



**HAL**  
open science

# The averaged control system of fast oscillating control systems

Alex Bombrun, Jean-Baptiste Pomet

► **To cite this version:**

Alex Bombrun, Jean-Baptiste Pomet. The averaged control system of fast oscillating control systems. [University works] 2011, pp.27. hal-00648330v2

**HAL Id: hal-00648330**

**<https://inria.hal.science/hal-00648330v2>**

Submitted on 5 Dec 2011 (v2), last revised 7 Dec 2012 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THE AVERAGED CONTROL SYSTEM OF FAST OSCILLATING CONTROL SYSTEMS\*

ALEX BOMBRUN<sup>†</sup> AND JEAN-BAPTISTE POMET<sup>‡</sup>

**Abstract.** For control systems that either have an explicit periodic dependence on time or have periodic solutions and small controls, we define an *average control system* that takes into account all possible variations of the control, and prove that its solutions approximate all solutions of the oscillating systems as oscillations go faster.

The dimension of its velocity set is characterised geometrically. When it is maximum the average system defines a Finsler metric, unfortunately not twice differentiable in general. Under particular assumptions, valid for the control two body system, this Finsler metric generates a Hamiltonian flow on the cotangent bundle. For minimum time control, this average system proves that averaging the Hamiltonian given by the maximum principle is a valid approximation.

**Key words.** Averaging, control systems, small control, optimal control, Finsler geometry.

**AMS subject classifications.** 34C29, 34H05, 49J15, 93B11, 93C15, 93C70, 53B40

**1. Introduction.** We consider either a “fast-oscillating control system” (1):

$$\dot{x} = u_1 X_1\left(\frac{t}{\varepsilon}, x\right) + \cdots + u_m X_m\left(\frac{t}{\varepsilon}, x\right), \quad \|u\| \leq 1,$$

where all  $X_i$ 's are  $2\pi$ -periodic with respect to  $t/\varepsilon$ , or a “Kepler control system” (43):

$$\dot{\xi} = f_0(\xi) + v_1 f_1(\xi) + \cdots + v_m f_m(\xi), \quad \|v\| \leq \varepsilon$$

where all solutions of  $\dot{\xi} = f_0(\xi)$  are periodic.

Averaging techniques to study perturbations of conservative (periodic or not) dynamical systems date back at least to H. Poincaré; see [2, §52] or [23] for modern expositions (where the formulation  $\dot{x} = \varepsilon F(t, x)$  on a time-interval of order  $1/\varepsilon$  is preferred to our equivalent  $\dot{x} = F(t/\varepsilon, x)$  on a fixed interval).

If  $u$  or  $v$  above is assigned to be a fixed function of state and time (or to be computed from additional state variables like  $u = \alpha(p, x)$ ,  $\dot{p} = g(p, x)$ ), then these techniques can be applied to give an approximation at first order (with respect to small  $\varepsilon$ ) of the movement of the slow variables. Averaging is usually used in this way in control theory: in vibrationnal control [19], fast oscillating controls are designed for possibly time-invariant systems, averaging techniques allows analysis and proof of stability; these have also been used, mostly in control of mechanical systems, both for stability and path planning (controllability), see for instance [8]; in [12, §5], high frequency control is used to approach a non-flat system by a flat one; one may also mention many applications to control [21, 18, 20] of the seminal work [17] that mimicks Lie brackets by highly oscillatory controls along the original vector fields. A common feature to these references is that the use of oscillations “creates” new independent controls used for design. The use of averaging in optimal control of oscillating systems [10, 13, 14, 7] is similar in spirit to the above, but closer to the framework of this paper because oscillations are present in the system instead of introduced by the

---

\* This work was partly supported by *Thales Alenia Space*, in 2004-2007.  
Submitted to *SIAM J. Control. Optim.*, December 5, 2011.

<sup>†</sup>Most of this work was done when this author was with INRIA. [Alex.Bombrun@gmail.com](mailto:Alex.Bombrun@gmail.com).

<sup>‡</sup>INRIA, B.P. 93, 06902 Sophia Antipolis cedex, France. Email: [Jean-Baptiste.Pomet@inria.fr](mailto:Jean-Baptiste.Pomet@inria.fr).

control. Very interesting results are obtained applying averaging to the Hamiltonian equations arising from Pontryagin Maximum principle. For instance, in [7], the authors have studied in this way the problem of minimal energy transfer between two elliptic orbits; extremals are the same as these giving the geodesics of a Riemannian metric. Note that we have again, indirectly, “new independant controls” after averaging, for geodesics of the Riemannian metric can go in any directions, whereas the velocity set of the original system at each point had positive codimension. The same averaging computation may be applied to the Hamiltonian differential equation obtained for minimum time, but, since this differential equation is discontinuous, there is no theoretical justification for averaging in that case.

Our contribution is to introduce a different way of averaging that takes into account all possible variations of the control —hence the control strategy can be decided *after* performing averaging— and to prove that it has satisfying properties in terms of being a good first order approximation of the above systems as  $\varepsilon \rightarrow 0$ , and in terms of regularity. This gives, as a side result, a justification of the use of averaging for minimum time in [13, 14]. This procedure also “creates new independent control”, i.e. increases the dimension of the velocity set, that we characterize in terms of the original vector fields. When this dimension is maximum, the average system defines a Finsler metric [3] on the manifold, whose geodesics are the limits of minimum time trajectories for the original systems as  $\varepsilon \rightarrow 0$ . This Finsler metric is in general not twice differentiable (hence not a Finsler metric in the sense of [3], indeed); we however prove that, at least in the less degenerate case, the Hamiltonian system governing extremals, although not locally Lipschitz, generates a flow on the cotangent bundle. We prove in §5 that this applies to the planar orbit transfer in minimum time.

The average control system may be used for other purposes than optimal control, for instance [4] designing a Lyapunov function for feedback control in the average system and using it for the oscillating systems; indeed the present work was developed out of comparing feedback control based on a priori chosen Lyapunov functions with minimum time-control for low thrust orbital transfer.

Preliminary versions of this work can be found in [5, 4].

*Organization of the paper.* The construction and results are developed for “fast-oscillating control system” in §3 and then transferred in §4 to “Kepler control systems”, and applied to minimum time orbit transfer in the planar 2-body problem in §5. Some computations on the smoothness of the average of the norm of an oscillating vector are postponed until the appendix, that comes after a short conclusion (§6).

**2. Notations and conventions.** **2.1.**  $M$  is a smooth connected manifold of dimension  $n$ ;  $TM$  and  $T^*M$  denote its tangent and cotangent bundles.

**2.2.** We assume that  $M$  is endowed with a distance  $d$ , that is locally Lipschitz equivalent, in any coordinates, to the Euclidean distance in these coordinates. Our results do *not* depend on the choice of  $d$  although the constants, like  $c$  in Theorem 3.4, do depend on it.  $d$  may be the distance associated to a Riemannian metric, may be obtained by patching the euclidean distance in a privileged set of coordinates, or may also, if  $M$  is an open subset of  $\mathbb{R}^n$ , be Euclidean distance.

**2.3.**  $S^1$  is  $\mathbb{R}/2\pi\mathbb{Z}$  and has length  $2\pi$ ; we denote the average of a map  $f: S^1 \rightarrow \mathbb{R}$  by  $\frac{1}{2\pi} \int_0^{2\pi} f(\theta) d\theta$ . More generally, we identify maps  $S^1 \rightarrow E$  (any set  $E$ ) with maps  $\mathbb{R} \rightarrow E$  that are periodic with period  $2\pi$ , and  $L^\infty(S^1, \mathbb{R}^m)$  with the subset of  $L^\infty(\mathbb{R}, \mathbb{R}^m)$  made of  $2\pi$ -periodic functions.

**2.4.** The Euclidean norm in  $\mathbb{R}^m$  or  $(\mathbb{R}^m)^*$  is denoted by  $\|\cdot\|$ , the scalar product by  $(\cdot|\cdot)$ , and the ball of radius one centered at the origin by  $B^m$ .

**2.5.** If  $E$  is a subset of a vector space  $V$ , then  $E^\perp$  is its annihilator, the vector subspace of its dual  $V^*$  made of all  $p$ 's such that  $\langle p, v \rangle = 0$  for all  $v$  in  $E$ .

### 3. Fast oscillating control systems.

**3.1. Systems under consideration.** We call *fast oscillating control system* on a smooth manifold  $M$  of dimension  $n$  a family of non-autonomous systems, linear in the control  $u \in \mathbb{R}^m$ :

$$\dot{x} = \mathcal{G}\left(\frac{t}{\varepsilon}, x\right) u = \sum_{i=1}^m \mathcal{G}_i\left(\frac{t}{\varepsilon}, x\right) u_i, \|u\| \leq 1 \quad (1)$$

indexed by a positive number  $\varepsilon$ ;  $\mathcal{G}$  can be viewed as a  $n \times m$  matrix, or as a mere notation: each  $\mathcal{G}_i \in C^\infty(\mathbb{R} \times M, \text{TM})$ , the  $i^{\text{th}}$  column of  $\mathcal{G}$ , is a smooth time-varying vector field,  $2\pi$ -periodic with respect to time:

$$\mathcal{G}_i(\theta + 2\pi, x) = \mathcal{G}_i(\theta, x), \quad \mathcal{G}(\theta + 2\pi, x) = \mathcal{G}(\theta, x). \quad (2)$$

An admissible control  $u(\cdot)$  is a bounded measurable function of time and the associated solution  $x(\cdot)$  is then absolutely continuous. The parameter  $\varepsilon$  is intended to be small, whence the fast oscillations, and we are interested in the limiting behavior when it tends to zero.

**3.2. Average control system of fast oscillating control systems.** Define the map  $\bar{\mathcal{G}} : M \times L^\infty([0, 2\pi], \mathbb{R}^m) \rightarrow \text{TM}$  by

$$\bar{\mathcal{G}}(x, \mathcal{U}) = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{G}(\theta, x) \mathcal{U}(\theta) d\theta. \quad (3)$$

It allows one to define, for all  $x \in M$ , the subset  $\mathcal{E}(x) \subset T_x M$  by:

$$\mathcal{E}(x) = \{\bar{\mathcal{G}}(x, \mathcal{U}), \mathcal{U} \in L^\infty([0, 2\pi], \mathbb{R}^m), \|\mathcal{U}\|_\infty \leq 1\} \subset T_x M, \quad (4)$$

and the *average control system* of (1) as follows:

DEFINITION 3.1. *The average control system of (1) is the differential inclusion*

$$\dot{x} \in \mathcal{E}(x). \quad (5)$$

A solution of (5) is an absolutely continuous  $x(\cdot) : [0, T] \rightarrow M$  such that  $\dot{x}(t) \in \mathcal{E}(x(t))$  for almost all  $t$ .

From the definition of  $\mathcal{E}(x)$ , a solution  $x(\cdot)$  of (5) is such that, for almost all  $t$ , there is a  $\mathcal{U}(t)$  in  $L^\infty([0, 2\pi], \mathbb{R}^m)$  such that  $\dot{x}(t) = \bar{\mathcal{G}}(x(t), \mathcal{U}(t))$ . For such a choice at random, the map  $(t, \theta) \mapsto \mathcal{U}(t)(\theta)$  will be measurable with respect to  $\theta$  but bear no regularity with respect to  $t$ , even measurability; the following proposition states that it may always be chosen jointly measurable with respect to  $(t, \theta)$ .

PROPOSITION 3.2. *A map  $x : [0, T] \rightarrow \mathbb{R}^n$  is a solution of the differential inclusion (5) if and only if there exists  $\hat{u} \in L^\infty([0, T] \times S^1, \mathbb{R}^m)$ ,  $\|\hat{u}\|_\infty \leq 1$  such that*

$$\dot{x}(t) = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{G}(\theta, x(\tau)) \hat{u}(\tau, \theta) d\theta \quad (6)$$

for almost all  $t$  in  $[0, T]$ .

*Proof.* We assume  $M = \mathbb{R}^n$  and use a Euclidean norm when useful; in general one may partition  $[0, T]$  and do the proof below in coordinates for each portion of curve remaining in the same coordinate chart.

Sufficiency is clear: from Fubini's theorem,  $\theta \mapsto \widehat{u}(t, \theta)$  is measurable for almost any fixed  $t$ , hence  $\dot{x}(t) \in \mathcal{E}(x(t))$  for almost all  $t$  and  $x(\cdot)$  is a solution of (5).

Conversely, assume  $x(\cdot)$  to be a solution of (5):  $\dot{x}(\cdot)$  is measurable and, for almost all  $t$ , there exists  $\tilde{u}_t \in L^\infty(S^1, \mathbb{R}^m)$ ,  $\|\tilde{u}_t\|_\infty \leq 1$  such that

$$\dot{x}(t) = \overline{\mathcal{G}}(x(t), \tilde{u}_t) = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{G}(s_1, x(t)) \tilde{u}_t(s_1) ds_1. \quad (7)$$

Let  $\phi : L^\infty([0, T] \times S^1, \mathbb{R}^m) \rightarrow L^2([0, T], \mathbb{R}^n)$  be the linear map defined by

$$\phi(u)(t) = \overline{\mathcal{G}}(x, u(t, \cdot)) = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{G}(s_1, x(t)) u(t, s_1) ds_1.$$

We shall prove that  $x(\cdot)$  belongs to  $\mathcal{J}$ , the image by  $\phi$  of the unit ball of  $L^\infty([0, T] \times S^1, \mathbb{R}^m)$ . Since  $\mathcal{J}$  is compact and convex in  $L^2([0, T], \mathbb{R}^n)$ , that contains  $\dot{x}(\cdot)$  (essentially bounded according to (7)), the distance from  $\dot{x}$  to  $\mathcal{J}$  is reached for a unique element  $\bar{\xi} \in \mathcal{J}$ :

$$\bar{\xi} = \phi(\bar{u}), \quad \bar{u} \in L^\infty([0, T] \times S^1, \mathbb{R}^m), \quad \|\bar{u}\|_{L^\infty} \leq 1.$$

Let us prove by contradiction that  $\bar{\xi} = \dot{x}$ . If  $\dot{x} \neq \bar{\xi}$ , one has, for all  $u$  in the unit ball of  $L^\infty([0, T] \times S^1, \mathbb{R}^m)$ ,

$$(\dot{x} - \bar{\xi} | \phi(u) - \phi(\bar{u}))_{L^2} \leq 0 \quad (8)$$

with equality only if  $\phi(u) = \phi(\bar{u})$ . Define  $\widehat{u}$  by  $\widehat{u}(t, s) = \mathcal{U}_{(\dot{x}(t) - \bar{\xi}(t))^\top, x(t)}^*(s)$  with  $\mathcal{U}_{p,x}^*$  defined by (27); clearly,  $\widehat{u}$  is in the unit ball of  $L^\infty([0, T] \times S^1, \mathbb{R}^m)$ , and, for all  $(t, s) \in [0, T] \times S^1$  and all  $\mathbf{u} \in \mathbb{R}^m$ ,

$$\|\mathbf{u}\| \leq 1 \quad \Rightarrow \quad (\dot{x}(t) - \bar{\xi}(t))^\top \mathcal{G}(s, x(t)) (\widehat{u}(t, s) - \mathbf{u}) \geq 0, \quad (9)$$

hence  $(\dot{x}(t) - \bar{\xi}(t))^\top \mathcal{G}(s_1, x(t)) (\widehat{u}(t, s_1) - \bar{u}(t, s_1))$  is non-negative for almost all  $(t, s_1)$  and, since it is the integrand of the left-hand side of (8), it must be zero; hence  $\bar{\xi} = \phi(\bar{u}) = \phi(\widehat{u})$  and  $\bar{\xi}(t) = \overline{\mathcal{G}}(x(t), \widehat{u}(t, \cdot))$  for almost all  $t$ .

In (7),  $\tilde{u}_t$  satisfies  $\|\tilde{u}_t(s_1)\| \leq 1$  for almost all  $s_1$ , hence, according to (9),

$$(\dot{x}(t) - \bar{\xi}(t))^\top \mathcal{G}(s_1, x(t)) (\widehat{u}(t, s_1) - \tilde{u}_t(s_1)) \geq 0.$$

Since  $\dot{x}(t) = \overline{\mathcal{G}}(x(t), \tilde{u}_t)$ ,  $\bar{\xi}(t) = \overline{\mathcal{G}}(x(t), \widehat{u}(t, \cdot))$ , integrating with respect to the variable  $s_1$  yields  $-\|\dot{x}(t) - \bar{\xi}(t)\|^2 \geq 0$  for almost all  $t$ ; this contradicts  $\dot{x} \neq \bar{\xi}$ .  $\square$

*Remark 3.3.* It would have been nicer to describe the average of (1) as a control system of the type  $\dot{x} = f(x, v)$ ,  $v \in U$  with  $U$  a subset of  $\mathbb{R}^r$  for some integer  $r$ , and if possible  $f$  one-to-one with respect to  $v$ , rather than the differential inclusion in (5). We are unfortunately not able to give such a one-to-one parameterization of the velocity set  $\mathcal{E}(x)$  by a finite-dimensional ‘‘control’’.

There is however a way of writing a control system equivalent to  $\mathcal{E}(x)$ , where the role of  $U$  is taken by the unit ball of  $L^\infty(S^1, \mathbb{R}^m)$ :

$$\dot{x} = \overline{\mathcal{G}}(x, \mathcal{U}), \quad \mathcal{U} \in L^\infty(S^1, \mathbb{R}^m), \quad \|\mathcal{U}\|_\infty \leq 1. \quad (10)$$

The map  $\overline{\mathcal{G}}$ , defined in (3), is obviously continuous, smooth with respect to  $x$  and linear with respect to  $\mathcal{U}$ . According to Proposition 3.2, it is natural to take as admissible controls for (10) the maps  $t \mapsto \mathcal{U}(t)$  such that  $(t, \theta) \mapsto \mathcal{U}(t)(\theta)$ , denoted  $\widehat{u}$  in the proposition, is jointly measurable with respect to  $(t, \theta)$ . The data of an initial condition  $x(0)$  and a admissible control do give rise to a unique solution  $t \mapsto x(t)$  on some interval. Of course, the parameterization of the velocity set  $\mathcal{E}(x)$  by  $\mathcal{U}$  is far from one-to-one: the control is infinite-dimensional and very “redundant”.

We do not pursue this point of view because it would make the exposition more difficult, but we could have kept the formulation (10) instead of (5); it would for instance have led us to use, instead of the Pontryagin maximum principle for (possibly nonsmooth) differential inclusions taken from [11] that we invoke before Theorem 3.16, a maximum principle for smooth systems with infinite-dimensional control.

**3.3. Convergence theorem.** The following is one of our main results. The use of the distance  $d$  on  $M$  was discussed in §2.2.

**THEOREM 3.4** (Convergence for fast-oscillating control systems).

1. For any solution  $x_0(\cdot) : [0, T] \rightarrow M$  of (5), there exist constants  $c > 0$  and  $\varepsilon_0 > 0$ , and, for each  $\varepsilon$ ,  $0 < \varepsilon < \varepsilon_0$ , a measurable function  $\overline{u}_\varepsilon(\cdot) : [0, T] \rightarrow B^m$  such that, for all positive  $\varepsilon$  smaller than  $\varepsilon_0$ , the solution  $x_\varepsilon(\cdot)$  of (1) with control  $u = \overline{u}_\varepsilon(t)$  and initialized at  $x_\varepsilon(0) = x_0(0)$  is defined on  $[0, T]$ , and one has

$$d(x_\varepsilon(t), x_0(t)) < c\varepsilon, \quad t \in [0, T], \quad (11)$$

thus  $x_\varepsilon(\cdot)$  converges uniformly on  $[0, T]$  to  $x_0(\cdot)$  when  $\varepsilon$  tends to zero.

2. Let  $\mathbb{K}$  be a compact subset of  $M$ ,  $(\varepsilon_n)_{n \in \mathbb{N}}$  a decreasing sequence of positive real numbers converging to zero, and  $(x_n(\cdot))_{n \in \mathbb{N}}$  a sequence of absolutely continuous maps  $[0, T] \rightarrow \mathbb{K}$  such that each  $x_n(\cdot)$  is a solution of system (1) with  $\varepsilon = \varepsilon_n$  and some control  $u = u_n(t)$ ,  $u_n(\cdot) \in L^\infty([0, T], \mathbb{R}^m)$ ,  $\|u_n(\cdot)\|_\infty \leq 1$ . Then the sequence  $(x_n(\cdot))_{n \in \mathbb{N}}$  is compact for the topology of uniform convergence on  $[0, T]$  and any accumulation point is a solution of the system (5).

This result is important because it relates solutions of the fast oscillating systems as  $\varepsilon$  tends to zero to solutions of the average system. To our knowledge, this kind of theorem where the control is not chosen prior to averaging has never been stated in the literature.

It does resemble the classical averaging theorem for ordinary differential equations, for instance [2]. Let alone the fact that the particular class of control systems (1) that we consider makes no sense when there is no control, if we simply translate absence of control in the above theorem by the fact that solutions depend on the initial condition only and not on a function of time called control, we do recover in Point 1 —where the measurable function  $\overline{u}_\varepsilon(\cdot) : [0, T] \rightarrow B^m$  now does not matter— the classical result that the solutions of the oscillating ODEs converge uniformly to the solution of the average ODE with same initial condition with an error of the order of  $\varepsilon$ ; Point 2 would be trivial if solutions would not depend on the control.

*Remark 3.5.* We restrain ourselves to systems (1), linear in the control without drift because they naturally occur when dealing with a time-invariant system whose drift has periodic solutions, see §4 below.

The above could however be proved for slightly more general systems than (1): for instance consider systems that are affine instead of linear in the control by adding a drift vector field  $\mathcal{G}_0(t/\varepsilon, x)$  to the right-hand side of (1). One then obtains an average control system where  $\mathcal{E}(x)$  is replaced by  $\overline{\mathcal{G}}_0(x) + \mathcal{E}(x)$ , with  $\overline{\mathcal{G}}_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{G}_0(\theta, x) d\theta$ .

We leave it to the reader to check whether next properties in this sections carry over or not. On a cultural level, note that, when the maximum rank assumption (31) made in §3.5 holds,  $\overline{\mathcal{G}}_0(x) + \mathcal{E}(x)$  is still a convex body in  $T^*M$  and Remark 3.13 is still valid in that the average system still defines, with the same reservations on smoothness, a Finsler metric, that is not longer “reversible”, *provided the origin belongs to the interior of  $\overline{\mathcal{G}}_0(x) + \mathcal{E}(x)$* . This property could have a big importance in generalizing the next results to systems with drift: we do use at many places in the sequel the fact that equilibrium at any point is a solution. The developments in §3.6 on application to the minimum time problem do not generalize easily, in any case.

Let us now prove Theorem 3.4. The following technical lemma is required.

LEMMA 3.6. *Let  $\varepsilon > 0$  and  $a < b$  be real numbers and  $\widehat{u} : [a - 2\pi\varepsilon, b] \times S^1 \rightarrow \mathbb{R}^m$  be measurable, one has the following identity<sup>1</sup>:*

$$\iint_{\substack{a \leq s \leq b \\ 0 \leq \theta \leq 2\pi}} \mathcal{G}(\theta, x(s)) \widehat{u}(\theta, s) \, d\theta \, ds = \iint_{\substack{a \leq s \leq b \\ 0 \leq \theta \leq 2\pi}} \mathcal{G}\left(\frac{s}{\varepsilon}, x(s)\right) \widehat{u}\left(s + \varepsilon\theta, \frac{s}{\varepsilon}\right) \, d\theta \, ds + \Delta_\varepsilon \quad (12)$$

with

$$\begin{aligned} \Delta_\varepsilon &= \iint_{T_\varepsilon^a} \mathcal{G}\left(\frac{s}{\varepsilon}, x(s + \varepsilon\theta)\right) \widehat{u}\left(s + \varepsilon\theta, \frac{s}{\varepsilon}\right) \, d\theta \, ds - \iint_{T_\varepsilon^b} \mathcal{G}\left(\frac{s}{\varepsilon}, x(s + \varepsilon\theta)\right) \widehat{u}\left(s + \varepsilon\theta, \frac{s}{\varepsilon}\right) \, d\theta \, ds \\ &\quad + \iint_{\substack{a \leq s \leq b \\ 0 \leq \theta \leq 2\pi}} \left[ \mathcal{G}\left(\frac{s}{\varepsilon}, x(s + \varepsilon\theta)\right) - \mathcal{G}\left(\frac{s}{\varepsilon}, x(s)\right) \right] \widehat{u}\left(s + \varepsilon\theta, \frac{s}{\varepsilon}\right) \, d\theta \, ds \end{aligned} \quad (13)$$

and the set  $T_\varepsilon^a$  defined by  $T_\varepsilon^a = \{(\theta, s), 0 \leq \theta \leq 2\pi, a - \varepsilon\theta \leq s \leq a\}$  and  $T_\varepsilon^b$  accordingly.

*Proof.* We constantly use the convention of footnote 1. The map  $(\tau, \phi) \mapsto (\tau - \varepsilon\phi, \tau/\varepsilon)$  is one-to-one from  $\{(\tau, \phi) \in \mathbb{R} \times S^1, 0 \leq \phi < 2\pi, a - \varepsilon\phi \leq \tau \leq b - \varepsilon\phi\}$  onto  $[a, b] \times S^1$ , with Wronskian equal to one. Performing the change of variables  $\theta = \tau/\varepsilon$ ,  $s = \tau + \varepsilon\phi$  in the left-hand side of (12) yields

$$\iint_{\substack{0 \leq \phi \leq 2\pi \\ a - \varepsilon\phi \leq \tau \leq b - \varepsilon\phi}} \mathcal{G}\left(\frac{\tau}{\varepsilon}, x(\tau + \varepsilon\phi)\right) \widehat{u}\left(\tau + \varepsilon\phi, \frac{\tau}{\varepsilon}\right) \, d\tau \, d\phi .$$

Keeping the names  $(s, \theta)$  instead of  $(\tau, \phi)$ , one gets (12), the correcting term  $\Delta_\varepsilon$  coming from the modified domain of integration and the argument of  $x$ .  $\square$

*Proof of Theorem 3.4, point 1.* Consider a solution  $x_0 : [0, T] \rightarrow M^n$  of (5). According to Proposition 3.2 there exists  $u_0 \in L^\infty([0, T] \times S^1, \mathbb{R}^m)$ ,  $\|u_0\|_\infty \leq 1$  satisfying (6). For each  $\varepsilon > 0$ , define  $\overline{u}_\varepsilon(\cdot) \in L^\infty([0, T], \mathbb{R}^m)$  (clearly,  $\|u_\varepsilon\|_\infty \leq 1$ ) by<sup>1</sup>

$$\overline{u}_\varepsilon(t) = \frac{1}{2\pi} \int_0^{2\pi} u_0\left(t + \varepsilon\theta, \frac{t}{\varepsilon}\right) \, d\theta , \quad (14)$$

where  $u_0$  is prolonged by zero outside  $[0, T]$ : by convention  $u_0(t + \varepsilon\theta, \frac{t}{\varepsilon}) = 0$  if  $t + \varepsilon\theta > T$ . Let us prove that this construction of  $\overline{u}_\varepsilon$  satisfies the two announced properties.

<sup>1</sup> In the right-hand side of (12) and (13), as well as along all this proof and the one of Theorem 3.4, there is a confusion between angles and real numbers, that should not be misleading. For instance: -  $\theta$  stands for the real number in  $[0, 2\pi)$  which is a measure of the angle when we write  $s + \varepsilon$ , while it is an angle (counted modulo  $2\pi$ ) in  $\mathcal{G}(\theta, x(s))$ , but we keep the same notation, - conversely  $\frac{s}{\varepsilon}$  is considered as an angle, i.e. it should be considered modulo  $2\pi$ , whereas  $s$  is a real number when we write  $x(s)$ .

Let us first assume that  $M$  is an open subset of  $\mathbb{R}^n$  and  $\mathcal{G}$  is zero outside a compact subset of  $M$ . Then  $\mathcal{G}(\theta, x)$  is a  $n \times m$  matrix for all  $(\theta, x)$  and, denoting by  $\|\cdot\|$  the Euclidean norm for vectors and the operator norm for matrices, there are global constants  $\text{Lip } \mathcal{G}$  and  $\text{sup } \mathcal{G}$  such that, for all  $x, x', \theta$  in  $M \times M \times S^1$ ,

$$\|\mathcal{G}(\theta, x) - \mathcal{G}(\theta, x')\| \leq (\text{Lip } \mathcal{G}) \|x - x'\|, \quad \|\mathcal{G}(\theta, x)\| \leq (\text{sup } \mathcal{G}). \quad (15)$$

Let  $b$  be a non-negative constant and consider, for each  $\varepsilon > 0$ , a solution  $x_\varepsilon(\cdot)$  of (1) with control  $u = \bar{u}_\varepsilon(t)$ , initialized at some  $x_\varepsilon(0)$  such that

$$\|x_\varepsilon(0) - x_0(0)\| \leq b\varepsilon. \quad (16)$$

By definition, expanding  $\bar{u}_\varepsilon(s)$  as in (14) and using Lemma 3.6, one has

$$\begin{aligned} x_\varepsilon(t) &= x_\varepsilon(0) + \frac{1}{2\pi} \int_0^t \int_0^{2\pi} \mathcal{G}\left(\frac{s}{\varepsilon}, x_\varepsilon(s)\right) u_0\left(s - \varepsilon\theta, \frac{s}{\varepsilon}\right) d\theta ds, \\ &= x_\varepsilon(0) + \frac{1}{2\pi} \left( \int_0^t \int_0^{2\pi} \mathcal{G}(\theta, x_\varepsilon(s)) u_0(s, \theta) d\theta ds - \Delta_\varepsilon \right) \end{aligned} \quad (17)$$

with  $\Delta_\varepsilon$  given by (13), that satisfies, since in particular  $\|u_0\| \leq 1$  and  $|\varepsilon\theta| < 2\pi\varepsilon$  and

$$\left\| \left( \mathcal{G}\left(\frac{s}{\varepsilon}, x_\varepsilon(s)\right) - \mathcal{G}\left(\frac{s}{\varepsilon}, x_\varepsilon(s + \varepsilon\theta)\right) \right) u_0\left(s + \varepsilon\theta, \frac{s}{\varepsilon}\right) \right\| \leq 2\pi (\text{Lip } \mathcal{G}) (\text{sup } \mathcal{G}) \varepsilon,$$

$\|\Delta_\varepsilon\| \leq 4\pi^2 (\text{Lip } \mathcal{G}) (1 + T \text{sup } \mathcal{G}) \varepsilon$ . Since, by definition,

$$x_0(t) = x_0(0) + \frac{1}{2\pi} \int_0^t \int_0^{2\pi} \mathcal{G}(\theta, x_0(s)) u_0(s, \theta) d\theta ds,$$

Using the above relations and (16), (17) and the above inequalities imply

$$\|x_\varepsilon(t) - x_0(t)\| \leq (b + 2\pi (\text{Lip } \mathcal{G}) (1 + T \text{sup } \mathcal{G})) \varepsilon + (\text{Lip } \mathcal{G}) \int_0^t \|x_\varepsilon(s) - x_0(s)\| ds \quad (18)$$

for all  $t$  in  $[0, T]$ , and finally, using the classical Gronwall lemma,

$$\|x_\varepsilon(t) - x(t)\| \leq [b + 2\pi (\text{Lip } \mathcal{G}) (1 + T \text{sup } \mathcal{G})] e^{T \text{Lip } \mathcal{G}} \varepsilon \quad (19)$$

for all  $t$  in  $[0, T]$  and  $\varepsilon$  in  $[0, \varepsilon_0]$ . This proves the theorem if  $M$  is an open subset of  $\mathbb{R}^n$  and  $\mathcal{G}$  is zero outside a compact subset, with an explicit constant  $c$  corresponding to the distance  $d$  defined from the Euclidean norm and with  $\varepsilon_0 = +\infty$ .

Now consider the general case without these specifications on  $M$  or  $\mathcal{G}$ . Define  $x_\varepsilon(\cdot)$  as the solution of (1) with control  $u = \bar{u}_\varepsilon(t)$  (defined in (14) from  $u_0$ ), with initial condition  $x_\varepsilon(0) = x_0(0)$ ; it is not necessarily defined on  $[0, T]$  but may have a maximum interval of definition  $[0, T_\varepsilon)$  with  $0 < T_\varepsilon < T$ . Let  $\tilde{T} \in [0, T]$  be the supremum of the set of numbers  $\tau \in [0, T]$  such that, for some  $\varepsilon_0$  and some  $c$  (that may depend on  $\tau$ ), the solution  $x_\varepsilon(\cdot)$  is defined on  $[0, \tau]$  and one has  $d(x_\varepsilon(t), x_0(0)) < c\varepsilon$  for all  $t \in [0, \tau]$  and  $\varepsilon \in [0, \varepsilon_0]$ .

Let us prove by contradiction that  $\tilde{T} = T$ . Indeed, if  $\tilde{T} < T$ , let  $\mathcal{O}$  be a coordinate neighborhood of  $x_0(\tilde{T})$ ; there is  $\alpha > 0$ , with  $0 \leq \tilde{T} - \alpha$  and  $T + \alpha \leq T$ , such that  $x_0([\tilde{T} - \alpha, \tilde{T} + \alpha]) \subset \mathcal{O}$ , and some  $c$  and  $\varepsilon_0$  such that  $d(x_\varepsilon(t), x_0(t)) < c\varepsilon$  for all  $t \in [0, \tilde{T} - \alpha]$  and  $\varepsilon \in [0, \varepsilon_0]$ ; taking  $\varepsilon_0$  possible smaller, one has  $x_\varepsilon(\tilde{T} - \alpha) \in \mathcal{O}$  for

$\varepsilon < \varepsilon_0$ ; let  $\mathbb{K} \subset \mathbb{K}' \subset \mathcal{O}$  be such that  $\mathbb{K}$  is a compact neighborhood of  $x_0([\tilde{T} - \alpha, \tilde{T} + \alpha])$  contained in the interior of  $\mathbb{K}'$ , itself compact, and let  $\rho : M \rightarrow [0, 1]$  be smooth, zero outside  $\mathbb{K}'$  and constant equal to one on  $\mathbb{K}$ ; define  $\mathcal{G}_\rho$  by  $\mathcal{G}_\rho(\theta, x) = \rho(x)\mathcal{G}(\theta, x)$ ; one may then apply step 1 in coordinates in  $\mathcal{O}$ , with  $\mathcal{G}_\rho$  instead of  $\mathcal{G}$ , on the interval  $[\tilde{T} - \alpha, \tilde{T} + \alpha]$  instead of  $[0, T]$  and with initial condition  $x_\varepsilon(\tilde{T} - \alpha)$  satisfying (16) with  $b$  deduced from  $c$  via Lipschitz equivalence of the distance  $d$  and the one induced by the norm in coordinates (see §2.2), getting some  $\varepsilon'_0 \leq \varepsilon_0$  and an inequality of the type (19), valid for  $t \in [\tilde{T} - \alpha, \tilde{T} + \alpha]$  and  $\varepsilon \in [0, \varepsilon'_0]$ ; hence  $x_\varepsilon([\tilde{T} - \alpha, \tilde{T} + \alpha]) \subset \mathbb{K}$  for some  $\varepsilon$  small enough and, possibly choosing a smaller  $\varepsilon'_0$ , the conclusion holds for  $\mathcal{G}$ , that coincide with  $\mathcal{G}_\rho$  in  $\mathbb{K}$ ; hence  $x_\varepsilon$  is defined on  $[0, \tilde{T} + \alpha]$  for all  $\varepsilon < \varepsilon'_0$  and satisfies  $d(x_\varepsilon(t) - x_0(t)) \leq c'\varepsilon$  with  $c' \geq c$  deduced from the explicit constant in (19) via Lipschitz equivalence of  $d$  and the Euclidean distance in coordinates. This contradicts the definition of  $\tilde{T}$  and ends the proof.  $\square$

*Proof of Theorem 3.4, point 2.* The sequence  $(x_n(\cdot))$  is equi-continuous because  $\mathcal{G}$  is bounded on  $S^1 \times \mathbb{K}$  (one may cover  $\mathbb{K}$  with a finite number of coordinate charts and define this bound in coordinates, yielding a common Lipschitz constant to all maps  $x_n(\cdot)$ ); hence it is compact from Ascoli-Arzelà Theorem: one may extract from any sub-sequence of  $(x_n(\cdot))$  an absolutely convergent sub-sequence.

Still denoting by  $(x_n(\cdot))_{n \in \mathbb{N}}$  such a converging sub-sequence and its (uniform) limit by  $x^*(\cdot)$ , define, for each  $n$ ,  $\hat{u}_n : [0, T] \times S^1 \rightarrow \mathbb{R}^m$  by

$$\hat{u}_n(t, \theta) = u_n(\beta_n(t, \theta)), \quad (20)$$

where  $u_n(\cdot) \in L^\infty([0, T], \mathbb{R}^m)$  is associated to  $x_n(\cdot)$  according to the assumption of the theorem and where the map  $\beta_n : [0, T] \times S^1 \rightarrow \mathbb{R}$  is the one such that

$$t - 2\pi\varepsilon_n < \beta_n(t, \theta) \leq t, \quad \frac{\beta_n(t, \theta)}{\varepsilon_n} \equiv \theta \text{ modulo } 2\pi. \quad (21)$$

Clearly  $\hat{u}_n$  is in  $L^\infty([0, T] \times S^1, \mathbb{R}^m)$  and  $\|\hat{u}_n\|_\infty \leq 1$ . Hence, after possibly extracting a sub-sequence,  $(\hat{u}_n)$  converges in the weak-\* topology to some  $\hat{u}^*$ . Let us prove that, for almost all  $t \in [0, T]$ ,

$$\dot{x}^*(t) = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{G}(\theta, x^*(t)) \hat{u}^*(t, \theta) d\theta. \quad (22)$$

Let  $\tilde{T} \in [0, T]$  be the supremum of the set of numbers  $\tau \in [0, T]$  such that this is true for almost all  $t$  in  $[0, \tau]$ , and let us prove by contradiction that  $\tilde{T} = T$ . Indeed, if  $\tilde{T} < T$ , let  $\mathcal{O}$  be a coordinate neighborhood of  $x_0(\tilde{T})$ ; there is  $\alpha > 0$  such that  $x_0([\tilde{T} - \alpha, \tilde{T} + \alpha]) \subset \mathcal{O}$ ; and from uniform convergence  $x_n([\tilde{T} - \alpha, \tilde{T} + \alpha]) \subset \mathcal{O}$  for  $n$  large enough; thus one has, in coordinates, for  $t \in [\tilde{T} - \alpha, \tilde{T} + \alpha]$ ,

$$x_n(t) - x_n(\tilde{T} - \alpha) = \int_{\tilde{T} - \alpha}^t \mathcal{G}\left(\frac{s}{\varepsilon}, x_n(s)\right) u_n(s) ds. \quad (23)$$

From (21), one has  $\beta_n(s - \varepsilon\theta, \frac{s}{\varepsilon}) = s$ , and hence, from (20),  $\hat{u}_n(s - \varepsilon\theta, \frac{s}{\varepsilon}) = u_n(s)$  for all  $\theta \in S^1$ ,  $s \in \mathbb{R}$ . Then, from Lemma 3.6, one has

$$\begin{aligned} \frac{1}{2\pi} \iint_{\substack{\tilde{T} - \alpha \leq s \leq t \\ 0 \leq \theta \leq 2\pi}} \mathcal{G}(\theta, x_n(s)) \hat{u}_n(\theta, s) d\theta ds &= \frac{1}{2\pi} \iint_{\substack{\tilde{T} - \alpha \leq s \leq t \\ 0 \leq \theta \leq 2\pi}} \mathcal{G}\left(\frac{s}{\varepsilon}, x_n(s)\right) u_n(s) d\theta ds \\ &= \int_{\tilde{T} - \alpha}^t \mathcal{G}\left(\frac{s}{\varepsilon}, x_n(s)\right) u_n(s) ds \end{aligned}$$

Substituting this in the right-hand side of (23), and using uniform convergence of  $x_n$  to  $x^*$  and weak convergence of  $\widehat{u}_n$  to  $\widehat{u}^*$ , one gets

$$x^*(t) - x^*(\widetilde{T} - \alpha) = \frac{1}{2\pi} \int_{\widetilde{T} - \alpha}^t \int_0^{2\pi} \mathcal{G}(\theta, x^*(s)) \widehat{u}^*(\theta, s) \, d\theta \, ds ,$$

for  $t$  in  $[\widetilde{T} - \alpha, \widetilde{T} + \alpha]$ , and finally that (22) hold for almost all  $t$  in  $[0, \widetilde{T} + \alpha]$ , thus contradicting the definition of  $\widetilde{T}$ .  $\square$

**3.4. Properties of the average system.** Let us further discuss the subset  $\mathcal{E}(x)$  of  $T_x M$  defined by (4), present in the right-hand side of the average system (5).

PROPOSITION 3.7. *For all  $x$  in  $M$ ,  $\mathcal{E}(x)$  is convex, compact and symmetric with respect to the origin.*

*Proof.* It is closed, convex and symmetric because it is the image of the unit ball on  $L^\infty(S^1, \mathbb{R}^m)$  by a linear map; it is compact because  $\mathcal{G}(x, \cdot)$  is bounded on  $S^1$ .  $\square$

If  $\mathcal{E}(x)$  had a nonempty interior, this would make it the unit ball for some norm on  $T_x M$ ; we shall see later that its dual norm on  $T_x^* M$  would be  $p \mapsto H(x, p)$  with  $H : T_x^* M \rightarrow [0, +\infty)$  defined by

$$H(x, p) = \frac{1}{2\pi} \int_0^{2\pi} \|\langle p, \mathcal{G}(\theta, x) \rangle\| \, d\theta . \quad (24)$$

where  $\langle p, \mathcal{G}(\theta, x) \rangle \in (\mathbb{R}^m)^*$  with  $\|\langle p, \mathcal{G}(\theta, x) \rangle\| = \left( \sum_{i=1}^m \langle p, \mathcal{G}_i(\theta, x) \rangle^2 \right)^{1/2}$ . Let us give, for now, a characterization of  $\mathcal{E}$  in terms of  $H$ , valid whether  $\mathcal{E}(x)$  has an empty interior or not.

PROPOSITION 3.8. *For all  $(x, p) \in T_x^* M$ , one has*

$$\mathcal{E}(x) = \left\{ v \in T_x M, \sup_{\substack{p \in T_x^* M \\ H(x, p) \leq 1}} \langle p, v \rangle \leq 1 \right\}, \quad (25)$$

$$H(x, p) = \sup_{v \in \mathcal{E}(x)} \langle p, v \rangle = \sup_{u \in L^\infty(S^1, \mathbb{R}^m), \|u\|_\infty \leq 1} \langle p, \overline{\mathcal{G}}(x, u) \rangle = \langle p, \overline{\mathcal{G}}(x, \mathcal{U}_{p,x}^*) \rangle, \quad (26)$$

with  $\mathcal{U}_{p,x}^* \in L^\infty(S^1, \mathbb{R}^m)$  is defined by (the transposition “ $\top$ ” takes place in  $\mathbb{R}^m$ ):

$$\mathcal{U}_{p,x}^*(\theta) = \begin{cases} 0 & \text{if } \langle p, \mathcal{G}(\theta, x) \rangle = 0, \\ \frac{\langle p, \mathcal{G}(\theta, x) \rangle^\top}{\|\langle p, \mathcal{G}(\theta, x) \rangle\|} & \text{if } \langle p, \mathcal{G}(\theta, x) \rangle \neq 0. \end{cases} \quad (27)$$

*Proof.* The last equality in (26) is a straightforward maximization, the second one comes from the definition (4) of  $\mathcal{E}(x)$  and a simple computation yields  $H(x, p) = \langle p, \overline{\mathcal{G}}(x, \mathcal{U}_{p,x}^*) \rangle$  is; this proves (26). Being closed and convex,  $\mathcal{E}(x)$  is the intersection of all its supporting half-spaces [24, Corollary 1.3.5]; according to (26), this yields the following relation, equivalent to (25):

$$\mathcal{E}(x) = \bigcap_{p \in T_x^* M} \{v \in T_x M, \langle p, v \rangle \leq H(x, p)\} . \quad \square$$

Let us now come back to the question whether  $\mathcal{E}(x)$  has a non-empty interior in  $T_x M$ ; more precisely, let us characterize its dimension.

Recall that, for a convex subset  $C$  of an affine space, its affine hull is the smallest affine subspace that contains  $C$ , its “relative interior” (interior of  $C$  as a subset of its affine hull) is always nonempty,  $\dim C$  is the dimension of this affine hull. In the case of  $\mathcal{E}(x)$ , containing the origin in a linear space, one may replace “affine” with “linear”; let us now characterize  $\dim \mathcal{E}(x)$ .

PROPOSITION 3.9.

1. *The linear hull of  $\mathcal{E}(x)$  satisfies the following two properties for all  $x$  in  $M$ , where the inclusion (29) is an equality if  $\mathcal{G}$  is real analytic:*

$$\text{Linear hull } \mathcal{E}(x) = \sum_{\theta \in S^1} \text{Range } \mathcal{G}(\theta, x), \quad (28)$$

$$\text{Linear hull } \mathcal{E}(x) \supset \sum_{j \in \mathbb{N}} \text{Range } \frac{\partial^j \mathcal{G}}{\partial \theta^j}(\theta, x) \quad \text{for all } \theta \in S^1. \quad (29)$$

2. *Define the integer  $r(\theta, x)$  by:*

$$r(\theta, x) = \text{rank} \left\{ \frac{\partial^j \mathcal{G}}{\partial \theta^j}(\theta, x), j \in \mathbb{N} \right\} = \dim \left( \sum_{j \in \mathbb{N}} \text{Range } \frac{\partial^j \mathcal{G}}{\partial \theta^j}(\theta, x) \right), \quad (30)$$

where the second equality is just an explanation of the notation  $\text{rank}\{.\}$ , the sum is a sum of linear subspaces of  $T_x M$  and, for a linear map  $A : E \rightarrow F$ ,  $\text{Range } A$  denotes its range  $\{Ax, x \in E\} \subset F$ .

One then has:

- if  $r(\theta, x) = n$  for at least one  $\theta$  in  $S^1$ , then  $\mathcal{E}(x)$  has a nonempty interior in  $T_x M$ , i.e.  $\dim \mathcal{E}(x) = n$ ,
- if the system (1) is real analytic, i.e. if the vector fields  $\mathcal{G}_i$  are real analytic with respect to  $x$  and  $\theta$ , then  $r(\theta, x)$  does not depend on  $\theta$  and  $r(\theta, x) = \dim \mathcal{E}(x)$ .

*Proof.* Point 2 is an easy consequence of 1, that we now prove. To prove the identity (28) between subspaces of  $T_x M$ , it is sufficient to prove the following one on their annihilators in  $T_x^* M$ :

$$\mathcal{E}(x)^\perp = \bigcap_{\theta \in S^1} (\text{Range } \mathcal{G}(\theta, x))^\perp.$$

If  $p$  is in  $(\text{Range } \mathcal{G}(\theta, x))^\perp$  for all  $\theta$ , then any  $v = \overline{\mathcal{G}}(x, \mathcal{U})$  in  $\mathcal{E}(x)$  satisfies  $\langle p, v \rangle = 0$  because  $\langle p, \mathcal{G}(\theta, x) \mathcal{U}(\theta) \rangle$  is identically zero on  $[0, 2\pi]$ . Conversely, let  $p$  be in  $\mathcal{E}(x)^\perp$ , and consider  $v = \overline{\mathcal{G}}(x, \mathcal{U}_{p,x}^*) \in \mathcal{E}(x)$ ; then  $\langle p, v \rangle = 0$  implies  $\langle p, \mathcal{G}(\theta, x) \rangle = 0$ , i.e.  $p \in (\text{Range } \mathcal{G}(\theta, x))^\perp$  for all  $\theta$ .

Similarly, and using (28), the following inclusion implies (29):

$$\bigcap_{\phi \in S^1} (\text{Range } \mathcal{G}(x, \phi))^\perp \subset \bigcap_{j \in \mathbb{N}} \left( \text{Range } \frac{\partial^j \mathcal{G}}{\partial \theta^j}(\theta, x) \right)^\perp.$$

If, for all  $\phi$ , one has  $p \in (\text{Range } \mathcal{G}(x, \phi))^\perp$ , i.e.  $p \mathcal{G}(x, \phi) = 0$ , then differentiating with respect to  $\phi$  an arbitrary number of times, one gets  $p \partial^j \mathcal{G} / \partial \phi^j(x, \phi) = 0$ ,  $j \in \mathbb{N}$ ; true in particular for  $\phi = \theta$ ; this proves the above inclusion, hence (29). To prove the reverse inclusion in the real analytic case, fix  $\theta \in S^1$  and  $p \in \bigcap_{j \in \mathbb{N}} \left( \frac{\partial^j \mathcal{G}}{\partial \theta^j}(\theta, x) \right)^\perp$ , and consider the real analytic mapping  $S^1 \rightarrow (\mathbb{R}^m)^*$  defined as  $\phi \mapsto p \mathcal{G}(x, \phi)$ ; the assumption on  $p$  implies that this map, as well as its derivatives at all orders vanish for  $\phi = \theta$ ; it must therefore be identically zero and this implies  $p \in \bigcap_{\phi \in S^1} (\text{Range } \mathcal{G}(x, \phi))^\perp$ .  $\square$

**3.5. Further properties in the full rank case.** We now assume that the mapping  $\mathcal{G}$  in (1) is such that the rank  $r(\theta, x)$  defined by (30) is maximal:

$$r(\theta, x) = n \quad \text{for all } x \text{ in } M \text{ and all } \theta \text{ in } \mathbb{R} \quad (31)$$

(or all  $\theta$  in  $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ , taking periodicity into account).

This condition is strongly related to controllability of the linear approximation of (1) around equilibria, i.e. around  $x$  constant and  $u$  identically zero. Indeed, the linear approximation of the time-varying nonlinear system (take  $\varepsilon = 1$  in (1)):

$$\dot{x} = \mathcal{G}(t, x)u \quad (32)$$

around the equilibrium  $x = x_1$  is the time-varying linear system  $\dot{\xi} = \mathcal{G}(t, x_1)u$ ; according to [16, p.614] or [25], it is ‘‘controllable with impulsive controls at any time’’ if and only if  $r(t, x_1) = n$  for all  $t$ . If this is true at all points  $x_1$  then all end-point mappings are submersions around zero controls; we shall need the following more precise result:

PROPOSITION 3.10. *Assume that (31) holds.*

1. *For all  $x_1 \in M$  and  $T > 0$ , there exist a coordinate neighborhood  $\mathcal{W}$  of  $x_1$  (the ball  $\mathcal{B}$  below refers to the Euclidean norm in these coordinates), positive constants  $\alpha_0, c_3$ , and, for all  $y \in \mathcal{W}$ , a smooth map  $\chi_y : \mathcal{B}(y, \alpha_0) \rightarrow L^\infty([0, T], \mathbb{R}^m)$  with Lipschitz constant  $c_3$ , which is a right inverse of the end-point mapping of (32) on  $[0, T]$  starting from  $y$ , i.e. for all  $y_f \in \mathcal{B}(y, \alpha_0)$ , the control  $\chi_y(y_f) : [0, T] \rightarrow \mathbb{R}^m$  is such that the solution of  $\dot{x} = \mathcal{G}(t, x)\chi_y(y_f)(t)$ ,  $x(0) = y$  satisfies  $x(T) = y_f$ .*

2. *For all  $\varepsilon > 0$ , the system (1) is fully controllable, i.e. there exists, for any  $\varepsilon > 0$  and any two point  $x_0, x_1$  in  $M$ , a time  $T$  and a measurable control  $u : [0, T] \rightarrow B^m$  such that the solution of (1) with  $x(0) = x_0$  satisfies  $x(T) = x_1$ .*

*Proof.* Let  $E_y : L^\infty([0, T], \mathbb{R}^m) \rightarrow M$  be the end-point mapping with starting point  $y$ . Condition (31) implies that the derivative of  $E_{x_1}$  at the zero control has rank  $n$ ; hence there exists an  $n$ -dimensional subspace  $V$  of  $L^\infty([0, T], \mathbb{R}^m)$  such that the restriction of  $E_{x_1}$ , and hence of  $E_y$  for  $y$  close enough, to  $V$  is a local diffeomorphism at zero; the  $\chi_y$ 's are the local inverses of these local diffeomorphisms; they depend smoothly on  $y$ , hence the common  $\alpha_0$  and  $c_3$  in Point 1.

This implies that the reachable set from any point at any positive or negative time contains a neighborhood of this point; a classical argument then tells us that the reachable set from a point  $x_0$  is  $M$ , assumed to be connected, for it is both open (obvious) and closed (if  $\bar{x}$  is in the closure of the reachable set, some points in the reachable set can be reached in negative time, hence  $\bar{x}$  can be reached from  $x_0$ ).  $\square$

Let us now prove further results on  $H : T^*M \rightarrow [0, +\infty)$  defined by (24).

PROPOSITION 3.11. *Under the rank condition (31),  $H(x, \cdot)$  defines a norm on  $T_x^*M$  for all  $x \in M$ , and the function  $(x, p) \mapsto H(x, p)^2$  is continuously differentiable.*

Let then  $N : TM \rightarrow [0, +\infty)$  be such that the dual norm of  $H(x, \cdot)$  is  $N(x, \cdot)$ :

$$N(x, v) = \max_{p \in T_x^*M, H(x, p) \leq 1} \langle p, v \rangle. \quad (33)$$

PROPOSITION 3.12. *Assume the rank condition (31).*

1. *The set  $\mathcal{E}(x)$  is the unit ball on the tangent space  $T_xM$  for the norm  $N(x, \cdot)$ :*

$$\mathcal{E}(x) = \{v, N(x, v) \leq 1\}. \quad (34)$$

2. *System (5) is fully controllable, i.e. there exists, for any points  $x_0, x_1$  in  $M$ , a time  $T$  and a solution  $x(\cdot) : [0, T] \rightarrow M$  of (5) such that  $x(0) = x_0$ ,  $x(T) = x_1$ .*

*Proof of Propositions 3.11 and 3.12.* From (24),  $H(x, p) = 0$  implies  $\langle p, \mathcal{G}(\theta, x) \rangle = 0$  for all  $\theta$  and, differentiating with respect to  $\theta$  and using (31), this implies  $p = 0$ ; this makes  $p \mapsto H(x, p)$  a norm for other properties are straightforward. Postponing differentiability of  $H$ , turn to Proposition 3.12: Point 1 is an obvious consequence of (25) in Proposition 3.8; to prove Point 2, take a continuously differentiable curve  $\gamma : [0, 1] \rightarrow M$  such that  $\gamma(0) = x_0$  and  $\gamma(1) = x_1$  and  $\sigma : [0, T] \rightarrow [0, 1]$  for some  $T > 0$ , differentiable, such that

$$t \geq \int_0^{\sigma(t)} N(\gamma(s), \frac{d\gamma}{ds}(s)) ds$$

( $N$  and  $H$  are obviously continuous), then  $t \mapsto x(t) = \gamma(\sigma(t))$  is a solution of (5) such that  $x(0) = x_0$  and  $x(T) = x_1$ .

To prove continuous differentiability of  $H^2$ , we operate in coordinates for this is anyway a local property. Apply Proposition A.1 with  $d = 2n$ ,  $X = (x, p) \in \mathbb{R}^{2n}$ ,  $V(\theta, X) = \langle p, \mathcal{G}(\theta, x) \rangle$  and  $O^d$  a neighborhood of a point where  $p \neq 0$ : the rank condition implies that derivatives of all order of the smooth map  $\theta \mapsto V(\theta, X)$  never vanish at the same point, so that its zeroes are isolated and the set  $\{\theta \in S^1, V(\theta, X) = 0\}$  is finite and *a fortiori* has measure zero; hence  $H$  is continuously differentiable outside  $\{p = 0\}$ . Since  $0 \leq H(x, p) \leq k\|p\|$  for some local constant  $k$  the derivative of  $H^2$  is zero at all points  $(x, 0)$  and, since (65) implies that the norm of  $dH(x, p)$  at neighboring points where  $p \neq 0$  is bounded, the derivative of  $H^2$  at these points tends to zero as  $p \rightarrow 0$ .  $H^2$  is therefore continuously differentiable everywhere.  $\square$

*Remark 3.13 (Finsler geometry).* A (reversible) Finsler metric [3] on  $M$  is the data of a map  $F : TM \rightarrow \mathbb{R}$ , such that  $\sqrt{F(x, \cdot)}$  defines, for each  $x$  in  $M$ , a norm on  $T_x M$  (reversible refers to the property  $F(x, v) = F(x, -v)$ , not assumed a priori). In all classical literature,  $F$  is assumed to be smooth and  $F(x, \cdot)$  to have a positive definite Hessian (needing at least two derivatives everywhere).

In the present case, defining  $F$  by  $F(x, v) = N(x, v)^2$  does not meet the smoothness requirements ( $C^2$  is not granted) and classical invariants involving many derivatives will not be available. One may however compute the length of a curve, and speak of geodesics. Minimizing trajectories for the time optimal problem for the differential inclusion (5), discussed in §3.6 below, coincide with geodesics. Their computation can be done using the Hamiltonian point of view, i.e. Pontryagin maximum principle, yielding a differential equation (Hamiltonian system) on the cotangent bundle, instead of the tangent bundle like the Lagrangian point of view. Because of the lack of smoothness of the metric (or of  $H$ ), this differential equation does not have a smooth right-hand side, and is even not locally Lipschitz; we shall prove however that, at least in the less degenerate cases, this Hamiltonian system gives raise to a flow.

**3.6. Application to the minimum time problem.** Fix two points  $x_0, x_1$  in  $M$  and consider the time optimal problem associated to (1) for  $\varepsilon > 0$ :

$$(\mathcal{P}_\varepsilon) : \quad \dot{x} = \mathcal{G}(t/\varepsilon, x) u, \quad u \in B^m, \quad x(0) = x_0, \quad x(T) = x_1, \quad \min T, \quad (35)$$

and the time optimal problem associated to the average system:

$$(\mathcal{P}_0) : \quad x(0) = x_0, \quad x(T) = x_1, \quad \dot{x} \in \mathcal{E}(x), \quad \min T. \quad (36)$$

Call  $T_\varepsilon(x_0, x_1)$  the minimum time in  $(\mathcal{P}_\varepsilon)$ ,  $\varepsilon > 0$  and  $T_0(x_0, x_1)$  the minimum time in  $(\mathcal{P}_0)$ . When no confusion arises, we write simply  $T_\varepsilon$  and  $T_0$ .

Controllability – Proposition 3.10, point 2 for  $\varepsilon > 0$  and Proposition 3.12 point 2 for  $(\mathcal{P}_0)$ – implies that these times are finite. However, without further properties there need not be a “solution” to each of these problems, i.e. an actual control and trajectory that realize the minimum.

**THEOREM 3.14.** *Assume that the rank condition (31) holds.*

1.  $T_\varepsilon$  is bounded as  $\varepsilon \rightarrow 0$  and  $\limsup_{\varepsilon \rightarrow 0} T_\varepsilon \leq T_0$ .
2. If, for  $\varepsilon > 0$  small enough, each  $(\mathcal{P}_\varepsilon)$  has a solution  $x_\varepsilon : [0, T_\varepsilon] \rightarrow M$  and there exists a compact  $\mathbb{K} \subset M$  such that  $x_\varepsilon([0, T_\varepsilon]) \subset \mathbb{K}$  for all  $\varepsilon > 0$  small enough, then all accumulation points (as  $\varepsilon \rightarrow 0$ ) of the compact family  $(x_\varepsilon(\cdot))_{\varepsilon > 0}$  in  $C^0([0, T_0], M)$  are solutions of  $(\mathcal{P}_0)$  and  $\lim_{\varepsilon \rightarrow 0} T_\varepsilon = T_0$ .

*Proof.* Consider a minimizing sequence for problem  $(\mathcal{P}_0)$ , i.e. solutions  $x^k : [0, T_0 + \beta_k] \rightarrow M$  of the average system (5) with  $(\beta_k)$  a sequence of positive numbers that tends to zero and  $x^k(0) = x_0$ ,  $x^k(T_0 + \beta_k) = x_1$  for all  $k$ . For each  $x^k$ , there is, according to Theorem 3.4, a family  $(x_\varepsilon^k(\cdot))_{\varepsilon > 0}$  such that each  $x_\varepsilon^k(\cdot)$  is a solution of (1) with  $x_\varepsilon^k(0) = x_0$  and  $d(x_\varepsilon^k(t), x^k(t)) < c_1\varepsilon$  for all  $t$  in  $[0, T_0 + \beta_k]$ . In particular  $d(x_\varepsilon^k(T_0 + \beta_k), x^k(T_0 + \beta_k)) \leq c_1\varepsilon$ . Now, from Lemma 3.15 stated below,  $T_\varepsilon(x_\varepsilon^k(T_0 + \beta_k), x_1) \leq (2\pi + c_1C_3)\varepsilon$ ; hence, from the second point of that lemma (with  $x'_1 = x_\varepsilon^k(T_0 + \beta_k)$ ), one has  $T_\varepsilon = T_\varepsilon(x_0, x_1) \leq T_0 + \beta_k + (4\pi + c_1C_3)\varepsilon$  and, letting  $k$  go to infinity,  $T_\varepsilon \leq T_0 + (4\pi + c_1C_3)\varepsilon$ ; this implies Point 1.

In Point 2, first prolong  $x_\varepsilon$  on  $[0, \overline{T}]$ , with  $\overline{T}$  an upperbound of  $T_\varepsilon$ , by taking  $x_\varepsilon(t) = x_1$  for all  $t \in [T_\varepsilon, \overline{T}]$ ; any sequence  $(x_{\varepsilon_k}(\cdot))_{k \in \mathbb{N}}$  with  $\lim \varepsilon_k = 0$  is compact in  $C^0([0, \overline{T}], M)$ ; take a convergent subsequence such that  $T_{\varepsilon_k}$  also converges to some  $T^*$ ; from Theorem 3.4, the uniform limit is a solution of the average system (5), and it goes through  $x_0$  at time 0 and  $x_1$  at time  $T^*$ , hence  $T^* \geq T_0$  by definition of  $T_0$ . Since  $T^*$  can be any accumulation point of  $(T_\varepsilon)$  as  $\varepsilon \rightarrow 0$ , this, together with Point 1 ends the proof of Point 2.  $\square$

**LEMMA 3.15.** 1. *There is a neighborhood  $\mathcal{W}$  of any  $x_1$  and two constants  $\alpha_0 > 0$  and  $C_3 > 0$  such that, for all  $y$  in  $\mathcal{W}$ ,  $T_\varepsilon(y, x_1) \leq 2\pi\varepsilon + C_3d(x_1, y)$ .*

2. *For any  $x_0, x'_1, x_1$  in  $M$ , one has  $T_\varepsilon(x_0, x_1) \leq T_\varepsilon(x_0, x'_1) + T_\varepsilon(x'_1, x_1) + 2\pi\varepsilon$ .*

*Proof.* Apply Proposition 3.10, point 1 with  $T = 2\pi$ , using as a distance in  $\mathcal{W}$  the Euclidean norm in some coordinates: for any two points  $y, y'$  in  $\mathcal{W}$  such that  $\|y - y'\| \leq \alpha_0$ , there is a control defined on  $[0, 2\pi]$ , with  $L^\infty$  norm smaller than  $c_3\|y - y'\|$  that brings  $y'$  at time 0 to  $y$  at time  $2\pi$  for system (32); rescaling time and control by  $\varepsilon$  yields, if  $c_3\|y - y'\| \leq \varepsilon$ , a control with  $L^\infty$  norm less than 1 that brings  $y'$  at time 0 to  $y$  at time  $2\pi\varepsilon$  for system (1) and hence, by concatenating controls and using periodicity of  $\mathcal{G}$ , for any positive integer  $k$ , a control with  $L^\infty$  norm less than 1 that brings  $y'$  at time 0 to  $y$  at time  $2k\pi\varepsilon$  for system (1) if  $c_3\|y - y'\| \leq k\varepsilon$ . In other words,  $T_\varepsilon(y', y) \leq 2\pi(\varepsilon + c_3\|y - y'\|)$ . Take  $y' = x_1$  and  $2\pi c_3/C_3$  the ratio between the Euclidean norm and the distance  $d$ ; this proves point 1. Point 2 follows from using periodicity of  $\mathcal{G}$  and concatenating controls while inserting a zero control between time  $T_\varepsilon(y', y)$  and the next multiple of  $2\pi$ .  $\square$

Let us now write the Pontryagin Maximum Principle (PMP) [22] both for  $(\mathcal{P}_\varepsilon)$ ,  $\varepsilon > 0$  and for  $(\mathcal{P}_0)$  and see how they are related.

**THEOREM 3.16.** *Problems  $(\mathcal{P}_\varepsilon)$  and  $(\mathcal{P}_0)$  are described in (35) and (36).*

*If an absolutely continuous  $t \mapsto \bar{x}(t)$  defined on  $[0, \overline{T}]$  is a solution of  $(\mathcal{P}_\varepsilon)$ ,  $\varepsilon > 0$ , (resp. of  $(\mathcal{P}_0)$ ), then there exists  $t \mapsto \bar{p}(t)$  defined on  $[0, \overline{T}]$  such that  $t \mapsto (\bar{p}(t), \bar{x}(t))$  is an extremal of  $(\mathcal{P}_\varepsilon)$ ,  $\varepsilon > 0$  (resp. of  $(\mathcal{P}_0)$ ).*

*The extremals of problem  $(\mathcal{P}_\varepsilon)$ ,  $\varepsilon > 0$ , are absolutely continuous  $t \mapsto (x(t), p(t))$*

solution, at least at times when  $\langle p, \mathcal{G}(t/\varepsilon, x) \rangle \neq 0$ , of

$$\dot{p} = -\frac{\partial H_\varepsilon}{\partial x}, \quad \dot{x} = \frac{\partial H_\varepsilon}{\partial p} \quad \text{with} \quad H_\varepsilon(t, p, x) = \|\langle p, \mathcal{G}(t/\varepsilon, x) \rangle\|, \quad (37)$$

whose right-hand side is discontinuous on  $\mathcal{S}_\varepsilon = \{(x, p, t), \langle p, \mathcal{G}(t/\varepsilon, x) \rangle = 0\}$  (the “switching surface”), where it is in fact not defined.

The extremals of  $(\mathcal{P}_0)$  are absolutely continuous  $t \mapsto (x(t), p(t))$  solution of

$$\dot{p} = -\frac{\partial H}{\partial x}, \quad \dot{x} = \frac{\partial H}{\partial p}. \quad (38)$$

with  $H$  given by (24). This differential equation has a continuous right-hand side.

*Remark 3.17.* This result and Theorem 3.14 have two interpretations:

1. They prove that the operations of *averaging* and *computing the Hamiltonian for the minimum time problem* commute. Indeed, the Hamiltonian  $H$  was obtained by applying the maximum principle to problem (36), i.e. minimum time for the average system (5), but it also the average of the one in (37) with respect to the fast variable.
2. They prove indirectly an averaging result for the minimum time control problem (36); the averaging techniques in [10] do not apply to minimum time for they require strict convexity of the Hamiltonian, while averaging is used in [14, 13] for minimum time with only partial theoretical justifications but numerical evidence of efficiency.

*Proof.* Problem  $(\mathcal{P}_\varepsilon)$ ,  $\varepsilon > 0$  deals with a classical smooth control system; according to [22, 1], the pseudo-Hamiltonian is  $h(t, x, p, u) = \langle p, \mathcal{G}(t/\varepsilon, x) u \rangle$ ; an extremal is a curve on the co-tangent bundle solution of the following dynamics in local coordinate chart:

$$\begin{aligned} \dot{p} &= -\frac{\partial h}{\partial x}(t, x, p, u^*) = -\langle p, \frac{\partial \mathcal{G}}{\partial x} u^* \rangle, \\ \dot{x} &= \frac{\partial h}{\partial p}(t, x, p, u^*) = \mathcal{G} u^*, \end{aligned} \quad (39)$$

with  $u^*(t)$  a control that maximizes the pseudo-Hamiltonian for almost all time; it is defined by  $u^* = \frac{\langle p, \mathcal{G} \rangle}{\|\langle p, \mathcal{G} \rangle\|}$  if  $\langle p, \mathcal{G}(t/\varepsilon, x) \rangle \neq 0$ ; the maximized Hamiltonian  $H_\varepsilon(t, p, x) = \max_u h(t, x, p, u)$  is the one in (37) and (39) is then the differential equation (37), whose right-hand side is discontinuous at points where  $\langle p, \mathcal{G}(t/\varepsilon, x) \rangle$  vanishes.

Let us now turn to the average system, i.e. to the minimisation problem  $(\mathcal{P}_0)$ . Since the set of admissible velocities is not priori smooth with respect to the state variable we have to use a non-smooth version of the Pontryagin maximum principle for differential inclusions<sup>2</sup>, that we recall for self-containedness:

*Theorem 9.1 in [11, Chapter 4]:* if  $\dot{x} \in \mathcal{E}(x)$  is a locally Lipschitz differential inclusion and  $t \mapsto \bar{x}(t)$  is an absolutely continuous function defined on  $[0, \bar{T}]$  solution of the problem (36), then there exists  $t \mapsto \bar{p}(t)$  defined on  $[0, \bar{T}]$  such that  $(-\bar{p}, \dot{\bar{x}}) \in \partial_C H(\bar{x}, \bar{p})$  for almost all  $t \in [0, \bar{T}]$  with  $H(x, p) = \max_{v \in \mathcal{E}(x)} \langle p, v \rangle$  and  $\partial_C H$  the generalized gradient of  $H$ .

The set-valued map  $\mathcal{E}(\cdot)$  in (4) is indeed locally Lipschitz: in local coordinates, for  $x_1, x_2$  in  $\mathbb{R}^n$ , denoting by  $\delta$  the Hausdorff distance between two sets, one has:

$$\begin{aligned} \delta(\mathcal{E}(x_1), \mathcal{E}(x_2)) &= \max \left\{ \sup_{v_1 \in \mathcal{E}(x_1)} \inf_{v_2 \in \mathcal{E}(x_2)} \|v_1 - v_2\|, \sup_{v_2 \in \mathcal{E}(x_2)} \inf_{v_1 \in \mathcal{E}(x_1)} \|v_1 - v_2\| \right\} \\ &= \max_{\|u_1\|_\infty \leq 1} \min_{\|u_2\|_\infty \leq 1} \|\bar{\mathcal{G}}(x_1, u_1) - \bar{\mathcal{G}}(x_2, u_2)\| \leq \text{Lip } \mathcal{G} \|x_1 - x_2\|. \end{aligned}$$

<sup>2</sup>If we had used, instead of the differential inclusion (5), the representation (10) of the average system as a smooth control system with controls of infinite dimension, we could use a maximum principle for this sort of systems, see Remark 3.3

According to (26), the map  $H$  defined in (24), continuously differentiable according to Proposition 3.11, coincides with the Hamiltonian  $H$  defined in the above quoted theorem, that therefore implies the second part of Theorem 3.16.  $\square$

Let us now focus on  $(\mathcal{P}_0)$ , and in particular on the differential equation (38) that governs its extremals. Its right-hand side is continuous because, from Proposition 3.11,  $H$  is continuously differentiable. This is not sufficient to imply uniqueness of solutions to the Cauchy problem, and existence of a Hamiltonian flow on  $T^*M$  for (38), which is of great importance when describing optimal solutions. Of course, from (24),  $H$  would be smooth as  $\mathcal{G}$ —hence infinitely differentiable— if the vector  $\langle p, \mathcal{G}(\theta, x) \rangle$  would not vanish; more precisely,  $H$  is smooth ( $C^\infty$ ) on  $T^*M \setminus \tilde{\mathcal{Z}}$  with

$$\tilde{\mathcal{Z}} = \{(x, p) \in T^*M, \exists \theta \in S^1, \langle p, \mathcal{G}(\theta, x) \rangle = 0\}. \quad (40)$$

Unfortunately,  $\tilde{\mathcal{Z}}$  is not empty: it is, generically and away from the zero section  $\{p = 0\}$  of  $T^*M$ , a  $2n - m + 1$  dimensional submanifold of  $T^*M$ . Let us give a result on the right-hand side of (38) in the less degenerated case possible; it does not give Local Lipschitz continuity of the right-hand side of (38), but a module of continuity which is slightly bigger than linear, but sufficient to ensure uniqueness of solutions:

**THEOREM 3.18** (Hamiltonian flow for the average minimum time problem). *Assume that the rank condition (31) holds and that*

(i) *for  $(x, p) \in T^*M$ ,  $p \neq 0$ , there is at most one  $\theta \in S^1$  such that  $\langle p, \mathcal{G}(\theta, x) \rangle = 0$ , and  $\langle p, \frac{\partial \mathcal{G}}{\partial \theta}(\theta, x) \rangle$  does not vanish at the same point,*

(ii) *for all  $(\theta, x) \in S^1 \times M$ , one has  $\text{rank } \mathcal{G}(\theta, x) = m$ ,*

*then the Cauchy problem (38) has unique solutions for any initial condition.*

*Indeed the right-hand side is smooth outside  $\tilde{\mathcal{Z}}$  and any point  $(\bar{x}, \bar{p})$  in  $\tilde{\mathcal{Z}}$  has a constant  $c$  and a coordinate neighborhood in  $T^*M$  such that for all  $X$  and  $Y$  in  $\mathbb{R}^{2n}$ , coordinates of points in the neighborhood,*

$$\|dH(Y) - dH(X)\| \leq c \|X - Y\| \left(1 + \left|\ln \frac{1}{\|X - Y\|}\right|\right). \quad (41)$$

*Proof.* Smoothness outside  $\tilde{\mathcal{Z}}$  is obvious from the expression (24) of  $H$ ; inequality (41) is a consequence of Proposition A.2 in the appendix, applied with  $d = 2n$ ,  $X = (x, p) \in \mathbb{R}^{2n}$ ,  $V(\theta, X) = \langle p, \mathcal{G}(\theta, x) \rangle$  and  $O^d$  a neighborhood of a point of  $\tilde{\mathcal{Z}} \setminus \{p = 0\}$ ; it is clear that points (i) and (ii) imply assumptions A1, A2 and A3.

Existence and uniqueness of solutions of the Cauchy problem follows from the Kamke uniqueness Theorem [15, chapter III, Theorem 6.1]; it is stated in [15] for time-dependent ODEs, but for an autonomous ODE  $\dot{z} = f(z)$  in a finite dimensional linear space, with  $f$  satisfying  $\|f(z_1) - f(z_2)\| \leq \omega(\|z_1 - z_2\|)$  with  $\omega$  a non-decreasing function  $[0, +\infty) \rightarrow [0, +\infty)$ , it states that uniqueness of solutions holds if  $\int_0^\alpha \frac{du}{\omega(u)} = +\infty$  for arbitrarily small  $\alpha > 0$ . We apply this with  $\omega(u) = u \ln(1/u)$  (and  $\int du/\omega(u) = -\ln(\ln(1/u))$ ).  $\square$

The sufficient conditions given by this theorem are certainly not the best possible, but they turn to be applicable to the control of orbit transfer with low thrust, see §5.

Point (ii) is very mild and only states that the control vector fields are linearly independent. Point (i) is more artificial: the fact that  $\langle p, \mathcal{G}(\theta, x) \rangle = 0$  has at most one solution  $\theta$  has to be checked by hand, while the fact that  $\partial \mathcal{G} / \partial \theta$  does not vanish at the

same time —true for the Kepler problem and used in [9] to show that the switchings in (37) are always “ $\pi$ -singularities”, i.e. the control  $u^*$  switches to its opposite— is implied, for example, by the rank condition

$$\text{rank} \left\{ \mathcal{G}(\theta, x), \frac{\partial \mathcal{G}}{\partial \theta}(\theta, x) \right\} = \dim \left( \text{Range } \mathcal{G}(\theta, x) + \text{Range } \frac{\partial \mathcal{G}}{\partial \theta}(\theta, x) \right) = n . \quad (42)$$

This turns out to be necessary as well: if, for some  $(\theta, x)$ , it does not hold, there is at least one nonzero  $p$  such that  $\langle p, \mathcal{G}(\theta, x) \rangle = \langle p, \frac{\partial \mathcal{G}}{\partial \theta}(\theta, x) \rangle = 0$ .

Proving existence of a flow for (38) in more general situations is obviously an interesting program to be pursued. In fact, the rank condition (42) seems to be sufficient by itself instead of point (i): it implies that the number of solutions  $\theta$  to  $\langle p, \mathcal{G}(\theta, x) \rangle = 0$  is finite instead of being at most 1; the proof (of Proposition A.2) would become more intricate but seems to follow through. Of course, (42) is much more restrictive than the rank assumption (31): the latter can be considered generic, whereas the former implies restrictions on the dimension: it is necessary that  $n \leq 2m$ . It is interesting to investigate whether, under the general rank assumption (31),  $dH$  has a module of continuity maybe slightly larger than (41), that will still guarantee uniqueness of solutions.

#### 4. Kepler control systems.

**4.1. Definition.** We call *Kepler control system* with small control a family of control system on  $S^1 \times M$  of the form

$$(\mathcal{K}_\varepsilon) \quad \begin{cases} \dot{\theta} &= \omega(\theta, x) + g(\theta, x) v \\ \dot{x} &= G(\theta, x) v \end{cases} , \quad \|v\| \leq \varepsilon , \quad (43)$$

where  $G$  and  $g$  can be viewed, with the same convention as in (1), as  $n \times m$  and  $1 \times m$  matrices smoothly depending on  $(\theta, x)$  and  $\omega$  is a smooth function  $S^1 \times M \rightarrow \mathbb{R}$  that remains larger than a strictly positive constant:

$$\omega(\theta, x) \geq k_\omega > 0 \quad \forall (\theta, x) \in S^1 \times M . \quad (44)$$

In fact, this is an affine control system on  $S^1 \times M$

$$\dot{\xi} = f_0(\xi) + \sum_{i=1}^m u_i f_i(\xi) \quad (45)$$

with  $\xi = (\theta, x)$ ,  $f_0 = \omega \frac{\partial}{\partial \theta}$  and, for  $1 \leq i \leq m$ , the smooth vector field  $f_i$  is represented by the  $i^{\text{th}}$  column of the matrix notations  $G$  and  $g$ .

Conversely, for a control system as (45) where  $\xi$  is not a priori decomposed as a product to have a representation like (43), not only all the solutions of  $\dot{\xi} = f_0(\xi)$  should be periodic, but the orbits should induce a nice foliation that splits the state manifold into a product  $M \times S^1$ .

**4.2. Transformation into a fast oscillating system.** First set as a new time the cumulated angle  $\theta$ , i.e.  $\Theta(t) \in \mathbb{R}$  with  $\Theta(t) \equiv \theta(t) \pmod{2\pi}$  for all  $t$  and, say,  $\theta_0 = \Theta(0) \in [0, 2\pi)$ , scaled by  $\varepsilon$ :

$$\lambda = \mathcal{R}(t) \stackrel{\Delta}{=} \varepsilon (\Theta(t) - \Theta(0)) . \quad (46)$$

Taking  $\varepsilon_0$  small enough that  $|\omega(\theta, x) + \varepsilon g(\theta, x) u| > k_\omega/2$  for  $x \in \mathbb{K}$ ,  $\|u\| \leq 1$ ,  $\varepsilon < \varepsilon_0$ , one has  $d\mathcal{R}/dt > \varepsilon k_\omega/2$  hence  $\mathcal{R}$  is strictly increasing and one-to-one, and also

$$\frac{k_\omega}{2} \varepsilon t \leq \mathcal{R}(t) \leq K_a \varepsilon t \quad \text{with } K_a = \sup_{S^1 \times \mathbb{K}} \omega + \varepsilon_0 \sup_{S^1 \times \mathbb{K}} \|g\|. \quad (47)$$

Define, from  $v(\cdot)$ , the control  $\lambda \mapsto \hat{u}(\lambda) = v(\mathcal{R}^{-1}(\lambda))/\varepsilon$ ; then  $\tilde{x}(\lambda) = x(\mathcal{R}^{-1}(\lambda))$  satisfies

$$(\tilde{\Sigma}_{\theta_0, \varepsilon}) \quad \frac{d\tilde{x}}{d\lambda} = \frac{G(\theta_0 + \frac{\lambda}{\varepsilon}, \tilde{x}) \hat{u}}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \tilde{x}) + \varepsilon g(\theta_0 + \frac{\lambda}{\varepsilon}, \tilde{x}) \hat{u}}, \quad \|\hat{u}\| \leq 1. \quad (48)$$

Discarding the term  $\varepsilon g \hat{u}$  in the denominator yields a system whose right-hand side differs only by a term of order  $\varepsilon$ :

$$(\hat{\Sigma}_{\theta_0, \varepsilon}) \quad \frac{d\hat{x}}{d\lambda} = \frac{G(\theta_0 + \frac{\lambda}{\varepsilon}, \hat{x}) \hat{u}}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \hat{x})}, \quad \|\hat{u}\| \leq 1. \quad (49)$$

This is a fast oscillating system of the type (1) with  $\mathcal{G}(\theta, x) = G(\theta_0 + \theta, x)/\omega(\theta_0 + \theta, x)$ .

**4.3. Average control system, convergence.** One may now, as in §3.2, define the average system of (49) according to (5)-(4)-(3); Theorem 3.4 gives convergence of solutions of  $(\tilde{\Sigma}_{\theta_0, \varepsilon})$  as  $\varepsilon \rightarrow 0$  (for fixed  $\theta_0$ ) to solutions of the average system on fixed intervals, say  $\lambda \in [0, \Lambda]$ ,  $\Lambda > 0$  fixed. This average system does not depend on  $\theta_0$ .

In order to have an average system where the time  $\tau$  will be equivalent to  $\varepsilon t$  as  $\varepsilon \rightarrow 0$ , we multiply the right-hand side of (5) by  $\bar{\omega}$  defined by

$$\frac{1}{\bar{\omega}(x)} = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta}{a(\theta, x)} : \quad (50)$$

DEFINITION 4.1. *The average control system of the Kepler control system (43) is the differential inclusion*

$$\dot{x} \in \mathcal{E}(x) \quad (51)$$

with  $\mathcal{E}$  defined by (4) using  $\bar{\mathcal{G}} : M \times L^\infty(S^1, \mathbb{R}^m) \rightarrow \text{TM}$  defined by

$$\bar{\mathcal{G}}(x, \mathcal{U}) = \bar{\omega}(x) \frac{1}{2\pi} \int_0^{2\pi} \frac{G(\theta, x)}{a(\theta, x)} \mathcal{U}(\theta) d\theta = \frac{\int_0^{2\pi} \frac{1}{a(\theta, x)} G(\theta, x) \mathcal{U}(\theta) d\theta}{\int_0^{2\pi} \frac{1}{a(\theta, x)} d\theta}. \quad (52)$$

*Solutions are defined as in Definition 3.1.*

We now have a convergence result, which is an adaptation to (43) of Theorem 3.4: THEOREM 4.2 (Convergence for Kepler control systems).

1. *For any solution  $x_0(\cdot) : [0, T] \rightarrow M$  of (51) and any  $\theta^0 \in S^1$ , there exist constants  $c > 0$  and  $\varepsilon_0 > 0$ , and, for each  $\varepsilon$ ,  $0 < \varepsilon < \varepsilon_0$ , a measurable function  $\bar{u}_\varepsilon(\cdot) : [0, T] \rightarrow B^m$  such that for all positive  $\varepsilon$  smaller than  $\varepsilon_0$ , the solution  $t \mapsto (\theta_\varepsilon(t), x_\varepsilon(t))$  of (43) with control  $u = \bar{u}_\varepsilon(t)$  and initialized at  $(\theta_\varepsilon(0), x_\varepsilon(0)) = (\theta^0, x_0(0))$  is defined on  $[0, T/\varepsilon]$ , and*

$$d(x_\varepsilon(\frac{t}{\varepsilon}), x_0(t)) < c\varepsilon, \quad t \in [0, T], \quad (53)$$

*thus  $t \mapsto x_\varepsilon(t/\varepsilon)$  converges uniformly on  $[0, T]$  to  $t \mapsto x_0(t)$  when  $\varepsilon$  tends to zero.*

2. Let  $\mathbb{K}$  be a compact subset of  $M$ ,  $(\varepsilon_n)_{n \in \mathbb{N}}$  a decreasing sequence of positive real numbers converging to zero, and, for each  $n$ ,  $(\theta_n(\cdot), x_n(\cdot)) : [0, T/\varepsilon_n] \rightarrow S^1 \times \mathbb{K}$  a solution of system (43) with  $\varepsilon = \varepsilon_n$  and some control  $u = u_n(t)$ ,  $u_n(\cdot) \in L^\infty([0, T/\varepsilon_n], \mathbb{R}^m)$ ,  $\|u_n(\cdot)\|_\infty \leq 1$ . Then the sequence  $(\tau \mapsto (x_n(\tau/\varepsilon_n)))_{n \in \mathbb{N}}$  is compact for the topology of uniform convergence on  $[0, T]$  and the limit of any converging sub-sequence is a solution  $x^*(\cdot)$  of the average differential inclusion (51).

*Proof.* This proof is done as if  $M$  was  $\mathbb{R}^n$ ,  $d$  the Euclidean distance and all vector fields zero outside a compact set, so that all maps have a global Lipschitz constant and a global bound; it is left to the care of the reader to verify, as in the proof of Theorem 3.4, that it translates on  $M$  with any distance  $d$  described in §2.2. The constants in the proof are global; in the theorem they depend on the chosen distance  $d$  and on the compact  $\mathbb{K}$  or the solution  $x_0(\cdot)$  we are approaching.

Since (48) can be re-written

$$\frac{dx}{d\lambda} = \left(1 - \frac{\varepsilon g(\theta_0 + \frac{\lambda}{\varepsilon}, x) u}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, x) (\omega(\theta_0 + \frac{\lambda}{\varepsilon}, x) + \varepsilon g(\theta_0 + \frac{\lambda}{\varepsilon}, x) u)}\right) \frac{G(\theta_0 + \frac{\lambda}{\varepsilon}, x) u}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, x)}, \quad (54)$$

the norm of the difference between the right-hand sides of  $(\widehat{\Sigma}_{\theta_0, \varepsilon})$  and  $(\widetilde{\Sigma}_{\theta_0, \varepsilon})$  is bounded by  $k\varepsilon$  and classical theorems on smooth dependence of solutions on “parameters” yields some  $c'$  such that, if two solutions  $\tilde{x}(\cdot)$  of  $(\widetilde{\Sigma}_{\theta_0, \varepsilon})$  and  $\hat{x}(\cdot)$  of  $(\widehat{\Sigma}_{\theta_0, \varepsilon})$  are defined —from (47),  $\lambda$  will never be larger than  $K_\omega T$ — on  $[0, K_\omega T]$ , with the same control  $\lambda \mapsto u(\lambda)$  and  $\tilde{x}(0) = \hat{x}(0)$ , then

$$d(\tilde{x}(\lambda), \hat{x}(\lambda)) \leq c' \varepsilon \quad \text{for all } \lambda \in [0, K_\omega T]. \quad (55)$$

Let  $\tau \mapsto x_0(\tau)$  be a solution of (51) on  $[0, T]$ . Then  $\lambda \mapsto \hat{x}_0(\lambda)$ , with

$$\lambda = \mathcal{P}(\tau) = \int_0^\tau \bar{\omega}(x_0(t)) dt \quad (56)$$

and  $x_0(\tau) = \hat{x}_0(\mathcal{P}(\tau))$  for all  $\tau$ , is a solution of

$$\frac{d\hat{x}_0}{d\lambda} \in \bar{\omega}(\hat{x}_0) \mathcal{E}(\hat{x}_0) \quad (57)$$

on  $[0, \mathcal{P}(T)]$ . This is, according to (52), the average system of the fast oscillating (49); applying Theorem 3.4 (Point 1) yields a family of controls  $\hat{u}_\varepsilon$  such that the solutions  $\hat{x}_\varepsilon(\cdot)$  of  $(\widehat{\Sigma}_{\theta_0, \varepsilon})$  with initial condition  $\hat{x}_0(0)$  and control  $\hat{u}_\varepsilon$  converge to  $\hat{x}_0(\cdot)$  uniformly on  $[0, \mathcal{P}(T)]$  with a distance less than  $c''\varepsilon$ . For each  $\varepsilon$ , the solution  $\tilde{x}_\varepsilon(\cdot)$  of  $(\widetilde{\Sigma}_{\theta_0, \varepsilon})$  with same initial condition and control satisfies, according to (55),  $d(\tilde{x}_\varepsilon(\lambda), \hat{x}_\varepsilon(\lambda)) < c'\varepsilon$ ; then define

$$t = \mathfrak{T}(\lambda) = \frac{1}{\varepsilon} \int_0^\lambda \frac{d\ell}{\omega(\theta_0 + \frac{\ell}{\varepsilon}, \tilde{x}_\varepsilon(\ell)) + \varepsilon g(\theta_0 + \frac{\ell}{\varepsilon}, \tilde{x}_\varepsilon(\ell)) \hat{u}_\varepsilon(\ell)} \quad (58)$$

and the controls  $t \mapsto \bar{u}_\varepsilon(t)$  by  $\hat{u}_\varepsilon(\lambda) = \bar{u}_\varepsilon(\mathfrak{T}(\lambda))$ ; the solutions  $x_\varepsilon(\cdot)$  of (43) with these controls are given by  $\tilde{x}_\varepsilon(\lambda) = x_\varepsilon(\mathfrak{T}(\lambda))$ , and one therefore has

$$d(x_\varepsilon(\mathfrak{T} \circ \mathcal{P}(\tau)), X_0(\tau)) < (c' + c'')\varepsilon, \quad \tau \in [0, T]. \quad (59)$$

Let us now study  $\mathfrak{T} \circ \mathcal{P}(\tau)$ . From (56) and (58), one has

$$\frac{d}{d\tau} \mathfrak{T} \circ \mathcal{P}(\tau) = \frac{1}{\varepsilon} \frac{\bar{\omega}(x_0(\tau))}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \tilde{x}_\varepsilon(\lambda)) + \varepsilon g(\theta_0 + \frac{\lambda}{\varepsilon}, \tilde{x}_\varepsilon(\lambda)) \hat{u}_\varepsilon(\lambda)} \quad \text{with } \lambda = \mathcal{P}(\tau);$$

writing the fraction as in (54), the fact that  $x_0(\tau) = \widehat{x}_0(\lambda)$ ,  $d(\widetilde{x}(\lambda), \widehat{x}(\lambda)) \leq c'' \varepsilon$ , the right-hand side above reads, , with  $|\rho|$  bounded independently of  $\varepsilon$ ,

$$\frac{1}{\varepsilon} + \frac{\overline{\omega}(x_0(\tau))}{\varepsilon} \left( \frac{1}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \widehat{x}_0(\lambda))} - \frac{1}{\overline{\omega}(\widehat{x}_0(\lambda))} \right) + \rho(\theta_0, \tau).$$

Now,

$$\begin{aligned} \int_0^\tau \frac{\overline{\omega}(x_0(t)) dt}{\omega(\theta_0 + \frac{\mathcal{P}(t)}{\varepsilon}, \widehat{x}_0(\mathcal{P}(t)))} &= \int_0^{\mathcal{P}(\tau)} \frac{d\lambda}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \widehat{x}_0(\lambda))} = \frac{1}{2\pi} \iint_{\substack{0 \leq \theta \leq 2\pi \\ 0 \leq \lambda \leq \mathcal{P}(\tau)}} \frac{d\lambda d\theta}{\omega(\theta_0 + \frac{\lambda}{\varepsilon}, \widehat{x}_0(\lambda))}, \\ \int_0^\tau \frac{\overline{\omega}(x_0(t)) dt}{\overline{\omega}(\widehat{x}_0(\mathcal{P}(t)))} &= \int_0^{\mathcal{P}(\tau)} \frac{d\lambda}{\overline{\omega}(\widehat{x}_0(\lambda))} = \frac{1}{2\pi} \iint_{\substack{0 \leq \theta \leq 2\pi \\ 0 \leq \lambda \leq \mathcal{P}(\tau)}} \frac{d\lambda d\theta}{\omega(\theta_0 + \frac{\lambda}{\varepsilon} + \theta, \widehat{x}_0(\lambda))} \\ &= \frac{1}{2\pi} \iint_{\substack{0 \leq \theta \leq 2\pi \\ \varepsilon\theta \leq \ell \leq \mathcal{P}(\tau) + \varepsilon\theta}} \frac{d\lambda d\theta}{\omega(\theta_0 + \frac{\ell}{\varepsilon}, \widehat{x}_0(\ell - \varepsilon\theta))}, \end{aligned}$$

hence the difference is less than  $k\varepsilon$  and this implies that  $|\mathfrak{T} \circ \mathcal{P}(\tau) - \frac{\tau}{\varepsilon}|$  is bounded by a constant and finally, since  $x_\varepsilon(\cdot)$  is Lipschitz continuous with constant  $2\varepsilon \sup \|G\|/k_\omega$ , one has  $d(x_\varepsilon(\mathfrak{T} \circ \mathcal{P}(\tau)), x_\varepsilon(\frac{\tau}{\varepsilon})) < c''' \varepsilon$  for some  $\varepsilon$  and finally (59) implies point 1 of the theorem, with  $c = c' + c'' + c'''$  in (58).

For point 2, consider  $(\theta_n(\cdot), x_n(\cdot)) : [0, T/\varepsilon_n] \rightarrow S^1 \times \mathbb{K}$  a solution of system (43) with  $\varepsilon = \varepsilon_n$  and some control  $u = u_n(t)$ . One associates to these, following (46)–(48) and setting  $\lambda = \mathcal{R}_n(t)$  (we write  $\mathcal{R}_n$  because  $\mathcal{R}$  in (46) is constructed for system  $(\widetilde{\Sigma}_{\theta_0, \varepsilon_n})$  and thus depends on  $n$ ), solutions and control  $\lambda \mapsto \widetilde{x}_n(\lambda)$  and  $\lambda \mapsto \widetilde{u}_n(\lambda)$  of  $(\widetilde{\Sigma}_{\theta_0, \varepsilon_n})$ . The solutions  $\lambda \mapsto \widehat{x}_n(\lambda)$  of  $(\widehat{\Sigma}_{\theta_0, \varepsilon_n})$  with same control and initial condition satisfy, according to (55),  $d(\widehat{x}_n(\lambda), \widetilde{x}_n(\lambda)) < c' \varepsilon_n$ . Applying Theorem 3.4 (Point 2) yields convergence of subsequences of  $(\widehat{x}_n)$  to solutions  $\lambda \mapsto \widehat{x}_0(\lambda)$  of (57), hence the same subsequences of  $(\widetilde{x}_n)$  converge as well, and, with  $\tau = \mathcal{Q}(\lambda) \triangleq \int_0^\lambda \frac{d\ell}{\overline{\omega}(\widehat{x}(\ell))}$ , the maps  $\tau \mapsto \widetilde{x}_n(\mathcal{Q}^{-1}(\tau)) = x_n((\mathcal{Q} \circ \mathcal{R}_n)^{-1}(\tau))$  converge to the solution  $\tau \mapsto x_0(\tau) = \widehat{x}_0(\mathcal{Q}^{-1}(\tau))$  of the average system (51). Using the same argument as in the end of the paragraph above for  $\mathfrak{T} \circ \mathcal{P}(\tau)$ , one may prove that  $|(\mathcal{Q} \circ \mathcal{R}_n)^{-1}(\tau) - \frac{\tau}{\varepsilon_n}|$  is bounded and that, for some  $c'''$ , one has  $d(x_n((\mathcal{Q} \circ \mathcal{R}_n)^{-1}(\tau)), x_n(\frac{\tau}{\varepsilon})) \leq c''' \varepsilon_n$ ; Point 2 is proved.  $\square$

**4.4. Further results.** Subsections 3.4 to 3.6 can be transposed *mutatis mutandis* from “fast oscillating control systems” (1) to “Kepler control systems” (43).

Instead of (24),  $H$  has to be taken as follows (Propositions 3.7, 3.8 and later on):

$$H(x, p) = \overline{\omega}(x) \frac{1}{2\pi} \int_0^{2\pi} \left\| \left\langle p, \frac{G(\theta, x)}{a(\theta, x)} \right\rangle \right\| d\theta = \frac{\int_0^{2\pi} \left\| \left\langle p, \frac{G(\theta, x)}{a(\theta, x)} \right\rangle \right\| d\theta}{\int_0^{2\pi} \frac{1}{a(\theta, x)} d\theta} \quad (60)$$

In Proposition 3.9,  $\mathcal{G}$  can simply be replaced with  $G$ . It is however interesting to give a more intrinsic characterisation of the dimension of  $\mathcal{E}(x)$  in terms of the vector fields  $f_i$  used in the expression (45) of system  $(\mathcal{K}_\varepsilon)$ :

$$\begin{aligned} r(\theta, x) &= -1 + \text{rank} \left( \{f_0(\theta, x)\} \cup \left\{ \text{ad}_{f_0}^j f_k(\theta, x), j \in \mathbb{N}, 1 \leq k \leq m \right\} \right) \\ &= \dim \left( \sum_{j \in \mathbb{N}} \text{Range} \frac{\partial^j G}{\partial \theta^j}(\theta, x) \right). \end{aligned} \quad (61)$$

where the second expression is merely (30). Recall that the maximum of  $r(\theta, x)$  over  $\theta \in S^1$  gives the dimension of  $\mathcal{E}(x)$  and in particular, from Proposition 3.11:

**PROPOSITION 4.3.** *If the vector fields  $f_0$  and  $\text{ad}_{f_0}^j f_k$  for  $1 \leq k \leq m$  and  $j \in \mathbb{N}$  span the whole tangent space of  $S^1 \times M$  at each point, then, for all  $x$ ,  $\mathcal{E}(x)$  has a nonempty interior and it is the unit ball of a norm on  $\text{TM}$ , dual of the norm  $p \mapsto H(p, x)$  on  $\text{T}^*M$ . Furthermore, the map  $H^2$  is continuously differentiable.*

*We call this case the full rank case.*

This rank condition also amounts to controllability of the linear approximation of the system (45), or (43), along the periodic solutions obtained for a zero control, see Proposition 3.10.

*Application to the minimum time problem.* As in §3.6, but for the Kepler system (43), let  $x_0, x_1$  be fixed, call  $T_\varepsilon$  the minimum time such that, from *some*  $\theta_0, \theta_1$ ,  $(\theta_1, x_1)$  can be reached from  $(\theta_0, x_0)$  in system  $(\mathcal{K}_\varepsilon)$  (obviously  $T_\varepsilon \rightarrow +\infty$  as  $\varepsilon \rightarrow 0$ ) and  $T_\varepsilon$  the minimum time such that  $x_1$  can be reached from  $x_0$  in the average system (51). The equivalent of Theorem 3.14, with a similar proof, using Theorem 4.2, is:

**THEOREM 4.4.** *In the full rank case, one has  $\limsup_{\varepsilon \rightarrow 0} \varepsilon T_\varepsilon \leq T_0$  (hence  $\varepsilon T_\varepsilon$  is bounded as  $\varepsilon \rightarrow 0$ ). If, for all  $\varepsilon > 0$  small enough, there is a minimizing solution  $(\theta_\varepsilon, x_\varepsilon) : [0, T_\varepsilon] \rightarrow S^1 \times M$  and they all remain in a common compact subset of  $M$ , then all accumulation points (as  $\varepsilon \rightarrow 0$ ) of the compact family  $(\tau \rightarrow x_\varepsilon(\frac{\tau}{\varepsilon}))_{\varepsilon > 0}$  in  $C^0([0, T_0], M)$  are minimizing for the average system and  $\lim_{\varepsilon \rightarrow 0} \varepsilon T_\varepsilon = T_0$ .*

The Hamiltonian for minimum time for the average system is given by (60); one has to perform the time scaling described in §4.2 to have a result like Theorem 3.18 and the simple “commutation between averaging and writing Hamiltonian” noted in Remark 3.17. Let us translate in terms of (43) the sufficient condition for existence of a Hamiltonian flow given by Theorem 3.18:

**THEOREM 4.5.** *In the full rank case, assume that  $\langle p, G(\theta, x) \rangle$  and  $\langle p, \partial G / \partial \theta(\theta, x) \rangle$  do not vanish simultaneously outside  $\{p = 0\}$ , that  $\theta \mapsto \langle p, G(\theta, x) \rangle$  vanishes at most once for each  $(x, p) \in \text{T}^*M$ ,  $p \neq 0$ , and that  $\text{rank } \mathcal{G}(\theta, x) = m$  for each  $(\theta, x) \in S^1 \times M$ .*

*Then the Cauchy problem (38), with  $H$  given by (60), has unique solutions for any initial condition.*

The discussion of Theorem 3.18 also applies to the above; let us mention that, once it has been checked that, for each  $(x, p)$ ,  $\langle p, G(\theta, x) \rangle$  vanishes for at most one  $\theta$ , the other conditions are guaranteed if, (42) holds with  $\mathcal{G}$  replaced by  $G$  or, in terms of the vector fields in (45), if, for all  $\xi = (\theta, x)$ ,

$$\text{rank}\{f_0(\xi), f_1(\xi), \dots, f_m(\xi), \text{ad}_{f_0} f_1(\xi), \dots, \text{ad}_{f_0} f_m(\xi)\} = n + 1. \quad (62)$$

We prove in next section that the above conditions are true at least for the planar control 2-body problem.

**5. Application to the control 2-body system.** In this section we study some properties of the planar control system and demonstrate that it satisfies the condition of Theorem 3.18 on the domain of non-degenerated elliptic orbits.

**5.1. Planar control 2-body system.** It is classically described by some first integrals of the free movement —here the semi-major axis  $a$  and the eccentricity vector  $(e_x, e_y)$ — and one angle  $L$  following the dynamics; we restrict to the set of non-degenerated elliptic orbits rotating in the direct sense, i.e. the state space is  $S^1 \times M$  with  $M = \{(a, e_x, e_y) \in \mathbb{R}^3, a > 0 \text{ and } e_x^2 + e_y^2 < 1\}$ . The control  $u = (u_t, u_n)$  is

expressed in the tangential-normal frame and the system reads:

$$\frac{d}{dt} \begin{pmatrix} a \\ e_x \\ e_y \\ L \end{pmatrix} = \frac{1}{a^{3/2}} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \mathbf{w}(e_x, e_y, L) \end{pmatrix} + \sqrt{a} \begin{pmatrix} 2a \mathbf{a}_a(e_x, e_y, L) & 0 \\ 2 \mathbf{a}_x(e_x, e_y, L) & \mathbf{b}_x(e_x, e_y, L) \\ 2 \mathbf{a}_y(e_x, e_y, L) & \mathbf{b}_y(e_x, e_y, L) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_t \\ u_n \end{pmatrix} \quad (63)$$

$$\begin{aligned} \text{with } \mathbf{w}(e_x, e_y, L) &= \frac{(1 + e_x \cos L + e_y \sin L)^2}{(1 - e^2)^{3/2}} \\ \mathbf{a}_a(e_x, e_y, L) &= \frac{\sqrt{1 + e^2 + 2e_x \cos L + 2e_y \sin L}}{\sqrt{1 - e^2}} \\ \mathbf{a}_x(e_x, e_y, L) &= \frac{\sqrt{1 - e^2}}{\sqrt{1 + e^2 + 2e_x \cos L + 2e_y \sin L}} (e_x + \cos L) \\ \mathbf{a}_y(e_x, e_y, L) &= \frac{\sqrt{1 - e^2}}{\sqrt{1 + e^2 + 2e_x \cos L + 2e_y \sin L}} (e_x + \cos L) \\ \mathbf{b}_x(e_x, e_y, L) &= \frac{\sqrt{1 - e^2}}{\sqrt{1 + e^2 + 2e_x \cos L + 2e_y \sin L}} \\ &\quad \times \frac{-2e_y + (e_x^2 - e_y^2 - 1) \sin L - 2e_x e_y \cos L}{1 + 2e_x \cos L + 2e_y \sin L} \\ \mathbf{b}_y(e_x, e_y, L) &= \frac{\sqrt{1 - e^2}}{\sqrt{1 + e^2 + 2e_x \cos L + 2e_y \sin L}} \\ &\quad \times \frac{2e_x + (e_x^2 - e_y^2 + 1) \cos L + 2e_x e_y \sin L}{1 + 2e_x \cos L + 2e_y \sin L} \end{aligned}$$

The eccentricity  $e$  is the norm of the eccentricity vector,  $e = \sqrt{e_x^2 + e_y^2}$ . Low thrust translates into  $\|u\| \leq \varepsilon$  for a small  $\varepsilon$ .

*Remark 5.1.* This is indeed a ‘‘Kepler control system’’ of the type (43) except that, with  $\omega = \mathbf{w}/a^{3/2}$  is strictly positive but not bounded from below on  $S^1 \times M$  by a positive constant with  $M$  defined above: to have such a positive lowerbond, one must take  $e_x^2 + e_y^2 < \bar{c}^2 < 1$  instead of  $e_x^2 + e_y^2 < 1$ . This is not a problem to apply our convergence results that require anyway to restrict oneself to a compact subset of  $M$  and on such a compact set such a positive  $\bar{\alpha}$  exists.

We may anyway compute the Hamiltonian that both defines the average system and yields the Hamiltonian system governing extremals for minimum time.

It can be expressed as  $H(a, e_x, e_y, p_a, p_{e_x}, p_{e_y}) = a^2 \mathcal{H}(e_x, e_y, ap_a, p_{e_x}, p_{e_y})$  with

$$\begin{aligned} \mathcal{H}(e_x, e_y, A, X, Y) &= \frac{1}{2\pi} \int_0^{2\pi} \|(A X Y) \mathbf{G}(e_x, e_y, L)\|, \\ \mathbf{G}(e_x, e_y, L) &= \begin{pmatrix} 2 \mathbf{a}_a/\mathbf{w} & 0 \\ 2 \mathbf{a}_x/\mathbf{w} & \mathbf{b}_x/\mathbf{w} \\ 2 \mathbf{a}_y/\mathbf{w} & \mathbf{b}_y/\mathbf{w} \end{pmatrix}. \end{aligned}$$

**5.2. Hamiltonian flow.** We want to make sure that the conditions of Theorem 4.5 are satisfied.

PROPOSITION 5.2. *Fore each  $(e_x, e_y, a)$  with  $e_x^2 + e_y^2 < 1$  and  $a > 0$ , and each  $(A, X, Y) \neq (0, 0, 0)$ , the vector  $(A, X, Y)\mathbf{G}(e_x, e_y, L)$  vanishes for at most one angle  $L$ .*

*Proof.* Removing denominators, the equations  $A\mathbf{a}_a + X\mathbf{a}_x + Y\mathbf{a}_y = 0$  and  $X\mathbf{b}_x + Y\mathbf{b}_y = 0$  can be written:

$$\begin{aligned} & (2e_x A + 2(1 - e^2)X) \cos L + (2e_y A + 2(1 - e^2)Y) \sin L \\ & \quad = -(1 + e^2)A - 2e_x(1 - e^2)X - 2e_y(1 - e^2)Y \\ (-2e_x e_y X + (e_x^2 - e_y^2 + 1)Y) \cos L + ((e_x^2 - e_y^2 - 1)X + 2e_x e_y Y) \sin L \\ & \quad = 2e_y X - 2e_x Y. \end{aligned}$$

$$\text{If } \Delta = \begin{vmatrix} 2e_x A + 2(1 - e^2)X & 2e_y A + 2(1 - e^2)Y \\ -2e_x e_y X + (e_x^2 - e_y^2 + 1)Y & (e_x^2 - e_y^2 - 1)X + 2e_x e_y Y \end{vmatrix} \text{ is nonzero,}$$

there is clearly at most one solution  $L$ . If  $\Delta = 0$ , there exists  $\lambda \neq 0$  such that

$$\begin{aligned} 2e_x A + 2(1 - e^2)X &= \lambda (-2e_x e_y X + (e_x^2 - e_y^2 + 1)Y), \\ 2e_y A + 2(1 - e^2)Y &= \lambda ((e_x^2 - e_y^2 - 1)X + 2e_x e_y Y), \end{aligned}$$

and there may be a solution to the system above only if

$$(1 + e^2)A + 2e_x(1 - e^2)X + 2e_y(1 - e^2)Y = -2\lambda(e_y X - e_x Y)$$

These three equations forms a linear system in  $(A, X, Y)$ ,  $M(A, X, Y)^T = 0$  with

$$M = \begin{pmatrix} 2e_x & 2(1 - e^2 + \lambda e_x e_y) & -\lambda(e_x^2 - e_y^2 + 1) \\ 2e_y & -\lambda(e_x^2 - e_y^2 - 1) & 2(1 - e^2 - \lambda e_x e_y) \\ (1 + e^2) & 2(e_x(1 - e^2) + \lambda e_y) & 2(e_y(1 - e^2) - \lambda e_x) \end{pmatrix}.$$

A brief computation gives  $\det M = (1 - e)^3(1 + e)^3(\lambda^2 + 4)$ , strictly positive when  $0 \leq e < 1$ . Hence  $M(A, X, Y)^T = 0$  implies  $(A, X, Y) = 0$ .  $\square$

Since the rank of  $\mathbf{G}$  is obviously 2 and the rank of  $\{\mathbf{G}, \partial\mathbf{G}/\partial L\}$  3 for any  $(e_x, e_y, L)$ , Theorem 4.5 can be applied to the planar control 2-body system and guarantees existence of a flow for the Hamiltonian system governing the extremals of minimum time for its average system.

**6. Conclusion.** We have defined an average control system in a way that is, we believe, different from the usual use of averaging in control theory and given interesting properties for it. It already allowed us to give (with restrictions on the eccentricities coming from the point raised in Remark 5.1) a proof that the minimum times between 2 ellipses grow like  $k/\varepsilon$  for the planar 2-body problem [6].

The present results are however mostly a starting point. The regularity of  $H$  has to be further explored when the conditions of Theorem 3.18 do not hold, see the last paragraph of §3. Explicit computation of the average system and its extremals for the 2-body problem has to be conducted, and compared with the average system for a quadratic cost that lead in [7] to a completely integrable Hamiltonian system.

#### REFERENCES

- [1] A. A. Agrachev and Y. L. Sachkov. *Control theory from the geometric viewpoint*, volume 87 of *Encyclopaedia Math. Sc.* Springer-Verlag, Berlin, 2004.

- [2] V. I. Arnold. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2nd edition, 1989.
- [3] D. Bao, S.-S. Chern, and Z. Shen. *An introduction to Riemann-Finsler geometry*, volume 200 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.
- [4] A. Bombrun. *Les Transferts Orbitaux à Faible Poussée : Optimalité et Feedback*. PhD thesis, Ecole des Mines de Paris, Mar. 2007.
- [5] A. Bombrun and J.-B. Pomet. On the averaged control system. In *Proceedings of the 17th MTNS*, pages 2912–2917, 2006.
- [6] A. Bombrun and J.-B. Pomet. Asymptotic behavior of time optimal orbital transfer for low thrust 2-body control system. *DCDS Supplements*, pages 122–129, 2007.
- [7] B. Bonnard and J.-B. Caillau. Riemannian metric of the averaged energy minimization problem in orbital transfer with low thrust. *Annales de l'Institut Henry Poincaré*, 2006.
- [8] F. Bullo. Averaging and vibrational control of mechanical systems. *SIAM J. Control Optim.*, 41(2):542–562, 2002.
- [9] J.-B. Caillau and J. Noailles. Coplanar control of a satellite around the Earth. *ESAIM Control Optim. Calc. Var.*, 6:239–258 (electronic), 2001.
- [10] F. Chaplais. Averaging and deterministic optimal control. *SIAM J. Control Optim.*, 25(3):767–780, 1987.
- [11] F. Clarke, Y. Ledyev, R. Stern, and P. Wolenski. *Nonsmooth Analysis and Control Theory*, volume 178 of *Graduate Texts in Mathematics*. Springer, 1998.
- [12] M. Fliess, J. Lévine, P. Martin, and P. Rouchon. Flatness and defect of nonlinear systems: Introductory theory and examples. *Internat. J. Control*, 61(6):1327–1361, 1995.
- [13] S. Geffroy. *Généralisation des techniques de moyennation en contrôle optimal - Application aux problèmes de rendez-vous orbitaux en poussée faible*. Thèse de doctorat, Institut National Polytechnique de Toulouse, Toulouse, France, Oct. 1997.
- [14] S. Geffroy and R. Epenoy. Optimal low-thrust transfers with constraints-generalization of averaging technics. *Acta Astronautica*, 41(3):133–149, 1997.
- [15] P. Hartman. *Ordinary differential equations*. Birkhauser, 1982. 2nd edition.
- [16] T. Kailath. *Linear systems*. Information and System Sciences. Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.
- [17] J. Kurzweil and J. Jarnik. Iterated Lie brackets in limit processes in ordinary differential equations. *Results in Mathematics*, 14:125–137, 1988.
- [18] W. Liu. Averaging theorems for highly oscillatory differential equations and iterated Lie brackets. *SIAM J. Control Optim.*, 35(6):1989–2020, Nov. 1997.
- [19] S. M. Meerkov. Principle of vibrational control: theory and applications. *IEEE Trans. Automat. Control*, 25(4):755–762, 1980.
- [20] P. Morin, J.-B. Pomet, and C. Samson. Design of homogeneous time-varying stabilizing control laws for driftless controllable systems via oscillatory approximation of lie brackets in closed loop. *SIAM J. Control Optim.*, 38(1):22–49, 1999.
- [21] R. M. Murray. Nilpotent bases for a class of nonintegrable distributions with applications to trajectory generation for nonholonomic systems. *Math. of Control, Signals & Systems*, 7:58–75, 1994.
- [22] L. S. Pontryagin, V. G. Boltjanskiĭ, R. V. Gamkrelidze, and E. Mitchenko. *Théorie mathématique des processus optimaux*. Editions MIR, Moscou, 1974.
- [23] J. A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems*, volume 56 of *Applied Mathematical Sciences*. Springer-Verlag, 1985.
- [24] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.
- [25] L. M. Silverman and H. E. Meadows. Controllability and observability in time-variable linear systems. *SIAM Journal on Control*, 5(1):64–73, 1967.

**Appendix. Results on averaging the norm of an oscillating vector.** Let  $O^d$  be an open subset of  $\mathbb{R}^d$ ,  $V: S^1 \times O^d \rightarrow \mathbb{R}^m$  a smooth map ( $C^\infty$ ),  $\tilde{\mathcal{Z}}$  the subset of  $O^d$  where  $V$  vanishes for some  $\theta$  and  $H: O^d \rightarrow [0, +\infty)$  the average of the norm of  $V$ :

$$H(X) = \frac{1}{2\pi} \int_0^{2\pi} \|V(\theta, X)\| d\theta, \quad \tilde{\mathcal{Z}} = \{X \in O^d, \exists \theta \in S^1, V(\theta, X) = 0\}. \quad (64)$$

This appendix states two regularity results for  $H$  (Propositions A.1 and A.2), only relevant on  $\tilde{\mathcal{Z}}$  for  $H$  is obviously smooth outside  $\tilde{\mathcal{Z}}$ .

PROPOSITION A.1. *If, for all  $X$  in  $O^d$ , the set  $\{\theta \in S^1, V(\theta, X) = 0\}$  has measure zero in  $S^1$ , then  $H$  is continuously differentiable and, for all  $X$ ,*

$$dH(X).h = \frac{1}{2\pi} \int_0^{2\pi} \left( V'_X(\theta, X).h \left| \frac{V(\theta, X)}{\|V(\theta, X)\|} \right. \right) d\theta. \quad (65)$$

*Proof.* Smoothness of  $V$  implies in particular that

$$\|V(\theta, X+h) - V(\theta, X) - V'_X(\theta, X).h\| \leq k \|h\|^2 \quad (66)$$

where  $V'_X(\theta, X)$  is smooth with respect to  $(\theta, X)$  and  $k$  is some local constant. Now, one has

$$\begin{aligned} \|V(\theta, x+h)\| - \|V(\theta, x)\| &= \left( V(\theta, x+h) - V(\theta, x) \left| \frac{V(\theta, X)}{\|V(\theta, X)\|} \right. \right) \\ &\quad + a(\theta, X, h) \frac{\|V(\theta, x+h) - V(\theta, x)\|^2}{\|V(\theta, x)\| + \|V(\theta, x+h)\|} \end{aligned}$$

with  $|a(\theta, X, h)| \leq 2$ . Hence, defining  $dH(X)$  by (65), one has, from (66), for  $\|h\|$  small enough and for some local constant  $k'$ ,

$$\frac{|H(X+h) - H(X) - dH(X).h|}{\|h\|} \leq \frac{k'}{2\pi} \int_0^{2\pi} \left( \|h\| + \frac{\|V(\theta, x+h) - V(\theta, x)\|}{\|V(\theta, x)\| + \|V(\theta, x+h)\|} \right) d\theta.$$

For fixed  $X$  and  $h \rightarrow 0$ , the integrand in the right-hand side is bounded by  $1 + \|h\|$  and converges to zero for  $\theta$  outside the set  $\{\theta \in S^1, V(\theta, x) = 0\}$ , that has measure zero because it is finite (see above); hence, from Lebesgue convergence theorem, the right-hand side tends to zero. This proves differentiability outside  $\{p = 0\}$ ; the derivative is given by (65) and is continuous from the same argument using Lebesgue convergence theorem.  $\square$

PROPOSITION A.2. *Let  $V$  satisfy the following assumptions:*

A1: *For all  $(\theta, X) \in S^1 \times O^d$  such that  $V(\theta, X) = 0$ , one has  $\text{rank } \frac{\partial V}{\partial X}(\theta, X) = m$ ,*

A2: *for any  $X \in O^d$ , there is at most one  $\theta$  such that  $V(\theta, X) = 0$ ,*

A3:  *$V$  and  $\partial V / \partial \theta$  do not vanish simultaneously,*

*and let  $\bar{X}$  be in  $\tilde{\mathcal{Z}}$ . There is a neighborhood  $U$  of  $\bar{X}$  in  $O^d$ , and a constant  $K > 0$  such that, for all  $X, Y$  in  $U$ ,*

$$\|H'(X) - H'(Y)\| \leq K \|X - Y\| \ln \frac{1}{\|X - Y\|}. \quad (67)$$

The proof is provided after stating and proving two lemmas.

LEMMA A.3. *Assume that  $\bar{X} \in \tilde{\mathcal{Z}}$  and A1–A3 are satisfied. There is a neighborhood  $U$  of  $\bar{X}$  in  $O^d$  and a smooth map  $\hat{\chi} : U \rightarrow S^1$  such that, for  $(\theta, X) \in U$ , one has  $V(\theta, X) = 0$  only if  $\theta = \hat{\chi}(X)$ , and*

$$\left( \frac{\partial V}{\partial \theta}(\hat{\chi}(X), X) \left| V(\hat{\chi}(X), X) \right. \right) = 0, \quad X \in U. \quad (68)$$

*Proof.* From A1,  $\mathcal{Z} = \{(\theta, X) \in S^1 \times O^d, V(\theta, X) = 0\}$  is a smooth submanifold of  $S^1 \times O^d$  and from A3,  $\tilde{\mathcal{Z}}$  given by (64) a smooth submanifold of  $O^d$ , both of

dimension  $d + 1 - m$ , and the projection  $\pi : S^1 \times O^d \rightarrow O^d$  induces a diffeomorphism  $\mathcal{Z} \rightarrow \tilde{\mathcal{Z}}$  whose inverse is of the form  $x \mapsto (\chi(x), x)$  with  $\chi$  a smooth map  $\tilde{\mathcal{Z}} \rightarrow S^1$  that satisfies, for all  $X \in \tilde{\mathcal{Z}}$ :  $V(\theta, X) = 0$  if and only if  $\theta = \chi(x)$ .

Consider the map  $T : S^1 \times O^d \rightarrow \mathbb{R}$  given by  $T(\theta, X) = \left( \frac{\partial V}{\partial \theta}(\theta, X) \mid V(\theta, X) \right)$ . Let  $\bar{X}$  be in  $\tilde{\mathcal{Z}}$ ; since  $V(\chi(\bar{X}), \bar{X}) = 0$ , one has  $T(\chi(\bar{X}), \bar{X}) = 0$  and  $\partial T / \partial \theta(\chi(\bar{X}), \bar{X}) = \left\| \frac{\partial V}{\partial \theta}(\chi(\bar{X}), \bar{X}) \right\|^2$ , nonzero from assumption A2: the implicit function theorem yields a unique map  $\hat{\chi}$  from a neighborhood  $U$  of  $\bar{X}$  in  $O^d$  to a neighborhood of  $\chi(\bar{X})$  in  $S^1$  such that  $\theta = \hat{\chi}(X)$  solves  $T(\theta, X) = 0$ ; it must therefore coincide with  $\chi$  in  $U \cap \tilde{\mathcal{Z}}$  and satisfies the lemma.  $\square$

LEMMA A.4. *Assume that  $\bar{X} \in \tilde{\mathcal{Z}}$  and A1–A3 are satisfied. There exist a neighborhood  $U$  of  $\bar{X}$  in  $O^d$ , local coordinates  $x_1, \dots, x_d$  defined on  $U$ , smooth maps*

$$P : U \rightarrow SO(m), \alpha : U \rightarrow \mathbb{R}, \text{ and } W : S^1 \times U \rightarrow \mathbb{R}^m \text{ such that, with } X_{\mathbf{I}} = \begin{pmatrix} x_1 \\ \vdots \\ x_{m-1} \end{pmatrix},$$

$$V(\theta, X) = P(X) \left[ \begin{pmatrix} X_{\mathbf{I}} \\ \alpha(X) (\theta - \hat{\chi}(X)) \end{pmatrix} + (\theta - \hat{\chi}(X))^2 W(\theta, X) \right] \quad (69)$$

$$= P(X) \left[ \begin{pmatrix} X_{\mathbf{I}} \\ 0 \end{pmatrix} + (\theta - \hat{\chi}(X)) W_1(\theta, X) \right] \quad (70)$$

$$\text{with } W_1(\theta, X) = \begin{pmatrix} 0_{m-1} \\ \alpha(X) \end{pmatrix} + (\theta - \hat{\chi}(X)) W(\theta, X), \quad (71)$$

in  $S^1 \times U$ , where  $\alpha$  is bounded from below:  $0 < \alpha_0 < \alpha(X)$ ,  $X \in U$ . Furthermore, for a constant  $K_3 > 0$ , one has, for all  $(\theta, X) \in S^1 \times U$ ,

$$\|V(\theta, X)\| \geq K_3 \sqrt{\|X_{\mathbf{I}}\|^2 + \alpha(X)^2 (\theta - \hat{\chi}(X))^2}, \quad (72)$$

$$\text{and } X_{\mathbf{I}} = 0 \Rightarrow \|W_1(\theta, X)\| \geq K_3. \quad (73)$$

*Proof.* The map  $X \mapsto \frac{\partial V}{\partial \theta}(\hat{\chi}(X), X)$  is nonzero for  $X = \bar{X}$ , hence it does not vanish on a sufficiently small neighborhood  $U$  of  $\bar{X}$ , and one may write

$$\frac{\partial V}{\partial \theta}(\hat{\chi}(X), X) = P(X) \begin{pmatrix} 0_{m-1} \\ \alpha(X) \end{pmatrix}, \quad \alpha(X) > \alpha_0 > 0. \quad (74)$$

Define  $v_1, \dots, v_m$ , smooth maps  $S^1 \times U \rightarrow \mathbb{R}$  by

$$\begin{pmatrix} v_1(\theta, X) \\ \vdots \\ v_m(\theta, X) \end{pmatrix} = P^{-1}(X) V(\theta, X). \quad (75)$$

For  $i$  between 1 and  $m - 1$ ,  $\frac{\partial v_i}{\partial \theta}(\hat{\chi}(X), X) = 0$  from (74), and  $v_i(\hat{\chi}(\bar{X}), \bar{X}) = 0$  from Lemma A.3 and, using Assumption A1, the rank of the map  $X \mapsto (v_1(\hat{\chi}(X), X), \dots, v_{m-1}(\hat{\chi}(X), X))$  is  $m - 1$  at  $X = \bar{X}$ : on a possibly smaller neighborhood  $U$ , there are local coordinates  $x_1, \dots, x_d$  such that  $v_i(\theta, X) = x_i + (\theta - \hat{\chi}(X))^2 W_i(\theta, X)$  for  $i \leq m - 1$  and for some smooth  $W_i$ ; substituting (74) and (75) in (68) implies  $v_m(\hat{\chi}(X), X) = 0$ , hence  $v_m(\theta, X) = \alpha(X) (\theta - \hat{\chi}(X)) + W_m(\theta, X) (\theta - \hat{\chi}(X))^2$  for a smooth  $W_m$ ; (69) is proved.

Possibly restricting  $U$  to a subset with compact closure,  $\|W(\theta, X)\|$  is bounded on  $S^1 \times U$ ; if  $|\theta - \widehat{\chi}(X)| \leq \frac{1}{2}\alpha_0/\max\|W\|$ , then (72) holds with  $K_3 = \frac{1}{2}$  according to (69); on the set where  $|\theta - \widehat{\chi}(X)| \geq \frac{1}{2}\alpha_0/\max\|W\|$ ,  $V$  does not vanish and hence  $(\|X_{\mathbf{I}}\|^2 + \alpha(X)^2(\theta - \widehat{\chi}(X))^2)^{1/2}/\|V(\theta, X)\|$  is bounded from below; (72) is proved, with  $K_3$  smaller than this bound and than  $\frac{1}{2}$ . From (71),  $W_1(\widehat{\chi}(\bar{X}), \bar{X}) \neq 0$  because  $\alpha$  does not vanish; from assumption A2 and (70) (where  $X_{\mathbf{I}} = 0$  if  $X = \bar{X}$ ),  $W_1(\theta, \bar{X}) \neq 0$  if  $\theta \neq \widehat{\chi}(\bar{X})$ , hence  $W_1$  does not vanish on  $S^1 \times \{\bar{X}\}$ ; it is therefore bounded from below on  $S^1 \times U$  with  $U$  a small enough neighborhood of  $\bar{X}$ : (73) holds with  $K_3$  smaller than this bound.  $\square$

*Proof of Proposition A.2.* We use  $[-\pi, \pi]$  instead of  $[0, 2\pi]$  as an interval of integration. Let  $h \in \mathbb{R}^d$ , with  $\|h\| = 1$ . From (65), one has, for some constant  $\tilde{K}$  using bounds on the derivatives of the smooth  $V$ ,

$$\begin{aligned} |H'(X).h - H'(Y).h| &\leq \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( V'_X(\theta, X).h - V'_X(\theta, Y).h \left| \frac{V(\theta, X)}{\|V(\theta, X)\|} \right. \right) d\theta \right| \\ &\quad + \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \widehat{V}'_X(\theta, Y).h \left| \frac{V(\theta, X)}{\|V(\theta, X)\|} \right. - \frac{V(\theta, Y)}{\|V(\theta, Y)\|} \right) d\theta \right| \\ &\leq \tilde{K}\|X - Y\| + \frac{\tilde{K}}{2\pi} \left\| \int_{-\pi}^{\pi} \frac{V(\theta, X)}{\|V(\theta, X)\|} d\theta - \int_{-\pi}^{\pi} \frac{V(\theta, Y)}{\|V(\theta, Y)\|} d\theta \right\|. \end{aligned}$$

Finally, defining

$$\widehat{V}(\varphi, X) = V(\widehat{\chi}(X) + \varphi, X), \quad \widehat{W}_1(\varphi, X) = W_1(\widehat{\chi}(X) + \varphi, X), \quad (76)$$

and making a different change of variables in the last two integrals,

$$\begin{aligned} |H'(X).h - H'(Y).h| &\leq \tilde{K}\|X - Y\| + \frac{\tilde{K}}{2\pi} \int_{-\pi}^{\pi} \left\| \frac{\widehat{V}(\varphi, X)}{\|\widehat{V}(\varphi, X)\|} - \frac{\widehat{V}(\varphi, Y)}{\|\widehat{V}(\varphi, Y)\|} \right\| d\varphi \\ &\leq \tilde{K}\|X - Y\| + \frac{\tilde{K}}{\pi} \int_{-\pi}^{\pi} \frac{\|\widehat{V}(\varphi, X) - \widehat{V}(\varphi, Y)\|}{\|\widehat{V}(\varphi, X)\|} d\varphi \quad (77) \end{aligned}$$

where the last inequality uses the identity  $\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\| \leq 2 \min\left\{ \frac{\|u-v\|}{\|u\|}, \frac{\|u-v\|}{\|v\|} \right\}$  and also holds with  $\|\widehat{V}(\varphi, Y)\|$  instead of  $\|\widehat{V}(\varphi, X)\|$  in the denominator. Now let us use Lemma A.4, let  $X = (x_1, \dots, x_d)$  and  $Y = (y_1, \dots, y_d)$  in these coordinates; from (70), one has, with  $\widehat{W}_1$  defined by (76),

$$\widehat{V}(\varphi, X) = P(X) \left[ \begin{pmatrix} X_{\mathbf{I}} \\ 0 \end{pmatrix} + \varphi \widehat{W}_1(\varphi, X) \right], \quad \widehat{V}(\varphi, Y) = P(Y) \left[ \begin{pmatrix} Y_{\mathbf{I}} \\ 0 \end{pmatrix} + \varphi \widehat{W}_1(\varphi, Y) \right]. \quad (78)$$

Hence

$$\begin{aligned} \widehat{V}(\varphi, X) - \widehat{V}(\varphi, Y) &= (P(X) - P(Y))P(X)^{-1}\widehat{V}(\varphi, X) \\ &\quad + P(Y) \left[ \varphi (W_1(\varphi, X) - W_1(\varphi, Y)) + \begin{pmatrix} X_{\mathbf{I}} - Y_{\mathbf{I}} \\ 0 \end{pmatrix} \right] \end{aligned}$$

and finally

$$\frac{\|\widehat{V}(\varphi, X) - \widehat{V}(\varphi, Y)\|}{\|\widehat{V}(\varphi, X)\|} \leq \|P(X) - P(Y)\| + \frac{|\varphi| \|W_1(\varphi, X) - W_1(\varphi, Y)\|}{\|\widehat{V}(\varphi, X)\|} + \frac{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|}{\|\widehat{V}(\varphi, X)\|}. \quad (79)$$

Two cases are to be distinguished:

(i) If  $X_{\mathbf{I}} = Y_{\mathbf{I}} = 0$ , then  $\varphi$  factors out of  $\widehat{V}(\varphi, X)$  and  $\widehat{V}(\varphi, Y)$  in (78) and the last term in (79) is zero: according to (73), the integrand in (77) is bounded by

$$\|P(X) - P(Y)\| + \frac{\|\widehat{W}_1(\varphi, X) - \widehat{W}_1(\varphi, Y)\|}{K_3},$$

and finally  $|H'(X).h - H'(Y).h| \leq K \|X - Y\|$  with a constant  $K$  that depends only on  $V$ , the open set  $U$  and the coordinates.

(ii) If  $X_{\mathbf{I}} \neq 0$  (or  $Y_{\mathbf{I}} \neq 0$ , interchanging  $X$  and  $Y$ ), then (79), using (72), implies that the integrand in (77) is bounded by

$$\|P(X) - P(Y)\| + \frac{1}{K_3} \frac{1}{\alpha_0} \|W_1(\varphi, X) - W_1(\varphi, Y)\| + \frac{1}{K_3} \sqrt{\frac{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|^2}{\|X_{\mathbf{I}}\|^2 + \alpha(X)\varphi^2}}.$$

but the same is also true replacing  $\alpha(X)$  with  $\alpha(Y)$  and  $\|X_{\mathbf{I}}\|^2$  with  $\|Y_{\mathbf{I}}\|^2$ ; hence, since  $\|a - b\|^2 \leq 4 \max\{\|a\|^2, \|b\|^2\}$ , the last term may be replaced by  $\frac{2}{K_3} \sqrt{\frac{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|^2}{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|^2 + 4\alpha_0\varphi^2}}$ , whose integral between  $-\pi$  and  $\pi$  is equal to

$$\frac{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|}{K_3 \sqrt{\alpha_0}} \ln\left(1 + \frac{4\pi\sqrt{\alpha_0}}{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|} + \frac{8\pi^2\alpha_0}{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|^2}\right),$$

which is less than  $\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|(k_1 + k_2 \ln \frac{1}{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|})$  for some  $k_1, k_2$  when, say,  $\frac{\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|}{2\sqrt{\alpha_0}} < 1$ , and finally, since  $\|X_{\mathbf{I}} - Y_{\mathbf{I}}\|$  is less than  $\|X - Y\|$  and  $u \mapsto u \ln(1/u)$  is nondecreasing, less than  $\|X - Y\|(k_1 + k_2 \ln \frac{1}{\|X - Y\|})$ .

Cases (i) and (ii) do imply (67) in general.  $\square$