Francesco Bullo and Andrew D. Lewis

Geometric Control of Mechanical Systems

Modeling, Analysis, and Design for Simple Mechanical Control Systems

– Supplementary Material –

August 1, 2014

Contents

S1

Tangent and cotangent bundle geometry

In Supplements 3 and 4 we will make use of certain tangent bundle structures arising from the presence of an affine connection. To keep the presentation interesting, certain ideas are presented in a general fiber bundle framework, although eventually we shall only utilize the tangent or cotangent bundle cases. The main references for the material in this supplement are [Yano and Ishihara 1973] and parts of [Kolář, Michor, and Slovák 1993]. All data in this chapter is of class C^∞, so we will not always explicitly state this hypothesis.

S1.1 Some things Hamiltonian

In the text, we avoided Hamiltonian presentations of mechanics, and of control theory for mechanical systems. Our main reason for doing this was to emphasize the Lagrangian point of view, while referring the reader to the literature for Hamiltonian treatments. However, in geometric optimal control, quite independent of anything "mechanical," the Hamiltonian point of view acquires great value, so we shall need to delve somewhat into the geometry associated with Hamiltonian mechanics. Our presentation will be far too brief to serve as a useful introduction, so we refer the reader to texts such as [Abraham and Marsden 1978, Arnol'd 1978, Bloch 2003, Guillemin and Sternberg 1990, Libermann and Marle 1987]. The books [Agrachev and Sachkov 2004, Bloch 2003, Jurdjevic 1997] have a description of the Hamiltonian aspects of optimal control.

S1.1.1 Differential forms

In the text we were able to omit treatment of an important class of tensors known as differential forms, since they did not come up in our development, except in the most elementary manner. However, now we shall need differential forms, so we give the briefest of introductions, referring to [Abraham, Marsden, and Ratiu 1988, Flanders 1989, Nelson 1967] for further details.

First we look at the linear case. A $(0, k)$-tensor α on a vector space V is ***skew-symmetric*** if, for every $\sigma \in S_k$ and for all $v_1, \dots, v_k \in \mathsf{V}$, we have

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

We now make the following definition.

Definition S1.1 (Exterior form). Let V be a finite-dimensional $\mathbb{R}$-vector space. An ***exterior k-form*** on V is a skew-symmetric $(0, k)$-tensor on V. We denote the set of exterior k-forms on V[1] by $\bigwedge_k(\mathsf{V})$. •

If $\alpha \in \bigwedge_k(\mathsf{V})$ and $\beta \in \bigwedge_l(\mathsf{V})$, then $\alpha \otimes \beta \in T^0_{k+l}(\mathsf{V})$, but it will not generally be the case that $\alpha \otimes \beta$ is an exterior $(k + l)$-form. However, one can define a product that preserves skew-symmetry of tensors. To do so, for each $k \in \mathbb{N}$, define a linear map $\operatorname{Alt}\colon T^0_k(\mathsf{V}) \to \bigwedge_k(\mathsf{V})$ by

$$\operatorname{Alt}(t)(v_1, \dots, v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) t(v_{\sigma(1)}, \dots, v_{\sigma(k)}).$$

Thus Alt "skew-symmetrizes" a tensor that is not skew-symmetric. One can verify that, if $\alpha \in \bigwedge_k(\mathsf{V})$, then $\operatorname{Alt}(\alpha) = \alpha$. Now, given $\alpha \in \bigwedge_k(\mathsf{V})$ and $\beta \in \bigwedge_l(\mathsf{V})$, we define $\alpha \wedge \beta \in \bigwedge_{k+l}(\mathsf{V})$ by

$$\alpha \wedge \beta = \frac{(k + l)!}{k!l!} \operatorname{Alt}(\alpha \otimes \beta).$$

This is the ***wedge product*** of α and β. We comment that the appearance of the factorial coefficient is not standard in the literature, so care should be exercised when working with different conventions. The wedge product is neither symmetric nor skew-symmetric, but obeys the relation $\alpha \wedge \beta = (-1)^{kl}\beta \wedge \alpha$.

Now let us see how to represent exterior k-forms in a basis $\{e_1, \dots, e_n\}$ for V. Since $\alpha \in \bigwedge_k(\mathsf{V})$ is a $(0, k)$-tensor, one can define its ***components*** in the usual manner:

$$\alpha_{i_1 \cdots i_k} = \alpha(e_{i_1}, \dots, e_{i_k}), \qquad i_1, \dots, i_k \in \{1, \dots, n\}.$$

Since α is skew-symmetric, the components obey

$$\alpha_{\sigma(i_1) \cdots \sigma(i_k)} = \operatorname{sgn}(\sigma)\alpha_{i_1 \cdots i_k},$$

for $\sigma \in S_k$. Now let us provide a basis for $\bigwedge_k(\mathsf{V})$.

Proposition S1.2 (Basis for $\bigwedge_k(\mathsf{V})$). *Let $\{e_1, \dots, e_n\}$ be a basis for the $\mathbb{R}$-vector space V. Then*

$$\left\{ e^{i_1} \wedge \cdots \wedge e^{i_k} \;\middle|\; i_1, \dots, i_k \in \{1, \dots, n\},\ i_1 < i_2 < \cdots < i_k \right\}$$

[1] ***Interior k-forms*** are defined similarly, but are $(k, 0)$-tensors.

is a basis for $\bigwedge_k(\mathsf{V})$*. In particular,*

$$\dim(\bigwedge_k(\mathsf{V})) = \begin{cases} \frac{n!}{(n-k)!k!}, & k \leq n, \\ 0, & k > n. \end{cases}$$

With the basis and the components defined as above, one can verify that

$$\alpha = \sum_{i_1 < \cdots < i_k} \alpha_{i_1 \cdots i_k} e^{i_k} \wedge \cdots \wedge e^{i_k}.$$

Let us illustrate this with some special cases.

Examples S1.3. 1. For $k = 0$, one takes the convention that $\bigwedge_0(\mathsf{V}) = \mathbb{R}$. Thus $\dim(\bigwedge_0(\mathsf{V})) = 1$.

2. Next we consider the case $k = 1$. One can easily see that $\bigwedge_1(\mathsf{V}) = \mathsf{V}^*$. Thus $\dim(\bigwedge_1(\mathsf{V})) = \dim(\mathsf{V})$.
3. Next consider the case when $k = n = \dim(\mathsf{V})$. In this case one can see that, if $\{e_1, \dots, e_n\}$ is a basis for V, then every exterior k-form is a multiple of $e^1 \wedge \cdots \wedge e^n$. Thus $\dim(\bigwedge_n(\mathsf{V})) = 1$.
4. Now take $k = 2$, and for concreteness, $n = 3$. Then a basis for $\bigwedge_2(\mathsf{V})$ is

   $$\{e^1 \wedge e^2, e^1 \wedge e^3, e^2 \wedge e^3\},$$

   and, given $\alpha \in \bigwedge_2(\mathsf{V})$, we have

   $$\alpha = \alpha_{12} e^1 \wedge e^2 + \alpha_{13} e^1 \wedge e^3 + \alpha_{23} e^2 \wedge e^3,$$

   where $\alpha_{12} = \alpha(e_1, e_2)$, $\alpha_{13} = \alpha(e_1, e_3)$, and $\alpha_{23} = \alpha(e_2, e_3)$. Therefore, $\dim(\bigwedge_1(\mathsf{V})) = \dim(\bigwedge_2(\mathsf{V})) = \dim(\mathsf{V})$ in the case that $\dim(\mathsf{V}) = 3$. •

Now let us turn to extending the algebraic setting to a differential geometric one on a manifold M. Of course, the basic step of defining

$$\bigwedge_k(\mathsf{TM}) = \overset{\circ}{\bigcup_{x \in \mathsf{M}}} \bigwedge_k(\mathsf{T}_x\mathsf{M})$$

is done in the usual way, with the push-forward of the overlap maps providing a vector bundle structure for $\bigwedge_k(\mathsf{TM})$.

Definition S1.4 (Differential form). Let M be a C^r-manifold for $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$. A C^r-section of the vector bundle $\bigwedge_k(\mathsf{TM})$ is a ***differential k-form***. •

Note that a differential one-form is what we, in the text, call a covector field. Since differential k-forms are tensor fields, all constructions concerning tensor fields—push-forward and pull-back, Lie differentiation, covariant differentiation with respect to an affine connection—are applicable to them.

However, differential forms come with an operation that is unique to them, and that is enormously useful in many areas of mathematics and mathematical physics, namely the exterior derivative. It is not easy to motivate the definition of this operator, so we sidestep this by simply not giving any motivation. We merely give the definition. The reader can find motivation in books such as [Abraham, Marsden, and Ratiu 1988, Bryant, Chern, Gardner, Goldschmidt, and Griffiths 1991, Flanders 1989, Nelson 1967]. To whet the appetite of the reader yet to be exposed to the charms of differential forms, we mention that the exterior derivative can be used to elegantly unify and generalize Stokes' Theorem, Gauss's Theorem, and the Divergence Theorem from vector calculus.

Definition S1.5 (Exterior derivative). For $\alpha \in \Gamma^\infty(\bigwedge_k(\mathsf{TM}))$, the ***exterior derivative*** of α is the element $\mathrm{d}\alpha \in \Gamma^\infty(\bigwedge_{k+1}(\mathsf{TM}))$ defined by

$$\begin{aligned}\mathrm{d}\alpha(X_0, X_1, \dots, X_k) = &\sum_{j=0}^{k}(-1)^j \mathscr{L}_{X_j}\alpha(X_0, \dots, \widehat{X}_j, \dots, X_k) \\ &+ \sum_{0\leq i<j\leq k}(-1)^{i+j}\alpha([X_i, X_j], X_0, \dots, \widehat{X}_i, \dots, \widehat{X}_j, \dots, X_k),\end{aligned}$$

where the $\widehat{\cdot}$ means that the argument is deleted. •

This definition may be shown to make sense, in that it defines a skew-symmetric tensor field. The definition belies the fact that the exterior derivative is simple (and natural) to compute in coordinates:

$$\mathrm{d}\alpha = \frac{\partial \alpha_{i_1 \cdots i_k}}{\partial x^j}\mathrm{d}x^j \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}.$$

The exterior derivative has the following useful properties, plus many others that we do not list here (see [Abraham, Marsden, and Ratiu 1988]):

1. "$\mathrm{d}f = \mathrm{d}f$": that is, the exterior derivative of a differential zero-form (i.e., a function) is the same as the differential defined in the text;
2. $\mathrm{dd}\alpha = 0$ for all $\alpha \in \Gamma^\infty(\bigwedge_k(\mathsf{TM}))$, $k \in \mathbb{Z}_+$;
3. $\mathrm{d}(\alpha\wedge\beta) = \mathrm{d}\alpha\wedge\beta + (-1)^k\alpha\wedge\mathrm{d}\beta$ if $\alpha \in \Gamma^\infty(\bigwedge_k(\mathsf{TM}))$ and $\beta \in \Gamma^\infty(\bigwedge_l(\mathsf{TM}))$, for $k, l \in \mathbb{Z}_+$;
4. $\mathrm{d}(f^*\alpha) = f^*(\mathrm{d}\alpha)$ for $f \in C^\infty(\mathsf{M};\mathsf{N})$ and $\alpha \in \Gamma^\infty(\bigwedge_k(\mathsf{TN}))$, $k \in \mathbb{Z}_+$;
5. $\mathscr{L}_X(\mathrm{d}\alpha) = \mathrm{d}(\mathscr{L}_X\alpha)$ for $X \in \Gamma^\infty(\mathsf{TM})$ and $\alpha \in \Gamma^\infty(\bigwedge_k(\mathsf{TM}))$, $k \in \mathbb{Z}_+$.

Readers having never seen the notion of differential forms or exterior derivative should now expect to do some extra reading if the rest of the section is to have any deep significance.

S1.1.2 Symplectic manifolds

In our treatment of Lagrangian mechanics, the primary role is played by Riemannian metrics and affine connections. In the typical Hamiltonian treatment of mechanics, the structure is provided by a symplectic structure (or, more generally, a Poisson structure). A symplectic structure is defined by a differential two-form with special properties. Let $\alpha \in \Gamma^\infty(\bigwedge_2(\mathsf{TM}))$. Since α is a $(0,2)$-tensor field, it has associated with it a vector bundle map $\alpha^\flat\colon \mathsf{TM} \to \mathsf{T}^*\mathsf{M}$ (see Section 2.3.4). The differential two-form α is ***nondegenerate*** if $\alpha^\flat$ is a vector bundle isomorphism. In such a case, we denote the inverse vector bundle isomorphism by $\alpha^\sharp$, of course. A differential k-form α is ***closed*** if $\mathrm{d}\alpha = 0$.

We may now state our main definition.

Definition S1.6 (Symplectic manifold). For $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$, a ***$C^r$-symplectic manifold*** is a pair (M, ω), where ω is a C^r-closed nondegenerate differential two-form on the C^r-manifold M. The differential two-form ω is called a ***symplectic form*** on M. •

One can verify (see [Abraham and Marsden 1978]) that nondegeneracy of ω for a symplectic manifold (M, ω) implies that the connected components of M are even-dimensional. The condition of closedness of symplectic forms is one that is not trivial to motivate, so we do not do so. We refer the reader to [Abraham and Marsden 1978, Arnol'd 1978, Guillemin and Sternberg 1990, Libermann and Marle 1987] for additional discussion.

It turns out that there is a natural symplectic form on the cotangent bundle of a manifold M, and it is this symplectic form that we shall exclusively deal within our treatment of optimal control. We first define a differential one-form on M. One may readily verify that there exists a unique differential one-form $\theta_0 \in \Gamma^\infty(\bigwedge_1(\mathsf{T}(\mathsf{T}^*\mathsf{M})))$ that satisfies $\beta^*\theta_0 = \beta$ for any differential one-form β on M (here one thinks of β as a map from M to $\mathsf{T}^*\mathsf{M}$). In verifying this, one also verifies that, in natural coordinates $((x^1, \dots, x^n), (p_1, \dots, p_n))$ for $\mathsf{T}^*\mathsf{M}$, we have $\theta_0 = p_i \mathrm{d}q^i$. The ***canonical symplectic form*** on $\mathsf{T}^*\mathsf{M}$ is then $\omega_0 = -\mathrm{d}\theta_0$. In natural coordinates one has $\omega_0 = \mathrm{d}q^i \wedge \mathrm{d}p_i$. That ω_0 is closed follows since $\mathrm{d}\omega_0 = -\mathrm{d}\mathrm{d}\theta_0 = 0$. The nondegeneracy of ω_0 follows from the fact that the matrix representative of ω_0 in natural coordinates is

$$[\omega_0] = \begin{bmatrix} \mathbf{0} & \boldsymbol{I}_n \\ -\boldsymbol{I}_n & \mathbf{0} \end{bmatrix}.$$

This matrix is invertible, and so too, then, is $\omega_0^\flat$.

Between symplectic manifolds, maps that preserve the symplectic form are important.

Definition S1.7 (Symplectic diffeomorphism). Let (M, ω) and (N, Ω) be C^r-symplectic manifolds, $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$. A C^r-diffeomorphism $\phi\colon \mathsf{M} \to \mathsf{N}$ is ***symplectic*** if $\phi^*\Omega = \omega$. •

S1.1.3 Hamiltonian vector fields

Now we give a little insight into what sorts of constructions one can make on a symplectic manifold. There is much that can be said here, some of it being quite deep. We shall stick to primarily simple matters, as these are all that we shall make immediate use of.

Let (M, ω) be a symplectic manifold. We let H be a function on $\mathbb{R} \times \mathsf{M}$ that is locally integrally of class C^{r+1} (see Section A.2.1). So that we are clear, if $H_t(x) = H(t,x)$, then $\mathrm{d}H$ is the locally integrally C^r-section of $\mathsf{T}^*\mathsf{M}$ defined by $\mathrm{d}H(t,x) = \mathrm{d}H_t(x)$.

Definition S1.8 (Hamiltonian vector field). Let (M, ω) be a symplectic manifold and let $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$. The ***Hamiltonian vector field*** associated to a locally integrally class C^{r+1}-function H on M is the locally integrally class C^r-vector field X_H on M defined by $X_H(t,x) = -\omega^\sharp(\mathrm{d}H(t,x))$. The function H is the ***Hamiltonian*** for the Hamiltonian vector field X_H. •

Our presentation is more general than the standard setup in Hamiltonian mechanics where one considers only time-independent functions; this is necessitated by our use of the Hamiltonian formulation of the Maximum Principle. If H is time-independent, then we note that

$$\mathscr{L}_{X_H} H = \langle X_H; \mathrm{d}H \rangle = -\omega(X_H, X_H) = 0.$$

Thus, for time-independent Hamiltonians, the Hamiltonian is a constant of motion for the corresponding Hamiltonian vector field. In many physical problems, this corresponds to conservation of energy. We shall see that in our optimal control setup, constancy of the Hamiltonian along trajectories of the control system also plays a role, although there is no physical energy associated with the problem.

Let us write the coordinate expression for a Hamiltonian vector field defined on a cotangent bundle using the canonical symplectic structure and using natural coordinates $((x^1, \ldots, x^n), (p_1, \ldots, p_n))$. In this case, a straightforward computation shows that

$$X_H = \frac{\partial H}{\partial p_i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p_i}, \tag{S1.1}$$

or, in terms of differential equations in the local coordinates,

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}, \qquad i \in \{1, \ldots, n\}.$$

These latter are called ***Hamilton's equations***, and are widely studied in the mathematical physics and dynamical systems communities, for example. We simply refer the reader to [Abraham and Marsden 1978, Arnol'd 1978, Guillemin and Sternberg 1990, Libermann and Marle 1987] for details and further references. The field of Hamiltonian mechanics is a very large one. Readers hoping to have more than a merely functional facility with things Hamiltonian can look forward to investing some time in this.

Remark S1.9. There are differing sign conventions in the literature that one should be aware of. For example, some authors define the cotangent bundle symplectic form with the opposite sign from what we choose. The thing that all conventions have in common is that the local representative of a Hamiltonian vector on a cotangent bundle, with the canonical symplectic form, will take the form (S1.1). Further confusing matters is the possibility of defining the associated map $\omega^\flat$ in two different ways, each differing by a sign. •

S1.2 Tangent and cotangent lifts of vector fields

In this section we introduce some of the necessary tangent bundle geometry that we shall use in subsequent supplements. Most of the constructions we make here are described by Yano and Ishihara [1973], at least in a time-independent setting.

The reader will observe in this section an alternation between the use of the letters M and Q to denote a generic manifold. There is some method behind this. We shall have occasion to use structure that, in the geometric constructions of Section S1.3, might appear on either a configuration manifold Q or its tangent bundle TQ. Such structures we will denote here as occurring on M. That is, when there appears an M in this section, it might refer to either Q or TQ in Supplement 4. We shall suppose both M and Q to be n-dimensional in this section.

We shall make heavy use in this section of the notion of sections of vector bundles that are locally integrally of class C^r for $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$. For convenience, we shall call such sections "LIC^r-sections." Similarly, for convenience we use the acronyms LAC and LAD to stand for "locally absolutely continuous" and "locally absolutely differentiable," respectively.

S1.2.1 More about the tangent lift

In this section we shall define the tangent lift for time-dependent vector fields, and give some further discussion that will be useful for interpreting our results in Supplement 4.

Let $X\colon I \times \mathsf{M} \to \mathsf{TM}$ be an LIC$^\infty$-vector field on a manifold M. Define an LIC$^\infty$-vector field X^T on TM by $X^T(t,x) = X_t^T(x)$, where X_t is the C^∞-vector field on M defined by $X_t(x) = X(t,x)$. The vector field X^T is the ***tangent lift*** of X. This is clearly a generalization of the tangent lift in the text to the time-dependent case.

One may verify in coordinates that

$$X^T = X^i \frac{\partial}{\partial x^i} + \frac{\partial X^i}{\partial x^j} v^j \frac{\partial}{\partial v^i}. \tag{S1.2}$$

From this coordinate expression, we may immediately assert a few useful facts. First we introduce some terminology. For manifolds M and N and for

$f \in C^\infty(\mathsf{M};\mathsf{N})$, an LIC$^\infty$-vector field Y on N is ***f*-related** to an LIC$^\infty$-vector field X on M if, for each $(t,x) \in \mathbb{R} \times \mathsf{M}$, $T_x f(X(t,x)) = Y(t, f(x))$. If $\pi\colon \mathsf{M} \to \mathsf{B}$ is a fiber bundle, an LIC$^\infty$-vector field X on M is ***π*-projectable** if $T\pi(X(t,x_1)) = T\pi(X(t,x_2))$ whenever $\pi(x_1) = \pi(x_2)$. One can easily verify that X is π-projectable if and only if there exists a vector field Y on B such that Y is π-related to X.

Remarks S1.10 (Properties of the tangent lift).

1. Note that X^T is a linear vector field on TM (see Definition S1.25). That is, X^T is π_{TM}-projectable and $X^T_t\colon \mathsf{TM} \to \mathsf{TTM}$ is a vector bundle map for each $t \in I$.
2. Since X^T is π_{TM}-projectable and projects to X, if $t \mapsto \Upsilon(t)$ is an integral curve for X^T, then this curve projects to the curve $t \mapsto \pi_{\mathsf{TM}} \circ \Upsilon(t)$, and this latter curve is further an integral curve for X. Thus integral curves for X^T may be thought of as vector fields along integral curves for X.
3. Let $x \in \mathsf{M}$ and let γ be the integral curve for X with initial condition x at time $t = a$. Let $v_{1,x}, v_{2,x} \in \mathsf{T}_x\mathsf{M}$ with Υ_1 and Υ_2 the integral curves for X^T with initial conditions $v_{1,x}$ and $v_{2,x}$, respectively, at time $t = a$. Then $t \mapsto \alpha_1\, \Upsilon_1(t) + \alpha_2\, \Upsilon_2(t)$ is the integral curve for X^T with initial condition $\alpha_1\, v_{1,x} + \alpha_2\, v_{2,x}$, for $\alpha_1, \alpha_2 \in \mathbb{R}$. That is to say, the family of integral curves for X^T that project to γ is a $\dim(\mathsf{M})$-dimensional vector space.
4. One may think of X^T as the "linearization" of X in the following sense. Let $\gamma\colon I \to \mathsf{M}$ be the integral curve of X through $x \in \mathsf{M}$ at time $t = a$, and let $\Upsilon\colon I \to \mathsf{TM}$ be the integral curve of X^T with initial condition $v_x \in \mathsf{T}_x\mathsf{M}$ at time $t = a$. Choose a variation $\sigma\colon I \times J \to$ of γ with the following properties:
   (a) J is an interval for which $0 \in \operatorname{int}(J)$;
   (b) $s \mapsto \sigma(t,s)$ is differentiable for $t \in I$;
   (c) for $s \in J$, $t \mapsto \sigma(t,s)$ is the integral curve of X through $\sigma(a,s)$ at time $t = a$;
   (d) $\sigma(t,0) = \gamma(t)$ for $t \in I$;
   (e) $v_x = \frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0}\sigma(a,s)$.

   We then have $\Upsilon(t) = \frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0}\sigma(t,s)$. Thus $X^T(v_x)$ measures the "variation" of solutions of X when perturbed by initial conditions lying in the direction of v_x. In cases where M has additional structure, as we shall see, we can make more precise statements about the meaning of X^T. •

S1.2.2 The cotangent lift of a vector field

There is also a cotangent version of X^T that we may define in a natural way. If X is an LIC$^\infty$-vector field on M, we define an LIC$^\infty$-vector field X^{T^*} on $\mathsf{T}^*\mathsf{M}$ by

$$X^{T^*}(t,\alpha_x) = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} T_x^* F_{t,-s}(\alpha_x).$$

This is the ***cotangent lift*** of X. In natural coordinates $((x^1,\dots,x^n),(p_1,\dots,p_n))$ for $\mathsf{T}^*\mathsf{M}$, we have

$$X^{T^*} = X^i \frac{\partial}{\partial x^i} - \frac{\partial X^j}{\partial x^i} p_j \frac{\partial}{\partial p_i}. \tag{S1.3}$$

As was the case with X^T, we may make some immediate useful remarks about the properties of X^{T^*}.

Remarks S1.11 (Properties of the cotangent lift).

1. X^{T^*} is the LIC$^\infty$-Hamiltonian vector field (with respect to the natural symplectic structure on $\mathsf{T}^*\mathsf{M}$) corresponding to the LIC$^\infty$-Hamiltonian $H_X\colon (t,\alpha_x) \mapsto \langle \alpha_x; X(t,x)\rangle$. One can verify this with a direct calculation.
2. Note that X^{T^*} is a linear vector field. That is, X^{T^*} is $\pi_{\mathsf{T}^*\mathsf{M}}$-projectable and $X^{T^*}\colon I \times \mathsf{T}^*\mathsf{M} \to \mathsf{T}\mathsf{T}^*\mathsf{M}$ is a vector bundle map.
3. If $t \mapsto \alpha(t)$ is an integral curve for X^{T^*}, then this curve covers the curve $t \mapsto \pi_{\mathsf{T}^*\mathsf{M}} \circ \alpha(t)$, and this latter curve is further an integral curve for X. Thus one may regard integral curves of X^{T^*} as covector fields along integral curves of X.
4. If $\gamma\colon I \to \mathsf{M}$ is the integral curve for X with initial condition $x \in \mathsf{M}$ at time $t = a \in I$, then the integral curves of X^{T^*} with initial conditions in $\mathsf{T}_x^*\mathsf{M}$ form a $\dim(\mathsf{M})$-dimensional vector space that is naturally isomorphic to $\mathsf{T}_x^*\mathsf{M}$ (in a manner entirely analogous to that described for X^T in Remark S1.10–3).

S1.2.3 Joint properties of the tangent and cotangent lift

By the very virtue of their definitions, together X^T and X^{T^*} should possess some joint properties. To formulate one of these common properties requires some small effort. Let $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ be the Whitney or direct sum of TM and $\mathsf{T}^*\mathsf{M}$.[2] As a manifold, this may be regarded as an embedded submanifold of $\mathsf{TM} \times \mathsf{T}^*\mathsf{M}$ by $v_x \oplus \alpha_x \mapsto (v_x, \alpha_x)$. If X is an LIC$^\infty$-vector field on M, then we define an LIC$^\infty$-vector field $X^T \times X^{T^*}$ on $\mathsf{TM} \times \mathsf{T}^*\mathsf{M}$ by

$$X^T \times X^{T^*}(t,v,\alpha) = (X^T(t,v), X^{T^*}(t,\alpha)).$$

Note that, in this definition, we do not require that $\pi_{\mathsf{TM}}(v) = \pi_{\mathsf{T}^*\mathsf{M}}(\alpha)$.

[2] We recall that the Whitney sum of two vector bundles $\pi_1\colon \mathsf{E}_1 \to \mathsf{M}$ and $\pi_2\colon \mathsf{E}_2 \to \mathsf{M}$ over the same base space can be thought of as the submanifold of $\mathsf{E}_1 \times \mathsf{E}_2$ given by

$$\mathsf{E}_1 \oplus \mathsf{E}_2 = \{\, (v_1,v_2) \in \mathsf{E}_1 \times \mathsf{E}_2 \mid \ \pi_1(v_1) = \pi_2(v_2)\} .$$

This may be verified to be a vector bundle over M with the fiber over $x \in \mathsf{M}$ being naturally isomorphic to $\pi_1^{-1}(x) \oplus \pi_2^{-1}(x)$.

Proposition S1.12 (Relationship between tangent and cotangent lifts I). *$X^T \times X^{T^*}$ is tangent to $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$.*

Proof. We denote natural coordinates for $\mathsf{TM} \times \mathsf{T}^*\mathsf{M}$ by $((\boldsymbol{x}, \boldsymbol{v}), (\boldsymbol{y}, \boldsymbol{p}))$. If we define an $\mathbb{R}^n$-valued function f in these coordinates by $f((\boldsymbol{x}, \boldsymbol{v}), (\boldsymbol{y}, \boldsymbol{p})) = (\boldsymbol{y} - \boldsymbol{x})$, then $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ is locally defined by $f^{-1}(\boldsymbol{0})$. Thus the result will follow if we can show that $X^T \times X^{T^*}$ is in the kernel of $T_{((\boldsymbol{x},\boldsymbol{v}),(\boldsymbol{y},\boldsymbol{p}))}f$ for each $((\boldsymbol{x}, \boldsymbol{v}), (\boldsymbol{y}, \boldsymbol{p})) \in f^{-1}(\boldsymbol{0})$. We compute

$$T_{((\boldsymbol{x},\boldsymbol{v}),(\boldsymbol{y},\boldsymbol{p}))}f((\boldsymbol{v}_1, \boldsymbol{v}_2), (\boldsymbol{v}_3, \boldsymbol{\alpha})) = \boldsymbol{v}_3 - \boldsymbol{v}_1.$$

From this computation, and the local coordinate expressions for X^T and X^{T^*}, the result follows. ■

In this way, the restriction of $X^T \times X^{T^*}$ to $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ makes sense, and we denote the restricted LIC$^\infty$-vector field by $X^T \oplus X^{T^*}$. The following result gives the desired joint property of X^T and X^{T^*}.

Proposition S1.13 (Relationship between tangent and cotangent lifts II). *If X is an LIC$^\infty$-vector field on M, then $X^T \oplus X^{T^*}$ leaves invariant the function $v_x \oplus \alpha_x \mapsto \alpha_x \cdot v_x$ on $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$.*

Proof. We employ a lemma.

Lemma. *If τ is a $(1, 1)$-tensor field on M, then the Lie derivative of the function $f_\tau \colon v_x \oplus \alpha_x \mapsto \tau(\alpha_x, v_x)$ on $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ with respect to the vector field $X^T \oplus X^{T^*}$ is the function $v_x \oplus \alpha_x \mapsto (\mathscr{L}_X\tau)(\alpha_x, v_x)$.*

Proof. We work in local coordinates where $f_\tau = \tau^i_j p_i v^j$. We then compute

$$\mathscr{L}_{X^T \oplus X^{T^*}} f_\tau = \frac{\partial \tau^i_j}{\partial x^k} X^k p_i v^j + \frac{\partial X^k}{\partial x^j} \tau^i_k p_i v^j - \frac{\partial X^i}{\partial x^k} \tau^k_j p_i v^j,$$

which we readily verify agrees with the coordinate expression for $(\mathscr{L}_X\tau)(\alpha_x, v_x)$. ▼

We now observe that the function $v_x \otimes \alpha_x \mapsto \alpha_x \cdot v_x$ is exactly $f_{\mathrm{id}_{\mathsf{TM}}}$ in the notation of the lemma. It thus suffices to show that $\mathscr{L}_X \mathrm{id}_{\mathsf{TM}} = 0$ for any vector field X. But, if we Lie differentiate the equality $\mathrm{id}_{\mathsf{TM}}(Y) = Y$ ($Y \in \Gamma^\infty(\mathsf{TM})$) with respect to X, then we obtain

$$(\mathscr{L}_X \mathrm{id}_{\mathsf{TM}})(Y) + \mathrm{id}_{\mathsf{TM}}([X, Y]) = [X, Y]$$

from which the proposition follows. ■

Remark S1.14. One may verify, in fact, that X^{T^*} is the unique linear vector field on $\mathsf{T}^*\mathsf{M}$ that projects to X and which satisfies Proposition S1.13. •

S1.2.4 The cotangent lift of the vertical lift

As mentioned in the introduction, we shall deal with systems whose state space is a tangent bundle, and whose control vector fields are vertical lifts. As a consequence of an application of the Maximum Principle to such systems, we will be interested in the cotangent lift of vertically lifted vector fields. So let Q be a C^∞-manifold, and let X be an LIC$^\infty$-vector field on Q with $\mathrm{vlft}(X)$ its vertical lift to TQ. One computes the local coordinate expression for $\mathrm{vlft}(X)^{T^*}$ to be

$$(\mathrm{vlft}(X))^{T^*} = X^i \frac{\partial}{\partial v^i} - \frac{\partial X^j}{\partial q^i} \beta_j \frac{\partial}{\partial \alpha_i}. \tag{S1.4}$$

Here we write natural coordinates for $\mathsf{T^*TQ}$ as $((\boldsymbol{q}, \boldsymbol{v}), (\boldsymbol{\alpha}, \boldsymbol{\beta}))$.

Remark S1.15. It is interesting to note the relationship between $(\mathrm{vlft}(X))^{T^*}$ and $\mathrm{vlft}(X^{T^*})$. The latter vector field has the coordinate expression

$$\mathrm{vlft}(X^{T^*}) = X^i \frac{\partial}{\partial u^i} - \frac{\partial X^j}{\partial q^i} p_j \frac{\partial}{\partial \gamma_i},$$

where we are writing natural coordinates for $\mathsf{TT^*Q}$ as $((\boldsymbol{q}, \boldsymbol{p}), (\boldsymbol{u}, \boldsymbol{\gamma}))$. Now we note that there is a canonical diffeomorphism ϕ_{Q} between $\mathsf{T^*TQ}$ and $\mathsf{TT^*Q}$ defined in coordinates by

$$((\boldsymbol{q}, \boldsymbol{v}), (\boldsymbol{\alpha}, \boldsymbol{\beta})) \mapsto ((\boldsymbol{q}, \boldsymbol{\beta}), (\boldsymbol{v}, \boldsymbol{\alpha})).$$

One easily verifies that $\mathrm{vlft}(X^{T^*}) = \phi_{\mathsf{Q}}^*(\mathrm{vlft}(X))^{T^*}$. We also remark that $\mathsf{T^*TQ}$ is a symplectic manifold, since it is a cotangent bundle. Tulczyjew [1977] demonstrates that the tangent bundle of a symplectic manifold is also a symplectic manifold. Thus, in particular, $\mathsf{TT^*Q}$ is a symplectic manifold. The symplectic structure on $\mathsf{TT^*Q}$ as defined by Tulczyjew is given in coordinates by

$$\omega_{\mathsf{TT^*Q}} = \mathrm{d}q^i \wedge \mathrm{d}\gamma_i + \mathrm{d}u^i \wedge \mathrm{d}p_i.$$

One then verifies that the diffeomorphism ϕ_{Q} is symplectic with respect to these symplectic structures. That is to say, $\phi_{\mathsf{Q}}^*\omega_{\mathsf{TT^*Q}}$ is the canonical symplectic form on $\mathsf{T^*TQ}$. Since $(\mathrm{vlft}(X))^{T^*}$ is a Hamiltonian vector field on $\mathsf{T^*TQ}$ by Remark S1.11–1, the vector field $\mathrm{vlft}(X^{T^*})$ must also be Hamiltonian on $\mathsf{TT^*Q}$ with the symplectic structure just described. The Hamiltonian, one readily computes, is given by $V_{\alpha_q} \mapsto \langle \alpha_q; X(q) \rangle$, where $V_{\alpha_q} \in \mathsf{TT^*Q}$.

An intrinsic definition of ϕ_{Q} is as follows.[3] We define a map $\rho\colon \mathsf{T^*TQ} \to \mathsf{T^*Q}$ as follows:

$$\left\langle \rho(\alpha_{v_q}); u_q \right\rangle = \left\langle \alpha_{v_q}; \mathrm{vlft}_{v_q}(u_q) \right\rangle.$$

We may then readily verify that ϕ_{Q} is the unique map that makes the diagram

[3] The authors thank Jerry Marsden for providing this definition.

$$\begin{tikzcd}
\mathsf{T}^*\mathsf{Q} & \mathsf{T}\mathsf{T}^*\mathsf{Q} \arrow[l, "\pi_{\mathsf{T}\mathsf{T}^*\mathsf{Q}}"'] \arrow[r, "T\pi_{\mathsf{T}^*\mathsf{Q}}"] & \mathsf{T}\mathsf{Q} \\
& \mathsf{T}^*\mathsf{T}\mathsf{Q} \arrow[ul, "\rho"] \arrow[u, "\phi_{\mathsf{Q}}"] \arrow[ur, "\pi_{\mathsf{T}^*\mathsf{T}\mathsf{Q}}"'] &
\end{tikzcd}$$

commute. •

S1.2.5 The canonical involution of TTQ

Let ρ_1 and ρ_2 be C^2 maps from a neighborhood of $(0,0) \in \mathbb{R}^2$ to Q. Let us denote by (t_1, t_2) coordinates for $\mathbb{R}^2$. We say two such maps are ***equivalent*** if $\rho_1(0,0) = \rho_2(0,0)$ and if

$$\frac{\partial \rho_1}{\partial t_1}(0,0) = \frac{\partial \rho_2}{\partial t_1}(0,0), \qquad \frac{\partial \rho_1}{\partial t_2}(0,0) = \frac{\partial \rho_2}{\partial t_2}(0,0),$$
$$\frac{\partial^2 \rho_1}{\partial t_1 \partial t_2}(0,0) = \frac{\partial^2 \rho_2}{\partial t_1 \partial t_2}(0,0).$$

To an equivalence class $[\rho]$ we associate a point in TTQ as follows. For fixed t_2, consider the curve γ_{t_2} at $q \triangleq \rho(0,0)$ given by $t_1 \mapsto \rho(t_1, t_2)$. Then $\gamma'_{t_2}(0)$ is a tangent vector in $\mathsf{T}_q\mathsf{Q}$. Therefore, $t_2 \mapsto \Upsilon_\rho(t_2) \triangleq \gamma'_{t_2}(0)$ is a curve in TQ at $\gamma'_{t_2}(0)$. To ρ we then assign the tangent vector $\Upsilon'_\rho(0)$. In natural coordinates $((\boldsymbol{q}, \boldsymbol{v}), (\boldsymbol{u}, \boldsymbol{w}))$ for TTQ, the point associated to $[\rho]$ is given by

$$\Big(\rho(0,0), \frac{\partial \rho}{\partial t_1}(0,0), \frac{\partial \rho}{\partial t_2}(0,0), \frac{\partial^2 \rho}{\partial t_1 \partial t_2}(0,0)\Big).$$

This then shows that the correspondence between equivalence classes and points in TTQ is bijective. Furthermore, the construction can be readily extended to give a construction of the higher-order tangent bundles $\mathsf{T}^k\mathsf{M}$, where $\mathsf{T}^1\mathsf{M} = \mathsf{TM}$ and, inductively, $\mathsf{T}^k\mathsf{M} = \mathsf{T}(\mathsf{T}^{k-1}\mathsf{M})$.

Associated with this representation of points in TTQ is an involution[4] $I_{\mathsf{Q}} \colon \mathsf{TTQ} \to \mathsf{TTQ}$. We define I_{Q} by saying how it acts on equivalence classes as given above. If ρ is a map from a neighborhood of $(0,0) \in \mathbb{R}^2$ to Q, then we define $\bar{\rho}(t_1, t_2) = \rho(t_2, t_1)$ which is also then a map from a neighborhood of $(0,0) \in \mathbb{R}^2$ into Q. We then define

$$I_{\mathsf{Q}}([\rho]) = [\bar{\rho}].$$

In coordinates,

$$I_{\mathsf{Q}}((\boldsymbol{q}, \boldsymbol{v}), (\boldsymbol{u}, \boldsymbol{w})) = ((\boldsymbol{q}, \boldsymbol{u}), (\boldsymbol{v}, \boldsymbol{w})).$$

Definition S1.16 (Canonical involution of TTQ). The map I_{Q} is the ***canonical involution*** of TTQ. •

[4] An ***involution*** on a set S is a map $f \colon S \to S$ with the property that $f \circ f = \mathrm{id}_S$.

S1.2.6 The canonical endomorphism of the tangent bundle

The final bit of tangent bundle geometry we discuss is that of a natural $(1,1)$-tensor field on the tangent bundle.

Definition S1.17 (Canonical endomorphism of TM). For a manifold M, the ***canonical endomorphism of*** TM is the $(1,1)$-tensor field J_{M} on TM defined by

$$J_{\mathsf{M}}(X_{v_x}) = \mathrm{vlft}_{v_x}(T_{v_x}\pi_{\mathsf{TM}}(X_{v_x}))$$

where $X_{v_x} \in \mathsf{T}_{v_x}\mathsf{TM}$. •

One verifies that, in natural coordinates, we have

$$J_{\mathsf{M}} = \frac{\partial}{\partial v^i} \otimes \mathrm{d}x^i.$$

The role of the canonical endomorphism in Lagrangian mechanics is discussed by Crampin [1983]. It is possible to talk about related structures in more general settings than tangent bundles; see [e.g., Thompson and Schwardmann 1991].

S1.3 Ehresmann connections induced by an affine connection

This section provides the essential ingredients for the development of linearization of affine connection control systems in Supplement 3, and of the Maximum Principle for affine connection control systems in Supplement 4. In actuality, the results of this section, particularly those of Section S1.3.10, represent the meat of this supplement since the linearization and optimal control results of Chapters 3 and 4 follow in a fairly straightforward way once one has at one's disposal the results that we now provide. The constructions are quite involved, so some motivation is in order.

S1.3.1 Motivating remarks

As we have stated several times already, we will be looking at control-affine systems whose drift vector field is the geodesic spray S for an affine connection. Readers familiar with linearization will recognize that the tangent lift of S will be important for us, and readers familiar with the geometry of the Maximum Principle will immediately realize that the cotangent lift of S will be important for us. One way to frame the objective of this section is to think about how one might represent S^T and S^{T^*} in terms of objects defined on Q, even though S^T and S^{T^*} are themselves vector fields on TTQ and $\mathsf{T^*TQ}$, respectively. That this ought to be possible seems reasonable as all the information used to describe S^T and S^{T^*} is contained in the affine connection ∇ on Q, along with some canonical tangent and cotangent bundle geometry. It turns out that it *is* possible to essentially represent S^T and S^{T^*} on Q, but to do so requires some effort. What is more, it is perhaps not immediately obvious how one should proceed. It turns out that a good way to motivate oneself is to think first about S^T, then use this to get to an understanding of S^{T^*}.

To understand the meaning of S^T, consider the following construction. Let $\gamma\colon I \to \mathsf{Q}$ be a geodesic for an affine connection ∇. Let $\sigma\colon I \times J \to \mathsf{Q}$ be a variation of γ. Thus

1. J is an interval for which $0 \in \operatorname{int}(J)$,
2. $s \mapsto \sigma(t, s)$ is differentiable for $t \in I$,
3. for $s \in J$, $t \mapsto \sigma(t, s)$ is a geodesic of ∇, and
4. $\sigma(t, 0) = \gamma(t)$ for $t \in I$.

If one defines $\xi(t) = \frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0}\sigma(t, s)$, then it can be shown (see Theorem 1.2 in Chapter VIII of volume 2 of [Kobayashi and Nomizu 1963]) that ξ satisfies the ***Jacobi equation***:[5]

$$\nabla^2_{\gamma'(t)}\xi(t) + R(\xi(t), \gamma'(t))\gamma'(t) + \nabla_{\gamma'(t)}(T(\xi(t), \gamma'(t))) = 0,$$

where T is the torsion tensor and R is the curvature tensor for ∇. Thus the Jacobi equation tells us how geodesics vary along γ as we vary their initial conditions.

With this as backdrop, a possible way to get moving in the right direction is as follows:

1. according to Remark S1.10–4 and the very definition of the Jacobi equation, we expect there to be some relationship between S^T and the Jacobi equation;
2. by Remark S1.14, there is a relationship between S^T and S^{T^*};
3. from 1 and 2, we may expect that, by coming to understand the relationship between S^{T^*} and the Jacobi equation, one may be able to see how to essentially represent S^{T^*} on Q.

One sees, then, that our approach to understanding S^{T^*} entails that we first understand S^T in terms of the Jacobi equation. The Jacobi equation itself is, evidently, useful in understanding linearization of affine connection control systems. Moreover, we shall see that once we have understood the relationship between the Jacobi equation and S^T, it is a simple matter to "dualize" our constructions to arrive at what we shall call the adjoint Jacobi equation, a covector field version of the Jacobi equation that appears in the Maximum Principle for affine connection control systems.

[5] We use the following notation. Let $\gamma\colon I \to \mathsf{Q}$ be an LAC curve and let $\tau\colon I \to T^r_s(\mathsf{TQ})$ be a section along γ that is sufficiently smooth that the constructions we are about to make are well-defined. For $k \in \mathbb{N}$, we define a section $t \mapsto \nabla^k_{\gamma'(t)}\tau(t)$ of $T^r_s(\mathsf{TQ})$ along γ by setting $\nabla^1_{\gamma'(t)}\tau(t) = \nabla_{\gamma'(t)}\tau(t)$, and inductively defining $\nabla^k_{\gamma'(t)}\tau(t) = \nabla_{\gamma'(t)}(\nabla^{k-1}_{\gamma'(t)}\tau(t))$.

In order to simultaneously understand S^T and S^{T^*}, we use various Ehresmann connections to provide splittings of the necessary tangent spaces. Note that, as maps, the vector fields S^T and S^{T^*} are $\mathsf{T}(\mathsf{TTQ})$- and $\mathsf{T}(\mathsf{T}^*\mathsf{TQ})$-valued, respectively. Also note that the tangent spaces to TTQ and $\mathsf{T}^*\mathsf{TQ}$ are $4\dim(\mathsf{Q})$-dimensional. Our goal is to break these tangent spaces up into four parts, each of dimension $\dim(\mathsf{Q})$. Some of the Ehresmann connections we describe here are well-known, but others might be new, even though they are straightforward to describe. We refer to [Kolář, Michor, and Slovák 1993, Chapter III] for a general discussion of Ehresmann connections. Along the way, we will point out various interesting relationships between the objects we encounter. Some of these relationships are revealed in the book of Yano and Ishihara [1973].

The reader wishing to cut to the chase and see the point of producing all of these Ehresmann connections is referred forward to Theorems S1.34 and S1.38.

S1.3.2 More about vector and fiber bundles

Our ensuing discussion will be aided by some additional terminology concerning fiber bundles in general, and vector bundles in particular. This section, therefore, supplements the material in Section 3.9.5 in the text.

For a fiber bundle, it is possible to choose a chart in such a way that it is natural with respect to the projection. For vector bundles, we see this in the form of vector bundle charts. For fiber bundles we have the following notion.

Definition S1.18 (Fiber bundle chart). Let $(\pi, \mathsf{M}, \mathsf{B}, \mathsf{F})$ be a locally trivial fiber bundle. A chart $(\mathcal{V}, \psi)$ for M is a ***fiber bundle chart*** if there exists a chart $(\mathcal{U}_0, \phi_0)$ for B and a chart $(\mathcal{U}_1, \phi_1)$ for F with the following properties:

(i) $\mathcal{V} \subset \pi^{-1}(\mathcal{U}_0)$;

(ii) there is a diffeomorphism χ from $\pi^{-1}(\mathcal{U}_0)$ to $\mathcal{U}_0 \times \mathsf{F}$ with the property that $\mathrm{pr}_1 \circ \chi(x) = \pi(x)$ for each $x \in \mathcal{V}$;

(iii) $\chi(\mathcal{V}) = \mathcal{U}_0 \times \mathcal{U}_1$.

If coordinates for $(\mathcal{U}_0, \phi_0)$ are denoted by $(x^1, \dots, x^m)$ and if coordinates for $(\mathcal{U}_1, \phi_1)$ are denoted by $(y^1, \dots, y^{n-m})$, then the coordinates $((x^1, \dots, x^m), (y^1, \dots, y^{n-m}))$ in the chart $(\mathcal{V}, \tilde{\psi})$, with $\tilde{\psi}(x) = (\phi_0 \circ \mathrm{pr}_1 \circ \chi(x), \phi_1 \circ \mathrm{pr}_2 \circ \chi(x))$, are called ***fiber bundle coordinates***. •

Note that the local representative of π in fiber bundle coordinates is $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x}$.

The next concept we discuss concerns vector bundles. It provides a useful way of constructing a vector bundle from an existing vector bundle and a map.

Definition S1.19 (Pull-back vector bundle). Let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a vector bundle and let $f\colon \mathsf{N} \to \mathsf{M}$ be a smooth map. The ***pull-back*** of E to N by f is the submanifold

$$f^*\mathsf{E} = \{\,(v,y) \in \mathsf{E}\times\mathsf{N} \mid \ \pi(v) = f(y)\}\,.$$

The map $f^*\pi\colon f^*\mathsf{E}\to\mathsf{N}$ is defined by $f^*\pi(v,y)=y$. •

It is not obvious, but it is true, that $f^*\pi\colon f^*\mathsf{E}\to\mathsf{N}$ is a vector bundle. The fiber over $y\in\mathsf{N}$ is diffeomorphic to $\pi^{-1}(f(y))$. Let us prove that there is a natural vector bundle structure on the pull-back of E to N by f.

Proposition S1.20 (Vector bundle structure for pull-back bundle). $f^*\pi\colon f^*\mathsf{E}\to\mathsf{N}$ *naturally possess the structure of a vector bundle.*

Proof. We shall construct vector bundle charts for $f^*\mathsf{E}$. Let $(\mathcal{U},\phi)$ be a chart for N and let $(\mathcal{V},\psi)$ be a vector bundle chart for E so that $f(\mathcal{U})\subset Z(\mathsf{E})\cap\mathcal{V}$. This defines an open set $\mathcal{V}\times\mathcal{U}\subset\mathsf{E}\times\mathsf{N}$. If M is modeled $\mathbb{R}^m$, N is modeled on $\mathbb{R}^n$, and if the fibers of E are isomorphic to $\mathbb{R}^k$, then we have

$$\begin{aligned}\psi\times\phi\colon\ \mathcal{V}\times\mathcal{U} &\to\ (\mathbb{R}^m\times\mathbb{R}^k)\times\mathbb{R}^n\\ (v,y)&\mapsto\ (\psi(v),\phi(y)).\end{aligned}$$

Denote by $\tilde f\colon\phi(\mathcal{U})\to\psi(\mathcal{V}\cap Z(\mathsf{E}))$ the local representative of f. With this notation, locally the subset $f^*\mathsf{E}$ of $\mathsf{E}\times\mathsf{N}$ is given by

$$\tilde{f}^*\mathsf{E}\triangleq\{((\boldsymbol{x},\boldsymbol{v}),\boldsymbol{y})\in\psi(\mathcal{V})\times\phi(\mathcal{U})\mid\ \boldsymbol{x}=\tilde f(\boldsymbol{y})\}.$$

Now define a map g from $\tilde f^*\mathsf{E}$ to $\psi(\mathcal{V})\times\mathbb{R}^k$ by $g((\boldsymbol{x},\boldsymbol{v}),\boldsymbol{y})=(\boldsymbol{y},\boldsymbol{v})$. We claim that

$$\begin{aligned}\{((\mathcal{V}\times\mathcal{U})\cap f^*\mathsf{E}, g\circ((\psi\times\phi)|f^*\mathsf{E}))\mid\\ (\mathcal{V},\psi)\text{ is a vector bundle chart for }\mathsf{E}\text{ and }(\mathcal{U},\phi)\text{ is a chart for }\mathsf{N}\\ \text{for which }f(\mathcal{U})\subset\mathcal{V}\cap Z(\mathsf{E})\}\end{aligned}$$

is a vector bundle atlas for $f^*\mathsf{E}$. We must verify the overlap conditions. We simplify things by assuming another chart for N of the form $(\mathcal{U},\phi')$ (i.e., the domain is the same as the chart $(\mathcal{U},\phi)$) and a vector bundle chart for E of the form $(\mathcal{V},\psi')$ (again the domain is the same). These simplifications can always be made by restriction if necessary. Since the charts $(\mathcal{U},\phi)$ and $(\mathcal{U},\phi')$ satisfy the overlap conditions, it holds that

$$\phi'\circ\phi^{-1}\colon\phi(\mathcal{U})\to\phi'(\mathcal{U})$$

is a diffeomorphism. Similarly, since $(\mathcal{V},\psi)$ and $(\mathcal{V},\psi')$ are vector bundle charts,

$$\psi'\circ\psi^{-1}(\boldsymbol{x},\boldsymbol{v})=(\sigma(\boldsymbol{x}),A(\boldsymbol{x})\cdot\boldsymbol{v})$$

where $\sigma\colon\psi(\mathcal{V}\cap Z(\mathsf{E}))\to\psi'(\mathcal{V}\cap Z(\mathsf{E}))$ is a diffeomorphism and $A\colon\psi(\mathcal{V}\cap Z(\mathsf{E}))\to\mathsf{GL}(k;\mathbb{R})$ is smooth.

Now we consider the two charts

$$((\mathcal{V}\times\mathcal{U})\cap f^*\mathsf{E}, g\circ((\psi\times\phi)|f^*\mathsf{E}))$$
$$((\mathcal{V}\times\mathcal{U})\cap f^*\mathsf{E}, g\circ((\psi'\times\phi')|f^*\mathsf{E}))$$

for $f^*\mathsf{E}$, and show that they satisfy the overlap conditions. Let $(v,y)\in(\mathcal{V}\times\mathcal{U})\cap f^*\mathsf{E}$. We write

$$\psi\times\phi(v,y) = ((\tilde{f}(\boldsymbol{y}),\boldsymbol{v}),\boldsymbol{y})$$

defining $\boldsymbol{y}\in\phi(\mathcal{U})$ and $\boldsymbol{v}\in\mathbb{R}^k$. If $\tilde{f}'\colon\phi'(\mathcal{U})\to\psi'(\mathcal{V}\cap Z(\mathsf{E}))$ is the local representative of f in the "primed" chart, then we may write

$$\psi'\times\phi'(v,y) = ((\tilde{f}'(\boldsymbol{y}'),\boldsymbol{v}'),\boldsymbol{y}')$$

defining $\boldsymbol{y}'\in\phi'(\mathcal{U})$ and $\boldsymbol{v}'\in\mathbb{R}^k$. Since $(\mathcal{U},\phi)$, $(\mathcal{U},\phi')$, $(\mathcal{V},\psi)$, and $(\mathcal{V},\psi')$ satisfy the overlap conditions, we must have

$$\boldsymbol{y}' = \phi'\circ\phi^{-1}(\boldsymbol{y}),\quad \boldsymbol{v}' = (A\circ\sigma(\boldsymbol{y}))\cdot\boldsymbol{v}.$$

This shows that the overlap condition is indeed satisfied. ■

S1.3.3 Ehresmann connections

In this section we introduce an important construction that can be associated to a fiber bundle $\pi\colon\mathsf{M}\to\mathsf{B}$. We recall that $\mathsf{VM}\triangleq\ker(T\pi)$ is the ***vertical subbundle*** of TM.

Definition S1.21 (Ehresmann connection). An ***Ehresmann connection*** on a locally trivial fiber bundle $\pi\colon\mathsf{M}\to\mathsf{B}$ is a complement HM to VM in TM, i.e., a subbundle of TM for which $\mathsf{TM}=\mathsf{HM}\oplus\mathsf{VM}$. We call HM the ***horizontal subbundle***. •

We also say that elements in VM are ***vertical*** and that elements in HM are ***horizontal***. We denote by $\mathrm{hor}\colon\mathsf{TM}\to\mathsf{TM}$ the horizontal projection, and by $\mathrm{ver}\colon\mathsf{TM}\to\mathsf{TM}$ the vertical projection. Note that, for each $x\in\mathsf{M}$, $T_x\pi|\mathsf{H}_x\mathsf{M}\colon\mathsf{H}_x\mathsf{M}\to\mathsf{T}_{\pi(x)}\mathsf{B}$ is an isomorphism. We denote its inverse by $\mathrm{hlft}_x\colon\mathsf{T}_{\pi(x)}\mathsf{B}\to\mathsf{H}_x\mathsf{M}$, which is called the ***horizontal lift***. If $((x^1,\dots,x^m),(y^1,\dots,y^{n-m}))$ are fiber bundle coordinates for M, then we have

$$\mathrm{hlft}_{(x,y)}\Big(\frac{\partial}{\partial x^a}\Big) = \frac{\partial}{\partial x^a} + C_a^\alpha(x,y)\frac{\partial}{\partial y^\alpha},\qquad a\in\{1,\dots,m\}.$$

This defines the ***connection coefficients*** C_a^α, $\alpha\in\{1,\dots,n-m\}$, $a\in\{1,\dots,m\}$.

We next associate two important objects to an Ehresmann connection. In order to define these, it is convenient to first give a general definition. For $k\in\mathbb{N}$, we denote

$$\underbrace{\mathsf{TM}\times_{\mathsf{M}}\cdots\times_{\mathsf{M}}\mathsf{TM}}_{k\text{ copies}} = \big\{(v_1,\dots,v_k)\in(\mathsf{TM})^k\ \big|\ \pi_{\mathsf{TM}}(v_1)=\dots=\pi_{\mathsf{TM}}(v_k)\big\}.$$

With this notation, we have the following definition.

Definition S1.22 (Bundle-valued differential form). Let $\rho\colon \mathsf{E} \to \mathsf{M}$ be a C^r-vector bundle, $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$. An ***$\mathsf{E}$-valued differential k-form of class C^r*** on M is a C^r-map

$$\omega\colon \underbrace{\mathsf{TM} \times_{\mathsf{M}} \cdots \times_{\mathsf{M}} \mathsf{TM}}_{k \text{ copies}} \to \mathsf{E}$$

with the property that, for each C^r-section α of E^*, the map that assigns to $X_1, \dots, X_k \in \Gamma^r(\mathsf{TM})$ the C^r-function

$$x \mapsto \langle \alpha(x); \omega(X_1(x), \dots, X_k(x)) \rangle$$

is a C^r-differential k-form (cf. the discussion at the end of Section 3.4.2). •

Now we define two bundle-valued differential forms associated to an Ehresmann connection.

Definition S1.23 (Connection and curvature form). Given an Ehresmann connection HM on a locally trivial fiber bundle $\pi\colon \mathsf{M} \to \mathsf{B}$,

(i) the ***connection form*** is the VM-valued differential one-form ω_{HM} defined by

$$\omega_{\mathsf{HM}}(v_x) = \operatorname{ver}(v_x),$$

and

(ii) the ***curvature form*** is the VM-valued differential two-form Ω_{HM} given by

$$\Omega_{\mathsf{HM}}(u_x, v_x) = -\omega_{\mathsf{HM}}([\operatorname{hor}(U), \operatorname{hor}(V)](x))$$

where U and V are vector fields that extend u_x and v_x, respectively. •

One verifies that Ω_{HM} does not depend on the extensions. It is a straightforward exercise to see that $\Omega_{\mathsf{HM}} = 0$ if and only if HM is integrable. In this case, one says that the Ehresmann connection HM is ***flat***.

S1.3.4 Linear connections and linear vector fields on vector bundles

In this section we generalize the constructions of Section S1.2.3.

If $\pi\colon \mathsf{E} \to \mathsf{B}$ is a vector bundle, then VE is isomorphic to the pull-back bundle $\pi^*\pi\colon \pi^*\mathsf{E} \to \mathsf{E}$ whose fiber over $e_b \in \mathsf{E}$ is exactly $\pi^{-1}(b)$. Thus the projection ver associated with an Ehresmann connection HE may be thought of as a map $\operatorname{ver}\colon \mathsf{TE} \to \mathsf{E}$. Further, we define a vector bundle isomorphism $\operatorname{vlft}\colon \pi^*\mathsf{E} \to \mathsf{VE}$ by

$$\operatorname{vlft}(\tilde{e}_b, e_b) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\tilde{e}_b + t e_b).$$

We adopt the standard notation and write $\operatorname{vlft}_{\tilde{e}_b}(e_b)$ rather than $\operatorname{vlft}(\tilde{e}_b, e_b)$.

We now give an important class of Ehresmann connections on vector bundles.

Definition S1.24 (Linear Ehresmann connection). An Ehresmann connection HE on a vector bundle $\pi\colon \mathsf{E} \to \mathsf{B}$ is ***linear*** if the vertical projection ver: $\mathsf{TE} \to \mathsf{E}$ is a vector bundle map with respect to the vector bundles $T\pi\colon \mathsf{TE} \to \mathsf{TB}$ and $\pi_{\mathsf{TE}}\colon \mathsf{TE} \to \mathsf{E}$. •

One verifies that an Ehresmann connection is linear if and only if the connection coefficients have the form $C^\alpha_a(\boldsymbol{x}, \boldsymbol{u}) = A^\alpha_{a\beta}(\boldsymbol{x})u^\beta$, $\alpha \in \{1, \dots, n - m\}$, $a \in \{1, \dots, m\}$, where $(\boldsymbol{x}, \boldsymbol{u})$ are vector bundle coordinates. This defines local functions $A^\alpha_{a\beta}$, $\alpha, \beta \in \{1, \dots, n-m\}$, $a \in \{1, \dots, m\}$, on the base space.

On vector bundles, one also has a distinguished class of vector fields.

Definition S1.25 (Linear vector field). Let X be a vector field on the base space B of a vector bundle $\pi\colon \mathsf{E} \to \mathsf{B}$. A ***linear vector field*** over X is a vector field $Y\colon \mathsf{E} \to \mathsf{TE}$ that is π-related to X and that is a vector bundle map that makes the following diagram commutative:

$$\begin{tikzcd} \mathsf{E} \arrow[r, "Y"] \arrow[d, "\pi"'] & \mathsf{TE} \arrow[d, "T\pi"] \\ \mathsf{B} \arrow[r, "X"'] & \mathsf{TB} \end{tikzcd}$$

•

In vector bundle coordinates $(\boldsymbol{x}, \boldsymbol{u})$, a linear vector field Y over X has the form

$$Y = X^a(\boldsymbol{x})\frac{\partial}{\partial x^a} + Y^\alpha_\beta(\boldsymbol{x})u^\beta \frac{\partial}{\partial u^\alpha}, \tag{S1.5}$$

where X^a, $a \in \{1, \dots, m\}$, are the components of X, and for some functions Y^α_β, $\alpha, \beta \in \{1, \dots, n - m\}$.

Next we define a vector field "dual" to a given linear vector field Y over X. We refer to [Kolář, Michor, and Slovák 1993] for the fairly straightforward details. The flow of a linear vector field is comprised of local vector bundle isomorphisms of $\pi\colon \mathsf{E} \to \mathsf{M}$. Thus, if $\pi^*\colon \mathsf{E}^* \to \mathsf{B}$ is the dual bundle, and if ν is a linear vector field on E^* over X, then we may define $Y \oplus \nu$ as the linear vector field on $\mathsf{E} \oplus \mathsf{E}^*$ whose flow is the family of vector bundle isomorphisms given by the direct sum of those generated by Y and by ν. We may define a function f_{E} on $\mathsf{E} \oplus \mathsf{E}^*$ by $f_{\mathsf{E}}(e_b \oplus \alpha_b) = \alpha_b \cdot e_b$.

Lemma S1.26 (Dual of a linear vector field). *If Y is a linear vector field over X on E, then there exists a unique linear vector field over X on E^*, denoted by Y^* and called the **dual** of Y, with the property that $\mathscr{L}_{Y \oplus Y^*} f_{\mathsf{E}} = 0$.*

In vector bundle coordinates $(\boldsymbol{x}, \boldsymbol{\rho})$ for E^*, if Y is as given by (S1.5), then

$$Y^* = X^a(\boldsymbol{x})\frac{\partial}{\partial x^a} - Y^\alpha_\beta(\boldsymbol{x})\rho_\alpha \frac{\partial}{\partial \rho_\beta}.$$

Of course, when $\mathsf{E} = \mathsf{TM}$ and $Y = X^T$, we see that $Y^* = X^{T^*}$, consistent with Remark S1.14.

Using the dual of a linear vector field, one can define the dual of a linear connection [Kolář, Michor, and Slovák 1993, Section 47.15].

Lemma S1.27 (Dual of a linear connection). *If* HE *is a linear connection on a vector bundle* $\pi\colon \mathsf{E} \to \mathsf{B}$*, then there exists a unique linear connection* HE^* *on the dual bundle* $\pi^*\colon \mathsf{E}^* \to \mathsf{B}$ *that satisfies the property*

$$\mathrm{hlft}(X)^* = \mathrm{hlft}^*(X), \tag{S1.6}$$

where X *is a vector field on* B*,* hlft *is the horizontal lift associated with* HE*, and* hlft^* *is the horizontal lift associated with* HE^**.*

If $A^{\alpha}_{a\beta}(\boldsymbol{x})u^{\beta}$, $\alpha \in \{1,\dots,n-m\}$, $a \in \{1,\dots,m\}$, are the connection coefficients for HE in vector bundle coordinates $(\boldsymbol{x},\boldsymbol{u})$, then the connection coefficients for HE^* in the dual vector bundle coordinates $(\boldsymbol{x},\boldsymbol{\rho})$ are $-A(\boldsymbol{x})^{\beta}_{a\alpha}\rho_{\beta}$, $a \in \{1,\dots,m\}$, $\alpha \in \{1,\dots,n-m\}$.

S1.3.5 The Ehresmann connection on $\pi_{\mathsf{TM}}\colon \mathsf{TM} \to \mathsf{M}$ associated with a second-order vector field on TM

Let us define the notion of a second-order vector field.

Definition S1.28 (Second-order vector field). A vector field S on TM is ***second-order*** if $T\pi_{\mathsf{TM}} \circ S = \mathrm{id}_{\mathsf{TM}}$. •

One can readily verify that a vector S is second-order if and only if, in natural coordinates for TM, the local representative of S is

$$S = v^i \frac{\partial}{\partial x^i} + S^i(\boldsymbol{x},\boldsymbol{v})\frac{\partial}{\partial v^i}, \tag{S1.7}$$

where S^i, $i \in \{1,\dots,n\}$, are smooth functions of the coordinates. For a second-order vector field S on TM, we define an Ehresmann connection on $\pi_{\mathsf{TM}}\colon \mathsf{TM} \to \mathsf{M}$ as follows [Crampin 1983]. Recall from Section S1.2.6 the canonical endomorphism J_{M} on TM. One may verify that the kernel of the vector bundle map $(\mathscr{L}_S J_{\mathsf{M}} + \mathrm{id}_{\mathsf{TT}M})\colon \mathsf{TTM} \to \mathsf{TTM}$ is a subbundle complementary to $\mathsf{VTM} = \ker(T\pi_{\mathsf{TM}})$. We denote this complementary distribution, which is thus an Ehresmann connection, by HTM. One then verifies that a local basis for HTM is given by the vector fields

$$\mathrm{hlft}\Big(\frac{\partial}{\partial x^i}\Big) = \frac{\partial}{\partial x^i} + \frac{1}{2}\frac{\partial S^j}{\partial v^i}\frac{\partial}{\partial v^j}, \qquad i \in \{1,\dots,n\}, \tag{S1.8}$$

where the functions S^i, $i \in \{1,\dots,n\}$, are as in (S1.7).

This Ehresmann connection gives a splitting of $\mathsf{T}_{v_x}\mathsf{TM}$ into a horizontal and a vertical part. The horizontal part is isomorphic to $\mathsf{T}_x\mathsf{M}$ via hlft_{v_x}, and the vertical part is isomorphic to $\mathsf{T}_x\mathsf{M}$ in the natural way (it is the tangent space to a vector space). Thus we have a natural isomorphism $\mathsf{T}_{v_x}\mathsf{TM} \simeq \mathsf{T}_x\mathsf{M} \oplus \mathsf{T}_x\mathsf{M}$, and we adopt the convention that the first part of this splitting will be horizontal, and the second will be vertical.

S1.3.6 The Ehresmann connection on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ associated with an affine connection on Q

If S is the geodesic spray defined by an affine connection ∇ on Q, then we may use the construction of the previous section to provide an Ehresmann connection on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$. One further verifies that, in local coordinates, a basis for HTQ is given by

$$\mathrm{hlft}\Big(\frac{\partial}{\partial q^i}\Big) = \frac{\partial}{\partial q^i} - \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})v^k \frac{\partial}{\partial v^j}, \qquad i \in \{1, \dots, n\}.$$

This defines "hlft" as the horizontal lift map for the connection we describe here. We shall introduce different notation for the horizontal lift associated with the other Ehresmann connections we define. We also denote by "vlft" the vertical lift map associated with this connection. Note that we have $S(v_q) = \mathrm{hlft}_{v_q}(v_q)$. Also note that this splitting $\mathsf{T}_{v_q}\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$ extends the splitting of Lemma 6.33 away from $Z(\mathsf{TQ})$. We depict the situation in Figure S1.1 to give the reader some intuition for what is going on.

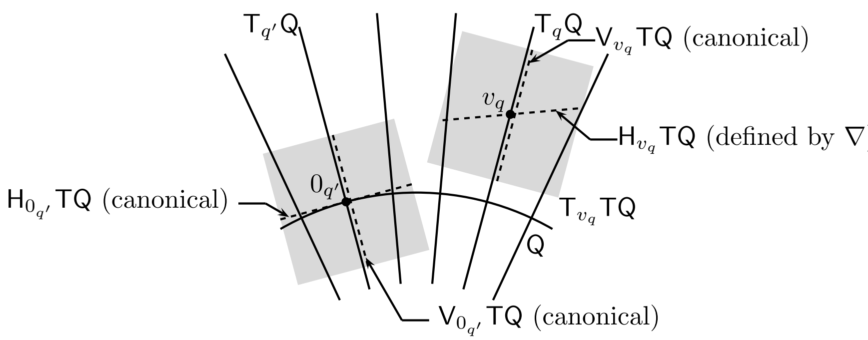

Figure S1.1. A depiction of the Ehresmann connection on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ associated with an affine connection on Q

The Ehresmann connection HTQ defines a connection form ω_{HTQ} and a curvature form Ω_{HTQ}, just as in Section S1.3.3. It will be useful to have a formula relating Ω_{HTQ} to the curvature tensor R and the torsion tensor T for ∇. As far as we are aware, this result does not appear in the literature.

Proposition S1.29 (Curvature form for Ehresmann connection associated to an affine connection). *Let ∇ be an affine connection on Q and let Ω_{HTQ} be the curvature form for the associated Ehresmann connection on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$. The following formula holds:*

$$\begin{aligned}\Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{v_q}(u_q), \mathrm{hlft}_{v_q}(w_q)) = \mathrm{vlft}_{v_q}\big(&R(u_q, w_q)v_q - \tfrac{1}{2}(\nabla_{u_q}T)(w_q, v_q)\\ &+ \tfrac{1}{2}(\nabla_{w_q}T)(u_q, v_q) - \tfrac{1}{2}T(T(u_q, w_q), v_q)\\ &+ \tfrac{1}{4}(T(T(u_q, v_q), w_q) - T(T(w_q, v_q), u_q))\big).\end{aligned}$$

In particular, if ∇ is torsion-free, then

$$\Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{v_q}(u_q), \mathrm{hlft}_{v_q}(w_q)) = \mathrm{vlft}_{v_q}\big(R(u_q, w_q)v_q\big).$$

Proof. The most straightforward, albeit tedious, proof is in coordinates. Let U and W be vector fields that extend u_q and w_q, respectively. A computation yields

$$\begin{aligned}\mathrm{ver}([\mathrm{hlft}(U), \mathrm{hlft}(W)]) = \frac{1}{2}\Big(&\frac{\partial \Gamma^i_{jk}}{\partial q^\ell} + \frac{\partial \Gamma^i_{kj}}{\partial q^\ell} - \frac{\partial \Gamma^i_{j\ell}}{\partial q^k} - \frac{\partial \Gamma^i_{\ell j}}{\partial q^k} \\ &+ \frac{1}{2}\Big(\Gamma^i_{m\ell}\Gamma^m_{kj} + \Gamma^i_{m\ell}\Gamma^m_{jk} + \Gamma^i_{\ell m}\Gamma^m_{kj} + \Gamma^i_{\ell m}\Gamma^m_{jk} \\ &- \Gamma^i_{mk}\Gamma^m_{\ell j} - \Gamma^i_{mk}\Gamma^m_{j\ell} - \Gamma^i_{km}\Gamma^m_{\ell j} - \Gamma^i_{km}\Gamma^m_{j\ell}\Big)\Big) v^j U^k W^l \frac{\partial}{\partial v^i}.\end{aligned}$$

One now employs the coordinate formulae for T and R (see Section 3.9.6), and the coordinate formula for ∇T (see Section 3.8.3) to directly verify that

$$\begin{aligned}&\mathrm{ver}([\mathrm{hlft}(X), \mathrm{hlft}(W)])(v_q) \\ &\quad = \mathrm{vlft}_{v_q}\big(R(w_q, u_q)v_q + R(w_q, v_q)u_q + R(v_q, u_q)w_q \\ &\qquad + (\nabla_{v_q}T)(w_q, u_q) + \tfrac{1}{2}T(T(u_q, v_q), w_q) + \tfrac{1}{2}T(T(v_q, w_q), u_q).\end{aligned}$$

The result may now be proved using the first Bianchi identity,

$$\begin{aligned}&\sum_{\substack{\sigma\in S_3\\ \mathrm{sgn}(\sigma)=1}} \big(R(X_{\sigma(1)}, X_{\sigma(2)})X_{\sigma(3)}\big) \\ &\qquad = \sum_{\substack{\sigma\in S_3\\ \mathrm{sgn}(\sigma)=1}} \big(T(T(X_{\sigma(1)}, X_{\sigma(2)}), X_{\sigma(3)}) + (\nabla_{X_{\sigma(1)}}T)(X_{\sigma(2)}, X_{\sigma(3)})\big),\end{aligned}$$

that is proved as Theorem 2.5 in Chapter III of [Kobayashi and Nomizu 1963, volume 1]. ■

Recall that the affine connection ∇ has associated with it an Ehresmann connection on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ that provides a natural isomorphism of $\mathsf{T}_{v_q}\mathsf{TQ}$ with $\mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$ as in Section S1.3.5. This in turn provides an isomorphism of $\mathsf{T}^*_{v_q}\mathsf{TQ}$ with $\mathsf{T}^*_q\mathsf{Q} \oplus \mathsf{T}^*_q\mathsf{Q}$. In coordinates the basis

$$\mathrm{d}q^i, \quad \mathrm{d}v^i + \frac{1}{2}(\Gamma^i_{jk} + \Gamma^i_{kj})v^k \mathrm{d}q^j, \qquad i \in \{1, \dots, n\}, \tag{S1.9}$$

is adapted to this splitting in that the first n vectors form a basis for the horizontal part of $\mathsf{T}^*_{v_q}\mathsf{TQ}$, and the second n vectors form a basis for the vertical part.

S1.3.7 The Ehresmann connection on $\pi_{\mathsf{T}^*\mathsf{Q}}\colon \mathsf{T}^*\mathsf{Q} \to \mathsf{Q}$ associated with an affine connection on Q

We can equip the cotangent bundle of Q with an Ehresmann connection induced by ∇. We do this by noting that the Ehresmann connection HTQ induced by an affine connection is a linear connection. Thus, we have an induced linear connection $\mathsf{HT}^*\mathsf{Q}$ on $\pi_{\mathsf{T}^*\mathsf{Q}}\colon \mathsf{T}^*\mathsf{Q} \to \mathsf{Q}$ whose horizontal lift is as defined by (S1.6). A coordinate basis for the $\mathsf{HT}^*\mathsf{Q}$ is given by

$$\mathrm{hlft}^*\Big(\frac{\partial}{\partial q^i}\Big) = \frac{\partial}{\partial q^i} + \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})p_j\frac{\partial}{\partial p_k}, \qquad i \in \{1,\dots,n\}.$$

Note that hlft^* denotes the horizontal lift for the Ehresmann connection on $\pi_{\mathsf{T}^*\mathsf{Q}}\colon \mathsf{T}^*\mathsf{Q} \to \mathsf{Q}$. The vertical subbundle has the basis

$$\mathrm{vlft}^*\big(\mathrm{d}q^i\big) = \frac{\partial}{\partial p_i}, \qquad i \in \{1,\dots,n\}$$

that, we remark, defines the vertical lift map $\mathrm{vlft}^*\colon \mathsf{T}^*_q\mathsf{Q} \to \mathsf{V}_{\alpha_q}\mathsf{T}^*\mathsf{Q}$.

S1.3.8 The Ehresmann connection on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$ associated with an affine connection on Q

In this section we induce a connection on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$ using a second-order vector field as described in Section S1.3.5. We do not yet *have* a second-order vector field on TTQ, so let us set about producing one from the geodesic spray S. Note that S^T is a vector field on TTQ, and we may determine its coordinate expression in natural coordinates $((\boldsymbol{x},\boldsymbol{v}),(\boldsymbol{u},\boldsymbol{w}))$ to be

$$\begin{aligned} S^T = v^i\frac{\partial}{\partial q^i} - \Gamma^i_{jk}v^jv^k\frac{\partial}{\partial v^i} + w^i\frac{\partial}{\partial u^i}& \\ - \Big(\frac{\partial\Gamma^i_{jk}}{\partial q^\ell}v^jv^ku^\ell + \Gamma^i_{jk}w^jv^k + \Gamma^i_{kj}w^jv^k\Big)\frac{\partial}{\partial w^i}&. \end{aligned} \tag{S1.10}$$

This vector field is "almost" second-order. To make it second-order, we use the canonical involution $I_{\mathsf{Q}}\colon \mathsf{TTQ} \to \mathsf{TTQ}$ described in Section S1.2.5 and the following lemma.

Lemma S1.30. *$I^*_{\mathsf{Q}}S^T$ is a second-order vector field on TTQ.*

Proof. A simple computation, using the coordinate expressions for I_{Q} and S^T, gives

$$\begin{aligned} I^*_{\mathsf{Q}}S^T = u^i\frac{\partial}{\partial q^i} + w^i\frac{\partial}{\partial v^i} - \Gamma^i_{jk}u^ju^k\frac{\partial}{\partial u^i}& \\ - \Big(\frac{\partial\Gamma^i_{jk}}{\partial q^\ell}v^\ell u^ju^k + \Gamma^i_{jk}u^kw^j + \Gamma^i_{kj}u^kw^j\Big)\frac{\partial}{\partial w^i}& \end{aligned} \tag{S1.11}$$

which verifies the lemma. ■

Now we may use the procedure of Section S1.3.5 to produce a connection on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$. Let us denote this connection by $\mathsf{H}(\mathsf{TTQ})$. Note that this connection provides a splitting

$$\mathsf{T}_{X_{v_q}}\mathsf{TTQ} \simeq \mathsf{T}_{v_q}\mathsf{TQ} \oplus \mathsf{T}_{v_q}\mathsf{TQ} \tag{S1.12}$$

for $X_{v_q} \in \mathsf{T}_{v_q}\mathsf{TQ}$. Also, the Ehresmann connection HTQ on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ described in Section S1.3.6 gives the splitting $\mathsf{T}_{v_q}\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$. Therefore, we have a resulting splitting

$$\mathsf{T}_{X_{v_q}}\mathsf{TTQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}. \tag{S1.13}$$

In this splitting, the first two components are the horizontal subspace and the second two components are the vertical subspace. Within each pair, the first part is horizontal and the second is vertical.

Let us now write a basis of vector fields on TTQ that is adapted to the splitting (S1.13). To obtain a coordinate expression for a basis of the Ehresmann connection on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$, we use the coordinate expression (S1.11) for $I_{\mathsf{Q}}^*S^T$. We write a basis that is adapted to the splitting of $\mathsf{T}_{v_q}\mathsf{TQ}$. The resulting basis vectors for $\mathsf{H}(\mathsf{TTQ})$ are

$$\begin{aligned}
\operatorname{hlft}^T\Big(\frac{\partial}{\partial q^i} - \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})v^k\frac{\partial}{\partial v^j}\Big) &= \frac{\partial}{\partial q^i} - \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})v^k\frac{\partial}{\partial v^j} \\
&\quad - \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})u^k\frac{\partial}{\partial u^j} - \frac{1}{2}\Big(\frac{\partial \Gamma^j_{i\ell}}{\partial q^k}u^\ell v^k + \frac{\partial \Gamma^j_{\ell i}}{\partial q^k}u^\ell v^k + (\Gamma^j_{ik} + \Gamma^j_{ki})w^k \\
&\quad - \frac{1}{2}(\Gamma^k_{i\ell} + \Gamma^k_{\ell i})(\Gamma^j_{km} + \Gamma^j_{mk})u^m v^\ell\Big)\frac{\partial}{\partial w^j}, \qquad i \in \{1,\dots,n\}, \\
\operatorname{hlft}^T\Big(\frac{\partial}{\partial v^i}\Big) &= \frac{\partial}{\partial v^i} - \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})u^k\frac{\partial}{\partial w^j}, \qquad i \in \{1,\dots,n\},
\end{aligned}$$

with the first n basis vectors forming a basis for the horizontal part of $\mathsf{H}_{X_{v_q}}(\mathsf{TTQ})$, and the second n vectors forming a basis for the vertical part of $\mathsf{H}_{X_{v_q}}(\mathsf{TTQ})$, with respect to the splitting $\mathsf{H}_{X_{v_q}}(\mathsf{TTQ}) \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$. Note that we use the notation hlft^T to refer to the horizontal lift for the connection on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$. We also denote by vlft^T the vertical lift on this vector bundle.

We may easily derive a basis for the vertical subbundle of $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$ that adapts to the splitting of $\mathsf{T}_{v_q}\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$. We may verify that the vector fields

$$\begin{aligned}
&\operatorname{vlft}^T\Big(\frac{\partial}{\partial q^i} - \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})v^k\frac{\partial}{\partial v^j}\Big) \\
&\qquad\qquad = \frac{\partial}{\partial u^i} - \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})v^k\frac{\partial}{\partial w^j}, \qquad i \in \{1,\dots,n\}, \\
&\operatorname{vlft}^T\Big(\frac{\partial}{\partial v^i}\Big) = \frac{\partial}{\partial w^i}, \qquad i \in \{1,\dots,n\},
\end{aligned}$$

have the property that the first n vectors span the horizontal part of $\mathsf{V}_{X_{v_q}}\mathsf{TTQ}$, and the second n span the vertical part of $\mathsf{V}_{X_{v_q}}\mathsf{TTQ}$.

Remarks S1.31. 1. The construction in this section may, in fact, be made with an arbitrary second-order vector field. That is, if S is a second-order vector field on TQ, then $I^*_{\mathsf{Q}}S^T$ is a second-order vector field on TTQ. This second-order vector field then induces a connection on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$. Clearly then, this construction can be iterated, and so provides a connection on the kth tangent bundle $\pi_{\mathsf{T}(\mathsf{T}^{k-1}\mathsf{Q})}\colon \mathsf{T}^k\mathsf{Q} \to \mathsf{T}^{k-1}\mathsf{Q}$ for each $k \geq 0$. This then provides an isomorphism of $\mathsf{T}_X\mathsf{T}^{k-1}\mathsf{Q}$ with the direct sum $\mathsf{T}_q\mathsf{Q} \oplus \cdots \oplus \mathsf{T}_q\mathsf{Q}$ of $2k$ copies of $\mathsf{T}_q\mathsf{Q}$ where $X \in \mathsf{T}^{k-1}\mathsf{Q}$ and $q = \pi_{\mathsf{TQ}} \circ \cdots \circ \pi_{\mathsf{TT}^{k-1}\mathsf{Q}}(X)$.

2. The vector field $I^*_{\mathsf{Q}}S^T$ is the geodesic spray of an affine connection on TQ. Thus we see how, given an affine connection on a manifold Q, it is possible to derive an affine connection on the kth tangent bundle $\mathsf{T}^k\mathsf{Q}$ for $k \geq 1$.
3. The vector field $I^*_{\mathsf{Q}}S^T$ is the geodesic spray of the affine connection on TQ that Yano and Ishihara [1973] call the "complete lift" of the affine connection ∇ on Q. Given that we are calling X^T the tangent lift of a vector field, let us also call ∇^T the ***tangent lift*** of ∇. This affine connection is defined as the unique affine connection ∇^T on TQ that satisfies $\nabla^T_{X^T}Y^T = \big(\nabla_X Y\big)^T$ for vector fields X and Y on Q. Yano and Ishihara also show that the geodesics of ∇^T are vector fields along geodesics of ∇. They further claim that these vector fields along geodesics are in fact Jacobi fields. This is indeed true as our results of Section S1.3.10 show, but the proof in Yano and Ishihara cannot be generally correct as, for example, the Jacobi equation they use lacks the torsion term that the actual Jacobi equation possesses. We shall have more to say about this in Remark S1.35. •

S1.3.9 The Ehresmann connection on $\pi_{\mathsf{T}^*\mathsf{TQ}}\colon \mathsf{T}^*\mathsf{TQ} \to \mathsf{TQ}$ associated with an affine connection on Q

As the tangent bundle of TQ comes equipped with an Ehresmann connection, so too does its cotangent bundle. To construct this connection, we note that the connection $\mathsf{H}(\mathsf{TTQ})$ on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$ is a linear connection, and thus there is a connection naturally induced on $\pi_{\mathsf{T}^*\mathsf{TQ}}\colon \mathsf{T}^*\mathsf{TQ} \to \mathsf{TQ}$ whose horizontal lift is as defined by (S1.6). We denote this connection by $\mathsf{H}(\mathsf{T}^*\mathsf{TQ})$, and note that it provides a splitting

$$\mathsf{T}_{\Lambda_{v_q}}\mathsf{T}^*\mathsf{TQ} \simeq \mathsf{T}_{v_q}\mathsf{TQ} \oplus \mathsf{T}^*_{v_q}\mathsf{TQ}$$

for $\Lambda_{v_q} \in \mathsf{T}^*\mathsf{TQ}$. In turn, the connection of Section S1.3.6 on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ gives a splitting $\mathsf{T}_{v_q}\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$, and so also a splitting $\mathsf{T}^*_{v_q}\mathsf{TQ} \simeq \mathsf{T}^*_q\mathsf{Q} \oplus \mathsf{T}^*_q\mathsf{Q}$. This then provides the splitting

$$\mathsf{T}_{\Lambda_{\alpha_q}}\mathsf{T}^*\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}^*_q\mathsf{Q} \oplus \mathsf{T}^*_q\mathsf{Q}.$$

The first two components of this splitting are the horizontal part of the subspace, and the second two are the vertical part. For each pair, the first component is horizontal and the second is vertical.

Let us now write a basis for vector fields on $\mathsf{T}^*\mathsf{TQ}$ that is adapted to the splitting we have just demonstrated. We use natural coordinates $((\boldsymbol{q},\boldsymbol{v}),(\boldsymbol{\alpha},\boldsymbol{\beta}))$. First we determine that a basis for the horizontal subbundle is

$$
\begin{aligned}
\operatorname{hlft}^{T^*}\Big(\frac{\partial}{\partial q^i}-\frac{1}{2}(\Gamma^j_{ik}+\Gamma^j_{ki})v^k\frac{\partial}{\partial v^j}\Big) &= \frac{\partial}{\partial q^i}-\frac{1}{2}(\Gamma^j_{ik}+\Gamma^j_{ki})v^m\frac{\partial}{\partial v^j}\\
&\quad+\frac{1}{2}\Big(\frac{\partial\Gamma^j_{ik}}{\partial q^\ell}v^\ell\beta_j+\frac{\partial\Gamma^j_{ki}}{\partial q^\ell}v^\ell\beta_j+(\Gamma^j_{ik}+\Gamma^j_{ki})\alpha_j\\
&\quad-\frac{1}{2}(\Gamma^\ell_{im}+\Gamma^\ell_{mi})(\Gamma^j_{\ell k}+\Gamma^j_{k\ell})v^m\beta_j\Big)\frac{\partial}{\partial\alpha_k}\\
&\quad+\frac{1}{2}(\Gamma^j_{ik}+\Gamma^j_{ki})\beta_j\frac{\partial}{\partial\beta_k},\qquad i\in\{1,\dots,n\},\\
\operatorname{hlft}^{T^*}\Big(\frac{\partial}{\partial v^i}\Big) &= \frac{\partial}{\partial v^i}+\frac{1}{2}(\Gamma^j_{ik}+\Gamma^j_{ki})\beta_j\frac{\partial}{\partial\alpha_k},\qquad i\in\{1,\dots,n\}.
\end{aligned}
$$

The first n basis vectors span the horizontal part of $\mathsf{H}_{\Lambda_{v_q}}\mathsf{T}^*\mathsf{TQ}$, and the second n vectors span the vertical part. We have introduced the notation $\operatorname{hlft}^{T^*}$ to refer to the horizontal lift map on the bundle $\pi_{\mathsf{T}^*\mathsf{TQ}}\colon \mathsf{T}^*\mathsf{TQ}\to\mathsf{TQ}$, and we shall denote the vertical lift map by $\operatorname{vlft}^{T^*}$.

One may also write a basis for $\mathsf{V}(\mathsf{T}^*\mathsf{TQ})=\ker(T\pi_{\mathsf{T}^*\mathsf{TQ}})$ that is adapted to the splitting of $\mathsf{T}^*_{v_q}\mathsf{TQ}$. We use (S1.9) to provide a basis

$$
\begin{aligned}
&\operatorname{vlft}^{T^*}(\mathrm{d}q^i)=\frac{\partial}{\partial\alpha_i},\qquad i\in\{1,\dots,n\},\\
&\operatorname{vlft}^{T^*}\Big(\mathrm{d}v^i+\frac{1}{2}(\Gamma^i_{jk}+\Gamma^i_{kj})v^k\mathrm{d}q^j\Big)\\
&\qquad\qquad=\frac{\partial}{\partial\beta_i}+\frac{1}{2}(\Gamma^i_{jk}+\Gamma^i_{kj})v^k\frac{\partial}{\partial\alpha_j},\qquad i\in\{1,\dots,n\},
\end{aligned}
$$

in which the first n vectors are a basis for the horizontal part of $\mathsf{V}_{\Lambda_{v_q}}\mathsf{T}^*\mathsf{TQ}$, and the second n are a basis for the vertical part.

S1.3.10 Representations of S^T and S^{T^*}

In the previous sections we have provided local bases for the various connections we have constructed. With these local bases in hand, and with the coordinate expressions (S1.10) and (S1.17) (see below) for S^T and S^{T^*}, it is a simple matter to determine the form of $S^T(X_{v_q})$ and $S^{T^*}(\Lambda_{v_q})$ in these splittings, where $X_{v_q}\in\mathsf{TTQ}$ and $\Lambda_{v_q}\in\mathsf{T}^*\mathsf{TQ}$. Let us merely record the results of these somewhat tedious computations.

First we look at S^T. In this case, recall that the connection $\mathsf{H}(\mathsf{TTQ})$ on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$ and the connection HTQ on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ combine to give a splitting

$$\mathsf{T}_{X_{v_q}}\mathsf{TTQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q},$$

where $X_{v_q} \in \mathsf{T}_{v_q}\mathsf{TQ}$. Here we maintain our convention that the first two components refer to the horizontal component for a connection $\mathsf{H}(\mathsf{TTQ})$ on $\pi_{\mathsf{TTQ}}\colon \mathsf{TTQ} \to \mathsf{TQ}$, and the second two components refer to the vertical component. Using the splitting (S1.12) let us write $X_{v_q} \in \mathsf{T}_{v_q}\mathsf{TQ}$ as $u_{v_q} \oplus w_{v_q}$ for some $u_{v_q}, w_{v_q} \in \mathsf{T}_q\mathsf{Q}$. Note that we depart from our usual notation of writing tangent vectors in $\mathsf{T}_q\mathsf{Q}$ with a subscript of q, instead using the subscript v_q. This abuse of notation is necessary (and convenient) to reflect the fact that these vectors depend on where we are in TQ, and not just in Q. A computation verifies the following result, where Ω_{HTQ} is the curvature form for the connection HTQ.

Proposition S1.32 (Representation of tangent lift of geodesic spray). *The following formula holds:*

$$S^T(u_{v_q} \oplus w_{v_q}) = v_q \oplus 0 \oplus w_{v_q} \oplus \big(-\Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{v_q}(u_{v_q}), \mathrm{hlft}_{v_q}(v_q))\big).$$

In writing this formula, we are regarding Ω_{HTQ} as taking values in $\mathsf{V}_{v_q}\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q}$.

We may use this representation of S^T to obtain a refined relationship between solutions of the Jacobi equation and integral curves of S^T. To do so, we first prove a simple lemma. We state a more general form of this lemma than we shall immediately use, but the extra generality will be useful in Supplement 4.

Lemma S1.33. *Let Y be a* LIC$^\infty$*-vector field on Q, suppose that $\gamma\colon I \to \mathsf{Q}$ is the* LAD *curve satisfying $\nabla_{\gamma'(t)}\gamma'(t) = Y(t, \gamma(t))$, and denote by $\Upsilon\colon I \to \mathsf{TQ}$ the tangent vector field of γ (i.e., $\Upsilon = \gamma'$). Let $X\colon I \to \mathsf{TTQ}$ be an* LAC *vector field along Υ, and denote $X(t) = X_1(t) \oplus X_2(t) \in \mathsf{T}_{\gamma(t)}\mathsf{Q} \oplus \mathsf{T}_{\gamma(t)}\mathsf{Q} \simeq \mathsf{T}_{\Upsilon(t)}\mathsf{TQ}$. Then the tangent vector field to the curve $t \mapsto X(t)$ is given by $\gamma'(t) \oplus Y(t, \gamma(t)) \oplus \tilde{X}_1(t) \oplus \tilde{X}_2(t)$, where*

$$\begin{aligned}\tilde{X}_1(t) &= \nabla_{\gamma'(t)}X_1(t) + \tfrac{1}{2}T(X_1(t), \gamma'(t)),\\ \tilde{X}_2(t) &= \nabla_{\gamma'(t)}X_2(t) + \tfrac{1}{2}T(X_2(t), \gamma'(t)).\end{aligned}$$

Proof. In coordinates, the curve $t \mapsto X(t)$ has the form

$$(q^i(t), \dot{q}^j(t), X_1^k(t), X_2^\ell(t) - \tfrac{1}{2}(\Gamma^\ell_{mr} + \Gamma^\ell_{rm})\dot{q}^m(t)X_1^r(t)).$$

The tangent vector to this curve is then given a.e. by

$$\begin{aligned}\dot{q}^i\frac{\partial}{\partial q^i} + (Y^i - \Gamma^i_{jk}\dot{q}^j\dot{q}^k)\frac{\partial}{\partial v^i} + \dot{X}_1^i\frac{\partial}{\partial u^i} + \Big(\dot{X}_2^i - \frac{1}{2}\frac{\partial\Gamma^i_{jk}}{\partial q^\ell}\dot{q}^k\dot{q}^\ell X_1^j - \frac{1}{2}\frac{\partial\Gamma^i_{kj}}{\partial q^\ell}\dot{q}^k\dot{q}^\ell X_1^j\\ - \frac{1}{2}(\Gamma^i_{jk} + \Gamma^i_{kj})(Y^k - \Gamma^k_{\ell m}\dot{q}^\ell\dot{q}^m)X_1^j - \frac{1}{2}(\Gamma^i_{jk} + \Gamma^i_{kj})\dot{q}^k\dot{X}_1^j\Big)\frac{\partial}{\partial w^i}.\end{aligned}$$

A straightforward computation shows that this tangent vector field has the representation

$$\gamma'(t) \oplus Y(t,\gamma(t)) \oplus \big(\nabla_{\gamma'(t)} X_1(t) + \tfrac{1}{2}T(X_1(t),\gamma'(t))\big) \\ \oplus \big(\nabla_{\gamma'(t)} X_2(t) + \tfrac{1}{2}T(X_2(t),\gamma'(t))\big),$$

which proves the lemma. ■

We may now prove our main result that relates the integral curves of S^T with solutions to the Jacobi equation.

Theorem S1.34 (Relationship between tangent lift of geodesic spray and Jacobi equation). *Let ∇ be an affine connection on Q with S the corresponding geodesic spray. Let $\gamma\colon I \to \mathsf{Q}$ be a geodesic with $t \mapsto \Upsilon(t) \triangleq \gamma'(t)$ the corresponding integral curve of S. Let $a \in I$, $u, w \in \mathsf{T}_{\gamma(a)}\mathsf{Q}$, and define vector fields $U, W\colon I \to \mathsf{TQ}$ along γ by asking that $t \mapsto U(t) \oplus W(t) \in \mathsf{T}_{\gamma(t)}\mathsf{Q} \oplus \mathsf{T}_{\gamma(t)}\mathsf{Q} \simeq \mathsf{T}_{\Upsilon(t)}\mathsf{TQ}$ be the integral curve of S^T with initial conditions $u \oplus w \in \mathsf{T}_{\gamma(a)}\mathsf{Q} \oplus \mathsf{T}_{\gamma(a)}\mathsf{Q} \simeq \mathsf{T}_{\Upsilon(a)}\mathsf{TQ}$. Then U and W have the following properties:*

(i) U satisfies the Jacobi equation

$$\nabla^2_{\gamma'(t)} U(t) + R(U(t),\gamma'(t))\gamma'(t) + \nabla_{\gamma'(t)}(T(U(t),\gamma'(t))) = 0;$$

(ii) $W(t) = \nabla_{\gamma'(t)} U(t) + \frac{1}{2}T(U(t),\gamma'(t))$.

Proof. Throughout the proof, we represent points in TTQ as the direct sum of tangent vectors to Q using the connection on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ induced by ∇. The tangent vector field to the curve $t \mapsto U(t) \oplus W(t)$ at t must equal $S^T(U(t) \oplus W(t))$. By Lemma S1.33 and Proposition S1.32, this means that

$$\begin{aligned} \nabla_{\gamma'(t)} U(t) &= W(t) - \tfrac{1}{2}T(U(t),\gamma'(t)), \\ \nabla_{\gamma'(t)} W(t) &= -\Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{\gamma'(t)}(U(t)), \mathrm{hlft}_{\gamma'(t)}(\gamma'(t))) - \tfrac{1}{2}T(W(t),\gamma'(t)). \end{aligned} \tag{S1.14}$$

The first of these equations proves (ii). To prove (i), we covariantly differentiate the first of equations (S1.14). This yields, using the second of equations (S1.14),

$$\nabla^2_{\gamma'(t)} U(t) = -\Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{\gamma'(t)}(U(t)), \mathrm{hlft}_{\gamma'(t)}(\gamma'(t))) \\ - \tfrac{1}{2}T(W(t),\gamma'(t)) - \tfrac{1}{2}\nabla_{\gamma'(t)}(T(U(t),\gamma'(t))). \tag{S1.15}$$

Now we see from Proposition S1.29 that

$$-\Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{\gamma'(t)}(U(t)), \mathrm{hlft}_{\gamma'(t)}(\gamma'(t))) = -R(U(t),\gamma'(t))\gamma'(t) \\ - \tfrac{1}{2}(\nabla_{\gamma'(t)} T)(U(t),\gamma'(t)) + \tfrac{1}{4}T(T(U(t),\gamma'(t)),\gamma'(t)). \tag{S1.16}$$

Combining (S1.15), (S1.16), and the first of equations (S1.14) shows that U satisfies the Jacobi equation. ■

Remark S1.35 (Comments on the tangent lift of an affine connection). Let us follow up on Remark S1.31–3 by showing that geodesics of ∇^T are indeed Jacobi fields. By Lemma S1.30, integral curves of S^T and of the geodesic spray for ∇^T are mapped to one another by the involution I_{Q}. Given

1. the representation $t \mapsto U(t) \oplus W(t)$ of integral curves of S^T as in Theorem S1.34,
2. the Ehresmann connection on $\pi_{\mathsf{TQ}}\colon \mathsf{TQ} \to \mathsf{Q}$ described in Section S1.3.6, and
3. the coordinate expression for I_{Q},

one verifies that the geodesics of ∇^T are exactly the vector fields $t \mapsto U(t)$ along geodesics as described in Theorem S1.34. But these are simply Jacobi fields according to the theorem. •

Now let us look at similar results relating to S^{T^*}. First we give the coordinate formula for this vector field in natural coordinates $((\boldsymbol{x},\boldsymbol{v}),(\boldsymbol{\alpha},\boldsymbol{\beta}))$ for $\mathsf{T}^*\mathsf{TQ}$:

$$S^{T^*} = v^i \frac{\partial}{\partial q^i} - \Gamma^i_{jk} v^j v^k \frac{\partial}{\partial v^i} + \frac{\partial \Gamma^\ell_{jk}}{\partial q^i} v^j v^k \beta_\ell \frac{\partial}{\partial \alpha_i} - \big(\alpha_i - \Gamma^\ell_{ij} v^j \beta_\ell - \Gamma^\ell_{ji} v^j \beta_\ell\big) \frac{\partial}{\partial \beta_i}. \tag{S1.17}$$

To provide the decomposition for S^{T^*}, we need an extra bit of notation. Fix $v_q \in \mathsf{T}_q\mathsf{Q}$ and note that, for $u_q \in \mathsf{T}_q\mathsf{Q}$, we have $\Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{v_q}(u_q), \mathrm{hlft}_{v_q}(v_q)) \in \mathsf{V}_{v_q}\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q}$. Thus we may regard $u_q \mapsto \Omega_{\mathsf{HTQ}}(\mathrm{hlft}_{v_q}(u_q), \mathrm{hlft}_{v_q}(v_q))$ as a linear map on $\mathsf{T}_q\mathsf{Q}$. Let us denote the dual linear map by $\beta_q \mapsto \Omega^*_{\mathsf{HTQ}}(\mathrm{vlft}_{v_q}(\beta_q), \mathrm{hlft}_{v_q}(v_q))$. The reason for this odd choice of notation for a dual linear map will become clear shortly.

We need more notation concerning the curvature and torsion tensors R and T. For $u_q, v_q \in \mathsf{T}_q\mathsf{Q}$ and $\alpha_q \in \mathsf{T}^*_q\mathsf{Q}$, define $R^*(\alpha_q, u_q)v_q \in \mathsf{T}^*_q\mathsf{Q}$ by

$$\langle R^*(\alpha_q, u_q)v_q; w_q\rangle = \langle \alpha_q; R(w_q, u_q)v_q\rangle, \qquad w_q \in \mathsf{T}_q\mathsf{Q},$$

and similarly define $T^*(\alpha_q, u_q) \in \mathsf{T}^*_q\mathsf{Q}$ by

$$\langle T^*(\alpha_q, u_q); w_q\rangle = \langle \alpha_q; T(w_q, u_q)\rangle, \qquad w_q \in \mathsf{T}_q\mathsf{Q}.$$

With these tensors defined, we say that a covector field $\alpha\colon I \to \mathsf{T}^*\mathsf{Q}$ along a geodesic $\gamma\colon I \to \mathsf{Q}$ of ∇ is a solution of the ***adjoint Jacobi equation*** if

$$\nabla^2_{\gamma'(t)}\alpha(t) + R^*(\alpha(t), \gamma'(t))\gamma'(t) - T^*(\nabla_{\gamma'(t)}\alpha(t), \gamma'(t)) = 0,$$

for $t \in I$.

Now let us recall the splittings associated with the connection $\mathsf{H}(\mathsf{T}^*\mathsf{TQ})$ on $\pi_{\mathsf{T}^*\mathsf{TQ}}\colon \mathsf{T}^*\mathsf{TQ} \to \mathsf{TQ}$ that is described in Section S1.3.9. For $\Lambda_{v_q} \in \mathsf{T}^*_{v_q}\mathsf{TQ}$ we have

$$\mathsf{T}_{\Lambda_{v_q}}\mathsf{T}^*\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q^*\mathsf{Q} \oplus \mathsf{T}_q^*\mathsf{Q}.$$

We then write $\Lambda_{v_q} \in \mathsf{T}^*_{v_q}\mathsf{TQ}$ as $\alpha_{v_q} \oplus \beta_{v_q}$ for some $\alpha_{v_q}, \beta_{v_q} \in \mathsf{T}_q^*\mathsf{Q}$, where we again make an abuse of notation. This then gives the following formula for S^{T^*} with respect to our splitting.

Proposition S1.36 (Representation of cotangent lift of geodesic spray). *The following formula holds:*

$$S^{T^*}(\alpha_{v_q} \oplus \beta_{v_q}) = v_q \oplus 0 \oplus \left(\Omega^*_{\mathsf{HTQ}}(\mathrm{vlft}_{v_q}(\beta_{v_q}), \mathrm{hlft}_{v_q}(v_q))\right) \oplus (-\alpha_{v_q}).$$

To demonstrate the relationship between integral curves of S^{T^*} and solutions to the adjoint Jacobi equation, we have the following analogue to Lemma S1.33.

Lemma S1.37. *Let Y be a* LIC^∞*-vector field on Q, suppose that $\gamma\colon I \to \mathsf{Q}$ is the* LAD *curve satisfying $\nabla_{\gamma'(t)}\gamma'(t) = Y(t, \gamma(t))$, and denote by $\Upsilon\colon I \to \mathsf{TQ}$ the tangent vector field of γ (i.e., $\Upsilon = \gamma'$). Let $\Lambda\colon I \to \mathsf{T}^*\mathsf{TQ}$ be an* LAC *covector field along Υ, and denote $\Lambda(t) = \Lambda^1(t) \oplus \Lambda^2(t) \in \mathsf{T}^*_{\gamma(t)}\mathsf{Q} \oplus \mathsf{T}^*_{\gamma(t)}\mathsf{Q} \simeq \mathsf{T}^*_{\Upsilon(t)}\mathsf{TQ}$. Then the tangent vector field to the curve $t \mapsto \Lambda(t)$ is given by $\gamma'(t) \oplus Y(t, \gamma(t)) \oplus \tilde{\Lambda}^1(t) \oplus \tilde{\Lambda}^2(t)$, where*

$$\begin{aligned}\tilde{\Lambda}^1(t) &= \nabla_{\gamma'(t)}\Lambda^1(t) - \tfrac{1}{2}T^*(\Lambda^1(t), \gamma'(t)),\\ \tilde{\Lambda}^2(t) &= \nabla_{\gamma'(t)}\Lambda^2(t) - \tfrac{1}{2}T^*(\Lambda^2(t), \gamma'(t)).\end{aligned}$$

Proof. In coordinates the curve $t \mapsto \Lambda(t)$ has the form

$$(q^i(t), \dot{q}^j(t), \Lambda^1_k(t) + \tfrac{1}{2}(\Gamma^m_{kr} + \Gamma^m_{rk})\dot{q}^r(t)\Lambda^2_m(t), \Lambda^2_i(t)).$$

The tangent vector to this curve is then given by

$$\begin{aligned}\dot{q}^i\frac{\partial}{\partial q^i} &+ (Y^i - \Gamma^i_{jk}\dot{q}^j\dot{q}^k)\frac{\partial}{\partial v^i} + \Big(\dot{\Lambda}^1_i + \frac{1}{2}\frac{\partial\Gamma^j_{ik}}{\partial q^\ell}\dot{q}^\ell\dot{q}^k\Lambda^2_j + \frac{1}{2}\frac{\partial\Gamma^j_{ki}}{\partial q^\ell}\dot{q}^\ell\dot{q}^k\Lambda^2_j\\ &+ \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})(Y^k - \Gamma^k_{\ell m}\dot{q}^\ell\dot{q}^m)\Lambda^2_j + \frac{1}{2}(\Gamma^j_{ik} + \Gamma^j_{ki})\dot{q}^k\dot{\Lambda}^2_j\Big)\frac{\partial}{\partial\alpha_i} + \dot{\Lambda}^2_i\frac{\partial}{\partial\beta^i}.\end{aligned}$$

A straightforward computation shows that this tangent vector field has the representation

$$\begin{aligned}\gamma'(t) \oplus Y(t, \gamma(t)) \oplus \left(\nabla_{\gamma'(t)}\Lambda^1(t) - \tfrac{1}{2}T^*(\Lambda^1(t), \gamma'(t))\right)&\\ \oplus \left(\nabla_{\gamma'(t)}\Lambda^2(t) - \tfrac{1}{2}T^*(\Lambda^2(t), \gamma'(t))\right)&,\end{aligned}$$

which proves the lemma. ■

We may now prove our main result that relates the integral curves of S^{T^*} with solutions to the adjoint Jacobi equation.

Theorem S1.38 (Relationship between cotangent lift of geodesic spray and adjoint Jacobi equation). *Let ∇ be an affine connection on Q with S the corresponding geodesic spray. Let $\gamma\colon I \to \mathsf{Q}$ be a geodesic with $t \mapsto \Upsilon(t) \triangleq \gamma'(t)$ the corresponding integral curve of S. Let $a \in I$, let $\theta, \lambda \in \mathsf{T}^*_{\gamma(a)}\mathsf{Q}$, and define covector fields $\Theta, \Lambda\colon I \to \mathsf{T}^*\mathsf{Q}$ along γ by asking that $t \mapsto \Theta(t) \oplus \Lambda(t) \in \mathsf{T}^*_{\gamma(t)}\mathsf{Q} \oplus \mathsf{T}^*_{\gamma(t)}\mathsf{Q} \simeq \mathsf{T}^*_{\Upsilon(t)}\mathsf{TQ}$ be the integral curve of S^{T^*} with initial conditions $\theta \oplus \lambda \in \mathsf{T}^*_{\gamma(a)}\mathsf{Q} \oplus \mathsf{T}^*_{\gamma(a)}\mathsf{Q} \simeq \mathsf{T}^*_{\Upsilon(a)}\mathsf{TQ}$. Then Θ and Λ have the following properties:*

(i) Λ *satisfies the adjoint Jacobi equation*

$$\nabla^2_{\gamma'(t)}\Lambda(t) + R^*(\Lambda(t), \gamma'(t))\gamma'(t) - T^*(\nabla_{\gamma'(t)}\Lambda(t), \gamma'(t)) = 0;$$

(ii) $\Theta(t) = -\nabla_{\gamma'(t)}\Lambda(t) + \frac{1}{2}T^*(\Lambda(t), \gamma'(t))$.

Proof. Throughout the proof, we represent points in $\mathsf{T}^*\mathsf{TQ}$ as the direct sum of cotangent vectors to Q using the connection on $\pi_{\mathsf{T}^*\mathsf{Q}}\colon \mathsf{T}^*\mathsf{Q} \to \mathsf{Q}$ induced by ∇. The tangent vector field to the curve $t \mapsto \Theta(t) \oplus \Lambda(t)$ at t must equal $S^{T^*}(\Theta(t) \oplus \Lambda(t))$. By Lemma S1.37 and Proposition S1.36, this means that

$$\begin{aligned} \nabla_{\gamma'(t)}\Theta(t) &= \Omega^*_{\mathsf{HTQ}}(\mathrm{vlft}_{\gamma'(t)}(\Lambda(t)), \mathrm{hlft}_{\gamma'(t)}(\gamma'(t))) + \tfrac{1}{2}T^*(\Theta(t), \gamma'(t)), \\ \nabla_{\gamma'(t)}\Lambda(t) &= -\Theta(t) + \tfrac{1}{2}T^*(\Lambda(t), \gamma'(t)). \end{aligned} \tag{S1.18}$$

The second of these equations proves (i). To prove (ii) we covariantly differentiate the second of equations (S1.18). This yields, using the first of equations (S1.18),

$$\begin{aligned} \nabla^2_{\gamma'(t)}\Lambda(t) = -\Omega^*_{\mathsf{HTQ}}&(\mathrm{hlft}_{\gamma'(t)}(\Lambda(t)), \mathrm{hlft}_{\gamma'(t)}(\gamma'(t))) - \tfrac{1}{2}T^*(\Theta(t), \gamma'(t)) \\ &+ \tfrac{1}{2}(\nabla_{\gamma'(t)}T^*)(\Lambda(t), \gamma'(t))) + \tfrac{1}{2}T^*(\nabla_{\gamma'(t)}\Lambda(t), \gamma'(t)). \end{aligned} \tag{S1.19}$$

Now we see from Proposition S1.29 that

$$\begin{aligned} \Omega^*_{\mathsf{HTQ}}(\mathrm{vlft}_{\gamma'(t)}&(\Lambda(t)), \mathrm{hlft}_{\gamma'(t)}(\gamma'(t))) = R^*(\Lambda(t), \gamma'(t))\gamma'(t) \\ &+ \tfrac{1}{2}(\nabla_{\gamma'(t)}T^*)(\Lambda(t), \gamma'(t)) - \tfrac{1}{4}T^*(T^*(\Lambda(t), \gamma'(t)), \gamma'(t)). \end{aligned} \tag{S1.20}$$

Combining (S1.19), (S1.20), and the second of equations (S1.18) shows that Λ satisfies the adjoint Jacobi equation. ■

Remarks S1.39. 1. Note that the adjoint Jacobi equation, along with the geodesic equations themselves, of course, contains the non-trivial dynamics of the Hamiltonian vector field S^{T^*}. Note also that the Hamiltonian in our splitting of $\mathsf{T}^*\mathsf{TQ}$ is simply given by $\alpha_{v_q} \oplus \beta_{v_q} \mapsto \alpha_{v_q} \cdot v_q$. Thus, while the Hamiltonian assumes a simple form in this splitting, evidently the symplectic form becomes rather complicated. However, since the Maximum Principle employs the Hamiltonian in its statement, the simple form of the Hamiltonian will be very useful for us.

2. We may express the content of Proposition S1.13, in the case when the vector field in question is the geodesic spray, as follows. We use the notation of Propositions S1.32 and S1.36. Let $u_{v_q} \oplus w_{v_q} \in \mathsf{T}_{v_q}\mathsf{T}\mathsf{Q}$ and $\alpha_{v_q} \oplus \beta_{v_q} \in \mathsf{T}^*_{v_q}\mathsf{T}\mathsf{Q}$. Let $\mathrm{ver}(S^T(u_{v_q} \oplus w_{v_q}))$ be the vertical part of $S^T(u_{v_q} \oplus w_{v_q})$ that we think of as a vector in $\mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$. In a similar manner, we think of $\mathrm{ver}(S^{T^*}(\alpha_{v_q} \oplus \beta_{v_q}))$ as a vector in $\mathsf{T}^*_q\mathsf{Q} \oplus \mathsf{T}^*_q\mathsf{Q}$. A straightforward computation shows that
$$\big\langle \mathrm{ver}(S^{T^*}(\alpha_{v_q} \oplus \beta_{v_q})); u_{v_q} \oplus w_{v_q} \big\rangle + \big\langle \alpha_{v_q} \oplus \beta_{v_q}; \mathrm{ver}(S^T(u_{v_q} \oplus w_{v_q}))\big\rangle = 0. \bullet$$

The Jacobi equation and the adjoint Jacobi equation have a closer relationship when ∇ is the Levi-Civita connection $\overset{\mathbb{G}}{\nabla}$ associated to a Riemannian metric $\mathbb{G}$. We have the following result.

Proposition S1.40 (Adjoint Jacobi equation for Levi-Civita affine connection). *Let $\mathbb{G}$ be a Riemannian metric on Q with $\overset{\mathbb{G}}{\nabla}$ the Levi-Civita affine connection. If $\gamma\colon I \to \mathsf{Q}$ is a geodesic of $\overset{\mathbb{G}}{\nabla}$, then a covector field $\lambda\colon I \to \mathsf{T}^*\mathsf{Q}$ along γ is a solution of the adjoint Jacobi equation if and only if the vector field $\mathbb{G}^\sharp \circ \lambda$ along γ is a solution of the Jacobi equation.*

Proof. Using the fact that $\overset{\mathbb{G}}{\nabla}\mathbb{G} = 0$, we compute
$$\nabla^2_{\gamma'(t)}\mathbb{G}^\sharp(\lambda(t)) = \mathbb{G}^\sharp(\nabla^2_{\gamma'(t)}\lambda(t)). \tag{S1.21}$$
Now, using the relation
$$\mathbb{G}(R(X_3, X_4)X_2, X_1) = \mathbb{G}(R(X_1, X_2)X_4, X_3),$$
$X_1, X_2, X_3, X_4 \in \Gamma^\infty(\mathsf{T}\mathsf{Q})$, (this is proved as Proposition 1.2 in Chapter V of [Kobayashi and Nomizu 1963, volume 1]), and, for $u \in \mathsf{T}_{\gamma(t)}\mathsf{Q}$, we compute
$$\begin{aligned}\langle R^*(\lambda(t), \gamma'(t))\gamma'(t); u\rangle &= \langle \lambda(t); R(u, \gamma'(t))\gamma'(t)\rangle\\ &= \mathbb{G}(R(u, \gamma'(t))\gamma'(t), \mathbb{G}^\sharp(\lambda(t)))\\ &= \mathbb{G}(R(\mathbb{G}^\sharp(\lambda(t)), \gamma'(t))\gamma'(t), u)\\ &= \langle \mathbb{G}^\flat(R(\mathbb{G}^\sharp(\lambda(t)), \gamma'(t))\gamma'(t)); u\rangle.\end{aligned}$$
This implies that
$$\mathbb{G}^\sharp(R^*(\lambda(t), \gamma'(t))\gamma'(t)) = R(\mathbb{G}^\sharp(\lambda(t)), \gamma'(t))\gamma'(t). \tag{S1.22}$$
Combining (S1.21) and (S1.22) gives
$$\begin{aligned}\nabla^2_{\gamma'(t)}\mathbb{G}^\sharp(\lambda(t)) &+ R(\mathbb{G}^\sharp(\lambda(t)), \gamma'(t))\gamma'(t)\\ &= \mathbb{G}^\sharp\big(\nabla^2_{\gamma'(t)}\lambda(t) + R^*(\lambda(t), \gamma'(t))\gamma'(t)\big)\end{aligned}$$
and the result now follows since $\overset{\mathbb{G}}{\nabla}$ is torsion-free, and since $\mathbb{G}^\sharp$ is a vector bundle isomorphism. ■

S1.3.11 The Sasaki metric

When the constructions of this section are applied in the case when ∇ is the Levi-Civita affine connection associated with a Riemannian metric $\mathbb{G}$ on Q, there is an important additional construction that can be made.

Definition S1.41 (Sasaki metric). Let $(\mathsf{Q}, \mathbb{G})$ be a Riemannian manifold, and for $v_q \in \mathsf{TQ}$, let $\mathsf{T}_{v_q}\mathsf{TQ} \simeq \mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$ denote the splitting defined by the affine connection $\overset{\mathbb{G}}{\nabla}$. The ***Sasaki metric*** is the Riemannian metric $\mathbb{G}^T$ on TQ given by

$$\mathbb{G}^T(u^1_{v_q} \oplus w^1_{v_q}, u^2_{v_q} \oplus w^2_{v_q}) = \mathbb{G}(u^1_{v_q}, u^2_{v_q}) + \mathbb{G}(w^1_{v_q}, w^2_{v_q}). \qquad \bullet$$

The Sasaki metric was introduced by Sasaki [1958, 1962]. Much research has been made into the properties of the Sasaki metric, beginning with the work of Sasaki who studied the curvature, geodesics, and Killing vector fields of the metric. Some of these results are also given in [Yano and Ishihara 1973].

S2

Controllability

This supplement provides extensions to some of the controllability results given in Chapter 7 of the text. In the first two sections, we extend the controllability results of Sections 7.3.2 and 7.3.3 to systems with external forces. We consider two special cases: a basic external force and an isotropic Rayleigh dissipative force. The third section in the supplement provides some controllability results in the presence of symmetry.

S2.1 Accessibility and controllability of systems with a basic external force

In Sections 7.3.2 and 7.3.3 we considered the accessibility and controllability of affine connection control systems, which model simple mechanical control systems without external forces, potential or otherwise, and possibly with constraints. In this section we consider the possibility of adding to such systems a basic external force. Thus we consider the external force supplied by a covector field F on Q to give a forced simple mechanical control system of the form $(\mathsf{Q}, \mathbb{G}, V, F, \mathcal{D}, \mathscr{F} = \{F^1, \dots, F^m\}, U)$. For such a system, let us define a vector field Y on Q by $Y(q) = P_{\mathcal{D}}\big(\mathbb{G}(q)^\sharp(F(q) - \mathrm{d}V(q))\big)$, and let us denote $Y_a = P_{\mathcal{D}}(\mathbb{G}^\sharp(F^a))$, in the usual way. Then the governing equations for this system are

$$\overset{\mathcal{D}}{\nabla}_{\gamma'(t)}\gamma'(t) = \sum_{a=1}^{m} u^a(t) Y_a(c(t)) + Y(\gamma(t)). \tag{S2.1}$$

These are the equations for a forced affine connection control system $(\mathsf{Q}, \nabla, Y, \mathcal{D}, \mathscr{Y}, U)$. With the preceding discussion providing mechanical motivation, in this section we work with a general forced affine connection control system, with a basic external force. Corresponding to this is the control-affine system $(\mathsf{M} = \mathsf{TQ}, \mathscr{C}_\Sigma = \{f_0 = S + \mathrm{vlft}(Y), f_1 = \mathrm{vlft}(Y_1), \dots, f_m = \mathrm{vlft}(Y_m)\}, U)$, where S is the geodesic spray for the affine connection ∇. We suppose the data to be analytic in this section.

S2.1.1 Accessibility results

To describe the accessibility of the system (S2.1) is not as elegant a matter as it is in the case when $Y = 0$. Indeed, all we can do is provide an inductive algorithm for determining two sequences of analytic distributions, denoted by $\mathsf{C}^{(k)}_{\mathrm{hor}}(\mathscr{Y}, Y)$ and $\mathsf{C}^{(k)}_{\mathrm{ver}}(\mathscr{Y}, Y)$, $k \in \mathbb{N}$, and defined in Algorithm S2.1. To really understand the algorithm requires delving into the proof of Theorem S2.2. Let us provide a guide so that the reader will be able to *use* the

Algorithm S2.1.

```
For i ∈ ℤ⁺, do
  For B′ ∈ Br^(i)(ξ′) primitive, do
    If |B′|_{m+1} = 0, then
      If B′ ∈ Br_{−1}(ξ′), then
        U ∈ C_ver^((1/2)(i+1))(𝒴, Y)_{q0} where Ev_{0_{q0}}^{𝒞′_Σ}(B′) = 0_{q0} ⊕ U
      else
        U ∈ C_hor^(i/2)(𝒴, Y)_{q0} where Ev_{0_{q0}}^{𝒞′_Σ}(B′) = U ⊕ 0_{q0}
      end
    else
      If B′ has no components of the form [ξ′_0, ξ′_{m+1}], then
        Compute B(B′) ∈ Br(ξ) by replacing every occurrence of ξ′_0 and ξ′_{m+1}
        in B′ with ξ_0 and by replacing every occurrence of ξ′_a in B′ with ξ_a, for
        a ∈ {1, ..., m}.
        Let L(B′) = 0.
        For B″ ∈ S(B(B′)) ∩ (Br_{−1}(ξ′) ∪ Br_0(ξ′)), do
          Write B″ as a finite sum of primitive brackets in Br(ξ′) by Lemma B.2.
          L(B′) = L(B′) + B″
        end
        If B′ ∈ Br_{−1}(ξ′), then
          U ∈ C_ver^((1/2)(i+1))(𝒴, Y)_{q0} where Ev_{0_{q0}}^{𝒞′_Σ}(L(B′)) = 0_{q0} ⊕ U
        else
          U ∈ C_hor^(i/2)(𝒴, Y)_{q0} where Ev_{0_{q0}}^{𝒞′_Σ}(L(B′)) = U ⊕ 0_{q0}
        end
      end
    end
  end
end
end
```

algorithm with a minimum of excursions into the details of the full proof. The algorithm uses two sets of indeterminates, $\boldsymbol{\xi} = \{\xi_0, \xi_1, \dots, \xi_m\}$ and $\boldsymbol{\xi}' = \{\xi_0, \xi_1, \dots, \xi_m, \xi_{m+1}\}$, defining the free Lie algebras $\mathrm{Lie}(\boldsymbol{\xi})$ and $\mathrm{Lie}(\boldsymbol{\xi}')$. These indeterminates correspond to the families of vector fields

$$\begin{aligned}\mathscr{C}_\Sigma &= \{S + \mathrm{vlft}(Y), \mathrm{vlft}(Y_1), \dots, \mathrm{vlft}(Y_m)\},\\ \mathscr{C}'_\Sigma &= \{S, \mathrm{vlft}(Y_1), \dots, \mathrm{vlft}(Y_m), \mathrm{vlft}(Y)\},\end{aligned}$$

respectively, on TQ. The algorithm also uses the notation $\boldsymbol{S}(B)$, for $B \in \mathrm{Br}(\boldsymbol{\xi})$, and this notation is explained in the preamble to the first lemma in the proof of Theorem S2.2. By understanding this minimum of notation, it should be possible to follow through Algorithm S2.1, and so compute $\mathcal{C}^{(k)}_{\mathrm{hor}}(\mathscr{Y}, Y)$, $k \in \mathbb{N}$, and $\mathcal{C}^{(k)}_{\mathrm{ver}}(\mathscr{Y}, Y)$, $k \in \mathbb{N}$. To illustrate the algorithm, let us provide generators for the first few terms in the sequences $\mathcal{C}^{(k)}_{\mathrm{hor}}(\mathscr{Y}, Y)$ and $\mathcal{C}^{(k)}_{\mathrm{ver}}(\mathscr{Y}, Y)$, $k \in \mathbb{N}$:

$\mathcal{C}^{(1)}_{\mathrm{hor}}(\mathscr{Y}, Y)$ is generated by $\{Y_1, \dots, Y_m\}$,

$\mathcal{C}^{(1)}_{\mathrm{ver}}(\mathscr{Y}, Y)$ is generated by $\{Y_1, \dots, Y_m, Y\}$,

$\mathcal{C}^{(2)}_{\mathrm{hor}}(\mathscr{Y}, Y)$ is generated by $\{\, \langle Y_a : Y_b \rangle \mid \ a, b \in \{1, \dots, m\}\}$
$\cup \{\, [Y_a, Y_b] \mid \ a, b \in \{1, \dots, m\}\} \cup \{\, 2\,\langle Y_a : Y \rangle + [Y_a, Y] \mid \ a \in \{1, \dots, m\}\}$,

$\mathcal{C}^{(2)}_{\mathrm{ver}}(\mathscr{Y}, Y)$ is generated by $\{\langle Y_a : Y_b \rangle \mid \ a, b \in \{1, \dots, m\}\}$
$\cup \{\langle Y_a : Y \rangle \mid \ a \in \{1, \dots, m\}\}$.

It would be interesting to be able to derive an inductive formula for generators for $\mathcal{C}^{(k)}_{\mathrm{ver}}(\mathscr{Y}, Y)$ and $\mathcal{C}^{(k)}_{\mathrm{hor}}(\mathscr{Y}, Y)$. However, such an inductive formula is not presently known. It would also be interesting to have a geometric interpretation for these distributions, along the lines of Theorem 3.108. Again, no such geometric description is presently known. For the objective of studying the controllability of certain multibody systems, Shen [2002] computes generators for $\mathcal{C}^{(3)}_{\mathrm{hor}}(\mathscr{Y}, Y)$ and a subset of the generators for $\mathcal{C}^{(4)}_{\mathrm{hor}}(\mathscr{Y}, Y)$.

The smallest analytic distribution containing each $\mathcal{C}^{(k)}_{\mathrm{hor}}(\mathscr{Y}, Y)$, $k \in \mathbb{N}$, is denoted by $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)$, and $\mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)$ is similarly defined (cf. Lemma 3.94). With this notation, we have the following result, recalling from Lemma 6.34 the natural decomposition of $\mathsf{T}_{0_q}\mathsf{TQ}$ as the direct sum $\mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$.

Theorem S2.2 (Accessibility of affine connection control systems with a basic external force). *Let* $\Sigma = (\mathsf{Q}, \nabla, Y, \mathcal{D}, \mathscr{Y})$ *be an analytic forced affine connection pre-control system with* Y *basic, and let* $\mathscr{C}_\Sigma = \{S + \mathrm{vlft}(Y), \mathrm{vlft}(Y_1), \dots, \mathrm{vlft}(Y_m)\}$*. The distributions* $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)$ *and* $\mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)$ *defined by Algorithm S2.1 satisfy*

$$\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma)_{0_{q_0}} \simeq \mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_{q_0} \oplus \mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)_{q_0} \subset \mathsf{T}_{q_0}\mathsf{Q} \oplus \mathsf{T}_{q_0}\mathsf{Q} \simeq \mathsf{T}_{0_{q_0}}\mathsf{TQ}.$$

In particular, the following statements hold:

(i) Σ *is accessible from* q_0 *if and only if* $\mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)_{q_0} = \mathcal{D}_{q_0}$ *and* $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$*;*

(ii) Σ *is configuration accessible from* q_0 *if and only if* $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$.

Proof. An outline of the proof we give is as follows.

1. In the proof we deal with two free Lie algebras:
   (a) one is generated by indeterminates $\boldsymbol{\xi} = \{\xi_0, \xi_1, \dots, \xi_m\}$ that are in correspondence with the family of vector fields $\mathscr{C}_\Sigma = \{S + \mathrm{vlft}(Y), \mathrm{vlft}(Y_1), \dots, \mathrm{vlft}(Y_m)\}$;
   (b) the other is generated by the indeterminates $\boldsymbol{\xi}' = \{\xi'_0, \xi'_1, \dots, \xi'_m, \xi'_{m+1}\}$ that are in correspondence with the family of vector fields $\mathscr{C}'_\Sigma = \{S, \mathrm{vlft}(Y_1), \dots, \mathrm{vlft}(Y_m), \mathrm{vlft}(Y)\}$.

   We provide a systematic way of relating these free Lie algebras using the structure of the Lie algebra generated by $\mathscr{C}_\Sigma$.
2. We recall from the proof of Theorem 7.36 the structure of the Lie algebra generated by $\mathscr{C}'_\Sigma$. The important fact is that one need only compute what we call "primitive" brackets, and these are evaluated in terms of brackets and symmetric products of the vector fields $\{Y_1, \dots, Y_m, Y\}$.
3. We then use the relationship between $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ to arrive at a form for $\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma)$.

Let $\boldsymbol{\xi} = \{\xi_0, \xi_1, \dots, \xi_m\}$. We formally set $\xi_0 = \xi'_0 + \xi'_{m+1}$ and $\xi_a = \xi'_a$, for $a \in \{1, \dots, m\}$. This corresponds to the relationship between the families of vector fields $\mathscr{C}'_\Sigma$ and $\mathscr{C}_\Sigma$. We may now write brackets in $\mathrm{Br}(\boldsymbol{\xi})$ as linear combinations of brackets in $\mathrm{Br}(\boldsymbol{\xi}')$ by $\mathbb{R}$-linearity of the bracket. We may, in fact, be even more precise about this. Let $B \in \mathrm{Br}(\boldsymbol{\xi})$. We define a subset, $\boldsymbol{S}(B)$, of $\mathrm{Br}(\boldsymbol{\xi}')$ by saying that $B' \in \boldsymbol{S}(B)$ if each occurrence of ξ_a in B is replaced with ξ'_a, for $a \in \{1, \dots, m\}$, and if each occurrence of ξ_0 in B is replaced with *either* ξ'_0 *or* ξ'_{m+1}. An example is illustrative. Suppose that

$$B = [[\xi_0, \xi_a], [\xi_b, [\xi_0, \xi_c]]].$$

Then

$$\begin{aligned}\boldsymbol{S}(B) = \{&[[\xi'_0, \xi'_a], [\xi'_b, [\xi'_0, \xi'_c]]], [[\xi'_0, \xi'_a], [\xi'_b, [\xi'_{m+1}, \xi'_c]]],\\ &[[\xi'_{m+1}, \xi'_a], [\xi'_b, [\xi'_0, \xi'_c]]], [[\xi'_{m+1}, \xi'_a], [\xi'_b, [\xi'_{m+1}, \xi'_c]]]\}.\end{aligned}$$

Now we may precisely state how we write brackets in $\mathrm{Br}(\boldsymbol{\xi})$.

Lemma. *Let* $B \in \mathrm{Br}(\boldsymbol{\xi})$*. Then*

$$B = \sum_{B' \in \boldsymbol{S}(B)} B'.$$

Proof. It suffices to prove the lemma for the case when B is of the form

$$B = [\xi_{a_k}, [\xi_{a_{k-1}}, [\cdots, [\xi_{a_2}, \xi_{a_1}]]]], \qquad a_1, \dots, a_k \in \{0, 1, \dots, m\}, \tag{S2.2}$$

since these brackets generate $\mathrm{Lie}(\boldsymbol{\xi})$ by Exercise E7.2. We proceed by induction on k. The lemma is clearly true for $k = 1$. Now suppose the lemma true for $k \in \{1, \dots, \ell\}$, where $\ell \geq 1$, and let B be of the form (S2.2) with $k = \ell + 1$. Then, either $B = [\xi_a, B'']$, $a \in \{1, \dots, m\}$, or $B = [\xi_0, B'']$ with B'' of the form (S2.2) with $k = \ell$. In the first case, by the induction hypotheses, we have

$$B = \sum_{B' \in \boldsymbol{S}(B'')} [\xi'_a, B'] = \sum_{B' \in \boldsymbol{S}(B)} B'.$$

In the second case we have

$$B = \sum_{B' \in \boldsymbol{S}(B'')} [\xi'_0 + \xi'_{m+1}, B'] = \sum_{B' \in \boldsymbol{S}(B)} B'.$$

This proves the lemma. ▾

By the preceding lemma, each vector field in $\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma)$ is a $\mathbb{R}$-linear sum of vector fields in $\mathrm{Lie}^{(\infty)}(\mathscr{C}'_\Sigma)$. Moreover, the characterization of $\mathrm{Lie}^{(\infty)}(\mathscr{C}'_\Sigma)$ has already been obtained, during the course of the proof of Theorem 7.36, in terms of Lie brackets and symmetric products of the vector fields $\{\mathrm{vlft}(Y_1), \dots, \mathrm{vlft}(Y_m), \mathrm{vlft}(Y)\}$. In this way, we arrive at a description of $\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma)$ in terms of Lie brackets and symmetric products of the vector fields $\{\mathrm{vlft}(Y_1), \dots, \mathrm{vlft}(Y_m), \mathrm{vlft}(Y)\}$. We claim that Algorithm S2.1 determines exactly which $\mathbb{R}$-linear combinations from $\mathrm{Lie}^{(\infty)}(\mathrm{Sym}^{(\infty)}(\mathscr{Y} \cup \{Y\}))$ we need to compute.

Lemma. *Let $q_0 \in \mathsf{Q}$. Then*

$$\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma)_{0_{q_0}} = \mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_{q_0} \oplus \mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)_{q_0}.$$

Proof. Studying Algorithm S2.1 that we have used to compute $\mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)$ and $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)$, the reader will notice that we have exactly taken each primitive bracket $B' \in \mathrm{Br}(\boldsymbol{\xi}')$ and computed which $\mathbb{R}$-linear combinations from $\mathrm{Br}(\boldsymbol{\xi}')$ appear along with B' in the decomposition of some $B \in \mathrm{Br}(\boldsymbol{\xi})$ given by the first lemma of the proof. Since it is only these primitive brackets that appear in $\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma) \mid Z(\mathsf{TQ})$, this will, by construction, generate $\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma) \mid Z(\mathsf{TQ})$.

We need to prove that, as stated in the first step of the algorithm, if $|B'|_{m+1} = 0$, then $\mathrm{Ev}^{\mathscr{C}'_\Sigma}_{0_{q_0}}(B') \in \mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma)_{0_{q_0}}$. To show that this is in fact the case, let $B \in \mathrm{Br}(\boldsymbol{\xi})$ be the bracket obtained from B' by replacing ξ'_a with ξ_a, for $a \in \{0, 1, \dots, m\}$. We claim that the only bracket in $\boldsymbol{S}(B)$ that contributes to $\mathrm{Ev}^{\mathscr{C}_\Sigma}(B)$ is B'. This is true since any other brackets in $\boldsymbol{S}(B)$ are obtained by replacing ξ'_0 in B' with ξ'_{m+1}. Such a replacement will result in a bracket that has at least one component that is in $\mathrm{Br}_{-\ell}(\boldsymbol{\xi}')$ for $\ell \geq 2$. These brackets evaluate to zero by Lemma B.5.

We also need to show that, if B' has components of the form $[\xi'_0, \xi'_{m+1}]$, then it will not contribute to $\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma) \mid Z(\mathsf{TQ})$. This is clear since, when constructing $B(B')$ in the algorithm, the component $[\xi'_0, \xi'_{m+1}]$ will become $[\xi_0, \xi_0]$, which means that $B(B')$ will be identically zero. ▾

The preceding lemma is none other than Theorem S2.2. ■

Clearly one should be able to recover Theorem 7.36 from Theorem S2.2. This can indeed by done by following Algorithm S2.1 in the case where $Y = 0$.

A non-obvious, but useful, fact is the following.

Proposition S2.3 (Involutivity of $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)$). *$\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)$ is involutive.*

Proof. Let $B_1, B_2 \in \mathrm{Br}(\boldsymbol{\xi})$ be brackets that, when evaluated under $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_q}$, give vector fields $U_1, U_2 \in \Gamma^\omega(\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y))$. Then, for $i \in \{1, 2\}$, the decomposition of B_i given by the first lemma of the proof of Theorem S2.2 has the form $B_i = B'_i + \tilde{B}_i$, where $B'_i \in \mathrm{Br}_0(\boldsymbol{\xi}')$ and $\tilde{B}_i$ is a sum of brackets in $\mathrm{Br}_j(\boldsymbol{\xi}')$, for $j \geq 2$. Therefore, $[B_1, B_2] = [B'_1, B'_2] + B''$ where B'' is a sum of brackets in $\mathrm{Br}_j(\boldsymbol{\xi}')$, for $j \geq 2$. This shows that $[U_1, U_2](0_q) \in \mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_q \oplus \{0_q\}$, using Lemma B.2. ■

S2.1.2 Controllability results

Next we state a sufficient condition for small-time local controllability of systems of the type under consideration. The result we give is an adaptation of Theorem 7.20 to the current setting. To state the theorem requires another redevelopment of the use of weights and obstructions to this setting. An understanding of the proof of Theorem S2.2 at the same level required to make sense of Algorithm S2.1 is required. Also, in order to eliminate the introduction of additional notation, we shall reuse some of the notation used in stating Theorem 7.40. The reader should be aware of this not so slight abuse of notation, and understand that the notation used from here to the end of the section applies only to the results contained therein.

We consider an analytic affine connection ∇ on a manifold Q, and analytic vector fields $\mathscr{Y}' = \{Y_1, \dots, Y_m, Y\}$ on Q. We work at a point $q_0 \in \mathsf{Q}$ for which $Y(q_0) = 0_{q_0}$. Thus q_0 is an equilibrium configuration for the forced affine connection control system $(\mathsf{Q}, \nabla, Y, \mathscr{Y}, U)$. The indeterminates $\boldsymbol{\xi} = \{\xi_0, \xi_1, \dots, \xi_m\}$ and $\boldsymbol{\xi}' = \{\xi'_0, \xi'_1, \dots, \xi'_m, \xi'_{m+1}\}$ have the same meaning as in Theorem S2.2. We also introduce indeterminates $\boldsymbol{\eta}' = \{\eta_1, \dots, \eta_m, \eta_{m+1}\}$ with the free symmetric algebra $\mathrm{Sym}(\boldsymbol{\eta}')$. An ***admissible weight*** for $\boldsymbol{\eta}'$ is an $(m+1)$-tuple $(w_1, \dots, w_m, w_{m+1})$ of nonnegative real numbers satisfying $w_a \geq w_{m+1}$, $a \in \{1, \dots, m\}$. The set of admissible weights is denoted by

$\mathrm{wgt}(\boldsymbol{\eta}')$. Let $B' \in \mathrm{Br}_{-1}(\boldsymbol{\xi}')$ be primitive and construct $L(B')$ as in Algorithm S2.1. There then exists $k(B') \in \mathbb{N}$ and $P_1(B'), \dots, P_{k(B')}(B') \in \mathrm{Pr}(\boldsymbol{\eta}')$ such that

$$\mathrm{Ev}_{0_{q_0}}^{\mathscr{C}_\Sigma}(L(B')) = \sum_{j=1}^{k(B')} \mathrm{vlft}(\mathrm{Ev}_{q_0}^{\mathscr{Y}'}(P_j(B'))). \tag{S2.3}$$

Note that $k(B')$ and $P_1(B'), \dots, P_{k(B')}(B') \in \mathrm{Pr}(\boldsymbol{\eta}')$ are uniquely defined by Algorithm S2.1. Also note that $|(P_j(B'))|_a$, $a \in \{1, \dots, m, m+1\}$, is independent of $j \in \{1, \dots, k(B')\}$ (this is a matter of checking the definition of $L(B')$). Now we define subsets $\mathrm{Sym}_0(\boldsymbol{\eta}')$ and $\overline{\mathrm{Sym}}_0(\boldsymbol{\eta}')$ of $\mathrm{Sym}(\boldsymbol{\eta}')$ by

$$\begin{aligned}
\mathrm{Sym}_0(\boldsymbol{\eta}') &= \left\{ P_1(B') + \cdots + P_{k(B')}(B') \;\middle|\; B' \in \mathrm{Br}_{-1}(\boldsymbol{\xi}') \text{ primitive} \right\}, \\
\overline{\mathrm{Sym}}_0(\boldsymbol{\eta}') &= \mathrm{span}_{\mathbb{R}} \left\{ \mathrm{Sym}_0(\boldsymbol{\eta}') \right\}.
\end{aligned}$$

For $\boldsymbol{w} \in \mathrm{wgt}(\boldsymbol{\eta}')$ and for $S \in \mathrm{Sym}_0(\boldsymbol{\eta}')$, define the ***w*-weight** of S by

$$\|S\|_{\boldsymbol{w}} = \sum_{a=1}^{m} w_a \, |S|_a + w_{m+1}(2\,|S|_{m+1} - 1),$$

where $|S|_a$ means $|P|_a$, where $P \in \mathrm{Pr}(\boldsymbol{\eta}')$ is any, it matters not which, of the summands of S. A member of $\overline{\mathrm{Sym}}_0(\boldsymbol{\eta}')$ is ***w*-homogeneous** if it is a sum of terms of the same $\boldsymbol{w}$-weight. For $\boldsymbol{w} \in \mathrm{wgt}(\boldsymbol{\eta}')$ and $k \in \mathbb{N}$, let us denote

$$\mathcal{V}_{\mathscr{Y}'}^{\boldsymbol{w},k} = \mathrm{span}_{\mathbb{R}} \left\{ \mathrm{Ev}_{q_0}^{\mathscr{Y}'}(S) \;\middle|\; S \in \mathrm{Sym}_0(\boldsymbol{\eta}'),\ \|S\|_{\boldsymbol{w}} \le k \right\}.$$

By convention we take $\mathcal{V}_{\mathscr{Y}'}^{\boldsymbol{w},0} = 0_{q_0}$. If $S \in \overline{\mathrm{Sym}}_0(\boldsymbol{\eta}')$ is $\boldsymbol{w}$-homogeneous, then it is ***w*-neutralized** by $\mathscr{Y}'$ at q_0 if there exists an integer $k < \|S\|_{\boldsymbol{w}}$ such that $\mathrm{Ev}_{q_0}^{\mathscr{Y}'}(S) \in \mathcal{V}_{\mathscr{Y}'}^{\boldsymbol{w},k}$. Denote

$$\mathcal{B}(\mathrm{Sym}_0(\boldsymbol{\eta}')) = \mathrm{span}_{\mathbb{R}}\{S \in \mathrm{Sym}_0(\boldsymbol{\eta}') \mid |S|_a \text{ is even, } a \in \{1, \dots, m\}\}.$$

For $\sigma \in S_m$ and $\boldsymbol{w} \in \mathrm{wgt}(\boldsymbol{\eta}')$, define

$$\sigma(\boldsymbol{w}) = (w_{\sigma(1)}, \dots, w_{\sigma(m)}, w_{m+1}),$$

and denote by $S_m^{\boldsymbol{w}}$ those permutations for which $\sigma(\boldsymbol{w}) = \boldsymbol{w}$. For $P \in \mathrm{Pr}(\boldsymbol{\eta}')$, denote by $\sigma(P)$ the product obtained by replacing η_a with $\eta_{\sigma(a)}$, $a \in \{1, \dots, m\}$. For $S \in \overline{\mathrm{Sym}}_0(\boldsymbol{\eta}')$, let $\sigma(S)$ be defined by applying σ to each of the summands from $\mathrm{Sym}_0(\boldsymbol{\eta}')$. For $\boldsymbol{w} \in \mathrm{wgt}(\boldsymbol{\eta}')$, define

$$\mathcal{B}_0^{\boldsymbol{w}}(\mathrm{Sym}_0(\boldsymbol{\eta}')) = \left\{ S \in \mathcal{B}(\mathrm{Sym}_0(\boldsymbol{\eta}')) \mid \sigma(S) = S \text{ for all } \sigma \in S_m^{\boldsymbol{w}} \right\}.$$

A ***w*-obstruction** is an element of $\mathcal{B}_0^{\boldsymbol{w}}(\mathrm{Sym}_0(\boldsymbol{\eta}'))$. Next define $\rho_{\mathscr{Y}'}^{q_0}(\boldsymbol{w}) \in \mathbb{N}$ to be the largest integer for which all $\boldsymbol{w}$-obstructions S satisfying $\|S\|_{\boldsymbol{w}} \le \rho_{\mathscr{Y}'}^{q_0}(\boldsymbol{w})$ are $\boldsymbol{w}$-neutralized by $\mathscr{Y}'$ at q_0.

Since $Y(q_0) = 0_{q_0}$, we may define $\mathrm{sym}_Y(q_0) \in L(\mathsf{T}_{q_0}\mathsf{Q}; \mathsf{T}_{q_0}\mathsf{Q})$ as in Exercise E7.23. Now define

$$\mathcal{V}^{\boldsymbol{w}}_{\mathscr{Y}'}(q_0) = \mathrm{span}_{\mathbb{R}}\big\{\mathrm{sym}_Y(q_0)^j(X)\big|\ X \in \mathcal{V}^{\boldsymbol{w},\rho^{q_0}_{\mathscr{Y}'}(\boldsymbol{w})}_{\mathscr{Y}'}(q_0),\ j \in \mathbb{N}_0\big\}.$$

We are now ready to state the adaptation of Theorem 7.20 to the systems we are considering.

Theorem S2.4 (Controllability of affine connection control systems with a basic external force). *Let* $\Sigma = (\mathsf{Q}, \nabla, Y, \mathcal{D}, \mathscr{Y} = \{Y_1, \dots, Y_m\})$ *be an analytic forced affine connection pre-control system with* Y *basic, and let* $\boldsymbol{\eta}' = \{\eta_1, \dots, \eta_m, \eta_{m+1}\}$*. If* $Y(q_0) = 0_{q_0}$ *and if*

$$\sum_{\boldsymbol{w}\in\mathrm{wgt}(\boldsymbol{\eta}')} \mathcal{V}^{\boldsymbol{w}}_{\mathscr{Y}}(q_0) = \mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)_{q_0},$$

then the following statements hold:

(i) if $\mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)_{q_0} = \mathcal{D}_{q_0}$ *and if* $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$*, then* Σ *is properly STLC from* q_0*;*

(ii) if $\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$*, then* Σ *is properly STLCC from* q_0*.*

Proof. We claim that, if $B \in \mathcal{B}(\boldsymbol{\xi})$, then $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B) = \mathrm{vlft}_{q_0}(\mathrm{Ev}^{\mathscr{Y}'}_{q_0}(S))$ where $S \in \mathcal{B}(\mathrm{Sym}_0(\boldsymbol{\eta}'))$. First note that $\mathrm{Sym}_0(\boldsymbol{\eta}')$ is specifically constructed such that, if $B \in \mathrm{Br}_{-1}(\boldsymbol{\xi})$ is primitive, then $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B) = \mathrm{vlft}_{q_0}(\mathrm{Ev}^{\mathscr{Y}'}_{q_0}(S))$ where $S \in \mathrm{Sym}_0(\boldsymbol{\eta}')$. Now let $B \in \mathcal{B}(\boldsymbol{\xi})$. Thus $|B|_a$ is even for $a \in \{1, \dots, m\}$ and $|B|_0$ is odd. When we evaluate $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B)$, the only terms that will remain in the decomposition of $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B)$ given by the first lemma in the proof of Theorem S2.2 are the terms obtained from brackets in $\boldsymbol{S}(B)$ that are in $\mathrm{Br}_0(\boldsymbol{\xi}') \cup \mathrm{Br}_{-1}(\boldsymbol{\xi}')$. Thus let $B' \in \boldsymbol{S}(B) \cap (\mathrm{Br}_0(\boldsymbol{\xi}') \cup \mathrm{Br}_{-1}(\boldsymbol{\xi}'))$. Since $B \in \mathcal{B}(\boldsymbol{\xi})$, we must have $|B'|_a$ even and $|B'|_0 + |B'|_{m+1}$ odd. We now have two cases.

1. If $|B'|_0$ is odd, then $|B'|_{m+1}$ must be even. In this case we get $\sum_{a=1}^{m+1} |B'|_a$ as even and $|B'|_0$ as odd. Thus the only such brackets $B' \in \boldsymbol{S}(B)$ that contribute to $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B)$ must be in $\mathrm{Br}_{-1}(\boldsymbol{\xi}')$.
2. If $|B'|_0$ is even for $B' \in \boldsymbol{S}(B)$, then $|B'|_{m+1}$ must be odd. In this case $\sum_{a=1}^{m+1} |B'|_a$ is odd and $|B'|_0$ is even and again, the only brackets in $\boldsymbol{S}(B)$ that contribute to $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B)$ must be in $\mathrm{Br}_{-1}(\boldsymbol{\xi}')$.

Thus we must have $\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B')$ as a linear combination of terms like $\mathrm{vlft}_{q_0}(\mathrm{Ev}^{\mathscr{Y}'}_{q_0}(P'))$, for $P' \in \mathcal{P}(\boldsymbol{\eta}')$. Since this is true for every $B' \in \boldsymbol{S}(B) \cap (\mathrm{Br}_0(\boldsymbol{\xi}') \cup \mathrm{Br}_{-1}(\boldsymbol{\xi}'))$, it is also true for B. Clearly, if $B \in \mathcal{B}(\boldsymbol{\xi})$, then $S \in \mathcal{B}(\mathrm{Sym}_0(\boldsymbol{\eta}'))$.

The above argument shows that all obstructions evaluate to vectors in $\mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)$. Since $Y(q_0) = 0_{q_0}$ it follows as in Exercise E7.23 that we may regard $\mathrm{sym}_Y(q_0)$ as a linear map on $\mathsf{T}_{q_0}\mathsf{Q}$. The hypotheses of the theorem ensure the existence of $\boldsymbol{w}_1, \dots, \boldsymbol{w}_k \in \mathrm{wgt}(\boldsymbol{\eta}')$ and elements $S_1, \dots, S_k \in \mathrm{Sym}_0(\boldsymbol{\eta}')$ such that $\|S_j\|_{\boldsymbol{w}_j} < \rho^{q_0}_{\mathscr{Y}'}(\boldsymbol{w}_j)$, $j \in \{1, \dots, k\}$, and such that

$$\left\langle \mathrm{sym}_Y(q_0), \mathrm{span}_{\mathbb{R}}\left\{\mathrm{Ev}^{\mathscr{Y}'}_{q_0}(S_j) \middle|\ j \in \{1, \dots, k\}\right\}\right\rangle = \mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)_{q_0}. \tag{S2.4}$$

For $j \in \{1, \dots, k\}$, write $\boldsymbol{w}_j = (w_{j,1}, \dots, w_{j,m}, w_{j,m+1})$, and define $\tilde{w}_j \in \mathrm{wgt}(\boldsymbol{\xi})$ by $\tilde{w}_j = (w_{j,m+1}, w_{j,1}, \dots, w_{j,m})$. Next we note that a direct computation may be used to verify that $A_{f_0}(0_{q_0})$ is the linear map on $\mathsf{T}_{0_{q_0}}\mathsf{TQ} \simeq \mathsf{T}_{q_0}\mathsf{Q} \oplus \mathsf{T}_{q_0}\mathsf{Q}$ represented by the block matrix

$$\begin{bmatrix} 0 & \mathrm{id}_{\mathsf{T}_{q_0}\mathsf{Q}} \\ \mathrm{sym}_Y(q_0) & 0 \end{bmatrix}.$$

Now, our definitions of $\mathrm{Sym}_0(\boldsymbol{\eta}')$, $\|\cdot\|_{\boldsymbol{w}}$, $\rho^{q_0}_{\mathscr{Y}'}(\boldsymbol{w}_j)$, $\mathcal{B}^{\boldsymbol{w}}_0(\mathrm{Sym}_0(\boldsymbol{\eta}))$, and $\mathrm{sym}_Y(q_0)$ ensure that (S2.4) implies the existence of $B_1, \dots, B_k \in \mathrm{Br}(\boldsymbol{\xi})$ such that

$$\left\langle A_{f_0}(0_{q_0}), \mathrm{span}_{\mathbb{R}}\left\{\mathrm{Ev}^{\mathscr{C}_\Sigma}_{0_{q_0}}(B_j) \middle|\ j \in \{1, \dots, k\}\right\}\right\rangle = \{0_{q_0}\} \oplus \mathcal{C}^{(\infty)}_{\mathrm{ver}}(\mathscr{Y}, Y)_{q_0}\big).$$

A similar (but easier since there are no obstructions to consider) argument may be used to show that

$$\mathcal{C}^{(\infty)}_{\mathrm{hor}}(\mathscr{Y}, Y)_{q_0} \oplus \{0_{q_0}\} \subset \sum_{\boldsymbol{w} \in \mathrm{wgt}(\boldsymbol{\xi})} \mathcal{V}^{\boldsymbol{w}}_{\mathscr{C}_\Sigma}(0_{q_0}),$$

and this gives the theorem. ■

We next state the "good/bad product" characterization for controllability resulting from an application of Corollary 7.24 to our present setting. We adopt the notation deployed in the buildup to Theorem S2.4. In particular, recall the subset $\mathrm{Sym}_0(\boldsymbol{\eta}')$ of $\mathrm{Sym}(\boldsymbol{\eta}')$ consisting of sums of products of the form $P_1(B'), \dots, P_{k(B')}(B')$ for a primitive $B' \in \mathrm{Br}_{-1}(\boldsymbol{\xi}')$. Recall that, for $a \in \{1, \dots, m, m+1\}$, $|S|_a$ is well-defined for $S \in \mathrm{Sym}_0(\boldsymbol{\eta}')$. We say that $S \in \mathrm{Sym}_0(\boldsymbol{\eta}')$ is ***bad*** if $|S|_a$ is even and ***good*** otherwise. The ***degree*** of $S \in \mathrm{Sym}_0(\boldsymbol{\eta}')$ is

$$\deg(S) = \sum_{a=1}^{m+1} |S|_a\,.$$

For $S \in \mathrm{Sym}_0(\boldsymbol{\eta}')$, define

$$\rho(S) = \sum_{\sigma \in S_m} \sigma(S).$$

As usual, the following result may be proved by a direct application of Theorem S2.4 (or of Theorem 7.20), and we leave the working out of this to the motivated reader.

Corollary S2.5 (Controllability of affine connection control systems with a basic external force). *Let $\Sigma = (\mathsf{Q}, \nabla, Y, \mathcal{D}, \mathscr{Y})$ be an analytic forced affine connection control system with $Y(q_0) = 0_{q_0}$. Suppose that, for every bad $S \in \mathrm{Sym}_0(\boldsymbol{\eta}')$, there exists good $Q_1, \dots, Q_k \in \mathrm{Sym}_0(\boldsymbol{\eta}')$ with $\deg(Q_j) < \deg(S)$, $j \in \{1, \dots, k\}$, and constants $a_1, \dots, a_k \in \mathbb{R}$ such that*

$$\mathrm{Ev}_{q_0}^{\mathscr{Y}'}(\rho(S)) = \sum_{j=1}^{k} a_j \mathrm{Ev}_{q_0}^{\mathscr{Y}}(Q_j).$$

Then the following statements hold:

(i) if $\mathcal{C}_{\mathrm{ver}}^{(\infty)}(\mathscr{Y}, Y)_{q_0} = \mathcal{D}_{q_0}$ and if $\mathcal{C}_{\mathrm{hor}}^{(\infty)}(\mathscr{Y}, Y)_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$, then Σ is properly STLC from q_0;

(ii) if $\mathcal{C}_{\mathrm{hor}}^{(\infty)}(\mathscr{Y}, Y)_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$, then Σ is properly STLCC from q_0.

Remark S2.6. Theorem 5.15 in [Lewis and Murray 1997], which is an attempt to state Corollary S2.5, is incorrect. While the statement in [Lewis and Murray 1997] is correct when specialized to systems without potential (this is Corollary 7.41 in the text), for systems with potential, the more complicated development of Corollary S2.5 is necessary. •

S2.2 Accessibility of systems with isotropic dissipation

The final general situation we consider is that when, to an affine connection control system, we add a dissipative force of a certain type. Specifically, we consider a simple mechanical control system $\Sigma = (\mathsf{Q}, \mathbb{G}, V = 0, F_{\mathrm{diss}}, \mathcal{D}, \mathscr{F} = \{F^1, \dots, F^m\}, U)$ where F_{diss} is a dissipative force satisfying $F_{\mathrm{diss}}(v_q)^\sharp = \delta v_q$, for some $\delta > 0$. Thus the dissipation is "isotropic," meaning that the dissipative force is independent of direction. This is not a very general model for dissipation in practice, and the results in this section merely provide a starting point for further investigation of controllability of systems with dissipation. The governing equations for a system of this type are

$$\overset{\mathcal{D}}{\nabla}_{\gamma'(t)}\gamma'(t) = \sum_{a=1}^{m} u^a(t) Y_a(c(t)) - \delta\gamma'(t). \qquad \text{(S2.5)}$$

We consider, therefore, the setting of a forced affine connection control system $(\mathsf{Q}, \nabla, -\delta\mathrm{id}_{\mathsf{TQ}}, \mathscr{Y}, U)$. This gives the corresponding control-affine system $(\mathsf{M} = \mathsf{TQ}, \mathscr{C}_\Sigma = \{f_0 = S - \delta V_{\mathrm{L}}, f_1 = \mathrm{vlft}(Y_1), \dots, f_m = \mathrm{vlft}(Y_m)\}, U)$, where V_{L} is the Liouville vector field. Cortés, Martínez, and Bullo [2003] provide the following characterization of the accessibility of the system (S2.5).

Theorem S2.7 (Accessibility of affine connection control systems with isotropic dissipation). *Consider the analytic forced affine connection control system $\Sigma = (\mathsf{Q}, \nabla, -\delta\mathrm{id}_{\mathsf{TQ}}, \mathscr{Y}, U)$. Then we have*

$$\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma)_{0_q} \simeq \mathrm{Lie}^{(\infty)}(\mathrm{Sym}^{(\infty)}(\mathcal{Y}))_{q_0} \oplus \mathrm{Sym}^{(\infty)}(\mathcal{Y})_{q_0}.$$

In particular, if U is almost proper, then the following statements hold:

(i) Σ *is accessible from* q_0 *if and only if* $\mathrm{Sym}^{(\infty)}(\mathcal{Y})_{q_0} = \mathcal{D}_{q_0}$ *and* $\mathrm{Lie}^{(\infty)}(\mathcal{D})_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$*;*

(ii) Σ *is configuration accessible from* q_0 *if and only if* $\mathrm{Lie}^{(\infty)}(\mathrm{Sym}^{(\infty)}(\mathcal{Y}))_{q_0} = \mathsf{T}_{q_0}\mathsf{Q}$.

Proof. We work with the families of vector fields

$$\begin{aligned}
\mathscr{C}'_\Sigma &= \{f'_0 = S, f'_1 = \mathrm{vlft}(Y_1), \dots, f'_m = \mathrm{vlft}(Y_m)\},\\
\mathscr{C}_\Sigma &= \{f_0 = S - \delta V_{\mathrm{L}}, f_1 = \mathrm{vlft}(Y_1), \dots, f_m = \mathrm{vlft}(Y_m)\}.
\end{aligned}$$

We claim that the ideal generated by $\{f'_1, \dots, f'_m\}$ in $\Gamma^\infty(\mathrm{Lie}^{(\infty)}(\mathscr{C}'_\Sigma))$ agrees with the ideal generated by $\{f_1, \dots, f_m\}$ in $\Gamma^\infty(\mathrm{Lie}^{(\infty)}(\mathscr{C}_\Sigma))$. These ideals are generated by vector fields of the form

$$[f'_{a_{k-1}}, [f'_{a_{k-2}}, \cdots, [f'_{a_1}, f_a]]],\qquad a_1, \dots, a_{k-1} \in \{0, 1, \dots, m\},\ a \in \{1, \dots, m\},\ k \in \mathbb{N},$$

and

$$[f_{a_{k-1}}, [f_{a_{k-2}}, \cdots, [f_{a_1}, f_a]]],\qquad a_1, \dots, a_{k-1} \in \{0, 1, \dots, m\},\ a \in \{1, \dots, m\},\ k \in \mathbb{N},$$

respectively. For fixed $k \in \mathbb{N}$, let $\mathscr{G}'_k$ and $\mathscr{G}_k$ denote these sets of generators. We will show by induction that, for each $k \in \mathbb{N}$,

1. $[V_{\mathrm{L}}, \mathscr{G}'_k] \subset \mathrm{span}_{\mathbb{R}}\{\mathscr{G}'_k\}$ and
2. $\mathrm{span}_{\mathbb{R}}\{\mathscr{G}_k\} = \mathrm{span}_{\mathbb{R}}\{\mathscr{G}'_k\}$.

These trivially hold when $k = 1$. One may readily compute $[V_{\mathrm{L}}, \mathrm{vlft}(Y_a)] = \mathrm{vlft}(Y_a)$, $a \in \{1, \dots, m\}$, and $[V_{\mathrm{L}}, S] = S$, thus giving Fact 1 for $k = 2$. Also, it is clear that Fact 2 holds when $k = 2$.

Now assume that Facts 1 and 2 hold for $k = \ell > 2$. To show that $[V_{\mathrm{L}}, \mathscr{G}'_{\ell+1}] \subset \mathrm{span}_{\mathbb{R}}\left\{\mathscr{G}'_{\ell+1}\right\}$, we show that

$$[V_{\mathrm{L}}, [X, Y]] \in \mathrm{span}_{\mathbb{R}}\left\{\mathscr{G}'_{\ell+1}\right\}, \qquad X \in \mathscr{C}'_\Sigma,\ Y \in \mathscr{G}'_\ell.$$

By the Jacobi identity, we have $[V_{\mathrm{L}}, [X, Y]] = -[Y, [V_{\mathrm{L}}, X]] - [X, [Y, V_{\mathrm{L}}]]$. Since Fact 1 holds when $k = 2$, we have $[V_{\mathrm{L}}, X] \in \mathrm{span}_{\mathbb{R}}\{\mathscr{G}'_1\}$ and so $[Y, [V_{\mathrm{L}}, X]] \in \mathrm{span}_{\mathbb{R}}\left\{\mathscr{G}'_{\ell+1}\right\}$. By the induction hypotheses, $[Y, V_{\mathrm{L}}] \in \mathrm{span}_{\mathbb{R}}\{\mathscr{G}'_\ell\}$, and so $[X, [Y, V_{\mathrm{L}}]] \in \mathrm{span}_{\mathbb{R}}\left\{\mathscr{G}'_{\ell+1}\right\}$. This shows that Fact 1 holds when $k = \ell + 1$.

Now we show that Fact 2 holds when $k = \ell + 1$. Let $X' \in \mathscr{C}'_\Sigma$ and $Y' \in \mathscr{G}'_\ell$. If $X' = \mathrm{vlft}(Y_a)$ for some $a \in \{1, \dots, m\}$, then $[X', Y'] \in \mathrm{span}_{\mathbb{R}}\{\mathscr{G}_{\ell+1}\}$ since $X' \in \mathscr{C}_\Sigma$ and since $Y' \in \mathrm{span}_{\mathbb{R}}\{\mathscr{G}_\ell\}$ by the induction hypothesis. If $X' = S$, then we write $X' = S + V_{\mathrm{L}} - V_{\mathrm{L}}$ and we have

$$[X', Y'] = [S + V_{\mathrm{L}}, Y'] - [V_{\mathrm{L}}, Y'] \in \operatorname{span}_{\mathbb{R}} \{\mathscr{G}_{\ell+1}\},$$

using the fact that both Facts 1 and 2 hold when $k = \ell$. This shows that $\operatorname{span}_{\mathbb{R}} \{\mathscr{G}'_{\ell+1}\} \subset \operatorname{span}_{\mathbb{R}} \{\mathscr{G}_{\ell+1}\}$. Now let $X \in \mathscr{C}_\Sigma$ and $Y \in \mathscr{G}_\ell$. If $X = \operatorname{vlft}(Y_a)$ for some $a \in \{1, \dots, m\}$, then $[X, Y] \in \operatorname{span}_{\mathbb{R}} \{\mathscr{G}'_{\ell+1}\}$ since $X \in \mathscr{C}'_\Sigma$ and since $Y \in \operatorname{span}_{\mathbb{R}} \{\mathscr{G}'_\ell\}$ by the induction hypothesis. If $X = S + V_{\mathrm{L}}$ then we have

$$[X, Y] = [S + V_{\mathrm{L}}, Y] = [S, Y] + [V_{\mathrm{L}}, Y] \in \operatorname{span}_{\mathbb{R}} \{\mathscr{G}'_{\ell+1}\},$$

using the fact that both Facts 1 and 2 hold when $k = \ell$. This shows that $\operatorname{span}_{\mathbb{R}} \{\mathscr{G}_{\ell+1}\} \subset \operatorname{span}_{\mathbb{R}} \{\mathscr{G}'_{\ell+1}\}$. Thus we have shown that the ideals stated as being equal are indeed equal. Thus $\operatorname{Lie}^{(\infty)}(\mathscr{C}'_\Sigma)$ and $\operatorname{Lie}^{(\infty)}(\mathscr{C}_\Sigma)$ differ only by $\operatorname{span}_{\mathbb{R}} \{f'_0 - f_0\} = \operatorname{span}_{\mathbb{R}} \{V_{\mathrm{L}}\}$. Since $V_{\mathrm{L}}(0_q) = 0_{0_q}$, the result follows. ■

S2.3 Controllability in the presence of symmetry

The results stated in Section 7.3 are general in nature. In this section we consider the effects of some additional structure on the configuration space, namely that when a Lie group G acts on Q in such a way that the problem data is invariant. We consider two cases of this. The first, considered in Section S2.3.1, is that when Q is itself a Lie group. In this case, the controllability results of Section 7.3 can be simplified to algebraic tests. This work was presented in the paper of Bullo, Leonard, and Lewis [2000]. In Section S2.3.2 we consider the situation where Q is a principal fiber bundle, and where the input vector fields $\mathscr{Y} = \{Y_1, \dots, Y_m\}$ are G-invariant and horizontal. A situation similar to this was considered by Bloch and Crouch [1992].

S2.3.1 Controllability of systems on Lie groups

We shall consider simple mechanical control systems for which the configuration space is a Lie group: $\mathsf{Q} = \mathsf{G}$. We ask that all problem data be invariant under left translations in the group. Since left-invariant functions are locally constant, we assume that the potential function is zero. Indeed, let us for simplicity assume that there are no external forces on the system, so that we may reduce to the case of an affine connection control system $\Sigma = (\mathsf{G}, \nabla, \mathcal{D}, \mathscr{Y} = \{Y_1, \dots, Y_m\}, U)$, where ∇ is a left-invariant affine connection, $\mathcal{D}$ is a left-invariant distribution, and $\mathscr{Y}$ is a collection of left-invariant vector fields. Let us write $Y_a(g) = T_e L_g(\eta_a)$, thus defining $\eta_a \in \mathfrak{g}$, $a \in \{1, \dots, m\}$. We define a bilinear map $\langle \cdot : \cdot \rangle_{\mathfrak{g}}$ on $\mathfrak{g}$ by

$$\langle \xi : \eta \rangle_{\mathfrak{g}} = \langle X : Y \rangle (e),$$

where X and Y are the left-invariant extensions of $\xi, \eta \in \mathfrak{g}$. Note that, if ∇ is the Levi-Civita connection for a left-invariant Riemannian metric, one may use Theorem 5.40 to deduce the following formula, which is given as part (b) in Exercise E7.19 in the text:

$$\langle \xi : \eta \rangle_{\mathfrak{g}} = -\mathbb{I}^\sharp\big(\mathrm{ad}^*_\xi \mathbb{I}^\flat(\eta) + \mathrm{ad}^*_\eta \mathbb{I}^\flat(\xi)\big), \tag{S2.6}$$

where $\mathbb{I}$ is the inner product on $\mathfrak{g}$ induced by the left-invariant Riemannian metric, and where $\xi, \eta \in \mathfrak{g}$ (this follows from part (iii) of Theorem 5.40).

The controllability of the system can now be stated, using Theorems 7.36 and 7.40, in purely algebraic terms. The following result characterizes the distributions needed to decide accessibility as per Theorem 7.36.

Proposition S2.8 (Accessibility of affine connection control systems on Lie groups). *Let $\Sigma = (\mathsf{G}, \nabla, \mathcal{D}, \mathscr{Y} = \{Y_, \dots, Y_m\}, U)$ be a left-invariant affine connection control system. The following statements hold:*

(i) $\mathrm{Sym}^{(\infty)}(\mathcal{Y})_e$ *is the smallest subspace of* $\mathfrak{g}$ *containing* $\mathcal{Y}_e$ *and closed under the product* $\langle \cdot : \cdot \rangle_{\mathfrak{g}}$*;*

(ii) $\mathrm{Lie}^{(\infty)}(\mathrm{Sym}^{(\infty)}(\mathcal{Y}))_e$ *is the smallest Lie subalgebra of* $\mathfrak{g}$ *containing* $\mathrm{Sym}^{(\infty)}(\mathcal{Y})_e$*;*

(iii) $\mathrm{Sym}^{(\infty)}(\mathcal{Y})_g = T_e L_g(\mathrm{Sym}^{(\infty)}(\mathcal{Y})_e)$*;*

(iv) $\mathrm{Lie}^{(\infty)}(\mathrm{Sym}^{(\infty)}(\mathcal{Y}))_g = T_e L_g(\mathrm{Lie}^{(\infty)}(\mathrm{Sym}^{(\infty)}(\mathcal{Y}))_e)$*.*

It is also possible to restate Theorem 7.40 for left-invariant systems on Lie groups. We leave the straightforward details to the reader in Exercise E7.20.

Remark S2.9. Some advantages of the methods we propose in this section are the following.

1. They reduce the computations to algebraic manipulations, and so eliminate any differentiations. Additionally, left-invariance allows one to check for accessibility or controllability only at the identity, and the conclusions will then automatically hold at any other configuration.
2. There is no need to introduce coordinates on G, a process which can be cumbersome on Lie groups such as $\mathsf{SO}(3)$ or $\mathsf{SE}(3)$. •

Example S2.10 (Section 7.4.2 cont'd). As we saw in Example 5.47, the planar body with a thruster can be regarded as a left-invariant system on the Lie group $\mathsf{G} = \mathsf{SE}(2)$. The Lie algebra is naturally isomorphic as a vector space to $\mathbb{R} \oplus \mathbb{R}^2 \simeq \mathbb{R}^3$, and we let $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ be the standard basis. The input vector fields Y_1 and Y_2 for the system are left-invariant extensions of $\eta_1 = \frac{1}{m}\boldsymbol{e}_2$ and $\eta_2 = -\frac{h}{J}\boldsymbol{e}_1 + \frac{1}{m}\boldsymbol{e}_3$. In Example 5.47 we give the Lie algebra structure constants in this basis (cf. (5.20)), and using this and (S2.6) we make the following computations:

$$\langle \eta_1 : \eta_1 \rangle_{\mathfrak{se}(2)} = 0, \quad \langle \eta_1 : \eta_2 \rangle_{\mathfrak{se}(2)} = -\frac{h}{mJ}\boldsymbol{e}_3, \quad \langle \eta_2 : \eta_2 \rangle_{\mathfrak{se}(2)} = \frac{2h}{mJ}\boldsymbol{e}_2,$$
$$\langle \eta_2 : \langle \eta_2 : \eta_2 \rangle \rangle_{\mathfrak{se}(2)} = -\frac{2h}{mJ^2}\boldsymbol{e}_3.$$

One may now use Proposition S2.8 and Exercise E7.20 to deduce the controllability conclusions we produced in Section 7.4.2. •

S2.3.2 Controllability of a class of systems on principal fiber bundles

We now look at a simple mechanical control system $\Sigma = (\mathsf{Q}, \mathbb{G}, V, \mathscr{F} = \{F^1, \dots, F^m\}, U)$, and we consider the situation where a Lie group G acts with an action Φ on the configuration manifold Q in such a way that $\pi\colon \mathsf{Q} \to \mathsf{B} = \mathsf{Q}/\mathsf{G}$ is a principal fiber bundle. We also suppose that $\mathbb{G}$ is a left-invariant Riemannian metric on Q. This setup is described in Section 5.4.3, where, in particular, the momentum map $J_\Phi\colon \mathsf{TQ} \to \mathfrak{g}^*$ is defined. We denote by VQ the subbundle of TQ defined by $\mathsf{V}_q\mathsf{Q} = \ker(T_q\pi)$. Thus VQ is the ***vertical subbundle***. The ***horizontal subbundle*** HQ of TQ is the $\mathbb{G}$-orthogonal complement of VQ.

Let us first record a useful fact about the momentum map.

Proposition S2.11 (Characterization of the horizontal subbundle). *The horizontal distribution HQ is smooth, regular, and geodesically invariant distribution on Q, and furthermore, $\mathsf{HQ} = J_\Phi^{-1}(0)$.*

Proof. By definition of J_Φ, it is clear that $J_\Phi(v_q) = 0$ if and only v_q is $\mathbb{G}$-orthogonal to $\xi_{\mathsf{Q}}(q)$ for every $\xi \in \mathfrak{g}$. That is to say, $J_\Phi^{-1}(0)$ is the $\mathbb{G}$-orthogonal complement to VQ. Since VQ is smooth and regular, so too is $J_\Phi^{-1}(0)$. Conservation of momentum translates to $J_\Phi^{-1}(0)$ being a submanifold of TQ that is invariant under the flow of S. The result now follows from fact 1 given in the proof of Theorem 3.108. ■

As usual, we let $\mathscr{Y} = \{Y_1 = \mathbb{G}^\sharp(F^1), \dots, Y_m = \mathbb{G}^\sharp(F^m)\}$ be the input vector fields for the system. We shall suppose that the input vector fields $\mathscr{Y}$

1. take values in the regular distribution $J_\Phi^{-1}(0)$ and
2. are G-invariant.

These assumptions, particularly the first, are quite restrictive. More generally, one can consider vector fields satisfying just the second assumption, although the situation here is not understood in as elegant manner as is likely possible (but see [Cortés, Martínez, Ostrowski, and Zhang 2002]). Note that this renders our treatment in this section "opposite" to that in Section S2.3.1. Indeed, in Section S2.3.1 the regular distribution $J_\Phi^{-1}(0)$ is trivial. Essentially, we deal in this section with input vector fields that are horizontal, whereas the situation of Section S2.3.1 uses input vector fields that are vertical (although the reduced space is a point in Section S2.3.1). The two assumptions we make on the input vector fields $\mathscr{Y}$ ensure that there are vector fields $\mathscr{Y}_{\mathsf{B}} = \{Y_{\mathsf{B},1}, \dots, Y_{\mathsf{B},m}\}$ on B with the property that $T_q\pi(Y_a(q)) = Y_{\mathsf{B},a}(b)$ for each $q \in \pi^{-1}(b)$ and for each $a \in \{1, \dots, m\}$. What is more, we also have a reduced Riemannian metric $\mathbb{G}_{\mathsf{B}}$ on B defined by

$$\mathbb{G}_{\mathsf{B}}(u_b, v_b) = \mathbb{G}(\mathrm{hlft}_q(u_b), \mathrm{hlft}_q(v_b))$$

for some (and so for all) $q \in \pi^{-1}(b)$. We also denote $\mathscr{F}_{\mathsf{B}} = \{F_{\mathsf{B}}^1 = \mathbb{G}_{\mathsf{B}}^\flat(Y_{\mathsf{B},1}), \dots, F_{\mathsf{B}}^m = \mathbb{G}_{\mathsf{B}}^\flat(Y_{\mathsf{B},m})\}$. Assuming that the potential function V is also G-invariant, this defines a function V_{B} on B by $V_{\mathsf{B}}(b) = V(q)$ for some (and so for all) $q \in \pi^{-1}(b)$. We then obtain a reduced simple mechanical control system $\Sigma_{\mathsf{B}} = (\mathsf{B}, \mathbb{G}_{\mathsf{B}}, V_{\mathsf{B}}, \mathscr{F}_{\mathsf{B}}, U)$. We refer the reader to Section 5.5 for related constructions. We wish to ascertain the relationship between the controllability properties of Σ and Σ_{B}.

The following result lists some of the conclusions that may be drawn.

Theorem S2.12 (Controllability of simple mechanical control systems on principal fiber bundles). *Let* $\Sigma = (\mathsf{Q}, \mathbb{G}, V, \mathscr{F}, U)$ *and* $\Sigma_{\mathsf{B}} = (\mathsf{B}, \mathbb{G}_{\mathsf{B}}, V_{\mathsf{B}}, \mathscr{F}_{\mathsf{B}}, U)$ *be as above, and suppose that all data are analytic. The following statement holds:*

(i) if $\dim(\mathsf{G}) > 0$*, then* Σ *is not accessible from any* $q \in \mathsf{Q}$*.*

Now suppose that $\mathrm{Lie}^{(\infty)}(J_{\Phi}^{-1}(0)) = \mathsf{TQ}$ *and that* $V = 0$*. Then the following statements hold:*

(ii) Σ *is configuration accessible from* $q \in \mathsf{Q}$ *if* Σ_{B} *is accessible from* $\pi(q) \in \mathsf{B}$*;*

(iii) Σ *satisfies the hypotheses of Theorem 7.40 at* $q \in \mathsf{Q}$ *if and only if* Σ_{B} *satisfies the hypotheses of Theorem 7.40 at* $\pi(q) \in \mathsf{B}$*.*

Proof. (i) As in the proof of Proposition S2.11, the geodesic spray S is tangent to $J_{\Phi}^{-1}(0) \subset \mathsf{TQ}$. Since the vector fields $\mathscr{Y}$ take values in $J_{\Phi}^{-1}(0)$, the vector fields $\mathrm{vlft}(Y_a)$, $a \in \{1, \dots, m\}$, are also tangent to $J_{\Phi}^{-1}(0)$ by fact 3 given in the proof of Theorem 3.108. Since V is G-invariant, V is constant on fibers of $\pi\colon \mathsf{Q} \to \mathsf{B}$. Therefore, $\mathrm{d}V \in \Gamma^{\infty}(\mathrm{ann}(\mathsf{VQ}))$, and from this it follows that $\mathbb{G}^{\sharp} \circ \mathrm{d}V \in \Gamma^{\infty}(J_{\Phi}^{-1}(0))$. Again, using fact 3 given in the proof of Theorem 3.108, it follows that $\mathrm{vlft}(\mathrm{grad}V)$ is tangent to $J_{\Phi}^{-1}(0)$. Therefore, the control-affine system corresponding to Σ has $J_{\Phi}^{-1}(0)$ as an invariant submanifold, precluding the possibility of accessibility, unless $J_{\Phi}^{-1}(0) = \mathsf{TQ}$. However, this can happen only if $\mathsf{VQ} = Z(\mathsf{TQ})$, from which the result follows.

(ii) By Theorem 7.36, Σ_{B} is accessible from $b \in \mathsf{B}$ if and only if $\mathrm{Sym}^{(\infty)}(\mathcal{Y}_{\mathsf{B}}) = \mathsf{T}_b\mathsf{B}$. Since the vector fields $\mathscr{Y}$ are G-invariant, and since symmetric products commute with Φ_g for each $g \in \mathsf{G}$, we have

$$T_q\pi(\mathrm{Sym}^{(\infty)}(\mathcal{Y})_q) = \mathrm{Sym}^{(\infty)}(\mathcal{Y}_{\mathsf{B}})_b$$

for each $q \in \pi^{-1}(b)$. Therefore, Σ_{B} is accessible at b if and only if $\mathrm{Sym}^{(\infty)}(\mathcal{Y})_q = J_{\Phi}^{-1}(0)_q$ for each $q \in \pi^{-1}(b)$. From the assumption that $\mathrm{Lie}^{(\infty)}(J_{\Phi}^{-1}(0)) = \mathsf{TQ}$, this part of the result now follows.

(iii) We introduce some terminology that makes connections with the work on reduction described in [Marsden 1992]. The computations here will also be seen to be strongly reminiscent of the computations of Section 5.5. The ***mechanical connection*** is the map $A_{\mathbb{G}}\colon \mathsf{TQ} \to \mathfrak{g}$ defined by

$$\mathbb{G}((A_{\mathbb{G}}(v_q))_{\mathsf{Q}}(q), \xi_{\mathsf{Q}}(q)) = \mathbb{G}(v_q, \xi_{\mathsf{Q}}(q)), \qquad v_q \in \mathsf{T}_q\mathsf{Q}.$$

Thus $A_{\mathbb{G}}(v_q)$ returns the Lie algebra element whose infinitesimal generator at q is the vertical component, relative to $\mathbb{G}$-orthogonality, of v_q. The ***curvature*** of the mechanical connection is the map $B_{\mathbb{G}}\colon \mathsf{TQ} \times \mathsf{TQ} \to \mathfrak{g}$ defined by

$$B_{\mathbb{G}}(u_q, v_q) = -A_{\mathbb{G}}([\mathrm{hor}(U), \mathrm{hor}(V)](q)),$$

with $\mathrm{hor}\colon \mathsf{TQ} \to \mathsf{TQ}$ being the $\mathbb{G}$-orthogonal projection onto $J_\Phi^{-1}(0)$, and where U and V are arbitrary vector fields extending u_q and v_q, respectively. One can show that

$$\overset{\mathbb{G}}{\nabla}_{Y_a} Y_b = \mathrm{hlft}\big(\overset{\mathbb{G}_{\mathsf{B}}}{\nabla}_{Y_{\mathsf{B},a}} Y_{\mathsf{B},b}\big) - \tfrac{1}{2} B_{\mathbb{G}}(Y_a, Y_b)_{\mathsf{Q}}.$$

Since $B_{\mathbb{G}}$ is skew-symmetric, we have

$$\langle Y_a : Y_b \rangle = \mathrm{hlft}(\langle Y_{\mathsf{B},a} : Y_{\mathsf{B},b} \rangle).$$

From this the result immediately follows. ∎

To illustrate the theorem, we consider the robotic leg example considered in the text at various points.

Example S2.13 (Example 5.71 and Section 7.4.1 cont'd). The robotic leg whose controllability was considered in Section 7.4.1 is a system of the type considered in this section. To illustrate this, we need to specify the symmetry group and its action on Q. We take $\mathsf{G} = \mathsf{SO}(2)$ and the action of G on $\mathsf{Q} = \mathbb{R}_+ \times \mathbb{S}^1 \times \mathbb{S}^1$ given by $\Phi(\boldsymbol{R}, (r, \boldsymbol{x}, \boldsymbol{y})) = (r, \boldsymbol{R}\boldsymbol{x}, \boldsymbol{R}\boldsymbol{y})$. Recall the coordinates (r, θ, ψ) used for this example. The quotient space $\mathsf{B} = \mathsf{Q}/\mathsf{G}$ is diffeomorphic to $\mathbb{R} \times \mathbb{S}^1$, and we identity the orbit through the point (r, θ, ψ) with $(r, \theta - \psi)$. That is to say, we use (r, ϕ) as coordinates for B, and the projection $\pi\colon \mathsf{Q} \to \mathsf{B}$ is given by $\pi(r, \theta, \psi) = (r, \theta - \psi)$. For $a \in \mathfrak{g} \simeq \mathbb{R}$, the infinitesimal generator is

$$a_{\mathsf{Q}}(r, \theta, \psi) = a\Big(\frac{\partial}{\partial \theta} + \frac{\partial}{\partial \psi}\Big).$$

We let $\{\boldsymbol{e}_1\}$ be the standard basis for $\mathfrak{g} \simeq \mathbb{R}$ with $\{\boldsymbol{e}^1\}$ the dual basis. With this, a direct computation yields

$$J_\Phi((r, \theta, \psi), (v_r, v_\theta, v_\psi)) = (J v_\psi + m r^2 v_\theta)\boldsymbol{e}^1,$$
$$\mathbb{G}_{\mathsf{B}}(r, \phi)((u_r, u_\phi), (v_r, v_\phi)) = m u_r v_r + \frac{J m r^2}{J + m r^2} u_\phi v_\phi.$$

The input vector fields Y_1 and Y_2 are horizontal lifts of the vector fields

$$Y_{\mathsf{B},1} = \frac{J + m r^2}{J m r^2} \frac{\partial}{\partial \phi}, \quad Y_{\mathsf{B},2} = \frac{\partial}{\partial r}$$

on B. Thus this system does indeed fit into the class of systems discussed in this section, and the results of Theorem S2.12 may be applied. To apply the results, we need a few computations. The nonzero Christoffel symbols for the Levi-Civita connection on B associated with $\mathbb{G}_{\mathsf{B}}$ are

$$\Gamma^r_{\phi\phi} = -\frac{J^2 r}{(J+mr^2)^2}, \quad \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = \frac{J}{r(J+mr^2)}.$$

This then gives the symmetric products

$$\langle Y_{\mathsf{B},1} : Y_{\mathsf{B},1}\rangle = -\frac{2}{m^2 r^3}\frac{\partial}{\partial r}, \quad \langle Y_{\mathsf{B},1} : Y_{\mathsf{B},2}\rangle = -\frac{2J}{mr(J+mr^2)}\frac{\partial}{\partial r},$$
$$\langle Y_{\mathsf{B},2} : Y_{\mathsf{B},2}\rangle = -\frac{2J^2 r}{(J+mr^2)^2}\frac{\partial}{\partial r}.$$

We also recall that

$$[Y_1, Y_2] = -\frac{2}{m^2 r^3}\frac{\partial}{\partial \theta},$$

from which we conclude that $\mathrm{Lie}^{(\infty)}(J_\Phi^{-1}(0)) = \mathsf{TQ}$.

From Theorem S2.12 we draw the following conclusions.

1. *All cases:* By part (i) of Theorem S2.12, we conclude that the system is not accessible from any $q \in \mathsf{Q}$.
2. Y_1 *only:* Since $\{Y_{\mathsf{B},1}, \langle Y_{\mathsf{B},1} : Y_{\mathsf{B},1}\rangle\}$ generate TB, we deduce from part (ii) of Theorem S2.12 that Σ is configuration accessible from any $q \in \mathsf{Q}$.

   The hypotheses of Theorem 7.40 do not hold for Σ_{B}, and so they also do not hold for Σ. Also, Σ_{B} can be directly seen to be neither STLC nor STLCC at any $b \in \mathsf{B}$. This alone does not allow us, using the results that we give here, to deduce anything concerning the controllability or configuration controllability of Σ. However, we have already seen that it is neither STLC nor STLCC from any $q \in \mathsf{Q}$.
3. Y_2 *only:* Σ_{B} is not accessible, so we conclude from part (ii) of Theorem S2.12 that Σ is not configuration accessible from any $q \in \mathsf{Q}$. This, of course, also precludes Σ from being STLC or STLCC.
4. Y_1 *and* Y_2 *only:* Since Σ_{B} is fully actuated, we deduce from part (ii) of Theorem S2.12 that Σ is configuration accessible from each $q \in \mathsf{Q}$. Also, since Σ_{B} is fully actuated, it satisfies the hypotheses of Theorem 7.40 for each $b \in \mathsf{B}$. Thus we conclude that Σ is STLCC from each $q \in \mathsf{Q}$.

These conclusions, of course, agree with the conclusions of Section 7.4.1. •

S3

Linearization and stabilization of relative equilibria

In Section 6.3 of the text, we considered the stability of relative equilibria. In this supplement, we apply this analysis, in combination with the analysis of Section 10.4 on nonlinear potential shaping, to obtain some results concerning the stabilization of relative equilibria. In the literature, this problem has been addressed by, for example, Bloch, Leonard, and Marsden [2001], Jalnapurkar and Marsden [1998, 2000] and Bullo [2000]. Papers dealing with applications include the work on underwater vehicles by Woolsey and Leonard [2004]. We do not discuss here systems with constraints, although preliminary work in this area includes [Zenkov, Bloch, and Marsden 2002].

Let us outline what we do in this supplement. Since we wish to consider stabilization using linear techniques, in Section S3.1 we consider the linearization around a relative equilibrium. The development here is significantly more complicated than linearization about equilibrium configurations. For example, it utilizes the Ehresmann connections of Section S1.3 that give rise to the Jacobi equation. The processes of reduction and linearization produce various natural energies, and these are related in Section S3.2. Linear stability of relative equilibria is studied in Section S3.3. In Section S3.4 we define the natural stabilization problems for relative equilibria. In Section S3.5 we then discuss stabilization of relative equilibria using linear and potential shaping techniques. The results in this last section are analogous to those in Sections 10.3 and 10.4 for the stabilization of equilibrium configurations. There are, however, some additional complications arising from the extra structure of the relative equilibrium.

S3.1 Linearization along relative equilibria

This long section contains many of the essential results in the paper. Our end objective is to relate the linearization of equilibria of the reduced equations of Theorem 5.83 to the linearization along the associated *relative* equilibria of the unreduced system. We aim to perform this linearization in a control theoretic setting so that our constructions will be useful not just for investigations of stability, but also for stabilization. To properly understand this process, we begin in Section S3.1.1 with linearization of general control-affine systems about general controlled trajectories. Then we specialize this discussion in Section S3.1.2 to the linearization of a general affine connection control system about a general controlled trajectory. We then, in Section S3.1.3, finally specialize to the case of interest, namely the linearization of the unreduced equations along a relative equilibrium. Here the main result is Theorem S3.11 which gives the geometry associated with linearization along a relative equilibrium. In particular this result, or more precisely its proof, makes explicit the relationship between the affine differential geometric concepts arising in the linearization of Section S3.1.2 and the usual concepts arising in reduction of mechanical systems, such as the curvature of the mechanical connection and the effective potential. Then, in Section S3.1.4 we turn to the linearization, in the standard sense, of the reduced equations about an equilibrium. The main result here is Theorem S3.14 which links the reduced and unreduced linearizations.

We let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-forced simple mechanical control system with F time-independent, X be a complete infinitesimal symmetry of Σ, and $\chi\colon \mathbb{R} \to \mathsf{Q}$ be a relative equilibrium for $(\mathsf{Q}, \mathbb{G}, V, F)$. Thus χ is an integral curve for X that is also an uncontrolled trajectory for the system. We let B denote the set of X-orbits, and following Assumption 5.78, we assume that B is a smooth manifold for which $\pi_{\mathsf{B}}\colon \mathsf{Q} \to \mathsf{B}$ is a surjective submersion. We let $Y_a = \mathbb{G}^\sharp \circ F^a$, $a \in \{1, \dots, m\}$, and if $u\colon I \to \mathbb{R}^m$ is a locally integrable control, we denote

$$Y_u(t,q) = \sum_{a=1}^m u^a(t) Y_a(q), \qquad t \in I,\ q \in \mathsf{Q}, \tag{S3.1}$$

for brevity. Define vector fields $Y_{\mathsf{B},a}$, $a \in \{1, \dots, m\}$, by $Y_{\mathsf{B},a}(b) = T_q\pi_{\mathsf{B}}(Y_a(q))$ for $q \in \pi_{\mathsf{B}}^{-1}(b)$. Since the vector fields $\mathscr{Y}$ are X-invariant, this definition is independent of $q \in \pi_{\mathsf{B}}^{-1}(b)$. Similarly to (S3.1), we denote

$$Y_{\mathsf{B},u}(t,b) = \sum_{a=1}^m u^a(t) Y_{\mathsf{B},a}(b).$$

In Theorem 5.83, along with Remark 5.84, we showed that the reduced system has $\mathsf{TB} \times \mathbb{R}$ as its state space, and satisfies the equations

$$\begin{aligned}
\overset{\mathbb{G}_{\mathsf{B}}}{\nabla}_{\eta'(t)}\eta'(t) &= -\mathrm{grad}_{\mathsf{B}}\big(V^{\mathrm{eff}}_{X,v(t)}\big)_{\mathsf{B}}(\eta(t)) + v(t)C_X(\eta'(t)) \\
&\qquad + T\pi_{\mathsf{B}}\circ\mathbb{G}^{\sharp}\circ F(\gamma'(t)) + Y_{\mathsf{B},u}(t,\eta(t)), \\
\dot{v}(t) &= -\frac{v(t)\langle \mathrm{d}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t));\eta'(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} + \frac{\langle F(\gamma'(t)); X(\gamma(t))\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} \\
&\qquad + \frac{\mathbb{G}(Y_u(t,\gamma(t)), X(\gamma(t)))}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))},
\end{aligned} \tag{S3.2}$$

where (γ, u) is the controlled trajectory on Q, $\eta = \pi_{\mathsf{B}}\circ\gamma$, and v is defined by $\mathrm{ver}(\gamma'(t)) = v(t)X(\gamma(t))$. As we saw in the proof of Theorem 6.56, the relative equilibrium χ corresponds to the equilibrium point $(T\pi_{\mathsf{B}}(\chi'(0)), 1)$ of the reduced equations (S3.2). Therefore, linearization of the relative equilibrium χ could be *defined* to be the linearization of the equations (S3.2) about the equilibrium point $(T\pi_{\mathsf{B}}(\chi'(0)), 1)$. This is one view of linearization of relative equilibria. Another view is that, since χ is a trajectory for the unreduced system, we could linearize along it in the manner described in Section S1.3.1 when describing the Jacobi equation. In this section we shall see how these views of linearization of relative equilibria tie together. We build up to this by first considering linearization in more general settings.

S3.1.1 Linearization of a control-affine system along a controlled trajectory

In order to talk about linearization along a relative equilibrium, we first discuss linearization along a general controlled trajectory. In order to do this, it is convenient to first consider the general control-affine case, then specialize to the mechanical setting.

First we need to extend our notion of a control-affine system to be time-dependent. This is because the linearization of a control-affine system will generally be time-dependent.

Definition S3.1 (Time-dependent control-affine system). For $r \in \mathbb{N}\cup\{\infty\}\cup\{\omega\}$, a ***$C^r$-time-dependent control-affine system*** is a triple $(\mathsf{M}, \mathscr{C} = \{f_0, f_1, \dots, f_m\}, U)$ where

(i) M is a C^r-manifold,
(ii) $f_0, f_1, \dots, f_m$ are locally integrally class C^r-vector fields on M, and
(iii) $U \subset \mathbb{R}^m$. •

The governing equations for a time-dependent control-affine system $(\mathsf{M}, \mathscr{F}, U)$ are then

$$\gamma'(t) = f_0(t,\gamma(t)) + \sum_{a=1}^{m} u^a(t) f_a(t,\gamma(t)).$$

Clearly, the only difference from our usual notion of a control-affine system is the dependence of the drift vector field and the control vector fields on time. Many of the same notions one has for control-affine systems, controlled trajectories, controllability, stability, etc., carry over to the time-dependent setting with appropriate modifications. However, we shall only use this more general time-dependent setup in a fairly specific manner.

Suppose that we have a (time-independent) control-affine system $(\mathsf{M}, \mathscr{F} = \{f_0, f_1, \dots, f_m\}, U)$ of class C^∞, and a controlled trajectory (γ_0, u_0) defined on an interval I. Let us take $U = \mathbb{R}^m$ to avoid unnecessary complication. We wish to linearize the system about this controlled trajectory. Linearization is to be done with respect to both state and control. Thus, speaking somewhat loosely for a moment, to compute the linearization, one should first fix the control at u_0 and linearize with respect to state, then fix the state and linearize with respect to control, and then add the results to obtain the linearization. Let us now be more formal about this.

If we fix the control at u_0, we obtain the LIC$^\infty$-vector field f_{u_0} on M defined by

$$f_{u_0}(t, x) = f_0(x) + \sum_{a=1}^{m} u_0^a(t) f_a(x).$$

We call f_{u_0} the ***reference vector field*** for the controlled trajectory (γ_0, u_0). The linearization of the reference vector field is exactly described by its tangent lift, as discussed in Remark S1.10–4. Thus one component of the linearization is $f_{u_0}^T$. The other component is computed by fixing the state, say at x, and linearizing with respect to the control. Thus we consider the map

$$\mathbb{R} \times \mathbb{R}^m \ni (t, u) \mapsto f_0(x) + \sum_{a=1}^{m} (u_0^a(t) + u^a) f_a(x) \in \mathsf{T}_x\mathsf{M},$$

and differentiate this with respect to u at $u = \mathbf{0}$. The resulting map from $\mathsf{T}_{\mathbf{0}}\mathbb{R}^m \simeq \mathbb{R}^m$ to $\mathsf{T}_{f_{u_0}(t,x)}(\mathsf{T}_x\mathsf{M}) \simeq \mathsf{T}_x\mathsf{M}$ is simply given by

$$u \mapsto \sum_{a=1}^{m} u^a f_a(x).$$

In order to add the results of the two computations, we regard $\mathsf{T}_x\mathsf{M}$ as being identified with $\mathsf{V}_{f_{u_0}(t,x)}\mathsf{TM}$. Thus the linearization with respect to the control yields the linearized control vector fields $\mathrm{vlft}(f_a)$, $a \in \{1, \dots, m\}$. In this way, we arrive at the C^∞-time-dependent control-affine system $\Sigma^T(\gamma_0, u_0) = (\mathsf{TM}, \{f_{u_0}^T, \mathrm{vlft}(f_1), \dots, \mathrm{vlft}(f_a)\}, \mathbb{R}^m)$, whose controlled trajectories (ξ, u) satisfy

$$\xi'(t) = f_{u_0}^T(t, \xi(t)) + \sum_{a=1}^{m} u^a(t) \mathrm{vlft}(f_a)(\xi(t)). \tag{S3.3}$$

The following result gives an important property of these controlled trajectories.

Lemma S3.2. *For every locally integrable control $t \mapsto u(t)$, the* LIC$^\infty$*-vector field*

$$(t, v_x) \mapsto f^T_{u_0}(t, v_x) + \sum_{a=1}^{m} u^a(t)\mathrm{vlft}(f_a)(v_x)$$

is a linear vector field over f_{u_0}.

Proof. This is easily proved in coordinates. ■

From our discussion of linear vector fields in Section S1.3.4, we then know that, if (ξ, u) is a controlled trajectory for $\Sigma^T(\gamma_0, u_0)$, then $\pi_{\mathsf{TM}} \circ \xi$ is an integral curve of f_{u_0}. In particular, if (ξ, u) is a controlled trajectory for $\Sigma^T(\gamma_0, u_0)$ that satisfies $\pi_{\mathsf{TM}} \circ \xi(t) = \gamma_0(t)$ for some $t \in I$, then ξ is a vector field along γ_0.

To formally define the linearization along (γ_0, u_0), we need an additional concept, following Sussmann [1997].

Definition S3.3 (Linear differential operator along a curve). Let M be a C^r-manifold, $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$, let $\gamma\colon I \to \mathsf{M}$ be an LAC curve, and let $\pi\colon \mathsf{E} \to \mathsf{M}$ be a vector bundle. A ***linear differential operator*** in E along γ assigns, to each LAC section ξ of E along γ, a locally integrable section $\mathscr{L}(\xi)$ along γ, and the assignment has the property that it is linear and, if $f \in C^r(\mathsf{M})$ and if $\Xi \in \Gamma^r(\mathsf{E})$, then

$$\mathscr{L}((f \circ \gamma(\Xi \circ \gamma))(t) = f \circ \gamma(t)\mathscr{L}(\Xi \circ \gamma)(t) + (\mathscr{L}_{\gamma'(t)} f)(\gamma(t))\Xi \circ \gamma(t). \qquad \bullet$$

Thus a linear differential operator simply "differentiates" sections of E along γ, with the differentiation rule satisfying the usual derivation property with respect to multiplication with respect to functions. Sussmann [1997] shows that, in coordinates $(x^1, \dots, x^n)$ for M, if $t \mapsto (\xi^1(t), \dots, \xi^k(t))$ are the fiber components of the local representative of an LAC section ξ of E, then the fiber components of the local representative of $\mathscr{L}(\xi)$ satisfy

$$(\mathscr{L}(\xi))^a(t) = \dot{\xi}^a(t) + \sum_{b=1}^{k} L^a_b(t)\xi^b(t), \qquad a \in \{1, \dots, k\},$$

for some locally integrable functions $t \mapsto L^b_a(t)$, $a, b \in \{a, \dots, k\}$. If $\gamma\colon I \to \mathsf{M}$ is an integral curve of an LIC^r-vector field X, then there is a naturally induced linear differential operator in TM along γ, denoted by $\mathscr{L}^{X,\gamma}$, and defined by

$$\mathscr{L}^{X,\gamma}(\xi) = [X_t, \Xi](\gamma(t)), \qquad \text{a.e. } t \in I,$$

where Ξ is a vector field satisfying $\xi = \Xi \circ \gamma$, and where X_t is the C^r-vector field defined by $X_t(x) = X(t, x)$. In coordinates this linear differential operator satisfies

$$\mathscr{L}^{X,\gamma}(\xi)^i(t) = \dot{\xi}^i(t) - \frac{\partial X^i}{\partial x^j}(\gamma(t))\xi^j(t), \qquad i \in \{1, \dots, n\}.$$

This linear differential operator is sometimes referred to as the "Lie drag" (see [Crampin and Pirani 1986, Section 3.5]).

A coordinate computation readily verifies the following result, and we refer to [Lewis and Tyner 2003, Sussmann 1997] for details.

Proposition S3.4 (Relationship between tangent lift and a linear differential operator). *Let $X \colon I \times M \to \mathsf{T}M$ be an* LIC^r *vector field, let $v_{x_0} \in \mathsf{T}_x M$, let $t_0 \in I$, and let $\gamma \colon I \to M$ be the integral curve of X satisfying $\gamma(t_0) = x_0$. For a vector field ξ along γ satisfying $\xi(t_0) = v_{x_0}$, the following statements are equivalent:*

(i) ξ is an integral curve for X^T;
(ii) there exists a variation σ of X along γ such that $\frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0}\sigma(t,s) = \xi(t)$ for each $t \in I$;
(iii) $\mathscr{L}^{X,\gamma}(\xi) = 0$.

With the preceding as motivation, we can define the linearization of a control-affine system.

Definition S3.5 (Linearization of a control-affine system about a controlled trajectory). Let $\Sigma = (\mathsf{M}, \mathscr{F} = \{f_0, f_1, \dots, f_m\}, \mathbb{R}^m)$ be a C^r-control-affine system with (γ_0, u_0) a controlled trajectory. The ***linearization*** of Σ about (γ_0, u_0) is given by $\{\mathscr{L}_\Sigma(\gamma_0, u_0), b_{\Sigma,1}(\gamma_0, u_0), \dots, b_{\Sigma,m}(\gamma_0, u_0)\}$, where

(i) $\mathscr{L}_\Sigma(\gamma_0, u_0)$ is the linear differential operator in TM along γ_0 defined by

$$\mathscr{L}_\Sigma(\gamma_0, u_0) = \mathscr{L}^{f_{u_0},\gamma_0},$$

and

(ii) $b_{\Sigma,a}$, $a \in \{1, \dots, m\}$, are the vector fields along γ_0 defined by

$$b_{\Sigma,a}(\gamma_0, u_0)(t) = \mathrm{vlft}(f_a(\gamma_0(t))), \qquad a \in \{1, \dots, m\}. \qquad \bullet$$

The equations governing the linearization are

$$\mathscr{L}_\Sigma(\gamma_0, u_0)(\xi)(t) = \sum_{a=1}^{m} u^a(t) b_{\Sigma,a}(\gamma_0, u_0),$$

which are thus equations for a vector field ξ along γ_0. By Proposition S3.4, these equations are exactly the restriction to image(γ_0) of the equations for the time-dependent control-affine system in (S3.3). In the special case where $f_0(x_0) = 0_{x_0}$, $u_0 = \mathbf{0}$, $\gamma_0 = x_0$ for some $x_0 \in \mathsf{M}$, one can readily check, along the lines of Proposition 3.75, that we recover the linearization of the system at x_0 as per Definition 7.27.

S3.1.2 Linearization of a forced affine connection control system along a controlled trajectory

After beginning our discussion of linearization in the context of control-affine systems, we next specialize to affine connection control systems. We let $\Sigma = (\mathsf{Q}, \nabla, Y, \mathscr{Y}, \mathbb{R}^m)$ be a C^∞-forced affine connection control system. In this section, we make the following assumption about the external force Y.

Assumption S3.6 (Form of external force for linearization along a controlled trajectory). Assume that the vector force Y is time-independent and decomposable as $Y(v_q) = \overline{Y}_0(q) + \overline{Y}_1(v_q)$, where $\overline{Y}_0$ is a basic vector force and where $\overline{Y}_1$ is a $(1,1)$-tensor field. •

This assumption will allow us to model potential forces and Rayleigh dissipative forces when we discuss stabilization in subsequent sections. The governing equations for the system are

$$\nabla_{\gamma'(t)}\gamma'(t) = \overline{Y}_0(\gamma(t)) + \overline{Y}_1(\gamma'(t)) + \sum_{a=1}^{m} u^a(t) Y_a(\gamma(t)).$$

To linearize these equations about *any* controlled trajectory (γ_0, u_0), following the development in the preceding section, we first need to compute the tangent lift for the LIC$^\infty$-vector field S_{u_0} on TQ defined by

$$S_{u_0}(t, v_q) = S(v_q) + \mathrm{vlft}(\overline{Y}_0(q) + \overline{Y}_1(v_q) + Y_{u_0}(t,q)),$$

where Y_{u_0} is defined as in (S3.1). The Jacobi equation, as we have seen in Theorem S1.34, contains the essential features of the tangent lift of S. Furthermore, the computations of Lemma S1.33 allow us to determine what needs to be added to the Jacobi equation to include the external and control forces. We recall the notation from Section S1.3.10 where points in TTQ are written as $u_{v_q} \oplus w_{v_q}$, relative to the splitting defined by the Ehresmann connection on $\pi_{\mathsf{TTQ}} \colon \mathsf{TTQ} \to \mathsf{TQ}$. The following result gives the linearization along (γ_0, u_0) using the Ehresmann connection of Section S1.3.8.

Proposition S3.7 (State linearization of an affine connection control system). *Let $\Sigma = (\mathsf{Q}, \nabla, Y, \mathscr{Y}, U)$ be a C^∞-forced simple mechanical control system where Y satisfies Assumption S3.6, let (γ_0, u_0) be a controlled trajectory for Σ defined on I, and let $t \mapsto \Upsilon_0(t) = \gamma_0'(t)$ be the tangent vector field of γ_0. For $a \in I$, let $u, w \in \mathsf{T}_{\gamma_0(a)}\mathsf{Q}$, and define vector fields $U, W \colon I \to \mathsf{TQ}$ along γ_0 by asking that $t \mapsto U(t) \oplus W(t) \in \mathsf{T}_{\gamma_0(t)}\mathsf{Q} \oplus \mathsf{T}_{\gamma_0(t)}\mathsf{Q} \simeq \mathsf{T}_{\Upsilon_0(t)}\mathsf{TQ}$ be the integral curve of $S_{u_0}^T$ with initial conditions $u \oplus w \in \mathsf{T}_{\gamma_0(a)}\mathsf{Q} \oplus \mathsf{T}_{\gamma_0(a)}\mathsf{Q} \simeq \mathsf{T}_{\Upsilon_0(a)}\mathsf{TQ}$. Then U and W satisfy the equations*

$$\begin{aligned}
&W(t) = \nabla_{\gamma_0'(t)} U(t) + \tfrac{1}{2} T(U(t), \gamma_0'(t)),\\
&\nabla^2_{\gamma_0'(t)} U(t) + R(U(t), \gamma_0'(t))\gamma_0'(t) + \nabla_{\gamma_0'(t)}(T(U(t), \gamma_0'(t)))\\
&\qquad = \nabla_{U(t)}(\overline{Y}_0 + Y_{u_0})(\gamma_0(t)) + (\nabla_{U(t)}\overline{Y}_1)(\gamma_0'(t)) + \overline{Y}_1(\nabla_{\gamma_0'(t)} U(t)).
\end{aligned}$$

Proof. Let us denote $X_{u_0} = \overline{Y}_0 + Y_{u_0}$, for brevity. A computation in coordinates readily shows that the tangent lift of the vertical lift of X_{u_0} is given by

$$\mathrm{vlft}(Y_{u_0})^T(u_{v_q} \oplus w_{v_q}) = 0 \oplus X_{u_0}(q) \oplus 0 \oplus (\nabla X_{u_0}(u_{v_q}) + \tfrac{1}{2}T(X_{u_0}(q), u_{v_q})).$$

A coordinate computation also gives

$$\begin{aligned}\mathrm{vlft}(\overline{Y}_1)^T(u_{v_q} \oplus w_{v_q}) = 0 \oplus \overline{Y}_1(v_q) \oplus 0 \\ \oplus \big(\nabla_{u_{v_q}}\overline{Y}_1(v) + \overline{Y}_1(w_{v_q}) + \tfrac{1}{2}T(\overline{Y}_1(v_q), u_{v_q}) + \tfrac{1}{2}\overline{Y}_1(T(u_{v_q}, v_q))\big).\end{aligned}$$

The tangent lift of S is given by Propositions S1.29 and S1.32 as

$$\begin{aligned}S^T(u_{v_q} \oplus w_{v_q}) = v_q \oplus 0 \oplus w_{v_q} \oplus (-R(u_{v_q}, v_q)v_q - \tfrac{1}{2}(\nabla_{v_q}T)(u_{v_q}, v_q) \\ + \tfrac{1}{4}T(T(u_{v_q}, v_q), v_q)).\end{aligned}$$

Thus, using Lemma S1.33, we have that U and W satisfy

$$\begin{aligned}&\nabla_{\gamma_0'(t)}U(t) + \tfrac{1}{2}T(U(t), \gamma_0'(t)) = W(t),\\ &\nabla_{\gamma_0'(t)}W(t) + \tfrac{1}{2}T(W(t), \gamma_0'(t)) = -R(U(t), \gamma_0'(t))\gamma_0'(t)\\ &\qquad - \tfrac{1}{2}(\nabla_{\gamma_0'(t)}T)(U(t), \gamma_0'(t)) + \tfrac{1}{4}T(T(U(t), \gamma_0'(t)), \gamma_0'(t))\\ &\qquad + \nabla X_{u_0}(U(t)) + \tfrac{1}{2}T(X_{u_0}(t, \gamma_0(t)), U(t))\\ &\qquad + \nabla_{U(t)}\overline{Y}_1(\gamma_0'(t)) + \overline{Y}_1(W(t)) + \tfrac{1}{2}T(\overline{Y}_1(\gamma_0'(t)), U(t))\\ &\qquad + \tfrac{1}{2}\overline{Y}_1(T(U(t), \gamma_0'(t))).\end{aligned}$$

The first of the equations is the first equation in the statement of the proposition. Differentiating this first equation, and substituting the second, gives the second equation in the statement of the proposition, after some simplification. ■

Next we linearize with respect to the controls. This is simpler, and following the procedure in the preceding section gives the control vector fields $\mathrm{vlft}(\mathrm{vlft}(Y_a))$, $a \in \{1, \dots, m\}$. Thus, we arrive at the time-dependent control-affine system $\Sigma^T(\gamma_0, u_0) = (\mathsf{TTQ}, \{S^T_{u_0}, \mathrm{vlft}(\mathrm{vlft}(Y_1)), \dots, \mathrm{vlft}(\mathrm{vlft}(Y_m))\}, \mathbb{R}^m)$. With respect to the splitting defined by the Ehresmann connection associated with ∇, it is easy to verify that

$$\mathrm{vlft}(\mathrm{vlft}(Y_a))(u_q \oplus w_q) = 0 \oplus 0 \oplus 0 \oplus Y_a(q).$$

If we write a controlled trajectory for $\Sigma^T(\gamma_0, u_0)$ as $(U \oplus W, u)$, reflecting the notation of Proposition S3.7, we see that the following equations govern this trajectory:

$$\begin{aligned}
&W(t) = \nabla_{\gamma_0'(t)}U(t) + \tfrac{1}{2}T(U(t),\gamma_0'(t)),\\
&\nabla^2_{\gamma_0'(t)}U(t) + R(U(t),\gamma_0'(t))\gamma_0'(t) + \nabla_{\gamma_0'(t)}(T(U(t),\gamma_0'(t)))\\
&\qquad = \nabla(\overline{Y}_0 + Y_{u_0})(U(t)) + (\nabla_{U(t)}\overline{Y}_1)(\gamma_0'(t)) + \overline{Y}_1(\nabla_{\chi'(t)}U(t))\\
&\qquad\quad + \sum_{a=1}^m u^a(t)Y_a(\gamma_0(t)).
\end{aligned}$$

With the above as backdrop, we make the following definition, and in so doing, hope the reader will forgive our using the same notation as was used for control-affine systems.

Definition S3.8 (Linearization of affine connection control system about a controlled trajectory). Let $\Sigma = (\mathsf{Q},\nabla,Y,\mathscr{Y},\mathbb{R}^m)$ be a C^∞-forced affine connection control system where Y satisfies Assumption S3.6, and let (γ_0,u_0) be a controlled trajectory. The ***linearization*** of Σ about (γ_0,u_0) is given by $\{A_\Sigma(\gamma_0,u_0), b_{\Sigma,1}(\gamma_0,u_0),\dots,b_{\Sigma,m}(\gamma_0,u_0)\}$, where

(i) $A_\Sigma(\gamma_0,u_0)$ is the linear differential operator in TQ along γ_0 defined by

$$\begin{aligned}
A_\Sigma(\gamma_0,u_0)(t)\cdot\xi(t) &= R(\xi(t),\gamma_0'(t))\gamma_0'(t) + \nabla_{\gamma_0'(t)}(T(\xi(t),\gamma_0'(t)))\\
&-\nabla_{\xi(t)}\overline{Y}_0(\gamma_0(t)) - \nabla_{\xi(t)}Y_{u_0}(t,\gamma_0(t)) + (\nabla_{\xi(t)}\overline{Y}_1)(\gamma_0'(t)) + \overline{Y}_1(\nabla_{\gamma_0'(t)}\xi(t)),
\end{aligned}$$

and

(ii) $b_{\Sigma,a}(\gamma_0,u_0)$, $a\in\{1,\dots,m\}$, are vector fields along γ_0 defined by

$$b_{\Sigma,a}(\gamma_0,u_0)(t) = Y_a(\gamma_0(t)).$$ •

The equations governing the linearization are then

$$\nabla^2_{\gamma_0'(t)}\xi(t) + A_\Sigma(\gamma_0,u_0)(t)\cdot\xi(t) = \sum_{a=1}^m u^a(t)Y_a(\gamma_0(t)). \tag{S3.4}$$

In particular, a ***controlled trajectory*** for the linearization of Σ along (γ_0,u_0) is a pair (ξ,u), where $u\colon I\to\mathbb{R}^m$ is a locally integrable control, and where $\xi\colon I\to\mathsf{TQ}$ is the LAD curve along γ_0 satisfying (S3.4).

Remarks S3.9. 1. Note that the structure of the Ehresmann connection induced by ∇ allows us to use a linear differential operator along γ_0 rather than along γ_0'.

2. If ∇ is torsion-free and if $\overline{Y}_1 = 0$, then $A_\Sigma(\gamma_0,u_0)$ is no longer a linear differential operator, but is actually a $(1,1)$-tensor field. In such a case, it is still possible to consider this as a linear differential operator, but one of "order zero." •

S3.1.3 Linearization of the unreduced equations along a relative equilibrium

With the work done in the preceding two sections, it is easy to give the form of the linearization along a relative equilibrium. We let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-forced simple mechanical control system. In this and the next section, we make the following assumption about the external force F.

Assumption S3.10 (Form of external force for linearization along a relative equilibrium). Assume that the force F is time-independent and that $F(v_q) = A^\flat(v_q - X(q))$ for an X-invariant $(0,2)$-tensor field A. •

This assumption will allow the inclusion of Rayleigh dissipative forces along the relative equilibrium in the stabilization results in subsequent sections. We suppose that X is an infinitesimal symmetry for Σ and that χ is a relative equilibrium. Then, according to Definition S3.8, a pair (ξ, u) is a controlled trajectory for the linearization of Σ along $(\chi, 0)$ if and only if

$$\begin{aligned}\overset{\mathbb{G}}{\nabla}{}^2_{\chi'(t)}\xi(t) + R(\xi(t), \chi'(t))\chi'(t) \\ = -\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\operatorname{grad} V)(\chi(t)) - \overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp \circ A^\flat \circ X)(\chi(t)) \\ + (\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp \circ A^\flat))(\chi'(t)) + \mathbb{G}^\sharp \circ A^\flat(\overset{\mathbb{G}}{\nabla}_{\chi'(t)}\xi(t)) + \sum_{a=1}^m u^a(t) Y_a(\chi(t)).\end{aligned} \tag{S3.5}$$

In order to facilitate making the connection between the preceding result and the reduced linearization given in the next section, we state the following characterization of the unreduced linearization.

Theorem S3.11 (Linearization of relative equilibrium before reduction). *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-simple mechanical control system satisfying Assumption S3.10, let X be a complete infinitesimal symmetry of Σ for which the projection $\pi_{\mathsf{B}}\colon \mathsf{Q} \to \mathsf{B}$ onto the set of X-orbits is a surjective submersion, and let $\chi\colon \mathbb{R} \to \mathsf{Q}$ be a regular relative equilibrium. For a vector field ξ along χ, let $x(t) = T_{\chi(t)}\pi_{\mathsf{B}}(\xi(t))$ and $\nu(t) = \dot{\zeta}(t)$, where* $\operatorname{ver}(\xi(t)) = \zeta(t)X(\chi(t))$. *Then the pair (ξ, u) is a controlled trajectory for the linearization of Σ along $(\chi, 0)$ if and only if*

$$\begin{aligned}\operatorname{hlft}_{\chi(t)}(\ddot{x}(t)) + \dot{\nu}(t)X(\chi(t)) = -\mathbb{G}^\sharp \circ \operatorname{Hess} V_X^\flat(\operatorname{hlft}_{\chi(t)}(x(t))) \\ - \frac{2\langle \mathrm{d}V(\chi(t)); \operatorname{hlft}_{\chi(t)}(\dot{x}(t))\rangle}{\|X\|_{\mathbb{G}}^2(\chi(t))} X(\chi(t)) \\ + \operatorname{hlft}_{\chi(t)}(C_X(\dot{x}(t))) + 2\nu(t)\operatorname{grad} V(\chi(t)) \\ + \mathbb{G}^\sharp \circ A^\flat(\operatorname{hlft}_{\chi(t)}(\dot{x}(t))) + \nu(t)\mathbb{G}^\sharp \circ A^\flat \circ X(\chi(t)) + \sum_{a=1}^m u^a(t) Y_a(\chi(t)),\end{aligned} \tag{S3.6}$$

where $b_0 = \pi_{\mathsf{B}}(\chi(0))$.

Proof. As in Proposition 5.64(ii), $\overset{\mathbb{G}}{\nabla}_X X = -\frac{1}{2}\mathrm{grad}\,\|X\|_{\mathbb{G}}^2$. Therefore,

$$\overset{\mathbb{G}}{\nabla}_X X + \mathrm{grad}V = \mathrm{grad}(V - \tfrac{1}{2}\,\|X\|_{\mathbb{G}}^2) = \mathrm{grad}V_X.$$

By Theorem 6.56(i), for each $t \in \mathbb{R}$, $\mathrm{grad}V_X(\chi(t)) = 0$. Using this fact, it is straightforward (e.g., using coordinates) to show that

$$\overset{\mathbb{G}}{\nabla}(\mathrm{grad}V_X)(\chi(t)) = \mathbb{G}^\sharp(\chi(t)) \circ \mathrm{Hess}\,V_X^\flat(\chi(t)).$$

Furthermore, since V_X is X-invariant, for any $x \in \mathsf{T}_{b_0}\mathsf{B}$, we have

$$\begin{aligned}\mathbb{G}^\sharp(\chi(t)) \circ \mathrm{Hess}\,V_X^\flat(\chi(t))(\mathrm{hlft}_{\chi(t)}(x))&\\ &= \mathrm{hlft}_{\chi(t)}(\mathbb{G}_{\mathsf{B}}^\sharp(b_0) \circ \mathrm{Hess}(V_X)_{\mathsf{B}}^\flat(b_0)(x)).\end{aligned}$$

For a vertical tangent vector $v_{\chi(t)} \in \mathsf{V}_{\chi(t)}\mathsf{Q}$ we have

$$\overset{\mathbb{G}}{\nabla}(\mathrm{grad}V_X)(v_{\chi(t)}) = 0,$$

using the fact that $\mathrm{grad}V_X(\chi(t)) = 0$ and using X-invariance of V_X. Summarizing the preceding computations is the following formula for a vector field ξ along χ:

$$\overset{\mathbb{G}}{\nabla}(\overset{\mathbb{G}}{\nabla}_X X + \mathrm{grad}V)(\xi(t)) = \mathrm{hlft}_{\chi(t)}\big(\mathbb{G}_{\mathsf{B}}^\sharp(b_0) \circ \mathrm{Hess}(V_X)_{\mathsf{B}}^\flat(b_0)(T\pi_{\mathsf{B}}(\xi(t)))\big). \tag{S3.7}$$

Now let ξ be a vector field along χ and let Ξ be a vector field extending ξ. Since $X(\chi(t)) = \chi'(t)$, we have, using the definition of the curvature tensor,

$$\begin{aligned}&\overset{\mathbb{G}}{\nabla}{}^2_{\chi'(t)}\xi(t) + R(\xi(t), \chi'(t))\chi'(t)\\ &\quad= \overset{\mathbb{G}}{\nabla}_X \overset{\mathbb{G}}{\nabla}_X \Xi(\chi(t)) + \overset{\mathbb{G}}{\nabla}_\Xi \overset{\mathbb{G}}{\nabla}_X X(\chi(t)) - \overset{\mathbb{G}}{\nabla}_X \overset{\mathbb{G}}{\nabla}_\Xi X(\chi(t)) - \overset{\mathbb{G}}{\nabla}_{[\Xi,X]} X(\chi(t)).\end{aligned}$$

A straightforward manipulation, using the fact that $\overset{\mathbb{G}}{\nabla}$ has zero torsion, gives

$$\overset{\mathbb{G}}{\nabla}_X \overset{\mathbb{G}}{\nabla}_X \Xi + \overset{\mathbb{G}}{\nabla}_\Xi \overset{\mathbb{G}}{\nabla}_X X - \overset{\mathbb{G}}{\nabla}_X \overset{\mathbb{G}}{\nabla}_\Xi X - \overset{\mathbb{G}}{\nabla}_{[\Xi,X]} X = \overset{\mathbb{G}}{\nabla}_\Xi \overset{\mathbb{G}}{\nabla}_X X + 2\overset{\mathbb{G}}{\nabla}_{[X,\Xi]} X + [X,[X,\Xi]]. \tag{S3.8}$$

Around a point $\chi(t_0) \in \mathrm{image}(\chi)$, let $(\mathcal{U}, \phi)$ be coordinates with the following properties:

1. $X = \frac{\partial}{\partial q^n}$;
2. $((q^1, \dots, q^{n-1}), (q^n))$ are fiber bundle coordinates for $\pi_{\mathsf{B}} \colon \mathsf{Q} \to \mathsf{B}$;
3. for any point $\chi(t) \in \mathcal{U}$, the basis $\{\frac{\partial}{\partial q^1}(\chi(t)), \dots, \frac{\partial}{\partial q^n}(\chi(t))\}$ for $\mathsf{T}_{\chi(t)}\mathsf{Q}$ is $\mathbb{G}$-orthogonal.

(To see that such coordinates exist, consider the following general construction. Let $((\tilde{q}^1,\dots,\tilde{q}^{n-1}),\tilde{q}^n)$ be principal bundle coordinates about $\chi(t_0)$ and be such that $\tilde{q}^n = X$. Also suppose that $\chi(t_0)$ is mapped to the origin in these coordinates. Now make a linear change of coordinates to coordinates $((q^1,\dots,q^{n-1}),q^n)$ such that $\frac{\partial}{\partial q^n} = \frac{\partial}{\partial \tilde{q}^n}$ and such that at $\chi(t_0)$ the coordinate vector fields $\frac{\partial}{\partial q^1},\dots,\frac{\partial}{\partial q^{n-1}}$ are orthogonal and orthogonal to $\frac{\partial}{\partial q^n}$.) In these coordinates one readily determines that

$$[X,\Xi](\chi(t)) = \dot{\xi}^i(t)\frac{\partial}{\partial q^i}, \quad [X,[X,\Xi]](\chi(t)) = \ddot{\xi}^i(t)\frac{\partial}{\partial q^i} \tag{S3.9}$$

for all values of t for which $\chi(t) \in \mathcal{U}$. In these coordinates it also holds that

$$\mathrm{hlft}_{\chi(t)}\frac{\partial}{\partial q^a}(b_0) = \frac{\partial}{\partial q^a}(\chi(t)), \qquad a \in \{1,\dots,n-1\}. \tag{S3.10}$$

Therefore, if ξ is as above and if $x(t) = T\pi_{\mathsf{B}}(\xi(t))$, then we have

$$\begin{aligned}
2\overset{\mathbb{G}}{\nabla}X([X,\Xi](\chi(t))) = {}& 2\big(\overset{\mathbb{G}}{\nabla}_{\mathrm{hlft}_{\chi(t)}(\dot{x}(t))}X(\chi(t))\big) + 2\overset{\mathbb{G}}{\nabla}X(\mathrm{ver}([X,\Xi])(\chi(t)))\\
= {}& -\mathrm{hlft}_{\chi(t)}(C_X(\dot{x}(t))) + 2\mathrm{ver}(\overset{\mathbb{G}}{\nabla}X(\mathrm{hlft}_{\chi(t)}(\dot{x}(t))))\\
&+ 2\overset{\mathbb{G}}{\nabla}X(\mathrm{ver}([X,\Xi])(\chi(t)))\\
= {}& -\mathrm{hlft}_{\chi(t)}(C_X(\dot{x}(t))) + 2\overset{\mathbb{G}}{\nabla}X(\mathrm{ver}([X,\Xi])(\chi(t)))\\
&+ \frac{2X(\chi(t))}{\|X\|^2_{\mathbb{G}}(\chi(t))}\mathbb{G}(\overset{\mathbb{G}}{\nabla}X(\mathrm{hlft}_{\chi(t)}(\dot{x}(t))), X(\chi(t)))\\
= {}& -\mathrm{hlft}_{\chi(t)}(C_X(\dot{x}(t))) + 2\overset{\mathbb{G}}{\nabla}X(\mathrm{ver}([X,\Xi])(\chi(t)))\\
&- \frac{2X(\chi(t))}{\|X\|^2_{\mathbb{G}}(\chi(t))}\mathbb{G}(\overset{\mathbb{G}}{\nabla}_X X(\chi(t)), \mathrm{hlft}_{\chi(t)}(\dot{x}(t)))\\
= {}& -\mathrm{hlft}_{\chi(t)}(C_X(\dot{x}(t))) + 2\overset{\mathbb{G}}{\nabla}X(\mathrm{ver}([X,\Xi])(\chi(t)))\\
&+ \frac{2X(\chi(t))}{\|X\|^2_{\mathbb{G}}(\chi(t))}\langle \mathrm{d}V(\chi(t)); \mathrm{hlft}_{\chi(t)}(\dot{x}(t))\rangle,
\end{aligned} \tag{S3.11}$$

using the fact that $\overset{\mathbb{G}}{\nabla}_X X = \mathrm{grad}V_X - \mathrm{grad}V$, and that $\mathrm{d}V_X(\chi(t)) = 0$ for all $t \in \mathbb{R}$. Also,

$$\mathrm{hlft}_{\chi(t)}(\ddot{x}(t)) = \mathrm{hor}([X,[X,\Xi]](\chi(t))), \qquad t \in \mathbb{R}. \tag{S3.12}$$

In the coordinates $(q^1,\dots,q^n)$, one also computes

$$\overset{\mathbb{G}}{\nabla}X = \overset{\mathbb{G}}{\Gamma}{}^i_{nj}\frac{\partial}{\partial q^i} \otimes \mathrm{d}q^j,$$

from which we ascertain that

$$2\overset{\mathbb{G}}{\nabla}X(\mathrm{ver}([X,\Xi])(\chi(t))) = 2\overset{\mathbb{G}}{\Gamma}{}^i_{nn}\dot{\xi}^n\frac{\partial}{\partial q^i},$$

where no summation is intended over the index "n." One readily verifies that, in our coordinates,

$$\overset{\mathbb{G}}{\Gamma}{}^i_{nn}\frac{\partial}{\partial q^i} = \overset{\mathbb{G}}{\nabla}_X X = \mathrm{grad}V_X - \mathrm{grad}V.$$

Since $\mathrm{d}V_X(\chi(t)) = 0$, we have

$$2\overset{\mathbb{G}}{\nabla}X(\mathrm{ver}([X,\Xi])(\chi(t))) = -2\mathrm{ver}([X,\Xi](\chi(t)))\mathrm{grad}V(\chi(t)). \tag{S3.13}$$

We also clearly have, by definition of $\mathscr{L}^{X,\chi}$,

$$\dot{\xi}^n(t)\frac{\partial}{\partial q^n} = \mathrm{ver}(\mathscr{L}^{X,\chi}(\xi(t))), \tag{S3.14}$$

where no summation is intended over "n."

To simplify the terms involving the external force, we note that

$$\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp\circ A^\flat(X(\chi(t)))) = (\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp\circ A^\flat))(X(\chi(t))) + \mathbb{G}^\sharp\circ A^\flat(\overset{\mathbb{G}}{\nabla}_{\xi(t)}X(\chi(t))).$$

Thus, using the fact that $\overset{\mathbb{G}}{\nabla}$ is torsion-free, we have

$$\begin{aligned} -\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp\circ A^\flat\circ X)(\chi(t)) + (\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp\circ A^\flat))(\chi'(t)) \\ + \mathbb{G}^\sharp\circ A^\flat(\overset{\mathbb{G}}{\nabla}_{\chi'(t)}\xi(t)) = \mathbb{G}^\sharp\circ A^\flat([X,\Xi](\chi(t))). \end{aligned}$$

Using (S3.9) and (S3.10) we arrive at

$$\begin{aligned} -\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp\circ A^\flat\circ X)(\chi(t)) + (\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp\circ A^\flat))(\chi'(t)) \\ + \mathbb{G}^\sharp\circ A^\flat(\overset{\mathbb{G}}{\nabla}_{\chi'(t)}\xi(t)) = \mathbb{G}^\sharp\circ A^\flat(\mathrm{hlft}_{\chi(t)}(\dot{x}(t))) + \dot{\xi}^n\mathbb{G}^\sharp\circ A^\flat(X(\chi(t))), \end{aligned} \tag{S3.15}$$

where $x(t) = T_{\chi(t)}\pi_{\mathsf{B}}(\xi(t))$.

Finally, for a vector field ξ along χ, let $x(t) = T\pi_{\mathsf{B}}(\xi(t))$ and let $\nu(t)X(t) = \mathrm{ver}(\mathscr{L}^{X,\chi}(\xi))$. In terms of our coordinates above, $\nu(t) = \dot{\xi}^n(t)$. One now combines equations (S3.7), (S3.8), (S3.9), (S3.11), (S3.12), (S3.13), (S3.14) and (S3.15) to get the result. ■

Remark S3.12. The preceding theorem is not very obvious; in particular, the equivalence of equations (S3.5) and (S3.6) is not transparent. Indeed, the relationship between the curvature tensor and the components of the system that appear in the theorem statement, C, $\mathrm{Hess}(V_X)$, and $\mathrm{grad}V$, is rather subtle. In this respect, the proof of the theorem bears study, if these relationships are to be understood. •

S3.1.4 Linearization of the reduced equations along a relative equilibrium

We again consider a C^∞-forced simple mechanical control system $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ satisfying Assumption S3.10, take X to be a complete infinitesimal symmetry for Σ, ask that $\pi_{\mathsf{B}} \colon \mathsf{Q} \to \mathsf{B}$ be a surjective submersion, and let χ be a relative equilibrium. In this section we provide the form of the linearization along a relative equilibrium by linearizing, in the usual manner, the reduced equations, which we reproduce here for convenience:

$$\begin{aligned}
\overset{\mathbb{G}_{\mathsf{B}}}{\nabla}_{\eta'(t)}\eta'(t) &= -\mathrm{grad}_{\mathsf{B}}\big(V^{\mathrm{eff}}_{X,v(t)}\big)_{\mathsf{B}}(\eta(t)) + v(t)C_X(\eta'(t)) \\
&\quad + T\pi_{\mathsf{B}} \circ \mathbb{G}^\sharp \circ A^\flat\big(\gamma'(t) - X(\gamma(t))\big) + Y_{\mathsf{B},u}(t, \eta(t)), \\
\dot{v}(t) &= -\frac{v(t)\langle \mathrm{d}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t)); \eta'(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} + \frac{\langle A^\flat(\gamma'(t) - X(\gamma(t))); X(\gamma(t))\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} \\
&\quad + \frac{\mathbb{G}(Y_u(t, \gamma(t)), X(\gamma(t)))}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))}.
\end{aligned} \tag{S3.16}$$

Here (γ, u) is a controlled trajectory for Σ, $\eta = \pi_{\mathsf{B}} \circ \gamma$, and v is defined by $\mathrm{ver}(\gamma'(t)) = v(t)X(\gamma(t))$.

The reduced equations are straightforward to linearize, since we are merely linearizing about an equilibrium point. To compactly state the form of the linearization requires some notation. Define a $(1,1)$-tensor field A_{B} on B by

$$A_{\mathsf{B}}(v_b) = T_q\pi_{\mathsf{B}} \circ \mathbb{G}^\sharp(q) \circ A^\flat(q) \circ \mathrm{hlft}_q(v_b),$$

for $q \in \pi_{\mathsf{B}}^{-1}(b)$. This definition can be shown to be independent of the choice of $q \in \pi_{\mathsf{B}}^{-1}(b)$ by virtue of the X-invariance of A. Define a vector field a_{B} on B by

$$a_{\mathsf{B}}(b) = T_q\pi_{\mathsf{B}} \circ \mathbb{G}^\sharp(q) \circ A^\flat(q)(X(q)),$$

where $q \in \pi_{\mathsf{B}}^{-1}(b)$, and again this definition can be shown to be well-defined. Finally, define a covector field α_{B} on B by

$$\langle \alpha_{\mathsf{B}}(b); v_b\rangle = \frac{\langle A^\flat(\mathrm{hlft}_q(v_b)); X(q)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b)},$$

where $q \in \pi_{\mathsf{B}}^{-1}(b)$ and $v_b \in \mathsf{T}_b\mathsf{B}$. This definition, too, is independent of the choice of q.

We may now state the form of the linearization of the reduced equations.

Proposition S3.13 (Linearization of relative equilibrium after reduction). *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-forced simple mechanical control system, let X be a complete infinitesimal symmetry of Σ for which the projection $\pi_{\mathsf{B}} \colon \mathsf{Q} \to \mathsf{B}$ onto the set of X-orbits is a surjective submersion, and let χ be a regular relative equilibrium with $b_0 = \pi_{\mathsf{B}} \circ \chi(0)$.*

The linearization of equations (S3.16) *about* $(0_{b_0}, 1)$ *is the linear control system* $(\mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}, A_\Sigma(b_0), B_\Sigma(b_0))$, *where*

$$A_\Sigma(b_0) = \begin{bmatrix} 0 & \mathrm{id}_{\mathsf{T}_{b_0}\mathsf{B}} & 0 \\ -\mathbb{G}_{\mathsf{B}}(b_0)^\sharp \circ \mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)^\flat & C_X(b_0) + A_{\mathsf{B}}(b_0) & 2\mathrm{grad}_{\mathsf{B}} V_{\mathsf{B}}(b_0) + a_{\mathsf{B}}(b_0) \\ 0 & -2\frac{\mathrm{d}V_{\mathsf{B}}(b_0)}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} + \alpha_{\mathsf{B}}(b_0) & \frac{\langle A^\flat(X(q_0)); X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} \end{bmatrix},$$

$$B_\Sigma(b_0) = \begin{bmatrix} 0 \\ B_{\Sigma,2}(b_0) \\ B_{\Sigma,3}(b_0) \end{bmatrix},$$

where $B_{\Sigma,2}(b_0) \in L(\mathbb{R}^m; \mathsf{T}_{b_0}\mathsf{B})$ *is defined by*

$$B_{\Sigma,2}(b_0)(u) = \sum_{a=1}^m u^a Y_{\mathsf{B},a}(b_0),$$

and where $B_{\Sigma,3}(b_0) \in L(\mathbb{R}^m; \mathbb{R})$ *is defined by*

$$B_{\Sigma,3}(b_0)(u) = \sum_{a=1}^m u^a \frac{\mathbb{G}(Y_a(\chi(0)), X(\chi(0)))}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)}.$$

Proof. Let us outline the computations that can be used to prove the result. For convenience, we break down $A_\Sigma(b_0)$ into nine entries which we call the (i,j)th entry, $i,j \in \{1,2,3\}$, based on the block form of $A_\Sigma(b_0)$. To linearize the equations (S3.16) we use three curves. The first curve is on $Z(\mathsf{TB})$, we denote it by η_1, and we require that $\eta_1(0) = 0_{b_0}$ and $\eta_1'(0) = v_{b_0} \oplus 0_{b_0}$. The second curve we use is a curve $t \mapsto \eta_2(t)$ on $\mathsf{T}_{b_0}\mathsf{B}$ for which $\eta_2(0) = 0_{b_0}$ and for which $\eta_2'(0) = 0_{b_0} \oplus v_{b_0} \in \mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B}$. In defining η_1 and η_2 we have used the identification of $\mathsf{T}_{0_{b_0}}\mathsf{TB}$ with $\mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B}$ as given in Lemma 6.33. The third curve is on $\mathbb{R}$, and is defined by $t \mapsto v(t) = 1 + t$. During the course of the proof, $\eta_1(t)$, $\eta_2(t)$, and $v(t)$ will always be taken to be so defined. Following the same arguments leading to Proposition 6.37, we arrive at the following conclusions.

1. The $(1,1)$, $(1,2)$, and the $(1,3)$ entries are directly verified to have the stated form.
2. One computes

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathrm{grad}_{\mathsf{B}}\big(V^{\mathrm{eff}}_{X,1}\big)_{\mathsf{B}}(\eta_1(t)) &= v_{b_0} \oplus 0_{b_0}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathrm{grad}_{\mathsf{B}}\big(V^{\mathrm{eff}}_{X,1}\big)_{\mathsf{B}}(\eta_2(t)) &= 0_{b_0} \oplus \mathbb{G}(b_0)^\sharp \circ (V_X)_{\mathsf{B}}(b_0)^\flat, \\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathrm{grad}_{\mathsf{B}}\big(V^{\mathrm{eff}}_{X,v(t)}\big)_{\mathsf{B}}(b_0) &= 0_{b_0} \oplus (-2\mathrm{grad}_{\mathsf{B}} V_{\mathsf{B}}(b_0)),
\end{aligned}$$

where we use the fact that $V^{\mathrm{eff}}_{X,\lambda} = V(q) - \frac{\lambda^2}{2}\|X\|^2_{\mathbb{G}}$ and that $(V_X)_{\mathsf{B}}$ has a critical point at b_0.

3. One computes

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} C_X(\eta_1(t)) &= v_{b_0} \oplus 0_{b_0},\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} C_X(\eta_2(t)) &= 0_{b_0} \oplus C_X(v_{b_0}),\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} v(t)C_X(0_{b_0}) &= 0_{b_0} \oplus 0_{b_0}.
\end{aligned}$$

This gives the first term in the $(2,2)$ entry of $A_\Sigma(b_0)$.

4. Let $q_0 \in \pi_{\mathsf{B}}^{-1}(b_0)$. Let γ_1 be the curve on $Z(\mathsf{TQ})$ satisfying $\gamma_1(0) = 0_{q_0}$ and with tangent vector field $\gamma_1'(t) = \mathrm{hlft}_{\gamma_1(t)}(\eta_1'(t)) - X(\gamma_1(t))$, where we make the identification of $Z(\mathsf{TB})$ with B and of $Z(\mathsf{TQ})$ with Q. Let γ_2 be the curve on $\mathsf{T}_{q_0}\mathsf{Q}$ with $\gamma_2(0) = 0_{q_0}$ and with $\gamma_2'(t) = \mathrm{hlft}_{q_0}(\eta_2'(t)) - X(q_0)$. Finally, let γ_3 be the curve in Q satisfying $\gamma_3(0) = q_0$, and with tangent vector field $\gamma_3'(t) = v(t)X(\gamma_3(t))$. To simplify notation, let $\tilde{A}\colon \mathsf{TQ} \to \mathsf{TB}$ be the vector bundle map over π_{B} defined by $\tilde{A} = T\pi_{\mathsf{B}} \circ \mathbb{G}^\sharp \circ A^\flat$. We then compute

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \tilde{A}(\gamma_1(t) - X \circ \pi_{\mathsf{TQ}}(\gamma_1(t))) &= v_{b_0} \oplus 0_{q_0},\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \tilde{A}(\gamma_2(t) - X \circ \pi_{\mathsf{TQ}}(\gamma_2(t))) &= 0_{b_0} \oplus \big(T_{q_0}\pi_{\mathsf{B}} \circ \mathbb{G}^\sharp(q_0) \circ A^\flat(q_0) \circ \mathrm{hlft}_{q_0}(v_{b_0})\big),\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \tilde{A}(\gamma_3'(t) - X(\gamma_3(t))) &= 0_{b_0} \oplus \big(T_{q_0}\pi_{\mathsf{B}} \circ \mathbb{G}^\sharp(q_0) \circ A^\flat(q_0)(X(q_0))\big).
\end{aligned}$$

This computation can be done directly using the coordinates $(q^1, \dots, q^n)$ introduced in the proof of Theorem S3.11. It gives the second parts of the $(2,2)$ and the $(2,3)$ entry in $A_\Sigma(b_0)$.

5. We compute

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \frac{\langle \mathrm{d}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\pi_{\mathsf{TB}} \circ \eta_1(t)); \eta_1(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\pi_{\mathsf{TB}} \circ \eta_1(t))} &= v_{b_0} \oplus 0_{b_0},\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \frac{\langle \mathrm{d}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0); \eta_2(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} &= 0_{b_0} \oplus \Big(-2\frac{\langle \mathrm{d}V_{\mathsf{B}}(b_0); v_{b_0}\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b)}\Big),\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \frac{\langle v(t)\mathrm{d}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0); 0_{b_0}\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} &= 0_{b_0} \oplus 0_{b_0},
\end{aligned}$$

using the fact that $V_X = V - \frac{1}{2}\|X\|^2_{\mathbb{G}}$, and that b_0 is a critical point for V_X. This gives the $(3,2)$ entry of $A_\Sigma(b_0)$.

6. Let $q_0 \in \pi_{\mathsf{B}}^{-1}(b_0)$ and define curves γ_i, $i \in \{1,2,3\}$, as in step 4 of the proof. We compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\frac{\langle A^\flat(\gamma_1(t)-X\circ\pi_{\mathsf{TQ}}(\gamma_1(t)));X(\pi_{\mathsf{TQ}}\circ\gamma_1(t))\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\pi_{\mathsf{TB}}\circ\eta_1(t))}=v_{q_0}\oplus 0_{q_0},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\frac{\langle A^\flat(\gamma_2(t)-X\circ\pi_{\mathsf{TQ}}(\gamma_2(t)));X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)}=\frac{\langle A^\flat(\mathrm{hlft}_{q_0}(v_{b_0}));X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\frac{\langle A^\flat(\gamma_3'(t)-X(\gamma_3(t)));X(\gamma_3(t))\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)}=\frac{\langle A^\flat(X(q_0));X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)},$$

so giving the $(3,3)$ entry in $A_\Sigma(q_0)$.

The linearization of the input vector fields is readily verified to give the stated form for $B_\Sigma(b_0)$, and this concludes the proof. ■

The equations governing controlled trajectories for the linearization of the reduced system are

$$\begin{aligned}\ddot{x}(t)=&-\mathbb{G}_{\mathsf{B}}(b_0)^\sharp\circ\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)^\flat(x(t))+C_X(b_0)(\dot{x}(t))\\&+2\nu(t)\mathrm{grad}_{\mathsf{B}}V_{\mathsf{B}}(b_0)+A_{\mathsf{B}}(b_0)(\dot{x}(t))+\nu(t)a_{\mathsf{B}}(b_0)+B_{\Sigma,2}(b_0)\cdot u(t),\\\dot{\nu}(t)=&-2\frac{\langle\mathrm{d}V_{\mathsf{B}}(b_0);\dot{x}(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)}+\alpha_{\mathsf{B}}(b_0)(\dot{x}(t))\\&+\nu(t)\frac{\langle A^\flat(X(q_0));X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)}+B_{\Sigma,3}(b_0)\cdot u(t).\end{aligned}$$

The following result gives the relationship between the reduced and the unreduced linearization.

Theorem S3.14 (Relationship between linearization before and after reduction). *Let $\Sigma=(\mathsf{Q},\mathbb{G},V,F,\mathscr{F},\mathbb{R}^m)$ be a C^∞-forced simple mechanical control system with F satisfying Assumption S3.10, let X be a complete infinitesimal symmetry of Σ for which the projection $\pi_{\mathsf{B}}\colon\mathsf{Q}\to\mathsf{B}$ onto the set of X-orbits is a surjective submersion, and let χ be a regular relative equilibrium with $b_0=\pi_{\mathsf{B}}\circ\chi(0)$.*

For a curve $t\mapsto x(t)\in\mathsf{T}_{b_0}\mathsf{B}$, a vector field ξ along χ, a function $\nu\colon\mathbb{R}\to\mathbb{R}$, and a locally integrable control $t\mapsto u(t)$, the following statements are equivalent:

(i) $t\mapsto(x(t)\oplus\dot{x}(t)\oplus\nu(t),u(t))$ *is a controlled trajectory for the linearization of the equations* (S3.16) *about* $(0_{b_0},1)$*, and in turn* $\mathrm{hor}(\xi(t))=\mathrm{hlft}_{\chi(t)}(x(t))$*, and* $\nu(t)=\dot{\zeta}(t)$*, where* $\mathrm{ver}(\xi(t))=\zeta(t)X(\chi(t))$*;*

(ii) (ξ,u) *is a controlled trajectory for the linearization of Σ about $(\chi,0)$, and in turn $x(t)=T\pi_{\mathsf{B}}(\xi(t))$, and $\nu(t)=\dot{\zeta}(t)$, where* $\mathrm{ver}(\xi(t))=\zeta(t)X(\chi(t))$.

Proof. This follows easily from Theorem S3.11 and Proposition S3.13. ■

The theorem is an important one, since it will allow us to switch freely between the reduced and unreduced linearizations. In some cases, it will be convenient to think of certain concepts in the unreduced setting, while computations are more easily performed in the reduced setting.

S3.2 Linearized effective energies

In many existing results concerning stability of relative equilibria, a central role is played by Hessian of the energy. This is a consequence of the fact that definiteness of the Hessian, restricted to certain subspaces, can easily deliver stability results in various forms. In this section we study the Hessian of the effective energy for a relative equilibria. In particular, we consider the interplay of the various natural energies with the reduction process and with linearization. Specifically, we spell out the geometry relating the processes of linearization and reduction.

S3.2.1 Some geometry associated to an infinitesimal isometry

The utility of the constructions in this section may not be immediately apparent, but will become clear in Proposition S3.20 below.

In this section we let $(\mathsf{Q}, \mathbb{G})$ be a Riemannian manifold with X an infinitesimal isometry satisfying Assumption 5.78. We denote by TTQ_X the restriction of the vector bundle $\pi_{\mathsf{TTQ}} \colon \mathsf{TTQ} \to \mathsf{TQ}$ to image(X). Thus TTQ_X is a vector bundle over image(X) whose fiber at $X(q)$ is $\mathsf{T}_{X(q)}\mathsf{TQ}$. We denote this fiber by $\mathsf{TTQ}_{X,X(q)}$. In like manner, HTQ_X and VTQ_X denote the restrictions of HTQ and TQ, respectively, to image(X). The Ehresmann connection on $\pi_{\mathsf{TQ}} \colon \mathsf{TQ} \to \mathsf{Q}$, defined by $\overset{\mathbb{G}}{\nabla}$ as in Section S1.3.6, gives a splitting of each fiber of TTQ_X as

$$\mathsf{TTQ}_{X,X(q)} = \mathsf{HTQ}_{X,X(q)} \oplus \mathsf{VTQ}_{X,X(q)}.$$

This gives a vector bundle isomorphism $\sigma_X \colon \mathsf{TTQ}_X \to \mathsf{HTQ}_X \oplus \mathsf{VTQ}_X$. Denote by $\Pi_{\mathsf{B}} \colon \mathsf{TQ} \to \mathsf{TQ}/\mathbb{R}$ the projection onto the set of X^T-orbits. We define $\phi_{\mathsf{B}} \colon \mathsf{TQ} \to \mathsf{TB} \times \mathbb{R}$ by

$$\phi_{\mathsf{B}}(w_q) = (T\pi_{\mathsf{B}}(w_q), \nu_X(w_q)),$$

where $\nu_X(w_q)$ is defined by $\mathrm{ver}(w_q) = \nu_X(w_q)X(q)$. Note that $\phi_{\mathsf{B}} \circ X(q) = (0_{\pi_{\mathsf{B}}(q)}, 1)$. Indeed, one can easily see that $\phi_{\mathsf{B},X} \triangleq \phi_{\mathsf{B}}|\mathrm{image}(X) \colon \mathrm{image}(X) \to Z(\mathsf{TB}) \times \{1\}$ is a surjective submersion. We next define a vector bundle map $\psi_{\mathsf{B}} \colon \mathsf{TTQ} \to \mathsf{T}(\mathsf{TB} \times \mathbb{R})$ over ϕ_{B} by $\psi_{\mathsf{B}} = T\phi_{\mathsf{B}}$. We denote $\psi_{\mathsf{B},X} = \psi_{\mathsf{B}}|\mathsf{TTQ}_X$, noting that this is a surjective vector bundle map from TTQ_X to the restricted vector bundle $\mathsf{T}(\mathsf{TB}\times\mathbb{R})|(Z(\mathsf{B})\times\{1\})$. Next we wish to give a useful description of the vector bundle $\mathsf{T}(\mathsf{TB} \times \mathbb{R})|(Z(\mathsf{B}) \times \{1\})$. We think of $\mathsf{TB} \times \mathbb{R}$ as a vector bundle over $\mathsf{B} \times \mathbb{R}$, and we let $\mathbb{R}_{\mathsf{B}\times\mathbb{R}}$ be the trivial vector bundle $(\mathsf{B} \times \mathbb{R}) \times \mathbb{R}$ over $\mathsf{B} \times \mathbb{R}$. We then note that $\mathsf{T}_{0_b}\mathsf{TB} \simeq \mathsf{T}_b \oplus \mathsf{T}_b\mathsf{B}$, as in Lemma 6.33. Thus we have a natural identification

$$\mathsf{T}(\mathsf{TB} \times \mathbb{R})|(Z(\mathsf{B}) \times \{1\}) \simeq (\mathsf{TB} \times \mathbb{R}) \oplus (\mathsf{TB} \times \mathbb{R}) \oplus \mathbb{R}_{\mathsf{B}\times\mathbb{R}} \tag{S3.17}$$

of vector bundles over $Z(\mathsf{TB}) \times \{1\} \simeq \mathsf{B} \times \{1\}$. The fiber over $(b, 1)$ is isomorphic to $\mathsf{T}_b\mathsf{B} \oplus \mathsf{T}_b\mathsf{B} \oplus \mathbb{R}$. We shall implicitly use the identification (S3.17) in the sequel. Next, we define a vector bundle map $\iota_{\mathsf{B}}\colon \mathsf{HTQ} \oplus \mathsf{VTQ} \to (\mathsf{TB} \times \mathbb{R}) \oplus (\mathsf{TB} \times \mathbb{R}) \oplus \mathbb{R}_{\mathsf{B}\times\mathbb{R}}$ by

$$\iota_{\mathsf{B}}(u_{v_q} \oplus w_{v_q}) = \big(T_q\pi_{\mathsf{B}}(u_{v_q}), T_q\pi_{\mathsf{B}}(w_{v_q} - \overset{\mathbb{G}}{\nabla}X(u_{v_q})), \nu_X(w_{v_q} - \overset{\mathbb{G}}{\nabla}X(u_{v_q}))\big).$$

We then let $\iota_{\mathsf{B},X}$ be the restriction of ι_{B} to $\mathsf{HTQ}_X \oplus \mathsf{VTQ}_X$.

The following result summarizes and ties together the above constructions.

Lemma S3.15. *The following statements hold:*

(i) σ_X *is a vector bundle isomorphism over* $\mathrm{id}_{\mathrm{image}(X)}$ *from* TTQ_X *to* $\mathsf{HTQ}_X \oplus \mathsf{VTQ}_X$*;*
(ii) $\phi_{\mathsf{B},X}$ *is a surjective submersion from* $\mathrm{image}(X)$ *to* $\mathsf{B} \times \{1\}$*;*
(iii) $\psi_{\mathsf{B},X}$ *is a surjective vector bundle map over* $\phi_{\mathsf{B},X}$ *from* TTQ_X *to* $(\mathsf{TB} \times \mathbb{R}) \oplus (\mathsf{TB} \times \mathbb{R}) \oplus \mathbb{R}_{\mathsf{B}\times\mathbb{R}}$*;*
(iv) $\iota_{\mathsf{B},X}$ *is a surjective vector bundle map over* $\phi_{\mathsf{B},X}$ *from* $\mathsf{HTQ}_X \oplus \mathsf{VTQ}_X$ *to* $(\mathsf{TB} \times \mathbb{R}) \oplus (\mathsf{TB} \times \mathbb{R}) \oplus \mathbb{R}_{\mathsf{B}\times\mathbb{R}}$*;*
(v) the following diagram commutes:

$$\begin{tikzcd}
\mathsf{TTQ}_X \arrow[rr, "\sigma_X"] \arrow[dr, "\psi_{\mathsf{B},X}"'] & & \mathsf{HTQ}_X \oplus \mathsf{VTQ}_X \arrow[dl, "\iota_{\mathsf{B},X}"] \\
& (\mathsf{TB} \times \mathbb{R}) \oplus (\mathsf{TB} \times \mathbb{R}) \oplus \mathbb{R}_{\mathsf{B}\times\mathbb{R}} &
\end{tikzcd}$$

Proof. This is most easily proved in an appropriate set of coordinates. Take coordinates $(q^1, \dots, q^n)$ for Q with the following properties:

1. $X = \frac{\partial}{\partial q^n}$;
2. for times t for which $\chi(t)$ is in the chart domain, $\{\frac{\partial}{\partial q^1}(\chi(t)), \dots, \frac{\partial}{\partial q^n}(\chi(t))\}$ is an orthogonal basis for $\mathsf{T}_{\chi(t)}\mathsf{Q}$.

This means that $(q^1, \dots, q^{n-1})$ are coordinates for B. These also form, therefore, coordinates for $Z(\mathsf{TB})$ and thus also for $Z(\mathsf{TB}) \times \{1\}$. Since a typical point in $\mathrm{image}(X)$ has the form

$$((q^1, \dots, q^n), (0, \dots, 0, 1))$$

in natural coordinates for TQ, we can use $(q^1, \dots, q^n)$ as coordinates for $\mathrm{image}(X)$. We denote natural coordinates for TTQ by $((\boldsymbol{q}, \boldsymbol{v}), (\boldsymbol{u}, \boldsymbol{w}))$. Then $(\boldsymbol{q}, \boldsymbol{u}, \boldsymbol{w})$ form a set of coordinates for TTQ_X.

The map ϕ_{B} from TQ to $\mathsf{TB} \times \mathbb{R}$ has the form

$$((q^1, \dots, q^n), (v^1, \dots, v^n)) \mapsto ((q^1, \dots, q^{n-1}), (v^1, \dots, v^{n-1}), v^n).$$

In the coordinates for $\mathrm{image}(X)$ and for $Z(\mathsf{TB}) \times \{1\}$, the map $\phi_{\mathsf{B},X}$ has the form

$$(q^1, \dots, q^n) \mapsto (q^1, \dots, q^{n-1}).$$

The coordinate form of ψ_{B} is then

$$\begin{aligned}&(((q^1,\dots,q^n),(v^1,\dots,v^n)),((u^1,\dots,u^n),(w^1,\dots,w^n)))\\ \mapsto\ &(((q^1,\dots,q^{n-1}),(v^1,\dots,v^{n-1}),v^n),((u^1,\dots,u^{n-1}),(w^1,\dots,w^{n-1}),w^n))\end{aligned}$$

and the coordinate form for $\psi_{\mathsf{B},X}$ is given by

$$\begin{aligned}&((q^1,\dots,q^n),(u^1,\dots,u^n),(w^1,\dots,w^n))\\ &\qquad\mapsto((q^1,\dots,q^{n-1}),(u^1,\dots,u^{n-1}),(w^1,\dots,w^{n-1}),w^n).\end{aligned}\tag{S3.18}$$

In coordinates, the map σ_X is given by

$$\begin{aligned}&((q^1,\dots,q^n),(u^1,\dots,u^n),(w^1,\dots,w^n))\\ &\qquad\mapsto((q^1,\dots,q^n),(u^1,\dots,u^n),(w^1+\overset{\mathbb{G}}{\Gamma}{}^i_{nj}u^j,\dots,w^n+\overset{\mathbb{G}}{\Gamma}{}^n_{nj}u^j)).\end{aligned}$$

Finally, in our above coordinates, the form of the map $\iota_{\mathsf{B},X}$ is

$$\begin{aligned}&((q^1,\dots,q^n),(u^1,\dots,u^n),(w^1,\dots,w^n))\\ \mapsto\ &((q^1,\dots,q^{n-1}),(u^1,\dots,u^n),(w^1-\overset{\mathbb{G}}{\Gamma}{}^1_{nj}u^j,\dots,w^{n-1}-\overset{\mathbb{G}}{\Gamma}{}^{n-1}_{nj}w^j),w^n-\overset{\mathbb{G}}{\Gamma}{}^n_{nj}u^j).\end{aligned}$$

All statements in the statement of the lemma follow directly from the preceding coordinate computations. ■

We shall see in Proposition S3.20 that $\iota_{\mathsf{B},X}$ relates two natural energies associated to a relative equilibrium.

S3.2.2 The effective energies and their linearizations

We let $\Sigma=(\mathsf{Q},\mathbb{G},V)$ be a C^∞-simple mechanical system with X a complete infinitesimal symmetry for Σ satisfying Assumption 5.78. First recall from (6.17) that the effective energy for a simple mechanical system $\Sigma=(\mathsf{Q},\mathbb{G},V)$ with complete infinitesimal symmetry X is

$$E_X(v_q)=\tfrac{1}{2}\,\|v_q-X(q)\|_{\mathbb{G}}^2+V_X(q),$$

where $V_X=V-\frac{1}{2}\,\|X\|_{\mathbb{G}}^2$ is the effective potential. The relative equilibria for Σ are then characterized by the critical points of E_X, as in part (i) of Theorem 6.56. As we shall see in Section S3.3, the Hessian of the effective energy at such critical points is useful for determining the stability of the corresponding relative equilibrium. The following result characterizes this Hessian in terms of the splitting of the fibers of TTQ using the Ehresmann connection on $\pi_{\mathsf{TTQ}}\colon\mathsf{TTQ}\to\mathsf{TQ}$ associated with $\overset{\mathbb{G}}{\nabla}$.

Lemma S3.16 (Hessian of effective energy). *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V)$ be a C^∞-simple mechanical system and let X be a complete infinitesimal symmetry for Σ. Let v_q be a critical point for the effective energy and let $\mathsf{T}_q\mathsf{Q} \oplus \mathsf{T}_q\mathsf{Q}$ be the splitting of $\mathsf{T}_{v_q}\mathsf{TQ}$ associated with $\overset{\mathbb{G}}{\nabla}$, as described in Section S1.3.6. Then*

$$\operatorname{Hess} E_X(u_1 \oplus w_1, u_2 \oplus w_2) = \mathbb{G}(w_1 - \overset{\mathbb{G}}{\nabla}_{u_1} X(q), w_2 - \overset{\mathbb{G}}{\nabla}_{u_2} X(q)) + \operatorname{Hess} V_X(u_1, u_2).$$

Proof. This is a messy, but straightforward, proof in coordinates. ■

With this as background, we make the following definition, recalling the notation $u_{v_q} \oplus w_{v_q}$ to denote a point in $\mathsf{T}_{v_q}\mathsf{TQ}$ relative to the splitting defined by $\overset{\mathbb{G}}{\nabla}$.

Definition S3.17 (Linearized effective energy). Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V)$ be a C^∞-simple mechanical system, let X be a complete infinitesimal isometry for Σ, and let χ be a relative equilibrium. The ***linearized effective energy*** is the function on $\mathsf{TTQ}|\mathrm{image}(\chi')$ defined by

$$E_\chi(u_{v_q} \oplus w_{v_q}) = \tfrac{1}{2}\|w_{v_q} - \overset{\mathbb{G}}{\nabla}_{u_{v_q}} X\|^2_{\mathbb{G}} + \tfrac{1}{2} \operatorname{Hess} V_X(u_{v_q}, u_{v_q}),$$

where $v_q = \chi'(0)$. •

Next we consider the linearized effective energy, but now for the reduced system. To do so, we assume that $\pi_{\mathsf{B}} \colon \mathsf{Q} \to \mathsf{B}$ is a surjective submersion, as in Assumption 5.78. The effective energy E_X is X^T-invariant and so drops to $\mathsf{TQ}/\mathbb{R} \simeq \mathsf{TB} \times \mathbb{R}$. During the course of the proof of Theorem 6.56, we further explicitly compute this "reduced effective energy," denoted here by E_X^{red}, as

$$E_X^{\mathrm{red}}(w_b, v) = \tfrac{1}{2}\mathbb{G}_{\mathsf{B}}(w_b, w_b) + (V_X)_{\mathsf{B}}(b) + \tfrac{1}{2}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b)(v-1)^2.$$

It then makes sense that the "reduced linearized effective energy" should be the Hessian of this function at a critical point, which corresponds, as we have seen, to a relative equilibrium. The following result records the form of the Hessian.

Lemma S3.18. *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V)$ be a C^∞-simple mechanical system, let X be a complete infinitesimal symmetry for Σ satisfying Assumption 5.78, and let $(0_b, 1) \in \mathsf{TB} \times \mathbb{R}$ be a critical point for E_X^{red}. Then*

$$\begin{aligned}\operatorname{Hess} E_X^{\mathrm{red}}&(0_b, 1)(u_1 \oplus v_1 \oplus \nu_1, u_2 \oplus v_2 \oplus \nu_2)\\ &= \mathbb{G}_{\mathsf{B}}(b)(v_1, v_2) + \operatorname{Hess}(V_X)_{\mathsf{B}}(b)(u_1, u_2) + (\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b)\nu_1\nu_2.\end{aligned}$$

Proof. This is a straightforward computation. ■

Based on this computation, let us make the following definition.

Definition S3.19 (Reduced linearized effective energy). Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V)$ be a C^∞-simple mechanical system, let X be a complete infinitesimal isometry for Σ satisfying Assumption 5.78, and let $b_0 = \pi_{\mathsf{B}}(\chi(0))$. The ***reduced linearized effective energy*** is the function on $\mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}$ defined by

$$E_\chi^{\mathrm{red}}(u_{b_0}, v_{b_0}, \nu) = \tfrac{1}{2}\,\|v_{b_0}\|^2_{\mathbb{G}_{\mathsf{B}}} + \tfrac{1}{2}\,\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)(u_{b_0}, u_{b_0}) + \tfrac{1}{2}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)\nu^2.$$

•

The preceding definition of the reduced linearized effective energy is obtained by "reducing" the effective energy, and then "linearizing" it. It should be possible to perform the operations in the opposite order to get to the same answer. To do this explicitly, we use the constructions of the preceding section. In particular, we use the vector bundle map $\iota_{\mathsf{B},X}$. As the following result indicates, one should think of this map as describing how the process of linearization is reduced when using the Ehresmann connection on $\pi_{\mathsf{TTQ}} \colon \mathsf{TTQ} \to \mathsf{TQ}$ associated with $\overset{\mathbb{G}}{\nabla}$.

Proposition S3.20 (Relating the linearized effective energies). *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V)$ be a C^∞-simple mechanical system with X a complete infinitesimal symmetry satisfying Assumption 5.78. If χ is a regular relative equilibrium, then $\iota_{\mathsf{B},X}(\chi'(t))^* E_\chi^{\mathrm{red}} = E_\chi$ for all $t \in \mathbb{R}$.*

Proof. Let $q = \chi(t)$ and let $b = \pi_{\mathsf{B}}(q)$. We compute

$$\begin{aligned}
\iota_{\mathsf{B},X}(\chi'(t))^* E_\chi^{\mathrm{red}}(u_{v_q} \oplus w_{v_q}) &= \tfrac{1}{2}\|T_q\pi_{\mathsf{B}}(w_{v_q} - \overset{\mathbb{G}}{\nabla} X(u_{v_q}))\|^2_{\mathbb{G}_{\mathsf{B}}} \\
&\quad + \tfrac{1}{2}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\nu(w_{v_q} - \overset{\mathbb{G}}{\nabla} X(u_q))) \\
&\quad + \tfrac{1}{2}\,\mathrm{Hess}(V_X)_{\mathsf{B}}(b)(T_q\pi_{\mathsf{B}}(u_b), T_q\pi_{\mathsf{B}}(u_b)) \\
&= \tfrac{1}{2}\|\mathrm{hor}(w_{v_q} - \overset{\mathbb{G}}{\nabla} X(u_{v_q}))\|^2_{\mathbb{G}} + \|\mathrm{ver}(w_{v_q} - \overset{\mathbb{G}}{\nabla} X(u_{v_q}))\|^2_{\mathbb{G}} \\
&\quad + \tfrac{1}{2}\,\mathrm{Hess}\,V_X(q)(u_{v_q}, u_{v_q}),
\end{aligned}$$

as desired. ■

S3.3 Linear stability of relative equilibria

In the text, we did not discuss linear stability of relative equilibria. Since we will be performing stabilization using linear methods, we now present this linear theory.

S3.3.1 Definitions and alternative characterizations

We let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F)$ be a C^∞-forced simple mechanical system where the force F satisfies Assumption S3.10, with X a complete infinitesimal symmetry for the system, and with $\chi \colon \mathbb{R} \to \mathsf{Q}$ a regular relative equilibrium. First we need a definition for linear stability of a relative equilibrium. The definition relies on the linearization along the relative equilibrium, which, from the developments of Section S3.1, satisfies an initial value problem of the form

$$\begin{aligned}\overset{\mathbb{G}}{\nabla}{}^2_{\chi'(t)}\xi(t) &+ R(\xi(t),\chi'(t))\chi'(t)\\ &= -\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathrm{grad}V)(\chi(t)) - \overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp \circ A^\flat \circ X)(\chi(t))\\ &\quad + (\overset{\mathbb{G}}{\nabla}_{\xi(t)}(\mathbb{G}^\sharp \circ A^\flat))(\chi'(t)) + \mathbb{G}^\sharp \circ A^\flat(\overset{\mathbb{G}}{\nabla}_{\chi'(t)}\xi(t)),\\ &\qquad \xi(0)=\xi_0,\ \mathscr{L}^{X,\chi}(\xi)(0) = v_{\xi,0}.\end{aligned} \tag{S3.19}$$

Remarks S3.21. 1. Note that it is immaterial that we specify the initial condition at $t=0$ due to X-invariance of the system.

2. Also note that we can specify the initial derivative condition for ξ by specifying $\overset{\mathbb{G}}{\nabla}_{\chi'(0)}\xi(0)$. Since both $\mathscr{L}^{X,\chi}$ and $\overset{\mathbb{G}}{\nabla}_{\chi'}$ are differential operators in TQ along χ, $\mathscr{L}^{X,\chi}(\xi)(t) - \overset{\mathbb{G}}{\nabla}_{\chi'(t)}\xi(t)$ depends only on $\xi(t)$. Thus specifying $\xi(0)$ and $\mathscr{L}^{X,\chi}(\xi)(0)$ is equivalent to specifying $\xi(0)$ and $\overset{\mathbb{G}}{\nabla}_{\chi'(0)}\xi(0)$. For our purposes, it is more convenient to specify the derivative initial condition in terms of $\mathscr{L}^{X,\chi}(\xi)(0)$. •

We may now state our stability definitions.

Definition S3.22 (Linear stability of relative equilibria). Let $\Sigma = (\mathsf{Q},\mathbb{G},V,F)$ be a C^∞-forced simple mechanical system with F satisfying Assumption S3.10, with X a complete infinitesimal symmetry for Σ satisfying Assumption 5.78, and with $\chi\colon \mathbb{R}\to\mathsf{Q}$ a relative equilibrium. For a vector field ξ along χ, let $\nu(t) = \dot{\zeta}(t)$, where $\mathrm{ver}(\xi(t)) = \zeta(t)X(\chi(t))$.

(i) The relative equilibrium χ is ***linearly base*** (resp. ***fiber***) ***stable*** if there exists $M>0$ such that the solution $t\mapsto\xi(t)$ to the initial value problem (S3.19) satisfies $\|\mathrm{hor}(\xi(t))\|_{\mathbb{G}} + \|\mathscr{L}^{X,\chi}(\mathrm{hor}(\xi))(t)\|_{\mathbb{G}} \le M(\|\mathrm{hor}(\xi_0)\|_{\mathbb{G}} + \|v_{\xi,0}\|_{\mathbb{G}})$ (resp. $|\nu(t)| \le M(\|\mathrm{hor}(\xi_0)\|_{\mathbb{G}} + \|v_{\xi,0}\|_{\mathbb{G}})$).

(ii) The relative equilibrium χ is ***linearly asymptotically base*** (resp. ***fiber***) ***stable*** if each solution $t\mapsto\xi(t)$ to the initial value problem (S3.19) satisfies $\lim_{t\to+\infty}(\|\mathrm{hor}(\xi(t))\|_{\mathbb{G}} + \|\mathscr{L}^{X,\chi}(\mathrm{hor}(\xi))(t)\|_{\mathbb{G}}) = 0$ (resp. $\lim_{t\to+\infty}\nu(t)=0$). •

Let us now give the relationship between these definitions of linear stability and the linear stability of the reduced system. To do this, let us write the equations governing the reduced linearization, following Proposition S3.13:

$$\begin{aligned}\ddot{x}(t) =&\ -\mathbb{G}_{\mathsf{B}}(b_0)^\sharp \circ \mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)^\flat(x(t)) + C_X(b_0)(\dot{x}(t))\\ &+ 2\nu(t)\mathrm{grad}_{\mathsf{B}}V_{\mathsf{B}}(b_0) + F_{\mathsf{B}}(b_0)(\dot{x}(t)) + \nu(t)f_{\mathsf{B}}(b_0),\\ \dot{\nu}(t) =&\ -2\frac{\langle \mathrm{d}V_{\mathsf{B}}(b_0);\dot{x}(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} + \alpha_{\mathsf{B}}(b_0)(\dot{x}(t)) + \nu(t)\frac{\langle A^\flat(X(q_0));X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)},\end{aligned} \tag{S3.20}$$

where we adopt the notation for F_{B}, f_{B}, and α_{B} as given before the statement of Proposition S3.13.

Proposition S3.23 (Base characterization of linear stability of relative equilibria). *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F)$ be a C^∞-forced simple mechanical system with F satisfying Assumption S3.10, with X a complete infinitesimal symmetry for the system, and with $\chi\colon \mathbb{R} \to \mathsf{Q}$ a regular relative equilibrium. Suppose that the projection $\pi_{\mathsf{B}}\colon \mathsf{Q} \to \mathsf{B}$ onto the set of X-orbits is a surjective submersion, and let $b_0 = \pi_{\mathsf{B}}(\chi(0))$. The following statements hold:*

(i) *χ is linearly base stable if and only if, for every solution $t \mapsto x(t) \oplus \dot{x}(t) \oplus \nu(t)$ of the equations* (S3.20)*, the function $\bar{\mathbb{R}}_+ \ni t \mapsto \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)}$ is bounded;*

(ii) *χ is linearly asymptotically base stable if and only if, for every solution $t \mapsto x(t) \oplus \dot{x}(t) \oplus \nu(t)$ of the equations* (S3.20)*, $\lim_{t\to+\infty} \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)} = 0$;*

(iii) *χ is linearly fiber stable if and only if, for every solution $t \mapsto x(t) \oplus \dot{x}(t) \oplus \nu(t)$ of the equations* (S3.20)*, the function $\bar{\mathbb{R}}_+ \ni t \mapsto |\nu(t)|$ is bounded;*

(iv) *χ is linearly asymptotically fiber stable if and only if, for every solution $t \mapsto x(t) \oplus \dot{x}(t) \oplus \nu(t)$ of the equations* (S3.20)*, $\lim_{t\to+\infty} |\nu(t)| = 0$.*

Proof. Since the equations (S3.20) are linear, their solution has the form

$$x(t) \oplus \dot{x}(t) \oplus \nu(t) = \exp(A_\Sigma(b_0)t)(x(0) \oplus \dot{x}(0) \oplus \nu(0)).$$

Let us abbreviate $\mathsf{V} = \mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}$, and define maps $p_1, p_2, p_3 \in L(\mathsf{V}; \mathsf{V})$ by

$$\begin{aligned} p_1(u_{b_0} \oplus v_{b_0} \oplus \nu) &= u_{b_0} \oplus 0_{b_0} \oplus 0, \\ p_2(u_{b_0} \oplus v_{b_0} \oplus \nu) &= 0_{b_0} \oplus v_{b_0} \oplus 0, \\ p_3(u_{b_0} \oplus v_{b_0} \oplus \nu) &= 0_{b_0} \oplus 0_{b_0} \oplus \nu. \end{aligned}$$

Define a norm $\|\cdot\|$ on V by

$$\|u_{b_0} \oplus v_{b_0} \oplus \nu\| = \|u_{b_0}\|_{\mathbb{G}_{\mathsf{B}}(b_0)} + \|v_{b_0}\|_{\mathbb{G}_{\mathsf{B}}(b_0)} + \frac{\nu^2}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)}.$$

Due to the nature of $\exp(A_\Sigma(b_0)t)$ (its components are linear combinations of products of polynomial, exponential, and trigonometric functions of t), the function $t \mapsto \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)}$ is bounded if and only if

$$\sup\left\{ \|p_1 \circ \exp(A_\Sigma(b_0)t)\| \;\middle|\; t \in \bar{\mathbb{R}}_+ \right\} < \infty,$$

where $\|\cdot\|$ is the operator norm (see Definition 3.15). Again due to the nature of the components of $\exp(A_\Sigma(b_0)t)$, if $t \mapsto \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)}$ is bounded, then so too is $t \mapsto \|\dot{x}(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)}$. Thus we have

$$\sup\left\{ \|p_2 \circ \exp(A_\Sigma(b_0)t)\| \;\middle|\; t \in \bar{\mathbb{R}}_+ \right\} < \infty.$$

Therefore, if $t \mapsto \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)}$ is bounded, it follows that, since $x(t) = p_1 \circ \exp(A_\Sigma(b_0)t)(x(0) \oplus \dot{x}(0) \oplus \nu(0))$ and $\dot{x}(t) = p_2 \circ \exp(A_\Sigma(b_0)t)(x(0) \oplus \dot{x}(0) \oplus \nu(0))$,

$$\|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)} + \|\dot{x}(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)} \leq M_1 \|x(0) \oplus \dot{x}(0) \oplus \nu(0)\|\,, \tag{S3.21}$$

for some $M_1 > 0$. Conversely, if (S3.21) holds for every solution of (S3.20), then it clearly holds that $t \mapsto \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)}$ is bounded for every solution of (S3.20). Thus we have shown that $t \mapsto \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)}$ is bounded for every solution of (S3.20) if and only if (S3.21) holds for every solution of (S3.20).

A similar argument can be used to show that $t \mapsto |\nu(t)|$ is bounded for every solution of (S3.20) if and only if the estimate

$$\frac{|\nu(t)|}{(\|X\|_{\mathbb{G}})_{\mathsf{B}}(b_0)} \leq M_2 \|x(0) \oplus \dot{x}(0) \oplus \nu(0)\| \tag{S3.22}$$

holds for every solution of (S3.20), for some $M_2 > 0$.

Parts (i) and (iii) now follow from (S3.21), (S3.22), and Theorem S3.14, along with the fact that the maps

$$\begin{gathered}\mathsf{T}_{b_0}\mathsf{B} \ni u_{b_0} \mapsto \mathrm{hlft}_{\chi(t)}(u_{b_0}) \in \mathsf{H}_{\chi(t)}\mathsf{Q},\\ \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R} \ni v_{b_0} \oplus \nu \mapsto \mathrm{hlft}_{\chi(t)}(v_{b_0}) + \nu X(\chi(t)) \in \mathsf{T}_{\chi(t)}\mathsf{Q}\end{gathered}$$

are isometries. Part (ii) follows from Theorem S3.14, along with the fact that, by properties of the components of $\exp(A_\Sigma(b_0))$, $\lim_{t\to+\infty} \|\dot{x}(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)} = 0$ if $\lim_{t\to+\infty} \|x(t)\|_{\mathbb{G}_{\mathsf{B}}(b_0)} = 0$. Part (iv) follows immediately from Theorem S3.14. ■

Remark S3.24. The definitions of base and fiber stability are examples of what is sometimes called “partial stability” in more general contexts. With this sort of stability, one is only interested in the behavior of some of the states of the system. This is studied in the text [Vorotnikov 1998], and, in particular, there one can find characterizations of partial stability of linear systems. •

S3.3.2 Sufficient conditions for linear stability of relative equilibria

Now let us give some natural sufficient conditions that we can use for the stabilization theory we give in the sections to follow. These conditions should be thought of as the linear analogue to the stability results of Section 6.3. As such, they use the preceding two notions of linearized effective energies. In this regard, the next result is the main result in this section.

Theorem S3.25 (Linear stability of relative equilibria). *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F)$ be a C^∞-forced simple mechanical system, with X a complete infinitesimal symmetry for Σ satisfying Assumption 5.78, and with $\chi\colon \mathbb{R} \to \mathsf{Q}$ a relative equilibrium. Suppose that $F(v_q) = -R^\flat_{\mathrm{diss}}(v_q - X(q))$, where R_{diss} is a Rayleigh dissipation function. For $b_0 = \pi_{\mathsf{B}} \circ \chi(0)$, the following statements hold:*

(i) *χ is linearly base and fiber stable if* $\operatorname{Hess}(V_X)_{\mathsf{B}}(b_0)$ *is positive-definite;*
(ii) *χ is linearly asymptotically base stable and linearly asymptotically fiber stable if* $\operatorname{Hess}(V_X)_{\mathsf{B}}(b_0)$ *is positive-definite and if* R_{diss} *is positive-definite.*

Proof. Note that combined linear (asymptotic) base and fiber stability of χ is equivalent to the linear (asymptotic) stability of the equilibrium point b_0 for the reduced system on $\mathsf{TB} \times \mathbb{R}$. Therefore, in the proof, we shall consider the stability in the reduced space, using the reduced linearized effective energy, E_χ^{red}, as a candidate Lyapunov function. First note that, under the hypothesis that $\operatorname{Hess}(V_X)_{\mathsf{B}}(b_0)$ is positive-definite, it follows that E_χ^{red} is positive-definite about $0_b \oplus 0_b \oplus 0$. Next, a straightforward computation, the details of which we omit, shows that

$$\begin{aligned}\frac{\mathrm{d}E_\chi^{\text{red}}}{\mathrm{d}t}(x(t) \oplus \dot{x}(t) \oplus \nu(t)) &= \mathbb{G}_{\mathsf{B}}(\dot{x}(t), F_{\mathsf{B}}(b_0)(\dot{x}(t))) + \nu(t)\mathbb{G}_{\mathsf{B}}(\dot{x}(t), f_{\mathsf{B}}(b_0)) \\ &\quad + \nu(t)(\|X\|_{\mathbb{G}}^2)_{\mathsf{B}}(b_0)\alpha_{\mathsf{B}}(b_0)(\dot{x}(t)) + \nu(t)^2\langle A^\flat(X(q_0)); X(q_0)\rangle,\end{aligned}$$

along a solution $t \mapsto x(t) \oplus \dot{x}(t) \oplus \nu(t)$ to equations (S3.20), where $q_0 \in \pi_{\mathsf{B}}^{-1}(b_0)$, and where A is as in Assumption S3.10. One now can easily show that

$$\begin{aligned}\mathbb{G}_{\mathsf{B}}(\dot{x}(t), F_{\mathsf{B}}(b_0)(\dot{x}(t))) &= A(\operatorname{hlft}_{q_0}(\dot{x}(t)), \operatorname{hlft}_{q_0}(\dot{x}(t))), \\ \nu(t)\mathbb{G}_{\mathsf{B}}(\dot{x}(t), f_{\mathsf{B}}(b_0)) &= A(\nu(t)X(q_0), \operatorname{hlft}_{q_0}(\dot{x}(t))), \\ \nu(t)(\|X\|_{\mathbb{G}}^2)_{\mathsf{B}}(b_0)\alpha_{\mathsf{B}}(b_0)(\dot{x}(t)) &= A(\operatorname{hlft}_{q_0}(\dot{x}(t)), \nu(t)X(q_0)).\end{aligned}$$

These computations allow us to conclude that

$$\begin{aligned}&\frac{\mathrm{d}E_\chi^{\text{red}}}{\mathrm{d}t}(x(t) \oplus \dot{x}(t) \oplus \nu(t)) \\ &\qquad = A(\operatorname{hlft}_{q_0}(\dot{x}(t)) + \nu(t)X(q_0), \operatorname{hlft}_{q_0}(\dot{x}(t)) + \nu(t)X(q_0)).\end{aligned}$$

In part (i), $A = -R_{\text{diss}}$ is negative-semidefinite, and in part (ii), $A = -R_{\text{diss}}$ is negative-definite, and the result then follows directly. ■

Alternatively, one can check the hypotheses of the theorem using the linearized effective energy. The following result contains the results of this transcription.

Corollary S3.26 (Linear stability of relative equilibria using unreduced data). *Let* $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F)$ *be a* C^∞*-forced simple mechanical system, with* X *a complete infinitesimal symmetry for* Σ *satisfying Assumption 5.78, and with* $\chi\colon \mathbb{R} \to \mathsf{Q}$ *a relative equilibrium. Suppose that* $F(v_q) = -R_{\text{diss}}^\flat(v_q - X(q))$*, where* R_{diss} *is a Rayleigh dissipation function. The following statements hold:*

(i) *χ is linearly base and fiber stable if* $\operatorname{Hess} V_X(\chi(t))$ *is positive-definite on any (and so every) complement to* $\operatorname{span}_{\mathbb{R}}\{X(\chi(t))\}$ *for some (and so for all)* $t \in \mathbb{R}$*;*

(ii) χ *is linearly asymptotically base stable and linearly asymptotically fiber stable if* $\operatorname{Hess} V_X(\chi(t))$ *is positive-definite on any (and so every) complement to* $\operatorname{span}_{\mathbb{R}}\{X(\chi(t))\}$ *for some (and so for all)* $t \in \mathbb{R}$*, and if* R_{diss} *is positive-definite.*

Remarks S3.27. 1. Note that since positive-definiteness of the Hessian of $(V_X)_{\mathsf{B}}$ at b_0 implies that b_0 is an isolated local minimum for $(V_X)_{\mathsf{B}}$, the satisfaction of the hypotheses of Theorem S3.25 implies the satisfaction of the hypotheses of Theorem 6.56.

2. The presence of gyroscopic forces in the reduced linearization makes it difficult to draw the sharpest possible conclusions regarding linear stability. This is to be contrasted with the linear stability of equilibrium points, where, in Theorem 6.42, we are able to give much sharper stability conditions in the presence of only dissipative forces.
3. In part (ii) of Theorem S3.25 we require that R_{diss} be positive-definite. As is the case with stability of equilibria for mechanical systems, this hypothesis is stronger than required. When we discuss stabilization in Theorems S3.41 and S3.47 below, we shall see that controllability of the linearization is all that is required to achieve asymptotic stabilization. •

S3.4 Stabilization problems for relative equilibria

The problems one encounters for stabilization of relative equilibria are, of course, similar to those one encounters for equilibria. However, the additional structure of a base space and a symmetry direction make for more variants of the sort of stability one can encounter (cf. Definitions 6.52 and S3.22), and these should also be accounted for in the stabilization problems.

First let us characterize feedback for relative equilibria.

Definition S3.28 (Feedback for relative equilibria). Let $\Sigma = (\mathsf{Q}, \nabla, V, F, \mathcal{D}, \mathscr{F} = \{F_1, \dots, F_m\}, U)$ be a C^∞-general simple mechanical control system for which F is time-independent, and let X be a complete infinitesimal symmetry for Σ.

(i) A ***controlled relative equilibrium*** for Σ is a pair (χ, u_0) where
  (a) $\chi\colon \mathbb{R} \to \mathsf{Q}$ is an integral curve of X and
  (b) $u_0 \in U$ has the property that

$$\overset{\mathbb{G}}{\nabla}_{\chi'(t)}\chi'(t) = -\operatorname{grad}V(\chi(t)) + \sum_{a=1}^{m} u_0^a \mathbb{G}^\sharp \circ F^a(\chi(t)).$$

(ii) An ***X-invariant state feedback*** (resp. ***X-invariant time-dependent state feedback***) for Σ is a map $u\colon \mathsf{TQ} \to U$ (resp. $u\colon \bar{\mathbb{R}}_+ \times \mathsf{TQ} \to U$) with the property that $u \circ \gamma'(s)$ (resp. $u(t, \gamma'(s))$) is independent of s for each integral curve γ of X.

(iii) For an X-invariant state feedback (resp. X-invariant time-dependent state feedback) u for Σ, the ***closed-loop system*** is the forced simple mechanical control system with constraints defined by the 5-tuple $\Sigma_{\text{cl}} = (\mathsf{Q}, \mathbb{G}, V, F_{\text{cl}}, \mathcal{D})$, where

$$F_{\text{cl}}(v_q) = F(v_q) + \sum_{a=1}^{m} u^a(v_q) F^a(q),$$
$$\Big(\text{resp. } F_{\text{cl}}(t, v_q) = F(v_q) + \sum_{a=1}^{m} u^a(t, v_q) F^a(q)\Big).$$

(iv) For $r \in \mathbb{Z}_+ \cup \{\infty\} \cup \{\omega\}$, a state feedback (resp. time-dependent state feedback) is $\boldsymbol{C^r}$ if the corresponding closed-loop system is of class C^r.

(v) For $r \in \mathbb{Z}_+ \cup \{\infty\} \cup \{\omega\}$ and $\chi\colon \mathbb{R} \to \mathsf{Q}$ an integral curve for X, a state feedback is ***almost*** $\boldsymbol{C^r}$ ***about*** $\boldsymbol{\chi}$ if there exists an X-invariant neighborhood $\mathcal{U}$ of image(χ) such that the corresponding closed-loop system is C^r on $\mathcal{U} \setminus \{\text{image}(\chi)\}$. •

Remark S3.29. Note that, for a controlled relative equilibrium, we take the control to be constant. More generally, one could allow time-dependent controls to maintain a relative equilibrium. However, the constant control has the advantage that a controlled relative equilibrium corresponds to a controlled equilibrium point for the reduced system. •

Note that we include constraints in the formulation here. Although relative equilibria, and therefore base and fiber stability, are not given in Definitions 6.51 and 6.52 for systems with constraints, the modifications to those definitions are straightforward. In particular, a relative equilibrium is still an integral curve of X that is a solution of the equations of motion. The definitions for stabilizability are as follows.

Definition S3.30 (Stabilization problems for relative equilibria). Let $\Sigma = (\mathsf{Q}, \nabla, V, F, \mathcal{D}, \mathscr{F} = \{F_1, \dots, F_m\}, U)$ be a C^∞-general simple mechanical control system for which F is time-independent, and let X be a complete infinitesimal symmetry for Σ.

(i) A controlled relative equilibrium (χ, u_0) is ***base*** (resp. ***fiber***) ***stabilizable by X-invariant state feedback*** (resp. ***stabilizable by X-invariant time-dependent state feedback***) if there exists an X-invariant state feedback (resp. X-invariant time-dependent state feedback) u for Σ with the property that the closed-loop system has χ as a base (resp. fiber) stable relative equilibrium.

(ii) A controlled relative equilibrium (χ, u_0) is ***locally asymptotically base*** (resp. ***fiber***) ***stabilizable by X-invariant state feedback*** (resp. ***by time-dependent state feedback***) if there exists an X-invariant state feedback (resp. X-invariant time-dependent state feedback) and an X-invariant neighborhood $\mathcal{U}$ of image(χ) with the properties that

(a) the closed-loop system leaves $\mathsf{T}\mathcal{U}$ invariant, and
(b) the restriction of the closed-loop system to $\mathsf{T}\mathcal{U}$ possesses χ as an asymptotically base (resp. fiber) stable relative equilibrium.

(iii) A controlled relative equilibrium (χ, u_0) is ***globally asymptotically base*** (resp. ***fiber***) ***stabilizable by X-invariant state feedback*** (resp. ***by X-invariant time-dependent state feedback***) if, in part (vii), one can take $\mathcal{U} = \mathsf{Q}$. •

Remark S3.31. As was the case with stabilization of equilibrium points, we shall use the language of "u ***base*** (resp. ***fiber***) ***stabilizes*** the controlled relative equilibrium $(\chi, \mathbf{0})$" to mean that the closed-loop system with the control u is base (resp. fiber) stable. Similar statements can be made, of course, for local asymptotic stability. •

S3.5 Relative equilibrium stabilization

In this section we produce results for relative equilibria, mirroring those in Sections 10.3 and 10.4 for equilibria. Before we give our stabilization results, we first discuss how stabilization of the unreduced equations is related to stabilization of the reduced equations. Since the latter case involves stabilization of an equilibrium point, it is simpler to prove stability in this case.

S3.5.1 The relationship between reduced and unreduced stabilization

In this section, we let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-simple mechanical control system with F satisfying Assumption S3.10, X be a complete infinitesimal symmetry for Σ satisfying Assumption 5.78, and $(\chi, \mathbf{0})$ be a controlled relative equilibrium. We let $b_0 = \pi_{\mathsf{B}}(\chi(0))$, so that $(0_{b_0}, 1)$ is the equilibrium point for the reduced equations

$$\begin{aligned}
\overset{\mathbb{G}_{\mathsf{B}}}{\nabla}_{\eta'(t)}\eta'(t) &= -\operatorname{grad}_{\mathsf{B}}\big(V^{\mathrm{eff}}_{X,v(t)}\big)_{\mathsf{B}}(\eta(t)) + v(t)C_X(\eta'(t)) \\
&\qquad + T\pi_{\mathsf{B}} \circ \mathbb{G}^\sharp \circ A^\flat\big(\gamma'(t) - X(\gamma(t))\big) + Y_{\mathsf{B},u}(t, \eta(t)), \\
\dot{v}(t) &= -\frac{v(t)\langle \mathrm{d}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t)); \eta'(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} + \frac{\langle A^\flat(\gamma'(t) - X(\gamma(t))); X(\gamma(t))\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} \\
&\qquad + \frac{\mathbb{G}(Y_u(t, \gamma(t)), X(\gamma(t)))}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))}.
\end{aligned}$$

Let us denote by Σ^{red} the control-affine system on $\mathsf{TB} \times \mathbb{R}$ defined by these equations. Let $\mathcal{U}_{\mathsf{B}}$ be a neighborhood of b_0 in B, and note that $\mathcal{U}_{\mathsf{Q}} \triangleq \pi_{\mathsf{B}}^{-1}(\mathcal{U}_{\mathsf{B}})$ is a neighborhood of χ that is X-invariant. Now let $u_{\mathsf{B}} \colon \mathsf{T}\mathcal{U}_{\mathsf{B}} \times \mathbb{R} \to \mathbb{R}^m$ be a state feedback for Σ^{red}, and define $u_{\mathsf{Q}} \colon \mathsf{T}\mathcal{U}_{\mathsf{Q}} \to \mathbb{R}^m$ by

$$u_{\mathsf{Q}}(v_q) = u_{\mathsf{B}}(T\pi_{\mathsf{B}}(v_q), \nu_X(v_q)),$$

where $\nu_X(v_q) \in \mathbb{R}$ is defined by $\operatorname{ver}(v_q) = \nu_X(v_q)X(q)$.

The following result says that u_{Q} is an X-invariant state feedback, as per Definition S3.28, and that stabilization of the reduced equations gives stabilization of the unreduced equations.

Proposition S3.32 (Relative equilibrium stabilization using reduced stabilization). *Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-simple mechanical control system with F satisfying Assumption S3.10, X be a complete infinitesimal symmetry for Σ satisfying Assumption 5.78, and $(\chi, \mathbf{0})$ a controlled relative equilibrium. If u_{B} and u_{Q} are defined as above, then u_{Q} is an X-invariant state feedback for Σ. Furthermore, if u_{B} is of class C^∞, then the following statements hold:*

(i) if u_{B} stabilizes the controlled equilibrium point $((0_{b_0}, 1), \mathbf{0})$ for Σ^{red}, then u_{Q} base and fiber stabilizes the controlled relative equilibrium $(\chi, \mathbf{0})$;

(ii) if u_{B} locally asymptotically stabilizes the controlled equilibrium point $((0_{b_0}, 1), \mathbf{0})$ for Σ^{red}, then u_{Q} locally asymptotically base and fiber stabilizes the controlled relative equilibrium $(\chi, \mathbf{0})$.

Proof. First let us show that u_{Q} is X-invariant. Let $\gamma\colon \mathbb{R} \to \mathcal{U}_{\mathsf{Q}}$ be an integral curve of X lying in $\mathcal{U}_{\mathsf{Q}}$. We then have

$$u_{\mathsf{Q}}(\gamma'(t)) = u_{\mathsf{B}}(T\pi_{\mathsf{B}}(\gamma'(t)), \nu(t)) = u_{\mathsf{B}}(0_{b_0}, 1) = u_{\mathsf{Q}}(\gamma'(0)),$$

using the fact that $\operatorname{ver}(\gamma'(t)) = X(\gamma(t))$.

The second two statements in the proposition follow directly from the definitions of base and fiber stability. ■

One can refine the result by separately considering base and fiber stability. However, the results we state below will be for joint base and fiber stability. We therefore leave the refined statements to the reader.

On the basis of Proposition S3.32, the results we give in the remainder of the section concern stabilization of the reduced system associated to a relative equilibrium.

S3.5.2 Stabilization of linearization of reduced system

In Section 10.3, we developed the theory of potential shaping for linear systems, and then applied this to the stabilization of not necessarily linear systems in the case when their linearization satisfied the hypotheses needed for linear stabilization. In this section, we carry out the same procedure, but now for relative equilibria. As we shall see, the ideas do not translate verbatim, as the extra structure of the relative equilibrium introduces some additional complications.

Let us recall that the reduced system is governed by the equations

$$
\begin{aligned}
\overset{\mathbb{G}_{\mathsf{B}}}{\nabla}_{\eta'(t)}\eta'(t) = &-\mathrm{grad}_{\mathsf{B}}\big(V^{\mathrm{eff}}_{X,v(t)}\big)_{\mathsf{B}}(\eta(t)) + v(t)C_X(\eta'(t)) \\
&+ T\pi_{\mathsf{B}} \circ \mathbb{G}^{\sharp} \circ A^{\flat}\big(\gamma'(t) - X(\gamma(t))\big) + Y_{\mathsf{B},u}(t,\eta(t)), \\
\dot{v}(t) = &-\frac{v(t)\langle \mathrm{d}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t)); \eta'(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} + \frac{\langle A^{\flat}(\gamma'(t) - X(\gamma(t))); X(\gamma(t))\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))} \\
&+ \frac{\mathbb{G}(Y_u(t,\gamma(t)), X(\gamma(t)))}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(\eta(t))}.
\end{aligned}
$$

Let us denote by Σ^{red} the control-affine system on $\mathsf{TB} \times \mathbb{R}$ defined by these equations. The linearization of Σ^{red} about the equilibrium point $(0_{b_0}, 1)$ is governed by the equations

$$
\begin{aligned}
\ddot{x}(t) = &- \mathbb{G}_{\mathsf{B}}(b_0)^{\sharp} \circ \mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)^{\flat}(x(t)) + C_X(b_0)(\dot{x}(t)) \\
&+ 2\nu(t)\mathrm{grad}_{\mathsf{B}} V_{\mathsf{B}}(b_0) + A_{\mathsf{B}}(b_0)(\dot{x}(t)) + \nu(t)a_{\mathsf{B}}(b_0) + B_{\Sigma,2}(b_0)\cdot u(t), \\
\dot{\nu}(t) = &- 2\frac{\langle \mathrm{d}V_{\mathsf{B}}(b_0); \dot{x}(t)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} + \alpha_{\mathsf{B}}(b_0)(\dot{x}(t)) \\
&+ \nu(t)\frac{\langle A^{\flat}(X(q_0)); X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} + B_{\Sigma,3}(b_0)\cdot u(t).
\end{aligned}
$$

Let us denote by $\Sigma^{\mathrm{red,lin}}$ the linear system on $\mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}$ defined by these equations. Recall the notation $A_\Sigma(b_0)$ and $B_\Sigma(b_0)$ from Proposition S3.13.

Let us give some definitions concerning the type of feedback we use to stabilize $\Sigma^{\mathrm{red,lin}}$. We first make some constructions concerning the control forces $\mathscr{F} = \{F^1, \dots, F^m\}$. For $a \in \{1, \dots, m\}$, define a covector field F^a_{B} on B by

$$\langle F^a_{\mathsf{B}}(b); w_b\rangle = \langle F^a(q); \mathrm{hlft}_q(w_b)\rangle,$$

for $q \in \pi_{\mathsf{B}}^{-1}(b)$. Denote $\mathscr{F}_{\mathsf{B}} = \{F^1_{\mathsf{B}}, \dots, F^m_{\mathsf{B}}\}$. Also define functions f^a_{B}, $a \in \{1, \dots, m\}$, on B by

$$f^a_{\mathsf{B}}(b) = \langle F^a(q); X(q)\rangle,$$

for $q \in \pi_{\mathsf{B}}^{-1}(b)$. Now define linear maps $F_{\mathsf{B}}(b) \in L(\mathbb{R}^m; \mathsf{T}^*_b\mathsf{B})$ and $f_{\mathsf{B}}(b) \in L(\mathbb{R}^m; \mathbb{R})$ by

$$F_{\mathsf{B}}(b)\cdot u = \sum_{a=1}^m u^a F^a_{\mathsf{B}}(b), \quad f_{\mathsf{B}}(b)\cdot u = \sum_{a=1}^m u^a f^a_{\mathsf{B}}(b).$$

With these definitions we have the following result that characterizes the linear maps $B_{\Sigma,2}(b_0)$ and $B_{\Sigma,3}(b_0)$ that arise in Proposition S3.13.

Lemma S3.33. *The following statements hold:*

(i) $B_{\Sigma,2}(b_0) = \mathbb{G}_{\mathsf{B}}(b_0)^{\sharp} \circ F_{\mathsf{B}}(b_0)$*;*

(ii) $B_{\Sigma,3}(b_0) = \frac{f_{\mathsf{B}}(b_0)}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)}$.

Proof. Let $q_0 \in \pi_{\mathsf{B}}^{-1}(b_0)$. We compute

$$\begin{aligned} B_{\Sigma,2}(b_0) \cdot u &= \sum_{a=1}^{m} u^a Y_{\mathsf{B},a}(b_0) = \sum_{a=1} u^a T_{q_0}\pi_{\mathsf{B}}(Y_a(q_0)) \\ &= \sum_{a=1} u^a T_{q_0}\pi_{\mathsf{B}}(\mathrm{hor}(Y_a(q_0))) = \sum_{a=1}^{m} u^a \mathrm{hor}(\mathbb{G}^\sharp(F^a(q_0))) \\ &= \sum_{a=1}^{m} \mathbb{G}_{\mathsf{B}}^\sharp(F_{\mathsf{B}}^a(b_0)) = \mathbb{G}_{\mathsf{B}}(b_0)^\sharp \circ F_{\mathsf{B}}(b_0) \cdot u, \end{aligned}$$

and

$$\begin{aligned} B_{\Sigma,3}(b_0) \cdot u &= \sum_{a=1}^{m} u^a \frac{\mathbb{G}(Y_a(q_0), X(q_0))}{(\|X\|_{\mathbb{G}}^2)_{\mathsf{B}}(b_0)} = \sum_{a=1}^{m} u^a \frac{\langle F^a(q_0); X(q_0)\rangle}{(\|X\|_{\mathbb{G}}^2)_{\mathsf{B}}(b_0)} \\ &= \frac{f_{\mathsf{B}}(b_0) \cdot u}{(\|X\|_{\mathbb{G}}^2)_{\mathsf{B}}(b_0)}. \end{aligned}$$

■

For PD control for the reduced linearization, it will be necessary to place restrictions on the proportional and derivative gains that respect the horizontal/vertical splitting of tangent spaces to Q; this is a complication added by the fact that we are stabilizing a *relative* equilibrium. The following definition indicates how this is done.

Definition S3.34 (Compatible gain matrix and control vector). Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-forced simple mechanical control system and let X be a vector field on Q. For $K \in \Sigma_2((\mathbb{R}^m)^*)$, define $A_K \in \Gamma^\infty(T_2^0(\mathsf{TQ}))$ by

$$A_K(u_q, v_q) = K(F_\Sigma(q)^*(u_q), F_\Sigma(q)^*(v_q)).$$

We make the following definitions:

(i) $K \in \Sigma_2((\mathbb{R}^m)^*)$ is ***compatible*** with X at $q \in \mathsf{Q}$ if $A_K^\flat(\mathsf{H}_q\mathsf{Q}) \subset \mathrm{ann}(\mathsf{V}_q\mathsf{Q})$;

(ii) $u \in \mathbb{R}^m$ is ***compatible*** with X at $q \in \mathsf{Q}$ if $F_\Sigma(q) \cdot u \in \mathrm{ann}(\mathsf{H}_q\mathsf{Q})$.

We denote by $\Sigma_2((\mathbb{R}^m)^*)_{X(q)}$ the subset of $\Sigma_2((\mathbb{R}^m)^*)$ consisting of tensors compatible with X at q, and we denote by $\mathbb{R}^m_{X(q)}$ $\mathbb{R}^m_{X(q)}$set of compatible control vectorsthe subset of $\mathbb{R}^m$ consisting of vectors compatible with X at q. •

It is fairly clear that $\Sigma_2((\mathbb{R}^m)^*)_{X(q)}$ is a subspace of $\Sigma_2((\mathbb{R}^m)^*)$ and that $\mathbb{R}^m_{X(q)}$ is a subspace of $\mathbb{R}^m$. Let us denote by $\mathrm{pr}_{X(q)} : \mathbb{R}^m \to \mathbb{R}^m_{X(q)}$ the $\mathbb{G}_{\mathbb{R}^m}$-orthogonal projection.

The following result gives a useful, alternative characterization of compatibility, in the sense of the preceding definition.

Lemma S3.35 (Characterization of compatibility). *Let* $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ *be a* C^∞*-forced simple mechanical control system* *and let X be a complete infinitesimal symmetry for Σ for which the projection $\pi_{\mathsf{B}} \colon \mathsf{Q} \to \mathsf{B}$ onto the set of X-orbits is a surjective submersion. The following statements hold:*

(i) $K \in \Sigma_2((\mathbb{R}^m)^*)_{X(q)}$ if and only if $f_{\mathsf{B}}(\pi_{\mathsf{B}}(q)) \circ K^\sharp \circ F^*_{\mathsf{B}}(\pi_{\mathsf{B}}(q)) = 0$;

(ii) $u \in \mathbb{R}^m_{X(q)}$ if and only if $u \in \ker(F_{\mathsf{B}}(\pi_{\mathsf{B}}(q)))$.

Proof. (i) Let $b = \pi_{\mathsf{B}}(q)$. Note that, by definition of f_{B}, $u \in \ker(f_{\mathsf{B}}(b))$ if and only if $F_\Sigma(q) \cdot u \in \operatorname{ann}(\mathsf{V}_q\mathsf{Q})$. Now, for any $u \in \mathbb{R}^m$ and $w_b \in \mathsf{T}_b\mathsf{B}$, we compute

$$\begin{aligned} \langle F_\Sigma(q) \circ K^\sharp \circ F_{\mathsf{B}}(b)^*(w_b); X(q)\rangle &= \langle F_\Sigma(q) \circ K^\sharp \circ F_\Sigma(q)^*(\operatorname{hlft}_q(w_b)); X(q)\rangle \\ &= K(F_\Sigma(q)^*(\operatorname{hlft}_q(w_b)), F_\Sigma(q)^*(X(q))) \\ &= A_K(\operatorname{hlft}_q(w_b), X(q)). \end{aligned}$$

Thus $K^\sharp \circ F_{\mathsf{B}}(b)^*(w_b) \in \ker(f_{\mathsf{B}}(b))$ for every $w_b \in \mathsf{T}_b\mathsf{B}$ if and only if $A^\sharp_K(\operatorname{hlft}_q(w_b)) \in \operatorname{ann}(\mathsf{V}_q\mathsf{Q})$ for every $w_b \in \mathsf{T}_b\mathsf{B}$. But this is exactly this part of the lemma.

(ii) By definition of F_{B}, $u \in \ker(F_{\mathsf{B}}(b))$ if and only if $F_\Sigma(q) \cdot u \in \operatorname{ann}(\mathsf{H}_q\mathsf{Q})$, which is the result. ■

Now, with Lemma 10.22 in mind, we make the following definition.

Definition S3.36 (PD control for linearization of reduced system). Let $\Sigma = (\mathsf{Q}, \mathbb{G}, V, F, \mathscr{F}, \mathbb{R}^m)$ be a C^∞-forced simple mechanical control system and let X be a complete infinitesimal symmetry for Σ for which the projection $\pi_{\mathsf{B}} \colon \mathsf{Q} \to \mathsf{B}$ onto the set of X-orbits is a surjective submersion. For the linear control system $\Sigma^{\mathrm{red,lin}}$, a ***linear proportional-derivative (PD) control law at b_0*** is a linear state feedback of the form

$$u(x, v, \nu) = -K^\sharp_{\mathrm{P}} \circ F^*_{\mathsf{B}}(b_0) \cdot x - K^\sharp_{\mathrm{D}} \circ F^*_{\mathsf{B}}(b_0) \cdot v - k_{\mathrm{D}} \mathrm{pr}_{X(q_0)} \circ f_{\mathsf{B}}(b_0)^* \nu,$$

where $K_{\mathrm{P}}, K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$, $q_0 \in \pi^{-1}_{\mathsf{B}}(b_0)$, and where $k_{\mathrm{D}} \in \mathbb{R}$. If K_{D} and k_{D} are both zero, then u is a ***linear proportional control law***. •

Remark S3.37. There are some potentially confusing identifications that arise in properly interpreting the term "$k_{\mathrm{D}} \mathrm{pr}_{X(q_0)} f_{\mathsf{B}}(b_0)^* \nu$" in the linear PD control law. Let us address these. First of all, we had defined $f_{\mathsf{B}}(b_0)$ as a linear map from $\mathbb{R}^m$ to $\mathbb{R}$. However, were we to be consistent with forces being cotangent space-valued, then we would more properly have defined $f_{\mathsf{B}}(b_0)$ as taking values in $\mathbb{R}^*$. However, this is resolved by noting that there is a natural isomorphism of $\mathbb{R}$ with $\mathbb{R}^*$. Therefore, although by our definition $f_{\mathsf{B}}(b_0)^* \in L(\mathbb{R}^*; (\mathbb{R}^m)^*)$, we should really think of $f_{\mathsf{B}}(b_0) \in L(\mathbb{R}; (\mathbb{R}^m)^*)$. In order that $k_{\mathrm{D}} \mathrm{pr}_{X(q_0)} \circ f_{\mathsf{B}}(b_0)^* \nu$ lie in $\mathbb{R}^m_{X(q_0)} \subset \mathbb{R}^m$, as it should, we must, therefore, regard $\mathrm{pr}_{X(q_0)}$ as a linear map from $(\mathbb{R}^m)^*$ to $\mathbb{R}^m_{X(q_0)}$. To do this, we would properly replace $\mathrm{pr}_{X(q_0)}$ with $\mathrm{pr}_{X(q_0)} \circ \iota_m$, where $\iota_m \in L((\mathbb{R}^m)^*; \mathbb{R}^m)$ is the canonical isomorphism (i.e., defined by the standard inner product on $\mathbb{R}^m$). The reader might find it useful to keep these identifications in mind, since they will be made tacitly in the calculations below. •

Let us give the form for the closed-loop system for linear PD control.

Lemma S3.38 (Closed-loop system for linear PD control). *Let $K_{\mathrm{P}}, K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$, $q_0 \in \pi_{\mathsf{B}}^{-1}(b_0)$, define a quadratic function V_{cl} on $\mathsf{T}_{b_0}\mathsf{B}$ by*

$$V_{\mathrm{cl}}(x) = \tfrac{1}{2}\operatorname{Hess}(V_X)_{\mathsf{B}}(b_0)(x,x) + \tfrac{1}{2}F_{\mathsf{B}}(b_0) \circ K_{\mathrm{P}}^{\sharp} \circ F_{\mathsf{B}}(b_0)^* \cdot (x,x),$$

and define a symmetric $(0,2)$-tensor field A_{cl} on Q by

$$\begin{aligned} A_{\mathrm{cl}}(u_q, v_q) = A(u_q, v_q) - F_{\Sigma}(q) \circ K_{\mathrm{D}}^{\sharp} \circ F_{\Sigma}(q)^* \cdot (\operatorname{hor}(u_q), \operatorname{hor}(v_q)) \\ - k_{\mathrm{D}} \|\operatorname{pr}_{X(q)}(f_{\mathsf{B}}(b)^*)\|^2_{\mathbb{R}^m} \nu_X(u_q)\nu_X(v_q), \end{aligned}$$

where $\nu_X(w_q) \in \mathbb{R}$ is defined by $\operatorname{ver}(w_q) = \nu_X(w_q)X(q)$ for $w_q \in \mathsf{TQ}$.

Then, the closed-loop system for the linear PD control law defined by K_{P} and K_{D} is the linear system on $\mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}$ defined by the linear map

$$A_{\Sigma,\mathrm{cl}}(b_0) = \begin{bmatrix} 0 & \mathrm{id}_{\mathsf{T}_{b_0}\mathsf{B}} & 0 \\ -\mathbb{G}_{\mathsf{B}}(b_0)^{\sharp} \circ \operatorname{Hess} V_{\mathrm{cl}}^{\flat} & C_X(b_0) + A_{\mathrm{cl},\mathsf{B}}(b_0) & 2\operatorname{grad}_{\mathsf{B}} V_{\mathsf{B}}(b_0) + a_{\mathrm{cl},\mathsf{B}}(b_0) \\ 0 & -2\frac{\mathrm{d}V_{\mathsf{B}}(b_0)}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} + \alpha_{\mathrm{cl},\mathsf{B}}(b_0) & \frac{\langle A_{\mathrm{cl}}^{\flat}(X(q_0)); X(q_0)\rangle}{(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b_0)} \end{bmatrix},$$

where $A_{\mathrm{cl},\mathsf{B}}$, $a_{\mathrm{cl},\mathsf{B}}$, and $\alpha_{\mathrm{cl},\mathsf{B}}$ are as defined preceding the statement of Proposition S3.13.

Proof. Let us first perform some computations. For brevity, we let $b = b_0$ and $q = q_0$. For $w_b \in \mathsf{T}_b\mathsf{B}$ and $\beta_b \in \mathsf{T}_b^*\mathsf{Q}$ we compute

$$\begin{aligned} A_{\mathrm{cl},\mathsf{B}}(w_b) &= T_q\pi_{\mathsf{B}} \circ \mathbb{G}(q)^{\sharp} \circ A_{\mathrm{cl}}(q)^{\flat} \circ \operatorname{hlft}_q(w_b) \\ &= A_{\mathsf{B}}(w_b) - T_q\pi_{\mathsf{B}} \circ \mathbb{G}(q)^{\sharp} \circ \operatorname{hor}^* \circ F_{\Sigma}(q) \circ K_{\mathrm{D}}^{\sharp} \circ F_{\Sigma}(q)^* \circ \operatorname{hlft}_q(w_b) \\ &= A_{\mathsf{B}}(w_b) - T_q\pi_{\mathsf{B}} \circ \mathbb{G}(q)^{\sharp} \circ F_{\Sigma}(q) \circ K_{\mathrm{D}}^{\sharp} \circ F_{\Sigma}(q)^* \circ \operatorname{hlft}_q(w_b) \\ &= A_{\mathsf{B}}(w_b) - \mathbb{G}_{\mathsf{B}}(b)^{\sharp} \circ F_{\mathsf{B}}(b) \circ K_{\mathrm{D}}^{\sharp} \circ F_{\mathsf{B}}(b)^*(w_b). \end{aligned}$$

Here hor^* denotes the projection onto $\operatorname{ann}(\mathsf{V}_q\mathsf{Q})$ associated with the decomposition $\mathsf{T}_q^*\mathsf{Q} = \operatorname{ann}(\mathsf{V}_q\mathsf{Q}) \oplus \operatorname{ann}(\mathsf{H}_q\mathsf{Q})$. In the third step in the calculation, we have used the fact that $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q)}$. We next compute

$$\begin{aligned} a_{\mathrm{cl},\mathsf{B}}(b) &= T_q\pi_{\mathsf{B}} \circ \mathbb{G}(q)^{\sharp} \circ A_{\mathrm{cl}}(q)^{\flat}(X(q)) \\ &= a_{\mathsf{B}}(b) - k_{\mathrm{D}} \|\operatorname{pr}_{X(q)}(f_{\mathsf{B}}(b)^*)\|^2_{\mathbb{R}^m} T_q\pi_{\mathsf{B}} \circ \mathbb{G}(q)^{\sharp}\nu \\ &= a_{\mathsf{B}}(b), \end{aligned}$$

where we think of ν as an element of $\operatorname{ann}(\mathsf{H}_q\mathsf{Q}) \subset \mathsf{T}_q^*\mathsf{Q}$. We compute

$$\begin{aligned}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b)\,\langle\alpha_{\mathrm{cl},\mathsf{B}}(b);w_b\rangle &= \langle A^\flat_{\mathrm{cl}}(\mathrm{hlft}_q(w_b));X(q)\rangle\\ &= (\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b)\langle\alpha_{\mathsf{B}}(b);w_b\rangle,\end{aligned}$$

using the definition of A_{cl}. Finally, we have

$$\langle A^\flat_{\mathrm{cl}}(X(q));X(q)\rangle = \langle A^\flat(X(q));X(q)\rangle - k_{\mathrm{D}}\|\mathrm{pr}_{X(q)}(f_{\mathsf{B}}(b)^*)\|^2_{\mathbb{R}^m}.$$

With these computations, and using Lemmas S3.33 and S3.35, the result follows from a direct computation. ■

Remark S3.39. In order that the closed-loop system have the desired form, it is essential to restrict the proportional gain K_{P} to lie in $\Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$. It is not, however, necessary to so restrict K_{D}, nor is it necessary to project $f_{\mathsf{B}}(b_0)^*$ onto $\mathbb{R}^m_{X(q_0)}$. However, by placing these restrictions on the form of the derivative feedback, we ensure that the horizontal and vertical components of the derivative feedback are separated. In practice, this may be useful. However, it is also possible to state Lemma S3.38 for the most general derivative feedback, and we leave the details of this to the reader. •

We may also make the corresponding definitions for stabilizability of the reduced system using linear PD control.

Definition S3.40 (Stabilizability of reduced linearized system using linear PD control). The system $\Sigma^{\mathrm{red,lin}}$ is

(i) ***stabilizable by proportional control*** if there exists a linear proportional control law that stabilizes the controlled equilibrium point $(0_{b_0}\oplus 0_{b_0}\oplus 0,\mathbf{0})$, and is

(ii) ***asymptotically stabilizable by proportional-derivative control*** if there exists a linear PD control law that asymptotically stabilizes the controlled equilibrium point $(0_{b_0}\oplus 0_{b_0}\oplus 0,\mathbf{0})$. •

We can now state conditions for stabilization of the reduced linearization, and of the reduced system, using linear PD control. First let us state a sufficient condition for stabilization of the linearization.

Theorem S3.41 (Stabilizability of reduced linearization by linear PD control). *The following statements hold:*

(i) if $\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)|\mathrm{coann}(\mathrm{image}(F_{\mathsf{B}}))$ *is positive-definite, then* $\Sigma^{\mathrm{red,lin}}$ *is stabilizable by proportional control, and* Σ^{red} *is stabilizable at* $(0_{b_0},1)$ *by linear proportional control;*

(ii) if

(a) $\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)|\mathrm{coann}(\mathrm{image}(F_{\mathsf{B}}(b_0)))$ *is positive-definite and*

(b) $\Sigma^{\mathrm{red,lin}}$ *is STLC from* $(0_{b_0}\oplus 0_{b_0}\oplus 0)$,

then $\Sigma^{\mathrm{red,lin}}$ *is asymptotically stabilizable by PD control, and* Σ^{red} *is locally asymptotically stabilizable at* $(0_{b_0},1)$ *by linear PD control.*

Proof. Let us define the closed-loop linearized reduced effective energy by

$$E_{\chi}^{\mathrm{red,cl}}(x,v,\nu) = \tfrac{1}{2}\mathbb{G}_{\mathsf{B}}(b_0)(v,v) + \tfrac{1}{2}V_{\mathrm{cl}}(x) + \tfrac{1}{2}(\|X\|_{\mathbb{G}}^2)_{\mathsf{B}}(b_0)\nu^2,$$

where V_{cl} is as defined in Lemma S3.38. For a general linear PD control defined by K_{P}, K_{D}, and k_{D}, we compute, following the proof of Theorem S3.25,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}&E_{\chi}^{\mathrm{red,cl}}(x(t)\oplus\dot{x}(t)\oplus\nu(t))\\ &= A_{\mathrm{cl}}(\mathrm{hlft}_{q_0}(\dot{x}(t))+\nu(t)X(q_0), \mathrm{hlft}_{q_0}(\dot{x}(t))+\nu(t)X(q_0)),\end{aligned}$$

where A_{cl} is as defined in Lemma S3.38, and where $t\mapsto x(t)\oplus\dot{x}(t)\oplus\nu(t)$ is a trajectory for the closed-loop system.

(i) In the case that $\mathrm{image}(F_{\Sigma}(q_0))\subset\mathrm{ann}(\mathsf{V}_{q_0}\mathsf{Q})$, it follows that $f_{\mathsf{B}}(b_0)=0$, and that $\Sigma_2((\mathbb{R}^m)^*)_{X(q_0)} = \Sigma_2((\mathbb{R}^m)^*)$. In this case, the result follows in the same manner as part (i) of Theorem 10.26. For simplicity, we therefore assume that $\mathrm{image}(F_{\Sigma}(q_0)) \not\subset \mathrm{ann}(\mathsf{V}_{q_0}\mathsf{Q})$. Let $q_0\in\pi_{\mathsf{B}}^{-1}(b_0)$ and define $\mathsf{V}=\mathrm{coann}(\mathrm{image}(F_{\Sigma}(q_0)))+\mathsf{V}_{q_0}\mathsf{Q}$. Note that, by our assumption that $\mathrm{image}(F_{\Sigma}(q_0))\not\subset\mathrm{ann}(\mathsf{V}_{q_0}\mathsf{Q})$, it holds that $\mathsf{V}=\mathrm{coann}(\mathrm{image}(F_{\Sigma}(q_0)))\oplus\mathsf{V}_{q_0}\mathsf{Q}$.

The following lemma contains the essential observation in this part of the proof.

Lemma. *Let $\phi\colon \mathsf{T}_{q_0}\mathsf{Q}\to\mathbb{R}$ be a quadratic function (i.e., one that satisfies $\phi(\lambda v_{q_0})=\lambda^2\phi(v_{q_0})$ for every $\lambda\in\mathbb{R}$ and $v_{q_0}\in\mathsf{T}_{q_0}\mathsf{Q}$) satisfying*

(i) $\mathrm{d}\phi(v_{q_0})\in\mathrm{ann}(\mathsf{V})$ for each $v_{q_0}\in\mathsf{T}_{q_0}\mathsf{Q}$ and

(ii) $\langle\mathrm{d}\phi(v_{q_0});X(q_0)\rangle=0$.

Then there exists $K\in\Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ such that $\mathrm{Hess}\,\phi = F_{\Sigma}(q_0)\circ K^{\sharp}\circ F_{\Sigma}(q_0)^*$.

Proof. Without loss of generality, suppose that $F_{\Sigma}(q_0)$ is injective. Let $\{e_1,\dots,e_n\}$ be a basis for $\mathsf{T}_{q_0}\mathsf{Q}$ with the following properties:

1. $\{e_{k+1},\dots,e_n\}$ is a basis for $\mathrm{coann}(\mathrm{image}(F_{\Sigma}(q_0)))$;
2. $e^k = X(q_0)$.

Let $(v^1,\dots,v^n)$ be the induced coordinates for $\mathsf{T}_{q_0}\mathsf{Q}$. Then the function ϕ in the statement of the lemma is a function of the coordinates $v^1,\dots,v^{k-1}$. The matrix representation for F_{Σ} in the basis is

$$[F_{\Sigma}(q_0)] = \begin{bmatrix}\boldsymbol{F}_1\\ \boldsymbol{F}_2\\ \boldsymbol{0}_{(n-k)\times m}\end{bmatrix},$$

where $\boldsymbol{F}_1\in\mathbb{R}^{(k-1)\times m}$ and $\boldsymbol{F}_2\in\mathbb{R}^{1\times m}$. To prove the lemma, we must solve the equation

$$\begin{bmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \\ \boldsymbol{0}_{(n-k)\times m} \end{bmatrix} \boldsymbol{K} \begin{bmatrix} \boldsymbol{F}_1^T & \boldsymbol{F}_2^T & \boldsymbol{0}_{m\times(n-k)} \end{bmatrix} \\ = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{0}_{(k-1)\times 1} & \boldsymbol{0}_{(k-1)\times(n-k)} \\ \boldsymbol{0}_{1\times(k-1)} & 0 & \boldsymbol{0}_{1\times(n-k)} \\ \boldsymbol{0}_{(n-k)\times(k-1)} & \boldsymbol{0}_{(n-k)\times 1} & \boldsymbol{0}_{(n-k)\times(n-k)} \end{bmatrix}, \quad \text{(S3.23)}$$

for $\boldsymbol{K} \in \mathbb{R}^{m\times m}$, where $\boldsymbol{\Phi} = [\operatorname{Hess}\phi] \in \mathbb{R}^{(k-1)\times(k-1)}$. Since $F_\Sigma(q_0)$ is assumed to be injective, the equation

$$\begin{bmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \end{bmatrix} \boldsymbol{K} \begin{bmatrix} \boldsymbol{F}_1^T & \boldsymbol{F}_2^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{0}_{(k-1)\times 1} \\ \boldsymbol{0}_{1\times(k-1)} & 0 \end{bmatrix}$$

can be solved for $\boldsymbol{K}$, and this same matrix clearly solves (S3.23). Next we need to show that, if $\boldsymbol{K}$ is the matrix representative for $K \in \Sigma_2((\mathbb{R}^m)^*)$, then it holds that $K \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$. This is true, however, since $A_K(q_0) = \operatorname{Hess}\phi$, and since it is clear that $(\operatorname{Hess}\phi)^\flat(\mathsf{H}_{q_0}\mathsf{Q}) \subset \operatorname{ann}(\mathsf{V}_{q_0}\mathsf{Q})$. ▼

To complete the proof of this part of the theorem, note that the hypotheses, along with the lemma, ensure that there exists $K_{\mathrm{P}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ such that $\operatorname{Hess} V_X + F_\Sigma(q_0) \circ K_{\mathrm{P}}^\sharp \circ F_\Sigma(q_0)^*$ is positive-definite on $\mathsf{H}_{q_0}\mathsf{Q}$. Since $K_{\mathrm{P}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$, it holds that

$$F_\Sigma(q_0) \circ K_{\mathrm{P}}^\sharp \circ F_\Sigma(q_0)^*(v_{q_0}, w_{q_0}) = F_{\mathsf{B}}(b_0) \circ K_{\mathrm{P}}^\sharp \circ F_{\mathsf{B}}(b_0)^*(T\pi_{\mathsf{B}}(v_{q_0}), T\pi_{\mathsf{B}}(w_{q_0})$$

for all $v_{q_0}, w_{q_0} \in \mathsf{T}_{q_0}\mathsf{Q}$. Thus $E_\chi^{\mathrm{red,cl}}$ is a Lyapunov function for the closed-loop system, and so stability follows from Theorem 6.14.

(ii) To prove this part of the theorem, we suppose that a linear proportional control law has already been designed as in part (i), and thus we may suppose that $\operatorname{Hess}(V_X)_{\mathsf{B}}$ is positive-definite. For simplicity, and without loss of generality, let us also assume that there is no open-loop dissipation. We next claim that the derivative part of the PD control law is dissipative as in Section 10.1.5. To see this, let $b_a \in \mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}$, $a \in \{1, \dots, m\}$, be defined such that

$$B_\Sigma(b_0) \cdot u = \sum_{a=1}^m b_a u^a.$$

Using Proposition S3.13 and Lemma S3.33, and thinking of b_a as a constant vector field on $\mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}$, we compute

$$\mathscr{L}_{b_a} E_\chi^{\mathrm{red,lin}}(x \oplus v \oplus \nu) = \langle F_{\mathsf{B}}^a(b_0); v\rangle + f_{\mathsf{B}}^a(b_0)\nu.$$

Now define $u_{\mathrm{diss}} \colon \mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R} \to \mathbb{R}^m$ by $u_{\mathrm{diss}}^a(x, v, \nu) = -\mathscr{L}_{b_a} E_\chi^{\mathrm{red,lin}}$. We next note that, if $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$,

$$\operatorname{pr}_{X(q_0)} \circ F_{\mathsf{B}}(b_0)^*(v) = 0, \qquad v \in \mathsf{T}_{b_0}\mathsf{B},$$

and

$$F_{\mathsf{B}}(b_0) \circ K_{\mathrm{D}}^{\sharp}(f_{\mathsf{B}}(b_0)^*) = 0,$$

with both equalities following from Lemma S3.35. These facts then give the following formula:

$$\begin{aligned} &F_{\mathsf{B}}(b_0)(K_{\mathrm{D}}^{\sharp} + k_{\mathrm{D}}\mathrm{pr}_{X(q_0)})(u_{\mathrm{diss}}(x \oplus v \oplus \nu)) \\ &\qquad = -F_{\mathsf{B}}(b_0) \circ K_{\mathrm{D}}^{\sharp} \circ F_{\mathsf{B}}(b_0)^*(v) - k_{\mathrm{D}} F_{\mathsf{B}}(b_0) \circ \mathrm{pr}_{X(q_0)}(f_{\mathsf{B}}(b_0)^*)\nu. \end{aligned}$$

This shows that the derivative control law associated with $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ and $k_{\mathrm{D}} \in \mathbb{R}$ is equivalent to the feedback

$$x \oplus v \oplus \nu \mapsto (K_{\mathrm{D}}^{\sharp} + k_{\mathrm{D}}\mathrm{pr}_{X(q_0)})(u_{\mathrm{diss}}(x \oplus v \oplus \nu)).$$

We next claim that it is possible to choose $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ and $k_{\mathrm{D}} \in \mathbb{R}$ such that the linear map $A_{K_{\mathrm{D}},k_{\mathrm{D}}}$ defined by

$$\mathbb{R}^m \ni u \mapsto (K_{\mathrm{D}}^{\sharp} + k_{\mathrm{D}}\mathrm{pr}_{X(q_0)})(u) \in (\mathbb{R}^m)^* \tag{S3.24}$$

is positive-definite with respect to the standard inner product on $\mathbb{R}^m$. This will follow if we can show that K_{D} can be chosen so that the linear map in (S3.24) is positive-definite on a complement to $\mathbb{R}^m_{X(q_0)}$ in $\mathbb{R}^m$. To prove this, we proceed indirectly. Denote by $F_{F_{\mathsf{B}},f_{\mathsf{B}}}$ the linear map from $\mathbb{R}^m$ to $\mathsf{T}^*_{b_0}\mathsf{B} \times \mathbb{R}$ defined by

$$F_{F_{\mathsf{B}},f_{\mathsf{B}}}(u) = (F_{\mathsf{B}} \cdot u) \oplus (f_{\mathsf{B}}(b_0) \cdot u).$$

We assume, without loss of generality, that this map is injective. On $\mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}$ consider the symmetric bilinear map defined by

$$v \oplus \nu \mapsto F_{F_{\mathsf{B}},f_{\mathsf{B}}} \circ A^{\sharp}_{K_{\mathrm{D}},k_{\mathrm{D}}} \circ F^*_{F_{\mathsf{B}},f_{\mathsf{B}}}(v \oplus \nu).$$

The computations of Lemma S3.38 show that this map can be expressed as

$$v \oplus \nu \mapsto (F_{\mathsf{B}}(b_0) \circ K_{\mathrm{D}}^{\sharp} \circ F_{\mathsf{B}}(b_0)^*(v)) \oplus (k_{\mathrm{D}}\|\mathrm{pr}_{X(q_0)}(f_{\mathsf{B}}(b_0)^*)\|^2_{\mathbb{R}^m}\nu). \tag{S3.25}$$

From the lemma used in the proof of the first part of the theorem, we know that it is possible to choose $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ such that the symmetric bilinear map defined by

$$v \mapsto F_{\mathsf{B}}(b_0) \circ K_{\mathrm{D}}^{\sharp} \circ F_{\mathsf{B}}(b_0)^*(v)$$

is positive-definite on a complement to coann(image($F_{\mathsf{B}}(b_0)$)). Therefore, it is possible to choose $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ and $k_{\mathrm{D}} > 0$ such that the symmetric bilinear map defined by (S3.25) is positive-definite on a complement to coann(image($F_{F_{\mathsf{B}},f_{\mathsf{B}}}$)). This, however, implies that $A_{K_{\mathrm{D}},k_{\mathrm{D}}}$ is positive-definite. That derivative feedback is dissipative now follows from Remark 10.18–2. That derivative control asymptotically stabilizes $\Sigma^{\mathrm{red,lin}}$ follows from Lemma 10.17, along with the assumption that $\Sigma^{\mathrm{red,lin}}$ is STLC from $0_{b_0} \oplus 0_{b_0} \oplus 0$. ■

Remark S3.42. Theorem 10.26, the result for stabilization of equilibria for linear mechanical systems using PD control, gives necessary and sufficient conditions for stabilizability and asymptotic stabilizability. This is because, in Theorem 6.42, we were able to give necessary and sufficient conditions for stability of linear mechanical systems with dissipative forces. However, since the linearization of the reduced equations for a relative equilibrium involve gyroscopic forces, it is more difficult to give sharp stability conditions for stability (cf. Exercise E6.10). Moreover, while the linear controllability condition for systems with no external forces can be simplified to a computation involving vector spaces of dimension equal to the dimension of the configuration manifold, rather than the dimension of its tangent bundle (see Theorem 7.31), no such results are known when gyroscopic forces are present. Thus there is the possibility of weakening, or improving the computability of, the hypotheses of Theorem S3.41 in multiple directions. •

S3.5.3 Stabilization of reduced system using linear control law

We may also define the notion of stabilizability of the nonlinear system Σ^{red} using linear state feedback. To do so, we recall from Definition 10.29 the notion of a near identity diffeomorphism. If $(\psi, \mathcal{U}_0, \mathcal{U}_1)$ is a near identity diffeomorphism at $b_0 \in \mathsf{B}$, then we denote by $T\psi^{-1} \times \mathrm{id}_{\mathbb{R}} \colon \mathsf{T}\mathcal{U}_1 \times \mathbb{R} \to (\mathcal{U}_0 \times \mathsf{T}_{b_0}\mathsf{B}) \times \mathbb{R}$ the map

$$T\psi^{-1} \times \mathrm{id}_{\mathbb{R}}(w_b, v) = (T\psi^{-1}(w_b), v).$$

We now make a definition.

Definition S3.43 (Stabilizability of reduced system using linear PD control). If u_{lin} is a linear PD control law for $\Sigma^{\mathrm{red,lin}}$, then an ***implementation*** of u^{lin} for Σ^{red} is a state feedback on $\mathsf{T}\mathcal{U}_1 \times \mathbb{R}$ given by $u_{\mathrm{nonlin}} = u_{\mathrm{lin}} \circ (T\psi^{-1} \times \mathrm{id}_{\mathbb{R}})$, where $(\psi, \mathcal{U}_0, \mathcal{U}_1)$ is a near identity diffeomorphism at b_0.

Furthermore, we say that Σ^{red} is

(i) ***stabilizable by linear proportional control*** at $(0_{b_0}, 1)$ if there exists a linear proportional control law u_{lin} and an implementation of this control law for Σ^{red} which stabilizes $((0_{b_0}, 1), \mathbf{0})$, and is

(ii) ***locally asymptotically stabilizable by linear proportional-derivative control*** at $(0_{b_0}, 1)$ if there exists a linear PD control law u_{lin} and an implementation of this control law for Σ^{red} which locally asymptotically stabilizes $((0_{b_0}, 1), \mathbf{0})$. •

The following result follows from Theorem S3.41 in the same manner in which Theorem 10.32 follows from Theorem 10.26.

Theorem S3.44. *For a near identity diffeomorphism $(\psi, \mathcal{U}_0, \mathcal{U}_1)$ at b_0 and a linear PD control law*

$$u_{\mathrm{lin}}(x, v, \nu) = -K_{\mathrm{P}}^{\sharp} \circ F_{\mathsf{B}}(b_0)^*(x) - K_{\mathrm{D}}^{\sharp} \circ F_{\mathsf{B}}(b_0)^*(v) - k_{\mathrm{D}}\mathrm{pr}_{X(q_0)} \circ f_{\mathsf{B}}(b_0)^*\nu,$$

for $\Sigma^{\mathrm{red,lin}}$*, define a state feedback for* Σ^{red} *on* $\mathsf{T}\mathcal{U}_1 \times \mathbb{R}$ *by* $u_{\mathrm{nonlin}} = u_{\mathrm{lin}} \circ (T\chi^{-1} \times \mathrm{id}_{\mathbb{R}})$*. The following statements hold.*

(i) *If* $\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)|\mathrm{coann}(\mathrm{image}(F_{\mathsf{B}}(b_0)))$ *is positive-definite, then there exists* $K_{\mathrm{P}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ *such that* $\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0) + F_{\mathsf{B}}(b_0) \circ K_{\mathrm{P}}^{\sharp} \circ F_{\mathsf{B}}(b_0)^* \in \Sigma_2(\mathsf{T}_{q_0}\mathsf{Q})$ *is positive-definite. Furthermore,* $((0_{b_0}, 1), \mathbf{0})$ *is stabilized by the state feedback* u_{nonlin} *with* K_{P} *so chosen and with* K_{D} *positive-semidefinite.*

(ii) *Suppose that* $\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)|\mathrm{coann}(\mathrm{image}(F_{\mathsf{B}}(b_0)))$ *is positive-definite, that* K_{P} *is chosen as in part (i), that* $\Sigma^{\mathrm{red,lin}}$ *is STLC from* $0_{b_0} \oplus 0_{b_0} \oplus 0$*, and that* $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ *is positive-definite. Then* $((0_{b_0}, 1), \mathbf{0})$ *is locally asymptotically stabilized by the state feedback* u_{nonlin} *with* K_{P} *and* K_{D} *so chosen.*

S3.5.4 Stabilization of reduced equations using potential shaping

In this section we define the notion of PD control for the system Σ^{red}, and we give sufficient conditions for stabilization of an equilibrium point for the reduced system using PD control. The results here are analogous to those in Section 10.4 in the text.

To get started, we need some notation. We let $\mathcal{F}_{\mathsf{B}}$ be the codistribution on B defined by $\mathcal{F}_{\mathsf{B},b} = \mathrm{image}(F_{\mathsf{B}}(b))$. Following our notation in Section 10.4, we denote by $C^\infty(\mathsf{B})_{\mathcal{F}_{\mathsf{B}}}$ the set of C^∞-functions ϕ on B for which $\mathrm{d}\phi \in \Gamma^\infty(\mathcal{F})$. By $\mathcal{F}_{\mathsf{B}}^{(\infty)}$ we denote the codistribution annihilating the involutive closure of $\mathrm{coann}(\mathcal{F}_{\mathsf{B}})$ (see the discussion in Section 10.4.1), and recall that $\mathcal{F}_{\mathsf{B}}$ is ***totally regular*** at b_0 if $\mathcal{F}_{\mathsf{B}}$ and $\mathcal{F}_{\mathsf{B}}^{(\infty)}$ are both regular at b_0.

With this notation, let us define what we mean by PD control for Σ^{red}.

Definition S3.45 (PD control for reduced system). Suppose that $\mathcal{F}_{\mathsf{B}}$ is totally regular at b_0. A ***proportional-derivative (PD) control law*** at $(0_{b_0}, 1)$ is a state feedback satisfying

$$F_{\mathsf{B}}(b) \cdot u(w_b, v) = -\mathrm{d}V_{\mathrm{P}}(b) - K_{\mathrm{D}}^{\sharp} \circ F_{\mathsf{B}}(b)^*(w_b) - k_{\mathrm{D}}\mathrm{pr}_{X(q)} \circ f_{\mathsf{B}}(b)^* v,$$

where $V_{\mathrm{P}} \in C^\infty(\mathsf{B})_{\mathcal{F}_{\mathsf{B}}}$, $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q)}$, $q \in \pi_{\mathsf{B}}^{-1}(b)$, and $k_{\mathrm{D}} \in \mathbb{R}$. If K_{D} and k_{D} are zero, then u is a ***proportional control law***. •

Correspondingly, one has the following notions of stabilizability.

Definition S3.46 (Stabilizability of reduced system using PD control). Suppose that $\mathcal{F}_{\mathsf{B}}$ is totally regular at b_0. Σ^{red} is

(i) ***stabilizable by proportional control*** at $(0_{b_0}, 1)$ if there exists a proportional control law at $(0_{b_0}, 1)$ such that the closed-loop system possesses $(0_{b_0}, 1)$ as a stable equilibrium point, and is

(ii) ***locally asymptotically stabilizable by proportional-derivative control*** at $(0_{b_0}, 1)$ if there exists a PD control law at $(0_{b_0}, 1)$ such that the closed-loop system possesses $(0_{b_0}, 1)$ as a locally asymptotically stable equilibrium point. •

It is now straightforward to state a sufficient condition for stabilization using PD control.

Theorem S3.47 (Stabilization of reduced system using PD control). *Suppose that* $\mathcal{F}_{\mathsf{B}}$ *is totally regular at* b_0 *and that* $\mathscr{F}_{\mathsf{B}} = \{\mathrm{d}\phi^1, \dots, \mathrm{d}\phi^k, F^{k+1}, \dots, F^m\}$ *is proportionally adapted at* b_0*. Then the following statements hold.*

(i) If $\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)|\mathrm{coann}(\mathcal{F}^{(\infty)}_{\mathsf{B},b_0})$ *is positive-definite, then there exists* $K_{\mathrm{P}} \in \Sigma_2((\mathbb{R}^k)^*)_{X(q_0)}$ *such that* $V_{\mathrm{cl}} = (V_X)_{\mathsf{B}} + \frac{1}{2}\sum_{a,c=1}^{k}(K_{\mathrm{P}})_{ac}\phi^a\phi^c$ *is locally positive-definite about* b_0*. Furthermore,* $(0_{b_0}, 1)$ *is stabilized by the proportional control law*

$$u(w_b, v) = -\sum_{a,c=1}^{k}(K_{\mathrm{P}})_{ac}\phi^c(b)\boldsymbol{e}_a,$$

where $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ *is the standard basis for* $\mathbb{R}^m$*.*

(ii) Suppose that $\mathrm{Hess}(V_X)_{\mathsf{B}}(b_0)|\mathrm{coann}(\mathcal{F}^{(\infty)}_{\mathsf{B},b_0})$ *is positive-definite, that* K_{P} *is chosen as in part (i), that* $\Sigma^{\mathrm{red,lin}}$ *is STLC from* $0_{b_0} \oplus 0_{b_0} \oplus 0$*, that* $K_{\mathrm{D}} \in \Sigma_2((\mathbb{R}^m)^*)_{X(q_0)}$ *is positive-definite, and that* $k_{\mathrm{D}} > 0$*. Then* $(0_{b_0}, 1)$ *is locally asymptotically stabilized by the PD control law*

$$u(w_b, v) = -\sum_{a,c=1}^{k}(K_{\mathrm{P}})_{ac}\phi^c(b)\boldsymbol{e}_a - K_{\mathrm{D}}^{\sharp} \circ F_{\mathsf{B}}(b)^*(w_b) - k_{\mathrm{D}}\mathrm{pr}_{X(q)} \circ f_{\mathsf{B}}(b)^* v.$$

Proof. Let us define a function E on $\mathsf{TB} \times \mathbb{R}$ by

$$E(w_b, v) = \frac{1}{2}\mathbb{G}_{\mathsf{B}}(w_b, w_b) + (V_X)_{\mathsf{B}}(b) + \frac{1}{2}\sum_{a,c=1}^{k}(K_{\mathrm{P}})_{ac}\phi^a(b)\phi^c(b) + \frac{1}{2}(\|X\|^2_{\mathbb{G}})_{\mathsf{B}}(b)(v-1)^2.$$

Let us compute the time-derivative of E along controlled trajectories $t \mapsto ((\eta'(t), v(t)), u(t))$ of Σ^{red} when the control is a general PD control law. We compute

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\eta'(t), v(t)) = A_{\mathrm{cl}}(\mathrm{hlft}_{\gamma(t)}(\eta'(t)) + v(t)X(\gamma(t)), \mathrm{hlft}_{\gamma(t)}(\eta'(t)) + v(t)X(\gamma(t)),$$

where A_{cl} is as defined in Lemma S3.38.

(i) That there exists K_{P} as asserted follows in the same manner as the analogous statement in Theorem 10.43. To prove the second assertion of this part of the theorem, one notes that $\frac{\mathrm{d}E}{\mathrm{d}t}(\eta'(t), v(t)) \leq 0$. The result then follows from Theorem 6.14.

(ii) For simplicity, let us suppose that the open-loop dissipation is zero, and let $\Sigma^{\mathrm{red,cl}}$ be the control-affine system on $\mathsf{TB} \times \mathbb{R}$ obtained after the potential has been shaped as in part (i). From the proof of part (ii) of Theorem S3.41, we know that the derivative control is dissipative. Thus the result will follow from Lemma 10.17 if we can show that $\Sigma^{\mathrm{red,cl}}$ has a controllable linearization at $(0_{b_0}, 1)$. We observe that the linearization of $\Sigma^{\mathrm{red,cl}}$ at $(0_{b_0}, 1)$ is given by the linear control system $(\mathsf{T}_{b_0}\mathsf{B} \oplus \mathsf{T}_{b_0}\mathsf{B} \oplus \mathbb{R}, A^{\mathrm{cl}}_{\Sigma}(b_0), B_{\Sigma}(b_0))$, where $B_{\Sigma}(b_0)$ is as defined in Proposition S3.13, where

$$A_{\Sigma}(b_0) + B_{\Sigma}(b_0)\begin{bmatrix}0 & -K^{\sharp}_{\mathsf{P}} \circ F^{*}_{\mathsf{B}}(b_0) & 0\end{bmatrix},$$

and where $A_{\Sigma}(b_0)$ is as defined in Proposition S3.13. The proof is completed by making the observation that if a linear system (V, A, B) is STLC from 0, then so too is the linear system $(\mathsf{V}, A + B \circ F, B)$, for any $F \in L(\mathsf{V}; \mathbb{R}^m)$. We leave the fairly straightforward proof of this fact to the reader to look up in [Wonham 1985]. ■

S4

Optimal control theory

In the field of control theory, optimal control theory is one of the more distinguished subjects. For example, the linear stabilization methods discussed in Section 10.5.1 have their roots in optimal control theory. However, the subject goes well beyond the linear theory. The seminal work in the subject is without question that which led to the publication of [Pontryagin, Boltyanskii, Gamkrelidze, and Mishchenko 1961]. This work came to fruition in the years following the Second World War. In this work appeared the first statement and proof of the Maximum Principle, to which a significant portion of this chapter is dedicated. Another cornerstone of optimal control theory is the Dynamic Programming Principle, the initial contributions to which are described in the book of Bellman [1957]. Since the appearance of these basic works, many fundamental contributions to the subject have been made, and it would be nearly impossible to give an accurate outline of the research literature. However, the contributions of Kalman [1960] to the linear theory have been very important, since they provide implementable design tools that see wide use in practice. Contributions to the geometric theory of optimal control, particular geometric formulations of the Maximum Principle, have been made by Sussmann in a series of papers in the late 1990's (see, for example, [Sussmann 1997, 2000]). Sussmann's work is devoted to giving versions of the Maximum Principle that hold with very weak hypotheses, and in quite general control theoretic frameworks. Beyond the Maximum Principle, the subject of higher-order conditions for optimality are important in optimal control theory. Significant contributions have been made here by Krener [1977]. Second-order conditions (thinking of the Maximum Principle as being a first-order condition) have been studied in great depth by Agrachev and coauthors, and these results are detailed in the book of Agrachev and Sachkov [2004], along with a nice geometric formulation of the Maximum Principle. For additional references, we refer to the books [Agrachev and Sachkov 2004, Jurdjevic 1997, Lee and Markus 1967].

One way to view the Maximum Principle in optimal control theory is as a generalization of the classical calculus of variations. This is the point of view with which we start this chapter, in Section S4.1, following the excellent account of Sussmann and Willems [1997]. After this motivation, we precisely state the Maximum Principle in Section S4.2, following the account of Sussmann [1997], although our treatment is far less general than that of Sussmann. We make use here of the Hamiltonian framework developed in Section S1.1. In Section S4.3 we discuss some ways in which one can interpret the Maximum Principle. Here we make connections between optimal control theory and controllability theory. This hints at something quite deep, and we refer to the treatment in [Agrachev and Sachkov 2004, Lee and Markus 1967] for further discussion along these lines. In Section S4.4 we give a version of the Maximum Principle for affine connection control systems. Here we make use of the tangent bundle geometry described in Sections S1.2 and S1.3. In Sections S4.5 and S4.6 we consider the special cases of optimal control problems that minimize the inputs (in a certain sense) and time, respectively. We conclude the chapter with an analysis, in Section S4.7, of force- and time-optimal control for the planar rigid body system discussed in the text.

S4.1 Going from the calculus of variations to the Maximum Principle

A profitable view of the Maximum Principle is that it is a generalization of the classical calculus of variations, as discussed briefly in the first two sections in Section 4.3. However, this connection is not obvious when one simply states the Maximum Principle and Hamilton's Principle side-by-side. Therefore, in this section we explicitly develop the calculus of variations in a manner that fairly obviously leads to the Maximum Principle. We suppose all data to be of class C^∞ in this section.

S4.1.1 Some additional discussion of the calculus of variations

In Section 4.3 we stated a problem, Problem 4.37, in the calculus of variations, and a necessary condition, Theorem 4.38, for the solution of this problem. In this section we give two additional necessary conditions that will be essential to understanding how the Maximum Principle arises from the calculus of variations. In this section, to simplify notation and to emphasize geometry, we deal exclusively with time-independent Lagrangians. The extension to time-dependent Lagrangians is straightforward. We do not prove the results in this section, but refer the reader to any somewhat more than basic account of the calculus of variations, e.g., [Giaquinta and Hildebrandt 1996].

Let $L\colon \mathsf{TM} \to \mathbb{R}$ be a C^∞-Lagrangian. To state the first of the two conditions, we need to associate to L a certain symmetric $(0,2)$-tensor field. More accurately, this tensor field will be a section of the vector bundle $T^0_2(\mathsf{VTM})$. First we define a map $\boldsymbol{F}L\colon \mathsf{TM} \to \mathsf{T}^*\mathsf{M}$ by

$$\langle \boldsymbol{F}L(v_x); w_x\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} L(v_x + tw_x).$$

In natural coordinates for TM and $\mathsf{T}^*\mathsf{M}$, $\boldsymbol{F}L$ has the local representative

$$((x^1,\dots,x^n),(v^1,\dots,v^n)) \mapsto \Big((x^1,\dots,x^n),\Big(\frac{\partial L}{\partial v^1},\dots,\frac{\partial L}{\partial v^n}\Big)\Big)$$

Note that $\boldsymbol{F}L$ is a fiber bundle map, but is not generally a vector bundle map. If $L(v_x) = \frac{1}{2}\mathbb{G}(v_x, v_x)$ for a Riemannian metric $\mathbb{G}$ on M, then $\boldsymbol{F}L = \mathbb{G}^\flat$. The map $\boldsymbol{F}L$ is called the ***fiber derivative*** of L, and plays an important role in the calculus of variations. Next one defines a section $\mathbb{G}_L$ of $T^0_2(\mathsf{VTM})$ by

$$\mathbb{G}_L(v_x)(U_{v_x}, V_{v_x}) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \langle \boldsymbol{F}L(v_x + tU_{v_x}); V_{v_x}\rangle\,, \tag{S4.1}$$

for $U_{v_x}, V_{v_x} \in \mathsf{V}_{v_x}\mathsf{TM}$. Note that, in writing (S4.1), we are making the natural identification of $\mathsf{V}_{v_x}\mathsf{TM}$ with $\mathsf{T}_x\mathsf{M}$. Under this identification, if $(\boldsymbol{x}, \boldsymbol{v})$ are natural coordinates for TM, then $\{\frac{\partial}{\partial x^1},\dots,\frac{\partial}{\partial x^n}\}$ can be thought of as local generators for VTM.[1] With this notation and with these identifications, one verifies that

$$\mathbb{G}_L = \frac{\partial^2 L}{\partial v^i \partial v^j}\mathrm{d}x^i \otimes \mathrm{d}x^j.$$

From this we see that $\mathbb{G}_L$ is a symmetric tensor field. Also, if $L(v_x) = \frac{1}{2}\mathbb{G}(v_x, v_x)$, then $\mathbb{G}_L(U_{v_x}, V_{v_x}) = \mathbb{G}(U_{v_x}, V_{v_x})$, where again we identify $\mathsf{V}_{v_x}\mathsf{TM}$ with $\mathsf{T}_x\mathsf{M}$.

We can now state a second necessary condition for a solution to Problem 4.37 in the calculus of variations.

Theorem S4.1 (Legendre Condition). *If $\gamma \in C^2([a,b], x_a, x_b)$ solves Problem 4.37, then $\mathbb{G}_L(\gamma'(t))$ is positive-semidefinite for each $t \in [a,b]$.*

Note that this necessary condition is automatically satisfied when $L(v_x) = \frac{1}{2}\mathbb{G}(v_x, v_x)$ for a Riemannian metric $\mathbb{G}$ on M. However, in the framework of the Maximum Principle, it is no longer natural to restrict attention to Lagrangians of this special form.

The third necessary condition we provide is due to Weierstrass, and relies on a new construction for its statement. We let $\mathsf{TM}\oplus\mathsf{TM}$ be the Whitney sum of TM with itself. The ***Weierstrass excess function*** is then the function $W_L \colon \mathsf{TM}\oplus\mathsf{TM} \to \mathbb{R}$ defined by

$$W_L(v_x \oplus u_x) = L(u_x) - L(v_x) - \boldsymbol{F}L(v_x)\cdot(u_x - v_x).$$

We now have the following result.

Theorem S4.2 (Weierstrass Side Condition). *If $\gamma \in C^2([a,b], x_a, x_b)$ solves Problem 4.37, then, for each $t \in [a,b]$, $W_L(\gamma'(t) \oplus u_{\gamma(t)}) \geq 0$ for all $u_{\gamma(t)} \in \mathsf{T}_{\gamma(t)}\mathsf{M}$.*

[1] Another way of understanding this, as elucidated in Section S1.3.4, is to note that VTM is isomorphic to the pull-back bundle $\pi^*_{\mathsf{TTM}}\mathsf{TM}$.

S4.1.2 A Hamiltonian setting for the calculus of variations

Now we provide a reorganization of Theorems 4.38, S4.1 and S4.2 in a Hamiltonian setting. This seems slightly contrived at this point. However, as we shall see, it provides just the right setting for the generalization of the calculus of variations to optimal control. We let $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ be the Whitney sum of TM and $\mathsf{T}^*\mathsf{M}$. We denote a typical point in $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ by $v_x \oplus \alpha_x$. Given a Lagrangian L as above, define a C^∞-function H_L on $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ by

$$H_L(v_x \oplus \alpha_x) = \langle \alpha_x; v_x \rangle - L(v_x).$$

Let us make some constructions associated to the function H_L. The notation suggests that H_L might serve as a Hamiltonian in some manner. This is indeed the case, although the standard Hamiltonian setup needs to be adapted to the Whitney sum $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$. We have a natural projection $\pi_2 \colon \mathsf{TM} \oplus \mathsf{T}^*\mathsf{M} \to \mathsf{T}^*\mathsf{M}$ defined by $\pi_2(v_x \oplus \alpha_x) = \alpha_x$. We then define a differential two-form on $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ by $\overline{\omega}_0 = \pi_2^*\omega_0$, where ω_0 is the canonical symplectic form on $\mathsf{T}^*\mathsf{M}$. In natural coordinates $(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{p})$ for $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$ we have

$$\overline{\omega}_0 = \mathrm{d}x^i \wedge \mathrm{d}p_i.$$

We now state a condition that is equivalent to Theorem 4.38.

Proposition S4.3 (Hamiltonian version of the Euler–Lagrange equations). *The following statements are equivalent for a curve $t \mapsto \Upsilon(t) \oplus \lambda(t)$ in $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$:*

(i) the Euler–Lagrange equations for $\pi_{\mathsf{TM}} \circ \Upsilon$ are satisfied, along with the equation $\lambda(t) = \boldsymbol{F}L(\Upsilon(t))$;

(ii) if $t \mapsto \Upsilon'(t) \oplus \lambda'(t)$ denotes the tangent vector field to the curve $t \mapsto \Upsilon(t) \oplus \lambda(t)$, then $\overline{\omega}_0^\flat(\Upsilon'(t) \oplus \lambda'(t)) = -\mathrm{d}H_L(\Upsilon(t) \oplus \lambda(t))$ for each t.

Proof. The most straightforward proof is done in natural coordinates $(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{p})$ for $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$. We denote by $t \mapsto (\boldsymbol{x}(t), \boldsymbol{v}(t), \boldsymbol{p}(t))$ the local representative of $t \mapsto \Upsilon(t) \oplus \lambda(t)$. One readily determines that the local representative of $t \mapsto \overline{\omega}_0^\flat(\Upsilon'(t) \oplus \lambda'(t))$ is

$$\dot{p}_i(t)\mathrm{d}x^i - \dot{x}^i(t)\mathrm{d}p_i.$$

From this, one readily ascertains that part (ii) is equivalent to the three equations

$$\begin{aligned} \dot{x}^i &= \frac{\partial H_L}{\partial p_i}(\boldsymbol{x}(t), \boldsymbol{v}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\}, \\ \dot{p}_i &= -\frac{\partial H_L}{\partial x^i}(\boldsymbol{x}(t), \boldsymbol{v}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\}, \\ 0 &= \frac{\partial H_L}{\partial v^i}(\boldsymbol{x}(t), \boldsymbol{v}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\}. \end{aligned} \tag{S4.2}$$

(i) $\Longrightarrow$ (ii) The equation additional to the Euler–Lagrange equation immediately implies the third of equations (S4.2). The Euler–Lagrange equations imply that $\boldsymbol{v}(t) = \dot{\boldsymbol{x}}(t)$. Since $\frac{\partial H_L}{\partial p_i} = v^i$, $i \in \{1, \dots, n\}$, this implies that

$$\dot{x}^i(t) = \frac{\partial H_L}{\partial p_i}(\boldsymbol{x}(t), \boldsymbol{v}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\},$$

which is the first of equations (ii). Taking the time derivative of the equation $\lambda(t) = \boldsymbol{F}L(\Upsilon(t))$, and using the relation

$$\frac{\partial H_L}{\partial x^i} = -\frac{\partial L}{\partial x^i}, \qquad i \in \{1, \dots, n\},$$

gives

$$\dot{p}_i(t) = -\frac{\partial H_L}{\partial x^i}(\boldsymbol{x}(t), \boldsymbol{v}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\},$$

which is the second of equations (S4.2).

(ii) $\Longrightarrow$ (i) The third of equations (S4.2) implies that

$$p_i(t) = \frac{\partial L}{\partial v^i}(\xi(t), \chi(t)), \qquad i \in \{1, \dots, n\},$$

which is the local form for $\lambda(t) = \boldsymbol{F}L(\Upsilon(t))$. The first of equations (S4.2) implies that $\dot{\boldsymbol{x}}(t) = \boldsymbol{v}(t)$, which means that Υ is the tangent vector field of some curve γ on M. Also, since

$$\frac{\partial H_L}{\partial x^i} = -\frac{\partial L}{\partial x^i}, \qquad i \in \{1, \dots, n\},$$

this shows that part (ii) implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\partial L}{\partial v^i}\Big)(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)) = \frac{\partial L}{\partial x^i}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)), \qquad i \in \{1, \dots, n\},$$

which are the Euler–Lagrange equations. ■

Part (ii) of Proposition S4.3 involves the classical Hamilton's equations, although they now involve a "parameter" $\boldsymbol{v}$. In natural coordinates these equations are as given in the proof as (S4.2). The importance of the third of equations (S4.2) becomes fully realized when one throws Theorem S4.1 into the mix. We now state how the three necessary conditions, Theorems 4.38, S4.1, and S4.2, can be expressed in Hamiltonian language.

Proposition S4.4 (Hamiltonian version of necessary conditions in calculus of variations). *The following statements are equivalent for $\gamma \in C^2([a, b], x_a, x_b)$:*

(i) γ satisfies the necessary conditions of Theorems 4.38, S4.1, and S4.2;

(ii) there exists a covector field λ along γ such that, for the curve $t \mapsto (\gamma'(t) \oplus \lambda(t))$, the following relations hold:

(a) if $t \mapsto \Upsilon'(t) \oplus \lambda'(t)$ denotes the tangent vector field to the curve $t \mapsto \Upsilon(t) \oplus \lambda(t)$, then $\overline{\omega}_0^\flat(\Upsilon'(t) \oplus \lambda'(t)) = -\mathrm{d}H_L(\Upsilon(t) \oplus \lambda(t))$ for each t;

(b) for each $t \in [a,b]$,

$$H_L(\gamma'(t) \oplus \lambda(t)) = \sup\left\{ H_L(v_{\gamma(t)} \oplus \lambda(t)) \;\middle|\; v_{\gamma(t)} \in \mathsf{T}_{\gamma(t)}\mathsf{M} \right\}.$$

Proof. As in the proof of Proposition S4.3, we use natural coordinates $(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{p})$ for $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$. In this case, the coordinate expression of part (ii) consists of the four equations

$$\begin{aligned}
\dot{x}^i(t) &= \frac{\partial H_L}{\partial p_i}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\}, \\
\dot{p}_i(t) &= -\frac{\partial H_L}{\partial x^i}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\}, \\
0 &= \frac{\partial H_L}{\partial v^i}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \boldsymbol{p}(t)), \qquad i \in \{1, \dots, n\}, \\
H_L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \boldsymbol{p}(t)) &= \sup\{ H_L(\boldsymbol{x}(t), \boldsymbol{u}, \boldsymbol{p}(t)) \mid \boldsymbol{u} \in \mathbb{R}^n \}.
\end{aligned} \tag{S4.3}$$

Clearly, these four equations are equivalent to the three equations remaining when the third is omitted, and therefore, part (ii) is equivalent in coordinates to the first, second, and fourth of equations (S4.3).

(i) $\Longrightarrow$ (ii) Define a covector field λ along γ by $\lambda(t) = \boldsymbol{F}L(\gamma'(t))$. A direct calculation shows that the excess function satisfies

$$W_L(\gamma'(t) \oplus v_{\gamma(t)}) = H_L(\gamma'(t) \oplus \lambda(t)) - H_L(v_{\gamma(t)} \oplus \lambda(t)),$$

for $t \in [a,b]$. Therefore the necessary condition of Theorem S4.2 translates to asserting that, for each $t \in [a,b]$,

$$H_L(\gamma'(t) \oplus \lambda(t)) - H_L(v_{\gamma(t)} \oplus \lambda(t)) \geq 0, \qquad v_{\gamma(t)} \in \mathsf{T}_{\gamma(t)}\mathsf{M},\ t \in [a,b]. \tag{S4.4}$$

This is exactly the fourth of equations (S4.3). By Proposition S4.3, the necessary condition of Theorem 4.38 implies the first two of equations (S4.3). This shows that (ii) holds since the first, second, and fourth of equations (S4.3) hold.

(ii) $\Longrightarrow$ (i) We suppose that the first, second, and fourth of equations (S4.3) hold for some covector field along γ whose local representative is $t \mapsto \boldsymbol{p}(t)$. The fourth of the equations (S4.3) implies that

$$\frac{\partial H_L}{\partial v^i}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \boldsymbol{p}(t)) = 0, \qquad i \in \{1, \dots, n\}.$$

The definition of H_L then gives $p_i(t) = \frac{\partial L}{\partial v^i}(\xi(t), \dot{\xi}(t))$, $i \in \{1, \dots, n\}$. This then shows that $\lambda(t) = \boldsymbol{F}L(\gamma'(t))$. By Proposition S4.3, this also implies that the necessary condition of Theorem 4.38 holds. The fourth of equations (S4.3) also implies that (S4.4) holds. Thus the necessary condition of Theorem S4.2 holds. Since (S4.4) holds, it follows that the matrix with components

$$\frac{\partial^2 H_L}{\partial v^i \partial v^j}(\xi(t), \dot{\xi}(t), \lambda(t)), \qquad i, j \in \{1, \dots, n\}$$

is negative-semidefinite for each $t \in [a, b]$. Since

$$\frac{\partial^2 H_L}{\partial v^i \partial v^j} = -\frac{\partial^2 L}{\partial v^i \partial v^j}, \qquad i, j \in \{1, \dots, n\},$$

it follows that the necessary condition of Theorem S4.1 holds. ■

Remark S4.5. Readers familiar with mechanics will recall that, in Lagrangian mechanics, there is a standard technique for going from the Euler–Lagrange equations on TQ to Hamilton's equations on $\mathsf{T}^*\mathsf{Q}$. In order for the equations to have a nice correspondence, there are conditions that must be placed on the Lagrangian, namely that $\boldsymbol{F}L\colon \mathsf{TQ} \to \mathsf{T}^*\mathsf{Q}$ be a diffeomorphism. In the setup in this section, no restrictions are placed on the Lagrangian. That this is possible is a consequence of the fact that, in the Hamiltonian setting in this section, we use, not $\mathsf{T}^*\mathsf{Q}$, but $\mathsf{TQ} \oplus \mathsf{T}^*\mathsf{Q}$. As we shall see, this way of doing things pays off when one looks at optimization problems in control theory. •

S4.1.3 From the calculus of variations to the Maximum Principle

We now make a transition to optimal control from the calculus of variations setting of the preceding section. We begin by introducing the "Maximum" in the Maximum Principle. As in the previous section, we let L be a C^∞-Lagrangian on M, and we let H_L be the corresponding Hamiltonian defined on $\mathsf{TM} \oplus \mathsf{T}^*\mathsf{M}$. We then define a function $H_L^{\max}$ on $\mathsf{T}^*\mathsf{M}$ by

$$H_L^{\max}(\alpha_x) = \sup \{ H_L(v_x \oplus \alpha_x) \mid \ v_x \in \mathsf{T}_x\mathsf{M} \}.$$

It is possible that $H_L^{\max}$ might take the value $+\infty$ at some or all points in $\mathsf{T}^*\mathsf{M}$. However, let us suppose that L has the property that $H_L^{\max}$ is well-defined as a $\mathbb{R}$-valued function. Let us also suppose that $H_L^{\max}$ is of class C^∞. In this case, we have the following theorem giving a necessary condition for solutions of Problem 4.37.

Theorem S4.6 (Maximum Principle version of necessary conditions in calculus of variations). *Assume that $H_L^{\max}$ is a C^∞-function that is well-defined on $\mathsf{T}^*\mathsf{M}$. If $\gamma \in C^2([a, b], x_a, x_b)$ solves Problem 4.37, then there exists a covector field λ along γ such that*

(i) $H_L(\gamma'(t) \oplus \lambda(t)) = H_L^{\max}(\lambda(t))$ and

(ii) λ is an integral curve of the C^∞-Hamiltonian vector field $X_{H_L^{\max}}$.

Proof. This will follow from the Maximum Principle that we state precisely in Section S4.2.2. We leave the translation to this special case to the reader. ■

Now we adapt the preceding theorem to a control-theoretic setting. It is convenient in this development to allow a slightly more general class of control system than the control-affine systems considered in the text.

Definition S4.7 (Control system). For $r \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$, a ***$C^r$-control system*** is a triple (M, f, U) where

(i) M is a C^r-manifold,

(ii) $f \colon \mathsf{M} \times \mathbb{R}^m \to \mathsf{TM}$ is a map with the following properties:

  (a) $f(x,u) \in \mathsf{T}_x\mathsf{M}$ for each $(x,u) \in \mathsf{M} \times \mathbb{R}^m$;

  (b) f is of class C^1;

  (c) the map $x \mapsto f(x,u)$ is of class C^r for each $u \in \mathbb{R}^m$,

 and

(iii) $U \subset \mathbb{R}^m$. •

Clearly, control-affine systems are examples of control systems as per the preceding definition. In our statement of the Maximum Principle, the extra structure afforded by control-affine systems will not be useful, and will simply complicate the notation. There are generalizations of the notion of a control system (e.g., [Nijmeijer and van der Schaft 1990, page 16]), but we shall not find these sorts of generalities illuminating, especially since Definition S4.7 includes control-affine systems.

A control system (M, f, U) gives rise to the control equations

$$\gamma'(t) = f(\gamma(t), u(t)). \tag{S4.5}$$

A ***controlled trajectory*** for a control system $\Sigma = (\mathsf{M}, f, U)$ is a pair (γ, u) where $\gamma \colon I \to \mathsf{M}$ and $u \colon I \to U$ have the property that u is locally integrable and γ satisfies (S4.5). Thus γ is locally absolutely continuous. We denote by $\mathrm{Ctraj}(\Sigma)$ the set of controlled trajectories for Σ. A ***controlled arc*** for Σ is a controlled trajectory defined on a compact interval, i.e., an interval of the form $I = [a,b]$. We denote by $\mathrm{Carc}(\Sigma)$ the set of controlled arcs for Σ.

We now suppose that we have a C^∞-control system $\Sigma = (\mathsf{M}, f, U)$. Note then that the admissible velocities at $x \in \mathsf{M}$ are now parameterized by the control set U. Therefore, a Lagrangian will now be a function, not on TM, but on $\mathsf{M} \times U$. Thus, for such a Lagrangian and for $(\gamma, u) \in \mathrm{Carc}(\Sigma)$ with u and γ defined on $[a,b]$, we define

$$A_{\Sigma,L}(\gamma, u) = \int_a^b L(u(t), \gamma(t))\,\mathrm{d}t.$$

Let us state a problem related with this setting.

Problem S4.8. Given $x_a, x_b \in \mathsf{M}$, find $(\gamma, u) \in \mathrm{Carc}(\Sigma)$ that minimizes $A_{\Sigma,L}$ subject to the constraint that $\gamma(a) = x_a$ and $\gamma(b) = x_b$. •

We now argue from Theorem S4.6 to a proposed (but imprecisely stated) solution to Problem S4.8. The objective is to establish a link from the calculus of variations to optimal control. The ***control Hamiltonian*** for Σ and L is the function on $\mathsf{T}^*\mathsf{M} \times U$ given by

$$H_{\Sigma,L}(\alpha_x, u) = \langle \alpha_x; f(x,u) \rangle - L(x,u).$$

Where in the calculus of variations we chose the velocity v_x to maximize the Hamiltonian with $\alpha_x \in \mathsf{T}^*\mathsf{M}$ fixed, we now fix $\alpha_x \in \mathsf{T}^*\mathsf{M}$ and maximize the control Hamiltonian:

$$H^{\max}_{\Sigma,L}(\alpha_x) = \sup\left\{ H_{\Sigma,L}(\alpha_x, u) \mid \ u \in U \right\}.$$

With Theorem S4.6 in mind, we state the following conjecture.[2]

Conjecture S4.9 (Naïve, incorrect version of the Maximum Principle). *If* $(\gamma, u) \in \mathrm{Carc}(\Sigma)$ *solves Problem S4.8, then there exists an* LAC *covector field* λ *along* γ *such that*

(i) $H_{\Sigma,L}(\lambda(t), u(t)) = H^{\max}_L(\lambda(t))$ *and*

(ii) λ *is an integral curve of the* C^∞*-Hamiltonian vector field* $X_{H^{\max}_{\Sigma,L}}$.

We devote the next section to stating the correct version of the preceding conjecture. As we shall see, many of the essential features of the conjecture are correct, but certain hypotheses are missing. Also, the complete statement of the Maximum Principle contains slightly more information that our conjecture.

S4.2 The Maximum Principle

In this section, we continue the development from the preceding section, but now we are significantly more precise in our statement of the optimal control problem, and in our necessary condition, the Maximum Principle, for a solution to this problem. In keeping with the flavor of the treatment in the text, we state a geometric version of the Maximum Principle. This differs somewhat from some standard treatments for which states typically live in an open subset of Euclidean space. It also bears mentioning that the Maximum Principle on Euclidean space *does not* imply the Maximum Principle on manifolds as we state it here. While it is natural to extend the *statement* of the Maximum Principle from Euclidean space to manifolds, the *proof* does not extend so easily. We rely on the extension of Sussmann [1997] of the Maximum Principle to manifolds. However, our framework is significantly less general than Sussmann's. We do, however, employ the notion of a control system from Definition S4.7 which generalizes the notion of a control-affine system.

We begin with a precise statement of the sorts of optimal control problems we consider.

[2] Actually, this is not a conjecture, since we know it to be false. We hope the reader can forgive the abuse of style.

S4.2.1 An optimal control problem

We consider a C^∞-control system $\Sigma = (\mathsf{M}, f, U)$ as in Definition S4.7, and additionally consider the added data of a ***cost function***[3] for Σ that is defined to be a continuous function $F\colon \mathsf{M} \times \mathbb{R}^m \to \mathbb{R}$ for which the function $x \mapsto F(x,u)$ is of class C^∞ for $u \in \mathbb{R}^m$. Although we are only interested in the value of F on $\mathsf{M} \times U$, it is convenient to suppose F to be the restriction of a continuous function on $\mathsf{M} \times \mathbb{R}^m$. Note that the function $(t,x) \mapsto F(x,u(t))$ is an LIC$^\infty$-function if $(\gamma, u) \in \mathrm{Carc}(\Sigma)$. We shall say that $(\gamma, u) \in \mathrm{Ctraj}(\Sigma)$ is ***F-acceptable*** if the function $t \mapsto F(\gamma(t), u(t))$ is locally integrable. We denote by $\mathrm{Ctraj}(\Sigma, F)$ the subset of $\mathrm{Ctraj}(\Sigma)$ consisting of F-acceptable controlled trajectories. We similarly denote by $\mathrm{Carc}(\Sigma, F)$ the subset of $\mathrm{Carc}(\Sigma)$ consisting of F-acceptable controlled arcs.

For $(\gamma, u) \in \mathrm{Carc}(\Sigma, F)$ with u and γ defined on $[a,b]$, define

$$A_{\Sigma,F}(\gamma, u) = \int_a^b F(\gamma(t), u(t))\,\mathrm{d}t.$$

The map $\mathrm{Carc}(\Sigma, F) \ni (\gamma, u) \mapsto A_{\Sigma,F}(\gamma, u)$ is the ***objective function***. Let S_0 and S_1 be disjoint submanifolds of M. We denote by

$$\begin{aligned}\mathrm{Carc}(\Sigma, F, S_0, S_1) = \{(\gamma, u) \in \mathrm{Carc}(\Sigma, F)|\ &\gamma(a) \in S_0 \text{ and } \gamma(b) \in S_1,\\ &u \text{ and } \gamma \text{ are defined on } [a,b] \text{ for some } a, b \in \mathbb{R}\}.\end{aligned}$$

In like fashion, for $a, b \in \mathbb{R}$ with $a < b$, we define

$$\begin{aligned}\mathrm{Carc}(\Sigma, F, S_0, S_1, [a,b]) = \{(\gamma, u) \in \mathrm{Carc}(\Sigma, F)|\ &u \text{ and } \gamma\\ &\text{are defined on } [a,b] \text{ and } \gamma(a) \in S_0 \text{ and } \gamma(b) \in S_1\}.\end{aligned}$$

The problems concerning the optimal path connecting two submanifolds are stated as follows.

Definition S4.10 (Optimal control problems). Let $\Sigma = (\mathsf{M}, f, U)$ be a C^∞-control system, let F be a cost function for Σ, and let S_0 and S_1 be disjoint submanifolds of M.

(i) A controlled arc $(\gamma_*, u_*) \in \mathrm{Carc}(\Sigma, F, S_0, S_1)$ is a ***solution of*** $\boldsymbol{\mathscr{P}(\Sigma, F, S_0, S_1)}$ if $A_{\Sigma,F}(\gamma_*, u_*) \le A_{\Sigma,F}(\gamma, u)$ for every $(\gamma, u) \in \mathrm{Carc}(\Sigma, F, S_0, S_1)$.

(ii) A controlled arc $(\gamma_*, u_*) \in \mathrm{Carc}(\Sigma, F, S_0, S_1, [a,b])$ is a ***solution of*** $\boldsymbol{\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)}$ if $A_{\Sigma,F}(\gamma_*, u_*) \le A_{\Sigma,F}(\gamma, u)$ for every $(\gamma, u) \in \mathrm{Carc}(\Sigma, F, S_0, S_1, [a,b])$. •

[3] Many authors calls such functions ***Lagrangians***. However, since we will be dealing with control systems whose dynamics themselves are sometimes Lagrangian, we refrain from using this notation as it might lead to one Lagrangian too many.

A special case of this problem occurs when $S_0 = \{x_0\}$ and $S_1 = \{x_1\}$ for two points $x_0, x_1 \in \mathsf{M}$. The problem $\mathscr{P}(\Sigma, F, S_0, S_1)$ is called a ***free interval problem***, since the interval of definition of solutions is left unspecified. Similarly, the problem $\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)$ is called a ***fixed interval problem***, since the interval of definition of solutions is part of the problem statement.

S4.2.2 The Maximum Principle

Now that we have stated clearly the optimal control problems we wish to investigate, let us state necessary conditions for solutions of this problem. Key to the necessary conditions of the Maximum Principle is the use of a Hamiltonian formalism. Let $\Sigma = (\mathsf{M}, f, U)$ be a C^∞-control system and let F be a cost function for Σ. We define the ***Hamiltonian*** $H_{\Sigma,F}$ as a function on $\mathsf{T}^*\mathsf{M} \times \mathbb{R}^m$ by

$$H_{\Sigma,F}(\alpha_x, u) = \alpha_x \cdot f(x, u) - F(x, u).$$

From this we define the ***maximum Hamiltonian*** as the function on $\mathsf{T}^*\mathsf{M}$ defined by

$$H^{\max}_{\Sigma,F}(\alpha_x) = \sup \{ H_{\Sigma,F}(\alpha_x, u) \mid \; u \in U \},$$

and we adopt the notation $H^{\max}_{\Sigma,F}(\alpha_x) = +\infty$ if, for each $C \in \mathbb{R}$, there exists $u \in U$ such that $H_{\Sigma,F}(\alpha_x, u) > C$. If $u \in U$ has the property that $H_{\Sigma,F}(\alpha_x, u) = H^{\max}_{\Sigma,F}(\alpha_x)$, then we say $H^{\max}_{\Sigma,F}$ is ***realized*** by u at α_x. If $(\gamma, u) \in \mathrm{Carc}(\Sigma, F)$, then the function $H^u_{\Sigma,F}(t, \alpha_x) \mapsto H_{\Sigma,F}(\alpha_x, u(t))$ is LIC^∞. Therefore, corresponding to this function will be an LIC^∞-Hamiltonian vector field $X_{H^u_{\Sigma,F}}$ on $\mathsf{T}^*\mathsf{M}$.

If $(\gamma, u) \in \mathrm{Carc}(\Sigma, F)$ with u and γ defined on an interval $[a, b]$, then an LAC covector field $\chi\colon [a, b] \to \mathsf{T}^*\mathsf{M}$ along γ is ***maximizing for (Σ, F) along u*** if

$$H_{\Sigma,F}(\chi(t), u(t)) \leq H^{\max}_{\Sigma,F}(\chi(t))$$

for almost every $t \in [a, b]$.

We now state the Maximum Principle as we shall employ it.

Theorem S4.11 (Maximum Principle). *Let $\Sigma = (\mathsf{M}, f, U)$ be a C^∞-control system with F a cost function for Σ, and let S_0 and S_1 be disjoint submanifolds of M. Suppose that $(\gamma, u) \in \mathrm{Carc}(\Sigma, F)$ is a solution of $\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)$. Then there exists an LAC covector field $\chi\colon [a, b] \to \mathsf{T}^*\mathsf{M}$ along γ and a constant $\chi_0 \in \{0, 1\}$ with the following properties:*

(i) $\chi(a) \in \mathrm{ann}(\mathsf{T}_{\gamma(a)}S_0)$ and $\chi(b) \in \mathrm{ann}(\mathsf{T}_{\gamma(b)}S_1)$;
(ii) $t \mapsto \chi(t)$ is an integral curve of $X_{H^u_{\Sigma,\chi_0 F}}$;
(iii) χ is maximizing for $(\Sigma, \chi_0 F)$ along u;
(iv) either $\chi_0 = 1$ or $\chi(a) \neq 0$;
(v) there exists a constant $C \in \mathbb{R}$ such that $H_{\Sigma,F}(\chi(t), u(t)) = C$ a.e.

If (γ, u) is a solution of $\mathscr{P}(\Sigma, F, S_0, S_1)$, then conditions (i)–(iv) hold, and condition (v) can be replaced with *(vi)* $H_{\Sigma,F}(\chi(t),u(t)) = 0$ *a.e.*

Remarks S4.12. 1. If $S_0 = \{x_0\}$, then $\chi(a)$ is unrestricted (modulo requirement (iv)). Similarly, if $S_1 = \{x_1\}$, then $\chi(b)$ is unrestricted.

2. Since the Hamiltonian vector field $X_{H^u_{\Sigma,F}}$ is linear (Remark S1.11–2), the condition (iv) in the statement of the Maximum Principle asserts that $(\chi_0, \chi(t))$ will be non-zero for $t \in [a,b]$. •

Let us write the equations for integral curves of $X^u_{H_{\Sigma,\chi_0 F}}$ in local coordinates $(\boldsymbol{x},\boldsymbol{p})$ for $\mathsf{T}^*\mathsf{M}$. First of all, the Hamiltonian in local coordinates has the local representative

$$(\boldsymbol{x},\boldsymbol{p},u) \mapsto \langle\!\langle \boldsymbol{p}, \tilde{f}(\boldsymbol{x},u(t))\rangle\!\rangle_{\mathbb{R}^n} - \chi_0 \tilde{F}(\boldsymbol{x},u(t)),$$

where $\tilde{f}$ and $\tilde{F}$ are the local representatives of f and F, respectively. From this, using the form (S1.1) for a Hamiltonian vector field on $\mathsf{T}^*\mathsf{M}$, we see that the governing differential equations for $X_{H^u_{\Sigma,\boldsymbol{\chi_0} F}}$ are

$$\begin{aligned} \dot{x}^i(t) &= \tilde{f}^i(\boldsymbol{x}(t),u(t)), \qquad i \in \{1,\dots,n\},\\ \dot{p}_i(t) &= -\frac{\partial \tilde{f}^j}{\partial x^i}(\boldsymbol{x}(t),u(t))p_j(t) + \chi_0 \frac{\partial \tilde{F}}{\partial x^i}(\boldsymbol{x}(t),u(t)), \qquad i \in \{1,\dots,n\}. \end{aligned}$$

The first n of these equations are simply the control equations. The second n of these equations are frequently called the ***adjoint equation***. A difficulty with this terminology is that there is no natural coordinate-invariant object associated to these equations. What is coordinate-invariant are both components of Hamilton's equations. To get around this matter, and come up with a version of the adjoint equation, one must introduce extra structure into the problem. For example, for systems on open subsets of Euclidean space, and for linear systems, one has a natural trivialization of the tangent and cotangent bundles of M, and so Hamilton's equations naturally decouple into an "$\boldsymbol{x}$-part" and a "$\boldsymbol{p}$-part." Sussmann [1997] uses the structure of a reference trajectory to talk about a differential operator on sections of $\mathsf{T}^*\mathsf{M}$ along the reference trajectory. For those familiar with the concept, this is entirely along the lines of the so-called "Lie drag" (see [Crampin and Pirani 1986, Section 3.5]). We shall see in Section S4.4 that the structure of an affine connection gives us a natural way of extracting a coordinate-invariant version of the adjoint equation from the vector field $X_{H^u_{\Sigma,\boldsymbol{\chi_0} F}}$.

S4.2.3 Extremals

In this section we introduce some useful terminology associated to controlled arcs satisfying the necessary conditions of the Maximum Principle.

Definition S4.13 (Extremal). Let (M, f, U) be a C^∞-control system with a cost function F. Let $\mathscr{P}$ be either $\mathscr{P}(\Sigma, F, S_0, S_1)$ or $\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)$.

(i) A ***controlled extremal*** for $\mathscr{P}$ is $(\gamma, u) \in \mathrm{Carc}(\Sigma, F)$ that satisfies corresponding necessary conditions of Theorem S4.11.
(ii) An absolutely continuous curve $\gamma\colon [a,b] \to \mathsf{M}$ is an ***extremal*** for $\mathscr{P}$ if there exists a control u such that (γ, u) is a controlled extremal for $\mathscr{P}$.
(iii) A integrable control $u\colon [a,b] \to U$ is an ***extremal control*** for $\mathscr{P}$ if there exists a curve γ on M such that (γ, u) is a controlled extremal for $\mathscr{P}$. •

Of course, controlled extremals need not be *solutions* of $\mathscr{P}(\Sigma, F, S_0, S_1)$ or $\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)$, but the converse is necessarily true.

It is sometimes also useful to give names to the other objects arising from the conditions in the Maximum Principle.

Definition S4.14 (Adjoint covector field and constant Lagrange multiplier). Let (M, f, U) be a C^∞-control system with a cost function F. Let (γ, u) be a controlled extremal for one of the problems $\mathscr{P}(\Sigma, F, S_0, S_1)$ or $\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)$, with χ a covector field along γ, and $\chi_0 \in \{0,1\}$ as in Theorem S4.11. Then χ is an ***adjoint covector field*** and χ_0 is a ***constant Lagrange multiplier***. •

Note that, for a given controlled extremal (γ, u), there may be multiple adjoint covector fields and constant Lagrange multipliers.

There is a fundamental dichotomy in classes of extremals, depending essentially on whether the constant Lagrange multiplier is zero or one.

Definition S4.15 (Normal and abnormal controlled extremals). Let (M, f, U) be a C^∞-control system with a cost function F. A controlled extremal (γ, u) for $\mathscr{P}(\Sigma, F, S_0, S_1)$ or $\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)$, satisfying the necessary conditions of Theorem S4.11 with $\chi_0 = 1$, is called ***normal***. A controlled extremal is ***abnormal*** if it satisfies the necessary conditions of Theorem S4.11 only for $\chi_0 = 0$.

An extremal γ is ***normal*** (resp. ***abnormal***) if there exists a control u such that (γ, u) is a normal (resp. abnormal) controlled extremal. •

The wording here must be correctly understood. An abnormal controlled extremal is *not* one that satisfies the necessary conditions of Theorem S4.11 with $\chi_0 = 0$. It is one that satisfies the necessary conditions of Theorem S4.11 with $\chi_0 = 0$, but *cannot* satisfy the necessary conditions of Theorem S4.11 with $\chi_0 = 1$. A discussion of abnormality is given in Section S4.3.2, and an example possessing abnormal controlled extremals is given in Section S4.3.3. That abnormal controlled extremals can be optimal in the case of sub-Riemannian geometry[4] has been shown by Montgomery [1994] and Liu and Sussmann [1994].

[4] In sub-Riemannian geometry, one studies distributions that possess a smooth assignment of an inner product to each fiber. One then considers curves whose tangent vector fields take values in the distribution. For such a curve, one can define its length in a manner entirely analogous to the Riemannian case. The

One of the remarkable features of the Maximum Principle is that the controlled extremal controls can sometimes be explicitly determined from the condition that the Hamiltonian is maximizing along an extremal. This will be illustrated in parts of the subsequent development, and in some examples to follow. However, cases can arise where the condition that the Hamiltonian be maximized gives no information about the control. This gives rise to the following definition.

Definition S4.16 (Regular and singular controlled extremals). Let (M, f, U) be a C^∞-control system with a cost function F. Let (γ, u) be a controlled extremal for $\mathscr{P}(\Sigma, F, S_0, S_1)$ or $\mathscr{P}_{[a,b]}(\Sigma, F, S_0, S_1)$, defined on $[a, b]$, with χ_0 the constant Lagrange multiplier and χ the adjoint covector field. We say that (γ, u) is ***singular*** if, for each $t \in [a, b]$, $H_{\Sigma,\chi_0 F}(\chi(t), \bar{u}) = H^{\max}_{\Sigma,\chi_0 F}(\chi(t))$ for all $\bar{u} \in U$. A controlled extremal that is not singular is ***regular***.

An extremal γ is ***singular*** (resp. ***regular***) if there exists a control u such that (γ, u) is a singular (resp. regular) controlled extremal. •

Thus the definition formalizes the notion that the maximization condition gives no information about the extremal control. Dunn [1967] proposes a more refined notion of singularity where one allows for extremals that are singular in our sense along one part of the extremal, and nonsingular on the rest. For control-affine systems, the classification of singular controlled extremals involves Lie brackets. We refer the reader to the book of Bonnard and Chyba [2003] for a detailed discussion of singular controlled extremals. In the paper of Chyba, Leonard, and Sontag [2003], it is shown that for mechanical systems, there can be singular extremals that are optimal.

S4.3 Coming to an understanding of the Maximum Principle

In Section S4.1 we illustrated the connection between the classical calculus of variations and the Maximum Principle. In this section we give additional interpretations of the Maximum Principle, but now in terms of concepts from control theory.

S4.3.1 The relationship between controllability and optimal control

There is a not-so-transparent link between optimal control and controllability that we explore in this section. First let us expose a property of solutions to optimal control problems.

---

optimization problem is then to find curves connecting two points with the minimum length. We refer to [Montgomery 2002] for a discussion of some aspects of sub-Riemannian geometry.

Proposition S4.17 (Subarcs of minimizers are minimizers). *Let $\Sigma = (\mathsf{M}, f, U)$ be a control system with F a cost function for Σ. Let $a < b \in \mathbb{R}$ and let $x_0, x_1 \in \mathsf{M}$. The following statements hold.*

(i) If $(\gamma, u) \in \mathrm{Carc}(\Sigma, F)$ solves $\mathscr{P}(\Sigma, F, \{x_0\}, \{x_1\})$ and is defined on $[a, b]$, then, for each $\tilde{a}, \tilde{b} \in [a, b]$, $\tilde{a} < \tilde{b}$, $(\gamma|[\tilde{a}, \tilde{b}], u|[\tilde{a}, \tilde{b}])$ solves $\mathscr{P}(\Sigma, F, \{\gamma(\tilde{a})\}, \{\gamma(\tilde{b})\})$.

(ii) If $(\gamma, u) \in \mathrm{Carc}(\Sigma, F)$ solves $\mathscr{P}_{[a,b]}(\Sigma, F, \{x_0\}, \{x_1\})$, then, for each $\tilde{a}, \tilde{b} \in [a, b]$ with $\tilde{a} < \tilde{b}$, $(\gamma|[\tilde{a}, \tilde{b}], u|[\tilde{a}, \tilde{b}])$ solves $\mathscr{P}_{[\tilde{a},\tilde{b}]}(\Sigma, F, \{\gamma(\tilde{a})\}, \{\gamma(\tilde{b})\})$.

Proof. (i) Suppose that $(\gamma|[\tilde{a}, \tilde{b}], u|[\tilde{a}, \tilde{b}])$ does not solve $\mathscr{P}(\Sigma, F, \{\gamma(\tilde{a})\}, \{\gamma(\tilde{b})\})$. Then there exists $(\bar{\gamma}, \bar{u}) \in \mathrm{Carc}(\Sigma, F, \{\gamma(\tilde{a})\}, \{\gamma(\tilde{b})\})$ such that $A_{\Sigma,F}(\gamma|[\tilde{a}, \tilde{b}], u|[\tilde{a}, \tilde{b}]) > A_{\Sigma,F}(\bar{\gamma}, \bar{u})$. Suppose that $(\bar{\gamma}, \bar{u})$ is defined on $[\bar{a}, \bar{b}]$. Then define a controlled trajectory $(\gamma_1 * \gamma_2 * \gamma_3, u_1 * u_2 * u_3)$ where

$$\begin{aligned} u_1(t) &= u(t), && t \in [a, \tilde{a}[\,, \\ u_2(t) &= \bar{u}(t + \bar{a} - \tilde{a}), && t \in [\tilde{a}, \tilde{a} + (\bar{b} - \bar{a})], \\ u_3(t) &= u(t + \tilde{b} - \tilde{a} - (\bar{b} - \bar{a})), && t \in]\tilde{a} + (\bar{b} - \bar{a}), b + (\bar{b} - \bar{a}) - (\tilde{b} - \tilde{a})], \end{aligned}$$

and $\gamma_1(0) = x_0$. Note that $(\gamma_1 * \gamma_2 * \gamma_3, u_1 * u_2 * u_3) \in \mathrm{Carc}(\Sigma, F, \{x_0\}, \{x_1\})$ and that

$$A_{\Sigma,F}(\gamma_1 * \gamma_2 * \gamma_3, u_1 * u_2 * u_3) < A_{\Sigma,F}(\gamma, u),$$

contradicting (γ, u) being a solution to $\mathscr{P}(\Sigma, F, \{x_0\}, \{x_1\})$. This proves (i).

(ii) Suppose that $(\gamma|[\tilde{a}, \tilde{b}], u|[\tilde{a}, \tilde{b}])$ does not solve $\mathscr{P}_{[\tilde{a},\tilde{b}]}(\Sigma, F, \{\gamma(\tilde{a})\}, \{\gamma(\tilde{b})\})$. Then there exists $(\tilde{\gamma}, \tilde{u}) \in \mathrm{Carc}(\Sigma, F, \{\gamma(\tilde{a})\}, \{\gamma(\tilde{b})\}, [\tilde{a}, \tilde{b}])$ such that $A_{\Sigma,F}(\gamma|[\tilde{a}, \tilde{b}], u|[\tilde{a}, \tilde{b}]) > A_{\Sigma,F}(\tilde{\gamma}, \tilde{u})$. Then define a controlled trajectory $(\gamma_1 * \gamma_2 * \gamma_3, u_1 * u_2 * u_3)$ where $u_1 = u|[a, \tilde{a}[\,$, $u_2 = \tilde{u}$, and $u_3 = u|\,]\tilde{b}, b]$ and $\gamma_1(0) = x_0$. Note that $(\gamma_1 * \gamma_2 * \gamma_3, u_1 * u_2 * u_3) \in \mathrm{Carc}(\Sigma, F, \{x_0\}, \{x_1\})$ and that

$$A_{\Sigma,F}(\gamma_1 * \gamma_2 * \gamma_3, u_1 * u_2 * u_3) < A_{\Sigma,F}(\gamma, u),$$

contradicting (γ, u) being a solution to $\mathscr{P}(\Sigma, F, \{x_0\}, \{x_1\})$. This proves (ii). ■

Suppose that we are in possession of a control system $\Sigma = (\mathsf{M}, f, U)$ and a cost function F for Σ. Associated with this is a new control system $\Sigma_F = (\mathsf{M}_F, f_F, U)$ where

1. $\mathsf{M}_F = \mathsf{M} \times \mathbb{R}$,
2. $f_F(x, \kappa) = (f(x, u), (\kappa, F(x, u))) \in \mathsf{TM} \times \mathbb{R} \times \mathbb{R} \simeq \mathsf{TM}_F$, and
3. $U = U$ (abuse of notation).

The equations governing this extended system are

$$\gamma'(t) = f(\gamma(t), u(t)), \quad \kappa'(t) = F(\gamma(t), u(t)).$$

Since $\kappa(t) = \int_a^t F(\gamma(\tau), u(\tau))\,\mathrm{d}\tau$, the additional state thus keeps a running tab on the objective function. We consider the fixed time problem $\mathscr{P} = \mathscr{P}_{[0,T]}(\Sigma, F, \{x_0\}, \{x_1\})$, and we suppose that (γ, u) solves $\mathscr{P}$. The optimal value of the objective function is then

$$A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P}) = \int_0^T F(\gamma(t), u(t))\,\mathrm{d}t.$$

In M_F consider the ray

$$\ell_{x_1,T} = \big\{(x_1, \kappa) \bigm| \kappa < A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})\big\}.$$

Recall that $\mathcal{R}_{\Sigma_F}((x_0, 0), T)$ denotes the set of points reachable from $(x_0, 0)$ by the extended system Σ_F in time T. Clearly, we must have $(\gamma(T) = x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})) \in \mathcal{R}_{\Sigma_F}((x_0, 0), T)$. One can say more, in fact, about the relationship between the terminal point and the reachable set.

Proposition S4.18 (Minimizing arcs lie on the boundary of the reachable set I). *With the above notation,* $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})) \in \mathrm{bd}(\mathcal{R}_{\Sigma_F}((x_0, 0), T))$. *Furthermore,* $\mathcal{R}_{\Sigma_F}((x_0, 0), T) \cap \ell_{x_1,T} = \emptyset$.

Proof. Let us prove the last statement first. Suppose there is a point $(x_1, \kappa) \in \mathcal{R}_{\Sigma_F}((x_0, 0), T) \cap \ell_{x_1,T}$. Since $(x_1, \kappa) \in \mathcal{R}_{\Sigma_F}((x_0, 0), T)$, there exists a controlled trajectory $(\tilde{\gamma} \times \sigma, \tilde{u})$ for Σ_F such that $(\tilde{\gamma}(T), \sigma(T)) = (x_1, \kappa)$. However, this contradicts (γ, u) being a solution to $\mathscr{P}$.

For the first assertion, note that, while $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})) \in \mathcal{R}_{\Sigma_F}((x_0, 0), T)$, every neighborhood of $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P}))$ contains a point in $\ell_{x_1,T}$. Thus by the first part of the proposition, every neighborhood of $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P}))$ contains a point not in $\mathcal{R}_{\Sigma_F}((x_0, 0), T)$. This means that $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})) \in \mathrm{bd}(\mathcal{R}_{\Sigma_F}((x_0, 0), T))$, as desired. ■

Now let us consider the optimal control problem $\mathscr{P} = \mathscr{P}(\Sigma, F, \{x_0\}, \{x_1\})$ with free time interval. We take a solution (γ, u) of $\mathscr{P}$ defined on $[0, T]$, so that the optimal value of the objective function is

$$A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P}) = \int_0^T F(\gamma(t), u(t))\,\mathrm{d}t.$$

The following result is analogous to Proposition S4.18.

Proposition S4.19 (Minimizing arcs lie on the boundary of the reachable set II). *With the above notation,* $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})) \in \mathrm{bd}(\mathcal{R}_{\Sigma_F}((x_0, 0), \leq T))$. *Furthermore,* $\mathcal{R}_{\Sigma_F}((x_0, 0), \leq T) \cap \ell_{x_1,T} = \emptyset$.

Proof. Let us prove the last statement first. Suppose there is a point $(x_1, \kappa) \in \mathcal{R}_{\Sigma_F}((x_0, 0), \leq T) \cap \ell_{x_1,T}$. Since $(x_1, \kappa) \in \mathcal{R}_{\Sigma_F}((x_0, 0), \leq T)$, there exists a controlled trajectory $(\tilde{\gamma} \times \sigma, \tilde{u})$ for Σ_F such that $(\tilde{\gamma}(\tilde{T}), \sigma(\tilde{T})) = (x_1, \kappa)$ for some $\tilde{T} \in \,]0, T]$. However, this contradicts (γ, u) being a solution to $\mathscr{P}$.

For the first assertion, note that, while $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})) \in \mathcal{R}_{\Sigma_F}((x_0, 0), \leq T)$, every neighborhood of $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P}))$ contains a point in $\ell_{x_1,T}$. Thus by the first part of the proposition, every neighborhood of $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P}))$ contains a point not in $\mathcal{R}_{\Sigma_F}((x_0, 0), T)$. This means that $(x_1, A^{\mathrm{opt}}_{\Sigma,F}(\mathscr{P})) \in \mathrm{bd}(\mathcal{R}_{\Sigma_F}((x_0, 0), T))$, as desired. ■

Combining Propositions S4.17, S4.18 and S4.19, the picture that emerges for a solution to an optimal control problem is that, at each time t, it should lie on the boundary of the set of points reachable by Σ_F in time t (resp. time at most t) for fixed interval (resp. free interval) problems. A picture one could have in mind regarding the proposition is presented in Figure S4.1. In the

Figure S4.1. The relationship between optimal trajectories and reachable sets for the extended system Σ_F

figure, $\mathcal{R}_{F,T}$ represents $\mathcal{R}_{\Sigma_F}((x_0, 0), T)$ or $\mathcal{R}_{\Sigma_F}((x_0, 0), \leq T)$, depending on whether we are looking at the fixed or free interval problem, respectively.

In the case of time-optimal control, the relationship between optimal trajectories and reachable sets can be made, not just for the reachable sets of Σ_F, but for those of Σ. Thus we take $F(x, u) = 1$, and consider the problem $\mathscr{P} = \mathscr{P}(\Sigma, F, \{x_0\}, \{x_1\})$ of steering from x_0 to x_1 in minimum time.

Proposition S4.20 (Time-optimal trajectories lie on the boundary of the reachable set). *With the notation of the preceding paragraph, if (γ, u) solves $\mathscr{P}$ and if the minimum time is T, then $x_1 \in \mathrm{bd}(\mathcal{R}_\Sigma(x_0, T))$.*

Proof. Note that

$$\mathcal{R}_{\Sigma_F}((x_0, 0), t) = \mathcal{R}_\Sigma(x_0, t) \times \{t\}, \qquad t \in [0, T].$$

If $x_1 \in \mathrm{int}(\mathcal{R}_\Sigma(x_0, T)$, then $(x_1, T) \in \mathrm{int}(\mathcal{R}_{\Sigma_F}((x_0, 0), \leq T)$. Now, since every neighborhood of $(x_1, T) \in \mathsf{M}_F$ contains a point in $\ell_{x_1,T}$, this means that there are points in $\mathcal{R}_{\Sigma_F}((x_0, 0), \leq T)$ that are also in $\ell_{x_1,T}$. But this contradicts Proposition S4.19. ■

S4.3.2 An interpretation of the Maximum Principle

The above relationship between solutions to optimal control problems and reachable sets actually leads to a fairly direct interpretation of the Maximum Principle. We suppose that (γ, u) is a solution to one of the optimal control problems $\mathscr{P}(\Sigma, F, \{x_0\}, \{x_1\})$ or $\mathscr{P}_{[a,b]}(\Sigma, F, \{x_0\}, \{x_1\})$. Let us simply denote this by $\mathscr{P}$. We suppose that the solution (γ, u) to $\mathscr{P}$ is defined on $[0, T]$ for concreteness. If $\mathscr{P}$ is the free interval problem, then let us denote $\mathcal{R}_{F,T} = \mathcal{R}_{\Sigma_F}((x_0, 0), \leq T)$, and if $\mathscr{P}$ is the fixed interval problem, then let us denote $\mathcal{R}_{F,T} = \mathcal{R}_{\Sigma_F}((x_0, 0), T)$. Let us similarly denote $\mathcal{R}_{F,T}$ as either $\mathcal{R}_\Sigma(x_0, \leq T)$ or $\mathcal{R}_\Sigma(x_0, T)$, respectively. Let us denote by (γ_F, u) the controlled trajectory of $\Sigma_F = (\mathsf{M}_F, f_F, U)$ corresponding to (γ, u). We know then that $\gamma_F(T) \in \mathrm{bd}(\mathcal{R}_{F,T})$. It turns out (this is nontrivial) that the boundary to the reachable set $\mathcal{R}_{F,T}$ possesses a normal as shown in Figure S4.2. This has to

Figure S4.2. The reachable set possesses a normal at its boundary

do somehow with the convexity of the reachable set. However, convexity in a manifold is not so easy to understand; one has to make the notion stand up using the "tangent space" to the boundary of the reachable set. In any case, the normal vector, denoted by $\Lambda(T)$ in Figure S4.2, should be thought of as being in $\mathsf{T}^*_{\gamma_F(T)}\mathsf{M}_F$, and $\mathrm{coann}(\Lambda(T))$ is then a hyperplane in $\mathsf{T}_{\gamma_F(T)}\mathsf{M}_F$

"tangent" to the boundary. But we know that, not only does $\gamma_F(T)$ lie on the boundary of $\mathcal{R}_{F,T}$, but, for any $\tau \in [0,T]$, we have $\gamma_F(\tau) \in \mathrm{bd}(\mathcal{R}_{F,\tau})$. Therefore, the normal construction at time T can be applied for all $\tau \in [0,T]$ (again see Figure S4.2). Thus we end up with a covector field Λ along γ_F. Now consider the reference flow $\Phi^{f_u}_{t_1,t_2}$ defined by the Carathéodory vector field $f_{F,u}\colon (t,(x,\kappa)) \mapsto f_F((x,\kappa),u(t))$. This flow will generate the extremal from $\gamma_F(0)$ to $\gamma_F(T)$ as in Figure S4.2 (this is the dashed arc in the figure). Since $\Phi^{f_{F,u}}_{\tau,T}(\mathcal{R}_{F,\tau}) \subset \mathcal{R}_{F,T}$, we can choose $\Lambda(\tau)$ such that $T^*_{\gamma_F(\tau)}(\Phi^{f_{F,u}}_{\tau,T})(\Lambda(\tau)) = \Lambda(T)$. What one can then show is that the act of defining the covector field Λ such that $T^*_{\gamma_F(\tau)}(\Phi^{f_{F,u}}_{\tau,T})(\Lambda(\tau)) = \Lambda(T)$ for each $\tau \in [0,T]$ is exactly equivalent to defining Λ as being an integral curve for the Hamiltonian vector field with LIC$^\infty$ Hamiltonian $H_{f_{F,u}}(t,\alpha_{(x,\kappa)}) = \alpha_{(x,\kappa)} \cdot f_F((x,\kappa),u(t))$. This comes about essentially for the following reason. The flow of the tangent lift of $f_{F,u}$ will have the property that it maps tangent spaces to the reachable set to other tangent spaces to the reachable. This requires proof, but is certainly believable if one refers to the discussion of the tangent lift in Section S1.2.1. Now, by understanding the relationship between the tangent lift and the cotangent lift (see Section S1.2.2), it becomes reasonable that the flow of the cotangent lift should map the coannihilator of a tangent space to the reachable set to another coannihilator of a tangent space to a reachable set. Now we note the cotangent lift of $f_{F,u}$ is exactly the Hamiltonian vector field with LIC$^\infty$-Hamiltonian $H_{f_{F,u}}$ (see Remark S1.11–1).

This relationship between the Maximum Principle and the reachable set also gives us a means of understanding what happens when we have an abnormal controlled extremal (γ,u). It turns out that the covector field Λ above can be written $\Lambda(\tau) = (\chi(\tau),\chi_0) \in \mathsf{T}^*_{\gamma_F(\tau)}\mathsf{M}_F \simeq \mathsf{T}^*_{\gamma(\tau)}\mathsf{M} \times \mathbb{R}$. (The not-so-obvious thing here is that χ_0 is independent of τ.) If the controlled extremal (γ,u) has the property that $\gamma'(T)$ is tangent to the boundary of $\mathcal{R}_{F,T}$, then it must be the case that $\chi_0 = 0$. From the Maximum Principle, implicit in this is the fact that, if $\gamma'(T)$ is tangent to $\mathcal{R}_{F,T}$, then $\gamma'(\tau)$ is tangent to $\mathcal{R}_{F,\tau}$ for each $\tau \in [0,T]$. Thus abnormal controlled extremals can be thought of as those that are tangent to the boundary of the reachable set for Σ. Note that it may be possible to choose $\chi_0 = 0$ even for normal controlled extremals. What distinguishes abnormal controlled extremals is that χ_0 *must* be zero.

Now let us briefly consider this relationship between optimal control and reachable sets as it pertains to the existence of solutions to optimal control problems. Again, we consider the extended system Σ_F with $\mathcal{R}_{F,T}$ the corresponding reachable sets, the exact nature of which depends on whether we are considering the fixed or free interval problem. Since a controlled extremal (γ,u) defined on $[0,T]$ and solving an optimal control problem $\mathscr{P}$ must satisfy $\gamma_F(T) \in \mathrm{bd}(\mathcal{R}_{F,T})$, this boundary of the reachable set must be nonempty. This will frequently preclude the existence of solutions in cases where the controls are not bounded. Certainly in time-optimal control, one very often bounds the controls in order to ensure that the problem possesses a solution.

S4.3.3 An example

We take the system $\Sigma = (\mathsf{M}, f, U)$ where

1. $\mathsf{M} = \mathbb{R}^2$,
2. $f((x^1, x^2), u) = x^2 \frac{\partial}{\partial x^1} + (-x^1 + u) \frac{\partial}{\partial x^2}$,
3. $U = [-1, 1] \subset \mathbb{R}$.

The cost function we choose is that associated with time-optimization; thus we take $F(\boldsymbol{x}, u) = 1$. We consider the problem $\mathscr{P}(\Sigma, F, \{(x_0^1, x_0^2)\}, \{x_1^1, x_1^2\})$. The Hamiltonian for this system is

$$H_{\Sigma,F}((\boldsymbol{x}, \boldsymbol{p}), u) = p_1 x^2 - p_2 x^1 + p_2 u - 1.$$

This gives the equations governing controlled extremals as

$$\dot{x}^1 = x^2, \quad \dot{x}^2 = -x^1 + u, \quad \dot{p}_1 = p_2, \quad \dot{p}_2 = -p_1.$$

We may solve the equations for the adjoint variables p_1 and p_2 directly:

$$p_1(t) = A \sin(t - \phi), \quad p_2 = A \cos(t - \phi) \tag{S4.6}$$

for some $A, \phi \in \mathbb{R}$.

The control u for a controlled extremal satisfies

$$p_2(t) u(t) = \max \{ p_2(t) \tilde{u} \mid \ \tilde{u} \in U \},$$

meaning that, when $p_2(t) < 0$, we have $u(t) = -1$, and when $p_2(t) > 0$ we have $u(t) = +1$. Thus $u(t)$ alternates between $+1$ and -1, depending on the sign of $p_2(t)$. However, given the form of $p_2(t)$, this means that $u(t)$ switches every π seconds.

This shows that controlled extremals will be concatenations of solutions of the two differential equations

1. $\dot{x}^1 = x^2$, $\dot{x}^2 = -x^1 + 1$ and
2. $\dot{x}^1 = x^2$, $\dot{x}^2 = -x^1 - 1$.

The solutions to the first equation are

$$x^1(t) = B_1 \sin(t - \psi_1) + 1, \quad x^2(t) = B_1 \cos(t - \psi_1), \tag{S4.7}$$

for constants $B_1, \psi_1 \in \mathbb{R}$. These are simply circles in the (x^1, x^2)-plane centered at $(1, 0)$. In like manner, the solutions for the other class of optimal arcs are determined by

$$x^1(t) = B_2 \sin(t - \psi_2) - 1, \quad x^2(t) = B_2 \cos(t - \psi_2), \tag{S4.8}$$

for constants $B_2, \psi_2 \in \mathbb{R}$. These are simply circles in the (x^1, x^2)-plane centered at $(-1, 0)$. Thus, to steer from (x_0^1, x_0^2) to (x_1^1, x_1^2) in a time-optimal

Figure S4.3. Two concatenations of circles to form an extremal. The solid line is the solution to optimal control problem, and the dashed line is another extremal.

manner, one would go from (x_0^1, x_0^2) to (x_1^1, x_1^2) along a curve consisting of a concatenation of circles centered at $(1, 0)$ and at $(-1, 0)$ (see Figure S4.3).

The question then arises, “What happens if $p_2(t) = 0$ on an interval so that we cannot determine $u(t)$ using the condition that the control Hamiltonian be maximized?” One can easily see that this is the case of a singular extremal. In such a case, we note that (S4.6) implies that $p_1(t)$ is also zero on the same interval. Since the differential equation for the (p_1, p_2) is linear, this means that they must be identically zero along the entire extremal. Furthermore, condition (vi) of Theorem S4.11 gives $\chi_0 = 0$, which violates condition (iv) of Theorem S4.11. Thus this problem possesses no singular extremals.

Next we look at the abnormal controlled extremals. In this case constancy (in fact, equality with zero) of the Hamiltonian, as guaranteed by the Maximum Principle, tells us that we must have

$$H_{\Sigma,0}((\boldsymbol{x}, \boldsymbol{p}), u) = p_1 x^2 - p_2 x^1 + p_2 u = 0.$$

A straightforward calculation, using (S4.6), (S4.7) or (S4.8), and the fact that $u(t) = \operatorname{sign}(p_2(t))$ gives

$$p_1 x^2 - p_2 x^1 + p_2 u = AB \sin(\psi - \phi).$$

Thus a controlled extremal is possibly abnormal if and only if $\psi - \phi = n\pi$, $n \in \mathbb{Z}$. Note that, to verify abnormality, one must also verify that there are no controlled extremals with $\chi_0 = 1$ that give the same controlled extremal, but this is easily done in this example.

For this problem, there exist time-optimal trajectories that are abnormal controlled extremals. For example, suppose that one wishes to go from $(x_0^1, x_0^2) = (0,0)$ to $(x_1^1, x_1^2) = (2,0)$. In this case the time-optimal control is given by $u(t) = 1$ that is applied for π seconds. The corresponding trajectory in state space is

$$x^1(t) = -\cos t + 1, \quad x^2(t) = \sin t.$$

That this is the time-optimal trajectory is intuitively clear: one pushes as hard as one can in the direction one wants to go until one gets there. However, this controlled extremal is abnormal. Let us see how this works. Since the controlled trajectory (γ, u) just described is minimizing, it must satisfy the conditions of the Maximum Principle. In particular, the maximization condition on the Hamiltonian must be realized. This means that $p_2(t)$ must be positive for $0 \le t \le \pi$, except possibly at the endpoints. If $p_2(t)$ changes sign in the interval $[0, \pi]$, then u must also change sign, but this cannot happen since $u(t) = 1$. This implies that $p_2(t) = A \sin t$, and so this immediately gives $p_1(t) = -A \cos t$. We see then that we may take $\phi = \frac{\pi}{2}$ and $\psi = \frac{\pi}{2}$. Given our characterization of abnormal controlled extremals, this shows that the time-optimal control we have found is only realizable as an abnormal controlled extremal.

Let us see if we can provide a geometric interpretation of what is going on here. In Figure S4.4 we show a collection of concatenated extremals that

Figure S4.4. The set of points reachable from $(0,0)$ in time π

emanate from the origin. From this picture it is believable that the set of points reachable from $(0,0)$ in time π is precisely the circle of radius 2 in the (x^1, x^2)-plane. Why are the points $(\pm 2, 0)$ distinguished? (We have only looked at the point $(2, 0)$, but the same arguments hold for $(-2, 0)$.) Well, look at how the extremals touch the boundary of the reachable set. Only at $(\pm 2, 0)$ do the extremals approach the boundary such that they are tangent to the supporting hyperplane. This is what we talked about at the conclusion of Section S4.3.2.

S4.4 The Maximum Principle for affine connection control systems

We now reap the benefits of the work in Supplement 1 to provide a concise translation of the Maximum Principle for systems whose drift vector field is the geodesic spray associated with an affine connection, and whose control vector fields are vertically lifted vector fields. As we shall see, one of the interesting features of the Maximum Principle for affine connection control systems is that the equations for the Hamiltonian vector field in the Maximum Principle decouple into the control equations, along with a separate adjoint equation for the covector field along the controlled arc. The development of this relies on the various splittings developed in Section S1.3. These splittings have the additional feature of simplifying the form of the Hamiltonian function.

We adopt the notation for affine connection control systems from the text. The only additional concept we require is that of a ***controlled arc*** for an affine connection control system $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$, by which we mean a controlled trajectory (γ, u) defined on a compact interval i.e., an interval of the form $[a, b]$. The set of controlled arcs for Σ is denoted by $\mathrm{Carc}(\Sigma)$. In this section, it will also be notationally convenient to apply the summation convention to the linear combinations of the input vector fields. That is to say, we will abbreviate $\sum_{a=1}^m u^a Y_a$ with $u^a Y_a$ when it is convenient to do so.

S4.4.1 Optimal control problems for affine connection control systems

Of course, since an affine connection control system defines a control-affine system, and so a control system as per Definition S4.7, one may simply formulate an optimal control problem on TQ exactly as was done in Section S4.2.1. However, we wish to choose a class of cost functions that reflects the fact that the problem data for an affine connection control system is defined on Q.

Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be an affine connection control system. An ***$\mathbb{R}^m$-dependent $(0, r)$-tensor field*** on Q is a map $A\colon \mathsf{Q} \times \mathbb{R}^m \to T^0_r(\mathsf{TQ})$ such that

1. A is continuous, and
2. $q \mapsto A(q, u)$ is a C^∞ $(0, r)$-tensor field for every $u \in \mathbb{R}^m$.

Note that we are really only interested in the value of $\mathbb{R}^m$-dependent tensor fields when evaluated at points $(q,u) \in \mathsf{Q} \times U$. However, for simplicity we suppose them to be defined on all of $\mathsf{Q} \times \mathbb{R}^m$. We let $r \geq 0$ and let $A\colon \mathsf{Q} \times \mathbb{R}^m \to T^0_r(\mathsf{TQ})$ be an $\mathbb{R}^m$-dependent symmetric $(0,r)$-tensor field on Q. We let $f\colon \mathbb{R} \to \mathbb{R}$ be a class C^∞ function, and let $\mathscr{A} = (A,f)$. A ***cost function*** for Σ is a function $F_{\mathscr{A}}\colon \mathsf{TQ} \times \mathbb{R}^m \to \mathbb{R}$ of the form

$$F_{\mathscr{A}}(v_q,u) = f(A(q,u)(v_q,\dots,v_q)).$$

Let $F_{\mathscr{A}}$ be a cost function for Σ as defined above. We say that $(\gamma,u) \in \mathrm{Ctraj}(\Sigma)$ is ***$F_{\mathscr{A}}$-acceptable*** if the function $t \mapsto F_{\mathscr{A}}(\gamma'(t),u(t))$ is locally integrable. We denote by $\mathrm{Ctraj}(\Sigma,F_{\mathscr{A}})$ the set of $F_{\mathscr{A}}$-acceptable controlled trajectories. Similarly, $\mathrm{Carc}(\Sigma,F_{\mathscr{A}})$ denotes the set of $F_{\mathscr{A}}$-acceptable controlled arcs.

For $(\gamma,u) \in \mathrm{Carc}(\Sigma,F_{\mathscr{A}})$, where u and γ are defined on $[a,b]$, we define

$$A_{\Sigma,F_{\mathscr{A}}}(\gamma,u) = \int_a^b F_{\mathscr{A}}(\gamma'(t),u(t))\,\mathrm{d}t.$$

For $q_0,q_1 \in \mathsf{Q}$, $v_{q_0} \in \mathsf{T}_{q_0}\mathsf{Q}$, and $v_{q_1} \in \mathsf{T}_{q_1}\mathsf{Q}$, we denote

$$\begin{aligned}\mathrm{Carc}(\Sigma,F_{\mathscr{A}},v_{q_0},v_{q_1}) = \{(\gamma,u) \in \mathrm{Carc}(\Sigma,F_{\mathscr{A}})|\ &\gamma'(a) = v_{q_0}\\ \text{and } \gamma'(b) = v_{q_1}, \text{ where } u \text{ and } \gamma \text{ are defined on } [a,b] &\text{ for some } a,b \in \mathbb{R}\}.\end{aligned}$$

For fixed $a,b \in \mathbb{R}$ with $a < b$, we define

$$\begin{aligned}\mathrm{Carc}(\Sigma,F_{\mathscr{A}},v_{q_0},v_{q_1},[a,b]) = \{(\gamma,u) \in \mathrm{Carc}(\Sigma,F_{\mathscr{A}})|\ &u \text{ and } \gamma\\ \text{are defined on } [a,b] \text{ and } \gamma'(a) = v_{q_0} &\text{ and } \gamma'(b) = v_{q_1}\}.\end{aligned}$$

The above subsets of controlled arcs correspond to fixing an initial and final configuration *and* velocity. For affine connection control systems, it also makes sense to consider only fixing the initial and final configuration while leaving the velocities free. Thus we define

$$\begin{aligned}\mathrm{Carc}(\Sigma,F_{\mathscr{A}},q_0,q_1) = \{(\gamma,u) \in \mathrm{Carc}(\Sigma,F_{\mathscr{A}})|\ &\gamma(a) = q_0 \text{ and } \gamma(b) = q_1\\ \text{where } u \text{ and } \gamma \text{ are defined on } [a,b] &\text{ for some } a,b \in \mathbb{R}\},\end{aligned}$$

and, for fixed $a,b \in \mathbb{R}$ with $a < b$, we define

$$\begin{aligned}\mathrm{Carc}(\Sigma,F_{\mathscr{A}},q_0,q_1,[a,b]) = \{(\gamma,u) \in \mathrm{Carc}(\Sigma,F_{\mathscr{A}})|\ &\text{where } u \text{ and } \gamma\\ \text{are defined on } [a,b] \text{ and } \gamma(a) = q_0 &\text{ and } \gamma(b) = q_1\}.\end{aligned}$$

Now we define the control problems we consider.

Definition S4.21 (Optimal control problems for affine connection control systems). Let $\Sigma = (\mathsf{Q},\nabla,\mathscr{Y},U)$ be an affine connection control system, let $F_{\mathscr{A}}$ be a cost function for Σ, let $q_0,q_1 \in \mathsf{Q}$, and let $v_{q_0} \in \mathsf{T}_{q_0}\mathsf{Q}$ and $v_{q_1} \in \mathsf{T}_{q_1}\mathsf{Q}$.

(i) A controlled arc $(\gamma_*, u_*) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1})$ is a ***solution of*** $\boldsymbol{\mathscr{P}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1})}$ if $A_{\Sigma,F_{\mathscr{A}}}(\gamma_*, u_*) \le A_{\Sigma,F_{\mathscr{A}}}(\gamma, u)$ for every $(\gamma, u) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1})$.

(ii) A controlled arc $(\gamma_*, u_*) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1}, [a, b])$ is a ***solution of*** $\boldsymbol{\mathscr{P}_{[a,b]}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1})}$ if $A_{\Sigma,F_{\mathscr{A}}}(\gamma_*, u_*) \le A_{\Sigma,F_{\mathscr{A}}}(\gamma, u)$ for every $(\gamma, u) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1}, [a, b])$.

(iii) A controlled arc $(\gamma_*, u_*) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, q_0, q_1)$ is a ***solution of*** $\boldsymbol{\mathscr{P}(\Sigma, F_{\mathscr{A}}, q_0, q_1)}$ if $A_{\Sigma,F_{\mathscr{A}}}(\gamma_*, u_*) \le A_{\Sigma,F_{\mathscr{A}}}(\gamma, u)$ for every $(\gamma, u) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, q_0, q_1)$.

(iv) A controlled arc $(\gamma_*, u_*) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, q_0, q_1, [a, b])$ is a ***solution of*** $\boldsymbol{\mathscr{P}_{[a,b]}(\Sigma, F_{\mathscr{A}}, q_0, q_1)}$ if $A_{\Sigma,F_{\mathscr{A}}}(\gamma_*, u_*) \le A_{\Sigma,F_{\mathscr{A}}}(\gamma, u)$ for every $(\gamma, u) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}}, q_0, q_1, [a, b])$. •

S4.4.2 Some technical lemmata

In our proof below of the Maximum Principle for affine connection control systems, we shall need some computations concerning the representation of various tensors in the splitting of $\mathsf{T}^*\mathsf{TQ}$ described in Section S1.3.9. We gather these in this section.

We shall need some notation involving symmetric $(0, r)$-tensor fields. Let A be such a tensor field on Q. For $v_1, \dots, v_{r-1} \in \mathsf{T}_q\mathsf{Q}$, we define $\hat{A}(v_1, \dots, v_{r-1}) \in \mathsf{T}^*_q\mathsf{Q}$ by

$$\langle \hat{A}(v_1, \dots, v_{r-1}); w\rangle = A(w, v_1, \dots, v_{r-1}), \qquad w \in \mathsf{T}_q\mathsf{Q}.$$

We adopt the convention that, if A is a $(0, 0)$-tensor field (i.e., A is a function), then $\hat{A} = 0$. Obviously this notation extends to tensor fields that are $\mathbb{R}^m$-dependent. The following lemma provides the form of a certain Hamiltonian vector field that we will encounter.

Lemma S4.22. *Let A be a symmetric $(0, r)$-tensor field on Q, let $f : \mathbb{R} \to \mathbb{R}$ be of class C^∞, and, in the splitting of $\mathsf{T}^*\mathsf{TQ}$ defined in Section S1.3.9, consider a function defined by*

$$\alpha_{v_q} \oplus \beta_{v_q} \mapsto f(A(v_q, \dots, v_q)).$$

The Hamiltonian vector field on $\mathsf{T}^*\mathsf{TQ}$ generated by this function has the decomposition

$$f'(A(v_q, \dots, v_q))\big(0 \oplus 0 \oplus \big(-\nabla A(v_q, \dots, v_q) - \tfrac{r}{2}T^*(\hat{A}(v_q, \dots, v_q), v_q)\big) \oplus (-r\hat{A}(v_q, \dots, v_q))\big).$$

Proof. In natural coordinates for $\mathsf{T}^*\mathsf{TQ}$, the Hamiltonian function defined in the lemma is given by

$$((\boldsymbol{q}, \boldsymbol{v}), (\boldsymbol{\alpha}, \boldsymbol{\beta})) \mapsto f(A_{j_1 \cdots j_r} v^{j_1} \cdots v^{j_r}).$$

The corresponding Hamiltonian vector field in natural coordinates is given by

$$f'(A_{j_1\cdots j_r}v^{j_1}\cdots v^{j_r})\Big(-\frac{\partial A_{j_1\cdots j_r}}{\partial q^i}v^{j_1}\cdots v^{j_r}\frac{\partial}{\partial \alpha_i}-rA_{ij_2\cdots j_r}v^{j_2}\cdots v^{j_r}\frac{\partial}{\partial \beta_i}\Big).$$

We now express this in a basis adapted to the splitting of $\mathsf{T}_{\Lambda_{v_q}}\mathsf{T}^*\mathsf{T}\mathsf{Q}$ to get

$$\begin{aligned}f'(A_{j_1\cdots j_r}v^{j_1}\cdots v^{j_r})\Big(-\Big(&\frac{\partial A_{j_1\cdots j_r}}{\partial q^i}v^{j_1}\cdots v^{j_r}\\&-\frac{r}{2}(\Gamma^{\ell}_{ik}+\Gamma^{\ell}_{ki})v^kA_{\ell j_2\cdots j_r}v^{j_2}\cdots v^{j_r}\Big)\frac{\partial}{\partial \alpha_i}\\&-rA_{ij_2\cdots j_r}v^{j_2}\cdots v^{j_r}\Big(\frac{\partial}{\partial\beta_i}+\frac{1}{2}(\Gamma^i_{k\ell}+\Gamma^i_{\ell k})v^{\ell}\frac{\partial}{\partial\alpha_k}\Big)\Big).\end{aligned}$$

From this we see that the representation of the Hamiltonian vector field is

$$\begin{aligned}\alpha_{v_q}\oplus\beta_{v_q}\mapsto f'(&A(v_q,\dots,v_q))\big(0\oplus 0\\&\oplus\big(-\nabla A(v_q,\dots,v_q)-\tfrac{r}{2}T^*(\hat{A}(v_q\dots,v_q),v_q)\big)\oplus(-r\hat{A}(v_q,\dots,v_q))\big).\end{aligned}$$

This completes the proof. ■

The lemma clearly extends to $\mathbb{R}^m$-dependent tensor fields.

This is an appropriate setting in which to present the following lemma, although it will not be used until Section S4.5. As we talk about symmetric $(0,r)$-tensor fields, we may also talk about symmetric $(r,0)$-tensor fields. And we generate for these latter some notation similar to that generated for the former. Precisely, if B is a symmetric $(r,0)$-tensor field and if $\alpha^1,\dots,\alpha^{r-1}\in\mathsf{T}^*_q\mathsf{Q}$, then we define $\hat{B}(\alpha^1,\dots,\alpha^{r-1})\in\mathsf{T}_q\mathsf{Q}$ by

$$\langle\beta;\hat{B}(\alpha^1,\dots,\alpha^{r-1})\rangle=B(\beta,\alpha^1,\dots,\alpha^{r-1}),\qquad \beta\in\mathsf{T}^*_q\mathsf{Q}.$$

We now state the lemma.

Lemma S4.23. *Let B be a symmetric $(r,0)$-tensor field on Q and define a function on $\mathsf{T}^*\mathsf{T}\mathsf{Q}$ by*

$$\alpha_{v_q}\oplus\beta_{v_q}\mapsto B(\beta_{v_q},\dots,\beta_{v_q}).$$

The Hamiltonian vector field generated by this function has the representation

$$0\oplus(r\hat{B}(\beta_{v_q},\dots,\beta_{v_q}))\oplus\big(\nabla B(\beta_{v_q},\dots,\beta_{v_q})-\tfrac{r}{2}T^*(\beta_{v_q},\hat{B}(\beta_{v_q},\dots,\beta_{v_q}))\big)\oplus 0.$$

Proof. In natural coordinates for $\mathsf{T}^*\mathsf{T}\mathsf{Q}$ the function in the lemma has the form

$$((\boldsymbol{q},\boldsymbol{v}),(\boldsymbol{\alpha},\boldsymbol{\beta}))\mapsto B^{j_1\cdots j_r}\beta_{j_1}\cdots\beta_{j_r}.$$

Thus the Hamiltonian vector field associated with this function is given in natural coordinates by

$$rB^{ij_2\cdots j_r}\beta_{j_2}\cdots\beta_{j_r}\frac{\partial}{\partial v^i}-\frac{\partial B^{j_1\cdots j_r}}{\partial q^i}\beta_{j_1}\cdots\beta_{j_r}\frac{\partial}{\partial\alpha_i}.$$

If we write this in the splitting of $\mathsf{T}_{\Lambda_{v_q}}\mathsf{T}^*\mathsf{TQ}$, then we obtain the decomposition of the Hamiltonian vector field as

$$0\oplus(r\hat{B}(\beta_{v_q},\dots,\beta_{v_q}))\oplus\big(\nabla B(\beta_{v_q},\dots,\beta_{v_q})-\tfrac{r}{2}T^*(\beta_{v_q},\hat{B}(\beta_{v_q},\dots,\beta_{v_q}))\big)\oplus 0,$$

just as we have asserted. ■

As a final technical lemma, we prove the form of $(\mathrm{vlft}(X))^{T^*}$ for a vector field X on Q. To give this formula, we need some notation. For $\alpha_q\in\mathsf{T}_q^*\mathsf{Q}$, define $(\nabla X)^*(\alpha_q)\in\mathsf{T}_q^*\mathsf{Q}$ by

$$\langle(\nabla X)^*(\alpha_q);w_q\rangle=\langle\alpha_q;\nabla_{w_q}X\rangle,\qquad w_q\in\mathsf{T}_q\mathsf{Q}.$$

We now adopt the same notation as used in Proposition S1.36. The proof is accomplished easily in coordinates using the formula (S1.4).

Lemma S4.24. *If X is a vector field on Q, then*

$$(\mathrm{vlft}(X))^{T^*}(\alpha_{v_q}\oplus\beta_{v_q})=0\oplus X(q)\oplus\big(\tfrac{1}{2}T^*(\beta_{v_q},X(q))-(\nabla X)^*(\beta_{v_q})\big)\oplus 0.$$

Remark S4.25. This representation of $(\mathrm{vlft}(X))^{T^*}$ has a further geometric interpretation as follows. Let X^{T^*} be the cotangent lift of X to a vector field on $\mathsf{T}^*\mathsf{Q}$. This vector field may then be written with respect to the decomposition corresponding to the connection on $\pi_{\mathsf{T}^*\mathsf{Q}}\colon\mathsf{T}^*\mathsf{Q}\to\mathsf{Q}$ given in Section S1.3.7. If we do so, then we have

$$X^{T^*}(\alpha_q)=X(q)\oplus\big(\tfrac{1}{2}T^*(\alpha_q,X(q))-(\nabla X)^*(\alpha_q)\big).$$

The interested reader will see that this is consistent with our explanation in Section S1.2.4 of the relationship between $(\mathrm{vlft}(X))^{T^*}$ and $\mathrm{vlft}(X^{T^*})$. •

S4.4.3 The Maximum Principle for affine connection control systems

Before stating the Maximum Principle for the systems we are investigating, let us look at the Hamiltonian for these systems. In doing so, we will use the splitting of $\mathsf{T}^*\mathsf{TQ}$ that we presented in Proposition S1.36. Thus we write $\Lambda_{v_q}\in\mathsf{T}^*_{v_q}\mathsf{TQ}$ as $\alpha_{v_q}\oplus\beta_{v_q}$ for some appropriately defined $\alpha_{v_q},\beta_{v_q}\in\mathsf{T}_q^*\mathsf{Q}$. The ***Hamiltonian*** for an affine connection control system Σ with cost function $F_{\mathscr{A}}$ is the function on $\mathsf{T}^*\mathsf{TQ}\times\mathbb{R}^m$ defined by

$$H_{\Sigma,F_{\mathscr{A}}}(\alpha_{v_q}\oplus\beta_{v_q},u)=\alpha_{v_q}\cdot v_q+u^a(\beta_{v_q}\cdot Y_a(q))-F_{\mathscr{A}}(v_q,u).$$

The ***maximum Hamiltonian*** is then defined in the usual manner:

$$H^{\max}_{\Sigma,F_{\mathscr{A}}}(\alpha_{v_q}\oplus\beta_{v_q})=\sup\big\{H_{\Sigma,F_{\mathscr{A}}}(\alpha_{v_q}\oplus\beta_{v_q},u)\bigm| u\in U\big\}.$$

Let $(\gamma, u) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}})$ with u and γ defined on an interval $[a, b]$. An LAD covector field $\lambda\colon [a, b] \to \mathsf{T}^*\mathsf{Q}$ along γ is ***maximizing for $(\Sigma, F_{\mathscr{A}})$ along u*** if

$$H_{\Sigma,F_{\mathscr{A}}}(\theta(t) \oplus \lambda(t)) \geq H^{\max}_{\Sigma,F_{\mathscr{A}}}(\theta(t) \oplus \lambda(t))$$

for almost every $t \in [a, b]$, and where

$$\begin{aligned}\theta(t) = \tfrac{1}{2}T^*(\lambda(t), \gamma'(t)) &- \nabla_{\gamma'(t)}\lambda(t)\\ &+ r\lambda_0 f'(A(v_q, \dots, v_q))\hat{A}(\gamma'(t), \dots, \gamma'(t)), \qquad t \in [a, b].\end{aligned}$$

Our main result in this chapter is the following.

Theorem S4.26 (Maximum Principle for affine connection control systems). *Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be an affine connection control system with $F_{\mathscr{A}}$ a cost function for Σ, where $\mathscr{A} = (A, f)$. Suppose that $(\gamma, u) \in$* $\mathrm{Carc}(\Sigma, F_{\mathscr{A}})$ *is a solution of $\mathscr{P}_{[a,b]}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1})$. Then there exists an* LAD *covector field $\lambda\colon [a, b] \to \mathsf{T}^*\mathsf{Q}$ along γ and a constant $\lambda_0 \in \{0, 1\}$ with the following properties:*

(i) for almost every $t \in [a, b]$ we have

$$\begin{aligned}&\nabla^2_{\gamma'(t)}\lambda(t) + R^*(\lambda(t), \gamma'(t))\gamma'(t) - T^*(\nabla_{\gamma'(t)}\lambda(t), \gamma'(t))\\ &= u^a(t)(\nabla Y_a)^*(\lambda(t)) - \lambda_0 f'(A(\gamma'(t), \dots, \gamma'(t)))\Big(\nabla A(\gamma'(t), \dots, \gamma'(t))\\ &- r(\nabla_{\gamma'(t)}\hat{A})(\gamma'(t), \dots, \gamma'(t)) - r(r-1)u^a(t)\hat{A}(Y_a(\gamma(t)), \gamma'(t), \dots, \gamma'(t))\\ &\qquad + rT^*(\hat{A}(\gamma'(t), \dots, \gamma'(t)), \gamma'(t))\Big)\\ &\quad + r\lambda_0 f''(A(\gamma'(t), \dots, \gamma'(t)))\Big(\nabla A(\gamma'(t), \dots, \gamma'(t); \gamma'(t))\\ &\qquad + ru^a(t)A(Y_a(\gamma(t)), \gamma'(t), \dots, \gamma'(t))\Big)\hat{A}(\gamma'(t), \dots, \gamma'(t));\end{aligned}$$

(ii) λ is maximizing for $(\Sigma, \lambda_0 F_{\mathscr{A}})$ along u;

(iii) either $\lambda_0 = 1$ or $\theta(a) \oplus \lambda(a) \neq 0$;

(iv) there exists a constant $C \in \mathbb{R}$ such that $H_{\Sigma,F_{\mathscr{A}}}(\theta(t) \oplus \lambda(t), u(t)) = C$ a.e.,

with

$$\begin{aligned}\theta(t) = \tfrac{1}{2}T^*(\lambda(t), \gamma'(t)) &- \nabla_{\gamma'(t)}\lambda(t)\\ &+ r\lambda_0 f'(A(v_q, \dots, v_q))\hat{A}(\gamma'(t), \dots, \gamma'(t)), \qquad t \in [a, b].\end{aligned}$$

If (γ, u) is a solution of $\mathscr{P}(\Sigma, F_{\mathscr{A}}, v_{q_0}, v_{q_1})$, then conditions (i)–(iii) hold, and condition (iv) can be replaced with

(v) $H_{\Sigma,F_{\mathscr{A}}}(\theta(t) \oplus \lambda(t)) = 0$ a.e.

If (γ, u) is a solution of $\mathscr{P}_{[a,b]}(\Sigma, F_{\mathscr{A}}, q_0, q_1)$, then conditions (i)–(iv) hold and, in addition, $\lambda(a) = 0$ and $\lambda(b) = 0$. If (γ, u) is a solution of $\mathscr{P}(\Sigma, F_{\mathscr{A}}, q_0, q_1)$, then conditions(i)–(iii) hold, condition (v) holds, and $\lambda(a) = 0$ and $\lambda(b) = 0$.

Proof. We will show the equivalence of the conditions in the theorem to those in the general Maximum Principle, stated as Theorem S4.11, in the case when $S_0 = \{v_{q_0}\}$ and $S_1 = \{v_{q_1}\}$. First we relate the covector field λ along γ to the integral curve χ of the Hamiltonian vector field with Hamiltonian $H_{\Sigma,F_{\mathscr{A}}}$ as asserted in the Maximum Principle. We claim that the curve

$$t \mapsto \Big(\tfrac{1}{2}T^*(\lambda(t),\gamma'(t)) - \nabla_{\gamma'(t)}\lambda(t) \\ + r\lambda_0 f'(A(v_q,\dots,v_q))\hat{A}(\gamma'(t),\dots,\gamma'(t))\Big) \oplus \lambda(t) \quad \text{(S4.9)}$$

exactly represents χ with respect to our splitting of $\mathsf{T}^*_{\gamma'(t)}\mathsf{TQ}$. To show this, we must show that (S4.9) is an integral curve of the time-dependent Hamiltonian vector field with Hamiltonian $(t,\alpha_{v_q}\oplus\beta_{v_q}) \mapsto H_{\Sigma,\lambda_0 F_{\mathscr{A}}}(\alpha_{v_q}\oplus\beta_{v_q},u(t))$. Note that $H_{\Sigma,\lambda_0 F_{\mathscr{A}}}$ is the sum of three functions: (1) $\Lambda_{v_q}\mapsto \Lambda_{v_q}\cdot S(v_q)$, (2) $\Lambda_{v_q}\mapsto u^a(t)(\Lambda_{v_q}\cdot \operatorname{vlft}(Y_a(q)))$, and (3) $\Lambda_{v_q}\mapsto -\lambda_0 F_{\mathscr{A}}(q,u(t))$. Thus the Hamiltonian vector field will be the sum of the three Hamiltonian vector fields corresponding to the three Hamiltonians. Let us write these three Hamiltonian vector fields in the splitting of $\mathsf{T}_{\Lambda_{v_q}}\mathsf{T}^*\mathsf{TQ}$. In each case we write, in the usual manner, $\Lambda_{v_q} = \alpha_{v_q}\oplus\beta_{v_q}$. By Propositions S1.29 and S1.36, the Hamiltonian vector field for the Hamiltonian (1) has the representation

$$\alpha_{v_q}\oplus\beta_{v_q} \mapsto v_q \oplus 0 \oplus \big(R^*(\beta_{v_q},v_q)v_q + \tfrac{1}{2}(\nabla_{v_q}T^*)(\beta_{v_q},v_q) \\ - \tfrac{1}{4}T^*(T^*(\beta_{v_q},v_q),v_q)\big) \oplus (-\alpha_{v_q}).$$

By Lemma S4.24, the Hamiltonian vector field associated with the Hamiltonian (2) has the representation

$$\alpha_{v_q}\oplus\beta_{v_q} \mapsto 0 \oplus u^a(t)Y_a(q) \\ \oplus \big(u^a(t)\tfrac{1}{2}T^*(\beta_{v_q},Y_a(q)) - u^a(t)(\nabla Y_a)^*(\beta_{v_q})\big) \oplus 0.$$

Now let us compute the representation of the Hamiltonian vector field associated with the Hamiltonian (3). In this case, we use Lemma S4.22 to see that the Hamiltonian vector field for the Hamiltonian (3) has the representation

$$\alpha_{v_q}\oplus\beta_{v_q} \mapsto -\lambda_0 f'(A(v_q,\dots,v_q))\big(0\oplus 0\oplus\big(-\nabla A(v_q,\dots,v_q) \\ - \tfrac{r}{2}T^*(\hat{A}(v_q\dots,v_q),v_q)\big)\oplus(-r\hat{A}(v_q,\dots,v_q))\big).$$

Here we have suppressed the explicit dependence of A on u, but it should be regarded as being implicit.

We now collect this all together. We write the integral curve of the Hamiltonian vector field as $\theta(t)\oplus\lambda(t)$, similar to Lemma S1.37. From this lemma we then have

$$
\begin{aligned}
\nabla_{\gamma'(t)}\theta(t) = {} & R^*(\lambda(t),\gamma'(t))\gamma'(t) + \tfrac{1}{2}(\nabla_{\gamma'(t)}T^*)(\lambda(t),\gamma'(t)) \\
& - \tfrac{1}{4}T^*(T^*(\lambda(t),\gamma'(t)),\gamma'(t)) + u^a(t)\tfrac{1}{2}T^*(\lambda(t),Y_a(t)) \\
& - u^a(t)(\nabla Y_a)^*(\lambda(t)) + \lambda_0 f'(A(\gamma'(t),\dots,\gamma'(t)))\nabla A(\gamma'(t),\dots,\gamma'(t)) \\
& + \tfrac{r}{2}\lambda_0 f'(A(\gamma'(t),\dots,\gamma'(t)))T^*(\hat{A}(\gamma'(t),\dots,\gamma'(t)),\gamma'(t)) \\
& + \tfrac{1}{2}T^*(\theta(t),\gamma'(t)), \\
\nabla_{\gamma'(t)}\lambda(t) = {} & -\theta(t) + r\lambda_0 f'(A(\gamma'(t),\dots,\gamma'(t)))\hat{A}(\gamma'(t),\dots,\gamma'(t)) \\
& + \tfrac{1}{2}T^*(\lambda(t),\gamma'(t)).
\end{aligned}
$$

Note that the right-hand side of the second equation is LAC since $t \mapsto \theta(t) \oplus \lambda(t)$ is the integral curve for an LIC$^\infty$-Hamiltonian vector field and since $t \mapsto \gamma'(t)$ is LAC. Thus λ satisfies a first-order time-dependent differential equation that is LAC in time. Therefore we may conclude that $t \mapsto \lambda(t)$ is LAD. Thus we may covariantly differentiate the second of these equations and substitute the first of the equations into the resulting expression. The result is

$$
\begin{aligned}
\nabla^2_{\gamma'(t)}\lambda(t) = {} & -R^*(\lambda(t),\gamma'(t))\gamma'(t) + T^*(\nabla_{\gamma'(t)}\lambda(t),\gamma'(t)) + u^a(t)(\nabla Y_a)^*(\lambda(t)) \\
& - \lambda_0 f'(A(\gamma'(t),\dots,\gamma'(t)))\Big(\nabla A(\gamma'(t),\dots,\gamma'(t)) - r(\nabla_{\gamma'(t)}\hat{A})(\gamma'(t),\dots,\gamma'(t)) \\
& - r(r-1)u^a(t)\hat{A}(Y_a(\gamma(t)),\gamma'(t),\dots,\gamma'(t)) + rT^*(\hat{A}(\gamma'(t),\dots,\gamma'(t)),\gamma'(t))\Big) \\
& + r\lambda_0 f''(A(\gamma'(t),\dots,\gamma'(t)))\Big(\nabla A(\gamma'(t),\dots,\gamma'(t);\gamma'(t)) \\
& + ru^a(t)A(Y_a(\gamma(t)),\gamma'(t),\dots,\gamma'(t))\Big)\hat{A}(\gamma'(t),\dots,\gamma'(t)),
\end{aligned}
$$

which holds a.e. From this we conclude that $t \mapsto \chi(t)$ as defined by (S4.9) has the property that $\chi'(t) = X_{H^u_{\Sigma,\lambda_0 F_{\mathscr{A}}}}$ a.e. Thus the existence of λ satisfying (i) is equivalent to the existence of the integral curve χ as asserted in Theorem S4.11.

We note that, for $u \in \mathbb{R}^m$, the vector field $v_q \mapsto S(v_q) + u^a Y_a(q)$ on TQ has the form $v_q \mapsto v_q \oplus u^a Y_a(q)$ in the splitting of TTQ defined in Section S1.3.8. This shows that the Hamiltonian $H_{\Sigma,F_{\mathscr{A}}}$ has the given form in the splitting of $\mathsf{T^*TQ}$.

It is then clear that the conditions (ii)–(v) are equivalent to the conditions (iii)–(vi) of Theorem S4.11.

The final assertions regarding solutions of $\mathscr{P}_{[a,b]}(\Sigma,F_{\mathscr{A}},q_0,q_1)$ and $\mathscr{P}(\Sigma,F_{\mathscr{A}},q_0,q_1)$ follow from Theorem S4.11 in the case where $S_0 = \mathsf{T}_{q_0}\mathsf{Q}$ and $S_1 = \mathsf{T}_{q_1}\mathsf{Q}$. Note that, in the splitting $\mathsf{T}^*_{v_{q_0}}\mathsf{TQ} = \mathsf{T}^*_{q_0}\mathsf{Q} \oplus \mathsf{T}^*_{q_0}\mathsf{Q}$, we have $\operatorname{ann}(\mathsf{T}_v(\mathsf{T}_{q_0}\mathsf{Q})) = \mathsf{T}^*_{q_0}\mathsf{Q} \oplus \{0\}$ for $v \in \mathsf{T}_{q_0}\mathsf{Q}$, and similarly for q_1. ■

Remarks S4.27. 1. Let us consider the import of the preceding theorem. Were we to simply apply the Maximum Principle of Theorem S4.11, we would obtain a first-order differential equation for a covector field along trajectories in TQ. Theorem S4.26 provides a second-order differential equation for a covector field along trajectories in Q. But, more importantly, the differential equation governing the evolution of this covector field on Q provides a clear indication of how the geometry of the control system enters into the optimal control problem. We shall exploit this knowledge in the next section to formulate an optimal control problem that clearly utilizes the geometry of the control system through its affine connection.

2. We call the second-order equation for λ in (i) of Theorem S4.26 the ***adjoint equation*** for $(\Sigma, F_{\mathscr{A}})$. Note that the left-hand side of this equation is none other than the adjoint Jacobi equation. One of the features of our constructions here is that we are able to intrinsically provide an equation for the adjoint covector field that is decoupled from the control system equations. This is generally not possible when talking about control systems on manifolds, but is possible here because of the existence of the myriad Ehresmann connections associated to the affine connection ∇.
3. It is interesting and useful that we may trivially incorporate into Theorem S4.26 systems with potential energy. Let us describe how this may be done. One has a potential function V on Q and defines a Lagrangian by $L(v_q) = \frac{1}{2}\mathbb{G}(v_q, v_q) - V(q)$, where $\mathbb{G}$ is a Riemannian metric on Q. The equations of motion for the system with control vector fields $\mathscr{Y} = \{Y_1, \dots, Y_m\}$ are then

$$\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t) = -\mathrm{grad}V(\gamma(t)) + \sum_{a=1}^{m} u^a(t) Y_a(\gamma(t)).$$

Clearly it makes no difference in the general scheme if we replace $\overset{\mathbb{G}}{\nabla}$ with an arbitrary affine connection ∇, and replace $\mathrm{grad}V$ with a general vector field Y_0 on Q. The question is how to incorporate the vector field Y_0 into our affine connection control system $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$. We do this by defining a new set $\tilde{\mathscr{Y}} = \{Y_0, Y_1, \dots, Y_m\}$ of control vector fields, and a new control set $\tilde{U} = \{1\} \times U \subset \mathbb{R} \times \mathbb{R}^m \simeq \mathbb{R}^{m+1}$. One may now apply verbatim Theorem S4.26 to the new affine connection control system $\tilde{\Sigma}_{\mathrm{aff}} = (\mathsf{Q}, \nabla, \tilde{\mathscr{Y}}, \tilde{U})$. •

S4.4.4 Extremals

The language of Section S4.2.3 can be specialized to the affine connection control system setting. A direct translation would give concepts defined in terms of objects on TQ. However, since the whole point of this section was to drop everything to Q, we should be precise and drop the language of extremals to Q as well. Note that in doing so, we make a slight abuse of terminology, since we use the same language as was used in Section S4.2.3.

Let us simply give the definitions, referring to Section S4.2.3 for discussion.

Definition S4.28 (Extremal). Let $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be a C^∞-affine connection control system with a cost function $F_{\mathscr{A}}$, and let $\mathscr{P}$ be one of the four problems of Definition S4.21.

(i) A ***controlled extremal*** for $\mathscr{P}$ is $(\gamma, u) \in \mathrm{Carc}(\Sigma, F_{\mathscr{A}})$ that satisfies corresponding necessary conditions of Theorem S4.26.

(ii) An absolutely differentiable curve $\gamma\colon [a,b] \to \mathsf{Q}$ is an ***extremal*** for $\mathscr{P}$ if there exists a control u such that (γ, u) is a controlled extremal for $\mathscr{P}$.

(iii) An integrable control $u\colon [a,b] \to U$ is an ***extremal control*** for $\mathscr{P}$ if there exists a curve γ on Q such that (γ, u) is a controlled extremal for $\mathscr{P}$. •

Definition S4.29 (Adjoint covector field and constant Lagrange multiplier). Let $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be a C^∞-affine connection control system with a cost function $F_{\mathscr{A}}$. Let (γ, u) be a controlled extremal for one of the four problems of Definition S4.21, with λ a covector field along γ and $\lambda_0 \in \{0, 1\}$ as in Theorem S4.26. Then λ is a ***adjoint covector field***, and λ_0 is a ***constant Lagrange multiplier***. •

Definition S4.30 (Normal and abnormal controlled extremals). Let $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be a C^∞-affine connection control system with a cost function $F_{\mathscr{A}}$. A controlled extremal (γ, u) for one of the four problems of Definition S4.21, satisfying the necessary conditions of Theorem S4.26 with $\lambda_0 = 1$, is called ***normal***. A controlled extremal is ***abnormal*** if it satisfies the necessary conditions of Theorem S4.26 only for $\lambda_0 = 0$.

An extremal γ is ***normal*** (resp. ***abnormal***) if there exists a control u such that (γ, u) is a normal (resp. abnormal) controlled extremal. •

Definition S4.31 (Regular and singular controlled extremals). Let $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be a C^∞-affine connection control system with a cost function $F_{\mathscr{A}}$. Let (γ, u) be a controlled extremal for one of the four problems of Definition S4.21, defined on $[a,b]$, with λ_0 the constant Lagrange multiplier and λ the adjoint covector field. We say that (γ, u) is ***singular*** if, for each $t \in [a,b]$, $H_{\Sigma,\lambda_0 F}(\theta(t) \oplus \lambda(t), \bar{u}) = H^{\max}_{\Sigma,\lambda_0 F}(\theta(t) \oplus \lambda(t))$ for all $\bar{u} \in U$, where θ is defined as in Theorem S4.26. A controlled extremal that is not singular is ***regular***.

An extremal γ is ***singular*** (resp. ***regular***) if there exists a control u such that (γ, u) is a singular (resp. regular) controlled extremal. •

In Section S4.7 we will classify all singular extremals for a planar rigid body example.

S4.5 Force minimizing controls

The version of the Maximum Principle stated in Theorem S4.26 is quite general, and much of the complexity in its statement is owed to that generality. However, when looking at specific classes of optimal problems, the general form of the theorem can often be reduced to something more appealing, and we now demonstrate this by beginning to look at a simple class of cost function. We consider an interesting special case of an optimal control problem involving minimizing a function of the inputs. Since the inputs are coefficients of the input vector fields, and the input vector fields are related to forces in physical systems, we dub this the force minimization problem. This problem was considered in the fully actuated case by Noakes, Heinzinger, and Paden [1989] and Crouch and Silva Leite [1991]. A variant of the underactuated case is considered by Silva Leite, Camarinha, and Crouch [2000].

S4.5.1 The force minimization problems

We suppose that Q is equipped with a Riemannian metric $\mathbb{G}$. The cost function we consider is

$$F_{\mathrm{force}}(v_q, u) = \tfrac{1}{2}\mathbb{G}(u^a Y_a(q), u^b Y_b(q)). \tag{S4.10}$$

Here, as in Section S4.4, we apply the summation convention to the expression $u^a Y_a$. In the parlance of Section S4.4.1, we use an $\mathbb{R}^m$-dependent $(0,0)$-tensor field and we choose $f = \mathrm{id}_{\mathbb{R}}$. We choose for our control set $U = \mathbb{R}^m$.

For the sake of formality, let us define precisely the problem we are solving.

Definition S4.32 (Force minimization problems). Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be an affine connection control system with $U = \mathbb{R}^m$ and with cost function F_{force} as defined by (S4.10). Let $q_0, q_1 \in \mathsf{Q}$, and let $v_{q_0} \in \mathsf{T}_{q_0}\mathsf{Q}$ and $v_{q_1} \in \mathsf{T}_{q_1}\mathsf{Q}$.

(i) A controlled arc (γ, u) is a ***solution of*** $\boldsymbol{\mathscr{F}(\Sigma, v_{q_0}, v_{q_1})}$ if it is a solution of $\mathscr{P}(\Sigma, F_{\mathrm{force}}, v_{q_0}, v_{q_1})$.

(ii) A controlled arc (γ, u) is a ***solution of*** $\boldsymbol{\mathscr{F}_{[a,b]}(\Sigma, v_{q_0}, v_{q_1})}$ if it is a solution of $\mathscr{P}_{[a,b]}(\Sigma, F_{\mathrm{force}}, v_{q_0}, v_{q_1})$.

(iii) A controlled arc (γ, u) is a ***solution of*** $\boldsymbol{\mathscr{F}(\Sigma, q_0, q_1)}$ if it is a solution of $\mathscr{P}(\Sigma, F_{\mathrm{force}}, q_0, q_1)$.

(iv) A controlled arc (γ, u) is a ***solution of*** $\boldsymbol{\mathscr{F}_{[a,b]}(\Sigma, q_0, q_1)}$ if it is a solution of $\mathscr{P}_{[a,b]}(\Sigma, F_{\mathrm{force}}, q_0, q_1)$. •

We shall sometimes find it convenient to refer to the force minimization problems as "force-optimal control."

S4.5.2 General affine connections

It will be helpful to make a few straightforward constructions given the data for the force minimization optimal control problems. We denote by $\mathcal{Y}$ the input distribution on Q generated by $\mathscr{Y} = \{Y_1, \dots, Y_m\}$, and we suppose this distribution to have constant rank (but not necessarily rank m). The map $i_{\mathcal{Y}} \colon \mathcal{Y} \to \mathsf{TQ}$ denotes the inclusion. The map $P_{\mathcal{Y}} \colon \mathsf{TQ} \to \mathsf{TQ}$ denotes the $\mathbb{G}$-orthogonal projection onto the distribution $\mathcal{Y}$, with $P_{\mathcal{Y}_q}$ being its restriction to $\mathsf{T}_q\mathsf{Q}$. We may then define a $(0,2)$-tensor field $\mathbb{G}_{\mathcal{Y}}$ on Q by

$\mathbb{G}_{\mathscr{Y}}|\mathsf{T}_q\mathsf{Q} = P^*_{\mathscr{Y}_q}(\mathbb{G}|\mathsf{T}_q\mathsf{Q})$. That is, $\mathbb{G}_{\mathscr{Y}}$ is the restriction to $\mathscr{Y}$ of $\mathbb{G}$. We also have the associated $(2,0)$-tensor field $h_{\mathscr{Y}}$ defined by

$$h_{\mathscr{Y}}(\alpha_q, \beta_q) = \mathbb{G}_{\mathscr{Y}}(\mathbb{G}^\sharp(\alpha_q), \mathbb{G}^\sharp(\beta_q)).$$

We define the vector bundle map $h^\sharp_{\mathscr{Y}} \colon \mathsf{T}^*\mathsf{Q} \to \mathsf{T}\mathsf{Q}$ by $\langle \alpha_q; h^\sharp_{\mathscr{Y}}(\beta_q)\rangle = h_{\mathscr{Y}}(\alpha_q, \beta_q)$.

We first look at the case when $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ is a general affine connection control system with $U = \mathbb{R}^m$. Thus, in particular, ∇ is not the Levi-Civita connection associated with the Riemannian metric $\mathbb{G}$ used to define the cost function.

The Hamiltonian function on $\mathsf{T}^*\mathsf{T}\mathsf{Q} \times \mathbb{R}^m$ is given by

$$H_{\Sigma,F_{\mathrm{force}}}(\alpha_{v_q} \oplus \beta_{v_q}, u) = \alpha_{v_q} \cdot v_q + u^a(\beta_{v_q} \cdot Y_a(q)) - \tfrac{1}{2}\mathbb{G}(u^a Y_a(q), u^b Y_b(q)).$$

Let us define $A_{\mathrm{sing}}(\Sigma) \subset \mathsf{T}^*\mathsf{T}\mathsf{Q}$ by

$$A_{\mathrm{sing}}(\Sigma) = \left\{ \alpha_{v_q} \oplus \beta_{v_q} \;\middle|\; \beta_{v_q} \in \mathrm{ann}(\mathscr{Y}_q \right\}).$$

Thus the restriction of $H_{\Sigma,0}$ to $A_{\mathrm{sing}}(\Sigma) \times \mathbb{R}^m$ is independent of u.

The following result gives the form of the maximum Hamiltonian and the values of u by which the maxima are realized.

Lemma S4.33. *The following statements hold.*

(i) The maximum Hamiltonian for the cost function F_{force} is given by

$$H^{\max}_{\Sigma,F_{\mathrm{force}}}(\alpha_{v_q} \oplus \beta_{v_q}) = \alpha_{v_q} \cdot v_q + \tfrac{1}{2} h_{\mathscr{Y}}(\beta_{v_q}, \beta_{v_q}).$$

If $u \in \mathbb{R}^m$ is a point at which $H^{\max}_{\Sigma,F_{\mathrm{force}}}$ is realized, then u is determined by

$$u^a Y_a(q) = P_{\mathscr{Y}_q}(\mathbb{G}^\sharp(\beta_{v_q})). \tag{S4.11}$$

(ii) The maximum Hamiltonian with zero cost function is

$$H^{\max}_{\Sigma,0}(\alpha_{v_q} \oplus \beta_{v_q}) = \begin{cases} \alpha_{v_q} \cdot v_q, & \alpha_{v_q} \oplus \beta_{v_q} \in A_{\mathrm{sing}}(\Sigma), \\ \infty, & \textit{otherwise}. \end{cases}$$

Proof. (i) We fix the state $\alpha_{v_q} \oplus \beta_{v_q}$, and determine u so as to maximize $H_{\Sigma,F_{\mathrm{force}}}(\alpha_{v_q} \oplus \beta_{v_q})$. We first note that, with the state fixed, we may think of $H_{\Sigma,F_{\mathrm{force}}}$ as being a function on the subspace $\mathscr{Y}_q$ of $\mathsf{T}_q\mathsf{Q}$. Let us denote a typical point in $\mathscr{Y}_q$ by w and note that $H_{\Sigma,F_{\mathrm{force}}}$, as a function on $\mathscr{Y}_q$, is

$$\begin{aligned} w \mapsto\ & \alpha_{v_q} \cdot v_q + \beta_{v_q} \cdot i_{\mathscr{Y}_q}(w) - \tfrac{1}{2}\mathbb{G}(i_{\mathscr{Y}_q}(w), i_{\mathscr{Y}_q}(w)) \\ &= \alpha_{v_q} \cdot v_q + \mathbb{G}(\mathbb{G}^\sharp(\beta_{v_q}), i_{\mathscr{Y}_q}(w)) - \tfrac{1}{2}\mathbb{G}(i_{\mathscr{Y}_q}(w), i_{\mathscr{Y}_q}(w)) \\ &= \alpha_{v_q} \cdot v_q + \mathbb{G}(P_{\mathscr{Y}_q}(\mathbb{G}^\sharp(\beta_{v_q})), i_{\mathscr{Y}_q}(w)) - \tfrac{1}{2}\mathbb{G}(i_{\mathscr{Y}_q}(w), i_{\mathscr{Y}_q}(w)). \end{aligned}$$

Since this is a negative-definite quadratic function of w, it will have a unique maximum. Differentiating $H_{\Sigma,F_{\mathrm{force}}}$ with respect to w, and setting the resulting expression to 0, shows that the maximum satisfies

$$i_{\mathscr{Y}}(w_{\max}) = P_{\mathscr{Y}_q}(\mathbb{G}^\sharp(\beta_{v_q})).$$

Controls u that give $w_{\max}$ are thus as specified by (S4.11). This part of the lemma is then proved by substituting the expression for $w_{\max}$ into $H_{\Sigma,F_{\mathrm{force}}}$.

(ii) For zero cost function, $H_{\Sigma,0}$ is an affine function of u. Thus it will be bounded above if and only if the linear part is zero. This happens if and only if $\alpha_{v_q} \oplus \beta_{v_q} \in A_{\mathrm{sing}}(\Sigma)$. ■

Remark S4.34. Note that the controls are uniquely determined by the state if the vector fields $Y_1, \dots, Y_m$ are linearly independent. Otherwise, there will be multiple vectors u that satisfy (S4.11), all of which give rise to the same maximum Hamiltonian $H^{\max}_{\Sigma,F_{\mathrm{force}}}$. Note that, in using Theorem S4.26, we allow $\lambda_0 = 0$ for and only for initial conditions lying in $A_{\mathrm{sing}}(\Sigma)$. We shall have more to say about this in Section S4.5.5. •

It is possible to determine a simplified form for the controlled extremals for the force minimization problem. To handle the abnormal case, we need to define a new tensor. Let $L(\mathrm{ann}(\mathscr{Y}) \times \mathscr{Y}; \mathsf{T}^*\mathsf{Q})$ denote the vector bundle of multilinear bundle maps from $\mathrm{ann}(\mathscr{Y}) \times \mathscr{Y}$ to $\mathsf{T}^*\mathsf{Q}$. To describe the abnormal controlled extremals, it will be helpful to define $B_{\mathscr{Y}} \in \Gamma^\infty(L(\mathrm{ann}(\mathscr{Y}) \times \mathscr{Y}; \mathsf{T}^*\mathsf{Q}))$ by

$$\langle B_{\mathscr{Y}}(\alpha, Y); X\rangle = \langle \alpha; \nabla_X Y\rangle .$$

That $B_{\mathscr{Y}}(\alpha, Y)$ does not depend on the derivative of Y follows since, for a function $f \in C^\infty(\mathsf{Q})$, we compute

$$\begin{aligned}\langle B_{\mathscr{Y}}(\alpha, fY); X\rangle &= \langle \alpha; \nabla_X(fY)\rangle \\ &= \langle \alpha; f\nabla_X Y\rangle + \langle \alpha; (\mathscr{L}_X f)Y\rangle \\ &= \langle f B_{\mathscr{Y}}(\alpha, Y); X\rangle ,\end{aligned}$$

since $Y \in \mathscr{Y}$ and $\lambda \in \mathrm{ann}(\mathscr{Y})$. We may now state the following theorem.

Theorem S4.35 (Maximum Principle for force minimization problem). *Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be an affine connection control system with $U = \mathbb{R}^m$. Suppose that (γ, u) is a controlled extremal for $\mathscr{F}_{[a,b]}(\Sigma, v_{q_0}, v_{q_1})$ or for $\mathscr{F}(\Sigma, v_{q_0}, v_{q_1})$ with u and γ defined on $[a, b]$, and with λ the adjoint covector field and λ_0 the Lagrange multiplier. We have the following two situations.*

(i) $\lambda_0 = 1$: In this case, it is necessary and sufficient that γ and λ together satisfy the differential equations

$$\begin{aligned}&\nabla_{\gamma'(t)}\gamma'(t) = h^\sharp_{\mathscr{Y}}(\lambda(t)),\\ &\nabla^2_{\gamma'(t)}\lambda(t) + R^*(\lambda(t), \gamma'(t))\gamma'(t) - T^*(\nabla_{\gamma'(t)}\lambda(t), \gamma'(t)) \\ &\qquad = T^*(\lambda(t), h^\sharp_{\mathscr{Y}}(\lambda(t))) - \tfrac{1}{2}\nabla h_{\mathscr{Y}}(\lambda(t), \lambda(t)).\end{aligned} \tag{S4.12}$$

(ii) $\lambda_0 = 0$*: In this case, it is necessary and sufficient that*

(a) $\nabla_{\gamma'(t)}\gamma'(t) = u^a(t)Y_a(\gamma(t))$*,*

(b) $\lambda(t) \in \operatorname{ann}(\mathcal{Y}_{\gamma(t)})$ *for* $t \in [a,b]$*, and*

(c) λ *satisfies the equation along* γ *given by:*

$$\begin{aligned}\nabla^2_{\gamma'(t)}\lambda(t) + R^*(\lambda(t),\gamma'(t))\gamma'(t) - T^*(\nabla_{\gamma'(t)}\lambda(t),\gamma'(t))&\\ = B_{\mathcal{Y}}(\lambda(t), u^a(t)Y_a(t)).&\end{aligned}$$

If (γ, u) *is a solution of* $\mathscr{F}_{[a,b]}(\Sigma, q_0, q_1)$ *or of* $\mathscr{F}(\Sigma, q_0, q_1)$*, then we additionally have* $\lambda(a) = 0$ *and* $\lambda(b) = 0$*.*

Proof. (i) We note that Lemma S4.33 provides for us a maximum Hamiltonian that is smooth. We may conclude that $H^{\max}_{\Sigma,F_{\text{force}}}$ is constant a.e. along the solutions. We conclude that, to compute the controlled extremals, it suffices to compute trajectories of the Hamiltonian vector field with Hamiltonian $H^{\max}_{\Sigma,F_{\text{force}}}$. We thus simply compute the equations corresponding to the Hamiltonian vector field with Hamiltonian

$$H^{\max}_{\Sigma,F_{\text{force}}}(\alpha_{v_q} \oplus \beta_{v_q}) = \alpha_{v_q} \cdot v_q + \tfrac{1}{2} h_{\mathcal{Y}}(\beta_{v_q}, \beta_{v_q}).$$

As usual, we use the notation corresponding to the splitting of $\mathsf{T}^*\mathsf{T}\mathsf{Q}$. We also write the vector field in the splitting of $\mathsf{T}(\mathsf{T}^*\mathsf{T}\mathsf{Q})$ as we have been doing all along. Thus the Hamiltonian vector field with Hamiltonian $H^{\max}_{\Sigma,F_{\text{force}}}$ is the sum of two Hamiltonians. By Theorem S1.38, the first Hamiltonian vector field has the representation

$$v_q \oplus 0 \oplus \big(R^*(\beta_{v_q}, v_q)v_q + \tfrac{1}{2}(\nabla_{v_q}T^*)(\beta_{v_q}, v_q) - \tfrac{1}{4}T^*(T^*(\beta_{v_q}, v_q), v_q)\big) \oplus (-\alpha_{v_q}).$$

By Lemma S4.23, the Hamiltonian vector field for the Hamiltonian $\frac{1}{2}h_{\mathcal{Y}}(\beta_{v_q}, \beta_{v_q})$ is

$$0 \oplus (h^\sharp_{\mathcal{Y}}(\beta_{v_q})) \oplus \big(\tfrac{1}{2}\nabla h_{\mathcal{Y}}(\beta_{v_q}, \beta_{v_q}) - \tfrac{1}{2}T^*(\beta_{v_q}, h^\sharp_{\mathcal{Y}}(\beta_{v_q}))\big) \oplus 0.$$

This immediately gives

$$\nabla_{\gamma'(t)}\gamma'(t) = h^\sharp_{\mathcal{Y}}(\lambda(t)),$$

which is the first of equations (S4.12). Now let $\theta(t) \oplus \lambda(t)$ be the integral curve over γ' of the Hamiltonian vector field. By Lemma S1.37, we have

$$\begin{aligned}\nabla_{\gamma'(t)}\theta(t) = {}& R^*(\lambda(t),\gamma'(t))\gamma'(t) + \tfrac{1}{2}(\nabla_{\gamma'(t)}T^*)(\lambda(t),\gamma'(t))\\ &- \tfrac{1}{4}T^*(T^*(\lambda(t),\gamma'(t))) + \tfrac{1}{2}\nabla h_{\mathcal{Y}}(\lambda(t),\lambda(t))\\ &- \tfrac{1}{2}T^*(\lambda(t), h^\sharp_{\mathcal{Y}}(\lambda(t))) + \tfrac{1}{2}T^*(\theta(t),\gamma'(t)),\\ \nabla_{\gamma'(t)}\lambda(t) = {}& -\theta(t) + \tfrac{1}{2}T^*(\lambda(t),\gamma'(t)).\end{aligned} \tag{S4.13}$$

Covariantly differentiating the second equation along γ gives

$$\begin{aligned}\nabla^2_{\gamma'(t)}\lambda(t) = -\nabla_{\gamma'(t)}\theta(t) + \tfrac{1}{2}(\nabla_{\gamma'(t)}T^*)(\lambda(t),\gamma'(t))\\ + \tfrac{1}{2}T^*(\nabla_{\gamma'(t)}\lambda(t),\gamma'(t)) + \tfrac{1}{2}T^*(\lambda(t),h^\sharp_{\mathcal{Y}}(\lambda(t))).\end{aligned}$$

Substituting the first of equations (S4.13) gives the second of equations (S4.12), which thus completes the proof of this part of the lemma.

(ii) We first note that λ can be maximizing only if $\theta(t)\oplus\lambda(t)\in A_{\mathrm{sing}}(\Sigma)$ for all $t\in[a,b]$. This means that $\lambda(t)$ must annihilate $\mathcal{Y}_{\gamma(t)}$. Since $\lambda_0=0$, our result follows from Theorem S4.26 and the definition of $B_{\mathcal{Y}}$. ■

Remarks S4.36. 1. The theorem implies that all normal controlled extremals for the force minimization problem are of class C^∞.

2. If one happens to choose $v_{q_0}, v_{q_1}\in \mathsf{TQ}$ with the property that there is a geodesic $\gamma\colon[0,T]\to\mathsf{Q}$ satisfying $\gamma'(0)=v_{q_0}$ and $\gamma'(T)=v_{q_1}$, then the optimal control for the problems $\mathscr{F}(\Sigma,v_{q_0},v_{q_1})$ and $\mathscr{F}_{[a,a+T]}(\Sigma,v_{q_0},v_{q_1})$, $a\in\mathbb{R}$, is the zero control.
3. The matter of investigating the existence of abnormal controlled extremals that are also minimizers would appear likely to take on a flavor similar to that of the sub-Riemannian case [see Liu and Sussmann 1994, Montgomery 1994].
4. If $(\gamma,u)\in\mathrm{Carc}(\Sigma)$, then

$$\begin{aligned}\nabla_{\gamma'(t)}\gamma'(t) = u^a(t)Y_a(\gamma(t)) \quad \Longrightarrow\\ \mathbb{G}(u^a(t)Y_a(\gamma(t)),u^b(t)Y_b(\gamma(t))) = \mathbb{G}(\nabla_{\gamma'(t)}\gamma'(t),\nabla_{\gamma'(t)}\gamma'(t)).\end{aligned}$$

Thus the force minimization problem may be seen as minimizing

$$\int_a^b \mathbb{G}(\nabla_{\gamma'(t)}\gamma'(t),\nabla_{\gamma'(t)}\gamma'(t))\,\mathrm{d}t$$

over all LAD curves $\gamma\colon[a,b]\to\mathsf{Q}$ subject to certain boundary conditions (fixed or free velocity), and subject to the constraint that $\nabla_{\gamma'(t)}\gamma'(t)\in\mathcal{Y}_{\gamma(t)}$ a.e. This may be thought of as a higher-order version of the sub-Riemannian geodesic problem. Indeed, note that the equations (S4.12) for the controlled extremals involve only the restriction of $\mathbb{G}$ to the distribution $\mathcal{Y}$. In the fully actuated case (see next section) we have a classical calculus of variations problem with a Lagrangian depending on first and second time-derivatives. This is the approach taken in [Crouch and Silva Leite 1991, Noakes, Heinzinger, and Paden 1989, Silva Leite, Camarinha, and Crouch 2000], for example. •

S4.5.3 The fully actuated case

As mentioned in the introduction, Crouch and Silva Leite [1991] and Noakes, Heinzinger, and Paden [1989] consider the force minimization problem with the Levi-Civita connection and with full actuation. Let us now consider the *general* case with full actuation. Thus in this section we let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be a fully actuated affine connection control system where $U = \mathbb{R}^m$.

Let us first show that all controlled extremals for the fully actuated force minimization problem are normal.

Proposition S4.37 (Normality of controlled extremals for fully actuated force minimization problem). *Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be a fully actuated affine connection control system with (γ, u) a controlled extremal for one of the four problems of Definition S4.32. The corresponding constant Lagrange multiplier λ_0 is nonzero.*

Proof. This follows from the Hamiltonian maximization condition. Since the Hamiltonian is

$$H_{\Sigma,\lambda_0 F_{\mathrm{force}}}(\alpha_{v_q} \oplus \beta_{v_q}, u) = \alpha_{v_q} \cdot v_q + u^a(\beta_{v_q} \cdot Y_a(q)) - \frac{\lambda_0}{2}\mathbb{G}(u^a Y_a(q), u^b Y_b(q)),$$

the only way for the Hamiltonian to be maximum with $\lambda_0 = 0$ is for β_{v_q} to be zero. This cannot happen since Theorem S4.26 asserts that both λ_0 and λ cannot be zero along an extremal. ■

We may now concentrate on the normal case of Theorem S4.35. The simplification here arises since $h_{\mathcal{Y}}$ becomes the vector bundle metric $\mathbb{G}^{-1}$ on $\mathsf{T}^*\mathsf{Q}$ induced by $\mathbb{G}$. From Theorem S4.35, if (γ, u) is a solution of $\mathscr{F}(\Sigma, v_{q_0}, v_{q_1})$, then we have

$$\begin{aligned}
&\nabla_{\gamma'(t)}\gamma'(t) = \mathbb{G}^\sharp(\lambda(t)),\\
&\nabla^2_{\gamma'(t)}\lambda(t) + R^*(\lambda(t), \gamma'(t))\gamma'(t) - T^*(\nabla_{\gamma'(t)}\lambda(t), \gamma'(t)) = T^*(\lambda(t), \mathbb{G}^\sharp(\lambda(t)))\\
&\qquad\qquad - \tfrac{1}{2}\nabla\mathbb{G}^{-1}(\lambda(t), \lambda(t)).
\end{aligned} \tag{S4.14}$$

We immediately see that the adjoint covector field λ is determined algebraically from the covariant derivative of γ along itself. This allows us to eliminate the adjoint covector field from the equations (S4.14) as the following result asserts.

Proposition S4.38 (Maximum Principle for fully actuated force minimization problem). *Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be a fully actuated affine connection control system with $U = \mathbb{R}^m$, let (γ, u) be a controlled extremal for one of the four problems of Definition S4.32, with u and γ defined on $[a, b]$, and let $\lambda\colon [a, b] \to \mathsf{T}^*\mathsf{Q}$ be the corresponding adjoint covector field. Then $\lambda(t) = \mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t))$ for $t \in [a, b]$, and γ satisfies the equation*

$$\begin{aligned}
&\nabla^3_{\gamma'(t)}\gamma'(t) + \mathbb{G}^\sharp(R^*(\mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t)),\gamma'(t))\gamma'(t)) \\
&\quad - \mathbb{G}^\sharp(T^*((\nabla_{\gamma'(t)}\mathbb{G}^\flat)(\nabla_{\gamma'(t)}\gamma'(t)),\gamma'(t))) - \mathbb{G}^\sharp(T^*(\mathbb{G}^\flat(\nabla^2_{\gamma'(t)}\gamma'(t)),\gamma'(t))) \\
&\qquad - \mathbb{G}^\sharp(T^*(\mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t)),\nabla_{\gamma'(t)}\gamma'(t))) \\
&\qquad + \tfrac{1}{2}\mathbb{G}^\sharp(\nabla\mathbb{G}^{-1}(\mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t)),\mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t)))) \\
&\quad - (\nabla^2_{\gamma'(t)}\mathbb{G}^\sharp)(\mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t))) - 2(\nabla_{\gamma'(t)}\mathbb{G}^\sharp)((\nabla_{\gamma'(t)}\mathbb{G}^\flat)(\nabla_{\gamma'(t)}\gamma'(t))) \\
&\qquad - 2(\nabla_{\gamma'(t)}\mathbb{G}^\sharp)(\mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t))) = 0.
\end{aligned}$$

If (γ,u) is a solution of either $\mathscr{F}(\Sigma,q_0,q_1)$ or $\mathscr{F}_{[a,b]}(\Sigma,q_0,q_1)$, then we additionally must have $\nabla_{\gamma'(a)}\gamma'(a)=0$ and $\nabla_{\gamma'(b)}\gamma'(b)=0$.

Proof. If we covariantly differentiate the first of equations (S4.14) along γ, then we get

$$\nabla^2_{\gamma'(t)}\gamma'(t) = (\nabla_{\gamma'(t)}\mathbb{G}^\sharp)(\lambda(t)) + \mathbb{G}^\sharp(\nabla_{\gamma'(t)}\lambda(t)),$$

and differentiating the same way again gives

$$\nabla^3_{\gamma'(t)}\gamma'(t) = (\nabla^2_{\gamma'(t)}\mathbb{G}^\sharp)(\lambda(t)) + 2(\nabla_{\gamma'(t)}\mathbb{G}^\sharp)(\nabla_{\gamma'(t)}\lambda(t)) + \mathbb{G}^\sharp(\nabla^2_{\gamma'(t)}\lambda(t)). \tag{S4.15}$$

Differentiating $\lambda(t)=\mathbb{G}^\flat(\nabla_{\gamma'(t)}\gamma'(t))$ gives

$$\nabla_{\gamma'(t)}\lambda(t) = (\nabla_{\gamma'(t)}\mathbb{G}^\flat)(\nabla_{\gamma'(t)}\gamma'(t)) + \mathbb{G}^\flat(\nabla^2_{\gamma'(t)}\gamma'(t)). \tag{S4.16}$$

Combining (S4.15), (S4.16), and the second of equations (S4.14), the result follows from a tedious computation. ■

S4.5.4 The Levi-Civita affine connection

Now we specialize the constructions of the previous sections to the case when $\nabla = \overset{\mathbb{G}}{\nabla}$, the Levi-Civita connection determined by the Riemannian metric $\mathbb{G}$ used in the definition of the cost function. In this case, matters simplify somewhat since $\overset{\mathbb{G}}{\nabla}\mathbb{G}=0$ and since $\overset{\mathbb{G}}{\nabla}$ is torsion-free. Let us state Theorem S4.35 for Levi-Civita connections.

Proposition S4.39 (Maximum Principle for force minimization problem for Levi-Civita affine connection). *Let $\mathbb{G}$ be a Riemannian metric on Q, and consider the affine connection control system $\Sigma = (\mathsf{Q},\overset{\mathbb{G}}{\nabla},\mathscr{Y},U)$ with $U=\mathbb{R}^m$ and with cost function F_{force} defined using $\mathbb{G}$. Suppose that (γ,u) is a controlled extremal for $\mathscr{F}_{[a,b]}(\Sigma,v_{q_0},v_{q_1})$ or for $\mathscr{F}(\Sigma,v_{q_0},v_{q_1})$ with u and γ defined on $[a,b]$, and with λ the adjoint covector field and λ_0 the Lagrange multiplier. Let $w=\mathbb{G}^\sharp\circ\lambda$. We have the following two situations.*

(i) $\lambda_0=1$: In this case, it is necessary and sufficient that γ and w together satisfy the differential equations

$$\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t) = P_{\mathcal{Y}}(w(t)),$$
$$\overset{\mathbb{G}}{\nabla}{}^2_{\gamma'(t)}w(t) + R(w(t),\gamma'(t))\gamma'(t) = -\tfrac{1}{2}\mathbb{G}^\sharp(\overset{\mathbb{G}}{\nabla}\mathbb{G}_{\mathcal{Y}}(w(t),w(t))).$$

(ii) $\lambda_0 = 0$*: In this case, it is necessary and sufficient that*
(a) $\nabla_{\gamma'(t)}\gamma'(t) = u^a(t)Y_a(\gamma(t))$,
(b) $w(t) \in \mathcal{Y}^\perp_{\gamma(t)})$ *for* $t \in [a,b]$*, and*
(c) w *satisfies the equation along* γ *given by:*

$$\nabla^2_{\gamma'(t)}w(t) + R(w(t),\gamma'(t))\gamma'(t) = B_{\mathcal{Y}}(\mathbb{G}^\flat \circ w(t), u^a(t)Y_a(t)).$$

If (γ, u) *is a solution of* $\mathscr{F}_{[a,b]}(\Sigma, q_0, q_1)$ *or of* $\mathscr{F}(\Sigma, q_0, q_1)$*, then we additionally have* $w(a) = 0$ *and* $w(b) = 0$*.*

Proof. (i) Since $\overset{\mathbb{G}}{\nabla}\mathbb{G} = 0$, we have

$$\overset{\mathbb{G}}{\nabla}{}^2_{\gamma'(t)}w(t) = \mathbb{G}^\sharp(\overset{\mathbb{G}}{\nabla}{}^2_{\gamma'(t)}\lambda(t)). \tag{S4.17}$$

A straightforward application of the definitions shows that

$$P_{\mathcal{Y}}(w(t)) = \mathbb{G}^\sharp(h^\sharp_{\mathcal{Y}}(\lambda(t))).$$

Thus the first equation in part (i) holds. By equation (S1.22) from the proof of Proposition S1.40, we have

$$R(w(t),\gamma'(t))\gamma'(t) = \mathbb{G}^\sharp(R^*(\lambda(t),\gamma'(t))\gamma'(t)). \tag{S4.18}$$

Now let $\beta \in \Gamma^\infty(\mathsf{T}^*\mathsf{Q})$ and $X = \mathbb{G}^\sharp(\beta) \in \Gamma^\infty(\mathsf{TQ})$. From the definition of $h_{\mathcal{Y}}$ we have

$$h_{\mathcal{Y}}(\beta(q),\beta(q)) = \mathbb{G}_{\mathcal{Y}}(X(q),X(q)).$$

Therefore, for $w \in \mathsf{T}_q\mathsf{Q}$, we have

$$\begin{aligned}\overset{\mathbb{G}}{\nabla}_w h_{\mathcal{Y}}(\beta(q),\beta(q)) &+ 2h_{\mathcal{Y}}(\overset{\mathbb{G}}{\nabla}_w\beta(q),\beta(q)) \\ &= \overset{\mathbb{G}}{\nabla}_w\mathbb{G}_{\mathcal{Y}}(X(q),X(q)) + 2\mathbb{G}_{\mathcal{Y}}(\overset{\mathbb{G}}{\nabla}_w X(q),X(q)).\end{aligned}$$

Since $\overset{\mathbb{G}}{\nabla}\mathbb{G} = 0$, we have $\nabla_w\beta(q) = \mathbb{G}^\sharp(\nabla_w X(q))$, from which we ascertain that

$$\overset{\mathbb{G}}{\nabla}_w h_{\mathcal{Y}}(\lambda(t),\lambda(t)) = \overset{\mathbb{G}}{\nabla}_w\mathbb{G}_{\mathcal{Y}}(w(t),w(t)). \tag{S4.19}$$

Bringing together equations (S4.17), (S4.18), and (S4.19), and the definition of $h_{\mathcal{Y}}$, gives the result by virtue of Theorem S4.35.

(ii) This is just a restatement of part (ii) of Theorem S4.35. ■

Now let us specialize to the fully actuated case. One applies Proposition S4.38 to show that the fully actuated extremals satisfy

$$\overset{\mathbb{G}}{\nabla}{}^3_{\gamma'(t)}\gamma'(t) + \mathbb{G}^\sharp\big(R^*(\mathbb{G}^\flat(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t)), \gamma'(t))\gamma'(t)\big) = 0.$$

Now we recall the equation (S1.22) from the proof of Proposition S1.40 to prove the following result that agrees with Crouch and Silva Leite [1991] and Noakes, Heinzinger, and Paden [1989].

Proposition S4.40 (Maximum Principle for fully actuated force minimization problem with Levi-Civita affine connection). *Let $\mathbb{G}$ be a Riemannian metric on Q and consider the fully actuated affine connection control system $\Sigma = (\mathsf{Q}, \overset{\mathbb{G}}{\nabla}, \mathscr{Y}, U)$ with cost function F_{force} defined using $\mathbb{G}$. A controlled extremal (γ, u) for one of the four problems of Definition S4.32, with u and γ defined on $[a, b]$, satisfies the differential equation*

$$\overset{\mathbb{G}}{\nabla}{}^3_{\gamma'(t)}\gamma'(t) + R(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t), \gamma'(t))\gamma'(t) = 0.$$

If (γ, u) is either a solution of $\mathscr{F}(\Sigma, q_0, q_1)$ or $\mathscr{F}_{[a,b]}(\Sigma, q_0, q_1)$, then we additionally have $\overset{\mathbb{G}}{\nabla}_{\gamma'(a)}\gamma'(a) = 0$ and $\overset{\mathbb{G}}{\nabla}_{\gamma'(b)}\gamma'(b) = 0$.

S4.5.5 Singular and abnormal extremals

In this section we state a result that brings together some of the statements made in the preceding section, and gives a clear picture of the nature of the singular and abnormal extremals for the force minimization problem.

Proposition S4.41 (Singular and abnormal extremals for the force minimization problem). *Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, \mathbb{R}^m)$ be a C^∞-affine connection control system, and let (γ, u) be a controlled extremal for one of the four problems of Definition S4.32, defined on $[a, b]$. Then the following statements are equivalent:*

(i) (γ, u) is abnormal;
(ii) (γ, u) is singular;

Either of the preceding two conditions implies the following:

(iii) any adjoint covector field λ along γ satisfies $\lambda(t) \in \mathrm{ann}(\mathcal{Y}_{\gamma(t)})$ for each $t \in [a, b]$.

Proof. Fix a state $\alpha_{v_q} \oplus \beta_{v_q}$ in $\mathsf{T}^*\mathsf{TQ}$, using the splitting of $\mathsf{T}^*\mathsf{TQ}$ described in Section S1.3.9. The Hamiltonian for the force minimization problem, with this state fixed and with u varying, can be thought of as a function on $\mathcal{Y}_q$ by

$$w \mapsto \alpha_{v_q} \cdot v_q + \beta_{v_q} \cdot w - \tfrac{1}{2}\lambda_0 \mathbb{G}(w, w).$$

The definition of the maximum Hamiltonian is achieved by maximizing this function of w.

(i) $\implies$ (ii) If (γ, u) is abnormal, then, for each $t \in [a,b]$, $u(t)$ satisfies

$$u^a(t)(\lambda(t) \cdot Y_a(\gamma(t))) = \sup \left\{ \lambda(t) \cdot w \mid \ w \in \mathcal{Y}_{\gamma(t)} \right\},$$

where λ is an adjoint covector field for (γ, u). Since the function $w \mapsto \lambda(t) \cdot w$ is linear, the value of the maximum Hamiltonian can be realized if and only if this linear function is zero, i.e., if and only if $\lambda(t) \in \operatorname{ann}(\mathcal{Y}_{\gamma(t)})$ for all $t \in [a,b]$. However, if this is the case, then it is clear that, for each $t \in [a,b]$, $H_{\Sigma,0}(\theta(t) \oplus \lambda(t), u) = H^{\max}_{\Sigma,0}(\theta(t) \oplus \lambda(t))$ is satisfied for each $u \in \mathbb{R}^m$. Thus (γ, u) is singular.

(ii) $\implies$ (i) If (γ, u) is singular, then, for each $t \in [a,b]$, the function

$$\mathcal{Y}_{\gamma(t)} \ni w \mapsto \theta(t) \cdot \gamma'(t) + \lambda(t) \cdot w - \tfrac{1}{2}\lambda_0 \mathbb{G}(w,w)$$

is always equal to its maximum. This immediately implies that $\lambda_0 = 0$, and so (γ, u) is abnormal.

(i) $\implies$ (iii) This was shown above while proving the implication (i) $\implies$ (ii). ■

In particular, it follows that any nontrivial (in the sense that the input force is nonzero) singular extremal must be abnormal, and so must satisfy condition (iii) of the proposition.

S4.6 Time-optimal control for affine connection control systems

Next we look at time-optimal control in the setting of affine connection control systems. Time-optimal control is one of the most basic of problems in optimal control theory. One reason for this is the intimate connection between controllability and time-optimal control as asserted in Proposition S4.20. Thus, the studying of the time-optimal control problem will reveal something about the system itself, as opposed to other sorts of optimal control problems, where the cost function contributes significantly to the character of solutions to the optimization problem.

S4.6.1 The time-optimal problem

We let $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be an affine connection control system. The cost function for time-optimal control is $F_{\mathrm{time}} = 1$. In the parlance of Section S4.4.1, we consider the pair $\mathscr{A} = (A, f)$, where A is the $\mathbb{R}^m$-dependent $(0,0)$-tensor field $A(q,u) = 1$ and $f = \mathrm{id}_{\mathbb{R}}$. Note then that minimizing

$$\int_0^T F_{\mathrm{time}}(\gamma'(t), u(t))\, \mathrm{d}t$$

means precisely minimizing the time along the controlled trajectory.

Definition S4.42 (Time-optimal control problems for affine connection control systems). Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be an affine connection control system, let $q_1, q_2 \in \mathsf{Q}$ and let $v_{q_1} \in \mathsf{T}_{q_1}\mathsf{Q}$ and $v_{q_2} \in \mathsf{T}_{q_2}\mathsf{Q}$.

(i) A controlled arc (γ, u) is a solution of $\mathscr{T}(\Sigma, v_{q_0}, v_{q_1})$ if it is a solution of $\mathscr{P}(\Sigma, F_{\mathrm{time}}, v_{q_0}, v_{q_1})$.

(ii) A controlled arc (γ, u) is a solution of $\mathscr{T}(\Sigma, q_0, q_1)$ if it is a solution of $\mathscr{P}(\Sigma, F_{\mathrm{time}}, q_0, q_1)$. •

It may be possible that solutions to the time-optimal problems do not exist. To see why this might be so, we consider the case when $U = \mathbb{R}^m$.

Proposition S4.43. *For an affine connection control system* $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$ *with* $U = \mathbb{R}^m$*, the problem* $\mathscr{T}(\Sigma, q_0, q_1)$ *has no solution.*

Proof. Let $(\gamma, u) \in \mathrm{Ctraj}(\Sigma)$ be defined on an interval I. For $\lambda > 0$, define $I_\lambda = \left\{ \frac{1}{\lambda}t \;\middle|\; t \in I \right\}$, and define $\tilde{u}\colon I_\lambda \to \mathbb{R}^m$ and $\tilde{\gamma}\colon I_\lambda \to \mathsf{Q}$ by $\tilde{u}(t) = \lambda^2 u(\lambda t)$ and $\tilde{\gamma}(t) = \gamma(\lambda t)$. Then $(\tilde{u}, \tilde{\gamma}) \in \mathrm{Carc}(\Sigma)$. Indeed, we directly compute

$$\nabla_{\tilde{\gamma}'(t)}\tilde{\gamma}'(t) = \lambda^2 \nabla_{\gamma'(\lambda t)}\gamma'(\lambda t) = \lambda^2 u(\lambda t) = \tilde{u}(t).$$

Therefore, if there exists a controlled arc connecting q_0 and q_1, then this controlled arc can be followed in arbitrarily small time, thus implying the lack of existence of a solution to the time minimization problem. ■

To ensure well-defined solutions to all time-optimal control problems, we must place bounds on the controls. For a control-affine system, control bounds are often specified by requiring that each control take values in a compact interval, typically symmetric about the origin. However, we shall use bounds that are elliptical. We do this for two reasons: (1) elliptical control bounds are more useful for fleshing out the geometry of the system since they do not rely on a specific choice of basis for the input vector fields, and (2) for the planar body system we study in Section S4.7, the elliptical control bounds are consistent with a thruster whose maximum output is independent of the direction in which it points. In any event, we introduce a Riemannian metric $\mathbb{G}$ on Q and ask that controls satisfy the bound

$$\mathbb{G}(u^a(t)Y_a(\gamma(t)), u^b(t)Y_b(\gamma(t))) \leq 1 \tag{S4.20}$$

along a controlled trajectory (γ, u). We again use the summation convention in the expression $u^a Y_a$. In order to ensure that this defines an affine connection control system as in our definition, the set of $u \in \mathbb{R}^m$ that satisfy the bound should not depend on the point along the controlled trajectory. This is possible, for example, if the vector fields in $\mathscr{Y}$ are $\mathbb{G}$-orthonormal. To simplify things, we shall make the assumption in the time-optimal problem that $\mathscr{Y}$ is a $\mathbb{G}$-orthonormal family of vector fields. In this case, the control set U is simply given by

$$U = \{ u \in \mathbb{R}^m \mid \|u\|_{\mathbb{R}^m} \leq 1 \}. \tag{S4.21}$$

Note that it cannot be expected to be able to choose a global basis of $\mathbb{G}$-orthonormal input vector fields. Obstructions can arise in two ways. First of all, the input distribution $\mathcal{Y}$ may not have constant rank. In this case, it will be impossible to choose an $\mathbb{G}$-orthonormal basis of input vector fields on any set containing singular points for $\mathcal{Y}$. Even if $\mathcal{Y}$ *does* have constant rank, it may well be the case that one cannot choose a global $\mathbb{G}$-orthonormal basis of vector fields for $\mathcal{Y}$. However, in this case, one can do this locally.

In summary, in this section, we make the following assumption.

Assumption S4.44. For the affine connection control system $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$ and the Riemannian metric $\mathbb{G}$ on Q, the vector fields $\mathscr{Y}$ are $\mathbb{G}$-orthogonal. •

S4.6.2 The Maximum Principle for time-optimal control

We have the following Maximum Principle for time-optimal control of affine connection control systems.

Theorem S4.45 (Maximum Principle for time-optimal control of affine connection control systems). *Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, U)$ be an affine connection control system satisfying Assumption S4.44. Suppose that $(\gamma, u) \in$ $\mathrm{Ctraj}(\Sigma)$ is a solution of $\mathscr{T}(\Sigma, v_{q_0}, v_{q_1})$ with u and γ defined on $[0, T]$. Then there exists an* LAD *covector field $\lambda\colon [0, T] \to \mathsf{T}^*\mathsf{Q}$ along γ and a constant $\lambda_0 \in \{0, 1\}$ with the following properties:*

(i) for almost every $t \in [0, T]$, we have

$$\nabla^2_{\gamma'(t)}\lambda(t) + R^*(\lambda(t), \gamma'(t))\gamma'(t) \\ - T^*(\nabla_{\gamma'(t)}\lambda(t), \gamma'(t)) = u^a(t)(\nabla Y_a)^*(\lambda(t));$$

(ii) either $\lambda_0 = 1$ or $\theta(0) \oplus \lambda(0) \neq 0$;

(iii) for almost every $t \in [0, T]$, we have

$$u^a(t)\langle\lambda(t); Y_a(\gamma(t))\rangle = \sup\left\{\, \tilde{u}^a\langle\lambda(t); Y_a(\gamma(t))\rangle \;\middle|\; \tilde{u} \in U \right\};$$

(iv) $\langle\theta(t); \gamma'(t)\rangle + u^a(t)\langle\lambda(t); Y_a(\gamma(t))\rangle = \lambda_0$,

with

$$\theta(t) = \tfrac{1}{2}T^*(\lambda(t), \gamma'(t)) - \nabla_{\gamma'(t)}\lambda(t), \qquad t \in [0, T].$$

If (γ, u) is a solution of $\mathscr{T}(\Sigma, q_0, q_1)$, then the conditions (i)–(iv) hold and, in addition, $\lambda(0) = 0$ and $\lambda(T) = 0$.

Proof. Follows directly from Theorem S4.26. ■

One can use part (iii) of Theorem S4.45 to determine the form of the control for time-optimal problems. This is quite simple to do as it involves minimizing a linear function of u subject to the constraint that u lie in a ball of unit radius. To express the result, we recall the notation $P_{\mathcal{Y}}$ and $h_{\mathcal{Y}}$ from

Section S4.5.2. One can readily show [see Coombs 2000] that the value of u that achieves the minimum satisfies

$$u^a Y_a(\gamma(t)) = -\frac{h^\sharp_{\mathcal{Y}}(\lambda(t))}{\left\|P^*_{\mathcal{Y}}(\lambda(t))\right\|_{\mathbb{G}}}, \tag{S4.22}$$

where $\|\cdot\|_{\mathbb{G}}$ denotes the norm with respect to the Riemannian metric $\mathbb{G}$. Note that (S4.22) gives a feedback control that gives the character of the controlled extremals by integrating the control equations with the equation from part (i) of Theorem S4.45.

Remark S4.46. It is true that (S4.22) defines the controls even when the set of input vector fields is not orthonormal with respect to the Riemannian metric $\mathbb{G}$. Indeed, for Theorem S4.45 to hold, it only needs to be *possible* to choose an orthonormal basis of input vector fields. One can then use (S4.22) to define the controls, even when the vector fields $\{Y_1, \dots, Y_m\}$ are not orthonormal. When it is not possible to choose an orthonormal basis for the input distribution (as, for example, with fully actuated systems on $\mathbb{S}^2$), then one loses condition (iv) in Theorem S4.45. •

S4.6.3 Singular extremals

In this section we state a simple result that classifies the singular extremals for the time-optimal problem for affine connection control systems.

Proposition S4.47 (Singular extremals for time-optimal control). *Let $\Sigma = (\mathsf{Q}, \nabla, \mathscr{Y}, \mathbb{R}^m)$ be a C^∞-affine connection control system satisfying Assumption S4.44, and let (γ, u) be a controlled extremal for one of the problems of Definition S4.42, defined on $[0, T]$. Then the following statements are equivalent:*

(i) (γ, u) is singular;

(ii) any adjoint covector field λ along γ satisfies $\lambda(t) \in \operatorname{ann}(\mathcal{Y}_{\gamma(t)})$ for each $t \in [0, T]$.

Proof. The Hamiltonian in this case is, using the notation as at the beginning of the proof of Proposition S4.41,

$$\mathcal{Y}_q \ni w \mapsto \alpha_{v_q} \cdot v_q + \beta_{v_q} \cdot w - \lambda_0.$$

Part (ii) is equivalent to the statement that this Hamiltonian be independent of w along trajectories in $\mathsf{T}^*\mathsf{TQ}$. This immediately implies, and is implied by, the assertion that, for each $t \in [0, T]$, $H_{\Sigma,\lambda_0 F}(\theta(t) \oplus \lambda(t), u) = H^{\max}_{\Sigma,\lambda_0 F}(\theta(t) \oplus \lambda(t))$ for all $u \in U$, where θ is defined as in Theorem S4.26. The result follows immediately. ■

For the planar rigid body example we look at in Section S4.7, it is possible to obtain a complete characterization of the singular controlled extremals.

S4.7 Force- and time-optimal control for a planar rigid body

We consider in this section the planar rigid body example considered in the text. The modeling for the system was carried out in Chapter 4, and we refer the reader there for details. The system is depicted in Figure S4.5. We use the

Figure S4.5. Coordinates and input forces for the planar rigid body

coordinates (x, y, θ) as indicated in the figure.

This is a left-invariant control system on a Lie group (see Example 5.47). We shall not take much advantage of this additional structure. However, we will occasionally make use of the following fact.

Lemma S4.48 ($\mathsf{SE}(2)$-invariance of controlled trajectories). *Let (γ, u) be a controlled trajectory for the planar rigid body through the point $q \in \mathsf{Q}$. If $\bar{q} = L_g(q)$ for $g \in \mathsf{SE}(2)$, and if $\bar{\gamma} = L_g \circ \gamma$, then $(\bar{\gamma}, u)$ is a controlled trajectory through the point $\bar{q}$.* •

S4.7.1 System data

We recall that the Riemannian metric for the system is

$$\mathbb{G} = m(\mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y) + J\mathrm{d}\theta \otimes \mathrm{d}\theta,$$

where m is the mass of the body and J is its moment of inertia about the center of mass, and that the input vector fields are

$$Y_1 = \frac{\cos\theta}{m}\frac{\partial}{\partial x} + \frac{\sin\theta}{m}\frac{\partial}{\partial y}, \qquad Y_2 = -\frac{\sin\theta}{m}\frac{\partial}{\partial x} + \frac{\cos\theta}{m}\frac{\partial}{\partial y} - \frac{h}{J}\frac{\partial}{\partial \theta}.$$

With this information, the equations of motion for the system are

$$\begin{aligned}\ddot{x} &= \frac{\cos\theta}{m}u^1 - \frac{\sin\theta}{m}u^2,\\ \ddot{y} &= \frac{\sin\theta}{m}u^1 + \frac{\cos\theta}{m}u^2,\\ \ddot{\theta} &= -\frac{h}{J}u^2.\end{aligned} \tag{S4.23}$$

One also computes

$$\begin{aligned}\overset{\mathbb{G}}{\nabla} Y_1 &= -\frac{\sin\theta}{m}\frac{\partial}{\partial x}\otimes \mathrm{d}\theta + \frac{\cos\theta}{m}\frac{\partial}{\partial y}\otimes \mathrm{d}\theta,\\ \overset{\mathbb{G}}{\nabla} Y_2 &= -\frac{\cos\theta}{m}\frac{\partial}{\partial x}\otimes \mathrm{d}\theta - \frac{\sin\theta}{m}\frac{\partial}{\partial y}\otimes \mathrm{d}\theta,\end{aligned}$$

We use the Riemannian metric $\mathbb{G}$ to define our time-optimal control bounds as in (S4.20), and to define our force-optimal cost function as in Section S4.5. It is possible, of course, to use other metrics, but in lieu of further information, we stick with the one given by the physics of the problem for the sake of naturality. We will need explicit representations for $P_{\mathcal{Y}}$ and $h_{\mathcal{Y}}$. Straightforward calculations give the matrix representation for $P_{\mathcal{Y}}$ as

$$\frac{1}{J+mh^2}\boldsymbol{R}(\theta)\begin{bmatrix} J+mh^2 & 0 & 0\\ 0 & J & -Jh\\ 0 & -mh & mh^2\end{bmatrix}\boldsymbol{R}^{-1}(\theta),$$

and the matrix representation for $h_{\mathcal{Y}}$ as

$$\frac{1}{J+mh^2}\boldsymbol{R}(\theta)\begin{bmatrix} \frac{J+mh^2}{m} & 0 & 0\\ 0 & \frac{J}{m} & -h\\ 0 & -h & \frac{mh^2}{J}\end{bmatrix}\boldsymbol{R}^{-1}(\theta),$$

where

$$\boldsymbol{R}(\theta) = \begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}.$$

For force-optimal control, we also need to know $\overset{\mathbb{G}}{\nabla} h_{\mathcal{Y}}$. In fact, we need only know the value of $\overset{\mathbb{G}}{\nabla} h_{\mathcal{Y}}$ when evaluated on a single covector in both arguments. Another straightforward, but tedious, computation gives

$$\overset{\mathbb{G}}{\nabla} h_{\mathcal{Y}}(\lambda,\lambda) = \frac{2h(\lambda_\theta + h\lambda_y\cos\theta - h\lambda_x\sin\theta)(\lambda_x\cos\theta + \lambda_y\sin\theta)}{J+mh^2}\,\mathrm{d}\theta.$$

S4.7.2 Nonsingular force-optimal control

For force-optimal control, we work with the normal case, that is also the nonsingular case. With Theorem S4.35 and the computations of Section S4.7.1, the equations governing the motion of the configurations and the adjoint covector field are readily computed to be

$$
\begin{aligned}
\ddot{x} &= -\frac{(2J+mh^2+mh^2\cos\theta)\lambda_x+2mh\sin\theta(\lambda_\theta+h\cos\theta\lambda_y)}{2m(J+mh^2)},\\
\ddot{y} &= \frac{2mh\cos\theta\lambda_\theta+(-2J-mh^2+mh^2\cos 2\theta)\lambda_y-mh^2\sin 2\theta\lambda_x}{2m(J+mh^2)},\\
\ddot{\theta} &= -\frac{h(mh\lambda_\theta-J\cos\theta\lambda_y+J\sin\theta\lambda_x)}{J(J+mh^2)},\\
\ddot{\lambda}_x &= 0,\\
\ddot{\lambda}_y &= 0,\\
\ddot{\lambda}_\theta &= = \frac{h(\lambda_\theta+h\lambda_y\cos\theta-h\lambda_x\sin\theta)(\lambda_x\cos\theta+\lambda_y\sin\theta)}{J+mh^2}.
\end{aligned}
\tag{S4.24}
$$

It is not our intention to give a complete analysis of the equations for controlled extremals. However, we can make some rather prosaic remarks about some of the more simple extremals. Perhaps the simplest extremals are those for which we undergo linear motion. To consider such motions, since the system is rotationally invariant, it suffices to consider linear motion in the $\boldsymbol{s}_1$-direction. Similarly, we may as well start at the initial configuration $(0,0,0)$. (Here we are using Lemma S4.48.) If we choose the other initial conditions such that $v_y(0)$, $v_\theta(0)$, $\lambda_y(0)$, $\lambda_\theta(0)$, $\dot{\lambda}_y(0)$, and $\dot{\lambda}_\theta(0)$ are all zero, then one can readily see from the controlled extremal equations (S4.24) that these quantities remain zero. The resulting motion is then along the line through $(0,0,0)$ in the $\boldsymbol{s}_1$-direction. One then verifies that the relevant equations governing these extremals are

$$
\ddot{x} = -\frac{\lambda_x}{m}, \qquad \ddot{\lambda}_x = 0. \tag{S4.25}
$$

These are readily solved to yield

$$
x(t) = -\frac{\dot{\lambda}_x(0)}{6m}t^3 - \frac{\lambda_x(0)}{2m}t^2 + \dot{x}(0)t + x(0).
$$

To join a state $(x(0),\dot{x}(0))$ with a state $(x(T),\dot{x}(T))$, one can readily design the initial conditions for λ_x to do the job. In Figure S4.6 we plot force-optimal controlled extremals for two boundary conditions. Note that, with force-optimal control, we may vary the final time, and we have chosen as final times the same times that will arise in the time-optimal analysis of Section S4.7.3, so that more useful comparisons may be made. Note that, in contrast with the time-optimal problem, for the force-optimal problem we must choose a time interval. If we do not, then for the situations depicted on the left in Figure S4.6, by stretching the time interval, we may make the value of the force-optimal objective function as low as we like. This would render the force-optimal problem one without a solution.

Figure S4.6. Two linear force-optimal controlled extremals with $m = 1$. (1) On the left we take $\dot{x}(0) = 0$, $x(T) = 1$, and $\dot{x}(T) = 0$. (2) On the right we take $\dot{x}(0) = 5$, $x(T) = 1$, and $\dot{x}(T) = -1$. The value of the objective function on the left is $A_{\Sigma,F_{\text{force}}}(\gamma, u) = \frac{3}{2}$ (compared with $A_{\Sigma,F_{\text{force}}}(\gamma, u) = 2$ for the corresponding time-optimal controlled extremal from Figure S4.7) and the value of the objective function on the right is $A_{\Sigma,F_{\text{force}}}(\gamma, u) = \frac{3(1604791+926528\sqrt{3})}{32(19+11\sqrt{3})} \approx 7.29$ (compared with $A_{\Sigma,F_{\text{force}}}(\gamma, u) = 4(\sqrt{3}+1) \approx 10.93$ for the corresponding time-optimal controlled extremal from Figure S4.7). In each plot, the top plot is $x(t)$ and the bottom is $u^1(t)$.

S4.7.3 Nonsingular time-optimal control

It is a simple matter to write down the equations governing the time-optimal controlled extremals, at least in the case when the controls may be determined from condition (ii) of Theorem S4.45. Indeed, one may use (S4.22) to derive

$$\begin{aligned} u^1 &= -\frac{\lambda_x \cos\theta + \lambda_y \sin\theta}{\left\|P_{\mathfrak{Y}}^*(\lambda)\right\|_{\mathbb{G}}}, \\ u^2 &= \frac{mh\lambda_\theta + J\lambda_x \sin\theta - J\lambda_y \cos\theta}{(J + mh^2)\left\|P_{\mathfrak{Y}}^*(\lambda)\right\|_{\mathbb{G}}}. \end{aligned} \tag{S4.26}$$

The expression for $\|P_{\mathfrak{Y}}^*(\lambda)\|_{\mathbb{G}}$ is a lengthy one, and we shall not give it here explicitly.

We may also express the equation of part (i) of Theorem S4.45 as

$$\begin{aligned}
\ddot{\lambda}_x &= 0,\\
\ddot{\lambda}_y &= 0,\\
\ddot{\lambda}_\theta &= -\frac{\sin\theta}{m}(\lambda_x u^1 + \lambda_y u^2) + \frac{\cos\theta}{m}(\lambda_y u^1 - \lambda_x u^2).
\end{aligned} \tag{S4.27}$$

Given the controls (S4.26), the nonsingular time-optimal controlled extremals satisfy (S4.27), along with the equations of motion (S4.23).

As with the force-optimal problem, we will not undergo a systematic investigation of these equations, but will merely look at a special case. We will again restrict ourselves to the situation where we have motion in the $\boldsymbol{s}_1$-direction through the initial point $(0,0,0)$. As with force-optimal control, if we choose the other initial conditions such that $v_y(0)$, $v_\theta(0)$, $\lambda_y(0)$, $\lambda_\theta(0)$, $\dot{\lambda}_y(0)$, and $\dot{\lambda}_\theta(0)$ are all zero, then these remain zero along controlled extremals. The resulting motion is then along the line through $(0,0,0)$ in the $\boldsymbol{s}_1$-direction, and the equations governing the resulting controlled extremals are

$$\ddot{x} = \frac{u^1}{m}, \qquad \ddot{\lambda}_x = 0. \tag{S4.28}$$

The control satisfies the bounds $u^1 \in [-\sqrt{m}, \sqrt{m}]$. One determines [see, for example, Jurdjevic 1997] that the time-optimal control that takes one from $(0,0,0)$ at rest to $(x_1,0,0)$ at rest is given by

$$u^1(t) = \begin{cases} \sqrt{m}, & t \in [0, T_s],\\ -\sqrt{m}, & t \in [T_s, 2T_s] \end{cases} \tag{S4.29}$$

where $T_s = \sqrt{\sqrt{m}x_1}$. Without loss of generality, we have supposed that $x_1 > 0$. It is possible to explicitly derive the time-optimal controls for nonzero initial and terminal velocity, but this is an inappropriate degree of generality for what we wish to accomplish here. In Figure S4.7 we represent two such linear extremals corresponding to various initial conditions, including one case where the initial and terminal velocity are nonzero. These motions are quite intuitive, and are essentially what one might affect by *ad hoc* methods. Observe that we see why bounds on control are necessary. For the situation on the left in Figure S4.7, if we had no control bounds, then by increasing the value of the control, we could execute the same maneuver in arbitrarily small time, making the time-optimal problem ill-defined.

S4.7.4 Characterization of the singular controlled extremals

Now we turn to looking at those controlled extremals that are singular for the planar rigid body system. Based on the characterizations of Propositions S4.41 and S4.47 of singular controlled extremals, we pose the following general problem as one that allows description of all nontrivial (in the sense that the input force is nonzero) force-optimal singular controlled extremals, and of all time-optimal singular extremals.

Figure S4.7. Two linear time-optimal controlled extremals with $m = 1$. (1) On the left we take $\dot{x}(0) = 0$, $x(T) = 1$, and $\dot{x}(T) = 0$. (2) On the right we take $\dot{x}(0) = 5$, $x(T) = 1$, and $\dot{x}(T) = -1$. The optimal time on the left is $T = 2$ and on the right is $T = 4(\sqrt{3} + 1)$. In each plot, the top plot is $x(t)$ and the bottom is $u^1(t)$.

Problem S4.49. For an affine connection control system $(\mathsf{Q}, \nabla, \mathscr{Y}, U)$, find curves $\gamma\colon I \to \mathsf{Q}$, $u\colon I \to \mathbb{R}^m$, $\lambda\colon I \to \mathsf{T}^*\mathsf{Q}$ with the following properties:

(i) u is locally integrable;

(ii) λ is a covector field along γ that is not identically zero;

(iii) the equations

$$\begin{aligned}
&\nabla_{\gamma'(t)}\gamma'(t) = u^a(t)Y_a(\gamma(t)),\\
&\nabla^2_{\gamma'(t)}\lambda(t) + R^*(\lambda(t), \gamma'(t))\gamma'(t)\\
&\qquad\qquad - T^*(\nabla_{\gamma'(t)}\lambda(t), \gamma'(t)) = u^a(t)(\nabla Y_a)^*(\lambda(t))
\end{aligned}$$

hold for a.e. $t \in I$;

(iv) $\lambda(t) \in \operatorname{ann}(\mathcal{Y}_{\gamma(t)})$ for a.e. $t \in I$. •

Nontrivial force-optimal singular controlled extremals, and time-optimal singular extremals, will be solutions to Problem S4.49 defined on appropriate intervals (i.e., intervals of the form $[a, b]$ for the force-optimal problem, and intervals of the form $[0, T]$ for the time-optimal problem).

In the analysis for the planar body, we shall encounter the quantity

$$(\dot{\lambda}_x(0)t + \lambda_x(0))^2 + (\dot{\lambda}_y(0)t + \lambda_y(0))^2, \tag{S4.30}$$

and we shall wish for this quantity to be nonzero for all $t \in \mathbb{R}$. We shall see that when this quantity *is* zero for some t, the singular controlled extremals reduce to a degenerate form. For now, let us make a statement equivalent to the expression (S4.30) being nonzero.

Lemma S4.50. *The expression* (S4.30) *is nonzero for all* $t \in \mathbb{R}$ *if and only if* $\dot{\lambda}_x(0)\lambda_y(0) \neq \dot{\lambda}_y(0)\lambda_x(0)$.

Proof. The expression (S4.30) is nonzero for all $t \in \mathbb{R}$ when and only when the equation

$$(\dot{\lambda}_x(0)t + \lambda_x(0))^2 + (\dot{\lambda}_y(0)t + \lambda_y(0))^2 = 0$$

has no real roots in t. This is a quadratic equation in t with discriminant $-(\dot{\lambda}_x(0)\lambda_y(0) - \dot{\lambda}_y\lambda_x(0))^2$. Therefore, it can have real roots when and only when $\dot{\lambda}_x(0)\lambda_y(0) - \dot{\lambda}_y\lambda_x(0) = 0$. ■

Let us now derive the equations governing solutions of Problem S4.49 that are defined on all of the real line. The condition that λ be in $\operatorname{ann}(\mathcal{Y})$ is given by

$$\lambda_x + \frac{mh}{J}\sin\theta\lambda_\theta = 0, \qquad \lambda_y - \frac{mh}{J}\cos\theta\lambda_\theta = 0. \tag{S4.31}$$

This means that λ must satisfy the condition

$$\lambda_\theta^2 = \Big(\frac{J}{mh}\Big)^2(\lambda_x^2 + \lambda_y^2).$$

Since λ_x and λ_y are determined from the controlled extremal equations (S4.24) (or equivalently from (S4.27)) to be simply

$$\lambda_x(t) = \dot{\lambda}_x(0)t + \lambda_x(0), \qquad \lambda_y(t) = \dot{\lambda}_y(0)t + \lambda_y(0),$$

this means that we are able to determine the adjoint covector field explicitly as a rational function of t. This in turn allows us to determine $\theta(t)$ from the equations (S4.31). From the equations of motion (S4.23), we then have $u^2(t) = -\frac{J}{h}\ddot{\theta}(t)$. To solve for $u^1(t)$, we employ the third of the equations (S4.27), as the remaining quantities are known as functions of t. Finally, to solve for $x(t)$ and $y(t)$, we go to the equations of motion (S4.23). The most relevant product of these computations for us is the resulting form of $(x(t), y(t), \theta(t))$, and these may be determined to be

$$\begin{aligned}
x(t) &= -\frac{J(\dot{\lambda}_y(0)t + \lambda_y(0))}{mh\sqrt{(\dot{\lambda}_y(0)t + \lambda_y(0))^2 + (\dot{\lambda}_x(0)t + \lambda_x(0))^2}} + C_{11}t + C_{10},\\
y(t) &= \frac{J(\dot{\lambda}_x(0)t + \lambda_x(0))}{mh\sqrt{(\dot{\lambda}_y(0)t + \lambda_y(0))^2 + (\dot{\lambda}_x(0)t + \lambda_x(0))^2}} + C_{21}t + C_{20},\\
\theta(t) &= \operatorname{atan}(\dot{\lambda}_y(0)t + \lambda_y(0)), -\dot{\lambda}_x(0)t - \lambda_x(0).
\end{aligned} \tag{S4.32}$$

It is possible to determine C_{11}, C_{10}, C_{21}, and C_{20} using the initial conditions $x(0)$, $\dot{x}(0)$, $y(0)$, and $\dot{y}(0)$. However, as we shall see, it is not advantageous to do so, so we will leave these constants as they are. A solution of Problem S4.49 is called ***stationary*** when it is of the form (S4.32) with $C_{11} = C_{21} = 0$. We may also explicitly represent the controls as

$$\begin{aligned} u^1(t) = &\big(J(\lambda_y(0)\dot{\lambda}_x(0) - \lambda_x(0)\dot{\lambda}_y(0))^2\big) \\ &/\big(h\big((\dot{\lambda}_y(0)t + \lambda_y(0))^2 + (\dot{\lambda}_x(0)t + \lambda_x(0))^2\big)^2\big), \\ u^2(t) = &-\big(2J(\lambda_y(0)\dot{\lambda}_x(0) - \lambda_x(0)\dot{\lambda}_y(0))\big((\dot{\lambda}_x^2(0) + \dot{\lambda}_y^2(0))t \\ &+ \lambda_x(0)\dot{\lambda}_x(0) + \lambda_y(0)\dot{\lambda}_y(0))\big)/\big(h\big((\dot{\lambda}_y(0)t + \lambda_y(0))^2 \\ &+ (\dot{\lambda}_x(0)t + \lambda_x(0))^2\big)^2\big). \end{aligned} \tag{S4.33}$$

In writing these equations, we make the assumption that the expression (S4.30) is nonzero for all $t \in \mathbb{R}$. Note that these same expressions hold, even when the singular controlled extremal does not include 0 in its domain—in this case, the constants in the expressions for $\lambda_x(t)$ and $\lambda_y(t)$ simply lose their interpretation as being values of the functions and their first derivatives at $t = 0$.

The above paragraph shows that it is possible to completely determine, as explicit functions of time, the solutions to Problem S4.49, at least when the expression (S4.30) is nowhere zero. Let us now investigate the case when the expression (S4.30) *can* vanish for some $t \in \mathbb{R}$, recalling our characterization of this situation in Lemma S4.50.

Lemma S4.51. *Let (γ, u) be a stationary solution to Problem S4.49 defined on $\mathbb{R}$. The following conditions are equivalent:*

(i) $\gamma'(t_0) = 0_{\gamma(t_0)}$ *for some* $t_0 \in \mathbb{R}$*;*
(ii) $\gamma'(t) = 0_{\gamma(t)}$ *for all* $t \in \mathbb{R}$*;*
(iii) $\dot{\lambda}_x(0)\lambda_y(0) - \dot{\lambda}_y(0)\lambda_x(0) = 0$*.*

Proof. Suppose that (iii) holds, and let $\bar{t}$ be the time at which the expression (S4.30) vanishes. First suppose we are at a point where $t \neq \bar{t}$. Then the equations (S4.32) are valid in a neighborhood of t, and one computes

$$\dot{\theta}(t) = \frac{\dot{\lambda}_x(0)\lambda_y(0) - \dot{\lambda}_y(0)\lambda_x(0)}{(\dot{\lambda}_y(0)t + \lambda_y(0))^2 + (\dot{\lambda}_x(0)t + \lambda_x(0))^2}.$$

This means that $\dot{\theta}(t) = 0$ for all $t \neq \bar{t}$. Similarly, one determines that $\dot{x}(t) = \dot{y}(t) = 0$ for all $t \neq \bar{t}$. Coupled with the fact that the trajectories in Q must be differentiable, we see that (iii) implies that $\gamma(t) = \gamma(t_0)$ as long as $t \neq \bar{t}$. When $t = \bar{t}$, then we have $\lambda_x(\bar{t}) = \lambda_y(\bar{t}) = 0$ and so $\lambda_\theta(\bar{t})$ is also zero. Since the equations governing the adjoint covector field are linear in the adjoint covector field, this implies that the adjoint covector field is identically zero for all t. This situation is in violation of the conditions on λ in Problem S4.49. The above arguments show that (iii) implies both (i) and (ii). These arguments are easily modified to show that (i) and (ii) also imply (iii), and that (i) and (ii) are equivalent. ∎

The following result essentially determines the character of the nontrivial singular controlled extremals.

Lemma S4.52 (Characterization of singular controlled extremals with nonzero control). *Suppose that* (γ, u) *is a stationary solution of Problem S4.49 for the planar rigid body system defined on the entire real line with, for some* $t_0 \in \mathbb{R}$, $(x(t_0), y(t_0))$ *lying on the circle of radius* $\frac{J}{mh}$ *with center at* $(0,0)$ *in the* (x,y)*-plane. If the expression* (S4.30) *is nowhere zero, then*

(i) $x^2(t) + y^2(t) = \left(\frac{J}{mh}\right)^2$ *for all* $t \in \mathbb{R}$,
(ii) $\theta(t) = \pi + \operatorname{atan}(x(t), y(t))$ *for all* $t \in \mathbb{R}$,
(iii) $\lim_{t\to\infty}(x(t), y(t)) = -\lim_{t\to-\infty}(x(t), y(t))$, *and*
(iv) $\lim_{t\to\infty}\theta(t) = \pi + \lim_{t\to-\infty}\theta(t)$.

Furthermore, the control u *is analytic.*

Proof. Choosing C_{11}, C_{10}, C_{21}, and C_{20} to be zero, it is apparent from (S4.32) that $x^2(t) + y^2(t) = \left(\frac{J}{mh}\right)^2$, and so (i) holds. It is also clear from (S4.32) that (ii), (iii), and (iv) hold. Analyticity of u follows from (S4.33). ■

In Figure S4.8 we show a typical solution of Problem S4.49 of the form

Figure S4.8. A solution of Problem S4.49 for the planar rigid body with $m = 1$, $J = 1$, and $h = \frac{1}{2}$. The initial conditions for the adjoint covector field are $\lambda_x(0) = 1$, $\lambda_y(0) = 5$, $\lambda_\theta(0) = 2\sqrt{26}$, $\dot{\lambda}_x(0) = 1$, and $\dot{\lambda}_y(0) = 1$. The time between plots of the body's position on the left is $\Delta t = 1$. On the right, the controls are shown.

described in Lemma S4.52. Note that the limiting initial and final angles are decided by the values of $\dot{\lambda}_x(0)$ and $\dot{\lambda}_y(0)$, as may be ascertained from (S4.32). It is also clear from the expressions for $x(t)$ and $y(t)$ from (S4.32) that *any* solution of Problem S4.49 will be a copy of a solution to Problem S4.49 from Lemma S4.52, but possibly translated away from $(0,0)$, and possibly moving with uniform velocity in the (x,y)-plane. To be succinct in stating the form of the general solution to Problem S4.49, it is convenient to introduce a equivalence relation on the set of curves on Q by saying that curves $\gamma_1 \colon I_1 \to \mathsf{Q}$ and $\gamma_2 \colon I_2 \to \mathsf{Q}$ are equivalent when $I_2 = I_1 + a$ for some $a \in \mathbb{R}$ and $\gamma_2(t) = \gamma_1(t-a)$ for each $t \in I_2$. When curves γ_1 and γ_2 are equivalent, we write $\gamma_1 \sim \gamma_2$. The following result makes this precise, and summarizes our description of the solutions to Problem S4.49 for the planar rigid body.

Proposition S4.53 (Characterization of singular controlled extremals). *If (γ,u) is a solution to Problem S4.49 for the planar rigid body defined on $I \subset \mathbb{R}$, then there exists*

(i) a solution $(\tilde{\gamma},\tilde{u})$ to Problem S4.49 with the property that $(\tilde{\gamma},\tilde{u}) \sim (\gamma,u)$ and

(ii) a solution $(\bar{\gamma},\bar{u})$ to Problem S4.49 that is either

(a) of the form described by Lemma S4.51 or

(b) of the form described by Lemma S4.52,

such that

(ii) $\tilde{x}(t) = x_0 + u_0 t + \bar{x}(t)$,

(iii) $\tilde{y}(t) = y_0 + u_0 t + \bar{y}(t)$, and

(iv) $\tilde{\theta}(t) = \bar{\theta}(t)$,

for some $x_0, y_0, u_0, v_0 \in \mathbb{R}$.

A "typical" solution to Problem S4.49 appears on the right in Figure S4.9. Note that, as per Proposition S4.53, it is a "superposition" of an unforced solution to Problem S4.49 with a stationary solution to Problem S4.49.

Remarks S4.54. 1. Note that the only unforced controlled extremals for the planar rigid body consist of linear motions of the center of mass with the angle θ remaining fixed. Such a motion is shown on the left in Figure S4.9.

2. The controlled extremals that we here name as singular would be "singular in all controls" for Chyba, Leonard, and Sontag [2003]. Because they use control bounds that are polyhedral, they also would consider the linear controlled extremals of Sections S4.7.3 and S4.7.2 to be singular. There would potentially be more controls that are singular in the sense of Chyba, Leonard, and Sontag.

3. While Proposition S4.53 does provide a complete description of the singular controlled extremals for the planar rigid body system, it does not illuminate the "reason" why these controlled extremals have the form they do. That is to say, there is in all likelihood a nice geometric description for these controlled extremals in terms of the affine connection $\overset{\mathbb{G}}{\nabla}$ and the input distribution $\mathscr{Y}$. However, at this point this description is unknown to the authors. •

Figure S4.9. (1) On the left is an unforced solution to Problem S4.49 for the planar rigid body with $m = 1$, $J = 1$, and $h = \frac{1}{2}$. (2) On the right is the solution to Problem S4.49 obtained by superimposing the linear motion on the left with the solution of Problem S4.49 of Figure S4.8.

S5

Mathematica® packages

While much of the methodology described in this book may be thought of as having a fairly sophisticated mathematical basis, much of it is easy to put into practice. The obstruction is often not conceptual, but rests in the fact that sometimes even simple examples can produce lengthy symbolic expressions when one carries out the analysis/design methods we describe. Therefore, in this chapter we document the use of Mathematica® packages for some of the more common computations that come up. While we do not presently support other symbolic manipulation programs, nor do we have any plans to do so, it is certainly true that, with a little effort, everything we do in Mathematica® can be done as well with any similarly-spirited program.

Our strategy is to devote one section to each package, and describe all functions in this package. In each section, a sample Mathematica® session will be given that illustrates all the functions defined by the package.

The reader is invited to download the software at

`http://penelope.mast.queensu.ca/smcs/Mma/`

The versions on the website will be updated, so there may be discrepancies with what is described here. A list of errata and changes will be maintained, along with a version of this chapter consistent with the software version. We do not claim to be sophisticated Mathematica® programmers, and we hope that some ambitious reader(s) will take it upon themselves to improve the code we have written, and make the improved code freely available.

S5.1 Tensors.m

There are several tensor manipulation packages for Mathematica® available. However, our needs are pretty limited, so we have made a version that covers these needs.

In[1]:= **<< Tensors.m**

```
\nPackage "Tensors" defines: ChangeBasis, ChangeCoordinates,
EvaluateTensor, InitializeTensor, ITen2Vec, IVec2Ten, LieDerivative,
Ten2Vec, TheJacobian, Vec2Ten.
```

To get help, type `?command`

In[2]:= **?ChangeCoordinates**

```
ChangeCoordinates[A,x,xp,xofxp,Type,Deriv] gives a tensor A of type
Type expressed originally in coordinates x, in coordinates xp. Here
xofxp gives x as a function of xp. If Type="Affine Connection" then
the input should be the Christoffel symbols of an affine connection
in coordinates x, and the result will be the Christoffel symbols in
coordinates xp.
```

S5.1.1 Tensor basics

Tensors of type (r, s) are stored as lists of depth $r + s$, with the basic list element being a component of the tensor. One can initialize a tensor to have all zero entries. The following command initializes a $(0, 2)$-tensor in a 2-dimensional vector space.

In[3]:= **g = InitializeTensor[{0, 2}, 2]**
Out[3]= {{0, 0}, {0, 0}}

Note that a tensor can be thought of as being on a vector space, or on the tangent space to a manifold. As far as how it is stored, they are the same thing.

Let us work with a specific tensor, namely the standard Riemannian metric on $\mathbb{R}^2$, first using Cartesian coordinates.

In[4]:= **g[[1, 1]] = g[[2, 2]] = 1; g**
Out[4]= {{1, 0}, {0, 1}}

A tensor can be evaluated on various of its arguments. For example, the Riemannian metric above can be evaluated on two vectors.

In[5]:= **u = {u1, u2}**
Out[5]= {u1, u2}

In[6]:= **v = {v1, v2}**
Out[6]= {v1, v2}

In[7]:= **EvaluateTensor[g, {u, v}, {0, 2}, {{}, {}}]**
Out[7]= u1 v1 + u2 v2

The syntax here bears explanation. The first argument is the tensor itself. The second argument is a list containing the vectors and covectors on which the tensor will be evaluated. The third argument is the type of the tensor.

The fourth argument consists of two lists. The first list is the contravariant (i.e., up) indices that will be left free, and the second is the covariant (i.e., down) indices that will be left free.

To see how this works, let us use the same tensor, but now evaluate it on only one argument. This corresponds in this case to the "flat map."

```
In[8]:= EvaluateTensor[g, {u}, {0, 2}, {{}, {1}}]
Out[8]= {u1, u2}
```

Since the tensor is symmetric, the answer will be the same if the second covariant index is left free.

```
In[9]:= EvaluateTensor[g, {u}, {0, 2}, {{}, {2}}]
Out[9]= {u1, u2}
```

It may be helpful to "flatten" a tensor, by which an (r, s)-tensor on a vector space of dimension n is converted to a list of length n^{r+s}. There are a few commands associated with this and related operations. First let us convert a list to a tensor.

```
In[10]:= Aten = Vec2Ten[{a11, a12, a21, a22}, {0, 2}, 2]
Out[10]= {{a11, a12}, {a21, a22}}
```

Now let us convert this back to a list.

```
In[11]:= Avec = Ten2Vec[Aten, {0, 2}, 2]
Out[11]= {a11, a12, a21, a22}
```

Specific entries can be grabbed as well. For example, one may want to grab from a long list the element corresponding to a certain tensor index. The following manipulations use the Mathematica® `Sequence` command.

```
In[12]:= IVec2Ten[3, {0, 2}, 2]
Out[12]= {2, 1}
```

```
In[13]:= Aten[[Sequence@@%]]
Out[13]= a21
```

One can also go the other way.

```
In[14]:= ITen2Vec[{1, 2}, 2]
Out[14]= 2
```

```
In[15]:= Avec[[%]]
Out[15]= a12
```

In the above commands, the "`I`" stands for "index," reflecting the fact that these commands have to do with manipulation of indices.

S5.1.2 Lie differentiation

One can Lie differentiate tensors of arbitrary type. The command takes as arguments, the tensor being Lie differentiated, the vector field with respect to which differentiation is being done, a list containing the coordinates, and the type of the tensor.

In[16]:= **X = {−y, x}**
Out[16]= $\{-y, x\}$

In[17]:= **q = {x, y}**
Out[17]= $\{x, y\}$

In[18]:= **LXg = LieDerivative[g, X, q, {0, 2}]**
Out[18]= $\{\{0, 0\}, \{0, 0\}\}$

Note that the vector field is Killing.

S5.1.3 Changes of coordinate

Coordinate changes can be done symbolically. First let us do a linear change of coordinates (i.e., a change of basis). We will work with the existing tensor `g`. One requires a change of basis matrix, and this is defined as follows. Suppose the existing basis is $\{e_1, \dots, e_n\}$ and the new basis is $\{f_1, \dots, f_n\}$. One may then write $f_i = P_i^j e_j$ (using the summation convention) for some invertible $n \times n$ matrix $\boldsymbol{P}$. The change of basis matrix in Mathematica® is defined so that `P[[i,j]]` is P_j^i. Thus, for us, if the new basis is

In[19]:= **f1 = {1, 1}; f2 = {0, 1};**

then we should define

In[20]:= **P = Transpose[{f1, f2}]**
Out[20]= $\{\{1, 0\}, \{1, 1\}\}$

We then have

In[21]:= **ChangeBasis[g, P, {0, 2}]**
Out[21]= $\{\{2, 1\}, \{1, 1\}\}$

Now let us change coordinates. We already have the coordinates `q` defined above. Let us introduce new coordinates which are polar coordinates.

In[22]:= **qp = {r, θ}**
Out[22]= $\{r, \theta\}$

What is needed for the change of basis function is the original coordinates expressed in terms of the new coordinates.

In[23]:= **qofqp = {r Cos[θ], r Sin[θ]}**

Out[23]= $\{r\ \cos[\theta], r\ \sin[\theta]\}$

Now we may make the change of coordinates by providing all of the above data, along with the type of the tensor. For example, for the vector field we have

In[24]:= **Xp = Simplify[ChangeCoordinates[X, q, qp, qofqp, {1, 0}]]**
Out[24]= $\{0, 1\}$

Also the metric.

In[25]:= **gp = Simplify[ChangeCoordinates[g, q, qp, qofqp, {0, 2}]]**
Out[25]= $\{\{1, 0\}, \{0, r^2\}\}$

The concept of a Killing vector field is coordinate invariant.

In[26]:= **LieDerivative[gp, Xp, qp, {0, 2}]**
Out[26]= $\{\{0, 0\}, \{0, 0\}\}$

One can also change coordinates for the Christoffel symbols of an affine connection. We shall do this for the Levi-Civita connection for the Riemannian metric g used above, noting that its Christoffel symbols are zero in Cartesian coordinates.

In[27]:= **conn = Table[0, {i, 2}, {j, 2}, {k, 2}]**
Out[27]= $\{\{\{0, 0\}, \{0, 0\}\}, \{\{0, 0\}, \{0, 0\}\}\}$

In[28]:= **Simplify[ChangeCoordinates[conn, q, qp, qofqp,**
Affine Connection]]
Out[28]= $\{\{\{0, 0\}, \{0, -r\}\}, \{\{0, \frac{1}{r}\}, \{\frac{1}{r}, 0\}\}\}$

One may recognize these as the Christoffel symbols for the standard metric in polar coordinates.

There is also a Jacobian function included. The Mathematica® Jacobian manipulations require too much setup to use conveniently.

In[29]:= **TheJacobian[qofqp, qp]**
Out[29]= $\{\{\cos[\theta], -r\ \sin[\theta]\}, \{\sin[\theta], r\ \cos[\theta]\}\}$

S5.2 Affine.m

The package Affine.m deals with things related to affine connections.

In[1]:= **<< Affine.m**

```
\nPackage "Tensors" defines: ChangeBasis, ChangeCoordinates,
EvaluateTensor, InitializeTensor, ITen2Vec, IVec2Ten, LieDerivative,
Ten2Vec, TheJacobian, Vec2Ten.
```

```
\nPackage "Affine" defines: AlongCurve, CovariantDerivative,
CovariantDifferential, CurvatureTensor, Grad, LeviCivita,
RicciCurvature, RiemannFlat, RiemannSharp, ScalarCurvature,
SectionalCurvature, Spray, SymmetricProduct, TorsionTensor.
```

Note that the package Tensors.m is loaded. Please see the documentation for that package to use its features. To get help, type `?command`

In[2]:= **?AlongCurve**

```
AlongCurve[A,Conn,c,t,Type,Deriv] returns the covariant derivative
of the tensor field A of type Type along the curve c. t is the time
parameter which c must depend upon.
```

S5.2.1 Riemannian geometry specifics

The metric can be used to convert vector fields to covector fields, and vice versa, in the usual manner. The rule is “Sharp raises the index (i.e., converts a covector field to a vector field) and flat lowers the index (i.e., converts a vector field to a covector field).”

In[3]:= **g = {{1, 0}, {0, r^2}}**
Out[3]= $\{\{1,0\},\{0,r^2\}\}$

In[4]:= $\mathbf{\{\{1,0\},\{0,r^2\}\}}$
Out[4]= $\{\{1,0\},\{0,r^2\}\}$

In[5]:= $\boldsymbol{\alpha} = \mathbf{\{0,1\}}$
Out[5]= $\{0,1\}$

In[6]:= **X = RiemannSharp[$\boldsymbol{\alpha}$, g]**
Out[6]= $\{0, \frac{1}{r^2}\}$

In[7]:= **RiemannFlat[X, g]**
Out[7]= $\{0,1\}$

A special instance of the flat map is the gradient, and there is a special purpose function for it.

In[8]:= **f = r Cos[$\boldsymbol{\theta}$]**
Out[8]= $r\ \cos[\theta]$

In[9]:= **q = {r, $\boldsymbol{\theta}$}**
Out[9]= $\{r,\theta\}$

In[10]:= **Grad[f, g, q]**
Out[10]= $\{\cos[\theta], -\frac{\sin[\theta]}{r}\}$

The Levi-Civita Christoffel symbols can be computed using a set of coordinates and the components of the Riemannian metric.

In[11]:= **conn = LeviCivita[g, q]**
Out[11]= $\{\{\{0,0\},\{0,-r\}\},\{\{0,\frac{1}{r}\},\{\frac{1}{r},0\}\}\}$

S5.2.2 Affine differential geometry basics

The standard covariant derivative of vector fields is computed as follows.

In[12]:= **X = {Cos[θ], r^2}**
Out[12]= $\{\cos[\theta], r^2\}$

In[13]:= **Y = {Sin[θ], 1/r}**
Out[13]= $\{\sin[\theta], \frac{1}{r}\}$

Then one computes the covariant derivative of Y with respect to X, using the Christoffel symbols for the connection.

In[14]:= **CovariantDerivative[X, Y, conn, q]**
Out[14]= $\{\frac{-r^3-\sin[\theta]}{r}, \frac{\cos[\theta]}{r^2}+3\,r\,\sin[\theta]\}$

The symmetric product is a useful operation for dealing with simple mechanical control systems, and its function works much like the covariant derivative.

In[15]:= **SymmetricProduct[X, Y, conn, q]**
Out[15]= $\{r^2\,(-1+\cos[\theta])+\frac{-r^3-\sin[\theta]}{r}, \frac{\cos[\theta]}{r^2}+4\,r\,\sin[\theta]\}$

The covariant derivative of a general tensor can also be computed. See the documentation for Tensors.m to see how tensors can be defined, and how they are stored by our packages. One should specify the tensor one is covariantly differentiating, the Christoffel symbols of the connection, the coordinates, and the type of the tensor.

In[16]:= **nablag = CovariantDifferential[g, conn, q, {0, 2}]**
Out[16]= $\{\{\{0,0\},\{0,0\}\},\{\{0,0\},\{0,0\}\}\}$

Note that, if the tensor is of type (r,s), then what comes out is a tensor of type $(r,s+1)$. To produce the covariant derivative of the tensor with respect to a vector field (i.e., a tensor field of type (r,s)), one can use the `EvaluateTensor` function that is part of the Tensors.m package.

In[17]:= **EvaluateTensor[nablag, {X}, {1, 2}, {{}, {1, 2}}]**
Out[17]= $\{\{0,0\},\{0,0\}\}$

In like manner one can compute the covariant derivative of a tensor field along a curve. This requires specifying the coordinate functions of time that define the curve. The variable parameterizing time is an argument, so can be whatever is desired.

In[18]:= $\boldsymbol{\gamma = \{r[t], \theta[t]\}}$
Out[18]= $\{r[t], \theta[t]\}$

In[19]:= **Upsilon = D[$\boldsymbol{\gamma}$, t]**
Out[19]= $\{r'[t], \theta'[t]\}$

In[20]:= **AlongCurve[Upsilon, conn, $\boldsymbol{\gamma}$, t, {1, 0}]**
Out[20]= $\left\{ - r\,\theta'[t]^2 + r''[t], \dfrac{2\,r'[t]\,\theta'[t]}{r} + \theta''[t]\right\}$

Note that the result in this case has been contrived to be the components of the geodesic equations in second-order form.

The geodesic spray can be computed by using velocity coordinates.

In[21]:= **v = D[$\boldsymbol{\gamma}$, t]**
Out[21]= $\{r'[t], \theta'[t]\}$

In[22]:= **Z = Spray[conn, $\boldsymbol{\gamma}$, v]**
Out[22]= $\left\{r'[t], \theta'[t], r\,\theta'[t]^2, -\dfrac{2\,r'[t]\,\theta'[t]}{r}\right\}$

Note that what is returned are the components of a vector field on the tangent bundle in natural coordinates.

S5.2.3 Torsion and curvature

These commands are all pretty basic. Some of them reflect mathematical constructions not defined in the text. We refer the reader to [Kobayashi and Nomizu 1963] for discussions of undefined terms.

In[23]:= **TorsionTensor[conn, q]**
Out[23]= $\{\{\{0,0\},\{0,0\}\},\{\{0,0\},\{0,0\}\}\}$

In[24]:= **CurvatureTensor[conn, q]**
Out[24]= $\{\{\{\{0,0\},\{0,0\}\},\{\{0,0\},\{0,0\}\}\},\{\{\{0,0\},\{0,0\}\},\{\{0,0\},\{0,0\}\}\}\}$

In[25]:= **RicciCurvature[conn, q]**
Out[25]= $\{\{0,0\},\{0,0\}\}$

Scalar and sectional curvature are defined only for Levi-Civita connections.

In[26]:= **ScalarCurvature[g, q]**
Out[26]= 0

The sectional curvature requires the specification of two orthonormal tangent vectors to prescribe a two-dimensional subspace.

In[27]:= **e1 = {1, 0}**

Out[27]= $\{1, 0\}$

In[28]:= **e2 = {0, 1/r}**

Out[28]= $\left\{0, \frac{1}{r}\right\}$

In[29]:= **SectionalCurvature[e1, e2, g, q]**

Out[29]= 0

S5.3 SMCS.m

The package SMCS.m deals with the modeling of simple mechanical control systems, and provides tools to perform the steps outlined in Chapter 4 for the modeling of kinetic energy, forces, and constraints. To illustrate the use of the package, we shall consider the rolling disk system described in Chapter 4, and depicted in Figure S5.1. The objective will be to systematically go through

Figure S5.1. Rolling disk

all of the modeling steps to arrive at all the components in the rolling disk model. The final step is a simulation of the resulting equations of motion.

In[1]:= **<< SMCS.m**

```
\nPackage "Tensors" defines: ChangeBasis, ChangeCoordinates,
EvaluateTensor, InitializeTensor, ITen2Vec, IVec2Ten, LieDerivative,
Ten2Vec, TheJacobian, Vec2Ten.
```

```
\nPackage "Affine" defines: AlongCurve, CovariantDerivative,
CovariantDifferential, CurvatureTensor, Grad, LeviCivita,
RicciCurvature, RiemannFlat, RiemannSharp, ScalarCurvature,
SectionalCurvature, Spray, SymmetricProduct, TorsionTensor.
```

```
\nPackage "SMCS" defines: ACCSequations, ACCSsimulate,
BodyAngularVelocity, ConstrainedConnection, Force,
GeneralizedCovariantDerivative, GetState, Hat, Hessian, KErot, KEtrans,
OrthogonalChristoffelSymbols, OrthogonalForce, OrthogonalProjection,
SetEqual, SMCSequations, SMCSsimulate, SpatialAngularVelocity, Unhat.
```

Note that the packages Tensors.m and Affine.m are loaded. We refer the reader to their documentation for instructions on using commands from these packages.

To get help, type `?command`

In[2]:= **?OrthogonalChristoffelSymbols**

```
OrthogonalChristoffelSymbols[X,g,conn,x] computes the generalized
Christoffel symbols for the orthogonal vector fields contained in the
columns of X. Here g is the matrix for the Riemannian metric, conn are
the Christoffel symbols of the Levi-Civita connection, and x are the
coordinates.
```

S5.3.1 Rigid body modeling

The rolling disk is comprised of a single body. Let us first define the inertia tensor of the body.

In[3]:= **Iten = {{Jspin, 0, 0}, {0, Jspin, 0}, {0, 0, Jroll}}**
Out[3]= {{Jspin, 0, 0}, {0, Jspin, 0}, {0, 0, Jroll}}

Now we define the forward kinematic map for the body by defining the position of the center of mass from the spatial origin, and by defining the orientation of the body frame relative to the spatial frame. Thus this step amounts to defining a vector in $\mathbb{R}^3$ and a matrix in $\mathsf{SO}(3)$. In specific examples, Mathematica® can be useful in obtaining these expressions. For the rolling disk, the derivation of the orientation matrix is not entirely trivial, and we refer the reader to Example 4.5 for details. First we define the configuration space coordinates and their velocities.

In[4]:= **conf = {x[t], y[t], θ[t], φ[t]}**
Out[4]= $\{x[t], y[t], \theta[t], \phi[t]\}$

In[5]:= **vel = D[conf, t]**
Out[5]= $\{x'[t], y'[t], \theta'[t], \phi'[t]\}$

We define the coordinates as “functions of time” in Mathematica®. We shall see that having the coordinates as functions is essential to using some of the macros defined in SMCS.m.

Now for the forward kinematic map.

In[6]:= $\mathbf{r = \{x[t], y[t], \rho\}}$
Out[6]= $\{x[t], y[t], \rho\}$

In[7]:= $\mathbf{R = \{\{Cos[\phi[t]]\ Cos[\theta[t]], Sin[\phi[t]]\ Cos[\theta[t]], Sin[\theta[t]]\},}$
$\mathbf{\{Cos[\phi[t]]\ Sin[\theta[t]], Sin[\phi[t]]\ Sin[\theta[t]],}$
$\mathbf{-\ Cos[\theta[t]]\}, \{-\ Sin[\phi[t]], Cos[\phi[t]], 0\}\}}$
Out[7]= $\{\{\cos[\phi[t]]\ \cos[\theta[t]], \cos[\theta[t]]\ \sin[\phi[t]], \sin[\theta[t]]\},$
$\{\cos[\phi[t]]\ \sin[\theta[t]], \sin[\phi[t]]\ \sin[\theta[t]], -\cos[\theta[t]]\},$
$\{-\sin[\phi[t]], \cos[\phi[t]], 0\}\}$

It is now possible to compute a multitude of things, since, as we emphasize in the text, the forward kinematic maps are key to much of our modeling. For example, one can compute body and spatial angular velocities.

In[8]:= **Simplify[BodyAngularVelocity[R, conf, t]]**
Out[8]= $\{-\sin[\phi[t]]\ \theta'[t], \cos[\phi[t]]\ \theta'[t], -\phi'[t]\}$

In[9]:= **Simplify[SpatialAngularVelocity[R, conf, t]]**
Out[9]= $\{-\sin[\theta[t]]\ \phi'[t], \cos[\theta[t]]\ \phi'[t], \theta'[t]\}$

Note that we do require the coordinates to be functions of time here, since time is one of the arguments of the angular velocity commands.

S5.3.2 Kinetic energy and the kinetic energy metric

Now we compute the kinetic energy, translational and rotational, for the body. Again, the forward kinematic map is key, and again, we do require the configuration space coordinates to be functions of time.

In[10]:= **ketran = KEtrans[r, conf, m, t]**
Out[10]= $\frac{1}{2}\, m\, (x'[t]^2 + y'[t]^2)$

In[11]:= **kerot = Simplify[KErot[R, conf, Iten, t]]**
Out[11]= $\frac{1}{2}\, (\mathrm{Jroll}\ \phi'[t]^2 + \mathrm{Jspin}\ \theta'[t]^2)$

Note that, in the above computations, the argument "`m`" is the mass. We can now obtain the total kinetic energy.

In[12]:= **KE = Simplify[ketran + kerot]**
Out[12]= $\frac{1}{2}\, (\mathrm{Jroll}\ \phi'[t]^2 + \mathrm{Jspin}\ \theta'[t]^2 + m\, (x'[t]^2 + y'[t]^2))$

Now we can compute the components of the kinetric energy metric.

In[13]:= **metric = Simplify[KE2Metric[KE, vel]]**
Out[13]= $\{\{m, 0, 0, 0\}, \{0, m, 0, 0\}, \{0, 0, \mathrm{Jspin}, 0\}, \{0, 0, 0, \mathrm{Jroll}\}\}$

We can also compute the Christoffel symbols for the associated Levi-Civita affine connection, although these are trivial in this case.

In[14]:= **lcgamma = LeviCivita[metric, conf]**
Out[14]= {{{0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}},
{{0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}},
{{0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}},
{{0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}, {0, 0, 0, 0}}}

S5.3.3 Force modeling

Now we consider the modeling of forces using the approach in the text. This is simple given the forward kinematic map. For the rolling disk, there are only control forces, and there are two of these. First we consider the force that spins the disk. The Newtonian force and torque are first defined.

In[15]:= **force1 = {{0, 0, 0}}**
Out[15]= {{0, 0, 0}}

We now do the same for torques.

In[16]:= **torque1 = {{0, 0, 1}}**
Out[16]= {{0, 0, 1}}

Note that the Newtonian force and torque are a list of vectors in $\mathbb{R}^3$. The length of the list is the number of bodies the force and torque act on, with each entry in the list corresponding to the force and torque exerted on a single one of the bodies. In this case, there is just one body, so the list has length one. See Section S5.4 for an example with multiple bodies.

Next we create the Lagrangian force.

In[17]:= **F1 = Simplify[Force[torque1, force1, {R}, {r}, conf, t]]**
Out[17]= {0, 0, 1, 0}

Note that the first four arguments are lists whose length is the number of bodies the Newtonian force and torque interact with. Let us do the same for the other control force that moves the wheels.

In[18]:= **force2 = {{0, 0, 0}}**
Out[18]= {{0, 0, 0}}

In[19]:= **torque2 = {{− Sin[θ[t]], Cos[θ[t]], 0}}**
Out[19]= $\{\{-\sin[\theta[t]], \cos[\theta[t]], 0\}\}$

In[20]:= **F2 = Simplify[Force[torque2, force2, {R}, {r}, conf, t]]**
Out[20]= {0, 0, 0, 1}

S5.3.4 Nonholonomic constraint modeling I

Next we turn to the modeling of nonholonomic constraints as described in the text. There are many ways one can do this. For example, one can use the constrained connection by computing its Christoffel symbols. Let us illustrate the steps. The first step is to compute an orthogonal basis of vector fields for which the first vector fields in the list are a basis for the constraint distribution. For the rolling disk, it turns out that there is a global orthogonal basis for $\mathcal{D}$. This is generally not the case. For example, it might be the case that the constraint distribution does not have constant rank. And, if the constraint distribution *does* have constant rank, there still might not be a global basis. However, since we are in luck here, we can proceed without misadventure. First we provide a set of covector fields that annihilate the constraint distribution.

In[21]:= $\mathbf{omega1\ =\ \{1, 0, 0, -\rho\ Cos[\theta[t]]\}}$

Out[21]= $\{1, 0, 0, -\rho\ \cos[\theta[t]]\}$

In[22]:= $\mathbf{omega2\ =\ \{0, 1, 0, -\rho\ Sin[\theta[t]]\}}$

Out[22]= $\{0, 1, 0, -\rho\ \sin[\theta[t]]\}$

By "sharping" these relative to the kinetic energy metric, we get two vector fields that are $\mathbb{G}$-orthogonal to the constraint distribution. Then we need to ensure that these are $\mathbb{G}$-orthogonal.

In[23]:= $\mathbf{X3\ =\ RiemannSharp[omega1, metric]}$

Out[23]= $\left\{\frac{1}{m}, 0, 0, -\frac{\rho\ \cos[\theta[t]]}{\mathrm{Jroll}}\right\}$

In[24]:= $\mathbf{X4t\ =\ RiemannSharp[omega2, metric]}$

Out[24]= $\left\{0, \frac{1}{m}, 0, -\frac{\rho\ \sin[\theta[t]]}{\mathrm{Jroll}}\right\}$

The next formula is the Gram–Schmidt Procedure to get an orthogonal basis.

In[25]:= $\mathbf{X4\ =\ Simplify[X4t - (X3.metric.X4t)X3/(X3.metric.X3)]}$

Out[25]= $\left\{-\frac{\rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2\ \cos[\theta[t]]^2}, \frac{1}{m}, 0, -\frac{\rho\ \sin[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2\ \cos[\theta[t]]^2}\right\}$

Note that we will not actually do much with `X3` and `X4`, but we produce them anyway, just to show how one does these orthogonal basis computations.

Now we use the two vector fields defined in the text as being a $\mathbb{G}$-orthogonal basis for the constraint distribution.

In[26]:= $\mathbf{X1\ =\ \{\rho\ Cos[\theta[t]], \rho\ Sin[\theta[t]], 0, 1\}}$

Out[26]= $\{\rho\ \cos[\theta[t]], \rho\ \sin[\theta[t]], 0, 1\}$

In[27]:= $\mathbf{X2\ =\ \{0, 0, 1, 0\}}$

Out[27]= $\{0, 0, 1, 0\}$

In[28]:= $\mathbf{X\ =\ \{X1, X2, X3, X4\}}$

Out[28]= $\{\{\rho\ \cos[\theta[t]], \rho\ \sin[\theta[t]], 0, 1\}, \{0, 0, 1, 0\}, \{\frac{1}{m}, 0, 0, -\frac{\rho\ \cos[\theta[t]]}{\text{Jroll}}\},$
$\{-\frac{\rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2 \cos[\theta[t]]^2}, \frac{1}{m}, 0, -\frac{\rho\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2 \cos[\theta[t]]^2}\}\}$

One can check that these vector fields are indeed orthogonal.

In[29]:= **Simplify[**
Table[(X[[i]].metric.X[[j]])/(X[[i]].metric.X[[i]]), {i, 4}, {j, 4}]]
Out[29]= $\{\{1, 0, 0, 0\}, \{0, 1, 0, 0\}, \{0, 0, 1, 0\}, \{0, 0, 0, 1\}\}$

One can now determine the components of the orthogonal projection onto $\mathcal{D}^{\perp}$. To do this, it is less cumbersome if we compute the orthogonal projection onto $\mathcal{D}$ first. Note that this only requires the basis for $\mathcal{D}$, and that this basis needs to be orthonormal for the macro `OrthogonalProjection`.

In[30]:= **P = Simplify[OrthogonalProjection[**
{X1/Sqrt[X1.metric.X1], X2/Sqrt[X2.metric.X2]}, metric]]
Out[30]= $\{\{\frac{m\ \rho^2 \cos[\theta[t]]^2}{\text{Jroll} + m\ \rho^2}, \frac{m\ \rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}, 0,$
$\frac{\text{Jroll}\ \rho\ \cos[\theta[t]]}{\text{Jroll} + m\ \rho^2}\}, \{\frac{m\ \rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2},$
$\frac{m\ \rho^2 \sin[\theta[t]]^2}{\text{Jroll} + m\ \rho^2}, 0, \frac{\text{Jroll}\ \rho\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}\}, \{0, 0, 1, 0\},$
$\{\frac{m\ \rho\ \cos[\theta[t]]}{\text{Jroll} + m\ \rho^2}, \frac{m\ \rho\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}, 0, \frac{\text{Jroll}}{\text{Jroll} + m\ \rho^2}\}\}$

Let us at least verify that this is actually the $\mathbb{G}$-orthogonal projection onto $\mathcal{D}$.

In[31]:= **Table[Simplify[P.X[[i]] − X[[i]]], {i, 2}]**
Out[31]= $\{\{0, 0, 0, 0\}, \{0, 0, 0, 0\}\}$

In[32]:= **Table[Simplify[P.X[[i]]], {i, 3, 4}]**
Out[32]= $\{\{0, 0, 0, 0\}, \{0, 0, 0, 0\}\}$

Now we define the projection onto $\mathcal{D}^{\perp}$.

In[33]:= **Pperp = Simplify[IdentityMatrix[4] − P]**
Out[33]= $\{\{1 - \frac{m\ \rho^2 \cos[\theta[t]]^2}{\text{Jroll} + m\ \rho^2}, -\frac{m\ \rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}, 0, -\frac{\text{Jroll}\ \rho\ \cos[\theta[t]]}{\text{Jroll} + m\ \rho^2}\},$
$\{-\frac{m\ \rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}, 1 - \frac{m\ \rho^2 \sin[\theta[t]]^2}{\text{Jroll} + m\ \rho^2},$
$0, -\frac{\text{Jroll}\ \rho\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}\}, \{0, 0, 0, 0\},$
$\{-\frac{m\ \rho\ \cos[\theta[t]]}{\text{Jroll} + m\ \rho^2}, -\frac{m\ \rho\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}, 0, \frac{m\ \rho^2}{\text{Jroll} + m\ \rho^2}\}\}$

Let us record the orthogonal basis for $\mathcal{D}$ for future use.

In[34]:= **Ddim = 2**
Out[34]= 2

In[35]:= **Dbasis = Table[X[[i]], {i, Ddim}]**
Out[35]= $\{\{\rho\ \cos[\theta[t]], \rho\ \sin[\theta[t]], 0, 1\}, \{0, 0, 1, 0\}\}$

Now we compute the Christoffel symbols for the constrained connection. The identity matrix in the second argument seems to be out of place here. The meaning of this second argument, along with an example of how it is used, can be found in Section S5.4.

In[36]:= **cgamma = ConstrainedConnection[**
lcgamma, IdentityMatrix[4], Pperp, conf]
Out[36]=
$$\Big\{\Big\{\Big\{0, 0, \frac{2\ m\ \rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}, \Big\{0, 0, -\frac{m\ \rho^2\ \cos[\theta[t]]^2}{\mathrm{Jroll} + m\ \rho^2} + \frac{m\ \rho^2\ \sin[\theta[t]]^2}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}, \{0, 0, 0, 0\}, \Big\{0, 0, \frac{\mathrm{Jroll}\ \rho\ \sin[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}\Big\},$$
$$\Big\{\Big\{0, 0, -\frac{m\ \rho^2\ \cos[\theta[t]]^2}{\mathrm{Jroll} + m\ \rho^2} + \frac{m\ \rho^2\ \sin[\theta[t]]^2}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}, \Big\{0, 0, -\frac{2\ m\ \rho^2\ \cos[\theta[t]]\ \sin[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}, \{0, 0, 0, 0\}, \Big\{0, 0, -\frac{\mathrm{Jroll}\ \rho\ \cos[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}\Big\},$$
$$\{\{0, 0, 0, 0\}, \{0, 0, 0, 0\}, \{0, 0, 0, 0\}, \{0, 0, 0, 0\}\},$$
$$\Big\{\Big\{0, 0, \frac{m\ \rho\ \sin[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}, \Big\{0, 0, -\frac{m\ \rho\ \cos[\theta[t]]}{\mathrm{Jroll} + m\ \rho^2}, 0\Big\}, \{0, 0, 0, 0\}, \{0, 0, 0, 0\}\Big\}\Big\}$$

S5.3.5 Nonholonomic constraint modeling II

In this section we illustrate the method for handling nonholonomic constraints that normally works best in practice, namely using the orthogonal Poincaré representation. Here we only compute the minimum number of Christoffel symbols. Fortunately, we have already done much of the work, namely the computation of a $\mathbb{G}$-orthogonal basis for the constraint distribution. Therefore, we can directly compute the 2^3 Christoffel symbols that appear in the orthogonal Poincaré representation.

In[37]:= **ogamma = OrthogonalChristoffelSymbols[**
Dbasis, metric, lcgamma, conf]
Out[37]= $\{\{\{0, 0\}, \{0, 0\}\}, \{\{0, 0\}, \{0, 0\}\}\}$

It is possible to covariantly differentiate vector fields taking values in the constraint distribution using the orthogonal Christoffel symbols. In fact, the command for doing this will work even if the vector fields forming the basis for the constraint distribution are not orthogonal. To execute the command, one needs to represent vector fields with values in the constraint distribution. This is done by giving their components relative to the basis vector fields. Let us define two such vector fields in general form.

In[38]:= **U = Table[Ucomp[i][x[t], x[y], θ[t], ϕ[t]], {i, Ddim}]**

Out[38]= $\{\mathrm{Ucomp}[1][x[t], x[y], \theta[t], \phi[t]], \mathrm{Ucomp}[2][x[t], x[y], \theta[t], \phi[t]]\}$

In[39]:= **V = Table[Vcomp[i][x[t], x[y], θ[t], ϕ[t]], {i, Ddim}]**

Out[39]= $\{\mathrm{Vcomp}[1][x[t], x[y], \theta[t], \phi[t]], \mathrm{Vcomp}[2][x[t], x[y], \theta[t], \phi[t]]\}$

Now we covariantly differentiate `V` with respect to `U`.

In[40]:= **GeneralizedCovariantDerivative[U, V, Dbasis, ogamma, conf]**

Out[40]=
$$\begin{aligned}\{&\mathrm{Ucomp}[2][x[t], x[y], \theta[t], \phi[t]]\ \mathrm{Vcomp}[1]^{(0,0,1,0)}[x[t], x[y], \theta[t], \phi[t]]+\\ &\mathrm{Ucomp}[1][x[t], x[y], \theta[t], \phi[t]]\ (\mathrm{Vcomp}[1]^{(0,0,0,1)}[x[t], x[y], \theta[t], \phi[t]]+\\ &\quad\rho\ \cos[\theta[t]]\ \mathrm{Vcomp}[1]^{(1,0,0,0)}[x[t], x[y], \theta[t], \phi[t]]),\\ &\mathrm{Ucomp}[2][x[t], x[y], \theta[t], \phi[t]]\ \mathrm{Vcomp}[2]^{(0,0,1,0)}[x[t], x[y], \theta[t], \phi[t]]+\\ &\mathrm{Ucomp}[1][x[t], x[y], \theta[t], \phi[t]]\ (\mathrm{Vcomp}[2]^{(0,0,0,1)}[x[t], x[y], \theta[t], \phi[t]]+\\ &\quad\rho\ \cos[\theta[t]]\ \mathrm{Vcomp}[2]^{(1,0,0,0)}[x[t], x[y], \theta[t], \phi[t]])\}\end{aligned}$$

It is possible to use the generalized covariant derivative to perform controllability computations. An example of this is given in Section S5.4.

We also need to model the forces in the framework of pseudo-velocities. Two things must be done to do this. First, the vector forces need to be projected onto the constraint distribution. Then the resulting vector forces need to be represented in terms of the (not necessarily $\mathbb{G}$-orthogonal) basis for the constraint distribution. In the case when the basis for $\mathcal{D}$ is $\mathbb{G}$-orthogonal, there is a command for this.

In[41]:= **Y1o = OrthogonalForce[F1, Dbasis, metric]**

Out[41]= $\left\{0, \frac{1}{\mathrm{Jspin}}\right\}$

In[42]:= **Y2o = Simplify[OrthogonalForce[F2, Dbasis, metric]]**

Out[42]= $\left\{\frac{1}{\mathrm{Jroll} + m\,\rho^2}, 0\right\}$

S5.3.6 Equations of motion I

We will compute equations of motion in two different ways. First we use the Christoffel symbols for the constrained connection as above, and just produce the full geodesic equations. First we need to give the input vector fields, properly projected onto the constraint distribution.

In[43]:= **Y1c = P.RiemannSharp[F1, metric]**

Out[43]= $\left\{0, 0, \frac{1}{\text{Jspin}}, 0\right\}$

In[44]:= **Y2c = P.RiemannSharp[F2, metric]**

Out[44]= $\left\{\frac{\rho\ \cos[\theta[t]]}{\text{Jroll} + m\ \rho^2}, \frac{\rho\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}, 0, \frac{1}{\text{Jroll} + m\ \rho^2}\right\}$

The total input is a linear combination of the two inputs, with the coefficients being the controls. Let us leave the controls as general for the moment.

In[45]:= **Yc = u1 Y1c + u2 Y2c**

Out[45]= $\left\{\frac{\rho\,\text{u2}\ \cos[\theta[t]]}{\text{Jroll} + m\ \rho^2}, \frac{\rho\,\text{u2}\ \sin[\theta[t]]}{\text{Jroll} + m\ \rho^2}, \frac{\text{u1}}{\text{Jspin}}, \frac{\text{u2}}{\text{Jroll} + m\ \rho^2}\right\}$

Now we can produce the equations of motion.

In[46]:= **eqmot1 = Simplify[ACCSequations[cgamma, Yc, conf, t]]**

Out[46]=
$$\Big\{\frac{1}{\text{Jroll} + m\ \rho^2}(-\rho\,\text{u2}\ \cos[\theta[t]] + \text{Jroll}\,\rho\ \sin[\theta[t]]\ \phi'[t]\ \theta'[t] + m\ \rho^2\ \theta'[t]\ (\sin[2\ \theta[t]]\ x'[t] - \cos[2\ \theta[t]]\ y'[t]) + \text{Jroll}\ x''[t] + m\ \rho^2\ x''[t]) == 0, 0 == \frac{1}{\text{Jroll} + m\ \rho^2}(\rho\,\text{u2}\ \sin[\theta[t]] + \text{Jroll}\,\rho\ \cos[\theta[t]]\ \phi'[t]\ \theta'[t] + m\ \rho^2\ \theta'[t]\ (\cos[2\ \theta[t]]\ x'[t] + \sin[2\ \theta[t]]\ y'[t]) - (\text{Jroll} + m\ \rho^2)\ y''[t]), \theta''[t] == \frac{\text{u1}}{\text{Jspin}}, \frac{1}{\text{Jroll} + m\ \rho^2}(-\text{u2} + m\,\rho\,\theta'[t]\ (\sin[\theta[t]]\ x'[t] - \cos[\theta[t]]\ y'[t]) + (\text{Jroll} + m\ \rho^2)\ \phi''[t]) == 0\Big\}$$

S5.3.7 Equations of motion II

Now we provide another means of producing the equations of motion, using the Poincaré representation. Since this representation, in principle, captures all possibilities, one must allow for both constrained and unconstrained cases. One of the differences will be that, in the unconstrained case with the natural Christoffel symbols, the dependent variables will be the configuration coordinates, and all equations will be second-order. For systems with constraints, and using generalized Christoffel symbols, there will be pseudo-velocities, and the equations will be first-order. Things are further complicated by the fact that, in some examples, some of the pseudo-velocities will be actual velocities. Thus the resulting equations of motion will be a mixture of first- and second-order equations. The difficulty is then to determine the correct state, taking into account that some pseudo-velocities are actual velocities. There is a command for this, whose usage we now illustrate. First one defines the "full" set of pseudo-velocities. In this case there are two.

In[47]:= **pv = {pv1[t], pv2[t]}**

Out[47]= $\{\mathrm{pv1}[t], \mathrm{pv2}[t]\}$

Then one extracts the state for the equations, properly taking into account that some of the pseudo-velocities are velocities. The following command does not require a $\mathbb{G}$-orthogonal basis for $\mathcal{D}$.

In[48]:= **state = GetState[Dbasis, pv, conf, t]**
Out[48]= $\{x[t], y[t], \theta[t], \phi[t]\}$

Note that, in the rolling disk, all pseudo-velocities are actual velocities, reflected by the fact that no pseudo-velocities appear in the list of states.

If the system were unconstrained and one wished to use the natural representation, then one would proceed as follows.

In[49]:= **GetState[IdentityMatrix[4],**
{pv1[t], pv2[t], pv3[t], pv4[t]}, conf, t]
Out[49]= $\{x[t], y[t], \theta[t], \phi[t]\}$

The first argument being the identity matrix corresponds to the fact that the pssudo-velocities are all real velocities. Then the state is correctly returned as simply the configuration coordinates. In such cases one may want to not bother with listing the pseudo-velocities, in which case an empty list will guarantee the correct result.

In[50]:= **GetState[IdentityMatrix[4], {}, conf, t]**
Out[50]= $\{x[t], y[t], \theta[t], \phi[t]\}$

A second difficulty arises with the treatment of forces. In unconstrained systems, one simply wants to use the natural representation of the force. For constrained systems using pseudo-velocities, one must properly represent vector forces as above. Therefore, the user is required to define a vector force being applied to the system by giving its components in the basis for $\mathcal{D}$. In this case, we have already done this.

In[51]:= **Yo = u1 Y1o + u2 Y2o**
Out[51]= $\left\{\frac{\mathrm{u2}}{\mathrm{Jroll} + m\,\rho^2}, \frac{\mathrm{u1}}{\mathrm{Jspin}}\right\}$

In the unconstrained case when using the natural representation, one would simply use the list comprised on the components of the vector force.

Now we can formulate the equations of motion. Note that one uses all pseudo-velocities. The program sorts out the state along the lines of the `GetState` command above. For an unconstrained system, an empty list of pseudo-velocities will give the desired result. Note that, for the following command, the basis for $\mathcal{D}$ need not be $\mathbb{G}$-orthogonal.

In[52]:= **eqmot2 = SMCSequations[ogamma, Yo, Dbasis, pv, conf, t]**
Out[52]= $\{x'[t] == \rho\, \cos[\theta[t]]\, \phi'[t], y'[t] == \rho\, \sin[\theta[t]]\, \phi'[t],$
$\phi''[t] == \frac{\mathrm{u2}}{\mathrm{Jroll} + m\,\rho^2}, \theta''[t] == \frac{\mathrm{u1}}{\mathrm{Jspin}}\}$

S5.3.8 Simulation

Once one has the equations of motion, one would like to be able to numerically solve the equations. In Mathematica® this is done using `NDSolve`, but an interface has been provided that simplifies certain things. First let us give numerical values for the parameters.

In[53]:= **params = {Jspin → 2, Jroll → 1, m → 1/2, ρ → 1}**
Out[53]= $\{\text{Jspin} \to 2, \text{Jroll} \to 1, m \to \frac{1}{2}, \rho \to 1\}$

Now define specific controls.

In[54]:= **u1 = 2 Sin[3t]**
Out[54]= $2\ \sin[3\,t]$

In[55]:= **u2 = 2 Sin[2t]**
Out[55]= $2\ \sin[2\,t]$

Next define the initial and final times for the simulation.

In[56]:= **Ti = 0**
Out[56]= 0

In[57]:= **Tf = 3π**
Out[57]= $3\,\pi$

Now the initial conditions.

In[58]:= **qinit = vinit = {0, 0, 0, 0}**
Out[58]= $\{0, 0, 0, 0\}$

It is assumed that the initial velocity satisfies the constraint.

Now simulate.

In[59]:= **sol1 = ACCSsimulate[(eqmot1/.params), conf, qinit, vinit, t, Ti, Tf]**
Out[59]= $\{\{x[t] \to \text{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$y[t] \to \text{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\theta[t] \to \text{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\phi[t] \to \text{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t]\}\}$

In[60]:= **Plot[x[t]/.sol1, {t, Ti, Tf}]**

Out[60]= -Graphics-

Now we simulate the system as a Poincaré representation. The initial condition is given as initial configuration, plus a complete list of initial pseudo-velocities. The program converts this into a state initial condition.

In[61]:= **pvinit = Table[vinit.metric.X[[i]]/(X[[i]].metric.X[[i]]), {i, Ddim}]**

Out[61]= $\{0, 0\}$

Now simulate.

In[62]:= **sol2 = SMCSsimulate[(eqmot2/.params), Dbasis, conf, pv, qinit, pvinit, t, Ti, Tf]**

Out[62]= $\{\{x[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$y[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\theta[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\phi[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t]\}\}$

In[63]:= **Plot[x[t]/.sol2, {t, Ti, Tf}]**

Out[63]= -Graphics-

The two solution methods give the same solutions to the differential equation, as expected.

S5.3.9 Other useful macros

The primary components of SMCS.m are illustrated above. But there are a few other macros that are implemented that might be useful. Let us indicate what these are and what they do.

There are macros that manage the isomorphism between $\mathfrak{so}(3)$ and $\mathbb{R}^3$.

In[64]:= **omegahat = Hat[{w1, w2, w3}]**
Out[64]= {{0, −w3, w2}, {w3, 0, −w1}, {−w2, w1, 0}}

In[65]:= **$\boldsymbol{\omega}$ = Unhat[omegahat]**
Out[65]= {w1, w2, w3}

There is also an implementation of the Hessian. The implementation supposes that the function is being evaluated at a critical point, where the matrix representative of the Hessian is simply the matrix of second partial derivatives.

In[66]:= **Hessian[f[x[t], y[t], $\boldsymbol{\theta}$[t], $\boldsymbol{\phi}$[t]], conf]**

Out[66]= $\{\{f^{(2,0,0,0)}[x[t], y[t], \theta[t], \phi[t]], f^{(1,1,0,0)}[x[t], y[t], \theta[t], \phi[t]],$
$f^{(1,0,1,0)}[x[t], y[t], \theta[t], \phi[t]], f^{(1,0,0,1)}[x[t], y[t], \theta[t], \phi[t]]\},$
$\{f^{(1,1,0,0)}[x[t], y[t], \theta[t], \phi[t]], f^{(0,2,0,0)}[x[t], y[t], \theta[t], \phi[t]],$
$f^{(0,1,1,0)}[x[t], y[t], \theta[t], \phi[t]], f^{(0,1,0,1)}[x[t], y[t], \theta[t], \phi[t]]\},$
$\{f^{(1,0,1,0)}[x[t], y[t], \theta[t], \phi[t]], f^{(0,1,1,0)}[x[t], y[t], \theta[t], \phi[t]],$
$f^{(0,0,2,0)}[x[t], y[t], \theta[t], \phi[t]], f^{(0,0,1,1)}[x[t], y[t], \theta[t], \phi[t]]\},$
$\{f^{(1,0,0,1)}[x[t], y[t], \theta[t], \phi[t]], f^{(0,1,0,1)}[x[t], y[t], \theta[t], \phi[t]],$
$f^{(0,0,1,1)}[x[t], y[t], \theta[t], \phi[t]], f^{(0,0,0,2)}[x[t], y[t], \theta[t], \phi[t]]\}\}$

A generally useful macro is `SetEqual,` which is used to set the components of two lists equal to one another in the form of an equation.

In[67]:= **list1 = Table[l1[i], {i, 3}]**
Out[67]= {l1[1], l1[2], l1[3]}

In[68]:= **list2 = Table[l2[i], {i, 3}]**
Out[68]= {l2[1], l2[2], l2[3]}

In[69]:= **SetEqual[list1, list2]**
Out[69]= {l1[1] == l2[1], l1[2] == l2[2], l1[3] == l2[3]}

S5.4 Snakeboard modeling using Mathematica®

In this section we illustrate the modeling of the snakeboard using our Mathematica® macros. The snakeboard is considered in Section 13.4 in the text. The snakeboard is a quite complicated example, so this section provides a good test for one's understanding of how to use the Mathematica® packages. For reference, we illustrate the model we use for the snakeboard in Figure S5.2.

In[1]:= **<< SMCS.m**

```
\nPackage "Tensors" defines: ChangeBasis, ChangeCoordinates,
EvaluateTensor, InitializeTensor, ITen2Vec, IVec2Ten, LieDerivative,
Ten2Vec, TheJacobian, Vec2Ten.
```

```
\nPackage "Affine" defines: AlongCurve, CovariantDerivative,
CovariantDifferential, CurvatureTensor, Grad, LeviCivita,
RicciCurvature, RiemannFlat, RiemannSharp, ScalarCurvature,
SectionalCurvature, Spray, SymmetricProduct, TorsionTensor.
```

```
\nPackage "SMCS" defines: ACCSequations, ACCSsimulate,
BodyAngularVelocity, ConstrainedConnection, Force,
GeneralizedCovariantDerivative, GetState, Hat, Hessian, KErot, KEtrans,
OrthogonalChristoffelSymbols, OrthogonalForce, OrthogonalProjection,
SetEqual, SMCSequations, SMCSsimulate, SpatialAngularVelocity, Unhat.
```

Figure S5.2. Model for the snakeboard

S5.4.1 Inertia tensors

The snakeboard is comprised of four components, the coupler, the rotor, the "back" wheels, and the "front" wheels. These will be denoted with the suffixes "`c`," "`r`," "`f`," and "`b`," respectively. In all cases, the frames we consider will be those described in the text. Let us first define the inertia tensors for the four bodies.

In[2]:= **Ic = {{a11, a12, 0}, {a12, a22, 0}, {0, 0, Jc}}**
Out[2]= {{a11, a12, 0}, {a12, a22, 0}, {0, 0, Jc}}

In[3]:= **Ir = {{a11, a12, 0}, {a12, a22, 0}, {0, 0, Jr}}**
Out[3]= {{a11, a12, 0}, {a12, a22, 0}, {0, 0, Jr}}

In[4]:= **Ib = If1 = {{a11, a12, 0}, {a12, a22, 0}, {0, 0, Jw}}**
Out[4]= {{a11, a12, 0}, {a12, a22, 0}, {0, 0, Jw}}

Note that we are forced to use "`If1`" since "`If`" is an internal Mathematica® symbol. Also note that the parameters `a11`, `a12`, and `a22` are dummies, and should not show up in the final expressions by virtue of the way in which the frames are defined.

S5.4.2 Forward kinematic maps

Now we define the forward kinematic maps for the four bodies by defining the positions of the centers of mass from the spatial origin, and by defining the orientation of the body frames relative to the spatial frame. Thus this step amounts to defining four vectors in $\mathbb{R}^3$ and four matrices in $\mathsf{SO}(3)$. First we need coordinates and their velocities. We use the ones from the text, of course.

In[5]:= $\mathbf{conf\ =\ \{x[t], y[t], \boldsymbol{\theta}[t], \boldsymbol{\psi}[t], \boldsymbol{\phi}[t]\}}$
Out[5]= $\{x[t], y[t], \theta[t], \psi[t], \phi[t]\}$

In[6]:= $\mathbf{vel\ =\ D[conf, t]}$
Out[6]= $\{x'[t], y'[t], \theta'[t], \psi'[t], \phi'[t]\}$

Now for the forward kinematic maps. The lowercase `r` is for the center of mass position, and the uppercase `R` is for the orientation matrices.

In[7]:= $\mathbf{rc\ =\ \{x[t], y[t], 0\}}$
Out[7]= $\{x[t], y[t], 0\}$

In[8]:= $\mathbf{rr\ =\ \{x[t], y[t], 0\}}$
Out[8]= $\{x[t], y[t], 0\}$

In[9]:= $\mathbf{rb\ =\ \{x[t] - l\ Cos[\boldsymbol{\theta}[t]], y[t] - l\ Sin[\boldsymbol{\theta}[t]], 0\}}$
Out[9]= $\{-l\ \cos[\theta[t]] + x[t], -l\ \sin[\theta[t]] + y[t], 0\}$

In[10]:= $\mathbf{rf\ =\ \{x[t] + l\ Cos[\boldsymbol{\theta}[t]], y[t] + l\ Sin[\boldsymbol{\theta}[t]], 0\}}$
Out[10]= $\{l\ \cos[\theta[t]] + x[t], l\ \sin[\theta[t]] + y[t], 0\}$

In[11]:= $\mathbf{Rc\ =\ \{\{Cos[\boldsymbol{\theta}[t]], -Sin[\boldsymbol{\theta}[t]], 0\}, \{Sin[\boldsymbol{\theta}[t]], Cos[\boldsymbol{\theta}[t]], 0\}, \{0, 0, 1\}\}}$
Out[11]= $\{\{\cos[\theta[t]], -\sin[\theta[t]], 0\}, \{\sin[\theta[t]], \cos[\theta[t]], 0\}, \{0, 0, 1\}\}$

In[12]:= $\mathbf{Rr\ =\ \{\{Cos[\boldsymbol{\theta}[t] + \boldsymbol{\psi}[t]], -Sin[\boldsymbol{\theta}[t] + \boldsymbol{\psi}[t]], 0\}, \{Sin[\boldsymbol{\theta}[t] + \boldsymbol{\psi}[t]], Cos[\boldsymbol{\theta}[t] + \boldsymbol{\psi}[t]], 0\}, \{0, 0, 1\}\}}$
Out[12]= $\{\{\cos[\psi[t] + \theta[t]], -\sin[\psi[t] + \theta[t]], 0\}, \{\sin[\psi[t] + \theta[t]], \cos[\psi[t] + \theta[t]], 0\}, \{0, 0, 1\}\}$

In[13]:= $\mathbf{Rb\ =\ \{\{Cos[\boldsymbol{\theta}[t] + \boldsymbol{\phi}[t]], -Sin[\boldsymbol{\theta}[t] + \boldsymbol{\phi}[t]], 0\}, \{Sin[\boldsymbol{\theta}[t] + \boldsymbol{\phi}[t]], Cos[\boldsymbol{\theta}[t] + \boldsymbol{\phi}[t]], 0\}, \{0, 0, 1\}\}}$
Out[13]= $\{\{\cos[\phi[t] + \theta[t]], -\sin[\phi[t] + \theta[t]], 0\}, \{\sin[\phi[t] + \theta[t]], \cos[\phi[t] + \theta[t]], 0\}, \{0, 0, 1\}\}$

In[14]:= $\mathbf{Rf\ =\ \{\{Cos[\boldsymbol{\theta}[t] - \boldsymbol{\phi}[t]], -Sin[\boldsymbol{\theta}[t] - \boldsymbol{\phi}[t]], 0\}, \{Sin[\boldsymbol{\theta}[t] - \boldsymbol{\phi}[t]], Cos[\boldsymbol{\theta}[t] - \boldsymbol{\phi}[t]], 0\}, \{0, 0, 1\}\}}$
Out[14]= $\{\{\cos[\phi[t] - \theta[t]], \sin[\phi[t] - \theta[t]], 0\}, \{-\sin[\phi[t] - \theta[t]], \cos[\phi[t] - \theta[t]], 0\}, \{0, 0, 1\}\}$

S5.4.3 Kinetic energy and the kinetic energy metric

Now we compute the kinetic energies, translational and rotational, for all of the bodies.

In[15]:= **ketranc = KEtrans[rc, conf, mc, t]**

Out[15]= $\frac{1}{2}\,\mathrm{mc}\,(x'[t]^2 + y'[t]^2)$

In[16]:= **kerotc = Simplify[KErot[Rc, conf, Ic, t]]**

Out[16]= $\frac{1}{2}\,\mathrm{Jc}\,\theta'[t]^2$

In[17]:= **ketranr = KEtrans[rr, conf, mr, t]**

Out[17]= $\frac{1}{2}\,\mathrm{mr}\,(x'[t]^2 + y'[t]^2)$

In[18]:= **kerotr = Simplify[KErot[Rr, conf, Ir, t]]**

Out[18]= $\frac{1}{2}\,\mathrm{Jr}\,(\psi'[t] + \theta'[t])^2$

In[19]:= **ketranb = KEtrans[rb, conf, mw, t]**

Out[19]= $\frac{1}{2}\,\mathrm{mw}\,((l\ \sin[\theta[t]]\,\theta'[t] + x'[t])^2 + (-l\ \cos[\theta[t]]\,\theta'[t] + y'[t])^2)$

In[20]:= **kerotb = Simplify[KErot[Rb, conf, Ib, t]]**

Out[20]= $\frac{1}{2}\,\mathrm{Jw}\,(\phi'[t] + \theta'[t])^2$

In[21]:= **ketranf = KEtrans[rf, conf, mw, t]**

Out[21]= $\frac{1}{2}\,\mathrm{mw}\,((-l\ \sin[\theta[t]]\,\theta'[t] + x'[t])^2 + (l\ \cos[\theta[t]]\,\theta'[t] + y'[t])^2)$

In[22]:= **kerotf = Simplify[KErot[Rf, conf, If1, t]]**

Out[22]= $\frac{1}{2}\,\mathrm{Jw}\,(\phi'[t] - \theta'[t])^2$

Note that, in the above computations, the arguments "`mc`," "`mr`," and "`mw`" are masses. We can now obtain the total kinetic energy.

In[23]:= **KE = Simplify[ketranc + kerotc + ketranr + kerotr + ketranb + kerotb + ketranf + kerotf]**

Out[23]= $$\frac{1}{2}\,(2\,\mathrm{Jw}\,\phi'[t]^2 + \mathrm{Jr}\,\psi'[t]^2 + 2\,\mathrm{Jr}\,\psi'[t]\,\theta'[t] + \mathrm{Jc}\,\theta'[t]^2 + \mathrm{Jr}\,\theta'[t]^2 + 2\,\mathrm{Jw}\,\theta'[t]^2 + 2\,l^2\,\mathrm{mw}\,\theta'[t]^2 + \mathrm{mc}\,x'[t]^2 + \mathrm{mr}\,x'[t]^2 + 2\,\mathrm{mw}\,x'[t]^2 + \mathrm{mc}\,y'[t]^2 + \mathrm{mr}\,y'[t]^2 + 2\,\mathrm{mw}\,y'[t]^2)$$

We can compute the components to the kinetric energy metric.

In[24]:= **metric = Simplify[KE2Metric[KE, vel]]**

Out[24]= $\{\{\mathrm{mc} + \mathrm{mr} + 2\,\mathrm{mw}, 0, 0, 0, 0\}, \{0, \mathrm{mc} + \mathrm{mr} + 2\,\mathrm{mw}, 0, 0, 0\}, \{0, 0, \mathrm{Jc} + \mathrm{Jr} + 2\,(\mathrm{Jw} + l^2\,\mathrm{mw}), \mathrm{Jr}, 0\}, \{0, 0, \mathrm{Jr}, \mathrm{Jr}, 0\}, \{0, 0, 0, 0, 2\,\mathrm{Jw}\}\}$

We can also compute the Christoffel symbols for the associated Levi-Civita affine connection.

In[25]:= **lcgamma = LeviCivita[metric, conf]**
Out[25]= {{{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0},
{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}},
{{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0},
{0, 0, 0, 0, 0}}, {{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0},
{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}},
{{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0},
{0, 0, 0, 0, 0}}, {{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0},
{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}}}

S5.4.4 Forces

For the snakeboard there are only control forces, and there are two of these. First we consider the force that spins the rotor. The corresponding Newtonian force is applied to the rotor and the coupler, and so the Lagrangian force will be a sum of these two Newtonian contributions. First we make a list of the Newtonian forces. In this case, there are two of these (one each for the coupler and the rotor), and they are both zero.

In[26]:= **force1 = {{0, 0, 0}, {0, 0, 0}}**
Out[26]= {{0, 0, 0}, {0, 0, 0}}

We now do the same for torques.

In[27]:= **torque1 = {{0, 0, −1}, {0, 0, 1}}**
Out[27]= {{0, 0, −1}, {0, 0, 1}}

Next we create the Lagrangian force.

In[28]:= **F1 = Simplify[Force[torque1, force1, {Rc, Rr}, {rc, rr}, conf, t]]**
Out[28]= {0, 0, 0, 1, 0}

Let us do the same for the other control force that moves the wheels.

In[29]:= **force2 = {{0, 0, 0}, {0, 0, 0}}**
Out[29]= {{0, 0, 0}, {0, 0, 0}}

In[30]:= **torque2 = {{0, 0, 1/2}, {0, 0, −1/2}}**
Out[30]= $\{\{0, 0, \frac{1}{2}\}, \{0, 0, -\frac{1}{2}\}\}$

In[31]:= **F2 = Simplify[Force[torque2, force2, {Rb, Rf}, {rb, rf}, conf, t]]**
Out[31]= {0, 0, 0, 0, 1}

S5.4.5 The constrained connection

The explicit expressions here are a little outrageous, so we suppress much of the output. The first step is to compute an orthogonal basis of vector fields for which the first vector fields in the list are a basis for the constraint distribution. This is generally not possible. For example, for the snakeboard, the constraint distribution does not have constant rank. Even when the constraint distribution *does* have constant rank, it is not always possible to find a global basis. We refer to the text for details surround such discussions. Here we merely note that, provided one omits consideration of the configurations where the constraint distribution gains rank, it is possible to find a global basis of vector fields for the snakeboard constraint distribution. First we provide a set of covector fields that annihilate the constraint distribution.

```
In[32]:= omega1 =
           {- Sin[φ[t] + θ[t]], Cos[φ[t] + θ[t]], -l Cos[φ[t]], 0, 0}
Out[32]= {- sin[φ[t] + θ[t]], cos[φ[t] + θ[t]], -l cos[φ[t]], 0, 0}
```

```
In[33]:= omega2 =
           {Sin[φ[t] - θ[t]], Cos[φ[t] - θ[t]], l Cos[φ[t]], 0, 0}
Out[33]= {sin[φ[t] - θ[t]], cos[φ[t] - θ[t]], l cos[φ[t]], 0, 0}
```

By "sharping" these relative to the kinetic energy metric, we get two vector fields that are $\mathbb{G}$-orthogonal to the constraint distribution. Then we need to ensure that these are $\mathbb{G}$-orthogonal.

```
In[34]:= X4 = RiemannSharp[omega1, metric];
```

```
In[35]:= X5t = RiemannSharp[omega2, metric];
```

The next equation is just the Gram–Schmidt Procedure.

```
In[36]:= X5 = X5t - (X4.metric.X5t)X4/(X4.metric.X4);
```

Now we define the three vector fields used in the text as a $\mathbb{G}$-orthogonal basis for the constraint distribution.

```
In[37]:= V1 = {Cos[θ[t]], Sin[θ[t]], 0, 0, 0}
Out[37]= {cos[θ[t]], sin[θ[t]], 0, 0, 0}
```

```
In[38]:= X1 = l Cos[φ[t]] V1 - Sin[φ[t]] {0, 0, 1, 0, 0}
Out[38]= {l cos[φ[t]] cos[θ[t]], l cos[φ[t]] sin[θ[t]], - sin[φ[t]], 0, 0}
```

```
In[39]:= X2t = {0, 0, 0, 1, 0}
Out[39]= {0, 0, 0, 1, 0}
```

```
In[40]:= X2 = Simplify[((X1.metric.X1)X2t
           -(X1.metric.X2t)X1)/((mc + mr + 2mw)l^2
             Cos[φ[t]]^2 + (Jc + Jr + 2(Jw + mw l^2)) Sin[φ[t]]^2)];
```

```
In[41]:= X3 = {0, 0, 0, 0, 1}
```

```
Out[41]= {0, 0, 0, 0, 1}

In[42]:= X = {X1, X2, X3, X4, X5};
```

Let us save the basis for $\mathcal{D}$ for later use.

```
In[43]:= Ddim = 3;

In[44]:= Dbasis = Table[X[[i]], {i, Ddim}];
```

One can check that these vector fields are indeed $\mathbb{G}$-orthogonal.

```
In[45]:= Simplify[
           Table[(X[[i]].metric.X[[j]])/(X[[i]].metric.X[[i]]), {i, 5}, {j, 5}]]
Out[45]= {{1, 0, 0, 0, 0}, {0, 1, 0, 0, 0}, {0, 0, 1, 0, 0}, {0, 0, 0, 1, 0}, {0, 0, 0, 0, 1}}
```

One can now determine the components of the orthogonal projection onto $\mathcal{D}^{\perp}$. To do this, we first compute the orthogonal projection onto $\mathcal{D}$.

```
In[46]:= P = OrthogonalProjection[{X1/Sqrt[X1.metric.X1],
             X2/Sqrt[X2.metric.X2], X3/Sqrt[X3.metric.X3]}, metric];
```

Since we have suppressed the somewhat lengthy explicit expression, let us at least verify that it is what it is supposed to be.

```
In[47]:= Table[Simplify[P.X[[i]] − X[[i]]], {i, 3}]
Out[47]= {{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}}

In[48]:= Table[Simplify[P.X[[i]]], {i, 4, 5}]
Out[48]= {{0, 0, 0, 0, 0}, {0, 0, 0, 0, 0}}
```

Now we define the projection onto $\mathcal{D}^{\perp}$.

```
In[49]:= Pperp = IdentityMatrix[5] − P;
```

Were one to examine the components of $P_{\mathcal{D}}$, one would find that each of them has as a denominator the following expression:

```
In[50]:= den = (Jc + Jr + 2Jw + l^2 mc + l^2 mr + 4l^2 mw−
                   (Jc + Jr + 2Jw − l^2 mc − l^2 mr) Cos[2ϕ[t]])^2
             (Jc + 2 Jw + l^2mc + l^2 mr + 4l^2 mw+
                 (−Jc − 2Jw + l^2(mc + mr)) Cos[2ϕ[t]])
```

$$\text{Out[50]= } (\mathrm{Jc} + \mathrm{Jr} + 2\,\mathrm{Jw} + l^2\,\mathrm{mc} + l^2\,\mathrm{mr} + 4\,l^2\,\mathrm{mw}$$
$$-(\mathrm{Jc} + \mathrm{Jr} + 2\,\mathrm{Jw} - l^2\,\mathrm{mc} - l^2\,\mathrm{mr})\ \cos[2\,\phi[t]])\hat{\ }2$$
$$(\mathrm{Jc} + 2\,\mathrm{Jw} + l^2\,\mathrm{mc} + l^2\,\mathrm{mr} + 4\,l^2\,\mathrm{mw}$$
$$+(-\mathrm{Jc} - 2\,\mathrm{Jw} + l^2\,(\mathrm{mc} + \mathrm{mr}))\ \cos[2\,\phi[t]])$$

Since we will be covariantly differentiating $P_{\mathcal{D}}^{\perp}$, the computations would simplify if we could get rid of this denominator (no messy quotient rule computations). It is not obvious how this can be done. The key is the following fact.

If A is an arbitrary invertible $(1,1)$-tensor field, then we may define the affine connection $\overset{A}{\nabla}$ on Q by

$$\overset{A}{\nabla}_X Y = \overset{\mathbb{G}}{\nabla}_X Y + A^{-1}(\overset{\mathbb{G}}{\nabla}_X (AP^{\perp}_{\mathcal{D}}))(Y).$$

It is easy to verify that this affine connection restricts to $\mathcal{D}$, and that its restriction to $\mathcal{D}$ agrees with $\overset{\mathcal{D}}{\nabla}$ (see [Lewis 2000]). We apply this formula with A being the identity tensor multiplied by the denominator. Since one covariantly differentiates $A \circ P^{\perp}_{\mathcal{D}}$ when computing $\overset{A}{\nabla}$, the denominator disappears, and the differentiations simplify.

```
In[51]:= cgamma = ConstrainedConnection[
              lcgamma, den IdentityMatrix[5], Pperp, conf];
```

Now, having computed the Christoffel symbols for the constrained connection, we will do nothing with them symbolically. They are just too unwieldy.

S5.4.6 The data for the orthogonal Poincaré representation

In this section we illustrate the use of the orthogonal Poincaré representation. Fortunately, we have already done much of the work, namely the computation of a $\mathbb{G}$-orthogonal basis for the constraint distribution. Therefore, we can directly compute the 3^3 Christoffel symbols that appear in the orthogonal Poincaré representation.

```
In[52]:= ogamma = Simplify[OrthogonalChristoffelSymbols[
              Dbasis, metric, lcgamma, conf]];
```

While we suppress the output of Mathematica® here, the explicit expressions are not that bad, and indeed are produced in the text in Section 13.4.1.

It is possible to use the generalized covariant derivative to perform controllability computations. Let us, for example, check that the snakeboard is STLC from points with zero initial velocity. In the text it is shown that the system is KC, and therefore STLCC. To show that the system is STLC, we need only show that $\mathrm{Sym}^{(1)}(\mathscr{Y}) = \mathcal{D}$ (cf. Theorem 8.9). First we need the components of the vector forces relative to the $\mathbb{G}$-orthogonal basis.

```
In[53]:= Y1o = OrthogonalForce[F1, Dbasis, metric];
```

```
In[54]:= Y2o = OrthogonalForce[F2, Dbasis, metric]
```

Out[54]= $\left\{0, 0, \frac{1}{2\,\mathrm{Jw}}\right\}$

Now we compute the symmetric products.

```
In[55]:= Y1symY1o = 2Simplify[GeneralizedCovariantDerivative[
              Y1o, Y1o, Dbasis, ogamma, conf]]
```

Out[55]= $\{0, 0, 0\}$

In[56]:= $\textbf{Y2symY2o} = \textbf{2Simplify[GeneralizedCovariantDerivative[}$
$\textbf{Y2o, Y2o, Dbasis, ogamma, conf]]}$

Out[56]= $\{0, 0, 0\}$

In[57]:= $\textbf{Y1symY2o} = \textbf{GeneralizedCovariantDerivative[}$
$\textbf{Y1o, Y2o, Dbasis, ogamma, conf]}$
$\textbf{+ GeneralizedCovariantDerivative[}$
$\textbf{Y2o, Y1o, Dbasis, ogamma, conf];}$

The vanishing of $\langle Y_1 : Y_1\rangle$ and $\langle Y_2 : Y_2\rangle$ gives the kinematic controllability of the system, as described in the text.

We now see that $\{Y_1, Y_2, \langle Y_1 : Y_2\rangle\}$ generate $\mathcal{D}$ at all points where $\{X_1, X_2, X_3\}$ generate the constraint distribution (i.e., at points where $\cos\phi \neq 0$).

In[58]:= **Simplify[Det[{Y1o, Y2o, Y1symY2o}]]**

Out[58]= $(l^2\ (\mathrm{mc} + \mathrm{mr} + 2\ \mathrm{mw})\ \cos[\phi[t]])$
$\big/\big(\mathrm{Jr}\ \mathrm{Jw}^2\ (\mathrm{Jc} + 2\ \mathrm{Jw} + l^2\ \mathrm{mc} + l^2\ \mathrm{mr} + 4\ l^2\ \mathrm{mw}$
$-(\mathrm{Jc} + 2\ \mathrm{Jw} - l^2\ (\mathrm{mc} + \mathrm{mr}))\ \cos[2\ \phi[t]])^2\big)$

S5.4.7 Affine connection control system equations

We will compute equations of motion in two different ways. First we use the Christoffel symbols for the constrained connection as above, and just produce the full geodesic equations. First we need to give the input vector fields, properly projected onto the constraint distribution.

In[59]:= **Y1c = P.RiemannSharp[F1, metric];**

In[60]:= **Y2c = P.RiemannSharp[F2, metric]**

Out[60]= $\left\{0, 0, 0, 0, \frac{1}{2\ \mathrm{Jw}}\right\}$

The total input is a linear combination of the two inputs, with the coefficients being the controls. Let us leave the controls as general for the moment.

In[61]:= **Yc = u1 Y1c + u2 Y2c;**

Now we can produce the equations of motion.

In[62]:= **eqmot1 = ACCSequations[cgamma, Yc, conf, t];**

S5.4.8 Poincaré equations

First one defines the "full" set of pseudo-velocities. In this case there are three.

In[63]:= **pv = {pv1[t], pv2[t], pv3[t]}**
Out[63]= $\{\mathrm{pv1}[t], \mathrm{pv2}[t], \mathrm{pv3}[t]\}$

Then one extracts the state for the equations, properly taking into account that some of the pseudo-velocities are velocities.

In[64]:= **state = GetState[Dbasis, pv, conf, t]**
Out[64]= $\{x[t], y[t], \theta[t], \psi[t], \phi[t], \mathrm{pv1}[t]\}$

Note that the second and third pseudo-velocities are actual velocities, and so do not appear in the list of states.

Next we give the vector force.

In[65]:= **Yo = u1 Y1o + u2 Y2o;**

Now we can formulate the equations of motion.

In[66]:= **eqmot2 = SMCSequations[ogamma, Yo, Dbasis, pv, conf, t];**

S5.4.9 Simulation

First let us give numerical values to the parameters.

In[67]:= **params = {Jc → 1/2, Jw → 1/8, Jr → 3/4,
mc → 1/2, mr → 3/4, mw → 1/4, l → 1/2}**
Out[67]= $\left\{\mathrm{Jc} \to \frac{1}{2}, \mathrm{Jw} \to \frac{1}{8}, \mathrm{Jr} \to \frac{3}{4}, \mathrm{mc} \to \frac{1}{2}, \mathrm{mr} \to \frac{3}{4}, \mathrm{mw} \to \frac{1}{4}, l \to \frac{1}{2}\right\}$

Now define specific controls.

In[68]:= **u1 = 2 Sin[3t]**
Out[68]= $2\ \sin[3\,t]$

In[69]:= **u2 = 2 Sin[2t]**
Out[69]= $2\ \sin[2\,t]$

Next define the initial and final times for the simulation.

In[70]:= **Ti = 0**
Out[70]= 0

In[71]:= **Tf = 3π**
Out[71]= $3\,\pi$

Now the initial conditions.

In[72]:= **qinit = vinit = {0, 0, 0, 0, 0}**

Out[72]= $\{0, 0, 0, 0, 0\}$

Now simulate.

In[73]:= **sol1 = ACCSsimulate[(eqmot1/.params), conf, qinit, vinit, t, Ti, Tf]**

Out[73]= $\{\{x[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$y[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\theta[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\psi[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\phi[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t]\}\}$

In[74]:= **Plot[x[t]/.sol1, {t, Ti, Tf}]**

Out[74]= -Graphics-

Now we simulate the system as a Poincaré representation. First we need to compute the initial pseudo-velocities.

In[75]:= **pvinit = Table[vinit.metric.X[[i]]/(X[[i]].metric.X[[i]]), {i, Ddim}]**

Out[75]= $\{0, 0, 0\}$

Now simulate.

In[76]:= **sol2 = SMCSsimulate[(eqmot2/.params), Dbasis, conf, pv, qinit, pvinit, t, Ti, Tf]**

Out[76]= $\{\{x[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$y[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\theta[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\psi[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\phi[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t],$
$\mathrm{pv1}[t] \to \mathrm{InterpolatingFunction}[\{\{0., 9.42478\}\}, <>][t]\}\}$

In[77]:= **Plot[x[t]/.sol2, {t, Ti, Tf}]**

Out[77]= -Graphics-

The two solution methods give the same solutions to the differential equation, as expected.

References

Abraham, R. and Marsden, J. E. [1978] *Foundations of Mechanics*, second edition, Addison Wesley, Reading, MA, ISBN 0-8053-0102-X.

Abraham, R., Marsden, J. E., and Ratiu, T. S. [1988] *Manifolds, Tensor Analysis, and Applications*, second edition, number 75 in Applied Mathematical Sciences, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-96790-7.

Agrachev, A. A. [1990] *Quadratic mappings in geometric control theory*, Journal of Soviet Mathematics, **51**(6), 2667–2734.

Agrachev, A. A. and Sachkov, Y. [2004] *Control Theory from the Geometric Viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 3-540-21019-9.

Arnol'd, V. I. [1978] *Mathematical Methods of Classical Mechanics*, first edition, number 60 in Graduate Texts in Mathematics, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-90314-3, second edition: [Arnol'd 1989].

— [1989] *Mathematical Methods of Classical Mechanics*, second edition, number 60 in Graduate Texts in Mathematics, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-96890-3.

Bellman, R. [1957] *Dynamic Programming*, Princeton University Press, Princeton, NJ, reprint: [Bellman 2003].

— [2003] *Dynamic Programming*, Dover Publications, Inc., New York, ISBN 0-486-42809-5, reprint of 1972 sixth edition.

Bloch, A. M. [2003] *Nonholonomic Mechanics and Control*, volume 24 of *Interdisciplinary Applied Mathematics*, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-095535-6.

Bloch, A. M. and Crouch, P. E. [1992] *Kinematics and dynamics of nonholonomic control systems on Riemannian manifolds*, in *Proceedings of the 32nd IEEE Conference on Decision and Control*, pages 1–5, Tucson, AZ.

Bloch, A. M., Leonard, N. E., and Marsden, J. E. [2001] *Controlled Lagrangians and the stabilization of Euler–Poincaré mechanical systems*, International Journal of Robust and Nonlinear Control, **11**(3), 191–214.

Bonnard, B. and Chyba, M. [2003] *Singular Trajectories and their Role in Control Theory*, number 40 in Mathematics & Applications, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 3-540-00838-1.

Bryant, R. L., Chern, S. S., Gardner, R. B., Goldschmidt, H. L., and Griffiths, P. A. [1991] *Exterior Differential Systems*, number 18 in Mathematical Sciences Research Institute Publications, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-97411-3.

Bullo, F. [2000] *Stabilization of relative equilibria for underactuated systems on Riemannian manifolds*, Automatica. The Journal of IFAC, **36**(12), 1819–1834.

Bullo, F., Leonard, N. E., and Lewis, A. D. [2000] *Controllability and motion algorithms for underactuated Lagrangian systems on Lie groups*, IEEE Transactions on Automatic Control, **45**(8), 1437–1454.

Chyba, M., Leonard, N. E., and Sontag, E. D. [2003] *Singular trajectories in multi-input time-optimal problems: Application to controlled mechanical systems*, Journal of Dynamical and Control Systems, **9**(1), 103–129.

Coombs, A. T. [2000] *Time-Optimal Control for Two Simple Mechanical Systems with Three Degrees of Freedom and Two Inputs*, Master's thesis, Queen's University, Kingston, ON.

Cortés, J., Martínez, S., and Bullo, F. [2003] *On nonlinear controllability and series expansions for Lagrangian systems with dissipative forces*, IEEE Transactions on Automatic Control, **47**(8), 1396–1401.

Cortés, J., Martínez, S., Ostrowski, J. P., and Zhang, H. [2002] *Simple mechanical control systems with constraints and symmetry*, SIAM Journal on Control and Optimization, **41**(3), 851–874.

Crampin, M. [1983] *Tangent bundle geometry for Lagrangian dynamics*, Journal of Physics. A. Mathematical and General, **16**(16), 3755–3772.

Crampin, M. and Pirani, F. A. E. [1986] *Applicable Differential Geometry*, number 59 in London Mathematical Society Lecture Note Series, Cambridge University Press, New York–Port Chester–Melbourne–Sydney, ISBN 0-521-23190-6.

Crouch, P. E. and Silva Leite, F. [1991] *Geometry and the dynamic interpolation problem*, in *Proceedings of the 1991 American Control Conference*, pages 1131–1136, Boston, MA.

Dunn, J. C. [1967] *On the classification of singular and nonsingular extremals for the Pontryagin maximum principle*, Journal of Mathematical Analysis and Applications, **17**, 1–36.

Flanders, H. [1989] *Differential Forms with Applications to the Physical Sciences*, Dover Publications, Inc., New York, ISBN 0-486-66169-5, reprint of 1963 edition by Academic Press.

Giaquinta, M. and Hildebrandt, S. [1996] *Calculus of Variations*, number 310 and 311 in Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 3-540-50625-X and 3-540-57961-3.

Guillemin, V. and Sternberg, S. [1990] *Symplectic Techniques in Physics*, second edition, Cambridge University Press, New York–Port Chester–Melbourne–Sydney, ISBN 0-521-38990-9.

Jalnapurkar, S. M. and Marsden, J. E. [1998] *Stabilization of relative equilibria. II*, Regular & Chaotic Dynamics, **3**(3), 161–179.

— [2000] *Stabilization of relative equilibria*, IEEE Transactions on Automatic Control, **45**(8), 1483–1491.

Jurdjevic, V. [1997] *Geometric Control Theory*, number 51 in Cambridge Studies in Advanced Mathematics, Cambridge University Press, New York–Port Chester–Melbourne–Sydney, ISBN 0-521-49502-4.

Kalman, R. E. [1960] *Contributions to the theory of optimal control*, Boletín de la Sociedad Matemática Mexicana. Segunda Serie, **5**, 102–119.

Kobayashi, S. and Nomizu, K. [1963] *Foundations of Differential Geometry, Volumes I and II*, number 15 in Interscience Tracts in Pure and Applied Mathematics, Interscience Publishers, New York, ISBN 0-470-49647-9 and 0-470-49648-7.

Kolář, I., Michor, P. W., and Slovák, J. [1993] *Natural Operations in Differential Geometry*, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-56235-4.

Krener, A. J. [1977] *The high order maximum principle and its application to singular extremals*, SIAM Journal on Control and Optimization, **15**(2), 256–293.

Lee, E. B. and Markus, L. [1967] *Foundations of Optimal Control Theory*, John Wiley and Sons, New York.

Lewis, A. D. [2000] *Simple mechanical control systems with constraints*, IEEE Transactions on Automatic Control, **45**(8), 1420–1436.

Lewis, A. D. and Murray, R. M. [1997] *Configuration controllability of simple mechanical control systems*, SIAM Journal on Control and Optimization, **35**(3), 766–790.

Lewis, A. D. and Tyner, D. R. [2003] *Jacobian linearisation in a geometric setting*, in *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 6084–6089, Maui, HI.

Libermann, P. and Marle, C.-M. [1987] *Symplectic Geometry and Analytical Mechanics*, number 35 in Mathematics and its Applications, D. Reidel Publishing Company, Dordrecht–Boston–Lancaster–Tokyo, ISBN 90-2772-438-5.

Liu, W. S. and Sussmann, H. J. [1994] *Abnormal sub-Riemannian minimizers*, in *Differential Equations, Dynamical Systems, and Control Science*, volume 152 of *Lecture Notes in Pure and Applied Mathematics*, pages 705–716, Dekker Marcel Dekker, New York.

Marsden, J. E. [1992] *Lectures on Mechanics*, number 174 in London Mathematical Society Lecture Note Series, Cambridge University Press, New York–Port Chester–Melbourne–Sydney, ISBN 0-521-42844-0.

Montgomery, R. [1994] *Abnormal minimizers*, SIAM Journal on Control and Optimization, **32**(6), 1605–1620.

— [2002] *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, number 91 in Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, ISBN 0-8218-1391-9.

Nelson, E. [1967] *Tensor Analysis*, Princeton University Press, Princeton, NJ.

Nijmeijer, H. and van der Schaft, A. J. [1990] *Nonlinear Dynamical Control Systems*, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-97234-X.

Noakes, L., Heinzinger, G., and Paden, B. [1989] *Cubic splines on curved spaces*, IMA Journal of Mathematical Control and Information, **6**(4), 465–473.

Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mishchenko, E. F. [1961] *Matematicheskaya teoriya optimal′ nykh protsessov*, Gosudarstvennoe izdatelstvo fiziko-matematicheskoi literatury, Moscow, reprint of translation: [Pontryagin, Boltyanskii, Gamkrelidze, and Mishchenko 1986].

— [1986] *The Mathematical Theory of Optimal Processes*, Classics of Soviet Mathematics, Gordon & Breach Science Publishers, New York, ISBN 2-88124-134-4, reprint of 1962 translation from the Russian by K. N. Trirogoff.

Sasaki, S. [1958] *On the differential geometry of tangent bundles of Riemannian manifolds*, The Tôhoku Mathematical Journal. Second Series, **10**, 338–354.

— [1962] *On the differential geometry of tangent bundles of Riemannian manifolds II*, The Tôhoku Mathematical Journal. Second Series, **14**, 146–155.

Shen, J. [2002] *Nonlinear Control of Multibody Systems with Symmetries via Shape Change*, Ph.D. thesis, University of Michigan, Ann Arbor, MI.

Silva Leite, F., Camarinha, M., and Crouch, P. E. [2000] *Elastic curves as solutions of Riemannian and sub-Riemannian control problems*, Mathematics of Control, Signals, and Systems, **13**(2), 140–155.

Sussmann, H. J. [1997] *An introduction to the coordinate-free maximum principle*, in *Geometry of Feedback and Optimal Control*, B. Jakubczyk and W. Respondek, editors, pages 463–557, Dekker Marcel Dekker, New York.

— [2000] *New theories of set-valued differentials and new versions of the maximum principle of optimal control theory*, in *Nonlinear Control in the Year 2000*, A. Isidori, F. Lamnabhi-Lagarrigue, and W. Respondek, editors, number 259 in Lecture Notes in Control and Information Sciences, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 1-85233-363-4.

Sussmann, H. J. and Willems, J. C. [1997] *300 years of optimal control: From the brachystochrone to the maximum principle*, IEEE Control Systems Magazine.

Thompson, G. and Schwardmann, U. [1991] *Almost tangent and cotangent structures in the large*, Transactions of the American Mathematical Society, **327**(1), 313–328.

Tulczyjew, W. M. [1977] *The Legendre transformation*, Annales de l'Institut Henri Poincarè. Section A. Physique Thèorique. Nouvelle Sèrie, **27**(1), 101–114.

Vorotnikov, V. I. [1998] *Partial Stability and Control*, Birkhäuser, Boston–Basel–Stuttgart, ISBN 0-8176-3917-9.

Wonham, W. M. [1985] *Linear Multivariable Control: A Geometric Approach*, third edition, number 10 in Applications of Mathematics, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-96071-6.

Woolsey, C. A. and Leonard, N. E. [2004] *Stabilizing underwater vehicle motion using internal rotors*, Automatica. The Journal of IFAC, **38**(12), 2053–2062.

Yano, K. and Ishihara, S. [1973] *Tangent and Cotangent Bundles*, number 16 in Pure and Applied Mathematics, Dekker Marcel Dekker, New York, ISBN 0-8247-6071-9.

Zenkov, D. V., Bloch, A. M., and Marsden, J. E. [2002] *The Lyapunov-Malkin theorem and stabilization of the unicycle with rider*, Systems & Control Letters, **45**(4), 293–302.

Symbol index