

Second order conditions for optimality and local controllability of discrete-time systems

M. Barbero-Liñán *

B. Jakubczyk †

Abstract

We study local controllability and optimal control problems for invertible discrete-time systems. We present second order necessary and sufficient conditions for optimality and for local controllability. The conditions are stated in geometric terms, using vector fields naturally associated to the system. The Hessian of the optimal problem is computed in terms of Lie brackets of vector fields of the system.

1 Introduction

Nonlinear discrete-time control systems

$$\Sigma : \quad x(t) = f(x(t-1), u(t))$$

are much less understood, comparing to continuous-time systems. This is due to the fact that algebraic and geometric tools available in the continuous-time case are not present or, at least, have not been used much for their analysis. Here we have in mind vector fields, Lie bracket and Lie algebraic techniques which are very helpful in the theory of continuous-time systems.

There is a class of discrete-time systems where such tools are available, however. These are *invertible systems*, where the state equations are solvable backwards in time, that is $x(t-1)$ is uniquely defined for given $x(t)$ and $u(t)$. Families of vector fields can be assigned to such systems

*Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain, and Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM). E-mail: mbarbero@icmat.es.

†Institute of Mathematics, Polish Academy of Sciences, Śniadeckich 8, 00-956 Warsaw, Poland. E-mail: b.jakubczyk@impan.pl.

Work partially supported by Ministry of Research and Higher Education (Poland), grant N201 607540; by the MICINN (Spain) MTM2010-21186-C02-02; 2009SGR1338 from the Catalan government; the European project IRSES-project GeoMech-246981 and the ICMAT Severo Ochoa project SEV-2011-0087.

Keywords: Nonlinear systems, discrete time, local controllability, optimal control, second order conditions, vector fields. **Mathematics Subject Classification 2000:** 49K99, 93B05, 93B29.

Date: January 25, 2013

(see [9, 10]) and can be used for analyzing their general controllability properties [4, 10]. Invertible systems appear e.g. in modeling of continuous-time systems, as in the case of sampling, and in numerical schemes approximating continuous-time systems. Codistributions can be also assigned to such systems and have been useful for characterizing their observability [3]. Recently, differential geometric tools have been used to analyze the accessibility of implicit discrete-time systems [13].

For continuous-time optimality of a trajectory is closely related to its local controllability property. Namely, for most optimal control problems, optimal trajectories lie on the boundary of the reachable set. On the other hand, it is a hard problem to find criteria in terms of the system vector fields which characterize local controllability along a trajectory (i.e., when the trajectory lies in the interior of the reachable set). Deep and far reaching results in this direction were obtained in the last forty years (see e.g. [5, 7, 11, 15]) but a complete characterization seems beyond the reach [1, 14]. General second order conditions for optimality in continuous-time can be found in [2].

In this paper we analyze invertible discrete-time systems using the formalism of vector fields from [10]. We present (Section 3) second order sufficient and necessary conditions for local controllability in terms of those vector fields and their Lie brackets. In Section 5 we present basic lemmata which are then used for proving the sufficiency result in Section 3. Second order optimality conditions are presented and proved in Sections 6 and 7, using the same lemmata. The proof of one version of these results is based on a geometric lemma in [2] on local openness of a nonlinear map at a singular point. We include an illustrating example in Section 4.

Our infinitesimal analysis of local controllability, with the use of Lie bracket, can be considered as a starting point to identifying higher order sufficient or necessary conditions for local controllability of discrete-time systems (analogous to conditions in the continuous-time case), in particular to identifying so called “bad brackets” (see [5, 7, 11, 12, 15] for the case of continuous-time).

2 Preliminaries

Let M and U be two sets called *state space* and *control space*, respectively. A map $f: M \times U \rightarrow M$ defines a nonlinear discrete-time control system with the dynamics

$$\Sigma: \quad x(t) = f(x(t-1), u(t)),$$

where x and u are called state and control, respectively, and $t \in \mathbb{Z}$.

We will assume that:

- (A1) M is an open subset of \mathbb{R}^n or a smooth differentiable manifold of dimension n ;
- (A2) U is a subset of \mathbb{R}^m , with nonempty interior, and the closure of the interior of U contains U ;
- (A3) the map $f: M \times U \rightarrow M$ is of class C^2 ;
- (A4) the system is *invertible*, which means that the maps $f_u: M \rightarrow M$, $u \in U$, are diffeomorphisms onto open images.

Condition (A3) means that f has a C^2 extension to $M \times \tilde{U}$, where $\tilde{U} \subset \mathbb{R}^m$ is an open superset of U . Above, and in the sequel, we denote

$$f_u(x) = f(x, u).$$

The map $f_u : M \rightarrow M$ defines the one-step transition defined by control u . The invertibility property (A4) means that $f_u(M) \subset M$ is open and each $f_u : M \rightarrow f_u(M)$ is a C^2 diffeomorphism.

Assumption (A4) is needed for associating natural vector fields to system Σ . Sometimes we will assume that the system is *strongly invertible*, which means that the maps $f_u : M \rightarrow M$, $u \in U$, are diffeomorphisms onto M .

It will be convenient to use the notation $t = i$ and write the system equations in the form

$$\Sigma : \quad x_i = f_{u_i}(x_{i-1}), \quad \text{where } x_i = x(i), \quad u_i = u(i).$$

Given an initial state x_0 and a control sequence u_1, \dots, u_N , the trajectory of Σ is defined by the sequence of states x_1, \dots, x_N , where x_i is given by the composition of maps $f_{u_i} \circ \dots \circ f_{u_2} \circ f_{u_1}$ applied to x_0 . We will usually omit the composition sign and write $x_i = f_{u_i} \dots f_{u_1}(x_0)$. The set of points reachable from x_0 in N forward steps is denoted by

$$\mathcal{R}^+(x_0, N) = \{x \in M \mid \exists (u_1, \dots, u_N) \in U^N \text{ such that } x = f_{u_N} \dots f_{u_1}(x_0)\}.$$

Given a subset $\bar{U} \subset U^N$ of control sequences (u_1, \dots, u_N) , we define the corresponding reachable set

$$\mathcal{R}^+(x_0, \bar{U}) = \{x \in M \mid \exists (u_1, \dots, u_N) \in \bar{U} \text{ such that } x = f_{u_N} \dots f_{u_1}(x_0)\}.$$

With the purpose to discuss local controllability of discrete-time systems we will provide an infinitesimal description of local deformations of admissible trajectories of Σ . We will use vector fields introduced in [10] (see also [6, 8] for earlier work), where they were needed for characterizing accessibility and controllability.

Assume first that the control is scalar, that it is $U \subset \mathbb{R}$. Then the following vector fields depending on u can be associated to the invertible system Σ (cf. [10]),

$$X_u^+(x) = \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} f_u^{-1} \circ f_{u+\epsilon}(x), \quad Y_u^+(x) = \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} f_{u+\epsilon}^{-1} \circ f_u(x), \quad (1)$$

$$X_u^-(x) = \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} f_u \circ f_{u+\epsilon}^{-1}(x), \quad Y_u^-(x) = \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} f_{u+\epsilon} \circ f_u^{-1}(x). \quad (2)$$

The vector fields in (1) can be used for the infinitesimal analysis of the variations of forward trajectories of Σ , whereas the vector fields in (2) play an analogous role for backward trajectories. Note that X_u^+ and Y_u^+ are well defined for all $x \in M$, if Σ is invertible (for small ϵ the point $f_{u+\epsilon}(x)$ is in the domain of f_u^{-1} , by the implicit function theorem), while the same holds for X_u^- and Y_u^- only when Σ is strongly invertible. Due to assumptions (A2) and (A3), these vector fields are well defined for any $u \in U$ and are of class C^1 .

We will mainly use the u -depending vector fields X_u^+ , however similar results can be obtained to describe local controllability and optimality by means of the vector fields in (2) when reachability is defined by backward trajectories. The vector fields X_u^+ can be alternatively defined as

$$X_u^+(x) = (df_u(x))^{-1} \frac{\partial}{\partial u} f_u(x).$$

In the case of multidimensional control these u -depending vector fields depend, additionally, on the index r of the component of $u = (u^1, \dots, u^m) \in U$,

$$X_{u,r}^+(x) = (df_u(x))^{-1} \frac{\partial}{\partial u^r} f_u(x) = \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} f_u^{-1} \circ f_{u+\epsilon e_r}(x),$$

where e_r is the r -th versor in \mathbb{R}^m .

Given a vector field Y and a control $u \in U$, we define another vector field using the diffeomorphism f_u on M ,

$$(\text{Ad}_u Y)(x) = (df_u(x))^{-1} Y(f_u(x)),$$

where $df_u(x)$ is the differential of f_u evaluated at the point x . (The above definition of Ad , used throughout the paper, is convenient for analyzing forward trajectories. Note that it does not match the usual notation of Lie group theory adopted to left actions.) More generally, denoting $f_{u_k \dots u_1}(x) = f_{u_k} \dots f_{u_1}(x)$ we define the following vector fields

$$(\text{Ad}_{u_k \dots u_1} Y)(x) = (\text{Ad}_{u_1} \dots \text{Ad}_{u_k} Y)(x) = (df_{u_k \dots u_1}(x))^{-1} Y(f_{u_k \dots u_1}(x)).$$

Note that, in the case of scalar control,

$$(\text{Ad}_{u_k \dots u_1} X_{u_0}^+)(x) = \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} f_{u_k \dots u_1}^{-1} f_{u_0}^{-1} f_{u_0+\epsilon} f_{u_k \dots u_1}(x).$$

For multidimensional control

$$(\text{Ad}_{u_k \dots u_1} X_{u_0,r}^+)(x) = \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} f_{u_k \dots u_1}^{-1} f_{u_0}^{-1} f_{u_0+\epsilon e_r} f_{u_k \dots u_1}(x),$$

where e_r is the r -th versor in \mathbb{R}^m .

Proposition 1. [10, Proposition 3.2] *For scalar control the following equalities hold for each $u \in U$:*

- (1) $X_u^+ = -Y_u^+, \quad X_u^- = -Y_u^-,$
- (2) $X_u^+ = -\text{Ad}_u X_u^-, \quad Y_u^+ = -\text{Ad}_u Y_u^-.$

For multidimensional control analogous equalities hold for the vector fields $X_{u,r}^+, Y_{u,r}^+, X_{u,r}^-$ and $Y_{u,r}^-$.

For so defined vector fields we can compute their Lie bracket $[\cdot, \cdot]$ in the usual way. We also denote the Lie bracket of Y and Z as $\text{ad}Y(Z) = [Y, Z]$.

Proposition 2. [10, Proposition 3.3] *The following equalities hold for any vector field Z and any scalar $u \in U$:*

$$\frac{\partial}{\partial u} \text{Ad}_u Z = \text{ad}X_u^+(\text{Ad}_u Z), \quad \frac{\partial}{\partial u} \text{Ad}_u^{-1} Z = \text{ad}X_u^-(\text{Ad}_u^{-1} Z).$$

In the multidimensional control case these equalities take the form

$$\frac{\partial}{\partial u^r} \text{Ad}_u Z = \text{ad}X_{u,r}^+(\text{Ad}_u Z), \quad \frac{\partial}{\partial u^r} \text{Ad}_u^{-1} Z = \text{ad}X_{u,r}^-(\text{Ad}_u^{-1} Z).$$

3 Local controllability and geometric optimality

We first assume that the control is scalar, $u \in U \subset \mathbb{R}$. Consider an admissible control sequence $\bar{u} = (u_1, \dots, u_N)$. Given an initial point x_0 , we say that system Σ is *N-step locally controllable* from x_0 along the trajectory $x(t)$ corresponding to \bar{u} (shortly, (x_0, \bar{u}) -locally controllable) if

$$x(N) \in \text{int}\mathcal{R}^+(x_0, N).$$

The system Σ will be called *strongly (x_0, \bar{u}) -locally controllable* if for any neighborhood $\bar{U} \subset U^N$ of \bar{u} we have

$$x(N) \in \text{int}\mathcal{R}^+(x_0, \bar{U}). \quad (3)$$

Then Σ has this property, with the same x_0 , for any control sequence \tilde{u} of length $N' > N$ with \bar{u} being its initial part.

In order to state a sufficient condition for local controllability we need the following notation. For a fixed control sequence $\bar{u} = (u_1, \dots, u_N)$ we introduce the *first variation vector fields*

$$Y_{\bar{u}}^i = Y_{u_1 \dots u_i}^i := \text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} X_{u_i}^+, \quad i = 1, \dots, N$$

(in particular, $Y_{\bar{u}}^1 = X_{u_1}^+$). Note that they have the recurrence property

$$Y_{u_1 \dots u_i}^i = \text{Ad}_{u_1} Y_{u_2 \dots u_i}^{i-1}.$$

For a given x we introduce the space of vectors in $T_x M$

$$L_{\bar{u}}(x) = \text{span}\{Y_{\bar{u}}^i(x), \quad i = 1, \dots, N\}. \quad (4)$$

The family of vector fields defining $L_{\bar{u}}(x)$ does not necessarily describe a minimal set of generators for such a subspace. We define a subspace, called the *kernel*, of the vector space of coefficients $a = (a_1, \dots, a_N)$,

$$K_{\bar{u}}(x) = \left\{ a \in \mathbb{R}^N : \sum_{i=1}^N a_i Y_{\bar{u}}^i(x) = 0 \right\}.$$

For $i, j \in \{1, \dots, N\}$ we define the *second variation vector fields*

$$Z_{\bar{u}}^{ij} = Z_{\bar{u}}^{ji} = \frac{1}{2} [Y_{\bar{u}}^i, Y_{\bar{u}}^j], \quad i < j, \quad (5)$$

$$Z_{\bar{u}}^{ii} = \frac{\partial}{\partial u_i} Y_{\bar{u}}^i, \quad (6)$$

where the square bracket denotes the Lie bracket of vector fields on M . Equivalently, $Z_{\bar{u}}^{ij}$ are given by

$$Z_{\bar{u}}^{ij} = Z_{\bar{u}}^{ji} = \frac{1}{2} \text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} [X_{u_i}^+, \text{Ad}_{u_i} \cdots \text{Ad}_{u_{j-1}} X_{u_j}^+], \quad i < j.$$

Given a point $x_0 \in M$, consider the vector-valued quadratic form on the space of parameters

$$H_{\bar{u}}(a) = H(a) = \sum_{i,j=1}^N a_i a_j Z_{\bar{u}}^{ij}(x_0) \quad (7)$$

(the subscript \bar{u} will be omitted). If λ is a covector in $T_{x_0}^*M$, the formula $\lambda H(a) = \sum_{i,j} a_i a_j \lambda(Z_{\bar{u}}^{ij}(x_0))$ defines a real-valued quadratic form. For a quadratic form Q on a finite dimensional real vector space V we denote by Ind^+Q (resp. Ind^-Q) the maximal dimension of a subspace $W \subset V$ such that Q restricted to W is strictly positive definite (resp. strictly negative definite). Recall also that Q is indefinite if there are vectors v and w such that $Q(v) > 0$ and $Q(w) < 0$.

Given a subspace $L \subset T_x M$ we denote by L^\perp its annihilator, $L^\perp = \{\lambda \in T_x^*M : \lambda|_L = 0\}$. We will often use a covector $\lambda \in T_{x_0}^*M$ annihilating all vector fields $Y_{\bar{u}}^i$ at x_0 , i.e.,

$$\lambda Y_{\bar{u}}^i(x_0) = 0, \quad i = 1, \dots, N.$$

This condition is shortly written as $\lambda \in (L_{\bar{u}}(x_0))^\perp$.

Theorem 3. *Assume that Σ satisfies (A1)-(A4). Given a fixed initial point $x_0 \in M$, consider an admissible control sequence $\bar{u} = (u_1, \dots, u_N)$ such that $\bar{u} \in (\text{int } U)^N$ and let $k = \text{codim } L_{\bar{u}}(x_0)$. Then the following statements hold.*

(a) *If $k = 1$ and λH restricted to $K_{\bar{u}}(x_0)$ is indefinite, for any $\lambda \in (L_{\bar{u}}(x_0))^\perp$, $\lambda \neq 0$, then system Σ is strongly (x_0, \bar{u}) -locally controllable.*

(b) *In general, if Σ satisfies the condition*

$$\text{Ind}^-(\lambda H)|_{K_{\bar{u}}(x_0)} \geq k, \quad \forall \lambda \in (L_{\bar{u}}(x_0))^\perp, \lambda \neq 0, \quad (8)$$

then it is strongly (x_0, \bar{u}) -locally controllable.

Note that replacing λ with $-\lambda$ gives the same inequality for Ind^+ . Thus condition (8) says that the quadratic form λH on \mathbb{R}^N , defined by the second variation vectors at x_0 , has at least k positive and k negative “eigenvalues”, when restricted to the subspace $K_{\bar{u}}(x_0) \subset \mathbb{R}^N$.

We will now state a similar result for the multidimensional control, i.e., for $U \subset \mathbb{R}^m$. We fix a control sequence $\bar{u} = (u_1, \dots, u_N) \in U^N$, where $u_i = (u_i^1, \dots, u_i^m)$. Analogously to the scalar control case we define *first variation vector fields*

$$Y_{\bar{u}}^{ir} = \text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} X_{u_i, r}^+,$$

for $i = 1, \dots, N$, $r = 1, \dots, m$. Given a point x we introduce the space of vectors in $T_x M$

$$L_{\bar{u}}(x) = \text{span}\{Y_{\bar{u}}^{ir}(x), \quad i = 1, \dots, N, \quad r = 1, \dots, m\}.$$

Consider the space of coefficients $a = (a_1, \dots, a_N)$, where $a_i = (a_i^1, \dots, a_i^m)$. We will use its subspace

$$K_{\bar{u}}(x) = \left\{ a \in \mathbb{R}^{mN} : \sum_{i=1}^N \sum_{r=1}^m a_i^r Y_{\bar{u}}^{ir}(x) = 0 \right\},$$

called the *kernel*. For $i, j \in \{1, \dots, N\}$, $r, s \in \{1, \dots, m\}$, we define *second variation vector fields*:

$$Z_{\bar{u}}^{ir, js} = Z_{\bar{u}}^{js, ir} = \frac{1}{2} [Y_{\bar{u}}^{ir}, Y_{\bar{u}}^{js}] = \frac{1}{2} \text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} [X_{u_i, r}^+, \text{Ad}_{u_i} \cdots \text{Ad}_{u_{j-1}} X_{u_j, s}^+], \quad i < j,$$

$$Z_{\bar{u}}^{ir, is} = \text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} \frac{\partial}{\partial u_i^r} X_{u_i, s}^+.$$

For $x_0 \in M$ consider the following vector-valued quadratic form on \mathbb{R}^{Nm}

$$H_{\bar{u}}(a) = H(a) = \sum_{\substack{i, j=1, \dots, N \\ r, s=1, \dots, m}} a_i^r a_j^s Z_{\bar{u}}^{ir, js}(x_0). \quad (9)$$

With the above definitions we have the following result.

Theorem 4. *Theorem 3 remains valid in the case of multidimensional control.*

The following converse result will be proved in Section 6 using Theorem 12.

Theorem 5. *If there exists $\lambda \in (L_{\bar{u}}(x_0))^\perp$ such that $\lambda H|_{K_{\bar{u}}(x_0)}$ is positive definite,*

$$\lambda H|_{K_{\bar{u}}(x_0)} > 0,$$

then Σ is not strongly (x_0, \bar{u}) -locally controllable.

Theorems 3 and 4 can be used for obtaining necessary conditions on geometric optimality. Recall that, given an initial point x_0 and control sequence $\bar{u} = (u_1, \dots, u_N)$, the corresponding trajectory $x(i)$, $i = 0, \dots, N$, of system Σ is called *geometrically optimal* if it lies on the boundary of the reachable set, i.e.,

$$x(i) \in \partial \mathcal{R}^+(x_0, i), \quad i = 1, \dots, N.$$

It follows from the invertibility of the system that condition $x(j) \in \text{int } \mathcal{R}^+(x_0, j)$ implies the inclusion $x(k) \in \text{int } \mathcal{R}^+(x_0, k)$, for any $k > j$. Thus, the above condition for geometric optimality is equivalent to $x(N) \in \partial \mathcal{R}^+(x_0, N)$. Theorems 3 and 4 trivially imply

Theorem 6. *If the trajectory of system Σ corresponding to initial point $x_0 \in M$ and an admissible control sequence $\bar{u} = (u_1, \dots, u_N)$ such that $\bar{u} \in (\text{int } U)^N$ is geometrically optimal, then there exists a nonzero covector $\lambda \in (L_{\bar{u}}(x_0))^\perp$ such that*

$$\text{Ind}^-(\lambda H)|_{K_{\bar{u}}(x_0)} < \text{codim } L_{\bar{u}}(x_0).$$

In particular, if $L_{\bar{u}}(x_0)$ is the whole tangent space then the trajectory is not geometrically optimal. If $\text{codim } L_{\bar{u}}(x_0) = 1$ then it is necessary for geometric optimality that there exists a nonzero $\lambda \in (L_{\bar{u}}(x_0))^\perp$ (unique up to a positive multiplier) such that the quadratic form λH restricted to the kernel $K_{\bar{u}}(x_0)$ is non-negative definite.

4 Example

Take $M = \mathbb{R}^3$ and let $U \subseteq \mathbb{R}$ be an open subset. Consider the discrete-time control system on M

$$f_u(x, y, z) = \begin{pmatrix} -x + z + \frac{u^2}{2} \\ xz - y \\ z + \frac{u^2}{2} \end{pmatrix}. \quad (10)$$

Denote the state $\mathbf{x} = (x, y, z)$. As the state is 3-dimensional and the control is scalar, the system is never locally controllable in two steps. It is 3-steps (\mathbf{x}, \bar{u}) -locally controllable if $Y_{\bar{u}}^1, Y_{\bar{u}}^2, Y_{\bar{u}}^3$ are linearly independent at \mathbf{x} (this simply follows from the inverse function theorem). If they are not linearly independent, the index condition from Theorem 3 does not help as it never holds here for 3-steps controls. The reader may analyze the cases where Theorem 4 is applicable.

We will consider 4-steps controls $\bar{u} = (u_1, u_2, u_3, u_4)$. Given an initial condition $\mathbf{x} = (x, y, z)$, we have

$$df_u(\mathbf{x}) = \begin{pmatrix} -1 & 0 & 1 \\ z & -1 & x \\ 0 & 0 & 1 \end{pmatrix}, \quad df_u^{-1}(\mathbf{x}) = \begin{pmatrix} -1 & 0 & 1 \\ -z & -1 & x+z \\ 0 & 0 & 1 \end{pmatrix}, \quad \frac{\partial f}{\partial u} = \begin{pmatrix} u \\ 0 \\ u \end{pmatrix}.$$

The first variation vector fields at \mathbf{x} are:

$$\begin{aligned} Y_{\bar{u}}^1(\mathbf{x}) &= Y_{u_1}^1(\mathbf{x}) = X_{u_1}^+(\mathbf{x}) = \begin{pmatrix} 0 \\ x \\ 1 \end{pmatrix} u_1, \\ Y_{\bar{u}}^2(\mathbf{x}) &= Y_{u_1 u_2}^2(\mathbf{x}) = \text{Ad}_{u_1} Y_{u_2}^1(\mathbf{x}) \\ &= df_{u_1}^{-1}(\mathbf{x}) Y_{u_2}^1(f_{u_1}(\mathbf{x})) = \begin{pmatrix} 1 \\ 2x - \frac{1}{2} u_1^2 \\ 1 \end{pmatrix} u_2, \\ Y_{\bar{u}}^3(\mathbf{x}) &= Y_{u_1 u_2 u_3}^3(\mathbf{x}) = \text{Ad}_{u_1} Y_{u_2 u_3}^2(\mathbf{x}) \\ &= df_{u_1}^{-1}(\mathbf{x}) Y_{u_2 u_3}^2(f_{u_1}(\mathbf{x})) = \begin{pmatrix} 0 \\ 3x - 2z + \frac{1}{2} u_2^2 - u_1^2 \\ 1 \end{pmatrix} u_3, \\ Y_{\bar{u}}^4(\mathbf{x}) &= Y_{u_1 u_2 u_3 u_4}^4(\mathbf{x}) = \text{Ad}_{u_1} Y_{u_2 u_3 u_4}^3(\mathbf{x}) \\ &= df_{u_1}^{-1}(\mathbf{x}) Y_{u_2 u_3 u_4}^3(f_{u_1}(\mathbf{x})) = \begin{pmatrix} 1 \\ 4x - \frac{1}{2} u_1^2 + u_2^2 - \frac{1}{2} u_3^2 \\ 1 \end{pmatrix} u_4. \end{aligned}$$

The second variation vector fields can be computed using the definition (6),

$$\begin{aligned} Z_{\bar{u}}^{11}(\mathbf{x}) &= \begin{pmatrix} 0 \\ x \\ 1 \end{pmatrix}, \\ Z_{\bar{u}}^{22}(\mathbf{x}) &= \begin{pmatrix} 1 \\ 2x - \frac{1}{2}u_1^2 \\ 1 \end{pmatrix}, \\ Z_{\bar{u}}^{33}(\mathbf{x}) &= \begin{pmatrix} 0 \\ 3x - 2z - u_1^2 + \frac{1}{2}u_2^2 \\ 1 \end{pmatrix}, \\ Z_{\bar{u}}^{44}(\mathbf{x}) &= \begin{pmatrix} 1 \\ 4x - \frac{1}{2}u_1^2 + u_2^2 - \frac{1}{2}u_3^2 \\ 1 \end{pmatrix}, \end{aligned}$$

and the definition (5): $Z_{\bar{u}}^{ij} = Z_{\bar{u}}^{ji} = \frac{1}{2}[Y_{\bar{u}}^i, Y_{\bar{u}}^j]$, for $i < j$. Computing the Lie brackets we find that

$$\begin{aligned} Z_{\bar{u}}^{12}(\mathbf{x}) &= \frac{1}{2} \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} u_1 u_2, & Z_{\bar{u}}^{13}(\mathbf{x}) &= \frac{1}{2} \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} u_1 u_3, & Z_{\bar{u}}^{14}(\mathbf{x}) &= \frac{1}{2} \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} u_1 u_4, \\ Z_{\bar{u}}^{23}(\mathbf{x}) &= \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} u_2 u_3, & Z_{\bar{u}}^{24}(\mathbf{x}) &= \frac{1}{2} \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} u_2 u_4, & Z_{\bar{u}}^{34}(\mathbf{x}) &= \frac{1}{2} \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} u_3 u_4. \end{aligned}$$

Case I. Assume that

$$u_1 \neq 0, \quad u_2 \neq 0, \quad u_3 \neq 0, \quad u_4 \neq 0. \quad (11)$$

Then the dimension of $L_{\bar{u}}(\mathbf{x}) = \text{span} \{Y_{\bar{u}}^1, Y_{\bar{u}}^2, Y_{\bar{u}}^3, Y_{\bar{u}}^4\}$ is 3 (then Σ is locally controllable) or it is equal to 2, if

$$x = \frac{1}{4}u_3^2 - \frac{1}{2}u_2^2, \quad z = \frac{1}{4}u_3^2 - \frac{1}{4}u_2^2 - \frac{1}{2}u_1^2. \quad (12)$$

Let $\mathbf{x} = (x, y, z)$ satisfy condition (12). Then $\text{codim } L_{\bar{u}}(\mathbf{x}) = 1$ and the annihilating space $L_{\bar{u}}^\perp(\mathbf{x})$ is generated by the covector

$$\lambda = \left(-x + \frac{1}{2}u_1^2, 1, -x \right). \quad (13)$$

The kernel is

$$K_{\bar{u}}(\mathbf{x}) = \left\{ (a_1, a_2, a_3, a_4) : a_3 = -\frac{u_1}{u_3} a_1, a_4 = -\frac{u_2}{u_4} a_2 \right\} \subseteq \mathbb{R}^4.$$

For λ defined by (13) the quadratic form λH is given by

$$\lambda H(a) = \frac{1}{2} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}^T \begin{pmatrix} 0 & -u_1 u_2 & -2u_1 u_3 & -u_1 u_4 \\ -u_1 u_2 & 0 & u_2 u_3 & 2u_2 u_4 \\ -2u_1 u_3 & u_2 u_3 & A & -u_3 u_4 \\ -u_1 u_4 & 2u_2 u_4 & -u_3 u_4 & B \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix},$$

where

$$\frac{1}{2}A = 2x - 2z - u_1^2 + \frac{1}{2}u_2^2, \quad \frac{1}{2}B = 2x + u_1^2 - \frac{1}{2}u_3^2.$$

Assuming (11), we have $A = 0 = B$ at points $\mathbf{x} \in M$ satisfying condition (12), where $\dim L_{\bar{u}}(\mathbf{x}) = 2$. Then

$$\lambda H(a) = \frac{1}{2} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix}^T \begin{pmatrix} 0 & -1 & -2 & -1 \\ -1 & 0 & 1 & 2 \\ -2 & 1 & 0 & -1 \\ -1 & 2 & -1 & 0 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix}, \quad \text{where } A_i = a_i u_i.$$

From the form of the kernel $K_{\bar{u}}$ we see that, for fixed \bar{u} satisfying (11), the vector (a_1, a_2, a_3, a_4) is in the kernel if and only if $A_1 + A_3 = 0$ and $A_2 + A_4 = 0$. Therefore, taking $A_3 = -A_1$ and $A_4 = -A_2$ we get

$$(\lambda H|_{K_{\bar{u}}})(a) = 4(A_1^2 - A_1 A_2 - A_2^2).$$

This expression is indefinite, treated as a quadratic function of the vector (a_1, a_2) parameterizing the kernel $K_{\bar{u}}$, which means that there exist values of a_1 and a_2 such that the quadratic form is positive and negative. Thus, by Theorem 3, for any 4-step control \bar{u} satisfying (11) and any initial point fulfilling (12) the system is strongly (\mathbf{x}, \bar{u}) -locally controllable, even if (12) means that it does not satisfy the first order sufficient condition for local controllability.

Case IIa. For other 4-steps controls, not satisfying (11), we consider only one case where $u_1 = u_3 = 0$ and $u_2 \neq 0, u_4 \neq 0$. In this case $\dim L_{\bar{u}}(\mathbf{x}) = 1$, if the initial condition satisfies $x = -\frac{1}{2}u_2^2$, and $\dim L_{\bar{u}}(\mathbf{x}) = 2$, otherwise. If $x \neq -\frac{1}{2}u_2^2$, we have $k = \text{codim } L_{\bar{u}}(\mathbf{x}) = 1$. Then $(L_{\bar{u}}(\mathbf{x}))^\perp$ is generated by the covector $\lambda = (-1, 0, 1)$, the kernel $K_{\bar{u}}(\mathbf{x})$ consists of vectors $(a_1, 0, a_3, 0)^T$, $a_1, a_3 \in \mathbb{R}$, and the quadratic form λH evaluated on vectors in $K_{\bar{u}}(\mathbf{x})$ is

$$\lambda H(a) = a_1^2 + a_3^2.$$

We see that λH restricted to $K_{\bar{u}}(\mathbf{x})$ is positive definite. It then follows from Theorem 5 that the system is not strongly (\mathbf{x}, \bar{u}) -locally controllable in this case.

Case IIb. If $u_1 = u_3 = 0, u_2 \neq 0, u_4 \neq 0$ and $x = -\frac{1}{2}u_2^2$, Theorem 5 can be used again. Namely, the annihilator $L_{\bar{u}}^\perp(\mathbf{x})$ is spanned by the covectors $\lambda_1 = (-1, 0, 1)$, $\lambda_2 = (2x, -2, 2x)$ and $\lambda \in L_{\bar{u}}^\perp(\mathbf{x})$ has the general form

$$\lambda = a\lambda_1 + b\lambda_2 = (-a + 2bx, -2b, a + 2bx).$$

The corresponding quadratic form is

$$\lambda H = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & 0 & 0 & -2b \\ 0 & 0 & a + b(4z + u_2^2) & 0 \\ 0 & -2b & 0 & 0 \end{pmatrix}.$$

The kernel $K_{\bar{u}}(\mathbf{x})$ is spanned by the vectors

$$V_1 = (1, 0, 0, 0)^T, \quad V_2 = (0, 0, 1, 0)^T, \quad V_3 = (0, 1, 0, -1)^T.$$

In the basis V_1, V_2, V_3 the matrix $\{\lambda H(V_i, V_j)\}$ of the quadratic form $\lambda H|_{K_{\bar{u}}(\mathbf{x})}$ is

$$\lambda H|_{K_{\bar{u}}(\mathbf{x})} = \text{diag}\{a, a + b(4z + u_2^2), 4b\}.$$

It is positive definite if $a \gg b > 0$. Using Theorem 5 we see that also in this case the system is not strongly (\mathbf{x}, \bar{u}) -locally controllable.

Conclusions. (a) For any 4-steps control \bar{u} satisfying (11) (thus, for any N-steps control with initial part \bar{u}) and any initial condition \mathbf{x} the system (10) is strongly (\mathbf{x}, \bar{u}) -locally controllable. (b) If \bar{u} satisfies $u_1 = u_3 = 0$ and $u_2 \neq 0 \neq u_4$ then (10) is not strongly (\mathbf{x}, \bar{u}) -locally controllable.

5 Proof of local controllability

We will prove the sufficiency result in Theorem 3, only (the proof of Theorem 4 is analogous). The proof, as well as further optimality results, are based on the following lemmata.

Given a control sequence $\bar{u} = (u_1, \dots, u_N)$ we define the composed map $f_{\bar{u}} : M \rightarrow M$,

$$f_{\bar{u}} = f_{u_N} \cdots f_{u_1},$$

which is a diffeomorphism. Let $W = \text{int } U \times \cdots \times \text{int } U \subset \mathbb{R}^N$ and consider the map $F : W \rightarrow \mathbb{R}^n$ defined for a fixed $x_0 \in M$ by

$$F(\bar{u}) = f_{\bar{u}}(x_0), \quad (14)$$

Lemma 7. *We have*

$$\text{Im } dF(\bar{u}) = df_{\bar{u}}(x_0)L_{\bar{u}}(x_0).$$

Lemma 8. *The kernel of $dF(\bar{u})$ is given by*

$$\ker dF(\bar{u}) = K_{\bar{u}}(x_0).$$

Lemma 9. *The second differential of F at \bar{u} , restricted to the kernel $\ker dF(\bar{u}) = K_{\bar{u}}(x_0)$, coincides with $df_{\bar{u}}(x_0)H$ restricted to this kernel,*

$$d^2F(\bar{u})|_{K_{\bar{u}}(x_0)} = df_{\bar{u}}(x_0)H|_{K_{\bar{u}}(x_0)}.$$

Proof of Lemma 7. For $x_0 \in M$ and a control sequence $\bar{u} = (u_1, \dots, u_N) \in (\text{int } U)^N$ denote $x_i = f_{u_i} \cdots f_{u_1}(x_0)$, $i = 1, \dots, N$. Then $x_i = f_{u_i}(x_{i-1})$ and the image of $dF(\bar{u})$ is spanned by the vectors

$$\begin{aligned} \frac{\partial}{\partial u_i} f_{u_N} \cdots f_{u_i} \cdots f_{u_1}(x_0) &= d(f_{u_N} \cdots f_{u_{i+1}})(x_i) \frac{\partial}{\partial u_i} f_{u_i}(f_{u_{i-1}} \cdots f_{u_1}(x_0)) \\ &= df_{\bar{u}}(x_0)(df_{u_1}(x_0))^{-1} \cdots (df_{u_i}(x_{i-1}))^{-1} \frac{\partial}{\partial u_i} f_{u_i}(f_{u_{i-1}} \cdots f_{u_1}(x_0)) \\ &= df_{\bar{u}}(x_0)(\text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} X_{u_i}^+)(x_0) = df_{\bar{u}}(x_0)Y_{\bar{u}}^i(x_0). \end{aligned}$$

This proves the lemma because of the definition of $L_{\bar{u}}(x_0)$ in (4). ■

Proof of Lemma 8. Given a control sequence $\bar{u} = (u_1, \dots, u_N) \in (\text{int } U)^N$ and a vector of parameters $a = (a_1, \dots, a_N) \in \mathbb{R}^N$, denote

$$u_i^\epsilon = u_i + a_i \epsilon, \quad i = 1, \dots, N,$$

where ϵ is a small real parameter. Fix $x_0 \in M$. In order to find the kernel of the differential $dF(\bar{u})$ it is enough to find such $a = (a_1, \dots, a_N)$ that $\partial f_{u_N}^\epsilon \cdots f_{u_1}^\epsilon(x_0)/\partial \epsilon = 0$, at $\epsilon = 0$. We compute

$$\frac{\partial}{\partial \epsilon} f_{u_N}^\epsilon \cdots f_{u_1}^\epsilon(x_0)|_{\epsilon=0} = \sum_{i=1}^N a_i \frac{\partial}{\partial u_i} f_{u_N} \cdots f_{u_i} \cdots f_{u_1}(x_0) = df_{\bar{u}}(x_0) \left(\sum_{i=1}^N a_i Y_{\bar{u}}^i(x_0) \right)$$

where we use the equality established in the preceding proof. Since $df_{\bar{u}}(x_0)$ is an isomorphism, the above sum is equal to zero if and only if $\sum_{i=1}^N a_i Y_{\bar{u}}^i(x_0) = 0$, which ends the proof. \blacksquare

Proof of Lemma 9. We will use the notation and the results from the preceding proofs. In order to compute $d^2 F(\bar{u})$ on the kernel $K_{\bar{u}}(x_0)$ it is enough to compute the second order derivative $\frac{\partial^2}{\partial \epsilon^2} f_{u_N}^\epsilon \cdots f_{u_1}^\epsilon(x_0)|_{\epsilon=0}$, for any vector $a = (a_1, \dots, a_N)$ satisfying $\sum_{i=1}^N a_i Y_{\bar{u}}^i(x_0) = 0$. We have

$$\begin{aligned} \frac{\partial^2}{\partial \epsilon^2} f_{u_N}^\epsilon \cdots f_{u_1}^\epsilon(x_0)|_{\epsilon=0} &= \sum_{i,j=1}^N a_i a_j \frac{\partial}{\partial u_i} \frac{\partial}{\partial u_j} f_{u_N} \cdots f_{u_1}(x_0) \\ &= \sum_{i=1}^N a_i \frac{\partial}{\partial u_i} \left(\sum_{j=1}^N a_j \frac{\partial}{\partial u_j} f_{u_N} \cdots f_{u_1}(x_0) \right) \\ &= \sum_{i=1}^N a_i \frac{\partial}{\partial u_i} \left(df_{\bar{u}}(x_0) \left(\sum_{j=1}^N a_j Y_{\bar{u}}^j(x_0) \right) \right) \\ &= df_{\bar{u}}(x_0) \left(\sum_{i=1}^N \sum_{j=1}^N a_i a_j \frac{\partial}{\partial u_i} Y_{\bar{u}}^j(x_0) \right), \end{aligned}$$

where in the last two equalities we use the equality shown in the proof of Lemma 8 and the fact that $\sum_j a_j Y_{\bar{u}}^j(x_0) = 0$. If $i < j$ then, using Proposition 2, we get

$$\begin{aligned} \frac{\partial}{\partial u_i} Y_{\bar{u}}^j(x_0) &= \frac{\partial}{\partial u_i} \left(\text{Ad}_{u_1} \cdots \text{Ad}_{u_{j-1}} X_{u_j}^+ \right)(x_0) \\ &= \left(\text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} [X_{u_i}^+, \text{Ad}_{u_i} \cdots \text{Ad}_{u_{j-1}} X_{u_j}^+] \right)(x_0) = [Y_{\bar{u}}^i, Y_{\bar{u}}^j](x_0) = 2Z_{\bar{u}}^{ij}(x_0). \end{aligned}$$

For $i = j$ we have

$$\frac{\partial}{\partial u_i} Y_{\bar{u}}^i(x_0) = Z_{\bar{u}}^{ii}(x_0).$$

Finally, $\frac{\partial}{\partial u_i} Y_{\bar{u}}^j(x_0) = 0$ if $i > j$. Thus, for $a \in K_{\bar{u}}(x_0)$, we obtain

$$(d^2 F(\bar{u}))(a) = \frac{\partial^2}{\partial \epsilon^2} f_{u_N}^\epsilon \cdots f_{u_1}^\epsilon(x_0)|_{\epsilon=0} = df_{\bar{u}}(x_0) \left(\sum_{i=1}^N \sum_{j=1}^N a_i a_j Z_{\bar{u}}^{ij}(x_0) \right) = df_{\bar{u}}(x_0) H(a)$$

and the proof is complete. \blacksquare

The proof of Theorem 3 is based on Lemmata 7, 8, 9 and the following result [2, Theorem 20.3].

Lemma 10. Let $F : W \rightarrow \mathbb{R}^n$ be a map of class C^2 , where $W \subset \mathbb{R}^k$ is an open subset. Fix a point $w_0 \in W$ and denote the kernel and the image of the differential $dF(w_0)$ of F at w_0 by

$$K = \ker dF(w_0) \quad \text{and} \quad L = \text{Im } dF(w_0).$$

If $k := \text{codim } L \geq 1$ and

$$\text{Ind}^- \lambda d^2 F(w_0)|_K \geq k \quad \forall \lambda \in L^\perp \subset (\mathbb{R}^n)^*, \quad \lambda \neq 0,$$

then $F(w_0) \in \text{int } F(W)$. (Here $\lambda d^2 F(w_0)|_K$ denotes the quadratic form $\lambda d^2 F(w_0)$ restricted to K .)

Proof of Theorem 3. Consider the map (14). By the definition of strong local controllability (3) it is enough to show that, for a control sequence $\bar{u} = (u_1, \dots, u_N)$, the point $F(\bar{u})$ is in the interior of the image $F(W)$ of some neighborhood W of \bar{u} . This is guaranteed if the assumptions of Lemma 10 are satisfied, for $w_0 = \bar{u}$. From the hypothesis of Theorem 3 it follows that we can use Lemmata 7, 8, 9 and we see that the assumption of Lemma 10 is fulfilled. This completes the proof. ■

6 Optimal control

The general local characteristics of the system, defined at a given initial state x_0 and corresponding to a given control sequence \bar{u} (see Section 3), may also be used for analyzing optimality of the control. In this case standard tools from optimization theory are available. Choosing simple optimal control problems we indicate the use in these problems of the geometric objects introduced earlier: the first variation vector fields $Y_{\bar{u}}^i$ and the space $L_{\bar{u}}(x_0)$ spanned by them at x_0 , the kernel $K_{\bar{u}}(x_0)$, the second variation vector fields $Z_{\bar{u}}^{ij}$ and the corresponding Hessian matrix H .

Consider the following optimal control problems. Given a system

$$\Sigma : \quad x(t) = f(x(t-1), u(t)), \quad x(0) = x_0, \quad x(t) \in M, \quad u(t) \in U \subseteq \mathbb{R}^m,$$

satisfying conditions (A1)-(A4) and given an initial point $x_0 \in M$, find a control sequence $u(t)$, $t = 1, \dots, N$, which minimizes a function

$$(P1) : \quad \varphi(x(N)).$$

The function $\varphi : M \rightarrow \mathbb{R}$, called *final cost*, is assumed of class C^2 . The number of steps N is assumed fixed. Another version of the problem is obtained if, instead, we minimize a *cost functional*

$$(P2) : \quad \varphi(x(N)) + \sum_{t=1}^N c(x(t-1), u(t)),$$

where $c : M \times U \rightarrow \mathbb{R}$ is called *cost function* and assumed of class C^2 . With analogy to continuous-time systems we will call (P1) Meyer problem and (P2) Bolza problem.

The Bolza problem can be reduced to the Meyer problem by introducing an additional state coordinate $x^0(t) = \sum_{i=1}^t c(x(i-1), u(i))$. This coordinate satisfies the additional state equation

$$x^0(t) = x^0(t-1) + c(x(t-1), u(t)), \quad x^0(0) = 0,$$

and then problem (P2) is equivalent to problem (P1) with augmented state $\hat{x} = (x^0, x)$ and the final cost $\hat{\varphi}(\hat{x}(N)) = \varphi(x(N)) + x^0(N)$.

We will state our results for problem (P1), only, using the notation introduced in Section 3. As earlier, we denote $u(i) = u_i$, $x(i) = x_i$, and $f_u(x) = f(x, u)$. A control sequence $\bar{u} = (u_1, \dots, u_N)$ is called *locally optimal* for (P1) or (P2) if it is optimal among all sequences in a neighborhood $\bar{U} \subset U^N$ of \bar{u} .

Theorem 11. *If \bar{u} is a locally optimal control sequence for problem (P1) and $\bar{u} \in (\text{int } U)^N$ then the covector*

$$\lambda = d\varphi(x_N) df_{u_N}(x_{N-1}) \cdots df_{u_1}(x_0) \quad (15)$$

satisfies

$$\lambda Y_{\bar{u}}^i(x_0) = 0, \quad i = 1, \dots, N, \quad (I)$$

and

$$\lambda H|_{K_{\bar{u}}(x_0)} \geq 0. \quad (II)$$

In condition (II) the inequality means that the quadratic form is non-negative definite. Conditions (I) and (II) can be called, respectively, first order and second order necessary conditions for local optimality. Clearly, (I) is equivalent to $\lambda \in (L_{\bar{u}}(x_0))^\perp$. A converse result is more complicated.

Theorem 12. *Given an initial point x_0 and a control $\bar{u} \in (\text{int } U)^N$, let λ be the covector defined by (15). Assume that (I) holds and condition (II) is strengthened to*

$$\lambda H|_{K_{\bar{u}}(x_0)} > 0. \quad (III)$$

Then there exists a quadratic form Q on the image $L \subset T_{x_N}M$ of the composed map

$$df_{u_N}(x_{N-1}) \cdots df_{u_1}(x_0)$$

such that if

$$d^2\varphi(x_N)|_L > Q \quad (IV)$$

then \bar{u} is locally optimal for problem (P1). Here Q depends on system Σ and x_0 , \bar{u} and $d\varphi(x_N)$.

Remark 13. Condition (IV) means that the quadratic form $d^2\varphi(x_N)|_L - Q$ is positive definite. The form Q will be determined by formulae (22), (23) in the proof. Note that if (IV) does not hold for a given φ then it is satisfied for another φ with the same $d\varphi(x_N)$ and suitable $d^2\varphi(x_N)$.

Remark 14. Clearly, the problem (P1) can be reduced to a standard optimization problem. Then standard second order conditions for optimality can be used, as will be seen at the beginning of the proof of Theorem 11. However, these conditions are impractical in use as they involve multiple compositions of nonlinear maps. Neither they give much geometric insight into the problem.

Proof of Theorem 5. The assumptions of the theorem imply that there is a covector λ at $x_0 \in M$ such that conditions (I) and $\lambda H|_{K_{\bar{u}}(x_0)} > 0$ hold. Define a covector $\tilde{\lambda}$ at $x_N = f_{u_N} \cdots f_{u_1}(x_0)$,

$$\tilde{\lambda} = \lambda (df_{u_1}(x_0))^{-1} \cdots (df_{u_N}(x_{N-1}))^{-1},$$

and the linear function $\varphi(x) = \tilde{\lambda}x$. Then x_0 , \bar{u} and φ satisfy assumptions (I) and (III) in Theorem 12. Let \tilde{Q} be a symmetric matrix such that $\tilde{Q}|_L$ satisfies condition (IV) in Theorem 12. Define another function

$$\tilde{\varphi}(x) = \tilde{\lambda}x + \frac{1}{2}(x - x_N)^T(\tilde{Q} + \varepsilon I)(x - x_N),$$

where I is identity matrix and $\varepsilon > 0$. Then $\tilde{\varphi}$ satisfies condition (IV), too, (with Q replaced by \tilde{Q}) and Theorem 12 implies that control \bar{u} is locally optimal, for problem (P1) with final cost $\tilde{\varphi}$.

The function $\tilde{\varphi}$ has regular level sets in a neighborhood of x_N . Since optimality of \bar{u} implies that the whole local reachable set from x_0 (obtained using controls \bar{v} in a neighborhood of \bar{u}) lies above or on the level set of the minimal value, no neighborhood of x_N is covered by the local N-step reachable set. Thus Σ is not strongly (x_0, \bar{u}) -locally controllable. ■

Proof of Theorem 11. We use the notation introduced in Section 5, formula (14), and assume that a local coordinate system is chosen in a neighborhood of $x_N \in M$. We have

$$x_N = F(\bar{u}) = f_{\bar{u}}(x_0),$$

where $f_{\bar{u}}$ is the composition $f_{\bar{u}} = f_{u_N} \cdots f_{u_1}$. Clearly, problem (P1) is equivalent to minimization of the composed function

$$\psi(\bar{u}) = \varphi \circ F(\bar{u}).$$

Since $d\psi(\bar{u}) = d\varphi(x_N) dF(\bar{u})$, the first order necessary condition for minimum can be written as

$$d\varphi(x_N) dF(\bar{u}) = 0 \tag{16}$$

and the second order condition

$$d^2\psi(\bar{u}) = d\varphi(x_N) d^2F(\bar{u}) + d^2\varphi(x_N)(dF(\bar{u}) \cdot, dF(\bar{u}) \cdot) \geq 0. \tag{17}$$

Above we treat $d^2F(\bar{u})$ as a vector-valued symmetric bilinear form and $d^2\varphi(x_N)$ as a symmetric bilinear form. The inequality means that the corresponding quadratic form is non-negative definite. If the above quadratic form is restricted to the kernel of $dF(\bar{u})$, equal to the space $K_{\bar{u}}(x_0)$ by Lemma 8, then the second term vanishes and we get the condition

$$d\varphi(x_N) d^2F(\bar{u})|_{K_{\bar{u}}(x_0)} \geq 0. \tag{18}$$

We will show that conditions (16) and (18) are equivalent to assertions (I) and (II) of the theorem.

We first prove equivalence of conditions (16) and (I). Note that (16) can be written in the form

$$d\varphi(x_N) \operatorname{Im} dF(\bar{u}) = 0. \tag{19}$$

Since the covector λ in (15) is equal to $\lambda = d\varphi(x_N) df_{\bar{u}}(x_0)$ and $\text{Im } dF(\bar{u}) = df_{\bar{u}}(x_0)L_{\bar{u}}(x_0)$, by Lemma 7, we can write (19) as

$$\lambda L_{\bar{u}}(x_0) = 0,$$

which is equivalent to condition (I).

The equality $d^2F(\bar{u})|_{K_{\bar{u}}(x_0)} = df_{\bar{u}}(x_0)H|_{K_{\bar{u}}(x_0)}$ from Lemma 9 implies that condition (18) is equivalent to

$$\lambda H|_{K_{\bar{u}}(x_0)} \geq 0 \quad (20)$$

which is condition (II) in the theorem. ■

Proof of Theorem 12. We use the notation from the above proof. Notice that condition (16) and condition (17) strengthened to strong inequality (positive definiteness) are standard sufficient conditions for local optimality of the point \bar{u} , for the function ψ . Thus it is enough to show that (16) and the strengthened version of (17) follow from the assumptions of the theorem.

We have seen in the first part of the proof of Theorem 11 that assumption (I) is equivalent to condition (16). Thus we should prove that, with appropriately chosen $d^2\varphi(x_N)$, the strengthened version of (17) holds. This means that, for any nonzero vector v , we should have

$$d\varphi(x_N) d^2F(\bar{u})(v, v) + d^2\varphi(x_N)(dF(\bar{u})v, dF(\bar{u})v) > 0. \quad (21)$$

In the standard basis in \mathbb{R}^{mN} we can identify the bilinear forms appearing in (21) with symmetric matrices

$$A \simeq d\varphi(x_N) d^2F(\bar{u}), \quad B \simeq d^2\varphi(x_N)(dF(\bar{u}) \cdot, dF(\bar{u}) \cdot).$$

Denote for brevity $K = K_{\bar{u}}(x_0)$. Define the subspace $\ker A = \{v \in \mathbb{R}^{mN} : Av = 0\} \subset \mathbb{R}^{mN}$ and notice that $\ker A \cap K = \{0\}$. This follows from condition (III). Namely, the quadratic form A restricted to K is equal to $\lambda H|_K$, by Lemma 9 and the definition (15) of λ . Thus $A|_K$ is positive definite. Since A restricted to $\ker A$ is zero, the intersection of $\ker A$ and K must be trivial.

We will bring the bilinear form A to a block-diagonal form. Since $\ker A \cap K = \{0\}$, we can choose a complement E of $\ker A$ in \mathbb{R}^{mN} so that $K \subset E$. Since A is nondegenerate on E and it is positive definite on K , the A -orthogonal complement $K^\perp = \{v \in E : v^T A w = 0 \ \forall w \in K\}$ of K in E has trivial intersection with K . Thus, $E = K \oplus K^\perp$ and \mathbb{R}^{mN} is the direct sum

$$\mathbb{R}^{mN} = K \oplus K^\perp \oplus \ker A.$$

From the definitions of $\ker A$ and K^\perp it follows that, in this decomposition,

$$A = \begin{pmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & B_{22} & B_{23} \\ 0 & B_{32} & B_{33} \end{pmatrix},$$

where the zeros in B appear because $v^T B w = 0$ if one of the vectors v, w belongs to $K = \ker dF(\bar{u})$.

From the definition of B and the fact that $dF(\bar{u})$ is injective on the complement $K^\perp \oplus \ker A$ of $K = \ker dF(\bar{u})$ we deduce that the blocks B_{ij} can be chosen arbitrarily, with an appropriate

choice of $d^2\varphi(x_N)$. Our aim is to characterize those $d^2\varphi(x_N)$ which make B satisfying (21), that is, $A + B > 0$.

The weaker condition $A + B \geq 0$ is satisfied, if we choose $B = \tilde{B}$ with $\tilde{B}_{22} = -A_{22}$ and the remaining $\tilde{B}_{ij} = 0$, since $A_{11} = A|_K = \lambda H|_K > 0$, as proved earlier. Such matrix \tilde{B} is obtained with choosing $\varphi = \tilde{\varphi}$ so that $d^2\tilde{\varphi}(x_N) = \tilde{Q}$, where \tilde{Q} is a symmetric matrix characterized via the equality

$$S^T \tilde{Q} S = \tilde{B} \quad (22)$$

and S is the matrix of the linear map $dF(\bar{u}) : \mathbb{R}^{mN} \rightarrow T_{x_N}M$, in a linear basis in $T_{x_N}M$.

Now we define

$$Q = \tilde{Q}|_L, \quad (23)$$

that is, Q is the quadratic form determined by the matrix \tilde{Q} in (22), restricted to the subspace $L = \text{Im } dF(\bar{u}) = \text{Im } S$, and φ is chosen so that $d^2\varphi(x_N)|_L > Q$. Then we have $A + B > 0$, where $B = S^T d^2\varphi(x_N) S$. Indeed, then $S^T d^2\varphi(x_N) S > S^T \tilde{Q} S$ (since the map defined by S is injective) and $A + B$ is block-diagonal, with two positive definite blocks, thus $A + B > 0$. ■

Remark 15. It follows from the above proof that the quadratic form Q defined by (22), (23) has the property that for any other quadratic form Q' having the property in the theorem we have $Q' \geq Q$. The form Q is uniquely defined by this minimality property.

Example 16. Consider the system given by the dynamics (10) and a final cost function $\varphi : M \rightarrow \mathbb{R}$. Then, for any initial condition, a control sequence $\bar{u} = (u_1, u_2, u_3, u_4)$ such that $u_i \neq 0$ for all i can not be optimal for problem (P1) if only $d\varphi(x_N) \neq 0$ at the final point x_N . To prove suppose that \bar{u} is optimal. Then the covector λ given by (15) satisfies condition (I), thus $\lambda \in (L_{\bar{u}}(\mathbf{x}))^\perp$. We have checked in Section 4 (Case I) that either $\dim L_{\bar{u}}(\mathbf{x}) = 3$ (then $\lambda = 0$ and $d\varphi(x_N) = 0$), or $\dim L_{\bar{u}}(\mathbf{x}) = 2$ and then any $\lambda \in (L_{\bar{u}}(\mathbf{x}))^\perp$ is unique up to a multiplier, thus it can be taken as in formula (15). It was checked in Section 4 that, for such nonzero λ , the quadratic form $\lambda H|_{K_{\bar{u}}(\mathbf{x})}$ is indefinite. Thus the necessary condition (III) in Theorem 11 is not satisfied and we conclude that \bar{u} is not an optimal control, if $d\varphi(x_N) \neq 0$.

Cases IIa and IIb in the example in Section 4 admit optimal control sequences. This can be seen from Theorem 12. Namely, in Case IIa we have $\text{codim } L_{\bar{u}}(\mathbf{x}) = 1$ and covectors $\lambda \in (L_{\bar{u}}(\mathbf{x}))^\perp$ (equivalently λ satisfying condition (I)) are unique, up to multiplier. Let $d\varphi(x_N)$ be such that the corresponding λ given by (15) satisfies the first order condition (I). Then $\lambda \in (L_{\bar{u}}(\mathbf{x}))^\perp$ and we have checked that, for such λ , the quadratic form λH restricted to the kernel is positive definite, thus it satisfies the sufficiency condition (III). From Theorem 12 we deduce that the control \bar{u} is optimal, provided that $d^2\varphi(x_N) > \tilde{Q}$, for some symmetric matrix \tilde{Q} depending on \bar{u} and the initial state (this matrix is determined from condition (22)). In fact, it is enough that $d^2\varphi(x_N)|_L > \tilde{Q}|_L$, where $L = \text{Im } dF(\bar{u})$. The conclusion in Case IIb is similar.

7 Optimality conditions using Hamiltonian

We will now state second order necessary conditions for optimal control problems applying a formalism similar to Hamiltonian formalism used for continuous-time systems. In order to be able to use duality of vectors and covectors we assume that M is an open subset of an affine or a vector space. In this case the difference $x(t) - x(t-1)$ of two consecutive states can be treated as a vector.

For simplicity we will assume that M is an open subset of a vector space,

$$M \subset \mathbb{R}^n.$$

In this case, we can canonically identify all the spaces of (tangent) vectors so that $T_x M = \mathbb{R}^n$, for any $x \in M$, and similarly identify the spaces of covectors, $T_x^* M = (\mathbb{R}^n)^*$.

As earlier, we consider an initialized system $\Sigma : x(t) = f(x(t-1), u(t))$, $x(0) = x_0$. We define the Hamiltonian $H : M \times (\mathbb{R}^n)^* \times U \rightarrow \mathbb{R}$ of Σ ,

$$H(x, p, u) = pf(x, u),$$

where we treat p as a linear function acting on $v = f(x, u)$, in coordinates $pv = \sum_i p_i v^i$.

Remark 17. If M is an open subset of an affine space then a more natural Hamiltonian is

$$H(x, p, u) = p(f(x, u) - x),$$

since $v = f(x, u) - x$ is a vector and pv is well defined by duality of covectors and vectors. Further considerations hold with both definitions, with necessary modifications stated in remarks.

In order to state the second order optimality conditions in terms of the Hamiltonian we consider a control sequence $\bar{u} = (u_1, \dots, u_N)$, where $u_t = u(t)$, $t = 1, \dots, N$, and the corresponding trajectory (x_0, \dots, x_N) , $x_t = x(t)$. We can restrict the control sequence to its initial part

$$\bar{u}^i = (u_1, \dots, u_i),$$

for any fixed $i = 1, \dots, N$. All the definitions from the previous sections work if the sequence \bar{u} is replaced with the restricted sequence \bar{u}^i . In particular, using definitions from Section 3, we denote the vector fields and the Hessian matrix corresponding to the restricted sequence by

$$L_{\bar{u}}^i(x_0) := L_{\bar{u}^i}(x_0), \quad K_{\bar{u}}^i(x_0) := K_{\bar{u}^i}(x_0), \quad H_{\bar{u}}^i := H_{\bar{u}^i}.$$

Then

$$L_{\bar{u}}^1(x_0) \subset \dots \subset L_{\bar{u}}^N(x_0)$$

and

$$K_{\bar{u}}^1(x_0) \subset \dots \subset K_{\bar{u}}^N(x_0)$$

where, in the latter case, we use the natural embeddings

$$\mathbb{R}^m \subset \mathbb{R}^m \times \mathbb{R}^m \subset \dots \subset \mathbb{R}^m \times \overset{N \text{ times}}{\dots} \times \mathbb{R}^m$$

given by $(u_1, \dots, u_i) \mapsto (u_1, \dots, u_i, 0, \dots, 0)$.

Consider again the optimality problem (P1) from Section 6, for system Σ satisfying (A1)-(A4). Theorem 11 can be reformulated in terms of the Hamiltonian in the following way.

Theorem 18. *If $(u(1), \dots, u(N)) \in (\text{int } U)^N$ is an optimal control sequence for problem (P1) and $x(0), \dots, x(N)$ is the corresponding trajectory, then there exists a sequence of covectors $p(0), \dots, p(N)$, with $p(N) = d\varphi(x_N)$, such that the following conditions are satisfied. The state equations hold*

$$x(t) = \frac{\partial H}{\partial p}(x(t-1), p(t), u(t)), \quad t = 1, \dots, N, \quad (\Sigma)$$

together with the adjoint equations

$$p(t-1) = \frac{\partial H}{\partial x}(x(t-1), p(t), u(t)), \quad t = 1, \dots, N, \quad (\Sigma^*)$$

the criticality condition

$$\frac{\partial H}{\partial u}(x(t-1), p(t), u(t)) = 0, \quad t = 1, \dots, N, \quad (CC)$$

and the second order necessary condition

$$p(0)H_{\bar{u}}^t|_{K_{\bar{u}}^t(x_0)} \geq 0, \quad t = 1, \dots, N. \quad (SO)$$

Remark 19. Note that equations (Σ) together with (Σ^*) are equivalent to

$$x(t) = f(x(t-1), u(t)), \quad t = 1, \dots, N, \quad (\Sigma)$$

$$p(t-1) = p(t) \frac{\partial f}{\partial x}(x(t-1), u(t)), \quad t = 1, \dots, N. \quad (\Sigma^*)$$

Remark 20. In the case of M being a subset of an affine space and the Hamiltonian of the form $H(x, p, u) = p(f(x, u) - x)$ the theorem holds with the modified state equations

$$x(t) - x(t-1) = \frac{\partial H}{\partial p}(x(t-1), u(t)), \quad t = 1, \dots, N, \quad (\Sigma)$$

and the adjoint equations

$$p(t-1) - p(t) = \frac{\partial H}{\partial x}(x(t-1), p(t), u(t)), \quad t = 1, \dots, N, \quad (\Sigma^*)$$

Proof of Theorem 18. Let $u(1) = u_1, \dots, u(N) = u_N$ be an optimal control sequence and $x(1) = x_1, \dots, x(N) = x_N$ the corresponding trajectory starting from $x(0) = x_0$. By Theorem 11 the covector

$$\lambda = d\varphi(x_N) df_{u_N}(x_{N-1}) \cdots df_{u_1}(x_0)$$

satisfies $\lambda Y_{\bar{u}}^i(x_0) = 0$ for $i = 1, \dots, N$ (equivalently, $\lambda \in (L_{\bar{u}}(x_0))^\perp$) and $\lambda H_{\bar{u}}|_{K_{\bar{u}}(x_0)} \geq 0$.

Denoting $p(i) = p_i$ we put

$$p_N = \lambda (df_{u_1}(x_0))^{-1} \cdots (df_{u_N}(x_{N-1}))^{-1}$$

and consider the sequence

$$p_t = p_N df_{u_N}(x_{N-1}) \cdots df_{u_{t+1}}(x_t) = \lambda (df_{u_1}(x_0))^{-1} \cdots (df_{u_t}(x_{t-1}))^{-1},$$

$t = 1, \dots, N-1$ (for $t = 0$ the latter equation does not hold and will not be used). Note that

$$p_0 = \lambda \quad \text{and} \quad p_N = d\varphi(x_N).$$

The above sequence satisfies $p_{t-1} = p_t df_{u_t}(x_{t-1}) = \partial H / \partial x(x_{t-1}, p_t, u_t)$ which is the adjoint equation (Σ^*) . We claim that this solution satisfies the other assertions of the theorem, too. The equation (Σ) is satisfied by the definitions of the sequences $u(t)$, $x(t)$ and $p(t)$ and Remark 19.

We will show that (CC) follows from condition $\lambda Y_{\bar{u}}^i(x_0) = 0$. Namely, for $i = 1$,

$$\begin{aligned} 0 &= \lambda Y_{\bar{u}}^1(x_0) = \lambda X_{u_1}^+(x_0) = p_0 (df_{u_1}(x_0))^{-1} \frac{\partial f_u}{\partial u}(x_0)|_{u=u_1} \\ &= p_1 \frac{\partial f_u}{\partial u}(x_0)|_{u=u_1} = \frac{\partial H}{\partial u}(x_0, p_1, u_1). \end{aligned}$$

For $i = 2, \dots, N$ we have

$$\begin{aligned} 0 &= \lambda Y_{\bar{u}}^i(x_0) = \lambda (\text{Ad}_{u_1} \cdots \text{Ad}_{u_{i-1}} X_{u_i}^+)(x_0) \\ &= \lambda (df_{u_1}(x_0))^{-1} \cdots (df_{u_{i-1}}(x_{i-2}))^{-1} X_{u_i}^+(x_{i-1}) \\ &= p_{i-1} X_{u_i}^+(x_{i-1}) = p_{i-1} (df_{u_i}(x_{i-1}))^{-1} \frac{\partial f_u}{\partial u}(x_{i-1})|_{u=u_i} \\ &= p_i \frac{\partial f_u}{\partial u}(x_{i-1})|_{u=u_i} = \frac{\partial H}{\partial u_i}(x_{i-1}, p_i, u_i), \end{aligned}$$

and we see that condition (CC) is satisfied.

In order to show that condition (SO) holds note first that it holds for $t = N$, by Theorem 11. Namely, with our construction of the adjoint sequence $p(N), \dots, p(0)$ we have $p_0 = p(N) = \lambda$, where λ is as in Theorem 11. We claim that for $1 \leq t < N$ condition (SO) follows from (SO) satisfied for $t = N$. This is rather obvious. Namely, under the natural embedding $\mathbb{R}^{im} \subset \mathbb{R}^{Nm}$ defined by $(u_1, \dots, u_i) \mapsto (u_1, \dots, u_i, 0, \dots, 0)$, we have $K_{\bar{u}}^i(x_0) \subset K_{\bar{u}}^N(x_0)$ which follows from the definition of $K_{\bar{u}}^i(x_0)$. The matrix $H_{\bar{u}}^i$ forms the upper-left block of $H_{\bar{u}}^N$, of size im , corresponding to the subspace $\mathbb{R}^{im} \subset \mathbb{R}^{Nm}$. Thus positive definiteness of the whole matrix implies the same property of the sub-matrix. The proof is complete. \blacksquare

Second order necessary conditions for geometric optimality can also be stated using Hamiltonian.

Theorem 21. *If $\bar{u} = (u(1), \dots, u(N)) \in (\text{int } U)^N$ is a geometrically optimal control sequence of an initialized invertible system Σ and $x(1), \dots, x(N)$ is the corresponding trajectory starting from $x(0) = x_0$, then $\dim L_{\bar{u}}(x_0) < n$ and, for any nonzero $\lambda \in (L_{\bar{u}}(x_0))^\perp$, there exists a sequence of nonzero covectors $p(0), \dots, p(N)$ with $p(0) = \lambda$ such that the state equations (Σ) , the adjoint equations (Σ^*) , the criticality condition (CC) and the following index condition*

$$\text{Ind}^-(\lambda H_{\bar{u}}^t)|_{K_{\bar{u}}^t(x_0)} < \text{codim } L_{\bar{u}}^t(x_0), \quad t = 1, \dots, N, \quad (IC)$$

hold. In particular, (IC) implies that if $\text{codim } L_u^t(x_0) = 0$ for some $t \in \{1, \dots, N\}$ then the trajectory is not geometrically optimal. If $\text{codim } L_u^t(x_0) = 1$, then (IC) means that the quadratic form $(\lambda H_u^t)|_{K_u^t(x_0)}$ is non-negative definite.

Proof of Theorem 21. The proof is analogous to the proof of Theorem 18 with the difference that one should use Theorem 6, instead of Theorem 11, for determining the covector λ with required properties. We leave the details to the reader. ■

References

- [1] A. A. Agrachev. Is it possible to recognize local controllability in a finite number of differentiations? In *Open problems in mathematical systems and control theory*, Comm. Control Engrg. Ser., pages 15–18. Springer, London, 1999.
- [2] A. A. Agrachev and Y. L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.
- [3] F. Albertini and D. D’Alessandro. Observability and forward-backward observability of discrete-time nonlinear systems. *Math. Control Signals Systems*, 15(4):275–290, 2002.
- [4] F. Albertini and E. D. Sontag. Discrete-time transitivity and accessibility: analytic systems. *SIAM J. Control Optim.*, 31(6):1599–1622, 1993.
- [5] R. M. Bianchini and G Stefani. Local controllability along a reference trajectory. In *Analysis and optimization of systems (Antibes, 1986)*, volume 83 of *Lecture Notes in Control and Inform. Sci.*, pages 342–353. Springer, Berlin, 1986.
- [6] M. Fliess and D. Normand-Cyrot. A group-theoretic approach to discrete-time nonlinear controllability. In *Proc. IEEE Conf. Dec. Control*. 1981.
- [7] H. Hermes. On local controllability. *SIAM J. Control Optim.*, 20(2):211–220, 1982.
- [8] B. Jakubczyk and D. Normand-Cyrot. Orbites de pseudogroupes de diffeomorphismes et commandabilite des systemes non-lineaires en temps discret. *Compt. Rend. Paris*, 298(11):257–260, 1984.
- [9] B. Jakubczyk and E. D. Sontag. Nonlinear discrete-time systems. Accessibility conditions. In *Modern optimal control*, volume 119 of *Lecture Notes in Pure and Appl. Math.*, pages 173–185. Dekker, New York, 1989.
- [10] B. Jakubczyk and E. D. Sontag. Controllability of nonlinear discrete-time systems: a Lie-algebraic approach. *SIAM J. Control Optim.*, 28(1):1–33, 1990.

- [11] M. Kawski. A necessary condition for local controllability. In *Differential geometry: the interface between pure and applied mathematics (San Antonio, Tex., 1986)*, volume 68 of *Contemp. Math.*, pages 143–155. Amer. Math. Soc., Providence, RI, 1987.
- [12] A. D. Lewis and R. M. Murray. Configuration controllability of simple mechanical control systems. *SIAM J. Control Optim.*, 35(3):766–790, 1997.
- [13] K. Rieger and K. Schlacher. Implicit discrete-time systems and accessibility. *Automatica J. IFAC*, 47(9):1849–1859, 2011.
- [14] E. D. Sontag. Controllability is harder to decide than accessibility. *SIAM J. Control Optim.*, 26(5):1106–1118, 1988.
- [15] H. J. Sussmann. A sufficient condition for local controllability. *SIAM J. Control Optim.*, 16(5):790–802, 1978.