Journal of Geometric Mechanics
2(4), 397–440, 2010
DOI: 10.3934/jgm.2010.2.397

Geometric Jacobian linearization and LQR theory

David R. Tyner* Andrew D. Lewis†

2010/12/20
Last updated: 2011/01/07

Abstract

The procedure of linearizing a control-affine system along a non-trivial reference trajectory is studied from a differential geometric perspective. A coordinate-invariant setting for linearization is presented. With the linearization in hand, the controllability of the geometric linearization is characterized using an alternative version of the usual controllability test for time-varying linear systems. The various types of stability are defined using a metric on the fibers along the reference trajectory and Lyapunov's second method is recast for linear vector fields on tangent bundles. With the necessary background stated in a geometric framework, linear quadratic regulator theory is understood from the perspective of the Maximum Principle. Finally, the resulting feedback from solving the infinite time optimal control problem is shown to uniformly asymptotically stabilize the linearization using Lyapunov's second method.

Keywords. linearization, linear controllability, linear quadratic optimal control, stability, stabilization

AMS Subject Classifications (2020). 37C10, 49K05, 93B05, 93B18, 93B27, 93B52.

1. Introduction and background

Jacobian linearization is a standard concept in control theory and is used to study controllability, stability, and stabilization of non-linear systems. Indeed, Jacobian linearization provides the setting for a significant number of the control algorithms implemented in practice for non-linear systems.

In this paper, the abstract setting of "affine systems" of is used to develop a geometric theory of linearization for control-affine systems evolving on a differentiable manifold. The objective is not so much to broaden the applicability of linearization techniques, but to better understand the structure of linearization and to make explicit some of the choices that are made without mention in the standard practice of linearization. The motivation, in part, comes from examples in mechanics. Given an affine connection, what it means to

*Mechanical and Aerospace Engineering, Carleton University, Ottawa, ON, Canada, K1S 5B6
Email: drtyner@gmail.com
Work performed while a graduate student at Queen's University.
†Professor, Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada
Email: andrew.lewis@queensu.ca, URL: http://www.mast.queensu.ca/~andrew/
Research supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada and by an Early Researcher Award from the Province of Ontario

linearize about a reference trajectory has a natural geometric interpretation provided by the Jacobi equation of geodesic variations. In the general setup of control-affine systems, a geometric setup is thus far not found in the literature. However, certain ideas presented here are implicit in the paper of Sussmann [1998], although the presented geometric framework is less abstract and so has more structure.

1.1. Linear systems and quadratic optimal control. In order to provide a point of reference for our geometric formulation of control systems and their linearizations, this section outlines the standard manner in which linearization and stabilization is normally carried out for control-affine systems on $\mathbb{R}^n$. This standard strategy is, of course, correct but it "sweeps under the rug" various issues, listed in Section 1.2, that must be addressed to develop a geometric theory.

Let $\Omega \subseteq \mathbb{R}^n$ be an open subset and let $f_0, f_1, \dots, f_m$ be smooth vector fields, possibly depending measurably on t, on Ω. Consider a control system with governing equations

$$\gamma'(t) = f_0(t, \gamma(t)) + \sum_{a=1}^{m} u^a(t) f_a(t, \gamma(t)), \tag{1.1}$$

where $\gamma\colon I \to \Omega$ is locally absolutely continuous and $u\colon I \to \mathbb{R}^m$ is bounded and measurable for some interval $I \subseteq \mathbb{R}$. For the purposes of linearization, fix a reference trajectory γ_{ref} corresponding to a reference control u_{ref}, both defined on $I \subseteq \mathbb{R}$. To define the linearization, for each $t \in I$ define $m+1$ smooth vector fields $f_{a,t}$, $a \in \{0, 1, \dots, m\}$, by $f_{a,t}(x) = f_a(t, x)$. The linearization of (1.1) along γ_{ref} is then defined by

$$\xi'(t) = A(t)\xi(t) + B(t)v(t), \tag{1.2}$$

where

$$\begin{aligned} A(t) &= \boldsymbol{D} f_{0,t}(\gamma_{\text{ref}}(t)) + \sum_{a=1}^{m} u^a_{\text{ref}}(t) \boldsymbol{D} f_{a,t}(\gamma_{\text{ref}}(t)), \\ B(t) &= \left[\ f_{1,t}(\gamma_{\text{ref}}(t)) \,\middle|\, \cdots \,\middle|\, f_{m,t}(\gamma_{\text{ref}}(t))\ \right]. \end{aligned}$$

Here $\boldsymbol{D} f_{a,t}$ denotes the Jacobian of the vector field $f_{a,t}$, $a \in \{0, 1, \dots, m\}$. Once the linearization (1.2) is obtained, its controllability properties can be investigated using the standard controllability Gramian (see Section 4.2 for restatements of the standard Gramian results). If the linearization (1.2) is ascertained to be controllable on $I = [0, \infty)$, then (1.1) can be locally stabilized along the reference trajectory by stabilizing the linearization (1.2) using linear feedback [Ikeda, Maeda, and Kodama 1972, Kalman 1960]. That is, if $\mathrm{L}(V; W)$ denotes the set of linear maps from a vector space V to a vector space W, a map $F\colon I \to \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ is chosen with the property that the closed-loop system

$$\xi'(t) = \big(A(t) + B(t)F(t)\big)\xi(t),$$

is uniformly asymptotically stable. If $F^a(t) \in (\mathbb{R}^n)^*$ is the ath row of $F(t)$, then the non-linear closed-loop system

$$\gamma'(t) = f_0(t, \gamma(t)) + \sum_{a=1}^{m} \left(u^a_{\text{ref}}(t) + F^a(t)(\gamma(t) - \gamma_{\text{ref}}(t))\right) f_a(t, \gamma(t)), \tag{1.3}$$

is locally uniformly asymptotically stable along the trajectory γ_{ref} [Vidyasagar 1993]. In practice one might design F through optimal control methods using a quadratic cost, the so-called linear quadratic regulator (**LQR**). We review this next.

Let $A\colon \mathbb{R} \to \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$ and $B\colon \mathbb{R} \to \mathrm{L}(\mathbb{R}^m;\mathbb{R}^n)$ be continuous maps and define a ***time-varying linear system on $\mathbb{R}^n$*** to be a pair (A, B) satisfying

$$x'(t) = A(t)x(t) + B(t)u(t). \tag{1.4}$$

The solution to (1.4) satisfying $x(t_0) = x_0$ for $t_0 \in I$ is given by the variations of constants formula,

$$x(t) = \Phi(t,t_0)x_0 + \int_{t_0}^{t} \Phi(t,\sigma)B(\sigma)u(\sigma)\mathrm{d}\sigma,$$

where $\Phi(t,t_0)$ is the state transition matrix. That is, $t \mapsto \Phi(t,t_0)$ is the solution to the homogeneous system $\Phi'(t,t_0) = A(t)\Phi(t,t_0)$ with initial condition $\Phi(t_0,t_0) = \mathrm{id}_{\mathbb{R}^n}$.

1.1 Problem: (Finite time LQR problem) For a time-varying linear system (A, B), find a pair $(x(t), u(t))$, satisfying the equation (1.4), defined on $I = [t_0, t_1]$ which minimizes the quadratic cost function

$$J(x(t_0), t_0, t_1) = \frac{1}{2}x^T(t_1)F(t_1)x(t_1) + \frac{1}{2}\int_{t_0}^{t_1} x^T(t)Q(t)x(t) + u^T(t)R(t)u(t)\,\mathrm{d}t,$$

where $F(t_1) \in L(\mathbb{R}^n;\mathbb{R}^n)$ is symmetric and positive-semidefinite, $Q\colon I \to L(\mathbb{R}^n;\mathbb{R}^n)$ is symmetric and positive-semidefinite for each $t \in I$, and $R\colon I \to L(\mathbb{R}^m;\mathbb{R}^m)$ is symmetric and positive-definite for each $t \in I$. •

Solutions to Problem 1.1 for a finite time t_1 can be obtained by variational methods or by applying the Maximum Principle [Athans and Falb 1966, Lee and Markus 1967]. The original presentation of the Maximum Principle is provided by Pontryagin, Boltyanskiĭ, Gamkrelidze, and Mishchenko [1961]. Using either method, the existence of a solution to Problem 1.1 is equivalent to the existence of a solution $K(t,t_1)$ to the ***differential Riccati equation***

$$-\frac{\mathrm{d}K}{\mathrm{d}t} = A^T K + KA - KBR^{-1}B^T K + Q, \qquad K(t_1,t_1) = F(t_1),$$

where the time dependence has been dropped for brevity. A solution to the Riccati equation then provides an optimal, in the sense of Problem 1.1, linear state feedback $u(t) = -R^{-1}(t)B^T(t)K(t,t_1)x(t)$ [Kalman 1960].

In the study of the infinite-time problem, the terminal cost $F(t_1)$ is considered to be zero.

1.2 Problem: (Infinite time LQR problem) For a time-varying linear system (A, B), find a pair $(x(t), u(t))$, satisfying the equation (1.4), defined on $I = [t_0, \infty)$ which minimizes the quadratic cost function

$$J(x(t_0), t_0, \infty) = \frac{1}{2}\int_{t_0}^{\infty} x^T(t)Q(t)x(t) + u^T(t)R(t)u(t)\,\mathrm{d}t,$$

where $Q\colon I \to L(\mathbb{R}^n;\mathbb{R}^n)$ is symmetric and positive-semidefinite for each $t \in I$, and $R\colon I \to L(\mathbb{R}^m;\mathbb{R}^m)$ is symmetric and positive-definite for each $t \in I$. •

If the time-varying linear system (A, B) is controllable, then solutions to the Riccati equation exist as $t_1 \to \infty$ and, similar to the finite time problem, the solution to the Riccati equation provides a linear feedback which is optimal in the sense of Problem 1.2 [Kalman 1960]. Furthermore, under suitable uniformity bounds on the time-varying linear system (A, B) and cost data, the uniform asymptotic stability of the closed-loop linear system follows from a Lyapunov argument. The mathematical details and proofs the standard time-varying case as the final time tends to infinity are found in Kalman [1960] and short survey of the standard case is provided in Tyner [2007].

1.2. Contribution and organization. This paper is a systematic investigation of Jacobian linearization and LQR theory from a differential geometric perspective. The above procedures by which (1.2), the linear system, is obtained and stabilized pose some problems when the Euclidean state space of the non-linear system (1.1) is replaced with a differentiable manifold. The main issues are outlined below and form the organizational basis of this paper.

1.3 Question: The two families of linear maps $\{A(t)\}$ and $\{B(t)\}$ are dependent on the choice coordinate chart since the Jacobian of the vector fields $f_{a,t}$, $a \in \{0, 1, \dots, m\}$, do not have a geometric analogue on a manifold. What replaces the Jacobian? •

To answer Question 1.3, time-dependent affine systems on M are introduced in Section 3 and serve as the base object on which our geometric theory is built. In Section 3.1, what it means to linearize an affine system along a reference trajectory is understood in terms of the tangent lift. In that sense, the tangent lift plays the role of the Jacobian in Section 1.1.

1.4 Question: Since the control system given by (1.1) has a state space that is an open subset of $\mathbb{R}^n$, there are several natural identifications that can be (unknowingly) made. The fact that the state space is naturally identified with each tangent space implies that (1.2) lives in a vector space. Where does the geometric version of (1.2) live? •

The geometry dictates that the linearization is an affine system on TM. Thus, in contrast to the standard case, an affine system and its linearization live on different manifolds. In Section 3.2, the complexity of the above development is seen to reduce significantly when the reference trajectory is chosen to be an equilibrium point.

1.5 Question: By virtue of (1.2) living in a vector space, its controllability can be checked using the controllability Gramian which makes use of the standard inner product and the coordinate-dependent family of maps $\{B(t)\}$. What does it mean for the geometric version of (1.2) to be controllable, and how can it be checked whether such a system has this property? •

In Section 4, Question 1.5 is addressed when the controllability of the linearization is considered. To begin, the standard controllability results for (1.2) are re-characterized in Section 4.2. These re-characterizations have the feature that they may be applied directly to the geometric setting of the linearization and this is done in Section 4.3.

1.6 Question: What does the geometric version of the linear quadratic regulator problem look like? What is the analogue of the Riccati equation? •

In Section 5.1 the geometric versions of both the finite and infinite time LQR problems are formulated for the linearization of an affine system. The bulk of this work involves characterizing solutions to the finite time LQR problem using the Maximum Principle. Given the geometric setup of the linearization along a reference trajectory as an affine system on TM, the regular Maximum Principle statements do not directly apply without reverting to working in a set of coordinates. Thus, a new Maximum Principle statement is provided by Theorem 5.2 and proved in Section 5.5. In Sections 5.2, 5.3, and 5.4, the key ingredients to prove the Maximum Principle are introduced. These include the variational and adjoint equations, needle variations, tangent cones, and, of course, the Hamiltonian. For readers familiar with the Hamiltonian in the standard setup, the Hamiltonian presented in Section 5.4 will look "different." However, it maintains the required maximization properties—see Lemmata 5.17 and 5.18—required to prove the Maximum Principle. In Section 5.6, the Maximum Principle is used, answering Question 1.6, to characterize solutions to the finite time LQR problem. In this characterization the geometric version of the Riccati equation is given.

In Section 5.7, the infinite time LQR problem is addressed. In particular, solutions to the Riccati equation are shown to exist as the final time in the LQR problem tends to infinity. To prove their existence, the geometric analogue of the classical minimum energy controller is developed. Finally, the trajectory corresponding to the solution of the Riccati equation, as the final time tends to infinity, is shown to be optimal in the sense of the infinite time LQR problem.

1.7 Question: Again, since the state space is naturally identified with each tangent space, the stability of both the non-linear system and its linearization are measured with respect to the standard Euclidean norm. What are the appropriate norms in a geometric setting? •

In Section 6, stability and stabilization by LQR methods of the linearization are formulated to complete the geometric picture of LQR theory. In Section 6.1, the stability definitions are provided for a fixed reference vector field X_{ref} and for linear vector fields over X_{ref}. (For example, the linearisation X_{ref}^T is a linear vector field over X_{ref}.) These definitions are made using both a metric on M and a metric on the fibres of TM over image(γ_{ref}). Such metrics are naturally induced by choosing a Riemannian metric, $\mathbb{G}$, on M. This answers Question 1.7 and contrasts with the standard setup of Section 1.1 where standard Euclidean metric on $\mathbb{R}^n$ is used for both the non-linear system and its linearization. It is noted that a metric $\mathbb{G}$ on the fibres over image(γ_{ref}), unlike the Euclidean norm, will in general be time-dependent. As a consequence, any stability definitions made in terms of $\mathbb{G}$, will be dependent on the choice of metric unless the state manifold is compact.

In Section 6.2, Lyapunov's direct method for linear vector fields on tangent bundles is introduced. As in the standard setup, the stability of the linear vector field is inferred from the properties of a Lyapunov candidate and its derivative along integral curves of the linear vector field. The derivative of the Lyapunov candidate along an integral curve is defined using the Lie derivative operator given by (2.1).

1.8 Question: What is a linear state feedback for the geometric version of (1.2)? •

In Section 6.3, Question 1.8 is addressed. After making geometric sense of the terms "linear state-feedback" and "closed-loop system," it is proved that the linearization of an affine system is uniformly asymptotically stabilized using the linear state-feedback provided by the infinite time LQR problem. The proof follows by showing that the solution to the Riccati equation, as the final time tends to infinity, is a suitable Lyapunov function.

1.9 Question: After stabilizing the linearisation, how can the stabilizing linear state feedback be implemented for the non-linear system? •

Finally, in Section 7, a rough answer to Question 1.9 is posed as future work.

2. Geometric constructions

The basic geometric notation follows that of [Abraham, Marsden, and Ratiu 1988]. Let M be an n-dimensional Hausdorff manifold with a C^∞ differentiable structure. The letter I will always denote an interval in $\mathbb{R}$. The set of class C^r functions on M is denoted by $C^r(M)$. The tangent bundle of M is denoted by $\tau_M\colon TM \to M$ and the cotangent bundle by $\pi_M\colon T^*M \to M$. If $\phi\colon M \to N$ is a differentiable map between manifolds, its derivative is denoted $T\phi\colon TM \to TN$. For a vector bundle $\pi\colon E \to M$, $\Gamma^r(E)$ denotes the sections of E that are of class C^r. The subbundle $VE \triangleq \ker(T\pi) \subseteq TE$ is the ***vertical bundle*** of E.

Let V and W be $\mathbb{R}$-vector spaces. The notation $\mathrm{L}(V;W)$ denotes the set of linear maps from V to W. The ***dual space*** to V is defined by $V^* = \mathrm{L}(V;\mathbb{R})$. For any nonempty set $U \subseteq V$, the ***annihilator*** of U is a subspace of V^* defined by

$$\mathrm{ann}(U) = \{\alpha \in V^* \mid \alpha(v) = 0,\ v \in U\}.$$

Similarly, for any nonempty set $S \subseteq V^*$, the ***coannihilator*** of S is a subspace of V defined by

$$\mathrm{coann}(S) = \{v \in V \mid \alpha(v) = 0,\ \alpha \in S\}.$$

For a bilinear map $T\colon V \times V \to \mathbb{R}$, the ***flat map*** $T^\flat\colon V \to V^*$ is defined by $\langle T^\flat(v); u\rangle = T(u,v)$ for all $u \in V$. If $T^\flat$ is invertible then its inverse, the ***sharp map***, is denoted by $T^\sharp\colon V^* \to V$.

2.1. Time-dependent objects on a manifold.

To define time-dependent vector fields on manifolds in a general way, following Sussmann [1998, §3] it is convenient to first introduce time-dependent functions (see also Aliprantis and Border [2006]). A ***Carathéodory function*** on M is a map $\phi\colon I \times M \to \mathbb{R}$ with the property that $\phi^t \triangleq \phi(t,\cdot)$ is continuous for each $t \in I$, and $\phi_x \triangleq \phi(\cdot,x)$ is Lebesgue measurable for each $x \in M$. A Carathéodory function ϕ is ***locally integrally bounded*** (**LIB**) if, for each compact subset $K \subseteq M$, there exists a positive locally integrable function $\psi_K\colon I \to \mathbb{R}$ such that $|\phi(t,x)| \leq \psi_K(t)$ for each $x \in K$. A Carathéodory function $\phi\colon I \times M \to \mathbb{R}$ is of ***class*** $\boldsymbol{C^r}$ if ϕ^t is of class C^r for each $t \in I$ and is ***locally integrally of class*** $\boldsymbol{C^r}$ (**LIC**^{r}) it is of class C^r and if $X_1 \cdots X_r \phi^t$ is LIB for all $t \in I$ and $X_1, \ldots, X_r \in \Gamma^\infty(TM)$.

A ***Carathéodory vector field*** on M is a map $X\colon I \times M \to TM$ with the property that $X(t,x) \in T_xM$ and with the property that the function $\alpha \cdot X\colon (t,x) \mapsto \alpha(x)\cdot X(t,x)$ is a Carathéodory function for each $\alpha \in \Gamma^\infty(T^*M)$. For a Carathéodory vector field X on M, denote by $X_t\colon M \to TM$ the map $X_t(x) = X(t,x)$. A Carathéodory vector field X on M is ***locally integrally of class*** $\boldsymbol{C^r}$ (**LIC**^{r}) if $\alpha \cdot X$ is LIC^r for every $\alpha \in \Gamma^\infty(T^*M)$. The set of LIC^r vector fields on M is denoted by LIC$^r(TM)$.

The classical theory of time-dependent vector fields with measurable time dependence gives the existence of locally absolutely continuous integral curves for LIC$^\infty$ vector fields [Sontag 1998, Appendix C]. An integral curve $\gamma\colon I \to M$ is ***locally absolutely***

continuous (**LAC**) if, for any $\phi \in C^\infty(M)$, the map $t \mapsto \phi \circ \gamma(t)$ is locally absolutely continuous. Let $\gamma'(t)$ denote the tangent vector to γ at $t \in I$, noting that this is defined for almost every $t \in I$. The flow of $X \in \mathrm{LIC}^\infty(TM)$ is denoted by $\Phi^X_{t_0,t}$ and the curve $\gamma\colon t \mapsto \Phi^X_{t_0,t}(x_0)$ is the integral curve for X with initial condition $\gamma(t_0) = x_0$.

Let $\gamma\colon I \to M$ be an LAC curve. A ***vector field along*** $\boldsymbol{\gamma}$ is a map $\xi\colon I \to TM$ with the property that $\xi(t) \in T_{\gamma(t)}M$. A vector field ξ along γ is ***locally absolutely continuous*** (**LAC**) if it is LAC as a curve in TM. A weaker notion than that of an LAC vector field along γ is that of a ***locally integrable*** (**LI**) vector field along γ, which is a vector field ξ along γ having the property that the function $t \mapsto \alpha(\gamma(t)) \cdot \xi(t)$ is locally integrable for every $\alpha \in \Gamma^\infty(T^*M)$.

Let $X \in \mathrm{LIC}^\infty(TM)$ and let $\gamma\colon I \to M$ be an integral curve for X. There is a naturally defined Lie derivative operator along γ that maps LAC sections of TM along γ to LI sections of TM along γ. This operator, denoted by $\mathscr{L}^{X,\gamma}$, is defined by

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

where $V \in \Gamma^1(TM)$ and V_γ is the LAC section of TM along γ defined by $V_\gamma(t) = V(\gamma(t))$. One easily verifies in coordinates that, for an LAC vector field ξ along γ, $\mathscr{L}^{X,\gamma}(\xi)$ is given in coordinates $(x^1, \dots, x^n)$ by

$$\mathscr{L}^{X,\gamma}(\xi)(t) = \Big(\frac{\mathrm{d}\xi^i}{\mathrm{d}t}(t) - \frac{\partial X^i_t}{\partial x^j}(\gamma(t))\xi^j(t)\Big)\frac{\partial}{\partial x^i}, \qquad \text{a.e. } t \in I, \tag{2.1}$$

where a summation over $i \in \{1, \dots, n\}$ is implied. The Lie differentiation of LAC vector fields along a curve will play an important role in future developments, particularly in Sections 2.2 and 4.3. The geometric details of Lie differentiation for vector fields that depend measurably on time are provided by Sussmann [1998, §4].

2.2. Tangent bundle geometry. The various ways to lift a vector field is a prominent geometric idea that arises frequently in future sections. These constructions are contained in [Yano and Ishihara 1973].

Let $\pi\colon E \to M$ be a vector bundle. An LIC^∞ vector field X on E is ***linear*** if, for each $t \in I$,

1. X_t is π-projectable (denote the resulting vector field on M by πX_t) and
2. X_t is a linear morphism of vector bundles relative to the following diagram:

$$\begin{array}{ccc} E & \xrightarrow{X_t} & TE \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle T\pi} \\ M & \xrightarrow[\pi X_t]{} & TM \end{array}$$

   That is, the induced mapping from $\pi^{-1}(x)$ to $T\pi^{-1}(\pi X_t(x))$ is a linear mapping of $\mathbb{R}$-vector spaces.

The flow of a linear vector field has the property that $\Phi^X_{t_0,t}|E_x\colon E_x \to E_{\Phi^{\pi X}_{t_0,t}}$ is a linear transformation.

A linear vector field on a vector bundle generalizes the notion of a time-varying differential equation in the following manner. Let V be a finite-dimensional $\mathbb{R}$-vector space and consider on V a linear differential equation

$$\xi'(t) = A(t)(\xi(t)),$$

where $A\colon \mathbb{R} \to \mathrm{L}(V;V)$ is locally integrable. Now define an LIC^∞ linear vector field on the trivial bundle $\mathrm{pr}_1\colon \mathbb{R}\times V \to \mathbb{R}$, where pr_1 is the projection onto the first factor, by $X_A(\tau, v) = ((\tau, v), (1, A(\tau)(v)))$. Here the projected vector field on the base space is simply $\pi X_A = \frac{\partial}{\partial\tau}$. This special case of a linear vector field has the feature that the vector bundle admits a natural global trivialization. The lack of this feature in general accounts for some of the additional complexity in our development.

Now consider the case when E is the tangent bundle of M. Let $X \in \Gamma^\infty(TM)$ and define the ***tangent lift*** of X as the vector field $X^T \in \Gamma^\infty(TTM)$ obtained by

$$X^T(v_x) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} T_x\Phi^X_{0,s}(v_x).$$

The definition of the tangent lift can be extended to time-varying vector fields as follows. For $X \in \mathrm{LIC}^\infty(TM)$, the ***tangent lift*** of X is the vector field $X^T \in \mathrm{LIC}^\infty(TTM)$ defined by $X^T(t, v_x) = X^T_t(v_x)$. In natural coordinates $(x^1, \dots, x^n, v^1, \dots, v^n)$ for TM, the vector field $X^T(t, v_x)$ is given by the coordinate expression

$$X^T(t, v_x) = X^i(t,x)\frac{\partial}{\partial x^i} + \frac{\partial X^i}{\partial x^j}(t,x)v^j\frac{\partial}{\partial v^i}. \tag{2.2}$$

To provide an interpretation of the tangent lift, let $\gamma\colon I \to M$ be an integral curve of $X \in \mathrm{LIC}^\infty(TM)$. A ***variation of X along γ*** is a map $\sigma\colon I\times J \to M$ satisfying

1. $J \subseteq \mathbb{R}$ is an interval for which $0 \in \mathrm{int}(J)$,
2. σ is continuous,
3. the map $I \ni t \mapsto \sigma_s(t) \triangleq \sigma(t,s) \in M$ is an integral curve for X for each $s \in J$,
4. the map $J \ni s \mapsto \sigma^t(s) \triangleq \sigma(t,s) \in M$ is LAC for each $t \in I$,
5. the map $I \ni t \mapsto \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \sigma^t(s) \in TM$ is LAC, and
6. $\sigma_0 = \gamma$.

Corresponding to a variation σ of X along γ is an LAC vector field V_σ along γ defined by

$$V_\sigma(t) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \sigma^t(s).$$

With this notation, the following result records some useful properties of the tangent lift.

2.1 Proposition: *Let $X\colon I\times M\to TM$ be an LIC^∞ vector field, let $v_{x_0}\in T_{x_0}M$, 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 Υ along γ satisfying $\Upsilon(t_0)=v_{x_0}$, the following statements are equivalent:*

(i) Υ is an integral curve for X^T;

(ii) there exists a variation σ of X along γ such that $V_\sigma=\Upsilon$;

(iii) $\mathscr{L}^{X,\gamma}(\Upsilon)=0$.

Proof: The equivalence of (i) and (ii) will follow from the more general Proposition 3.1 below. Thus only the equivalence of (i) and (iii) needs to be proved. This, however, follows directly from the coordinate expressions (2.1) and (2.2). ■

2.2 Corollary: *For $X\in\mathrm{LIC}^\infty(TM)$, X^T is a linear vector field on $\tau_M\colon TM\to M$ and $\pi X^T=X$.*

The cotangent version of X^T, used in Section 5, is defined in a similar manner. For $X\in\mathrm{LIC}^\infty(TM)$, the ***cotangent lift*** of X is the vector field $X^{T^*}\in\mathrm{LIC}^\infty(TT^*M)$ defined by

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

In natural coordinates $(x^1,\dots,x^n,p_1,\dots,p_n)$ for T^*M, the vector field $X^{T^*}(t,\alpha_x)$ is given by the coordinate expression

$$X^{T^*}(t,\alpha_x)=X^i(t,x)\frac{\partial}{\partial x^i}-\frac{\partial X^j}{\partial x^i}(t,x)p_j\frac{\partial}{\partial p_i}. \tag{2.3}$$

The LIC^∞ vector fields X^T and X^{T^*} define an LIC^∞ vector field $X^T\times X^{T^*}$ on $TM\times T^*M$ by

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

The Whitney sum $TM\oplus T^*M$ is an embedded submanifold of $TM\times T^*M$ with embedding $v_x\oplus\alpha_x\mapsto(v_x,\alpha_x)$. Since the LIC^∞ vector field $X^T\times X^{T^*}$ is tangent to $TM\oplus T^*M$, its restriction to $TM\oplus T^*M$, denoted by $X^T\oplus X^{T^*}$, is well-defined. An interesting joint property of X^T and X^{T^*} is that the LIC^∞ vector field $X^T\oplus X^{T^*}$ leaves invariant the function $v_x\oplus\alpha_x\mapsto v_x\cdot\alpha_x$ on $TM\oplus T^*M$.

Corresponding to $X\in\mathrm{LIC}^\infty(TM)$ there is also a natural vertical vector field $\mathrm{vlft}(X)$ on $\tau_M\colon TM\to M$ defined by

$$\mathrm{vlft}(X)(t,v_x)=\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(v_x+sX(t,x)).$$

In natural coordinates $(x^1,\dots,x^n,v^1,\dots,v^n)$ for TM, the coordinate expression for the vertical lift is

$$\mathrm{vlft}(X)(t,v_x)=X^i(t,x)\frac{\partial}{\partial v^i}.$$

3. Affine systems and their linearization

In this section time-dependent affine systems on M are introduced. In Section 3.1 the linearization of an affine system on M along a non-trivial reference trajectory is obtained using the tangent lift. The resulting linearization has the structure of an affine system on TM. In Section 3.2, it is seen that the complexity of the above development reduces significantly at an equilibrium point.

A ***time-dependent distribution*** on M is a subset $\mathcal{D} \subseteq \mathbb{R}\times TM$ with the property that, for each $x_0 \in M$, there exist a neighbourhood $\mathcal{N}$ and LIC$^\infty$ vector fields $\mathscr{X} = \{X_1, \dots, X_k\}$ on $\mathcal{N}$ such that

$$\mathcal{D}_{(t,x)} \triangleq \mathcal{D} \cap \big(\{t\} \times T_xM\big) = \Big\{ \sum_{j=1}^{k} u^j X_j(t,x) \ \Big|\ u \in \mathbb{R}^k \Big\}.$$

The vector fields $\mathscr{X}$ are called ***local generators*** for $\mathcal{D}$. A ***time-dependent affine subbundle*** on M is a subset $\mathcal{A} \subseteq \mathbb{R} \times TM$ with the property that, for each $x_0 \in M$, there exists a neighbourhood $\mathcal{N}$ and LIC$^\infty$ vector fields $\mathscr{X} = \{X_0, X_1, \dots, X_k\}$ on $\mathcal{N}$ such that

$$\mathcal{A}_{(t,x)} \triangleq \mathcal{A} \cap \big(\{t\} \times T_xM\big) = \Big\{ X_0(t,x) + \sum_{j=1}^{k} u^j X_j(t,x) \ \Big|\ u \in \mathbb{R}^k \Big\}.$$

The vector fields $\mathscr{X}$ are called ***local generators*** for $\mathcal{A}$. The ***linear part*** of a time-dependent affine subbundle is the time-dependent distribution $L(\mathcal{A})$ defined by $L(\mathcal{A})_{(t,x)}$ being the subspace of T_xM upon which the affine subspace $\mathcal{A}_{(t,x)}$ is modelled. If $\mathscr{X}$ are local generators for $\mathcal{A}$ as above, then the vector fields $\{X_1, \dots, X_k\}$ are ***local linear generators*** for $L(\mathcal{A})$. The next step is to define an "affine system" in $\mathcal{A}$ to be an assignment to each $(t,x) \in \mathbb{R} \times M$ a subset $\mathscr{A}(t,x)$ of $\mathcal{A}_{(t,x)}$. This amounts to specifying the control set for the system. However, in order to focus on the geometry associated with an affine system and its linearization, it is assumed that $\mathscr{A}_{(t,x)} = \mathcal{A}_{(t,x)}$. This essentially means that the controls are unrestricted. Accepting a slight abuse of notation, a time-dependent affine subbundle $\mathcal{A}$ will be called a ***time-dependent affine system***. A ***trajectory*** for $\mathcal{A}$ is then an LAC curve $\gamma\colon I \to M$ with the property that $\gamma'(t) \in \mathcal{A}_{(t,\gamma(t))}$.

Note that the specification of an affine system does not provide a natural notion of a drift vector field and control vector fields. It can be seen that the basic properties like controllability can depend on the choice of a drift vector field. For the geometric development of the linearization, this is a non-issue since it is natural to assume the presence of a reference vector field, cf. the discussion of Section 1.1. To be formal about this, a ***reference vector field*** for an affine system $\mathcal{A}$ is an LIC$^\infty$ vector field $X_{\text{ref}} \in \text{LIC}^\infty(TM)$ with the property that $X_{\text{ref}}(t,x) \in \mathcal{A}_{(t,x)}$ for all $x \in M$ and almost every $t \in \mathbb{R}$. Of course, integral curves of X_{ref} are trajectories for $\mathcal{A}$. If $\gamma\colon I \to M$ is a trajectory for $\mathcal{A}$, then there exists a reference vector field X_{ref} for $\mathcal{A}$ for which γ is an integral curve [Sussmann 1998, Proposition 4.1].

3.1. Linearization about a reference trajectory. Let $\mathcal{A}$ be a time-dependent affine system and let X_{ref} be a reference vector field with corresponding LAC reference trajectory $\gamma_{\text{ref}}\colon I \to M$. The embedding of γ_{ref} as an integral curve of a reference vector field gives additional useful structure and corresponds more naturally to the standard case, cf. Section 1.1.

What it means to linearize about γ_{ref} is captured in the following definition. An ***$\mathcal{A}$-variation*** of γ_{ref} is a map $\sigma\colon I \times J \to M$ with the following properties:

1. $J \subseteq \mathbb{R}$ is an interval for which $0 \in \text{int}(J)$;
2. σ is continuous;
3. the map $I \ni t \mapsto \sigma_s(t) \triangleq \sigma(t, s) \in M$ is a trajectory of $\mathcal{A}$ for each $s \in J$;
4. the map $J \ni s \mapsto \sigma^t(s) \triangleq \sigma(t, s) \in M$ is LAC for each $t \in I$;
5. the map $I \ni t \mapsto \frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0} \sigma^t(s) \in TM$ is LAC;
6. $\sigma_0 = \gamma_{\text{ref}}$.

Given an $\mathcal{A}$-variation σ of γ_{ref}, a vector field V_σ along γ_{ref} is defined by

$$V_\sigma(t) = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \sigma^t(s). \tag{3.1}$$

The vector field V_σ should be thought of as being the result of linearizing in the direction of the $\mathcal{A}$-variation σ. Using the geometric constructions of Section 2.2, these vector fields along γ_{ref} arise as trajectories for a time-dependent affine system on TM. Such a time-dependent affine subbundle $\mathcal{A}^T_{\text{ref}}$ on TM is defined as follows. Let $(t, v_x) \in \mathbb{R} \times TM$ and define

$$\mathcal{A}^T_{\text{ref},(t,v_x)} = \{X^T_{\text{ref}}(t, v_x) + \text{vlft}(X) \mid X \in L(\mathcal{A})_{(t,x)}\}.$$

This is a time-dependent affine subbundle since it possesses local generators $\{X^T_{\text{ref}}, \text{vlft}(X_1), \dots, \text{vlft}(X_k)\}$ where $\{X_{\text{ref}}, X_1, \dots, X_k\}$ are local generators for $\mathcal{A}$.

3.1 Proposition: *Let $\mathcal{A}$ be a time-dependent affine system, and let X_{ref} be a reference vector field with LAC differentiable reference trajectory γ_{ref}, as above. For a vector field Υ along γ_{ref}, the following statements are equivalent:*

(i) Υ is a trajectory for $\mathcal{A}^T_{\text{ref}}$;

(ii) there exists an $\mathcal{A}$-variation σ of γ_{ref} such that $V_\sigma = \Upsilon$.

Proof: (ii)⇒(i) Let σ be an $\mathcal{A}$-variation giving rise to the vector field V_σ along γ_{ref}. Using a set of local generators $\{X_{\text{ref}}, X_1, \dots, X_k\}$ for $\mathcal{A}$,

$$\sigma_s'(t) = X_{\text{ref}}(t, \sigma_s(t)) + \sum_{j=1}^k u^j(s,t) X_j(t, \sigma_s(t)),$$

since σ_s is a trajectory for $\mathcal{A}$. Differentiating with respect to s at $s = 0$ gives

$$V_\sigma'(t) = X^T_{\text{ref}}(t, \gamma_{\text{ref}}'(t)) + \sum_{j=1}^k \Big(v^j(t) X_j(t, \gamma_{\text{ref}}(t)) + u^j(0,t) \, \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} X_j(t, \sigma_s(t)) \Big)$$

for almost every t, where $v^j(t) = \frac{\partial u^j}{\partial s}(0,t)$. Since $\sigma_0 = \gamma_{\text{ref}}$ it follows that $u^j(0,t) = 0$, and so it follows that $V'_\sigma(t) \in \mathcal{A}^T_{\text{ref},(t,\gamma_{\text{ref}}(t))}$, as desired.

(i)⇒(ii) Let $\sigma_{j,s}$ be the $\mathcal{A}$-variation of γ_{ref} satisfying

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_{j,s}(t) = X_{\text{ref}}(t, \sigma_{j,s}(t)) + sX_j(t, \sigma_{j,s}(t)),$$

noting that the corresponding infinitesimal variation is

$$\frac{\mathrm{d}}{\mathrm{d}t}V_{\sigma_j}(t) = X^T_{\text{ref}}(t, \gamma'_{\text{ref}}(t)) + \operatorname{vlft}(X_j(t))(\gamma'_{\text{ref}}(t)), \qquad \text{ad } t.$$

The convexity of the set of variations of a given order (see [Bianchini and Stefani 1993]) now ensures the existence of a variation for any trajectory Υ that covers γ_{ref}. ■

3.2. Linearization about an equilibrium point. The above developments concerning linearization about a reference trajectory simplify significantly when dealing with an equilibrium point. Here the development looks a lot more like the standard non-geometric setup.

Let $\mathcal{A}$ be a time-dependent affine subbundle on M and let $X_{\text{ref}}\colon I \times M \to TM$ be a reference vector field for $\mathcal{A}$. A point $x_0 \in M$ is an ***equilibrium point*** for X_{ref} if $X_{\text{ref}}(t, x_0) = 0_{x_0}$ for each $t \in I$. Thus the curve $I \ni t \mapsto x_0 \in M$ is an integral curve for X_{ref}. The tangent lift X^T_{ref} at an equilibrium point for X_{ref} has the following properties.

3.2 Proposition: *If $x_0 \in M$ is an equilibrium point for an LIC^∞ vector field $X_{\text{ref}}\colon I \times M \to TM$, then $X^T_{\text{ref}}(t, v_{x_0})$ is vertical for each $v_{x_0} \in T_{x_0}M$. Furthermore, for each $t \in I$ there exists a unique $A(t) \in \mathrm{L}(T_{x_0}M; T_{x_0}M)$ such that $X^T_{\text{ref}}(t, v_{x_0}) = \operatorname{vlft}(A(t)(v_{x_0}))$, and the map $I \ni t \mapsto A(t) \in \mathrm{L}(T_{x_0}M; T_{x_0}M)$ is Lebesgue measurable.*

Proof: This follows directly from the coordinate representation (2.2) for the tangent lift. ■

Thus the tangent lift is vertical-valued on $T_{x_0}M$. Since $V_{v_{x_0}}TM \simeq T_{x_0}M$ this means that the linearization is a time-dependent linear affine system on $T_{x_0}M$ for which

$$\mathcal{A}^T_{\text{ref}}(t, v_{x_0}) = \{A(t)(v_{x_0}) + b \mid\ b \in L(\mathcal{A})_{(t,x_0)}\}.$$

Trajectories $\xi\colon I \to T_{x_0}M$ of the linearization then satisfy

$$\xi'(t) = A(t)(\xi(t)) + b(t), \tag{3.2}$$

for some measurable curve $b\colon I \to T_{x_0}M$ having the property that $b(t) \in L(\mathcal{A})_{(t,x_0)}$. To make this look more like the usual notion of a time-varying linear system, for each $t \in I$ let U be a finite-dimensional $\mathbb{R}$-vector space and let $B(t) \in \mathrm{L}(U; T_{x_0}M)$ have the property that $\operatorname{image}(B(t)) = L(\mathcal{A})_{(t,x_0)}$. Then the equation governing trajectories become

$$\xi'(t) = A(t)(\xi(t)) + B(t)(u(t)),$$

for a measurable curve $u\colon I \to U$. This then recovers the usual notion of a time-dependent linear system.

4. Linear controllability

In Section 3, a time-dependent affine subbundle $\mathcal{A}^T_{\text{ref}}$ on TM was constructed by linearizing a time-dependent affine subbundle $\mathcal{A}$ on M along a reference trajectory γ_{ref}. In this section notions of the reachable set for both $\mathcal{A}$ and $\mathcal{A}^T_{\text{ref}}$ are defined, as well as the associated versions of controllability along γ_{ref}. In Section 4.2 the standard results for a time-varying linear system are cast in a geometric manner. In Section 4.3 this geometric characterization is used to state a completely geometric result that describes the reachable sets for the linearization.

4.1. Controllability definitions. In this section the reachable sets for both $\mathcal{A}$ and $\mathcal{A}^T_{\text{ref}}$ are defined and, for the latter, two equivalent characterizations of the reachable set are provided. Then controllability along a reference trajectory is defined for $\mathcal{A}$ and $\mathcal{A}^T_{\text{ref}}$. Recall that a trajectory for a time-dependent affine system $\mathcal{A}$ is an LAC curve $\gamma\colon I \to M$ such that $\gamma'(t) \in \mathcal{A}_{(t,\gamma(t))}$. The set of trajectories defined on $[t_0, t_1]$ is denoted by $\operatorname{Traj}(\mathcal{A}, t_1, t_0)$ and

$$\operatorname{Traj}(\mathcal{A}, t_0) = \bigcup_{t_1 \geq t_0} \operatorname{Traj}(\mathcal{A}, t_1, t_0).$$

For $x_0 \in M$ and $t \geq t_0$, the ***reachable set*** of $\mathcal{A}$ from x_0 is defined by

$$\mathcal{R}_{\mathcal{A}}(x_0, t, t_0) = \{\gamma(t) \mid \ \gamma \in \operatorname{Traj}(\mathcal{A}, t, t_0),\ \gamma(t_0) = x_0\}.$$

Similarly, a ***trajectory*** for the linearized time-dependent affine system $\mathcal{A}^T_{\text{ref}}$ is an LAC curve $\Upsilon\colon I \to TM$ such that $\Upsilon'(t) \in \mathcal{A}^T_{\text{ref},(t,\Upsilon(t))}$. Then the set of trajectories defined on $[t_0, t_1]$ is denoted by $\operatorname{Traj}(\mathcal{A}^T_{\text{ref}}, t_1, t_0)$ and $\operatorname{Traj}(\mathcal{A}^T_{\text{ref}}, t_0) = \bigcup_{t_1 \geq t_0} \operatorname{Traj}(\mathcal{A}^T_{\text{ref}}, t_1, t_0)$. For $v_{x_0} \in T_{x_0}M$ and $t \geq t_0$, the ***reachable set*** from v_{x_0} is defined by

$$\mathcal{R}_{\mathcal{A}^T_{\text{ref}}}(v_{x_0}, t, t_0) = \{\Upsilon(t) \mid \ \Upsilon \in \operatorname{Traj}(\mathcal{A}^T_{\text{ref}}, t, t_0),\ \Upsilon(t_0) = v_{x_0}\}.$$

With these notions of the reachable sets, the controllability of each system is defined as follows.

4.1 Definition: Let X_{ref} be a reference vector field for $\mathcal{A}$ and let $\gamma_{\text{ref}}\colon I \to M$ be a reference trajectory. Let $x_0 \in M$ and $\gamma_{\text{ref}}(t_0) = x_0$. The system $\mathcal{A}$ is

(i) ***controllable at t_0*** along γ_{ref} if $\gamma_{\text{ref}}(t) \in \operatorname{int} \mathcal{R}_{\mathcal{A}}(x_0, t, t_0)$ for each $t > t_0$ and is

(ii) ***linearly controllable at t_0*** along γ_{ref} if $\mathcal{R}_{\mathcal{A}^T_{\text{ref}}}(0_{x_0}, t, t_0) = T_{\gamma_{\text{ref}}(t)}M$ for each $t > t_0$. •

4.2. Recasting the standard results. In a step toward a geometric theory of Jacobian linearization, the standard setup of Brockett [1970] is recast on general $\mathbb{R}$-vector spaces. In doing so, the extra structure available with Euclidean spaces, in particular the standard inner product, is removed. Let U and V be $\mathbb{R}$-vector spaces with $\dim(U) = m$ and $\dim(V) = n$. Let $A\colon \mathbb{R} \to \mathrm{L}(V; V)$ and $B\colon \mathbb{R} \to \mathrm{L}(U; V)$ be continuous and define a time-varying affine subbundle $\mathcal{A}_{(A,B)}$ on V by

$$\mathcal{A}_{(A,B),(t,v)} = \{A(t)v + B(t)u \mid \ u \in U\}.$$

A trajectory ξ of $\mathcal{A}_{(\mathrm{A},\mathrm{B})}$ satisfies

$$\xi'(t) = A(t)\xi(t) + B(t)u(t). \tag{4.1}$$

The solution to (4.1) satisfying $\xi(t_0) = \xi_0$ for $t_0 \in I$ is given by,

$$\xi(t) = \Phi(t, t_0)\xi_0 + \int_{t_0}^{t} \Phi(t, \sigma)B(\sigma)u(\sigma)\mathrm{d}\sigma, \tag{4.2}$$

where $\Phi(t, t_0)$ is the state transition matrix. That is, $t \mapsto \Phi(t, t_0)$ is the solution to the homogeneous system $\Phi'(t, t_0) = A(t)\Phi(t, t_0)$ with initial condition $\Phi(t_0, t_0) = \mathrm{id}_V$. The transition matrix then has the following properties:

1. $\Phi(t, \tau)\Phi(\tau, t_0) = \Phi(t, t_0)$;
2. $\Phi(t, \tau)^{-1} = \Phi(\tau, t)$.

The subbundle $\mathcal{A}_{(A,B)}$ is ***controllable at t_0*** if, for each $\xi_0, \xi_1 \in V$, there exists a control $u\colon [t_0, t_1] \to U$ which steers from ξ_0 at time t_0 to ξ_1 at time t_1. In the standard case the controllability of a time-varying linear system is equivalent to the controllability Gramian,

$$W(t_0, t) = \int_{t_0}^{t} \Phi(t_0, \sigma)B(\sigma)B^T(\sigma)\Phi^T(t_0, \sigma)\mathrm{d}\sigma, \tag{4.3}$$

having full rank for $t > t_0$. This definition makes use of the standard inner product on $\mathbb{R}^m$ to identify $\mathbb{R}^m$ with $(\mathbb{R}^m)^*$. Indeed, the domain of $B\colon \mathbb{R} \to L(\mathbb{R}^m; \mathbb{R}^n)$ and image of $B^T\colon \mathbb{R} \to L((\mathbb{R}^n)^*; (\mathbb{R}^m)^*)$ for each t are different spaces. Inducing an inner product on U by a symmetric map $R\colon I \to L(U; U^*)$ which is positive-definite for each $t \in I$ yields a Gramian of the form

$$W(t_0, t) = \int_{t_0}^{t} \Phi(t_0, \sigma)B(\sigma)R^{\sharp}(t)B^T(\sigma)\Phi^T(t_0, \sigma)\mathrm{d}\sigma. \tag{4.4}$$

The derivation of (4.4) follows directly from the standard case in [Brockett 1970] and the time-varying affine subbundle $\mathcal{A}_{(A,B)}$ is controllable at t_0 if and only if $W(t_0, t)$ is surjective for $t > t_0$. Later, in Section 5.7, the quadratic cost in the LQR problem provides a natural choice of an inner product.

The notion of a controllability Gramian does not make sense in the geometric framework of Section 3. There is no natural way to construct the analogue of $W(t_0, t)$ for the linearization of a reference vector field X_{ref} along a reference trajectory γ_{ref} since (4.4) is an integral of maps $\{t \mapsto B(t)\}$ that depend on a choice of coordinates. Therefore, an alternative characterization of controllability that can be applied in the geometric setting is needed. The following result gives one such characterization.

4.2 Theorem: *Let V, U, A, and B be as above. Then*

$$\mathrm{image}(W(t_0, t)) = \mathrm{span}_{\mathbb{R}}\Bigg(\bigcup_{\substack{\tau \in [t_0, t] \\ b_\tau \in \mathrm{image}(B(\tau))}} \Phi(t_0, \tau)b_\tau\Bigg).$$

Proof: For notational convenience define

$$\mathcal{S}_{\mathcal{A}_{(\mathrm{A,B})}}(t_0, t) = \mathrm{span}_{\mathbb{R}}\Bigg(\bigcup_{\substack{\tau \in [t_0, t] \\ b_\tau \in \mathrm{image}(B(\tau))}} \Phi(t_0, \tau)b_\tau\Bigg).$$

Let $v \in \operatorname{image}(W(t_0, t))$. Then there exists a continuous control $u \colon [t_0, t] \to U$ such that

$$v = \int_{t_0}^{t} \Phi(t_0, \sigma) B(\sigma) u(\sigma) \mathrm{d}\sigma$$

[Brockett 1970]. Since A, B, and u are continuous, there exists a sequence of partitions $P_i = \{t_0 = t_{1,i}, \dots, t_{k_i,i} = t\}$ of $[t_0, t]$ such that, if

$$v_i = \sum_{j=2}^{k_i} \Phi(t_0, t_{j,i}) B(t_{j,i}) u(t_{j,i})(t_{j,i} - t_{j-1,i}),$$

then $\lim_{i \to \infty} v_i = v$. It is clear that, for each $i \in \mathbb{N}$, $v_i \in \mathcal{S}_{\mathcal{A}_{(\mathrm{A},\mathrm{B})}}(t_0, t)$. Since $\mathcal{S}_{\mathcal{A}_{(\mathrm{A},\mathrm{B})}}(t_0, t)$ is closed it follows that $v \in \mathcal{S}_{\mathcal{A}_{(\mathrm{A},\mathrm{B})}}(t_0, t)$.

Now assume that $v \in \mathcal{S}_{\mathcal{A}_{(\mathrm{A},\mathrm{B})}}(t_0, t)$. Choose $t_1, \dots, t_k \in [t_0, t]$ and $b_{t_j} \in \operatorname{image}(B(t_j))$, $j \in \{1, \dots, k\}$, such that

$$\mathcal{S}_{\mathcal{A}_{(\mathrm{A},\mathrm{B})}}(t_0, t) = \operatorname{span}_{\mathbb{R}}(\Phi(t_0, t_1) b_{t_1}, \dots, \Phi(t_0, t_k) b_{t_k}).$$

Then v can be written as

$$v = \sum_{j=1}^{k} c_j \Phi(t_0, t_j) b_{t_j}.$$

A useful characterization of points in $\operatorname{image}(W(t_0, t))$ is provided by the next lemma.

Lemma $\operatorname{image}(W(t_0, t)) = \{\Phi(t_0, t)\tilde{v} \in V \mid \exists\, u \colon [t_0, t] \to U \text{ steering zero to } \tilde{v}\}$.

Proof: By (4.2), the set of points reachable from $0 \in V$ in time t from t_0 is

$$\Big\{ \int_{t_0}^{t} \Phi(t, \sigma) B(\sigma) u(\sigma) \mathrm{d}\sigma \;\Big|\; u \colon [t_0, t] \to U \text{ continuous} \Big\}.$$

Using the composition property of the transition matrix, apply $\Phi(t_0, t)$ to any point in this set:

$$\Phi(t_0, t) \int_{t_0}^{t} \Phi(t, \sigma) B(\sigma) u(\sigma) \mathrm{d}\sigma = \int_{t_0}^{t} \Phi(t_0, \sigma) B(\sigma) u(\sigma) \mathrm{d}\sigma.$$

The lemma now follows by comparison with (4.3). ▼

If the system can be steered from 0 to $\Phi(t, t_0)v$, this part of the theorem will follow from the lemma. Let $\mu_j \in U$ have the property that $B(t_j)\mu_j = b_{t_j}$ $j \in \{1, \dots, k\}$. Now consider the distributional control $u = \sum_{j=1}^{k} c_j \delta_{t_j} \mu_j$, where δ_{t_j} is the delta-distribution with support $\{t_j\}$. Applying this control, by (4.2), yields

$$\int_{t_0}^{t} \Phi(t, \sigma) B(\sigma) u(\sigma)\, \mathrm{d}\sigma = \sum_{j=1}^{k} c_j \Phi(t, t_j) b_{t_j} = \Phi(t, t_0) v. \tag{4.5}$$

Thus the distributional control u steers from 0 to $\Phi(t, t_0)v$, as desired. To show the distributional control u can be replaced with a sequence of piecewise continuous controls, consider the following lemma.

Lemma There exists a sequence of controls $\{u_i\}_{i\in\mathbb{N}}$ such that

$$\lim_{i\to\infty}\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)u_i(\sigma)\mathrm{d}\sigma=\sum_{j=1}^{k}c_j\Phi(t,t_j)b_{t_j}.$$

Proof: For $j\in\{1,\dots,k\}$ and $i\in\mathbb{N}$ define

$$u_{j,i}(t)=\begin{cases} i\,c_j\,\mu_j, & t\in[t_j,t_j+\frac{1}{i}],\\ 0, & \text{otherwise.}\end{cases}$$

Now note that, using the Peano–Baker series,

$$\begin{aligned}\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)u_{j,i}(\sigma)\mathrm{d}\sigma &= \Phi(t,t_j+\tfrac{1}{i})\int_{t_0}^{t}\Phi(t_j+\tfrac{1}{i},\sigma)B(\sigma)u_{j,i}(\sigma)\mathrm{d}\sigma\\ &= \Phi(t,t_j+\tfrac{1}{i})\int_{t_j}^{t_j+\frac{1}{i}}\Phi(t_j+\tfrac{1}{i},\sigma)B(\sigma)u_{j,i}(\sigma)\mathrm{d}\sigma\\ &= i\,c_j\,\Phi(t,t_j+\tfrac{1}{i})\int_{t_j}^{t_j+\frac{1}{i}}\Big(\mathrm{id}_V+\int_{\sigma}^{t_j+\frac{1}{i}}A(\sigma_1)\mathrm{d}\sigma_1\\ &\quad+\int_{\sigma}^{t_j+\frac{1}{i}}A(\sigma_1)\int_{\sigma}^{\sigma_1}A(\sigma_2)\mathrm{d}\sigma_2\mathrm{d}\sigma_1+\dots\Big)B(\sigma)\mu_j\mathrm{d}\sigma.\end{aligned}$$

Because A is continuous, all terms in the Peano–Baker series go to zero at least as fast as $(\frac{1}{i})^2$. Thus only the first term remains in the limit, giving

$$\lim_{i\to\infty}\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)u_{j,i}(\sigma)\mathrm{d}\sigma=c_j\Phi(t,t_j)b_{t_j}.$$

The result now follows by taking $u_i=\sum_{j=1}^{k}u_{j,i}$. ▼

Let $\{u_i\}_{i\in\mathbb{N}}$ be a sequence of controls defined by the preceding lemma. For each $i\in\mathbb{N}$,

$$\Phi(t_0,t)\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)u_i(\sigma)\mathrm{d}\sigma\in\mathrm{image}(W(t_0,t))$$

by the first of the above lemmata. Therefore, the limit as $i\to\infty$ is also in $\mathrm{image}(W(t_0,t))$. But, by (4.5),

$$\lim_{i\to\infty}\Phi(t_0,t)\int_{t_0}^{t}\Phi(t,\sigma)B(\sigma)u_i(\sigma)\mathrm{d}\sigma=v,$$

giving the result. ■

4.3. Controllability of linearizations. To provide a geometric characterization, using Theorem 4.2, of the reachable set for the linearization of a reference vector field X_{ref} along a reference trajectory γ_{ref} requires the following definition. For an LAC curve $\gamma\colon I\to M$, a ***distribution along*** γ is a subset $D\subseteq TM|\,\mathrm{image}(\gamma)$ with the property that, for each $t_0\in I$, there exists a neighbourhood $J\subseteq I$ of t_0 and LAC vector fields $\xi_1,\dots,\xi_k$ along $\gamma|J$ such that $D_{\gamma(t)}=\mathrm{span}_{\mathbb{R}}(\xi_1(t),\dots,\xi_k(t))$ for each $t\in J$. Let $t_0\in\mathrm{int}(I)$ and denote $T=\sup I$, allowing that $T=\infty$. Let $I_{t_0}=[t_0,T)$. Denote by γ_{t_0} the restriction of γ_{ref} to I_{t_0}. Recall from Section 2.1 the definition of $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{t_0}}$ and denote by $\langle\mathscr{L}^{X_{\mathrm{ref}},\gamma_{t_0}},L(\mathcal{A})_{t_0}\rangle$ the smallest $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{t_0}}$-invariant distribution along γ_{t_0} that agrees with $L(\mathcal{A})$ at $\gamma_{\mathrm{ref}}(t_0)$.

4.3 Theorem: *Let $\mathcal{A}$ be a time-dependent affine system on M with X_{ref} a reference vector field and $\gamma_{\mathrm{ref}}\colon I \to M$ an LAC reference trajectory. For $t_0 \in I$ and $t > t_0$ the following sets are equal:*

(i) $\mathcal{R}_{\mathcal{A}^T_{\mathrm{ref}}}(0_{x_0}, t, t_0)$;

(ii) $\operatorname{span}_{\mathbb{R}}\left(\bigcup_{\substack{\tau\in[t_0,t]\\ v_\tau\in L(\mathcal{A})_{\tau,\gamma_{\mathrm{ref}}(\tau)}}} \Phi^{X^T_{\mathrm{ref}}}_{\tau,t}(v_\tau)\right)$;

(iii) $\langle \mathscr{L}^{X_{\mathrm{ref}},\gamma_{t_0}}, L(\mathcal{A})_{t_0}\rangle_{\gamma_{\mathrm{ref}}(t)}$.

Proof: Recall that

$$\gamma^*_{\mathrm{ref}}TM = \{(t,v) \mid \gamma_{\mathrm{ref}}(t) = \tau_M(v)\},$$

where $\gamma^*_{\mathrm{ref}}\tau_M(t,v) = t$. Thus $\gamma^*_{\mathrm{ref}}TM$ is a vector bundle over I with fibre over $t \in I$ being $T_{\gamma_{\mathrm{ref}}(t)}M$. This bundle may be trivialized since I is contractible. Denote by $\rho\colon \gamma^*_{\mathrm{ref}}\tau_M \to I \times V$ a particular trivialization with a commuting diagram

$$\begin{array}{ccc} \gamma^*_{\mathrm{ref}}\tau_M & \xrightarrow{\ \rho\ } & I\times V \\ & {\scriptstyle \gamma^*_{\mathrm{ref}}\tau_M}\searrow \quad \swarrow{\scriptstyle \mathrm{pr}_1} & \\ & I & \end{array}$$

where pr_1 the projection onto the first factor.

The following lemma records some useful properties of the representation of trajectories of $\mathcal{A}^T_{\mathrm{ref}}$.

Lemma The following statements hold:

(i) there exists a vector bundle endomorphism $A\colon I \times V \to I \times V$ over id_I with the property that $T_{(t,\gamma_{\mathrm{ref}}(t))}\rho(1, X^T_{\mathrm{ref}}(t, v_{\gamma_{\mathrm{ref}}(t)})) = (1, A(t)\cdot\rho(v_{\gamma_{\mathrm{ref}}(t)}))$;

(ii) if $X_{\mathrm{ref}} \in \Gamma^\infty(TM)$ then there exists a section $\xi_{X_{\mathrm{ref}}}$ of $\mathrm{pr}_1\colon I \times V \to I$ such that $T_{(t,\gamma_{\mathrm{ref}}(t))}\rho(0, \mathrm{vlft}(X_{\mathrm{ref}})(\gamma_{\mathrm{ref}}(t))) = \xi_X(t)$.

Proof: The first assertion follows since X^T_{ref} is a vector bundle mapping over X_{ref}. The second part of the lemma is merely the definition of $\xi_{X_{\mathrm{ref}}}$. ▼

The lemma says that, if $v(t) = T\rho(\Upsilon(t))$ for a trajectory Υ for $\mathcal{A}^T_{\mathrm{ref}}$, then we have

$$v'(t) = A(t)v(t) + b(t),$$

where $b(t) \in \operatorname{image}\big(\rho(\mathrm{vlft}(\mathcal{A}_{(t,\gamma_{\mathrm{ref}}(t))}))\big)$. Therefore, the equality of the sets (i) and (ii) follows from Theorem 4.2.

From the definition of the set in (ii), the equivalence of (ii) and (iii) will follow if it can be shown that the notion of a distribution along γ_{t_0} being invariant under $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{t_0}}$ is equivalent to the notion of being invariant under the flow of X^T_{ref} along γ_{t_0}. Let $\mathcal{D}$ be a distribution along γ_{t_0}. The distribution $\mathcal{D}$ is invariant under the flow of X^T_{ref} if and only if it is invariant under $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{t_0}}$. This claim follows from the characterization of the flow of X_{ref} in Proposition 2.1 in terms of $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{t_0}}$. That result states that the flow of X^T_{ref} through $v \in \mathcal{D}_{\gamma_{t_0}(t_0)}$ is obtained by transporting v along γ_{t_0}. ■

The set described in part (iii) of Theorem 4.3 should be thought of as the analogue of "the smallest A-invariant subspace containing image(B)" in the time-invariant linear theory. Finally, Theorem 4.3 immediately gives the following corollary, the second part of which follows from the variational cone results of [Bianchini and Stefani 1993].

4.4 Corollary: *Let $\mathcal{A}$, X_{ref}, and γ_{ref} be as in Theorem 4.3. Then the following statements hold for $t_0 \in I$:*

(i) $\mathcal{A}$ is linearly controllable at t_0 along γ_{ref} if the smallest $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}$-invariant distribution along γ_{ref} containing $L(\mathcal{A})|\,\mathrm{image}(\gamma_{\mathrm{ref}})$ is equal to $TM|\,\mathrm{image}(\gamma_{\mathrm{ref}})$;

(ii) if $\mathcal{A}$ is linearly controllable at t_0 along γ_{ref} then it is controllable at t_0 along γ_{ref}.

5. LQR and the Maximum Principle

In this section the main geometric structure for LQR theory is presented by characterizing solutions to the finite time LQR problem using the Maximum Principle as stated in Theorem 5.2. Although Theorem 5.2 is a new formulation of the Maximum Principle, the ideas required to prove it come for the existing formulations of the Maximum Principle. Thus, many of the technicalities follow from the standard versions of the Maximum Principle found in [Lee and Markus 1967].

After providing the geometric versions of both the finite and infinite time LQR problems in Section 5.1, the bulk of this section builds the tools to prove Theorem 5.2 in Section 5.5. In Sections 5.2, 5.3, and 5.4, the variational and adjoint equations, needle variations, tangent cones, and, of course, the Hamiltonian are defined. Again, it is noted that the Hamiltonian presented in Section 5.4 will look "different" from the standard case but maintains the required maximization properties—see Lemmata 5.17 and 5.18—required to prove the Maximum Principle. In Section 5.6, the Maximum Principle is used to characterize solutions to the finite time LQR problem and the geometric version of the Riccati equation is given. In Section 5.7, solutions to the Riccati equation are shown to exist as the final time in the LQR problem tends to infinity. In arriving at this result, the geometric analogue of the minimum energy controller is defined. Finally, the trajectory corresponding to the solution of the Riccati equation, as the final time tends to infinity, is shown to be optimal in the sense of the infinite time LQR problem.

5.1. LQR problem definitions. Let X_{ref} be a reference vector field, with a reference trajectory γ_{ref} defined on I, for an affine system $\mathcal{A}$. The linearization, a time-dependent affine system $\mathcal{A}^T_{\mathrm{ref}}$ on TM, is defined as in Section 3:

$$\mathcal{A}^T_{\mathrm{ref},(t,\,v_x)} = \{X^T_{\mathrm{ref}}(t, v_x) + \mathrm{vlft}(X) \mid \ X \in L(\mathcal{A})_{(t,x)}\}.$$

To formulate an LQR problem for the linearization $\mathcal{A}^T_{\mathrm{ref}}$ requires the following constructions. For a vector bundle $\pi\colon E \to B$, we denote by $\Sigma_2(E)$ be the bundle of symmetric $(0,2)$-tensors on B. To define the cost function along the reference trajectory, let Q be an LI section of $\Sigma_2(TM|_{\mathrm{image}(\gamma_{\mathrm{ref}})})$ with the property that $Q(t)$ is positive-semidefinite for each $t \in I$. Also, let R be an LI section of $\Sigma_2(L(\mathcal{A})|_{\mathrm{image}(\gamma_{\mathrm{ref}})})$ with the property that $R(t)$ is positive-definite for each $t \in I$.

5.1 Problem: (i) (Finite time, fixed interval, free endpoint problem) Find a vector field $\Upsilon\colon [t_0,t_1]\to TM$ along γ_{ref} such that $\Upsilon\in \text{Traj}(\mathcal{A}^T_{\text{ref}},t_1,t_0)$ and minimizes the cost

$$J(\Upsilon(t_0),t_0,t_1)=\frac{1}{2}F(t_1)(\Upsilon(t_1),\Upsilon(t_1))+\int_{t_0}^{t_1}\frac{1}{2}Q(t)(\Upsilon(t),\Upsilon(t))+\frac{1}{2}R(t)(X(t),X(t))\mathrm{d}t,$$

where $F(t_1)$ is a symmetric positive-semidefinite $(0,2)$-tensor on $T_{\gamma_{\text{ref}}(t_1)}M$.

(ii) (Infinite time, free endpoint problem) Find a vector field $\Upsilon\colon [t_0,\infty)\to TM$ along γ_{ref} such that $\Upsilon\in\text{Traj}(\mathcal{A}^T_{\text{ref}},t_0)$ and minimizes the cost

$$J(\Upsilon(t_0),t_0,\infty)=\int_{t_0}^{\infty}\frac{1}{2}Q(t)(\Upsilon(t),\Upsilon(t))+\frac{1}{2}R(t)(X(t),X(t))\mathrm{d}t.$$ •

The ***Lagrangian*** associated with Problem 5.1 is the map $L\colon TM\times L(\mathcal{A})\to\mathbb{R}$ defined by

$$L(\Upsilon(t),X(t))=\frac{1}{2}Q(t)(\Upsilon(t),\Upsilon(t))+\frac{1}{2}R(t)(X(t),X(t)). \tag{5.1}$$

Let $\bar{L}\colon TM\to\mathbb{R}$ be a smooth map and define the ***fiber derivative*** as the map $F\bar{L}\colon TM\to T^*M$ given by

$$\langle F\bar{L}(v_x);w_x\rangle=\frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0}\bar{L}(v_x+sw_x).$$

In the natural coordinates for TM and T^*M, the local representative of the fibre derivative is given by

$$((x^1,\dots,x^n),(v^1,\dots,v^n))\mapsto\left((x^1,\dots,x^n),\left(\frac{\partial\bar{L}}{\partial v^1},\dots,\frac{\partial\bar{L}}{\partial v^n}\right)\right).$$

For a fixed section X of $L(\mathcal{A})$, the restriction of the fiber derivative along γ_{ref} applied to (5.1) yields $FL(\Upsilon(t))=Q^\flat(t)\Upsilon(t)$.

Theorem 5.2, stated below, is used to characterize solutions to Problem 5.1(i) in Section 5.6. The infinite time problem, Problem 5.1(ii), is addressed in Section 5.7. The proof of Theorem 5.2 is found in Section 5.5 after building the necessary background in Sections 5.2, 5.3 and 5.4. Let $\iota(t)\colon L(\mathcal{A})_{(t,\gamma_{\text{ref}}(t))}\to T_{\gamma_{\text{ref}}(t)}M$ be the natural inclusion with $\iota^*(t)\colon T^*_{\gamma_{\text{ref}}(t)}M\to L(\mathcal{A})^*_{(t,\gamma_{\text{ref}}(t))}$ its dual.

5.2 Theorem: *Let $\mathcal{A}$ be an affine system with reference vector field X_{ref} and a controllable linearization $\mathcal{A}^T_{\text{ref}}$ along γ_{ref}. Let $t_0,t_1\in\mathbb{R}$ satisfy $t_0<t_1$ and let $v_{\gamma_{\text{ref}}(t_0)}\in T_{\gamma_{\text{ref}}(t_0)}M$. Suppose that (Υ_*,X_*) solves Problem 5.1(i). Then there exists a covector field $\Lambda_*\colon[t_0,t_1]\to T^*M$ along γ_{ref} such that the following equations on $TM\oplus T^*M$ hold:*

$$\begin{aligned}\dot{\Upsilon}_*(t)&=X^T_{\text{ref}}(t,\Upsilon_*(t))+\text{vlft}(R^\sharp(t)\iota^*(t)\Lambda_*(t))(\Upsilon(t)),\\ \dot{\Lambda}_*(t)&=X^{T^*}_{\text{ref}}(t,\Lambda_*(t))-\text{vlft}(FL(\Upsilon_*(t)))(\Lambda_*(t)),\end{aligned}$$

where $\Upsilon_*(t_0)=v_{\gamma_{\text{ref}}(t_0)}$ and $\Lambda_*(t_1)=F^\flat(t_1)\Upsilon(t_1)$.

5.2. The variational and adjoint equations. In the standard theory of optimal control for non-linear controls systems on a manifold M, the variational equations are given by the linearization of the dynamics. A trajectory of the variational equations is interpreted as an infinitesimal variation arising from varying the initial conditions of a fixed trajectory on M. In the present geometric framework, the varying of initial conditions for a trajectory for $\mathcal{A}^T_{\text{ref}}$ corresponds to a variation in the fiber over γ_{ref} at the initial time. In other words, the trajectories of the variational equation will be vertical.

5.3 Definition: Let $\mathcal{A}$ be an affine system with linearization $\mathcal{A}^T_{\text{ref}}$ along γ_{ref} and reference vector field X_{ref}. The ***variational equation*** for $\mathcal{A}^T_{\text{ref}}$ is

$$\dot{\Upsilon}(t) = X^T_{\text{ref}}(t, \Upsilon(t)) + \text{vlft}(X(t))(\Upsilon(t)),$$
$$\dot{\Xi}(t) = (X^T_{\text{ref}})^T(t, \Xi(t)),$$

where $\Xi(t)$ is a vertical vector field along $\Upsilon(t)$. •

The trajectories of the variational equations can be interpreted as infinitesimal variations in the following way. Fix a section X of $L(\mathcal{A})$ along γ_{ref} and consider the variations of the resulting trajectory Υ. An ***$\mathcal{A}^T_{\text{ref}}$-variation*** of Υ is a map $\Sigma\colon J \times I \to TM$ with the following properties:

1. $J \subseteq \mathbb{R}$ for which $0 \in \text{int}(J)$;
2. Σ is continuous;
3. $\Sigma(s,t)$ is a trajectory of $\mathcal{A}^T_{\text{ref}}$ for each $s \in J$;
4. $\Sigma(0,t) = \Upsilon(t)$;
5. $\pi_{TM}\Sigma(s,t) = \gamma_{\text{ref}}$;
6. the map $I \ni t \mapsto \frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0} \Sigma(s,t) \in TTM$ is LAC.

For an $\mathcal{A}^T_{\text{ref}}$-variation Σ of Υ, a vector field V_Σ along Υ is defined by

$$V_\Sigma(t) = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \Sigma(s,t). \tag{5.2}$$

As a consequence of property 5, the vector field $V_\Sigma(t)$ along Υ is vertical. In the natural coordinates (x, v, u, w) for TTM, $V_\Sigma(t)$ is given by

$$\mathbf{V}_\Sigma(t) = (\gamma_{\text{ref}}(t), \Upsilon(t), 0, V_\Sigma(t)).$$

5.4 Proposition: *Let $\mathcal{A}$ be an affine system on M with reference vector field X_{ref} and LAC reference trajectory γ_{ref}. Let Υ be a trajectory of $\mathcal{A}^T_{\text{ref}}$. For a vertical vector field Ξ along Υ, the following statements are equivalent:*

(i) Ξ is a integral curve of $(X^T_{\text{ref}})^T$;

(ii) there exists an $\mathcal{A}^T_{\text{ref}}$-variation Σ of Υ such that $V_\Sigma = \Xi$.

Proof: (ii)⇒(i) Let Σ be an $\mathcal{A}^T_{\text{ref}}$-variation Σ of Υ for which $V_\Sigma = \Xi$. Since $\Sigma(s,t)$ is a trajectory of $\mathcal{A}^T_{\text{ref}}$ for each $s \in J$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Sigma(s,t) = X^T_{\text{ref}}(t, \Sigma(s,t)) + \text{vlft}(X(t))(\Sigma(s,t)), \tag{5.3}$$

using a set of local generators $\{X^T_{\text{ref}}, \text{vlft}(X_1), \dots, \text{vlft}(X_k)\}$ for $\mathcal{A}^T_{\text{ref}}$ where $\{X_{\text{ref}}, X_1, \dots, X_k\}$ are local generators for $\mathcal{A}$. Differentiating (5.3) in coordinates with respect to s at $s = 0$ yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{V}_\Sigma(t) &= \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} X^T_{\text{ref}}(t, \Sigma(s,t)) \\ &= (X_{\text{ref}}(t, \gamma_{\text{ref}}(t)), \mathbf{D}X_{\text{ref}}(t, \gamma_{\text{ref}}(t))(\Upsilon(t)), 0, \mathbf{D}X_{\text{ref}}(t, \gamma_{\text{ref}}(t))(V_\Sigma(t))).\end{aligned}$$

For $w_{v_x} \in T_{v_x}TM$ the principle part of $(X^T_{\text{ref}})^T(t, w_{v_x})$ in the natural coordinates for TTM is $(X_{\text{ref}}(t,x), \mathbf{D}X_{\text{ref}}(t,x)v, \mathbf{D}X_{\text{ref}}(t,x)u, \mathbf{D}X_{\text{ref}}(t,x)w)$ which gives $\frac{\mathrm{d}}{\mathrm{d}t}V_\Sigma(t) = (X^T_{\text{ref}})^T(t, V_\Sigma(t))$ as desired.

(i)⇒(ii) This follows by choosing Σ to be the $\mathcal{A}^T_{\text{ref}}$-variation of Υ such that

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \Sigma(s, t_0) = \Xi(t_0). \qquad \blacksquare$$

By making the natural identification of the fibers of the vertical subbundle of TTM with the fibers of the tangent bundle of M, it is possible to view trajectories of the variational equation as vector fields along γ_{ref}. The end effect is that the curves $t \mapsto \Upsilon(t) \oplus \Xi(t)$ in the Whitney sum $TM \oplus TM$, a vector bundle over M, are used to formulate the optimal control problem. The refined variational definitions are as follows.

5.5 Definition: The ***variational equation*** for $\mathcal{A}^T_{\text{ref}}$ is

$$\begin{aligned}\dot{\Upsilon}(t) &= X^T_{\text{ref}}(t, \Upsilon(t)) + \text{vlft}(X(t))(\Upsilon(t)), \\ \dot{\Xi}(t) &= X^T_{\text{ref}}(t, \Xi(t)),\end{aligned}$$

where Ξ is a vector field along γ_{ref}. •

Note that Definition 5.5 agrees with the statement "the linearization of a linear system is the original linear system." The upshot is that the adjoint equation will evolve on $TM \oplus T^*M$, which allows for the effect of the cost to be incorporated into the adjoint equations for the extended system; see Definition 5.10.

5.6 Definition: The ***adjoint equation*** for $\mathcal{A}^T_{\text{ref}}$ is

$$\begin{aligned}\dot{\Upsilon}(t) &= X^T_{\text{ref}}(t, \Upsilon(t)) + \text{vlft}(X(t))(\Upsilon(t)), \\ \dot{\Lambda}(t) &= X^{T*}_{\text{ref}}(t, \Lambda(t)),\end{aligned}$$

where Λ is a covector field along γ_{ref}. •

The adjoint equations will play an important role in the statement and proof of the Maximum Principle. The relationship between the adjoint equations and variational equations is provided by Proposition 5.7.

5.7 Proposition: *If the curve $t \mapsto \Upsilon(t) \oplus \Lambda(t)$ (resp. $t \mapsto \Upsilon(t) \oplus \Xi(t)$) satisfies the adjoint equations (resp. variational equations), then the pairing $\langle \Lambda(t); \Xi(t) \rangle$ is constant along γ_{ref} (i.e., $\langle \Lambda(t); \Xi(t) \rangle = \langle \Lambda(t_0); \Xi(t_0) \rangle$).*

Proof: This follows from a direct computation of $\frac{\mathrm{d}}{\mathrm{d}t}\langle \Lambda(t); \Xi(t) \rangle$ using the coordinate versions of the adjoint and variational equations. ■

A geometric interpretation of the adjoint equations is that they describe the evolution of a hyperplane in TM along γ_{ref}.

5.8 Corollary: *Let $\Lambda(t_0) \in T^*_{\gamma_{\mathrm{ref}}(t_0)}M$ and let $P_{t_0} = \mathrm{coann}(\Lambda(t_0))$. For each $t \in [t_0, t_1]$ define $P_t \subseteq T_{\gamma_{\mathrm{ref}}(t)}M$ and $\Lambda(t) \in T^*_{\gamma_{\mathrm{ref}}(t)}M$ by asking that*

$$P_t = \{\Xi(t) \mid t \mapsto \Upsilon(t) \oplus \Xi(t) \text{ satisfies the variation equation with } \Xi(t_0) \in P_{t_0}\}$$

and that $t \mapsto \Upsilon(t) \oplus \Lambda(t)$ is a solution of the adjoint equations with initial condition $\Lambda(t_0) \in T^*_{\gamma_{\mathrm{ref}}(t_0)}M$. Then $P_t = \mathrm{coann}(\Lambda(t))$ for every $t \in [t_0, t_1]$.

5.3. Needle variations and tangent cones. Roughly speaking the tangent cone is constructed by pushing forward needle variations. Its properties are instrumental in proving the Maximum Principle. The key property of the tangent cone is convexity. The main role of the tangent cone is to approximate the reachable set and it is interpreted as the set of "directions" from which a trajectory can start. In the case of a linear system, both the reachable set $\mathcal{R}(v_{\gamma_{\mathrm{ref}}(t_0)}, t, t_0)$ and the tangent cone at time t are contained in the tangent space $T_{\gamma_{\mathrm{ref}}(t)}M$. In fact, they are equal [Lee and Markus 1967]. However, since the proof of the Maximum Principle makes use of the extended system in Definition 5.9, which is not linear because of the cost being quadratic, this means that the general setup to construct the tangent cone is still required.

To prove the Maximum Principle, it is advantageous to include the cost as a state variable by defining the extended system.

5.9 Definition: The ***extended system***, denoted by $\hat{A}^T_{\mathrm{ref}} \subseteq TM \times \mathbb{R}$, is defined by asking that a trajectory $\hat{\Upsilon} = (\Upsilon, \Upsilon^0)$ satisfies the differential equations

$$\begin{aligned}\dot{\Upsilon}(t) &= X^T_{\mathrm{ref}}(t, \Upsilon(t)) + \mathrm{vlft}(X(t))(\Upsilon(t)),\\ \dot{\Upsilon}^0(t) &= L(\Upsilon(t), X(t)).\end{aligned}$$ •

The adjoint and variational equations can be obtained as before from the linearization of the extended system along a trajectory that projects to the reference trajectory. The effect of the cost enters the adjoint and variational equations using the fiber derivative of the Lagrangian.

5.10 Definition: (i) The ***extended variational equation*** is defined by

$$\begin{aligned}\dot{\Upsilon}(t) &= X^T_{\mathrm{ref}}(t, \Upsilon(t)) + \mathrm{vlft}(X(t))(\Upsilon(t)),\\ \dot{\Upsilon}^0(t) &= L(\Upsilon(t), X(t)),\\ \dot{\Xi}(t) &= X^T_{\mathrm{ref}}(t, \Xi(t)),\\ \dot{\Xi}^0(t) &= FL(\Upsilon(t))\Xi(t),\end{aligned}$$

where Ξ is a vector field along γ_{ref}.

(ii) The ***extended adjoint equation*** is defined by

$$\begin{aligned}
\dot{\Upsilon}(t) &= X^T_{\mathrm{ref}}(t, \Upsilon(t)) + \mathrm{vlft}(X(t))(\Upsilon(t)),\\
\dot{\Upsilon}^0(t) &= L(\Upsilon(t), X(t)),\\
\dot{\Lambda}(t) &= X^{T*}_{\mathrm{ref}}(t, \Lambda(t)) + \Lambda^0(t)\,\mathrm{vlft}(FL(\Upsilon(t)))\Lambda(t),\\
\dot{\Lambda}^0(t) &= 0,
\end{aligned}$$

where $\Lambda(t)$ is a covector field along γ_{ref}. •

The first step toward constructing the tangent cone involves defining needle variations for the extended system, Definition 5.9. The motivation for using needle variations versus some other variety of variations is that the constructions involving needle variations are enough prove to the Maximum Principle.

5.11 Definition: Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Let $\hat{\mathcal{A}}^T_{\mathrm{ref}}$ be an extended system with initial conditions $\hat{\Upsilon}(t_0)$ and X a section of $L(\mathcal{A})$ along γ_{ref}.

(i) ***Fixed interval needle variation data*** is a triple $\theta = (\tau_\theta, \ell_\theta, Z_\theta)$, where
   (a) $\tau_\theta \in (t_0, t_1]$,
   (b) $\ell_\theta \in \mathbb{R}_{\geq 0}$, and
   (c) Z_θ is a section of $L(\mathcal{A})$.

(ii) The ***variation of X*** associated to the fixed interval needle variation data $\theta = (\tau_\theta, \ell_\theta, Z_\theta)$ is a map $X_\theta \colon J \times [t_0, t_1] \to L(\mathcal{A})$ defined by

$$X_\theta(s,t) = \begin{cases} Z_\theta(t), & t \in [\tau_\theta - s\ell_\theta, \tau_\theta], \\ X(t), & \text{otherwise}, \end{cases} \tag{5.4}$$

where $J = [0, s_0]$ is an interval sufficiently small such that $X_\theta(s, \cdot) \colon t \to X_\theta(s,t)$ is a section of $L(\mathcal{A})$ along γ_{ref} for each $s \in J$.

(iii) Let $t \mapsto \Sigma(X_\theta(s,t), \hat{\Upsilon}(t_0), t_0, t)$ be the trajectory of $\hat{\mathcal{A}}^T_{\mathrm{ref}}$ corresponding to $X_\theta(s, \cdot)$ with the fixed interval needle variation data $\theta = (\tau_\theta, \ell_\theta, Z_\theta)$. The ***fixed interval needle variation*** associated with X is defined by

$$v_\theta = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \Sigma(X_\theta(s,\cdot), \hat{\Upsilon}(t_0), t_0, \cdot) \tag{5.5}$$

and is a vertical curve in $VTM \times \mathbb{R}$ which projects to γ_{ref}. •

The existence of the derivative in (5.5) is guaranteed when τ_θ is a Lebesgue point [Lee and Markus 1967, §4.1].

5.12 Definition: Let X_{ref} be an LIC$^\infty$ reference vector field for $\mathcal{A}$ with reference trajectory γ_{ref} and let X be an LI-section of $L(\mathcal{A})$ along γ_{ref}. A point $\tau \in I$ is a ***Lebesgue point of*** $\hat{\mathcal{A}}^T_{\mathrm{ref}}$ if in local coordinates

$$\lim_{s \to 0} \frac{1}{s} \int_{\tau - s}^{\tau} |X^i(t) - X^i(\tau)| \,\mathrm{d}t = 0,$$

where $i \in \{1, \ldots, n\}$. •

It is noted that almost every $\tau_\theta \in (t_0, t_1]$ is a Lebesgue point. For the fixed interval needle variation data $\theta = (\tau_\theta, \ell_\theta, Z_\theta)$, where τ_θ is a Lebesgue point, the fixed interval needle variation v_θ has the form,

$$v_\theta(\tau_\theta) = \ell_\theta(\mathrm{vlft}(Z_\theta(\tau_\theta) - X(\tau_\theta)), L(\Upsilon(\tau_\theta), Z_\theta(\tau_\theta) - X(\tau_\theta))), \tag{5.6}$$

[Lee and Markus 1967, §4.1]. In light of (5.6), fixed interval needle variations will be considered as elements of $L(\mathcal{A}) \times \mathbb{R}$. The set of fixed interval needle variations at Lebesgue points form a cone in $L(\mathcal{A}) \times \mathbb{R}$. More precisely, if v_θ is a fixed interval needle variation with data $\theta = (\tau_\theta, \ell_\theta, Z_\theta)$ and $\lambda \in \mathbb{R}_{\geq 0}$, then λv_θ is a fixed interval needle variation with data $(\tau_\theta, \lambda\ell_\theta, Z_\theta)$. Assigning the notation $\lambda\theta = (\tau_\theta, \lambda\ell_\theta, Z_\theta)$ implies the relation $v_{\lambda\theta} = \lambda v_\theta$.

The above constructions are now extended to allow for multiple variations of X to contribute to corresponding fixed interval needle variations.

5.13 Definition: Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Let $\hat{\mathcal{A}}^T_{\mathrm{ref}}$ be an extended system with initial conditions $\hat{\Upsilon}(t_0)$ and X a section of $L(\mathcal{A})$.

(i) ***Fixed interval multi-needle variation data*** is a collection $\Theta = \{\theta_1, \dots, \theta_k\}$ of fixed interval needle variation data $\theta_j = (\tau_j, \ell_j, Z_j)$, $j \in \{1, \dots, k\}$, such that the times $\tau_1, \dots, \tau_k$ are distinct Lebesgue points.

(ii) The ***variation of X*** associated to the fixed interval multi-needle variation data $\Theta = \{\theta_1, \dots, \theta_k\}$ is a map $X_\Theta \colon J \times [t_0, t_1] \to L(\mathcal{A})$ defined by

$$X_\Theta(s,t) = \begin{cases} Z_j(t), & t \in [\tau_j - s\ell_j, \tau_j],\ j \in \{1, \dots, k\}, \\ X(t), & \text{otherwise}, \end{cases} \tag{5.7}$$

where $J = [0, s_0]$ is an interval sufficiently small that $X_\Theta(s,\cdot) \colon t \to X_\theta(s,t)$ is a section of $L(\mathcal{A})$ along γ_{ref} for each $s \in J$.

(iii) Let $\Sigma(X_\Theta(s,\cdot), \hat{\Upsilon}(t_0), t_0, \cdot)$ be the trajectory corresponding to $X_\Theta(s,\cdot)$ with the fixed interval multi-needle variation data $\Theta = \{\theta_1, \dots, \theta_k\}$. The ***fixed interval multi-needle variation*** associated with X is defined by

$$v_\Theta = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \Sigma(X_\Theta(s,\cdot), \hat{\Upsilon}(t_0), t_0, \cdot) \tag{5.8}$$

and projects to γ_{ref}. •

If $\tau_j < t$, $j \in \{1, \dots, k\}$, associated with the fixed interval multi-needle variation data $\Theta = \{\theta_1, \dots, \theta_k\}$, then v_θ exists and is given by

$$v_\Theta(t) = \sum_{j=1}^{k} \Phi_{\tau_j,t} v_{\theta_j} \tag{5.9}$$

where v_{θ_j} is the fixed interval needle variation for θ_j and where $\Phi_{\tau_j,t}$ is the flow, from τ_j to t, of the linear part of the extended variational equation, Definition 5.10(i), with fixed section $X(t)$, [Lee and Markus 1967, §4.1]. That is, $\Phi_{\tau_j,t}$ is the flow corresponding to

$$\begin{aligned} \dot{\Xi}(t) &= X^T_{\mathrm{ref}}(t, \Xi(t)), \\ \dot{\Xi}^0(t) &= FL(\Upsilon(t))\Xi(t). \end{aligned}$$

If $\Theta = \{\theta_1, \dots, \theta_k\}$ is fixed interval multi-needle variation data and if $\lambda = \{\lambda_1, \dots, \lambda_k\}$, $\lambda_j \in \mathbb{R}_{\geq 0}$, then denote $\lambda\Theta = \{\lambda_1\theta_1, \dots, \lambda_k\theta_k\}$. With this notation, a coned convex combination of fixed interval multi-needle variations takes the form

$$v_{\lambda\Theta}(t) = \sum_{j=1}^{k} \Phi_{\tau_j,t}\lambda_j v_{\theta_j}. \tag{5.10}$$

Furthermore, if $\sum_{j=1}^{k} \lambda_j = 1$, then the limit $\frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0} \Sigma(X_{\lambda\Theta}(s,\cdot), \hat{\Upsilon}(t_0), t_0, \cdot)$ exists uniformly in λ.

Finally we define fixed interval tangent cones for the extended system.

5.14 Definition: Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Let $\hat{\mathcal{A}}^T_{\mathrm{ref}}$ be an extended system with initial conditions $\hat{\Upsilon}(t_0)$ and X a section of $L(\mathcal{A})$. For $t \in [t_0, t_1]$ define the ***fixed interval tangent cone at t***, denoted by $K(X, \hat{\Upsilon}(t_0), t_0, t)$, as the closure of the coned convex hull of the set

$$\bigcup \{\Phi_{\tau,t} v \mid v \text{ is a fixed interval needle variation at a Lebesgue point } \tau\}. \qquad \bullet$$

The next lemma tells us that points in the interior of the fixed interval tangent cone are in the reachable set.

5.15 Lemma: ([Lee and Markus 1967, §4.1, Lemma 2]) *Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$. Let $\hat{\mathcal{A}}^T_{\mathrm{ref}}$ be an extended system with initial conditions $\hat{\Upsilon}(t_0)$ and X a section of $L(\mathcal{A})$. If $v_0 \in \operatorname{int}(K(X, \hat{\Upsilon}(t_0), t_0, t))$ for $t \in [t_0, t_1]$, then there exists a convex cone $K \subseteq K(X, \hat{\Upsilon}(t_0), t_0, t)$ such that*

(i) $v_0 \in \operatorname{int}(K)$ and

(ii) $\{\Sigma(X(t), \hat{\Upsilon}(t_0), t_0, t) + \epsilon v \mid v \in K,\} \subseteq \mathcal{R}_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t_0), t, t_0)$ where $\epsilon \geq 0$ is sufficiently small.

5.4. The Hamiltonian.

The next required ingredient is a Hamiltonian. The Hamiltonian, $H_{\mathcal{A}^T_{\mathrm{ref}}}$, defined below differs from that of the general setup because of the previous identification of the fibers of the vertical subbundle of TTM with the fibers of TM. Note that since there is not a natural notion of exterior differentiation along a curve, this Hamiltonian does not give rise to Hamilton's equations. However, the following Hamiltonian does maintain the required maximization properties, see Lemmata 5.17 and 5.18, required to prove the Maximum Principle. Also, the Hamilton–Jacobi equations, using the analogue of $H_{\mathcal{A}^T_{\mathrm{ref}},\Lambda^0 L}$ in the standard setup, provide the desired value/form of the optimal cost and its relationship with the Riccati equation. Recall that $\iota^*(t)\colon T^*_{\gamma_{\mathrm{ref}}(t)}M \to L(\mathcal{A})^*_{(t,\gamma_{\mathrm{ref}}(t))}$ is the dual of the inclusion map.

5.16 Definition: Let $\mathcal{A}$ be an affine system with linearization $\mathcal{A}^T_{\mathrm{ref}}$ along γ_{ref}. Let $L\colon TM \times L(\mathcal{A}) \to \mathbb{R}$ be the Lagrangian defined by (5.1) and let $\hat{\mathcal{A}}^T_{\mathrm{ref}}$ be the extended system.

(i) The ***Hamiltonian*** is the function

$$\begin{aligned} H_{\mathcal{A}^T_{\mathrm{ref}}}\colon (TM \oplus T^*M) \times L(\mathcal{A}) &\to \mathbb{R} \\ (\Upsilon(t) \oplus \Lambda(t), X(t)) &\mapsto \langle \iota^*(t)\Lambda(t); X(t)\rangle. \end{aligned}$$

(ii) The ***maximum Hamiltonian*** is the function

$$H^{\max}_{\mathcal{A}^T_{\mathrm{ref}}}\colon TM \oplus T^*M \to \mathbb{R}$$
$$(\Upsilon(t) \oplus \Lambda(t)) \mapsto \sup\{H_{\mathcal{A}^T_{\mathrm{ref}}}(\Upsilon(t) \oplus \Lambda(t), X(t)) \mid X(t) \in L(\mathcal{A})_{t,\gamma_{\mathrm{ref}}(t)}\}.$$

(iii) The ***extended Hamiltonian*** is the function

$$H_{\mathcal{A}^T_{\mathrm{ref}},\Lambda^0 L}\colon ((TM \times \mathbb{R}) \oplus (T^*M \times \mathbb{R})) \times L(\mathcal{A}) \to \mathbb{R}$$
$$(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), X(t)) \mapsto \langle \iota^*(t)\Lambda(t); X(t)\rangle + \Lambda^0(t)L(\Upsilon(t), X(t)).$$

(iv) The ***extended maximum Hamiltonian*** is the function

$$H^{\max}_{\hat{\mathcal{A}}^T_{\mathrm{ref}},\Lambda^0 L}\colon (TM \times \mathbb{R}) \oplus (T^*M \times \mathbb{R}) \to \mathbb{R}$$
$$(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t)) \mapsto \sup\{H_{\hat{\mathcal{A}}^T_{\mathrm{ref}},\Lambda^0 L}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), X(t)) \mid X(t) \in L(\mathcal{A})_{t,\gamma_{\mathrm{ref}}(t)}\}.$$

•

The following lemmata provide a relationship between the Hamiltonian and the tangent cones of Section 5.3. It is interesting to note that these maximization statements only involve properties of tangent cones and do not rely on the optimal control problem data and, although they are stated for the extended system, they hold for general non-linear systems.

5.17 Lemma: (Hamiltonian maximization property) *Let $\mathcal{A}$ be an affine system with linearization $\mathcal{A}^T_{\mathrm{ref}}$ along γ_{ref} and extended system $\hat{\mathcal{A}}^T_{\mathrm{ref}}$. Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$ and let $(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), \bar{X}(t)) \in ((TM \times \mathbb{R}) \oplus (T^*M \times \mathbb{R})) \times L(\mathcal{A})$. Then $H_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), \bar{X}(t)) = H^{\max}_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t))$ if and only if,*

$$\langle (\iota^*(t)\Lambda(t), \Lambda^0(t)); v(t)\rangle \leq 0,$$

where $v(t) \in \{(X(t) - \bar{X}(t), L(\Upsilon(t), X(t) - \bar{X}(t))) \mid X(t) \in L(\mathcal{A})_{t,\gamma_{\mathrm{ref}}(t)}\}$.

Proof: Let X be a section of $L(\mathcal{A})$ along γ_{ref}. Then

$$\begin{aligned}
& H_{\mathcal{A}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), \bar{X}(t)) = H^{\max}_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t)) \\
\iff\ & H_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), \bar{X}(t)) \geq H_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), X(t)), \\
\iff\ & H_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), \bar{X}(t)) - H_{\hat{\mathcal{A}}^T_{\mathrm{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), X(t)) \geq 0 \\
\iff\ & \langle \iota^*(t)\Lambda(t); X(t) - \bar{X}(t)\rangle + \Lambda^0(t)L(\Upsilon(t), X(t) - \bar{X}(t)) \leq 0 \\
\iff\ & \langle (\iota^*(t)\Lambda(t), \Lambda^0(t)); v(t)\rangle \leq 0,
\end{aligned}$$

where $v(t) \in \{(X(t) - \bar{X}(t), L(\Upsilon(t), X(t) - \bar{X}(t))) \mid X(t) \in L(\mathcal{A})_{t,\gamma_{\mathrm{ref}}(t)}\}$. ■

5.18 Lemma: (Hamiltonian maximization and the fixed interval tangent cone) *Let $\mathcal{A}$ be an affine system with linearization $\mathcal{A}^T_{\text{ref}}$ along γ_{ref} and extended system $\hat{\mathcal{A}}^T_{\text{ref}}$. Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$ and let $X(t) \in L(\mathcal{A})_{t,\gamma_{\text{ref}}(t)}$. For each $t \in [t_0, t_1]$ let κ_t be a convex cone in $L(\mathcal{A})_{t,\gamma_{\text{ref}}(t)}$ containing $K(X(t), \hat{\Upsilon}(t_0), t_0, t)$ and suppose, for some time $\tau \in [t_0, t_1]$, that there exists a covector $(\Lambda(\tau), \Lambda^0(\tau)) \in T^*_{\gamma_{\text{ref}}(\tau)}M \times \mathbb{R}$ such that*

$$\langle (\iota^*(t)\Lambda(\tau), \Lambda^0(\tau)); v_\tau \rangle \leq 0, \qquad v_\tau \in \kappa_\tau.$$

Let $t \mapsto \hat{\Upsilon}(t) \oplus \hat{\Lambda}(t)$ be a solution to the extended adjoint equation for $\hat{\mathcal{A}}^T_{\text{ref}}$ along γ_{ref} with the above property at time τ. Then, for almost every $t \in [t_0, \tau]$,

$$H_{\hat{\mathcal{A}}^T_{\text{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t), X(t)) = H^{\max}_{\hat{\mathcal{A}}^T_{\text{ref}}}(\hat{\Upsilon}(t) \oplus \hat{\Lambda}(t)).$$

Proof: Let $\chi_t \in L(\mathcal{A})_{(t,\gamma_{\text{ref}}(t))}$. Then, by definition of the fixed interval tangent cone,

$$\Phi_{t,\tau}(\chi_t - X(t), L(\Upsilon, \chi_t - X(t))) \in K(X(\tau), \hat{\Upsilon}(t_0), t_0, \tau) \subseteq \kappa_\tau.$$

By hypothesis,

$$\langle (\iota^*(t)\Lambda(\tau), \Lambda^0(\tau)); \Phi_{t,\tau}(\chi_t, L(\Upsilon(t), \chi_t) \rangle \\ - \langle (\iota^*(t)\Lambda(\tau), \Lambda^0(\tau)); \Phi_{t,\tau}(X(t), L(\Upsilon(t), X(t))) \rangle \leq 0.$$

Now use the definition of the adjoint equations to obtain

$$\langle (\iota^*(t)\Lambda(t), \Lambda^0(t)); (\chi_t - X(t), L(\Upsilon(t), \chi_t - X(t)) \rangle \leq 0.$$

Since this holds for every $\chi_t \in L(\mathcal{A})_{t,\gamma_{\text{ref}}(t)}$, the lemma follows from Lemma 5.17. ■

5.5. Proof of Theorem 5.2. With the constructions of Sections 5.2, 5.3, and 5.4 for the extended system, the proof of the Maximum Principle is as follows. Note that $\hat{\mathcal{R}}_{\hat{\mathcal{A}}^T_{\text{ref}}}(v_x, t_0, t_1)$ denotes the reachable set from v_x for the extended system and is defined in a similar manner to $\mathcal{R}_{\mathcal{A}^T_{\text{ref}}}(v_x, t_0, t_1)$.

Theorem 5.2: *Let $\mathcal{A}$ be an affine system with a controllable linearisation $\mathcal{A}^T_{\text{ref}}$ along γ_{ref} and reference vector field X_{ref}. Let $t_0, t_1 \in \mathbb{R}$ satisfy $t_0 < t_1$ and let $v_{\gamma_{\text{ref}}(t_0)} \in T_{\gamma_{\text{ref}}(t_0)}M$. Suppose that (Υ_*, X_*) solves Problem 5.1(i). Then there exists a covector field $\Lambda_*\colon [t_0, t_1] \to T^*M$ along γ_{ref} such that the following equations on $TM \oplus T^*M$ hold:*

$$\begin{aligned}\dot{\Upsilon}_*(t) &= X^T_{\text{ref}}(t, \Upsilon_*(t)) + \text{vlft}(R^\sharp(t)\iota^*(t)\Lambda_*(t))(\Upsilon(t)),\\ \dot{\Lambda}_*(t) &= X^{T^*}_{\text{ref}}(t, \Lambda_*(t)) - \text{vlft}(FL(\Upsilon_*(t)))\Lambda_*(t),\end{aligned}$$

where $\Upsilon_*(t_0) = v_{\gamma_{\text{ref}}(t_0)}$ and $\Lambda_*(t_1) = F^\flat(t_1)\Upsilon(t_1)$.

Proof: Let (Υ_*, X_*) be an optimal trajectory for Problem 5.1(i). The proof relies on the construction of a hyperplane which is used to define the final condition for the extended adjoint equations.

Lemma $(0_{\gamma_{\text{ref}}(t_1)}, -1) \in T_{\gamma_{\text{ref}}(t_1)}M \times \mathbb{R}$ is not in the interior of $K(X_*, \hat{\Upsilon}(t_0), t_0, t_1)$.

Proof: At each time $t \in [t_0, t_1]$ the reachable set at time t for the extended system is contained in $T_{\gamma_{\text{ref}}(t)}M \times \mathbb{R}$, as is the fixed interval tangent cone. To prove the lemma, the important non-trivial fact that the fixed interval tangent cone is contained in the reachable set is utilized, Lemma 5.15. Suppose that $(0_{\gamma_{\text{ref}}(t_1)}, -1) \in \text{int}(K(X_*, \hat{\Upsilon}(t_0), t_0, t_1))$. Then $(0_{\gamma_{\text{ref}}(t_1)}, -1) \in \hat{\mathcal{R}}_{\hat{\mathcal{A}}^T_{\text{ref}}}(v_{\gamma_{\text{ref}}(t_0)}, t_0, t_1)$. This implies there are points in $\hat{\mathcal{R}}_{\hat{\mathcal{A}}^T_{\text{ref}}}(v_{\gamma_{\text{ref}}(t_0)}, t_0, t_1)$ whose final cost are lower than $\Upsilon^0_*(t_1)$ which contradicts the hypothesis that (Υ_*, X_*) is optimal. ▼

Since $K(X_*, \hat{\Upsilon}(t_0), t_0, t_1)$ is convex and $(0_{\gamma_{\text{ref}}(t_1)}, -1) \notin \text{int}(K(X_*, \hat{\Upsilon}(t_0), t_0, t_1))$, there exist a hyperplane $\hat{P}(t_1)$ separating $K(X_*, \hat{\Upsilon}(t_0), t_0, t_1)$ and $(0_{\gamma_{\text{ref}}(t_1)}, -1)$ [Lee and Markus 1967]. Let $\hat{\Lambda}_*(t_1) \in \text{ann}(\hat{P}(t_1))$ and note that

$$\begin{aligned}&\langle \hat{\Lambda}_*(t_1); (0_{\gamma_{\text{ref}}(t_1)}, -1)\rangle \geq 0,\\ &\langle \hat{\Lambda}_*(t_1); \hat{v}_{\gamma_{\text{ref}}(t_1)}\rangle \leq 0, \qquad \hat{v}_{\gamma_{\text{ref}}(t_1)} \in K(X_*, \hat{v}_{\gamma_{\text{ref}}(t_0)}, t_0, t_1).\end{aligned}$$

The first expression implies that $\Lambda^0_*(t_1) \leq 0$. Let $\hat{\Lambda}_*$ be the adjoint response with final condition $\hat{\Lambda}_*(t_1)$ which implies that $\Lambda^0_*(t)$ is also constant. If $\Lambda^0_* \neq 0$ then, without loss of generality, set $\Lambda^0_* = -1$ by redefining $\hat{\Lambda}_*$ as $-(\Lambda^0_*)^{-1}\hat{\Lambda}_*$. The linearity of the adjoint equation and a non-zero initial condition implies that $\hat{\Lambda}_* \neq (0_{\gamma_{\text{ref}}(t)}, 0)$ for every $t \in [t_0, t_1]$ which implies that either $\Lambda^0_* = -1$ or $\Lambda_*(t_0) \neq 0_{\gamma_{\text{ref}}(t_0)}$.

The next step is to prove by contradiction, using the controllability assumption, that $\Lambda^0_* \neq 0$. Suppose that $\Lambda^0_* = 0$. Then $(\Lambda_*, 0)$ satisfies the adjoint equations:

$$\begin{aligned}\dot{\Upsilon}(t) &= X^T_{\text{ref}}(t, \Upsilon(t)) + \text{vlft}(X(t))(\Upsilon(t)),\\ \dot{\Upsilon}^0(t) &= L(\Upsilon(t), X(t)),\\ \dot{\Lambda}(t) &= X^{T^*}_{\text{ref}}(t, \Lambda(t)),\\ \dot{\Lambda}^0(t) &= 0,\end{aligned}$$

where $\Lambda(t)$ is a covector field along γ_{ref}. The Hamiltonian for the extended system becomes

$$H_{\hat{\mathcal{A}}^T_{\text{ref}}, 0L}(\hat{\Upsilon}_*(t) \oplus \hat{\Lambda}_*(t), X_*(t)) = \langle \iota^*(t)\Lambda_*(t); X_*(t)\rangle.$$

The function $H_{\hat{\mathcal{A}}^T_{\text{ref}}, 0L}$ is a linear function of X_* and is maximized if $\iota^*(t)\Lambda_*(t) = 0$. Thus $\Lambda_*(t) \in \text{ann}(L(\mathcal{A})_{t, \gamma_{\text{ref}}(t)})$ for all $t \in [t_0, t_1]$. The controllability hypothesis then implies that $\Lambda_*(t) = 0_{\gamma_{\text{ref}}(t)}$ for all $t > t_0$. Since $\Lambda_*(t)$ satisfies the linear differential equation $\dot{\Lambda}(t) = X^{T^*}_{\text{ref}}(t, \Lambda(t))$, it follows that $\Lambda_*(t_0) = 0_{\gamma_{\text{ref}}(t_0)}$, which contradicts the non-triviality condition that either $\Lambda^0_* = -1$ or $\Lambda_*(t_0) \neq 0_{\gamma_{\text{ref}}(t_0)}$.

Assume $\Lambda^0_* = -1$. By Lemma 5.18, $\hat{\Lambda}_*$ maximizes the Hamiltonian $H_{\hat{\mathcal{A}}^T_{\text{ref}}, \Lambda^0 L}$. Since $H_{\hat{\mathcal{A}}^T_{\text{ref}}, \Lambda^0 L}$ is a quadratic function of X and $R(t)$ is positive-definite, the Hamiltonian is maximized if

$$\iota^*(t)\Lambda_*(t) + \Lambda^0_* R^\flat(t) X_*(t) = 0.$$

The above equation can be solved for the optimal control in terms of the costate variable to obtain

$$X_*(t) = -\frac{1}{\Lambda^0_*} R^\sharp(t)\iota^*(t)\Lambda_*(t).$$

The proof now follows from the form of the extended adjoint equations. ■

5.6. Characterizing the finite time LQR problem. The next theorem characterizes solutions to Problem 5.1(i) using Theorem 5.2 and introduces the geometric version of the differential Riccati equation.

5.19 Theorem: *The following are equivalent:*

(i) there exists a unique solution K, a symmetric $(0,2)$–tensor along γ_{ref} such that $K(t)$ is positive-definite for each $t \in [t_0, t_1]$, to the ***Riccati equation***

$$\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}} K(t) - K^{\flat}(t)\iota(t)R^{\sharp}(t)\iota^*(t)K^{\flat}(t) + Q(t) = 0, \qquad K(t_1) = F(t_1),$$

for each $t \in [t_0, t_1]$;

(ii) there exists a solution to Problem 5.1(i);

(iii) the pair $(\Upsilon(t), \Lambda(t))$ satisfy

$$\begin{aligned}\dot{\Upsilon}(t) &= X^{T}_{\mathrm{ref}}(t, \Upsilon(t)) + \mathrm{vlft}(R^{\sharp}(t)\iota^*(t)\Lambda(t))(\Upsilon(t)),\\ \dot{\Lambda}(t) &= X^{T^*}_{\mathrm{ref}}(t, \Lambda(t)) - \mathrm{vlft}(FL(\Upsilon(t)))\Lambda(t),\end{aligned}$$

where $\Upsilon(t_0) = v_{\gamma_{\mathrm{ref}}(t_0)}$ and $\Lambda(t_1) = F^{\flat}(t_1)\Upsilon(t_1)$.

Proof: (i)⇒(ii) Let $\Upsilon \in \mathrm{Traj}(\mathcal{A}^{T}_{\mathrm{ref}}, t_1, t_0)$ and let K be a symmetric $(0,2)$–tensor field along γ_{ref} such that $K(t)$ is positive-definite for each $t \in [t_0, t_1]$ and that K satisfies the Riccati equation with final condition $K(t_1) = F(t_1)$. Note that $K^{\flat}(t)\colon T_{\gamma_{\mathrm{ref}}(t)}M \to T^*_{\gamma_{\mathrm{ref}}(t)}M$ is a linear map. Now integrate both sides of

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle K^{\flat}(t)\Upsilon(t); \Upsilon(t)\rangle = \frac{1}{2}\left(\langle \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(K^{\flat}(t)\Upsilon(t)); \Upsilon(t)\rangle + \langle K^{\flat}(t)\Upsilon(t); \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(\Upsilon(t))\rangle\right),$$

over the interval $[t_0, t_1]$ and add the result to the cost to obtain

$$\begin{aligned} J(\Upsilon(t_0), t_0, t_1) &+ \frac{1}{2}\langle K^{\flat}(t_1)\Upsilon(t_1); \Upsilon(t_1)\rangle - \frac{1}{2}\langle K^{\flat}(t_0)\Upsilon(t_0); \Upsilon(t_0)\rangle\\ &= \frac{1}{2}\int_{t_0}^{t_1} Q(t)(\Upsilon(t), \Upsilon(t)) + R(t)(X, X) + \langle \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(K^{\flat}(t)\Upsilon(t)); \Upsilon(t)\rangle\\ &\qquad + \langle K^{\flat}(t)\Upsilon(t); \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(\Upsilon(t))\rangle \,\mathrm{d}t + \frac{1}{2}F(t_1)(\Upsilon(t_1), \Upsilon(t_1)).\end{aligned}$$

Now in local coordinates—for brevity the time dependence is no longer indicated—the right hand side is

$$\begin{aligned} &\frac{1}{2}\int_{t_0}^{t_1} Q_{ij}\Upsilon^i\Upsilon^j + R_{ij}X^iX^j + \left(\frac{\mathrm{d}}{\mathrm{d}t}(K_{ij}\Upsilon^i) + \frac{\partial X^{\ell}_{\mathrm{ref}}}{\partial x^j}K_{i\ell}\Upsilon^i\right)\Upsilon^j + K_{ij}\Upsilon^i\iota(t)X^j \,\mathrm{d}t\\ &= \frac{1}{2}\int_{t_0}^{t_1} Q_{ij}\Upsilon^i\Upsilon^j + R_{ij}X^iX^j + \frac{\mathrm{d}K_{ij}}{\mathrm{d}t}\Upsilon^i\Upsilon^j + K_{ij}\frac{\partial X^{i}_{\mathrm{ref}}}{\partial x^{\ell}}\Upsilon^{\ell}\Upsilon^j\\ &\quad + K_{ij}\iota(t)X^i\Upsilon^j\frac{\partial X^{\ell}_{\mathrm{ref}}}{\partial x^j}K_{i\ell}\Upsilon^i\Upsilon^j + K_{ij}\Upsilon^i\iota(t)X^j \,\mathrm{d}t\\ &= \frac{1}{2}\int_{t_0}^{t_1} R_{ij}(X^i + R^{is}\iota^*(t)K_{s\ell}\Upsilon^{\ell})(X^j + R^{js}\iota^*(t)K_{s\ell}\Upsilon^{\ell})\\ &\quad + \left(\frac{\mathrm{d}K_{ij}}{\mathrm{d}t} + K_{sj}\frac{\partial X^{s}_{\mathrm{ref}}}{\partial x^i} + \frac{\partial X^{\ell}_{\mathrm{ref}}}{\partial x^j}K_{i\ell} - K_{\ell i}\iota(t)R^{\ell s}\iota^*(t)K_{sj} + Q_{ij}\right)\Upsilon^i\Upsilon^j \,\mathrm{d}t.\end{aligned}$$

Using the hypothesis, the cost becomes

$$J(\Upsilon(t_0), t_0, t_1) = \frac{1}{2}\langle K^\flat(t_0)\Upsilon(t_0); \Upsilon(t_0)\rangle + \frac{1}{2}\int_{t_0}^{t_1} \|X(t) + R^\sharp(t)\iota^*(t)K^\flat(t)\Upsilon(t)\|^2_{R(t)}\, \mathrm{d}t$$

and the cost is then minimized by choosing a trajectory Υ such that

$$X(t) = -R^\sharp(t)\iota^*(t)K^\flat(t)\Upsilon(t). \tag{5.11}$$

(ii)⇒(iii) By the Maximum Principle, Theorem 5.2, this follows.

(iii)⇒(i) The following lemma will aid in the construction of a solution to the Riccati equation.

Lemma Let $\Upsilon\colon I \to TM$ and $\Lambda\colon I \to T^*M$ be vector and covector fields along γ_{ref}, respectively. Consider the following statements:

(i) the pair (Υ, Λ) satisfy

$$\begin{aligned}\dot{\Upsilon}(t) &= X^T_{\text{ref}}(t, \Upsilon(t)) + \text{vlft}(R^\sharp(t)\iota^*(t)\Lambda(t))(\Upsilon(t)),\\ \dot{\Lambda}(t) &= X^{T^*}_{\text{ref}}(t, \Lambda(t)) - \text{vlft}(FL(\Upsilon(t)))\Lambda(t),\end{aligned}$$

where $\Upsilon(t_0) = v_{\gamma_{\text{ref}}(t_0)}$ and $\Lambda(t_1) = F^\flat(t_1)\Upsilon(t_1)$;

(ii) $\Upsilon(t) = K_1(t)\Phi^{X^T_{\text{ref}}}_{t_1,t}(\Upsilon(t_1))$, $\Lambda(t) = K_2(t)\Phi^{X^T_{\text{ref}}}_{t_1,t}(\Upsilon(t_1))$ where $K_1(t)\colon T_{\gamma_{\text{ref}}(t)}M \to T_{\gamma_{\text{ref}}(t)}M$ and $K_2(t)\colon T_{\gamma_{\text{ref}}(t)}M \to T^*_{\gamma_{\text{ref}}(t)}M$ are linear maps for each $t \in [t_0, t_1]$ and satisfy

$$\begin{aligned}\mathscr{L}^{X_{\text{ref}},\gamma_{\text{ref}}}K_1(t) &= \iota(t)R^\sharp(t)\iota^*(t)K_2(t),\\ \mathscr{L}^{X_{\text{ref}},\gamma_{\text{ref}}}K_2(t) &= Q^\flat(t)K_1(t),\end{aligned}$$

where K_1 is invertible, $K_1(t_1) = \text{id}_{TM}$, and $K_2(t_1) = F(t_1)$.

Then (i)⇒(ii).

Proof: Note that $FL(\Upsilon(t)) = Q^\flat(t)\Upsilon(t)$. Assuming (i), let (Υ, Λ) be the pair satisfying

$$\begin{aligned}\mathscr{L}^{X_{\text{ref}},\gamma_{\text{ref}}}(\Upsilon(t)) &= \iota(t)R^\sharp(t)\iota^*(t)\Lambda(t),\\ \mathscr{L}^{X_{\text{ref}},\gamma_{\text{ref}}}(\Lambda(t)) &= -Q^\flat(t)\Upsilon(t),\end{aligned}$$

with initial conditions $\Upsilon(t_0) = v_{\gamma_{\text{ref}}(t_0)}$ and $\Lambda(t_1) = F^\flat(t_1)\Upsilon(t_1)$. Let $\eta(t) = \Phi^{X^T_{\text{ref}}}_{t_1,t}(\Upsilon(t_1))$. That is, η is the integral curve of X^T_{ref} such that $\eta(t_1) = \Upsilon(t_1)$. Next define the pair $(\hat{\Upsilon}(t), \hat{\Lambda}(t))$ by $\hat{\Upsilon}(t) = K_1(t)\eta(t)$ and $\hat{\Lambda}(t) = K_2(t)\eta(t)$, respectively. The coordinate calculations

$$\begin{aligned}(\mathscr{L}^{X_{\text{ref}},\gamma_{\text{ref}}}(\hat{\Upsilon}(t))^i &= (\mathscr{L}^{X_{\text{ref}},\gamma_{\text{ref}}}(K_1(t)\eta(t)))^i\\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t}K^i_{1,\ell}(t)\right)\eta^l(t) + K^i_{1,s}(t)\frac{\partial X^s_{\text{ref}}}{\partial x^\ell}(\eta(t))\eta^\ell(t) - \frac{\partial X^i_{\text{ref}}}{\partial x^j}(\eta(t))K^j_{1,\ell}(t)\eta^\ell(t)\\ &= \frac{\partial K^i_{1,\ell}(t)}{\partial x^s}X^s_{\text{ref}}(\eta(t))\eta^l(t) + K^i_{1,s}(t)\frac{\partial X^s_{\text{ref}}}{\partial x^\ell}(\eta(t))\eta^\ell(t)\\ &\quad - \frac{\partial X^i_{\text{ref}}}{\partial x^j}(\eta(t))K^j_{1,\ell}(t)\eta^\ell(t)\\ &= (\mathscr{L}^{X_{\text{ref}},\gamma_{\text{ref}}}K_1(t)\eta(t))^i = (\iota(t)R^\sharp(t)\iota^*(t)\hat{\Lambda}(t))^i,\end{aligned}$$

and similarly

$$\begin{aligned}(\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(\hat{\Lambda}(t))^i &= (\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(K_2(t)\eta(t)))^i \\ &= (\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}K_2(t)\eta(t))^i = -(Q^\flat(t)\hat{\Upsilon}(t))^i,\end{aligned}$$

show that $(\hat{\Upsilon}(t), \hat{\Lambda}(t))$ and $(\Upsilon(t), \Lambda(t))$ satisfy the same differential equation. Since $K_1(t_1) = \mathrm{id}_{TM}$, and $K_2(t_1) = F(t_1)$, it follows that $(\hat{\Upsilon}(t_1), \hat{\Lambda}(t_1)) = (\Upsilon(t_1), F^\flat(t_1)\Upsilon(t_1))$, and the lemma follows by the uniqueness of solutions to differential equations. ▼

It is now shown that, given any $w_{\gamma_{\mathrm{ref}}(\tau)} \in T_{\gamma_{\mathrm{ref}}(\tau)}M$, $w_{\gamma_{\mathrm{ref}}(\tau)} \in \mathrm{image}(K_1(\tau))$. Let (Υ_*, Λ_*) satisfy (iii) with $\Upsilon(\tau) = w_{\gamma_{\mathrm{ref}}(\tau)}$ and $\Lambda_*(t_1) = F^\flat(t_1)\Upsilon(t_1)$. By the lemma it follows that

$$w_{\gamma_{\mathrm{ref}}(\tau)} = \Upsilon_*(\tau) = K_1(\tau)\Phi^{X^T_{\mathrm{ref}}}_{t_1,\tau}(\Upsilon_*(t_1))$$

which implies that $K_1(\tau)$ is invertible.

If (Υ, Λ) satisfy (iii) and $K_1(t)$ is invertible for all $t \in [t_0, t_1]$, then by the lemma the following linear relationship holds:

$$\Lambda(t) = K_2(t)K_1^{-1}(t)\Upsilon(t).$$

To show that $K(t) = K_2(t)K_1^{-1}(t)$ satisfies the Riccati equation in statement (i), it is first observed that $K(t_1) = F(t_1)$ by construction. Using the linear relationship $\Lambda(t) = K(t)\Upsilon(t)$, the costate equation is

$$\langle Q^\flat(t)\Upsilon(t); \Upsilon(t)\rangle + \langle \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(K^\flat(t)\Upsilon(t)); \Upsilon(t)\rangle = 0.$$

In coordinates the above equation becomes

$$\begin{aligned}\dot{K}_{ij}(t)\Upsilon^i(t)\Upsilon^j(t) + K_{ij}(t)\left(\frac{\partial X^i_{\mathrm{ref}}}{\partial x^\ell}(\gamma_{\mathrm{ref}}(t)) - \iota(t)R^{si}(t)\iota^*(t)\Lambda_s(t)\right)\Upsilon^j(t)& \\ + \frac{\partial X^\ell_{\mathrm{ref}}}{\partial x^j}(\gamma_{\mathrm{ref}}(t))K_{i\ell}(t)\Upsilon^i(t)\Upsilon^j(t) + Q_{ij}(t)\Upsilon^i(t)\Upsilon^j(t) &= 0.\end{aligned}$$

Again, using the linear relationship of state and costate implies that

$$\begin{aligned}\dot{K}_{ij}(t) + K_{\ell j}(t)\frac{\partial X^\ell_{\mathrm{ref}}}{\partial x^i}(\gamma_{\mathrm{ref}}(t)) + \frac{\partial X^\ell_{\mathrm{ref}}}{\partial x^j}(\gamma_{\mathrm{ref}}(t))K_{i\ell}(t)-& \\ K_{\ell j}(t)\iota(t)R^{s\ell}(t)\iota^*(t)K_{is}(t) + Q_{ij}(t) &= 0.\end{aligned}$$

Thus $K(t)$ satisfies

$$\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}K(t) - K^\flat(t)\iota(t)R^\sharp(t)\iota^*(t)K^\flat(t) + Q(t) = 0.$$ ■

5.20 Remark: The coordinate expression

$$\begin{aligned}\dot{K}_{ij}(t) + K_{\ell j}(t)\frac{\partial X^\ell_{\mathrm{ref}}}{\partial x^i}(\gamma_{\mathrm{ref}}(t)) + \frac{\partial X^\ell_{\mathrm{ref}}}{\partial x^j}(\gamma_{\mathrm{ref}}(t))K_{i\ell}(t)& \\ - K_{\ell j}(t)\iota(t)R^{s\ell}(t)\iota^*(t)K_{is}(t) + Q_{ij}(t) &= 0,\end{aligned}$$

recovers the standard Riccati equation

$$\dot{K}(t) + K(t)A(t) + A^T(t)K(t) - K(t)B(t)R^{-1}(t)B^T(t)K(t) + Q(t) = 0.$$ •

5.7. Infinite time LQR problems. In this section the solution to Problem 5.1(ii) is constructed by extending the ideas of Theorem 5.19 to the infinite time case. This will require various uniformity bounds on the problem data and ensuring the existence of solutions to the Riccati equation as the final time tends to infinity.

The idea is to construct the analogue of the minimum energy controller in the present geometric framework. We quickly recall from [Brockett 1970] the development of this controller. We suppose that we have a controllable time-varying linear system (A, B). Given an initial state $x_0 \in \mathbb{R}^n$ and $t_0, t_1 \in \mathbb{R}$ satisfying $t_0 < t_1$, we seek a control $u\colon I \to \mathbb{R}^m$ that steers x_0 to the origin at time t_1 while minimizing the energy

$$\int_{t_0}^{t_1} \|u(t)\|_{\mathbb{R}^m} \mathrm{d}t.$$

One can show that the control is given by

$$u(\sigma) = -B^T(\sigma)\Phi^T(t_0, \sigma)\eta,$$

where η satisfies $W(t_0, t_1)\eta = x_0$. The trajectory corresponding to this control is

$$x(\tau) = \Phi(\tau, t_0)\left(x_0 - W(t_0, \tau)W^{-1}(t_0, t_1)x_0\right).$$

To develop our analogue to this minimum energy control law, define $W(t, t_1)$, a $(2, 0)$–tensor on $T_{\gamma_{\mathrm{ref}}(t)}M$, as the solution to

$$\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}} W^\flat(t, t_1) = \iota(t)R(t)^\sharp \iota^*(t). \tag{5.12}$$

The differential equation (5.12) should be thought of as the geometric analogue of the formula

$$\frac{\mathrm{d}}{\mathrm{d}t}W(t, t_1) = A(t)W(t, t_1) + W(t, t_1)A^T(t) - B(t)B^T(t) \qquad W(t_1, t_1) = 0,$$

for the derivative of the standard controllability Gramian (4.3). The rule for differentiating with respect to second parameter is provided by the following lemma.

5.21 Lemma: *If $W(t, t_1)$ is a $(2, 0)$–tensor on $T_{\gamma_{\mathrm{ref}}(t)}M$ satisfying $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}} W^\flat(t, \tau) = \iota(t)R(t)^\sharp\iota^*(t)$, then*

$$\frac{\mathrm{d}}{\mathrm{d}\tau}W(t, \tau) = \Phi_{\tau,t}^{X_{\mathrm{ref}}^T} \iota(\tau)R^\sharp(\tau)\iota^*(\tau)\Phi_{t,\tau}^{X_{\mathrm{ref}}^{T*}}.$$

Proof: In coordinates $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}} W^\flat(t, \tau) = \iota(t)R(t)^\sharp\iota^*(t)$ is given by

$$\begin{aligned}
&\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}} W^{ij}(t, \tau) - R^{ij}(t) = 0\\
\Rightarrow\quad & X_{\mathrm{ref}}^k(\gamma_{\mathrm{ref}}(t))\frac{\partial W^{ij}}{\partial x^k}(t, \tau) - W^{kj}(t, \tau)\frac{\partial X_{\mathrm{ref}}^i}{\partial x^k} - W^{ik}(t, \tau)\frac{\partial X_{\mathrm{ref}}^j}{\partial x^k} - R^{ij}(t) = 0\\
\Rightarrow\quad & \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}}{\mathrm{d}\tau}W^{ij}(t, \tau)\right) - \frac{\mathrm{d}}{\mathrm{d}\tau}\left(W^{kj}(t, \tau)\right)\frac{\partial X_{\mathrm{ref}}^i}{\partial x^k} - \frac{\mathrm{d}}{\mathrm{d}\tau}\left(W^{ik}(t, \tau)\right)\frac{\partial X_{\mathrm{ref}}^j}{\partial x^k} = 0.
\end{aligned} \tag{5.13}$$

To complete the proof it is shown that $\Phi^{X^T_{\rm ref}}_{\tau,t}\iota(\tau)R^\sharp(\tau)\iota^*(\tau)\Phi^{X^{T^*}_{\rm ref}}_{t,\tau}$ also satisfies the differential equation (5.13). Applying the "backward differentiation lemma" [Abraham, Marsden, and Ratiu 1988],

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi^{X^T_{\rm ref}}_{t,\tau}(v_x) = -T\Phi^{X^T_{\rm ref}}_{t,\tau}\circ X^T_{\rm ref}(v_x)$$

yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi^{X^T_{\rm ref}}_{\tau,t}\iota(\tau)R^\sharp(\tau)\iota^*(\tau)\Phi^{X^{T^*}_{\rm ref}}_{t,\tau}\right) = X^T_{\rm ref}(\Phi^{X^T_{\rm ref}}_{\tau,t}\iota(\tau)R^\sharp(\tau)\iota^*(\tau)\Phi^{X^{T^*}_{\rm ref}}_{t,\tau}) \\ - T_{\gamma_{\rm ref}(\tau))}\Phi^{X^T_{\rm ref}}_{\tau,t}\iota(\tau)R^\sharp(\tau)\iota^*(\tau)X^{T^*}_{\rm ref}\Phi^{X^{T^*}_{\rm ref}}_{t,\tau},$$

which in coordinates becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\Phi^{X^T_{\rm ref}}_{\tau,t}\iota(\tau)R^\sharp(\tau)\iota^*(\tau)\Phi^{X^{T^*}_{\rm ref}}_{t,\tau}\right)^{sj} = \frac{\partial X^s_{\rm ref}}{\partial x^\ell}(\Phi^{X^T_{\rm ref}}_{\tau,t_0})^\ell_k R^{ki}(\tau)(\Phi^{X^{T^*}_{\rm ref}}_{t_0,\tau})^j_i \\ + (\Phi^{X^T_{\rm ref}}_{\tau,t_0})^s_k R^{ki}(\tau)(\Phi^{X^{T^*}_{\rm ref}}_{t_0,\tau})^j_\ell\frac{\partial X^\ell_{\rm ref}}{\partial x^i},$$

providing the desired result. ■

The next lemma plays a central role in the rest of the proof.

5.22 Lemma: *If $\mathcal{A}^T_{\rm ref}$ is controllable from $v_0 \in T_{\gamma_{\rm ref}(t_0)}M$, then there exists a section X_1 of $L(\mathcal{A})$, linear in v_0, such that, for the resulting trajectory $\Upsilon_1 \in \mathrm{Traj}(\mathcal{A}^T_{\rm ref}, t_1, t_0)$ and time $t_2(v_0, t_0) \leq t_1$, the following hold:*

(i) $\Upsilon_1(t_0) = v_0$ *and*

(ii) $\Upsilon_1(t) = 0_{\gamma_{\rm ref}(t)}$ *for all* $t \geq t_2(v_0, t_0)$.

Proof: In line with the standard minimum energy controller [Brockett 1970], define a vector field along $\gamma_{\rm ref}$ by

$$\Upsilon(\tau) = \Phi^{X^T_{\rm ref}}_{t_0,\tau}(\Upsilon(t_0) - W^\flat(t_0,\tau)W^\sharp(t_0,t_1)\Upsilon(t_0)), \tag{5.14}$$

where $\Upsilon(t_0) = v_0$. The lemma will follow if (5.14) is a trajectory of $\mathcal{A}^T_{\rm ref}$. We claim, this is the trajectory prescribed by the following section of $L(\mathcal{A})$:

$$X(t) = -R(t)^\sharp\iota^*(t)\Phi^{X^{T^*}}_{t_0,t}\,\eta,$$

where $\eta \in T_{\gamma_{\rm ref}(t_0)}M$ satisfies $W(t_0,t_1)\eta = \Upsilon(t_0)$. If $\Upsilon(\tau)$, as defined in (5.14), is a trajectory, then $\Upsilon(\tau)$ must satisfy

$$\mathscr{L}^{\gamma_{\rm ref},X_{\rm ref}}(\Upsilon(\tau)) = -\iota(\tau)R(\tau)^\sharp\iota^*(\tau)\Phi^{X^{T^*}}_{t_0,\tau}\,\eta.$$

Consider the following coordinate computations:

$$
\begin{aligned}
&\mathscr{L}^{\gamma_{\mathrm{ref}},X_{\mathrm{ref}}}((\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}W^\flat(t_0,\tau)\eta))^i\\
&\quad= \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}W^\flat(t_0,\tau)\eta\right)^i - \frac{\partial X^i_{\mathrm{ref}}}{\partial x^k}(\gamma_{\mathrm{ref}}(\tau))\left(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}W^\flat(t_0,\tau)\eta\right)^k\\
&\quad= \frac{\mathrm{d}}{\mathrm{d}\tau}\left((\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^i_k W^{kj}(t_0,\tau)\eta_j\right) - \frac{\partial X^i_{\mathrm{ref}}}{\partial x^k}(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^k_s W^{sj}(t_0,\tau)\eta_j\\
&\quad= \frac{\partial X^i_{\mathrm{ref}}}{\partial x^k}(\gamma_{\mathrm{ref}}(\tau))\left((\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^k_s W^{sj}(t_0,\tau)\eta_j\right)\\
&\qquad+ (\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^i_k\frac{\mathrm{d}}{\mathrm{d}\tau}\left(W^{kj}(t_0,\tau)\right)\eta_j - \frac{\partial X^i_{\mathrm{ref}}}{\partial x^k}(\gamma_{\mathrm{ref}}(\tau))\left(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^k_s W^{sj}(t_0,\tau)\eta_j\right)\\
&\quad= (\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^i_k\frac{\mathrm{d}}{\mathrm{d}\tau}\left(W^{kj}(t_0,\tau)\right)\eta_j\\
&\quad= (\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^i_k(\Phi^{X^T_{\mathrm{ref}}}_{\tau,t_0})^k_\ell R^{\ell s}(\tau)(\Phi^{X^{T^*}_{\mathrm{ref}}}_{t_0,\tau})^j_s\eta_j.\\
&\quad= R^{is}(\tau)(\Phi^{X^{T^*}_{\mathrm{ref}}}_{t_0,\tau})^j_s\eta_j = (\iota(\tau)R(\tau)^\sharp\iota^*(\tau)\Phi^{X^{T^*}_{\mathrm{ref}}}_{t_0,\tau}\eta)^i
\end{aligned}
$$

and

$$
\begin{aligned}
\mathscr{L}^{\gamma_{\mathrm{ref}},X_{\mathrm{ref}}}(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}\Upsilon(t_0)) &= \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}\Upsilon(t_0)\right)^i - \frac{\partial X^i_{\mathrm{ref}}}{\partial x^k}\left(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}\Upsilon(t_0)\right)^k\\
&= \frac{\partial X^i_{\mathrm{ref}}}{\partial x^k}(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^k_j\Upsilon^j(t_0) - \frac{\partial X^i_{\mathrm{ref}}}{\partial x^k}\left((\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau})^k_j\Upsilon^j(t_0)\right) = 0.
\end{aligned}
$$

Combining the above coordinate calculations gives

$$
\begin{aligned}
\mathscr{L}^{\gamma_{\mathrm{ref}},X_{\mathrm{ref}}}(\Upsilon(\tau)) &= \mathscr{L}^{\gamma_{\mathrm{ref}},X_{\mathrm{ref}}}(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}(\Upsilon(t_0) - W^\flat(t_0,\tau)W^\sharp(t_0,t_1)\Upsilon(t_0)))\\
&= \mathscr{L}^{\gamma_{\mathrm{ref}},X_{\mathrm{ref}}}(\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}\Upsilon(t_0)) - \mathscr{L}^{\gamma_{\mathrm{ref}},X_{\mathrm{ref}}}((\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}W^\flat(t_0,\tau)\eta))\\
&= -\iota(\tau)R(\tau)^\sharp\iota^*(\tau)\Phi^{X^{T^*}_{\mathrm{ref}}}_{t_0,\tau}\eta,
\end{aligned}
$$

as desired. ■

Now Lemmata 5.21 and 5.22 are used to prove that solutions to the Riccati equation exist as the final time in the LQR problem tends to infinity.

5.23 Proposition: *For fixed t_1, let $K_{t_1}(t)$ be a solution to*

$$
\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}K(t) - K^\flat(t)\iota(t)R^\sharp(t)\iota^*(t)K^\flat(t) + Q(t) = 0, \qquad K(t_1) = 0, \tag{5.15}
$$

for each $t \in [t_0, t_1]$. If $\mathcal{A}^T_{\mathrm{ref}}$ is controllable, then

(i) $\lim_{t_1\to\infty} K_{t_1}(t) = \bar{K}(t)$ *and*

(ii) $\bar{K}(t)$ *satisfies (5.15).*

Proof: (i) Let $\Upsilon_* \in \mathrm{Traj}(\mathcal{A}^T_{\mathrm{ref}}, t_1, t_0)$ be a vector field along γ_{ref} that minimizes the cost

$$
J(\Upsilon(t_0), t_0, t_1) = \frac{1}{2}\int_{t_0}^{t_1} Q(t)(\Upsilon(t),\Upsilon(t)) + R(t)(X(t),X(t))\,\mathrm{d}t
$$

provided by Theorem 5.19. Since $\mathcal{A}_{\text{ref}}^T$ is controllable, there exists $\Upsilon_1 \in \operatorname{Traj}(\mathcal{A}_{\text{ref}}^T, t_1, t_0)$ and a time $t_2(\Upsilon_*(t_0), t_0) \leq t_1$ such that

1. $\Upsilon_1(t_0) = \Upsilon_*(t_0)$ and
2. $\Upsilon_1(t) = 0_{\gamma_{\text{ref}}(t)}$ for all $t \geq t_2(\Upsilon_*(t_0), t_0)$.

The idea is to choose, by Lemma 5.22, a control that steers the system to zero along the reference curve in time $t_2 \leq t_1$ and which depends on both the initial condition $\Upsilon_*(t_0)$ and t_0. Once the system is at zero, the control is set to zero and because the system is linear the state remains zero along the reference curve.

The calculations in the proof of Theorem 5.19 show that the value of the minimal cost and the solution to the Riccati equation are related by

$$K_{t_1}(t_0)(\Upsilon_*(t_0), \Upsilon_*(t_0)) = J(\Upsilon_*(t_0), t_0, t_1).$$

The optimality of the trajectory $\Upsilon_*(t)$ implies that

$$\begin{aligned} K_{t_1}(t_0)(\Upsilon_*(t_0), \Upsilon_*(t_0)) = J(\Upsilon_*(t_0), t_0, t_1) &\leq J(\Upsilon_1(t_0), t_0, t_1) \\ &= \frac{1}{2}\int_{t_0}^{t_2} Q(\tau)(\Upsilon_1(\tau), \Upsilon_1(\tau)) + R(\tau)(X_1(\tau), X_1(\tau))\, \mathrm{d}\tau \\ &\quad + \frac{1}{2}\int_{t_2}^{t_1} Q(\tau)(\Upsilon_1(\tau), \Upsilon_1(\tau)) + R(\tau)(X_1(\tau), X_1(\tau))\, \mathrm{d}\tau \\ &= \frac{1}{2}\int_{t_0}^{t_2} Q(\tau)(\Upsilon_1(\tau), \Upsilon_1(\tau)) + R(\tau)(X_1(\tau), X_1(\tau))\, \mathrm{d}\tau. \end{aligned}$$

The trajectory

$$\Upsilon_1(\tau) = \Phi_{t_0,\tau}^{X_{\text{ref}}^T}(\Upsilon(t_0) - W^\flat(t_0, \tau) W^\sharp(t_0, t_2)\Upsilon(t_0)),$$

and the control

$$X_1(\tau) = -R(\tau)^\sharp \iota^*(\tau) \Phi_{t_0,\tau}^{X^{T^*}} W^\sharp(t_0, t_2)\Upsilon(t_0),$$

are linear in the initial condition, which implies that

$$\begin{aligned} K_{t_1}(t_0)(\Upsilon_*(t_0), \Upsilon_*(t_0)) &\leq \frac{1}{2}\int_{t_0}^{t_2} Q(\tau)(\Upsilon_1(\tau), \Upsilon_1(\tau)) + R(\tau)(X_1(\tau), X_1(\tau))\, \mathrm{d}\tau \\ &\leq f(t_0, t_2)(\Upsilon_*(t_0), \Upsilon_*(t_0)), \end{aligned}$$

where $f(t_0, t_2)$ is a non-negative (0,2)-tensor on $T_{\gamma_{\text{ref}}(t_0)}M$. This ensures that, for each t_0, the bilinear map $K_{t_1}(t_0)\colon T_{\gamma_{\text{ref}}(t_0)}M \times T_{\gamma_{\text{ref}}(t_0)}M \to \mathbb{R}$ satisfies,

$$\sup_{t_1 \in \mathbb{R}} |K_{t_1}(t_0)(\Upsilon_*(t_0), \Upsilon_*(t_0))| \leq \sup_{t_1 \in \mathbb{R}} f(t_0, t_2)(\Upsilon_*(t_0), \Upsilon_*(t_0)) < \infty.$$

Since $K_{t_1}(t_0)(\Upsilon(t_0), \Upsilon(t_0)) = J(\Upsilon(t_0), t_0, t_1)$ and $J(\Upsilon(t_0), t_0, t_1) \leq J(\Upsilon(t_0), t_0, t_2)$ for all $t_1 \leq t_2$, the map $t_1 \mapsto K_{t_1}(t_0)(\Upsilon(t_0), \Upsilon(t_0))$ is non-decreasing as $t_1 \to \infty$. Using the above facts, the limit $\bar{K}(t_0)$ is constructed as follows. Define a quadratic form $K(t_0)$ by

$$K(t_0)(v, v) = \lim_{t_1 \to \infty} \langle K_{t_1}^\flat(t_0)v, v\rangle, \qquad v \in T_{\gamma_{\text{ref}}(t_0)}M. \tag{5.16}$$

The limit in (5.16) exists because it is a bounded non-deceasing function of t_1. The polarization identity,

$$4\bar{K}(t_0)(v,w) = K(t_0)(v+w, v+w) - K(t_0)(v-w, v-w),$$

then defines a symmetric $(0,2)$–tensor $\bar{K}(t_0)$ for all $v, w \in T_{\gamma_{\text{ref}}(t_0)}M$.

(ii) Let $t \in [t_0, t_1]$ and let $K_{t_1}(t)$ be a solution to the Riccati equation with final condition $K_{t_1}(t_1) = 0$. By part (i) of the proof

$$\bar{K}(t) = \lim_{t_1 \to \infty} K_{t_1}(t),$$

exists. Now suppose, for $t \in [t_0, \tau_1]$, $\tau_1 \le t_1$, that $P_{\tau_1}(t)$ is a solution to the Riccati equation with final condition $P_{\tau_1}(\tau_1) = K_{t_1}(\tau_1)$. Then, the continuity of solutions to differential equations implies that $P_{\tau_1}(t) = K_{t_1}(t)$ for $t \le \tau_1 \le t_1$. Thus, it follows that

$$\bar{K}(t) = \lim_{t_1 \to \infty} K_{t_1}(t) = \lim_{t_1 \to \infty} P_{\tau_1}(t) = P_{\tau_1}(t),$$

with final condition $\bar{K}(\tau_1) = P_{\tau_1}(\tau_1)$. ■

5.24 Theorem: *The trajectory corresponding to the section of $L(\mathcal{A})$ defined by*

$$X(t) = -R^\sharp(t)\iota^*(t)\bar{K}^\flat(t)\Upsilon(t) \tag{5.17}$$

is optimal in the sense of Problem 5.1(ii).

Proof: Let $\bar{J}(\Upsilon(t_0), t_0, t_1)$ be the cost associated with (5.17) on the interval $[t_0, t_1]$.

Lemma The trajectory $\Upsilon(t)$ corresponding to (5.17) has the associated cost

$$\bar{J}(\Upsilon(t_0), t_0, \infty) := \lim_{t_1 \to \infty} \bar{J}(\Upsilon(t_0), t_0, t_1) = \bar{K}(t_0)(\Upsilon(t_0), \Upsilon(t_0)).$$

Proof: Let $J(\Upsilon(t_0), t_0, t_1)$ be the optimal cost. By Proposition 5.23, for $\epsilon > 0$ there exists $T > t_0$ such that

$$\bar{J}(\Upsilon(t_0), t_0, t_1) \ge J(\Upsilon(t_0), t_0, t_1) = K_{t_1}(t_0)(\Upsilon(t_0), \Upsilon(t_0)) \ge \bar{K}(t_0)(\Upsilon(t_0), \Upsilon(t_0)) - \epsilon,$$

for all $t_1 \ge T$. On the other hand, $\bar{K}(t)$ is a solution to the Riccati equation and thus, by Theorem 5.19,

$$\bar{J}(\Upsilon(t_0), t_0, t_1) = \bar{K}(t_0)(\Upsilon(t_0), \Upsilon(t_0)) - \bar{K}(t_1)(\Upsilon(t_1), \Upsilon(t_1)) \le \bar{K}(t_0)(\Upsilon(t_0), \Upsilon(t_0)). \quad ▼$$

Now, by contradiction, $\bar{J}(\Upsilon(t_0), t_0, \infty)$ is proved to be the optimal cost. By assumption $J(\Upsilon(t_0), t_0, \infty) \le \bar{J}(\Upsilon(t_0), t_0, \infty)$ and, if the inequality is strict, then there exist a positive constant C such that

$$0 < C \le \bar{J}(\Upsilon(t_0), t_0, \infty) - J(\Upsilon(t_0), t_0, \infty).$$

To obtain a contradiction, choose a section of $L(\mathcal{A})$ whose cost $J_1(\Upsilon(t_0), t_0, \infty)$ has the property that

$$0 < \frac{C}{2} \le \bar{J}(\Upsilon(t_0), t_0, \infty) - J_1(t_0, \infty, \Upsilon(t_0)). \tag{5.18}$$

For $\epsilon = \frac{C}{4} > 0$ there exists $T > t_0$ such that

$$\begin{aligned} J(\Upsilon(t_0), t_0, t_1) + \frac{C}{4} &= K_{t_1}(t_0)(\Upsilon(t_0), \Upsilon(t_0)) + \frac{C}{4} \\ &> \bar{K}(t_0)(\Upsilon(t_0), \Upsilon(t_0)) \\ &= \bar{J}(\Upsilon(t_0), t_0, t_1) \geq J_1(\Upsilon(t_0), t_0, t_1) + \frac{C}{2}, \quad \forall\, t_1 \geq T, \end{aligned}$$

which contradicts the fact that $J(\Upsilon(t_0), t_0, t_1)$ is the optimal cost. ■

6. Stability and Stabilization

In this section, stability and stabilization by LQR methods of the linearization are addressed to complete the geometric picture of LQR theory. In Section 6.1, the stability definitions for a fixed reference vector field X_{ref} and for linear vector fields over X_{ref} are defined by using a metric on M and a metric on the fibres of TM over image(γ_{ref}), respectively. Such metrics are naturally induced by choosing a Riemannian metric $\mathbb{G}$ on M. Note that, unless the state manifold is compact, these stability definitions depend on the choice of metric. In Section 6.2, Lyapunov's direct method for linear vector fields on tangent bundles is introduced. In Section 6.3, after making geometric sense of the terms "linear state-feedback" and "closed-loop system," it is proved that the linearization of an affine system is uniformly asymptotically stabilized using the linear state-feedback provided by the infinite time LQR problem.

6.1. Stability definitions.

Let X_{ref} be an LIC^∞ reference vector field for a time-dependent affine system $\mathcal{A}$ with a reference trajectory $\gamma_{\mathrm{ref}}\colon I \to M$. Let $\mathbb{G}$ be a Riemannian metric on M. The Riemannian metric induces a metric on M denoted by $d_{\mathbb{G}}\colon M \times M \to \bar{\mathbb{R}}_+$. If $\mathrm{LAC}(x_1, x_2, [0,1])$ denotes the set of absolutely continuous curves γ defined on $[0,1]$ and satisfying $\gamma(0) = x_0$ and $\gamma(1) = x_1$, then

$$d_{\mathbb{G}}(x_0, x_1) = \inf_{\gamma \in \mathrm{LAC}(x_0, x_1, [0,1])} \int_0^1 \sqrt{\mathbb{G}(t)(\gamma'(t), \gamma'(t))}\,\mathrm{d}t.$$

The stability definitions for X_{ref} along γ_{ref} and linear vector fields over X_{ref} are as follows.

6.1 Definition: Let $X_{\mathrm{ref}}\colon I \times M \to TM$ be an LIC^∞ reference vector field with $\sup I = \infty$ and let $\gamma_{\mathrm{ref}}\colon I \to M$ be an integral curve for X_{ref} with the property that $\gamma_{\mathrm{ref}}\colon I \to M$ is defined on all of I. Let $\mathbb{G}$ be a Riemannian metric on M. With respect to γ_{ref}, X_{ref} is

(i) ***locally $d_{\mathbb{G}}$-stable*** if, for each $t_0 \in I$ and for each $\epsilon > 0$, there exists $\delta > 0$ such that

$$d_{\mathbb{G}}(x_0, \gamma_{\mathrm{ref}}(t_0)) < \delta \implies d_{\mathbb{G}}\big(\Phi^{X_{\mathrm{ref}}}_{t_0,t}(x_0), \gamma_{\mathrm{ref}}(t)\big) < \epsilon, \quad t \geq t_0;$$

(ii) ***locally uniformly $d_{\mathbb{G}}$-stable*** if, for each $\epsilon > 0$, there exists $\delta > 0$ such that

$$d_{\mathbb{G}}(x_0, \gamma_{\mathrm{ref}}(t_0)) < \delta \implies d_{\mathbb{G}}\big(\Phi^{X_{\mathrm{ref}}}_{t_0,t}(x_0), \gamma_{\mathrm{ref}}(t)\big) < \epsilon, \quad t_0 \in I,\ t \geq t_0;$$

(iii) ***locally asymptotically $d_{\mathbb{G}}$-stable*** if it is locally $d_{\mathbb{G}}$-stable and if, for each $t_0 \in I$ and $\epsilon > 0$, there exists $\delta, T > 0$ such that

$$d_{\mathbb{G}}(x_0, \gamma_{\mathrm{ref}}(t_0)) < \delta \implies d_{\mathbb{G}}\big(\Phi^{X_{\mathrm{ref}}}_{t_0,t}(x_0), \gamma_{\mathrm{ref}}(t)\big) < \epsilon, \quad t \geq t_0 + T;$$

(iv) ***locally uniformly asymptotically $d_{\mathbb{G}}$-stable*** if it is locally uniformly $d_{\mathbb{G}}$-stable and if, for each $\epsilon > 0$, there exists $\delta, T > 0$ such that

$$d_{\mathbb{G}}(x_0, \gamma_{\mathrm{ref}}(t_0)) < \delta \implies d_{\mathbb{G}}\big(\Phi^{X_{\mathrm{ref}}}_{t_0,t}(x_0), \gamma_{\mathrm{ref}}(t)\big) < \epsilon, \quad t_0 \in I,\ t \geq T + t_0;$$

(v) ***locally exponentially $d_{\mathbb{G}}$-stable*** if, for each $t_0 \in I$, there exists $\delta, c_1, c_2 > 0$ such that

$$d_{\mathbb{G}}(x_0, \gamma_{\mathrm{ref}}(t_0)) < \delta \implies d_{\mathbb{G}}\big(\Phi^{X_{\mathrm{ref}}}_{t_0,t}(x_0), \gamma_{\mathrm{ref}}(t)\big) \leq c_1 \exp(-c_2(t - t_0)), \quad t \geq t_0;$$

(vi) ***locally uniformly exponentially $d_{\mathbb{G}}$-stable*** if there exists $\delta, c_1, c_2 > 0$ such that

$$d_{\mathbb{G}}(x_0, \gamma_{\mathrm{ref}}(t_0)) < \delta \implies d_{\mathbb{G}}\big(\Phi^{X_{\mathrm{ref}}}_{t_0,t}(x_0), \gamma_{\mathrm{ref}}(t)\big) \leq c_1 \exp(-c_2(t - t_0)), \quad t_0 \in I,\ t \geq t_0.$$

•

For the linear stability definitions, we denote by $\|\cdot\|_{\mathbb{G}}$ the induced norm on tangent spaces, i.e.,

$$\|X_x\|_{\mathbb{G}} = \sqrt{\mathbb{G}(X_x, X_x)}$$

for $X_x \in T_xM$. We can then make the following definitions for stability of a linear vector field.

6.2 Definition: Let X_{ref} be an LIC^∞ reference vector field with a reference trajectory $\gamma_{\mathrm{ref}}\colon I \to M$ where $\sup I = \infty$. Let $\mathbb{G}$ be a Riemannian metric on M. The linear vector field $Y\colon I \times TM \to TTM$ over X_{ref} with respect to γ_{ref} is

(i) ***$\mathbb{G}$-stable*** if, for each $t_0 \in I$, there exists $c > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} < c\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}$$

for $t \geq t_0,\ v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$;

(ii) ***uniformly $\mathbb{G}$-stable*** if there exists $c > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} < c\,\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}$$

for $t_0 \in I,\ t \geq t_0,\ v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$;

(iii) ***asymptotically $\mathbb{G}$-stable*** if it is $\mathbb{G}$-stable and if, for each $t_0 \in I$ and for each $\epsilon > 0$, there exists and $T > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} < \epsilon(t_0)\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}$$

for $t \geq t_0 + T,\ v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$;

(iv) ***uniformly asymptotically $\mathbb{G}$-stable*** if it is uniformly $\mathbb{G}$-stable and if, for each $\epsilon > 0$, there exists $T > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} < \epsilon\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}$$

for $t_0 \in I,\ t \geq t_0 + T,\ v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$;

(v) ***exponentially $\mathbb{G}$-stable*** if, for each $t_0 \in I$, there exists $c_1, c_2 > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq c_1\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \exp(-c_2(t - t_0))$$

for $t \geq t_0, v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$;

(vi) ***uniformly exponentially $\mathbb{G}$-stable*** if there exists $c_1, c_2 > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq c_1\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \exp(-c_2(t - t_0))$$

for $t_0 \in I,\ t \geq t_0,\ v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$. •

6.3 Example: To illustrate how stability along a non-compact reference trajectory depends on the choice metric, we take $M = \mathbb{R}_+$ and denote by x the standard coordinate for M and by (x, v) the natural coordinates for TM. We take $X = x\frac{\partial}{\partial x}$ so that

$$X^T = x\frac{\partial}{\partial x} + v\frac{\partial}{\partial v}.$$

We take the reference trajectory $\gamma_{\text{ref}} \colon \mathbb{R} \to M$ given by $\gamma_{\text{ref}}(t) = \mathsf{e}^t$. An integral curve of X^T over this reference trajectory with initial condition $(1, v_0)$ at $t = 0$ has the form

$$\Phi^{X^T}_{0,t}(1, v_0) = (\mathsf{e}^t, v_0\mathsf{e}^t).$$

On M we consider the family of Riemannian metrics

$$\mathbb{G}_\lambda = x^{\lambda-2}\mathrm{d}x \otimes \mathrm{d}x,$$

where $\lambda \in \mathbb{R}$. We have

$$\|\Phi^{X^T}_{0,t}(1, v_0)\|_{\mathbb{G}_\lambda} = v_0\mathsf{e}^{\lambda t/2}.$$

We see, therefore, that X^T is

1. $\mathbb{G}_\lambda$-stable over X with respect to γ_{ref} if $\lambda = 0$,
2. uniformly exponentially $\mathbb{G}_\lambda$-stable if $\lambda < 0$, and
3. not $\mathbb{G}_\lambda$-stable if $\lambda > 0$. •

The above definitions, being for linear vector fields, hold both at a global and local level and are analogous with the standard stability definitions for linear time-varying ordinary differential equations. The following diagram provides the correspondence between each type of stability:

$$\begin{array}{ccccc} UES & \rightleftarrows & UAS & \longrightarrow & US \\ \downarrow & & \downarrow & & \downarrow \\ ES & \longrightarrow & AS & \longrightarrow & S \end{array}$$

The implications given by transitivity of "if-then" are left off for the sake of clarity. The only non-trivial implication is $UAS \implies UES$, and we show this in Proposition 6.4 below. In the statement of the result, let Y be a linear vector field over X_{ref} and let $|||\Phi^Y_{t_0,t}|||_{\mathbb{G}}$ denote the norm of the linear map $\Phi^Y_{t_0,t}|T_{\gamma_{\text{ref}}(t_0)}M \colon T_{\gamma_{\text{ref}}(t_0)}M \to T_{\gamma_{\text{ref}}(t)}M$ induced by the norms on $T_{\gamma_{\text{ref}}(t_0)}M$ and $T_{\gamma_{\text{ref}}(t)}M$.

6.4 Proposition: *The following statements are equivalent:*

(i) Y is uniformly asymptotically $\mathbb{G}$-stable with respect to γ_{ref};

(ii) Y is uniformly exponentially $\mathbb{G}$-stable with respect to γ_{ref};

(iii) there exists constants $c_1, c_2 > 0$ such that $|||\Phi^Y_{t_0,t}|||_{\mathbb{G}} \leq c_1 \exp(-c_2(t - t_0))$.

Proof: (i)⇒(ii) Since the system is uniformly $\mathbb{G}$-stable, there exists $c > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\text{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\text{ref}}(t))} \leq c\|v_{\gamma_{\text{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\text{ref}}(t_0))}$$

for all $t_0 \in I$, $t \geq t_0$, and $v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$. Therefore,

$$|||\Phi^Y_{t_0,t}|||_{\mathbb{G}} = \sup_{\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}=1} \|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq c$$

for all $t_0 \in I$ and $t \geq t_0$. Now choose $T > 0$ such that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq \frac{\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}}{\mathsf{e}}, \quad t \geq t_0 + T,\ v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M,$$

where e denotes the base of the natural logarithm. Define a function $N\colon [t_0,\infty) \to \mathbb{Z}_+$ by asking that $N(t)$ is the smallest integer for which $t \leq t_0 + N(t)T$. Divide the interval $[t_0, (N(t)-1)T]$ into $N(t)-1$ subintervals of length T, and compute

$$\Phi^Y_{t_0,t} = \Phi^Y_{t_0+(N(t)-1)T,t} \circ \Phi^Y_{t_0+(N(t)-2)T,t_0+(N(t)-1)T} \circ \cdots \circ \Phi^Y_{t_0,t_0+T}.$$

Thus,

$$\begin{aligned}|||\Phi^Y_{t_0,t}|||_{\mathbb{G}} &\leq \Big|\Big|\Big|\Phi^Y_{t_0+(N(t)-1)T,t}\Big|\Big|\Big|_{\mathbb{G}} \prod_{j=1}^{N(t)-1} \Big|\Big|\Big|\Phi^Y_{t_0+(j-1)T,t_0+jT}\Big|\Big|\Big|_{\mathbb{G}} \\ &\leq (c\mathsf{e})\mathsf{e}^{-N(t)} \leq (c\mathsf{e})\mathsf{e}^{-(t-t_0)/T} = c_1 \exp(-c_2(t-t_0)),\end{aligned}$$

with $c_1 = c\mathsf{e}$ and $c_2 = \frac{1}{T}$. Therefore,

$$\begin{aligned}\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} &\leq |||\Phi^Y_{t_0,t}|||_{\mathbb{G}} \|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \\ &\leq c_1 \|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \exp(-c_2(t-t_0)),\end{aligned}$$

as desired.

(ii)⇒(iii) Compute

$$|||\Phi^Y_{t_0,t}|||_{\mathbb{G}} = \sup_{\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}=1} \|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} = c_1 \exp(-c_2(t-t_0)).$$

(iii)⇒(i) Let $\epsilon > 0$, take $T = \max\{-\ln(\epsilon/c_1)/c_2, 1/c_2\}$ and $t \geq t_0 + T$, then compute

$$\begin{aligned}\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} &\leq |||\Phi^Y_{t_0,t}|||_{\mathbb{G}} \|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \\ &\leq c_1 \|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \exp(-c_2(t-t_0)) \\ &\leq c_1 \|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \exp(-c_2 T) \leq \epsilon \|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))},\end{aligned}$$

as desired. ■

6.2. Lyapunov theory for linear vector fields on tangent bundles. In this section Lyapunov's direct method for linear vector fields on tangent bundles is introduced. The objective of the direct or "second" method is to infer the stability of a linear vector field Y with respect to a reference trajectory γ_{ref} without explicit knowledge of the flow of the vector field in question. As in the preceding section, a reference vector field $X_{\mathrm{ref}} \in \mathrm{LIC}^\infty(TM)$ with reference trajectory γ_{ref} is fixed. Let $Y \in \mathrm{LIC}^\infty(TTM)$ be a linear vector field over $X_{\mathrm{ref}} \in \mathrm{LIC}^\infty(TM)$ with integral curve Υ. In the classical linear systems theory, a Lyapunov candidate is defined by a quadratic form. In turn its derivative along trajectories of the linear system is also a quadratic form. The geometric definition that is provided here is analogous. A ***Lyapunov candidate*** V for Y is a symmetric $(0,2)$-tensor field along γ_{ref} such that, for each $t \geq t_0$, $V(t)$ is positive-definite. The derivative of V along an integral curve Υ of Y is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}V(t)(\Upsilon(t)),\Upsilon(t)) = \langle \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(V^\flat(t)\Upsilon(t));\Upsilon(t)\rangle + \langle V^\flat(t)\Upsilon(t);\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(\Upsilon(t))\rangle. \quad (6.1)$$

For notational convenience, define a symmetric quadratic form Q_V along γ_{ref} by $-Q_V(t)(\Upsilon(t),\Upsilon(t)) = \frac{\mathrm{d}}{\mathrm{d}t}V(t)(\Upsilon(t)),\Upsilon(t))$. By choosing the natural coordinates for the tangent bundle, a linear vector field Y over X has the form

$$Y(t,x,v) = X^i(t,x)\frac{\partial}{\partial x^i} + Y^i_j(t,x)v^j\frac{\partial}{\partial v^i}.$$

Then (6.1) as a coordinate expression becomes

$$-Q_{V,ij}(t) = \dot{V}_{ij}(t) + V_{mj}(t)Y^m_i(t,x) + V_{im}(t)Y^m_j(t,x),$$

which resembles the classical differential Lyapunov equation. The stability of Y with respect to γ_{ref}, in the sense of Section 6.1, may be characterized in terms of V and Q_V.

6.5 Proposition: *Let $X_{\mathrm{ref}},\gamma_{\mathrm{ref}},Y,\Upsilon,V$ and Q_V be as above. Then a linear vector field $Y : I\times TM \to TTM$ with respect to γ_{ref} is*

(i) ***$\mathbb{G}$-stable*** *if $Q_V(t)(\Upsilon(t),\Upsilon(t)) \geq 0$ and there exist a positive constant α such that*

$$\alpha\|v_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq V(t)(v_{\gamma_{\mathrm{ref}}(t)},v_{\gamma_{\mathrm{ref}}(t)}),$$

for all $v_{\gamma_{\mathrm{ref}}(t)} \in T_{\gamma_{\mathrm{ref}}(t)}M,\ t\geq t_0$;

(ii) ***uniformly $\mathbb{G}$-stable*** *if Y is $\mathbb{G}$-stable and there exists a positive constant β such that*

$$V(t)(v_{\gamma_{\mathrm{ref}}(t)},v_{\gamma_{\mathrm{ref}}(t)}) \leq \beta\|v_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))},$$

for all $v_{\gamma_{\mathrm{ref}}(t)} \in T_{\gamma_{\mathrm{ref}}(t)}M$, $t\geq t_0$;

(iii) ***uniformly asymptotically $\mathbb{G}$-stable*** *if Y is uniformly $\mathbb{G}$-stable and there exists a positive constant η such that*

$$Q_V(t)(\Upsilon(t),\Upsilon(t)) > \eta\|\Upsilon(t)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))}, \qquad \forall\ t\geq t_0.$$

Proof: (i) Fix $t_0 \in I$ and let $v_{\gamma_{\mathrm{ref}}(t_0)} \in T_{\gamma_{\mathrm{ref}}(t_0)}M$. Let Υ be an integral curve of Y such that $\Upsilon(t_0) = v_{\gamma_{\mathrm{ref}}(t_0)}$. Since V is a Lyapunov candidate,

$$\begin{aligned}\int_{t_0}^t \langle \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(V^\flat(\sigma)\Upsilon(\sigma));\Upsilon(\sigma)\rangle + \langle V^\flat(\sigma)\Upsilon(\sigma);\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(\Upsilon(\sigma))\rangle\ \mathrm{d}\sigma \\ = -\int_{t_0}^t Q_V(\sigma)(\Upsilon(\sigma),\Upsilon(\sigma))\ \mathrm{d}\sigma \\ = V(t)(\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)}),\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})) - V(t_0)(v_{\gamma_{\mathrm{ref}}(t_0)},v_{\gamma_{\mathrm{ref}}(t_0)}).\end{aligned}$$

Given the hypothesis,

$$Q_V(t)(\Upsilon(t),\Upsilon(t)) \geq 0 \text{ and } \alpha\|v_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq V(t)(v_{\gamma_{\mathrm{ref}}(t)}, v_{\gamma_{\mathrm{ref}}(t)})$$

for all $v_{\gamma_{\mathrm{ref}}(t)} \in T_{\gamma_{\mathrm{ref}}(t)}M,\ t \geq t_0$. Thus, the following inequality holds:

$$\alpha\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq V(t)(\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)}), \Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})) \leq V(t_0)(v_{\gamma_{\mathrm{ref}}(t_0)}, v_{\gamma_{\mathrm{ref}}(t_0)}).$$

For fixed t_0, $V(t_0)(v_{\gamma_{\mathrm{ref}}(t_0)}, v_{\gamma_{\mathrm{ref}}(t_0)})$ is a positive-definite symmetric bilinear form on the vector space $T_{\gamma_{\mathrm{ref}}(t_0)}M$ which implies that there exists $\kappa_{t_0} > 0$ such that

$$V(t_0)(v_{\gamma_{\mathrm{ref}}(t_0)}, v_{\gamma_{\mathrm{ref}}(t_0)}) \leq \kappa_{t_0}\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}.$$

The result follows:

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq \sqrt{\frac{\kappa_{t_0}}{\alpha}}\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}.$$

(ii) Using similar arguments as in part (i), for $t \geq t_0$ and the added hypothesis yield

$$\alpha\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq V(t_0)(v_{\gamma_{\mathrm{ref}}(t_0)}, v_{\gamma_{\mathrm{ref}}(t_0)}) \leq \beta\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))},$$

and result follows:

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq \sqrt{\frac{\beta}{\alpha}}\|v_{\gamma_{\mathrm{ref}}(t_0)}\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}.$$

(iii) For $\epsilon > 0$ it is shown that for $T = \frac{\beta}{\eta\epsilon}$ we have

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} < \epsilon\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))},$$

for all $t \geq t_0 + T$. Let $\epsilon > 0$, $T = \frac{\beta}{\eta\epsilon}$, and suppose that

$$\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \geq \epsilon\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))},$$

for $t \geq t_0 + T$. Using the arguments, from the proof of (i),

$$\begin{aligned}
0 < \alpha\ &\|\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)})\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq V(t)(\Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)}), \Phi^Y_{t_0,t}(v_{\gamma_{\mathrm{ref}}(t_0)}))\\
&= V(t_0)(v_{\gamma_{\mathrm{ref}}(t_0)}, v_{\gamma_{\mathrm{ref}}(t_0)}) - \int_{t_0}^{t} Q_V(\sigma)(\Phi^Y_{t_0,\sigma}(v_{\gamma_{\mathrm{ref}}(t_0)}), \Phi^Y_{t_0,\sigma}(v_{\gamma_{\mathrm{ref}}(t_0)}))\,\mathrm{d}\sigma\\
&< \beta\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} - \int_{t_0}^{t}\eta\|\Phi^Y_{t_0,\sigma}(v_{\gamma_{\mathrm{ref}}(t_0)})\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(\sigma))}\,\mathrm{d}\sigma\\
&\leq \beta\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} - \int_{t_0}^{t_0+T}\eta\|\Phi^Y_{t_0,\sigma}(v_{\gamma_{\mathrm{ref}}(t_0)})\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(\sigma))}\,\mathrm{d}\sigma\\
&\leq \beta\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} - \int_{t_0}^{t_0+T}\eta\epsilon\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}\,\mathrm{d}\sigma\\
&\leq\ (\beta - \eta\epsilon T)\,\|v_{\gamma_{\mathrm{ref}}(t_0)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}\\
&= 0,
\end{aligned}$$

giving a contradiction. ■

6.3. Stabilization of the linearization. Let X_{ref} be a reference vector field for the affine system $\mathcal{A}$ with $\gamma_{\mathrm{ref}}\colon I \to M$ a reference trajectory. To avoid complications, suppose that $L(\mathcal{A})$ has constant rank and so is a vector bundle over M. Let $\mathrm{L}(TM; L(\mathcal{A}))$ be the set of vector bundle mappings from TM to $L(\mathcal{A})$ over id_M. A ***linear state feedback*** along γ_{ref} is then a section F of the bundle consisting of the fibers of $\mathrm{L}(TM; L(\mathcal{A}))$ over γ_{ref}. Thus F assigns to each point $t \in I$ a linear map $F(t)\colon T_{\gamma_{\mathrm{ref}}(t)}M \to L(\mathcal{A})_{(t,\gamma_{\mathrm{ref}}(t))}$. For such a linear state feedback, the ***closed-loop system*** is then the linear LIC$^\infty$ vector field on TM defined by $X^F_{\mathrm{ref}}(t, v_x) = X^T_{\mathrm{ref}}(t, v_x) + \mathrm{vlft}(F(t)(v_x))$. Note that the integral curves of X^F_{ref} with initial conditions projecting to γ_{ref} will project to γ_{ref}. Therefore, given a linear state feedback F, the stability of the linear vector field X^F_{ref} relative to $\mathbb{G}$ is as defined in Section 6.1.

To obtain uniform asymptotic stability, various uniformity bounds on the problem data are required and, as in Kalman [1960], the class of linear systems is restricted to those that are uniformly controllable.

6.6 Definition: Let $\mathcal{A}$ be a time-dependent affine system on M with X_{ref} a reference vector field and $\gamma_{\mathrm{ref}}\colon I \to M$ a reference trajectory. Let $\mathbb{G}$ be a Riemannian metric on M.

(i) The linearisation $\mathcal{A}^T_{\mathrm{ref}}$ along γ_{ref} is ***$\mathbb{G}$-uniformly controllable*** if there exist strictly increasing functions $\alpha, \beta\colon \mathbb{R}_+ \to \mathbb{R}_+$ with $\alpha(0) = 1$, $\beta(0) = 1$, such that

(a)

$$\|\Phi^{X^T_{\mathrm{ref}}}_{\tau,t}(v_\tau)\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq \alpha(|t-\tau|)\|v_\tau\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(\tau))},$$
$$\|\Phi^{X^{T*}_{\mathrm{ref}}}_{\tau,t}(\eta_\tau)\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq \beta(|t-\tau|)\|\eta_\tau\|_{\mathbb{G}(\gamma_{\mathrm{ref}}(\tau))},$$

for all $t, \tau \in I$, $v_\tau \in T_{\gamma_{\mathrm{ref}}(\tau)}M$, and $\eta_\tau \in T^*_{\gamma_{\mathrm{ref}}(\tau)}M$, and

(b) there exists a constant σ such that, for each $t \in I$,

$$0 < \alpha_3(\sigma)\|\eta_t\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} < W(t, t+\sigma)(\eta_t, \eta_t) < \alpha_4(\sigma)\|\eta_t\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))},$$

where $\eta_t \in T^*_{\gamma_{\mathrm{ref}}(t)}M$ and $W(t,\tau)$ is defined in (5.12).

(ii) A Riemannian metric $\mathbb{G}$ is ***X_{ref}-compatible*** if the linearisation $\mathcal{A}^T_{\mathrm{ref}}$ along γ_{ref} is $\mathbb{G}$-uniformly controllable. •

Consider the following optimal closed-loop system $X^{\bar{K}}_{\mathrm{ref}}\colon I \times TM \to TTM$ defined by

$$X^{\bar{K}}_{\mathrm{ref}}(t, v_x) = X^T_{\mathrm{ref}}(t, v_x) - \mathrm{vlft}((\iota(t)R^\sharp(t)\iota^*(t)\bar{K}^\flat(t))(v_x)),$$

where $\bar{K}^\flat$ is defined as in Proposition 5.23. Is $X^{\bar{K}}_{\mathrm{ref}}$ with respect to γ_{ref} uniformly asymptotically $\mathbb{G}$-stable? In the classical approach, Kalman [1960] proves this using the optimal cost associate with the infinite time LQR problem as a Lyapunov function. The approach here is analogous.

The cost data Q and R in the statement of Theorem 6.7 are defined as in Section 5.1. That is, let Q be an LI section of $\Sigma_2(TM|_{\mathrm{image}(\gamma_{\mathrm{ref}})})$ with the property that $Q(t)$ is positive-semidefinite for each $t \in I$. Also, let R be an LI section of $\Sigma_2(L(\mathcal{A})|_{\mathrm{image}(\gamma_{\mathrm{ref}})})$ with the property that $R(t)$ is positive-definite for each $t \in I$.

6.7 Theorem: (LQR stability) *Let X_{ref} be a reference vector field for an affine system $\mathcal{A}$ along with an LAC reference trajectory γ_{ref}. Let Q and R be as above with the property that there exists constants $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$ such that*

$$\alpha_1 \|v_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq Q(t)(v_{\gamma_{\mathrm{ref}}(t)}, v_{\gamma_{\mathrm{ref}}(t)}) \leq \alpha_2 \|v_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))},$$
$$\beta_1 \|\iota(t) w_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq R(t)(w_{\gamma_{\mathrm{ref}}(t)}, w_{\gamma_{\mathrm{ref}}(t)}) \leq \beta_2 \|\iota(t) w_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))},$$

where $v_{\gamma_{\mathrm{ref}}(t)} \in T_{\gamma_{\mathrm{ref}}(t)} M$, $w_{\gamma_{\mathrm{ref}}(t)} \in L(\mathcal{A})_{(t, \gamma_{\mathrm{ref}}(t))}$. If the Riemannian metric $\mathbb{G}$ is X_{ref}-compatible, then the closed-loop system

$$\mathcal{A}^{\bar{K}}_{\mathrm{ref}}(t, v_x) = \{X^T_{\mathrm{ref}}(t, v_x) - \mathrm{vlft}((\iota(t) R^\sharp(t) \iota^*(t) \bar{K}^\flat(t))(v_x)) \mid (t, v_x) \in \mathbb{R} \times TM\}$$

is uniformly asymptotically $\mathbb{G}$-stable.

Proof: It suffices to show the existence of a Lyapunov candidate for the linear vector field $X^{\bar{K}}_{\mathrm{ref}} \colon I \times TM \to TTM$ over X_{ref} defined by

$$X^{\bar{K}}_{\mathrm{ref}}(t, v_x) = X^T_{\mathrm{ref}}(t, v_x) - \mathrm{vlft}((\iota(t) R^\sharp(t) \iota^*(t) \bar{K}^\flat(t))(v_x)).$$

Recall that a Lyapunov candidate for $X^{\bar{K}}_{\mathrm{ref}}$ is a symmetric $(0,2)$-tensor field along γ_{ref} that is positive-definite for each $t \geq t_0$. Thus the optimal feedback $\bar{K}$ is a Lyapunov candidate for $X^{\bar{K}}_{\mathrm{ref}}$. Thus, the vector field $X^{\bar{K}}_{\mathrm{ref}}$ with integral curve Υ is uniformly asymptotically $\mathbb{G}$-stable if

1. there exists positive constants c_1, c_2, and c such that

$$c_1 \|v_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \leq \bar{K}(v_{\gamma_{\mathrm{ref}}(t)}, v_{\gamma_{\mathrm{ref}}(t)}) \leq c_2 \|v_{\gamma_{\mathrm{ref}}(t)}\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))}$$

   and

2. $\langle \mathscr{L}^{X_{\mathrm{ref}}, \gamma_{\mathrm{ref}}}(\bar{K}^\flat(t)\Upsilon(t)); \Upsilon(t)\rangle + \langle \bar{K}^\flat(t)\Upsilon(t; \mathscr{L}^{X_{\mathrm{ref}}, \gamma_{\mathrm{ref}}}(\Upsilon(t))\rangle < -c\|\Upsilon(t)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))}$.

We thus verify these two facts.

(1) The first step is to obtain an upper bound on $\bar{K}$. Using Lemma 5.22 with $t_2 = t_0 + \sigma$, where σ is the constant prescribed by $\mathbb{G}$-uniform controllability, and using the added hypotheses on the cost data gives

$$\bar{K}(t_0)(\Upsilon_*(t_0), \Upsilon_*(t_0)) \leq \frac{1}{2} \int_{t_0}^{t_2} \alpha_2 \|\Upsilon_1(\tau)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} + \beta_2 \|\iota(\tau) X_1(\tau)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t))} \, \mathrm{d}\tau, \tag{6.2}$$

where the trajectory is given by

$$\Upsilon_1(\tau) = \Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}(\Upsilon(t_0) - W^\flat(t_0, \tau) W^\sharp(t_0, t_2)\Upsilon(t_0)).$$

From Lemma 5.22 the corresponding section of $L(\mathcal{A})$ is

$$X_1(\tau) = -R(\tau)^\sharp \iota^*(\tau) \Phi^{X^{T^*}}_{t_0,\tau} \eta,$$

where $\eta \in T^*_{\gamma_{\mathrm{ref}}(t_0)}M$ satisfies $W^\flat(t_0,t_2)\eta = \Upsilon(t_0)$. First consider an upper bound on the trajectory given the following computation:

$$\begin{aligned}
\|\Upsilon_1(\tau)\|^2 &= \|\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}(\Upsilon(t_0) - W^\flat(t_0,\tau)W^\sharp(t_0,t_2)\Upsilon(t_0))\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(\tau))} \\
&\leq \|\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}\|^2_{\mathbb{G}_{\gamma_{\mathrm{ref}}},\mathrm{op}}\|(W^\flat(t_0,t_2) - W^\flat(t_0,\tau))W^\sharp(t_0,t_2)\|^2_{\mathbb{G}_{\gamma_{\mathrm{ref}}},\mathrm{op}}\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \\
&\leq \|\Phi^{X^T_{\mathrm{ref}}}_{t_0,\tau}\|^2_{\mathbb{G}_{\gamma_{\mathrm{ref}}},\mathrm{op}}\|W^\sharp(t_0,t_2)\|^2_{\mathbb{G}_{\gamma_{\mathrm{ref}}},\mathrm{op}}\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \\
&\leq c_3(|\tau - t_0|)\alpha_3(\sigma)\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))},
\end{aligned}$$

where

(a) line 3 follows from the observations that the operator $(W^\flat(t_0,t_2) - W^\flat(t_0,\tau))$, is non-negative, zero at $\tau = t_2$, takes the form $W^\flat(t_0,t_2)$ at $\tau = t_0$, and has a derivative with respect to τ that is negative for all $\tau \in [t_0,t_2]$;

(b) line 4 follows from $\mathbb{G}$-uniform controllability since $t_2 = t_0 + \sigma$.

Similar calculations provide an upper bound on $X_1(\tau)$. Consider the following computation:

$$\begin{aligned}
\|\iota(\tau)X_1(\tau)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(\tau))} &= \|-\iota(\tau)R(\tau)^\sharp\iota^*(\tau)\Phi^{X^{T^*}}_{t_0,\tau}\eta\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(\tau))} \\
&\leq c_4(|t_0 - \tau|)\|\frac{\mathrm{d}}{\mathrm{d}\tau}W(t_0,\tau)\eta\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \\
&\leq c_4(|t_0 - \tau|)\frac{\mathrm{d}}{\mathrm{d}\tau}\alpha_4(|\tau - t_0|)\|\eta\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))},
\end{aligned}$$

where

(a) line 2 uses Lemma 5.21 and the $\mathbb{G}$-uniform controllability hypothesis;

(b) line 3 uses the derivative of the $\mathbb{G}$-uniform controllability condition.

The above computations, together with (6.2), imply that

$$\begin{aligned}
\bar{K}(t_0)(\Upsilon_*(t_0),\Upsilon_*(t_0)) &\leq \frac{1}{2}\int_{t_0}^{t_2}\alpha_2 c_3(|\tau - t_0|)\alpha_3(\sigma)\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \\
&\quad + \beta_2\alpha_3(\sigma)c_4(|t_0 - \tau|)\|\frac{\mathrm{d}}{\mathrm{d}\tau}W(t_0,\tau)\|^2_{\mathbb{G}_{\gamma_{\mathrm{ref}}},\mathrm{op}}\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}\,\mathrm{d}\tau \\
&= \frac{1}{2}\int_0^{t_2 - t_0}\alpha_2\alpha_3 c_3(s)\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))} \\
&\quad + \beta_2\alpha_3 c_4(s)\frac{\mathrm{d}}{\mathrm{d}s}\alpha_4(s)\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}\,\mathrm{d}s \\
&\leq c_2(\sigma)\|\Upsilon(t_0)\|^2_{\mathbb{G}(\gamma_{\mathrm{ref}}(t_0))}.
\end{aligned}$$

The lower bound is constructed using the same procedure as above since the inverse of $\bar{K}(t)$ exists for each t and satisfies the Riccati equation,

$$\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}P(t) - P(t)Q(t)P(t) + \iota(t)R^\sharp(t)\iota^*(t) = 0,$$

of the “dual system.”

(2) The derivative of $\bar{K}$ along a trajectory of the closed looped system,

$$\begin{aligned}\frac{\mathrm{d}\bar{K}}{\mathrm{d}t}(\Upsilon(t),\Upsilon(t)) &= \langle \mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(\bar{K}^\flat(t)\Upsilon(t));\Upsilon(t)\rangle + \langle \bar{K}^\flat(t)\Upsilon(t);\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}(\Upsilon(t))\rangle \\ &= -\langle Q^\flat(t)\Upsilon(t);\Upsilon(t)\rangle - \langle \bar{K}^\flat(t)\Upsilon(t);\iota(t)R^\sharp(t)\iota^*(t)\bar{K}^\flat(t)\Upsilon(t)\rangle,\end{aligned}$$

is negative as required. ■

7. Discussion and future work

In this paper a geometric understanding of Jacobian linearization along a non-trivial reference trajectory and LQR theory is given for affine systems on a differentiable manifold M. The basis of this geometric formulation involved embedding the reference trajectory γ_{ref} for an affine system $\mathcal{A}$ on M as an integral curve of a LIC$^\infty$ reference vector field X_{ref} on M. Given X_{ref}, differentiation along the reference trajectory is defined by the Lie derivative operator, $\mathscr{L}^{X_{\mathrm{ref}},\gamma_{\mathrm{ref}}}$. This differential operator, together with the tangent, cotangent, and vertical lifts, were used to clarify the geometric structure corresponding to what is commonly done in the standard approach as outlined in Section 1.1. The geometry of Jacobian linearization as presented in this paper provides a framework in which to approach other standard concepts in control theory associated with linearization and optimal control.

7.1. Open questions about stability and stabilization.

Using the definitions in Section 6.1, consider the stability X^T_{ref} along γ_{ref} relative to the metric $d_{\mathbb{G}}$ and consider the stability of the linearisation X^T_{ref} relative to $\mathbb{G}$.

7.1 Question: Does uniform asymptotic stability of the linearisation imply uniform exponential stability of X_{ref}? •

In the standard case, this is of course well known and follows from Lyapunov's second method. The required Lyapunov function to show uniform exponential stability, in fact, is the same function used to guarantee the uniform asymptotic stability of the linearization [Vidyasagar 1993]. For a time-varying linear system that is uniformly asymptoticly stable, such a Lyapunov function is defined by

$$V(t) = \int_t^\infty \Phi^T(\sigma,t)M(\sigma)\Phi(\sigma,t)\,\mathrm{d}\sigma, \tag{7.1}$$

where $M\colon I \to L(\mathbb{R}^n;\mathbb{R}^n)$ is bounded and symmetric for each $t \in I$ [Brockett 1970]. To answer Question 7.1 in a geometric context requires first making sense of (7.1). Then, the next hurdle is to "transfer" a Lyapunov function for the linear vector field on TM to Lyapunov function for X_{ref} on M.

If X_{ref} is not stable with respect to some choice of metric, then perhaps it is possible to stabilize it under feedback. In Section 6.3 the linearization $\mathcal{A}^T_{\mathrm{ref}}$ of an affine system $\mathcal{A}$ along a non-trivial reference trajectory γ_{ref} was uniformly asymptotically stabilized using a linear state-feedback obtained from solving an infinite time LQR problem. In the setting of Section 1.1, a stabilizing linear state-feedback would be then implemented to locally uniformly exponentially stabilize the non-linear system along the reference trajectory. This is easily done in the standard setup since the state space is naturally identified with each tangent space. From a geometric point of view this raises the following question.

7.2 Question: For a linear state-feedback $F\colon I \to \mathrm{L}(TM; L(\mathcal{A}))$ which uniformly asymptotically stabilizes the linearization $\mathcal{A}^T_{\mathrm{ref}}$, how can F be implemented with the affine system $\mathcal{A}$? And once a method of implementation is understood, is it ensured that the affine system $\mathcal{A}$ is locally uniformly exponentially stabilized along the reference trajectory? •

The feedback implementation problem amounts to interpreting geometrically the process of choosing coordinates on a neighbourhood of the reference trajectory. In the classical case, this is carried out by Vidyasagar [1993]. The geometrization of this classical approach is a subject of future work.

References

Abraham, R., Marsden, J. E., and Ratiu, T. S. [1988] *Manifolds, Tensor Analysis, and Applications*, number 75 in Applied Mathematical Sciences, Springer-Verlag: New York/-Heidelberg/Berlin, ISBN: 978-0-387-96790-5.

Aliprantis, C. D. and Border, K. C. [2006] *Infinite-Dimensional Analysis, A Hitchhiker's Guide*, 3rd edition, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-32696-0.

Athans, M. and Falb, P. L. [1966] *Optimal Control, An Introduction to the Theory and its Applications*, McGraw-Hill: New York, NY, Reprint: [Athans and Falb 2006].

— [2006] *Optimal Control, An Introduction to the Theory and its Applications*, Dover Publications, Inc.: New York, NY, ISBN: 978-0-486-45328-6, Original: [Athans and Falb 1966].

Bianchini, R. M. and Stefani, G. [1993] *Controllability along a trajectory: A variational approach*, SIAM Journal on Control and Optimization, **31**(4), pages 900–927, ISSN: 0363-0129, DOI: 10.1137/0331039.

Brockett, R. W. [1970] *Finite Dimensional Linear Systems*, John Wiley and Sons: NewYork, NY, ISBN: 978-0-471-10585-5.

Ikeda, M., Maeda, H., and Kodama, S. [1972] *Stabilization of linear systems*, Journal of the Society of Industrial and Applied Mathematics, Series A Control, **10**(4), pages 716–729, DOI: 10.1137/0310052.

Kalman, R. E. [1960] *Contributions to the theory of optimal control*, Boletín de la Sociedad Matemática Mexicana. Segunda Serie, **5**, pages 102–119, URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.26.4070 (visited on 07/10/2014).

Lee, E. B. and Markus, L. [1967] *Foundations of Optimal Control Theory*, John Wiley and Sons: NewYork, NY, Reprint: [Lee and Markus 1986].

— [1986] *Foundations of Optimal Control Theory*, Robert E. Krieger Publishing Company: Huntington/New York, ISBN: 978-0-89874-807-9, Original: [Lee and Markus 1967].

Pontryagin, L. S., Boltyanskiĭ, V. G., Gamkrelidze, R. V., and Mishchenko, E. F. [1961] *Matematicheskaya teoriya optimal′ nykh protsessov*, Gosudarstvennoe izdatelstvo fizikomatematicheskoi literatury: Moscow, Translation: [Pontryagin, Boltyanskiĭ, Gamkrelidze, and Mishchenko 1986].

— [1986] *The Mathematical Theory of Optimal Processes*, translated by K. N. Trirogoff, Classics of Soviet Mathematics, Gordon & Breach Science Publishers: New York, NY, ISBN: 978-2-88124-077-5, Original: [Pontryagin, Boltyanskiĭ, Gamkrelidze, and Mishchenko 1961].

Sontag, E. D. [1998] *Mathematical Control Theory, Deterministic Finite Dimensional Systems*, 2nd edition, number 6 in Texts in Applied Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-98489-6.

Sussmann, H. J. [1998] *An introduction to the coordinate-free maximum principle*, in *Geometry of Feedback and Optimal Control*, edited by B. Jakubczyk and W. Respondek, Monographs and Textbooks in Pure and Applied Mathematics, pages 463–557, Dekker Marcel Dekker: New York, NY, ISBN: 978-0-8247-9068-4.

Tyner, D. R. [2007] "Geometric Jacobian Linearisation", Doctoral thesis, Kingston, ON, Canada: Queen's University, Kingston.

Vidyasagar, M. [1993] *Nonlinear Systems Analysis*, Prentice-Hall: Englewood Cliffs, NJ, ISBN: 0-13-623463-1, Reprint: [Vidyasagar 2002].

— [2002] *Nonlinear Systems Analysis*, 2nd edition, number 42 in Classics in Applied Mathematics, Society for Industrial and Applied Mathematics: Philadelphia, PA, ISBN: 978-0-89871-526-2, Original: [Vidyasagar 1993].

Yano, K. and Ishihara, S. [1973] *Tangent and Cotangent Bundles*, number 16 in Pure and Applied Mathematics, Dekker Marcel Dekker: New York, NY, ISBN: 978-0-8247-6071-7.