neq 0$, it follows that λ is also nowhere zero since it is a solution to a linear differential equation with a nonzero final condition.

(ii) $\Longrightarrow$ (i) Let $\lambda \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$ be nowhere zero and satisfy

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\lambda + S^*_{\mathsf{D}}(\gamma')(\lambda) = 0.$$

Then, by (5.7), we have

$$\mathbb{G}(\xi(t_1), \lambda(t_1)) = 0, \qquad \xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};\gamma(t_0)).$$

This implies that $\mathrm{image}(T_\gamma \mathrm{ev}_{t_1})$ is orthogonal to $\lambda(t_1) \neq 0$, and so $T_\gamma \mathrm{ev}_{t_1}$ is not surjective. ■

5.1.4. A summary of classes of curves and their tangent spaces. We can summarise the preceding constructions with the following definitions.

5.10 Definition: (Tangent space to spaces of curves in a distribution with endpoint constraints) Let $(\mathsf{M}, \mathbb{G})$ be a smooth manifold, let $\mathsf{D} \subseteq \mathsf{TM}$ be a smooth subbundle, let $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$, and let $x_0, x_1 \in \mathsf{M}$.

(i) The ***tangent space*** to $\mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0)$ at γ is

$$\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0) = \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{TM}; x_0).$$

(ii) The ***tangent space*** to $\mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1)$ at γ is

$$\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1) = \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{TM}; x_0, x_1).$$

(iii) The ***tangent space*** to $\mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$ at γ is

$$\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}) = \{\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}) \mid P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\xi) - S_{\mathsf{D}}(\xi, \gamma') = 0\}.$$

(iv) The ***tangent space*** to $\mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0)$ at γ is

$$\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0) = \{\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}) \mid \xi(t_0) = 0\}.$$

(v) If γ is additionally a D-regular curve, then the ***tangent space*** to $\mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1)$ at γ is

$$\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1) = \{\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0) \mid \xi(t_1) = 0\}.$$ •

5.2. Energies and their derivatives. We shall work with the natural kinetic energy function associated with a Riemannian metric as a Lagrangian, as well as the simpler potential energy function. We shall also need to characterise an appropriate derivative of the actions associated with these energy functions using our derivative from Definition 3.16.

First we consider kinetic energy.

5.11 Definition: (Kinetic energy function, kinetic energy action, constrained kinetic energy action) Let $(\mathsf{M}, \mathbb{G})$ be a smooth Riemannian manifold and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$.

(i) The ***kinetic energy function*** is the mapping

$$\begin{aligned} K_{\mathbb{G}} \colon \mathsf{TM} &\to \mathbb{R} \\ v_x &\mapsto \frac{1}{2}\mathbb{G}(v_x, v_x). \end{aligned}$$

(ii) The ***kinetic energy action*** is the mapping

$$\begin{aligned} A_{\mathbb{G}} \colon \mathrm{H}^1([t_0, t_1]; \mathsf{M}) &\to \mathbb{R} \\ \gamma &\mapsto \int_{t_0}^{t_1} K_{\mathbb{G}}(\gamma'(t))\,\mathrm{d}t. \end{aligned}$$

If, additionally, $\mathsf{D} \subseteq \mathsf{TM}$ is a smooth subbundle,

(iii) the ***constrained kinetic energy action*** is the mapping

$$\begin{aligned} A_{\mathbb{G},\mathsf{D}} \colon \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}) &\to \mathbb{R} \\ \gamma &\mapsto A_{\mathbb{G}}(\gamma). \end{aligned}$$ •

We should ensure that the kinetic energy action is well-defined.

5.12 Lemma: (Well-definedness of kinetic energy action) *If* $(\mathsf{M}, \mathbb{G})$ *is a smooth Riemannian manifold, then the kinetic energy action is a continuous function on* $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$.

Proof: To show that $A_{\mathbb{G}}$ is well-defined on $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$, note that, since $\gamma' \in \mathrm{H}^0([t_0, t_1]; \mathsf{TM})$, there are $X_1, \dots, X_N \in \Gamma^\infty(\mathsf{TM})$ such that we can write

$$\gamma'(t) = \gamma^1(t)X_1(\gamma(t)) + \cdots + \gamma^N(t)X_N(\gamma(t))$$

for $\gamma^1, \dots, \gamma^N \in \mathrm{L}^2([t_0, t_1]; \mathbb{R})$, cf. Sublemma 1 from the proof of Lemma 5.4. Then we have

$$A_{\mathbb{G}}(\gamma) = \frac{1}{2} \sum_{l,m=1}^{N} \int_{t_0}^{t_1} \gamma^l(t)\gamma^k(t)\mathbb{G}(X_l(\gamma(t)), X_m(\gamma(t)))\, \mathrm{d}t.$$

Since the function $x \mapsto \mathbb{G}(X_l(x), X_m(x))$ is smooth, it is bounded on any compact subset of M containing $\operatorname{image}(\gamma)$. Thus

$$t \mapsto \sum_{l,m=1}^{N} \gamma^l(t)\gamma^k(t)\mathbb{G}(X_l(\gamma(t)), X_m(\gamma(t)))$$

is integrable since the product of square integrable functions is integrable by Hölder's inequality. This shows that $A_{\mathbb{G}}$ is well-defined.

To see that it is continuous, let $(\gamma_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$ converging to γ. First we calculate, for $l, m \in \{1, \dots, N\}$,

$$\int_{t_0}^{t_1} |\gamma_j^l(t)\gamma_j^m(t) - \gamma^l(t)\gamma^m(t)|\, \mathrm{d}t \\ \leq \int_{t_0}^{t_1} |\gamma_j^l(t) - \gamma^l(t)||\gamma_j^m(t)|\, \mathrm{d}t + \int_{t_0}^{t_1} |\gamma^l(t)||\gamma_j^m(t) - \gamma^m(t)|\, \mathrm{d}t.$$

Since the sequences $(\gamma_j^l)_{j \in \mathbb{Z}_{>0}}$, $l \in \{1, \dots, N\}$, are bounded, we arrive at the conclusion that

$$\lim_{j \to \infty} \int_{t_0}^{t_1} |\gamma_j^l(t)\gamma_j^m(t) - \gamma^l(t)\gamma^m(t)|\, \mathrm{d}t = 0.$$

Finally, since

$$t \mapsto \mathbb{G}(X_l(\gamma(t)), X_m(\gamma(t)))$$

is bounded, we have

$$\lim_{j \to \infty} |A_{\mathbb{G}}(\gamma_j) - A_{\mathbb{G}}(\gamma)| = 0,$$

giving the desired continuity. ■

The following result characterises the derivative of $A_{\mathbb{G}}$.

5.13 Lemma: (Derivative of kinetic energy action) *Let* $(\mathsf{M}, \mathbb{G})$ *be a smooth Riemannian manifold, let* $t_0, t_1 \in \mathbb{R}$ *satisfy* $t_0 < t_1$*, let* $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$*, and let* $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$*. Then* $A_{\mathbb{G}}$ *is differentiable at* γ *and*

$$T_\gamma A_{\mathbb{G}}(\xi) = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \xi(t), \gamma'(t))\, \mathrm{d}t.$$

Proof: By Lemma 3.12, let σ be a variation of γ with $\delta\sigma(0) = \xi$. We then have

$$A_{\mathbb{G}} \circ \sigma(s) = \frac{1}{2} \int_{t_0}^{t_1} \mathbb{G}(\hat{\sigma}'_s(t), \hat{\sigma}'_s(t))\, \mathrm{d}t.$$

According to the definition of variational derivative following Definition 3.15, we have

$$\begin{aligned}
\delta A_{\mathbb{G}}(\gamma; \xi)(t) &= \delta(A_{\mathbb{G}} \circ \sigma)(0)(t) \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \frac{1}{2} \int_{t_0}^{t_1} \mathbb{G}(\nu\hat{\sigma}(s,t), \nu\hat{\sigma}(s,t))\, \mathrm{d}t \\
&= \left.\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{(\hat{\sigma}^t)'(s)} \nu\hat{\sigma}^t(s), \nu\hat{\sigma}(s,t))\, \mathrm{d}t\right|_{s=0} \\
&= \left.\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\hat{\sigma}'_s(t)} \delta\hat{\sigma}_s(t), \nu\hat{\sigma}(s,t))\, \mathrm{d}t\right|_{s=0} \\
&= \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \xi(t), \gamma'(t))\, \mathrm{d}t,
\end{aligned}$$

where we have used Lemma 3.14. This shows that $A_{\mathbb{G}}$ is differentiable at γ and $T_\gamma A_{\mathbb{G}}$ is as asserted. ■

Because $A_{\mathbb{G}}$ is $\mathbb{R}$-valued, we denote

$$T_\gamma A_{\mathbb{G}}(\xi) = \langle \mathrm{d}A_{\mathbb{G}}(\gamma); \xi \rangle, \qquad \xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}).$$

Let us next consider potential energy.

5.14 Definition: (Potential energy function, potential energy action, constrained potential energy action) Let $(\mathsf{M}, \mathbb{G})$ be a smooth Riemannian manifold and let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$.

(i) A ***potential energy function*** is $V \in \mathrm{C}^\infty(\mathsf{M})$.

(ii) The ***potential energy action*** is the mapping

$$\begin{aligned}
A_V \colon \mathrm{H}^1([t_0, t_1]; \mathsf{M}) &\to \mathbb{R} \\
\gamma &\mapsto \int_{t_0}^{t_1} V \circ \gamma(t)\, \mathrm{d}t.
\end{aligned}$$

If, additionally, $\mathsf{D} \subseteq \mathsf{TM}$ is a smooth subbundle,

(iii) the ***constrained potential energy action*** is the mapping

$$\begin{aligned}
A_{V,\mathsf{D}} \colon \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}) &\to \mathbb{R} \\
\gamma &\mapsto A_V(\gamma).
\end{aligned}$$

•

Just by the definition of continuity in $\mathrm{H}^1([t_0, t_1]; \mathsf{M})$, A_V is continuous. We can also easily characterise its differentiability.

5.15 Lemma: (Derivative of potential energy action) *Let* $(\mathsf{M}, \mathbb{G})$ *be a smooth Riemannian manifold, let* $V \in \mathrm{C}^\infty(\mathsf{M})$ *be a potential energy function, let* $t_0, t_1 \in \mathbb{R}$ *satisfy* $t_0 < t_1$*, let* $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$*, and let* $\xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$*. Then* A_V *is differentiable at* γ *and*

$$T_\gamma A_V(\xi) = \int_{t_0}^{t_1} \mathbb{G}(\operatorname{grad} V \circ \gamma(t), \xi(t))\, \mathrm{d}t.$$

Proof: Let σ be a variation of γ such that $\delta\sigma(0) = \xi$. Using our definitions of variational derivative, we calculate

$$\begin{aligned} T_\gamma A_V(\xi) &= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \int_{t_0}^{t_1} V \circ \hat{\sigma}^t(s)\, \mathrm{d}t \\ &= \left.\int_{t_0}^{t_1} \langle \mathrm{d}V(\hat{\sigma}(s,t)); \delta\hat{\sigma}(s,t)\rangle\, \mathrm{d}t\right|_{s=0} \\ &= \int_{t_0}^{t_1} \mathbb{G}(\operatorname{grad} V \circ \gamma(t), \xi(t))\, \mathrm{d}t. \end{aligned}$$

Thus A_V is differentiable with the asserted derivative. ■

As with the derivative of the kinetic energy action, we denote

$$T_\gamma A_V(\xi) = \langle \mathrm{d}A_V(\gamma); \xi\rangle, \qquad \xi \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}).$$

Now we can combine the two energies, and we record a formula for the derivative in the direction of fixed endpoint variations in the following result.

5.16 Lemma: (Derivative of kinetic minus potential energy action in the direction of a fixed endpoint variation) *Let* $(\mathsf{M}, \mathbb{G})$ *be a smooth Riemannian manifold, let* V *be a potential energy function, let* $t_0, t_1 \in \mathbb{R}$ *satisfy* $t_0 < t_1$*, let* $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$*, and let* $\delta \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1)$*. Then*

$$\langle \mathrm{d}(A_{\mathbb{G}} - A_V)(\gamma); \delta\rangle = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\, \mathrm{d}t,$$

for any $\beta_V \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$ *satisfying* $\overset{\mathbb{G}}{\nabla}_{\gamma'}\beta_V = \operatorname{grad} V \circ \gamma$.

Proof: We first state an elementary sublemma, which can be seen as "integration by parts."

1 Sublemma: *If* $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M})$*, if* $\delta \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1)$*, and if* $\alpha \in \mathrm{L}^2([t_0, t_1]; \gamma^*\mathsf{TM})$*, then*

$$\int_{t_0}^{t_1} \mathbb{G}(\delta(t), \alpha(t))\, \mathrm{d}t = -\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \beta(t))\, \mathrm{d}t$$

for every $\beta \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$ *satisfying* $\overset{\mathbb{G}}{\nabla}_{\gamma'}\beta = \alpha$.

Proof: We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{G}(\delta(t), \beta(t)) = \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \beta(t)) + \mathbb{G}(\delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\beta(t)),$$

and so

$$\begin{aligned}0 &= \mathbb{G}(\delta(t_1), \beta(t_1)) - \mathbb{G}(\delta(t_0), \beta(t_0)) \\ &= \int_{t_0}^{t_1} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{G}(\delta(t), \beta(t))\,\mathrm{d}t \\ &= \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \beta(t))\,\mathrm{d}t + \int_{t_0}^{t_1} \mathbb{G}(\delta(t), \alpha(t))\,\mathrm{d}t,\end{aligned}$$

as desired. ▼

Now we can execute the proof. Combining Lemmata 5.13 and 5.15, we have

$$\langle \mathrm{d}(A_{\mathbb{G}} - A_V)(\gamma); \delta\rangle = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t))\,\mathrm{d}t - \int_{t_0}^{t_1} \mathbb{G}(\operatorname{grad} V \circ \gamma(t), \delta(t))\,\mathrm{d}t.$$

Let $\beta_V \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$ satisfy $\overset{\mathrm{G}}{\nabla}_{\gamma'}\beta_V = \operatorname{grad} V \circ \gamma$. By the sublemma we have

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \operatorname{grad} V \circ \gamma(t))\,\mathrm{d}t = -\int_{t_0}^{t_1} \mathbb{G}(\delta(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\beta_V(t))\,\mathrm{d}t$$

and so

$$\langle \mathrm{d}(A_{\mathbb{G}} - A_V)(\gamma); \delta\rangle = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\,\mathrm{d}t,$$

as claimed. ■

5.3. Nonholonomic mechanics. The translation of the Newton–Euler force/moment balance equations to a geometric setting gives the following two essential features of the equations: (1) the motions must satisfy the constraints; (2) there is a force, orthogonal to the constraint distribution, required to maintain the constraints [e.g., Lewis 2017]. This idea is often referred to "d'Alembert's Principle," among other things. Let us translate this into language amenable to our present needs.

5.17 Definition: (Nonholonomic trajectory) Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Let $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$ and let $x_0, x_1 \in \mathsf{M}$. A curve $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1)$ is a ***nonholonomic trajectory*** for Σ if

$$\langle \mathrm{d}(A_{\mathbb{G}} - A_V); \delta\rangle = 0, \qquad \delta \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}; x_0, x_1). \qquad \bullet$$

The following result characterises nonholonomic trajectories. The section λ along γ in the second of the statements of the theorem is to be thought of as a "constraint force," orthogonal to the constraints.

5.18 Theorem: (Characterisation of nonholonomic trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Let $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$ and let $x_0, x_1 \in \mathsf{M}$. For $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1)$, the following statements are equivalent:*

(i) γ is a nonholonomic trajectory;

(ii) $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$ *and there exists* $\lambda \in \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^\perp)$ *such that*

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma = \lambda;$$

(iii) $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$ *and satisfies*

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0.$$

Proof: (i) $\Longrightarrow$ (ii) Let $\beta_V \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D})$ satisfy

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\beta_V = P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma.$$

Thus

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\beta_V = \overset{\mathsf{D}}{\nabla}_{\gamma'}\beta_V + S_{\mathsf{D}}(\gamma',\beta_V) = P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma + S_{\mathsf{D}}(\gamma',\beta_V).$$

Then, for $\delta \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0,x_1)$, we have

$$\begin{aligned}
\int_{t_0}^{t_1} \mathbb{G}(\delta(t), \operatorname{grad} V \circ \gamma(t))\,\mathrm{d}t &= \int_{t_0}^{t_1} \mathbb{G}(\delta(t), P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma(t))\,\mathrm{d}t \\
&= \int_{t_0}^{t_1} \mathbb{G}(\delta(t), P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma(t) + S_{\mathsf{D}}(\gamma'(t),\beta_V(t)))\,\mathrm{d}t \\
&= \int_{t_0}^{t_1} \mathbb{G}(\delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\beta_V(t))\,\mathrm{d}t \\
&= -\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \beta_V(t))\,\mathrm{d}t,
\end{aligned}$$

using Sublemma 1 from the proof of Lemma 5.16. Therefore, combining Lemmata 5.13 and 5.15, the current hypothesis is that

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\,\mathrm{d}t = 0$$

for all $\delta \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0,x_1)$. Let us show that this implies that $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$.

Define $\zeta_0,\zeta_1 \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ as solutions to the initial value problems

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\zeta_0 = \gamma' + \beta_V, \quad \zeta_0(t_0) = 0,$$

and

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\zeta_1 = 0, \quad \zeta_1(t_1) = \zeta_0(t_1),$$

respectively. Since $\overset{\mathsf{D}}{\nabla}$ is a connection in D, we have $\zeta_0,\zeta_1 \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D})$. Denote

$$\delta(t) = \zeta_0(t) - \left(\frac{t-t_0}{t_1-t_0}\right)\zeta_1(t)$$

so that $\delta(t_0)=0$ and $\delta(t_1)=0$. Also, for almost every $t\in[t_0,t_1]$,

$$\begin{aligned}\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t) &= \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\zeta_0(t)-\left(\frac{t_0}{t_1-t_0}\right)\zeta_1(t)-\left(\frac{t-t_0}{t_1-t_0}\right)\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\zeta_1(t)\\ &= \overset{\mathsf{D}}{\nabla}_{\gamma'(t)}\zeta_0(t)+(S_{\mathsf{D}}(\gamma'(t),\zeta_0(t))-\left(\frac{t_0}{t_1-t_0}\right)\zeta_1(t)\\ &\quad -\frac{t-t_0}{t_1-t_0}\overset{\mathsf{D}}{\nabla}_{\gamma'(t)}\zeta_1(t)-\frac{t-t_0}{t_1-t_0}S_{\mathsf{D}}(\gamma'(t),\zeta_1(t))\\ &= \gamma'(t)+\beta_V(t)-\left(\frac{t_0}{t_1-t_0}\right)\zeta_1(t)+S_{\mathsf{D}}(\gamma'(t),\delta(t)).\end{aligned}$$

By Lemma 2.36(i) we have

$$S_{\mathsf{D}}(\gamma',\delta)\in\mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^{\perp})$$

since $\delta\in\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D})$.

We have

$$\begin{aligned}0 &= \mathbb{G}(\zeta_1(t_1),\delta(t_1))-\mathbb{G}(\zeta_1(t_0),\delta(t_0))\\ &= \int_{t_0}^{t_1}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{G}(\zeta_1(t),\delta(t))\\ &= \int_{t_0}^{t_1}\left(\mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\zeta_1(t),\delta(t))+\mathbb{G}(\zeta_1(t),\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t))\right)\mathrm{d}t\\ &= \int_{t_0}^{t_1}\mathbb{G}(\overset{\mathsf{D}}{\nabla}_{\gamma'(t)}\zeta_1(t)+S_{\mathsf{D}}(\gamma'(t),\zeta_1(t)),\delta(t))\,\mathrm{d}t\\ &\quad+\int_{t_0}^{t_1}\mathbb{G}\left(\zeta_1(t),\gamma'(t)+\beta_V(t)-\left(\frac{t_0}{t_1-t_0}\right)\zeta_1(t)\right)\mathrm{d}t\\ &= \int_{t_0}^{t_1}\mathbb{G}\left(\zeta_1(t),\gamma'(t)+\beta_V(t)-\left(\frac{t_0}{t_1-t_0}\right)\zeta_1(t)\right)\mathrm{d}t,\end{aligned}$$

noting that

$$S_{\mathsf{D}}(\gamma',\zeta_1),S_{\mathsf{D}}(\gamma',\delta)\in\mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^{\perp})$$

since $\zeta_1,\delta\in\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D})$.

Since $\delta\in\mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0,x_1)$, we have

$$\begin{aligned}0 &= \int_{t_0}^{t_1}\mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t),\gamma'(t)+\beta_V(t))\,\mathrm{d}t\\ &= \int_{t_0}^{t_1}\mathbb{G}\left(\gamma'(t)+\beta_V(t)-\left(\frac{t_0}{t_1-t_0}\right)\zeta_1(t),\gamma'(t)+\beta_V(t)\right)\mathrm{d}t,\end{aligned}$$

again using the fact that

$$S_{\mathsf{D}}(\gamma',\delta)\in\mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^{\perp}).$$

Combining the preceding two computations gives

$$\int_{t_0}^{t_1}\left\|\gamma'(t)+\beta_V(t)-\frac{t_0}{t_1-t_0}\zeta_1(t)\right\|_{\mathbb{G}}^2\mathrm{d}t=0,$$

whence $\gamma' = \frac{t_0}{t_1-t_0}\zeta_1 - \beta_V$. Since $\zeta_1, \beta_V \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M})$, we conclude that $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$, as desired.

Having now shown that $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$, it is straightforward to complete the proof of this part of the theorem. By assumption we have $\langle \mathrm{d}(A_{\mathbb{G}} - A_V);\delta\rangle = 0$ for every $\delta \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0,x_1)$. Thus, for $\delta \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D};x_0,x_1)$, by Lemmata 5.13 and 5.15, and by Sublemma 1 from the proof of Lemma 5.16, we have

$$\begin{aligned} 0 &= \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t))\,\mathrm{d}t - \int_{t_0}^{t_1} \mathbb{G}(\delta(t), \operatorname{grad} V \circ \gamma(t))\,\mathrm{d}t \\ &= -\int_{t_0}^{t_1} \mathbb{G}(\delta(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\gamma'(t) + \operatorname{grad} V \circ \gamma(t))\,\mathrm{d}t. \end{aligned}$$

Thus

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma \in \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^\perp),$$

showing the existence of λ as asserted in part (iii) by taking $\lambda = \overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma$.

(ii) $\Longrightarrow$ (iii) Note that part (ii), along with the condition that $\gamma'(t) \in \mathsf{D}_{\gamma'(t)}$ for $t \in [t_0,t_1]$, can be expressed as

$$\begin{aligned} &\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma = \lambda, \\ &P_{\mathsf{D}^\perp}(\gamma') = 0 \end{aligned}$$

for $\lambda \in \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^\perp)$. Differentiating the second of these equalities gives

$$(\overset{\mathrm{G}}{\nabla}_{\gamma'}P_{\mathsf{D}^\perp})(\gamma') + P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma') = 0.$$

From the first of the above equalities we have

$$(\overset{\mathrm{G}}{\nabla}_{\gamma'}P_{\mathsf{D}^\perp})(\gamma') = -P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma') = -\lambda + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma,$$

whence

$$\lambda = P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma.$$

Thus, by Lemma 2.36(ii),

$$0 = \overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma - \lambda = \overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma,$$

which is the desired conclusion.

(iii) $\Longrightarrow$ (i) Lemma 2.36(ii) and the fact that

$$P_{\mathsf{D}}(\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma) = \overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0$$

gives

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma \in \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^\perp).$$

Thus

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\gamma'(t) + \operatorname{grad} V \circ \gamma(t), \delta(t))\,\mathrm{d}t = 0$$

for every $\delta \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}; x_0, x_1)$. Using Sublemma 1 from the proof of Lemma 5.16 gives

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t))\,\mathrm{d}t - \int_{t_0}^{t_1} \mathbb{G}(\operatorname{grad} V \circ \gamma(t), \delta(t))\,\mathrm{d}t = 0$$

for every $\delta \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}; x_0, x_1)$. This part of the theorem now follows by Lemmata 5.13 and 5.15. ■

Based on the theorem, let us extend our notion of nonholonomic trajectories to arbitrary intervals.

5.19 Definition: (Nonholonomic trajectory on general interval) Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Let $I \subseteq \mathbb{R}$ be an interval. A curve $\gamma \colon I \to \mathsf{M}$ is a ***nonholonomic trajectory*** for Σ if it satisfies

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0. \qquad \bullet$$

Of course, nonholonomic trajectories for a C^r-constrained simple mechanical system are of class C^r.

5.20 Remark: (Nonholonomic trajectories and geodesics) We note that, when the potential energy function is zero, the nonholonomic trajectories are geodesics of the constrained connection, restricted to initial conditions in D. This observation seems to have been first made by Synge [1928], and further observations are made by Lewis [1998]. •

5.4. Constrained variational mechanics.

Now we consider a variational problem associated with nonholonomic mechanics. We begin with the definition.

5.21 Definition: (Constrained variational trajectory) Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Let $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$ and let $x_0, x_1 \in \mathsf{M}$. A curve $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1)$ is a ***constrained variational trajectory*** for Σ if

$$\langle \mathrm{d}(A_{\mathsf{G},\mathsf{D}} - A_{V,\mathsf{D}}); \delta\sigma(0)\rangle = 0, \qquad \sigma \colon J \to \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1) \text{ a variation of } \gamma. \qquad \bullet$$

Now we can state a few equivalent characterisations of a constrained variational trajectory.

5.22 Theorem: (Characterisation of constrained variational trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Let $t_0, t_1 \in \mathbb{R}$ with $t_0 < t_1$ and let $x_0, x_1 \in \mathsf{M}$. For $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1)$, the following statements are equivalent:*

(i) γ is a constrained variational trajectory;

(ii) at least one of the following conditions holds:

(a) there exists a nowhere zero $\lambda \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}^{\perp})$ such that

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda + S_{\mathsf{D}}^*(\gamma')(\lambda) = 0;$$

(b) $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$ *and there exists* $\lambda \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$ *such that*

$$\overset{\scriptscriptstyle G}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma - \overset{\scriptscriptstyle G}{\nabla}_{\gamma'}\lambda - S^*_{\mathsf{D}}(\gamma')(\lambda) = 0;$$

(iii) at least one of the following conditions holds:

(a) there exists a nowhere zero $\lambda \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$ *such that*

$$\begin{aligned} &F^*_{\mathsf{D}}(\gamma')(\lambda) = 0,\\ &\overset{\scriptscriptstyle \mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda);\end{aligned}$$

(b) $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$ *and there exists* $\lambda \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$ *such that*

$$\begin{aligned} &\overset{\scriptscriptstyle \mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = F^*_{\mathsf{D}}(\gamma')(\lambda),\\ &\overset{\scriptscriptstyle \mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G_{\mathsf{D}}(\gamma',\gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma + \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda).\end{aligned}$$

Proof: (i) $\Longrightarrow$ (ii) The proof of this implication has two cases.

Case I: γ *is a* D*-singular curve*

In this case, from Proposition 5.9 we have

$$\overset{\scriptscriptstyle G}{\nabla}_{\gamma'}\lambda + S^*_{\mathsf{D}}(\gamma')(\lambda) = 0$$

for some nowhere zero $\lambda \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$, as asserted in part (ii)(a).

Case II: γ *is a* D*-regular curve*

Let us denote

$$\begin{aligned} \Delta^\gamma_{\mathsf{D}}\colon \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1) &\to \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^\perp)\\ \delta &\mapsto P_{\mathsf{D}^\perp}(\overset{\scriptscriptstyle G}{\nabla}_{\gamma'}\delta) - S_{\mathsf{D}}(\delta,\gamma').\end{aligned}$$

As we saw in Lemma 5.4, the kernel of $\Delta^\gamma_{\mathsf{D}}$ is the tangent space at γ to $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$ as a submanifold of $\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$.

Let us now state a couple of technical lemmata. The relevance of these is only understood when they are used in last part of the proof of this part of the theorem. Thus they are best referred back to, rather than read in order.

1 Lemma: *If* $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$*, then the orthogonal complement to* $\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ *in* $\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ *with respect to the Dirichlet semi-inner product (3.3) is*

$$\{\eta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M}) \mid \ \eta(t) = \mu_0(t) + (t-t_0)\mu_1(t),\ t \in [t_0,t_1],$$
$$\textit{for some } \mu_0,\mu_1 \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M}) \textit{ satisfying } \overset{\scriptscriptstyle G}{\nabla}_{\gamma'}\mu_0 = \overset{\scriptscriptstyle G}{\nabla}_{\gamma'}\mu_1 = 0\}.$$

Proof: First let

$$\eta(t) = \mu_0(t) + (t-t_0)\mu_1(t), \qquad t \in [t_0,t_1],$$

where $\mu_0, \mu_1 \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$ satisfy

$$\overset{\mathbb{G}}{\nabla}_{\gamma'} \mu_0 = \overset{\mathbb{G}}{\nabla}_{\gamma'} \mu_1 = 0.$$

Also let $\delta \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1)$. We compute, using Sublemma 1 from the proof of Lemma 5.16,

$$\begin{aligned}\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \eta(t)) \, \mathrm{d}t &= \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \delta(t), \mu_1(t)) \, \mathrm{d}t \\ &= -\int_{t_0}^{t_1} \mathbb{G}(\delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \mu_1(t)) \, \mathrm{d}t = 0,\end{aligned}$$

and so η is orthogonal to δ with respect to the Dirichlet semi-inner product.

Now suppose that η is orthogonal to $\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1)$ with respect to the Dirichlet semi-inner product:

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \eta(t)) \, \mathrm{d}t = 0, \qquad \delta \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1).$$

We shall use the covariant derivative for distributional sections of $\gamma^* \mathsf{TM}$ as in Section 3.8. Thus we write the preceding equality, restricting to δ's in $\gamma^* \mathscr{D}(\mathsf{TM})$, as

$$\langle \overset{\mathbb{G}}{\nabla}_{\gamma'} \theta_\eta ; \overset{\mathbb{G}}{\nabla}_{\gamma'} \gamma^*(\mathbb{G}^\sharp \circ \delta) \rangle = 0, \qquad \delta \in \gamma^* \mathscr{D}(\mathsf{TM}).$$

Using the definition of the distributional covariant derivative, we have

$$\langle \overset{\mathbb{G}}{\nabla}{}^2_{\gamma'} \theta_\eta ; \gamma^*(\mathbb{G}^\sharp \circ \delta) \rangle = 0, \qquad \delta \in \gamma^* \mathscr{D}(\mathsf{TM}).$$

which gives $\overset{\mathbb{G}}{\nabla}{}^2_{\gamma'} \theta_\eta = 0$. Let $\theta_1 = \overset{\mathbb{G}}{\nabla}_{\gamma'} \theta_\eta$ and note that $\overset{\mathbb{G}}{\nabla}_{\gamma'} \theta_1 = 0$. Since $\eta \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$, we have $\theta_1 = \theta_{\mu_1}$ for some $\mu_1 \in \mathrm{L}^2([t_0, t_1]; \gamma^* \mathsf{TM})$. Note that the equality $\overset{\mathbb{G}}{\nabla}_{\gamma'} \theta_1 = 0$ means that

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \delta(\gamma(t)), \mu_1(t)) \, \mathrm{d}t = 0, \qquad \delta \in \gamma^* \mathscr{D}(\mathsf{TM}).$$

Since $\gamma^* \mathscr{D}(\mathsf{TM})$ is dense in $\mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1)$, this means that

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)} \delta(\gamma(t)), \mu_1(t)) \, \mathrm{d}t = 0, \qquad \delta \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1). \tag{5.9}$$

We claim that this implies that $\mu_1 \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$. To see this, let $\zeta_0 \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$ satisfy the initial value problem

$$\overset{\mathbb{G}}{\nabla}_\gamma \zeta_0 = \mu_1, \quad \zeta_0(t_0) = 0,$$

and let $\zeta_1 \in \mathsf{T}_\gamma \mathrm{H}^1([t_0, t_1]; \mathsf{M})$ satisfy the initial value problem

$$\overset{\mathbb{G}}{\nabla}_\gamma \zeta_1 = 0, \quad \zeta_1(t_1) = \zeta_0(t_1).$$

Denote

$$\delta(t) = \zeta_0(0) - \left(\frac{t-t_0}{t_1-t_0}\right)\zeta_1(t)$$

and note that $\delta(t_0)=0$ and $\delta(t_1)=0$. We also have

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\delta = \mu_1 - \frac{t_0}{t_1-t_0}\zeta_1.$$

We then calculate

$$\begin{aligned}\int_{t_0}^{t_1} \mathbb{G}\left(\zeta_1(t), \mu_1(t) - \frac{t_0}{t_1-t_0}\zeta_1(t)\right) \mathrm{d}t &= \int_{t_0}^{t_1} \mathbb{G}(\zeta_1(t), \overset{\mathbb{G}}{\nabla}_{\gamma(t)}\delta(t))\,\mathrm{d}t \\ &= -\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\zeta_1(t), \delta(t))\,\mathrm{d}t = 0,\end{aligned}$$

using Sublemma 1 from the proof of Lemma 5.16 and the definition of ζ_1. We also compute

$$\int_{t_0}^{t_1} \mathbb{G}\left(\mu_1(t), \mu_1(t) - \frac{t_0}{t_1-t_0}\zeta_1(t)\right) \mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\mu_1(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t))\,\mathrm{d}t = 0,$$

using (5.9). Combining the preceding two computations gives

$$\int_{t_0}^{t_1} \left\|\mu_1(t) - \frac{t_0}{t_1-t_0}\zeta_1(t)\right\|^2 \mathrm{d}t = 0,$$

and so $\mu_1 = \frac{t_0}{t_1-t_0}\zeta_1$, and thus $\mu_1 \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M})$, as claimed.

Now, with $\mu_1 \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M})$ prescribed, we claim that

$$\eta(t) = (t-t_0)\mu_1(t) + \mu_0(t)$$

for some μ_0 satisfying $\overset{\mathbb{G}}{\nabla}_{\gamma'}\mu_0 = 0$. Indeed, let μ_0 be the unique solution to the initial value problem

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\mu_0 = 0, \quad \mu_0(t_0) = \eta(t_0).$$

Then we have

$$\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}((t-t_0)\mu_1(t) + \mu_0(t)) = \mu_1(t) = \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\eta(t)$$

and

$$((t-t_0)\mu_1(t) + \mu_0(t))|_{t=t_0} = \mu_0(t) = \eta(t_0),$$

and so the sections over γ

$$t \mapsto (t-t_0)\mu_1(t) + \mu_0(t), \quad t \mapsto \eta(t)$$

satisfy the same initial value problem, and so are equal. ▼

2 Lemma: *Let $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M})$, let $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$, and let $\alpha \in \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{TM})$. Then there exists $\beta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ such that $\overset{\mathrm{G}}{\nabla}_{\gamma'}\beta = \alpha + \zeta$, where $\zeta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ satisfies $\overset{\mathrm{G}}{\nabla}_{\gamma'}\zeta = 0$, and such that*

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t),\alpha(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t),\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\beta(t))\,\mathrm{d}t.$$

Proof: Let $\beta_0 \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ satisfy $\overset{\mathrm{G}}{\nabla}_{\gamma'}\beta_0 = \alpha$ and define $\mu_0,\mu_1 \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ by their satisfying the initial value problems

$$\begin{aligned}
&\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\mu_0(t) = 0, \quad \mu_0(t_0) = -\beta_0(t_0),\\
&\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\mu_1(t) = 0, \quad \mu_1(t_1) = -\frac{1}{t_1-t_0}(\beta_0(t_1)+\mu_0(t_1)).
\end{aligned}$$

Then define

$$\beta(t) = \beta_0(t) + \mu_0(t) + (t-t_0)\mu_1(t), \qquad t \in [t_0,t_1],$$

so that $\beta(t_0) = 0$ and $\beta(t_1) = 0$. We have

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\beta = \overset{\mathrm{G}}{\nabla}_{\gamma'}\beta_0 + \mu_1 = \alpha + \mu_1,$$

and $\overset{\mathrm{G}}{\nabla}_{\gamma'}\mu_1 = 0$. Also note that, using Lemma 1, we have

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t),\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\beta(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t),\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\beta_0(t))\,\mathrm{d}t,$$

showing that β satisfies the assertion of the lemma. ▼

The following lemma will be used to understand the kernel of $\Delta^\gamma_{\mathsf{D}}$ as the orthogonal complement to the image of $\Delta^{\gamma,*}_{\mathsf{D}}$ with respect to the Dirichlet semi-inner product.

3 Lemma: *If $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ and if $\lambda \in \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^\perp)$, then*

$$\int_{t_0}^{t_1} \mathbb{G}(\Delta^\gamma_{\mathsf{D}}(\delta)(t),\lambda(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t),\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\beta_\lambda(t))\,\mathrm{d}t,$$

where

(i) $\beta_\lambda \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ satisfies $\overset{\mathrm{G}}{\nabla}_{\gamma'}\beta_\lambda = \lambda + \alpha_\lambda + \eta_\lambda$, and where

(ii) $\eta_\lambda \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ satisfies $\overset{\mathrm{G}}{\nabla}_{\gamma'}\eta_\lambda = 0$ and

(iii) $\alpha_\lambda \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ satisfies $\overset{\mathrm{G}}{\nabla}_{\gamma'}\alpha_\lambda = S^*_{\mathsf{D}}(\gamma')(\lambda)$.

Proof: We have

$$\begin{aligned}
\int_{t_0}^{t_1} \mathbb{G}(\Delta^\gamma_{\mathsf{D}}(\delta)(t),\lambda(t))\,\mathrm{d}t &= \int_{t_0}^{t_1} \Big(\mathbb{G}(P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t)),\lambda(t)) - \mathbb{G}(S_{\mathsf{D}}(\delta(t),\gamma'(t)),\lambda(t))\Big)\,\mathrm{d}t\\
&= \int_{t_0}^{t_1} \Big(\mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t),\lambda(t)) - \mathbb{G}(S^*_{\mathsf{D}}(\gamma'(t))(\lambda(t)),\delta(t))\Big)\,\mathrm{d}t\\
&= \int_{t_0}^{t_1} \Big(\mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t),\lambda(t)) - \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\alpha_\lambda(t),\delta(t))\Big)\,\mathrm{d}t.
\end{aligned}$$

Now, by Sublemma 1 from the proof of Lemma 5.16, we have

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\alpha_\lambda(t), \delta(t))\,\mathrm{d}t = -\int_{t_0}^{t_1} \mathbb{G}(\alpha_\lambda(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t))\,\mathrm{d}t,$$

which gives

$$\int_{t_0}^{t_1} \mathbb{G}(\Delta^\gamma_{\mathsf{D}}(\delta)(t), \lambda(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}\left(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \lambda(t) + \alpha_\lambda(t)\right)\,\mathrm{d}t.$$

By Lemma 2, there exists $\beta_\lambda \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ satisfying

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\beta_\lambda = \lambda + \alpha_\lambda + \eta_\lambda,$$

where $\overset{\mathrm{G}}{\nabla}_{\gamma'}\eta_\lambda = 0$, and also satisfying

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \lambda(t) + \alpha_\lambda(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\beta_\lambda(t))\,\mathrm{d}t.$$

Thus

$$\int_{t_0}^{t_1} \mathbb{G}(\Delta^\gamma_{\mathsf{D}}(\delta)(t), \lambda(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}\left(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\beta_\lambda(t)\right)\,\mathrm{d}t,$$

as desired. ▼

Now note that, by Lemma 5.16, if γ is a D-regular constrained variational trajectory, then there exists $\beta_V \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ such that $\overset{\mathrm{G}}{\nabla}_{\gamma'}\beta_V = \operatorname{grad} V \circ \gamma$ and such that

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\,\mathrm{d}t = 0$$

for every $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$. By Lemma 2, there exists $\tau_{\gamma,V} \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ such that

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\tau_{\gamma,V} = \gamma' + \beta_V + \zeta_{\gamma,V},$$

where $\overset{\mathrm{G}}{\nabla}_{\gamma'}\zeta_{\gamma,V} = 0$, and such that

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{D}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\tau_{\gamma,V}(t))\,\mathrm{d}t$$

for all $\delta \in \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$. Thus $\tau_{\gamma,V}$ is orthogonal to $\ker(\Delta^\gamma_{\mathsf{D}})$ in $\mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ with respect to the Dirichlet inner product, and so in $\operatorname{image}(\Delta^{\gamma,*}_{\mathsf{D}})$. Thus there exists $\lambda \in \mathrm{L}^2([t_0,t_1];\gamma^*\mathsf{D}^\perp)$ such that $\tau_{\gamma,V} = \Delta^{\gamma,*}_{\mathsf{D}}(\lambda)$. Note that $\Delta^{\gamma,*}_{\mathsf{D}}$ is defined by the condition that

$$\int_{t_0}^{t_1} \mathbb{G}(\Delta^\gamma_{\mathsf{D}}(\delta)(t), \lambda(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\Delta^{\gamma,*}_{\mathsf{D}}(\lambda)(t))\,\mathrm{d}t,$$

and so, from Lemma 3, we have

$$\tau_{\gamma,V} = \Delta_{\mathsf{D}}^{\gamma,*}(\lambda) = \beta_\lambda,$$

where $\beta_\lambda \in \mathsf{T}_\gamma \mathrm{H}^1([t_0,t_1];\mathsf{M};x_0,x_1)$ satisfies

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\beta_\lambda = \lambda + \alpha_\lambda + \eta_\lambda,$$

and where $\eta_\lambda \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ satisfies $\overset{\mathbb{G}}{\nabla}_{\gamma'}\eta_\lambda = 0$ and $\alpha_\lambda \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M})$ satisfies $\overset{\mathbb{G}}{\nabla}_{\gamma'}\alpha_\lambda = S^*_{\mathsf{D}}(\gamma')(\lambda)$.

The relation $\tau_{\gamma,V} = \beta_\lambda$ implies

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\tau_{\gamma,V} = \overset{\mathbb{G}}{\nabla}_{\gamma'}\beta_\lambda \quad\Longrightarrow\quad \gamma' + \beta_V + \zeta_{\gamma,V} = \lambda + \alpha_\lambda + \eta_\lambda. \tag{5.10}$$

Therefore,

$$\begin{aligned}\gamma' &= P_{\mathsf{D}} \circ (\alpha_\lambda + \eta_\lambda - \beta_V - \zeta_{\gamma,V}),\\ \lambda &= P_{\mathsf{D}^\perp} \circ (\beta_V - \zeta_{\gamma,V} - \alpha_\lambda - \eta_\lambda),\end{aligned}$$

and so $\gamma \in \mathrm{H}^2([t_0,t_1];\mathsf{M})$ and $\lambda \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$. Thus we can differentiate the right-hand side of (5.10) to get

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma - \overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda - S^*_{\mathsf{D}}(\gamma')(\lambda) = 0,$$

giving this part of the theorem.

(ii) $\Longleftrightarrow$ (iii) We shall consider two cases, the D-singular case and the D-regular case. In both cases, we shall compute the projection of the equations from part (ii) to both D and $\mathsf{D}^\perp$. Clearly the original equations hold if and only if both of the projected equations hold.

Case I: γ is a D-singular curve

The equation we work with in this case is

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda + S^*_{\mathsf{D}}(\gamma')(\lambda) = 0, \tag{5.11}$$

for nowhere zero $\lambda \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{D}^\perp)$.

Let us take the projection of this equation onto D. Let $X, Y \in \Gamma^\infty(\mathsf{D})$ and $\alpha \in \Gamma^\infty(\mathsf{D}^\perp)$. We then have

$$\begin{aligned}&\mathbb{G}(\alpha, Y) = 0\\ \Longrightarrow\ &\mathbb{G}(\overset{\mathbb{G}}{\nabla}_X\alpha, Y) + \mathbb{G}(\alpha, \overset{\mathbb{G}}{\nabla}_X Y) = 0.\end{aligned}$$

We may then compute

$$\begin{aligned}\mathbb{G}(P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_X\alpha + S^*_{\mathsf{D}}(X)(\alpha)), Y) &= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_X\alpha, Y) + \mathbb{G}(S^*_{\mathsf{D}}(X)(\alpha), Y)\\ &= -\mathbb{G}(\alpha, \overset{\mathbb{G}}{\nabla}_X Y) + \mathbb{G}(\alpha, S_{\mathsf{D}}(Y,X))\\ &= -\mathbb{G}(\alpha, \overset{\mathbb{G}}{\nabla}_X Y) + \mathbb{G}(\alpha, P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_Y X))\\ &= -\mathbb{G}(\alpha, \overset{\mathbb{G}}{\nabla}_X Y - \overset{\mathbb{G}}{\nabla}_Y X)\\ &= \mathbb{G}(\alpha, F_{\mathsf{D}}(Y,X)) = \mathbb{G}(F^*_{\mathsf{D}}(X)(\alpha), Y).\end{aligned}$$

Thus

$$P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_X\alpha) + P_{\mathsf{D}}(S^*_{\mathsf{D}}(X)(\alpha)) = F^*_{\mathsf{D}}(X)(\alpha), \qquad X \in \Gamma^\infty(\mathsf{D}),\ \alpha \in \Gamma^\infty(\mathsf{D}^\perp). \tag{5.12}$$

This gives

$$P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda) + P_{\mathsf{D}}(S^*_{\mathsf{D}}(\gamma')(\lambda)) = F^*_{\mathsf{D}}(\gamma')(\lambda),$$

which is the first of the equations from part (iii)(a).

Now let us compute the projection of (5.11) onto $\mathsf{D}^\perp$. First, by Lemma 2.36(ii), we have

$$P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda) = \overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda. \tag{5.13}$$

Now, for $X \in \Gamma^\infty(\mathsf{D})$ and $\alpha, \beta \in \Gamma^\infty(\mathsf{D}^\perp)$, we have

$$\begin{aligned} &\mathbb{G}(\alpha, X) = 0 \\ \Longrightarrow\ &\mathbb{G}(\overset{\mathbb{G}}{\nabla}_\beta\alpha, X) + \mathbb{G}(\alpha, \overset{\mathbb{G}}{\nabla}_\beta X) = 0. \end{aligned}$$

This then can be used to compute

$$\begin{aligned} \mathbb{G}(P_{\mathsf{D}^\perp}(S^*_{\mathsf{D}}(X)(\alpha)), \beta) &= \mathbb{G}(S^*_{\mathsf{D}}(X)(\alpha), \beta) = \mathbb{G}(\alpha, S_{\mathsf{D}}(\beta, X)) \\ &= \mathbb{G}(\alpha, P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_\beta X)) = \mathbb{G}(\alpha, \overset{\mathbb{G}}{\nabla}_\beta X) = -\mathbb{G}(\overset{\mathbb{G}}{\nabla}_\beta\alpha, X) \\ &= -\mathbb{G}(P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_\beta\alpha), X) = -\mathbb{G}(S_{\mathsf{D}^\perp}(\beta, \alpha), X) \\ &= -\frac{1}{2}\mathbb{G}(G_{\mathsf{D}^\perp}(\beta, \alpha), X) - \frac{1}{2}\mathbb{G}(F_{\mathsf{D}^\perp}(\beta, \alpha), X) \\ &= -\frac{1}{2}\mathbb{G}(G^\star_{\mathsf{D}^\perp}(X)(\alpha), \beta) - \frac{1}{2}\mathbb{G}(F^\star_{\mathsf{D}^\perp}(X)(\alpha), \beta), \end{aligned}$$

and so we conclude that

$$P_{\mathsf{D}^\perp}(S^*_{\mathsf{D}}(X)(\alpha)) = -\frac{1}{2}G^\star_{\mathsf{D}^\perp}(X)(\alpha) - \frac{1}{2}F^\star_{\mathsf{D}^\perp}(X)(\alpha), \qquad X \in \Gamma^\infty(\mathsf{D}),\ \alpha \in \Gamma^\infty(\mathsf{D}^\perp).$$

This gives

$$P_{\mathsf{D}^\perp}(S^*_{\mathsf{D}}(\gamma')(\lambda)) = -\frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) - \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda).$$

Combining this with (5.13) gives the second of equations from part (iii)(a).

Case II: γ is a D-regular curve

In this case, the equation we work with is

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma = \overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda + S^*_{\mathsf{D}}(\gamma')(\lambda), \tag{5.14}$$

for $\lambda \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}^\perp)$.

Let us first project this equation onto D. An application of P_{D} to (5.14) gives

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda) + P_{\mathsf{D}}(S^*_{\mathsf{D}}(\gamma')(\lambda)),$$

using Lemma 2.36(ii). As we saw in the proof of the D-singular case above,

$$P_{\mathsf{D}}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda) + P_{\mathsf{D}}(S^*_{\mathsf{D}}(\gamma')(\lambda)) = F^*_{\mathsf{D}}(\gamma')(\lambda).$$

Combining the preceding two equations gives the first of the equations from part (iii)(b).

Next, an application of $P_{\mathsf{D}^\perp}$ to (5.14) gives

$$\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma - P_{\mathsf{D}^\perp}(S^*_{\mathsf{D}}(\gamma')(\lambda)),$$

using Lemma 2.36(ii). By Lemma 2.36(i) and (vi), we have

$$P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma') = S_{\mathsf{D}}(\gamma', \gamma') = \frac{1}{2} G_{\mathsf{D}}(\gamma', \gamma').$$

As we saw in the proof of the D-singular case above,

$$P_{\mathsf{D}^\perp}(S^*_{\mathsf{D}}(\gamma')(\lambda)) = -\frac{1}{2} G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) - \frac{1}{2} F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda).$$

This, combined with the preceding two equations, gives the second of equations from part (iii)(b).

(ii) $\implies$ (i) Again we consider two cases.

Case I: γ is a D-singular curve

In this case, the conclusion follows immediately from Proposition 5.9 and the definition of D-singular curves.

Case II: γ is a D-regular curve

We shall essentially reverse the computations above for the converse implication, but give the details for clarity and completeness.

Since $\lambda \in \mathrm{H}^1([t_0, t_1]; \gamma^*\mathsf{D}^\perp)$, by Lemma 3 we have

$$\int_{t_0}^{t_1} \mathbb{G}(\Delta^\gamma_{\mathsf{D}}(\delta)(t), \lambda(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\beta_\lambda(t))\,\mathrm{d}t$$

for all $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D}; x_0, x_1)$, where

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\beta_\lambda = \lambda + \alpha_\lambda + \zeta_\lambda,$$

and

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\alpha_\lambda = S^*_{\mathsf{D}}(\gamma')(\lambda), \quad \overset{\mathbb{G}}{\nabla}_{\gamma'}\zeta_\lambda = 0.$$

Now let $\beta_V \in \mathsf{T}_\gamma\mathrm{H}^1([t_0, t_1]; \mathsf{M})$ be such that $\overset{\mathbb{G}}{\nabla}_{\gamma'}\beta_V = \operatorname{grad} V \circ \gamma$. By Lemma 2, there exists $\tau_{\gamma,V} \in \mathsf{T}_\gamma\mathrm{H}^1([t_0, t_1]; \mathsf{M}; x_0, x_1)$ such that

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\tau_{\gamma,V} = \gamma' + \beta_V + \zeta_{\gamma,V},$$

where $\overset{\mathbb{G}}{\nabla}_{\gamma'}\zeta_{\gamma,V} = 0$, and such that

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathsf{D}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\,\mathrm{d}t = \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\tau_{\gamma,V}(t))\,\mathrm{d}t$$

for all $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$.

By hypothesis, we have

$$\begin{aligned}
&\overset{\mathbb{G}}{\nabla}_{\gamma'}(\gamma' + \beta_V - \lambda - \alpha_\lambda) = 0,\\
\implies &\overset{\mathbb{G}}{\nabla}_{\gamma'}\tau_{\gamma,V} - \overset{\mathbb{G}}{\nabla}_{\gamma'}\beta_\lambda - \zeta_{\gamma,V} + \zeta_\lambda = 0.
\end{aligned}$$

Since

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\zeta_{\gamma,V}(t) - \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\zeta_\lambda(t))\,\mathrm{d}t = 0$$

for all $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$ by Lemma 1, we have

$$\begin{aligned}
0 &= \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\tau_{\gamma,V}(t) - \overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\beta_\lambda(t))\,\mathrm{d}t\\
&= \int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\,\mathrm{d}t - \int_{t_0}^{t_1} \mathbb{G}(\Delta^\gamma_{\mathsf{D}}(\delta)(t), \lambda(t))\,\mathrm{d}t
\end{aligned}$$

for all $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$. For $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$ we have $\Delta^\gamma_{\mathsf{D}}(\delta) = 0$, giving

$$\int_{t_0}^{t_1} \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\delta(t), \gamma'(t) + \beta_V(t))\,\mathrm{d}t = 0$$

for all $\delta \in \mathsf{T}_\gamma\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$. By Lemma 5.16 this part of the result follows. ■

We note that, as illustrated in the proof, the equivalence of parts (ii) and (iii) is obtained by projecting the equations from part (ii) into D and $\mathsf{D}^\perp$.

We can use the previous result to give sense to the following notion of a constrained variational trajectory defined on an arbitrary interval.

5.23 Definition: (Constrained variational trajectory on general interval) Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Let $I \subseteq \mathbb{R}$ be an interval.

(i) A locally absolutely continuous curve $\gamma\colon I \to \mathsf{M}$ is a ***D-singular constrained variational trajectory*** for Σ if there exists a nowhere zero locally absolutely continuous $\lambda\colon I \to \mathsf{D}^\perp$ such that $\pi_{\mathsf{D}^\perp} \circ \lambda = \gamma$ and such that

$$\begin{aligned}
&F^*_{\mathsf{D}}(\gamma')(\lambda) = 0,\\
&\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda).
\end{aligned}$$

(ii) A curve $\gamma\colon I \to \mathsf{M}$ is a ***D-regular constrained variational trajectory*** for Σ if there exists a locally absolutely continuous $\lambda\colon I \to \mathsf{D}^\perp$ such that $\pi_{\mathsf{D}^\perp} \circ \lambda = \gamma$ and such that

$$\begin{aligned}
&\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = F^*_{\mathsf{D}}(\gamma')(\lambda),\\
&\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G_{\mathsf{D}}(\gamma',\gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma + \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda).
\end{aligned}$$

•

5.24 Definition: (Adjoint field of a constrained variational trajectory) If $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ is a C^r-constrained simple mechanical system, $r \in \{\infty, \omega\}$, and if the pair (γ, λ), $\gamma \colon I \to \mathsf{M}$ and $\lambda \colon I \to \mathsf{D}^{\perp}$, satisfy the conditions for γ to be a (either D-singular or D-regular) constrained variational trajectory, then λ is called the ***adjoint field*** along γ. •

A D-regular constrained variational trajectory for a C^r-constrained simple mechanical system is of class C^r. However, there is no *a priori* requirement that D-singular constrained variational trajectories be anything but absolutely continuous.

5.25 Remark: (Constrained variational trajectories that are both D-singular and D-regular) Note that it is possible that a constrained variational trajectory will simultaneously satisfy both of the conditions (ii)(a) and (ii)(b) (or, equivalently, conditions (iii)(a) and (iii)(b)) of the Theorem 5.22. Moreover, all possible situations can be actualised.

1. It may be the case that no trajectories are D-singular. This happens, for example, if $\mathsf{D} = \mathsf{TM}$. Indeed, if $\mathsf{D} = \mathsf{TM}$, then $\hat{F}^*_{\mathsf{D}} = 0$ and so there can be no nowhere zero sections λ of $\mathsf{D}^{\perp}$ along a curve γ satisfying the conditions of Theorem 5.22(iii) for D-singular trajectories.
2. It may be the case that some constrained variational trajectories are both D-singular and D-regular, but others are just D-singular or just D-regular.
3. It may be the case that all constrained variational trajectories are both D-singular and D-regular. For example, one can see that this situation arises when
   (a) $\mathsf{D} \subset \mathsf{TM}$ is integrable,
   (b) D is geodesically invariant,
   (c) $V = 0$.

This suggests introducing the notion of a curve $\gamma \in \mathrm{H}^1([t_0, t_1]; \mathsf{M}; \mathsf{D})$ being ***strictly* D*-singular***, meaning that it is D-singular but not D-regular. This is indeed an interesting notion to explore, but we shall not do so here. •

6. Connections to sub-Riemannian geometry

As we mentioned in the introduction, there is a connection between constrained variational mechanics and sub-Riemannian geometry. In this section we establish this connection and make some comments about sub-Riemannian geometry using our approach. We note that work related to our approach can be found in the paper of Langerok [2003].

We shall give an overview of sub-Riemannian geometry, and we refer to [e.g., Agrachev, Barilari, and Boscain 2018] for details that we omit.

In our development we shall make use of an adaptation of a part of the theory of time-varying vector fields presented by Jafarpour and Lewis [2014]. Specifically, for an interval $I \subseteq \mathbb{R}$ and a smooth vector bundle $\pi\colon \mathsf{E} \to \mathsf{M}$, we shall denote by $\mathrm{L}^2\Gamma^\infty(I;\mathsf{E})$ the collection of time-varying sections $\xi\colon I \times \mathsf{M} \to \mathsf{E}$ that are smooth for each fixed $t \in I$ and for which the induced mapping $\hat{\xi}\colon I \to \Gamma^\infty(\mathsf{E})$ defined by $\hat{\xi}(t)(x) = \xi(t,x)$ is square Bochner integrable if $\Gamma^\infty(\mathsf{E})$ is equipped with its weak C^∞-topology.

6.1. Sub-Riemannian geometry. Sub-Riemannian geometry is the study of a smooth manifold M equipped with a smooth subbundle $\mathsf{D} \subseteq \mathsf{TM}$ and a smooth fibre metric $\mathbb{G}_\mathsf{D}$ on D, called a ***sub-Riemannian metric***. The triple $(\mathsf{M},\mathsf{D},\mathbb{G}_\mathsf{D})$ is called a ***sub-Riemannian manifold***. It is possible and interesting to formulate sub-Riemannian geometry in case D is not a subbundle; however, we shall not do this but refer to [Agrachev, Barilari, and Boscain 2018] for considerations of this nature. The constant rank situation we consider is called by Agrachev, Barilari, and Boscain a "regular" sub-Riemannian manifold. Some authors tacitly only consider this regular case.

A sub-Riemannian metric $\mathbb{G}_\mathsf{D}$ on a distribution D defines a D-valued vector bundle morphism $\mathbb{G}^\sharp_\mathsf{D}\colon \mathsf{T}^*\mathsf{M} \to \mathsf{TM}$ by requiring that

$$\mathbb{G}_\mathsf{D}(\mathbb{G}^\sharp_\mathsf{D}(\alpha_x), u_x) = \langle \alpha_x; u_x\rangle, \qquad u_x \in \mathsf{D}_x, \tag{6.1}$$

where $\alpha_x \in \mathsf{T}^*_x\mathsf{M}$. Correspondingly, we can associate to $\mathbb{G}_\mathsf{D}$ a tensor field of type $(2,0)$ on M, denoted by $\mathbb{G}^{-1}_\mathsf{D}$ and defined by

$$\mathbb{G}^{-1}_\mathsf{D}(\alpha_x, \beta_x) = \langle \alpha_x; \mathbb{G}^\sharp_\mathsf{D}(\beta_x)\rangle. \tag{6.2}$$

Note that $\mathbb{G}^{-1}_\mathsf{D}$ is symmetric since

$$\begin{aligned}\mathbb{G}^{-1}_\mathsf{D}(\alpha_x, \beta_x) &= \langle \alpha_x; \mathbb{G}^\sharp_\mathsf{D}(\beta_x)\rangle = \mathbb{G}_\mathsf{D}(x)(\mathbb{G}^\sharp_\mathsf{D}(\alpha_x), \mathbb{G}^\sharp_\mathsf{D}(\beta_x))\\ &= \mathbb{G}_\mathsf{D}(x)(\mathbb{G}^\sharp_\mathsf{D}(\beta_x), \mathbb{G}^\sharp_\mathsf{D}(\alpha_x)) = \langle \alpha_x; \mathbb{G}^\sharp_\mathsf{D}(\beta_x)\rangle = \mathbb{G}^{-1}_\mathsf{D}(\beta_x, \alpha_x).\end{aligned}$$

Note that $\mathbb{G}^{-1}_\mathsf{D}$ is positive-semidefinite, and is positive-definite only when $\mathsf{D} = \mathsf{TM}$.

This correspondence between "distributions with a sub-Riemannian metric" and "symmetric positive-semidefinite constant rank $(2,0)$-tensor fields" is one-to-one as, given a symmetric positive-semidefinite constant rank $(2,0)$-tensor field $\mathbb{H}$, we define a distribution $\mathsf{D}_x = \mathrm{image}(\mathbb{H}^\sharp_x)$ and an inner product $\mathbb{G}_\mathsf{D}$ on D_x by

$$\mathbb{G}_\mathsf{D}(u_x, v_x) = \mathbb{H}(\alpha_x, \beta_x),$$

where $\mathbb{H}^\sharp(\alpha_x) = u_x$ and $\mathbb{H}^\sharp(\beta_x) = v_x$. One can show that this gives a well-defined inner product, and the corresponding distribution and sub-Riemannian metric inherit the regularity of the $(2,0)$-tensor field $\mathbb{H}$.

Typically one assumes that M is connected and that D is bracket generating since this ensures that, given $x_0, x_1 \in \mathsf{M}$, $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1) \neq \varnothing$ [Chow 1940/1941]. We can then define the ***length action*** by

$$\begin{aligned} \ell_{\mathbb{G}_\mathsf{D}} \colon \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D}) &\to \mathbb{R} \\ \gamma &\mapsto \int_{t_0}^{t_1} \sqrt{\mathbb{G}_\mathsf{D}(\gamma'(t),\gamma'(t))}\,\mathrm{d}t. \end{aligned}$$

Then we make M into a metric space by the metric

$$\mathrm{d}_{\mathbb{G}_\mathsf{D}}(x_1,x_2) = \inf\{\ell_{\mathbb{G}_\mathsf{D}}(\gamma) \mid \gamma \in \mathrm{H}^1([0,1];\mathsf{M};\mathsf{D})\}.$$

We say that $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$ is a ***sub-Riemannian geodesic*** if there is a partition

$$t_0 = s_0 < s_1 < \cdots < s_k = t_1$$

of $[t_0,t_1]$ such that

$$\ell_{\mathbb{G}_\mathsf{D}}(\gamma|[s_{j-1},s_j]) = \mathrm{d}_{\mathbb{G}_\mathsf{D}}(\gamma(s_{j-1}),\gamma(s_j)), \qquad j \in \{1,\dots,k\}.$$

To determine sub-Riemannian geodesics, one first determines the ***extremals*** which are the critical points of the length function $\ell_{\mathbb{G}_\mathsf{D}}$ on $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$. One can imagine, by comparing the length action to the kinetic energy action, that there might be some correspondence between sub-Riemannian geodesics and constrained variational trajectories. For now, we restrict to the case of sub-Riemannian geodesics, and define the ***energy action*** in this setting by

$$\begin{aligned} A_{\mathbb{G}_\mathsf{D}} \colon \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D}) &\to \mathbb{R} \\ \gamma &\mapsto \int_{t_0}^{t_1} \mathbb{G}_\mathsf{D}(\gamma'(t),\gamma'(t))\,\mathrm{d}t. \end{aligned}$$

We then have the following result.

6.1 Lemma: (Relationship between minimisers of the length action and of the energy action) *Let* $(\mathsf{M},\mathsf{D},\mathbb{G}_\mathsf{D})$ *be a sub-Riemannian manifold, let* $t_0,t_0',t_1,t_1' \in \mathbb{R}$ *satisfy* $t_0 < t_1$ *and* $t_0' < t_1'$*, and let* $x_0,x_1 \in \mathsf{M}$ *be distinct. Then the following statements hold:*

(i) if $\tau\colon [t_0',t_1'] \to [t_0,t_1]$ *is a reparameterisation,*[8] *then* $\ell_{\mathbb{G}_\mathsf{D}}(\gamma\circ\tau) = \ell_{\mathbb{G}_\mathsf{D}}(\gamma)$ *for every* $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$*;*

(ii) for $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$*, the following statements are equivalent:*

- *(a)* γ *is a minimiser for* $A_{\mathbb{G}_\mathsf{D}}|\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$*;*
- *(b)* γ *is a minimiser for* $\ell_{\mathbb{G}_\mathsf{D}}|\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$ *and the function* $t \mapsto \|\gamma'(t)\|_{\mathbb{G}_\mathsf{D}}$ *is constant.*

Proof: (i) Note that γ is absolutely continuous if and only if $f\circ\gamma$ is absolutely continuous for every $f \in \mathrm{C}^\infty(\mathsf{M})$.[9] Therefore, $\gamma\circ\tau$ is absolutely continuous if and only if $f\circ\gamma\circ\tau$ is absolutely continuous. Thus the absolute continuity of $\gamma\circ\tau$ follows from the fact that the composition

[8] Meaning that τ is Lipschitz and monotonically increasing.

[9] Indeed, this makes a nice definition of absolute continuity.

of an absolutely continuous and Lipschitz function is absolutely continuous [cf. Ziemer 1989, Theorem 2.1.11]. Since τ is continuous, we also have $f \circ \gamma \circ \tau \in \mathrm{L}^2([t_0', t_1']; \mathbb{R})$. Since τ is Lipschitz, by Rademacher's Theorem [Federer 1969, Theorem 3.1.5] τ is differentiable almost everywhere, and its derivative is bounded by any Lipschitz constant for τ. Thus, since $F \circ \gamma' \in \mathrm{L}^2([t_0, t_1]; \mathbb{R})$, we have

$$F \circ (\gamma \circ \tau)' = F \circ (\gamma' \circ \tau)\tau' \in \mathrm{L}^2([t_0, t_1]; \mathbb{R}).$$

Thus $\gamma \circ \tau \in \mathrm{H}^1([t_0', t_1']; \mathsf{M}; \mathsf{D})$.

Finally,

$$\begin{aligned} \ell_{\mathbb{G}_\mathsf{D}}(\gamma \circ \tau) &= \int_{t_0'}^{t_1'} \sqrt{\mathbb{G}((\gamma \circ \tau)'(t), (\gamma \circ \tau)'(t))}\,\mathrm{d}t \\ &= \int_{t_0'}^{t_1'} \tau'(t)\sqrt{\mathbb{G}(\gamma' \circ \tau(t), \gamma' \circ \tau(t))}\,\mathrm{d}t \\ &= \int_{t_0}^{t_1} \sqrt{\mathbb{G}(\gamma'(s), \gamma'(s))}\,\mathrm{d}s = \ell_{\mathbb{G}_\mathsf{D}}(\gamma), \end{aligned}$$

by the change of variable $s = \tau(t)$.

(ii) By Cauchy–Schwarz, we have

$$\begin{aligned} \ell_{\mathbb{G}_\mathsf{D}}(\gamma) &= \left(\int_{t_0}^{t_1} \sqrt{\mathbb{G}_\mathsf{D}(\gamma'(t), \gamma'(t))}\,\mathrm{d}t\right)^2 \\ &\le \left(\int_{t_0}^{t_1} \mathbb{G}_\mathsf{D}(\gamma'(t), \gamma'(t))\,\mathrm{d}t\right)\left(\int_{t_0}^{t_1} \mathrm{d}t\right) \\ &= 2A_{\mathbb{G}_\mathsf{D}}(\gamma)(t_1 - t_0). \end{aligned}$$

Also, we have equality in the above inequality if and only if the functions

$$t \mapsto \sqrt{\mathbb{G}_\mathsf{D}(\gamma'(t), \gamma'(t))}, \quad t \mapsto 1$$

are collinear, i.e., if and only if the function $t \mapsto \|\gamma'(t)\|_{\mathbb{G}_\mathsf{D}}$ is constant. Thus the energy action and the length action agree, up to a constant, on curves parameterised with a constant speed. Since the length action is independent of reparameterisation, this part of the lemma follows. ∎

The lemma allows us to determine extremals for the length action by computing critical points for the energy action. A common way of doing this is to utilise the Maximum Principle of Pontryagin. In this formulation, one works with the ***Hamiltonian***

$$\begin{aligned} H_{\mathsf{D},\mu_0} \colon \mathsf{D} \oplus \mathsf{T}^*\mathsf{M} &\to \mathbb{R} \\ v \oplus \alpha &\mapsto \langle \alpha; v\rangle - \frac{\mu_0}{2}\mathbb{G}_\mathsf{D}(v, v), \end{aligned}$$

for $\mu_0 \in \mathbb{R}_{\ge 0}$. We then define the ***maximum Hamiltonian*** by

$$H^{\max}_{\mathsf{D},\mu_0}(\alpha) = \sup\{H_{\mathsf{D},\mu_0}(v \oplus \alpha) \mid v \in \mathsf{D}_{\pi_{\mathsf{T}^*\mathsf{M}}(\alpha)}\}.$$

The following result gives the maximum Hamiltonian.

6.2 Lemma: (The maximum Hamiltonian) *For a sub-Riemannian manifold* $(\mathsf{M}, \mathsf{D}, \mathbb{G}_{\mathsf{D}})$,

$$H^{\max}_{\mathsf{D},\mu_0}(\alpha) = \begin{cases} \frac{1}{2\mu_0}\mathbb{G}_{\mathsf{D}}^{-1}(\alpha,\alpha), & \mu_0 \neq 0, \\ 0, & \mu_0 = 0,\ \alpha \in \Lambda(\mathsf{D}), \\ \infty, & \alpha \notin \Lambda(\mathsf{D}),\ \mu_0 = 0. \end{cases}$$

Proof: Note that we can write

$$H_{\mathsf{D},\mu_0}(v \oplus \alpha) = \mathbb{G}_{\mathsf{D}}\left(\mathbb{G}^{\sharp}_{\mathsf{D}}(\alpha) - \frac{\mu_0}{2}v, v\right).$$

To extremise this as a function of v, we differentiate in the direction of $u \in \mathsf{D}_{\pi_{\mathsf{D}}(v)}$ to get

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} H_{\mathsf{D},\mu_0}((v+tu)\oplus\alpha) = \mathbb{G}_{\mathsf{D}}\left(\mathbb{G}^{\sharp}_{\mathsf{D}}(\alpha) - \mu_0 v, u\right).$$

This vanishes for every $u \in \mathsf{D}_{\pi_{\mathsf{D}}(v)}$ if and only if $\mu_0 v = \mathbb{G}^{\sharp}_{\mathsf{D}}(\alpha)$. One also checks that, when $\alpha \neq 0$, the second derivative is negative-definite, and so this prescription of v gives a maximum of H_{D,μ_0}. Substituting v into H_{D,μ_0} gives the desired conclusion when $\mu_0 \neq 0$. Let us now consider the case of $\mu_0 = 0$. When $\alpha \in \Lambda(\mathsf{D})$, then $H_{\mathsf{D},0}(v \oplus \alpha) = 0$, giving the conclusion in this case. Finally, when $\alpha \notin \Lambda(\mathsf{D})$, then $H_{\mathsf{D},0}$ is unbounded both above and below, giving the result in this case. ■

According to the lemma, let us denote $\mathsf{P}_{\mathsf{D},\mu_0} \subseteq \mathsf{T}^*\mathsf{M}$ by

$$\mathsf{P}_{\mathsf{D},\mu_0} = \begin{cases} \mathsf{T}^*\mathsf{M}, & \mu_0 \neq 0, \\ \Lambda(\mathsf{D}), & \mu_0 = 0. \end{cases}$$

The idea is that $H^{\max}_{\mathsf{D},\mu_0}$ is a well-defined $\mathbb{R}$-valued function on $\mathsf{P}_{\mathsf{D},\mu_0}$. Let us denote by $\rho_{\mathsf{D},\mu_0} \colon \mathsf{P}_{\mathsf{D},\mu_0} \to \mathsf{M}$ the restriction of the cotangent bundle projection.

The Maximum Principle of Pontryagin (originating in [Pontryagin, Boltyanskiĭ, Gamkrelidze, and Mishchenko 1961]) gives necessary conditions for a length minimising curve, according to the next definition. We do not prove that the conditions given are, indeed, necessary conditions for minimisation of the energy action as this is a bit of a project, and is carried out nicely by [Agrachev, Barilari, and Boscain 2018, Theorem 3.59]. (In any case, this will follow from Theorem 5.22 and Proposition 6.7 below.)

6.3 Definition: (Pontryagin extremal in sub-Riemannian geometry) A curve $\gamma \in \mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D};x_0,x_1)$ is a ***Pontryagin extremal*** if there exists $\mu_0 \in \{0,1\}$ and $\mu \in \mathrm{H}^1([t_0,t_1];\gamma^*\mathsf{P}_{\mathsf{D},\mu_0})$ such that the following conditions hold:

(i) either $\mu_0 = 1$ or μ is nowhere zero;

(ii) there exists a time-varying section $\xi \in \mathrm{L}^2\Gamma^\infty([t_0,t_1];\rho^*_{\mathsf{D},\mu_0}\mathsf{D})$ that satisfies

 (a) $H_{\mathsf{D},\mu_0}(\xi(t,\alpha)\oplus\alpha) = H^{\max}_{\mathsf{D},\mu_0}(\alpha)$ for $(t,\alpha) \in [t_0,t_1]\times\mathsf{P}_{\mathsf{D},\mu_0}$ and

 (b) $\gamma'(t) = \xi(t,\mu(t))$

 for almost every $t \in [t_0,t_1]$;

(iii) the curve μ in $\mathsf{T}^*\mathsf{M}$ is an integral curve for the (possibly time-varying) Hamiltonian vector field associated with the (possibly time-varying) Hamiltonian

$$(t,\alpha) \mapsto H_{\mathsf{D},\mu_0}(\xi(t,\pi_{\mathsf{T}^*\mathsf{M}}(\alpha))\oplus\alpha). \qquad \bullet$$

Let us make some observations about these conditions.

6.4 Remarks: (Properties of Pontryagin extremals)

1. When $\mu_0 = 1$, then, according to Lemma 6.2, the maximisation condition from part (ii) of the definition uniquely determines ξ by

$$\xi(t, \pi_{\mathsf{T}^*\mathsf{M}}(\alpha)) = \frac{1}{2}\mathbb{G}_{\mathsf{D}}^{\sharp}(\alpha).$$

   Thus ξ is not time-varying in this case, and the Hamiltonian from part (iii) of the definition is simply the associated maximum Hamiltonian from Lemma 6.2:

$$H_{\mathsf{D},1}^{\max}(\alpha) = \frac{1}{2}\mathbb{G}_{\mathsf{D}}^{-1}(\alpha, \alpha).$$

   The Pontryagin extremals in this case we call ***normal***. A consequence is that normal extremals are smooth.

   Note that we can then extend the notion of a normal extremal as a curve $\gamma\colon I \to \mathsf{M}$ defined on an arbitrary interval that is a projection to M of an integral curve of the Hamiltonian vector field associated to the maximum Hamiltonian in this case.

2. When $\mu_0 = 0$, then it is required that μ be nowhere zero along γ. More importantly, however, the maximisation condition from part (ii) of the definition requires that $\mu(t) \in \Lambda(\mathsf{D})$, according to Lemma 6.2. Also by Lemma 6.2, the maximum Hamiltonian is zero, and so places no constraints on the time-varying section ξ of $\rho_{\mathsf{D},\mu_0}^*\mathsf{D}$. Thus one obtains no information about the velocity γ' along the extremal γ directly from the maximisation condition. Indeed, one must look elsewhere, beyond the conditions for Pontryagin extremals, to get useful conditions on velocities. One such condition will arise in Corollary 6.9. Such conditions are also studied in detail in Chapter 12 of [Agrachev, Barilari, and Boscain 2018]. We note, however, that part (iii) of the definition is not vacuous, and gives conditions on the curve μ in $\Lambda(\mathsf{D})$.

   The Pontryagin extremals in this case we call ***abnormal***.

   As with normal extremals, we can extend the notion of an abnormal extremal to arbitrary intervals. Thus, an abnormal extremal in this case is the projection to M of an integral curve of the restriction to $\Lambda(\mathsf{D})$ of the Hamiltonian vector field X_ξ associated with the Hamiltonian $H_\xi(t, \alpha) = \langle \alpha; \xi(t, \pi_{\mathsf{T}^*\mathsf{M}}(\alpha))\rangle$, where ξ is a time-varying D-valued vector field with appropriate regularity, i.e., in $\mathrm{L}^2\Gamma^\infty(\rho_{\mathsf{D},0}^*\mathsf{D})$.

   An important open question in sub-Riemannian geometry is whether there are abnormal sub-Riemannian geodesics that are not smooth. At present, there is no general proof of this, but there are also no counterexamples. •

6.2. The connection between sub-Riemannian geometry and constrained variational mechanics. In this section we establish the connections between the theory of sub-Riemannian geometry, as described in the preceding section, and the theory of constrained variational trajectories, as described in Section 5.4. For the purposes of the current presentation, when we say "constrained simple mechanical system," we shall always take the potential function V to be zero, so the data is a triple $(\mathsf{M}, \mathbb{G}, \mathsf{D})$.

Let us begin by showing how the data of a sub-Riemannian manifold arises from restricting the data of a constrained simple mechanical system.

6.5 Lemma: (Sub-Riemannian manifolds from constrained simple mechanical systems) *If* $(\mathsf{M},\mathsf{D},\mathbb{G}_{\mathsf{D}})$ *is a sub-Riemannian manifold, then there exists a Riemannian metric* $\mathbb{G}$ *on* M *such that* $\mathbb{G}_{\mathsf{D}} = \mathbb{G}|\mathsf{D}$.

Proof: Let M be properly embedded in $\mathbb{R}^N$ for sufficiently large N. Let $\mathsf{D}^\perp \subseteq \mathsf{TM}$ be the orthogonal complement to $\mathsf{D} \subseteq \mathsf{TM}$ with respect to the Euclidean Riemannian metric in $\mathbb{R}^N$. Then D has the fibre metric $\mathbb{G}_{\mathsf{D}}$ while we can equip $\mathsf{D}^\perp$ with the restriction of the Euclidean Riemannian metric, for example, which we denote by $\mathbb{G}_{\mathsf{D}^\perp}$. We can then define $\mathbb{G} = \mathbb{G}_{\mathsf{D}} + \mathbb{G}_{\mathsf{D}^\perp}$. ■

An immediate consequence of the lemma is that the energy action $A_{\mathbb{G}_{\mathsf{D}}}$ in sub-Riemannian geometry is the same as the restriction of the energy action $A_{\mathbb{G}}$ for a constrained simple mechanical system to $\mathrm{H}^1([t_0,t_1];\mathsf{M};\mathsf{D})$, provided that the kinetic energy is chosen so as to agree with the energy of the sub-Riemannian manifold on D. Note that this correspondence relies on the regularity of D; if D is not regular, then there may not be an extension of $\mathbb{G}_{\mathsf{D}}$ from D to a Riemannian metric on M. But in the regular case, the sub-Riemannian extremals and the constrained variational trajectories agree, being critical points of the same actions. However, somewhat more than this is true, as we shall now explore.

Let us begin by clarifying how the Riemannian metric $\mathbb{G}$ relates to the objects associated with $\mathbb{G}_{\mathsf{D}}$. Given a smooth Riemannian manifold $(\mathsf{M},\mathbb{G})$ and a smooth subbundle, denote

$$\mathbb{G}_{\mathsf{D}} = \mathbb{G}|\mathsf{D}, \quad \mathbb{G}_{\mathsf{D}^\perp} = \mathbb{G}|\mathsf{D}^\perp.$$

Note that both $\mathbb{G}_{\mathsf{D}}$ and $\mathbb{G}_{\mathsf{D}^\perp}$ define smooth $(0,2)$-tensor fields on M. We shall, of course, think of $(\mathsf{M},\mathsf{D},\mathbb{G}_{\mathsf{D}})$ as a sub-Riemannian manifold. As such, associated with $\mathbb{G}_{\mathsf{D}}$ is the vector bundle mapping $\mathbb{G}_{\mathsf{D}}^\sharp\colon \mathsf{T}^*\mathsf{M} \to \mathsf{TM}$ and the $(2,0)$-tensor field $\mathbb{G}_{\mathsf{D}}^{-1}$ on M, as described above. Let us describe these in terms of the Riemannian metric $\mathbb{G}$. To do so, we denote by $\mathbb{G}^{-1}$ the vector bundle metric on $\mathsf{T}^*\mathsf{M}$ associated to $\mathbb{G}$ and we define subbundles $\Lambda(\mathsf{D})$ and $\Lambda(\mathsf{D}^\perp)$ of $\mathsf{T}^*\mathsf{M}$ by

$$\begin{aligned}\Lambda(\mathsf{D})_x &= \{\alpha_x \in \mathsf{T}^*_x\mathsf{M} \mid \langle\alpha_x;u_x\rangle = 0,\ u_x \in \mathsf{D}_x\},\\ \Lambda(\mathsf{D}^\perp)_x &= \{\alpha_x \in \mathsf{T}^*_x\mathsf{M} \mid \langle\alpha_x;w_x\rangle = 0,\ w_x \in \mathsf{D}^\perp_x\}.\end{aligned}$$

We note that $\Lambda(\mathsf{D})$ and $\Lambda(\mathsf{D}^\perp)$ are $\mathbb{G}^{-1}$-orthogonal. We denote

$$\mathbb{G}^{-1}_{\Lambda(\mathsf{D})} = \mathbb{G}^{-1}|\Lambda(\mathsf{D}), \quad \mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)} = \mathbb{G}^{-1}|\Lambda(\mathsf{D}^\perp).$$

Note that both $\mathbb{G}^{-1}_{\Lambda(\mathsf{D})}$ and $\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)}$ define smooth $(2,0)$-tensor fields on M.

We then have the following elementary result.

6.6 Lemma: (Riemannian characterisations of sub-Riemannian tensors) *With the above notation, we have*

(i) $\mathbb{G}^\sharp_{\mathsf{D}} = (\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp$ *and*

(ii) $\mathbb{G}^{-1}_{\mathsf{D}} = \mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)}$.

Proof: First of all, since

$$\mathbb{G} = \mathbb{G}_{\mathsf{D}} + \mathbb{G}_{\mathsf{D}^\perp},$$

since

$$\mathbb{G}^{-1} = \mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)} + \mathbb{G}^{-1}_{\Lambda(\mathsf{D})},$$

since D and $\mathsf{D}^\perp$ are $\mathbb{G}$-orthogonal, since $\Lambda(\mathsf{D}^\perp)$ and $\Lambda(\mathsf{D})$ are $\mathbb{G}^{-1}$-orthogonal, and by definition of $\Lambda(\mathsf{D}^\perp)$ and $\Lambda(\mathsf{D})$, we have vector bundle isomorphisms

$$\mathbb{G}^\flat_{\mathsf{D}}|\mathsf{D}\colon \mathsf{D} \to \Lambda(\mathsf{D}^\perp), \quad \mathbb{G}^\flat_{\mathsf{D}^\perp}|\mathsf{D}^\perp\colon \mathsf{D}^\perp \to \Lambda(\mathsf{D})$$

with inverses

$$(\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp|\Lambda(\mathsf{D}^\perp)\colon \Lambda(\mathsf{D}^\perp) \to \mathsf{D}, \quad (\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp|\Lambda(\mathsf{D})\colon \Lambda(\mathsf{D}) \to \mathsf{D}^\perp,$$

respectively.

(i) For $u_x \in \mathsf{D}_x$ and $\alpha_x \in \mathsf{T}^*_x\mathsf{M}$, we have

$$\begin{aligned}\mathbb{G}_{\mathsf{D}}((\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\alpha_x), u_x) &= \mathbb{G}((\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\alpha_x) + (\mathbb{G}^{-1}_{\Lambda(\mathsf{D})})^\sharp(\alpha_x), u_x)\\ &= \mathbb{G}(\mathbb{G}^\sharp(\alpha_x), u_x) = \langle \alpha_x; u_x\rangle.\end{aligned}$$

Thus $(\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp$ satisfies the defining conditions (6.1) of $\mathbb{G}^\sharp_{\mathsf{D}}$.

(ii) For $\alpha_x, \beta_x \in \mathsf{T}^*_x\mathsf{M}$, we compute

$$\begin{aligned}\langle \alpha_x; (\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\beta_x)\rangle &= \mathbb{G}(\mathbb{G}^\sharp(\alpha_x), (\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\beta_x))\\ &= \mathbb{G}((\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\alpha_x) + (\mathbb{G}^{-1}_{\Lambda(\mathsf{D})})^\sharp(\alpha_x), (\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\beta_x))\\ &= \mathbb{G}_{\mathsf{D}}((\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\alpha_x), (\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)})^\sharp(\beta_x))\\ &= \mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)}(\alpha_x, \beta_x).\end{aligned}$$

Thus $\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)}$ satisfies the defining conditions (6.2) for $\mathbb{G}^{-1}_{\mathsf{D}}$. ■

We can now state the precise relationship between extremals for the sub-Riemannian problem and constrained variational trajectories.

6.7 Proposition: (Correspondence between Pontryagin extremals and constrained variational trajectories) *Let* $(\mathsf{M}, \mathsf{D}, \mathbb{G}_{\mathsf{D}})$ *be a sub-Riemannian manifold and let* $(\mathsf{M}, \mathbb{G}, \mathsf{D})$ *be a smooth constrained simple mechanical system for which* $\mathbb{G}_{\mathsf{D}} = \mathbb{G}|\mathsf{D}$*. Then, for* $\gamma\colon I \to \mathsf{M}$*, the following statements hold:*

(i) the following statements are equivalent:

(a) γ *is a normal Pontryagin extremal;*

(b) γ *is a* D*-regular constrained variational trajectory;*

(ii) the following statements are equivalent:

(a) γ *is an abnormal Pontryagin extremal;*

(b) γ *is a* D*-singular constrained variational trajectory.*

Proof: (i) Let us determine the pull-back of the maximum Hamiltonian

$$H^{\max}_{\mathsf{D},1}(\alpha) = \frac{1}{2}\mathbb{G}_{\mathsf{D}}(\alpha,\alpha) = \frac{1}{2}\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)}(\alpha,\alpha)$$

by the diffeomorphism $\mathbb{G}^\flat\colon \mathsf{TM} \to \mathsf{T}^*\mathsf{M}$. We have

$$\begin{aligned} H^{\max}_{\mathbb{G},\mathsf{D}}(v) \triangleq (\mathbb{G}^\flat)^* H^{\max}_{\mathsf{D},1}(v) &= \frac{1}{2}\mathbb{G}^{-1}_{\Lambda(\mathsf{D}^\perp)}(\mathbb{G}^\flat(v),\mathbb{G}^\flat(v)) \\ &= \frac{1}{2}\mathbb{G}_{\mathsf{D}}(v,v) = \frac{1}{2}\mathbb{G}(P_{\mathsf{D}}(v),P_{\mathsf{D}}(v)). \end{aligned}$$

Now let us compute the Hamiltonian vector field for the Hamiltonian $H^{\max}_{\mathbb{G},\mathsf{D}}$ with respect to the symplectic form $\omega_{\mathbb{G}}$ on TM defined in (2.6). This, by definition, is the vector field $X^{\max}_{\mathbb{G},\mathsf{D}}$ on TM satisfying

$$\langle \mathrm{d}H^{\max}_{\mathbb{G},\mathsf{D}}(v_x); X_{v_x}\rangle = \omega_{\mathbb{G}}(X^{\max}_{\mathbb{G},\mathsf{D}}(v_x), X_{v_x}), \qquad v_x \in \mathsf{TM},\ X_{v_x} \in \mathsf{T}_{v_x}\mathsf{TM}.$$

As in the proof of Lemma 2.12, let $X_0, Y_0, X_1, Y_1 \in \Gamma^\infty(\mathsf{TM})$ be such that

$$X_{v_x} = \mathrm{hlft}(X_0(x), v_x) + \mathrm{vlft}(X_1(x), v_x),\ \ X^{\max}_{\mathbb{G},\mathsf{D}}(v_x) = \mathrm{hlft}(Y_0(x), v_x) + \mathrm{vlft}(Y_1(x), v_x).$$

Let $\gamma\colon [0,T] \to \mathsf{M}$ be the integral curve of X_0 through x and let $\Upsilon\colon [0,T] \to \mathsf{TM}$ be the vector field along γ obtained by parallel translation of v_x. We then have

$$\begin{aligned} \langle \mathrm{d}H^{\max}_{\mathbb{G},\mathsf{D}}(v_x); X_0^{\mathrm{h}}(v_x)\rangle &= \frac{1}{2}\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathbb{G}(P_{\mathsf{D}}\circ\Upsilon(t), P_{\mathsf{D}}\circ\Upsilon(t)) \\ &= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'}(P_{\mathsf{D}}\circ\Upsilon)(0), P_{\mathsf{D}}\circ\Upsilon(0)) \\ &= \mathbb{G}((\overset{\mathbb{G}}{\nabla}_{X_0}P_{\mathsf{D}})(v_x), P_{\mathsf{D}}(v_x)). \end{aligned}$$

We also compute

$$\langle \mathrm{d}H^{\max}_{\mathbb{G},\mathsf{D}}(v_x); X_1^{\mathrm{v}}(v_x)\rangle = \mathbb{G}(P_{\mathsf{D}}\circ X_1(x), P_{\mathsf{D}}(v_x)).$$

Therefore,

$$\langle \mathrm{d}H^{\max}_{\mathbb{G},\mathsf{D}}(v_x); X_{v_x}\rangle = \mathbb{G}((\overset{\mathbb{G}}{\nabla}_{X_0}P_{\mathsf{D}})(v_x) + P_{\mathsf{D}}\circ X_1(x), P_{\mathsf{D}}(v_x)).$$

By Lemma 2.12 we have

$$\begin{aligned} \omega_{\mathbb{G}}(X^{\max}_{\mathbb{G},\mathsf{D}}(v_x), X_{v_x}) &= \mathbb{G}(T_{v_x}\pi_{\mathsf{TM}}(X^{\max}_{\mathbb{G},\mathsf{D}}(v_x)), K_{\mathbb{G}}(X_{v_x})) - \mathbb{G}(K_{\mathbb{G}}(X^{\max}_{\mathbb{G},\mathsf{D}}(v_x)), T_{v_x}\pi_{\mathsf{TM}}(X_{v_x})) \\ &= \mathbb{G}(Y_0(x), X_1(x)) - \mathbb{G}(Y_1(x), X_0(x)). \end{aligned}$$

We conclude that

$$\begin{aligned} \mathbb{G}(Y_0(x), X_1(x)) &= \mathbb{G}(X_1(x), P_{\mathsf{D}}(v_x)) \\ \mathbb{G}(Y_1(x), X_0(x)) &= -\mathbb{G}((\overset{\mathbb{G}}{\nabla}_{X_0}P_{\mathsf{D}})(v_x), P_{\mathsf{D}}(v_x)). \end{aligned} \tag{6.3}$$

Now write $v_x = v_x^{\parallel} + v_x^{\perp}$ for $v_x^{\parallel} \in \mathsf{D}_x$ and $v_x^{\perp} \in \mathsf{D}_x^{\perp}$. We then have

$$Y_0(x) = v_x^{\parallel} \tag{6.4}$$

from the first of equations (6.3). Now consider $Y \in \Gamma^\infty(\mathsf{D})$ and $\alpha \in \Gamma^\infty(\mathsf{D}^\perp)$. Then

$$\begin{aligned}&\mathbb{G}(\alpha, Y) = 0\\ \Longrightarrow\ &\mathbb{G}(\overset{\mathrm{G}}{\nabla}_{X_0}\alpha, Y) + \mathbb{G}(\alpha, \overset{\mathrm{G}}{\nabla}_{X_0}Y) = 0.\end{aligned}$$

Using this, we compute

$$\begin{aligned}\mathbb{G}(S_{\mathsf{D}^\perp}(X_0,\alpha), Y) = \mathbb{G}(P_{\mathsf{D}}(\overset{\mathrm{G}}{\nabla}_{X_0}\alpha), Y) &= \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{X_0}\alpha, Y)\\ &= -\mathbb{G}(\alpha, \overset{\mathrm{G}}{\nabla}_{X_0}Y) = -\mathbb{G}(\alpha, P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{X_0}Y))\\ &= -\mathbb{G}(\alpha, S_{\mathsf{D}}(X_0, Y)) = -\mathbb{G}(X_0, S^*_{\mathsf{D}}(Y)(\alpha)).\end{aligned}$$

Using this computation and Lemma 2.36(iii), we have

$$\begin{aligned}\mathbb{G}((\overset{\mathrm{G}}{\nabla}_{X_0}P_{\mathsf{D}})(v_x), P_{\mathsf{D}}(v_x)) &= -\mathbb{G}((\overset{\mathrm{G}}{\nabla}_{X_0}P_{\mathsf{D}})(v_x^\perp), v_x^\parallel)\\ &= -\mathbb{G}(S_{\mathsf{D}^\perp}(X_0(x), v_x^\perp), v_x^\parallel) = \mathbb{G}(X_0, S^*_{\mathsf{D}}(v_x^\parallel)(v_x^\perp)).\end{aligned}$$

By the second of equations (6.3), we have

$$Y_1(x) = -S^*_{\mathsf{D}}(v_x^\parallel)(v_x^\perp). \tag{6.5}$$

Combining (6.4) and (6.5), we have

$$X^{\max}_{\mathbb{G},\mathsf{D}}(v_x) = \mathrm{hlft}(v_x^\parallel, v_x) - \mathrm{vlft}(S^*_{\mathsf{D}}(v_x^\parallel)(v_x^\perp)).$$

Now let $\Upsilon\colon I \to \mathsf{T}\mathsf{M}$ be an integral curve for $X^{\max}_{\mathbb{G},\mathsf{D}}$:

$$\Upsilon'(t) = X^{\max}_{\mathbb{G},\mathsf{D}} \circ \Upsilon(t), \qquad t \in I.$$

Denote $\gamma = \pi_{\mathsf{TM}} \circ \Upsilon$. From (6.4) we have

$$\gamma'(t) = T_{\Upsilon(t)}\pi_{\mathsf{TM}}(\Upsilon'(t)) = P_{\mathsf{D}} \circ \Upsilon(t), \qquad t \in I.$$

Thus we can write

$$\Upsilon(t) = \gamma'(t) - \lambda(t), \qquad t \in I,$$

where $\gamma'(t) \in \mathsf{D}_{\gamma(t)}$ and $\lambda(t) \in \mathsf{D}^\perp_{\gamma(t)}$. By (2.3), (6.5), and since Υ is an integral curve for $X^{\max}_{\mathbb{G},\mathsf{D}}$, we have

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\Upsilon = S^*_{\mathsf{D}}(\gamma')(\lambda).$$

But we also have

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\Upsilon = \overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' - \overset{\mathrm{G}}{\nabla}_{\gamma'}\lambda.$$

Thus we obtain the conclusion that Υ is an integral curve for $X^{\max}_{\mathbb{G},\mathsf{D}}$ if and only if

1. $\gamma' = P_{\mathsf{D}} \circ \Upsilon(t)$ for $\gamma = \pi_{\mathsf{TM}} \circ \Upsilon$, and
2. there exists a $\mathsf{D}^\perp$-valued section λ along γ so that $\Upsilon = \gamma' - \lambda$ and

$$\overset{\mathrm{G}}{\nabla}_{\gamma'}\gamma' - \overset{\mathrm{G}}{\nabla}_{\gamma'}\lambda - S^*_{\mathsf{D}}(\gamma')(\lambda) = 0.$$

This is gives what we are required to prove in this part of the proof, given Theorem 5.22(ii).

(ii) In this case we work with the Hamiltonian

$$(t,\alpha)\mapsto H_{\mathsf{D},0}(\xi(t,\pi_{\mathsf{T}^*\mathsf{M}}(\alpha))\oplus\alpha)=\langle\alpha;\xi(t,\pi_{\mathsf{T}^*\mathsf{M}}(\alpha))\rangle$$

associated with a D-valued time-varying vector field ξ. We shall denote $\xi_t(x)=\xi(t,x)$, We can pull-back this Hamiltonian to TM by the diffeomorphism $\mathbb{G}^\flat\colon \mathsf{TM}\to\mathsf{T}^*\mathsf{M}$:

$$H_\xi(t,v)=\langle\mathbb{G}^\flat(v);\xi(t,\pi_{\mathsf{TM}}(v))\rangle=\mathbb{G}(v,\xi(t,\pi_{\mathsf{TM}}(v))).$$

We then denote by X_ξ the corresponding time-varying Hamiltonian vector field defined by

$$\langle \mathrm{d}H_\xi(t,v_x);X_{v_x}\rangle=\omega_{\mathbb{G}}(X_\xi(t,v_x),X_{v_x}),\qquad v_x\in\mathsf{TM},\ X_{v_x}\in\mathsf{T}_{v_x}\mathsf{TM}.$$

We denote $X_{\xi,t}(v)=X_\xi(t,v)$ and fix t for the moment. As above, let $X_0,Y_0,X_1,Y_1\in\Gamma^\infty(\mathsf{TM})$ be such that

$$X_{v_x}=\operatorname{hlft}(X_0(x),v_x)+\operatorname{vlft}(X_1(x),v_x),\ \ X_{\xi,t}(v_x)=\operatorname{hlft}(Y_0(x),v_x)+\operatorname{vlft}(Y_1(x),v_x).$$

Let $\gamma\colon[0,T]\to\mathsf{M}$ be the integral curve of X_0 through x and let $\Upsilon\colon[0,T]\to\mathsf{TM}$ be the vector field along γ obtained by parallel translation of v_x. Then

$$\langle\mathrm{d}H_\xi(t,v_x);X_0^{\mathrm{h}}(v_x)\rangle=\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}\mathbb{G}(\Upsilon(s),\xi_t(\gamma(s)))=\mathbb{G}(v_x,\overset{\mathrm{G}}{\nabla}_{X_0(x)}\xi_t)$$

and

$$\langle\mathrm{d}H_\xi(v_x);X_1^{\mathrm{v}}(v_x)\rangle=\mathbb{G}(X_1(x),\xi_t(x)).$$

Thus

$$\langle\mathrm{d}H_\xi(t,v_x);X_{v_x}\rangle=\mathbb{G}(v_x,\overset{\mathrm{G}}{\nabla}_{X_0(x)}\xi_t)+\mathbb{G}(X_1(x),\xi_t(x)).$$

Since

$$\omega_{\mathbb{G}}(X^{\max}_{\mathbb{G},\mathsf{D}}(v_x),X_{v_x})=\mathbb{G}(Y_0(x),X_1(x))-\mathbb{G}(Y_1(x),X_0(x)),$$

we have

$$X_\xi(t,v_x)=\operatorname{hlft}(\xi(t,x),v_x)-\operatorname{vlft}((\overset{\mathrm{G}}{\nabla}\xi_t)^*(v_x),v_x),$$

where $(\overset{\mathrm{G}}{\nabla}\xi_t)^*$ denotes the $\mathbb{G}$-adjoint of $\overset{\mathrm{G}}{\nabla}\xi_t$. Note that, by the definition of abnormal extremals, we are only interested in the restriction of this vector field to $\mathsf{D}^\perp$. Let us determine the restriction of X_ξ to $\mathsf{D}^\perp$. Let $Y\in\Gamma^\infty(\mathsf{D})$ and let $\alpha\in\Gamma^\infty(\mathsf{D}^\perp)$. Then

$$\begin{aligned}
&\mathbb{G}(\alpha,Y)=0\\
\Longrightarrow\ &\mathbb{G}(\overset{\mathrm{G}}{\nabla}_{X_0}\alpha,Y)+\mathbb{G}(\alpha,\overset{\mathrm{G}}{\nabla}_{X_0}Y)\\
\Longrightarrow\ &\mathbb{G}(\alpha,\overset{\mathrm{G}}{\nabla}_{X_0}Y)=-\mathbb{G}(\alpha,P_{\mathsf{D}^\perp}(\overset{\mathrm{G}}{\nabla}_{X_0}(Y)))\\
\Longrightarrow\ &\mathbb{G}(\alpha,\overset{\mathrm{G}}{\nabla}_{X_0}Y)=\mathbb{G}(\alpha,S_{\mathsf{D}}(X_0,Y))\\
\Longrightarrow\ &\mathbb{G}(\alpha,\overset{\mathrm{G}}{\nabla}_{X_0}Y)=\mathbb{G}(X_0,S^*_{\mathsf{D}}(Y)(\alpha)).
\end{aligned}$$

Thus, for $v_x\in\mathsf{D}^\perp$, we have

$$X_\xi(t,v_x)=\operatorname{hlft}(\xi(t,x),v_x)-\operatorname{vlft}(S^*_{\mathsf{D}}(\xi_t(x))(v_x),v_x).\tag{6.6}$$

Let $\lambda\colon I \to \mathsf{D}^\perp$ be an integral curve of X_ξ and define $\gamma = \pi_{\mathsf{TM}} \circ \lambda$. Then

$$\gamma'(t) = T_{\Upsilon(t)}\pi_{\mathsf{TM}}(\lambda'(t)) = \xi(t, \gamma(t)), \qquad t \in I.$$

Then we immediately deduce that γ is an integral curve of ξ. We then have, by (2.3) and (6.6),

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda = -S^*_{\mathsf{D}}(\gamma')(\lambda),$$

and this gives this part of the result, according to Theorem 5.22(ii). ■

This result is asserted, but not proved, by Kupka and Oliva [2001]; they likely had a coordinate proof in mind since a coordinate proof is straightforward, if messy. Langerok [2003] gives a related result, proved partly in coordinates. We give, for what we believe is the first time, a coordinate-independent proof of an affine connection characterisation of extremals in sub-Riemannian geometry. Moreover, directly from Theorem 5.22(iii) we have the following characterisation of sub-Riemannian extremals.

6.8 Corollary: (Affine connection characterisation of Pontryagin extremals) *Let $(\mathsf{M}, \mathsf{D}, \mathbb{G}_{\mathsf{D}})$ be a sub-Riemannian manifold and let $(\mathsf{M}, \mathbb{G}, \mathsf{D})$ be a smooth constrained simple mechanical system for which $\mathbb{G}_{\mathsf{D}} = \mathbb{G}|\mathsf{D}$. Then, for $\gamma\colon I \to \mathsf{M}$, the following statements hold:*

(i) the following statements are equivalent:

(a) γ is a normal Pontryagin extremal;

(b) there exists a section $\lambda\colon I \to \mathsf{D}^\perp$ over γ so that γ and λ together satisfy

$$\begin{aligned}
&\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' = F^*_{\mathsf{D}}(\gamma')(\lambda),\\
&\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G_{\mathsf{D}}(\gamma', \gamma') + \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda);
\end{aligned}$$

(ii) the following statements are equivalent:

(a) γ is an abnormal Pontryagin extremal;

(b) there exists a nowhere zero section $\lambda\colon I \to \mathsf{D}^\perp$ over γ so that γ and λ together satisfy

$$\begin{aligned}
&F^*_{\mathsf{D}}(\gamma')(\lambda) = 0,\\
&\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda).
\end{aligned}$$

There are a few interesting conclusions one can draw from the preceding results. First of all, in the normal case, our result shows that extremals are projections of integral curves of a smooth linear vector field, and so are smooth. Second of all, we can use our geometric structure to give the following characterisation of abnormal extremals.

6.9 Corollary: (Property of abnormal extremals in sub-Riemannian geometry) *Let $(\mathsf{M}, \mathsf{D}, \mathbb{G}_{\mathsf{D}})$ be a sub-Riemannian manifold. If $\gamma\colon I \to \mathsf{M}$ is an abnormal extremal with adjoint field $\lambda\colon I \to \mathsf{D}^\perp$ over γ, then $\lambda(t) \in \ker(\hat{F}^*_{\mathsf{D}})_{\gamma'(t)}$ for every $t \in I$. In particular, if $\hat{F}^*_{\mathsf{D}}$ is injective on fibres of $\pi^*_{\mathsf{D}}\mathsf{D}^\perp$, then there are no abnormal extremals.*

7. When are nonholonomic trajectories variational (and vice versa)?

In this section we address the question of when nonholonomic and constrained variational trajectories for a constrained simple mechanical system $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ coincide in some way. We recall from Theorems 5.18 and 5.22 that the equations governing nonholonomic trajectories are

$$\boxed{\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0,} \tag{NH}$$

while the equations governing constrained variational trajectories are

$$\boxed{\begin{aligned} &F^*_{\mathsf{D}}(\gamma')(\lambda) = 0,\\ &\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda)\end{aligned}} \tag{SCV}$$

in the D-singular case and

$$\boxed{\begin{aligned} &\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = F^*_{\mathsf{D}}(\gamma')(\lambda),\\ &\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G_{\mathsf{D}}(\gamma',\gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma + \frac{1}{2}G^\star_{\mathsf{D}}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}}(\gamma')(\lambda)\end{aligned}} \tag{RCV}$$

in the D-regular case.

We consider various versions of the question of coincidence of trajectories:

1. what are the initial conditions in D through which solutions of (NH) are also solutions of either (SCV) or (RCV)?
2. when are all solutions of equations (SCV) and (RCV) also solutions of (NH)?
3. when is a given solution of (NH) also a solution of either (SCV) or (RCV)?
4. when are all solutions of (NH) also solutions of either (SCV) or (RCV)?

Let us say a few words about the questions we ask, and some we do not ask.

7.1 Remarks: (Regarding questions of comparison of solutions) One of the important differences between the nonholonomic equation (NH), and the variational equations (SCV) and (RCV) is that the latter equations require an initial condition for the section λ. This distinction leads to some important interpretations of the above questions that we now address.

1. The first question should be read as, "Determine all solutions of (NH) for which there exists *some choice* of initial condition for λ so that the solution to (NH) is also a solution of either (SCV) or (RCV)." Physically, this means that we are determining all physical motions that are also solutions to a constrained variational problem.
2. The second question should be read as, "When is it true that, *for any choice* of initial condition for λ, the solutions of the equations (SCV) and (RCV) also solutions of (NH)?" Physically, the question can be interpreted as, "When are all solutions of the constrained variational problem also solutions to the physical equations of motion?" As we shall see, this question is easily answered. Indeed, it is well known that the answer to the second question is, "when and only when D is integrable." We shall see that this is particularly easily proved in our framework; indeed, the reader can see that it holds virtually by inspection.

3. The third question should be read as, "When does there exist *some choice* of initial condition for λ so that a given solution of (NH) is also a solution of the equations (SCV) or (RCV)?" Physically, this question is tantamount to asking, "When is a particular physical motion also a solution to the constrained variational problem?"
4. The fourth question has the interpretation, "When is it possible, for every solution of (NH), to make *some choice* of initial condition for λ so that the resulting trajectory also satisfies (SCV) or (RCV)?" The physical question can be phrased as, "When is every physical motion also a solution to the constrained variational problem?"
5. Note that we do not ask the question, "When is a given solution of either (SCV) or (RCV) also a solution of (NH)?" The answer to this question is easy, however. One merely must verify whether $F^*_{\mathsf{D}}(\gamma')(\lambda) = 0$. •

7.1. Pulling back equations to D. Throughout the following discussion, we let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. In order to compare the nonholonomic and D-regular constrained variational equations, we shall pull the equations back to equations from M to D. As we shall see, doing this will allow us to make use of the methods of Section 4 for determining subbundles invariant under an affine vector field. In particular, these pulled-back equations will be described by an affine vector field. In [Langerok 2003] the constructions we give here are presented in the context of "connections over bundle maps."

Note that we have the following vector bundles over M:

$$\pi_{\mathsf{TM}} \colon \mathsf{TM} \to \mathsf{M}, \quad \pi_{\mathsf{D}} \colon \mathsf{D} \to \mathsf{M}, \quad \pi_{\mathsf{D}^\perp} \colon \mathsf{D}^\perp \to \mathsf{M}.$$

Associated with these, we have the pull-back vector bundles over D:

$$\pi^*_{\mathsf{D}}\pi_{\mathsf{TM}} \colon \pi^*_{\mathsf{D}}\mathsf{TM} \to \mathsf{D}, \quad \pi^*_{\mathsf{D}}\pi_{\mathsf{D}} \colon \pi^*_{\mathsf{D}}\mathsf{D} \to \mathsf{D}, \quad \pi^*_{\mathsf{D}}\pi_{\mathsf{D}^\perp} \colon \pi^*_{\mathsf{D}}\mathsf{D}^\perp \to \mathsf{D}.$$

Explicitly and to fix notation,

$$\begin{aligned} \pi^*_{\mathsf{D}}\mathsf{TM} &= \{(v,u) \in \mathsf{TM} \times \mathsf{D} \mid \pi_{\mathsf{TM}}(v) = \pi_{\mathsf{D}}(u)\}, \\ \pi^*_{\mathsf{D}}\mathsf{D} &= \{(v,u) \in \mathsf{D} \times \mathsf{D} \mid \pi_{\mathsf{D}}(v) = \pi_{\mathsf{D}}(u)\}, \\ \pi^*_{\mathsf{D}}\mathsf{D}^\perp &= \{(v,u) \in \mathsf{D}^\perp \times \mathsf{D} \mid \pi_{\mathsf{D}^\perp}(v) = \pi_{\mathsf{D}}(u)\}. \end{aligned}$$

We can pull back $v_x \in \mathsf{T}_x\mathsf{M}$ to $u_x \in \mathsf{D}_x$ according to the formula

$$\pi^*_{\mathsf{D}} v_x = (v_x, u_x) \in \pi^*_{\mathsf{D}}\mathsf{TM}.$$

In particular, if $X \in \Gamma^r(\mathsf{TM})$, we can pull this back to a section $\pi^*_{\mathsf{D}}X$ of $\pi^*_{\mathsf{D}}\mathsf{TM}$ by

$$\pi^*_{\mathsf{D}}X(v) = (X(\pi_{\mathsf{TM}}(v)), v), \qquad v \in \mathsf{D}.$$

Finally, if $\gamma \colon I \to \mathsf{M}$ is such that γ' is locally absolutely continuous and satisfies $P_{\mathsf{D}^\perp} \circ \gamma' = 0$, and if $\xi \colon I \to \mathsf{TM}$ is a vector field along γ, then we can pull back ξ to a section over γ' according to

$$\pi^*_{\mathsf{D}}\xi(t) = (\xi(t), \gamma'(t)) \in \pi^*_{\mathsf{D}}\mathsf{TM}.$$

Note that $\pi^*_{\mathsf{D}}\xi$ is locally absolutely continuous. Similar constructions hold, of course, for $\pi^*_{\mathsf{D}}\mathsf{D}$ and $\pi^*_{\mathsf{D}}\mathsf{D}^\perp$.

We wish to pull-back some of the bundle maps from Section 2.11 to D. We remind the reader of the tensor constructions following Lemma 2.36, and, particularly, remind them that there is a (minor) difference between the superscript $*$ and the superscript $\star$. Keeping this in mind, we have vector bundle maps

$$\begin{aligned}
\hat{F}_{\mathsf{D}}\colon \pi_{\mathsf{D}}^*\mathsf{D} &\to \pi_{\mathsf{D}}^*\mathsf{D}^{\perp}\\
(u_x, v_x) &\mapsto (F_{\mathsf{D}}(v_x)(u_x), v_x),\\
\hat{F}_{\mathsf{D}}^*\colon \pi_{\mathsf{D}}^*\mathsf{D}^{\perp} &\to \pi_{\mathsf{D}}^*\mathsf{D}\\
(\alpha_x, v_x) &\mapsto (F_{\mathsf{D}}^*(v_x)(\alpha_x), v_x),\\
\hat{F}_{\mathsf{D}^{\perp}}^{\star}\colon \pi_{\mathsf{D}}^*\mathsf{D}^{\perp} &\to \pi_{\mathsf{D}}^*\mathsf{D}^{\perp}\\
(\alpha_x, v_x) &\mapsto (F_{\mathsf{D}^{\perp}}^{\star}(v_x)(\alpha_x), v_x),\\
\hat{G}_{\mathsf{D}^{\perp}}^{\star}\colon \pi_{\mathsf{D}}^*\mathsf{D}^{\perp} &\to \pi_{\mathsf{D}}^*\mathsf{D}^{\perp}\\
(\alpha_x, v_x) &\mapsto (G_{\mathsf{D}^{\perp}}^{\star}(v_x)(\alpha_x), v_x).
\end{aligned}$$

The vector bundle map $\hat{F}_{\mathsf{D}}^*$, particularly its kernel, is important to us. Indeed, we give this kernel a name, referring to Definition 2.42 for the background for the terminology.

7.2 Definition: (Cocharacteristic subbundle) Let $r \in \{\infty, \omega\}$, let $(\mathsf{M}, \mathbb{G})$ be a C^r-Riemannian manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle. The ***cocharacteristic subbundle*** is the C^r-cogeneralised subbundle $\ker(\hat{F}_{\mathsf{D}}^*)$ of $\pi_{\mathsf{D}}^*\mathsf{D}^{\perp}$. •

The connection $\overset{\mathsf{D}^{\perp}}{\nabla}$ on the vector bundle $\pi_{\mathsf{D}^{\perp}}\colon \mathsf{D}^{\perp} \to \mathsf{M}$ can be pulled back to a connection $\overset{\mathsf{D}^{\perp}}{\nabla}{}^*$ on $\pi_{\mathsf{D}}^*\pi_{\mathsf{D}^{\perp}}\colon \pi_{\mathsf{D}}^*\mathsf{D}^{\perp} \to \mathsf{D}$ by requiring that

$$\overset{\mathsf{D}^{\perp}}{\nabla}{}^*_w \pi_{\mathsf{D}}^*\alpha = (\overset{\mathsf{D}^{\perp}}{\nabla}_{T_v\pi_{\mathsf{D}}(w)}\alpha, v), \qquad v \in \mathsf{D},\ w \in \mathsf{T}_v\mathsf{D},\ \alpha \in \Gamma^r(\mathsf{D}^{\perp}).$$

Suppose that $\gamma\colon I \to \mathsf{M}$ is such that γ' is locally absolutely continuous and satisfies $P_{\mathsf{D}^{\perp}} \circ \gamma' = 0$. Denote $\Upsilon = \gamma'$ and define

$$b_{\Upsilon}\colon I \to \pi_{\mathsf{D}}^*\mathsf{D}^{\perp}, \quad A_{\Upsilon}\colon I \to \mathrm{End}(\pi_{\mathsf{D}}^*\mathsf{D}^{\perp})$$

by

$$\begin{aligned}
b_{\Upsilon}(t) &= \frac{1}{2}\pi_{\mathsf{D}}^* G_{\mathsf{D}}(\Upsilon(t), \Upsilon(t)) + \pi_{\mathsf{D}}^* P_{\mathsf{D}^{\perp}} \circ \operatorname{grad} V \circ \gamma(t),\\
A_{\Upsilon}(t)(\pi^*\alpha_{\gamma(t)}) &= \frac{1}{2}\hat{G}_{\mathsf{D}^{\perp}}^{\star}(\pi^*\alpha_{\gamma(t)}) + \frac{1}{2}\hat{F}_{\mathsf{D}^{\perp}}^{\star}(\pi^*\alpha_{\gamma(t)}).
\end{aligned}$$

We can now give the form of the evolution of the adjoint field λ in the constrained variational equations. First we consider the case of D-singular curves.

7.3 Proposition: (Lifted D-singular constrained variational equations) *Let $r \in \{\infty, \omega\}$, let $(\mathsf{M}, \mathbb{G})$ be a C^r-Riemannian manifold, and let $\mathsf{D} \subseteq \mathsf{TM}$ be a C^r-subbundle. Let $\gamma\colon I \to \mathsf{M}$ be such that γ' is locally absolutely continuous and such that $P_{\mathsf{D}^{\perp}} \circ \gamma' = 0$. Denote $\Upsilon = \gamma'$. Then, for a locally absolutely continuous $\lambda\colon I \to \mathsf{D}^{\perp}$ satisfying $\pi_{\mathsf{D}^{\perp}} \circ \lambda = \gamma$, the following are equivalent:*

(i) $\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda)$;

(ii) $\overset{\mathsf{D}^\perp}{\nabla}{}^*_{\Upsilon'}\hat{\lambda} = A_\Upsilon \circ \hat{\lambda}$,

where $\hat{\lambda}\colon I \to \pi^*_{\mathsf{D}}\mathsf{D}^\perp$ *is defined by* $\hat{\lambda}(t) = (\lambda(t), \Upsilon(t))$.

Proof: We have

$$\hat{G}^\star_{\mathsf{D}^\perp}(\pi^*_{\mathsf{D}}\lambda) = \hat{G}^\star_{\mathsf{D}^\perp}(\lambda, \gamma') = (G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda), \gamma')$$

and

$$\hat{F}^\star_{\mathsf{D}^\perp}(\pi^*_{\mathsf{D}}\lambda) = \hat{F}^\star_{\mathsf{D}^\perp}(\lambda, \gamma') = (F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda), \gamma').$$

Note that $\pi_{\mathsf{D}} \circ \Upsilon = \gamma$ and so

$$T_{\Upsilon(t)}\pi_{\mathsf{D}}(\Upsilon'(t)) = \gamma'(t), \qquad \text{a.e. } t \in I.$$

Therefore, we also have

$$\overset{\mathsf{D}^\perp}{\nabla}{}^*_{\Upsilon'(t)}\pi^*_{\mathsf{D}}\lambda(t) = (\overset{\mathsf{D}^\perp}{\nabla}_{T_{\Upsilon(t)}\pi_{\mathsf{D}}(\Upsilon'(t))}\lambda(t), \Upsilon(t)) = (\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'(t)}\lambda(t), \gamma'(t)), \qquad \text{a.e. } t \in I.$$

The proposition now follows immediately. ■

Next we consider the evolution of the adjoint field in the case of D-regular curves for the constrained variational equations.

7.4 Proposition: (Lifted D-regular constrained variational equations) *Let* $r \in \{\infty, \omega\}$ *and let* $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ *be a* C^r*-constrained simple mechanical system. Let* $\gamma\colon I \to \mathsf{M}$ *be such that* γ' *is locally absolutely continuous and such that* $P_{\mathsf{D}^\perp} \circ \gamma' = 0$. *Denote* $\Upsilon = \gamma'$. *Then, for a locally absolutely continuous* $\lambda\colon I \to \mathsf{D}^\perp$ *satisfying* $\pi_{\mathsf{D}^\perp} \circ \lambda = \gamma$, *the following are equivalent:*

(i) $\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G_{\mathsf{D}}(\gamma', \gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma + \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda)$;

(ii) $\overset{\mathsf{D}^\perp}{\nabla}{}^*_{\Upsilon'}\hat{\lambda} = A_\Upsilon \circ \hat{\lambda} + b_\Upsilon$,

where $\hat{\lambda}\colon I \to \pi^*_{\mathsf{D}}\mathsf{D}^\perp$ *is defined by* $\hat{\lambda}(t) = (\lambda(t), \Upsilon(t))$.

Proof: Here, in addition to the computations from the proof of Proposition 7.3, we note that

$$\pi^*_{\mathsf{D}}G_{\mathsf{D}}(\Upsilon, \Upsilon) = (G_{\mathsf{D}}(\gamma', \gamma'), \gamma')$$

and

$$\pi^*_{\mathsf{D}}P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma = (P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma, \gamma'),$$

and from this, combined with the computations from the proof of Proposition 7.3, the result follows. ■

Now let us make these constructions "global," rather than concentrating on a single curve. As ever, let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. We let $X^{\mathrm{nh}}_{\mathsf{D}} \in \Gamma^r(\mathsf{TD})$ be the vector field on D whose integral curves are curves $\Upsilon = \gamma'$, where $\gamma\colon I \to \mathsf{M}$ is a nonholonomic trajectory:

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0, \quad \gamma'(t_0) \in \mathsf{D}_{\gamma(t_0)},$$

for some $t_0 \in I$. That is, $X^{\mathrm{nh}}_{\mathsf{D}}$ is the restriction of $Z_{\mathbb{G},\mathsf{D}} + (P_{\mathsf{D}} \circ \operatorname{grad} V)^{\mathrm{v}}$ to $\mathsf{D} \subseteq \mathsf{T}\mathsf{M}$, where $Z_{\mathbb{G},\mathsf{D}}$ is the geodesic spray of $\overset{\mathsf{D}}{\nabla}$. This makes sense since D is geodesically invariant under $\overset{\mathsf{D}}{\nabla}$, cf. [Bullo and Lewis 2004, Theorem 4.87]. We denote by $(X^{\mathrm{nh}}_{\mathsf{D}})^{\mathrm{h}} \in \Gamma^r(\mathsf{T}(\pi^*_{\mathsf{D}}\mathsf{D}^{\perp}))$ the horizontal lift of $X^{\mathrm{nh}}_{\mathsf{D}} \in \Gamma^r(\mathsf{D})$ to $\pi^*_{\mathsf{D}}\mathsf{D}^{\perp}$ by the connection $\overset{\mathsf{D}^{\perp}}{\nabla}{}^*$. Now define $b_{\mathsf{D}} \in \Gamma^r(\pi^*_{\mathsf{D}}\mathsf{D}^{\perp})$ and $A_{\mathsf{D}} \in \Gamma^r(\operatorname{End}(\pi^*_{\mathsf{D}}\mathsf{D}^{\perp}))$ by

$$\begin{aligned} b_{\mathsf{D}}(u_x) &= \left(\frac{1}{2}G_{\mathsf{D}}(u_x,u_x) + P_{\mathsf{D}^{\perp}} \circ \operatorname{grad} V, u_x\right), \\ A_{\mathsf{D}}(\alpha_x,u_x) &= \left(\frac{1}{2}\hat{G}^{\star}_{\mathsf{D}^{\perp}}(\alpha_x,u_x) + \frac{1}{2}\hat{F}^{\star}_{\mathsf{D}^{\perp}}(\alpha_x,u_x), u_x\right), \end{aligned}$$

for $u_x \in \mathsf{D}$ and $(\alpha_x,u_x) \in \pi^*_{\mathsf{D}}\mathsf{D}^{\perp}$. We can then define the linear vector field $X^{\mathrm{sing}}_{\mathsf{D}} \in \Gamma^r(\mathsf{T}(\pi^*_{\mathsf{D}}\mathsf{D}^{\perp}))$ by

$$X^{\mathrm{sing}}_{\mathsf{D}} = (X^{\mathrm{nh}}_{\mathsf{D}})^{\mathrm{h}} + A^{\mathrm{e}}_{\mathsf{D}}$$

in the D-singular case and the affine vector field $X^{\mathrm{reg}}_{\mathsf{D}} \in \Gamma^r(\mathsf{T}(\pi^*_{\mathsf{D}}\mathsf{D}^{\perp}))$

$$X^{\mathrm{reg}}_{\mathsf{D}} = (X^{\mathrm{nh}}_{\mathsf{D}})^{\mathrm{h}} + A^{\mathrm{e}}_{\mathsf{D}} + b^{\mathrm{v}}_{\mathsf{D}}$$

in the D-regular case.

Let us record the significance of the vector fields $X^{\mathrm{sing}}_{\mathsf{D}}$ and $X^{\mathrm{reg}}_{\mathsf{D}}$, starting with the D-singular case. Note that, since we are considering constrained variational trajectories that project to nonholonomic trajectories, we can assume all curves to be of class C^r, $r \in \{\infty,\omega\}$.

7.5 Proposition: (D-singular constrained variational trajectories along nonholonomic trajectories) *Let $r \in \{\infty,\omega\}$ and let $\Sigma = (\mathsf{M},\mathbb{G},V,\mathsf{D})$ be a C^r-constrained simple mechanical system. For a C^r-curve $\hat{\lambda}\colon I \to \pi^*_{\mathsf{D}}\mathsf{D}^{\perp}$, write $\hat{\lambda}(t) = (\lambda(t),\Upsilon(t))$. Then the following statements are equivalent:*

- *(i) Υ is an integral curve of $X^{\mathrm{nh}}_{\mathsf{D}}$ and λ is such that $\gamma = \pi_{\mathsf{D}} \circ \Upsilon$ and λ together satisfy the conditions for a D-singular constrained variational trajectory from Theorem 5.22(ii)(a) (or (iii)(a));*
- *(ii) the following conditions hold:*
  - *(a) $\lambda(t) \neq 0$ for every $t \in I$;*
  - *(b) $\hat{\lambda}(t) \in \ker(\hat{F}^*_{\mathsf{D}})_{\Upsilon(t)}$ for every $t \in I$;*
  - *(c) $\hat{\lambda}$ is an integral curve of $X^{\mathrm{sing}}_{\mathsf{D}}$.*

Proof: Note that (i) is equivalent to the three equations

$$\begin{aligned} &\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0, \\ &F^*_{\mathsf{D}}(\gamma')(\lambda) = 0, \\ &\overset{\mathsf{D}^{\perp}}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G^{\star}_{\mathsf{D}^{\perp}}(\gamma')(\lambda) + \frac{1}{2}F^{\star}_{\mathsf{D}^{\perp}}(\gamma')(\lambda), \end{aligned}$$

along with the condition that λ be nowhere zero. By Lemma 2.4, the conditions of part (ii) are equivalent to the three equations

$$\begin{aligned}
&F^*_{\mathsf{D}}(\gamma')(\lambda) = 0,\\
&\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0,\\
&\overset{\mathsf{D}^\perp}{\nabla}_{\Upsilon'}\hat{\lambda} = A_{\mathsf{D}} \circ \hat{\lambda},
\end{aligned}$$

along with the condition that λ be nowhere zero. The proposition follows immediately from the definition of A_{D}. ■

Now let us consider the D-regular case.

7.6 Proposition: (D-regular constrained variational trajectories along nonholonomic trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. For a C^r-curve $\hat{\lambda}\colon I \to \pi^*_{\mathsf{D}}\mathsf{D}^\perp$, write $\hat{\lambda}(t) = (\lambda(t), \Upsilon(t))$. Then the following statements are equivalent:*

(i) Υ is an integral curve of $X^{\mathrm{nh}}_{\mathsf{D}}$ and λ is such that $\gamma = \pi_{\mathsf{D}} \circ \Upsilon$ and λ together satisfy the conditions for a D-regular constrained variational trajectory from Theorem 5.22(ii)(b) (or (iii)(b));

(ii) the following conditions hold:

(a) $\hat{\lambda}(t) \in \ker(\hat{F}^*_{\mathsf{D}})_{\Upsilon(t)}$;

(b) $\hat{\lambda}$ is an integral curve of $X^{\mathrm{reg}}_{\mathsf{D}}$.

Proof: Note that (i) is equivalent to the three equations

$$\begin{aligned}
&\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0,\\
&F^*_{\mathsf{D}}(\gamma')(\lambda) = 0,\\
&\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G_{\mathsf{D}}(\gamma',\gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma + \frac{1}{2}G^\star_{\mathsf{D}^\perp}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}^\perp}(\gamma')(\lambda).
\end{aligned}$$

By Lemma 2.4, the conditions of part (ii) are equivalent to the three equations

$$\begin{aligned}
&F^*_{\mathsf{D}}(\gamma')(\lambda) = 0,\\
&\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0,\\
&\overset{\mathsf{D}^\perp}{\nabla}_{\Upsilon'}\hat{\lambda} = A_{\mathsf{D}} \circ \hat{\lambda} + b_{\mathsf{D}} \circ \gamma.
\end{aligned}$$

The proposition follows immediately from the definitions of A_{D} and b_{D}. ■

7.2. Main results. Now we assemble the preceding developments of the paper to prove the main results, giving answers to the questions posed at the beginning of this section.

7.2.1. When are all constrained variational trajectories also nonholonomic trajectories?. Our first result is one that has been observed by many authors, sometimes for more general Lagrangians than we consider here [e.g., Cortés, de León, Martín de Diego, and Martínez 2002, Fernandez and Bloch 2008, Jóźwikowski and Respondek 2019, Kupka and Oliva 2001, Lewis and Murray 1995, Terra 2018]. The result does not rely on our results about invariant cogeneralised distributions or affine subbundle varieties from Section 4. Instead, it is proved just by direct comparison of the nonholonomic and constrained variational equations.

The result is the following.

7.7 Theorem: (When all D-regular constrained variational trajectories are nonholonomic trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Then the following statements are equivalent:*

(i) every D-regular constrained variational trajectory is a nonholonomic trajectory;

(ii) D is integrable.

Proof: (i) $\Longrightarrow$ (ii) Suppose that every D-regular constrained variational trajectory is a nonholonomic trajectory and that D is not integrable, i.e., that F_{D} is nonzero, cf. Remark 2.37–3. Then there exists $x \in \mathsf{M}$ and $u_x, v_x \in \mathsf{D}_x$ such that $F_{\mathsf{D}}(u_x, v_x) \neq 0$. Thus there exists $\alpha_x \in \mathsf{D}_x^{\perp}$ such that

$$0 \neq \mathbb{G}(\alpha_x, F_{\mathsf{D}}(u_x, v_x)) = \mathbb{G}(u_x, F_{\mathsf{D}}^*(v_x)(\alpha_x)).$$

Then we have $F_{\mathsf{D}}^*(v_x)(\alpha_x) \neq 0$. Therefore, if $\gamma\colon I \to \mathsf{M}$ is a D-regular constrained variational trajectory satisfying $\gamma'(t_0) = v_x$ for some $t_0 \in I$ and if $\lambda\colon I \to \mathsf{D}^{\perp}$ is the corresponding section over γ satisfying $\lambda(t_0) = \alpha_x$, then we have

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma'(t_0) + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma(t_0) = F_{\mathsf{D}}^*(v_x)(\alpha_x) \neq 0$$

by Theorem 5.22(iii)(b). By continuity,

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma'(t) + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma(t) \neq 0$$

for t sufficiently close to t_0. This, then prohibits γ from being a nonholonomic trajectory.

(ii) $\Longrightarrow$ (i) If D is integrable, then the Frobenius curvature F_{D} vanishes, as above. It then follows from Theorem 5.22(iii) that, if γ is a D-regular constrained variational trajectory, then

$$\overset{\mathsf{D}}{\nabla}_{\gamma'}\gamma' + P_{\mathsf{D}} \circ \operatorname{grad} V \circ \gamma = 0,$$

and so γ is a nonholonomic trajectory by Theorem 5.18(iii). ■

Note that the question of when all D-singular constrained variational trajectories are nonholonomic trajectories does not make sense in our context, since the condition (SCV) for D-singular variational trajectories does not determine conditions for γ. It is possible that there are constrained simple mechanical systems, all of whose D-singular constrained variational trajectories are nonholonomic trajectories, but the determining of conditions for this would require studying higher-order conditions beyond the essentially first-order conditions yielded by Theorem 5.22. This is not something we do here.

7.2.2. The classification of all nonholonomic trajectories that are also constrained variational trajectories. Now we turn to results that make use of our invariance results of Section 4. So suppose that $r \in \{\infty, \omega\}$ and that we have a C^r-constrained mechanical system $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ giving rise to the following data:

1. the vector field $X_{\mathsf{D}}^{\mathrm{nh}}$ on D;
2. the linear connection $\overset{\mathsf{D}^\perp}{\nabla}{}^*$ in the vector bundle $\pi_{\mathsf{D}}^*\pi_{\mathsf{D}^\perp} \colon \pi_{\mathsf{D}}^*\mathsf{D}^\perp \to \mathsf{D}$;
3. the section $b_{\mathsf{D}} \in \Gamma^r(\pi_{\mathsf{D}}^*\mathsf{D}^\perp)$;
4. the section $A_{\mathsf{D}} \in \Gamma^r(\mathrm{End}(\pi_{\mathsf{D}}^*\mathsf{D}^\perp))$;
5. the C^r-cogeneralised subbundle $\ker(\hat{F}_{\mathsf{D}}^*)$ of $\pi_{\mathsf{D}}^*\mathsf{D}^\perp$.

We apply the results of Section 4.6, taking

1. $\mathsf{M} = \mathsf{D}$,
2. $\mathsf{E} = \pi_{\mathsf{D}}^*\mathsf{D}^\perp$,
3. $\pi = \pi_{\mathsf{D}}^*\pi_{\mathsf{D}^\perp}$,
4. $\nabla = \overset{\mathsf{D}^\perp}{\nabla}{}^*$,
5. $X_0 = X_{\mathsf{D}}^{\mathrm{nh}}$,
6. $b = b_{\mathsf{D}}$,
7. $A = A_{\mathsf{D}}$, and
8. $\mathsf{F} = \ker(\hat{F}_{\mathsf{D}}^*)$

(to the left of the equals sign are the objects from Section 4 and to the right of the equals sign are the objects in our current setting).

We wish to find all initial conditions $v \in \mathsf{D}$ with the property that there exists an initial condition $(\alpha, v) \in \pi_{\mathsf{D}}^*\mathsf{D}^\perp$ such that, if $t \mapsto \Upsilon(t)$ is the integral curve of $X_{\mathsf{D}}^{\mathrm{nh}}$ through v, the integral curve $t \mapsto \hat{\lambda} = (\lambda(t), \tilde{\Upsilon}(t))$ of either $X_{\mathsf{D}}^{\mathrm{sing}}$ or $X_{\mathsf{D}}^{\mathrm{reg}}$ through (α, v) satisfies $\hat{\lambda}(t) \in \ker(\hat{F}_{\mathsf{D}}^*)$ for each t. If this is so, then we automatically have $\tilde{\Upsilon} = \Upsilon$. Thus this gives a solution for either (SCV) or (RCV) that projects to a solution for (NH). Thus the classification of solutions to either (SCV) or (RCV) that projects to a solution for (NH) amounts to finding suitable initial conditions (α, v) for either $X_{\mathsf{D}}^{\mathrm{sing}}$ or $X_{\mathsf{D}}^{\mathrm{reg}}$. Our observations just preceding, combined with Theorems 4.22 and 4.23, are that the set of all such initial conditions is the largest cogeneralised subbundle or affine subbundle variety contained in $\ker(\hat{F}_{\mathsf{D}}^*)$ and invariant under either $X_{\mathsf{D}}^{\mathrm{sing}}$ or $X_{\mathsf{D}}^{\mathrm{reg}}$, respectively.

Now, Remarks 4.27 and 4.28 are culminants of our lengthy discussion characterising exactly such invariant cogeneralised subbundles and affine subbundle varieties. The notions, from these remarks, of a $(X_{\mathsf{D}}^{\mathrm{sing}}, \ker(\hat{F}_{\mathsf{D}}^*))$-admissible cogeneralised subbundle and of a $(X_{\mathsf{D}}^{\mathrm{reg}}, \ker(\hat{F}_{\mathsf{D}}^*))$-admissible defining subbundle are ones we shall make reference to in our results below. These are merely literal transcriptions of the conditions of Remarks 4.27 and 4.28 using our existing notation.

Let us state two results, one for the regular case and one for the singular case, that relate $(X_{\mathsf{D}}^{\mathrm{reg}}, \ker(\hat{F}_{\mathsf{D}}^*))$-admissible affine subbundle varieties and $(X_{\mathsf{D}}^{\mathrm{sing}}, \ker(\hat{F}_{\mathsf{D}}^*))$-admissible cogeneralised subbundles to flow-invariant subbundles for either $X_{\mathsf{D}}^{\mathrm{reg}}$ or $X_{\mathsf{D}}^{\mathrm{sing}}$.

First we consider the regular case.

7.8 Theorem: (When some nonholonomic trajectories are D-regular constrained variational trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Consider the following statements:*

(i) some nonholonomic trajectories are D-regular constrained variational trajectories;

(ii) there exists a C^r-affine subbundle variety $\mathsf{A} \subseteq \ker(\hat{F}^*_{\mathsf{D}})$ that is flow-invariant under $X^{\mathrm{reg}}_{\mathsf{D}}$;

(iii) there exists a partial $(X^{\mathrm{reg}}_{\mathsf{D}}, \ker(\hat{F}^*_{\mathsf{D}}))$-admissible C^r-defining subbundle $\Delta \subseteq (\pi^*_{\mathsf{D}}\mathsf{D}^{\perp})^* \oplus \mathbb{R}_{\mathsf{D}}$.

Then

(iv) (i) $\Longrightarrow$ (ii) if $r = \infty$ or if $X^{\mathrm{nh}}_{\mathsf{D}}$ is complete,

(v) (ii) $\Longrightarrow$ (i),

(vi) (ii) $\Longrightarrow$ (iii), and

(vii) (iii) $\Longrightarrow$ (ii) if either (a) $r = \omega$ or (b) $r = \infty$ and $\ker(\hat{F}^*_{\mathsf{D}})$ has locally constant rank.

Proof: If (i) holds, then, by Proposition 7.6, there is an integral curve $\hat{\Upsilon}$ of $X^{\mathrm{reg}}_{\mathsf{D}}$ over Υ for which $\mathrm{image}(\hat{\Upsilon}) \subseteq \ker(\hat{F}^*_{\mathsf{D}})$. If $r = \infty$ or if $X^{\mathrm{nh}}_{\mathsf{D}}$ is complete, by Theorem 4.23 we conclude that there is a C^r-cogeneralised affine subbundle of $\ker(\hat{F}^*_{\mathsf{D}})$ that is flow-invariant under $X^{\mathrm{reg}}_{\mathsf{D}}$. This shows that (ii) holds.

If (ii) holds, then there is an integral curve $\hat{\Upsilon}$ of $X^{\mathrm{reg}}_{\mathsf{D}}$ over $X^{\mathrm{nh}}_{\mathsf{D}}$ with values in $\ker(\hat{F}^*_{\mathsf{D}})$. By Proposition 7.6, it follows that this integral curve is a D-regular constrained variational trajectory, showing that (i) holds.

The conditions of part (iii) are just those of Theorem 4.26 that are equivalent to the existence of a C^r-defining subbundle $\Delta \subseteq \Lambda(\widehat{\ker(\hat{F}^*_{\mathsf{D}})})$ that is flow-invariant under the linear vector field on $(\pi^*_{\mathsf{D}}\mathsf{D}^{\perp})^* \oplus \mathbb{R}_{\mathsf{D}}$ associated with the affine vector field $X^{\mathrm{reg}}_{\mathsf{D}}$. Since Δ is partial, by Proposition 4.20, $\mathsf{A}(\Delta)$ is nonempty. By Lemma 4.18(ii), $\mathsf{A}(\Delta)$ is flow-invariant. This gives (ii). ■

Now we consider when all nonholonomic trajectories are D-singular constrained variational trajectories.

7.9 Theorem: (When some nonholonomic trajectories are D-singular constrained variational trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Consider the following statements:*

(i) some nonholonomic trajectories are D-singular constrained variational trajectories;

(ii) there exists a C^r-cogeneralised subbundle $\mathsf{L} \subseteq \ker(\hat{F}^*_{\mathsf{D}})$ that is flow-invariant under X^{sing};

(iii) there exists a $(X^{\mathrm{sing}}_{\mathsf{D}}, \ker(\hat{F}^*_{\mathsf{D}}))$-admissible C^r-cogeneralised subbundle $\mathsf{L} \subseteq \pi^*_{\mathsf{D}}\mathsf{D}^{\perp}$.

Then

(iv) (i) $\Longrightarrow$ (ii) if $r = \infty$ or if $X^{\mathrm{nh}}_{\mathsf{D}}$ is complete,

(v) (ii) $\Longrightarrow$ (i),

(vi) (ii) $\Longrightarrow$ (iii), and

(vii) (iii) $\Longrightarrow$ (ii) if either (a) $r = \omega$ or (b) $r = \infty$ and $\ker(\hat{F}^*_{\mathsf{D}})$ has locally constant rank.

Proof: The proof follows from Theorems 4.22 and 4.24, and Proposition 7.5, in the same way as Theorem 7.8 follows from Theorems 4.23 and 4.26, and Proposition 7.6, noting that $X_{\mathsf{D}}^{\mathrm{sing}}$ is a linear vector field over $X_{\mathsf{D}}^{\mathrm{nh}}$. ■

7.2.3. When is a given nonholonomic trajectory also a constrained variational trajectory?. We now consider the matter of when single nonholonomic trajectories are constrained variational trajectories. One can certainly make use of the general constructions in the preceding section, asking whether the initial condition for the nonholonomic trajectory is covered by a suitable initial condition for the constrained variational trajectory. However, because we are focussing on a single trajectory, the problem can be reduced, and so this should be done.

We start with the D-regular case. Some words about this are printed in [Terra 2018], but a conclusive statement such as we now give is not quite given by Terra.

7.10 Theorem: (When a nonholonomic trajectory is a D-regular constrained variational trajectory) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. For a nonholonomic trajectory $\gamma \colon I \to \mathsf{M}$ with $\Upsilon = \gamma'$, consider the following statements:*

(i) γ is a D-regular constrained variational trajectory;

(ii) there exists a C^r-affine subbundle variety $\mathsf{A} \subseteq \Upsilon^* \ker(\hat{F}_{\mathsf{D}}^*)$ that is invariant under $\Upsilon^* X_{\mathsf{D}}^{\mathrm{reg}}$;

(iii) there exists a partial $(\Upsilon^* X_{\mathsf{D}}^{\mathrm{reg}}, \Upsilon^* \ker(\hat{F}_{\mathsf{D}}^*))$-admissible C^r-defining subbundle $\Delta \subseteq \Upsilon^*(\pi_{\mathsf{D}}^* \mathsf{D}^{\perp} \oplus \mathbb{R}_{\mathsf{D}})$.

Then

(iv) (i) $\Longrightarrow$ (ii) if $r = \infty$ or if X^{nh} is complete,

(v) (ii) $\Longrightarrow$ (i),

(vi) (ii) $\Longrightarrow$ (iii), and

(vii) (iii) $\Longrightarrow$ (ii) if either (a) $r = \omega$ or (b) $r = \infty$ and $\Upsilon^* \ker(\hat{F}_{\mathsf{D}}^*)$ has locally constant rank.

Proof: We shall use Theorems 4.23 and 4.26, after pulling all data back from D to I by Υ. We begin by performing all of the required pull-backs, and giving the properties of these.

In order to temporarily unburden the notation, we shall make use of the abbreviations suggested above:

$$\mathsf{M} = \mathsf{D},\ \ \mathsf{E} = \pi_{\mathsf{D}}^* \mathsf{D}^{\perp},\ \ \pi = \pi_{\mathsf{D}}^* \pi_{\mathsf{D}^{\perp}},\ \ \nabla = \overset{\mathsf{D}^{\perp}}{\nabla}{}^*,$$
$$X_0 = X_{\mathsf{D}}^{\mathrm{nh}},\ \ b = b_{\mathsf{D}},\ \ A = A_{\mathsf{D}},\ \ \mathsf{F} = \ker(\hat{F}_{\mathsf{D}}^*),\ \ X^{\mathrm{aff}} = X_{\mathsf{D}}^{\mathrm{reg}}.$$

With this notation, we make the following constructions.

1. Let $\Upsilon^*\pi \colon \Upsilon^*\mathsf{E} \to I$ and $\Upsilon^*\pi^* \colon \Upsilon^*\mathsf{E}^* \to I$ be the pull-back bundles:

$$\Upsilon^*\mathsf{E} = \{(e,t) \mid \Upsilon(t) = \pi(e)\}, \quad \Upsilon^*\mathsf{E}^* = \{(\alpha,t) \mid \Upsilon(t) = \pi^*(\alpha)\}.$$

2. Let $\tau \colon I \to \mathsf{T}I$ be the standard vector field $\tau(s) = (t,1)$. We work with the integral curve σ of τ given by $\sigma(t) = t$. Thus $\Upsilon = \Upsilon \circ \sigma$ and so $\Upsilon' = T\Upsilon \circ \sigma'$.

3. Let $\overset{\Upsilon}{\nabla}$ be the pull-back of ∇ to I by Υ. Thus

$$\overset{\Upsilon}{\nabla}_{\sigma'(t)}\pi^*\xi(t) = \nabla_{T_t\Upsilon(\sigma'(t))}\xi \circ \Upsilon(t) = \nabla_{\Upsilon'(t)}\xi \circ \Upsilon(t).$$

4. Define $\Upsilon^*b \in \Gamma^r(\Upsilon^*\mathsf{E})$ by

$$\Upsilon^*b(t) = (b \circ \Upsilon(t), t), \qquad t \in I.$$

The vertical lift of Υ^*b is then

$$(\Upsilon^*b)^{\mathrm{v}}(e,t) = ((b^{\mathrm{v}}(e), 0), (e,t)),$$

noting that

$$\mathsf{T}(\Upsilon^*\mathsf{E}) \subseteq \mathsf{T}(I \times \mathsf{E}).$$

5. Define $\Upsilon^*A \in \Gamma^r(\Upsilon^*\mathrm{End}(\mathsf{E}))$ by

$$\Upsilon^*A(t) = (A \circ \Upsilon(t), t), \qquad t \in I.$$

The vertical evaluation of Υ^*A is then

$$(\Upsilon^*A)^{\mathrm{e}}(e,t) = ((A^{\mathrm{e}}(e), 0), (e,t)).$$

6. Define $\Upsilon^*X_0^{\mathrm{h}} \in \Gamma^r(\mathsf{T}(\Upsilon^*\mathsf{E}))$ by

$$\Upsilon^*X_0^{\mathrm{h}}(e,t) = ((X_0^{\mathrm{h}}(e), 1), (e,t)).$$

We claim that $\Upsilon^*X_0^{\mathrm{h}}$ is the horizontal lift of τ. Indeed, since

$$\overset{\Upsilon}{\nabla}_{\sigma'(t)}\pi^*\xi \circ \Upsilon(t) = \nabla_{\Upsilon'(t)}\xi \circ \Upsilon(t),$$

the image of parallel translation along σ is parallel translation along Υ [Kobayashi and Nomizu 1963, §III.1], our claim follows from Lemma 2.3(iv).

7. If we define $\Upsilon^*X^{\mathrm{aff}} \in \Gamma^r(\mathsf{T}(\Upsilon^*\mathsf{E}))$ by

$$\Upsilon^*X^{\mathrm{aff}}(e,t) = ((X^{\mathrm{aff}}(e), 1), (e,t)),$$

then we have

$$\Upsilon^*X^{\mathrm{aff}} = \Upsilon^*X_0^{\mathrm{h}} + (\Upsilon^*A)^{\mathrm{e}} + (\Upsilon^*b)^{\mathrm{v}}.$$

Thus $\Upsilon^*X^{\mathrm{aff}}$ is an affine vector field on $\Upsilon^*\mathsf{E}$.

8. The integral curves of $\Upsilon^*X^{\mathrm{aff}}$ are of the form

$$t \mapsto (\hat{\Upsilon}(t), 1) \triangleq \Upsilon^*\hat{\Upsilon}(t),$$

where $\hat{\Upsilon}$ is an integral curve of X^{aff} that projects to Υ.

9. Denote

$$\Upsilon^*\mathsf{F} = \{(e,t) \in \Upsilon^*\mathsf{E} \mid \ e \in \mathsf{F}\}$$

and

$$\Upsilon^*\Lambda(\mathsf{F}) = \{(\alpha,t) \in \Upsilon^*\mathsf{E}^* \mid \ \alpha \in \Lambda(\mathsf{F})\}.$$

10. If $\hat{\Upsilon}\colon I \to \mathsf{E}$ is a curve for which $\pi \circ \hat{\Upsilon} = \Upsilon$, then $\mathrm{image}(\Upsilon^*\hat{\Upsilon}) \in \Upsilon^*\mathsf{F}$ if and only if $\mathrm{image}(\hat{\Upsilon}) \subseteq \mathsf{F}$.

We now revert back to the unabbreviated notation. Note that $\Upsilon^*A = A_\Upsilon$ and $\Upsilon^*b = b_\Upsilon$.

If (i) holds, then, by Proposition 7.6, there is an integral curve $\hat{\Upsilon}$ of $X^{\mathrm{reg}}_{\mathsf{D}}$ over Υ for which $\mathrm{image}(\hat{\Upsilon}) \subseteq \ker(\hat{F}^*_{\mathsf{D}})$. As pointed out in 10, this implies that $\mathrm{image}(\Upsilon^*\hat{\Upsilon}) \subseteq \Upsilon^*\ker(\hat{F}^*_{\mathsf{D}})$. If $r = \infty$ or if $X^{\mathrm{nh}}_{\mathsf{D}}$ is complete, by Theorem 4.23 we conclude that there is a C^r-cogeneralised affine subbundle of $\Upsilon^*\ker(\hat{F}^*_{\mathsf{D}})$ that is flow-invariant under $\Upsilon^*X^{\mathrm{reg}}_{\mathsf{D}}$. This shows that (ii) holds.

If (ii) holds, then there is an integral curve $\Upsilon^*\hat{\Upsilon}$ of $\Upsilon^*X^{\mathrm{reg}}_{\mathsf{D}}$ over τ with values in $\Upsilon^*\ker(\hat{F}^*_{\mathsf{D}})$. By the observation of 10, this implies that $\hat{\Upsilon}$ is an integral curve of $X^{\mathrm{reg}}_{\mathsf{D}}$ with values in $\ker(\hat{F}^*_{\mathsf{D}})$. By Proposition 7.6, it follows that this integral curve is a D-regular constrained variational trajectory.

The conditions of part (iii) are just those of Theorem 4.26 that are equivalent to the existence of a C^r-defining subbundle $\Delta \subseteq \Upsilon^*\Lambda(\widehat{\ker(\hat{F}^*_{\mathsf{D}})})$ that is flow-invariant under the linear vector field on $\Upsilon^*((\pi^*_{\mathsf{D}}\mathsf{D}^\perp)^* \oplus \mathbb{R}_{\mathsf{D}})$ associated with the affine vector field $\Upsilon^*X^{\mathrm{reg}}_{\mathsf{D}}$. Since Δ is partial, by Proposition 4.20, $\mathsf{A}(\Delta)$ is nonempty. By Lemma 4.18(ii), $\mathsf{A}(\Delta)$ is flow-invariant. This gives (ii). ■

Now we consider the D-singular case, of which there is no discussion in the existing literature.

7.11 Theorem: (When a nonholonomic trajectory is a D-singular constrained variational trajectory) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. For a nonholonomic trajectory $\gamma\colon I \to \mathsf{M}$ with $\Upsilon = \gamma'$, consider the following statements:*

(i) γ is a D-singular constrained variational trajectory;

(ii) there exists a C^r-cogeneralised subbundle $\mathsf{L} \subseteq \Upsilon^*\ker(\hat{F}^*_{\mathsf{D}})$ that is flow-invariant under $\Upsilon^*X^{\mathrm{sing}}_{\mathsf{D}}$;

(iii) there exists a partial $(\Upsilon^*X^{\mathrm{sing}}_{\mathsf{D}}, \Upsilon^*\ker(\hat{F}^*_{\mathsf{D}}))$-admissible C^r-cogeneralised subbundle $\mathsf{L} \subseteq \Upsilon^*\pi^*_{\mathsf{D}}\mathsf{D}^\perp$.

Then

(iv) (i) $\Longrightarrow$ (ii) if $r = \infty$ or if $X^{\mathrm{nh}}_{\mathsf{D}}$ is complete,

(v) (ii) $\Longrightarrow$ (i),

(vi) (ii) $\Longrightarrow$ (iii), and

(vii) (iii) $\Longrightarrow$ (ii) if either (a) $r = \omega$ or (b) $r = \infty$ and $\Upsilon^*\ker(\hat{F}^*_{\mathsf{D}})$ has locally constant rank.

Proof: The proof follows from Theorems 4.22 and 4.24, and Proposition 7.5, in the same way as Theorem 7.10 follows from Theorems 4.23 and 4.26, and Proposition 7.6, noting that $X^{\mathrm{sing}}_{\mathsf{D}}$ is a linear vector field over $X^{\mathrm{nh}}_{\mathsf{D}}$. ■

7.2.4. When are all nonholonomic trajectories also constrained variational trajectories?. Now we consider the situation where all nonholonomic trajectories are constrained variational trajectories. The results here follow easily along the lines of Theorems 7.8 and 7.9.

We consider first the D-regular case.

7.12 Theorem: (When all nonholonomic trajectories are D-regular constrained variational trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Consider the following statements:*

(i) all nonholonomic trajectories are D*-regular constrained variational trajectories;*

(ii) there exists a C^r-affine subbundle variety $\mathsf{A} \subseteq \ker(\hat{F}^*_{\mathsf{D}})$ that is flow-invariant under $X^{\mathrm{reg}}_{\mathsf{D}}$ and such that $\mathsf{S}(\mathsf{A}) = \mathsf{M}$;

(iii) there exists a total $(X^{\mathrm{reg}}_{\mathsf{D}}, \ker(\hat{F}^*_{\mathsf{D}}))$-admissible C^r-defining subbundle $\Delta \subseteq \pi^*_{\mathsf{D}}\mathsf{D}^{\perp} \oplus \mathbb{R}_{\mathsf{D}}$.

Then

(iv) (i) $\Longrightarrow$ (ii) if $r = \infty$ or if $X^{\mathrm{nh}}_{\mathsf{D}}$ is complete,

(v) (ii) $\Longrightarrow$ (i),

(vi) (ii) $\Longrightarrow$ (iii), and

(vii) (iii) $\Longrightarrow$ (ii) if either (a) $r = \omega$ or (b) $r = \infty$ and $\Upsilon^* \ker(\hat{F}^*_{\mathsf{D}})$ has locally constant rank.

Proof: This is an obvious modification of the proof of Theorem 7.8. ■

Now we consider when all nonholonomic trajectories are D-singular constrained variational trajectories.

7.13 Theorem: (When all nonholonomic trajectories are D-singular constrained variational trajectories) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Consider the following statements:*

(i) all nonholonomic trajectories are D*-singular constrained variational trajectories;*

(ii) there exists a C^r-cogeneralised subbundle $\mathsf{L} \subseteq \ker(\hat{F}^*_{\mathsf{D}})$ that is flow-invariant under X^{sing};

(iii) there exists a total $(X^{\mathrm{sing}}_{\mathsf{D}}, \ker(\hat{F}^*_{\mathsf{D}}))$-admissible C^r-cogeneralised subbundle $\mathsf{L} \subseteq \pi^*_{\mathsf{D}}\mathsf{D}^{\perp}$.

Then

(iv) (i) $\Longrightarrow$ (ii) if $r = \infty$ or if $X^{\mathrm{nh}}_{\mathsf{D}}$ is complete,

(v) (ii) $\Longrightarrow$ (i),

(vi) (ii) $\Longrightarrow$ (iii), and

(vii) (iii) $\Longrightarrow$ (ii) if either (a) $r = \omega$ or (b) $r = \infty$ and $\ker(\hat{F}^*_{\mathsf{D}})$ has locally constant rank.

Proof: This is an obvious modification of the proof of Theorem 7.9. ■

7.3. Recovery of existing results. In this section we give an overview of some known results giving conditions under which some/every nonholonomic trajectory is a constrained variational trajectory. Some of the results we cite are proved in the references for more general Lagrangians, but our proofs only apply for kinetic energy minus potential energy Lagrangians. We also transform all existing results into our language, sometimes as a consequence making the result trivial.

Here are the results in chronological order.

In the first result by Favretti [1998], we recall the terminology that, given a Riemannian manifold $(\mathsf{M}, \mathbb{G})$ and a foliation $\mathscr{F}$, the metric $\mathbb{G}$ is ***bundle-like*** for $\mathscr{F}$ if the distribution orthogonal to $\mathscr{F}$ is geodesically invariant for the Levi-Civita connection $\overset{\mathbb{G}}{\nabla}$.

7.14 Corollary: (**[Favretti 1998, Theorem 3.2]**) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, 0, \mathsf{D})$ be a C^r-constrained simple mechanical system and suppose that D is the orthogonal distribution for a foliation $\mathscr{F}$ of M for which $\mathbb{G}$ is bundle-like. Then every nonholonomic trajectory is a D-regular constrained variational trajectory.*

Proof: Under the hypothesis that D is geodesically invariant for $\overset{\mathbb{G}}{\nabla}$ and that $V = 0$, $b_{\mathsf{D}} = 0$. We can then apply Theorem 7.12 with A the zero section. ■

Note that Favretti actually requires more than is needed, since there is no need for D to be the orthogonal distribution of a foliation for which $\mathbb{G}$ is bundle-like. All we require is that D be geodesically invariant for $\overset{\mathbb{G}}{\nabla}$.

Our next result has to do with certain nonholonomic trajectories that are also constrained variational trajectories.

7.15 Corollary: (**[Cortés, de León, Martín de Diego, and Martínez 2002, Proposition 6.2]**) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. If γ is a nonholonomic trajectory for Σ that is also a nonholonomic trajectory for the unconstrained system $\Sigma' = (\mathsf{M}, \mathbb{G}, V, \mathsf{TM})$, then γ is a D-regular constrained variational trajectory.*

Proof: The hypotheses are that $P_{\mathsf{D}^\perp} \circ \gamma = 0$ and

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V \circ \gamma = 0.$$

Therefore, applying $P_{\mathsf{D}^\perp}$ to the preceding equation, we have

$$\frac{1}{2} G_{\mathsf{D}}(\gamma', \gamma') + P_{\mathsf{D}^\perp} \circ \operatorname{grad} V \circ \gamma = 0.$$

Thus $b_\Upsilon = 0$, with $\Upsilon = \gamma'$. We can then apply Theorem 7.10 with A the zero section. ■

The next result we state is one that has been observed by many authors in many different ways. It is an essentially obvious result, but it is worth pointing to a few occurrences of it in order to make connections between various approaches. One such statement is given by Fernandez and Bloch [2008], and requires substantial translation to get from the stated result to something in our terminology. Crampin and Mestdag [2010] give a version of the result as their Corollary 1, although their setup is rather different than ours. Another occurrence is in the paper of Langerok [2003], and is given in a setting more reminiscent of our approach. The result of Fernandez and Bloch is stated for general Lagrangians, while the result of Langerok is given in the setting of sub-Riemannian geometry, and so applies only to kinetic energy Lagrangians. Fernandez and Bloch also attribute the result to Rumiantsev [1978], but we could not locate such a statement in Rumiantsev's paper.[10]

7.16 Corollary: (**e.g., [Langerok 2003, Proposition 37], [Fernandez and Bloch 2008, Proposition 2]**) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D})$ be a C^r-constrained simple mechanical system. Then a nonholonomic trajectory $\gamma\colon I \to \mathsf{M}$ is a D-regular constrained variational trajectory if and only if there exists a smooth section $\lambda\colon I \to \mathsf{D}^\perp$ along γ that satisfies*

[10] This is different than saying the statement is not there.

(i) $\overset{\mathsf{D}^\perp}{\nabla}_{\gamma'}\lambda = \frac{1}{2}G_{\mathsf{D}}(\gamma',\gamma') + P_{\mathsf{D}^\perp}\circ \operatorname{grad} V\circ\gamma + \frac{1}{2}G^\star_{\mathsf{D}}(\gamma')(\lambda) + \frac{1}{2}F^\star_{\mathsf{D}}(\gamma')(\lambda)$ *and*

(ii) $\mathbb{G}(\lambda, F_{\mathsf{D}}(\gamma',\xi)) = 0$ *for every smooth section* ξ *of* D *along* γ*.*

We note that Langerok uses a particular linear connection for which the second and third term on the right in the equation for the adjoint field are absorbed by the covariant derivative term on the left.

The next result from the literature we consider concerns a special class of constrained simple mechanical systems.

7.17 Definition: (Chaplygin system) Let $r\in\{\infty,\omega\}$. A ***C***^{r}***-Chaplygin system*** is a quintuple $(\mathsf{M},\mathbb{G},V,\mathsf{D},\mathsf{G})$ where

(i) $(\mathsf{M},\mathbb{G},V,\mathsf{D})$ is a C^r-constrained simple mechanical system and

(ii) G is a Lie group equipped with a C^r-left-action $\Psi\colon \mathsf{G}\times\mathsf{M}\to\mathsf{M}$ for which

  (a) if $\pi\colon \mathsf{M}\to\mathsf{M}/\mathsf{G}$ is the projection onto the orbit space, this is a principal G-bundle,

  (b) $\mathbb{G}$ is G-invariant in that $\Psi_g^*\mathbb{G}=\mathbb{G}$ for every $g\in\mathsf{G}$,

  (c) V is G-invariant in that $\Psi_g^*V = V$ for every $g\in\mathsf{G}$, and

  (d) D is a principal connection on $\pi\colon\mathsf{M}\to\mathsf{M}/\mathsf{G}$ in that

    I. $\mathsf{TM}=\mathsf{D}\oplus\ker(T\pi)$ and

    II. $\mathsf{D}_{\Psi_g(x)} = T\Psi_g(\mathsf{D}_x)$ for every $x\in\mathsf{M}$ and $g\in\mathsf{G}$. •

Chaplygin systems are treated in a few places in the literature. The idea seems to have originated in the paper of Koiller [1992], and we can recommend the presentation in [Cantrijn, Cortés, de León, and Martín de Diego 2002] as it gives an affine connection formalism resembling ours.

Let us overview the structure that arises from a Chaplygin system. We denote by ver, hor: $\mathsf{TM}\to\mathsf{TM}$ the projections onto $\ker(T\pi)$ and D, respectively. We denote the Lie algebra of G by $\mathfrak{g}$. For each $\xi\in\mathfrak{g}$, its ***infinitesimal generator*** is the vector field $\xi_{\mathsf{M}}\in\Gamma^r(\mathsf{TM})$ defined by

$$\xi_{\mathsf{M}}(x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\Psi_{\exp(t\xi)}(x).$$

The ***connection form*** is the $\mathfrak{g}$-valued Ad-equivariant one-form ω on M defined by

$$\operatorname{ver}(v) = (\omega(v))_{\mathsf{M}}(\pi_{\mathsf{TM}}(v)).$$

We note that

$$\mathsf{D} = \ker(\omega) = \ker(\operatorname{ver}) = \ker(P_{\mathsf{D}^\perp}),$$

although $\operatorname{image}(\operatorname{ver})\neq\operatorname{image}(P_{\mathsf{D}^\perp})$, in general. The ***curvature form*** is $\Omega\in\mathfrak{g}\otimes\bigwedge^2(\mathsf{T}^*\mathsf{M})$ defined by

$$\Omega(U,V) = \omega([\operatorname{hor}(U),\operatorname{hor}(V)]),\qquad U,V\in\Gamma^r(\mathsf{TM}).$$

Correspondingly with the constructions of Section 2.11, for $Y\in\Gamma^r(\mathsf{D})$ and for $\mu\in\mathfrak{g}^*$, we denote by $\Omega^*(Y)(\mu)\in\Gamma^r(\mathsf{D})$ the D-valued vector field defined by

$$\langle\mu;\Omega(X,Y)\rangle = \mathbb{G}(\Omega^*(Y)(\mu),X),\qquad X\in\Gamma^r(\mathsf{D}).$$

We also have the ***momentum map*** which is the Ad^*-equivariant function

$$J\colon \mathsf{TM} \to \mathfrak{g}^*$$

defined by

$$\langle J(v_x); \xi\rangle = \mathbb{G}(v_x, \xi_{\mathsf{M}}(x)), \qquad v_x \in \mathsf{T}_x\mathsf{M},\ x \in \mathsf{M}.$$

Given $X \in \Gamma^r(\mathsf{T}(\mathsf{M}/\mathsf{G}))$, we denote by X^{h} the unique D-valued vector field on M for which $T_x\pi(X^{\mathrm{h}}(x)) = X(x)$ for $x \in \mathsf{M}$.

In the literature, for Chaplygin systems the adjoint field is not presented as a section of $\mathsf{D}^{\perp}$ along γ, but rather as a $\mathfrak{g}^*$-valued function along γ. This corresponds to the choice in the literature of $\ker(T\pi)$ as a complement to D, as opposed to our choice of $\mathsf{D}^{\perp}$. The following lemma is useful for comparing known results to our approach.

7.18 Lemma: (Constrained variational trajectories via the momentum map) *Let $r \in \{\infty, \omega\}$ and let $(\mathsf{M}, \mathbb{G}, V, \mathsf{D}, \mathbb{G})$ be a C^r-Chaplygin system. Then, for a smooth curve $\gamma\colon I \to \mathsf{M}$, the following statements are equivalent:*

(i) γ is a D-regular constrained variational trajectory;

(ii) there exists a smooth function $\mu\colon I \to \mathfrak{g}^*$ such that

$$\begin{aligned}&\mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\gamma'(t) + \operatorname{grad} V \circ \gamma(t), X^{\mathrm{h}} \circ \gamma(t)) = \mathbb{G}(\Omega^*(\gamma'(t))(\mu(t)), X^{\mathrm{h}} \circ \gamma(t)),\\ &\frac{\mathrm{d}}{\mathrm{d}t}\langle J \circ \lambda(t); \xi\rangle = \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)}\gamma'(t), \xi_{\mathsf{M}} \circ \gamma(t))\end{aligned}$$

for every $t \in I$, $X \in \Gamma^r(\mathsf{T}(\mathsf{M}/\mathsf{G}))$, and $\xi \in \mathfrak{g}$.

Moreover, if μ satisfies (ii), then $\lambda\colon I \to \mathsf{D}^{\perp}$ defined by $\mu = J \circ \lambda$ is an adjoint field for γ from (i).

Proof: Let us make some preliminary calculations.

Note that the curve γ is horizontal, and so there exists $X \in \Gamma^r(\mathsf{T}(\mathsf{M}/\mathsf{G}))$ so that γ is an integral curve for X^{h}. We claim that the Lie bracket $[X^{\mathrm{h}}, \xi_{\mathsf{M}}]$ is zero. First of all, for $g \in \mathsf{G}$, note that both sides of the equation

$$X^{\mathrm{h}}(\Psi_g(x)) = T_x\Psi_g(X^{\mathrm{h}}(x))$$

are horizontal vectors (since the horizontal distribution is G-invariant) at $\Psi_g(x)$ that project to $X(\pi(x))$. Thus $(\Psi_g)_*X^{\mathrm{h}} = X^{\mathrm{h}}$. Now, by [Abraham, Marsden, and Ratiu 1988, Theorem 4.2.19],

$$[\xi_{\mathsf{M}}, X^{\mathrm{h}}](x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\Phi_t^{\xi_{\mathsf{M}}})^* X^{\mathrm{h}}(x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \Psi^*_{\exp(t\xi)} X^{\mathrm{h}}(x) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} X^{\mathrm{h}}(x) = 0.$$

Now we have

$$\overset{\mathrm{G}}{\nabla}_{X^{\mathrm{h}}}\xi_{\mathsf{M}} = \overset{\mathrm{G}}{\nabla}_{\xi_{\mathsf{M}}}X^{\mathrm{h}} + [X^{\mathrm{h}}, \xi_{\mathsf{M}}] = \overset{\mathrm{G}}{\nabla}_{\xi_{\mathsf{M}}}X^{\mathrm{h}}$$

since $\overset{\mathrm{G}}{\nabla}$ is torsion-free.

Note that

$$\operatorname{ver} \circ F_{\mathsf{D}}(U, V) = \Omega(U, V)_{\mathsf{M}} \quad \Longrightarrow \quad F_{\mathsf{D}}(U, V) = P_{\mathsf{D}^{\perp}}(\Omega(U, V)_{\mathsf{M}})$$

for every $U, V \in \Gamma^r(\mathsf{D})$. (This uses the easily verified fact that

$$\mathrm{ver}\,|\mathsf{D}_x^\perp\colon \mathsf{D}_x^\perp \to \ker(T_x\pi), \quad P_{\mathsf{D}^\perp}|\ker(T_x\pi)\colon \ker(T_x\pi) \to \mathsf{D}_x^\perp \tag{7.1}$$

are both isomorphisms with one being the inverse of the other.)

Next we compute, for $\xi \in \mathfrak{g}$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle J\circ\lambda(t);\xi\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{G}(\lambda(t),\xi_{\mathsf{M}}\circ\gamma(t)) \\
&= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{X^{\mathrm{h}}\circ\gamma(t)}\lambda(t),\xi_{\mathsf{M}}\circ\gamma(t)) + \mathbb{G}(\lambda(t),\overset{\mathbb{G}}{\nabla}_{X^{\mathrm{h}}\circ\gamma(t)}\xi_{\mathsf{M}}\circ\gamma(t)) \\
&= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{X^{\mathrm{h}}\circ\gamma(t)}\lambda(t),\xi_{\mathsf{M}}\circ\gamma(t)) + \mathbb{G}(\lambda(t),P_{\mathsf{D}^\perp}(\overset{\mathbb{G}}{\nabla}_{\xi_{\mathsf{M}}\circ\gamma(t)}X^{\mathrm{h}}\circ\gamma(t)) \\
&= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\lambda(t) + S^*_{\mathsf{D}}(\gamma'(t))(\lambda(t)),\xi_{\mathsf{M}}\circ\gamma(t)),
\end{aligned}$$

by virtue of Lemma 2.36(i).

Since V is G-invariant, we have

$$\begin{aligned}
&V\circ\Psi_{\exp(t\xi)}(x) = V(x) \\
\implies &\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} V\circ\Psi_{\exp(t\xi)}(x) = 0 \\
\implies &\langle \mathrm{d}V(x);\xi_{\mathsf{M}}(x)\rangle = 0.
\end{aligned}$$

Now let us use the preceding calculations (without explicitly indicating which ones we use where) to prove the lemma.

Note that, by Theorem 5.22(ii), γ is a D-regular constrained variational trajectory if and only if there exists a smooth section $\lambda\colon I \to \mathsf{D}^\perp$ along γ such that

$$\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V\circ\gamma - \overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda - S^*_{\mathsf{D}}(\gamma')(\lambda) = 0.$$

This equation holds if and only if its inner product with both X^{h} and ξ_{M} are zero for every $X \in \Gamma^r(\mathsf{T}(\mathsf{M}/\mathsf{G}))$ and $\xi \in \mathfrak{g}$.

So let us take said inner products, first with X^{h}. We get

$$\begin{aligned}
0 &= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V\circ\gamma, X^{\mathrm{h}}\circ\gamma) - \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\lambda + S^*_{\mathsf{D}}(\gamma')(\lambda), X^{\mathrm{h}}\circ\gamma) \\
&= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V\circ\gamma, X^{\mathrm{h}}\circ\gamma) - \mathbb{G}(F^*_{\mathsf{D}}(\gamma')(\lambda), X^{\mathrm{h}}\circ\gamma) \\
&= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V\circ\gamma, X^{\mathrm{h}}\circ\gamma) - \mathbb{G}(\lambda, P_{\mathsf{D}^\perp}(\Omega(X^{\mathrm{h}}\circ\gamma,\gamma')_{\mathsf{M}})) \\
&= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'}\gamma' + \operatorname{grad} V\circ\gamma, X^{\mathrm{h}}\circ\gamma) - \mathbb{G}(\Omega^*(\gamma')(J\circ\lambda), X^{\mathrm{h}}\circ\gamma),
\end{aligned}$$

using the computation (5.12). This gives the first of the equations of part (ii).

Now take the inner product with ξ_{M}:

$$\begin{aligned}
0 = \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t) + \operatorname{grad} V\circ\gamma(t), \xi_{\mathsf{M}}\circ\gamma(t)) - \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\lambda(t) + S^*_{\mathsf{D}}(\gamma'(t))(\lambda(t)), \xi_{\mathsf{M}}\circ\gamma(t)) \\
= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t), \xi_{\mathsf{M}}\circ\gamma(t)) - \frac{\mathrm{d}}{\mathrm{d}t}\langle J\circ\lambda(t);\xi(t)\rangle,
\end{aligned}$$

which gives the second of the equations from part (ii), taking $\mu = J \circ \lambda$.

The above gives the implication (i) $\implies$ (ii). For the converse implication, we need to specify $\lambda\colon I \to \mathsf{D}^\perp$ given $\mu\colon I \to \mathfrak{g}^*$. For this, we claim that λ is uniquely prescribed by the equation $\mu = J \circ \lambda$. This, however, follows from the fact that we have the second of the isomorphisms (7.1), since the defining equality

$$\langle \mu; \xi \rangle = \mathbb{G}(\lambda, \xi_{\mathsf{M}} \circ \gamma) = \mathbb{G}(\lambda, P_{\mathsf{D}^\perp} \circ \xi_{\mathsf{M}} \circ \gamma)$$

then uniquely determines λ given μ. ■

In the literature, this observation has appeared, in different contexts and using different notation, as [Crampin and Mestdag 2010, page 175], [Favretti 1998, Proposition 4.1], [Fernandez and Bloch 2008, Proposition 3(1)], and [Jóźwikowski and Respondek 2019, Corollary 4.13].

With all of the preceding development, we can now state the result.

7.19 Corollary: ([Fernandez and Bloch 2008, Proposition 3], [Jóźwikowski and Respondek 2019, Corollary 4.13], [Favretti 1998, Theorem 3.1]) *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, V, \mathsf{D}, \mathsf{G})$ be a C^r-Chaplygin system. Then, for a nonholonomic trajectory $\gamma\colon I \to \mathsf{M}$, the following statements are equivalent:*

(i) γ is a D-regular constrained variational trajectory;

(ii) there exists a smooth function $\mu\colon I \to \mathfrak{g}^*$ that satisfies

$$\begin{aligned} &\langle \mu(t); \Omega(\gamma'(t), X^{\mathrm{h}} \circ \gamma) \rangle = 0, \\ &\frac{\mathrm{d}}{\mathrm{d}t} \langle \mu(t); \xi \rangle = \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)} \gamma'(t), \xi_{\mathsf{M}} \circ \gamma(t)), \end{aligned}$$

for every $t \in I$, $X \in \Gamma^r(\mathsf{T}(\mathsf{M}/\mathsf{G}))$, and $\xi \in \mathfrak{g}$.

Moreover, if either (and so both) of the preceding conditions hold, then

(iii) $J \circ \lambda = J \circ \gamma' + \mathrm{constant}$.

Proof: By Theorem 5.22(ii) and Corollary 7.16, (i) holds if and only if there is a section $\lambda\colon I \to \mathsf{D}^\perp$ along γ such that

$$\begin{aligned} &\mathbb{G}(\lambda, F_{\mathsf{D}}(X^{\mathrm{h}} \circ \gamma, \gamma')), \\ &\overset{\mathrm{G}}{\nabla}_{\gamma'} \gamma' + \operatorname{grad} V \circ \gamma - \overset{\mathrm{G}}{\nabla}_{\gamma'} \lambda - S^*_{\mathsf{D}}(\gamma')(\lambda) = 0 \end{aligned}$$

for every $X \in \Gamma^r(\mathsf{T}(\mathsf{M}/\mathsf{G}))$. Taking inner products of the second of these equations with X^{h} and ξ_{M}, we then show that the preceding equation is equivalent to

$$\begin{aligned} &\langle \mu(t); \Omega(\gamma'(t), X^{\mathrm{h}} \circ \gamma) \rangle = 0, \\ &\mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)} \gamma'(t) + \operatorname{grad} V \circ \gamma(t), X^{\mathrm{h}} \circ \gamma(t)) = 0, \\ &\frac{\mathrm{d}}{\mathrm{d}t} \langle J \circ \lambda(t); \xi \rangle = \mathbb{G}(\overset{\mathrm{G}}{\nabla}_{\gamma'(t)} \gamma'(t), \xi_{\mathsf{M}} \circ \gamma(t)) \end{aligned}$$

for every $t \in I$, $X \in \Gamma^r(\mathsf{T}(\mathsf{M}/\mathsf{G}))$, and $\xi \in \mathfrak{g}$, using the computations from the proof of Lemma 7.18. Since the second of these equations is vacuous as γ is a nonholonomic trajectory, we get the equivalence of (i) and (ii) under the correspondence of μ and $J \circ \lambda$.

(iii) For $\xi \in \mathfrak{g}$, we compute

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle J\circ\gamma'(t);\xi\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{G}(\gamma'(t),\xi_{\mathsf{M}}\circ\gamma(t)) \\ &= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t),\xi_{\mathsf{M}}\circ\gamma(t)) + \mathbb{G}(\gamma'(t),\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\xi_{\mathsf{M}}\circ\gamma(t)) \\ &= \mathbb{G}(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t),\xi_{\mathsf{M}}\circ\gamma(t)),\end{aligned}$$

using the fact that $\overset{\mathbb{G}}{\nabla}\xi_{\mathsf{M}}$ is skew-symmetric [Petersen 2006, Proposition 27]. Thus

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle J\circ\gamma'(t);\xi\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\langle J\circ\lambda(t);\xi\rangle, \qquad t\in I,$$

and so $\langle J\circ(\gamma'-\lambda);\xi\rangle$ is constant, which gives this part of the corollary. ■

The third condition of the corollary should be regarded as, up to a constant, determining the adjoint field from the trajectory.

Note that the equivalence of parts (i) and (ii) of the result is simply a literal translation of Corollary 7.16 to Chaplygin systems. Thus it falls into the category of "the obvious condition," but with the extra structure taken into account. Jóźwikowski and Respondek give a generalisation of this result that retains the structure of M as a principal G-bundle, but the G-invariance of $\mathbb{G}$, V, and D are relaxed. A development of this result would take us a little far afield, and will again be an adaptation of Corollary 7.16 to the available structure. In the case when G is Abelian, Fernandez and Bloch [2008, Proposition 3] give some conditions on the curvature Ω that must be satisfied in order that all nonholonomic trajectories be constrained variational trajectories. Again, we will not develop these results here.

The next result we give involves the generalised subbundle

$$\mathsf{D}^{(1)}_x = \{X(x)+[Y,Z](x) \mid X,Y,Z\in\Gamma^r(\mathsf{D})\}$$

of TM. Note that

$$\mathsf{D}\subseteq\mathsf{D}^{(1)} \implies (\mathsf{D}^{(1)})^{\perp}\subseteq\mathsf{D}^{\perp}.$$

The result is the following.

7.20 Corollary: **([Terra 2018, Theorem 1])** *Let $r\in\{\infty,\omega\}$ and let $\Sigma=(\mathsf{M},\mathbb{G},V,\mathsf{D})$ be a C^r-constrained simple mechanical system. Assume that either (1) $r=\omega$ or that (2) $r=\infty$ and $\ker(\hat{F}^*_{\mathsf{D}})$ is a subbundle. Suppose that the following conditions hold:*

(i) $b_{\mathsf{D}}\in\Gamma^r(\pi^*_{\mathsf{D}}(\mathsf{D}^{(1)})^{\perp})$;

(ii) $A_{\mathsf{D}}(\pi^*_{\mathsf{D}}(\mathsf{D}^{(1)})^{\perp})\subseteq\pi^*_{\mathsf{D}}(\mathsf{D}^{(1)})^{\perp}$;

(iii) $\nabla_{X_0}(\mathscr{G}^r_{\pi^*_{\mathsf{D}}\mathsf{D}^{(1)}})\subseteq\mathscr{G}^r_{\pi^*_{\mathsf{D}}(\mathsf{D}^{(1)})}$.

Then every nonholonomic trajectory is a D-regular constrained variational trajectory.

Proof: We first note that, by Proposition 4.13, the hypotheses are that $\pi^*_{\mathsf{D}}(\mathsf{D}^{(1)})^{\perp}$ is flow-invariant under $X^{\mathrm{reg}}_{\mathsf{D}}$. Next we claim that $\operatorname{image}(\hat{F}_{\mathsf{D}})\subseteq\pi^*_{\mathsf{D}}\mathsf{D}^{(1)}$. Indeed, let $(\alpha_x,u_x)\in\operatorname{image}(\hat{F}_{\mathsf{D}})$. Thus

$$(\alpha_x,u_x)=(F_{\mathsf{D}}(u_x)(w_x),u_x)$$

for some $(w_x, u_x) \in \pi_{\mathsf{D}}^*\mathsf{D}$, and the claim follows by definition of F_{D}. Thus

$$\ker(\hat{F}_{\mathsf{D}}^*)^{\perp} \subseteq \pi_{\mathsf{D}}^*\mathsf{D}^{(1)} \quad \Longrightarrow \quad \pi_{\mathsf{D}}^*(\mathsf{D}^{(1)})^{\perp} \subseteq \ker(\hat{F}_{\mathsf{D}}^*).$$

Thus, if $\pi_{\mathsf{D}}^*(\mathsf{D}^{(1)})^{\perp}$ is flow-invariant under $X_{\mathsf{D}}^{\mathrm{reg}}$, then so too is $\ker(\hat{F}_{\mathsf{D}}^*)$. Thus $\ker(\hat{F}_{\mathsf{D}}^*)$ contains a C^r-cogeneralised affine subbundle that is invariant under $X_{\mathsf{D}}^{\mathrm{reg}}$. Since a C^r-cogeneralised affine subbundle is a special example of a C^r-affine subbundle variety whose base variety is M, the result then follows by Theorems 4.24 and 7.12. ■

Our hypotheses are not the same as those of Terra, but are equivalent to them by Proposition 4.13. Note that we are able to relax the assumption of Terra that $\mathsf{D}^{(1)}$ be a subbundle.

Our final result is simply Corollary 7.14, stripped of the extraneous requirement that D be orthogonal to a foliation for which $\mathbb{G}$ is bundle-like.

7.21 Corollary: **([Terra 2018, Corollary 1])** *Let $r \in \{\infty, \omega\}$ and let $\Sigma = (\mathsf{M}, \mathbb{G}, 0, \mathsf{D})$ be a C^r-constrained simple mechanical system. If D is geodesically invariant, then every nonholonomic trajectory is a D-regular constrained variational trajectory.*

We note that all of the results quoted above are either reformulations of the "obvious" condition of Corollary 7.16 or they give conditions under which the purely algebraic conditions of Theorems 7.10 or 7.12 apply, without needing to resort to the differential conditions. It would be interesting to have physical examples—or even mathematical examples—of constrained simple mechanical systems for which every nonholonomic trajectory is a constrained variational trajectory, but for which a verification of this requires one to use the differential conditions of Theorem 7.12.

References

Abraham, R., Marsden, J. E., and Ratiu, T. S. [1988] *Manifolds, Tensor Analysis, and Applications*, number 75 in Applied Mathematical Sciences, Springer-Verlag: New York/-Heidelberg/Berlin, ISBN: 978-0-387-96790-5.

Adams, R. A. and Fournier, J. J. F. [2003] *Sobolev Spaces*, 2nd edition, number 140 in Pure and Applied Mathematics, Dekker Marcel Dekker: New York, NY, ISBN: 978-0-12-044143-3.

Agrachev, A. A., Barilari, D., and Boscain, U. [2018] *A Comprehensive Introduction to Sub-Riemannian geometry*, number 181 in Cambridge Studies in Advanced Mathematics, Cambridge University Press: New York/Port Chester/Melbourne/Sydney, ISBN: 978-1-108-67732-5.

Agrachev, A. A. and Gamkrelidze, R. V. [1978] *The exponential representation of flows and the chronological calculus*, Mathematics of the USSR-Sbornik, **107**(4), pages 467–532, ISSN: 0025-5734, DOI: 10.1070/SM1979v035n06ABEH001623.

Aliprantis, C. D. and Border, K. C. [2006] *Infinite-Dimensional Analysis, A Hitchhiker's Guide*, 3rd edition, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-32696-0.

Berger, M. [1987] *Geometry I*, Universitext, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-11658-5.

Borisov, A. V., Mamaev, I. S., and Bizyaev, I. A. [2017] *Dynamical systems with non-integrable constraints, vakonomic mechanics, sub-Riemannian geometry, and non-holonomic mechanics*, Russian Mathematical Surveys, Translation of *Rossiĭskaya Akademiya Nauk. Moskovskoe Matematicheskoe Obshchestvo. Uspekhi Matematicheskikh Nauk*, **72**(5), pages 783–840, ISSN: 0036-0279, DOI: 10.1070/RM9783.

Bullo, F. and Lewis, A. D. [2004] *Geometric Control of Mechanical Systems, Modeling, Analysis, and Design for Simple Mechanical Systems*, number 49 in Texts in Applied Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-22195-3.

Cantrijn, F., Cortés, J., de León, M., and Martín de Diego, D. [2002] *On the geometry of generalised Chaplygin systems*, Mathematical Proceedings of the Cambridge Philosophical Society, **132**(2), pages 323–351, ISSN: 0305-0041, DOI: 10.1017/S0305004101005679.

Cardin, F. and Favretti, M. [1996] *On nonholonomic and vakonomic dynamics of mechanical systems with nonintegrable constraints*, Journal of Geometry and Physics, **18**(4), pages 295–325, ISSN: 0393-0440, DOI: 10.1016/0393-0440(95)00016-X.

Cartan, H. [1957] *Variétés analytiques réelles et variétés analytiques complexes*, Bulletin de la Société Mathématique de France, **85**, pages 77–99, ISSN: 0037-9484, URL: http://www.numdam.org/item?id=BSMF_1957__85__77_0 (visited on 07/10/2014).

Chow, W.-L. [1940/1941] *Über Systemen von linearen partiellen Differentialgleichungen erster Ordnung*, Mathematische Annalen, **117**(1), pages 98–105, ISSN: 0025-5831, DOI: 10.1007/BF01450011.

Convent, A. and Van Schaftingen, J. [2016a] *Geometric partial differentiability on manifolds: The tangential derivative and the chain rule*, Journal of Mathematical Analysis and Applications, **435**(2), pages 1672–1681, ISSN: 0022-247X, DOI: 10.1016/j.jmaa.2015.11.036.

— [2016b] *Intrinsic colocal weak derivatives and Sobolev spaces between manifolds*, Annali della Scuola Normale Superiore di Pisa. Serie III, **16**(1), pages 97–128, DOI: 10.2422/2036-2145.201312_005.

— [2019] *Higher order weak differentiability and Sobolev spaces between manifolds*, Advances in Calculus of Variations, **12**(3), pages 303–332, ISSN: 1864-8258, DOI: 10.1515/acv-2017-0008.

Cortés, J., de León, M., Martín de Diego, D., and Martínez, S. [2002] *Geometric description of vakonomic and nonholonomic dynamics. Comparison of solutions*, SIAM Journal on Control and Optimization, **41**(5), pages 1389–1412, ISSN: 0363-0129, DOI: 10.1137/S036301290036817X.

Crampin, M. and Mestdag, T. [2010] *Anholonomic frames in constrained dynamics*, Dynamical Systems, An International Journal, **25**(2), pages 159–187, ISSN: 1468-9367, DOI: 10.1080/14689360903360888.

Favretti, M. [1998] *Equivalence of dynamics for nonholonomic systems with transverse constraints*, Journal of Dynamics and Differential Equations, **10**(4), pages 511–536, ISSN: 1040-7294, DOI: 10.1023/A:1022667307485.

Federer, H. [1969] *Geometric Measure Theory*, Grundlehren der Mathematischen Wissenschaften, Springer-Verlag: New York/Heidelberg/Berlin, Reprint: [Federer 1996].

— [1996] *Geometric Measure Theory*, Classics in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-60656-7, Original: [Federer 1969].

Fefferman, C. L. and Kollár, J. [2013] *Continuous solutions of linear equations*, in *From Fourier Analysis and Number Theory to Radon Transforms and Geometry*, edited by J. Farkas, R. C. Gunning, M. I. Knopp, and B. A. Taylor, 28 Developments in Mathematics, pages 233–282, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-1-4614-4074-1.

Fernandez, O. E. and Bloch, A. M. [2008] *Equivalence of the dynamics of nonholonomic and variational nonholonomic systems for certain initial data*, Journal of Physics. A. Mathematical and Theoretical, **41**(34), page 344005, ISSN: 1751-8113, DOI: 10.1088/1751-8113/41/34/344005.

Goldschmidt, H. L. [1967a] *Existence theorems for analytic linear partial differential equations*, Annals of Mathematics. Second Series, **86**(2), pages 246–270, ISSN: 0003-486X, DOI: 10.2307/1970689.

— [1967b] *Integrability criteria for systems of nonlinear partial differential equations*, Journal of Differential Geometry, **1**(3-4), pages 269–307, ISSN: 0022-040X, URL: http://projecteuclid.org/euclid.jdg/1214428094 (visited on 07/12/2014).

Gràcia, X., Marin-Solano, J., and Muñoz-Lecanda, M. C. [2003] *Some geometric aspects of variational calculus in constrained systems*, Reports on Mathematical Physics, **51**(1), pages 127–148, ISSN: 0034-4877, DOI: 10.1016/S0034-4877(03)80006-X.

Grauert, H. [1958] *On Levi's problem and the imbedding of real-analytic manifolds*, Annals of Mathematics. Second Series, **68**(2), pages 460–472, ISSN: 0003-486X, DOI: 10.2307/1970257.

Grauert, H. and Remmert, R. [1984] *Coherent Analytic Sheaves*, number 265 in Grundlehren der Mathematischen Wissenschaften, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-13178-8.

Jafarpour, S. and Lewis, A. D. [2014] *Time-Varying Vector Fields and Their Flows*, Springer Briefs in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-319-10138-5.

Jost, J. [2005] *Postmodern Analysis*, 3rd edition, Universitext, Springer-Verlag: New York/-Heidelberg/Berlin, ISBN: 978-3-540-25830-8.

Jóźwikowski, M. and Respondek, W. [2019] *A comparison of vakonomic and nonholonomic dynamics with applications to non-invariant Chaplygin systems*, Journal of Geometric Mechanics, **11**(1), pages 77–122, ISSN: 1941-4889, DOI: 10.3934/jgm.2019005.

Kharlomov, P. V. [1992] *A critique of some mathematical models of mechanical systems with differential constraints*, Journal of Applied Mathematics and Mechanics, Translation of *Rossiĭskaya Akademiya Nauk. Prikladnaya Matematika i Mekhanika*, **56**(4), pages 584–594, ISSN: 0021-8928, DOI: 10.1016/0021-8928(92)90016-2.

Klingenberg, W. [1995] *Riemannian Geometry*, 2nd edition, number 1 in Studies in Mathematics, Walter de Gruyter: Berlin/New York, ISBN: 978-3-11-014593-9.

Kobayashi, S. and Nomizu, K. [1963] *Foundations of Differential Geometry*, volume 1, number 15 in Interscience Tracts in Pure and Applied Mathematics, Interscience Publishers: New York, NY, Reprint: [Kobayashi and Nomizu 1996].

— [1996] *Foundations of Differential Geometry*, volume 1, Wiley Classics Library, John Wiley and Sons: NewYork, NY, ISBN: 978-0-471-15733-5, Original: [Kobayashi and Nomizu 1963].

Koiller, J. [1992] *Reduction of some classical nonholonomic systems with symmetry*, Archive for Rational Mechanics and Analysis, **118**(2), pages 113–148, ISSN: 0003-9527, DOI: 10.1007/BF00375092.

Kolář, I., Michor, P. W., and Slovák, J. [1993] *Natural Operations in Differential Geometry*, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-56235-1.

Kozlov, V. V. [1992] *The problem of realizing constraints in dynamics*, Journal of Applied Mathematics and Mechanics, Translation of *Rossiĭskaya Akademiya Nauk. Prikladnaya Matematika i Mekhanika*, **56**(4), pages 594–600, ISSN: 0021-8928, DOI: 10.1016/0021-8928(92)90017-3.

Kupka, I. and Oliva, W. M. [2001] *The nonholonomic mechanics*, Journal of Differential Equations, **169**(1), pages 169–189, ISSN: 0022-0396, DOI: 10.1006/jdeq.2000.3897.

Langerok, B. [2003] *A connection theoretic approach to sub-Riemannian geometry*, Journal of Geometry and Physics, **46**(3-4), pages 203–230, ISSN: 0393-0440, DOI: 10.1016/S0393-0440(02)00026-8.

Lee, J. M. [2018] *Riemannian Manifolds, An Introduction to Curvature*, number 176 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-319-91754-2.

Lewis, A. D. [1998] *Affine connections and distributions with applications to nonholonomic mechanics*, Reports on Mathematical Physics, **42**(1/2), pages 135–164, ISSN: 0034-4877, DOI: 10.1016/S0034-4877(98)80008-6.

— [2012] *Generalised Subbundles, Distributions, and Families of Vector Fields, A Comprehensive Review*, Lecture notes, Madrid, Spain: ICMAT, URL: http://www.mast.queensu.ca/~andrew/notes/abstracts/2011a.html (visited on 05/12/2017).

— [2017] *The physical foundations of geometric mechanics*, Journal of Geometric Mechanics, **9**(4), pages 487–574, ISSN: 1941-4889, DOI: 10.3934/jgm.2017019.

Lewis, A. D. and Murray, R. M. [1995] *Variational principles for constrained systems: Theory and experiment*, International Journal of Non-Linear Mechanics, **30**(6), pages 793–815, ISSN: 0020-7462, DOI: 10.1016/0020-7462(95)00024-0.

Liu, W. S. and Sussmann, H. J. [1994] *Abnormal sub-Riemannian minimizers*, in *Differential Equations, Dynamical Systems, and Control Science*, Festschrift in honor of Lawrence Markus, (1994), edited by K. D. Elworthy, W. N. Everitt, and E. B. Lee, 152 Lecture Notes in Pure and Applied Mathematics, pages 705–716, Dekker Marcel Dekker: New York, NY, ISBN: 978-0-8247-8904-6.

Michor, P. W. [1980] *Manifolds of Differentiable Mappings*, number 3 in Shiva Mathematics Series, Shiva Publishing Limited: Orpington, UK, ISBN: 978-0-906812-03-7.

Minguzzi, E. [2015] *The equality of mixed partial derivatives under weak differentiability conditions*, Real Analysis Exchange, **40**(1), pages 81–98, ISSN: 0147-1937, DOI: 10.14321/realanalexch.40.1.0081.

Nestruev, J. [2003] *Smooth Manifolds and Observables*, number 220 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-95543-8.

Paternain, G. P. [1999] *Geodesic Flows*, number 180 in Progress in Mathematics, Birkhäuser: Boston/Basel/Stuttgart, ISBN: 978-0-8176-4144-3.

Petersen, P. [2006] *Riemannian Geometry*, number 171 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0387-29246-5.

Pontryagin, L. S., Boltyanskiĭ, V. G., Gamkrelidze, R. V., and Mishchenko, E. F. [1961] *Matematicheskaya teoriya optimal′ nykh protsessov*, Gosudarstvennoe izdatelstvo fizikomatematicheskoi literatury: Moscow, Translation: [Pontryagin, Boltyanskiĭ, Gamkrelidze, and Mishchenko 1986].

— [1986] *The Mathematical Theory of Optimal Processes*, translated by K. N. Trirogoff, Classics of Soviet Mathematics, Gordon & Breach Science Publishers: New York, NY, ISBN: 978-2-88124-077-5, Original: [Pontryagin, Boltyanskiĭ, Gamkrelidze, and Mishchenko 1961].

Ramanan, S. [2005] *Global Calculus*, number 65 in Graduate Studies in Mathematics, American Mathematical Society: Providence, RI, ISBN: 978-0-8218-3702-3.

Rumiantsev, V. V. [1978] *On Hamilton's principle for nonholonomic systems*, Journal of Applied Mathematics and Mechanics, Translation of *Rossiĭskaya Akademiya Nauk. Prikladnaya Matematika i Mekhanika*, **42**(3), pages 387–399, ISSN: 0021-8928, DOI: 10.1016/0021-8928(78)90108-9.

Sasaki, S. [1958] *On the differential geometry of tangent bundles of Riemannian manifolds*, The Tôhoku Mathematical Journal. Second Series, **10**(3), pages 338–354, ISSN: 0040-8735, DOI: 10.2748/tmj/1178244668.

Sontag, E. D. [1998] *Mathematical Control Theory, Deterministic Finite Dimensional Systems*, 2nd edition, number 6 in Texts in Applied Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-98489-6.

Sussmann, H. J. [2008] *Smooth distributions are globally finitely spanned*, in *Analysis and Design of Nonlinear Control Systems*, edited by A. Astolfi and L. Marconi, In honor of Alberto Isidori, pages 3–8, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-74357-6.

Synge, J. L. [1928] *Geodesics in nonholonomic geometry*, Mathematische Annalen, **99**(1), pages 738–751, ISSN: 0025-5831, DOI: 10.1007/BF01459122.

Terra, G. [2018] *Vakonomic versus nonholonomic mechanics revisited*, São Paulo Journal of Mathematical Science, University of São Paulo. São Paulo, **12**(1), pages 136–145, ISSN: 1982-6907, DOI: 10.1007/s40863-017-0062-z.

Terra, G. and Kobayashi, M. H. [2004a] *On the classical mechanical systems with non-linear constraints*, Journal of Geometry and Physics, **49**(3-4), pages 385–417, issn: 0393-0440, doi: 10.1016/j.geomphys.2003.08.005.

— [2004b] *On the variational mechanics with non-linear constraints*, Journal de Mathématiques Pures et Appliquées. Neuvième Série, **83**(5), pages 629–671, issn: 0021-7824, doi: 10.1016/S0021-7824(03)00069-2.

Vershik, A. M. and Gershkovich, V. Y. [1990] *Nonholonomic dynamical systems, geometry of distributions and variational problems*, in *Dynamical Systems*, edited by V. I. Arnol′d and S. Novikov, volume 7, Encyclopedia of Mathematical Sciences, pages 4–79, Springer-Verlag: New York/Heidelberg/Berlin, isbn: 978-3-540-18176-7.

Wells Jr., R. O. [2008] *Differential Analysis on Complex Manifolds*, 3rd edition, number 65 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, isbn: 978-0-387-90419-0.

Whitney, H. [1936] *Differentiable manifolds*, Annals of Mathematics. Second Series, **37**(3), pages 645–680, issn: 0003-486X, doi: 10.2307/1968482.

Zampieri, G. [2000] *Nonholonomic versus vakonomic mechanics*, Journal of Differential Equations, **163**(2), pages 335–347, issn: 0022-0396, doi: 10.1006/jdeq.1999.3727.

Ziemer, W. P. [1989] *Weakly Differentiable Functions, Sobolev Spaces and Functions of Bounded Variation*, number 120 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, isbn: 978-0-387-97017-2.