RINGMATHEMATICALENGINEERINGMATHEMAT

Jean Lévine

Analysis and Control of Nonlinear Systems

A Flatness-based Approach

Springer

Analysis and Control of Nonlinear Systems

Mathematical Engineering

For further volumes:
http://www.springer.com/series/8445

Jean Lévine

Analysis and Control of Nonlinear Systems

A Flatness-based Approach

Springer

Prof. Dr. Jean Lévine
Directeur de Recherche
CAS, Unité Mathématiques et Systèmes
MINES-Paris Tech
35, rue Saint-Honoré
77300 Fontainebleau, FRANCE
jean.levine@mines-paristech.fr

ISBN 978-3-642-00838-2 e-ISBN 978-3-642-00839-9
DOI 10.1007/978-3-642-00839-9
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009926861

Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Martine, Sonia and Benjamin

Preface

The present book is a translation and an expansion of lecture notes corresponding to a course of Mathematics of Control delivered during four years at the École Nationale des Ponts et Chaussées (Marne-la-Vallée, France) to Master students. A reduced version of this course has also been given at the Master level at the University of Paris-Sud since eight years. It may therefore serve as lecture notes for teaching at the Master or PhD level but also as a comprehensive introduction to researchers interested in flatness and more generally in the mathematical theory of finite dimensional systems and control.

This book may be seen as an outcome of the applied research policy pioneered by the École des Mines de Paris (now MINES-ParisTech), France, aiming not only at academic excellence, but also at collaborating with industries on specific innovative projects to enhance technological innovation using the most advanced know-how. This influence, though indirectly visible, mainly concerns the originality of some of the topics addressed here which are, in a sense, a theoretic synthesis of the author's applied contributions and viewpoints in the control field, continuously elaborated and modified in contact with the industrial realities. Such a synthesis wouldn't have been made possible without the scientific trust and financial support of many companies during periods ranging from two to ten years. Particular thanks are due to Elf, Shell, Ifremer, Sextant Avionique, Valeo, PSA, IFP and Micro-Controle/Newport, and to all the outstanding engineers of these companies, from which the author could learn so much. The author particularly wishes to express his gratitude to Frédéric Autran and Bernard Rémond (Valeo), Alain Danielo and Roger Desailly (Micro-Controle), and Emmanuel Sedda (PSA).

The largest part of this book, dealing with flatness and applications, is inspired by works in collaboration, successively, with Benoît Charlet and Riccardo Marino, and then with Michel Fliess, Philippe Martin and Pierre Rouchon. The author addresses his warmest thanks to all of them for many fruitful discussions, in particular those in which the notion of differential flatness could be brought to light. Some of the material used in the Singular Perturbation Chapter has been elaborated with Pierre Rouchon and Yann Creff, starting with a collaboration with Elf on distillation control. Their contributions are warmly acknowledged.

The author is also indebted to all his former PhD students, and particularly Michel Cohen de Lara, Guchuan Zhu, Régis Baron, Jean-Christophe Ponsart, Philippe Müllhaupt, Balint Kiss, Rida Sabri, Thierry Miquel, Thomas Devos and Jérémy Malaizé, in addition to the previously cited ones, Benoît Charlet, Philippe Martin, Pierre Rouchon and Yann Creff, for their skillful help to develop various applications of flatness in particularly interesting directions.

The author also wishes to warmly thank all his colleagues of the Centre Automatique et Systèmes, and more particularly Guy Cohen, Pierre Carpentier and Laurent Praly for their constant scientific trust and friendly encouragements during more than twenty years.

He would also like to especially acknowledge a recent fruitful collaboration with Felix Antritter of the Bundeswehr University, München, on symbolic computation of flatness conditions.

The first part of this manuscript was translated into english when the author visited the Department of Mathematics and Statistics of the University of Kuopio, Finland, from April to June 2006, as an Invited Professor funded by a Marie Curie Host Fellowship for the Transfer of Knowledge (project PARAMCOSYS, MTKD-CT-2004-509223), and was used as lecture notes for a course delivered during this period. The author is not only indebted to Markku Nihtilä, Chairman of this department, for his kind invitation, but also for his stimulating discussions and encouragements without which this book would not yet be finished. Many thanks are also addressed to Petri Kokkonen for his most efficient and enthusiastic help in the exercise sessions and in his careful reading of a draft version of this manuscript.

This manuscript has also been used as lecture notes for a two months intensive course given in March and April 2007 at the School of Electrical Engineering and Computer Science of the University of Newcastle, Australia, at the invitation of Jose De Dona and his group, where the decision to append a second part, dealing with industrial applications, has been taken. The author particularly wishes to express his profound thanks to Jose De Dona, Maria Seron, Jaqui Ramage and Graham Goodwin.

The author is also deeply indebted to Prof. Claus Hillermeier of the Bundeswehr University, München, for his kind invitation to publish this manuscript in the Springer collection he is supervising.

J. Lévine

Contents

Chapter 2
Introduction to Differential Geometry

This Chapter aims at introducing the reader to the basic concepts of differential geometry such as *diffeomorphism*, *tangent* and *cotangent space*, *vector field*, *differential form*. Special emphasis is put on the *integrability of a family of vector fields*, or *distribution*[1], according to its role in nonlinear system theory,

For simplicity's sake, we have defined a *manifold* as the solution set to a system of implicit equations expressed in a given coordinate system, according to the *implicit function theorem*. One can then get rid of the coordinate choice thanks to the notion of *diffeomorphism* or *curvilinear coordinates*. Particular interest is given to the notion of *straightening out* coordinates, that allow to express manifolds, vector fields or distributions in a trivial way.

The interested reader may find a more axiomatic presentation e.g. in Anosov and Arnold [1980], Arnold [1974, 1980], Boothby [1975], Chevalley [1946], Choquet-Bruhat [1968], Demazure [2000], Godbillon [1969], Pham [1992]. The implicit function theorem, the constant rank theorem and the existence and uniqueness of integral curves of a differential equation, which are part of the foundations of analysis, are given without proof. Excellent proofs may be found in Arnold [1974], Cartan [1967], Dieudonné [1960], Marino [1986], Pham [1992], Pontriaguine [1975].

Some applications of these methods to Mechanics may be found in Abraham and Marsden [1978], Godbillon [1969] and, in Isidori [1995], Khalil [1996], Nijmeijer and van der Schaft [1990], Sastry [1999], Slotine and Li [1991], Vidyasagar [1993], other approaches and developments of the theory of control of nonlinear systems.

---

[1] a geometric object not to be confused with the functional analytic notion of distribution developed by L. Schwartz.

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_2, © Springer-Verlag Berlin Heidelberg 2009

2.1 Manifold, Diffeomorphism

Recall that, given a coordinate system $(x_1, \ldots, x_n)$ and a k-times continuously differentiable mapping Φ from an open set $U \subset \mathbb{R}^n$ to $\mathbb{R}^{n-p}$ with $0 \leq p < n$, the tangent linear mapping $D\Phi(x)$, also called Jacobian matrix of Φ, is the matrix whose entry of row i and column j is $\frac{\partial \Phi_i}{\partial x_j}(x)$.

We start with the following fundamental theorem:

Theorem 2.1. ***(Implicit Function Theorem)*** *Let Φ be a k-times continuously differentiable mapping, with $k \geq 1$, from an open set $U \subset \mathbb{R}^n$ to $\mathbb{R}^{n-p}$ with $0 \leq p < n$. We assume that there exists at least an $x_0 \in U$ such that $\Phi(x_0) = 0$. If for every x in U the tangent linear mapping $D\Phi(x)$ has full rank (equal to $n - p$), there exists a neighborhood $V = V_1 \times V_2 \subset U$ of x_0 in $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^{n-p}$, with $V_1 \in \mathbb{R}^p$ and $V_2 \in \mathbb{R}^{n-p}$, and a k-times continuously differentiable mapping ψ from V_1 to V_2 such that the two sets $\{x \in V_1 \times V_2 \,|\Phi(x) = 0\}$ and $\{(x_1, x_2) \in V_1 \times V_2 \,|x_2 = \psi(x_1)\}$ are equal.*

In other words, the function ψ locally satisfies $\Phi(x_1, \psi(x_1)) = 0$ and the "dependent variable" $x_2 = \psi(x_1)$ is described by the p (locally) independent variables x_1.

Example 2.1. Consider the function Φ from $\mathbb{R}^2$ to $\mathbb{R}$ defined by $\Phi(x_1, x_2) = x_1^2 + x_2^2 - R^2$ where R is a positive real. Clearly, a solution to the equation $\Phi = 0$ is given by $x_1 = \pm\sqrt{R^2 - x_2^2}$ for $|x_2| \leq R$. The implicit function Theorem confirms the existence of a local solution around a point $(x_{1,0}, x_{2,0})$ such that $x_{1,0}^2 + x_{2,0}^2 = R^2$ (e.g. $x_{1,0} = R$, $x_{2,0} = 0$), since the tangent linear mapping of Φ at such point is: $D\Phi(x_{1,0}, x_{2,0}) = (2x_{1,0}, 2x_{2,0}) \neq (0, 0)$, and has rank 1.

Note that there are two local solutions according to whether we consider the point $(x_{1,0}, x_{2,0})$ equal to $\left(\sqrt{R^2 - x_{2,0}^2}, x_{2,0}\right)$ or to $\left(-\sqrt{R^2 - x_{2,0}^2}, x_{2,0}\right)$.

The notion of manifold is a direct consequence of Theorem 2.1:

Definition 2.1. Given a differentiable mapping Φ from $\mathbb{R}^n$ to $\mathbb{R}^{n-p}$ $(0 \leq p < n)$, we assume that there exists at least an x_0 satisfying $\Phi(x_0) = 0$ and that the tangent linear mapping $D\Phi(x)$ has full rank $(n - p)$ in a neighborhood V of x_0. The set X defined by the implicit equation $\Phi(x) = 0$, is called *differentiable manifold of dimension p*. Otherwise stated:

$$X = \{x \in V | \Phi(x) = 0\}. \tag{2.1}$$

The fact that this set is non empty is a direct consequence of Theorem 2.1. If in addition Φ is k-times differentiable (resp. analytic), we say that X is a C^k differentiable manifold, $k = 1, \ldots, \infty$ (resp. analytic –or C^ω–).

If non ambiguous, we simply say *manifold*.

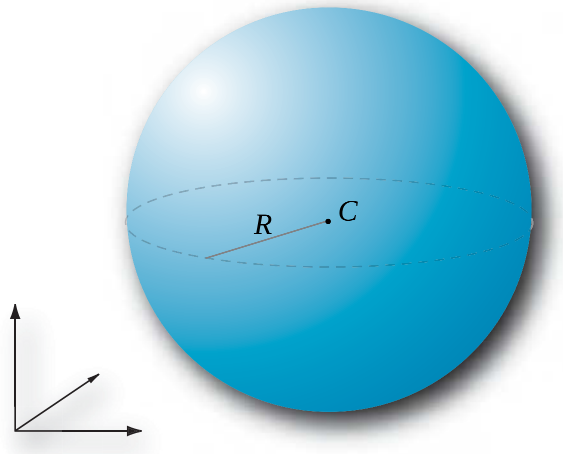

Fig. 2.1 The sphere of $\mathbb{R}^3$

Example 2.2. The affine (analytic) manifold: $\{x \in \mathbb{R}^n | Ax - b = 0\}$ has dimension p if rank $(A) = n - p$ and $b \in \mathrm{Im} A$.

Example 2.3. The sphere of $\mathbb{R}^3$ centered at C, of coordinates (x_C, y_C, z_C), and of radius R, given by $\{(x, y, z) \in \mathbb{R}^3 | (x - x_C)^2 + (y - y_C)^2 + (z - z_C)^2 - R^2 = 0\}$, is a 2-dimensional analytic manifold (see Fig. 2.1).

The concept of *local diffeomorphism* is essential to describe manifolds in an intrinsic way, namely independently of the choice of coordinates in which the implicit equation $\Phi(x) = 0$ is stated).

Definition 2.2. Given a mapping φ from an open subset $U \subset \mathbb{R}^p$ to an open subset $V \subset \mathbb{R}^p$, of class C^k, $k \geq 1$ (resp. analytic), we say that φ is a *local diffeomorphism* of class C^k (resp. analytic) in the neighborhood $U(x_0)$ of a point x_0 of U if φ is invertible from $U(x_0)$ to a neighborhood $V(\varphi(x_0))$ of $\varphi(x_0)$ of V and if its inverse φ^{-1} is also C^k (resp. analytic).

Indeed, if we consider the manifold X defined by (2.1), and if we introduce the change of coordinates $x = (x_1, x_2) = \varphi(z) = (\varphi_1(z_1), \varphi_2(z_2))$ where $\varphi = (\varphi_1, \varphi_2)$ is a local diffeomorphism of $\mathbb{R}^n$, with φ_1 (resp. φ_2) local diffeomorphism of $\mathbb{R}^p$ (resp. $\mathbb{R}^{n-p}$), the expression $x_2 = \psi(x_1)$ becomes $\varphi_2(z_2) = \psi(\varphi_1(z_1))$, or $z_2 = \left(\varphi_2^{-1} \circ \psi \circ \varphi_1\right)(z_1)$, which means that the same manifold can be equivalently represented by $x_2 = \psi(x_1)$, in the x-coordinates, or by $z_2 = \tilde{\psi}(z_1)$, with $\tilde{\psi} = \varphi_2^{-1} \circ \psi \circ \varphi_1$, in the z-coordinates. It results that the notion of manifold doesn't depend on the choice of coordinates, if the coordinate changes are diffeomorphisms.

We also introduce the slightly weaker notion of *local homeomorphism*. We say that φ is a local C^k (resp. analytic) homeomorphism if φ is of class C^k (resp. analytic), locally invertible and if its inverse is continuous.

Local diffeomorphisms are characterized by the following classical result:

Theorem 2.2. ***(of local inversion)*** *A necessary and sufficient condition for φ to be a local C^k diffeomorphism ($k \geq 1$) in a neighborhood of x_0 is that its tangent linear mapping $D\varphi(x_0)$ is one-to-one.*

We also recall:

Theorem 2.3. ***(constant rank)*** *Let φ be a C^k mapping ($k \geq 1$) from a m-dimensional C^k manifold X to a r-dimensional C^k manifold Y.*

(i) for every $y \in \varphi(U) \subset Y$, $\varphi^{-1}(\{y\})$ is a $m-q$-dimensional C^k submanifold of X;
(ii) $\varphi(U)$ is a q-dimensional C^k submanifold of Y.

In particular,

(i)' if $m \leq r$, φ is injective from U to Y if and only if $\operatorname{rank}(D\varphi(x)) = m$ for every $x \in U$ (thus φ is a homeomorphism from U to $\varphi(U)$).
(ii)' if $m \geq r$, φ is onto from U to V, an open subset of Y, if and only if $\operatorname{rank}(D\varphi(x)) = r$.

The notion of curvilinear coordinates provide a remarkable geometric interpretation of a diffeomorphism. In particular, one can find (locally) an adapted system of curvilinear coordinates in which the manifold X given by (2.1) is expressed as a vector subspace of $\mathbb{R}^p$. It suffices, indeed, to introduce the curvilinear coordinates:

$$y_1 = \Phi_1(x), \ldots, y_{n-p} = \Phi_{n-p}(x), y_{n-p+1} = \Psi_1(x), \ldots, y_n = \Psi_p(x),$$

the independent functions $\Psi_1, \ldots, \Psi_p$ being chosen such that the mapping

$$x \mapsto (\Phi_1(x), \ldots, \Phi_{n-p}(x), \Psi_1(x), \ldots, \Psi_p(x))$$

is a local diffeomorphism. In that case, we say that we have (locally) "straightened out" the coordinates of X since

$$X = \{y | y_1 = \cdots = y_{n-p} = 0\}.$$

Example 2.4. We go back to the sphere of example 2.3 and introduce the polar coordinates (ρ, θ, φ) corresponding to the transformation Γ from $\mathbb{R}_+ \times \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}/2\pi\mathbb{Z}$ to $\mathbb{R}^3$, given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \Gamma(\rho, \theta, \varphi) = \begin{pmatrix} x_C + \rho \cos\varphi \cos\theta \\ y_C + \rho \cos\varphi \sin\theta \\ z_C + \rho \sin\varphi \end{pmatrix}.$$

Clearly Γ is invertible in any of the two open sets defined by the intersection of $\mathbb{R}_+ \times \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}/2\pi\mathbb{Z}$ with $\{\cos\varphi > 0\}$ or $\{\cos\varphi < 0\}$, and where the closed subset $\{\rho\cos\varphi = 0\}$, whose image by Γ is the pair of points of cartesian coordinates $x = x_C, y = y_C, z = z_C \pm \rho$, is excluded. Γ is of class C^∞, and its local inverse is given (e.g. for $\cos\varphi > 0$) by

$$\begin{pmatrix} \rho \\ \theta \\ \varphi \end{pmatrix} = \Gamma^{-1}(x,y,z) = \begin{pmatrix} \sqrt{(x-x_C)^2 + (y-y_C)^2 + (z-z_C)^2} \\ \arctan\left(\dfrac{y-y_C}{x-x_C}\right) \\ \arctan\left(\dfrac{z-z_C}{\sqrt{(x-x_C)^2+(y-y_C)^2}}\right) \end{pmatrix}.$$

Θ is also of class C^∞ in the open set

$$\Gamma(\mathbb{R}_+ \times \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}/2\pi\mathbb{Z} \cap \{\cos\varphi > 0\}) = \mathbb{R}^2 \times \{z > z_C\} - \{(x_C, y_C, z_C + \rho)\}$$

and thus Γ is a local diffeomorphism.

In polar coordinates, the implicit equation defining the sphere becomes $\rho - R = 0$. Therefore, the sphere of $\mathbb{R}^3$ is locally equal to the set $\{\rho = R\}$.

One can check that the tangent linear mapping of Γ is given by

$$D\Gamma = \begin{pmatrix} \cos\varphi\cos\theta & -\rho\cos\varphi\sin\theta & -\rho\sin\varphi\cos\theta \\ \cos\varphi\sin\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \sin\varphi & 0 & \rho\cos\varphi \end{pmatrix}$$

and that $\det(D\Gamma) = \rho^2\cos\varphi$, which precisely vanishes on the closed subset $\{\rho\cos\varphi = 0\}$ where Γ is not injective, in accordance with Theorem 2.2.

2.2 Vector Fields

2.2.1 Tangent space, Vector Field

Assume, as before, that we are given a differentiable mapping Φ from $\mathbb{R}^n$ to $\mathbb{R}^{n-p}$ $(0 \leq p < n)$, with at least an x_0 satisfying $\Phi(x_0) = 0$. The tangent linear mapping $D\Phi(x)$ of Φ at x, expressed in the local coordinate system $(x_1, \ldots, x_n)$, is thus the matrix $\left(\frac{\partial\Phi_j}{\partial x_i}(x)\right)_{1\leq i\leq n,\, 1\leq j\leq n-p}$.
It is also assumed that $D\Phi(x)$ has full rank $(n-p)$ in a neighborhood V of x_0, so that the implicit equation $\Phi(x) = 0$ defines a p-dimensional manifold denoted by X.

We easily check that a normal vector at the point x to the manifold X is "carried" by $D\Phi(x)$, or more precisely, is a linear combination of the rows

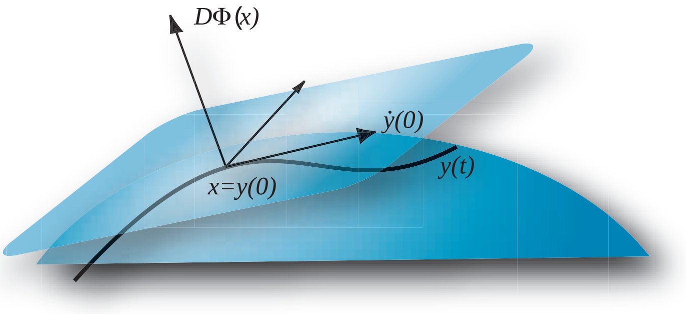

Fig. 2.2 Tangent and normal spaces to a manifold at a point.

of $D\Phi(x)$. Indeed, let $y(t)$ be a differentiable curve contained in X for all $t \in [0, \tau[$,with $\tau > 0$ sufficiently small, such that $y(0) = x$ (the existence of such a curve results from the implicit function Theorem). We therefore have $\Phi(y(t)) = 0$ for all $t \in [0, \tau[$ and thus $\frac{\Phi(y(t))-\Phi(x)}{t} = 0$. Letting t converge to 0, we get $D\Phi(x).\dot{y}(0) = 0$, where $\dot{y}(0) \stackrel{\text{def}}{=} \frac{dy}{dt}_{|t=0}$ (see Fig.2.2), which proves that the vector $\dot{y}(0)$, tangent to X at the point x, belongs to the kernel of $D\Phi(x)$. Doing the same for every curve contained in X and passing through x, it immediately results that every element of the range of $D\Phi(x)$ is orthogonal to every tangent vector to X at the point x, *Q.E.D.*

This motivates the following:

Definition 2.3. The *tangent space* to X at the point $x \in X$ is the vector space

$$\mathrm{T}_x X = \ker D\Phi(x).$$

The *tangent bundle* $\mathrm{T}X$ is the set $\mathrm{T}X = \bigcup_{x \in X} \mathrm{T}_x X$.

Taking into account the fact that $D\Phi(x)$ has rank $n-p$ in V,

$$\dim \mathrm{T}_x X = \dim \ker D\Phi(x) = p \ , \forall x \in V.$$

Example 2.5. Going back to example 2.3, the tangent space to the sphere of $\mathbb{R}^3$ at the point $(x, y, z) \neq (x_C, y_C, z_C \pm R)$ is

$$\ker\left((x - x_C)\ (y - y_C)\ (z - z_C)\right) = \text{span}\left\{ \begin{pmatrix} (y - y_C) \\ -(x - x_C) \\ 0 \end{pmatrix}, \begin{pmatrix} (z - z_C) \\ 0 \\ -(x - x_C) \end{pmatrix} \right\}$$

and is clearly 2-dimensional.

Definition 2.4. A *vector field* f (of class C^k, analytic) on X is a mapping (of class C^k, analytic) that maps every $x \in X$ to the vector $f(x) \in \mathrm{T}_x X$.

Definition 2.5. An *integral curve* of the vector field f is a local solution of the differential equation $\dot{x} = f(x)$.

The local existence and uniqueness of integral curves of f results from the fact that f is of class C^k, $k \geq 1$, and thus locally Lipschitzian[2].

2.2.2 Flow, Phase Portrait

We denote by $X_t(x)$ the point of the integral curve of the vector field f at time t, starting from the initial state x at time 0. Recall that if f is of class C^k (resp. C^∞, analytic) there exists a unique *maximal* integral curve $t \mapsto X_t(x)$ of class C^{k+1} (resp. C^∞, analytic) for every initial condition x in a given neighborhood, in the sense that the interval of time I on which it is defined is maximal.

As a consequence of existence and uniqueness, the mapping $x \mapsto X_t(x)$, noted X_t, is a local diffeomorphism for every t for which it is defined: $X_t(X_{-t}(x)) = x$ for every x and t in a suitable neighborhood $U \times I$ of $X \times \mathbb{R}$, and thus $X_t^{-1}\big|_U = X_{-t}\big|_U$, where we have denoted by $\varphi\big|_U$ the restriction of a function φ to U.

When the integral curves of f are globally defined on $\mathbb{R}$, we say that the vector field f is *complete*. In this case, X_t exists for all $t \in \mathbb{R}$, and defines a one-parameter group of local diffeomorphisms, namely:

1. the mapping $t \mapsto X_t$ is C^∞,
2. $X_t \circ X_s = X_{t+s}$ for all $t, s \in \mathbb{R}$ and $X_0 = Id_X$.

As already remarked, the items 1 and 2 imply that X_t is a local diffeomorphism for all t.

The mapping $t \mapsto X_t$ is called the *flow associated to the vector field* f. It is also often called the *flow associated to the differential equation* $\dot{x} = f(x)$.

It is straightforward to verify that the flow satisfies the differential equation

$$\frac{d}{dt} X_t(x) = f(X_t(x)) \tag{2.2}$$

for all t and every initial condition x such that $X_t(x)$ is defined.

In the time-varying case, namely for a system

---

[2] Recall that a function f from $\mathbb{R}^p$ to $\mathbb{R}^p$ is locally Lipschitzian if and only if for every open set U of $\mathbb{R}^p$ and every x_1, x_2 in U, there exists a real K such that $\|f(x_1) - f(x_2)\| \leq K\|x_1 - x_2\|$.
The differential equation $\dot{x} = f(x)$, with f locally Lipschitzian, admits, in a neighborhood of every point x_0, an integral curve passing through x_0 at $t = 0$, *i.e.* a mapping $t \mapsto x(t)$ satisfying $\dot{x}(t) = f(x(t))$ and $x(0) = x_0$ for all $t \in I$, I being an open interval of $\mathbb{R}$ containing 0.

$$\dot{x} = f(t, x) \tag{2.3}$$

the corresponding notion of flow is deduced from what precedes by adding a new differential equation describing the time evolution $\dot{t} = 1$, and augmenting the state $\tilde{x} = (x, t)$, which amounts to work with the new vector field $\tilde{f}(\tilde{x}) = (f(t, x), 1)$, which is now a stationary one on the augmented manifold $X \times \mathbb{R}$ of dimension $p + 1$.

We call *orbit* of the vector field f an equivalence class for the equivalence relation "$x_1 \sim x_2$ *if and only if there exists* t *such that* $X_t(x_1) = x_2$ *or* $X_t(x_2) = x_1$".

In other words, $x_1 \sim x_2$ if and only if x_1 and x_2 belong to the same maximal integral curve of f. We also call *orbit of a point* the maximal integral curve passing through this point and its *oriented orbit* the orbit of this point along with its sense of motion.

The *phase portrait* of the vector field f is defined as the partition of the manifold X into oriented orbits.

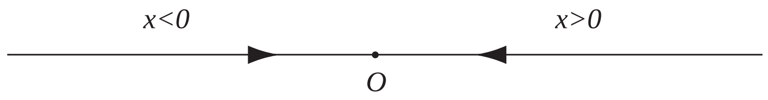

Fig. 2.3 The 3 orbits of system (2.4).

Example 2.6. The flow of the differential equation on $\mathbb{R}$

$$\dot{x} = -x \tag{2.4}$$

is $X_t(x_0) = e^{-t}x_0$. Since e^{-t} is positive for all t, two arbitrary points of $\mathbb{R}$ belong to the same integral curve if and only if they belong to the same half-line ($\mathbb{R}_+$ or $\mathbb{R}_-$) or they are both 0, *i.e.* $x_1 \sim x_2$ is equivalent to $\text{sign}\,(x_1) = \text{sign}\,(x_2)$ or $x_1 = x_2 = 0$. The system (2.4) thus admits 3 orbits: $\mathbb{R}_+$, $\mathbb{R}_-$ and $\{0\}$, as indicated on Fig. 2.3.

The same conclusion holds for the system $\dot{x} = +x$, the only difference being the orientation of the orbits, opposite to the one of (2.4).

Indeed, the flow and phase portrait do not depend on the choice of coordinates of X: if φ is a local diffeomorphism and if we note

$$z = \varphi(x)$$

we have

$$\dot{z} = \frac{\partial \varphi}{\partial x} f(\varphi^{-1}(z)). \tag{2.5}$$

Thus, denoting by g the vector field on $\varphi(X) \subset X$ defined by

$$g(z) = \frac{\partial \varphi}{\partial x} f(\varphi^{-1}(z))$$

and Z_t the local flow associated to g, one immediately sees that Z_t is deduced from the flow X_t by the formula $Z_t(\varphi(x)) = \varphi(X_t(x))$, or:

$$Z_t \circ \varphi = \varphi \circ X_t. \tag{2.6}$$

It results that if $x_1 \sim x_2$, then $z_1 = \varphi(x_1) \sim z_2 = \varphi(x_2)$, which proves that the orbits of g are the orbits of f transformed by φ and the same for their respective phase portraits.

2.2.3 Lie Derivative

Consider a system of local coordinates $(x_1, \ldots, x_p)$ in an open set $U \subset \mathbb{R}^p$. The components of the vector field f in these coordinates are denoted by $(f_1, \ldots, f_p)^T$. We now show that to f one can associate in a one-to-one way a first order differential operator called *Lie derivative along f*.

Denote, as before, by $t \longmapsto X_t(x)$ the integral curve of f in U passing through x at $t = 0$.

Definition 2.6. Let h be a function of class C^1 from $\mathbb{R}^p$ to $\mathbb{R}$ and $x \in U$. We call *Lie derivative of h along f at x*, noted $L_f h(x)$, the time derivative, at $t = 0$, of $h(X_t(x))$, *i.e.*:

$$L_f h(x) = \frac{d}{dt} h(X_t(x))_{|t=0} = \sum_{i=1}^{p} f_i(x) \frac{\partial h}{\partial x_i}(x) \ .$$

We also call *Lie derivative of h along f*, denoted by $L_f h$, the mapping $x \mapsto L_f h(x)$ from U to $\mathbb{R}$.

According to this formula, every vector field f may be identified to the linear differential operator of the first order

$$L_f = \sum_{i=1}^{p} f_i(x) \frac{\partial}{\partial x_i}.$$

It results that, in local coordinates, we can use indifferently the component-wise or the differential operator expression of f, namely

$$f = (f_1, \ldots, f_p)^T \sim \sum_{i=1}^{p} f_i(x) \frac{\partial}{\partial x_i}$$

the sign $\sim$ meaning "identified with". In the sequel, we systematically make the abuse of notation $f = (f_1, \dots, f_p)^T = \sum_{i=1}^p f_i(x) \frac{\partial}{\partial x_i}$.

Note that it results from this definition (exercise) that the Lie derivative formula is not affected by changes of coordinates[3]:

$$L_f h(x) = L_{\varphi_* f}(h \circ \varphi^{-1})(y). \tag{2.7}$$

Example 2.7. In an open set U of $\mathbb{R}^2$ with coordinates (x,t), we consider the vector field $f(t,x) = \begin{pmatrix} v \\ 1 \end{pmatrix}$ where v is an arbitrary real number, and the function $h(t,x) = x - vt$ from U to $\mathbb{R}$. The Lie derivative of h along the vector field f is given by: $L_f h(x,t) = \frac{dh}{dt} = \frac{\partial h}{\partial x} v + \frac{\partial h}{\partial t} 1 = v - v = 0$.

2.2.4 Image of a Vector Field

As before, one can introduce local coordinates in which a vector field is *straightened out*. Beforehand, we need to introduce the notions of image of a vector field by a diffeomorphism and of first integral.

To this aim, let us introduce the following simple computation: let φ be a local diffeomorphism and set $y = \varphi(x)$. We also denote by $y(t) = \varphi(X_t(x))$ the image by φ of the integral curve $X_t(x)$ of f passing through x at $t = 0$. We have

$$\frac{d}{dt} y_i(t) = \frac{d}{dt} \varphi_i(X_t(x)) = L_f \varphi_i(\varphi^{-1}(y(t))) \quad \forall i = 1, \dots, p.$$

Thus, the curve $t \mapsto y(t)$ satisfies the system of differential equations

$$\dot{y}_i = L_f \varphi_i(\varphi^{-1}(y(t))), \quad i = 1, \dots, p,$$

whose right-hand side properly defines the desired image of the vector field f by φ. We therefore have:

Definition 2.7. The image by the diffeomorphism φ of the vector field f, noted $\varphi_* f$ is the vector field given by

$$\varphi_* f = (L_f \varphi_1(\varphi^{-1}(y)), \dots, L_f \varphi_p(\varphi^{-1}(y)))^T .$$

Example 2.8. Going back to the examples 2.3, 2.4 and 2.5, we first compute the tangent space to the sphere in polar coordinates. Recall that its equation is $\rho - R = 0$. The tangent linear approximation to this mapping is $(1, 0, 0)$ and

---

[3] Hint: use the fact that $\left(\varphi^{-1}\right)_k(\varphi(x)) = x_k$ and thus, differentiating with respect to x_i, $\sum_{j=1}^p \frac{\partial (\varphi^{-1})_k}{\partial y_j} \frac{\partial \varphi_j}{\partial x_i} = \delta_{k,i}$ for all $i, k = 1 \dots, p$.

the tangent space to the sphere is its kernel, namely span $\left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$.

According to the previous notations, $v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ can be identified with $\frac{\partial}{\partial \theta}$ and $v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ with $\frac{\partial}{\partial \varphi}$.

Let us compute the image of these two vectors by the local diffeomorphism Γ defined in example 2.4. According to Definition 2.7, we have

$$\Gamma_* v_1 = \begin{pmatrix} -\rho \cos\varphi \sin\theta \\ \rho \cos\varphi \cos\theta \\ 0 \end{pmatrix} = \begin{pmatrix} -(y - y_C) \\ (x - x_C) \\ 0 \end{pmatrix}$$

and

$$\Gamma_* v_2 = \begin{pmatrix} -\rho \sin\varphi \cos\theta \\ -\rho \sin\varphi \sin\theta \\ 0 \end{pmatrix} = \begin{pmatrix} -\dfrac{(x - x_C)(z - z_C)}{\sqrt{(x - x_C)^2 + (y - y_C)^2}} \\ -\dfrac{(y - y_C)(z - z_C)}{\sqrt{(x - x_C)^2 + (y - y_C)^2}} \\ \sqrt{(x - x_C)^2 + (y - y_C)^2} \end{pmatrix}.$$

It is readily verified that the two vectors $\Gamma_* v_1$ and $\Gamma_* v_2$ are independent and belong to the kernel of the mapping $(x, y, z) \mapsto (x - x_C)^2 + (y - y_C)^2 + (z - z_C)^2$, and thus that the image of the tangent space to the sphere (in polar coordinates) at a given point is the tangent space to the sphere (in cartesian coordinates) at the image of this point. This property also results from the invariance formula (2.7) with $f = v_i$, $i = 1, 2$, and $h(\rho, \theta, \varphi) = \rho - R$. By construction, $L_{v_i} h = 0$, $i = 1, 2$, thus $L_{\Gamma_* v_i}(h \circ \Gamma^{-1}) = 0$, $i = 1, 2$.

2.2.5 *First Integral, Straightening Out of a Vector Field*

Definition 2.8. A *first integral* of f is a function γ satisfying $L_f \gamma = 0$. In other words, γ remains constant along the integral curves of f.

We say that f is *straightened out by diffeomorphism* (see Fig.2.4) if

$$L_f \varphi_p(\varphi^{-1}(y)) = 1 \text{ and } L_f \varphi_i(\varphi^{-1}(y)) = 0 \quad \forall\, i = 1, \ldots, p - 1.$$

Thus the transformed flow $Y_t = \varphi(X_t)$ of X_t by φ satisfies the set of differential equations

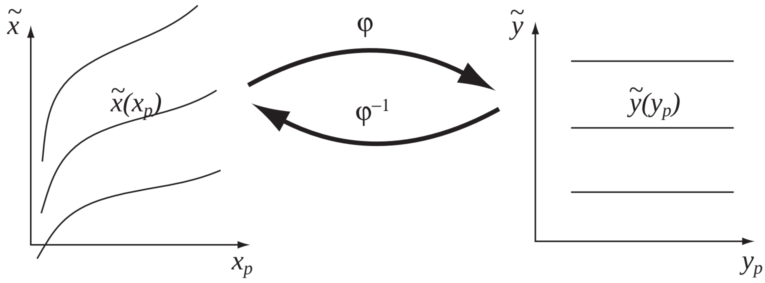

Fig. 2.4 The image by a straightening out diffeomorphism φ of the flow of a vector field in the phase space (showing $\tilde{x} = (x_1, \ldots, x_{p-1})$ in function of x_p). The transformed phase portrait (in coordinates $\tilde{y} = (y_1, \ldots, y_{p-1})$ in function of y_p) is made of parallel horizontal half-lines of equations $y_i(y_p) = Cte$, $i = 1, \ldots, p-1$.

$$\dot{y}_1 = 0, \dot{y}_2 = 0, \ldots, \dot{y}_{p-1} = 0, \dot{y}_p = 1$$

which means that the transformed vector field is $(0, 0, \ldots, 0, 1)^T$ (straightened out) and that the corresponding orbits $Y_i(y_p)$, $i = 1, \ldots, p-1$ are parallel lines to the y_p axis.

We say that $x_0 \in X$ is a *transient*, or *regular*, point of the vector field f if $f(x_0) \neq 0$. Clearly, at such a point, we have $\dot{x} = f(x_0) \neq 0$, which means that the integral curve passing through x_0 does not stop there, which justifies the word *transient.*

Proposition 2.1. *Let x_0 be a transient point of the vector field f, of class C^k, $k \geq 1$, in X. There exists a system of local coordinates $(\xi_1, \ldots, \xi_p)$ of class C^k at x_0 for which f is straightened out, i.e. such that $L_f \xi_i = 0$ for $i = 1, \ldots, p-1$ and $L_f \xi_p = 1$. Otherwise stated, f admits a system of $p-1$ independent first integrals in a neighborhood of x_0, defined by $L_f \xi_i = 0$ for $i = 1, \ldots, p-1$.*

Example 2.9. Consider the differential system in $\mathbb{R}^2$ with coordinates (x_1, x_2):

$$\dot{x}_1 = \cos^2 x_1, \qquad \dot{x}_2 = 1. \tag{2.8}$$

Since a primitive of $\frac{\dot{x}}{cos^2 x}$ is $\tan x + C$, we immediately verify that the integral curves of (2.8) passing through (x_1^0, x_2^0) at time t_0 with $x_1^0 \neq \pm\frac{\pi}{2} + 2k\pi$, are such that $\tan x_1(t) = t - t_0 + \tan x_1^0$, $x_2(t) = t - t_0 + x_2^0$. Thus, if we set $y_1 = \tan x_1(t) - x_2$, $y_2 = x_2$, we immediately verify that $\dot{y}_1 = 0$ and $\dot{y}_2 = 1$. The local diffeomorphism $(y_1, y_2) : (x_1, x_2) \mapsto (\tan x_1 - x_2, x_2)$ is thus a straightening out of the field $\cos^2 x_1 \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}$ and, consequently, $y_1 = \tan x_1 - x_2$ is a first integral of (2.8).

Example 2.10. Consider now the system

$$\begin{aligned} \dot{x}_1 &= x_1 x_2 \cos^2 x_2 \\ \dot{x}_2 &= \cos^2 x_2. \end{aligned} \tag{2.9}$$

The vector field $x_1 x_2 \cos^2 x_2 \frac{\partial}{\partial x_1} + \cos^2 x_2 \frac{\partial}{\partial x_2}$ associated to (2.9) is straightened out by the local diffeomorphism $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \log x_1 - \frac{1}{2}x_2^2 \\ \tan x_2 \end{pmatrix}$.

Example 2.11. Let U be a function of class C^1 from $\mathbb{R}^p$ to $\mathbb{R}$. We say that the system

$$m\ddot{x}_i = -\frac{\partial U}{\partial x_i}, \quad i = 1, \ldots, p \tag{2.10}$$

is *of gradient type* relatively to the potential U. We easily check that the function

$$V(x, \dot{x}) = \frac{1}{2} m \sum_{i=1}^{p} \dot{x}_i^2 + U(x)$$

(often called mechanical energy) is a first integral of the vector field $\sum_{i=1}^{p} \left(\dot{x}_i \frac{\partial}{\partial x_i} - \frac{1}{m} \frac{\partial U}{\partial x_i} \frac{\partial}{\partial \dot{x}_i} \right)$.

Example 2.12. Let H be a function of class C^1 from $\mathbb{R}^r \times \mathbb{R}^r$ to $\mathbb{R}$. Consider the Hamiltonian system

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad i = 1, \ldots, r. \tag{2.11}$$

We easily check that the function H is a first integral of the vector field $\sum_{i=1}^{r} \left(\frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i} \right)$.

Moreover, if we set $q_i = x_i$, $p_i = m\dot{q}_i = m\dot{x}_i$ and $H = \frac{1}{2m} \sum_{i=1}^{r} p_i^2 + U(q)$, we recover the previous example. In Hamiltonian Mechanics, q is interpreted as the *generalized position* and p as the *generalized impulsion.*

2.2.6 Lie Bracket

Consider, as in the previous section, a vector field f and a smooth function h. The operator L_f, the Lie derivative along f, can be iterated. Indeed, for every smooth function h from X to $\mathbb{R}$, one can define $L_f^k h$ for all $k \geq 0$ as follows:

$$L_f^0 h = h \text{ and } L_f^k h = L_f(L_f^{k-1} h) \; \forall k \geq 1 .$$

For instance, $L_f^2 h = \sum_{i,j=1}^{n} (f_i \frac{\partial f_j}{\partial x_i} \frac{\partial h}{\partial x_j} + f_i f_j \frac{\partial^2 h}{\partial x_i \partial x_j})$. The iterated Lie derivative of order k thus defines a continuous differential operator of order k on the set of functions of class C^k.

Accordingly, if $g_1, \ldots, g_k$ are k vector fields on X, following the same lines, we can inductively define the Lie derivative of order $r_1 + \ldots + r_k$ by:

$$L_{g_1}^{r_1} \cdots L_{g_k}^{r_k} h = L_{g_1}^{r_1} (L_{g_2}^{r_2} \cdots L_{g_k}^{r_k} h) .$$

If now f and g are two vector fields, let us compute, in a local coordinate system, the following expression:

$$L_f L_g h - L_g L_f h = \sum_{i=1}^{p} \left(\sum_{j=1}^{p} \left(f_j \frac{\partial g_i}{\partial x_j} - g_j \frac{\partial f_i}{\partial x_j} \right) \right) \frac{\partial h}{\partial x_i} .$$

This expression is skew symmetric with respect to f and g, and defines a new differential operator of order 1, since the symmetric second order terms of $L_f L_g h$ and $L_g L_f h$ cancel. It thus constitutes a new vector field, denoted by $[f, g]$, called the Lie bracket of f and g.

Definition 2.9. The *Lie bracket* of the vector fields f and g is the vector field defined by:

$$L_{[f,g]} = L_f L_g - L_g L_f.$$

In local coordinates:

$$[f, g] = \sum_{i=1}^{p} \left(\sum_{j=1}^{p} \left(f_j \frac{\partial g_i}{\partial x_j} - g_j \frac{\partial f_i}{\partial x_j} \right) \right) \frac{\partial}{\partial x_i} . \tag{2.12}$$

The Lie bracket enjoys the following properties:

- skew symmetry : $[f, g] = -[g, f]$;
- $[\alpha f, \beta g] = \alpha\beta[f, g] + (\alpha L_f \beta) g - (\beta L_g \alpha) f$, for every pair (α, β) of C^∞ functions;
- Jacobi identity: $[f_1, [f_2, f_3]] + [f_2, [f_3, f_1]] + [f_3, [f_1, f_2]] = 0$.

It also satisfies:

Proposition 2.2. *Given a diffeomorphism φ from a manifold X to a manifold Y and f_1 and f_2 two arbitrary vector fields of X, denoting by $\varphi_* f_1$ and $\varphi_* f_2$ their images by φ in Y, we have:*

$$[\varphi_* f_1, \varphi_* f_2] = \varphi_* [f_1, f_2]. \tag{2.13}$$

Proof. By definition, we have $L_{\varphi_* f_i} h(\varphi(x)) = L_{f_i}(h \circ \varphi)(x)$, $i = 1, 2$, for every differentiable function h on Y. Thus $L_{\varphi_* f_2} L_{\varphi_* f_1} h(\varphi(x)) = L_{f_2}((L_{\varphi_* f_1} h) \circ \varphi(x)) = L_{f_2} L_{f_1}(h \circ \varphi)(x)$. We immediately deduce that $L_{[\varphi_* f_1, \varphi_* f_2]} h(\varphi(x)) = L_{[f_1, f_2]} h(\varphi(x))$, which proves the Proposition.

The bracket$[f, g]$ has the following geometric interpretation: let us denote by $X_t(x) \stackrel{\mathrm{def}}{=} \exp t f(x)$, by analogy with the solution of a linear differential equation, the point of the integral curve of f at time t passing through the point x at time 0. This notation allows to precise which vector field is considered when several vector fields may be integrated. Thus, the point of the integral curve of g at time t passing through x at time 0 is noted $\exp t g(x)$. For ε sufficiently small, let us consider the expression

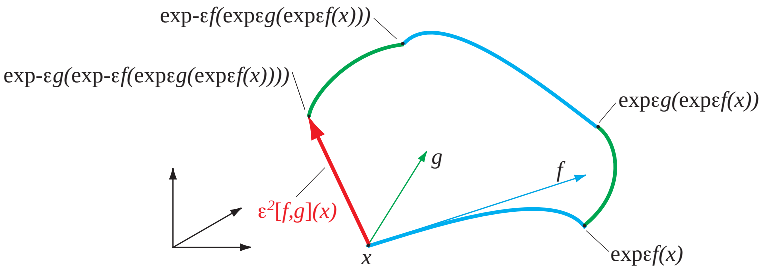

Fig. 2.5 Geometric interpretation of the bracket $[f, g]$. The vector field $[f, g]$ does not necessarily belong to the plane spanned by f and g.

$$\exp(-\varepsilon g) \circ \exp(-\varepsilon f) \circ \exp \varepsilon g \circ \exp \varepsilon f(x)$$

whose graphic representation is given in Fig.2.5. We now establish the following classical result (particular case of the well-known Baker-Campbell-Hausdorff formula):

Proposition 2.3.

$$\exp(-\varepsilon g) \circ \exp(-\varepsilon f) \circ \exp \varepsilon g \circ \exp \varepsilon f(x) = x + \varepsilon^2 [f, g](x) + 0(\varepsilon^3) \ .$$

Proof. Consider the differential equation $\dot{x} = f(x)$. The integral curve passing through x_0 at time $t = 0$ satisfies $x(t) = x_0 + \int_0^t f(x(s))ds$. We pose $F(t) = \int_0^t f(x(s))ds$. Its Taylor development for t sufficiently small is given by

$$F(t) = F(0) + tF'(0) + \frac{t^2}{2} F''(0) + 0(t^3)$$

with $F'(t) = f(x(t))$ and $F''(t) = \frac{\partial f}{\partial x}(x(t)) f(x(t))$. Thus $F(0) = 0$, $F'(0) = f(x_0)$ and $F''(0) = \frac{\partial f}{\partial x}(x_0) f(x_0)$. A Taylor development of $x(t)$ for $t = \varepsilon$ sufficiently small is thus given, at the order 2 with respect to ε, by

$$x(\varepsilon) = \exp \varepsilon f(x_0) = x_0 + \varepsilon f(x_0) + \frac{\varepsilon^2}{2} \frac{\partial f}{\partial x}(x_0) f(x_0) + 0(\varepsilon^3). \qquad (2.14)$$

By the same method, we get

$$\begin{aligned} x(2\varepsilon) &= \exp \varepsilon g \circ \exp \varepsilon f(x_0) = x_0 + \varepsilon \left(f(x_0) + g(x_0) \right) \\ &+ \frac{\varepsilon^2}{2} \left(\frac{\partial f}{\partial x}(x_0) f(x_0) + \frac{\partial g}{\partial x}(x_0) g(x_0) + 2 \frac{\partial g}{\partial x}(x_0) f(x_0) \right) + 0(\varepsilon^3), \qquad (2.15) \end{aligned}$$

then

$$\begin{aligned}x(3\varepsilon) &= \exp(-\varepsilon f)\circ\exp\varepsilon g\circ\exp\varepsilon f(x_0)\\ &= x_0+\varepsilon g(x_0)+\frac{\varepsilon^2}{2}\frac{\partial g}{\partial x}(x_0)g(x_0)\\ &+\varepsilon^2\left(\frac{\partial g}{\partial x}(x_0)f(x_0)-\frac{\partial f}{\partial x}(x_0)g(x_0)\right)+0(\varepsilon^3),\end{aligned} \tag{2.16}$$

and finally

$$\begin{aligned}x(4\varepsilon) &= \exp(-\varepsilon g)\circ\exp(-\varepsilon f)\circ\exp\varepsilon g\circ\exp\varepsilon f(x)\\ &= x_0+\varepsilon^2\left(\frac{\partial g}{\partial x}(x_0)f(x_0)-\frac{\partial f}{\partial x}(x_0)g(x_0)\right)+0(\varepsilon^3)\end{aligned} \tag{2.17}$$

which is the required result since $\frac{\partial g}{\partial x}(x_0)f(x_0)-\frac{\partial f}{\partial x}(x_0)g(x_0)$ is equal to the vector $[f,g](x_0)$, according to (2.12).

It results from this Proposition that if for every x in a given neighborhood we consider the vector subspace $E(x)$ of T_xX generated by $f(x)$ and $g(x)$, the bracket $[f,g](x)$ indicates if the integral curves of f and g remain close to E for t sufficiently small (which is the case if $[f,g]\in E$) or not ($[f,g]\notin E$). In particular, one may hope to find a submanifold Ξ of X for which $E(x)$ is its tangent space at every x of the considered neighborhood if $[f,g]\in E$, whereas $[f,g]\notin E$, seems to imply that such a submanifold doesn't exist, since the integral curves are going out of it.

Before addressing the problem of the existence of integral submanifolds, and consequently of the straightening out of a family of vector fields, we need to introduce the notion of distribution of vector fields.

2.2.7 Distribution of Vector Fields

Definition 2.10. Assume as before that X is a p-dimensional C^∞ manifold, and that for every $x\in X$, we are given a vector subspace $\mathcal{D}(x)$ of T_xX. A *distribution of vector fields* $\mathcal{D}$ is a mapping for which to every point $x\in X$ there corresponds the vector subspace $\mathcal{D}(x)$ of T_xX.

Let V be an open subset of X. The distribution $\mathcal{D}$ is said *regular with constant rank* $k\le p$ *in* V if there exists C^∞ vector fields $g_1,\dots,g_k$ such that:

- $\mathrm{rank}\,(g_1(x),\dots,g_k(x))=k$ for all $x\in V$.
- $\mathcal{D}(x)=\mathrm{span}\,\{g_1(x),\dots,g_k(x)\}$ for all $x\in V$.

The vector fields made up with the velocities and angular velocities of two rigid bodies rolling without slipping with permanent contact constitutes a classical example of distribution of vector fields. For instance, the velocity v of the contact point of a coin rolling without slipping on a fixed plane with coordinates $\{x_1,x_2\}$ must remain parallel to this plane: $v\in\mathrm{span}\left\{\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2}\right\}$.

Remark that, according to the previous definitions, the notion of distribution of vector fields is independent of the choice of coordinates.

Definition 2.11. The distribution $\mathcal{D}$ is said *involutive* if and only if for every pair of vector fields f and g in $\mathcal{D}$ we have $[f, g] \in \mathcal{D}$.

An involutive distribution is thus characterized by $[\mathcal{D}, \mathcal{D}] \subset \mathcal{D}$.

If $\mathcal{D}$ is not involutive, one can define its involutive closure:

Definition 2.12. The *involutive closure* $\overline{\mathcal{D}}$ of a distribution $\mathcal{D}$ is the smallest involutive distribution containing $\mathcal{D}$.

A constructive algorithm to compute $\overline{\mathcal{D}}$ exists and is based on the computation of the iterated brackets of the vector fields $g_1, \ldots, g_k$ (see Isidori [1995]).

Example 2.13. **(Non Involutive Distribution)** Consider the distribution $\mathcal{D}$ in $T\mathbb{R}^3$ generated by $g_1 = \frac{\partial}{\partial x_1}$ and $g_2 = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_3}$. Clearly, $[g_1, g_2] = \frac{\partial}{\partial x_3}$. The three vectors g_1, g_2 and $[g_1, g_2]$ are linearly independent, which proves that $\mathcal{D}$ is non involutive. Its involutive closure $\overline{\mathcal{D}}$ is equal to $T\mathbb{R}^3$. ■

2.2.8 Integral Manifolds

We now address the following problem:

"Given a regular distribution $\mathcal{D}$ with constant rank k, to which condition does there exist a diffeomorphism $\xi = \varphi(x)$ such that, in the ξ-coordinates, the image distribution $\varphi_* \mathcal{D}$ is given by $\varphi_* \mathcal{D} = \text{span} \left\{ \frac{\partial}{\partial \xi_1}, \ldots, \frac{\partial}{\partial \xi_k} \right\}$?"

This problem is often referred to as the problem of the existence of an *integral manifold* to the distribution $\mathcal{D}$, or as the *straightening out of the distribution* $\mathcal{D}$ by diffeomorphism. The answer is given by:

Theorem 2.4. ***(Frobenius) :*** *Let $\mathcal{D}$ be a regular distribution with constant rank k. A necessary and sufficient condition for the existence of a diffeomorphism that straightens out $\mathcal{D}$ is that $\mathcal{D}$ is involutive.*

In this case, by every point $x \in U$ open dense subset of X, there passes one submanifold of dimension k which is everywhere tangent to $\mathcal{D}$.

Such a submanifold of X everywhere tangent to $\mathcal{D}$ is called an *integral manifold* of $\mathcal{D}$.

Proof. If $\mathcal{D}$ is straightened out by diffeomorphism, its involutivity is clear: let φ be the corresponding diffeomorphism. If v_1 and v_2 are arbitrary vector fields in $\mathcal{D}$, they may be expressed as $\varphi_* v_i = \sum_{j=1}^{k} a_{i,j} \frac{\partial}{\partial \xi_j}$, $i = 1, 2$ and, according to Proposition 2.2, $\varphi_*[v_1, v_2] = [\varphi_* v_1, \varphi_* v_2]$. Using the fact that $[\frac{\partial}{\partial \xi_i}, \frac{\partial}{\partial \xi_j}] = 0$ for every i, j, one easily checks that $[\varphi_* v_1, \varphi_* v_2] = \sum_{i,j=1}^{k} [a_{1,i} \frac{\partial}{\partial \xi_i}, a_{2,j} \frac{\partial}{\partial \xi_j}] =$

$\sum_{i=1}^{k}\left(\sum_{j=1}^{k} a_{1,j}\frac{\partial a_{2,j}}{\partial \xi_i} - a_{2,j}\frac{\partial a_{1,j}}{\partial \xi_i}\right)\frac{\partial}{\partial \xi_i}$. We conclude that $\varphi_*[v_1, v_2]$ is a linear combination of $\frac{\partial}{\partial \xi_i}$, $i = 1, \ldots, k$, which proves that $[v_1, v_2] \in \mathcal{D}$. The involutivity follows.

The proof of the converse is by far more complicated. It uses the property of the Lie bracket shown in Fig. 2.5. The interested reader may refer to Chevalley [1946].

The second part of the theorem is a straightforward consequence of the implicit function Theorem: denoting by $\xi_i = \varphi_i(x)$ for every $i = 1, \ldots, p$, the set defined by $\xi_{k+1} = \xi^0_{k+1}, \ldots, \xi_p = \xi^0_p$, with $\xi^0 = (\xi^0_1, \ldots, \xi^0_p)$ in an open dense subset $\varphi(U)$ where the tangent mapping $D\varphi(x)$ has full rank equal to p, is clearly a k-dimensional submanifold of X, tangent to $\mathcal{D}$ at every point $(\xi_1, \ldots, \xi_k, \xi^0_{k+1}, \ldots, \xi^0_p) \in \varphi(U)$, and thus, by definition, an integral manifold of $\mathcal{D}$, for every ξ^0, *Q.E.D.*

2.2.9 First Order Partial Differential Equations

We now consider the application of the previous results to the integration of a set of k first order partial differential equations in a given open set V of the manifold X:

$$\begin{cases} L_{g_1} y = 0 \\ \vdots \\ L_{g_k} y = 0 \end{cases} \tag{2.18}$$

where y is the unknown function and where $g_1, \ldots, g_k$ are regular vector fields on X such that

$$\operatorname{rank}(g_1(x), \ldots, g_k(x)) = k$$

for all $x \in V$.

Let us introduce the distribution $\mathcal{D}$ given by $\mathcal{D}(x) = \operatorname{span}\{g_1(x), \ldots, g_k(x)\}$ for all $x \in V$. Clearly, if y is solution to the system (2.18), since $L_{[g_i,g_j]}y = L_{g_i}(L_{g_j}y) - L_{g_j}(L_{g_i}y) = 0 - 0 = 0$, we have $L_g y = 0$ for every $g \in \overline{\mathcal{D}}$. Conversely, since $g_1, \ldots, g_k$ are linearly independent, one can find regular vector fields $g_{k+1}, \ldots, g_r$ such that $g_1, \ldots, g_r$ form a basis of $\overline{\mathcal{D}}(x)$ at every point of V and $L_g y = 0$ for every $g \in \overline{\mathcal{D}}$ implies that $L_{g_i} y = 0$ for $i = 1, \ldots, k$ which proves that y is solution to the system (2.18).

By Frobenius Theorem, a necessary and sufficient condition for $\mathcal{D}$ to be straightened out in V is that $\mathcal{D}$ is involutive. Note that $\overline{\mathcal{D}}$ can always be straightened out, as far as it has constant rank in V. Assume, in a first step, that $\mathcal{D}$ est involutive. Then there exists a diffeomorphism φ such that, if we note $\xi = \varphi(x)$,

$$\varphi_*\mathcal{D}(\xi) = \operatorname{span}\{\varphi_* g_1(\xi), \ldots, \varphi_* g_k(\xi)\} = \operatorname{span}\left\{\frac{\partial}{\partial \xi_1}, \ldots, \frac{\partial}{\partial \xi_k}\right\}$$

in $\varphi(V)$. Setting $z = y \circ \varphi^{-1}$, we immediately check that $L_g y = 0$ for all $g \in \mathcal{D}$ is equivalent to

$$\frac{\partial z}{\partial \xi_1} = 0, \ldots, \frac{\partial z}{\partial \xi_k} = 0.$$

It results that z is independent of $(\xi_1, \ldots, \xi_k)$ or, otherwise stated, $z(\xi) = z(\overline{\xi})$ for all $\overline{\xi} = (\xi_{k+1}, \ldots, \xi_p)$ such that $\xi \in \varphi(V)$. The solution y is thus immediately deduced from z by $y = z \circ \varphi$.

In the case where $\mathcal{D}$ is not involutive and if $\overline{\mathcal{D}}$ has constant rank r, we follow the same lines with $\overline{\mathcal{D}}$ in place of $\mathcal{D}$. We nevertheless may remark that the function y is not completely determined by the equations (2.18). The supplementary equations generated by the vector fields of $\overline{\mathcal{D}}$ that are not in $\mathcal{D}$ are often called *compatibility conditions*. It results that $\mathcal{D}$ cannot be straightened out without simultaneously straightening out $\overline{\mathcal{D}}$, which implies that $z(\xi) = z(\overline{\xi})$ for all $\overline{\xi} = (\xi_{r+1}, \ldots, \xi_p)$ such that $\xi \in \varphi(V)$.

Let us also remark that it often happens that the involutive closure of $\mathcal{D}$ is the tangent space of the whole $\mathbb{R}^p$, which implies that the only possible solution y is a constant in V.

We have thus shown that solving a set of first order partial differential equations is equivalent to the straightening out of the corresponding distribution of vector fields, or of its involutive closure if need be. However, the straightening out diffeomorphism may be itself obtained by solving the system $L_{\gamma_i}\varphi_j = \delta_{i,j}$ for $i = 1, \ldots, k$, $j = 1, \ldots, p$, with $\{\gamma_1, \ldots, \gamma_k\}$ a regular basis of $\overline{\mathcal{D}}$ and $\delta_{i,j}$ the Kronecker symbol. Therefore, this method only applies in practice if we know a particular basis of $\overline{\mathcal{D}}$ for which the expression of the diffeomorphism is simpler. There also exists other constructive methods as the method of bicharacteristics. A thorough review of this question may be found, e.g., in Arnold [1974, 1980] .

Example 2.14. **(Non Involutive Distribution, continued)** Consider the distribution $\mathcal{D}$ of example 2.13, generated by $g_1 = \frac{\partial}{\partial x_1}$ and $g_2 = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_3}$. This distribution being non involutive, the set of first order partial differential equations

$$\frac{\partial y}{\partial x_1} = 0, \quad \frac{\partial y}{\partial x_2} + x_1 \frac{\partial y}{\partial x_3} = 0$$

is equivalent to

$$\frac{\partial y}{\partial x_1} = 0, \quad \frac{\partial y}{\partial x_2} = 0, \quad \frac{\partial y}{\partial x_3} = 0,$$

which means that $y(x_1, x_2, x_3) = y_0$ with y_0 arbitrary constant. ■

2.3 Differential Forms

In this section, our aim is to give a dual form of Frobenius Theorem, stated in section 2.2.8 for a family of vector fields, in terms of the so-called *complete integrability* of a family of 1-forms.

For this purpose, we need to introduce the notions of *differential form*, *exterior derivative of a differential form*, *reciprocal image of a differential form*, or *induced differential form*, or, more shortly, *image of a differential form*, of *Lie derivative* of differential forms, *codistributions* and *exterior differential systems.*

2.3.1 Cotangent Space, Differential Form, Duality

Let us consider the distribution $\mathrm{T}X$ for which to every point of the manifold X there corresponds the tangent space T_xX. It is immediately seen that $\mathrm{T}X$ is involutive (by definition of the tangent space) and according to Frobenius Theorem, one can find local coordinates such that a basis of $\mathrm{T}X$ is given by $\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_p}\}$. Let us introduce the dual space of T_xX, noted T_x^*X and called the *cotangent space* to the manifold X at the point x, by defining a dual basis $\{dx_1, \ldots, dx_p\}$ by the duality pairing

$$< \frac{\partial}{\partial x_i}, dx_j >= \delta_{i,j} \quad , \quad \forall i, j = 1, \ldots, p \tag{2.19}$$

where $\delta_{i,j}$ is the usual Kronecker δ-symbol ($\delta_{i,j} = 1$ if $i = j$ and 0 otherwise).

The *cotangent bundle*, noted T^*X, is thus defined by $\mathrm{T}^*X = \bigcup_{x\in X} \mathrm{T}_x^*X$.

If we now consider a function h of class C^∞ from X to $\mathbb{R}$, one can define its differential by expressing it in the above basis as $dh = \sum_{i=1}^{p} \frac{\partial h}{\partial x_i} dx_i$. Note that, according to this latter formula, the differential of a function may be identified to its tangent linear mapping. Thus, at every point $x \in X$, $dh(x)$ belongs to T_x^*X and if f is a vector field ($f(x) \in \mathrm{T}_xX$ for all x), the value of the duality pairing $< dh, f >$ is deduced from the previous formula by

$$< dh, f >= \sum_{i,j=1}^{p} f_i \frac{\partial h}{\partial x_j} < \frac{\partial}{\partial x_i}, dx_j >= \sum_{i=1}^{p} f_i \frac{\partial h}{\partial x_i} = L_f h. \tag{2.20}$$

Example 2.15. Consider again the function $h(t, x) = x - vt$ with $v \in \mathbb{R}$ a given constant. We have $dh = -vdt + dx$ and, if $f = \frac{\partial}{\partial t} + v\frac{\partial}{\partial x}$, we get $< dh, f >= -v + v = 0$. We therefore recover the result of example 2.7.

In fact, the differential dh of h may be more “globally” defined, *i.e.* without reference to its value $dh(x)$ at each point $x \in X$.

To this aim, let us define the projection π from T^*X to X by the formula $\pi(dh(x)) = x$ for all $x \in X$. Then $dh(x) = \sum_{i=1}^{p} \frac{\partial h}{\partial x_i}(x)dx_i \in \pi^{-1}(x)$ for all x, its coefficients $\frac{\partial h}{\partial x_i}(x)$ being C^∞ functions on X. We say that dh is a C^∞-*section* of the cotangent bundle T^*X relatively to the canonical projection π.

One easily verifies that this definition is independent of the choice of coordinates of X and, in local coordinates, coincides with the above one.

More generally:

Definition 2.13. We call *differential form of degree 1*, or 1-form, a C^∞-section ω of the cotangent bundle T^*X, *i.e.* a mapping for which, to each point $x \in X$, there corresponds an element $\omega(x) \in \mathrm{T}^*_xX$, $\omega(x)$ being a linear combination of the local basis covectors of T^*_xX with C^∞ coefficients on X. The set of C^∞-sections of T^*X is a vector space noted $\Lambda^1(X)$.

The duality pairing between a 1-form $\omega = \sum_{i=1}^{p} \omega_i dx_i$ and a vector field $f = \sum_{i=1}^{p} f_i \frac{\partial}{\partial x_i}$ is given by

$$< \omega, f >= \sum_{i=1}^{p} f_i \omega_i. \tag{2.21}$$

A 1-form is not generally the differential of a function, as we now show, and consequently, $\Lambda^1(X)$ contains more than the differentials of functions.

Definition 2.14. A 1-form ω is said *exact* if there exists a differentiable function h such that $\omega = dh$.

A sufficient condition is given by:

Proposition 2.4. *For a 1-form ω to be exact, it is necessary that ω satisfies*

$$\frac{\partial \omega_i}{\partial x_j} = \frac{\partial \omega_j}{\partial x_i} \quad \forall i, j. \tag{2.22}$$

Proof. Clearly, if $\omega = \sum_{i=0}^{p} \omega_i dx_i = dh = \sum_{i=1}^{p} \frac{\partial h}{\partial x_i} dx_i$, we have $\omega_i = \frac{\partial h}{\partial x_i}$ for all i and, differentiating with respect to x_j, we get:

$$\frac{\partial \omega_i}{\partial x_j} = \frac{\partial^2 h}{\partial x_j \partial x_i} = \frac{\partial^2 h}{\partial x_i \partial x_j} = \frac{\partial \omega_j}{\partial x_i}$$

which proves the Proposition.

Example 2.16. Consider the differential form $\omega = \frac{xdy - ydx}{x^2+y^2} = \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy$. Set $\omega_x = \frac{-y}{x^2+y^2}$ and $\omega_y = \frac{x}{x^2+y^2}$. One easily sees that $\frac{\partial \omega_x}{\partial y} = \frac{\partial \omega_y}{\partial x}$ and that $h(x, y) = \arctan\left(\frac{y}{x}\right)$ satisfies $dh = \omega$. However, h and the 1-form ω are not defined at the origin and one can prove that ω doesn't have a "primitive" in a neighborhood of $(0, 0)$ (see the Theorem 2.5 below, often called Poincaré's Lemma).

We also can define differential forms of degree 2 (and more generally of degree greater than 1) by introducing the skew-symmetric (and more generally alternating) tensor product of cotangent spaces, called *wedge product* $\mathrm{T}_x^* X \wedge \mathrm{T}_x^* X$ whose basis (in local coordinates) is formed by the products $dx_i \wedge dx_j$ for $i < j$, with

$$dx_i \wedge dx_j = -dx_j \wedge dx_i, \quad \forall i, j.$$

The space $\mathrm{T}_x^* X \wedge \mathrm{T}_x^* X$ may be seen as the dual of $\mathrm{T}_x X \wedge \mathrm{T}_x X$, the associated pairing being defined by

$$\begin{aligned} < dx_i \wedge dx_j, \frac{\partial}{\partial x_k} \wedge \frac{\partial}{\partial x_l} > &= \det \begin{pmatrix} < dx_i, \frac{\partial}{\partial x_k} > & < dx_i, \frac{\partial}{\partial x_l} > \\ < dx_j, \frac{\partial}{\partial x_k} > & < dx_j, \frac{\partial}{\partial x_l} > \end{pmatrix} \\ &= \delta_{i,k}\delta_{j,l} - \delta_{j,k}\delta_{i,l} \end{aligned}$$

$\delta_{i,j}$ being the Kronecker δ-symbol.

A 2-form $\varpi(x)$ at the point x thus reads, in these coordinates:

$$\varpi(x) = \sum_{i,j} \varpi_{i,j}(x) dx_i \wedge dx_j$$

with $\varpi_{i,j} \in C^\infty(X)$ for all i, j and the pairing between ϖ and a 2-vector $f \wedge g$ with $f = \sum_{k=1}^p f_k \frac{\partial}{\partial x_k}$ and $g = \sum_{l=1}^p g_l \frac{\partial}{\partial x_l}$, according to the bilinearity and skew-symmetry of the product, is

$$< \varpi, f \wedge g >= \sum_{i,j} \varpi_{i,j} \det \begin{pmatrix} f_i & g_i \\ f_j & g_j \end{pmatrix}.$$

As before, a 2-form is a C^∞-section of $\mathrm{T}^* X \wedge \mathrm{T}^* X$, *i.e.* a mapping for which to every $x \in X$ there corresponds $\varpi(x) \in \mathrm{T}_x^* X \wedge \mathrm{T}_x^* X$, linear combination of the basis covectors of $\mathrm{T}_x^* X \wedge \mathrm{T}_x^* X$ with C^∞ coefficients. The vector space of the 2-forms on X is denoted by $\Lambda^2(X)$.

By a similar construction, we define the space $\Lambda^k(X)$ of k-forms on X as the set of C^∞-sections of the alternating tensor product of $\mathrm{T}^* X \wedge \ldots \wedge \mathrm{T}^* X$ made of k copies of $\mathrm{T}^* X$, for $k \in \mathbb{N}$. A k-form is thus a mapping for which to every $x \in X$ there corresponds the k-covector

$$\varpi(x) = \sum_{i_1, i_2, \ldots, i_k} \varpi_{i_1, i_2, \ldots, i_k}(x) dx_{i_1} \wedge dx_{i_2} \wedge \ldots \wedge dx_{i_k}$$

with $\varpi_{i_1, i_2, \ldots, i_k} \in C^\infty$ for all $i_1, i_2, \cdots, i_k$.

Again, the pairing

$$< dx_{i_1} \wedge \ldots \wedge dx_{i_k}, \frac{\partial}{\partial x_{j_1}} \wedge \ldots \wedge \frac{\partial}{\partial x_{j_k}} >$$
$$= \det \begin{pmatrix} < dx_{i_1}, \frac{\partial}{\partial x_{j_1}} > \ldots < dx_{i_1}, \frac{\partial}{\partial x_{j_k}} > \\ \vdots & \vdots \\ < dx_{i_k}, \frac{\partial}{\partial x_{j_1}} > \ldots < dx_{i_k}, \frac{\partial}{\partial x_{j_k}} > \end{pmatrix} \quad (2.23)$$

makes $\mathrm{T}_x^* X \wedge \ldots \wedge \mathrm{T}_x^* X$ the dual space of $\mathrm{T}_x X \wedge \ldots \wedge \mathrm{T}_x X$.

Moreover, if $\omega = \sum_{i_1,\ldots,i_k} \omega_{i_1,\ldots,i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k} \in \Lambda^k(X)$ and $\theta = \sum_{i_1,\ldots,i_r} \theta_{i_1,\ldots,i_r} dx_{i_1} \wedge \ldots \wedge dx_{i_r} \in \Lambda^r(X)$, according to the previous rules associated to the wedge product, we have

$$\omega \wedge \theta = \sum_{i_1,\ldots,i_k,j_1,\ldots,j_r} \omega_{i_1,\ldots,i_k} \theta_{j_1,\ldots,j_r} dx_{i_1} \wedge \ldots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \ldots \wedge dx_{j_r} \in \Lambda^{k+r}(X).$$

An obvious consequence of the skew-symmetry is that every form of degree higher than p on a p-dimensional manifold vanishes identically.

2.3.2 Exterior differentiation

Some 2-forms may also be deduced from 1-forms by the *exterior derivative* operator d from $\Lambda^1(X)$ to $\Lambda^2(X)$, by the formula

$$d\omega = \sum_{i,j} \frac{\partial \omega_i}{\partial x_j} dx_j \wedge dx_i = \sum_{i<j} \left(\frac{\partial \omega_j}{\partial x_i} - \frac{\partial \omega_i}{\partial x_j} \right) dx_i \wedge dx_j \quad (2.24)$$

where ω is a 1-form defined, as before, by $\omega = \sum_i \omega_i dx_i$. The previous formula shows that if h is a C^∞ function from X to $\mathbb{R}$ and if $\omega \in \Lambda^1(X)$, the (usual) product $h\omega$ belongs to $\Lambda^1(X)$ and the exterior derivative of this product satisfies $d(h\omega) = dh \wedge \omega + hd\omega$.

The exterior derivative operator d may be extended to an operator from $\Lambda^2(X)$ to $\Lambda^3(X)$ by the following *anti-derivation* formula:

$$d(\omega \wedge \theta) = d\omega \wedge \theta - \omega \wedge d\theta \quad (2.25)$$

for all $\omega, \theta \in \Lambda^1(X)$, since then $d\omega \wedge \theta$ and $\omega \wedge d\theta$ are 3-forms.

This results from the fact that $\omega \wedge \theta = \left(\sum_j \omega_j dx_j\right) \wedge \left(\sum_k \theta_k dx_k\right) = \sum_{j,k} \omega_j \theta_k dx_j \wedge dx_k$. By the exterior differentiation formula (2.24), we get $d(\omega \wedge \theta) = \sum_{i,j,k} \left(\frac{\partial \omega_j}{\partial x_i} \theta_k + \omega_j \frac{\partial \theta_j}{\partial x_i} \right) dx_i \wedge dx_j \wedge dx_k$.

Moreover, $d\omega \wedge \theta = \left(\sum_{i,j} \frac{\partial \omega_j}{\partial x_i} dx_i \wedge dx_j \right) \wedge \left(\sum_k \theta_k dx_k\right) = \sum_{i,j,k} \frac{\partial \omega_j}{\partial x_i} \theta_k dx_i \wedge dx_j \wedge dx_k$.

Accordingly, $\omega \wedge d\theta = \left(\sum_j \omega_j dx_j\right) \wedge \left(\sum_{i,k} \frac{\partial \theta_k}{\partial x_i} dx_i \wedge dx_k \right) =$

$\sum_{i,j,k} \omega_j \frac{\partial \theta_k}{\partial x_i} dx_j \wedge dx_i \wedge dx_k = -\sum_{i,j,k} \omega_j \frac{\partial \theta_k}{\partial x_i} dx_i \wedge dx_j \wedge dx_k$ by skew-symmetry. The formula (2.25) is thus immediately deduced by subtracting these two latter expressions.

More generally, d may be seen as an operator from $\Lambda^{p+q}(X)$ to $\Lambda^{p+q+1}(X)$ for all integers p and q that is additive ($d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$) and such that

$$d(\omega \wedge \theta) = d\omega \wedge \theta + (-1)^p \omega \wedge d\theta$$

for all $\omega \in \Lambda^p(X)$ et $\theta \in \Lambda^q(X)$.

Let us go back to condition (2.22), which is now clearly equivalent to:

$$d\omega = 0. \tag{2.26}$$

Remark that this latter expression is now independent of the choice of coordinates.

We also can deduce from what precedes that every 1-form ω satisfies $d^2\omega = d(d\omega) = 0$.

Definition 2.15. A 1-form satisfying (2.26) is called a *closed form.*

The converse of Proposition 2.4 only holds true under topological restrictions:

Theorem 2.5. (Poincaré's Lemma) *A closed form on a contractible*[4] *open set is exact.*

Example 2.17. The 1-form $\omega = \frac{xdy - ydx}{x^2+y^2}$ of example 2.16 is defined on the non contractible open set $\mathbb{R}^2 \setminus \{(0,0)\}$. It is therefore non exact in any neighborhood of the origin, though it is exact in a neighborhood of any other point of the plane $\mathbb{R}^2$ that doesn't contain the origin.

2.3.3 Image of a Differential Form

Definition 2.16. Given a 1-form ω on the n-dimensional manifold X and a local diffeomorphism φ from an n-dimensional manifold Y to X, we call *reciprocal image of* ω *by* φ, or induced form, noted $\varphi^*\omega$, the 1-form on Y defined by

$$\langle \varphi^*\omega, v \rangle = \langle \omega, \varphi_* v \rangle \tag{2.27}$$

for every vector field $v \in \mathrm{T}Y$.

[4] An open set U is contractible if there exists a point $x_U \in U$ and a continuous mapping γ from $U \times [0,1]$ to U that "continuously distorts" U to the point x_U, *i.e.* such that $\gamma(\cdot, 0) = Id_U$ ($\gamma(x,0) = x \; \forall x \in U$) and $\gamma(x,1) = x_U$ for all $x \in U$.

In local coordinates $(x_1, \ldots, x_p)$ of X and $(y_1, \ldots, y_p)$ of Y, with $\omega = \sum_{i=1}^{p} \omega_i dx_i$, the previous definition yields:

$$\varphi^* \omega = \sum_{i,j=1}^{p} (\omega_i \circ \varphi) \frac{\partial \varphi_i}{\partial y_j} dy_j. \tag{2.28}$$

Remark that this formula may be extended to an arbitrary differentiable mapping φ (not necessarily a diffeomorphism) and for a manifold Y of dimension r arbitrary:

$$\varphi^* \omega = \sum_{j=1}^{r} \sum_{i=1}^{p} (\omega_i \circ \varphi) \frac{\partial \varphi_i}{\partial y_j} dy_j$$

this latter definition being in accordance with (2.27) if we also slightly extend the definition of the image of a vector field $v = \sum_{j=1}^{r} v_j(y) \frac{\partial}{\partial y_j}$ on Y by

$$\varphi_* v_{|\varphi(Y)} = \sum_{i=1}^{p} (L_v \varphi_i)_{|\varphi(Y)} \frac{\partial}{\partial x_i}.$$

This definition also easily extends to higher degree forms by:

$$\varphi^* (\omega \wedge \theta) = \varphi^* \omega \wedge \varphi^* \theta$$

for every 1-forms ω and θ.

We then easily check that

Proposition 2.5. *For every form ω on X and every differentiable mapping φ from Y to X we have*

$$d\varphi^* \omega = \varphi^* d\omega.$$

2.3.4 Pfaffian System, Complete Integrability

Definition 2.17. Given a collection of r independent 1-forms[5] $\{\omega_1, \ldots, \omega_r\}$, a manifold Y is called *integral manifold* of the exterior differential system

$$\omega_1 = 0, \quad \ldots, \quad \omega_r = 0 \tag{2.29}$$

if and only if there exists a differentiable mapping φ from Y to X such that $\varphi^* \omega_1$, ..., $\varphi^* \omega_r$ identically vanish on Y.

Two exterior differential systems are said *algebraically equivalent* if and only if any form of the first system can be expressed as a linear combination

[5] such a collection of 1-forms is often called a *codistribution* in reference to the duality with distributions of vector fields.

of the forms of the second one and conversely. More precisely, consider two exterior differential systems defined by the forms $(\omega_1, \ldots, \omega_r)$ and $(\omega'_1, \ldots, \omega'_{r'})$ where ω_i and ω'_i have arbitrary degrees. ω_i is a linear combination of the ω'_j's if and only if there exists r' forms $\theta_1, \ldots, \theta_{r'}$ such that $\omega_i = \sum_{j=1}^{r'} \omega'_j \wedge \theta_j$.

When the system (2.29) is only made of 1-forms, it is called *Pfaffian.*

Definition 2.18. The (algebraic) closure of the system (2.29) is the exterior differential system

$$\omega_1 = 0, \quad d\omega_1 = 0, \quad \ldots, \quad \omega_r = 0, \quad d\omega_r = 0. \tag{2.30}$$

We say that the system (2.29) is (algebraically) closed if and only if it is algebraically equivalent to its closure.

The next Proposition is obvious:

Proposition 2.6. *Every integral manifold of the Pfaffian system (2.29) is also an integral manifold of its closure (2.30).*

Definition 2.19. The exterior differential system (2.29) of rank r is said *completely integrable* if and only if there exists an open set U of X and r independent differentiable functions $y_1, \ldots, y_r$ such that (2.29) is algebraically equivalent in U to

$$dy_1 = 0, \ldots, dy_r = 0.$$

In other words, if it locally admits r independent first integrals.

We thus have the dual version of Frobenius Theorem:

Theorem 2.6. *A necessary and sufficient condition for the Pfaffian system (2.29) of rank r to be completely integrable in an open set $U \subset X$ is that it is algebraically closed in U.*

In this case, there exists an open dense $U_1 \subset U$ such that by every point of U_1 there passes a local integral manifold of (2.29) of dimension $p - r$ which is a submanifold of X.

Proof. We only sketch the simplest parts of the proof, namely those concerning the necessary condition and the second part of the Theorem. For the sufficiency part, the interested reader may refer to Choquet-Bruhat [1968].

If the Pfaffian system (2.29) is completely integrable, there exists r locally independent first integrals $y_1, \ldots, y_r$ such that $\omega_i = \sum_{j=1}^{r} a_{i,j} dy_j$ and conversely $dy_j = \sum_{i=1}^{r} b_{j,i}\omega_i$. Thus we get

$$\begin{aligned} d\omega_i &= \sum_{j=1}^{r} \left(da_{i,j} \wedge dy_j + a_{i,j} d^2 y_j \right) = \sum_{j=1}^{r} da_{i,j} \wedge dy_j \\ &= \sum_{j=1}^{r} da_{i,j} \wedge \left(\sum_{k=1}^{r} b_{j,k}\omega_k \right) = \sum_{k=1}^{r} \theta_{i,k} \wedge \omega_k \end{aligned}$$

with $\theta_{i,k} = \sum_{j=1}^{r} b_{j,k} da_{i,j}$, which proves that the Pfaffian system (2.29) is algebraically equivalent to its closure and is thus closed.

Concerning the second part, if the system is completely integrable, it admits r locally independent first integrals $y_1, \ldots, y_r$. Let U_1 be an open dense subset of U in which the Jacobian matrix $\frac{\partial y_i}{\partial x_j}$ has full rank equal to r. For every $x_0 \in U_1$, the system $dy_1 = \cdots = dy_r = 0$ is equivalent to the implicit system made of the r independent equations $y_1(x) = y_1(x_0), \ldots, y_r(x) = y_r(x_0)$, thus admitting a local solution in U, which constitutes a submanifold of X of dimension $n - r$, and the result is proven.

Let us now compare the notions of integrability that result from Theorems 2.6 and 2.4: for every vector field v of $\mathrm{T}Y$, where Y is an integral manifold of the Pfaffian system (2.29), and if φ is the corresponding differentiable mapping from Y to X, we have:

$$\langle \varphi^* \omega_i, v \rangle = 0 = \langle \omega_i, \varphi_* v \rangle .$$

Thus the set of vector fields $\varphi_* v$ for $v \in \mathrm{T}Y$ defines a distribution $\mathcal{D}$ such that $\langle dy_i, w \rangle = 0$ for all $i = 1, \ldots, r$ and all $w \in \mathcal{D}$. This distribution thus admits a basis at every regular point, noted $\left(\frac{\partial}{\partial \xi_1}, \ldots, \frac{\partial}{\partial \xi_{p-r}}\right)$, supplementary of $\left(\frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_r}\right)$, involutive by construction, and such that the manifold Y, everywhere tangent to $\mathcal{D}$, is effectively an integral manifold in the sense of Theorem 2.4.

2.3.5 Lie Derivative of a 1-Form

Let us close this introduction to differential geometry by a generalization to differential forms of the notion of Lie derivative.

We want to define, on T^*X, an operation that may be seen as the dual of the Lie bracket, called the *Lie derivative of a 1-form.* We proceed in several steps, starting first from the Lie derivative of the differential of a function (closed form).

The exterior derivative of $L_f h$ is given by:

$$dL_f h = \sum_{i=1}^{n} \frac{\partial L_f h}{\partial x_i} dx_i = \sum_{i=1}^{n} \left(L_f \left(\frac{\partial h}{\partial x_i} \right) + < dh, \frac{\partial f}{\partial x_i} > \right) dx_i. \tag{2.31}$$

This formula defines a new 1-form, noted $L_f dh$, called Lie derivative of dh along f. If g is another vector field, one easily verifies that:

$$< L_f dh, g >= \sum_{i=1}^{n} (L_f(\frac{\partial h}{\partial x_i}) + < dh, \frac{\partial f}{\partial x_i} >) g_i$$
$$= L_f < g, dh > - < [f, g], dh > . \quad (2.32)$$

To define the Lie derivative of a general 1-form, we extend this relation:

Definition 2.20. The Lie derivative of a 1-form ω is defined by

$$< L_f \omega, g >= L_f < \omega, g > - < \omega, [f, g] > . \quad (2.33)$$

for every pair of vector fields f and g.

An easy computation shows that

$$L_f \omega = \sum_{i=1}^{n} \left(L_f(\omega_i) + < \omega, \frac{\partial f}{\partial x_i} > \right) dx_i. \quad (2.34)$$

If we also call the Lie bracket of f and g the Lie derivative of g along f, $L_f g \overset{\text{def}}{=} [f, g]$ and if we rewrite (2.33):

$$L_f < \omega, g >=< L_f \omega, g > + < \omega, L_f g >$$

we obtain the usual rule of differentiation of a bilinear form.

As previously seen, according to the pairing between 2-forms and pairs of vector fields, we have:

$$< d\omega, f \wedge g > = \sum_{i,j} \frac{\partial \omega_j}{\partial x_i} \det \begin{pmatrix} f_i & g_i \\ f_j & g_j \end{pmatrix} = \sum_{i,j} \frac{\partial \omega_j}{\partial x_i} (f_i g_j - f_j g_i)$$
$$= \sum_{i,j} \left(\frac{\partial \omega_j}{\partial x_i} - \frac{\partial \omega_i}{\partial x_j} \right) f_i g_j$$

We thus have the following result:

Proposition 2.7. *For every pair (f, g) of vector fields and every 1-form ω, we have*

$$< d\omega, f \wedge g >= L_f < \omega, g > - L_g < \omega, f > + < \omega, [f, g] > . \quad (2.35)$$

Proof. We have

$$< d\omega, f \wedge g >= \sum_{i,j} \left(\frac{\partial \omega_j}{\partial x_i} - \frac{\partial \omega_i}{\partial x_j} \right) f_i g_j ,$$

$$L_f < \omega, g >= \sum_{i,j} \left(f_i \frac{\partial \omega_j}{\partial x_i} g_j + f_i \omega_j \frac{\partial g_j}{\partial x_i} \right) ,$$

$$L_g < \omega, f >= \sum_{i,j} \left(g_i \frac{\partial \omega_j}{\partial x_i} f_j + g_i \omega_j \frac{\partial f_j}{\partial x_i} \right),$$

$$< \omega, [f, g] >= \sum_{i,j} \omega_i \left(f_j \frac{\partial g_i}{\partial x_j} - g_j \frac{\partial f_i}{\partial x_j} \right),$$

which proves the result by combining the three last relations and comparing with the first one.

2.3.6 Back to Frobenius Theorem

We deduce another version of Frobenius Theorem:

Corollary 2.1. *If the Pfaffian system (2.29) is completely integrable, and if f and g are two tangent vector fields to an integral manifold, then $[f, g]$ is tangent to the same integral manifold.*

Conversely, if $\mathcal{D}$ is an involutive distribution of vector fields and ω a 1-form vanishing on $\mathcal{D}$, then ω is closed.

Proof. This is a direct application of (2.35): if (2.29) is completely integrable, if f and g are tangent to an integral manifold, every 1-form ω of the system is an annihilator of f and g, *i.e.* $< \omega, f >=< \omega, g >= 0$, and the same for $d\omega$ by closedness of the Pfaffian system. We deduce that $< \omega, [f, g] >= 0$ for every 1-form of (2.29) and thus $[f, g]$ is tangent to the same integral manifold.

Conversely, if ω annihilates f and g, for arbitrary f and g in $\mathcal{D}$, according to the involutivity of $\mathcal{D}$, ω also annihilates $[f, g]$, and thus $< d\omega, f \wedge g >= 0$ for every pair $f, g \in \mathcal{D}$. The 2-form $d\omega$ vanishes also on $\mathcal{D}$, which proves that ω is closed.

Chapter 3
Introduction to Dynamical Systems

This chapter is devoted to the study of the dynamical behaviors of nonlinear uncontrolled systems: stability, instability of flows around an equilibrium or a periodic orbit and comparison to their tangent linear approximation.

We consider the set of (uncontrolled) differential equation, or differential system

$$\dot{x} = f(x) \tag{3.1}$$

on a manifold X of class C^∞ and of dimension n, f being a vector field of class C^∞[1]

In analogy to physics, we will sometimes interpret a vector field f as a field of velocities and $f(x)$ will be sometimes called the *velocity vector* at the point x.

We will see later that, for controlled systems, once the control law has been chosen, we may end up with a system of that form.

The set of equations (3.1) is said *stationary* or *time-invariant*, which means that the velocity vector $f(x)$ at every point x doesn't depend on the passing time: if an integral curve of (3.1) passes through x at two different instants, the velocity vector $f(x)$ will be the same.

Remark 3.1. The case of *non stationary* or *time-varying* differential equations, *i.e.* for which f depends on time:

$$\dot{x} = f(t, x) \tag{3.2}$$

will be addressed later.

---

[1] In the previous chapter, we had to distinguish between the dimension n of the "external vector space" where the equation of the manifold was stated, and p, the dimension of the manifold. Here, we only consider systems defined by vector fields on manifolds without reference to the "external space" which doesn't appear anymore. Therefore, the introduction of two different symbols, namely n and p, is no more necessary. *From now on, in order to comply with the usual conventions and notations in Control Theory, we use n as the dimension of the manifold X, and therefore of the vector field f too.*

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_3,

Recall from section 2.2.2 that an instationary vector field on a manifold X may be seen as a stationary one on the augmented manifold $X \times \mathbb{R}$. However, we'll see that many properties of stationary vector fields do not extend to time-varying ones. In particular, an arbitrary change of coordinates would change (x, t) in $(z, s) = \varphi(x, t)$, s playing the role of a new clock, the sign of $\dot{s}$ being arbitrary. Therefore, if the sign of $\dot{s}$ changes along an integral curve, we may lose the possibility of capturing the asymptotic properties of the system (when $t \to \pm\infty$) in the transformed coordinates.

We first study the flow associated to (3.1), its orbits and its associated phase portrait and introduce the notion of *singular* or *equilibrium point*. In the case of closed orbits, or periodic orbits, we introduce the notion of *Poincaré's mapping*, or *first return mapping*, for which a periodic orbit is a fixed point. We also recall basic facts on *Lyapunov* and *Chetaev functions*.

In a second step, we study the qualitative behavior of the solutions of (3.1) around an equilibrium point. We recall the classification of linear systems around an equilibrium point in function of their eigenvalues, and we discuss the following question:
"Can we compare the asymptotic behavior of a nonlinear system around an equilibrium point to the one of its tangent linear approximation?"
To precise the word *compare*, it is important to remark that the stability or instability is not affected if we *distort* the trajectories in a differentiable way. The previous question may thus be reformulated as follows:
"At which condition does there exist a smooth change of coordinates that locally transforms the trajectories of the nonlinear system into the trajectories of its linear tangent approximation?"

Note that our concern is of the same nature as in the previous chapter since we aim at revealing some coordinates that make the analysis (stability or instability) particularly easy.

If these changes of coordinates are chosen in the class of homeomorphisms, the classification is provided by the Hartman-Grobman Theorem under the so-called *hyperbolicity* assumption. Finally, for a *non hyperbolic* system, the Shoshitaishvili Theorem shows that the stability or instability is not determined by its linear tangent approximation only, but depends on the stability or instability of its projection on its *center manifold*.

The proofs of the main results, generally difficult, are omitted. For more details, the reader may refer to Anosov and Arnold [1980], Arnold [1974, 1980], Carr [1981], Demazure [2000], Fenichel [1979], Guckenheimer and Holmes [1983], Hirsch and Smale [1974], Hirsch et al. [1977], LaSalle and Lefschetz [1961], Liapounoff [1907], Milnor [1978], Ruelle [1989], Saunders and Verhulst [1987], Thom [1977], Tikhonov et al. [1980]. Numerous examples of application to Mechanics may be found in Anosov and Arnold [1980], Arnold [1974, 1980], Liapounoff [1907].

3.1 Recalls on Flows and Orbits

Recall (Proposition 2.1) that, given a C^∞ vector field f in a neighborhood of a *regular* or *transient point* x_0, namely such that $f(x_0) \neq 0$, there exists a local diffeomorphism that straightens out the vector field f. The transformed flow is given by

$$z_1(t) = z_1^0, \quad \ldots, \quad z_{n-1}(t) = z_{n-1}^0, \quad z_n(t) = t + z_n^0$$

with

$$z_i^0 = \varphi_i(x_0), \quad i = 1, \ldots, n$$

and thus

$$x(t) = \varphi^{-1}(\varphi_1(x_0), \ldots, \varphi_{n-1}(x_0), t + \varphi_n(x_0)).$$

In the time-varying case (3.2), one can prove that the straightening out diffeomorphism may be chosen in such a way that it respects the time orientation.

Consequently, in the neighborhood of a regular point, all the transient behaviors (of arbitrary nonlinear systems in n dimensions) are equivalent, namely they can be described by $n-1$ constants (or n in the instationary case) and an affine function with respect to time. It also amounts to say that all their orbits are parallel straight half-lines (see Fig 2.4). It results that all the nonlinear complexity may be found elsewhere, namely in neighborhoods of singular points, as opposed to regular ones, *i.e.* satisfying $f(x_0) = 0$. These points are also often called *equilibrium points*.
There are also more global aspects in the case of *periodic orbits* since the straightening out diffeomorphism in this case cannot be defined over a whole period, since then a straight line would have to be a closed curve. In fact, we will see later on that periodic orbits are fixed points of the *Poincaré's mapping*, thus playing an analogous role to the one of equilibrium points.

3.1.1 Equilibrium Point, Variational Equation

We are now interested in the study of the so-called *asymptotic* phenomena. More precisely we consider the flow of (3.1) on an unbounded interval of time and we study its attraction or repulsion with respect to some limit points, orbits or sets.

A *singular point* or *equilibrium point* of the vector field f or, by extension, of system (3.1), is a point $\bar{x}$ such that $f(\bar{x}) = 0$, or, equivalently, such that $X_t(\bar{x}) = \bar{x}$ (fixed point of the flow).

A vector field doesn't necessarily admit an equilibrium point: this is the case, for instance, of a constant vector field $f(x) = 1$, or more generally of a time-varying vector field $(f(t,x), 1)$, even if its projection $x \mapsto f(t,x)$

may vanish, namely if there exists a point $\bar{x}$ such that $f(t, \bar{x}) = 0$ for all t. Consequently, as with flows, stationary or time-varying systems deserve a specific treatment as far as equilibrium points are concerned.

An approximation of the flow $X_t(x)$ in a neighborhood of an equilibrium point $\bar{x}$ may be obtained itself as the solution of a differential equation called the *variational equation* or *tangent linear approximation*.

The Taylor development of $X_t(x)$ around $\bar{x} = X_t(\bar{x})$ at the first order in $x - \bar{x}$ is given by

$$X_t(x) - \bar{x} = X_t(x) - X_t(\bar{x}) = \frac{\partial X_t}{\partial x}(\bar{x})(x - \bar{x}) + 0(\|x - \bar{x}\|^2)$$

and $\frac{\partial X_t}{\partial x}(\bar{x})$ satisfies the *tangent linear* or *variational equation*:

$$\frac{\partial}{\partial x}\frac{dX_t}{dt}(\bar{x}) = \frac{d}{dt}\frac{\partial X_t}{\partial x}(\bar{x}) = \frac{\partial f}{\partial x}(\bar{x})\frac{\partial X_t}{\partial x}(\bar{x})$$

or, noting $A \stackrel{\mathrm{def}}{=} \frac{\partial f}{\partial x}(\bar{x})$ and $z \stackrel{\mathrm{def}}{=} \frac{\partial X_t}{\partial x}(\bar{x})$:

$$\dot{z} = Az.$$

The eigenvalues of the matrix A are often called *characteristic exponents* of the equilibrium point $\bar{x}$. They indeed don't depend on the choice of coordinates since the matrix associated to the tangent linearization in any other system of local coordinates is similar to A and its eigenvalues are thus unchanged.

We say that $\bar{x}$ is *non degenerate* if A is invertible (i.e. has no 0 eigenvalues). In this case, the origin is the unique equilibrium point of the variational system and $\bar{x}$ is a locally unique equilibrium point of the vector field f.

We say that $\bar{x}$ is *hyperbolic* if A has no eigenvalues on the imaginary axis. Note that a generic[2] matrix is precisely a matrix without eigenvalues on the imaginary axis.

Example 3.1. Consider the classical differential equation of the simple pendulum

$$\ddot{\theta} = -\frac{g}{l}\sin\theta \tag{3.3}$$

or, using the notations $x_1 = \theta$ and $x_2 = \dot{\theta}$,

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{g}{l}\sin x_1 \ . \end{cases} \tag{3.4}$$

We easily verify that this system admits two singular points in the cylinder $\mathbb{S}^1 \times \mathbb{R}$: $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (\pi, 0)$ (a point $x_1 \in \mathbb{S}^1$ being, by

[2] in the sense of "almost all" matrices. For a rigorous definition, the reader may refer to Arnold [1980], Demazure [2000].

definition, the unique element in the interval $[0, 2\pi]$ of the set of reals of the form $x_1 + 2k\pi$, with $k \in \mathbb{Z}$).

The tangent linearization of the system at $(0, 0)$ is $\dot{z} = \begin{pmatrix} 0 & 1 \\ -\frac{g}{l} & 0 \end{pmatrix} z$ whose characteristic exponents are $\pm i\sqrt{\frac{g}{l}}$. The origin is thus non hyperbolic.

The tangent linearization of the system at $(\pi, 0)$ is $\dot{z} = \begin{pmatrix} 0 & 1 \\ \frac{g}{l} & 0 \end{pmatrix} z$. Its characteristic exponents are $\pm\sqrt{\frac{g}{l}}$ and $(\pi, 0)$ is hyperbolic.

Remark 3.2. One can always consider that the singular point is the origin 0 since if $\bar{x}$ satisfies $f(\bar{x}) = 0$, it suffices to make the change the coordinates as $\xi = x - \bar{x}$ which transforms the vector field f in $g(\xi) = f(\xi + \bar{x})$. We therefore have $g(0) = 0$ and $\frac{\partial g}{\partial \xi}(0) = \frac{\partial f}{\partial x}(\bar{x})$.

3.1.2 Periodic Orbit

We call *cycle* or *periodic orbit* an integral curve of (3.1) which is not reduced to a point and closed (diffeomorphic to a circle), *i.e.* such that there exists $T > 0$ satisfying $X_T(x) = x$.

One can easily prove that the integral curves of (3.1) are either points or curves diffeomorphic to a line or a circle. Therefore, if an orbit is neither constant nor injective, it is periodic.

Some remarkable classes of systems cannot possess periodic orbits. This is the case for instance of *gradient systems*: a system is said *gradient* if its vector field f is given by

$$f(x) = -\frac{\partial V}{\partial x}(x) \ , \quad \forall x \in X$$

where V is a C^2 function from X to $\mathbb{R}$.

Obviously, a necessary and sufficient condition for f to be locally a gradient is that

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i} \ , \quad \forall i, j.$$

We have the following result:

Proposition 3.1. *A gradient system has no periodic orbit.*

Proof. By definition, we have

$$\frac{d}{dt}V(x(t)) = \frac{\partial V}{\partial x}(x(t))\left(-\frac{\partial V}{\partial x}(x(t))\right) = -\|\frac{\partial V}{\partial x}(x(t))\|^2 \leq 0 \ , \quad \forall t. \tag{3.5}$$

Let us assume that the integral curve $x(t)$ is periodic, *i.e.* there exists $T > 0$ such that $x(t + T) = x(t)$ for all t. If $\frac{\partial V}{\partial x}(x(t)) \neq 0$ for at least one t, we have $V(x(t+s)) < V(x(t))$ for all $s \geq 0$ according to the inequality (3.5) and thus in particular $V(x(t+T)) < V(x(t))$, which contradicts the fact that $x(t+T) = x(t)$ (since then one should have $V(x(t+T)) = V(x(t))$). Therefore $\frac{\partial V}{\partial x}(x(t)) = 0$ for all t and, according to the uniqueness of the solution of $\dot{x} = -\frac{\partial V}{\partial x}(x)$, we get $x(t+s) = x(t)$ for all s, in other words, the corresponding orbit is reduced to a point which contradicts the fact that it is periodic.

Remark 3.3. The wording *gradient system* is somewhat misleading since systems corresponding to this definition cover a very small part of physical systems whose force field is a gradient: if a force field is a gradient, with potential U, a smooth function from $\mathbb{R}^n$ to $\mathbb{R}$, in place of $\dot{x} = -\frac{\partial V}{\partial x}$, we should have $m\ddot{x} = -\frac{\partial U}{\partial x}$, or, setting $v = \dot{x}$, $x \in \mathbb{R}^n$, $v \in \mathbb{R}^n$,

$$\dot{x} = v = -\frac{\partial V}{\partial x}$$
$$\dot{v} = -\frac{1}{m}\frac{\partial U}{\partial x} = -\frac{\partial V}{\partial v}.$$

The corresponding vector field, $\sum_{i=1}^{n}\left(v_i \frac{\partial}{\partial x_i} - \frac{1}{m}\frac{\partial U}{\partial x_i}\frac{\partial}{\partial v_i}\right)$, is a potential gradient, with potential V, in the sense of Proposition 3.1, if and only if $\frac{\partial^2 V}{\partial x_i \partial v_j} = -\delta_{i,j} = \frac{1}{m}\frac{\partial^2 U}{\partial x_i \partial x_j}$, $i,j = 1,\ldots,n$, or $U(x) = -\frac{1}{2}m\|x\|^2 + C$ (C being an arbitrary constant and $\|.\|$ the Euclidean norm of $\mathbb{R}^n$), which implies that $V(x,v) = -(x-b)^T v + C$ (where b is an arbitrary vector of $\mathbb{R}^n$) and that the only possible $2n$-dimensional gradient dynamics are given by

$$\dot{x} = v, \quad \dot{v} = x - b.$$

Example 3.2. Let us give a direct proof that the system of the previous remark doesn't have a periodic orbit, which we know by construction according to Proposition 3.1.

Integrating $\dot{x} = v, \quad \dot{v} = x - b$, we get $x_i(t) = \alpha_i e^t + \beta_i e^{-t} + b_i$, $v_i(t) = \alpha_i e^t - \beta_i e^{-t}$, $i = 1,\ldots,n$.

If this integral curve had a periodic orbit, there would exist $T > 0$ such that $x(t+T) = x(t)$ and $v(t+T) = v(t)$ for all t.Thus

$$\begin{pmatrix} \alpha_i & \beta_i \\ \alpha_i & -\beta_i \end{pmatrix}\begin{pmatrix} e^{t+T} \\ e^{-t-T} \end{pmatrix} = \begin{pmatrix} \alpha_i & \beta_i \\ \alpha_i & -\beta_i \end{pmatrix}\begin{pmatrix} e^{t} \\ e^{-t} \end{pmatrix}, \quad i = 1,\ldots,n,$$

which immediately implies that, for all t, the non zero vector $\begin{pmatrix} (e^T - 1)e^t \\ (e^{-T} - 1)e^{-t} \end{pmatrix}$ (we both have $e^T - 1 \neq 0$ and $e^{-T} - 1 \neq 0$ by assumption) is an eigenvector of the matrix $\begin{pmatrix} \alpha_i & \beta_i \\ \alpha_i & -\beta_i \end{pmatrix}$, whose determinant must be 0, *i.e.* $-2\alpha_i\beta_i = 0$. Thus

$\alpha_i = 0$ or $\beta_i = 0$, $i = 1, \ldots, n$. But if $\alpha_i = 0$ or $\beta_i = 0$, the integral curve is given by $x_i(t) = \beta_i e^{-t} + b_i$, $v_i(t) = -\beta_i e^{-t}$ or $x_i(t) = \alpha_i e^t + b_i$, $v_i(t) = \alpha_i e^t$, and it cannot be a periodic orbit according the strict monotonicity of the exponential. Therefore, the above system has no periodic orbit.

Let us now state the following non existence result of periodic orbits for a large class of physical systems:

Proposition 3.2. *Let U be a differentiable function from $\mathbb{R}^n$ to $\mathbb{R}$ and $\psi(x)$ an arbitrary function from $\mathbb{R}^n$ to $\mathbb{R}^n$. We denote by $\langle , \rangle$ the scalar product of $\mathbb{R}^n$ and by m a positive real number. The dynamics*

$$m\ddot{x} = -\frac{\partial U}{\partial x}(x) - \langle \psi(x), \dot{x} \rangle \, \psi(x) \tag{3.6}$$

have no periodic orbit. If in additions U has a unique minimum $\overline{x}$ on $\mathbb{R}^n$ and if ψ doesn't vanish on $\mathbb{R}^n$, then all the integral curves of (3.6) converge to $\overline{x}$.

Clearly, the right-hand side represents the sum of forces deriving from the potential U and the dissipative ones $(- \langle \psi(x), \dot{x} \rangle \, \psi(x))$.

Proof. Set $V(x, \dot{x}) = \frac{1}{2} m \langle \dot{x}, \dot{x} \rangle + U(x)$ (mechanical energy). We have

$$\begin{aligned} \frac{d}{dt} V(x(t), \dot{x}(t)) &= \langle m\ddot{x}, \dot{x} \rangle + \left\langle \frac{\partial U}{\partial x}(x), \dot{x} \right\rangle \\ &= -\left\langle \frac{\partial U}{\partial x}(x), \dot{x} \right\rangle - \langle \psi(x), \dot{x} \rangle \langle \psi(x), \dot{x} \rangle + \left\langle \frac{\partial U}{\partial x}(x), \dot{x} \right\rangle \\ &= - \langle \psi(x), \dot{x} \rangle^2 \leq 0. \end{aligned} \tag{3.7}$$

We then easily adapt the argument of the proof of Proposition 3.1.

If in addition $\overline{x}$ satisfies $U(\overline{x}) = \min_{x \in \mathbb{R}^n} U(x)$, and if ψ doesn't vanish, since V is strictly decreasing along the integral curves of (3.6), all the latter converge to the minimum of V which is precisely $\overline{x}$: one can easily check that the point $(x = \overline{x}, \dot{x} = 0)$ is an equilibrium point since in the opposite case we would have $U(\overline{x}) = V(\overline{x}, 0) > \min_{(x,\dot{x})} V(x, \dot{x}) \geq U(\overline{x})$, which is absurd.

Remark 3.4. Going back to remark 3.3, the system (3.6) is a Hamiltonian system with a dissipative term. Indeed, setting $H = \frac{1}{2m} \|x_2\|^2 + U(x_1)$ with $x_1 = x$, $x_2 = m\dot{x}_1$ and $\|x_2\|^2 = \langle x_2, x_2 \rangle$., we have $\frac{\partial H}{\partial x_1} = \frac{\partial U}{\partial x_1}$ and $\frac{\partial H}{\partial x_2} = \frac{1}{2m} x_2$ thus $\dot{x}_1 = \frac{\partial H}{\partial x_2}$ and $\dot{x}_2 = -\frac{\partial H}{\partial x_1} - \frac{1}{m} \langle \psi(x_1), x_2 \rangle \, \psi(x_1)$. Remark that if there is no dissipation ($\psi \equiv 0$), the system is Hamiltonian and may well have periodic orbits.

3.1.3 Poincaré's Map

Given a periodic orbit γ of (3.1), of period T, a point $x \in \gamma$ and a submanifold W of dimension $n-1$ transverse to γ at the point x (*i.e.* such that the tangent space $\mathrm{T}_x W$ to W at the point x and the line $\mathbb{R}.f(x)$ tangent at x to the orbit γ are supplementary), we call *Poincaré's map* or *first return map* associated to W and x, the mapping P that, to every $z \in W$ close enough to x, makes correspond the point $P(z)$ obtained by taking the first intersection of the orbit of z with W (see Fig. 3.1).

The expression *first intersection* makes sense since, by continuity, the time $T(z)$ on the orbit of z to come back to W is close to the period T and is thus bounded by below by a strictly positive number.

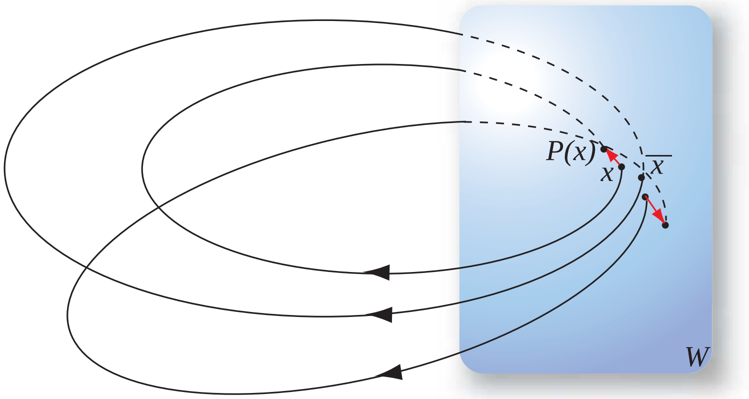

Fig. 3.1 The Poincaré map P around the fixed point $\bar{x}$. Here, the tangent linear mapping of P at $\bar{x}$ is unstable.

The Poincaré map doesn't depend on the choice of the transverse submanifold W or on the point x: if we choose another transverse submanifold W' and a point x' of γ, the corresponding mapping P' is the image of P by the diffeomorphism that transforms W in W' and x in x'. We leave the verification of this property to the reader.

Recall that, given a diffeomorphism f on a manifold X, the point $x \in X$ is called a *fixed point* of f if and only if $f(x) = x$.

Remark that Poincaré's map P maps W, of dimension $n - 1$, into itself and is a local diffeomorphism. We can also define inductively

$$z_{k+1} = P(z_k)$$

which admits x as fixed point. It results that the study of (3.1) in a neighborhood of the periodic orbit γ is reduced to the study of the above discrete-time equation.

Remark that for discrete-time systems of the form

$$x_{k+1} = f(x_k)$$

where f is a diffeomorphism of class C^∞, equilibrium points are replaced by fixed points of f, and the tangent linear, or variational, system around the fixed point $\bar{x}$ is obtained as in the continuous case:

$$X_k(x) - \bar{x} = X_k(x) - X_k(\bar{x}) = \frac{\partial X_k}{\partial x}(\bar{x})(x - \bar{x}) + 0(\|x - \bar{x}\|^2)$$

and $\frac{\partial X_k}{\partial x}(\bar{x})$ satisfies the *tangent linear* or *variational equation*:

$$\frac{\partial X_{k+1}}{\partial x}(\bar{x}) = \frac{\partial f}{\partial x}(\bar{x})\frac{\partial X_k}{\partial x}(\bar{x})$$

or, noting as in the continuous-time case $A \stackrel{\text{def}}{=} \frac{\partial f}{\partial x}(\bar{x})$ and $z_k \stackrel{\text{def}}{=} \frac{\partial X_k}{\partial x}(\bar{x})$:

$$z_{k+1} = Az_k.$$

In this discrete-time case, we call the eigenvalues of the tangent linear matrix A the *characteristic multipliers* of $\bar{x}$.

For a periodic orbit, the $(n-1)$ characteristic multipliers of P (the $n-1$ eigenvalues of the tangent linear mapping of P at x) play an important role. We say that the orbit γ is *hyperbolic* if P doesn't have its characteristic multipliers on the unit circle of the complex plane.

Example 3.3. Let us go back to the simple pendulum of example 3.1. We recall that it has two singular points in the cylinder $\mathbb{S}^1 \times \mathbb{R}$: $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (\pi, 0)$. The tangent linearized system at $(0, 0)$ is $\dot{z} = \begin{pmatrix} 0 & 1 \\ -\frac{g}{l} & 0 \end{pmatrix} z$ and its characteristic exponents are $\pm i\sqrt{\frac{g}{l}}$. Thus, the singular point $(0, 0)$ is not hyperbolic.

The tangent linearized system at $(\pi, 0)$ is given by $\dot{z} = \begin{pmatrix} 0 & 1 \\ \frac{g}{l} & 0 \end{pmatrix} z$. Its characteristic exponents are real, given by $\pm\sqrt{\frac{g}{l}}$ which means that the singular point $(\pi, 0)$ is a hyperbolic saddle (the juggler's unstable equilibrium).

According to Proposition 3.1, the above vector field, $f(x_1, x_2) = (x_2, -\frac{g}{l}\sin x_1)^T$, is not a gradient since

$$\frac{\partial f_1}{\partial x_2} = 1 \neq \frac{\partial f_2}{\partial x_1} = -\frac{g}{l}\cos x_1.$$

On the contrary, it is Hamiltonian since, if we set as before $H = \frac{1}{2}x_2^2 + \frac{g}{l}(1 - \cos x_1)$, we have $\dot{x}_1 = \frac{\partial H}{\partial x_2}$ and $\dot{x}_2 = -\frac{\partial H}{\partial x_1}$. We thus have here a periodic Hamiltonian system.

It is well-known that the pendulum movements are periodic as far as air friction or other damping forces are neglected. The mechanical energy is equal to H (up to the multiplicative constant ml^2) and it is easy to verify that its Lie derivative along the pendulum vector field is equal to 0, which means that the Hamiltonian H is a first integral. The pendulum integral curves are thus characterized by the equation

$$\frac{1}{2}x_2^2 + \frac{g}{l}(1 - \cos x_1) = \frac{g}{l}(1 - \cos \theta_0)$$

the right-hand side corresponding to the energy at initial time, the pendulum being released at 0 initial velocity (any other initial condition can be reduced to this one by a proper choice of initial position θ_0). The corresponding orbits, of equation

$$x_2 = \pm\sqrt{\frac{2g}{l}(\cos x_1 - \cos \theta_0)}$$

are closed and their associated period $T(\theta_0)$ is obtained by the integration of $dt = \frac{1}{x_2}dx_1$ on a quarter period:

$$T(\theta_0) = 2\sqrt{\frac{2l}{g}} \int_{\theta_0}^{0} \frac{d\zeta}{\sqrt{\cos \zeta - \cos \theta_0}}.$$

One can prove (exercise) that this expression, which is an elliptic integral, is approximated, for a small enough initial angle θ_0, by the usual formula $T = 2\pi\sqrt{\frac{l}{g}}$.

Let us now intersect the orbits with the half-line $\{ x_1 \ \geq \ 0 \ , \ x_2 \ = \ 0 \}$ (any other choice of transverse half-line would give the same result). We verify that the restriction of the above orbit equation to this half-line is given by $\cos x_1 = \cos \theta_0$, which amounts to say that each orbit crosses the axis $x_2 = 0$ at $x_1 = \theta_0 + 2k\pi$ for $x_1 \ \geq \ 0$ or that each orbit winds up indefinitely. We thus deduce, without calculation, that the Poincaré map is locally equal to the identity of $\mathbb{R}$ and thus admits 1 as unique characteristic multiplier, which proves that the pendulum orbits are not hyperbolic. One also can prove that they are not persistent in presence of perturbations: for instance, if we consider the air friction as a viscous friction, namely a force proportional to the velocity with opposite orientation, the pendulum equation becomes

$$\ddot{\theta} = -\frac{g}{l}\sin \theta - \epsilon\dot{\theta}$$

with $\epsilon > 0$, or

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{g}{l} \sin x_1 - \epsilon x_2. \end{cases}$$

Thus system is now of the form (3.6) with $\psi(x_1) \equiv \epsilon$.

The two previous equilibrium points remain but the origin is now hyperbolic (and stable as will be seen later) and the equilibrium $(\pi, 0)$ remains hyperbolic unstable. However, even for ϵ arbitrarily small, there is no more periodic orbit according to Proposition 3.2.

3.2 Stability of Equilibrium Points and Orbits

3.2.1 Attractor

The set U is called *invariant* (resp. *positively invariant*) if it contains its image by the flow for all t (resp. for all $t \geq 0$), in other words if $X_t(U) \subset U$ for all $t \in \mathbb{R}$ (resp. $t \geq 0$).
Thus, if U is invariant (resp. positively invariant), the integral curves starting from U remain in U for all times positive or negative (resp. for all positive times).

We say that U is globally positively invariant if it is positively invariant and if all the integral curves enter U after some positive finite time. Also, an invariant set which is a submanifold of X is called an *invariant manifold*.

Note that an equilibrium point, or an equilibrium set

$$f^{-1}(0) = \{x \in X | f(x) = 0\}$$

is always a globally invariant set since $X_t(x) = x$ for all $x \in f^{-1}(0)$, which means that $X_t(f^{-1}(0))) = f^{-1}(0)$ for all t.

In discrete time, the same definitions hold if we change $t \in \mathbb{R}$ in $t \in \mathbb{Z}$ and $t \geq 0$ in $t \in \mathbb{N}$.
The same remark applies to the next definitions and, for clarity's sake, these straightforward adaptations are left to the reader.

Example 3.4. For the system $\dot{x} = -x$ in $\mathbb{R}$, every compact neighborhood of 0 is globally positively invariant. On the contrary, for the system $\dot{x} = +x$, only the equilibrium point $\{0\}$ is positively invariant.

Example 3.5. The orbit of a globally defined first integral is an invariant manifold: in $\mathbb{R}^2$, an arbitrary circle $x_1^2 + x_2^2 = R^2$ is an invariant manifold of the system $\dot{x}_1 = x_2, \ \dot{x}_2 = -x_1$.

Example 3.6. In continuous time, one easily verifies that if $X = \mathbb{R}^n$, if $U \subset X$ is a compact with non empty interior whose boundary ∂U is differentiable and orientable, U is positively invariant if and only if the vector field f points inwards on ∂U. If we note ν the normal to ∂U pointing outwards and $< .,. >$ a scalar product on X, we say that f is inwards on ∂U if $< f, \nu >_{|\partial U} < 0$.

For example, in $\mathbb{R}^2$, the ellipsoid $U = \{(x_1, x_2) \in \mathbb{R}^2 | kx_1^2 + x_2^2 \leq r^2\}$ is positively invariant for the vector field $f = x_2 \frac{\partial}{\partial x_1} - (kx_1 + x_2) \frac{\partial}{\partial x_2}$ since a normal vector ν on the boundary $\partial U = \{(x_1, x_2) \in \mathbb{R}^2 | kx_1^2 + x_2^2 = r^2\}$ of U has the form $(kx_1, x_2)^T$ and the scalar product $< f, \nu >$ is equal to $-kx_1x_2 - x_2^2 + kx_1x_2 = -x_2^2 \leq 0$. Moreover, one can verify that if $x_2 = 0$, $\dot{x}_2 = -kx_1 < 0$ (x_1 and x_2 cannot vanish simultaneously on ∂U) which proves that f is strictly inwards on ∂U. In fact, a direct computation of the flow shows that the integral curves of f not only remain in U but also converge to the origin.

We now introduce *limit sets* which are both more precise and more intrinsic than invariant sets. We denote by $\bar{U}$ the (topological) closure of U.

- $\bigcap_{t \in \mathbb{R}} X_t(\bar{U})$ is the set made of the limit points of integral curves remaining in U when t tends to $\pm\infty$;
- $\bigcap_{t \geq 0} X_t(\bar{U})$ is the set made of the limit points of integral curves remaining in U when t tends to $+\infty$;
- $\bigcap_{t \leq 0} X_t(\bar{U})$ is the set made of the limit points of integral curves remaining in U when t tends to $-\infty$.

Among these limit sets, attractors are those to which the integral curves converge in positive time. Note that time inversion transforms divergence into convergence and vice versa.

We call *attractor* the set $V = \bigcap_{t \geq 0} X_t(\bar{U})$ where $\bar{U}$ is a compact invariant set.

If U is globally positively invariant, the attractor V is called *maximal.*

We easily see that an attractor is an invariant set: $V = X_t(V)$ for all t.

Example 3.7. All singular points of a vector field f (*i.e.* elements of $f^{-1}(0)$) are not necessarily attractors. Only stable ones are attractors, in particular those whose characteristic exponents have strictly negative real part, as will be seen later. Instable points are indeed attractors when we change t in $-t$, as mentioned above, which also amounts to consider the vector field $-f$ in place of f.

There exist invariant sets that are attractors for both f and $-f$. Consider the system

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -x_1,$$

The orbits are given by $x_1^2 + x_2^2 = R^2$ where R is an arbitrary real. They are compact invariant sets whatever the sense of motion, which means that the orbits are attractors both for f and $-f$.

Remark 3.5. Note that an attractor needs not be a manifold. It may have a complicated geometric structure. For instance, an attractor which is not made of a finite union of submanifolds of X is called *strange attractor* or *fractal set*. Its dimension, in the sense of Hausdorff, needs not be an integer.

3.2.2 Lyapunov Stability

The notion of attractor is rather vague and several refinements are possible. Among them, Lyapunov's stability[3] is a key concept. More precisely, we introduce the notions of *Lyapunov's stability*, or *L-stability*, and of *asymptotic Lyapunov's stability*, or *L-asymptotic stability*, for equilibrium points or periodic orbits.

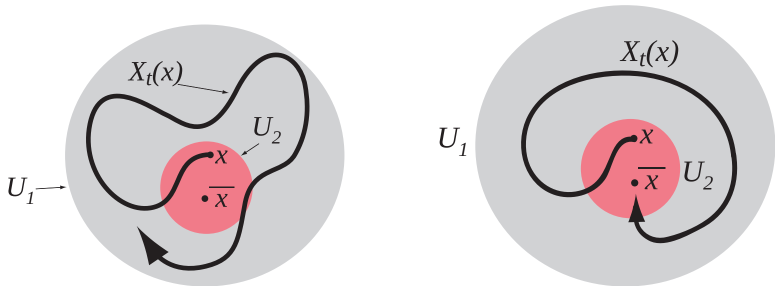

Fig. 3.2 L-stability (left) and L-asymptotic stability (right) of the point $\bar{x}$.

We say that the point $\bar{x}$ is *Lyapunov-stable*, or *L-stable*, if for every neighborhood U_1 of $\bar{x}$ there exists a neighborhood U_2 of $\bar{x}$ included in U_1 such that all the integral curves starting from U_2 at time $t = 0$ remain in U_1 for all $t \geq 0$ (Fig. 3.2, left).

We say that the point $\bar{x}$ is *Lyapunov-asymptotically stable*, or *L-asymptotically stable*, if it is L-stable and if every integral curve starting from a neighborhood of $\bar{x}$ at time $t = 0$ converges to $\bar{x}$ when $t \to +\infty$ (Fig. 3.2, right).

The main difference between these two definitions is that a small perturbation on the initial state of a system around an L-stable equilibrium point may generate small undamped oscillations, whereas these oscillations are damped and asymptotically vanish in the case of an L-asymptotic stable equilibrium.

[3] in reference to the work of Alexander Lyapunov (1854–1918) one of the founders, with Henri Poincaré (1854–1912), and then Ivar Bendixson (1861–1936), George Birkhoff (1884–1944) and many other mathematicians and physicists, of the qualitative analysis of differential equations. Most of their researches in this field were motivated by problems of celestial mechanics.

Before stating the next stability Theorem, let us recall that, given an equilibrium point $\bar{x}$ of the vector field f and denoting the matrix of the linear tangent mapping of f at $\bar{x}$ by $A = \frac{\partial f}{\partial x}(\bar{x})$, $\bar{x}$ is said to be non degenerate if A is invertible (*i.e.* A has no 0 eigenvalues), and is said *hyperbolic* if A has no eigenvalue on the imaginary axis.

Theorem 3.1. (Stability of an equilibrium point, continuous case) *Let $\bar{x}$ be a non degenerate equilibrium point of the vector field f.*

1. *If all its characteristic exponents at $\bar{x}$ have a strictly negative real part, then $\bar{x}$ is L-asymptotically stable.*
2. *If at least one of the characteristic exponents has a strictly positive real part, then $\bar{x}$ is not L-stable.*

Remark 3.6. To study the properties of the characteristic exponents, it is generally useful to put the matrix A in Jordan form (see e.g. Gantmacher [1966]). Let P be the matrix of the change of basis that transforms A into its Jordan form J, *i.e.* such that $PAP^{-1} = J$. If A has only non 0 distincts eigenvalues, P is the matrix of the eigenvectors associated to the eigenvalues of A. Setting $\tilde{x} = Px$ and $\tilde{f}(\tilde{x}) = Pf(P^{-1}\tilde{x})$, the system (3.1) is transformed in $\dot{\tilde{x}} = \tilde{f}(\tilde{x})$. One readily verifies that the linear tangent approximation of $\tilde{f}$ at 0 is equal to $\dot{\tilde{z}} = J\tilde{z}$ with $\tilde{z} = Pz$. We generally choose these coordinates for which the statements and results are significantly simplified.

In the plane $\mathbb{R}^2$, non degenerate equilibrium points are classified into four categories according to the sign of the real part of their characteristic exponents: *saddle point* (2 eigenvalues with real part of opposite signs), the *node* (2 real eigenvalues of same sign), the *focus* (2 complex eigenvalues with non zero real parts of the same sign) and the *centre* (2 imaginary eigenvalues) (see Fig. 3.3).

Example 3.8. Considering again the Hamiltonian system with dissipation (3.6) of Proposition 3.2, in dimension $n = 1$ for simplicity's sake, assume that the potential U admits a unique minimum at the point $\overline{x}$. Recall that the necessary conditions for $\overline{x}$ to be a unique minimum are given by

$$\frac{\partial U}{\partial x}(\overline{x}) = 0, \quad \frac{\partial^2 U}{\partial x^2}(\overline{x}) > 0.$$

To compute the characteristic exponents of $\overline{x}$, let us set, as in the Remark 3.4, $x_1 = x$, $x_2 = m\dot{x}_1$. The system thus reads, in two dimensions:

$$\begin{aligned}
\dot{x}_1 &= \frac{1}{m} x_2 \\
\dot{x}_2 &= -\frac{\partial U}{\partial x_1}(x_1) - \frac{1}{m}\psi^2(x_1) x_2
\end{aligned}$$

and has $(\overline{x}, 0)$ as equilibrium point.

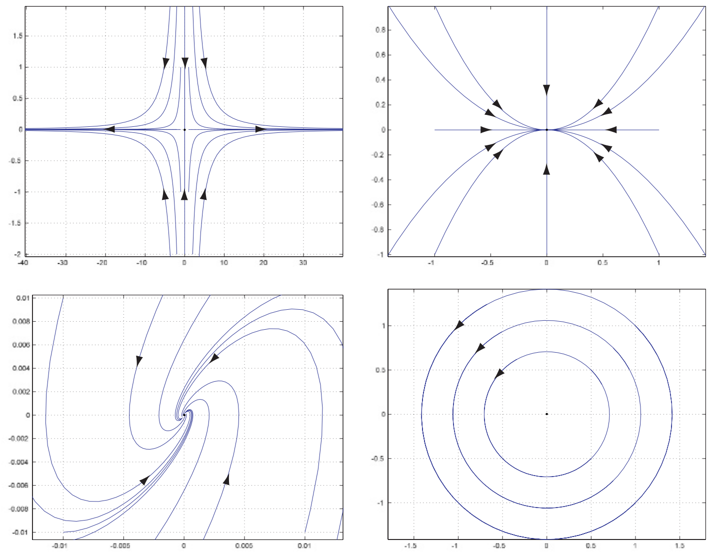

Fig. 3.3 The origin is a saddle (top left), equilibrium point of the system $\dot{x}_1 = x_1,\ \dot{x}_2 = -2x_2$; a stable node (top right), equilibrium point of the system $\dot{x}_1 = -x_1,\ \dot{x}_2 = -2x_2$; a stable focus (bottom left), equilibrium point of the system $\dot{x}_1 = -x_2,\ \dot{x}_2 = 2x_1 - 2x_2$; a centre (bottom right), equilibrium point of the system $\dot{x}_1 = x_2,\ \dot{x}_2 = -x_1$. In these four cases, the equilibrium point is non degenerate but not hyperbolic in the case of a centre.

Its tangent linear system around $(\overline{x}, 0)$ is given by

$$\dot{\xi}_1 = \frac{1}{m}\xi_2$$
$$\dot{\xi}_2 = -\frac{\partial^2 U}{\partial x_1^2}(\overline{x})\xi_1 - \frac{1}{m}\psi^2(\overline{x})\xi_2$$

and its characteristic exponents are the roots of the characteristic polynomial (of the complex variable λ):

$$\det\begin{pmatrix} \lambda & -\frac{1}{m} \\ \frac{\partial^2 U}{\partial x_1^2}(\overline{x}) & \lambda + \frac{1}{m}\psi^2(\overline{x}) \end{pmatrix} = \lambda^2 + \frac{1}{m}\psi^2(\overline{x})\lambda + \frac{1}{m}\frac{\partial^2 U}{\partial x_1^2}(\overline{x}) = 0$$

Since, by assumption, $\frac{\partial^2 U}{\partial x_1^2}(\overline{x}) > 0$, the product of the roots is positive, and since $\psi^2(\overline{x}) > 0$, the sum of the roots is strictly negative. Thus the 2 roots have a strictly negative real part, which allows to recover the convergence result of Proposition 3.2 and makes it even more precise since it proves the L-asymptotic stability of the equilibrium point $\overline{x}$, according to the first item of Theorem 3.1.

The conditions of Theorem 3.1 are generally referred to as *Lyapunov's first method*. They are sufficient but not necessary since they cannot conclude, for instance, in the case of purely imaginary or, a fortiori, null, eigenvalues. The next example shows the difficulties to establish a general result in this case.

Example 3.9. Consider the two scalar differential equations $\dot{x} = x^3$ and $\dot{x} = -x^3$. They both admit $x = 0$ as equilibrium point with 0 as characteristic exponent. The general solution of $\dot{x} = ax^3$ being $x(t) = (x_0^{-2} - 2at)^{-\frac{1}{2}}$, we easily see that for $a = -1$ (second system) all the integral curves are well defined for $t \to +\infty$ and converge to 0, which proves that 0 is an attractor, whereas for $a = 1$ (first system) the integral curves are not defined after $t = \frac{x_0^{-2}}{2}$, though they all start from 0 at $t = -\infty$, which proves that 0 is an attractor in backward time. Therefore, the two systems have opposite behaviors whereas they have the same eigenvalue 0 at the same equilibrium point 0.

Note that this problem is not specific of nonlinear systems: it already exists in the linear case in dimension larger than or equal to 2, as shown by the next example:

Example 3.10. The origin $(0, 0)$ of $\mathbb{R}^2$ for the system $\dot{x} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} x$ is L-stable since $x(t) = x_0$ for all t, but is not L-asymptotically stable. On the contrary, for the system $\dot{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x$, it is not L-stable since the integral curves are given byt $x_1(t) = x_2(0)t + x_1(0)$, $x_2(t) = x_2(0)$. In both cases, 0 is a double eigenvalue.

To analyze the L-stability of the Poincaré map around a fixed point, which corresponds to the stability of a periodic orbit, Theorem 3.1 is adapted as follows:

Theorem 3.2. (Stability of a fixed point, discrete case)
Let $\bar{x}$ be a fixed point of the diffeomorphism f.

1. *If all the characteristic multipliers of $\bar{x}$ have their modulus strictly smaller than 1, then $\bar{x}$ is L-asymptotically stable.*
2. *If at least one of the characteristic multipliers of $\bar{x}$ has modulus strictly larger than 1, then $\bar{x}$ is not L-stable.*

3.2.3 Remarks on the Stability of Time-Varying Systems

Let us now stress on a major difference between time-varying and stationary systems concerning stability: in the next example, we exhibit an *unstable* time-varying system whose eigenvalues of the tangent linear system at the equilibrium point, at every time, have *strictly negative real part*, which would contradict Theorem 3.1, assuming it would apply to this context.

We consider the linear time-varying system:

$$\dot{x} = \begin{pmatrix} -1 + \frac{3}{2}\cos^2 t & 1 - \frac{3}{2}\sin t \cos t \\ -1 - \frac{3}{2}\sin t \cos t & -1 + \frac{3}{2}\sin^2 t \end{pmatrix} x \ .$$

Its unique equilibrium point is the origin $x = 0$ and, by linearity, the system is equal to its tangent linear approximation.
We immediately see that, for all fixed t, its eigenvalues are equal to $-\frac{1}{4} \pm i\frac{\sqrt{7}}{4}$ (thus independent of time and with strictly negative real part $-\frac{1}{4}$).
However, it is straightforward to verify that the unique integral curve passing through the point $(-a, 0)$, $a \in \mathbb{R}$, at time $t = 0$, is given by

$$x(t) = \begin{pmatrix} -ae^{\frac{t}{2}} \cos t \\ ae^{\frac{t}{2}} \sin t \end{pmatrix}$$

and thus, for all $a \neq 0$, $\lim_{t \to +\infty} \|x(t)\| = +\infty$, which proves that the origin is not L-stable.

In fact, for some classes of time-varying systems, say $\dot{x}(t) = A(t)x(t)$, one can find a time-varying transformation $x = B(t)y$ such that the transformed system becomes stationary: $\dot{y} = \Lambda y$. In the periodic case, Floquet's Theorem (see e.g.. Arnold [1974, 1980]) asserts that the matrix function $B(.)$ is periodic of period possibly double of the one of $A(.)$. The stability of the original system at an equilibrium point is thus equivalent to the stability of Λ at the corresponding equilibrium point since $x(t) = B(t)e^{\Lambda t}B(0)^{-1}x(0)$ with $B(t)$ bounded (because periodic). In our example, the matrices $B(t)$ and Λ are given by:

$$B(t) = \begin{pmatrix} -\cos t & \sin t \\ \sin t & \cos t \end{pmatrix} \ , \quad \Lambda = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -1 \end{pmatrix} \ .$$

Λ thus has $\frac{1}{2}$ as eigenvalue, strictly positive real, which proves the non L-stability.

3.2.4 Lyapunov's and Chetaev's Functions

Originally, Lyapunov's functions were introduced to study the stability of an equilibrium point. We present here a slightly more general definition that applies to invariant manifolds.

Assume that system (3.1) admits a bounded invariant manifold X_0. A *Lyapunov's function* associated to X_0 is a mapping V of class C^1 on an

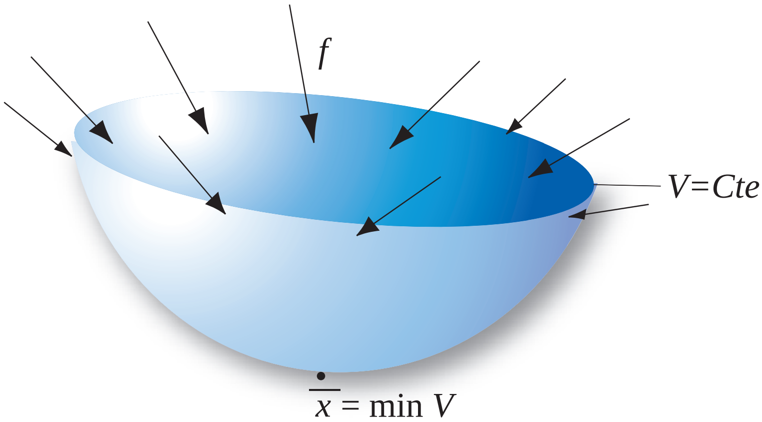

Fig. 3.4 Lyapunov's function V for the vector field f: f is inwards on each level set of V and the integral curves of f converge to the minimum of V which is the equilibrium point $\bar{x}$ of f.

open U of X containing X_0, with values in $\mathbb{R}_+$, and satisfying the following properties (see Fig. 3.4):

(i) V attains its minimum in U;
(ii) V is non increasing along the integral curves of (3.1), *i.e.* $L_f V \leq 0$ in U [4].

When the open set U is equal to X, if X is not compact, we call *proper Lyapunov's function* a Lyapunov's function satisfying the additional condition

(iii) $\lim_{\|x\| \to \infty,\ x \in X} V(x) = +\infty.$

Note that the differentiability property of the Lyapunov function V is not necessary and can be weakened by asking V to be of bounded variations, which suffices to define the Lie derivative $L_f V$ as a negative Radon measure on X, or in other words, as the sum of a non positive continuous function and a countable combination, with non positive coefficients, of Dirac masses concentrated on a collection of closed sets of X with empty interior. The use of such a technical machinery may be necessary when the vector field f is not everywhere differentiable or, worse, discontinuous (see e.g.. Filippov Filippov [1988]).

In the time-varying case, the previous definition must be adapted as follows: in (i), the Lyapunov function V is defined on $U \times \mathbb{R}$, admits a minimum, uniformly with respect to t, in U and, in (ii), the Lie derivative of V along

[4] In the case of a discrete dynamics $x_{k+1} = f(x_k)$ where f is a diffeomorphism, that includes the case of Poincaré's map of a periodic orbit, the latter inequality must be replaced by $V(f(x)) \leq V(x)$ in U.

$\tilde{f} = (f, 1)$ must be understood as: $L_{\tilde{f}} V = \frac{\partial V}{\partial t} + L_f V$. In this case too, the regularity of V may be weakened.

For simplicity's sake, we assume that $X = \mathbb{R}^n$ for the rest of this chapter.

Theorem 3.3. (LaSalle) *Let C be a compact set, positively invariant by the vector field f, contained in an open set U of X, and let V be a differentiable function satisfying $L_f V \leq 0$ in U. Let $W_0 = \{x \in U | L_f V = 0\}$ and X_0 the largest invariant set by f contained in W_0. Then, for every initial condition in C, X_0 is an attractor, i.e. $\bigcap_{t \geq 0} X_t(C) \subset X_0$.*

This result is called *LaSalle's invariance principle* and, if the manifold W_0 is reduced to a single point (which is therefore necessarily an equilibrium point of f), it is called *Lyapunov's second method*[5].

If the function V is such that the set $V^{-1}(]-\infty, c]) = \{x \in X | V(x) \leq c\}$, called level set, is bounded for a well-chosen real c, and if $L_f V \leq 0$ in all of X, one can choose $C = \overline{V^{-1}(]-\infty, c])}$ which is compact and positively invariant since V decreases along the integral curves of f.

Until now we have presented criteria to decide if a system is stable or not. But the nature of the convergence of the corresponding flow remains rather vague. More precisely, nothing is said to compare it with the exponential (resp. geometric) convergence. For this purpose, the following variants of condition (ii) are often introduced:

(ii)' V vanishes on X_0 and $L_f V < 0$ (resp. $V(f(x)) < V(x)$) in $U \setminus X_0$;
(ii)" V vanishes on X_0 and there exists $\alpha > 0$ such that $L_f V \leq -\alpha V$ (resp. there exists $\alpha \in]0, 1[$ such that $V(f(x)) \leq (1 - \alpha)V(x)$) in $U \setminus X_0$.

A *strict (resp. strong) Lyapunov's function* is a Lyapunov's function that satisfies (ii)' (resp. (ii)") in place of (ii).

Proposition 3.3. *Under the assumptions of Theorem 3.3, we have:*

- *if (ii)' holds true, then X_0 is an attractor;*
- *if (ii)" holds true, X_0 is an attractor and $V(X_t(x)) \leq e^{-\alpha t} V(x)$ for all x in a neighborhood of X_0 (resp. $V(f^k(x)) \leq (1 - \alpha)^k V(x)$).*

The adaptation of these results in the case where the set X_0 is reduced to an equilibrium point (resp. a fixed point) or a periodic orbit doesn't raise particular difficulties.

---

[5] also called Lyapunov's direct method since it doesn't require to solve the differential equation (3.1). The difficulty is nevertheless replaced by the determination of the function V! For some classes of systems, e.g. the dissipative Hamiltonian systems, the function V may be obtained according to physical considerations, as in Proposition 3.2, where V is the Hamiltonian (mechanical energy). Recall that Lyapunov's first method consists in looking at the sign of the characteristic exponents (Theorem 3.1).

We say that the equilibrium (resp. fixed) point $\bar{x}$ is *exponentially stable*[6] if it is L-asymptotically stable and if it locally admits a strong Lyapunov's function equivalent to a squared norm, in other words if there exists $\alpha > 0$ such that $\|X_t(x) - \bar{x}\|^2 \leq e^{-\alpha t}\|x - \bar{x}\|^2$ (resp. if there exists $\alpha \in]0, 1[$ such that $\|X_k(x) - \bar{x}\|^2 \leq (1 - \alpha)^k \|x - \bar{x}\|^2$) for all x in a suitable neighborhood of $\bar{x}$.

Note that Assumption (ii)" (strong Lyapunov's function) doesn't suffice to guarantee the exponential stability, as shown by the following example:

Example 3.11. Consider the scalar system $\dot{x} = -x^3$. One easily sees that $V(x) = e^{-\frac{\alpha}{2x^2}}$ is a strong Lyapunov's function for all $\alpha > 0$, but the integral curves, given by: $x(t) = (2t + C)^{-\frac{1}{2}}$, are not exponentially convergent to 0. Note that V is not polynomial, whereas the existence of a *quadratic* strong Lyapunov's function would suffice to guarantee the exponential stability.

In the hyperbolic case, one can more easily describe the links between L-asymptotic stability, existence of a strong Lyapunov's function and exponential stability:

Theorem 3.4. *Consider a hyperbolic equilibrium point (resp. fixed point) $\bar{x}$ of the vector field f. The following conditions are equivalent:*

1. *The equilibrium (resp. fixed) point $\bar{x}$ has all its characteristic exponents (resp. multipliers) with negative real part (resp. inside the unit disk)*
2. *There exists a strong Lyapunov's function in a neighborhood of $\bar{x}$.*

Moreover, if one of these two conditions is satisfied, $\bar{x}$ is exponentially stable.

In the latter case, one can choose as Lyapunov's function the squared norm of the distance to the equilibrium point in the basis of the eigenvectors associated to the Jordan form of A: if P is the transformation matrix (*i.e.* such that $PAP^{-1} = \Delta$ with Δ diagonal in the case where A is diagonalizable), then the function $V(x) = \|P^{-1}(x - \bar{x})\|^2$ is a strong Lyapunov's function in a sufficiently small neighborhood of $\bar{x}$. This is an easy consequence of the Theorems 3.1 and 3.2.

Example 3.12. Let us go back to the example 3.3 of the pendulum with small friction:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{g}{l}\sin x_1 - \epsilon x_2. \end{cases}$$

We denote by $f(x_1, x_2) = x_2 \frac{\partial}{\partial x_1} - (\frac{g}{l}\sin x_1 + \epsilon x_2)\frac{\partial}{\partial x_2}$ the corresponding vector field and we recall that a Lyapunov's function is given by:

$$V(x_1, x_2) = \frac{1}{2}x_2^2 + \frac{g}{l}(1 - \cos x_1).$$

---

[6] In the discrete-time case, the exponential convergence is replaced by the geometric one and thus, the term "exponential stability" is incorrect. However, for simplicity's sake, and since there is no ambiguity, we keep the same terminology as in the continuous-time case.

Indeed, $L_f V = -\epsilon x_2^2 \leq 0$ in $\mathbb{S}^1 \times \mathbb{R}$ and the set defined by $W_0 = \{(x_1, x_2) | L_f V = 0\}$ is equal to the subspace $\{x_2 = 0\} = \mathbb{R}$. Since $V \leq c$, for c real positive, it turns out that $|x_2| \leq \sqrt{2c}$, and thus one can choose $C = \overline{(]-\theta, +\theta[\times\mathbb{R}) \bigcap V^{-1}(]-\infty, c])}$, for $\theta \in]0, \pi[$, which is compact and positively invariant. Let us compute the largest invariant set by f contained in W_0. If we restrict f to W_0, we get $f_{|W_0} = -(\frac{g}{l} \sin x_1) \frac{\partial}{\partial x_2}$ for which the largest invariant set is $\{(x_1, x_2) | \sin x_1 = 0, x_2 = 0\} = \{(0,0), (\pi, 0)\}$. But, since the point $(\pi, 0)$ is excluded according to $\theta < \pi$, we converge to the equilibrium point $(0,0)$ corresponding to the pendulum at rest in downward vertical position, for all initial conditions in C.

It is often more difficult to prove the instability of a system than its stability. In the continuous time case, the Chetaev's functions may be useful to analyze unstable equilibrium points.

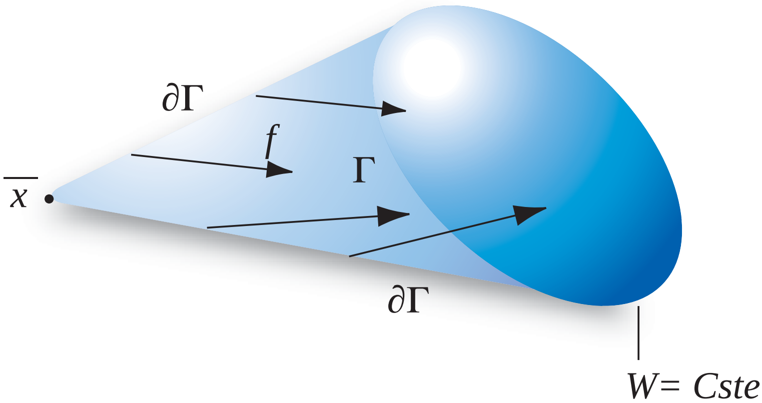

Fig. 3.5 Chetaev's Function W for the vector field f: f is inwards with respect to the cone Γ on the boundary $\partial\Gamma$ and W is increasing along the integral curves of f in Γ.

Let $\bar{x}$ be an equilibrium point of the vector field f. We say that the map W of class C^1 on a neighborhood U of $\bar{x}$ to $\mathbb{R}_+$ is a *Chetaev's function* (see Fig. 3.5) if

(i) U contains a cone with non empty interior Γ, with vertex $\bar{x}$ and piecewise smooth boundary $\partial\Gamma$, such that f is oriented inward Γ on $\partial\Gamma$;
(ii) $\lim_{x \to \bar{x}, x \in \Gamma} W(x) = 0$, $W > 0$ and $L_f W > 0$ in Γ.

Theorem 3.5. *An equilibrium point $\bar{x}$ for which a Chetaev's function exists is unstable. In particular, if $\bar{x}$ is hyperbolic and has at least one characteristic exponent with positive real part, the function obtained by taking the squared* *norm of the projection of $x - \bar{x}$ on the eigenspace corresponding to the eigenvalues with positive real part, in a suitable conic neighborhood, is a Chetaev's function.*

3.2.5 Hartman-Grobman's and Shoshitaishvili's Theorems, Centre Manifold

Definition 3.1. We say that the vector field (resp. the diffeomorphism) f having the origin[7] as equilibrium (resp. fixed) point, is *topologically equivalent* to its linear tangent approximation Az if there exists a homeomorphism ψ from a neighborhood U of 0 to itself that maps every orbit of f in an orbit of its tangent linear system and which preserves its sense of motion, *i.e.* such that $X_t(h(z)) = \psi(e^{A\tau(t,z)}z)$ (resp. $f^k(h(z)) = \psi(A^{\kappa(k,z)}z)$) for all $z \in U$ with τ a strictly increasing real function for all z (resp. κ a strictly increasing integer function for all z).

Remark that the scalar fields $-x$ et $-kx$ with $k > 0$ are topologically equivalent since, on the corresponding orbits, it suffices to change t in $\tau(t, x) = kt$ which is indeed an increasing function of time. On the contrary, the vector fields $-x$ and x are not topologically equivalent since, to map orbits into orbits we need to change t in $-t$ and the sense of motion is not preserved.

We have seen that if the origin is a hyperbolic equilibrium (resp. fixed) point, its stability can be determined by inspection of the sign of the real part (resp. the magnitude of the modulus) of the eigenvalues of the matrix A.

If A admits eigenvalues with both positive and negative real parts (resp. with modulus both smaller and larger than 1), the eigenspace corresponding to the negative ones (resp. with modulus smaller than 1), called the stable eigenspace, is the set of initial conditions from which the flow exponentially converges to 0 and the eigenspace corresponding to the positive ones, called the unstable eigenspace, the set of initial conditions from which the flow exponentially diverges from 0.

To this space decomposition, corresponds also a flow decomposition according to the invariance by A of the respective eigenspaces.

This decomposition may be locally extended, in a neighborhood af an equilibrium (resp. fixed) point, to stable and unstable manifolds of the vector field f, and to their corresponding invariant flows, by means of the above topological equivalence.

More precisely, we assume that A is hyperbolic and admits $k < n$ eigenvalues with strictly positive real part (resp. with modulus strictly larger than 1), counted with their multiplicity, and $n - k$ eigenvalues with strictly negative

[7] We have already remarked (see Remark 3.2 of section 3.1.1) that, by a suitable change of coordinates, any equilibrium point may be mapped to the origin.

real part (resp. of modulus strictly smaller than 1), again counted with their multiplicity. The total number of eigenvalues so involved is necessarily equal to n according to the hyperbolicity assumption.

To the eigenvalues of A with strictly positive real part (resp. of modulus strictly larger than 1) is associated the eigenspace E^+, of dimension k, and to the eigenvalues of A with strictly negative real part (resp. of modulus strictly smaller than 1) is associated E^-, of dimension $n-k$. E^+ and E^- are indeed supplementary vector subspaces that are invariant by A by definition of the eigenspaces: $E^+ \oplus E^- = \mathbb{R}^n$, $AE^+ \subset E^+$ and $AE^- \subset E^-$.

Thus, for an initial condition $z_0 \in E^-$, we have $\lim_{t\to+\infty} e^{At}z_0 = 0$ (resp. $\lim_{j\to+\infty} A^j z_0 = 0$) which means that E^- is the stable eigenspace of A. Accordingly, if $z_0 \in E^+$, changing t in $-t$ (resp. j in $-j$), we have $\lim_{t\to+\infty} e^{-At}z_0 = 0$ (resp. $\lim_{j\to+\infty} A^{-j} z_0 = 0$) which means that E^+ is the unstable eigenspace of A.

Moreover, in the coordinates corresponding to an adapted basis (e^+, e^-) of the eigenspace $E^+ \oplus E^-$, the system $\dot{x} = Ax$ (resp. $x_{j+1} = Ax_j$) is decomposed in two decoupled subsystems $\dot{x}^+ = A_+x^+$, $\dot{x}^- = A_-x^-$ (resp. $x^+_{j+1} = A_+x^+_j$, $x^-_{j+1} = A_-x^-_j$ for all j) where x^+ and x^- are the components of x in this basis.

To extend this decomposition to a vector field f, we need the following definition:

Definition 3.2. We call *stable manifold* of the vector field f at the equilibrium point 0, the submanifold

$$W^-_{loc}(0) = \{x \in U \mid \lim_{t\to+\infty} X_t(x) = 0 \text{ and } X_t(x) \in U \ \forall t \geq 0\}. \tag{3.8}$$

We call *local unstable manifold* of the vector field f at the equilibrium point 0, the submanifold

$$W^+_{loc}(0) = \{x \in U \mid \lim_{t\to+\infty} X_{-t}(x) = 0 \text{ and } X_{-t}(x) \in U \ \forall t \geq 0\}. \tag{3.9}$$

The submanifold $W^-_{loc}(0)$, which is obviously a positively invariant manifold of f, thus corresponds to the set of initial conditions in a neighborhood U for which the integral curves of f stay in U and converge to the origin. Accordingly, $W^+_{loc}(0)$ is a positively invariant manifold of $-f$, and corresponds to the set of initial conditions of U for which the integral curves of f diverge from the origin.

Proposition 3.4. *Assume that $W^-_{loc}(0)$ is a $(n-k)$-dimensional smooth manifold and that $W^+_{loc}(0)$ is a k-dimensional smooth manifold. Then $W^-_{loc}(0)$ is tangent at 0 to E^- and $W^+_{loc}(0)$ is tangent at 0 to E^+.*

Proof. We prove this assertion for $W^-_{loc}(0)$, the proof for $W^+_{loc}(0)$ following the same lines.

By assumption, there exists a smooth mapping h_s from X to $\mathbb{R}^k$ such that $W^-_{loc}(0) = \{x \in U | h_s(x) = 0\}$. By the invariance property, we have $L_f h_s(x) =$

$\frac{\partial h_s}{\partial x}(x)f(x) = 0$ for all $x \in W^-_{loc}(0)$. Then, using the fact that $f(x) = Ax + 0(\|x\|^2)$ and $\frac{\partial h_s}{\partial x}(x) = \frac{\partial h_s}{\partial x}(0) + 0(\|x\|)$, we get $\frac{\partial h_s}{\partial x}(0)Ax + 0(\|x\|^2) = 0$ for all $x \in W^-_{loc}(0) \cap U_0$ where U_0 is a sufficiently small neighborhood of 0. Thus, identifying the monomials in powers of $\|x\|$ in both sides, we get

$$\frac{\partial h_s}{\partial x}(0)Ax = 0 \quad \text{for all } x \in W^-_{loc}(0) \cap U_0. \tag{3.10}$$

If we denote by $\{e_1^+, \ldots, e_k^+\}$ a basis of E^+ made by eigenvectors of A associated to the eigenvalues λ_i^+, $i = 1, \ldots, k$, with positive real part, and by $\{e_1^-, \ldots, e_{n-k}^-\}$ a basis of E^- made of eigenvectors of A associated to the eigenvalues λ_i^-, $i = 1, \ldots, n-k$, with negative real part, we have $x = \sum_{i=1}^{n-k} x_i^- e_i^-$ for all $x \in E^- \cap U_0$. Thus, restricting (3.10) to $x \in E^- \cap U_0$, we get

$$\begin{aligned} 0 = \frac{\partial h_s}{\partial x}(0)Ax &= \frac{\partial h_s}{\partial x}(0)\left(\sum_{i=1}^{n-k} Ax_i^- e_i^-\right) = \frac{\partial h_s}{\partial x}(0)\left(\sum_{i=1}^{n-k} x_i^- Ae_i^-\right) \\ &= \frac{\partial h_s}{\partial x}(0)\left(\sum_{i=1}^{n-k} x_i^- \lambda_i^- e_i^-\right) \end{aligned}$$

where we have used the relation $Ae_i^- = \lambda_i^- e_i^-$, which proves that every element of E^- (generated by the combinations $\sum_{i=1}^{n-k} x_i^- \lambda_i^- e_i^-$, with arbitrary x_i^-'s) belongs to the kernel of $\frac{\partial h_s}{\partial x}(0)$. This kernel has dimension k by assumption and $\dim E^- = k$, thus $\ker \frac{\partial h_s}{\partial x}(0) = E^-$. Since, by definition, $\ker \frac{\partial h_s}{\partial x}(0) = \mathrm{T}_0 W^-_{loc}(0)$, the tangent space at 0 to $W^-_{loc}(0)$, the result is proven.

An example, in dimension 3, of 2-dimensional stable and 1-dimensional unstable manifolds, as in Proposition 3.4, are depicted in Figure 3.6.

Example 3.13. The system

$$\begin{cases} \dot{x}_1 = x_1 + x_1 x_2 \\ \dot{x}_2 = -x_2 \end{cases} \tag{3.11}$$

has the origin as equilibrium point and its tangent linear system at 0 is $\dot{z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} z$, its two eigenvalues being $+1$ and -1.

Clearly, starting from an initial condition such that $x_1(0) = 0$, we have $\dot{x}_1(0) = 0$, and thus $x_1(t) = 0$ for all t. Moreover, since $x_2 = e^{-t}x_2(0)$, the flow $(x_1(t), x_2(t))$ converges to the origin as $t \to +\infty$, thus $\{(0, x_2) | x_2 \in \mathbb{R}\} \subset W^-_{loc}(0)$. But since we are looking for a manifold $W^-_{loc}(0)$ of the same dimension as the eigenspace associated to the eigenvalue -1, namely 1, $W^-_{loc}(0)$ must

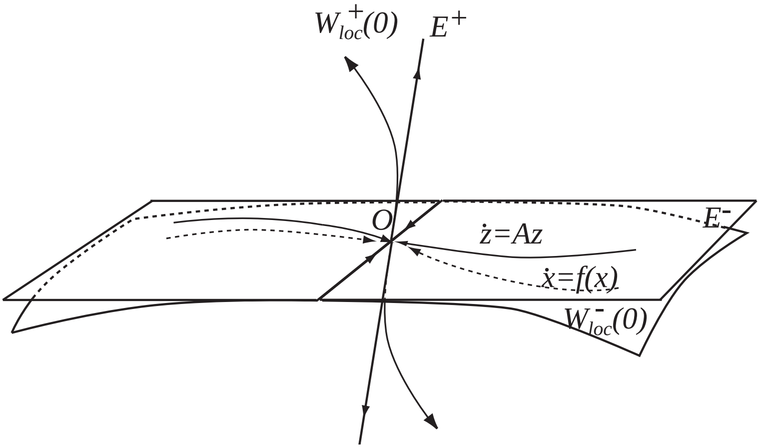

Fig. 3.6 The topological equivalence between the integral curves of f and those of its tangent linear approximation. The 2-dimensional stable manifold $W_{loc}^{-}(0)$ is tangent to E^{-} at 0 and the 1-dimensional unstable manifold $W_{loc}^{+}(0)$ is tangent to E^{+} at 0

be tangent to this eigenspace, which is precisely $\{(0,x_2)|x_2 \in \mathbb{R}\}$, therefore we have $\{(0,x_2)|x_2 \in \mathbb{R}\} = W_{loc}^{-}(0)$.

Accordingly, starting from an initial condition satisfying $x_2(0) = 0$, we have $x_2(t) = 0$ for all t and $\dot{x}_1 = x_1$, or $x_1 = e^t x_1(0)$. Thus $W_{loc}^{+}(0) = \{(x_1,0)|x_1 \in \mathbb{R}\}$.

Let us prove that (3.11) is topologically equivalent to its tangent linear system by posing $\tau(t,z_2) = t-(e^{-t}-1)z_2$ for $z_2 > -1$ and $\psi(z_1,z_2) = (z_1,z_2)$. Indeed, rewriting (3.11) as: $\frac{\dot{x}_1}{x_1} = 1 - \dot{x}_2$, or $x_1 = x_1(0)e^{t-(x_2-x_2(0))}$, and, combining this expression with $x_2 = e^{-t}x_2(0)$ obtained from $\dot{x}_2 = -x_2$, we get $x_1 = x_1(0)e^{\tau(t,x_2(0))}$ and $x_2 = x_2(0)e^{-t}$. Comparing with the integral curves of the tangent linear system: $z_1 = e^t z_1(0)$, $z_2 = e^{-t}z_2(0)$, we have: $x_1(t) = \psi_1(z_1(\tau(t,z_2(0))), z_2(t)) = z_1(\tau(t,z_2(0)) = e^{\tau(t,z_2(0)}z_1(0)$ and $x_2(t) = \psi_2(z_1(\tau(t,z_2(0))), z_2(t)) = z_2(t) = e^{-t}z_2(0)$, which achieves to prove the assertion by remarking that the sense of motion is preserved if and only if $\frac{\partial\tau}{\partial t} = 1 + e^{-t}z_2 > 0$, in other words if $z_2 > -1$. The equivalence is thus valid in the neighborhood $\mathbb{R}\times]-1,+\infty[$ of the origin.

In the case of a diffeomorphism, the previous definition becomes:

Definition 3.3. We call *local stable manifold* of the diffeomorphism f at the fixed point 0, the submanifold

$$W_{loc}^{-}(0) = \{x \in U \mid \lim_{k\to+\infty} f^k(x) = 0 \text{ and } f^k(x) \in U\ \forall k \geq 0\} . \qquad (3.12)$$

We call *local unstable manifold* of the diffeomorphism f at the fixed point 0, the submanifold

$$W^+_{loc}(0) = \{x \in U \mid \lim_{k \to +\infty} f^{-k}(x) = 0 \text{ and } f^{-k}(x) \in U \ \forall k \geq 0\} \ . \qquad (3.13)$$

We have the following result:

Theorem 3.6. (Hartman and Grobman) *If 0 is a hyperbolic equilibrium (resp. fixed) point of f, then f is topologically equivalent to its tangent linear system and the corresponding homeomorphism preserves the sense of motion. Moreover, there exist local stable and unstable manifolds of f at 0 with* $\dim\ W^+_{loc}(0) = \dim\ E^+$ *and* $\dim\ W^-_{loc}(0) = \dim\ E^-$, *tangent at the origin to E^+ and E^- respectively, and having the same regularity as f.*

The second part of the Theorem is often called Hadamard–Perron's Theorem as in the linear case.

Remark that these manifolds may be prolonged using the following formulas, in the continuous-time case:

$$W^-(0) = \bigcup_{t \geq 0} X_{-t}(W^-_{loc}(0)) \ , \quad W^+(0) = \bigcup_{t \geq 0} X_t(W^+_{loc}(0)) \qquad (3.14)$$

and in the discrete-time case

$$W^-(0) = \bigcup_{k \geq 0} f^{-k}(W^-_{loc}(0)) \ , \quad W^+(0) = \bigcup_{k \geq 0} f^k(W^+_{loc}(0)) \ . \qquad (3.15)$$

This decomposition is not valid for a matrix A having 0 or imaginary eigenvalues (resp. of modulus 1). The extension to such cases is non trivial since it isn't only based on the tangent linear mapping, and may be done thanks to the introduction of the notion of *centre manifold*.

If 0 is a non hyperbolic equilibrium (resp. fixed) point of f, we consider the eigenspace E^0 associated to the eigenvalues with 0 real part (resp. of modulus 1). We thus have $E^+ \oplus E^0 \oplus E^- = \mathbb{R}^n$ and E^0 is also invariant by A.

Theorem 3.7. (Shoshitaishvili) *If the vector field (resp. diffeomorphism) f is of class C^r and admits 0 as equilibrium (resp. fixed) point, it admits local stable, unstable and centre manifolds, noted $W^-_{loc}(0)$, $W^+_{loc}(0)$ and $W^0_{loc}(0)$, of class C^r, C^r and C^{r-1} respectively, tangent at the origin to E^-, E^+ and E^0. $W^-_{loc}(0)$ and $W^+_{loc}(0)$ are uniquely defined whereas $W^0_{loc}(0)$ is not necessarily unique.*
Moreover, f is topologically equivalent, in a neighborhood of the origin, to the vector field $-x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2} + f_0(x_3) \frac{\partial}{\partial x_3}$ (resp. to the diffeomorphism $(x_1, x_2, x_3) \longrightarrow (\frac{1}{2}x_1, 2x_2, f_0(x_3))$) where x_1, x_2 and x_3 are systems of local coordinates of $W^-_{loc}(0)$, $W^+_{loc}(0)$ and $W^0_{loc}(0)$ respectively, and $f_0(x_3)$ is the projection of f on $W^0_{loc}(0)$.

Remark 3.7. It should be noted that, as opposed to the fact that the stable and unstable dynamics that copy their tangent linear dynamics, the dynamics in the centre manifold, if non trivial, is nonlinear and therefore concentrates the nonlinear aspects of f. Thus, locally, for the topological equivalence relation, the only nonlinear dynamics are those corresponding to centre manifolds.

To compute centre manifolds, the following general principle may be applied. We assume, for simplicity's sake, that the vector field f is of class C^∞. We start by the decomposition of A in Jordan form. In the corresponding coordinates, the system reads

$$\begin{cases} \dot{x} = A_\pm x + f_1(x, y) \\ \dot{y} = A_0 y + f_2(x, y) \end{cases} \tag{3.16}$$

where $A_\pm$ is the diagonal part corresponding to the hyperbolic eigenvalues of A, and A_0 the block grouping all the eigenvalues with 0 real part of A, that correspond to the subspace E^0. Since the origin is an equilibrium point, we may have $f_i(0, 0) = 0$, $i = 1, 2$ and since the above system must have

$$\begin{aligned} \dot{z}_1 &= A_\pm z_1 \\ \dot{z}_2 &= A_0 z_2 \end{aligned}$$

as its tangent linear system, it results that $\dfrac{\partial f_i}{\partial x}(0, 0) = \dfrac{\partial f_i}{\partial y}(0, 0) = 0$, $i = 1, 2$.

Since the coordinates y are adapted to the centre manifold $W^0_{loc}(0)$, its equation must be of the form

$$x = h(y), \tag{3.17}$$

where h is a C^∞ function to be determined. We indeed must have $h(0) = 0$ due to the equilibrium at the origin. To express the fact that $W^0_{loc}(0)$ must be tangent to E^0 at 0, we set $\varphi(x, y) = x - h(y)$. The tangent space to $W^0_{loc}(0)$ at the origin is

$$\ker D\varphi(0, 0) = \{(v_1, v_2) | v_1 - \frac{\partial h}{\partial y}(0) v_2 = 0\} = E_0 = \{(0, v_2) | v_2 \in \mathbb{R}^{k_0}\}$$

with $k_0 = \dim E_0$, which immediately implies that $\dfrac{\partial h}{\partial y}(0) = 0$.

The dynamics on the centre manifold must satisfy:

$$\dot{y} = A_0 y + f_2(h(y), y) \tag{3.18}$$

where $f_2(h(y), y) \stackrel{\text{def}}{=} \tilde{f}_2(y)$, similarly to h, satisfies $\tilde{f}_2(0) = 0$ and $\frac{\partial \tilde{f}_2}{\partial y}(0) = 0$. In other words, if we develop $\tilde{f}_2$ and h in Taylor's series around 0, the lowest degree non zero terms must be of the second degree with respect to y. We note this property $f_2(h(y), y) = 0(\|y\|^2)$ and $h(y) = 0(\|y\|^2)$ in a suitable neighborhood of 0. The tangent vector field to $W^0_{loc}(0)$ of (3.18) is thus given, up to the second order in y, by $A_0 y + f_2(0, y) + 0(\|y\|^3)$.

To identify the successive terms of the development of h, we use the invariance of $W^0_{loc}(0)$ by the flow, *i.e.* $\frac{d}{dt}(x - h(y)) = 0$, or:

$$A_\pm h(y) + f_1(h(y), y) - \frac{\partial h}{\partial y}(A_0 y + f_2(h(y), y)) = 0. \tag{3.19}$$

We immediately deduce

Proposition 3.5. *Consider the solution h_0 of the implicit equation*

$$A_\pm h_0(y) + f_1(h_0(y), y) = 0 \tag{3.20}$$

such that $h_0(0) = 0$, in a suitable neighborhood of the origin[8]*. Then h, defined by (3.17), is approximated by h_0 at the order 2 with respect to y in a sufficiently small neighborhood $\mathcal{Y}_0$ of $y = 0$, i.e. $\|h(y) - h_0(y)\| = 0(\|y\|^3)$ for every $y \in \mathcal{Y}_0$.*

Proof. According to (3.19), since $\frac{\partial h}{\partial y}(0) = 0$ by assumption, $\frac{\partial h}{\partial y}(y) = \frac{\partial h}{\partial y}(0) + 0(\|y\|^2) = 0(\|y\|^2)$, and we have $A_\pm h(y) + f_1(h(y), y) - \frac{\partial h}{\partial y}(y)(A_0 y + f_2(h(y), y)) = 0 = A_\pm h(y) + f_1(h(y), y) + 0(\|y\|^3)$. Thus $A_\pm h(y) + f_1(h(y), y) = 0(\|y\|^3)$, which proves that h_0 satisfying (3.20) coincides with h up to the second order in y, hence the result.

Remark 3.8. To compute an approximation of the center manifold, it suffices to solve (3.20), which represents the set of equilibria of the hyperbolic part of (3.16) for all fixed y in a suitable neighborhood of the origin.
Assuming, for simplicity's sake, that $A_\pm = A_-$ has only eigenvalues with negative real part, this leads to the following important interpretation: the hyperbolic part, which is topologically equivalent to its stable tangent dynamics $\dot{z}_1 = A_- z_1$, exponentially converges to $x = h_0(y)$, close to the center manifold $x = h(y)$, while y, whose dynamics in not exponential, moves so slowly in comparison to x that it can be approximated by a constant.
On the other hand, the dynamics of the non hyperbolic part on the center manifold is locally approximated by $A_0 y + f_2(h_0(y), y)$ and is independent of the hyperbolic dynamics.

Theorem 3.7 has two major consequences, one concerning the stability analysis, and the other one the possibility to reduce the system's dimension:

Even if we assume, as in the previous Remark, that the hyperbolic dynamics are stable, the overall stability depends on the stability of the central dynamics (projection of the dynamics on the centre manifold):

---

[8] this solution exists and is locally unique since $A_\pm$ is invertible

- the central flow, if stable, converges to $y = 0$. Thus x first locally converges to a point y close to the center manifold and then slowly converges to 0 while remaining close to this manifold.
- in the unstable case, y diverges, implying that x first converges to a point y close to the center manifold and then is slowly repelled from the origin while remaining close to this manifold.

If we are interested in approximating the system, its projection on the center manifold provides a model of reduced dimension which faithfully reproduces the slow behavior of the system. We will see in the next chapters that such model reduction may be helpful, for instance, if the hyperbolic (fast) dynamics are stable and if we want to locally modify the stability of the central flow only. This approach may be seen as a "surgical operation" on the system since we only modify the unwanted behavior on a submanifold almost without affecting the rest.

If we set $h(y) = h^{(2)}y^{\otimes 2} + 0(\|y\|^3)$, $f_i(0,y) = f_i^{(2)}y^{\otimes 2} + 0(\|y\|^3)$, where $y^{\otimes 2}$ represents the vector made up with all the monomials of the form $y_i y_j$ for $1 \leq i \leq j \leq \dim E_0$, we easily see that $h^{(2)}$ satisfies the equation:

$$A_{\pm}h^{(2)}y^{\otimes 2} + f_1^{(2)}y^{\otimes 2} = \frac{\partial h^{(2)}y^{\otimes 2}}{\partial y}A_0 y$$

and we go on, by the same method, identifying the successive coefficients of the development of h (see Example 3.14 below).

Remark 3.9. This process gives only a polynomial approximation of h and thus of the centre manifold, and, even if we can compute an infinite number of terms, the corresponding series is not necessarily convergent. Nevertheless, this approach is constructive, and therefore useful in applications, and directly shows the key role played by the invariance.

Example 3.14. Consider the following system in 2 dimensions

$$\begin{cases} \dot{x}_1 = -x_1 + f_1(x_1,x_2) \\ \dot{x}_2 = f_2(x_1,x_2) \end{cases} \tag{3.21}$$

where f_1 and f_2 satisfy: $f_i(0,0) = 0$ and $\dfrac{\partial f_i}{\partial x_j}(0,0) = 0$ for $i,j = 1,2$.
The tangent linear mapping is thus given by

$$\dot{z} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} z \stackrel{\text{def}}{=} Az. \tag{3.22}$$

The two eigenvalues of A are $\lambda_1 = -1$ and $\lambda_2 = 0$. According to Theorem 3.7, there is a centre manifold $x_1 = h(x_2)$ that we now construct. As before, let us set $h(x_2) = h_2 x_2^2 + h_3 x_2^3 + O(x_2^4)$ in a neighborhood of 0. We have:

$$\dot{x}_1 = -x_1 + f_1(x_1, x_2) = \frac{dh}{dx_2} f_2(x_1, x_2).$$

Replacing x_1 by $h(x_2)$, we get

$$-h(x_2) + f_1(h(x_2), x_2) = \frac{dh}{dx_2} f_2(h(x_2), x_2). \tag{3.23}$$

The expansion at the order 3 in x_2 of $f_i((h(x_2), x_2)$ is given by:

$$f_i((h(x_2), x_2) = \frac{1}{2}\frac{\partial^2 f_i}{\partial x_2^2}(0)x_2^2 + \left(\frac{\partial^2 f_i}{\partial x_1 \partial x_2}(0)h_2 + \frac{1}{6}\frac{\partial^3 f_i}{\partial x_2^3}(0)\right)x_2^3 + O(x_2^4)$$
$$i = 1, 2.$$

Identifying the monomials of degree 2 and 3 in (3.23), we obtain (after simple but long computations which are omitted):

$$x_1 = h(x_2) = \frac{1}{2}\frac{\partial^2 f_1}{\partial x_2^2}(0)x_2^2 + \left(\frac{1}{2}\left(\frac{\partial^2 f_1}{\partial x_1 \partial x_2}(0) - \frac{\partial^2 f_2}{\partial x_2^2}(0)\right)\frac{\partial^2 f_1}{\partial x_2^2}(0) + \frac{1}{6}\frac{\partial^3 f_1}{\partial x_2^3}(0)\right)x_2^3 + O(x_2^4)$$

and the dynamics of x_2 (central dynamics) are given by:

$$\dot{x}_2 = \frac{1}{2}\frac{\partial^2 f_2}{\partial x_2^2}(0)x_2^2 + \left(\frac{1}{2}\frac{\partial^2 f_2}{\partial x_1 \partial x_2}(0)\frac{\partial^2 f_1}{\partial x_2^2}(0) + \frac{1}{6}\frac{\partial^3 f_2}{\partial x_2^3}(0)\right)x_2^3 + O(x_2^4).$$

The central dynamics thus has a quadratic term as lowest degree term. Let us denote by $a = \frac{1}{2}\frac{\partial^2 f_2}{\partial x_2^2}(0)$ and assume that $a \neq 0$. We have

$$x_2(t) = (x_2(0)^{-1} - at)^{-1}$$

for $x_2(0)$ sufficiently small. This solution is not L-stable in a neighborhood of 0 since the integral curves converge to 0 only if the sign of $x_2(0)$ is the opposite of the one of a, and blow up in finite time $T = \frac{1}{ax_2(0)}$ in the opposite case. If $a = 0$, The central dynamics are given at the order 3 in x_2 by

$$\dot{x}_2 = \left(\frac{1}{2}\frac{\partial^2 f_2}{\partial x_1 \partial x_2}(0)\frac{\partial^2 f_1}{\partial x_2^2}(0) + \frac{1}{6}\frac{\partial^3 f_2}{\partial x_2^3}(0)\right)x_2^3 + O(x_2^4)$$

and is stable if $\left(\frac{\partial^2 f_2}{\partial x_1 \partial x_2}(0)\frac{\partial^2 f_1}{\partial x_2^2}(0) + \frac{1}{3}\frac{\partial^3 f_2}{\partial x_2^3}(0)\right) < 0$ and unstable otherwise.

It results from this analysis and from Theorem 3.7 that if $a > 0$ (resp. $a < 0$), the system (3.21) is not L-stable, and if $a = 0$ and $\left(\frac{\partial^2 f_2}{\partial x_1 \partial x_2}(0)\frac{\partial^2 f_1}{\partial x_2^2}(0) + \frac{1}{3}\frac{\partial^3 f_2}{\partial x_2^3}(0)\right) < 0$ the system (3.21) is stable for all x_2 sufficiently small.

As noted in Remark 3.8, there is a natural interpretation of the centre manifold in terms of time scales, which also suggests a strong link with the theory of singular perturbations as outlined in the next section.

3.3 Singularly Perturbed Systems

We consider a family of vector fields

$$\{f_\varepsilon : (x,\varepsilon) \in X \times [0,\varepsilon_0[\mapsto f(x,\varepsilon) \in \mathrm{T}_x X\}$$

where f_ε depends in a C^∞ way of the scalar parameter ε. We say that this family is a family of perturbations of $f = f_0$. The corresponding family of perturbed systems is thus the family of systems

$$\dot{x} = f(x,\varepsilon) \tag{3.24}$$

for every $\varepsilon \in [0,\varepsilon_0[$.

Augmenting the state by adding ε as a new state satisfying $\dot{\epsilon} = 0$, we get

$$\begin{aligned} \dot{x} &= f(x,\varepsilon) \\ \dot{\varepsilon} &= 0 \end{aligned} \tag{3.25}$$

Let us assume that $f(0,0) = 0$ and that, for $\epsilon = 0$, the tangent linear mapping of f admits the eigenspaces E^-, E^+ and E^0. The previous construction of local invariant manifolds therefore can be applied in a neighborhood of $(0,0)$: in the local coordinates $(Px,\epsilon) = (x_1,x_2,\epsilon)$, with P the matrix of eigenvectors of the linear tangent mapping of f, system (3.25) reads

$$\begin{cases} \dot{x}_1 = A_\pm x_1 + f_1(x_1,x_2,\epsilon) \\ \dot{x}_2 = A_0 x_2 + f_2(x_1,x_2,\epsilon) \\ \dot{\epsilon} \;\; = 0 \end{cases} \tag{3.26}$$

Then the center manifold is given by

$$x_1 = h(x_2,\epsilon) \tag{3.27}$$

with h given by its Taylor's series as before. According to Theorem 3.7, the system (3.25) is topologically equivalent to the system

$$\begin{cases} \dot{x}_1 = A_\pm x_1 \\ \dot{x}_2 = A_0 x_2 + f_2(h(x_2,\epsilon),x_2,\epsilon) \end{cases} \tag{3.28}$$

where the "slow variable" is x_2 (ϵ is indeed slowly varying since it is constant!) and the"fast variable" x_1. If the latter variable is stable for all x_2 in the centre manifold, the stability of the equilibrium point $(0,0)$ depends on the stability of its projection on the centre manifold. In addition, for ϵ sufficiently small, the centre manifold may be approximated by

$$x_1 = h(x_2, 0) + \frac{\partial h}{\partial \epsilon}(x_2, 0)\epsilon + \frac{\partial^2 h}{\partial \epsilon^2}(x_2, 0)\epsilon^2 + \dots$$

and the slow dynamics by

$$\dot{x}_2 = A_0 x_2 + f_2(h(x_2, 0), x_2, 0) + 0(\epsilon^2).$$

We now focus attention on a more general class of controlled systems (see *e.g.* Lévine and Rouchon [1994]), which, in adapted coordinates (not necessarily those in which the model has been obtained) is given by:

$$\begin{aligned} \dot{x}_1 &= \varepsilon f_1(x_1, x_2, u, \varepsilon) \\ \dot{x}_2 &= f_2(x_1, x_2, u, \varepsilon) \end{aligned} \tag{3.29}$$

where x_1 has dimension n_1, x_2 has dimension n_2 and ε is a "small" parameter, expressing the fact that the vector field f_2, assumed of class C^∞ with respect to all its arguments, is, in a neighborhood to be made precise later, much larger than f_1, also assumed of class C^∞. The system (3.29) is said to be in *standard form* (Tikhonov et al. [1980]).

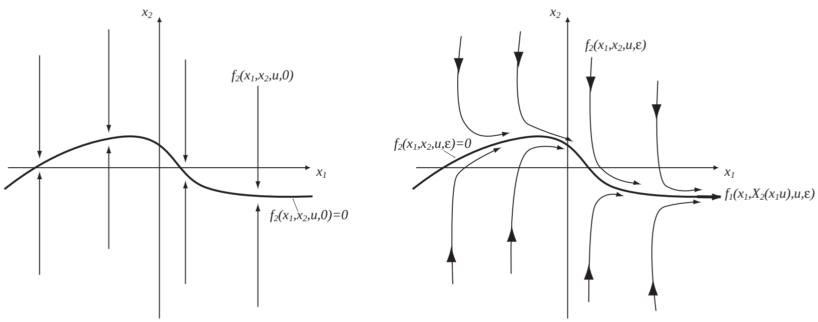

Fig. 3.7 The evolution of the variable x_2 is faster than the one of x_1. The left figure corresponds to $\varepsilon = 0$ in (3.29) and the right one corresponds to its deformation for $\varepsilon > 0$ small enough.

Let us introduce the new time variable $\tau = \varepsilon t$. τ is obviously a slower time variable than t since for all finite t, $\lim_{\varepsilon \to 0} \tau(t) = 0$. τ is often called *slow time-scale*.

We have $\frac{dx_1}{d\tau} = \frac{dx_1}{dt}\frac{dt}{d\tau} = \frac{1}{\varepsilon}\dot{x}_1 = f_1(x_1, x_2, u, \varepsilon)$ and $\frac{dx_2}{d\tau} = \frac{dx_2}{dt}\frac{dt}{d\tau} = \frac{1}{\varepsilon}\dot{x}_2 = \frac{1}{\varepsilon}f_2(x_1, x_2, u, \varepsilon)$, or

$$\frac{dx_1}{d\tau} = f_1(x_1, x_2, u, \varepsilon)$$
$$\frac{dx_2}{d\tau} = \frac{1}{\varepsilon} f_2(x_1, x_2, u, \varepsilon)$$

which means that, in this slow time scale, the second dynamics is of order $\frac{1}{\varepsilon}$, thus fast, whereas the first one is of order 0 in ε.

Before going deeper into the subject, let us remark that the standard form (3.29) crucially depends on the choice of coordinates. Indeed, if we introduce the diffeomorphic change of variables:

$$z_1 = x_1 + x_2, \quad z_2 = x_1 - x_2,$$

the resulting dynamics read

$$\dot{z}_1 = \varepsilon f_1(\frac{z_1 + z_2}{2}, \frac{z_1 - z_2}{2}, u, \varepsilon) + f_2(\frac{z_1 + z_2}{2}, \frac{z_1 - z_2}{2}, u, \varepsilon)$$
$$\dot{z}_2 = \varepsilon f_1(\frac{z_1 + z_2}{2}, \frac{z_1 - z_2}{2}, u, \varepsilon) - f_2(\frac{z_1 + z_2}{2}, \frac{z_1 - z_2}{2}, u, \varepsilon)$$

and in the two components of $z = (z_1, z_2)$, fast and slow terms are mixed up. These coordinates are therefore not suitable to directly detect if the system is decomposable in slow and fast parts. The questions of the existence of coordinates in which the system is in standard form (3.29) and their computation are beyond the scope of this book and are not addressed here.

Coming back to (3.29), the fact that x_2's dynamics are fast means that x_2 is vigorously attracted to an equilibrium point of f_2, in a neighborhood to be specified, if the latter is stable, or vigorously repelled from this point in the opposite case. The first step of our analysis thus naturally concerns the submanifold where the fast dynamics converge, which, analogously to Proposition 3.5, provides a suitable approximation of the center manifold. The latter, in this context, is called *slow manifold* and is constructed using its invariance, as in the previous section.

3.3.1 Invariant Slow Manifold

When $\varepsilon = 0$, the set of equilibrium points of (3.29) is the manifold given by

$$\Sigma_0 = \{(x_1, x_2, u) | f_2(x_1, x_2, u, 0) = 0\}$$

called the *equilibrium manifold*, whereas, for $\varepsilon \neq 0$, this set is given by

$$\Sigma_\varepsilon = \{(x_1, x_2, u) | f_1(x_1, x_2, u, \varepsilon) = 0, f_2(x_1, x_2, u, \varepsilon) = 0\}.$$

It is remarkable that Σ_ε is lower dimensional than Σ_0 since, for $\varepsilon = 0$, the condition $f_1 = 0$ disappears. This justifies the expression *singular pertur-*

bation: when $\varepsilon \to 0$, the manifold Σ_ε degenerates to Σ_0 who is therefore singular.

By the implicit function Theorem, if, in a neighborhood of Σ_0, $\frac{\partial f_2}{\partial x_2}$ has full rank, equal to n_2, the dimension of x_2, one has

$$\Sigma_0 = \{(x_1, x_2, u) | x_2 = X_2(x_1, u)\}.$$

The equilibrium manifold Σ_0 obviously enjoys the following property: for initial conditions (x_1^0, x_2^0, u_0) in Σ_0, and if $\varepsilon = 0$, the integral curves of the system remain in Σ_0 since $(\dot{x}_1, \dot{x}_2) = (0, 0)$. Σ_0 is thus an *invariant manifold* for $\varepsilon = 0$.

We now study the persistance of this invariant manifold when $\varepsilon \neq 0$, sufficiently small, and when u is slowly varying.

3.3.2 Persistence of the Invariant Slow Manifold

We will show that for all $\varepsilon \neq 0$ sufficiently small, there exists an invariant manifold denoted by $\Sigma_{0,\varepsilon}$, close to Σ_0, called *slow manifold*, on which the dynamics of the system is of order ε, *i.e.* slow.

From now on, we assume that u is slowly varying, *i.e.* $\dot{u} = \varepsilon v$ with v a bounded, but otherwise arbitrary, time function, satisfying $\sup_{t \in \mathbb{R}} \|v(t)\| < +\infty$. We set

$$\Sigma_{0,\varepsilon} = \{(x_1, x_2) | f_2(x_1, x_2, u, \varepsilon) = 0\}.$$

Clearly, the regularity of f_2 with respect to ε implies that $\Sigma_{0,\varepsilon}$ remains close to Σ_0 for small ε.

Let us assume that $\text{rank}\left(\frac{\partial f_2}{\partial x_2}\right) = n_2$ for all ε sufficiently small. By the implicit function theorem, there exists a function $\overline{X}_2$, C^∞ in all its arguments, such that for all u satisfying $\dot{u} = \varepsilon v$, $(x_1, x_2) \in \Sigma_{0,\varepsilon}$ is equivalent to $x_2 = \overline{X}_2(x_1, u, \varepsilon)$ for all x_1 and u in a suitable neighborhood, or, in other words, that $f_2(x_1, \overline{X}_2(x_1, u, \varepsilon), u, \varepsilon) = 0$.

Let us prove that $\Sigma_{0,\varepsilon}$ is invariant by the *slow dynamics*

$$\begin{aligned} \dot{x}_1 &= \varepsilon f_1(x_1, \overline{X}_2(x_1, u, \varepsilon), u, \varepsilon) \\ \dot{u} &= \varepsilon v. \end{aligned} \tag{3.30}$$

Since $\frac{d}{dt} x_2 = f_2(x_1, \overline{X}_2(x_1, u, \varepsilon), u, \varepsilon) = 0 = \frac{d}{dt}\left(\overline{X}_2(x_1, u, \varepsilon)\right)$ on $\Sigma_{0,\varepsilon}$, we get:

$$\frac{\partial \overline{X}_2}{\partial x_1} \varepsilon f_1(x_1, \overline{X}_2(x_1, u, \varepsilon), u, \varepsilon) + \frac{\partial \overline{X}_2}{\partial u} \varepsilon v = 0$$

and thus $\overline{X}_2$ is a first integral of $\varepsilon\left(f_1 \frac{\partial}{\partial x_1} + v \frac{\partial}{\partial u}\right)$, which proves the invariance with respect to the slow dynamics.

Therefore, the invariant manifold $x_2 = X_2(x_1, u)$ obtained, for $\varepsilon = 0$, as a zero of the vector field f_2, still exists for $\varepsilon > 0$ sufficiently small and remains invariant with respect to the slow dynamics (3.30). This property is called *persistence.*

Moreover, we can obtain an approximation of $\overline{X}_2$ and of the slow dynamics (3.30) polynomial in ε. Let us sketch their computation.

We set $\overline{X}_2(x_1, u, \varepsilon) = X_2(x_1, u) + \varepsilon X_2'(x_1, u) + 0(\varepsilon^2)$. At the first order in ε, the slow dynamics is given by

$$\dot{x}_1 = \varepsilon f_1(x_1, X_2(x_1, u), u, 0) + 0(\varepsilon^2). \tag{3.31}$$

Theorem 3.8. *For every ε sufficiently small, if u satisfies $\dot{u} = \varepsilon v$ with $\sup_{t\in\mathbb{R}} \|v(t)\| < +\infty$, and if the matrix $\frac{\partial f_2}{\partial x_2}$ has all its eigenvalues with strictly negative real part at every point $(x_1, x_2, u, 0)$ of an open neighborhood $V(\Sigma_0)$ of Σ_0, a first order approximation in ε of the invariant manifold $\Sigma_{0,\varepsilon}$, in this neighborhood, is given by:*

$$\begin{aligned} x_2 &= X_2(x_1, u) \\ &\quad + \varepsilon \left(\frac{\partial f_2}{\partial x_2}\right)^{-1} \left(\frac{\partial X_2}{\partial x_1} f_1 + \frac{\partial X_2}{\partial u} v - \frac{\partial f_2}{\partial \varepsilon}\right)(x_1, X_2(x_1, u), u, 0) \\ &\quad + 0(\varepsilon^2) \end{aligned} \tag{3.32}$$

a first order approximation of the slow dynamics in ε being given by (3.31).

Proof. Assuming that $\Sigma_{0,\varepsilon}$ is given, in a suitable neighborhood of Σ_0, at the first order in ε, by $x_2 = X_2(x_1, u) + \varepsilon X_2'(x_1, u, v) + 0(\varepsilon^2)$ with X_2' sufficiently differentiable in all its arguments in the considered neighborhood. By invariance at the first order in ε, we mean that the function $x_2 - X_2(x_1, u) - \varepsilon X_2'(x_1, u, v)$ is of order $0(\varepsilon^2)$ along the system flow:

$$\begin{aligned} \dot{x}_1 &= \varepsilon f_1(x_1, X_2(x_1, u), u, 0) + 0(\varepsilon^2) \\ \dot{x}_2 &= f_2(x_1, X_2(x_1, u) + \varepsilon X_2'(x_1, u, v), u, \varepsilon) + 0(\varepsilon^2) \\ \dot{u} &= \varepsilon v. \end{aligned} \tag{3.33}$$

It results from this property, and the fact that $f_2(x_1, X_2(x_1, u), u, 0) = 0$, that:

$$\begin{aligned} \dot{x}_2 = \varepsilon \frac{\partial X_2}{\partial x_1}(x_1, u) f_1(x_1, X_2(x_1, u), u, 0) + \varepsilon \frac{\partial X_2}{\partial u}(x_1, u) v + 0(\varepsilon^2) \\ = \varepsilon \frac{\partial f_2}{\partial x_2}(x_1, X_2(x_1, u), u, 0) X_2'(x_1, u, v) \\ + \varepsilon \frac{\partial f_2}{\partial \varepsilon}(x_1, X_2(x_1, u), u, 0) + 0(\varepsilon^2) \end{aligned}$$

from which we get, by identification of the terms of order 1 in ε:

$$\frac{\partial X_2}{\partial x_1} f_1 + \frac{\partial X_2}{\partial u} v = \frac{\partial f_2}{\partial x_2} X_2' + \frac{\partial f_2}{\partial \varepsilon}$$

which, thanks to the invertibility of $\frac{\partial f_2}{\partial x_2}$, gives X_2' and (3.32), which proves the result.

Going back to the interpretation in terms of slow and fast dynamics: in the slow time scale $\tau = \varepsilon t$, the system (3.33) reads:

$$\begin{aligned} \frac{dx_1}{d\tau} &= f_1(x_1, X_2(x_1, u), u, 0) + 0(\varepsilon) \\ \frac{dx_2}{d\tau} &= \frac{1}{\varepsilon} f_2(x_1, X_2(x_1, u) + \varepsilon X_2'(x_1, u, v), u, 0) + 0(\varepsilon). \end{aligned}$$

The invariant manifold, assumed attractive, thus plays the following role: for every initial condition close to $\Sigma_{0,\varepsilon}$, the fast dynamics (of order $\frac{1}{\varepsilon}$) converge exponentially fast to $\Sigma_{0,\varepsilon}$ and then give way to the slow dynamics whose flow, at least in a small time interval, remains on $\Sigma_{0,\varepsilon}$ (invariant manifold). Moreover, the fast and slow flows, for small ε, are locally small deformations of the respective fast and slow flows for $\varepsilon = 0$ (see Figure 3.7). This interpretation may also be reformulated as the so-called *Shadow Lemma*: for every initial condition in a neighborhood of $\Sigma_{0,\varepsilon}$, the system flow converges to $\Sigma_{0,\varepsilon}$, without necessarily belonging to $\Sigma_{0,\varepsilon}$, but is approximated by a flow remaining in $\Sigma_{0,\varepsilon}$ (its shadow), to which it converges.

We now address the question of the robustness of the stability of the system with respect to ε.

3.3.3 Robustness of the Stability

In the limiting case $\varepsilon = 0$, even if the fast dynamics (corresponding to the x_2 coordinate) is stable, the system is not hyperbolic, though it is for $\varepsilon \neq 0$. If we assume in addition that the slow dynamics is stable, one may want to know if the overall system is asymptotically stable on the one hand, and if one can compare it with the overall dynamics for $\varepsilon \neq 0$, which is hyperbolically stable, on the other hand. If the stability is unchanged when $\varepsilon \to 0$, we say that the stability is *robust*.

The following result is given without proof:

Theorem 3.9. *If, as before, the matrix* $\frac{\partial f_2}{\partial x_2}$ *has all its eigenvalues with strictly negative real part at every point* $(x_1, x_2, u, 0)$ *of an open neighborhood* $V(\Sigma_0)$ *of* Σ_0 *and if, denoting by*

$$F_1(x_1, u) = f_1(x_1, X_2(x_1, u), u, 0),$$

the matrix $\frac{\partial F_1}{\partial x_1}$ *has all its eigenvalues with strictly negative real part at every point* (x_1, u) *of a neighborhood* V_1 *such that* $V_1 \times X_2(V_1)$ *contains* Σ_0*, then* *the system (3.29) is asymptotically stable for all $\varepsilon \geq 0$ sufficiently small, i.e. its stability is robust.*

Let us stress on the fact that this result is not as obvious as it seems to be: if two systems are independent, the stability of each one of them implies the overall stability. Nevertheless, this result is no longer valid if the two systems are coupled since the coupling, even small, may become dominant in some neighborhood and may be in conflict with the stable uncoupled part of the dynamics. We are here in an "almost" decoupled case in a neighborhood of Σ_0, which allows us to use the fact that a stable system remains stable under small perturbations.

3.3.4 An Application to Modelling

The persistence of the invariant manifold and the robustness of its stability have an important consequence concerning the modelling of a system with fast and stable dynamics. Indeed, in the slow dynamics, in a first order approximation, we do not use the information on the fast dynamics. The only required knowledge of the latter is an approximation of the equilibrium manifold if we use a sufficiently slow control ($\dot{u} = \varepsilon v$). In this case, the system dimension may be *reduced* by keeping only the slow dynamics (3.31), the original manifold being locally replaced by an approximation at the order 0, Σ_0, of the slow manifold.

This approach has at least two advantages if one restricts to sufficiently slow control: the system dynamics dimension is lowered and a precise knowledge of the fast dynamics is not required.

Example 3.15. We consider a DC motor made of permanent magnets (stator) and a coil (rotor) of inductance L and resistance R, in which there flows a current I, and fed by a variable voltage U. The electromotive force is assumed linear in the current and its torque constant is denoted by K. The rotor inertia is denoted by J and the resultant of the exterior torques applied to the motor shaft is denoted by C_r. Finally, we denote by K_v the constant of viscous friction, the latter being assumed linear with respect to the motor rotation speed ω. The electrical model of the coil is obtained thanks to Ohm's law, combined with the Lenz law for the electromotive force, and the mechanical one, expressing the torque balance on the motor shaft, thanks to the second principle of the dynamics:

$$\begin{aligned} L\frac{dI}{dt} &= U - RI - K\omega \\ J\frac{d\omega}{dt} &= KI - K_v\omega - C_r \ . \end{aligned} \tag{3.34}$$

For most of the motors of this type, the inductance L is small compared to the other constants. We thus set $L = \varepsilon$ and rewrite (3.34) in the fast time scale $\tau = \frac{t}{\varepsilon}$.

$$\frac{dI}{d\tau} = U - RI - K\omega$$
$$J\frac{d\omega}{d\tau} = \varepsilon(KI - K_v\omega - C_r)\ .$$

We immediately deduce that $x_1 = \omega$, $x_2 = I$, $u = U$, and, if we restrict to $\varepsilon = 0$, the manifold Σ_0 is given by $U - RI - K\omega = 0$, *i.e.* $I = \frac{U-K\omega}{R}$. The slow dynamics is thus given by $J\frac{d\omega}{d\tau} = \varepsilon(K(\frac{U-K\omega}{R}) - K_v\omega - C_r)$ or

$$J\frac{d\omega}{d\tau} = \varepsilon\left(-\left(\frac{K^2}{R} + K_v\right)\omega - C_r + \frac{K}{R}U\right).$$

In the slow time scale, we obtain the reduced slow model (of dimension 1 in place of 2 for the original model) which doesn't use the inductance L, but which requires, on the other hand, a precise knowledge of the torque constant K and of the ratio $\frac{K}{R}$:

$$J\frac{d\omega}{dt} = -\left(\frac{K^2}{R} + K_v\right)\omega - C_r + \frac{K}{R}U\ .$$

This slow approximation may be refined to deduce an identification method of the resisting torque if the inertia and the torque constant of the motor are sufficiently precisely known and if we measure the current in the coil, the applied voltage and the rotation speed of the motor.

The invariant manifold $\Sigma_{0,\varepsilon}$ at the order 1 in $\varepsilon = L$ is given by

$$I = \frac{U - K\omega}{R} + \frac{L}{R^2}\left[\frac{K}{J}\left(K\left(\frac{U-K\omega}{R}\right) - K_v\omega - C_r\right) - \dot{U}\right] + 0(L^2)$$

Applying the voltage $U = K\omega$, corresponding to the zero equilibrium current if L were equal to 0. The current I converges to the current I_0 of order 1 in L given by:

$$I_0 = -\frac{LK}{JR^2}\left(K_v\omega + C_r\right) - \frac{L}{R^2}\dot{U} + 0(L^2)$$

Thus, to compensate for the viscous friction and the resisting torque, we must correct the voltage U by tuning its rate $\dot{U}$: to ensure $I_0 = 0$, one must apply the slow voltage variation:

$$\dot{U} = -\frac{K}{J}\left(K_v\omega + C_r\right) + 0(L)\ .$$

Thus, even without a precise knowledge of L and R, this relation may be used to estimate the resisting torque by tuning $\dot{U}$ such that the current I_0

remains as small as possible and that the rotation speed ω vanishes:

$$C_r = \frac{J}{K}\dot{U} + 0(L).$$

This method may be useful, in particular, to estimate the dry friction force at motor start.

3.4 Application to Hierarchical Control

We propose now two control applications of the previous theory. The first one concerns the control of the slow dynamics of a singularly perturbed system, its fast dynamics being stable in a neighborhood of the equilibrium manifold and the second one consists in creating a singularly perturbed situation by the design of a suitable feedback and to control the fast and slow dynamics by a so-called hierarchical or cascaded controller.

3.4.1 Controlled Slow Dynamics

We consider the singularly perturbed control system (3.29) with the dimension of x_1 equal to m, the dimension of the input vector u. We assume that all the eigenvalues of $\frac{\partial f_2}{\partial x_2}(x_1, x_2, u, \varepsilon)$ in a neighborhood of Σ_0 have their real part smaller than $-\alpha$, where α is a strictly positive real number, for every $\varepsilon \in]0, \varepsilon_0[$. Its slow dynamics, for u satisfying $\dot{u} = \varepsilon v$, is given by

$$\begin{aligned}\dot{x}_1 &= \varepsilon f_1(x_1, X_2(x_1, u, \varepsilon), u, \varepsilon)\\ \dot{u} &= \varepsilon v\end{aligned}$$

where $x_2 = X_2(x_1, u, \varepsilon)$ is the equation of the invariant manifold $\Sigma_{0,\varepsilon} = \{f_2(x_1, x_2, u, \varepsilon) = 0\}$.

We assume that $\left(\frac{\partial f_1}{\partial x_2}\frac{\partial X_2}{\partial u} + \frac{\partial f_1}{\partial u}\right)$ is everywhere invertible in the considered neighborhood (it makes sense since $\dim x_1 = \dim u$).

Thus, given an at least twice differentiable reference trajectory $t \mapsto x_1^*(t)$, with $t \in [0, T]$, such that $\sup_{t\in[0,T]} \|\dot{x}_1^*(t)\| \leq C\varepsilon$ and $\sup_{t\in[0,T]} \|\ddot{x}_1^*(t)\| \leq C\varepsilon^2$, where C is a finite positive real number, according to the above invertibility assumption, possibly after decreasing T, one can find u^* and v^* such that

$$\begin{aligned}\dot{x}_1^* &= \varepsilon f_1(x_1^*, X_2(x_1^*, u^*, \varepsilon), u^*, \varepsilon)\\ \dot{u}^* &= \varepsilon v^*.\end{aligned}$$

Indeed, by the implicit function Theorem, there exists u^* such that

$$\frac{\dot{x}_1^*}{\varepsilon} = f_1(x_1^*, X_2(x_1^*, u^*, \varepsilon), u^*, \varepsilon)$$

and, differentiating this last expression, we find that

$$v^* = \left(\frac{\partial f_1}{\partial x_2}\frac{\partial X_2}{\partial u} + \frac{\partial f_1}{\partial u}\right)^{-1} \cdot \left(\frac{\ddot{x}_1^*}{\varepsilon^2} - \left(\frac{\partial f_1}{\partial x_2}\frac{\partial X_2}{\partial x_1} + \frac{\partial f_1}{\partial x_1}\right) f_1(x_1^*, X_2(x_1^*, u^*, \varepsilon), u^*, \varepsilon)\right)$$

these expressions being finite according to the assumption on the norm of the successive derivatives of x_1^*.

We now construct a feedback controller to follow x_1^*. To this aim, it suffices to set

$$\begin{aligned}\dot{x}_1 - \dot{x}_1^* &= -k\varepsilon(x_1 - x_1^*) \\ &= \varepsilon\left(f_1(x_1, X_2(x_1, u, \varepsilon), u, \varepsilon) - f_1(x_1^*, X_2(x_1^*, u^*, \varepsilon), u^*, \varepsilon)\right)\end{aligned}$$

to deduce the controller $u = U(x_1, x_1^*, \varepsilon)$ again by the implicit function Theorem and $v = V(x_1, u, x_1^*, v^*, \varepsilon)$ by differentiation of U.

It results that, as far as the obtained controller remains bounded, the slow variable x_1 exponentially tracks its (slow enough) reference and, according to Theorem 3.9, since U doesn't depend on the fast variable x_2, and thus doesn't modify the eigenvalues of the fast dynamics, we can conclude to the stability of the overall system.

A variant of this approach in the context of distillation columns may be found in Lévine and Rouchon [1991].

3.4.2 Hierarchical Feedback Design

Consider, for the sake of simplicity, the 2-dimensional system with a single input, of the form:

$$\begin{aligned}\dot{x}_1 &= f_1(x_1, x_2) \\ \dot{x}_2 &= f_2(x_1, x_2) + u.\end{aligned} \tag{3.35}$$

We assume that our objective is to assign the variable x_2 to a given reference trajectory $x_2^\star$, while remaining in a domain where $f_2(x_1, x_2)$ is bounded. We also assume that $\frac{\partial f_1}{\partial x_2}$ is everywhere invertible in this domain.

We use the *high gain* control $u = -\frac{k}{\varepsilon}(x_2 - \xi_2)$, for ε small enough and with k a finite positive real. The system (3.35) reads, after introducing the fast time scale $\tau = \frac{t}{\varepsilon}$,

$$\frac{dx_1}{d\tau} = \varepsilon f_1(x_1, x_2)$$
$$\frac{dx_2}{d\tau} = \varepsilon f_2(x_1, x_2) - k(x_2 - \xi_2)$$

which makes, according to Theorem 3.8, the manifold of equation $x_2 = \xi_2 + \frac{\varepsilon}{k} f_2(x_1, \xi_2)$ invariant at the order 1 in ε. Moreover, since $k > 0$, this manifold is attractive and the resulting slow dynamics is given, at the order 1 in ε, by

$$\frac{dx_1}{d\tau} = \varepsilon f_1(x_1, \xi_2)$$

which amounts to control the slow dynamics through the reference ξ_2, which may be thus interpreted as a fictitious control variable for the slow subsystem. Furthermore, the invariant manifold may approach the curve $x_2 = \xi_2$ arbitrarily close since ε, that is here a design variable, can be chosen as small as we want.

We have thus transformed a control problem for a 2-dimensional system into two control problems for 2 "cascaded" 1-dimensional subsystems. The fast control loop is generally called the *low-level control loop*, whereas the indirect feedback realized via the reference ξ_2, that will be specified below, is called the *high-level control loop*.

We complete the design of the high-level loop as in the slow control section: we choose a reference x_1^* to be tracked, using ξ_2 as the input, by requiring that

$$\frac{dx_1}{d\tau} - \frac{dx_1^*}{d\tau} = -K\varepsilon\,(x_1 - x_1^*)$$

and we deduce that ξ_2 must satisfy $-K(x_1 - x_1^*) = f_1(x_1, \xi_2) - f_1(x_1^*, x_2^*)$ with x_2^* such that $\dot{x}_1^* = f_1(x_1^*, x_2^*)$.

Remark that the high-level loop must be slow enough to preserve this feedback created slow-fast decoupling. Clearly, this construction extends to systems of arbitrary dimension if they have a triangular structure similar to (3.35).

3.4.3 Practical Applications

The main applications of hierarchical control concern the robust control design of actuators supposed to work in a wide range of situations, not necessarily planned by the manufacturer.

Take for instance the case of an electric drive supposed to automatically open a door. This actuator must produce a torque to produce a prescribed rotation of the door around its vertical axis. However, it is designed and manufactured independently of this particular application, and is supposed to work satisfactorily whatever the mechanical characteristics of the door (inertia, frictions, *etc.*) in a given power range. Thus, to open a door of given mass, inertia, resisting torque, in less than 1 second with a positioning precision of ± 1 degree of angle, the drive must be able to deliver the reference torques required for the positioning with a much smaller time constant to reduce the positioning error of the door's position with respect to its reference at every time.

An analogous design can be found, for instance,

- for the control of hydraulic jacks driving the orientation of the wing flaps of an aircraft, that control the aircraft attitude;
- for the control of electromagnetic valves that inject a fluid in the active dampers of a vehicle suspension, or the inlet products of a chemical reactor, at a reference flow rate;
- for the control of the current in the coils of an electromagnet used to levitate a rotating shaft of a vacuum pump;
- for the control of electric drives (AC or DC, synchronous or asynchronous) for the active positioning of mechanical systems such as electric windows, windshield wipers, positioning tables in 2 or 3 dimensions, cranes, robots, *etc.*
- *etc.*

Example 3.16. Consider a mixing unit for a chemical reactor driven by a DC motor. We aim at varying the rotation speed of the mixer in function of the slowly varying volume of reactants in the reactor.

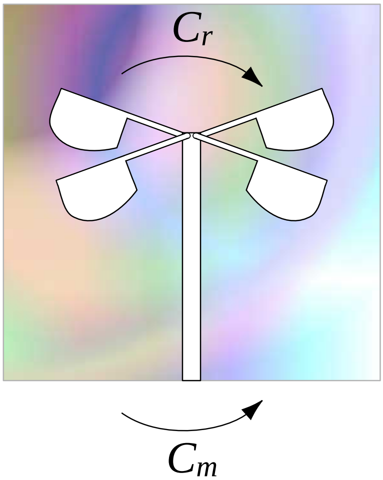

Fig. 3.8 Mixer in a chemical reactor

The model of the set motor-mixer consists of two differential equations, the first one describing the current evolution in the motor (see (3.34)) and the second one describing the mechanical torque balance exerted on the motor shaft. The equations are the same as those of (3.34), except that the resisting torque, which results from the friction of the helix on the fluid, is assumed proportional to the square of the angular speed of the helix: $C_r = K_f\omega^2$ where K_f is the viscous friction coefficient.

$$\begin{aligned} L\frac{dI}{dt} &= U - RI - K\omega \\ J\frac{d\omega}{dt} &= KI - K_f\omega^2. \end{aligned} \tag{3.36}$$

As before, we assume that the inductance L is much smaller than the other coefficients in the system. Let ω^* be the reference angular velocity, assumed constant or slowly varying.

We start with the low-level loop design:

$$U = -\frac{k_1}{L}(I - v_I)$$

with $k_1 > 0$ of order 0 in L, so that the current I fast converges (at the first order in L) to a reference v_I to be determined.

The resulting slow dynamics is thus

$$J\frac{d\omega}{dt} = Kv_I - K_f\omega^2.$$

The high-level loop, which is aimed at compensating the error between ω and its reference ω^*, consists in writing:

$$v_I = v_I^* - k_2(\omega - \omega^*)$$

with $Kv_I^* = K_f(\omega^*)^2$, assuming that K_f is precisely known. Indeed:

$$J\frac{d\omega}{dt} = -Kk_2(\omega - \omega^*) - K_f(\omega + \omega^*)(\omega - \omega^*) = -\left(Kk_2 + K_f(\omega + \omega^*)\right)(\omega - \omega^*)$$

which ensures the exponential convergence of ω to ω^* if $k_2 > 0$.

Thus, the angular velocity is directly controlled by the current, assuming that the latter is instantaneously equal to its reference, a behavior which results from the low-level voltage loop.

In a practical point of view, the low-level controller may be integrated in the motor's electronic driver since it only requires at every time the measurement of the current I and the input v_I to feed the voltage loop, the corresponding gain k_1 being tuned once for all, independently of this precise application. It suffices, for the user, to enter the the value of the current reference v_I at every time, that results from the high-level loop design.

Accordingly, the control electronics of the high-level loop may be designed separately, and the gain k_2 may be tuned in order to produce a slow enough behavior compared to the one of the electrical part of the motor. Note that the tuning margin is generally large enough since the electrical time constant of the motor is generally smaller than 10^{-3} seconds whereas, for the mechanical time constant, we rarely need to react below 10^{-2} seconds.

Such a design is indeed extendable to all the above mentioned applications.

Chapter 4
Controlled Systems, Controllability

4.1 Linear System Controllability

Controllability is one of the so-called *structural* properties of systems depending on inputs. It has initially been studied in the framework of linear systems to describe the possibility of generating arbitrary motions. For nonlinear systems, several extensions are possible. They are outlined in the second part of this chapter.

Though controllability is not of direct importance in applications, it serves as an introduction to numerous practical problems such as, e.g. the motion planning problem.

As in the previous chapters, we insist on the existence of particular coordinates where the controllability property is directly understood, coordinates for which the system is expressed in a so-called canonical form. Though such coordinates always exist for linear controllable systems, the situation is much more difficult for nonlinear systems. Several notions of controllability in the nonlinear context are studied in the second part of this chapter. The links between some Lie algebraic properties and controllability are outlined.

For a more detailed presentation of the structural aspects of nonlinear systems, and in particular their controllability, observability and applications to feedback design, the reader may refer to Kailath [1980], Sontag [1998], Wolovich [1974], Wonham [1974] and for the underlying linear algebraic questions to Gantmacher [1966].

4.1.1 Kalman's Criterion

We consider the linear system

$$\dot{x} = Ax + Bu \tag{4.1}$$

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_4,

where $x \in \mathbb{R}^n$ is the state vector and $u \in \mathbb{R}^m$ is the input vector. Clearly, the matrix A is of size $n \times n$ and B of size $n \times m$.

Definition 4.1. We say that the pair (A, B) or, equivalently, that the system is *controllable*of linear systems if and only if, given a duration $T > 0$ and two arbitrary points x_0 and x_T of $\mathbb{R}^n$, there exists a piecewise continuous time function $t \mapsto \bar{u}(t)$ from $[0, T]$ to $\mathbb{R}^m$ such that the solution $\bar{x}(t)$ of (4.1) generated by $\bar{u}$ and with initial conditions $\bar{x}(0) = x_0$, satisfies $\bar{x}(T) = x_T$.

In other words:

$$e^{AT}x_0 + \int_0^T e^{A(T-t)}B\bar{u}(t)dt = x_T. \tag{4.2}$$

This property, which concerns the integral curves of the system, only depends in fact on the pair of matrices A and B as shown by the following criterion obtained by R.E. Kalman:

Theorem 4.1. (Kalman) *A necessary and sufficient condition for the system (4.1) to be controllable is that the rank of the matrix*

$$\mathcal{C} = \left(B \vdots AB \vdots \cdots \vdots A^{n-1}B \right) \tag{4.3}$$

is equal to n.

the matrix $\mathcal{C}$ is called *Kalman's controllability matrix*. Its size is $n \times nm$.

Proof. We can, without loss of generality, consider that $x_0 = 0$. It indeed suffices to change x_T in $y_T = x_T - e^{AT}x_0$ since then one looks for $\bar{u}$ such that

$$\int_0^T e^{A(T-t)}B\bar{u}(t)dt = y_T = x_T - e^{AT}x_0$$

which is precisely (4.2).

Consider the matrix $C(t) = e^{A(T-t)}B$, where A' and B' are the transposed matrices of A and B, and assume for a moment that the matrix $\mathcal{G} = \int_0^T C(t)C'(t)dt$, of size $n \times n$, is invertible. It suffices then to set

$$\bar{u}(t) = B'e^{A'(T-t)}\mathcal{G}^{-1}y_T. \tag{4.4}$$

Indeed, one easily verifies that $\bar{u}$ is continuous and that

$$\int_0^T e^{A(T-t)}B\bar{u}(t)dt = \int_0^T e^{A(T-t)}BB'e^{A'(T-t)}\mathcal{G}^{-1}y_T dt = \mathcal{G}\mathcal{G}^{-1}y_T = y_T$$

and thus that $\bar{u}$ generates the required trajectory.

It remains thus to prove that the invertibility of $\mathcal{G}$ is equivalent to $\text{rank}\,(\mathcal{C}) = n$. The proof is by contradiction, assuming that $\mathcal{G}$ is not invertible.

In this case, there exists a vector $v \in \mathbb{R}^n$, $v \neq 0$, such that $v'\mathcal{G} = 0$. Right multiplying by v, we get

$$v'\mathcal{G}v = \int_0^T v'C(t)C'(t)vdt = 0.$$

Since $v'C(t)C'(t)v$ is a non negative quadratic form, the integral cannot vanish unless $v'C(t)C'(t)v = 0$ for all $t \in [0,T]$, and thus unless $v'C(t) = 0$ for all $t \in [0,T]$. But this last relation reads $v'e^{A(T-t)}B = 0$ and, by definition of the exponential of a matrix,

$$v'\left(I + \sum_{i\geq 1} A^i \frac{(T-t)^i}{i!}\right) B = 0.$$

This polynomial of $T-t$ being identically 0, each monomial is 0 too, so that $v'B = 0$ and $v'A^iB = 0$ for all $i \geq 1$. This implies that $v'\mathcal{C} = 0$ with $v \neq 0$ and therefore $\mathcal{C}$ cannot be of rank n.

Conversely, if the rank of $\mathcal{C}$ is strictly smaller than n, there exists a non zero vector v such that $v'\mathcal{C} = 0$ and thus $v'A^iB = 0$ for all $i = 0, \ldots, n-1$. Using the Cayley-Hamilton Theorem (Gantmacher [1966]), all the successive powers A^k for $k \geq n$ are thus linear combinations of the A^k's for $0 \leq k \leq n-1$. Since $v'B = v'AB = \ldots = v'A^{n-1}B = 0$, it results that $v'A^iB = 0$ for all i and thus, making the previous calculations backwards, we get $v'\mathcal{G} = 0$, which implies that $\mathcal{G}$ is not invertible, which completes the proof.

Example 4.1. The single input, 2-dimensional system

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= u \end{aligned}$$

is controllable: $B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\mathcal{C} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\text{rank}\,(\mathcal{C}) = 2$.

Example 4.2. The single input, 2-dimensional system

$$\begin{aligned} \dot{x}_1 &= u \\ \dot{x}_2 &= u \end{aligned}$$

is not controllable: $B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\mathcal{C} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ and $\text{rank}\,(\mathcal{C}) = 1$. Note that the vector v of the proof of Theorem 4.1 can be chosen as $v' = (1, -1)$ and gives a combination of states which is independent of the input: $v'x = x_1 - x_2$ is such that $\frac{d}{dt}v'x = \dot{x}_1 - \dot{x}_2 = 0$, *i.e.* the expression $x_1 - x_2$ only depends on the initial conditions and cannot be modified by the input. This is precisely what characterizes a non controllable system.

4.1.2 Controllability Canonical Form

The notion of *controllability canonical form*, or *Brunovský canonical form*, is defined in reference to the SLF-equivalence defined below: a partition of the considered set of systems is induced by the equivalence classes, and each equivalence class is represented by a canonical form.

Definition 4.2. We say that 2 systems $\dot{x} = Ax + Bu$ and $\dot{z} = Fz + Gv$ are equivalent by linear change of coordinates and (non degenerate) feedback, or shortly *static linear feedback equivalence* or *SLF-equivalence* if there exist two invertible matrices M and L and a matrix K such that if x and u satisfy $\dot{x} = Ax + Bu$ and if $z = Mx$ and $v = Kx + Lu$, then z and v satisfy $\dot{z} = Fz + Gv$, and conversely.

M is the invertible matrix of change of coordinates of size $n \times n$, and K and L are the feedback matrices, with L of size $m \times m$, invertible, and K of size $m \times n$.

From the invertibility of M and L, one immediately deduces that 2 SLF-equivalent systems have the same state and input dimensions. To express that this equivalence depends only on the matrices defining the 2 systems, we also say that the pairs (A, B) and (F, G) are SLF-equivalent.

The definition 4.2 indeed corresponds to an equivalence relation: it is trivially reflexive and transitive. Let us prove the symmetry: if there exist M, K and L such that $\dot{x} = Ax+Bu$ is SLF-equivalent to $\dot{z} = Fz+Gv$, with $z = Mx$ and $v = Kx + Lu$, the symmetric transformation is given by $x = M^{-1}z$ and $u = -L^{-1}KM^{-1}z + L^{-1}v$.

Note that M, K and L are characterized by the formulas

$$F = M(A - BL^{-1}K)M^{-1} \ , \quad G = MBL^{-1}.$$

Indeed, $\dot{z} = M\dot{x} = M(Ax + Bu) = M(Ax + BL^{-1}(v - Kx)) = M(A - BL^{-1}K)M^{-1}z+MBL^{-1}v$. The inverse formulas give: $A = M^{-1}(FM+GK)$, $B = M^{-1}GL$.

To simplify, in a first step, we study the single input case ($m = 1$). We will show that every single input controllable system is SLF-equivalent to a particularly simple system, called *canonical*. In this case, B is a vector noted b to avoid any confusion.

The system thus reads $\dot{x} = Ax + bu$.

We know that it is controllable if and only if the rank of $\mathcal{C} = \left(b \vdots Ab \vdots \ldots \vdots A^{n-1}b\right)$ is equal to n. Since this matrix has exactly n columns, they are necessarily independent, which means that b is a *cyclic vector of* A (see Gantmacher [1966]), *i.e.*, the vectors $\{b, Ab, \ldots, A^{n-1}b\}$ form a base of $\mathbb{R}^n$, isomorphic to the canonical orthonormal base of $\mathbb{R}^n$ whose ith vector is made of $n-1$ zeroes and of a single 1 in the ith position. This canonical basis is itself generated by the cyclic vector g of the matrix F given by

$$g = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{4.5}$$

Let us set $\dot{z} = Fz + gv$ and let us show that this latter system is SLF-equivalent to the original system, in other words, that there exist matrices M and K and a scalar L such that if $z = Mx$ and $v = Kx + Lu$, then z satisfies $\dot{z} = Fz + gv$, or

$$\begin{aligned} \dot{z}_1 &= z_2 \\ &\vdots \\ \dot{z}_{n-1} &= z_n \\ \dot{z}_n &= v. \end{aligned} \tag{4.6}$$

We pose $M_1 = \begin{pmatrix} 1 \ 0 \dots 0 \end{pmatrix} M$, the first line of the required matrix M, and choose it such that

$$M_1 b = 0, \ M_1 A b = 0, \dots, \ M_1 A^{n-2} b = 0, \ M_1 A^{n-1} b = 1 \tag{4.7}$$

i.e. $M_1 \begin{pmatrix} b \ Ab \dots A^{n-1} b \end{pmatrix} = M_1 \mathcal{C} = \begin{pmatrix} 0 \ 0 \dots 1 \end{pmatrix}$. But, since $\mathcal{C}$ is invertible, M_1 exists and is given by

$$M_1 = \begin{pmatrix} 0 \ 0 \dots 1 \end{pmatrix} \mathcal{C}^{-1}. \tag{4.8}$$

Thus, (4.6) reads $z_1 = M_1 x$, $\dot{z}_1 = M_1(Ax + bu) = M_1 Ax = z_2, \dots, \dot{z}_{n-1} = z_1^{(n-1)} = M_1 A^{n-2}(Ax+bu) = M_1 A^{n-1} x = z_n$ and $z_1^{(n)} = \dot{z}_n = M_1 A^{n-1}(Ax+bu) = M_1 A^n x + M_1 A^{n-1} bu = M_1 A^n x + u$, which proves that

$$M = \begin{pmatrix} M_1 \\ M_1 A \\ \vdots \\ M_1 A^{n-1} \end{pmatrix} \tag{4.9}$$

and $\dot{z}_i = z_1^{(i)} = z_{i+1}$ for $i = 1, \dots, n-1$, and $\dot{z}_n = M_1 A^n x + u$. Let us denote by

$$v = M_1 A^n x + u$$

and $K = M_1 A^n$, $L = 1$. We thus have constructed the desired matrices M, K and L that transform the system $\dot{x} = Ax + bu$ into the so-called *canonical* system, $\dot{z} = Fz + gv$, with F and g given by (4.5).

Remark that this construction is effective: we first compute the matrix $\mathcal{C}$*, then* $M_1 = \begin{pmatrix} 0 \ 0 \dots 1 \end{pmatrix} \mathcal{C}^{-1}$*, and then each row of* M*, namely* $M_1 A^i$*, obtained by right-multiplying the previous one by* A*, to finally get* $K = M_1 A^n$ *and* $L = 1$*.*

In the multi-input case ($m > 1$), the canonical form is constructed by blocks, each one having the same structure as F and g and the same dimension as the cyclic subspace generated by each column of B.

Let us precise, before stating the next Theorem, how the dimension of each block is obtained.

Let us denote by b_i the ith column of B and let us rewrite the controllability matrix as follows:

$$\mathcal{C} = \left(b_1 \dots b_m \; Ab_1 \dots Ab_m \dots A^{n-1}b_1 \dots A^{n-1}b_m \right).$$

We establish the list of all the independent columns of $\mathcal{C}$ starting from b_1 and going forward to the right, by elimination of the columns that are linearly dependent of the previously selected ones on the left. Since $\operatorname{rank}(B) = m$, the m first columns of $\mathcal{C}$ necessarily belong to this list. Once complete, the list gives rise to the matrix noted $\overline{\mathcal{C}}$,

$$\overline{\mathcal{C}} = \left(b_1 \; Ab_1 \dots A^{n_1-1}b_1 \dots b_m \; Ab_m \dots A^{n_m-1}b_m \right)$$

having exactly n independent columns if the pair (A, B) is controllable. The integers $n_1, \dots, n_m$ so defined are thus such that $1 \leq n_i \leq n$ for all $i = 1, \dots, m$ and $n_1 + \dots + n_m = n$.

They can be proven invariant by linear change of coordinates and (non degenerate) feedback. In other words, two SLF-equivalent systems have the same sequence of integers $\{n_1, \dots, n_m\}$, up to permutation.

These remarkable integers n_i are called the *controllability indices*, or *Brunovský indices.*

Theorem 4.2. (Brunovský) *Every linear controllable system with n states and m inputs, given by a pair (A, B), is SLF-equivalent to its canonical form $F = \operatorname{diag}\{F_1, \dots, F_m\}$, $G = \operatorname{diag}\{g_1, \dots, g_m\}$ where each pair F_i, g_i is of the form (4.5), $i = 1, \dots, m$, with F_i of size $n_i \times n_i$ and g_i of size $n_i \times 1$, the integers $n_1, \dots, n_m$ being the* controllability indices *of (A, B), and satisfying $1 \leq n_i \leq n$ and $\sum_{i=1}^{m} n_i = n$.*

The proof of this Theorem with the construction of the corresponding matrices M, L, K, comparable to the one of the single-input case ($m = 1$), may be found in Brunovský [1970], Kailath [1980], Sontag [1998], Wolovich [1974], Wonham [1974].

This result has very important consequences since, via canonical forms, it is particularly easy to design reference trajectories and feedbacks. We now sketch these design approaches in the single-input case for the sake of simplicity.

4.1.3 Motion Planning

Let us go back to the single input, n states, controllable system $\dot{x} = Ax + bu$. According to its SLF-equivalence to the system in canonical form $\dot{z} = Fz + gv$ with $z = Mx$, $v = Kx + Lu$, if we want to find a trajectory starting from $x(0) = x_0$ at time 0 with the initial control $u(0) = u_0$, and arriving at $x(T) = x_T$ at time T with final control $u(T) = u_T$, we may express these conditions on z and v:

$$z(0) = Mx_0, \quad v(0) = Kx_0 + Lu_0, \quad z(T) = Mx_T, \quad v(T) = Kx_T + Lu_T, \tag{4.10}$$

and remark, from (4.5), that the first component of z, that we rename y for clarity's sake, satisfies

$$y^{(i)} = z_{i+1}, \quad i = 0, \ldots, n-1, \quad y^{(n)} = v$$

with the notation $y^{(i)} = \frac{d^i y}{dt^i}$. Thus, the conditions (4.10) may be interpreted as conditions on the successive derivatives of y, up to order n, at times 0 and T. Consequently, if we are given an n times differentiable *scalar* curve $t \in [0, T] \mapsto y_{ref}(t) \in \mathbb{R}$, satisfying the initial and final conditions (4.10), all the other system variables may be deduced by differentiation, *without integration of the system differential equations.* In particular, the input v may be obtained as the nth order derivative of y_{ref} with respect to time and the original control u_{ref} is deduced by $u_{ref} = -L^{-1}KM^{-1}z_{ref} + L^{-1}v_{ref}$, with $z_{ref} = (y_{ref}, \dot{y}_{ref}, \ldots, y_{ref}^{(n-1)})$. Accordingly, the trajectory x_{ref} is obtained by $x_{ref} = M^{-1}z_{ref}$, the control u_{ref} so obtained, exactly satisfying $\dot{x}_{ref} = Ax_{ref} + bu_{ref}$.

It remains to find such a curve y_{ref}. Using *interpolation theory*, one can find a polynomial with respect to time, of degree at least equal to $2n+1$, at least n times differentiable by construction, such that the $n+1$ initial conditions and the $n+1$ final ones are satisfied:

$$y_{ref}(t) = \sum_{i=0}^{2n+1} a_i \left(\frac{t}{T}\right)^i .$$

The coefficients $a_0, \ldots, a_{2n+1}$ are computed by equating the successive derivatives of y_{ref} at times 0 and T to their corresponding initial and final values, obtained from the initial and final conditions (4.10):

$$y_{ref}^{(k)}(t) = \frac{1}{T^k} \sum_{i=k}^{2n+1} i(i-1)\cdots(i-k+1)a_i \left(\frac{t}{T}\right)^{i-k}$$

thus, at $t = 0$:

$$
\begin{gathered}
M_1 x_0 = y_{ref}(0) = a_0, \quad M_{k+1} x_0 = y_{ref}^{(k)}(0) = \frac{k!}{T^k} a_k, \; k = 1, \ldots, n-1, \\
K x_0 + L u_0 = v_{ref}(0) = \frac{n!}{T^n} a_n,
\end{gathered} \tag{4.11}
$$

and at $t = T$:

$$
\begin{aligned}
&M_1 x_T = y_{ref}(T) = \sum_{i=0}^{2n+1} a_i, \\
&M_{k+1} x_T = y_{ref}^{(k)}(T) = \frac{1}{T^k} \sum_{i=k}^{2n+1} \frac{i!}{(i-k)!} a_i, \quad k = 1, \ldots, n-1, \\
&K x_T + L u_T = v_{ref}(T) = \frac{1}{T^n} \sum_{i=n}^{2n+1} \frac{i!}{(i-n)!} a_i
\end{aligned}
$$

which finally gives $2n+2$ linear equations in the $2n+2$ coefficients $a_0, \ldots, a_{2n+1}$. In fact, one can reduce this system to $n+1$ linear equations in the $n+1$ unknown coefficients $a_{n+1}, \ldots, a_{2n+1}$ only, since the $n+1$ first equations (4.11) are solved in $a_0, \ldots, a_n$:

$$
a_0 = y_{ref}(0) \;, \quad a_k = \frac{T^k}{k!} y_{ref}^{(k)}(0) \;, k = 1, \ldots, n-1, \quad a_n = \frac{T^n}{n!} v_{ref}(0).
$$

Note that the linear system (**??**) always has a unique solution:

$$
\begin{pmatrix}
1 & 1 & \cdots & 1 \\
n+1 & n+2 & & 2n+1 \\
(n+1)n & (n+2)(n+1) & & (2n+1)2n \\
\vdots & & & \vdots \\
(n+1)! & \frac{(n+2)!}{2} & \cdots & \frac{(2n+1)!}{(n+1)!}
\end{pmatrix}
\begin{pmatrix}
a_{n+1} \\ \vdots \\ a_{2n+1}
\end{pmatrix}
=
\begin{pmatrix}
y_{ref}(T) - \sum_{i=0}^{n} a_i \\
\vdots \\
T^k y_{ref}^{(k)}(T) - \sum_{i=k}^{n} \frac{i!}{(i-k)!} a_i \\
\vdots \\
T^n v_{ref}(T) - n! a_n
\end{pmatrix}
$$

the matrix of the left-hand side having all its columns independent, which achieves the construction of the reference trajectory.

4.1.4 Trajectory Tracking, Pole Placement

We now show how the canonical form may be used for feedback design: we go back to the system in canonical form

$$y^{(n)} = v.$$

We assume that the whole state x is measured at every time. If we want to follow the reference trajectory y_{ref}, with $y_{ref}^{(n)} = v_{ref}$, that we just designed in the previous section, and if the system is submitted to unmodelled disturbances, the deviation between the measured trajectory and its reference is given by $e = y - y_{ref}$ and satisfies the nth-order differential equation $e^{(n)} = v - v_{ref}$. Note that this deviation, together with its $n-1$ first derivatives, can be computed at every time from the measured state x. Our aim is to guarantee the convergence to 0 of e and its $n-1$ successive derivatives.

Let us set $v - v_{ref} = -\sum_{i=0}^{n-1} K_i e^{(i)}$, the gains K_i being arbitrarily chosen. The deviation dynamics becomes

$$e^{(n)} = -\sum_{i=0}^{n-1} K_i e^{(i)}$$

or, in matrix form,

$$\begin{pmatrix} \dot{e} \\ \vdots \\ e^{(n)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & 0 & 0 & & 1 \\ -K_0 & -K_1 & -K_2 & \dots & -K_{n-1} \end{pmatrix} \begin{pmatrix} e \\ \dot{e} \\ \vdots \\ e^{(n-1)} \end{pmatrix}.$$

We easily check that the gains K_i are the coefficients of the characteristic polynomial of the above matrix, so that its eigenvalues[1] may be placed arbitrarily in the complex plane, by a suitable choice of the gains. If the gains are chosen so that all the roots of the characteristic polynomial have negative real part, applying the results of the previous Chapter, the deviation dynamics is exponentially stable.

We thus have proved, in the case $m = 1$, the following important result:

Theorem 4.3. ***(pole placement))*** *If the system $\dot{x} = Ax + Bu$ is controllable, the eigenvalues of the closed loop matrix $A + BK$ may be arbitrarily placed in the complex plane by a suitable state feedback $u = Kx$.*

Proof. The generalization of the above construction to the multi-input case (m arbitrary) follows the same lines. Its proof is left to the reader. We now deduce the existence of a gain matrix $\overline{K}$ such that if $v = \overline{K}z$, the closed loop

[1] often called the system poles in reference to the transfer function representation

matrix $F+G\overline{K}$ has all its eigenvalues arbitrarily placed in the complex plane. Since $v = \overline{K}z = Kx + Lu$, we get, with $z = Mx$, that $u = L^{-1}(\overline{K}M - K)x$ and thus that the closed loop system in the original coordinates is given by $\dot{x} = (A + BL^{-1}(\overline{K}M - K))x$. Using again the fact that $z = Mx$, we have $M(A + BL^{-1}(\overline{K}M - K))M^{-1} = F + G\overline{K}$ and thus $A + BL^{-1}(\overline{K}M - K)$ and $F + G\overline{K}$ are similar. They thus have the same eigenvalues, which proves that given an arbitrary sequence of n eigenvalues, the gain $L^{-1}(\overline{K}M - K)$ is such that the closed loop matrix $A + BL^{-1}(\overline{K}M - K)$ admits exactly this sequence as eigenvalues.

Corollary 4.1. *A controllable linear system is stabilizable and, by a suitable choice of state feedback, its characteristic exponents may be arbitrarily placed.*

4.2 Nonlinear System Controllability

We now consider the nonlinear system

$$\dot{x} = f(x, u) \tag{4.12}$$

where the state x lies in an open subset of $\mathbb{R}^n$ and the input u is m-dimensional.

Several notions of controllability may be defined in this context. The notions of local controllability around an equilibrium point $(\bar{x}, \bar{u})$, *i.e.* such that $f(\bar{x}, \bar{u}) = 0$, and of first-order controllability, though relatively restrictive, are directly inspired by the previous linear analysis.

4.2.1 First Order and Local Controllability

The tangent linear system around the equilibrium point $(\bar{x}, \bar{u})$ is given by

$$\dot{x} = Ax + Bu\ , \quad A = \frac{\partial f}{\partial x}(\bar{x}, \bar{u})\ , \quad B = \frac{\partial f}{\partial u}(\bar{x}, \bar{u}). \tag{4.13}$$

Definition 4.3. We say that the system (4.12) is *first-order controllable* at the equilibrium point $(\bar{x}, \bar{u})$ if the rank of $\mathcal{C}$, defined by (4.3) for the tangent linear system (4.13), is equal to n.

The local controllability definition is the following:

Definition 4.4. The system (4.12) is *locally controllable* at the equilibrium point $(\bar{x}, \bar{u})$ if for all real $\varepsilon > 0$ there exists a real $\eta > 0$ such that for every pair of points $(x_0, x_1) \in \mathbb{R}^n \times \mathbb{R}^n$ close enough to the equilibrium point, namely satisfying $\|x_0 - \bar{x}\| < \eta$ and $\|x_1 - \bar{x}\| < \eta$, there exists a piecewise continuous control $\tilde{u}$ on $[0,\varepsilon]$ such that $\|\tilde{u}(t)\| < \varepsilon \ \forall t \in [0,\varepsilon]$ and $X_\varepsilon(x_0,\tilde{u}) = x_1$, where we have denoted by $X_\varepsilon(x_0,\tilde{u})$ the integral curve of (4.12) at time ε, generated by $\tilde{u}$ from x_0 at time 0.

Otherwise stated, the system is locally controllable at the equilibrium point $(\bar{x},\bar{u})$ if we can join two arbitrary points in a small neighborhood of the equilibrium, in sufficiently small duration and with a sufficiently small control.

One can prove the following:

Theorem 4.4. *If system (4.12) is first-order controllable at the equilibrium point $(\bar{x},\bar{u})$, it is locally controllable at $(\bar{x},\bar{u})$.*

Remark 4.1. A nonlinear system (4.12) can be locally controllable without being first-order controllable. The scalar system: $\dot{x} = u^3$ is locally controllable. To join x_0 and x_1 in duration $T = \varepsilon$, x_0 and x_1 arbitrarily chosen close to 0, one can, using the motion planning approach of section 4.1.2, construct a differentiable enough function of time $\overline{x}(t)$, joining the two points at the respective times 0 and ε and arriving at zero speed: $\overline{x}(t) = x_0 + (x_1 - x_0)\left(\frac{t}{\varepsilon}\right)^2\left(3 - 2\frac{t}{\varepsilon}\right)$, and deduce $\overline{u}(t) = \left(\dot{\overline{x}}(t)\right)^{\frac{1}{3}} = \left(6\left(\frac{x_1-x_0}{\varepsilon}\right)\left(\frac{t}{\varepsilon}\right)\left(1-\frac{t}{\varepsilon}\right)\right)^{\frac{1}{3}}$. One easily checks that if $|x_0| < \eta$ and $|x_1| < \eta$ with $\eta < \frac{\varepsilon^4}{3}$ then $|\overline{u}(t)| < \varepsilon$, which proves the local controllability. On the contrary, at the equilibrium point $(0,0)$, the tangent linear system is $\dot{x} = 0$, and is indeed not first-order controllable.

4.2.2 Local Controllability and Lie Brackets

We now assume, for simplicity's sake, that the vector field f of (4.12) is affine in the control, *i.e.*

$$\dot{x} = f_0(x) + \sum_{i=1}^{m} u_i f_i(x) \tag{4.14}$$

with $f_0(0) = 0$, so that $(\bar{x},\bar{u}) = (0,0)$ is an equilibrium point.

From the vector fields $f_0,\ldots,f_m$, we construct the following sequence of distributions:

$$\mathcal{D}_0 = \operatorname{span}\{f_1,\ldots,f_m\} \ ,\mathcal{D}_{i+1} = [f_0,\overline{\mathcal{D}}_i] + \overline{\mathcal{D}}_i \ , \quad i \geq 1 \tag{4.15}$$

where $\overline{\mathcal{D}}_i$ is the involutive closure of the distribution $\mathcal{D}_i$.

In the sequel, we assume that there exists an open subset of $X = \mathbb{R}^n$ where all the considered distributions have constant rank.

Proposition 4.1. *The sequence of distributions $\overline{\mathcal{D}}_i$ is non decreasing, i.e. $\overline{\mathcal{D}}_i \subset \overline{\mathcal{D}}_{i+1}$ for all i, and there exists an integer k^* and an involutive distribution $\mathcal{D}^*$ such that $\overline{\mathcal{D}}_{k^*} = \overline{\mathcal{D}}_{k^*+r} = \mathcal{D}^*$ for all $r \geq 0$. Moreover, $\mathcal{D}^*$ enjoys the two following properties:*

(i) $\text{span}\{f_1, \ldots, f_m\} \subset \mathcal{D}^*$
(ii) $[f_0, \mathcal{D}^*] \subset \mathcal{D}^*$.

Proof. The non decrease of the sequence $\{\overline{\mathcal{D}}_i\}$ is obvious. Since moreover, $\overline{\mathcal{D}}_i \subset \mathrm{T}X$ for all i, this sequence admits a largest element $\mathcal{D}^*$, which is involutive by construction. In addition, $\dim \mathrm{T}_x X = n$ and, in a suitably chosen open U where all the $\overline{\mathcal{D}}_i$'s have constant rank, if the vector space $\overline{\mathcal{D}}_i(x)$ is not equal to $\overline{\mathcal{D}}_{i+1}(x)$, we have $\dim \overline{\mathcal{D}}_{i+1}(x) \geq \dim \overline{\mathcal{D}}_i(x) + 1$ in U. We immediately deduce that there exists an integer $k^* \leq n$ such that $\overline{\mathcal{D}}_{k^*+1} = \overline{\mathcal{D}}_{k^*} = \mathcal{D}^*$. The property (i) follows immediately from $\mathcal{D}_0 \subset \mathcal{D}^*$. Concerning (ii), we have $\mathcal{D}_{k^*+1} = [f_0, \overline{\mathcal{D}}_{k^*}] + \overline{\mathcal{D}}_{k^*}$, thus $\mathcal{D}^* = [f_0, \mathcal{D}^*] + \mathcal{D}^*$, which proves (ii) and the Proposition is proven.

Property (ii) is called the *invariance of* $\mathcal{D}^*$ *by* f_0 and $\mathcal{D}^*$ is called the *strong accessibility distribution*.

We thus have the following characterization of $\mathcal{D}^*$:

Proposition 4.2. *Let $\mathcal{D}$ be an involutive distribution with constant rank equal to k in an open U satisfying (i) and (ii). Then, there exists a diffeomorphism φ such that, if we note $\xi_i = \varphi_i(x)$ for $i = 1, \ldots, k$ and $\zeta_j = \varphi_{k+j}(x)$ for $j = 1, \ldots, n-k$, the vector field $\varphi_* f_0$ admits the following decomposition:*

$$\varphi_* f_0(\xi) = \sum_{i=1}^{k} \gamma_i(\xi, \zeta) \frac{\partial}{\partial \xi_i} + \sum_{i=1}^{n-k} \gamma_{k+i}(\zeta) \frac{\partial}{\partial \zeta_i} \tag{4.16}$$

where the γ_i's are C^∞ functions in all their variables.

Proof. SInce $\mathcal{D}$ is involutive with constant rank equal to k in U, By Frobenius' Theorem, there exists a diffeomorphism φ such that $\mathcal{D}$ is transformed into

$$\varphi_* \mathcal{D} = \text{span}\left\{\frac{\partial}{\partial \xi_1}, \ldots, \frac{\partial}{\partial \xi_k}\right\}.$$

In these coordinates, $\varphi_* f_0$ reads

$$\varphi_* f_0(\xi, \zeta) = \sum_{i=1}^{k} \gamma_i(\xi, \zeta) \frac{\partial}{\partial \xi_i} + \sum_{i=1}^{n-k} \gamma_{k+i}(\xi, \zeta) \frac{\partial}{\partial \zeta_i} \tag{4.17}$$

where the functions γ_i, $i = 1, \ldots, n$ are C^∞ with respect to (ξ, ζ). Since $[f_0, \mathcal{D}] \subset \mathcal{D}$, the bracket $[\varphi_* f_0, \frac{\partial}{\partial \xi_i}]$ must be a linear combination of the $\frac{\partial}{\partial \xi_j}$ for $j = 1, \ldots, k$, thus

$$\left[\varphi_* f_0, \frac{\partial}{\partial \xi_i}\right] = \sum_{j=1}^{k} \alpha_{i,j}(\xi, \zeta) \frac{\partial}{\partial \xi_j} \tag{4.18}$$

using (4.17) with, as a consequence of the straightening out, $[\frac{\partial}{\partial \xi_i}, \frac{\partial}{\partial \xi_j}] = 0$ for all i, j and $[\frac{\partial}{\partial \xi_i}, \frac{\partial}{\partial \zeta_j}] = 0$ for all i, j, we have

$$\left[\varphi_* f_0, \frac{\partial}{\partial \xi_i}\right] =$$

$$\sum_{j=1}^{k} \gamma_j \left[\frac{\partial}{\partial \xi_j}, \frac{\partial}{\partial \xi_i}\right] + \sum_{j=1}^{n-k} \gamma_{k+j} \left[\frac{\partial}{\partial \zeta_j}, \frac{\partial}{\partial \xi_i}\right] - \sum_{j=1}^{k} \frac{\partial \gamma_j}{\partial \xi_i} \frac{\partial}{\partial \xi_j} - \sum_{j=1}^{n-k} \frac{\partial \gamma_{k+j}}{\partial \xi_i} \frac{\partial}{\partial \zeta_j}$$

$$= - \sum_{j=1}^{k} \frac{\partial \gamma_j}{\partial \xi_i} \frac{\partial}{\partial \xi_j} - \sum_{j=1}^{n-k} \frac{\partial \gamma_{k+j}}{\partial \xi_i} \frac{\partial}{\partial \zeta_j}.$$

Comparing the latter formula to (4.18), the part of the vector field made of the linear combination of the $\frac{\partial}{\partial \zeta_j}$ must vanish:

$$\frac{\partial \gamma_{k+j}}{\partial \xi_i} = 0 \ , \forall j = 1, \ldots, n-k \ , \forall i = 1, \ldots, k$$

which proves that γ_{k+j} doesn't depend on $\xi_1, \ldots, \xi_k$, which proves (4.16), for $j = 1, \ldots, n-k$.

We immediately deduce the following necessary condition for local controllability:

Theorem 4.5. *A necessary condition for the system (4.14) with $f_0(0) = 0$, to be locally controllable around the origin is that the strong accessibility distribution $\mathcal{D}^*$ constructed by the induction (4.15), satisfies*

$$\operatorname{rank}(\mathcal{D}^*(x)) = n \ , \quad \forall x \in U$$

where U is a neighborhood of the origin.

Proof. Assume that $\operatorname{rank}(\mathcal{D}^*(x)) = k < n$, for all $x \in U$. According to Proposition 4.2, there exists a local diffeomorphism φ such that, if we note $(\xi, \zeta) = \varphi(x)$, as in Proposition 4.2, with $\dim \xi = k$ and $\dim \zeta = n-k$, the vector field f_0 reads as in (4.16), and, since $f_i \in \mathcal{D}^*$, $\varphi_* f_i = \sum_{j=1}^{k} \eta_{i,j}(\xi, \zeta) \frac{\partial}{\partial \xi_j}$, $i = 1, \ldots, m$. Thus, the vector field $\varphi_* \left(f_0 + \sum_{i=1}^{m} u_i f_i\right)$ reads

$$\varphi_* \left(f_0 + \sum_{i=1}^{m} u_i f_i \right) = \sum_{j=1}^{k} \left(\gamma_j + \sum_{i=1}^{m} u_i \eta_{i,j} \right) \frac{\partial}{\partial \xi_j} + \sum_{j=1}^{n-k} \gamma_{k+j} \frac{\partial}{\partial \zeta_j}$$

and consequently, the system (4.14), in the local coordinates (ξ, ζ) becomes:

$$\begin{aligned} \dot{\xi}_i &= \gamma_i(\xi, \zeta) + \sum_{j=1}^{m} u_j \eta_{i,j}(\xi, \zeta), \quad i = 1, \ldots k, \\ \dot{\zeta}_i &= \gamma_{k+i}(\zeta), \quad i = 1, \ldots n-k. \end{aligned}$$

It is thus clear that ζ's dynamics, independent of ξ and u, cannot be modified by u, so that from any initial point (ξ_0, ζ_0), every reachable final point is necessarily of the form $\zeta_T = Z_T(\zeta_0)$, where $Z_T(\zeta_0)$ is the point of the integral curve of $\dot{\zeta} = \hat{\gamma}(\zeta)$ at $t = T$, and with $\hat{\gamma}$ the vector whose components are $\gamma_{k+1}, \ldots, \gamma_n$. Thus, any final point outside the closed subset $\zeta_T = Z_T(\zeta_0)$ is therefore not reachable, which contradicts the local controllability.

We denote by $\mathrm{Lie}\{f_0, \ldots, f_m\}$ the Lie algebra generated by the linear combinations of $f_0, \ldots, f_m$ and all their Lie brackets, often called the *weak accessibility distribution* and $\mathrm{Lie}\{f_0, \ldots, f_m\}(x)$ the vector space generated by the vectors of $\mathrm{Lie}\{f_0, \ldots, f_m\}$ at the point x.

By construction,

$$\mathcal{D}^* \subset \mathrm{Lie}\{f_0, \ldots, f_m\}$$

but the equality doesn't hold true in general. In the analytic case, we have the following variant of Theorem 4.5.

Theorem 4.6. *Assume that the $m+1$ vector fields $f_0, \ldots, f_m$ are analytic. If the system (4.14) is locally controllable at $x = 0$, $u = 0$, then*

$$\mathrm{Lie}\{f_0, \ldots, f_m\}(x) = \mathrm{T}_x\mathbb{R}^n, \quad \forall x \in X.$$

Proof. Since (4.14) is locally controllable at $x = 0$, $u = 0$, we have $\mathrm{rank}(\mathcal{D}^*(0)) = n$ and, according to the analyticity of the vector fields $f_0, \ldots, f_m$ this rank remains constant on X which is diffeomorphic to an open subset of $\mathbb{R}^n$. Thus $\mathcal{D}^*(x)$ may be identified with $\mathrm{T}_x\mathbb{R}^n$ and, necessarily, $f_0(x) \in \mathcal{D}^*(x)$ for all x. We conclude that $\mathrm{Lie}\{f_0, \ldots, f_m\}(x) = \mathcal{D}^*(x)$ for all x.

Before studying the converse of this Theorem, which is more complicated, let us give the following simple example:

Example 4.3. The 2-dimensional , single input system given by

$$\dot{x}_1 = x_2^2, \quad \dot{x}_2 = u$$

admits every point of $\mathbb{R}^2$ such that $x_2 = 0$ as equilibrium point for $u = 0$. In a neighborhood of such a point $X_0 = (x_1^0, 0)$, the system is affine in the control with $f_0(x_1, x_2) = x_2^2 \frac{\partial}{\partial x_1}$ and $f_1(x_1, x_2) = \frac{\partial}{\partial x_2}$. Remark that, at X_0, we have $f_0 = 0$. Let us compute the algorithm (4.15). We have $\mathcal{D}_0 = \mathrm{span}\{f_1\} = \mathrm{span}\left\{\frac{\partial}{\partial x_2}\right\}$ and $\mathcal{D}_0 = \overline{\mathcal{D}}_0$; $\mathcal{D}_1 = \mathrm{span}\{f_1, [f_0, f_1]\} = \mathrm{span}\left\{\frac{\partial}{\partial x_2}, x_2\frac{\partial}{\partial x_1}\right\}$, which has rank 2, is involutive as far as $x_2 \neq 0$, but is singular at X_0: $\mathcal{D}_1(X_0) = \mathcal{D}_0$, of rank 1. On the contrary, $[f_1, [f_0, f_1]] = 2\frac{\partial}{\partial x_1} \in \overline{\mathcal{D}}_1(X_0) \neq \mathcal{D}_1(X_0)$, and $\mathcal{D}^* = \overline{\mathcal{D}}_1(X_0) = \mathrm{span}\left\{\frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_1}\right\}$ which is nothing else than the tangent space of $\mathbb{R}^2$. The rank condition is thus satisfied. However, since $\dot{x}_1 = x_2^2 \geq 0$, every trajectory is such that $x_1(t)$ is non decreasing, and, from X_0, it is impossible to join a point $X_1 = (\tilde{x}_1, x_2)$ such that $\tilde{x}_1 < x_1$ whatever the duration T is. The system is therefore not locally controllable at X_0. However, The set of points in an arbitrary neighborhood of X_0 that can be reached from X_0 by an integral curve of the system using any piecewise continuous input and an arbitrary duration T has non empty interior (*i.e.* contains an open of $\mathbb{R}^2$). The condition $\mathrm{rank}\,(\mathcal{D}^*)\,(X_0) = 2$ thus implies a weaker property than the local controllability. We call this latter property *local reachability.*

4.2.3 Reachability

Definition 4.5. The reachable set at time T, noted $R_T(x_0)$, is the set of points $x \in X$ (X is an n-dimensional manifold) such that $X_T(x_0, u) = x$ with u piecewise continuous on $[0, T]$, where $X_t(x_0, u)$ is the integral curve of (4.14) at time t generated from x_0 at time 0 with the input u. We say that (4.14) is *locally reachable* if for every neighborhood V of x_0 in X, $R_T(x_0) \cap V$ has non empty interior.

We admit the following:

Theorem 4.7. *If for every neighborhood V of $x_0 \in X$, $\mathrm{rank}\,(\mathcal{D}^*(x)) = n$ for all $x \in V$, then the system (4.14) is locally reachable. If furthermore $f_0 \equiv 0$, then (4.14) is locally controllable.*

There are more general results, that apply in particular when $f_0 \neq 0$, but whose complexity goes far beyond the scope of this course. The interested reader may refer to Sussmann [1987].

Let us also precise that the results of this section, where we have restricted to systems affine in the control, of the form (4.14), can easily be extended to the more general class of systems of the form (4.12) using the following extension trick:

Setting $\dot{u}_i = v_i$, $i = 1, \ldots, m$, the extended system reads

$$\begin{aligned} \dot{x} &= f(x, u) \\ \dot{u}_1 &= v_1 \\ &\vdots \\ \dot{u}_m &= v_m \end{aligned} \tag{4.19}$$

and has the required form $\dot{X} = F_0(X) + \sum_{i=1}^{m} v_i F_i(X)$ with $X = (x, u)$, $F_0(X) = f(x, u)\frac{\partial}{\partial x}$, $F_i(X) = \frac{\partial}{\partial u_i}$, $i = 1, \ldots, m$. Note that, since the manifold of definition is now $X \times U$ where U is an open set of $\mathbb{R}^m$, the dimension n of the rank condition must be replaced by $n + m$.

Proposition 4.3. *If the extended system (4.19) is locally reachable or locally controllable, the same holds true for the original system.*

Proof. If the reachable set of the extended system at time T has non empty interior in $X \times U$, its projection on X, which is precisely the reachable set of the original system at time T, has also non empty interior. Accordingly, if the extended system is locally controllable, there exists a trajectory joining two arbitrary points (x_1, u_1) and (x_2, u_2) of a suitably chosen neighborhood, with an input $\dot{u}$ piecewise continuous and with small enough norm. Integrating this trajectory, one obtains the input u, continuous, that generates a trajectory of the original system joining x_1 and x_2, which achieves the proof.

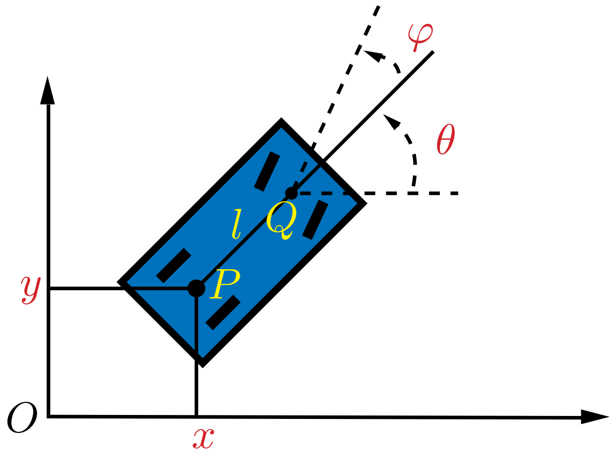

Fig. 4.1 Vehicle rolling without slipping on the plane.

Example 4.4. We consider a vehicle with 4 wheels rolling without slipping on the horizontal plane (O, X, Y). We denote by (x, y) the coordinates of the point P, middle of the rear axle, Q the middle point of the front axle, $\|\overrightarrow{PQ}\| = l$, θ the angle between the longitudinal axis of the vehicle and the Ox axis, and φ the angle of the front wheels (see Figure 4.1). The rolling without slipping condition, that geometrically corresponds to a specific kind of constraint called *non holonomic constraint* (see *e.g.* Neĭmark and Fufaev [1972]), which justifies the name of *non holonomic vehicle* given to this car idealization, reads: $\frac{d\overrightarrow{OP}}{dt}$ is parallel to $\overrightarrow{PQ}$ and $\frac{d\overrightarrow{OQ}}{dt}$ is parallel to the front wheels. Let us denote by $u = \left\|\frac{d\overrightarrow{OP}}{dt}\right\| \cdot \frac{\overrightarrow{PQ}}{\left\|\overrightarrow{PQ}\right\|}$ the modulus of the car's speed.

An elementary kinematic calculation yields:

$$\begin{aligned} \dot{x} &= u\cos\theta \\ \dot{y} &= u\sin\theta \\ \dot{\theta} &= \frac{u}{l}\tan\varphi. \end{aligned} \tag{4.20}$$

The two control variables are u and φ. However, since φ doesn't appear linearly in (4.20), and in order to apply the previous controllability results, we use the extension trick of (4.19) by setting $\dot{\varphi} = v$. The complete model now reads

$$\begin{aligned} \dot{x} &= u\cos\theta \\ \dot{y} &= u\sin\theta \\ \dot{\theta} &= \frac{u}{l}\tan\varphi \\ \dot{\varphi} &= v. \end{aligned} \tag{4.21}$$

Therefore, denoting by $X = (x, y, \theta, \varphi)^T$ the state of this model, (4.21) has the form $\dot{X} = f(X)u + g(X)v$ with $f(X) = \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y} + \frac{1}{l}\tan\varphi\frac{\partial}{\partial\theta}$ and $g(X) = \frac{\partial}{\partial\varphi}$.

Let us first compute the tangent linear approximation at an equilibrium point. Equilibrium points are characterized by $u\cos\theta = u\sin\theta = \frac{u}{l}\tan\varphi = v = 0$, or $u = v = 0$. Therefore any state $\overline{X} = (\overline{x}, \overline{y}, \overline{\theta}, \overline{\varphi})^T$ is an equilibrium point. The corresponding variational equation is (recall that $u = v = 0$), with the usual notation of $\delta X =: X - \overline{X}$

$$\begin{aligned} \dot{\delta x} &= \cos\theta\delta u \\ \dot{\delta y} &= \sin\theta\delta u \\ \dot{\delta\theta} &= \frac{1}{l}\tan\varphi\delta u \\ \dot{\delta\varphi} &= \delta v. \end{aligned} \tag{4.22}$$

or $\dot{\delta X} = 0\delta X + B\delta U$ with $B = \begin{pmatrix} \cos\theta & 0 \\ \sin\theta & 0 \\ \frac{1}{l}\tan\varphi & 0 \\ 0 & 1 \end{pmatrix}$ and $\delta U = (\delta u, \delta v)^T$.

It is readily seen that the Kalman controllability matrix is equal to $(B\ 0_{2\times 4})$ and has rank 2. Thus, the system isn't first-order controllable.

To check the local controllability, we compute the sequence of distributions $\mathcal{D}_i$, $i \geq 0$, defined by (4.15). Note that, since $f_0 = 0$, this sequence reduces to $\mathcal{D}_i = \mathcal{D}_0$ for all $i \geq 0$.

The distribution $\mathcal{D}_0 = \text{span}\{f, g\}$ is clearly non involutive since $[f, g] = -\frac{1}{l}(1 + \tan^2\varphi)\frac{\partial}{\partial\theta} \notin \mathcal{D}_0$, and again $[f, [f, g]] = \frac{1}{l}(1 + \tan^2\varphi)\left(-\sin\theta\frac{\partial}{\partial x} + \cos\theta\frac{\partial}{\partial y}\right) \notin \mathcal{D}_0$. The vector fields $f, g, [f, g]$ and $[f, [f, g]]$ are linearly independent at every point $\overline{X}$ and therefore generate the involutive closure of $\mathcal{D}_0$, which is thus of dimension 4 at every point $\overline{X}$. Finally, the local controllability of (4.20) results from Theorem 4.7 and Proposition 4.3.

The interpretation is clear: when the car is stopped, no matter how we turn the wheels, the car's position doesn't change. For arbitrary small speed and wheel turn angular speed, the first-order approximations of the car's motion remain in the 2-dimensional subspace spanned by B and the system is not first-order controllable. But if we consider higher-order approximations, the car can move in every direction.

4.2.4 Lie Brackets and Kalman's Criterion for Linear Systems

To conclude this section, we show that, for linear systems, the two notions of first order controllability and local controllability coincide and are boil down to Kalman's criterion.

We consider the linear system (4.1) that is rewritten in the form

$$\dot{x} = Ax + \sum_{i=1}^{m} u_i b_i$$

with the notations of (4.14).

Clearly, since the system is linear, it is equal to its tangent linear system and thus Kalman's criterion applied to the system itself or to its tangent linear one indeed give the same result.

We now evaluate the rank of $\mathcal{D}^*$. We set $Ax = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} A_{i,j}x_j\right)\frac{\partial}{\partial x_i}$ and $b_i = \sum_{j=1}^{n} b_{i,j}\frac{\partial}{\partial x_j}$, $i = 1,\ldots,m$. We have $f_0(x) = Ax$ and $f_i(x) = b_i$, $i = 1,\ldots,m$.

Let us compute the algorithm (4.15). We have $\mathcal{D}_0 = \text{span}\{b_1,\ldots,b_m\}$ and since the vector fields b_i are constant, $[b_i, b_j] = 0$ for all i, j, which proves that $\overline{\mathcal{D}}_0 = \mathcal{D}_0$.

Further, we have $[Ax, b_i] = -\sum_{j=1}^{n} b_{i,j}\frac{\partial}{\partial x_j}\left(\sum_{k=1}^{n}\sum_{l=1}^{n} A_{k,l}x_l\frac{\partial}{\partial x_k}\right) = -\sum_{j,k=1}^{n} b_{i,j}A_{k,j}\frac{\partial}{\partial x_k}$, or $[Ax, b_i] = -\sum_{j=1}^{n}(Ab_i)_j\frac{\partial}{\partial x_j} = Ab_i$. Thus

$$\mathcal{D}_1 = [Ax, \mathcal{D}_0] + \mathcal{D}_0 = \text{span}\{b_1,\ldots,b_m, Ab_1,\ldots,Ab_m\}.$$

As previously, $\mathcal{D}_1$ is involutive since it is only made of constant vector fields. The same iterated computation leads to

$$\mathcal{D}_k = \text{span}\left\{b_1,\ldots,b_m, Ab_1,\ldots,Ab_m,\ldots,A^k b_1,\ldots,A^k b_m\right\}$$

for all $k \geq 1$.

According to Proposition 4.1, there exists an integer $k^* < n$ such that

$$\mathcal{D}^{k^*} = \mathcal{D}^*$$

and the controllability implies that $\operatorname{rank}(\mathcal{D}^*(x)) = n$ for all $x \in U$, *i.e.*, again with matrix notations, $B = \left(b_1 \vdots \ldots \vdots b_m\right)$:

$$\operatorname{rank}\left(B, AB, \ldots, A^{k^*}B\right) = n$$

with $k^* < n$, which immediately implies Kalman's criterion.

Chapter 5
Jets of Infinite Order, Lie-Bäcklund's Equivalence

5.1 An Introductory Example of Crane

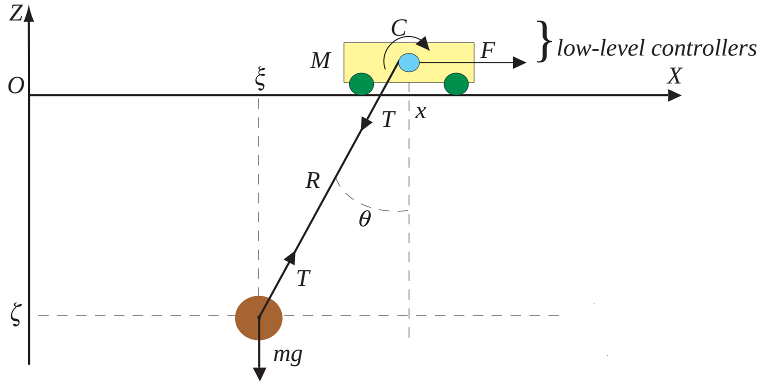

Fig. 5.1 A 2-dimensional overhead crane.

A cart of mass M rolls along the axis OX of the overhead crane. Its position is denoted by x. It is actuated by a motor that produces a horizontal force of intensity F. Moreover, the cart carries a winch of radius ρ around which is winding a cable hoisting the load attached at its end. The position of the load in the fixed frame XOZ is denoted by (ξ, ζ) and its mass is m. The torque exerted on the winch by a second motor is denoted by C. The cable length, its tension and the angle of the cable with respect to the vertical are denoted by R, T and θ respectively. The working space is limited to $R < R_0$ to avoid that the load touches the ground, and we assume that the tension T of the cable is always positive. We use the angular sign convention for θ : $\theta \leq 0$ if $\xi \leq x$ and $\theta > 0$ otherwise.

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_5, © Springer-Verlag Berlin Heidelberg 2009

We also assume that a viscous friction force, noted $\gamma_1(\dot{x})$, including the aerodynamic friction of the cable, is opposing to the cart's displacement and that another viscous friction torque, noted $\gamma_2(\dot{R})$, resists to the winch motion. The functions γ_1 and γ_2 are non negative and such that $\gamma_i(0) = 0$, $i = 1, 2$.

A second principle model may be easily obtained by considering the two bodies, cart and load, and assuming that the cable is rigid:

$$\begin{aligned} &m\ddot{\xi} = -T\sin\theta \\ &m\ddot{\zeta} = T\cos\theta - mg \\ &M\ddot{x} = -\gamma_1(\dot{x}) + F + T\sin\theta \\ &\frac{J}{\rho}\ddot{R} = -\gamma_2(\dot{R}) - C + T\rho \end{aligned} \tag{5.1}$$

the geometric constraints between the cart and load coordinates being given by:

$$\begin{aligned} \xi &= x + R\sin\theta \\ \zeta &= -R\cos\theta. \end{aligned} \tag{5.2}$$

Note that this model is not given in explicit form. It contains variables (T and θ) whose time derivatives don't explicitly appear, and both algebraic and differential equations relating ξ and ζ to x, R, T and θ. Some of the unknowns may thus be eliminated, though this operation is not needed in the sequel.

In fact, an explicit 6-dimensional representation of the system may be obtained by considering $(x, \dot{x}, R, \dot{R}, \theta, \dot{\theta})$ as state variables:

$$\begin{pmatrix} (\frac{M}{m} + \sin^2\theta) & \sin\theta & 0 \\ \sin\theta & (\frac{J}{m\rho^2} + 1) & 0 \\ \cos\theta & 0 & R \end{pmatrix} \begin{pmatrix} \ddot{x} \\ \ddot{R} \\ \ddot{\theta} \end{pmatrix} = \begin{pmatrix} (R\dot{\theta}^2 + g\cos\theta)\sin\theta - \frac{1}{m}\gamma_1(\dot{x}) + \frac{F}{m} \\ R\dot{\theta}^2 + g\cos\theta - \frac{1}{m\rho}\gamma_2(\dot{R}) - \frac{C}{m\rho} \\ -2\dot{R}\dot{\theta} - g\sin\theta \end{pmatrix}$$

after inversion of the left-hand side matrix.

This representation is obtained as follows. We first twice differentiate the equations (5.2):

$$\begin{aligned} \ddot{\xi} &= \ddot{x} + \ddot{R}\sin\theta + 2\dot{R}\dot{\theta}\cos\theta + R\ddot{\theta}\cos\theta - R\dot{\theta}^2\sin\theta \\ \ddot{\zeta} &= -\ddot{R}\cos\theta + 2\dot{R}\dot{\theta}\sin\theta + R\ddot{\theta}\sin\theta + R\dot{\theta}^2\cos\theta. \end{aligned} \tag{5.3}$$

Then, we combine the 2 first equations of (5.1), first, by multiplying the first one by $\cos\theta$, the second one by $\sin\theta$ and adding them, and second, by multiplying the first one by $\sin\theta$, the second one by $\cos\theta$ and subtracting them. We get:

$$\begin{aligned} &\ddot{\xi}\cos\theta + (\ddot{\zeta} + g)\sin\theta = 0 \\ &m\ddot{\xi}\sin\theta - m(\ddot{\zeta} + g)\cos\theta = -T. \end{aligned} \tag{5.4}$$

It suffices then to eliminate T, ξ and ζ by replacing, in the two last equations of (5.1), T by its expression in (5.4) and $\ddot{\xi}$ and $\ddot{\zeta}$ by their expressions obtained in (5.3), and the first equation of (5.4).

5.2 Description of the System Trajectories

We have seen in 4.1.2, that for every n-dimensional linear controllable system, one can find a change of coordinates such that the system reads $z_i^{(n_i)} = v_i$, $i = 1, \ldots, m$, with $\sum_{i=1}^m n_i = n$. It is easily seen that the linear system state vector x may be represented as

$$x = \Phi(z_1, \ldots, z_1^{(n_1-1)}, \ldots, z_m, \ldots, z_m^{(n_m-1)})$$

where Φ is a linear invertible mapping.

Accordingly, the control vector u may be represented as

$$u = \Psi(z_1, \ldots, z_1^{(n_1)}, \ldots, z_m, \ldots, z_m^{(n_m)})$$

where the partial mapping $(z_1^{(n_1)}, \ldots, z_m^{(n_m)}) \mapsto (u_1, \ldots, u_m)$ is linear invertible for all $(z_1, \ldots, z_1^{(n_1-1)}, \ldots, z_m, \ldots, z_m^{(n_m-1)})$. We now show that this property may be generalized to the context of nonlinear systems (unfortunately not all!), if the condition that Φ is a diffeomorphism of an n-dimensional manifold is weakened. We will only require that Φ is onto from some $\mathbb{R}^{n+q}$ to $\mathbb{R}^n$, with $q \geq 0$, and invertible in a sense that will be made precise later.

Let us show on the crane example how this transformation naturally appears in the description of the system trajectories.

A most common problem for a crane operator consists in bringing a load from one point to another, with the cable in vertical position at rest at the start and at the end.

Remark that the verticality of the cable can be naturally read on the variables ξ and ζ: denoting by x_i, R_i and x_f, R_f the cart position and the cable length at initial and final times t_i and t_f respectively, we must have, at time t_i,

$$\xi(t_i) = x_i \ , \quad \zeta(t_i) = R_i \ , \quad \dot{\xi}(t_i) = 0 \ , \quad \dot{\zeta}(t_i) = 0 \ , \quad \ddot{\xi}(t_i) = 0 \ , \quad \ddot{\zeta}(t_i) = 0$$

and, at time t_f,

$$\xi(t_f) = x_f \ , \quad \zeta(t_f) = R_f \ , \quad \dot{\xi}(t_f) = 0 \ , \quad \dot{\zeta}(t_f) = 0 \ , \quad \ddot{\xi}(t_f) = 0 \ , \quad \ddot{\zeta}(t_f) = 0.$$

We indeed immediately check that these conditions imply, for $\tilde{t} = t_i, t_f$, that $\theta(\tilde{t}) = 0$, $T(\tilde{t}) = mg$, $\ddot{x}(\tilde{t}) = 0$, $\ddot{R}(\tilde{t}) = 0$, $F(\tilde{t}) = 0$ and $C(\tilde{t}) = -mg\rho$, which implies that the system is at rest at t_i and t_f.

A natural question concerns the possibility to describe all the crane trajectories, corresponding to the system (5.1)-(5.2), through the variables ξ and ζ, for all times between t_i and t_f.

Let us show that the answer is positive. For this purpose, we eliminate T from the two first equations of (5.1):

$$\tan\theta = -\frac{\ddot{\xi}}{\ddot{\zeta}+g} \tag{5.5}$$

and use (5.2):

$$\tan\theta = -\frac{\xi - x}{\zeta}\ , \quad (\xi - x)^2 + \zeta^2 = R^2. \tag{5.6}$$

Finally, eliminating $\tan\theta$ from (5.5) and (5.6), we get the differential-algebraic system:

$$\begin{cases} (\ddot{\zeta}+g)(\xi - x) = \ddot{\xi}\zeta \\ (\xi - x)^2 + \zeta^2 = R^2 \end{cases} \tag{5.7}$$

that yields

$$\begin{aligned} & x = \xi - \frac{\ddot{\xi}\zeta}{\ddot{\zeta}+g}, \qquad R^2 = \zeta^2 + \left(\frac{\ddot{\xi}\zeta}{\ddot{\zeta}+g}\right)^2, \\ & \theta = \arctan\left(\frac{\ddot{\xi}}{\ddot{\zeta}+g}\right), T^2 = m^2\left(\ddot{\xi}^2 + (\ddot{\zeta}+g)^2\right). \end{aligned} \tag{5.8}$$

Then, using the two last equations of (5.1), we get

$$\begin{aligned} F &= M\ddot{x} + \gamma_1(\dot{x}) - T\sin\theta \\ &= (M+m)\ddot{\xi} - M\frac{d^2}{dt^2}\left(\frac{\ddot{\xi}\zeta}{\ddot{\zeta}+g}\right) + \gamma_1\left(\dot{\xi} - \frac{d}{dt}\left(\frac{\ddot{\xi}\zeta}{\ddot{\zeta}+g}\right)\right) \end{aligned} \tag{5.9}$$

and, accordingly,

$$\begin{aligned} C &= -\frac{J}{\rho}\ddot{R} - \gamma_2(\dot{R}) + T\rho \\ &= -\frac{J}{\rho}\frac{d^2}{dt^2}\sqrt{\zeta^2 + \left(\frac{\ddot{\xi}\zeta}{\ddot{\zeta}+g}\right)^2} - \gamma_2\left(\frac{d}{dt}\sqrt{\zeta^2 + \left(\frac{\ddot{\xi}\zeta}{\ddot{\zeta}+g}\right)^2}\right) \\ &\quad + m\rho\sqrt{\ddot{\xi}^2 + (\ddot{\zeta}+g)^2} \end{aligned} \tag{5.10}$$

which proves that all the system variables $x, R, \theta, \xi, \zeta, T, F, C$, including the inputs, may be expressed as functions of ξ and ζ (the load coordinates) and their time derivatives up to the order 4.

Setting

$$\begin{aligned} \xi^{(4)} &= v_1 \\ \zeta^{(4)} &= v_2 \end{aligned} \tag{5.11}$$

we obtain a form similar to the linear canonical forms studied in section 4.1.2.

Moreover, if we consider the set of variables $(\xi, \dot{\xi}, \ldots, \xi^{(4)}, \zeta, \dot{\zeta}, \ldots, \zeta^{(4)})$ as a parameterization of the pair of time functions $t \mapsto (\xi(t), \zeta(t))$, the mapping Φ for which to every $(\xi, \dot{\xi}, \ldots, \xi^{(3)}, v_1, \zeta, \dot{\zeta}, \ldots, \zeta^{(3)}, v_2)$ there corresponds $(x, \dot{x}, R, \dot{R}, \xi, \dot{\xi}, \zeta, \dot{\zeta}, T, \theta, F, C)$ given by the previous formulas, is invertible in the sense that it maps trajectories to trajectories in a one to one way: we have just shown that to every trajectory $t \mapsto (\xi(t), \zeta(t))$ there corresponds in a locally unique way a trajectory $t \mapsto (x(t), R(t), \theta(t), T(t), F(t), C(t))$. Conversely, to every trajectory $t \mapsto (x(t), R(t), \theta(t), T(t), F(t), C(t))$ there corresponds a locally unique trajectory $t \mapsto (\xi(t), \zeta(t))$ by $\xi(t) = x(t) + R(t)\sin\theta(t)$ and $\zeta(t) = -R(t)\cos\theta(t)$ (see 5.2). In addition, these formulas are compatible with the operator of differentiation with respect to time: noting $F(X, \dot{X}) = 0$ the set of equations (5.1)-(5.2) with $X = (x, \dot{x}, R, \dot{R}, \xi, \dot{\xi}, \zeta, \dot{\zeta}, \theta, T, F, C)$, we say that Φ is compatible with $\frac{d}{dt}$ if

$$F(\Phi, \dot{\Phi}) = 0$$

identically.

Remark, moreover, that the trajectories $t \mapsto (\xi(t), \zeta(t))$ may be arbitrarily chosen in the set of 4 times differentiable functions since the variables $(\xi, \dot{\xi}, \ldots, \xi^{(4)}, \zeta, \dot{\zeta}, \ldots, \zeta^{(4)})$ are not required to satisfy any other differential equation than $\frac{d}{dt}\xi^{(j)} = \xi^{(j+1)}$ and $\frac{d}{dt}\zeta^{(j)} = \zeta^{(j+1)}$.

Let us briefly sketch the interest to deal with such transformations: first, we have transformed a complicated system, namely (5.1)-(5.2), in the much simpler one (5.11). In addition, we have shown that the trajectories of (5.1)-(5.2) are parameterized by the coordinates (ξ, ζ) of the load. Therefore, the computation of a trajectory of the load that goes from one point at rest to another one at rest becomes an elementary exercise, analogously to what we have presented in the linear case in section 4.1.3. There remains thus to deduce, using the previous formulas, the cart position, the length of the cable and its angle with respect to the vertical, as well as the force F and torque C that generate this trajectory. Note that we nowhere need to integrate the system differential equations (5.1)-(5.2). We will see later on that these properties are shared by a class of systems called *differentially flat systems*, or shortly *flat systems*. The variables ξ and ζ, in this formalism, constitute the two components of a *flat output*.

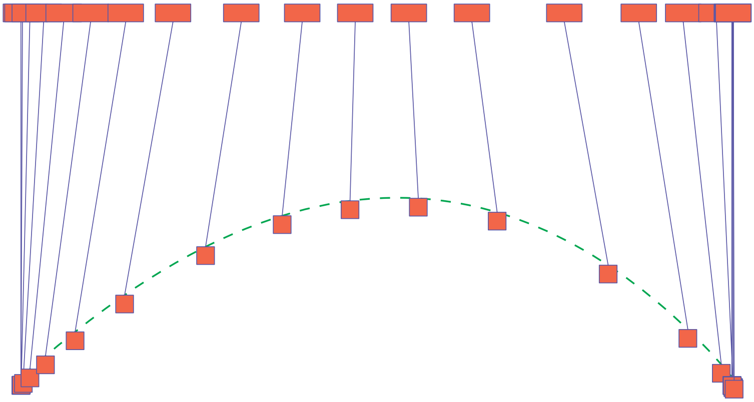

Fig. 5.2 Stroboscopic view of the displacement of the set cart-cable-load for a rest-to-rest trajectory of the load, with obstacle avoidance: in the first half of the trajectory, the cart pulls the load, then the latter overtakes the cart which then slows down the load to stop at rest at the required position.

A graphical example of a rest-to-rest displacement of the load along a parabola for obstacle avoidance is presented in Fig. 5.2, where the positions of the set cart-cable-load are decomposed along the trajectory.

For other results on various types of cranes, the reader may refer e.g. to Fliess et al. [1991, 1993], Lévine et al. [1997], Kiss [2001], Kiss et al. [1999, 2000c,b], Lévine [1999].

To conclude, we have exhibited an invertible mapping in the sense that it maps trajectories in a one-to-one way, though it doesn't preserve the state dimension: system (5.1)-(5.2) lives in a 6-dimensional manifold, whereas system (5.11) lives in an 8-dimensional space. Such a mapping is not a diffeomorphism between these manifolds since, by the constant rank theorem (theorem 2.3) it would imply that their dimensions are equal, which is not the case.

This apparent paradox may be eluded by replacing the considered finite dimensional coordinate systems by infinite dimensional extensions: in place of the mapping Φ from $(\xi, \dot{\xi}, \ldots, \xi^{(4)}, \zeta, \dot{\zeta}, \ldots, \zeta^{(4)})$ to $(x, \dot{x}, R, \dot{R}, \xi, \dot{\xi}, \zeta, \dot{\zeta}, T, \theta, F, C)$, we need to consider its extension $\tilde{\Phi}$ from the countably infinite sequence of coordinates $(\xi, \zeta, \dot{\xi}, \dot{\zeta}, \ddot{\xi}, \ddot{\zeta}, \ldots)$ to the other countably infinite sequence of coordinates $(x, \dot{x}, R, \dot{R}, \xi, \dot{\xi}, \zeta, \dot{\zeta}, T, \theta, F, C, \dot{F}, \dot{C}, \ddot{F}, \ddot{C}, \ldots)$. In this context, $\tilde{\Phi}$ is called a *Lie-Bäcklund isomorphism*. Note that $\tilde{\Phi}$ now maps spaces having the same (infinite) dimension so that the previous paradox is no more present.

5.3 Jets of Infinite Order, Change of coordinates, Equivalence

The present approach is mostly inspired by Martin [1992], Fliess et al. [1999] and, to a lesser extent, by Lévine [2006]. The reader may find detailed presentations of the geometry of jets of infinite order in Krasil'shchik et al. [1986], Zharinov [1992] (see also Anderson and Ibragimov [1979], Ibragimov [1985]). This relatively recent theory has been elaborated to study several mathematical and physical problems such as invariance groups, symmetries, conservation laws, or various questions of partial differential equations. We will only use this theory in the particular case of ordinary differential equations, where only one differential operator (with respect to time) is considered. In other respects, the notion of system equivalence based on changes of coordinates in the space of jets of infinite order, goes back to Hilbert [1912] and then to Cartan [1914]. Some variants of the present approach may be found in Fliess et al. [1995], Jakubczyk [1993], Pomet [1993, 1997], Pomet et al. [1992], Shadwick [1990], Sluis [1993], van Nieuwstadt et al. [1994].

In the previous section, we have shown that some transformations, which are not diffeomorphisms but nevertheless invertible, and take into account not only the coordinates of the original manifold but also a finite number of their successive derivatives with respect to time, this number being unknown *a priori*, are useful to describe the trajectories of the crane system.

These transformations have some important properties:

- every component of the transformation only depends on a finite, but otherwise not known in advance, number of successive time derivatives of the coordinates,
- the transformations preserve differentiation with respect to time,
- they are "invertible", in the sense that one can go back to the original coordinates by transformations of the same type.

Transformations satisfying these three properties are called *Lie-Bäcklund isomorphisms*. They will be useful to define a general equivalence relation between systems, analogously to what we have suggested, in the crane's context, to transform the crane model (5.1)-(5.2) into (5.11).

This approach is presented as well for nonlinear implicit systems of the form

$$F(x, \dot{x}) = 0 \tag{5.12}$$

where x belongs to a manifold X of dimension n and F is a C^∞ mapping from $\mathrm{T}X$ to $\mathbb{R}^{n-m}$, whose Jacobian matrix $\frac{\partial F}{\partial \dot{x}}$ has an everywhere constant rank equal to $n-m$ in a suitable open dense subset of $\mathrm{T}X$, as for explicit systems of the form

$$\dot{x} = f(x, u) \tag{5.13}$$

still with x in a manifold X of dimension n, u in an open subset U of $\mathbb{R}^m$, and with rank $\left(\frac{\partial f}{\partial u}\right)$ constant equal to m.

5.3.1 Jets of Infinite Order, Global Coordinates

An apparent paradox has been raised in the crane example: we have exhibited an invertible differentiable mapping whose inverse was also differentiable, which was not a diffeomorphism since it doesn't preserve the dimension of the original manifold. The reason is that we are dealing with a mapping from an infinite dimensional manifold whose coordinates are not only made of the original variables but also of all their jets of arbitrary order, to a comparable manifold. To make this point clear, we introduce global coordinates made of countably many components, noted as if they were successive time derivatives for reasons that will be given later, of the form

$$\overline{x} = \left(x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n, \ldots, x_1^{(k)}, \ldots, x_n^{(k)}, \ldots\right). \tag{5.14}$$

Remark that it is quite common in physics to use *generalized coordinates*, position-impulse or position-velocity, which are sufficient to describe the dynamics of a system. On the contrary, we don't restrict here to first order derivatives.

In the explicit case, where the input u is specified, we can introduce the global coordinates

$$(x, \overline{u}) = \left(x_1, \ldots, x_n, u_1, \ldots, u_m, \dot{u}_1, \ldots, \dot{u}_m, \ldots, u_1^{(k)}, \ldots, u_m^{(k)}, \ldots\right). \tag{5.15}$$

Before defining a control system in these coordinates, we must introduce a topology and a differential structure on the corresponding spaces, and then introduce vector fields and flows on these spaces in order to justify the notations (5.14) and (5.15).

5.3.2 Product Manifolds, Product Topology

We consider the set of jets of infinite order $X \times \mathbb{R}^n_\infty$, product of the manifold X and of countably infinitely many copies of $\mathbb{R}^n$:

$$X \times \mathbb{R}^n_\infty = X \times \mathbb{R}^n \times \mathbb{R}^n \times \cdots$$

whose coordinates are made of a sequence of coordinates of every term of this product (which formally correspond to (5.14)). This set, endowed with the countably infinite product topology, namely the topology generated by open sets of the form $\mathcal{O} \times \mathbb{R}^n_\infty$ where $\mathcal{O}$ is an arbitrary open set of $X \times (\mathbb{R}^n)^N$ for some N, is a manifold.

Recall that, in this topology, open sets of $X \times \mathbb{R}^n_\infty$ are generated by products made of infinitely many copies of $\mathbb{R}^n$ and a finite number of non trivial open sets.

A continuous function from $X \times \mathbb{R}^n_\infty$ to $\mathbb{R}$ is a function whose inverse image of an open set of $\mathbb{R}$ is an open set of $X \times \mathbb{R}^n_\infty$. Its local structure is given by the next proposition.

Proposition 5.1. *A function φ from $X \times \mathbb{R}^n_\infty$ to $\mathbb{R}$ is continuous if and only if, around every point, φ depends only on a finite, but otherwise arbitrary, number of coordinates of $X \times \mathbb{R}^n_\infty$ and is continuous with respect to these coordinates.*

Proof. Given a continuous function φ from $X \times \mathbb{R}^n_\infty$ to $\mathbb{R}$ and a real number r in the range of φ, for every integer k, the backward image $\varphi^{-1}(]r - \frac{1}{k}, r + \frac{1}{k}[)$ is of the form $\mathcal{O}_k \times \mathbb{R}^n_\infty$ where $\mathcal{O}_k$ is an open subset of X and finitely many copies of $\mathbb{R}^n$. Therefore, every point of the form (ξ, ζ) with $\xi \in \mathcal{O}_k$ and $\zeta \in \mathbb{R}^n_\infty$ is such that $\varphi(\xi, \zeta) \in]r - \frac{1}{k}, r + \frac{1}{k}[$. Since, for all $k' \geq k$ we have $]r - \frac{1}{k'}, r + \frac{1}{k'}[\subset]r - \frac{1}{k}, r + \frac{1}{k}[$ and thus $\varphi^{-1}(]r - \frac{1}{k'}, r + \frac{1}{k'}[) \subset \varphi^{-1}(]r - \frac{1}{k}, r + \frac{1}{k}[)$, we easily deduce that the set of elements (ξ_0, ζ) such that $\varphi(\xi_0, \zeta) = r$ satisfies $\xi_0 \in \mathcal{O}_k$ and $\zeta \in \mathbb{R}^n_\infty$ arbitrary. Consequently, φ is independent of ζ and depends only on a finite number of components of $X \times \mathbb{R}^n_\infty$ in a neighborhood of r.

The converse is immediate.

The set of continuous functions from $X \times \mathbb{R}^n_\infty$ to $\mathbb{R}$ is denoted by $C^0(X \times \mathbb{R}^n_\infty; \mathbb{R})$.

Accordingly, for every $k \geq 1$, we say that φ is of class C^k, which we note $\varphi \in C^k(X \times \mathbb{R}^n_\infty; \mathbb{R})$, if φ is k times differentiable in the usual sense *i.e.* if and only if it is k times differentiable with respect to every variable, in finite number, it depends on.

We define in a similar way a continuous, differentiable, infinitely differentiable, analytic function from $X \times \mathbb{R}^n_\infty$ to another manifold $Y \times \mathbb{R}^p_\infty$. Such a function is thus made of countably many components, each one depending only on a finite number of coordinates.

For a manifold of the type $X \times U \times \mathbb{R}^m_\infty$, formally corresponding to the coordinates (5.15), the associated product topology and the continuity and differentiability of functions defined on $X \times U \times \mathbb{R}^m_\infty$ are defined in a similar way.

5.3.3 Cartan Vector Fields, Flows, Control Systems

Let us denote by

$$\overline{x} \stackrel{def}{=} (x_1^{(0)}, \ldots, x_n^{(0)}, x_1^{(1)}, \ldots, x_n^{(1)}, \ldots)$$

the global coordinates of the manifold $X \times \mathbb{R}_\infty^n$.

A C^∞ vector field on $X \times \mathbb{R}_\infty^n$ is, as in finite dimension, a first order differential operator of the form:

$$v = \sum_{i \geq 0} \sum_{j=1}^{n} v_{i,j} \frac{\partial}{\partial x_j^{(i)}}$$

whose components, the $v_{i,j}$'s, are C^∞ functions from $X \times \mathbb{R}_\infty^n$ to $\mathbb{R}$, and thus depend on a finite number of coordinates.

Note that an arbitrary vector field on $X \times \mathbb{R}_\infty^n$ doesn't necessarily generate a flow as demonstrated by the following example:

Consider the vector field $v = v_0(x)\frac{\partial}{\partial x} + \sum_{j \geq 1} x^{(j)} \frac{\partial}{\partial x^{(j)}}$ on $\mathbb{R}_\infty$ with $\frac{dv_0}{dx} \neq 1$ for all x. Its flow, if it exists, is made of the integral curves of the system of differential equations

$$\dot{x} = v_0(x), \quad \ddot{x} = \dot{x}, \quad \ldots, \quad \dot{x}^{(j)} = x^{(j)}, \quad \ldots$$

But integral curves don't exist since $\dot{x} = v_0(x)$ and $\ddot{x} = \frac{dv_0}{dx}\dot{x} \neq \dot{x}$ according to the assumption $\frac{dv_0}{dx} \neq 1$.

Such a situation never appears in finite dimension since the coordinates are differentially independent (there is no identically satisfied relation between x_i and $x_j, \dot{x}_j, \ldots, x_j^{(k)}$ for $j \neq i$ and $k \geq 0$), whereas here, each coordinate is the derivative of the preceding one.

On the contrary, a vector field constructed by prolongation of a vector field of X always generates a flow: let $v_0 \in \mathrm{T}X$ where X is a n-dimensional manifold. Its prolongation on $X \times \mathbb{R}_\infty^n$ is defined by

$$\overline{v} = v_0(\overline{x})\frac{\partial}{\partial x} + \sum_{j \geq 0} x^{(j+1)} \frac{\partial}{\partial x^{(j)}} \tag{5.16}$$

where $x = (x_1, \ldots, x_n)$, $x^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)})$ and with the notation $\alpha \frac{\partial}{\partial x} = \sum_{i=1}^n \alpha_i \frac{\partial}{\partial x_i}$. In this case, we easily verify that nothing has been changed with respect to the differential equation on X:

$$\dot{x} = v_0(\overline{x}), \quad \ddot{x} = \ddot{x}, \quad \ldots, \quad \dot{x}^{(j)} = x^{(j+1)}, \quad \ldots$$

The vector fields of the form (5.16) are called *Cartan vector fields.*

This definition is indeed in accordance with the notion of Lie derivative since, given an arbitrary function h of class C^∞ from $X \times \mathbb{R}_\infty^n$ to $\mathbb{R}$, its Lie derivative along the Cartan field $\overline{v}$ is given by:

$$L_{\overline{v}} h(\overline{x}) = v_0(\overline{x})\frac{\partial h}{\partial x}(\overline{x}) + \sum_{j \geq 0} x^{(j+1)} \frac{\partial h}{\partial x^{(j)}}(\overline{x})$$

where all the terms of this series but a finite number, corresponding to the coordinates on which h effectively depends, vanish identically.

Accordingly, global coordinates of the manifold $X \times U \times \mathbb{R}^m_\infty$ are denoted by

$$(x, \overline{u}) \stackrel{def}{=} (x_1, \ldots, x_n, u_1^{(0)}, \ldots, u_m^{(0)}, u_1^{(1)}, \ldots, u_m^{(1)}, \ldots)$$

A C^∞ vector field on $X \times U \times \mathbb{R}^m_\infty$ is a first order differential operator of the form:

$$w = \sum_{i=1}^{n} \tilde{w}_i \frac{\partial}{\partial x_i} + \sum_{i \geq 0} \sum_{j=1}^{m} w_{i,j} \frac{\partial}{\partial u_j^{(i)}}$$

whose components, the $\tilde{w}_i, w_{i,j}$'s are C^∞ functions from $X \times U \times \mathbb{R}^m_\infty$ to $\mathbb{R}$.

Again, a *Cartan vector field* is defined as a prolonged vector field on X that depends on the m independent variables $u = (u_1, \ldots, u_m)$ and their higher order jets:

$$\overline{f} = \sum_{i=1}^{n} f_i(x, \overline{u}) \frac{\partial}{\partial x_i} + \sum_{j \geq 0} \sum_{i=1}^{m} u_i^{(j+1)} \frac{\partial}{\partial u_i^{(j)}} \tag{5.17}$$

5.3.3.1 Some Remarkable Cartan Vector Fields

- The so-called *trivial* vector field on $X \times \mathbb{R}^n_\infty$ is defined by:

$$\tau_X = \sum_{i \geq 0} \sum_{j=1}^{n} x_j^{(i+1)} \frac{\partial}{\partial x_j^{(i)}}. \tag{5.18}$$

  To the trivial vector field τ_X there corresponds the trivial system $\dot{x}^{(j)} = x^{(j+1)}$ for all j, for which any infinitely differentiable function $t \mapsto x(t)$ on X is an integral curve. Moreover, h being an arbitrary function, its Lie derivative along τ_X is given by

$$L_{\tau_X} h = \sum_{i \geq 0} \sum_{j=1}^{n} x_j^{(i+1)} \frac{\partial h}{\partial x_j^{(i)}} = \frac{dh}{dt}.$$

  and we can thus identify τ_X as the differential operator $\frac{d}{dt}$. The associated implicit system is empty, *i.e.* given by $F \equiv 0$.
- A linear Cartan field on $X \times U \times \mathbb{R}^m_\infty$ is given by

$$\lambda_{A,B} = \sum_{i=1}^{n} (Ax + Bu)_i \frac{\partial}{\partial x_i} + \sum_{j \geq 0} \sum_{i=1}^{m} u_i^{(j+1)} \frac{\partial}{\partial u_i^{(j)}} \tag{5.19}$$

where $(Ax+Bu)_i$ is the ith component of the vector $Ax+Bu$. One immediately verifies that it corresponds to the linear system $\dot{x} = Ax + Bu$. The corresponding implicit system is thus given by $C(\dot{x} - Ax) = 0$ where C is a matrix of rank $n-m$ such that $CB = 0$.

5.3.3.2 Jets of Infinite Order, Internal Formulation

We are now in order to justify the name of manifold of jets: if we associate, to the manifold $X \times U \times \mathbb{R}^m_\infty$ a Cartan field of the form:

$$\overline{f} = \sum_{i=1}^{n} f_i(x, \overline{u}) \frac{\partial}{\partial x_i} + \sum_{j \geq 0} \sum_{i=1}^{m} u_i^{(j+1)} \frac{\partial}{\partial u_i^{(j)}} \tag{5.20}$$

the Lie derivative of a function h of class C^∞ from $X \times U \times \mathbb{R}^m_\infty$ to $\mathbb{R}$, along $\overline{f}$ is given by

$$\frac{dh}{dt} = L_{\overline{f}} h.$$

In the following successive particular cases $h = x_i$ and $h = u_i^{(j)}$, the latter Lie derivative reads:

$$\frac{dx_i}{dt} = L_{\overline{f}} x_i = f_i(x, \overline{u}) \frac{\partial x_i}{\partial x_i} = f_i(x, \overline{u})$$

and

$$\frac{du_i^{(j)}}{dt} = L_{\overline{f}} u_i^{(j)} = u_i^{(j+1)} \frac{\partial u_i^{(j)}}{\partial u_i^{(j)}} = u_i^{(j+1)}$$

therefore, the coordinates, which are indeed independent, satisfy $\frac{du_i^{(j)}}{dt} = u_i^{(j+1)}$ for all $i = 1, \ldots, m$ and all $j \geq 0$ along the flow of the vector field (5.20).

Remark that the first term f of the Cartan field $\overline{f}$ is allowed to depend not only on u but also on a finite number of derivatives of u, *i.e.* $f(x, u, \dot{u}, \ldots, u^{(\alpha)})$.

5.3.3.3 Jets of Infinite Order, External Formulation

The notion of manifold of jets of infinite order on $X \times \mathbb{R}^n_\infty$ may be similarly justified: consider the trivial Cartan field τ_X defined by (5.18). As was just remarked, we have $\dot{x}_i^{(j)} = x_i^{(j+1)}$ for all j and all $i = 1, \ldots, n$, which lets us recover the natural meaning of $x_i^{(j)}$ as the jth order time derivative of x_i.

5.3.3.4 Control Systems, Internal Formulation

In our formalism, a control system is given by the particularly simple definition:

Definition 5.1. A control system is the data of a pair $(\mathfrak{X}, \overline{f})$ where $\mathfrak{X}$ is a manifold of jets of infinite order whose Cartan vector field is $\overline{f}$, of the form (5.20).

In other words, a system is given, in the coordinates $(x, u, \dot{u}, \ldots)$ by

$$\begin{aligned} \dot{x} &= f(x, \overline{u}) \\ \dot{u} &= \dot{u} \\ \ddot{u} &= \ddot{u} \\ &\vdots \end{aligned}$$

In this approach, the control variables appear in the Cartan field as an infinite sequence $\overline{u} = (u, \dot{u}, \ldots)$ of vectors of $\mathbb{R}^m$, at every time t, and the function $t \mapsto u(t)$ results from the integration of the flow, as well as the trajectory $t \mapsto x(t)$. On the other hand, in the classical finite dimensional approach, the control, that may be seen as parameterizing the vector field $f(x, \overline{u})$, is *a priori* given as a function $t \mapsto u(t)$ on a given interval of time, thus belonging to a function space, of infinite dimension. Therefore, the two formalisms handle the same "amount" of information.

Note also that the Cartan vector field $\sum_{i=1}^{m} \left(u_i \frac{\partial}{\partial x_i} + \sum_{j \geq 0} u_i^{(j+1)} \frac{\partial}{\partial u_i^{(j)}} \right)$ corresponds to the m-dimensional system with m inputs $\dot{x}_i = u_i$, $i = 1, \ldots, m$, and that this vector field is precisely the trivial field τ_X for $X = \mathbb{R}^m$, in the coordinates $(x_1, \ldots, x_m, \dot{x}_1 = u_1, \ldots, \dot{x}_m = u_m, \ldots, x_1^{(j+1)} = u_1^{(j)}, \ldots, x_m^{(j+1)} = u_m^{(j)}, \ldots)$.

In the same spirit, if, in the Cartan field $\overline{f}$ defined by (5.20), f depends on u and derivatives up to $u^{(r)}$ for some finite r, renaming $\tilde{x} = (x, u, \ldots, u^{(r-1)})$, $u^{(r)} = v$ and $\tilde{f}(\tilde{x}, v) = f(x, u, \ldots, u^{(r)})$, we get $\overline{f} = \tilde{f}(\tilde{x}, v) \frac{\partial}{\partial \tilde{x}} + \sum_{j \geq 0} \sum_{i=1}^{m} v_i^{(j+1)} \frac{\partial}{\partial v_i^{(j)}}$ as a Cartan field in the coordinates $(\tilde{x}, \overline{v})$, this Cartan field now depending only on v and not on its derivatives, though it describes the same differential equations. Therefore, we only need to consider Cartan fields of the form

$$\overline{f} = \sum_{i=1}^{n} f_i(x, u) \frac{\partial}{\partial x_i} + \sum_{j \geq 0} \sum_{i=1}^{m} u_i^{(j+1)} \frac{\partial}{\partial u_i^{(j)}} \tag{5.21}$$

5.3.3.5 Control Systems, External Formulation

In the manifold of jets of infinite order $X \times \mathbb{R}^n_\infty$ endowed with the trivial Cartan field $\tau_X = \frac{d}{dt} = \sum_{j\geq 0}\sum_{i=1}^n x_i^{(j+1)}\frac{\partial}{\partial x_i^{(j)}}$, let us show that to the implicit system (5.12), one can associate a flow generated by a vector field of the form (5.20).

Indeed, if the coordinates (5.14) satisfy the constraint (5.12), since $\operatorname{rank}\left(\frac{\partial F}{\partial \dot{x}}\right) = n-m$, according to the implicit function Theorem, there exists a mapping $\tilde{f}$ of class C^∞ from an open subset of $X \times \mathbb{R}^m$ to $\mathbb{R}^{n-m}$ such that

$$
\begin{aligned}
&\dot{x}_{m+1} = \tilde{f}_1(x, \dot{x}_1, \ldots, \dot{x}_m)\\
&\vdots\\
&\dot{x}_n = \tilde{f}_{n-m}(x, \dot{x}_1, \ldots, \dot{x}_m)
\end{aligned}
$$

and, setting $u_1 = \dot{x}_1, \ldots, u_m = \dot{x}_m$,

$$
f(x,u) = \begin{pmatrix} u_1 \\ \vdots \\ u_m \\ \tilde{f}_1(x,u) \\ \vdots \\ \tilde{f}_{n-m}(x,u) \end{pmatrix}
$$

we obtain the vector field (5.21) by noticing that, since $u_i = \dot{x}_i$, $i = 1, \ldots, m$, we get $u_i^{(j+1)} = x_i^{(j)}$ for all $j \geq 0$. The coordinates defined by (5.15) are thus recovered by replacing $\dot{x}, \ddot{x}, \ldots$ by $f(x,u), L_{\tau_X} f(x,u) = \frac{d}{dt}(f(x,u)), \ldots$. We thus say that the vector field $\overline{f}$ is *compatible* with the implicit system (5.12), or, in other words, $\frac{d^k}{dt^k} F(x, \overline{f}(x, \overline{u})) = 0$ for all $\overline{u} = (u, \dot{u}, \ddot{u}, \ldots)$ and all $k \geq 0$.

Conversely, one can view the explicit vector field expressed in the internal coordinates as an implicit system expressed in a set of external coordinates. Indeed, consider the vector field (5.21) expressed in the coordinates (5.15). Since $\operatorname{rank}\left(\frac{\partial f}{\partial u}\right) = m$, let us first permute the rows of f, if necessary, such that its Jacobian matrix of the m first rows with respect to u have rank m. Let us call $\tilde{x}$ the vector obtained after the above permutation, if any, and $\tilde{f}(\tilde{x}, u)$ the corresponding vector field. Again by the implicit function Theorem, we get $u = U(\tilde{x}, \dot{\tilde{x}}_1, \ldots, \dot{\tilde{x}}_m)$. Reinjecting this expression in $\dot{\tilde{x}} = \tilde{f}(\tilde{x}, u)$ and eliminating the m first equations which are thus identically satisfied, the $n-m$ remaining equations are of the form $F_{i-m}(\tilde{x}, \dot{\tilde{x}}) = \dot{\tilde{x}}_i - \tilde{f}_i(\tilde{x}, U(\tilde{x}, \dot{\tilde{x}}_1, \ldots, \dot{\tilde{x}}_m)) = 0$, $i = m+1, \ldots, n$, with $\operatorname{rank}\left(\frac{\partial F}{\partial \dot{x}}\right) = n-m$. Moreover, in the coordinates (5.15), replacing u and its successive derivatives by their expression in function of x and its derivatives, these new expressions are clearly expressed in the coordinates (5.14), which proves the result.

We also can remark that the description of the solutions of $F = 0$ doesn't require the knowledge of an explicit compatible vector field. The equation $F(x, \dot{x}) = 0$ suffices to express $n - m$ components of $\dot{x}$ in function of the remaining m ones, which are free, and of x. Thus, the successive derivatives of x and $\dot{x}$ may be seen as solution of the countably many equations

$$L_{\tau_X}^k F(x, \dot{x}) = \frac{d^k}{dt^k}\left(F(x, \dot{x})\right) = 0, \quad \forall k \geq 0. \tag{5.22}$$

Indeed, from the last equation, one can extract $n-m$ components of $x^{(k+1)}$ in function of the m other ones, variables substituted by their values obtained at the previous steps. One can easily verify that this elimination process is the same as the one presented above to construct a compatible vector field. The manifold $\mathfrak{X}_0$ of the jets of infinite order included in $X \times \mathbb{R}_\infty^n$ and satisfying $F = 0$ is thus given by

$$\mathfrak{X}_0 = \{\overline{x} \in X \times \mathbb{R}_\infty^n | \frac{d^k}{dt^k}\left(F(x, \dot{x})\right) = 0, \forall k \geq 0\}. \tag{5.23}$$

To summarize, the manifold $X \times U \times \mathbb{R}_\infty^m$ endowed with the vector field (5.20), or the manifold $X \times \mathbb{R}_\infty^n$, with the constraint $F(x, \dot{x}) = 0$, using (5.23), equally define a manifold of jets of infinite order.

Definition 5.2. We call *implicit control system*, a triplet $(\mathfrak{X}, \tau_X, F)$, where $\mathfrak{X}$ is a manifold of the type $X \times \mathbb{R}_\infty^n$ endowed with its trivial Cartan field $\tau_X = \sum_{j \geq 0} \sum_{i=1}^n x_i^{(j+1)} \frac{\partial}{\partial x_i^{(j)}}$, and with the countably many equations $L_{\tau_X}^k F(x, \dot{x}) = \frac{d^k}{dt^k} F(x, \dot{x}) = 0$ for all $k \geq 0$.

By construction, there clearly exists a vector field $\overline{f}$ of the type (5.20) compatible with $L_{\tau_X}^k F = 0$ for all $k \geq 0$ and such that the flow associated to the triplet $(\mathfrak{X}, \tau_X, F)$ is identical to the one associated to the pair $(X \times U \times \mathbb{R}_\infty^m, \overline{f})$ and conversely, which achieves to prove the equivalence of the internal and external formulations.

5.3.4 Lie-Bäcklund Equivalence

5.3.4.1 Image of a Cartan Field by a Mapping

We consider two manifolds of jets of infinite order, $\mathfrak{X} = X \times U \times \mathbb{R}^m$ with its Cartan field $\overline{f}$ and $\mathfrak{Y} = Y \times V \times \mathbb{R}^q$ with its Cartan field $\overline{g}$. Let Φ be a mapping of class C^∞ from $\mathfrak{Y}$ to $\mathfrak{X}$, invertible in the sense that there exists a mapping Ψ of class C^∞ from $\mathfrak{X}$ to $\mathfrak{Y}$ such that $(x, \overline{u}) = \Phi(y, \overline{v})$ is equivalent to $(y, \overline{v}) = \Psi(x, \overline{u})$.

One can define the image $\Phi_*\overline{g}$ of the Cartan field $\overline{g}$ as in finite dimension, by setting $(x,\overline{u}) = \Phi(y,\overline{v})$ and computing $(\dot{x},\dot{u},\ddot{u},\ldots) = \frac{d}{dt}\Phi(y,\overline{v})$.

We get $\dot{x} = L_{\overline{g}}\varphi_0(y,\overline{v})$, $\dot{u} = L_{\overline{g}}\varphi_1(y,\overline{v})$, $\ldots$, and thus $\overline{g}$ is transformed in the vector field

$$\Phi_*\overline{g} = (L_{\overline{g}}\varphi_0)\circ\Psi\frac{\partial}{\partial x} + \sum_{j\geq 0}(L_{\overline{g}}\varphi_{j+1})\circ\Psi\frac{\partial}{\partial u^{(j)}} \tag{5.24}$$

which is the analogue, in infinite dimensions, of the formula of Definition 2.7.

Let us give a simple necessary condition for the image of a Cartan field on $\mathfrak{Y}$ to be a Cartan field on $\mathfrak{X}$. For all j, we must have

$$\frac{du^{(j)}}{dt} = L_{\Phi_*\overline{g}}u^{(j)} = u^{(j+1)} = \varphi_{j+2}\circ\Psi$$

or, according to (5.24),

$$L_{\Phi_*\overline{g}}u^{(j)} = (L_{\overline{g}}\varphi_0)\circ\Psi\frac{\partial u^{(j)}}{\partial x} + \sum_{k\geq 0}(L_{\overline{g}}\varphi_{k+1})\circ\Psi\frac{\partial u^{(j)}}{\partial u^{(k)}} = (L_{\overline{g}}\varphi_{j+1})\circ\Psi.$$

Thus the required necessary condition is

$$L_{\overline{g}}\varphi_{j+1} = \varphi_{j+2}, \quad \forall j \geq 0.$$

5.3.4.2 Equivalence

We are now in order to define the announced equivalence relation between two systems $(\mathfrak{X},\overline{f})$ and $(\mathfrak{Y},\overline{g})$ with $\mathfrak{X} = X\times U\times\mathbb{R}^m_\infty$, X being a manifold of dimension n, and $\mathfrak{Y} = Y\times V\times\mathbb{R}^q_\infty$, Y being a manifold of dimension p.

The external point of view, *i.e.* the equivalence between the two triplets $(\mathfrak{X},\tau_X,F)$ with $\mathfrak{X} = X\times\mathbb{R}^n_\infty$, and $(\mathfrak{Y},\tau_Y,G)$ with $\mathfrak{Y} = Y\times\mathbb{R}^p_\infty$, will be treated in a second step.

Definition 5.3. Let $(y_0,\overline{v}_0)\in\mathfrak{Y}$ and $(x_0,\overline{u}_0)\in\mathfrak{X}$. We say that the two systems $(\mathfrak{X},\overline{f})$ and $(\mathfrak{Y},\overline{g})$ are locally equivalent at $((y_0,\overline{v}_0),(x_0,\overline{u}_0))$ in the sense of Lie-Bäcklund, or shortly, L-B equivalent, if there exists

- a mapping Φ from $\mathfrak{Y}$ to $\mathfrak{X}$ such that $(x_0,\overline{u}_0) = \Phi(y_0,\overline{v}_0)$, of class C^∞ from a neighborhood of $(y_0,\overline{v}_0)\in\mathfrak{Y}$ to a neighborhood of $(x_0,\overline{u}_0)\in\mathfrak{X}$ and invertible in these neighborhoods, such that $\Phi_*\overline{g} = \overline{f}$,
- and conversely, if there exists a mapping Ψ of class C^∞ from a neighborhood of $(x_0,\overline{u}_0)\in\mathfrak{X}$ to a neighborhood of $(y_0,\overline{v}_0)\in\mathfrak{Y}$ and invertible in these neighborhoods, such that $\Psi_*\overline{f} = \overline{g}$.

Such a mapping Φ (or Ψ) is called *local Lie-Bäcklund isomorphism*.

Φ is thus a Lie-Bäcklund isomorphism if for all $(x, \overline{u}) = (x, u, \dot{u}, \ldots)$, in a neighborhood of $(x_0, \overline{u}_0)$ in $\mathfrak{X}$, there exists $(y, \overline{v}) = (y, v, \dot{v}, \ldots) \in \mathfrak{Y}$ in a neighborhood of $(y_0, \overline{v}_0)$, with $(x_0, \overline{u}_0) = \Phi(y_0, \overline{v}_0)$, such that $(x, \overline{u}) = \Phi(y, \overline{v})$, or

$$x = \varphi_0(y, \overline{v}), u = \varphi_1(y, \overline{v}), \dot{u} = \varphi_2(y, \overline{v}), \ldots$$

and conversely, if for all $(y, \overline{v}) = (y, v, \dot{v}, \ldots) \in \mathfrak{Y}$ in a neighborhood of $(y_0, \overline{v}_0)$, there exists $(x, \overline{u}) = (x, u, \dot{u}, \ldots)$ in a neighborhood of $(x_0, \overline{u}_0)$ in $\mathfrak{X}$, such that $(y, \overline{v}) = \Psi(x, \overline{u})$, or

$$y = \psi_0(x, \overline{u}), v = \psi_1(x, \overline{u}), \dot{v} = \psi_2(x, \overline{u}), \ldots$$

Moreover, Φ and Ψ must locally preserve the derivation with respect to time both on $\mathfrak{X}$ and on $\mathfrak{Y}$: $\Phi_* \overline{g} = \overline{f}$ and $\Psi_* \overline{f} = \overline{g}$. In other words, $\frac{d}{dt}\Phi(y, \overline{v}) = L_{\overline{f}}(x, \overline{u})$, or

$$(L_{\overline{g}}\varphi_0) \circ \Psi = f, \quad (L_{\overline{g}}\varphi_j) \circ \Psi = u^{(j)}, \ \forall j \geq 1,$$

and $\frac{d}{dt}\Psi(x, \overline{u}) = L_{\overline{g}}(y, \overline{v})$, or

$$\left(L_{\overline{f}}\psi_0\right) \circ \Phi = g, \quad \left(L_{\overline{f}}\psi_j\right) \circ \Phi = v^{(j)}, \ \forall j \geq 1.$$

Thus, we must have

$$f(\varphi_0(y, \overline{v}), \varphi_1(y, \overline{v})) = L_{\overline{g}}\varphi_0(y, \overline{v}), \quad \varphi_{j+1}(y, \overline{v}) = L_{\overline{g}}\varphi_j(y, \overline{v}), \ \forall j \geq 1$$

and

$$g(\psi_0(x, \overline{u}), \psi_1(x, \overline{u})) = L_{\overline{f}}\psi_0(x, \overline{u}), \quad \psi_{j+1}(x, \overline{u}) = L_{\overline{f}}\psi_j(x, \overline{u}), \ \forall j \geq 1.$$

Example 5.1. Consider the linear scalar system $\dot{x} = u$ ($x \in \mathbb{R}$, $u \in \mathbb{R}$). There corresponds the manifold $\mathbb{R}_\infty$ and the trivial Cartan field $\overline{f} = u\frac{\partial}{\partial x} + \sum_{j \geq 0} u^{(j+1)} \frac{\partial}{\partial u^{(j)}}$. It is L-B equivalent to any system of the form $\ddot{y} = g(y, \dot{y}) + v$ defined on $(\mathbb{R}_\infty, \overline{g})$ with $\overline{g} = \dot{y}\frac{\partial}{\partial y} + (g(y, \dot{y}) + v)\frac{\partial}{\partial \dot{y}} + \sum_{j \geq 0} v^{(j+1)} \frac{\partial}{\partial v^{(j)}}$. Indeed, it suffices to set $y = x$ and $v = \dot{u} - g(x, u)$, and conversely, $x = y$ and $u = \dot{y}$, or, equivalently, $\Phi(y, \overline{v}) = (y, \dot{y}, \ddot{y}, \ldots)$ and $\Psi(x, \overline{u}) = (x, u, \dot{u} - g(x, u), L_{\overline{f}}(\dot{u} - g(x, u)), \ldots)$. We easily verify that $\Phi_* \overline{g} = \overline{f}$ and $\Psi_* \overline{f} = \overline{g}$. Note in particular that the dimension of the first system is equal to 1, whereas the dimension of the second one is equal to 2. Therefore, the state dimension is not preserved by Lie-Bäcklund isomorphisms, which, contrarily to diffeomorphisms, doesn't prevent them from being invertible.

In fact, this example illustrates a more general result. We call *integrator of order* q, with q arbitrary integer, an integrator of the form $z^{(q)} = v$ and $u = z$, or, equivalently, $u^{(q)} = v$.

Proposition 5.2. *A system and the system obtained by prolonging itself by an integrator of arbitrary order are L-B equivalent.*

Proof. If the system is given by $(\mathfrak{X}, \overline{f})$ with $\overline{f}$ given by (5.20), the prolongation by an integrator transforms $\mathfrak{X} = X \times U \times \mathbb{R}^m_\infty$ in $X \times \mathbb{R}^{mq} \times U \times \mathbb{R}^m_\infty$ and the vector field $\overline{f}$ in itself. The equivalence is therefore obvious.

We now state the L-B equivalence in the external formulation:

Proposition 5.3. *Consider two implicit systems $(\mathfrak{X}, \tau_X, F)$, $\mathfrak{X} = X \times \mathbb{R}^n_\infty$, in a neighborhood V of $\overline{x}_0 \in \mathfrak{X}_0$ with $\mathfrak{X}_0$ defined by*

$$\mathfrak{X}_0 = \{\overline{x} \in X \times \mathbb{R}^n_\infty | L^k_{\tau_X}(F(x, \dot{x})) = 0, \forall k \geq 0\}, \tag{5.25}$$

where F satisfies $\operatorname{rank}\left(\frac{\partial F}{\partial \dot{x}}\right) = n - m$ *in V, and $(\mathfrak{Y}, \tau_Y, G)$, $\mathfrak{Y} = \mathbb{R} \times Y \times \mathbb{R}^p_\infty$, in a neighborhood W of $\overline{y}_0 \in \mathfrak{Y}_0$, with*

$$\mathfrak{Y}_0 = \{\overline{y} \in \mathbb{R} \times Y \times \mathbb{R}^p_\infty | L^k_{\tau_Y}(G(y, \dot{y})) = 0, \forall k \geq 0\}, \tag{5.26}$$

where G satisfies $\operatorname{rank}\left(\frac{\partial G}{\partial \dot{y}}\right) = p - q$ *in W.*

The two following properties are locally equivalent:

(i) The systems (5.25) and (5.26) are such that

- *there exists a mapping Φ of class C^∞ and invertible from W to V such that $\Phi(\overline{y}_0) = \overline{x}_0$ and $\Phi_* \tau_Y = \tau_X$,*
- *and a mapping Ψ of class C^∞ and invertible from V to W such that $\Psi(\overline{x}_0) = \overline{y}_0$ and $\Psi_* \tau_X = \tau_Y$.*

(ii) For every Cartan field $\overline{f}$ on $X \times \mathbb{R}^m_\infty$ compatible with (5.25) in V and every Cartan field $\overline{g}$ on $Y \times \mathbb{R}^q_\infty$ compatible with (5.26) in W, the systems $(X \times \mathbb{R}^m_\infty, \overline{f})$ and $(Y \times \mathbb{R}^q_\infty, \overline{g})$ are locally L-B equivalent at $(x_0, \overline{u}_0)$ and $(y_0, \overline{v}_0)$, for $\overline{u}_0$ and $\overline{v}_0$ suitably chosen.

Proof. **(i) implies (ii) :** Let Φ and Ψ satisfy **(i)**. For all $\overline{g}$ compatible with (5.26) and if $\overline{y} \in W$ we have $G(y, g(y, v)) = 0$ for all v in a suitable open subset of $\mathbb{R}^q$. Since by assumption $\overline{x} = \Phi(\overline{y}) \in V \subset \mathfrak{X}_0$, with $\Phi = (\varphi_0, \varphi_1, \ldots)$, we have $x = (\varphi_0(y, g(y, v)), \frac{dg}{dt}(y, v, \dot{v}), \ldots) \stackrel{\Delta}{=} \tilde{\varphi}_0(y, \overline{v})$, and $\overline{x}$ satisfies $L^k_{\tau_X} F(\overline{x}) = 0$ for all $k \geq 0$. Let then $\overline{v}_0$ be such that $\overline{x}_0 = \Phi(y_0, g(y_0, v_0), \frac{d}{dt} g(y_0, v_0, \dot{v}_0), \ldots)$. For all $\overline{f}$ compatible with (5.25) we have $\dot{x} = f(x, u)$ for $u = \mu(x, \dot{x})$, or $u = \mu(\tilde{\varphi}_0(y, \overline{v}), \frac{d}{dt}\tilde{\varphi}_0(y, \overline{v})) \stackrel{\Delta}{=} \tilde{\varphi}_1(y, \overline{v})$. Using $\Phi_* \tau_Y = \tau_X$, we get

$$\begin{aligned}
&f(x, u)\Big|_{(\tilde{\varphi}_0(y,\overline{v}), \tilde{\varphi}_1(y,\overline{v}))} = L_{\tau_X} x\Big|_{(\tilde{\varphi}_0(y,\overline{v}), \tilde{\varphi}_1(y,\overline{v}))} = L_{\tau_Y} \varphi_0(y, L_{\overline{g}} y, L^2_{\overline{g}} y, \ldots) \\
&= \sum_{j \geq 0} \frac{\partial \varphi_0}{\partial y^{(j)}} L_{\overline{g}}\left(L^j_{\overline{g}} y\right) = g \frac{\partial \varphi_0}{\partial y^{(j)}} \frac{\partial L^j_{\overline{g}} y}{\partial y} + \sum_{k \geq 0} v^{(k+1)} \frac{\partial \varphi_0}{\partial y^{(j)}} \frac{\partial L^j_{\overline{g}} y}{\partial v^{(k)}} \\
&= g \frac{\partial \tilde{\varphi}_0}{\partial y} + \sum_{k \geq 0} v^{(k+1)} \frac{\partial \tilde{\varphi}_0}{\partial v^{(k)}} = (\tilde{\varphi}_0)_* \overline{g}(y, \overline{v}).
\end{aligned}$$

Similarly, $\dot{u} = \frac{d}{dt}u = L_{\tau_X}\mu(\overline{x}) = \frac{d}{dt}\tilde{\varphi}_1(y, \overline{v}) = g(y,v)\frac{\partial \tilde{\varphi}_1}{\partial y} + \sum_{j\geq 0} v^{(j+1)} \frac{\partial \tilde{\varphi}_1}{\partial v^{(j)}} = (\tilde{\varphi}_1)_* \overline{g}$, which proves that $(x, \overline{u}) = \tilde{\Phi}(y, \overline{v})$ with $\tilde{\Phi} = (\tilde{\varphi}_0, \tilde{\varphi}_1, \ldots)$ and $\overline{f} = \tilde{\Phi}_* \overline{g}$. Let us define $\overline{u}_0$ by $(x_0, \overline{u}_0) = \tilde{\Phi}(y_0, \overline{v}_0)$. Symmetrically, we have $(y, \overline{v}) = \tilde{\Psi}(x, \overline{u})$ with $\tilde{\Psi} = (\tilde{\psi}_0, \tilde{\psi}_1, \ldots)$ et $\tilde{\psi}_0(x, \overline{u}) = \psi_0(x, f(x,u), \frac{df}{dt}(x, u, \dot{u}), \ldots)$, $\tilde{\psi}_1(x, \overline{u}) = \nu(\tilde{\psi}_0(x, \overline{u}), \frac{d}{dt}\tilde{\psi}_0(x, \overline{u}))$, $v = \nu(x, \dot{x})$, and thus $\tilde{\Phi}$ has inverse $\tilde{\Psi}$ with $(y_0, \overline{v}_0) = \tilde{\Psi}(x_0, \overline{u}_0)$, and $\overline{g} = \tilde{\Psi}_* \overline{f}$. We have thus proven that if the implicit systems $(X \times \mathbb{R}^n_\infty, \tau_X, F)$ et $(Y \times \mathbb{R}^p_\infty, \tau_Y, G)$ satisfy **(i)**, then the implicit systems $(X \times \mathbb{R}^m_\infty, \overline{f})$ and $(Y \times \mathbb{R}^q_\infty, \overline{g})$, for all Cartan fields $\overline{f}$ compatible with $F = 0$ and $\overline{g}$ compatible with $G = 0$, are locally L-B equivalent at $(x_0, \overline{u}_0)$, $(y_0, \overline{v}_0)$, for $\overline{u}_0$ and $\overline{v}_0$ suitably chosen (their choice depending on f and g).

(ii) implies (i) : The proof follows the same lines and is left to the reader.

Definition 5.4. We say that the implicit systems (5.25) and (5.26) satisfying the condition ***(i)*** of Proposition 5.3 are locally L-B equivalent at $\overline{x}_0$ and $\overline{y}_0$.

Remark 5.1. If we generalize to nonlinear systems the equivalence relation that we introduced, for linear systems, in definition 4.2, we get:

Two systems $\dot{x} = f(x,u)$, with $x \in X$, and $\dot{y} = g(y,v)$, with $y \in Y$, are equivalent by diffeomorphism and state feedback, or shortly, static feedback equivalent *if there exists a diffeomorphism φ from Y to X and a state feedback $v = \alpha(y,u)$ where α is invertible with respect to u for all y, such that the closed loop system $\dot{y} = g(y, \alpha(y,u))$, transformed by $x = \varphi(y)$, satisfies $\dot{x} = f(x,u)$.*

According to the invertibility of φ and α, this clearly defines an equivalence relation.

However, if two systems are static feedback equivalent, the dimensions of x and y must be equal as well as the dimensions of u and v. Moreover, since $(x,u) = (\varphi(y), \alpha^{-1}(y,v))$ and $(y,v) = (\varphi^{-1}(x), \alpha(x,u))$, we immediately deduce that the two systems are also L-B equivalent.

Therefore the L-B equivalence is strictly coarser than the static feedback equivalence. Nevertheless, L-B equivalence possesses nice enough properties as will be seen later.

5.3.5 Properties of the L-B Equivalence

As we just have seen, L-B equivalence doesn't preserve the state dimension. However, the dimension of the input vector is preserved. The number of independent inputs, if $\mathfrak{X} = X \times U \times \mathbb{R}^m_\infty$, is simply the dimension of the space $\mathbb{R}^m$ copied countably many times. In the case $\mathfrak{X} = X \times \mathbb{R}^n_\infty$ with $F(x, \dot{x}) = 0$, the number of independent inputs is the integer equal to the difference between the dimension of X, *i.e.* n, and the number of independent equations, *i.e.* $n - m$, since indeed $m = n - (n - m)$.

Theorem 5.1. *If two systems are L.B. equivalent, they have the same number of independent inputs.*

The proof is based on the following lemma:

Lemma 5.1. *Let us denote by Φ_k the mapping made of the $k+1$ first components of Φ, i.e.*

$$\Phi_k(y,\overline{v}) = (\varphi_0(y,\overline{v}),\ldots,\varphi_k(y,\overline{v}))\,.$$

If Φ is a Lie-Bäcklund isomorphism, then Φ_k is locally onto for all k. Similarly, if Ψ is the inverse isomorphism, Ψ_k is locally onto for all k.

Proof. If $\Phi_0 = \varphi_0$ is not onto, there exists at least one $\tilde{x} \in X$ such that $\varphi_0(y,\overline{v}) \neq \tilde{x}$ for all $(y,\overline{v}) \in \mathfrak{Y}$. Clearly, $\Phi = (\varphi_0,\varphi_1,\ldots,\varphi_k,\ldots)$ cannot be onto, which contradicts the fact that it is a Lie-Bäcklund isomorphism. The same lines apply for Φ_k and Ψ_k for all k.

Let us now go back to the proof of Theorem 5.1:

Proof. Let Φ be a Lie-Bäcklund isomorphism between the two systems $(\mathfrak{X},\overline{f})$ and $(\mathfrak{Y},\overline{g})$. According to Lemma 5.1, the mapping $\Phi_k : (y,\overline{v}) \mapsto (x,u,\dot{u},\ldots,u^{(k)}) = \Phi_k(y,\overline{v})$ is onto. Since Φ_k depends only on a finite number of variables, Φ_k is a surjection between finite dimensional spaces and the dimension of its range must be smaller than or equal to the dimension of the source. Remark that if φ_1 depends only on the variables $(y,v,\dot{v},\ldots,v^{(\alpha+1)})$, then, since $\dot{u} = \varphi_2(y,\overline{v}) = \frac{d}{dt}\varphi_1(y,v,\dot{v},\ldots,v^{(\alpha+1)})$, φ_2 depends only on $(y,v,\dot{v},\ldots,v^{(\alpha+2)})$. Similarly, φ_k at most depends on $(y,v,\dot{v},\ldots,v^{(\alpha+k)})$. Let us denote by q the dimension of y, n the dimension of x, r the dimension of v and m the dimension of u. We have, for all k, the inequalities:

$$q + (\alpha+k+1)r \geq \operatorname{rank}\left(\frac{\partial \Phi_k}{\partial(y,\overline{v})}\right) = n + (k+1)m$$

or

$$(k+1)(r-m) \geq n - q - \alpha r \tag{5.27}$$

which implies that $r-m \geq 0$ since, otherwise, for k sufficiently large, the left-hand side of (5.27) would become arbitrarily negative and would contradict the inequality since $n-q-\alpha r$ is a constant independent of k.

Following the same lines for Ψ_k, where Ψ is the inverse of Φ, we get $m-r \geq 0$, and thus $m = r$, which achieves the proof.

In the linear case, the previous result reads:

Proposition 5.4. *Two linear controllable systems are L-B equivalent if and only if they have the same number of independent inputs.*

Proof. The condition is necessary by Theorem 5.1. Let us show that it is sufficient. Since each system is L-B equivalent to its canonical form (see section 4.1.2), it suffices to show that two arbitrary canonical forms having the same number of independent inputs are L-B equivalent. In fact, since their associated Cartan fields are both the trivial Cartan field with m inputs, they are obviously L-B equivalent. *Q.E.D.*

Lie-Bäcklund isomorphisms also enjoy an important property concerning equilibrium points:

Theorem 5.2. *The L-B equivalence preserves equilibrium points.*

Proof. Let us assume that the systems $\dot{x} = f(x, u)$ and $\dot{y} = g(y, v)$ are L-B equivalent. Let (x_0, u_0) be an equilibrium point, *i.e.* $f(x_0, u_0) = 0$, $u_0^{(j)} = 0$ for all $j \geq 1$. Since $y_0 = \psi_0(x_0, u_0, \dot{u}_0, \ldots, u_0^{(s)}) = \psi_0(x_0, u_0, 0, \ldots, 0)$ and $v_0^{(j)} = \psi_{j+1}(x_0, u_0, \dot{u}_0, \ldots, u_0^{(s+j+1)}) = \psi_{j+1}(x_0, u_0, 0, \ldots, 0)$ for all j, we have $\dot{y}_0 = \frac{\partial \psi_0}{\partial x}\dot{x}_0 + \frac{\partial \psi_0}{\partial u}\dot{u}_0 + \ldots + \frac{\partial \psi_0}{\partial u^{(s)}} u_0^{(s+1)} = 0$ and $v_0^{(j+1)} = \frac{\partial \psi_{j+1}}{\partial x}\dot{x}_0 + \frac{\partial \psi_{j+1}}{\partial u}\dot{u}_0 + \ldots + \frac{\partial \psi_{j+1}}{\partial u^{(s+j+1)}} u_0^{(s+j+2)} = 0$ for all $j \geq 1$, which proves that (y_0, v_0) is an equilibrium point of $\dot{y} = g(y, v)$.

We proceed in the same way to prove that the image of an equilibrium point of $\dot{y} = g(y, v)$ is an equilibrium point of $\dot{x} = f(x, u)$.

5.3.6 Endogenous Dynamic Feedback

We now generalize the previously introduced notion of integrator to a class of feedback called endogenous dynamic feedback.

Let us consider a system $(\mathfrak{X}, \overline{f})$ whose representation in finite dimension is $\dot{x} = f(x, u)$. A *dynamic feedback* is the data of a differential equation $\dot{z} = \beta(x, z, v)$ and a feedback $u = \alpha(x, z, v)$.

The closed-loop system is thus

$$\begin{aligned} \dot{x} &= f(x, \alpha(x, z, v)) \\ \dot{z} &= \beta(x, z, v). \end{aligned}$$

Such a dynamic feedback may have unexpected properties such as the uncontrollability of the closed-loop system or its non invertibility, *i.e.* that it is not possible to recover the original system by applying another dynamic feedback.

Let us illustrate this by simple examples:

Example 5.2. Consider the system $\dot{x} = u$ with the dynamic feedback $\dot{z} = v$, $u = v$. The closed-loop system is $\dot{x} = v, \ \dot{z} = v$, which implies that $\dot{x} - \dot{z} = 0$ or $x(t) - z(t) = x_0 - z_0$ for all t. The closed-loop system is thus no more controllable.

Example 5.3. Consider now the 2-dimensional linear system: $\dot{x}_1 = x_2,\ \dot{x}_2 = u$ and let us apply the same dynamical feedback as in the previous example: $\dot{z} = v$, $u = v$. The closed-loop system reads $\dot{x}_1 = x_2,\ \dot{x}_2 = \dot{z}$, or $\dot{x}_1 = z + c$ where c is a constant, since $\dot{x}_2 = \dot{z}$ implies $x_2 = z + c$. Thus, the knowledge of $v = \dot{z}$ is not sufficient to deduce z, because of the missing initial condition z_0. Thus x_1 and x_2, and a fortiori u, cannot be recovered so that the closed-loop system contains less information than the original system $\ddot{x}_1 = u$.

This loss of information comes from the fact that $u = \dot{z}$, so that z cannot be expressed as a function of $(x, \overline{u}) = (x, u, \dot{u}, \dots,)$ only, because the initial condition z_0 cannot be deduced from $(x, \overline{u})$. We say that such a dynamic feedback is exogenous since it depends on the extra variable z_0 that didn't appear in the original system, which prevents from making the original system and the closed-loop one L-B equivalent.

On the contrary, if we restrict to endogenous dynamic feedback, *i.e.* such that the closed-loop variables can be expressed as functions of the original system variables, these pathologies disappear, as will be shown later. We therefore introduce the following definition:

Definition 5.5. Consider the system $(\mathfrak{X}, \overline{f})$. We call *endogenous dynamic feedback* a dynamic feedback of the form

$$\dot{z} = \beta(x, z, v), \quad u = \alpha(x, z, v) \tag{5.28}$$

such that the closed-loop system is L-B equivalent to the system $(\mathfrak{X}, \overline{f})$.

Otherwise stated, there exists a Lie-Bäcklund isomorphism Φ, of inverse Ψ, such that $(x, \overline{u}) = \Phi(x, z, \overline{v})$ and $(x, z, \overline{v}) = \Psi(x, \overline{u})$, which implies that z, v, $\dot{v}$, ..., can be expressed as functions of x, u and a finite number of successive derivatives of u.

It is therefore obvious from this definition that an integrator of arbitrary order is an endogenous dynamic feedback.

The next result clarifies the links between this type of feedback and L-B equivalence.

We consider two systems $(\mathfrak{X}, \overline{f})$ and $(\mathfrak{Y}, \overline{g})$ defined by $\dot{x} = f(x, u)$ and $\dot{y} = g(x, v)$ respectively.

Theorem 5.3 (Martin [1992]). *Assume that the systems $(\mathfrak{X}, \overline{f})$ and $(\mathfrak{Y}, \overline{g})$ are L-B equivalent. Then there exists an endogenous dynamic feedback (5.28) such that the closed-loop system*

$$\begin{aligned} \dot{x} &= f(x, \alpha(x, z, v)) \\ \dot{z} &= \beta(x, z, v) \end{aligned} \tag{5.29}$$

is diffeomorphic to the prolonged system

$$\begin{aligned} \dot{y} &= g(y, w) \\ w^{(r+1)} &= v \end{aligned} \tag{5.30}$$

with r a large enough integer.

Proof. We note $\overline{w}^r = (w, w^{(1)}, \ldots, w^{(r)})$. Let $\tilde{y} = (y, \overline{w}^r) = (y, w, w^{(1)}, \ldots, w^{(r)})$ and $v = w^{(r+1)}$. According to the L-B equivalence, we have $(x, \overline{u}) = \Phi(y, \overline{w}) = (\varphi_0(y, \overline{w}), \varphi_1(y, \overline{w}), \ldots)$.

For r sufficiently large, φ_0 depends only on $\tilde{y}$ and φ_1 on $(\tilde{y}, v)$, *i.e.* the Lie-Bäcklund isomorphism Φ is of the form

$$\Phi(\tilde{y}, v, v^{(1)}, \ldots) = (\varphi_0(\tilde{y}), \varphi_1(\tilde{y}, v), \varphi_2(\tilde{y}, \overline{v}^1), \ldots),$$

still with the notation $\overline{v}^k = (v, v^{(1)}, \ldots, v^{(k)})$ for all $k \geq 1$, and the system $\dot{x} = f(x, u)$ reads

$$\dot{x} = f(\varphi_0(\tilde{y}), \varphi_1(\tilde{y}, v)) = g(\tilde{y}) \frac{\partial \varphi_0}{\partial y} + \sum_{j=0}^{r-1} w^{(j+1)} \frac{\partial \varphi_0}{\partial w^{(j)}} + v \frac{\partial \varphi_0}{\partial w^{(r)}}. \tag{5.31}$$

Denote by

$$\tilde{g} = g \frac{\partial}{\partial y} + \sum_{j=0}^{r-1} w^{(j+1)} \frac{\partial}{\partial w^{(j)}} + v \frac{\partial}{\partial w^{(r)}}$$

the prolonged vector field corresponding to (5.30). The identity (5.31) reads

$$f(\varphi_0(\tilde{y}), \varphi_1(\tilde{y}, v)) = \tilde{g}(\tilde{y}, v) \frac{\partial \varphi_0}{\partial \tilde{y}}. \tag{5.32}$$

Applying Lemma 5.1, φ_0 is onto. Thus, there exists a surjective mapping γ such that $\tilde{y} \mapsto \begin{pmatrix} \varphi_0(\tilde{y}) \\ \gamma(\tilde{y}) \end{pmatrix} = K(\tilde{y})$ is a local diffeomorphism of $Y \times \mathbb{R}^{m(r+1)}$.

Let us set $z = \gamma(\tilde{y})$. We have $\dot{z} = \gamma_* \tilde{g}(\tilde{y}, v)$ by definition of the image of the vector field $\tilde{g}$ by γ. Since $x = \varphi_0(\tilde{y})$, we have $(x, z) = K(\tilde{y})$. Applying the dynamic feedback

$$\begin{aligned} u &= \varphi_1(K^{-1}(x, z), v) \\ \dot{z} &= \gamma_* \tilde{g}(K^{-1}(x, z), v), \end{aligned}$$

we thus obtain the closed-loop system

$$\begin{pmatrix} \dot{x} \\ \dot{z} \end{pmatrix} = \tilde{f}(x, z, v) = \begin{pmatrix} f(x, \varphi_1(K^{-1}(x, z), v)) \\ \gamma_* \tilde{g}(K^{-1}(x, z), v) \end{pmatrix} \tag{5.33}$$

and, using (5.32),

$$\tilde{f}(K(\tilde{y}), v) = \begin{pmatrix} f(\varphi_0(\tilde{y}),\ \varphi_1(\tilde{y}, v)) \\ \gamma_* \tilde{g}(\tilde{y}, v) \end{pmatrix} = \begin{pmatrix} \frac{\partial \varphi_0}{\partial \tilde{y}}(\tilde{y}) \\ \frac{\partial \gamma}{\partial \tilde{y}}(\tilde{y}) \end{pmatrix} .\tilde{g}(\tilde{y}, v) = \frac{\partial K}{\partial \tilde{y}}(\tilde{y}).\tilde{g}(\tilde{y}, v)$$

which implies that (5.33) and (5.30) are diffeomorphic.

Finally, by the inverse diffeomorphism, since $y = \psi_0(x, \overline{u}^s)$ and $v = \psi_1(x, \overline{u}^{s+1})$ for at least one s, we get $z = \gamma(\tilde{y}) = \gamma(\tilde{\psi}(x, \overline{u}^{s+r+1}))$, where $\tilde{\psi}$ is the mapping $(\psi_0, \psi_1, \dots, \psi_{r+1})$, which proves that (z, v) can be expressed in function of $(x, \overline{u})$ and thus that $\dot{x} = f(x, u)$ is Lie-Bäcklund equivalent to the closed-loop system (5.33), which proves that the constructed dynamic feedback is endogenous.

Chapter 6
Differentially Flat Systems

6.1 Flat System, Flat Output

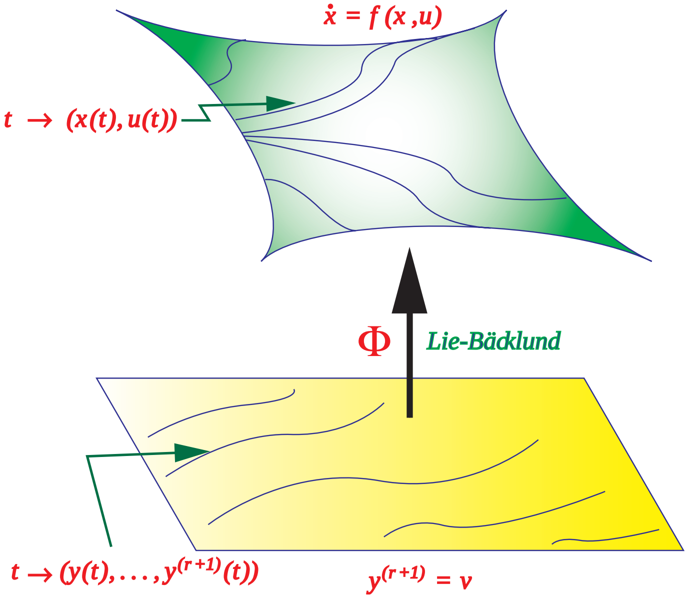

Fig. 6.1 Graphic representation of the flatness property: L-B equivalence between the trajectories of the trivial system (below) an those of the nonlinear system (above).

Definition 6.1. We say that the system $(X \times U \times \mathbb{R}^m_\infty, \overline{f})$ (resp. $(X \times \mathbb{R}^n_\infty, \tau_X, F)$), with m inputs, is *differentially flat*, or, shortly, *flat*, if and only if

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_6, © Springer-Verlag Berlin Heidelberg 2009

it is L-B equivalent to the trivial system $(\mathbb{R}^m_\infty, \tau_m)$ (resp. $(\mathbb{R}^m_\infty, \tau_m, 0)$), where τ_m is the trivial Cartan field of $\mathbb{R}^m_\infty$ with coordinates $(y, \dot{y}, \ddot{y}, \ldots)$:

$$\tau_m = \sum_{j \geq 0} \sum_{i=1}^{m} y_i^{(j+1)} \frac{\partial}{\partial y_i^{(j)}}. \tag{6.1}$$

The vector $y = (y_1, \ldots, y_m)$ is called a *flat output.*

Let us interpret this definition in the two previously studied cases:

Explicit Case. The explicit system $\dot{x} = f(x, u)$ with m inputs is flat if and only if there exists a flat output y of dimension m, two integers r and s and mappings ψ from $X \times (\mathbb{R}^m)^{s+1}$ to $\mathbb{R}^m$, of rank m in a suitably chosen open subset, and (φ_0, φ_1) from $\mathbb{R}^{(m+2)r}$ to $\mathbb{R}^n \times \mathbb{R}^m$, of rank $n+m$ in a suitable open subset, such that $y = (y_1, \ldots, y_m) = \psi(x, u, \dot{u}, \ldots, u^{(s)})$ implies that $x = \varphi_0(y, \dot{y}, \ldots, y^{(r)})$, $u = \varphi_1(y, \dot{y}, \ldots, y^{(r+1)})$, the differential equation $\frac{d\varphi_0}{dt} = f(\varphi_0, \varphi_1)$ being identically satisfied.
It is easily verified that if $x = \varphi_0(\overline{y})$ and $u = \varphi_1(\overline{y})$ with $\frac{d\varphi_0}{dt} = f(\varphi_0, \varphi_1)$ we have $(x, \overline{u}) = \Phi(\overline{y}) = (\varphi_0(\overline{y}), \varphi_1(\overline{y}), \frac{d\varphi_1}{dt}(\overline{y}), \frac{d^2\varphi_1}{dt^2}(\overline{y}), \ldots)$, and $\frac{dy^{(j)}}{dt} = \frac{d^j y}{dt^j} = y^{(j+1)}$.

Implicit Case. The implicit system $F(x, \dot{x}) = 0$ with rank $\left(\frac{\partial F}{\partial \dot{x}}\right) = n - m$, is flat if and only if there exists an integer s and a mapping ψ from $X \times (\mathbb{R}^n)^s$ to $\mathbb{R}^m$, of rank m in a suitably chosen open subset, such that $y = (y_1, \ldots, y_m) = \psi(x, \dot{x}, \ldots, x^{(s)})$ implies $x = \varphi_0(y, \dot{y}, \ldots, y^{(r)})$ for a suitable integer r, the implicit equation $F(\varphi_0, \frac{d\varphi_0}{dt}) = 0$ being identically satisfied, with $\frac{d\varphi_0}{dt} = L_{\tau_m}\varphi_0$.

In all cases, one can express all the system variables in function of the flat output and a finite number of its successive derivatives.

A flat output is indeed non unique: if (y_1, y_2) is a flat output of a system with 2 inputs, then the output $(z_1, z_2) = (y_1 + y_2^{(k)}, y_2)$ for k arbitrary integer, is also a flat output since the so defined mapping $y \mapsto z$ is invertible ($y_1 = z_1 - z_2^{(k)}, y_2 = z_2$) and preserves the trivial Cartan field of $\mathbb{R}^2_\infty$. By the transitivity of the L-B equivalence, the assertion is proven.

Remark 6.1. Note that flatness may be traced back to Hilbert and Cartan Hilbert [1912], Cartan [1914]. In fact, using the definition of the implicit case, flatness may be seen as a generalization in the framework of manifolds of jets of infinite order of the *uniformization of analytic functions* of Hilbert's 22nd problem Hilbert [1901], solved by Poincaré Poincaré [1907] in 1907 (see Bers [1976] for a modern presentation of this subject and recent extensions and results). This problem consists, roughly speaking, given a set of complex polynomial equations in one complex variable, in finding an open dense subset D of the complex plane $\mathbb{C}$ and a holomorphic function s from D to $\mathbb{C}$ such that

s is surjective and $s(p)$ identically satisfies the given equations for all values of the "parameter" $p \in D$. In our setting, $\mathbb{C}$ is replaced by a (real) manifold of jets of infinite order, a flat output $y_1, \ldots, y_m$ plays the role of the parameter p and s is the associated Lie-Bäcklund isomorphism $s = (\varphi_0, \varphi_1, \dot{\varphi}_1, \ddot{\varphi}_1, \ldots)$ with φ_0 and φ_1 defined above.

The following result is a straightforward consequence of Theorem 5.1.

Proposition 6.1. *Given a flat system, the number of components of a flat output is equal to the number of independent inputs.*

6.2 Examples

6.2.1 Mass-Spring System

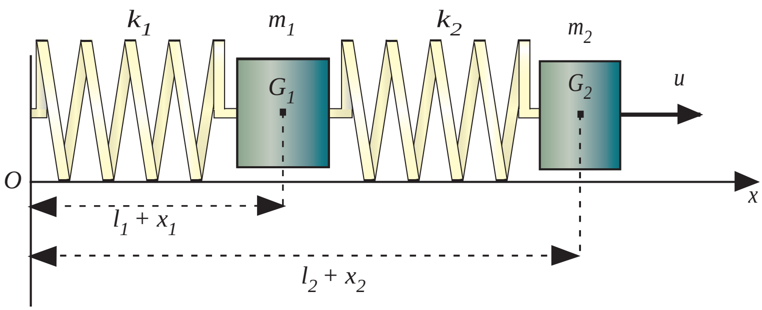

Fig. 6.2 A mass-spring system.

Consider a system made of two bodies of respective mass m_1 and m_2, related by two springs of respective stiffness k_1 and k_2, moving along the axis Ox (see Figure 6.2). The respective position of the gravity centres G_1 and G_2 of the two bodies are denoted by $l_1 + x_1$ and $l_2 + x_2$, where l_1 and l_2 are the equilibrium positions of G_1 and G_2. We assume that the two bodies are submitted to viscous frictions $\gamma_1(\dot{x}_1)$ and $\gamma_2(\dot{x}_2)$ respectively. The functions γ_1 and γ_2 are assumed non negative, twice continuously differentiable, and $\gamma_1(0) = \gamma_2(0) = 0$. Finally, a force u is applied to G_2.

The second principle reads:

$$\begin{cases} m_1\ddot{x}_1 + k_1 x_1 + \gamma_1(\dot{x}_1) = k_2(x_2 - x_1) \\ m_2\ddot{x}_2 + k_2(x_2 - x_1) + \gamma_2(\dot{x}_2) = u \ . \end{cases} \tag{6.2}$$

To represent (6.2) in implicit form, it suffices to eliminate the last equation, *i.e.*

$$m_1\ddot{x}_1 + k_1x_1 + \gamma_1(\dot{x}_1) - k_2(x_2 - x_1) = 0. \tag{6.3}$$

Let us show that we can express x_2 and u in function of x_1. From the first equation of (6.2), we get:

$$x_2 = \frac{1}{k_2}\left(m_1\ddot{x}_1 + (k_1 + k_2)\,x_1 + \gamma_1(\dot{x}_1)\right) \tag{6.4}$$

and, differentiating, with the notation γ_1' for the first derivative of γ_1,

$$\dot{x}_2 = \frac{1}{k_2}\left(m_1x_1^{(3)} + (k_1 + k_2))\dot{x}_1 + \gamma_1'(\dot{x}_1)\ddot{x}_1\right), \tag{6.5}$$

then, using the second equation of (6.2) and the above expressions of x_2 and $\dot{x}_2$

$$\begin{aligned} u = \frac{m_1m_2}{k_2}x_1^{(4)} + \left(\frac{m_2k_1}{k_2} + m_2 + m_1\right)\ddot{x}_1 + k_1x_1 + \gamma_1(\dot{x}_1) \\ + \frac{m_2}{k_2}\left(\gamma_1''(\dot{x}_1)(\ddot{x}_1)^2 + \gamma_1'(\dot{x}_1)x_1^{(3)}\right) \\ + \gamma_2\left(\frac{m_1}{k_2}x_1^{(3)} + \frac{1}{k_2}((k_1 + k_2)\dot{x}_1 + \gamma_1'(\dot{x}_1)\ddot{x}_1)\right) \end{aligned} \tag{6.6}$$

which proves that x_2 and u can be expressed as functions of x_1 and a finite number of its derivatives, and thus that the system (6.2) is flat with x_1 as flat output.

The obtained transformation from the flat output $y = x_1$ boils down to express the system in its controllable canonical form (see section 4.1.2), which reads here $y^{(4)} = v$.

6.2.2 Robot Control

We consider a robot arm with n degrees of freedom and n actuators. Its dynamical equations are of the form:

$$\Gamma_0(q)\ddot{q} + \Gamma_1(q,\dot{q}) = Q(q,\dot{q})u \tag{6.7}$$

where q is the vector of generalized coordinates (angular coordinates for a rigid robot with rotoid joints), $\dim q = \dim u = n$, $\mathrm{rank}\,(Q) = n$, $\Gamma_0(q)$ being the inertia matrix, assumed everywhere invertible, $\Gamma_1(q,\dot{q})$ the vector of centrifugal and Coriolis forces and the matrix $Q(q,\dot{q})$ characterizing the actuators (localization, direct or indirect drives, presence of reduction units, ...).

We set $x_1 = q$, $x_2 = \dot{q}$ and the system reads:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -(\Gamma_0(x_1))^{-1} \left(\Gamma_1(x_1, x_2) + Q(x_1, x_2)u\right). \end{cases} \tag{6.8}$$

Since $\operatorname{rank}(Q) = n$, we immediately verify that all the equations of (6.7) can be eliminated, so that the implicit system reduces to $0 = 0$ (trivial).

As previously, the vector x_1 is a flat output of (6.8): $x_2 = \dot{x}_1$ and

$$u = Q^{-1}(x_1, \dot{x}_1) \left(\Gamma_0(x_1)\ddot{x}_1 + \Gamma_1(x_1, \dot{x}_1)\right)$$

formula called *computed torque*.

6.2.3 Pendulum

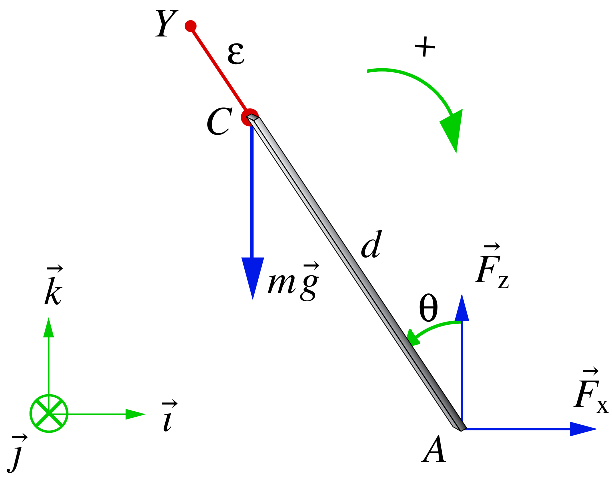

Fig. 6.3 The inverted pendulum in the vertical plane

Consider an inverted pendulum, in the plane of coordinates (x, z), of mass m, controlled by the exterior force $\overrightarrow{F}$ applied to the point A located at a distance d of the centre of mass C of the pendulum (see Figure 6.3). We consider the inertial frame $(\overrightarrow{\imath}, \overrightarrow{\jmath}, \overrightarrow{k})$. The force $\overrightarrow{F}$, in this frame, reads: $\overrightarrow{F} = F_x \overrightarrow{\imath} + F_z \overrightarrow{k}$. The resulting force applied to the centre of mass is the sum of $\overrightarrow{F}$ and of the weight $-mg\overrightarrow{k}$: $F_x \overrightarrow{\imath} + (F_z - mg)\overrightarrow{k}$. Let us denote by (x_C, O, z_C) the coordinates of the point C. We have

$$m\ddot{x}_C = F_x \ , \quad m\ddot{z}_C = F_z - mg \ .$$

We denote by θ the angle between the pendulum and the vertical (parallel to $\overrightarrow{k}$) and J the inertia of the pendulum. The vector $\overrightarrow{CA}$ reads $\overrightarrow{CA} = -d(\sin\theta \ \overrightarrow{\imath} + \cos\theta \ \overrightarrow{k})$ and the angular moment around C is given by

$$J\ddot{\theta}\overrightarrow{\jmath} = \overrightarrow{CA} \times \overrightarrow{F} = d(F_z \sin\theta - F_x \cos\theta)\overrightarrow{\jmath} \ .$$

We thus have

$$\begin{cases} m\ddot{x}_C = F_x \\ m\ddot{z}_C = F_z - mg \\ \dfrac{J}{d}\ddot{\theta} = F_z \sin\theta - F_x \cos\theta \ . \end{cases}$$

Setting

$$x = \frac{x_C}{g}, \ z = \frac{z_C}{g}, \ u_1 = \frac{F_x}{mg}, \ u_2 = \frac{F_z}{mg} - 1, \ \varepsilon = \frac{J}{mgd}$$

we obtain the pendulum dynamical equations:

$$\begin{cases} \ddot{x} = u_1 \\ \ddot{z} = u_2 \\ \varepsilon\ddot{\theta} = -u_1 \cos\theta + (u_2 + 1)\sin\theta \end{cases} \tag{6.9}$$

where u_1 and u_2 are the two components of the input vector u.

An implicit representation of (6.9) is given by:

$$\varepsilon\ddot{\theta} = -\ddot{x}\cos\theta + (\ddot{z} + 1)\sin\theta. \tag{6.10}$$

We now consider another implicit differential system with 4 unknown functions and 2 equations:

$$\begin{cases} (\xi - x)^2 + (\zeta - z)^2 = \varepsilon^2 \\ \ddot{\xi}(\zeta - z) - (\xi - x)(\ddot{\zeta} + 1) = 0 \ . \end{cases} \tag{6.11}$$

This system admits the following geometric interpretation: let (ξ, ζ) be the position of the point Y located at the distance ε of the point C, *i.e.* $(\xi - x)^2 + (\zeta - z)^2 = \varepsilon^2$, whose acceleration minus the constant vector $-\overrightarrow{k}$, corresponding to the normalized acceleration of gravity, is colinear to the vector $\overrightarrow{YC}$:

$$\frac{\ddot{\xi}}{\ddot{\zeta} + 1} = \frac{\xi - x}{\zeta - z}. \tag{6.12}$$

The point Y, satisfying $(\xi - x)^2 + (\zeta - z)^2 = \varepsilon^2$, or, in polar coordinates:

$$\xi = x + \varepsilon \sin\theta \ , \quad \zeta = z + \varepsilon\cos\theta \ , \tag{6.13}$$

is called the *oscillation centre* or the *Huygens oscillation centre* Fliess et al. [1999], Whittaker [1937].

Let us show that (6.11) is equivalent (in the usual sense) to (6.9), in other words that (6.11) and (6.10) are equivalent.

Starting from (6.11) and differentiating two times (6.13), we get

$$\ddot{\xi} = \ddot{x} + \varepsilon\ddot{\theta}\cos\theta - \varepsilon\dot{\theta}^2\sin\theta\ , \quad \ddot{\zeta} = \ddot{z} - \varepsilon\ddot{\theta}\sin\theta - \varepsilon\dot{\theta}^2\cos\theta. \tag{6.14}$$

Taking into account the fact that $\xi - x = \varepsilon\sin\theta$ and $\zeta - z = \varepsilon\cos\theta$, we have

$$\ddot{\xi}(\zeta-z)-(\ddot{\zeta}+1)(\xi-x) = \varepsilon(\ddot{\xi}\cos\theta-(\ddot{\zeta}+1)\sin\theta) = \varepsilon(\varepsilon\ddot{\theta}+\ddot{x}\cos\theta-(\ddot{z}+1)\sin\theta).$$

Thus, if (6.11) holds true, (6.10) also, and conversely, which proves the announced equivalence.

In (6.11), the control variables are x and z, which implies, by (6.13), that $\ddot{x} = u_1$ and $\ddot{z} = u_2$, where (u_1, u_2) is the input of the pendulum model (6.9). Thus, (6.11) may be interpreted as a reduced system where the two double integrators describing the acceleration of the centre of mass of the inverted pendulum have been removed.

Proposition 6.2. *The system (6.9) is flat with the coordinates* (ξ, ζ) *of the oscillation centre as flat output.*

Proof. From (6.13) and (6.12), we get

$$\tan\theta = \frac{\xi - x}{\zeta - z} = \frac{\ddot{\xi}}{\ddot{\zeta}+1}$$

or

$$\theta = \arctan\left(\frac{\ddot{\xi}}{\ddot{\zeta}+1}\right)$$

i.e.

$$\sin\theta = \frac{\ddot{\xi}}{\sqrt{(\ddot{\xi})^2+(\ddot{\zeta}+1)^2}}\ , \quad \cos\theta = \frac{\ddot{\zeta}+1}{\sqrt{(\ddot{\xi})^2+(\ddot{\zeta}+1)^2}}.$$

Moreover

$$x = \xi - \varepsilon\frac{\ddot{\xi}}{\sqrt{(\ddot{\xi})^2+(\ddot{\zeta}+1)^2}}\ , \quad z = \zeta - \varepsilon\frac{\ddot{\zeta}+1}{\sqrt{(\ddot{\xi})^2+(\ddot{\zeta}+1)^2}} \tag{6.15}$$

and the force can be easily deduced by twice differentiating the expressions of x and z:

$$
\begin{aligned}
u_1 &= \frac{d^2}{dt^2}\left(\xi - \varepsilon \frac{\ddot{\xi}}{\sqrt{(\ddot{\xi})^2 + (\ddot{\zeta}+1)^2}}\right) \\
u_2 &= \frac{d^2}{dt^2}\left(\zeta - \varepsilon \frac{\ddot{\zeta}+1}{\sqrt{(\ddot{\xi})^2 + (\ddot{\zeta}+1)^2}}\right),
\end{aligned}
\tag{6.16}
$$

expression that contains the derivatives of ξ and ζ up to the order four.

It results that all the system variables of (6.9) can be expressed, in an invertible way, as functions of ξ, ζ and a finite number of their derivatives with respect to time $\dot{\xi}, \ddot{\xi}, \ldots, \dot{\zeta}, \ddot{\zeta}, \ldots$. The system flatness is immediately deduced with (ξ, ζ) as flat output.

6.2.4 Non Holonomic Vehicle

We go back to the car of the example 4.4. Recall that the car's model is given by

$$
\begin{aligned}
\dot{x} &= u \cos\theta \\
\dot{y} &= u \sin\theta \\
\dot{\theta} &= \frac{u}{l} \tan\varphi \ .
\end{aligned}
\tag{6.17}
$$

In implicit form, this system becomes:

$$\dot{x} \sin\theta - \dot{y} \cos\theta = 0. \tag{6.18}$$

Let us show that this system is flat with (x, y) as flat output. The two first equations of (6.17) give

$$\tan\theta = \frac{\dot{y}}{\dot{x}} \ , \quad u^2 = \dot{x}^2 + \dot{y}^2. \tag{6.19}$$

Differentiating the expression of $\tan\theta$, we get $\dot{\theta}(1 + \tan^2\theta) = \dfrac{\ddot{y}\dot{x} - \dot{y}\ddot{x}}{\dot{x}^2}$ from which we deduce

$$\dot{\theta} = \frac{\ddot{y}\dot{x} - \dot{y}\ddot{x}}{\dot{x}^2 + \dot{y}^2}$$

and, by the third equation of (6.17),

$$\tan\varphi = \frac{l\dot{\theta}}{v} = l \frac{\ddot{y}\dot{x} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}}. \tag{6.20}$$

All the system variables x, y, θ, u, φ are thus expressed as functions of $x, \dot{x}, \ddot{x}, y, \dot{y}, \ddot{y}$, which proves our assertion.

6.2.5 Vehicle with Trailers

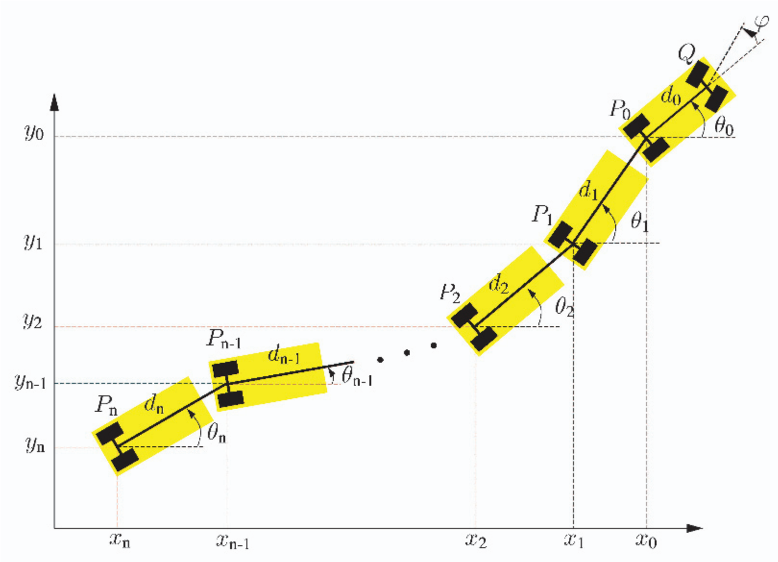

Fig. 6.4 The n-trailer vehicle.

We consider a vehicle with several trailers as described in Figure 6.4, each trailer being attached to the previous one at the middle of the rear axle of this trailer. The vehicle and the trailers are assumed to roll without slipping.

The kinematic equations, that include the rolling without slipping conditions, or non holonomic constraints, are given by

$$
\begin{aligned}
\dot{x}_0 &= u \cos\theta_0 \\
\dot{y}_0 &= u \sin\theta_0 \\
\dot{\theta}_0 &= \frac{u}{d_0} \tan\varphi \\
\dot{\theta}_i &= \frac{u}{d_i}\left(\prod_{j=1}^{i-1} \cos(\theta_{j-1}-\theta_j)\right) \sin(\theta_{i-1}-\theta_i) \ , \quad i = 1, \ldots, n \ .
\end{aligned}
\tag{6.21}
$$

with the convention, for $i = 1$, $\prod_{j=1}^{0} \cos(\theta_{j-1}-\theta_j) \stackrel{\text{def}}{=} 1$.

Recall that u is the modulus of the velocity of the front vehicle, and is one of the controls, the second one being the angle φ of the front wheels.

In implicit form, this system becomes:

$$\begin{aligned} &\dot{x}_0 \sin\theta_0 - \dot{y}_0 \cos\theta_0 = 0 \\ &d_i\dot{\theta}_i - (\dot{x}_0 \cos\theta_0 + \dot{y}_0 \sin\theta_0)\left(\prod_{j=1}^{i-1}\cos(\theta_{j-1}-\theta_j)\right)\sin(\theta_{i-1}-\theta_i) = 0, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1,\ldots,n\ . \end{aligned} \tag{6.22}$$

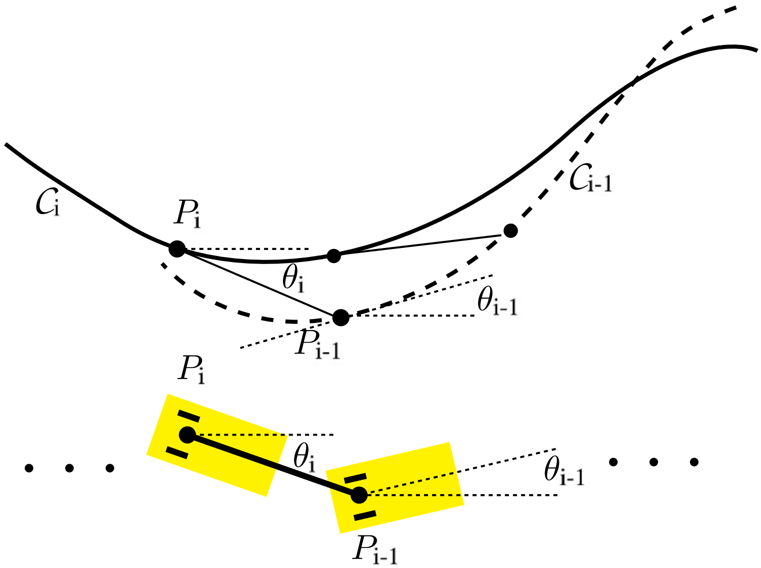

Fig. 6.5 Geometric interpretation of the rolling without slipping conditions.

Let us show that the coordinates (x_n, y_n) of the middle of the rear axle of the last trailer constitutes a flat output. One easily verifies, by induction, that the modulus of the velocity u_i of the ith trailer satisfies

$$u_i = u\prod_{j=1}^{i}\cos(\theta_{j-1}-\theta_j)\ , \quad i = 1,\ldots,n$$

and that

$$\dot{x}_i = u_i\cos\theta_i\ , \quad \dot{y}_i = u_i\sin\theta_i$$

thus

$$\tan\theta_n = \frac{\dot{y}_n}{\dot{x}_n}$$

or $\sin\theta_n = \dfrac{\dot{y}_n}{(\dot{x}_n^2+\dot{y}_n^2)^{\frac{1}{2}}}$ and $\cos\theta_n = \dfrac{\dot{x}_n}{(\dot{x}_n^2+\dot{y}_n^2)^{\frac{1}{2}}}$. Since, moreover, we have $x_{n-1} = x_n + d_n\cos\theta_n$ and $y_{n-1} = y_n + d_n\sin\theta_n$, on the one hand, and

$\tan\theta_{n-1} = \dfrac{\dot{y}_{n-1}}{\dot{x}_{n-1}}$ on the other hand, we immediately see that x_{n-1} and y_{n-1} are expressed in function of x_n, y_n and their first order derivatives, and that θ_{n-1} is expressed in function of x_n, y_n and derivatives up to second order.

By induction, we show that all the x_i and y_i can be expressed in function of x_n, y_n and derivatives of to order $n-i$, and that the angles θ_i can also be expressed in function of x_n, y_n and derivatives up to order $n-i+1$, $i=0,\ldots n$. Finally, we deduce that u is a function of x_n, y_n and derivatives up to order $n+1$ and the same for φ up to order $n+2$, which achieves the proof of our assertion.

This property may be directly verified on Figure 6.5.

6.3 Flatness and Controllability

Theorem 6.1. *A flat system is locally reachable.*

Proof. Consider the flat system $(\mathfrak{X},\overline{f})$. corresponding to $\dot{x} = f(x,u)$. According to Proposition 4.3, it is locally reachable if the extended system $\dot{X} = F(X,\dot{u})$, with $X = (x,u)$ and $F(X,\dot{u}) = \begin{pmatrix} f(x,u) \\ \dot{u} \end{pmatrix}$, is also locally reachable. Consider then the vector fields $F_0(X,u) = f(x,u)\frac{\partial}{\partial x}$ and $F_i(X,u) = \frac{\partial}{\partial u_i}$, $i=1,\ldots,m$. They generate a distribution $\mathcal{D}^*$ according to the algorithm 4.15. Let us assume that $\dot{x} = f(x,u)$ is not reachable, and thus the same for its extended system. It results that the dimension of $\mathcal{D}^*$ is, at every point of a suitable open subset, equal to $N_0 < n+m$. According to Theorem 4.7, the extended system may be represented by $(x,u) = \chi(\xi,\zeta)$ with $\dim\xi = N_0$, $\dim\zeta = n+m-N_0$ and

$$\begin{aligned}\dot{\xi} &= \gamma_1(\xi,\zeta) + \sum_{i=1}^{m} \dot{u}_i\eta_i(\xi,\zeta)\\ \dot{\zeta} &= \gamma_2(\zeta)\end{aligned}$$

where ζ corresponds to the non controllable part.

Since the system is flat, one must have $\zeta = \sigma(y,\dot{y},\ldots,y^{(r)})$ for some integer r. Differentiating this last expression, we get:

$$\dot{\zeta} = \frac{\partial\sigma}{\partial y}\dot{y} + \ldots + \frac{\partial\sigma}{\partial y^{(r)}}y^{(r+1)} = \gamma_2(\zeta) = \gamma_2(\sigma(y,\dot{y},\ldots,y^{(r)})).$$

But $\gamma_2\circ\sigma$ doesn't depend on $y^{(r+1)}$, which proves that $\frac{\partial\sigma}{\partial y^{(r)}} = 0$, so that σ doesn't depend on $y^{(r)}$, or $\zeta = \sigma(y,\dot{y},\ldots,y^{(r-1)})$. Iterating this process, we conclude that σ neither depends on y nor on its derivatives, which contradicts the flatness property, and the system reachability follows.

Consequently:

Corollary 6.1. *A linear system is flat if and only if it is controllable.*

Proof. By Theorem 6.1, a flat system is locally reachable, and, since the dimension of $\mathcal{D}^*$, for a linear system, is equal to the Kalman controllability matrix rank (see the end of section 4.2.2), a flat linear system is controllable.

Conversely, a linear controllable system is equivalent by change of basis and feedback to its Brunovský canonical form, and thus L-B equivalent to a trivial system, which proves that it is flat.

We now study the properties of the tangent linear approximation of a flat system at an equilibrium point:

Theorem 6.2. *A system is flat at an equilibrium point if and only if it is L-B equivalent to its linear tangent system, the latter being controllable. Moreover, a system is flat at an equilibrium point if it is first order controllable at this point.*

Proof. Let the system $(\mathfrak{X}, \overline{f})$ be flat at the equilibrium point $(x_0, u_0, 0, \ldots)$. There exists Φ of class C^∞ invertible such that for all $\overline{y}$ in a neighborhood of $(y_0, 0, \ldots)$ of $\mathbb{R}^m_\infty$,there exists $(x, \overline{u})$ in a neighborhood of $(x_0, u_0, 0, \ldots)$ in $\mathfrak{X}$ with $\Phi(x, \overline{u}) = \overline{y}$ and $\Phi(x_0, u_0, 0, \ldots) = (y_0, 0, \ldots)$. We denote by $X^u_t(x_0, t_0)$ the flow of the vector field $f(\cdot, u)$ corresponding to the smooth function $t \mapsto u(t)$ on a suitable interval of time including the initial time t_0, that remains in a suitable neighborhood of x_0. We also denote by $D_\nu X^u_t(x_0, t_0) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left(X^{u+\varepsilon\nu}_t(x_0, t_0) - X^u_t(x_0, t_0))\right)$. Setting $\xi(t) = D_\nu X^{u_0}_t(x_0, t_0)$, by a standard variational computation, we have

$$\dot{\xi} = \frac{\partial f}{\partial x}(x_0, u_0)\xi + \frac{\partial f}{\partial u}(x_0, u_0)\nu$$

which means that ξ is an integral curve of the tangent linear system at the equilibrium point (x_0, u_0). But, according to flatness, we have $\Phi(X^{u_0+\varepsilon\nu}_t(x_0, t_0), u_0 + \varepsilon\nu, \ldots) = (y_\varepsilon(t), \dot{y}_\varepsilon(t), \ldots)$ and $\Phi(X^{u_0}_t(x_0, t_0), u_0, 0, \ldots) = \Phi(x_0, u_0, 0, \ldots) = (y_0, 0, \ldots)$. Clearly, since Φ is differentiable, the limit $\overline{\zeta}(t) = (\zeta(t), \dot{\zeta}(t), \ldots) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left(\Phi(X^{u_0+\varepsilon\nu}_t(x_0, t_0), u_0 + \varepsilon\nu, \ldots) - \Phi(X^{u_0}_t(x_0, t_0), u_0, 0, \ldots)\right) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(y_\varepsilon(t) - y_0, \dot{y}_\varepsilon(t), \ldots)$ exists. Therefore, $\zeta(t) = \frac{\partial\varphi_0}{\partial x}\xi(t) + \frac{\partial\varphi_0}{\partial u}\nu(t) + \frac{\partial\varphi_0}{\partial\dot{u}}\dot{\nu}(t) + \ldots$. The inverse relations giving $\xi(t)$ and $\nu(t)$ in function of $\overline{\zeta}$ may be obtained in the same way. Clearly the tangent manifold of coordinates $\overline{\zeta} = (\zeta, \dot{\zeta}, \ldots)$ is the trivial manifold, which proves that the system is L-B equivalent to its tangent linear system and that the latter is controllable, according to Corollary 6.1. Thus the system $(\mathfrak{X}, \overline{f})$ is first-order controllable.

The converse, namely L-B equivalence to its controllable linear tangent system implies flatness, is an obvious consequence of Corollary 6.1 since the controllability of the linear tangent system implies L-B equivalence to a trivial system and, by transitivity, that the original system is L-B equivalent to the trivial system, hence flat.

Remark 6.2. Fatness may be seen as a generalization to control systems of the hyperbolicity property of equilibrium points of (non controlled) systems, with L-B equivalence in place of topological equivalence, since, according to Hartman-Grobman's Theorem (Theorem 3.6), at such an equilibrium point, a system is topologically equivalent to is tangent linear system.

6.4 Flatness and Linearization

To precise the Theorem 5.3 in the flat case, we introduce the following definition:

Definition 6.2. A system is said to be *endogenous dynamic feedback linearizable* if there exists an endogenous dynamic feedback of the form (5.28) and a diffeomorphism of the extended manifold $X \times Z$ that transforms the corresponding closed-loop system in a linear controllable system.

Corollary 6.2. *Every flat system is endogenous dynamic feedback linearizable. Conversely, every endogenous dynamic feedback linearizable system is flat.*

Moreover, if the system is given by $\dot{x} = f(x,u)$ *with* x *in a* n*-dimensional manifold and* u *of dimension* m*, there exists integers* $r_1, \ldots, r_m$ *with* $\sum_{i=1}^{m} r_i \geq n$ *such that* x *and* u *are given by*

$$\begin{aligned} x &= \varphi_0(y_1, \dot{y}_1, \ldots, y_1^{(r_1)}, \ldots, y_m, \dot{y}_m, \ldots, y_m^{(r_m)}) \\ u &= \varphi_1(y_1, \dot{y}_1, \ldots, y_1^{(r_1+1)}, \ldots, y_m, \dot{y}_m, \ldots, y_m^{(r_m+1)}) \end{aligned} \tag{6.23}$$

and such that the closed-loop system is diffeomorphic to the linear controllable system in canonical form

$$\begin{aligned} & y_1^{(r_1+1)} = v_1 \\ & \vdots \\ & y_m^{(r_m+1)} = v_m. \end{aligned} \tag{6.24}$$

Proof. Since the system $(\mathfrak{X}, \overline{f})$ is L-B equivalent to the trivial system $(\mathbb{R}^m_\infty, \tau_m)$, by Theorem 5.3, there exist an endogenous dynamic feedback and a diffeomorphism that transform the system in the linear controllable system $y^{(\rho)} = v$ for ρ large enough.

The converse is obvious since the closed-loop system is L-B equivalent to the trivial system $(\mathbb{R}^m_\infty, \tau_m)$ and since the dynamic feedback is endogenous: x and u can be expressed as functions of y and a finite number of successive derivatives.

Since in (6.23) only appear derivatives of y up to order $r_1+1,\ldots,r_m+1$, and since the mapping (φ_0,φ_1) is onto, we get $n+m \leq \sum_{i=1}^m (r_i+1)$, thus $n \leq \sum_{i=1}^m r_i$. Moreover, following the same lines as in Theorem 5.3 in the flat case, it suffices to set $v_i = y_i^{(r_i+1)}$ to construct the corresponding diffeomorphism and endogenous dynamic feedback, which completes the proof.

Remark 6.3. The set of diffeomorphisms and static state feedback being obviously a strict subset of the set of endogenous dynamic feedbacks, the systems that are linearizable by diffeomorphism and static state feedback (often simply called *static feedback linearizable*) are thus a strict subset of the set of differentially flat systems. In the single input case ($m = 1$), one can prove that flatness is equivalent to static feedback linearization Charlet et al. [1989, 1991]. The flatness property thus only reveals its richness in the multi-input case, where static state feedback linearization and endogenous dynamic feedback linearization are no more equivalent. The interested reader may verify, e.g., that the pendulum and the non holonomic vehicle examples are flat but not static feedback linearizable (one may use the Jakubczyk-Respondek criterion Jakubczyk and Respondek [1980]. See also Hunt et al. [1983a], Isidori [1995], Lévine [1997], Marino [1986], Marino and Tomei [1995], Nijmeijer and van der Schaft [1990]).

We go back to the previous examples and construct the associated linearizing endogenous dynamic feedbacks.

6.4.1 Mass-Spring System (followed)

According to (6.2), (6.4), (6.5), (6.6), x_1 is a flat output and the input u can be expressed in function of x_1 and derivatives up to order 4. Let us set:

$$x_1^{(4)} = v \tag{6.25}$$

and let us show that systems (6.2) and (6.25) are L-B equivalent. According to (6.6), xe have:

$$x_1^{(4)} = \frac{k_2}{m_1 m_2}\left(u - X(x_1, \dot{x}_1, \ddot{x}_1, x_1^{(3)})\right) \tag{6.26}$$

with

$$
\begin{aligned}
X(x_1, \dot{x}_1, \ddot{x}_1, x_1^{(3)}) = \\
&\left(\frac{m_2 k_1}{k_2} + m_2 + m_1\right) \ddot{x}_1 + k_1 x_1 + \gamma_1(\dot{x}_1) \\
&\quad + \frac{m_2}{k_2} \left(\gamma_1''(\dot{x}_1)(\ddot{x}_1)^2 + \gamma_1'(\dot{x}_1) x_1^{(3)}\right) \\
&\qquad + \gamma_2 \left(\frac{m_1}{k_2} x_1^{(3)} + \frac{1}{k_2}((k_1 + k_2)\dot{x}_1 + \gamma_1'(\dot{x}_1)\ddot{x}_1)\right)
\end{aligned}
$$

and the linearizing feedback is given by $v = \frac{k_2}{m_1 m_2}\left(u - X(x_1, \dot{x}_1, \ddot{x}_1, x_1^{(3)})\right)$. Remark that, in this case, we obtain a static feedback.

6.4.2 Robot Control (followed)

We verify, as in the previous example, that (6.8) is L-B equivalent to $\ddot{x}_1 = v$ (by static feedback as in the previous example).

6.4.3 Pendulum (followed)

According to (6.16), it suffices to set

$$\xi^{(4)} = v_1 \ , \quad \zeta^{(4)} = v_2. \tag{6.27}$$

The endogenous dynamic feedback computation is done by identification of the 4th order derivatives of ξ and ζ: we have

$$
\begin{aligned}
\ddot{\xi} &= \left(u_1 \sin\theta + (u_2 + 1)\cos\theta - \varepsilon\dot{\theta}^2\right) \sin\theta, \\
\ddot{\zeta} &= \left(u_1 \sin\theta + (u_2 + 1)\cos\theta - \varepsilon\dot{\theta}^2\right) \cos\theta - 1
\end{aligned}
$$

or, with the change of inputs:

$$w_1 = u_1 \sin\theta + (u_2 + 1)\cos\theta - \varepsilon\dot{\theta}^2, \tag{6.28}$$

$$\ddot{\xi} = w_1 \sin\theta \ , \quad \ddot{\zeta} = w_1 \cos\theta - 1.$$

differentiating again 2 times and setting

$$w_2 = -u_1 \cos\theta + (u_2 + 1)\sin\theta \tag{6.29}$$

we get:

$$\xi^{(4)} = \ddot{w}_1 \sin\theta + \frac{1}{\varepsilon} w_1 w_2 \cos\theta + 2\dot{w}_1 \dot{\theta} \cos\theta - w_1 \dot{\theta}^2 \sin\theta = v_1$$
$$\zeta^{(4)} = \ddot{w}_1 \cos\theta - \frac{1}{\varepsilon} w_1 w_2 \sin\theta - 2\dot{w}_1 \dot{\theta} \sin\theta - w_1 \dot{\theta}^2 \cos\theta = v_2.$$

inverting this linear system with respect to $\ddot{w}_1$ and w_2, we get

$$\begin{aligned} \ddot{w}_1 &= v_1 \sin\theta + v_2 \cos\theta + w_1 \dot{\theta}^2 \\ w_2 &= \frac{\varepsilon}{w_1} \left(v_1 \cos\theta - v_2 \sin\theta - 2\dot{w}_1 \dot{\theta} \right) \end{aligned} \tag{6.30}$$

with

$$\begin{aligned} u_1 &= (w_1 + \varepsilon\dot{\theta}^2) \sin\theta - w_2 \cos\theta \\ u_2 &= (w_1 + \varepsilon\dot{\theta}^2) \cos\theta + w_2 \sin\theta - 1. \end{aligned} \tag{6.31}$$

The endogenous dynamic feedback is thus made of a compensator whose state is $(w_1, \dot{w}_1, w_2)$ and for which the system (6.9) is equivalent to (6.27).

6.4.4 Non Holonomic Vehicle (followed)

As before, we show that (6.17) is L-B equivalent to

$$\ddot{x} = v_1 \ , \quad \ddot{y} = v_2.$$

An elementary computation yields:

$$\ddot{x} = \dot{u} \cos\theta - \frac{u^2}{l} \sin\theta \tan\varphi = v_1$$
$$\ddot{y} = \dot{u} \sin\theta + \frac{u^2}{l} \cos\theta \tan\varphi = v_2$$

and, after inversion of this linear system with respect to $\dot{u}$ and $\tan\varphi$, we obtain the endogenous dynamic compensator:

$$\dot{u} = v_1 \cos\theta + v_2 \sin\theta \ , \quad \tan\varphi = \frac{l}{u^2} \left(-v_1 \sin\theta + v_2 \cos\theta \right)$$

hence the result.

6.5 Flat Output Characterization

We consider the implicit system:

$$F(x, \dot{x}) = 0 \tag{6.32}$$

with rank $\left(\frac{\partial F}{\partial \dot{x}}\right) = n - m$ in a given neighborhood of an arbitrary trajectory. Recall that a flat output y is such that $y = (y_1, \ldots, y_m)^T = \psi(x, \dot{x}, \ldots, x^{(s)})$ implies $x = \varphi(y, \dot{y}, \ldots, y^{(r)})$ for some suitably chosen multi-integers r and s, the equation $F(\varphi, \frac{d\varphi}{dt}) = 0$ being locally identically satisfied, with $\frac{d\varphi}{dt} = L_{\tau_m}\varphi$, τ_m being the trivial Cartan field of order m.

The implicit representation (6.32), as opposed to the explicit representation

$$\dot{x} = f(x, u) \tag{6.33}$$

is invariant by endogenous dynamic feedback extension:

Proposition 6.3. *System (6.32) is invariant by endogenous dynamic feedback extension. In other words, if the explicit system (6.33) admits the locally equivalent implicit representation (6.32), and if we are given the endogenous dynamic feedback*

$$u = a(x, z, v), \quad \dot{z} = b(x, z, v) \tag{6.34}$$

then the closed-loop system

$$\dot{x} = f(x, a(x, z, v)), \quad \dot{z} = b(x, z, v) \tag{6.35}$$

also admits the locally equivalent implicit representation (6.32).

Proof. We assume that (6.33) admits the locally equivalent implicit representation (6.32) and that the endogenous dynamic feedback (6.34) is given, with z remaining in Z, a given finite dimensional smooth manifold. By definition of an endogenous dynamic feedback, the closed-loop system (6.35) is Lie-Bäcklund equivalent to (6.33). Thus, there exist finite integers α and β and locally onto smooth mappings Φ and Ψ such that

$$(x, z, v) = \Phi(x, u, \ldots, u^{(\alpha)}), \quad (x, u) = \Psi(x, z, v, \ldots, v^{(\beta)}). \tag{6.36}$$

According to the implicit function Theorem, there exists a locally defined smooth mapping μ such that (6.33) is locally equivalent to (6.32) with $u = \mu(x, \dot{x})$. Denoting by $u^{(k)} = \mu^{(k)}(x, \dot{x}, \ldots, x^{(k+1)})$ and $\tilde{\Phi}(x, \ldots, x^{(\alpha+1)}) = \Phi(x, \mu(x, \dot{x}), \ldots, \mu^{(\alpha)}(x, \ldots, x^{(\alpha+1)}))$, we immediately deduce that for every smooth local integral curve $t \mapsto x(t)$ of (6.32) passing through an arbitrary point $(x_0, \dot{x}_0)$, using the first relation of (6.36), $t \mapsto (x(t), z(t))$, given by $(x(t), z(t), v(t)) = \tilde{\Phi}(x(t), \ldots, x^{(\alpha+1)}(t))$, is also a smooth local integral curve of the closed-loop system (6.35) passing through (x_0, z_0) such that $(x_0, z_0, v_0) = \tilde{\Phi}(x_0, \ldots, x_0^{(\alpha+1)})$.

Conversely, if $t \mapsto (x(t), z(t), v(t))$ is a smooth local integral curve of (6.35) passing through (x_0, z_0), by the second relation of (6.36), we get that $(x(t), u(t)) = \Psi(x(t), z(t), v(t), \ldots, v^{(\beta)}(t))$ is a smooth local integral curve of (6.33) and, according to the assumption, $t \mapsto x(t)$ is also a local smooth integral curve of (6.32) passing through the corresponding point $(x_0, \dot{x}_0)$, which achieves the proof.

6.5.1 The Ruled Manifold Necessary Condition

Theorem 6.3 (Sluis [1993], Rouchon [1994]). *If System (6.32) is flat, there exists q non zero independent vector fields $g_1, \ldots, g_q$, with $1 \le q \le m$, such that*

$$F(x, \dot{x} + \sum_{j=1}^{q} \lambda_j g_j) = 0 \tag{6.37}$$

for all $(x, \dot{x}, \ldots)$ in an open dense subset of $\mathfrak{X}$, and all $\lambda = (\lambda_1, \ldots, \lambda_q)$ in a neighborhood of the origin of $\mathbb{R}^q$.

Proof. Consider, on the time interval $[t_0, t_1]$, a flat output y such that for every integral curve x of (6.32), we have

$$x = \varphi_0(y_1, \ldots, y_1^{(r_1)}, \ldots, y_m, \ldots, y_m^{(r_m)})$$

for some multi-integer $r = (r_1, \ldots, r_m)$.

We also consider a curve y^* such that every component y_i^* is at least C^{r_i+1}, $i = 1, \ldots, m$, and a perturbed curve $\tilde{y}$ whose components are C^{r_i}, defined by

$$\tilde{y}_i^{(j)}(t) = \begin{cases} (y_i^*)^{(j)}(t) & \text{if } j \le r_i,\ \forall t \in [t_0, t_0 + \tau] \\ (y_i^*)^{(j)}(t) & \text{if } j = r_i + 1\ ,\ \forall t \in [t_0, t_0 + \tau[\\ (y_i^*)^{(j)}(t) + \lambda_i & \text{if } j = r_i + 1\ ,\ t = t_0 + \tau \end{cases}$$

with τ such that $t_0 + \tau < t_1$ and $\lambda_i \in] - l_i, l_i[$, $i = 1, \ldots, m$.

The flatness property implies that to y^* and $\tilde{y}$ there corresponds

$$x^* = \varphi_0(y^*, \dot{y}^*, \ldots, (y^*)^{(r)}), \quad \tilde{x} = \varphi_0(\tilde{y}, \dot{\tilde{y}}, \ldots, \tilde{y}^{(r)}) \tag{6.38}$$

which are integral curves of (6.32), *i.e.* such that

$$F(x^*(t), \dot{x}^*(t)) = 0, \quad F(\tilde{x}(t), \dot{\tilde{x}}(t)) = 0 \tag{6.39}$$

for all $t_0 \le t \le t_0 + \tau$. Conversely, we have

$$y^* = \psi_0(x^*, \dot{x}^*, \ldots, (x^*)^{(s)}), \qquad (y^*)^{(j)} = \psi_j(x^*, \dot{x}^*, \ldots, (x^*)^{(s+j)}), \ \ j \ge 0$$

for some multi-integer s, and

$$\tilde{y} = \psi_0(\tilde{x}, \dot{\tilde{x}}, \ldots, \tilde{x}^{(s)}), \qquad \tilde{y}^{(j)} = \psi_j(\tilde{x}, \dot{\tilde{x}}, \ldots, \tilde{x}^{(s+j)}), \ \ j \ge 0.$$

It is easily seen that, for $t_0 \le t < t_0 + \tau$, we have $\tilde{x}(t) = x^*(t)$, and differentiating the second relation of (6.38) with respect to t, we get

$$\dot{\tilde{x}}(t) = \dot{x}^*(t) \qquad \forall t \in [t_0, t_0 + \tau[$$

and

$$\dot{\bar{x}}(t_0+\tau) = \dot{x}^*(t_0+\tau) + \sum_{j=1}^{m} \lambda_j \frac{\partial \varphi_0}{\partial y_j^{(r_j)}} (\psi(x^*(t_0+\tau), \dot{x}^*(t_0+\tau), \ldots, (x^*)^{(s+r)}(t_0+\tau)))$$

where $\psi(x, \dot{x}, \ldots, x^{(s+r)})$ stands for the vector $(\psi_0(x, \dot{x}, \ldots, x^{(s)}), \ldots, \psi_r(x, \dot{x}, \ldots, x^{(s+r)}))$. Thus, denoting by

$$g_j(x, \dot{x}, \ldots, x^{(s+r)}) = \sum_{i=1}^{n} \frac{\partial \varphi_{0,i}}{\partial y_j^{(r_j)}} (\psi(x, \dot{x}, \ldots, x^{(s+r)})) \frac{\partial}{\partial \dot{x}_i}, \quad j = 1, \ldots, m$$

the g_j's are not all equal to zero and independent by the surjectivity of φ_0, and, using (6.39), condition (6.37) readily follows. *Q.E.D.*

The geometric interpretation of (6.37) is that at every point $(x, \dot{x})$ of the manifold of equation $F(x, \dot{x}) = 0$, an open subset of the plane generated by the vectors $(g_1, \ldots, g_q)$ entirely lies in this manifold. Therefore, the manifold $F(x, \dot{x}) = 0$ is locally obtained by rolling the plane generated by $(g_1, \ldots, g_q)$ on a submanifold. If $q = 1$, this plane is just a line, rolling on a hypersurface, hence the name *ruled manifold.*

Example 6.1. We consider the 6th-order dynamical system, with states $\{\psi, \phi, \dot{\psi}, \dot{\phi}, \omega_m, \omega_r\}$ and two inputs v_m and v_r, representing a toycopter model studied in Müllhaupt et al. [2008]:

$$\begin{aligned} I_\psi \ddot{\psi} + I_r \dot{\omega}_r = {} & C_m \omega_m \mid \omega_m \mid - C_{r1} \omega_r \mid \omega_r \mid \\ & + G_s \sin\psi + G_c \cos\psi \\ & + \frac{1}{2} I_c \dot{\phi}^2 \sin(2\psi) + I_m \omega_m \dot{\phi} \cos\psi \end{aligned} \tag{6.40}$$

$$\begin{aligned} (I_\phi &+ I_c \sin^2(\psi))\ddot{\phi} + I_m \dot{\omega}_m \sin\psi \\ &= C_r \omega_r \mid \omega_r \mid \sin\psi - C_{m1} \omega_m \mid \omega_m \mid \sin\psi \\ &\quad - I_c \dot{\psi}\dot{\phi} \sin(2\psi) - I_m \omega_m \dot{\psi} \cos\psi \end{aligned} \tag{6.41}$$

$$\dot{\omega}_m = v_m \tag{6.42}$$

$$\dot{\omega}_r = v_r \tag{6.43}$$

We next show that this model is not flat by application of Theorem 6.3. For this purpose, we change the state variables ω_m and ω_r to the generalized momenta

$$\eta_m = I_\psi \dot{\psi} + I_r \omega_r \tag{6.44}$$

$$\eta_r = (I_\phi + I_c \ \sin^2\psi)\dot{\phi} + I_m \sin\psi \omega_m \tag{6.45}$$

The Toycopter dynamics reads, when operating such that both $\omega_m > 0$ and $\omega_r > 0$[1]:

$$\begin{aligned}\dot{\eta}_m &= C_m\omega_m^2 - C_{r1}\omega_r^2 + \frac{1}{2}I_c\sin(2\ \psi)\dot{\phi}^2 + I_m\omega_m\dot{\phi}\ \cos\psi\\ &\quad + G_s\sin\psi + G_c\cos\psi\\ \dot{\eta}_r &= C_r\omega_r^2\sin\psi - C_{m1}\omega_m^2\sin\psi \end{aligned} \tag{6.46}$$

Replacing ω_m and ω_r in the above equations by the generalized momenta η_m and η_r using (6.44) and (6.45) leads to the implicit system $\tilde{F}(\tilde{x}, \dot{\tilde{x}}) = 0$ with the state variables $\tilde{x} = \{\psi, \phi, \eta_m, \eta_r\}$. Replacing $\dot{\tilde{x}}$ by $\dot{\tilde{x}} + \lambda g$ where $g = [g_\psi, g_\phi, g_m, g_r]^T$ and $\dot{\tilde{x}} = [\dot{\psi}, \dot{\phi}, \dot{\eta}_m, \dot{\eta}_r]^T$, and using the fact that $\tilde{F}(\tilde{x}, \dot{\tilde{x}}) = 0$ and $\tilde{F}(\tilde{x}, \dot{\tilde{x}} + \lambda g) = 0$, gives two polynomial equations in λ, $\lambda\xi_1 + \lambda^2\xi_2 = 0$ and $\lambda\xi_3 + \lambda^2\xi_4$ with:

$$\begin{aligned}\xi_1 = g_m &- \frac{2C_{r1}I_\psi}{I_r^2}(\eta_m - I_\psi p_\psi)g_\psi \\ &+ \frac{2C_m}{I_m^2}\left[\left(I_\phi + I_c\sin^2\psi - \frac{I_m^2}{4C_m}\sin(2\psi)\right)\eta_r \right.\\ &\left. - \left((I_\phi + I_c\sin^2\psi)^2 + \frac{I_m^2}{2C_m}\sin(2\ \psi)\right)\dot{\phi}\right] g_\phi\end{aligned} \tag{6.47}$$

$$\xi_2 = \left[-\frac{C_m}{I_m^2}\left(\frac{I_\phi}{\sin\psi} + I_c\sin\ \ \psi\right)^2 + I_\phi\frac{\cos\psi}{\sin\psi}\right]g_\phi^2 + \frac{C_{r1}}{I_r^2}I_\psi^2 g_\psi^2$$

$$\begin{aligned}\xi_3 = g_r &+ \frac{2C_r\sin\psi}{I_r^2}I_\psi(\eta_m - I_\psi p_\psi)g_\psi \\ &- \frac{2C_{m1}}{I_m^2\sin\psi}(I_\phi + I_c\sin^2\psi)[\eta_r \\ &- (I_\phi + I_c\sin^2\psi)p_\phi]g_\phi\end{aligned} \tag{6.48}$$

$$\xi_4 = \left[\frac{C_{m1}}{I_m^2}\left(\frac{I_\phi}{\sin\psi} + I_c\sin\psi\right)^2 g_\phi^2 - \frac{C_r}{I_r^2}I_\psi^2 g_\psi^2\right]\sin\psi$$

These four functions should identically vanish since the associated polynomial equations in λ must be valid independently of the value of λ. It can be verified that $\xi_2 = 0$ and $\xi_4 = 0$ form a system of two equations in the two unknowns g_ψ^2 and g_ϕ^2 whose only solution is $g_\psi^2 = 0$ and $g_\phi^2 = 0$, thus $g_\psi = g_\phi = 0$. The remaining coefficients ξ_1 and ξ_3 then force $g_m = 0$ and $g_r = 0$. Therefore, there exists no g vector different from 0 such that $\tilde{F}(\tilde{x}, \dot{\tilde{x}} + \lambda g) = 0$ is satisfied for all λ in an open subset of $\mathbb{R}$. Hence, the Toycopter is not flat.

[1] This is not restrictive and the other cases lead to the same result.

6.5.2 Variational Characterization

We now follow the approach developed in Lévine [2004], Lévine [2006]. The interested reader may find different approaches in Aranda-Bricaire et al. [1995], Avanessoff and Pomet [2007], Chetverikov [2001], Fossas and Franch [1999], Franch [1999], Martin and Rouchon [1993], van Nieuwstadt et al. [1998], Pereira da Silva [2000], Pomet [1997], Pommaret and Quadrat [1999], Rathinam and Murray [1998], Schlacher and Schöberl [2007], Shadwick [1990], Sluis and Tilbury [1996].

Theorem 6.4. *The implicit system $F(x, \dot{x}) = 0$ is locally flat at $(\overline{x}_0, \overline{y}_0)$ with $\overline{x}_0 \in \mathfrak{X}_0$ and $\overline{y}_0 \in \mathbb{R}^m_\infty$ if and only if there exists a locally invertible mapping Φ from a neighborhood of $\overline{y}_0$ in $\mathbb{R}^m_\infty$ to a neighborhood of $\overline{x}_0$ in $\mathfrak{X}_0$, of class C^∞, satisfying $\Phi(\overline{y}_0) = \overline{x}_0$, and such that locally*

$$\Phi^* dF = 0. \tag{6.49}$$

Proof. First remark that the implicit representation of a trivial system is given by $G \equiv 0$ since there is no relation between the components of y and their successive derivatives.

If system $(X \times \mathbb{R}^n_\infty, \tau_X, F)$ is flat at $(\overline{x}_0, \overline{y}_0)$, we have $F(\varphi_0(\overline{y}), \varphi_1(\overline{y})) = 0$ for all $\overline{y}$ in a neighborhood of $\overline{y}_0$ in $\mathbb{R}^m_\infty$. Thus $\frac{\partial F}{\partial x}\frac{\partial \varphi_0}{\partial \overline{y}} d\overline{y} + \frac{\partial F}{\partial \dot{x}}\frac{\partial \varphi_1}{\partial \overline{y}} d\overline{y} = \Phi^* dF = 0$, by (2.28), hence the necessary condition.

Conversely, since $\Phi^* dF_i = d\left(F_i \circ \Phi\right)$, the Pfaffian system $\Phi^* dF_i = 0$, $i = 1, \ldots, n-m$, is integrable by construction. If we set $\overline{x} = \Phi(\overline{y})$, we get that x can be expressed in function of y and a finite number of derivatives and that $F_i(x, \dot{x}) = c_i$, the c_i's being arbitrary constants. Since $F(x_0, \dot{x}_0) = F \circ \Phi(\overline{y}_0) = 0$, these constants vanish. Furthermore, since Φ is invertible, we conclude that the system $F(x, \dot{x}) = 0$ is flat at $(\overline{x}_0, \overline{y}_0)$. *Q.E.D.*

This result shows that the mapping Φ, if it exists, is characterized by its tangent mapping since the latter's range must be contained in the kernel of dF: if we denote by $\Phi = (\varphi_{0,1}, \ldots, \varphi_{0,n}, \varphi_{1,1}, \ldots, \varphi_{1,n}, \ldots)$ (recall that $x_i = \varphi_{0,i}(\overline{y})$ and $x_i^{(k)} = \varphi_{k,i}(\overline{y})$ for all $i = 1, \ldots, n$ and $k \geq 1$, with $\varphi_{k,i} = \frac{d^k \varphi_{0,i}}{dt^k}$, and, by (2.28), $\Phi^* dF = \frac{\partial F}{\partial x}\frac{\partial \varphi_0}{\partial \overline{y}} d\overline{y} + \frac{\partial F}{\partial \dot{x}}\frac{\partial \varphi_1}{\partial \overline{y}} d\overline{y}$), the condition $\Phi^* dF = 0$ reads, in matrix form:

$$\sum_{j\geq 0}\begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} & \frac{\partial F_1}{\partial \dot{x}_1} & \cdots & \frac{\partial F_1}{\partial \dot{x}_n} \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial F_{n-m}}{\partial x_1} & \cdots & \frac{\partial F_{n-m}}{\partial x_n} & \frac{\partial F_{n-m}}{\partial \dot{x}_1} & \cdots & \frac{\partial F_{n-m}}{\partial \dot{x}_n} \end{pmatrix} \begin{pmatrix} \frac{\partial \varphi_{0,1}}{\partial y_1^{(j)}} & \cdots & \frac{\partial \varphi_{0,1}}{\partial y_m^{(j)}} \\ \vdots & & \vdots \\ \frac{\partial \varphi_{0,n}}{\partial y_1^{(j)}} & \cdots & \frac{\partial \varphi_{0,n}}{\partial y_m^{(j)}} \\ \frac{\partial \varphi_{1,1}}{\partial y_1^{(j)}} & \cdots & \frac{\partial \varphi_{1,1}}{\partial y_m^{(j)}} \\ \vdots & & \vdots \\ \frac{\partial \varphi_{1,n}}{\partial y_1^{(j)}} & \cdots & \frac{\partial \varphi_{1,n}}{\partial y_m^{(j)}} \end{pmatrix} \begin{pmatrix} dy_1^{(j)} \\ \vdots \\ dy_m^{(j)} \end{pmatrix} = 0.$$

6.5.3 The Polynomial Matrix Approach

We introduce the polynomial matrix notations

$$P(F) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} + \frac{\partial F_1}{\partial \dot{x}_1}\frac{d}{dt} & \cdots & \frac{\partial F_1}{\partial x_n} + \frac{\partial F_1}{\partial \dot{x}_n}\frac{d}{dt} \\ \vdots & & \vdots \\ \frac{\partial F_{n-m}}{\partial x_1} + \frac{\partial F_{n-m}}{\partial \dot{x}_1}\frac{d}{dt} & \cdots & \frac{\partial F_{n-m}}{\partial x_n} + \frac{\partial F_{n-m}}{\partial \dot{x}_n}\frac{d}{dt} \end{pmatrix}$$

and

$$P(\varphi_0) = \begin{pmatrix} \sum_{j\geq 0} \frac{\partial \varphi_{0,1}}{\partial y_1^{(j)}}\frac{d^j}{dt^j} & \cdots & \sum_{j\geq 0} \frac{\partial \varphi_{0,1}}{\partial y_m^{(j)}}\frac{d^j}{dt^j} \\ \vdots & & \vdots \\ \sum_{j\geq 0} \frac{\partial \varphi_{0,n}}{\partial y_1^{(j)}}\frac{d^j}{dt^j} & \cdots & \sum_{j\geq 0} \frac{\partial \varphi_{0,n}}{\partial y_m^{(j)}}\frac{d^j}{dt^j} \end{pmatrix}.$$

Thus $\Phi^* dF = 0$ reads

$$P(F).P(\varphi_0)dy = 0 \tag{6.50}$$

with $dy = \begin{pmatrix} dy_1 \\ \vdots \\ dy_m \end{pmatrix}$.

Indeed, since $dy_k^{(l)} = \frac{d^l}{dt^l}dy_k$ and since $x_i = \varphi_{0,i}(\overline{y})$ simultaneously implies

$$dx_i = \sum_{l\geq 0}\sum_{k=1}^{m} \frac{\partial \varphi_{0,i}}{\partial y_k^l} dy_k^{(l)} \text{ and } \dot{x}_i = \frac{d\varphi_{0,i}}{dt}(\overline{y}) = \varphi_{1,i}(\overline{y})$$

we get

$$
\begin{aligned}
d\dot{x}_i &= \sum_{l\geq 0}\sum_{k=1}^{m}\left(\frac{\partial}{\partial y_k^{(l)}}\left(\frac{d\varphi_{0,i}}{dt}\right)dy_k^{(l)} + \frac{\partial \varphi_{0,i}}{\partial y_k^{(l)}}dy_k^{(l+1)}\right)\\
&= \sum_{l\geq 0}\sum_{k=1}^{m}\left(\frac{\partial}{\partial y_k^{(l)}}\left(\frac{d\varphi_{0,i}}{dt}\right) + \frac{\partial \varphi_{0,i}}{\partial y_k^{(l-1)}}\right)dy_k^{(l)}\\
&= \sum_{l\geq 0}\sum_{k=1}^{m}\frac{\partial \varphi_{1,i}}{\partial y_k^{(l)}}dy_k^{(l)}
\end{aligned}
$$

thus the ith row of $P(F).P(\varphi_0)dy$ is

$$
\begin{aligned}
P(F_i).P(\varphi_0)dy &= \sum_{k=1}^{m}\sum_{j=1}^{n}\left(\frac{\partial F_i}{\partial x_j} + \frac{\partial F_i}{\partial \dot{x}_j}\frac{d}{dt}\right)\left(\sum_{l\geq 0}\frac{\partial \varphi_{0,j}}{\partial y_k^{(l)}}\frac{d^l}{dt^l}dy_k\right)\\
&= \sum_{l\geq 0}\sum_{k=1}^{m}\sum_{j=1}^{n}\left(\frac{\partial F_i}{\partial x_j}\frac{\partial \varphi_{0,j}}{\partial y_k^{(l)}}dy_k^{(l)} + \frac{\partial F_i}{\partial \dot{x}_j}\frac{\partial}{\partial y_k^{(l)}}\left(\frac{d\varphi_{0,j}}{dt}\right)dy_k^{(l)} + \frac{\partial F_i}{\partial \dot{x}_j}\frac{\partial \varphi_{0,j}}{\partial y_k^{(l)}}dy_k^{(l+1)}\right)\\
&= \sum_{l\geq 0}\sum_{k=1}^{m}\sum_{j=1}^{n}\left(\frac{\partial F_i}{\partial x_j}\frac{\partial \varphi_{0,j}}{\partial y_k^{(l)}} + \frac{\partial F_i}{\partial \dot{x}_j}\frac{\partial \varphi_{1,j}}{\partial y_k^{(l)}}\right)dy_k^{(l)} = \Phi^* dF_i
\end{aligned}
$$

hence the result.

Moreover, since the components of dy are independent by definition, the equation (6.50) is equivalent to

$$
P(F)_{\big|\Phi(\overline{y})}.P(\varphi_0)_{\big|\overline{y}} = P(F)_{\big|\overline{x}}.P(\varphi_0)_{\big|\Psi(\overline{x})} = 0. \tag{6.51}
$$

This equation is made of the product of two matrices whose entries are polynomials of the operator $\frac{d}{dt}$ with coefficients in the space of C^∞ functions on $\mathfrak{X}_0$.

For a matrix A whose entries are real numbers, in order to solve the equation $AX = 0$, where the unknown matrix X is a real matrix with maximal rank, it is well-known that X is obtained by the singular values decomposition of A. Clearly, X is the product of a matrix made of the vectors forming a basis of $\ker(A)$ completed by a suitable number of columns of zeros. This kind of solution remains valid for matrices on a ring (see e.g. Cohn [1985]). Unfortunately, the set of polynomials whose coefficients are C^∞ functions doesn't form a ring (see e.g. Kostrikin and Shafarevich [1980]). This is why we now restrict to the ring of polynomials whose coefficients are meromorphic functions on $\mathfrak{X}_0$ (recall that a real function is meromorphic if and only if it may be represented as a rational fraction of analytic functions).

We thus assume from now on that F is meromorphic on $\mathfrak{X}$ (endowed with the trivial Cartan field $\frac{d}{dt} = \tau_X$), as well as Ψ, and Φ on $\mathbb{R}^m_\infty$ (endowed with the trivial Cartan field τ_m).

We denote by $\mathfrak{K}$ the ring of meromorphic functions from $\mathfrak{X}$ to $\mathbb{R}$, $\mathfrak{K}[\frac{d}{dt}]$ the ring of polynomials in $\frac{d}{dt}$ with coefficients in $\mathfrak{K}$, and $\mathcal{M}_{n,m}[\frac{d}{dt}]$ the module of the matrices of size $n \times m$ on $\mathfrak{K}[\frac{d}{dt}]$.

Note that $\mathfrak{K}[\frac{d}{dt}]$ is non commutative: first recall that its elements are differential operators on $\mathfrak{K}$ of the form $\sum_{i\geq 0} a_i \frac{d^i}{dt^i} = \sum_{i\geq 0} a_i L^i_{\tau_X}$ with $a_i \in \mathfrak{K}$ for all i. Thus, if p and q belong to $\mathfrak{K}[\frac{d}{dt}]$, they are two such operators and their products pq and qp, which are also in $\mathfrak{K}[\frac{d}{dt}]$ by construction, are equal if and only if $pqf = qpf$ for every function f in $\mathfrak{K}$. Taking $p = \frac{d}{dt}$ and $q = x\frac{d}{dt}$, we have $pqf = \frac{d}{dt}(x\frac{df}{dt}) = \dot{x}\frac{df}{dt} + x\frac{d^2 f}{dt^2} = \left(\dot{x}\frac{d}{dt} + x\frac{d^2}{dt^2}\right) f$, or $pq = \left(\dot{x}\frac{d}{dt} + x\frac{d^2}{dt^2}\right)$, whereas $qpf = x\frac{d}{dt}(\frac{df}{dt}) = x\frac{d^2 f}{dt^2}$, or $qp = x\frac{d^2}{dt^2} \neq pq$. We thus have to make a clear distinction between right and left products.

The matrices of $\mathcal{M}_{n,n}[\frac{d}{dt}]$ (square matrices of size n on $\mathfrak{K}[\frac{d}{dt}]$), even if they are invertible in the usual sense (all rows or columns linearly independent), don't necessarily possess an inverse in $\mathcal{M}_{n,n}[\frac{d}{dt}]$ since their entries are rational fractions, but not polynomials in general, of $\frac{d}{dt}$. Those matrices of $\mathcal{M}_{n,n}[\frac{d}{dt}]$ that possess an inverse in $\mathcal{M}_{n,n}[\frac{d}{dt}]$ are called *unimodular matrices*. They form a subgroup, noted $\mathcal{U}_n[\frac{d}{dt}]$, of $\mathcal{M}_{n,n}[\frac{d}{dt}]$.

The matrices of $\mathcal{M}_{n,m}[\frac{d}{dt}]$, in spite of their poor algebraic properties compared to matrices on $\mathbb{R}$, nevertheless can be diagonalized according to the so-called *Smith decomposition* or *diagonal decomposition* process. More precisely, if $M \in \mathcal{M}_{n,m}[\frac{d}{dt}]$, there exist unimodular matrices $U \in \mathcal{U}_m[\frac{d}{dt}]$, $V \in \mathcal{U}_n[\frac{d}{dt}]$ and a diagonal matrix Δ of size $p \times p$ with $p = \min(n, m)$, whose ith diagonal element $d_{i,i} \in \mathfrak{K}[\frac{d}{dt}]$ divides $d_{j,j}$ for all $0 \leq i \leq j \leq p$, such that

$$VMU = \begin{cases} (\Delta, 0_{n,m-n}) \text{ if } n < m \\ \begin{pmatrix} \Delta \\ 0_{n-m,m} \end{pmatrix} \text{ if } n > m \end{cases} \tag{6.52}$$

with $0_{r,s}$ the matrix of size $r \times s$ whose entries are all 0. This decomposition is non unique. Only the entries of Δ are uniquely defined up to multiplication by a diagonal matrix with non zero elements of $\mathfrak{K}$.

In the special case where, in the Smith decomposition process, we obtain a matrix Δ whose entries are all in $\mathfrak{K}$, we say that M is *hyper-regular*. In this case, up to elementary modifications of the corresponding matrices U and V, we end up with $\Delta = I_p$, the identity matrix of size $p \times p$.
If M is square of size n, saying that M is hyper-regular is equivalent to saying that M is unimodular, *i.e.* $M \in \mathcal{U}_n[\frac{d}{dt}]$.

Given $M \in \mathcal{M}_{n,m}[\frac{d}{dt}]$, we note $U \in \mathsf{R}-\mathsf{Smith}\,(M)$ if there exists $V \in \mathcal{U}_n[\frac{d}{dt}]$ such that (6.52) holds true, and $V \in \mathsf{L}-\mathsf{Smith}\,(M)$ if there exists $U \in \mathcal{U}_m[\frac{d}{dt}]$ such that (6.52) holds true.

From now on, we assume that $P(F) \in \mathcal{M}_{n-m,n}[\frac{d}{dt}]$ (with $n > m$) is hyper-regular. This property can in fact be interpreted as the controllability of the tangent linear system at every point of $\mathfrak{X}_0$ (see Fliess [1990, 1992], Lévine [2006]).

6.5.4 Practical Computation of the Smith Decomposition

Let us precise the conditions of Theorem 6.6. To this aim, we recall some basic properties of polynomial matrices and the Smith decomposition algorithm.

We consider matrices of size $p \times q$, for arbitrary integers p and q, over the principal ideal domain $\mathfrak{K}[\frac{d}{dt}]$, here the non commutative ring of polynomials of $\frac{d}{dt}$ with coefficients in the field $\mathfrak{K}$ of meromorphic functions on a suitable time interval $\mathfrak{I}$. The set of all such matrices is denoted by $\mathcal{M}_{p,q}[\frac{d}{dt}]$. For arbitrary $p \in \mathbb{N}$, the set $\mathcal{U}_p[\frac{d}{dt}]$ of unimodular matrices of size $p \times p$ is the subgroup of $\mathcal{M}_{p,p}[\frac{d}{dt}]$ of invertible elements, namely the set of invertible polynomial matrices whose inverse is also polynomial.

The following fundamental result on the transformation of a polynomial matrix over a principal ideal domain to its Smith form (or diagonal reduction) may be found in [Cohn, 1985, Chap.8]):

Theorem 6.5. *Given a $(\mu \times \nu)$ polynomial matrix A over the non commutative ring $\mathfrak{K}[\frac{d}{dt}]$, with $\mu \leq \nu$, there exist matrices $V \in \mathcal{U}_\mu[\frac{d}{dt}]$ and $U \in \mathcal{U}_\nu[\frac{d}{dt}]$ such that $VAU = (\Delta, 0)$ where Δ is a $\mu \times \mu$ (resp. $\nu \times \nu$) diagonal matrix whose diagonal elements, $(\delta_1, \ldots, \delta_\sigma, 0, \ldots, 0)$, are such that δ_i is a non zero $\frac{d}{dt}$-polynomial for $i = 1, \ldots, \sigma$, and is a divisor of δ_j for all $\sigma \geq j \geq i$.*

The group of unimodular matrices admits a finite set of generators corresponding to the following *elementary right and left actions*:

- *right actions* consist of permuting two columns, right multiplying a column by a non zero function of $\mathfrak{K}$, or adding the jth column right multiplied by an arbitrary polynomial to the ith column, for arbitrary i and j;
- *left actions* consist, analogously, of permuting two rows, left multiplying a row by a non zero function of $\mathfrak{K}$, or adding the jth row left multiplied by an arbitrary polynomial to the ith row, for arbitrary i and j.

Every elementary action may be represented by an *elementary unimodular matrix* of the form $T_{i,j}(p) = I_\nu + 1_{i,j}p$ with $1_{i,j}$ the matrix made of a single 1 at the intersection of row i and column j, $1 \leq i, j \leq \nu$, and zeros elsewhere, with p an arbitrary polynomial, and with $\nu = m$ for right actions and $\nu = n$ for left actions. One can easily prove that:

- right multiplication $AT_{i,j}(p)$ consists of adding the ith column of A right multiplied by p to the jth column of A, the remaining part of A remaining unchanged,

- left multiplication $T_{i,j}(p)A$ consists of adding the jth row of A left multiplied by p to the ith row of A, the remaining part of A remaining unchanged,
- $T_{i,j}^{-1}(p) = T_{i,j}(-p)$,
- $T_{i,j}(1)T_{j,i}(-1)T_{i,j}(1)A$ (resp. $AT_{i,j}(1)T_{j,i}(-1)T_{i,j}(1)$) is the permutation matrix replacing the jth row of A by the ith one and replacing the jth one of A by the ith one multiplied by -1, all other rows remaining unchanged (resp. the permutation matrix replacing the ith column of A by the jth one multiplied by -1 and replacing the jth one by the ith one, all other columns remaining unchanged).

Every unimodular matrix V (left) and U (right) may be obtained as a product of such elementary unimodular matrices, possibly with a diagonal matrix $D(\alpha) = \mathrm{diag}\{\alpha_1, \ldots, \alpha_\nu\}$ with $\alpha_i \in \mathfrak{K}$, $\alpha_i \neq 0$, $i = 1, \ldots, \nu$, at the end since $T_{i,j}(p)D(\alpha) = D(\alpha)T_{i,j}(\frac{1}{\alpha_i}p\alpha_j)$.

In addition, every unimodular matrix U is obtained by such a product: its decomposition yields $VU = I$ with V finite product of the $T_{i,j}(p)$'s and a diagonal matrix. Thus, since the inverse of any $T_{i,j}(p)$ is of the same form, namely $T_{i,j}(-p)$, and since the inverse of a diagonal matrix is diagonal, it results that $V^{-1} = U$ is a product of elementary matrices of the same form, which proves the assertion. Moreover, if U has degree K with respect to $\frac{d}{dt}$, *i.e.* K is the maximum polynomial degree of the entries of U, then it can be proved that V has at most degree $K(\nu - 1)$ (see Ritt [1935], or, in a more general context, Ollivier [1990] using the Jacobi bound of Kondratieva et al. [1982]. The interested reader may also refer to Ollivier and Brahim [2007]).

Going back to the algorithm of decomposition of the matrix A, it consists first in permuting columns (resp. rows) to put the element of lowest degree in upper left position, denoted by $a_{1,1}$, or creating this element by Euclidean division of two or more elements of the first row (resp. column) by suitable right actions (resp. left actions). Then right divide all the other elements $a_{1,k}$ (resp. left divide the $a_{k,1}$) of the new first row (resp. first column) by $a_{1,1}$. If one of the rests is non zero, say $r_{1,k}$ (resp. $r_{k,1}$), subtract the corresponding column (resp. row) to the first column (resp. row) right (resp. left) multiplied by the corresponding quotient $q_{1,k}$ defined by the right Euclidean division $a_{1,k} = a_{1,1}q_{1,k} + r_{1,k}$ (resp. $q_{k,1}$ defined by $a_{k,1} = q_{k,1}a_{1,1} + r_{k,1}$). Then right multiplying all the columns by the corresponding quotients $q_{1,k}$, $k = 2, \ldots, \nu$ (resp. left multiplying rows by $q_{k,1}$, $k = 2, \ldots, \mu$), we iterate this process with the transformed first row (resp. first column) until it becomes $(a_{1,1}, 0, \ldots, 0)$ (resp. $(a_{1,1}, 0, \ldots, 0)^T$ where ^{T} means transposition). We then apply the same algorithm to the second row starting from $a_{2,2}$ and so on. To each transformation of rows and columns correspond a left or right elementary unimodular matrix and the unimodular matrix V (resp. U) is finally obtained as the product of all left (resp. right) elementary unimodular matrices so constructed.

6.5.5 Flatness Necessary and Sufficient Conditions

Let us outline our approach: assume that the system (6.32) is flat at some $\overline{x}_0 \in \mathfrak{X}_0$. According to (6.51), the $n \times m$ polynomial matrix $P(\varphi_0)$ maps the vector 1-form dy, differential of a flat output y, to the kernel of $P(F)$, namely $P(F)P(\varphi_0)dy = 0$, and has a "Lie-Bäcklund inverse" of size $n \times m$, $P(\psi_0)$, such that $dy = P(\psi_0)dx$, with $P(\psi_0)P(\varphi_0) = I_m$, the identity matrix of size m. Therefore, the rows of $P(\psi_0)dx$ generate an ideal of 1-forms, denoted by Ω, that contains the integrable basis $\{dy_1, \ldots, dy_m\}$. To summarize, $P(\varphi_0)$ is solution of the following polynomial matrix equation, in the unknown Θ,

$$P(F)\Theta = 0, \tag{6.53}$$

$P(\psi_0)$ is solution of the following polynomial matrix equation, in the unknown Q,

$$Q\Theta = I_m, \tag{6.54}$$

and the ideal Ω generated by the rows of Qdx, and which is the same for all solution Q of (6.54), contains an integrable basis. More precisely, we introduce the definition:

Definition 6.3. if $\tau_1, \ldots, \tau_r$ are r independent 1-forms, the $\mathfrak{K}[\frac{d}{dt}]$-ideal generated by $\tau_1, \ldots, \tau_r$ is the set made of the combinations with coefficients in $\mathfrak{K}[\frac{d}{dt}]$ of the forms $\eta \wedge \tau_i$ with η arbitrary form of arbitrary degree on $\mathfrak{X}_0$ and $i = 1, \ldots, r$.
We say that this ideal is *strongly closed* if and only if there exists $M \in \mathfrak{U}_r[\frac{d}{dt}]$ such that $d(M\tau) = 0$ with $\tau = (\tau_1, \ldots, \tau_r)^T$.

We now prove the following:

Theorem 6.6. *Assume that $P(F)$ is hyper-regular in a neighborhood of the point $\overline{x}_0 \in \mathfrak{X}_0$.*

1. *Every hyper-regular solution Θ of size $n \times m$ of (6.53) is given by*

$$\Theta = U \begin{pmatrix} 0_{n-m,m} \\ I_m \end{pmatrix} W = \hat{U} W \tag{6.55}$$

   with $U \in \mathsf{R-Smith}\,(P(F))$, $\hat{U} = U \begin{pmatrix} 0_{n-m,m} \\ I_m \end{pmatrix}$, and arbitrary $W \in \mathfrak{U}_m[\frac{d}{dt}]$.

2. *There exists $Q \in \mathsf{L-Smith}\left(\hat{U}\right)$ of size $n \times n$ and $Z \in \mathfrak{U}_m[\frac{d}{dt}]$ such that*

$$Q \cdot \Theta = \begin{pmatrix} I_m \\ 0_{n-m,m} \end{pmatrix} Z \tag{6.56}$$

   and that the submatrix $\hat{Q} = (0_{n-m,m}, I_{n-m})\,Q$ is equivalent to $P(F)$, i.e. $\exists L \in \mathfrak{U}_{n-m}[\frac{d}{dt}]$ such tha $P(F) = L\hat{Q}$.

3. *A necessary and sufficient condition for the system (6.32) to be flat at* $\overline{x}_0 \in \mathfrak{X}_0$ *is that the* $\mathfrak{K}[\frac{d}{dt}]$*-ideal* Ω*, generated by the 1-forms* $\omega_1, \ldots, \omega_m$ *defined by*

$$\omega = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_m \end{pmatrix} = (I_m, 0_{m,n-m})\, Qdx, \tag{6.57}$$

is strongly closed in a neighborhood of $\overline{x}_0 \in \mathfrak{X}_0$*. Moreover, a flat output is given by the integration of* $dy = M\omega$*, with* M *such that* $d(M\omega) = 0$*.*

Proof. The point **1** is a direct consequence of the Smith decomposition of $P(F)$: since $P(F)$ is hyper-regular, there exist $V \in \mathsf{L-Smith}\,(P(F))$ and $U \in \mathsf{R-Smith}\,(P(F))$ such that $V \cdot P(F) \cdot U = (I_{n-m}, 0_{n-m,m})$, thus $V \cdot P(F) \cdot U \cdot \begin{pmatrix} 0_{n-m,m} \\ I_m \end{pmatrix} W = 0$ for every matrix $W \in \mathfrak{U}[\frac{d}{dt}]$, which proves (6.55).

The point **2** is obtained similarly by decomposing $\hat{U}$, who is hyper-regular as the product of a unimodular matrix by an hyper-regular one: let thus $Q \in \mathsf{L-Smith}\left(\hat{U}\right)$ and $R \in \mathsf{R-Smith}\left(\hat{U}\right)$. be such that $Q\hat{U}R = \begin{pmatrix} I_m \\ 0_{n-m,m} \end{pmatrix}$. Since $\Theta = \hat{U}W$, thus $Q\Theta = Q\hat{U}W = Q\hat{U}R\left(R^{-1}W\right) = \begin{pmatrix} I_m \\ 0_{n-m,m} \end{pmatrix} Z$ with $Z = R^{-1}W$ unimodular since R, and therefore R^{-1}, and W are. Moreover, multiplying the latter identity by $(0_{n-m,m}, I_{n-m})$, we get $\hat{Q}\Theta = 0$ which proves, comparing with $DF\Theta = 0$, that $\hat{Q}$ and DF are equivalent.

Concerning the point **3**, let $\omega = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_m \end{pmatrix} = (I_m, 0_{m,n-m})\, Qdx$ where $dx = \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}$ and let us show that the strong closedness of Ω, ideal generated by $\{\omega_i | i = 1, \ldots, m\}$, is necessary. Assume that the system $(\mathfrak{X}, \tau_X, F)$, with $\mathfrak{X} = X \times \mathbb{R}^n_\infty$ and τ_X the associated trivial Cartan field, is flat at $(\overline{x}_0, \overline{y}_0)$ with $\overline{x}_0 \in \mathfrak{X}_0 = \{\overline{x} \in \mathfrak{X} | L^k_{\tau_X} F = 0 \ \forall k \geq 0\}$ and $\overline{y}_0 \in \mathbb{R}^m_\infty$. There exists a Lie-Bäcklund isomorphism Φ from $\mathbb{R}^m_\infty$ to $\mathfrak{X}_0$, with inverse Ψ, satisfying (6.51). The matrix $P(\varphi_0)$ is necessarily hyper-regular: since its size is $n \times m$, its Smith decomposition is equal to $\begin{pmatrix} \Delta_m \\ 0_{n-m,m} \end{pmatrix}$. Assume that Δ_m contains polynomial terms of degree ≥ 1. There would exist $w \neq 0$, linear combination of dy_i, $i = 1, \ldots, m$, such that $\Delta_m w = 0$, since the latter is a linear differential equation that always has a local non zero integral curve, which would imply that there exists a differential equation relating the dy_i's who are independent by definition, which is absurd, and proves that $P(\varphi_0)$ is hyper-regular. According to (6.55) and (6.76), one can take $\Theta = P(\varphi_0)$. According to **2**, we have

$Q \cdot P(\varphi_0) = \begin{pmatrix} I_m \\ 0_{n-m,m} \end{pmatrix} Z$. Left-multiplying this identity by $(I_m, 0_{m,n-m})$ and setting $\tilde{Q} = (I_m, 0_{m,n-m})Q$, yields: $\tilde{Q} \cdot P(\varphi_0) = Z$. Moreover, since $x = \varphi_0(\overline{y})$, taking the differentials of both sides, we get $dx = P(\varphi_0)dy$. We easily verify that $\omega = \tilde{Q}dx = \tilde{Q} \cdot P(\varphi_0)dy = Zdy$, or in other words, $dy = Z^{-1}\omega$. Taking the exterior derivative of this last expression, since $d^2y = 0$, we get $d(Z^{-1}\omega) = 0$, which proves that the $\mathfrak{K}[\frac{d}{dt}]$-ideal Ω is strongly closed.

Conversely, assume that $U \in \mathsf{R-Smith}(P(F))$, $Q \in \mathsf{L-Smith}\left(\hat{U}\right)$ and that Ω, generated by $\omega = \tilde{Q}dx$, is strongly closed in a neighborhood V_0 of $\overline{x}_0 \in \mathfrak{X}_0$. Let $M \in \mathfrak{U}_m[\frac{d}{dt}]$ be such that $d(M\omega) = 0$ and set $\eta = M\omega$. The m 1-forms $\eta_1, \ldots, \eta_m$ are independent in V_0 and generate Ω. Since these 1-forms depend only on a finite number of derivatives of x, in the corresponding finite dimensional jet manifold, since $d\eta_j = 0$ for all $j = 1, \ldots, m$, according to Poincaré's Lemma, there exists a mapping $\psi_0 \in C^\infty(\mathfrak{X}_0; \mathbb{R}^m)$ such that $d\psi_0 = \eta = M\omega = M\tilde{Q}dx$. Moreover ψ_0 is meromorphic in V_0 since its differential is. Setting $y = \psi_0(\overline{x})$ for all $\overline{x} \in V_0$, and $\Psi = \left(\psi_0, \frac{d\psi_0}{dt}, \frac{d^2\psi_0}{dt^2}, \ldots\right) = (\psi_0, \psi_1, \psi_2, \ldots)$, it remains to prove that Ψ is the desired Lie-Bäcklund isomorphism.

Since $Q \in \mathsf{L-Smith}\left(\hat{U}\right)$, noting as before $Q = \begin{pmatrix} \tilde{Q} \\ \hat{Q} \end{pmatrix}$ with $\tilde{Q} = (I_m, 0_{m,n-m})Q$ and $\hat{Q} = (0_{n-m,m}, I_{n-m})\,Q$, let $R \in \mathsf{R-Smith}\left(\hat{U}\right)$ be such that $\tilde{Q}\hat{U}R = I_m$ and $\hat{Q}\hat{U}R = 0_{n-m,m}$. We set $W = RM^{-1}$. Thus $\tilde{Q}\hat{U}W = M^{-1}$ and $\hat{Q}\hat{U}W = 0_{n-m,m}$, which, combined with $d\psi_0 = M\tilde{Q}dx$, yield $\tilde{Q}\hat{U}Wd\psi_0 = \tilde{Q}dx$ and $\hat{Q}\hat{U}Wd\psi_0 = 0_{n-m}$. Thus, there exists $dz \in \ker(\tilde{Q})$ such that $\hat{U}Wd\psi_0 = dx + dz$. But, since $\hat{Q}$ is equivalent to $P(F)$, we get $\hat{Q}dx = LP(F)dx = 0$ on $\mathfrak{X}_0$ and thus $\hat{Q}\hat{U}Wd\psi_0 = 0 = \hat{Q}dx + \hat{Q}dz = \hat{Q}dz$, which implies that $dz \in \ker(\hat{Q}) \cap \ker(\tilde{Q}) = \{0\}$. We have thus proven that $dx = \hat{U}Wd\psi_0$ with $W = RM^{-1}$. Denoting by σ_i the highest polynomial degree in $\frac{d}{dt}$ of the ith column of $\hat{U}W$. We must have $n \leq m + \sigma_1 + \ldots + \sigma_m$ to guarantee that the matrix $\hat{U}W$, considered as the matrix Ξ whose entries are $(\hat{U}W)^k_{i,j}$, the index k corresponding to the kth order term in $\frac{d}{dt}$ of the polynomial $(\hat{U}W)_{i,j}$, and that maps $\mathbb{R}^{\sigma_1+1} \times \ldots \times \mathbb{R}^{\sigma_m+1}$ to $\mathbb{R}^n$, is onto. Furthermore, denoting by $\sigma = \max(\sigma_i | i = 1, \ldots, m)$ and $\overline{y}^\sigma = \left(y_1^{(0)}, \ldots, y_1^{(\sigma_1)}, \ldots, y_m^{(0)}, \ldots, y_m^{(\sigma_m)}\right)$, we have $\Xi = \frac{\partial x}{\partial \overline{y}^\sigma}$, and $\text{rank}\,(\Xi) = n$. Thus, noting $\psi_j = \frac{d^j\psi_0}{dt^j}$ for all j, the implicit system

$$\begin{aligned} y &= \psi_0(\overline{x}) \\ \dot{y} &= \psi_1(\overline{x}) \\ &\vdots \\ y^{(\sigma)} &= \psi_\sigma(\overline{x}) \end{aligned}$$

has rank n with respect to x since its Jacobian matrix is the pseudo-inverse of Ξ. According to the implicit function Theorem, there exists a local solution given by $x = \varphi_0(y, \ldots, y^{(\sigma)}, \dot{x}, \ldots, x^{(\rho)})$ for some suitably chosen integer ρ. But, taking the differential of φ_0, and using the fact that dF and all its derivatives vanish on $\mathfrak{X}_0$, and thus $P(F)dx = 0$, comparing to (6.53), with $\Theta = P(\varphi_0)$, we find that φ_0 is independent of $(\dot{x}, \ldots, x^{(\rho)})$, or $x = \varphi_0(\overline{y})$. Thus the inverse of $\Phi = (\varphi_0, \frac{d}{dt}\varphi_0, \ldots)$ is Ψ, and y so obtained is clearly a flat output, which completes the proof.

Corollary 6.3. *The vector 1-form ω, defined by (6.57), is a flat output of the variational system $P(F)dx = 0$.*

Proof. Assume that dx satisfies $P(F)dx = 0$. Denoting by $\tilde{Q} = (I_m, 0_{m,n-m})\,Q$, (6.57) reads $\omega = \tilde{Q}dx$. By Theorem 6.6, point 2, since $\tilde{Q}\hat{U}R = I_m$ for some $R \in \mathfrak{U}_m[\frac{d}{dt}]$, we have $\tilde{Q}\left(dx - \hat{U}R\omega\right) = 0$, which proves that there exists a 1-form $\zeta \in \ker \tilde{Q}$ such that $dx = \hat{U}R\omega + \zeta$. But, again by Theorem 6.6, point 2, we have $P(F) = L\hat{Q}$ with $L \in \mathfrak{U}_{n-m}[\frac{d}{dt}]$ and $\hat{Q} = (0_{n-m,m}, I_{n-m})\,Q$, which yields that $0 = L^{-1}P(F)dx = \hat{Q}dx = \hat{Q}\hat{U}R\omega + \hat{Q}\zeta$. Since $\hat{Q}\hat{U} = 0$, we immediately get that $\zeta \in \ker \tilde{Q} \cap \ker \hat{Q} = \{0\}$ and $dx = \hat{U}R\omega$. Therefore, dx can be expressed as a function of ω and successive derivatives, and ω may be expressed as a function of dx and successive derivatives.

6.5.6 The Operator $\mathfrak{d}$

We now want to obtain a necessary and sufficient condition for strong closedness of the $\mathfrak{K}[\frac{d}{dt}]$-ideal Ω. However, we first need to introduce some new tools to establish a Leibnitz-like formula for the exterior derivative $d(M\tau)$ of the product of a matrix $M \in \mathfrak{U}_m[\frac{d}{dt}]$ with a vector 1-form τ for the following reason:

If M, in Definition 6.3, is a matrix on $\mathfrak{K}$, *i.e.* with entries that are meromorphic functions, or, otherwise stated, 0-degree polynomials w.r. t. $\frac{d}{dt}$, the Leibnitz formula holds true

$$d(M\tau) = dM \wedge \tau + Md\tau \tag{6.58}$$

where dM is the matrix of 1-forms whose entries are the exterior derivatives of the entries of M. But, since we consider here polynomial matrices M (*i.e.* matrices on $\mathfrak{K}[\frac{d}{dt}]$), their exterior derivatives remain to be defined.

We thus introduce a new operator $\mathfrak{d}$, extending the exterior derivative operator d to polynomial matrices, such that the Leibnitz formula (6.58) is satisfied.

For this purpose, let us introduce some notations. First, let us denote by $\Lambda^p(\mathfrak{X})$ the module of all the p-forms on $\mathfrak{X}$, by $(\Lambda^p(\mathfrak{X}))^m$ the space of all the m-dimensional vector p-forms on $\mathfrak{X}$, by $(\Lambda(\mathfrak{X}))^m$ the space of all the m-dimensional vector forms of arbitrary degree on $\mathfrak{X}$, and by $\mathcal{L}_q\left((\Lambda(\mathfrak{X}))^m\right) = \mathcal{L}\left((\Lambda^p(\mathfrak{X}))^m, (\Lambda^{p+q}(\mathfrak{X}))^m, p \geq 1\right)$, the space of all linear operators from $(\Lambda^p(\mathfrak{X}))^m$ to $(\Lambda^{p+q}(\mathfrak{X}))^m$ for all $p \geq 1$, where $\mathcal{L}(\mathcal{P}, \mathcal{Q})$ denotes the set of linear mappings from a given space $\mathcal{P}$ to a given space $\mathcal{Q}$.

We define, as announced, the operator $\mathfrak{d}$ by:

$$\mathfrak{d}(H)\kappa = d(H\kappa) - Hd\kappa \tag{6.59}$$

for all m-dimensional vector p-form κ in $(\Lambda^p(\mathfrak{X}))^m$ and all $p \geq 1$. Therefore (6.59) uniquely defines $\mathfrak{d}(H)$ as an element of $\mathcal{L}_1\left((\Lambda(\mathfrak{X}))^m\right)$.

We can prolong $\mathfrak{d}$ for all $\mu \in \mathcal{L}_q\left((\Lambda(\mathfrak{X}))^m\right)$ and for all $\kappa \in (\Lambda^p(\mathfrak{X}))^m$ and all $p \geq 1$ by the formula:

$$\mathfrak{d}(\mu)\kappa = d(\mu\ \kappa) - (-1)^q \mu\ d\kappa. \tag{6.60}$$

To interpret the latter formula, we remark that an arbitrary element $\mu \in \mathcal{L}_q\left((\Lambda(\mathfrak{X}))^m\right)$ is an $m \times m$ matrix whose (i,j)-th entry reads

$$\mu_{i,j} = \sum_{k \geq 0} \mu_{i,j,k} \wedge \frac{d^k}{dt^k}, \quad i,j = 1,2,\ldots,m \tag{6.61}$$

where $\mu_{i,j,k} \in \Lambda^q(\mathfrak{X})$ for every $i,j = 1,\ldots,m$, $k \geq 0$, and that, for every $\kappa \in (\Lambda^p(\mathfrak{X}))^m$, the ith component of $\mu\kappa$ is given by

$$(\mu\kappa)_i = \sum_{k \geq 0} \sum_{j=1}^m \mu_{i,j,k} \wedge L^k_{\tau_{\mathfrak{X}}} \kappa_j$$

which is a $(p+q)$-form for every $i = 1,\ldots,m$.

It is thus straightforward to check that the (i,j)th entry $\mathfrak{d}(\mu)_{i,j}$ of $\mathfrak{d}(\mu)$ is given by

$$\mathfrak{d}(\mu)_{i,j} = \sum_{k \geq 0} d\mu_{i,j,k} \wedge \frac{d^k}{dt^k}. \tag{6.62}$$

Indeed, by (6.60), we have, for every $i = 1,\ldots,m$,

$$\begin{aligned}
(\mathfrak{d}(\mu)\kappa)_i &= d\left(\sum_{k \geq 0} \sum_{j=1}^m \mu_{i,j,k} \wedge L^k_{\tau_{\mathfrak{X}}} \kappa_j\right) - (-1)^q \left(\sum_{k \geq 0} \sum_{j=1}^m \mu_{i,j,k} \wedge L^k_{\tau_{\mathfrak{X}}} d\kappa_j\right) \\
&= \left(\sum_{k \geq 0} \sum_{j=1}^m d\mu_{i,j,k} \wedge L^k_{\tau_{\mathfrak{X}}} \kappa_j + (-1)^q \sum_{k \geq 0} \sum_{j=1}^m \mu_{i,j,k} \wedge L^k_{\tau_{\mathfrak{X}}} d\kappa_j\right) \\
&\quad -(-1)^q \left(\sum_{k \geq 0} \sum_{j=1}^m \mu_{i,j,k} \wedge L^k_{\tau_{\mathfrak{X}}} d\kappa_j\right)
\end{aligned}$$

hence the result.

The operator $\mathfrak{d}$ enjoys the following properties:

Proposition 6.4. *For all* $\mu \in \mathcal{L}_q\left((\Lambda(\mathfrak{X}))^m\right)$*, all* $q \in \mathbb{N}$ *and all* $\kappa \in (\Lambda^p(\mathfrak{X}))^m$*, with* $p \geq 1$ *arbitrary, we have*

$$\mathfrak{d}\left(\mathfrak{d}\left(\mu\right)\right)\kappa = 0. \tag{6.63}$$

In other words $\mathfrak{d}^2 = 0$*, i.e.* $\mathfrak{d}$ *is a complex.*

Proof. According to (6.60), replacing μ by $\mathfrak{d}\left(\mu\right) \in \mathcal{L}_{q+1}\left((\Lambda(\mathfrak{X}))^m\right)$, we get

$$\mathfrak{d}\left(\mathfrak{d}\left(\mu\right)\right)\kappa = d\left(\mathfrak{d}\left(\mu\right)\kappa\right) - (-1)^{q+1}\mathfrak{d}\left(\mu\right)d\kappa. \tag{6.64}$$

Since, again with (6.60), $d\left(\mathfrak{d}\left(\mu\right)\kappa\right) = d^2\left(\mu\ \kappa\right) - (-1)^q d\left(\mu\ d\kappa\right)$ and, since $d^2 = 0$, we have

$$\begin{aligned}\mathfrak{d}\left(\mathfrak{d}\left(\mu\right)\right)\kappa &= -(-1)^q\left(\mathfrak{d}\left(\mu\right)\ d\kappa + (-1)^q\mu\ d^2\kappa\right) - (-1)^{q+1}\mathfrak{d}\left(\mu\right)\ d\kappa \\ &= -(-1)^q\mathfrak{d}\left(\mu\right)\ d\kappa + (-1)^q\mathfrak{d}\left(\mu\right)\ d\kappa = 0\end{aligned}$$

the result is proven.

We also have:

Proposition 6.5. *For all* $H \in \mathfrak{U}_m[\frac{d}{dt}]$ *and all* $\mu \in \mathcal{L}_q\left((\Lambda(\mathfrak{X}))^m\right)$*, with arbitrary* $q \geq 0$*, we have*

$$\mathfrak{d}\left(H\right)\mu + H\mathfrak{d}\left(\mu\right) = \mathfrak{d}\left(H\mu\right). \tag{6.65}$$

In particular, if $\mu = -H^{-1}\mathfrak{d}\left(H\right) \in \mathcal{L}_1\left((\Lambda(\mathfrak{X}))^m\right)$*, we have*

$$\mathfrak{d}\left(\mu\right) = \mu^2. \tag{6.66}$$

Proof. Let $H \in \mathfrak{U}_m[\frac{d}{dt}]$, $\kappa \in \left(\Lambda^p_{\mathfrak{K}}(\mathfrak{X})\right)^m$ and $\mu \in \mathcal{L}_q\left((\Lambda(\mathfrak{X}))^m\right)$, with $q \geq 0$ arbitrary. (6.59) and (6.60), yield

$$\mathfrak{d}\left(H\right)\mu\kappa = d(H\mu\kappa) - Hd(\mu\kappa), \quad d(\mu\kappa) = \mathfrak{d}\left(\mu\right)\kappa + (-1)^q\mu d\kappa$$

or

$$\mathfrak{d}\left(H\right)\mu\kappa = d(H\mu\kappa) - H\mathfrak{d}\left(\mu\right)\kappa - (-1)^q H\mu d\kappa.$$

In other words:

$$\left(\mathfrak{d}\left(H\right)\mu + H\mathfrak{d}\left(\mu\right)\ \right)\kappa = d(H\mu\kappa) - (-1)^q H\mu d\kappa = \mathfrak{d}\left(H\mu\right)\kappa.$$

This relation being valid for all $\kappa \in \left(\Lambda^p_{\mathfrak{K}}(\mathfrak{X})\right)^m$ and all $p \geq 1$, we immediately deduce (6.65).

If now $\mu = -H^{-1}\mathfrak{d}\left(H\right) \in \mathcal{L}_1\left(\left(\Lambda^p_{\mathfrak{K}}(\mathfrak{X})\right)^m\right)$, we get $-H\mu = \mathfrak{d}\left(H\right)$ and thus, according to what precedes,

$$\mathfrak{d}\,(H\mu) = \mathfrak{d}\,(H)\,\mu + H\mathfrak{d}\,(\mu) = -\mathfrak{d}^2(H) = 0$$

or $H\mathfrak{d}\,(\mu) = -\mathfrak{d}\,(H)\,\mu$, or also $\mathfrak{d}\,(\mu) = -H^{-1}\mathfrak{d}\,(H)\,\mu = \mu^2$, which achieves the proof.

6.5.7 Strong Closedness Necessary and Sufficient Conditions

Theorem 6.7. *The $\mathfrak{K}[\frac{d}{dt}]$-ideal Ω generated by the 1-forms $\omega_1, \ldots, \omega_m$ defined by (6.57) is strongly closed in $\mathfrak{X}_0$ (or, equivalently, the system $(\mathfrak{X}, \tau_{\mathfrak{X}}, F)$ is flat) if and only if there exists $\mu \in \mathcal{L}_1\left((\Lambda(\mathfrak{X}))^m\right)$, and a matrix $M \in \mathfrak{U}_m[\frac{d}{dt}]$ such that*

$$d\omega = \mu\;\omega, \qquad \mathfrak{d}\,(\mu) = \mu^2, \qquad \mathfrak{d}\,(M) = -M\mu. \tag{6.67}$$

where ω is the vector of 1-forms defined by (6.57).

In addition, if (6.67) holds true, a flat output y is obtained by integration of $dy = M\omega$.

Proof. If $d(M\omega) = 0$, according to (6.59), we have $\mathfrak{d}\,(M)\,\omega = -Md\omega$ or $d\omega = -M^{-1}\mathfrak{d}\,(M)\,\omega$. Setting $\mu = -M^{-1}\mathfrak{d}\,(M) \in \mathcal{L}_1\left((\Lambda_{\mathfrak{K}}(\mathfrak{X}))^m\right)$, which is equivalent to the last relation of (6.67), we have $d\omega = -M^{-1}\mathfrak{d}\,(M)\,\omega = \mu\;\omega$. The conditions (6.67) are thus immediately deduced from (6.66).

Conversely, from the last identity of (6.67), we get $-M^{-1}\mathfrak{d}\,(M) = \mu$. Its combination with the first one yields $d\omega = -M^{-1}\mathfrak{d}\,(M)\,\omega$, or $Md\omega = -\mathfrak{d}\,(M)\;\omega$ and, according to (6.59), we immediately get that $d(M\omega) = 0$, i.e. Ω is strongly closed.

Finally, if there exists a matrix $M \in \mathfrak{U}_m[\frac{d}{dt}]$ such that $d(M\omega) = 0$, By Poincaré's Lemma, there exist m functions $y_1, \ldots, y_m$ such that $dy = M\omega$, which achieves the proof.

Remark 6.4. Condition (6.67) may be seen as a generalization in the framework of manifolds of jets of infinite order of the well-known moving frame structure equations (see e.g. Chern et al. [2000]).

Proposition 6.6. *The differential system*

$$d\omega = \mu\;\omega, \qquad \mathfrak{d}\,(\mu) = \mu^2 \tag{6.68}$$

always admits a solution $\mu \in \mathcal{L}_1\left((\Lambda(\mathfrak{X}))^m\right)$.

Proof. μ may be represented as an $m \times m$ matrix whose (i,j)th entry has the form

$$\mu_{i,j} = \sum_{k=0}^{K} \mu_{i,j}^k \wedge \frac{d^k}{dt^k}$$

with $\mu_{i,j}^k \in \Lambda^1(\mathfrak{X})$ for $i,j = 1, \ldots, m$ and $k = 0, \ldots, K$ for some finite K.

Thus, $d\omega = \mu\omega$ reads

$$d\omega_i = \sum_{j=1}^{m}\sum_{k=0}^{K} \mu_{i,j}^k \wedge \omega_j^{(k)}, \quad i = 1, \ldots, m \tag{6.69}$$

and the (i,j)th entry of $\mu^2 = \mu\mu$ is given by

$$(\mu^2)_{i,j} = \sum_{l=1}^{m}\sum_{k=0}^{K}\sum_{r=0}^{K}\sum_{s=1}^{k+r} \begin{pmatrix} k \\ k+r-s \end{pmatrix} \mu_{i,l}^k \wedge \left(\mu_{l,j}^r\right)^{(k+r-s)} \wedge \frac{d^s}{dt^s}$$

with the notation $\begin{pmatrix} p \\ q \end{pmatrix} = \frac{p!}{q!(p-q)!}$. On the other hand, using (6.62), $\mathfrak{d}(\mu)$ is the $m \times m$ matrix whose (i,j)th entry is

$$\mathfrak{d}(\mu)_{i,j} = \sum_{k=0}^{K} d\mu_{i,j}^k \wedge \frac{d^k}{dt^k}$$

where $d\mu_{i,j}^k$ is the "ordinary" exterior derivative of the 1-form $\mu_{i,j}^k$. Therefore, the second equation of (6.68) reads, after identification of the polynomial terms on both sides:

$$\begin{aligned} d\mu_{i,j}^k &= \sum_{l=1}^{m}\sum_{s=0}^{K}\sum_{r=\max(k-s,0)}^{k} \begin{pmatrix} s \\ s+r-k \end{pmatrix} \mu_{i,l}^s \wedge \left(\mu_{l,j}^r\right)^{(s+r-k)}, \\ &\qquad i,j = 1,\ldots,m,\ k = 0,\ldots,K \\ 0 &= \sum_{l=1}^{m}\sum_{s=k-K}^{K}\sum_{r=k-s}^{K} \begin{pmatrix} s \\ s+r-k \end{pmatrix} \mu_{i,l}^s \wedge \left(\mu_{l,j}^r\right)^{(s+r-k)}, \\ &\qquad i,j = 1,\ldots,m,\ k = K+1,\ldots,2K \end{aligned} \tag{6.70}$$

which proves that the system (6.68) is equivalent to (6.69)-(6.70).

This differential system is algebraically closed: applying the operator $\mathfrak{d}$ to the right-hand side of both equations of (6.68), we indeed obtain $\mathfrak{d}(\mu\omega) = d(\mu\omega) = \mathfrak{d}(\mu)\omega - \mu d\omega = \mu^2\omega - \mu^2\omega = 0$, and $\mathfrak{d}(\mu^2) = \mathfrak{d}(\mu)\mu - \mu\mathfrak{d}(\mu) = \mu^3 - \mu^3 = 0$. Expressing the polynomial coefficients of the latter expressions and using the fact that d and $\frac{d}{dt}$ commute, they turn out to be identically equal to the exterior differentials of the right-hand sides of (6.69)-(6.70), which are thus equal to 0. This completes the proof of the algebraic closedness of (6.69)-(6.70).

It remains to precise the coordinate system in which it is expressed. Let us remark, on the one hand, that the coordinates of $\mathrm{T}^*_{\overline{x}}\mathfrak{X}$ are such that

$$\sum_{j=1}^{n} \frac{d^k}{dt^k}\left(\frac{\partial F_i}{\partial x_j} dx_j + \frac{\partial F_i}{\partial \dot{x}_j} d\dot{x}_j\right) = 0, \quad i = 1, \ldots, n-m, \quad k \geq 0. \tag{6.71}$$

Thus, by the implicit function theorem and up to a permutation of $\{1,\ldots,n\}$, the set $\{dx_1,\ldots,dx_n,d\dot{x}_1,\ldots,d\dot{x}_m,\ldots,dx_1^{(k)},\ldots,dx_m^{(k)}\ldots\}$ forms a basis of $\mathrm{T}^*_{\overline{x}}\mathfrak{X}$.
On the other hand, since $\omega=\tilde{Q}dx$ (using the notations of Corollary 6.3) for dx satisfying (6.71), the ith component ω_i of ω has the form

$$\omega_i=\sum_{j=1}^{n}\omega_{i,j}^0 dx_j+\sum_{j=1}^{m}\sum_{k=1}^{K_\omega}\omega_{i,j}^k dx_j^{(k)}$$

for some finite integer K_ω and where the $\omega_{i,j}^k$'s are meromorphic functions of $(x_1,\ldots,x_n,\ldots,x_1^{(\alpha)},\ldots,x_n^{(\alpha)})$ for some α. Thus $d\omega_i$ is a 2-form that, after substitution of the $dx_i^{(k)}$ for $i=m+1,\ldots,n$ and $k=1,\ldots,\alpha$ using (6.71), is a combination of $dx_i\wedge dx_j,\ i<j,\ dx_i\wedge dx_j^{(k)}, i=1,\ldots,n,\ j=1,\ldots,m,\ k=1,\ldots,\max(\alpha,K_\omega)$, and $dx_i^{(k)}\wedge dx_j^{(l)},\ i,j=1,\ldots,m,\ k,l=1,\ldots,\max(\alpha,K_\omega),\ k\neq l$ whose coefficients are meromorphic functions of the coordinates $(x_1,\ldots,x_n,\ldots,x_1^{(\alpha)},\ldots,x_n^{(\alpha)})$. According to (6.69)-(6.70), it results that there exists some finite integer $\overline{K}$ such that all the coefficients of the $\mu_{i,j}^k$'s and $d\mu_{i,j}^k$'s may be considered as meromorphic functions of $(x,\ldots,x^{(\overline{K})})$. Identifying in (6.69)-(6.70) the coefficients of all the monomial forms $dx_i^{(k)}\wedge dx_j^{(l)}$, we obtain a set of first order partial differential equations (PDE's) in the unknown functions $\mu_{i,j}^k,\ i,j=1,\ldots,m,\ k=0,\ldots,K$ of the finite jet coordinates $(x,\ldots,x^{(\overline{K})})$. This equivalent differential system (6.69)-(6.70) being algebraically closed, according to Frobenius Theorem, the system of PDE's admits a local solution in an open dense subset of the jet manifold of coordinates $(x,\ldots,x^{(\overline{K})})$.

Corollary 6.4. *The differential system (6.67) always admits a solution* $\mu\in\mathcal{L}_1\left((\Lambda(\mathfrak{X}))^m\right)$ *and* $M\in\mathcal{M}_{m,m}[\frac{d}{dt}]$.

Proof. The proof follows the same lines as Proposition 6.6, the only difference relying on the closedness of the associated exterior system. The latter property follows the fact that, applying the operator $\mathfrak{d}$ to the right-hand side of $\mathfrak{d}(M)=-M\mu$, we get $-\mathfrak{d}(M)\mu-M\mathfrak{d}(\mu)=M\mu^2-M\mu^2=0$ and the result is proven.

Remark 6.5. In the previous Corollary, the existence of $M\in\mathcal{M}_{m,m}[\frac{d}{dt}]$ is not sufficient for flatness since it doesn't guarantee that M is unimodular. Therefore, the only restrictive condition in Theorem 6.7 is that $M\in\mathcal{U}_m[\frac{d}{dt}]$.

Note that the conditions (6.67) provide an effective algorithm to compute the matrix M and hence a flat output:

Algorithm.

1. We first compute a vector 1-form ω defined by (6.57).
2. We compute the operator μ such that $d\omega = \mu\omega$ by componentwise identification. Such μ always exist in virtue of Corollary 6.3: the ω_j's are naturally obtained as combinations of the dx_i's and derivatives. By exterior differentiation, every component $d\omega_j$ of $d\omega$ can also be expressed in terms of the dx_i's and derivatives. Since the dx_i's satisfying the variational system can be expressed in terms of the ω_j's and derivatives, it results that the $d\omega_j$'s are combinations of the ω_j's and derivatives, hence the result.
3. Among the possible μ's, only those satisfying $\mathfrak{d}(\mu) = \mu^2$ are kept. Such μ's always exist by Proposition 6.6.
4. We then compute M such that $\mathfrak{d}(M) = -M\mu$, still by componentwise identification.
5. Finally, only those matrices M which are unimodular are kept. If there are no such M, the system is non flat. In the opposite case, a flat output is obtained by the integration of $dy = M\omega$ which is possible since $d(M\omega) = 0$.

Remark 6.6. A **non flat system** is characterized by $d(M\omega) \neq 0$ for all $M \in \mathfrak{U}_m[\frac{d}{dt}]$. In this case, there exists a non zero m-dimensional vector 2-form $\tau(M)$ such that $d(M\omega) = \tau(M)$ for all $M \in \mathfrak{U}_m[\frac{d}{dt}]$. Setting $\varpi = M^{-1}\tau(M)$, the 2-form ϖ is clearly uniquely defined modulo $(\Omega, d\Omega)$: indeed, if M_1 and M_2 are in $\mathfrak{U}_m[\frac{d}{dt}]$ and if we denote by $\tau_i = d(M_i\omega)$, $\varpi_i = M_i^{-1}\tau_i$ and $\mu_i = -M_i^{-1}\mathfrak{d}(M_i)$, $i = 1, 2$, we have $d(M_i\omega) = \mathfrak{d}(M_i)\,\omega + M_i d\omega = \tau_i$, $i = 1, 2$, or $d\omega = M_i^{-1}(\tau_i - \mathfrak{d}(M_i)\,\omega) = \varpi_i + \mu_i\omega$, so that $\varpi_1 - \varpi_2 = (\mu_2 - \mu_1)\,\omega$, or $\varpi_1 - \varpi_2 = 0$ modulo $(\Omega, d\Omega)$, since, from (6.59), it is easily seen that $(\mu_2 - \mu_1)\,\omega$ is a combination of elements of Ω and $d\Omega$.

We thus have :

Theorem 6.8. *A system is non flat if and only if for every matrix $M \in \mathfrak{U}_m[\frac{d}{dt}]$ there exists a non zero $\varpi(M) \in \left(\Lambda^2(\mathfrak{X})\right)^m$ such that:*

$$d\omega = \mu(M)\omega + \varpi(M), \quad d\varpi(M) = \mu(M)\varpi(M), \quad \mathfrak{d}(\mu(M)) = \mu^2(M) \tag{6.72}$$

with $\mu(M) = -M^{-1}\mathfrak{d}(M)$.

Proof. Since $\mathfrak{d}(M)\,\omega + Md\omega = \tau(M) \neq 0$ for all M, we get $d\omega = M^{-1}\tau(M) + \mu(M)\omega = \varpi(M) + \mu(M)\omega$ after having noted $M^{-1}\tau(M) = \varpi(M)$, the first relation of (6.72) follows. By exterior differentiation, we get: $0 = d\varpi(M) + \mathfrak{d}(\mu(M))\,\omega - \mu(M)d\omega$. According to Proposition 6.5, we deduce that $\mathfrak{d}(\mu(M)) = \mu^2(M)$ (last relation of (6.72)) since, by definition, $\mu(M) = -M^{-1}\mathfrak{d}(M)$. Thus $0 = d\varpi(M) + \mu^2(M) - \mu(M)\,(\varpi(M) + \mu(M)\omega)$, which proves the second relation of (6.72), and (6.72) is proven.

Conversely, if (6.72) is valid with $\varpi(M) \neq 0$ for all M, we have $d(M\omega) = \mathfrak{d}(M)\,\omega + Md\omega$ or

$$d(M\omega) = -M\mu(M)\omega + M\,(\varpi(M) + \mu(M)\omega) = M\varpi(M) = \tau(M) \neq 0$$

for all $M \in \mathfrak{U}_m[\frac{d}{dt}]$, which achieves the proof.

Therefore, one may introduce the equivalence relation *"two systems are equivalent if and only if they have the same ideal Ω and their 2-forms ϖ are equal modulo $(\Omega, d\Omega)$"*, which suggests that it is possible to classify the non flat systems. In particular, a system has a non zero defect (see Fliess et al. [1999]) if and only if it admits a non zero 2-form ϖ modulo $(\Omega, d\Omega)$, that might be interpreted as a generalized curvature (as opposed to flatness).

6.5.8 Flat Outputs of Linear Controllable Systems

Theorem 6.6 indeed applies to linear systems. Moreover, in the case of linear controllable systems, we are assured that the flatness necessary and sufficient conditions are fulfilled, with one simplification: if we are looking for a *linear flat output*, *i.e.* such that x can be expressed as a linear combination of the components of y and a finite number of its derivatives, which we know always exists (see *e.g.* the Brunovský canonical form), the strong closedness condition is always satisfied since a linear combination (with constant coefficients) of the dx_i's, say $dz = \sum_{i=1}^n a_i dx_i$, is always closed (its integral is $z = \sum_{i=1}^n a_i x_i$).

Remark 6.7. Note that, even in the linear case, there exists nonlinear flat outputs *i.e.* such that $x = \Phi(\overline{y})$ with Φ nonlinear. For instance, the linear system $\dot{x}_1 = u_1, \ \dot{x}_2 = u_2$ is flat with $y_1 = x_1, \ y_2 = x_2$ as flat output, but we can choose as well $x_1 = y_1' + (y_2')^2, \ x_2 = y_2'$ since $u_1 = \dot{y}_1' + 2y_2\dot{y}_2', \ u_2 = \dot{y}_2'$ and $y_1' = x_1 - x_2^2, \ y_2' = x_2$.

Let us now specialize Theorem 6.6 in the linear case (see also Lévine and Nguyen [2003]). We are given a linear system in the form

$$\dot{x} = Ax + Bu \tag{6.73}$$

with A of size $n \times n$, B of size $n \times m$ and $\text{rank}\,(B) = m$. According to our approach, we first eliminate the input u. For this purpose, let us introduce a matrix C of size $n \times (n-m)$ with $\text{rank}\,(C) = n - m$, such that $C^T B = 0$. Left-multiplying both sides of the system equations by C^T we get

$$C^T(\dot{x} - Ax) = 0$$

which plays the role of (6.32), and in polynomial form:

$$C^T(I_n \frac{d}{dt} - A)x = 0. \tag{6.74}$$

Thus, $P(F) = C^T(I_n \frac{d}{dt} - A)$.

Theorem 6.9. *Assume that the linear system (6.73) is controllable. Every solution Θ of $P(F)\Theta = 0$ is given by*

$$\Theta = U \begin{pmatrix} 0_{n-m,m} \\ I_m \end{pmatrix} W = \hat{U} W \tag{6.75}$$

with $U \in \mathsf{R-Smith}\left(C^T(I_n \frac{d}{dt} - A)\right)$, $\hat{U} = U \begin{pmatrix} 0_{n-m,m} \\ I_m \end{pmatrix}$, *and* $W \in \mathfrak{U}_m[\frac{d}{dt}]$ *arbitrary. Moreover, let* $Q \in \mathsf{L-Smith}\left(\hat{U}\right)$ *of size* $n \times n$, *we have*

$$Q \cdot \Theta = \begin{pmatrix} I_m \\ 0_{n-m,m} \end{pmatrix} Z \tag{6.76}$$

with $Z \in \mathfrak{U}_m[\frac{d}{dt}]$, *and a flat output* y *is given by*

$$y = (I_m, 0_{m,n-m})\, Qx, \quad x = \hat{U} y.$$

Proof. Direct consequence of Theorem 6.6 and of linearity.

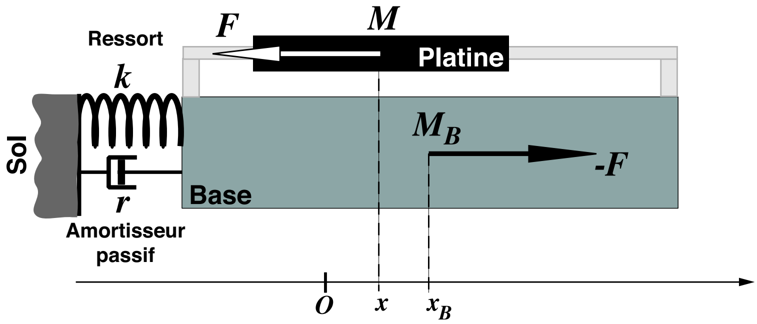

Fig. 6.6 Base-stage high-precision positionning system

Example 6.2. We consider a motorized stage[2] (see figure 6.6) of mass M moving without friction along a rail fixed to the base, whose mass is M_B, itself related to the ground in an elastic way, with stiffness coefficient k and damping r. The base is assumed to move along a parallel axis to the rails.

Let us denote by x_B the abscissa of the center of mass of the base and by x the abscissa of the center of mass of the stage in a fixed coordinate frame. We note F the force applied to the stage, delivered by the motor. F is the

[2] This application was the subject of the European Patent N. 00400435.4-2208 and US Patent N. 09/362,643, registered by Newport Corporation.

control input. The position of the center of mass of the setup stage+base is $x_G = \frac{M_B x_B + M x}{M_B + M}$. According to Newton's second principle, we have:

$$(M_B + M)\ddot{x}_G = -k x_B - r\dot{x}_B$$

or:

$$\begin{aligned} M\ddot{x} &= F \\ M_B\ddot{x}_B &= -F - k x_B - r\dot{x}_B \end{aligned} \tag{6.77}$$

which yields

$$\begin{pmatrix} M\frac{d^2}{dt^2} & 0 \\ 0 & M_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k \end{pmatrix} \begin{pmatrix} x \\ x_B \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} F.$$

We set $A[\frac{d}{dt}] = \begin{pmatrix} M\frac{d^2}{dt^2} & 0 \\ 0 & M'_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k \end{pmatrix}$ and $B = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Note that this system is made of second degree polynomials with respect to $\frac{d}{dt}$ as a consequence of the second principle, whereas the above Theorem is stated for first-order systems $(I_n\frac{d}{dt} - A)x = Bu$. However, the reader may easily verify that Theorem 6.9 holds true for systems of any degree with respect to $\frac{d}{dt}$. Moreover, this second-order form is more convenient for our computations since we are dealing with smaller matrices (n=2, whereas n=4 for the system expressed as a first-degree one).

To eliminate the input F, we compute a matrix C orthogonal to B: $C^T = \begin{pmatrix} 1 & 1 \end{pmatrix}$. The implicit system equivalent to (6.77) is thus

$$C^T A[\frac{d}{dt}]\begin{pmatrix} x \\ x_B \end{pmatrix} = \begin{pmatrix} M\frac{d^2}{dt^2} & M_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k \end{pmatrix}\begin{pmatrix} x \\ x_B \end{pmatrix} = 0. \tag{6.78}$$

We are thus looking for a polynomial matrix $\Theta = \begin{pmatrix} \Theta_1 \\ \Theta_2 \end{pmatrix}$, solution of

$$C^T A[\frac{d}{dt}]\Theta = \begin{pmatrix} M\frac{d^2}{dt^2} & M_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k \end{pmatrix}\begin{pmatrix} \Theta_1 \\ \Theta_2 \end{pmatrix} = 0 \tag{6.79}$$

or

$$M\frac{d^2}{dt^2}\Theta_1 = -\left(M_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k\right)\Theta_2.$$

We could apply the Smith decomposition of Theorem 6.9, but we leave it to the reader since, in this case, a direct computation turns out to be much simpler and gives a slightly different perspective in terms of Euclidean division.

Since the polynomials $M\frac{d^2}{dt^2}$ et $M_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k$ are mutually prime, By the Gauss divisibility Theorem, we get

$$\Theta_1 = \frac{1}{k}\left(M_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k\right), \quad \Theta_2 = -\frac{1}{k}\left(M\frac{d^2}{dt^2}\right). \tag{6.80}$$

x and x_B are thus given by $\begin{pmatrix} x \\ x_B \end{pmatrix} = \Theta y$, or

$$\begin{aligned} x &= \frac{1}{k}\left(M_B\frac{d^2}{dt^2} + r\frac{d}{dt} + k\right) y = \frac{M_B}{k}\ddot{y} + \frac{r}{k}\dot{y} + y, \\ x_B &= -\left(\frac{M}{k}\frac{d^2}{dt^2}\right) y = -\frac{M}{k}\ddot{y} \end{aligned} \tag{6.81}$$

and $F = M\left(\frac{M_B}{k}y^{(4)} + \frac{r}{k}y^{(3)} + \ddot{y}\right)$.

Finally, inverting (6.81), we get

$$y = x - \frac{r}{k}\dot{x} + \frac{1}{M}(M_B - \frac{r^2}{k})x_B - \frac{M_B r}{Mk}\dot{x}_B.$$

6.5.9 Examples

Example 6.3. Let us go back to the non holonomic vehicle:

$$F = \dot{x}\sin\theta - \dot{y}\cos\theta = 0.$$

Its differential is

$$P(F) = \begin{pmatrix} \sin\theta\frac{d}{dt} & -\cos\theta\frac{d}{dt} & \dot{x}\cos\theta + \dot{y}\sin\theta \end{pmatrix} \tag{6.82}$$

and

$$P(F)\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix} = 0.$$

The Smith decomposition calculations are as follows: at each step, the matrices on the left part of the page result from a previous computation and the matrices on the right part are the corresponding transformations. We have:

1.

$$\begin{pmatrix} \sin\theta\frac{d}{dt} & -\cos\theta\frac{d}{dt} & \dot{x}\cos\theta + \dot{y}\sin\theta \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = U_1$$

2.

$$\begin{pmatrix} \dot{x}\cos\theta + \dot{y}\sin\theta & -\cos\theta\frac{d}{dt} & \sin\theta\frac{d}{dt} \end{pmatrix} \begin{pmatrix} 1 & \frac{\cos\theta}{\dot{x}\cos\theta+\dot{y}\sin\theta}\frac{d}{dt} \\ -\frac{\sin\theta}{\dot{x}\cos\theta+\dot{y}\sin\theta}\frac{d}{dt} & \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = U_2$$

3. $\begin{pmatrix} \dot{x}\cos\theta + \dot{y}\sin\theta & 0 & 0 \end{pmatrix}$.

Thus

$$P(F)U = \begin{pmatrix} \Delta & 0 & 0 \end{pmatrix}$$

with

$$U = U_1 U_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & \frac{\cos\theta}{\dot{x}\cos\theta+\dot{y}\sin\theta}\frac{d}{dt} & -\frac{\sin\theta}{\dot{x}\cos\theta+\dot{y}\sin\theta}\frac{d}{dt} \end{pmatrix}$$

and

$$\Delta = \dot{x}\cos\theta + \dot{y}\sin\theta.$$

Note that here(n-m=1) the matrix V of the triplet (U, V, Δ) is useless and that Δ is a scalar.

We thus have $\hat{U} = U \begin{pmatrix} 0_{1,2} \\ I_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ \frac{\cos\theta}{\Delta}\frac{d}{dt} & -\frac{\sin\theta}{\Delta}\frac{d}{dt} \end{pmatrix}$ where I_2 is the identity matrix of $\mathbb{R}^2$.

The same algorithm is applied (this time on the left) to compute $Q \in \mathsf{L-Smith}\left(\hat{U}\right)$:

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ \frac{\sin\theta}{\Delta}\frac{d}{dt} & -\frac{\cos\theta}{\Delta}\frac{d}{dt} & 1 \end{pmatrix}.$$

Let us multiply Q by $\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix}$. The last row is $\frac{1}{\Delta}(\sin\theta d\dot{x} - \cos\theta d\dot{y} + (\dot{x}\cos\theta + \dot{y}\sin\theta)d\theta) = \frac{1}{\Delta}d(\dot{x}\sin\theta - \dot{y}\cos\theta)$ and is precisely the system itself, thus identically 0 on $\mathfrak{X}_0$.

The two first rows of $Q\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix}$ are $\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix} = \begin{pmatrix} dy \\ dx \end{pmatrix}$. The ideal Ω is thus generated by (dx, dy) and is trivially strongly closed, choosing the matrix $M = I_2$. We thus find the flat output components $y_1 = y$ and $y_2 = x$, already found in section 6.2.4.

Example 6.4. The non holonomic car (continued).

Other decompositions of $P(F)$, given by (6.82), may indeed be obtained, but they are all equivalent in the sense that one decomposition may be deduced from another one by multiplication by a unimodular matrix. However, the resulting vector 1-form ω, contrarily to what happens in the previous example, may not be integrable. Our aim is here to show how the generalized moving frame structure equations (6.67) may be used to obtain an integrable $M\omega$. Such an example is provided by restarting the right-Smith decomposition of $P(F)$ by right-multiplying it by $\begin{pmatrix} \cos\theta & 0 & 0 \\ \sin\theta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and using the formula $\sin\theta \frac{d}{dt}(\cos\theta) - \cos\theta \frac{d}{dt}(\sin\theta) = -\dot{\theta}$, we obtain

$$U = \begin{pmatrix} \cos\theta & -\frac{1}{\dot\theta}\cos^2\theta \frac{d}{dt} & \frac{1}{\dot\theta}(\dot x\cos\theta + \dot y\sin\theta)\cos\theta \\ \sin\theta & 1 - \frac{1}{\dot\theta}\sin\theta\cos\theta\frac{d}{dt} & \frac{1}{\dot\theta}(\dot x\cos\theta + \dot y\sin\theta)\sin\theta \\ 0 & 0 & 1 \end{pmatrix},$$

$$\hat U = \begin{pmatrix} -\frac{1}{\dot\theta}\cos^2\theta \frac{d}{dt} & \frac{1}{\dot\theta}(\dot x\cos\theta + \dot y\sin\theta)\cos\theta \\ 1 - \frac{1}{\dot\theta}\sin\theta\cos\theta\frac{d}{dt} & \frac{1}{\dot\theta}(\dot x\cos\theta + \dot y\sin\theta)\sin\theta \\ 0 & 1 \end{pmatrix}$$

and then $Q \in \mathsf{L} - \mathsf{Smith}\left(\hat U\right)$:

$$Q = \begin{pmatrix} -\tan\theta & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\dot\theta}\sin\theta\cos\theta\frac{d}{dt} & -\frac{1}{\dot\theta}\cos^2\theta\frac{d}{dt} & -\frac{1}{\dot\theta}(\dot x\cos\theta + \dot y\sin\theta)\cos\theta \end{pmatrix}.$$

The vector 1-form $\omega = (\omega_1, \omega_2)^T$ is obtained by multiplying the two first lines of Q by $(dx, dy, d\theta)^T$: $\omega_1 = -\tan\theta dx + dy$ and $\omega_2 = d\theta$. We have $d\omega_1 = -\frac{1}{\cos^2\theta} d\theta \wedge dx$ and $d\omega_2 = 0$. Hence, in order that $d\omega = \mu\omega$, a possible choice is $\mu = \begin{pmatrix} 0 & \left(\frac{1}{\cos^2\theta}dx + \eta d\theta\right)\wedge \\ 0 & 0 \end{pmatrix}$ where η is an arbitrary meromorphic function. One can verify that μ is admissible if $\eta = -2\frac{x\sin\theta}{\cos^3\theta}$ to ensure that $\mathfrak{d}(\mu) = 0 = \mu^2$. Again, by componentwise identification, one finds $M = \begin{pmatrix} 1 & -\frac{x}{\cos^2\theta} \\ 0 & 1 \end{pmatrix}$ and we have

$$M\omega = \begin{pmatrix} -\tan\theta dx - \frac{x}{\cos^2\theta} d\theta + dy \\ d\theta \end{pmatrix}, \quad d(M\omega) = 0.$$

Thus, setting $\begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix} = M\omega$,

$$y_1 = y - x\tan\theta, \quad y_2 = \theta$$

which is another possible flat output: it is easily checked that the inverse L-B isomorphism is given by $x = -\frac{\dot{y}_1}{\dot{y}_2}\cos^2 y_2$, $y = y_1 - \frac{\dot{y}_1}{\dot{y}_2}\sin y_2 \cos y_2$, $\theta = y_2$.

Example 6.5. We go back to the pendulum example of section 6.2.3, in the implicit form (6.10):

$$\varepsilon\ddot{\theta} + \ddot{x}\cos\theta - (\ddot{z}+1)\sin\theta = 0.$$

We have

$$P(F) = \begin{pmatrix} \cos\theta\frac{d^2}{dt^2} & -\sin\theta\frac{d^2}{dt^2} & \varepsilon\frac{d^2}{dt^2} - (\ddot{x}\sin\theta + (\ddot{z}+1)\cos\theta) \end{pmatrix}$$

and

$$P(F)\begin{pmatrix} dx \\ dz \\ d\theta \end{pmatrix} = 0.$$

To compute the Smith decomposition of $P(F)$, we first prove the relation

$$\cos\theta\frac{d^2}{dt^2}(\cos\theta) + \sin\theta\frac{d^2}{dt^2}(\sin\theta) - \frac{d^2}{dt^2} = -\dot{\theta}^2.$$

We have $\frac{d}{dt}(\cos\theta.h) = \cos\theta.\dot{h} - \dot{\theta}\sin\theta.h$ for every differentiable function h, thus $\frac{d}{dt}(\cos\theta) = \cos\theta\frac{d}{dt} - \dot{\theta}\sin\theta$. Similarly, $\frac{d}{dt}(\sin\theta) = \sin\theta\frac{d}{dt} + \dot{\theta}\cos\theta$. Iterating this calculation, we get $\frac{d^2}{dt^2}(\cos\theta) = \cos\theta\frac{d^2}{dt^2} - 2\dot{\theta}\sin\theta\frac{d}{dt} - \ddot{\theta}\sin\theta - \dot{\theta}^2\cos\theta$, $\frac{d^2}{dt^2}(\sin\theta) = \sin\theta\frac{d^2}{dt^2} + 2\dot{\theta}\cos\theta\frac{d}{dt} + \ddot{\theta}\cos\theta - \dot{\theta}^2\sin\theta$, hence the announced relation.

Let us go back to the matrix $P(F)$ and set $U_1 = \begin{pmatrix} 1 & 0 & \varepsilon\cos\theta \\ 0 & 1 & -\varepsilon\sin\theta \\ 0 & 0 & -1 \end{pmatrix}$. We have $P(F).U_1 = \begin{pmatrix} \cos\theta\frac{d^2}{dt^2} & -\sin\theta\frac{d^2}{dt^2} & -\varepsilon\dot{\theta}^2 + A \end{pmatrix}$ with

$$A = (\ddot{x}\sin\theta + (\ddot{z}+1)\cos\theta).$$

Since the lowest degree term is in the 3rd column, we right multiply $P(F).U_1$ by $U_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ and $P(F).U_1.U_2 = \begin{pmatrix} -\varepsilon\dot{\theta}^2 + A & -\sin\theta\frac{d^2}{dt^2} & \cos\theta\frac{d^2}{dt^2} \end{pmatrix}$. Finally, right-multiplying the result by $U_3 = \begin{pmatrix} 1 & \frac{\sin\theta}{A-\varepsilon\dot{\theta}^2}\frac{d^2}{dt^2} & -\frac{\cos\theta}{A-\varepsilon\dot{\theta}^2}\frac{d^2}{dt^2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, we obtain the required form: $\begin{pmatrix} A - \varepsilon\dot{\theta}^2 & 0 & 0 \end{pmatrix}$. The matrix U is thus given by

$$U = U_1.U_2.U_3 = \begin{pmatrix} \varepsilon\cos\theta & \frac{\varepsilon\sin\theta\cos\theta}{A-\varepsilon\dot\theta^2}\frac{d^2}{dt^2} & -\frac{\varepsilon\cos^2\theta}{A-\varepsilon\dot\theta^2}\frac{d^2}{dt^2}+1 \\ -\varepsilon\sin\theta & -\frac{\varepsilon\sin^2\theta}{A-\varepsilon\dot\theta^2}\frac{d^2}{dt^2}+1 & \frac{\varepsilon\sin\theta\cos\theta}{A-\varepsilon\dot\theta^2}\frac{d^2}{dt^2} \\ -1 & -\frac{\sin\theta}{A-\varepsilon\dot\theta^2}\frac{d^2}{dt^2} & \frac{\cos\theta}{A-\varepsilon\dot\theta^2}\frac{d^2}{dt^2} \end{pmatrix}.$$

We set $E = A - \varepsilon\dot\theta^2$. Thus

$$\hat U = \begin{pmatrix} \frac{\varepsilon\sin\theta\cos\theta}{E}\frac{d^2}{dt^2} & -\frac{\varepsilon\cos^2\theta}{E}\frac{d^2}{dt^2}+1 \\ -\frac{\varepsilon\sin^2\theta}{E}\frac{d^2}{dt^2}+1 & \frac{\varepsilon\sin\theta\cos\theta}{E}\frac{d^2}{dt^2} \\ -\frac{\sin\theta}{E}\frac{d^2}{dt^2} & \frac{\cos\theta}{E}\frac{d^2}{dt^2} \end{pmatrix}.$$

The calculation of $Q \in \mathsf{L-Smith}\left(\hat U\right)$ yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{\sin\theta}{E}\frac{d^2}{dt^2} & \frac{\cos\theta}{E}\frac{d^2}{dt^2} & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -a\sin\theta \\ 1 & 0 & a\cos\theta \\ 0 & 0 & 1 \end{pmatrix} \hat U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

or

$$Q = \begin{pmatrix} 0 & 1 & -a\sin\theta \\ 1 & 0 & a\cos\theta \\ \frac{\cos\theta}{E}\frac{d^2}{dt^2} & -\frac{\sin\theta}{E}\frac{d^2}{dt^2} & \frac{a}{E}\frac{d^2}{dt^2} - \frac{b}{E} \end{pmatrix}$$

and

$$Q\begin{pmatrix} dx \\ dz \\ d\theta \end{pmatrix} = \begin{pmatrix} dz - a\sin\theta d\theta \\ dx + a\cos\theta d\theta \\ \frac{1}{E}\left(\cos\theta d\ddot x - \sin\theta d\ddot z + a d\ddot\theta - b d\theta\right) \end{pmatrix}$$
$$= \begin{pmatrix} dz - a\sin\theta d\theta \\ dx + a\cos\theta d\theta \\ 0 \end{pmatrix} = \begin{pmatrix} d(z + a\cos\theta) \\ d(x + a\sin\theta) \\ 0 \end{pmatrix}$$

The strong closedness condition of the ideal Ω generated by $(d(z + a\cos\theta), d(x + a\sin\theta))$ is thus satisfied. Let us set

$$d(z + a\cos\theta) = dy_1, \quad d(x + a\sin\theta) = dy_2$$

we thus get

$$y_1 = z + a\cos\theta, \quad y_2 = x + a\sin\theta$$

that correspond (up to a permutation) to the coordinates of the Huygens oscillation centre previously exhibited in section 6.2.3.

Example 6.6 (Chetverikov [2001]. See also Schlacher and Schöberl [2007] for a different approach). We consider the following academic 3-dimensional example with 2 inputs:

$$\begin{aligned}\dot{x}_1 &= u_1 \\ \dot{x}_2 &= u_2 \\ \dot{x}_3 &= \sin\left(\frac{u_1}{u_2}\right)\end{aligned} \tag{6.83}$$

or, after input elimination:

$$F = \dot{x}_3 - \sin\left(\frac{\dot{x}_1}{\dot{x}_2}\right) = 0. \tag{6.84}$$

The polynomial matrix of its variational system is given

$$P(F) = \begin{pmatrix} -\frac{1}{\dot{x}_2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & \frac{\dot{x}_1}{(\dot{x}_2)^2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & \frac{d}{dt}\end{pmatrix} \tag{6.85}$$

To compute $U \in \mathsf{R}-\mathsf{Smith}\,(P(F))$, we right-multiply $P(F)$ successively by U_1 and U_2 given by

$$U_1 = \begin{pmatrix} \frac{\dot{x}_1}{\dot{x}_2} & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad U_2 = \begin{pmatrix} \frac{1}{a} & \frac{1}{a\dot{x}_2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & -\frac{1}{a}\frac{d}{dt} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where $a = -\frac{1}{\dot{x}_2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\left(\frac{\ddot{x}_1\dot{x}_2 - \dot{x}_1\ddot{x}_2}{(\dot{x}_2)^2}\right)$.

Thus

$$U = U_1U_2 = \begin{pmatrix} \frac{\dot{x}_1}{a\dot{x}_2} & 1 + \frac{\dot{x}_1}{a\dot{x}_2^2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & \frac{\dot{x}_1}{a\dot{x}_2}\frac{d}{dt} \\ \frac{1}{a} & \frac{1}{a\dot{x}_2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & -\frac{1}{a}\frac{d}{dt} \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\hat{U} = \begin{pmatrix} 1 + \frac{\dot{x}_1}{a\dot{x}_2^2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & \frac{\dot{x}_1}{a\dot{x}_2}\frac{d}{dt} \\ \frac{1}{a\dot{x}_2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & -\frac{1}{a}\frac{d}{dt} \\ 0 & 1 \end{pmatrix} \tag{6.86}$$

The reader may easily verify that $P(F)U = (1\ 0\ 0)$ and $P(F)\hat{U} = (0\ 0)$.

Next, we compute $Q \in \mathsf{L}-\mathsf{Smith}\left(\hat{U}\right)$ by left-multiplying $\hat{U}$ successively by Q_1 and Q_2 defined by

$$Q_1 = \begin{pmatrix} 1 & -\frac{\dot{x}_1}{\dot{x}_2} & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad Q_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{a\dot{x}_2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & \frac{1}{a}\frac{d}{dt} & 1 \end{pmatrix}$$

and

$$Q = Q_2 Q_1 = \begin{pmatrix} 1 & -\frac{\dot{x}_1}{\dot{x}_2} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{a\dot{x}_2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & \frac{\dot{x}_1}{a\dot{x}_2^2}\cos\left(\frac{\dot{x}_1}{\dot{x}_2}\right)\frac{d}{dt} & \frac{1}{a}\frac{d}{dt} \end{pmatrix}, \tag{6.87}$$
$$\tilde{Q} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} Q = \begin{pmatrix} 1 & -\frac{\dot{x}_1}{\dot{x}_2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

which yields, as required, $Q\hat{U} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$, and $\tilde{Q}\hat{U} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Then, in accordance with (6.57) of Theorem 6.6, we set

$$\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \tilde{Q}\begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} = \begin{pmatrix} dx_1 - \frac{\dot{x}_1}{\dot{x}_2}dx_2 \\ dx_3 \end{pmatrix} \tag{6.88}$$

Clearly, ω_1 is not closed since $d\omega_1 = -d\left(\frac{\dot{x}_1}{\dot{x}_2}\right) \wedge dx_2 \neq 0$, which, according to (6.84), reads

$$d\omega_1 = -\frac{1}{\sqrt{1-\dot{x}_3^2}} d\dot{x}_3 \wedge dx_2 = \frac{1}{\sqrt{1-\dot{x}_3^2}} dx_2 \wedge \dot{\omega}_2 \tag{6.89}$$

From now on, there are two possible ways to tackle a solution: integrate ω_1 modulo the ideal generated by $\omega_2, \dot{\omega}_2, \ldots$, or apply the systematic algorithm of section 6.5.7. We present both approaches.

Direct integration

We start from the remark that, by (6.88), combined with the fact that, by (6.84), $d\left(\frac{\dot{x}_1}{\dot{x}_2}\right) = \frac{1}{\sqrt{1-\dot{x}_3^2}} d\dot{x}_3$,

$$\omega_1 = dx_1 - \frac{\dot{x}_1}{\dot{x}_2}dx_2 = d\left(x_1 - \frac{\dot{x}_1}{\dot{x}_2}x_2\right) + \frac{x_2}{\sqrt{1-\dot{x}_3^2}} d\dot{x}_3$$

Thus, since $d\dot{x}_3 = \dot{\omega}_2$, we have

$$d\left(x_1 - \frac{\dot{x}_1}{\dot{x}_2}x_2\right) = \omega_1 - \frac{x_2}{\sqrt{1-\dot{x}_3^2}}\dot{\omega}_2 = \begin{pmatrix} 1 & -\frac{x_2}{\sqrt{1-\dot{x}_3^2}}\frac{d}{dt} \end{pmatrix}\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$$

Therefore, if we set

$$\begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix} = \begin{pmatrix} d\left(x_1 - \frac{\dot{x}_1}{\dot{x}_2}x_2\right) \\ dx_3 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{x_2}{\sqrt{1-\dot{x}_3^2}}\frac{d}{dt} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \stackrel{\text{def}}{=} M\omega$$

the matrix M is unimodular, with inverse $M^{-1} = \begin{pmatrix} 1 & \frac{x_2}{\sqrt{1-\dot{x}_3^2}}\frac{d}{dt} \\ 0 & 1 \end{pmatrix}$, which means that the ideal Ω is strongly closed and that

$$y_1 = x_1 - \frac{\dot{x}_1}{\dot{x}_2}x_2, \quad y_2 = x_3 \tag{6.90}$$

is a flat output.

Application of the general algorithm

In order to identify μ such that $d\omega = \mu\omega$ with $\mathfrak{d}(\mu) = \mu^2$, we use the expression (6.89):

$$d\omega_1 = \frac{1}{\sqrt{1-\dot{x}_3^2}} dx_2 \wedge \dot{\omega}_2$$

Therefore, for some meromorphic function η to be determined:

$$d\omega = \begin{pmatrix} d\omega_1 \\ d\omega_2 \end{pmatrix} = \begin{pmatrix} 0 & \left(\frac{1}{\sqrt{1-\dot{x}_3^2}}dx_2 + \eta\dot{\omega}_2\right) \wedge \frac{d}{dt} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \stackrel{\text{def}}{=} \mu\omega \tag{6.91}$$

We then compute $\mathfrak{d}(\mu)$ and μ^2:

$$\mathfrak{d}(\mu) = \begin{pmatrix} 0 & \left(\frac{\dot{x}_3}{(1-\dot{x}_3^2)^{\frac{3}{2}}}d\dot{x}_3 \wedge dx_2 + d\eta \wedge \dot{\omega}_2\right) \wedge \frac{d}{dt} \\ 0 & 0 \end{pmatrix} = \mu^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

which yields

$$\frac{\dot{x}_3}{(1-\dot{x}_3^2)^{\frac{3}{2}}}d\dot{x}_3 \wedge dx_2 + d\eta \wedge \dot{\omega}_2 = -\frac{\dot{x}_3}{(1-\dot{x}_3^2)^{\frac{3}{2}}}dx_2 \wedge \dot{\omega}_2 + d\eta \wedge \dot{\omega}_2 = 0$$

or, with γ a suitable meromorphic function:

$$d\eta = \frac{\dot{x}_3}{(1-\dot{x}_3^2)^{\frac{3}{2}}}dx_2 + \gamma d\dot{x}_3$$

This last differential equation is clearly integrable if and only if

$$\gamma = \frac{\partial}{\partial \dot{x}_3}\left(\frac{x_2\dot{x}_3}{(1-\dot{x}_3^2)^{\frac{3}{2}}} + \sigma(\dot{x}_3)\right)$$

where σ is an arbitrary meromorphic function of $\dot{x}_3$ only, and yields

$$\eta = \frac{x_2 \dot{x}_3}{(1-\dot{x}_3^2)^{\frac{3}{2}}} + \sigma(\dot{x}_3)$$

We thus obtain

$$\begin{aligned} \mu &= \begin{pmatrix} 0 & \left(\frac{1}{\sqrt{1-\dot{x}_3^2}} dx_2 + \frac{x_2 \dot{x}_3}{(1-\dot{x}_3^2)^{\frac{3}{2}}} d\dot{x}_3 + \sigma(\dot{x}_3) d\dot{x}_3 \right) \wedge \frac{d}{dt} \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & d\left(\frac{x_2}{\sqrt{1-\dot{x}_3^2}} + \sigma_1(\dot{x}_3) \right) \wedge \frac{d}{dt} \\ 0 & 0 \end{pmatrix} \end{aligned} \tag{6.92}$$

where σ_1 is a primitive of σ.

Finally, assuming that M has the form $M = \begin{pmatrix} 1 & m_{1,2} \frac{d}{dt} \\ 0 & 1 \end{pmatrix}$, we must have $\mathfrak{d}(M) = -M\mu$, or

$$\begin{pmatrix} 0 & dm_{1,2} \wedge \frac{d}{dt} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -d\left(\frac{x_2}{\sqrt{1-\dot{x}_3^2}} + \sigma_1(\dot{x}_3) \right) \wedge \frac{d}{dt} \\ 0 & 0 \end{pmatrix}$$

or $m_{1,2} = -\frac{x_2}{\sqrt{1-\dot{x}_3^2}} - \sigma_1(\dot{x}_3)$. Thus we get

$$M = \begin{pmatrix} 1 & -\left(\frac{x_2}{\sqrt{1-\dot{x}_3^2}} + \sigma_1(\dot{x}_3) \right) \frac{d}{dt} \\ 0 & 1 \end{pmatrix} \tag{6.93}$$

and setting

$$\begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix} = \begin{pmatrix} 1 & -\left(\frac{x_2}{\sqrt{1-\dot{x}_3^2}} + \sigma_1(\dot{x}_3) \right) \frac{d}{dt} \\ 0 & 1 \end{pmatrix}$$

we get the flat output

$$y_1 = x_1 - \frac{\dot{x}_1}{\dot{x}_2} x_2 + \sigma_2(\dot{x}_3), \quad y_2 = x_3 \tag{6.94}$$

where σ_2 is an arbitrary meromorphic function of $\dot{x}_3$ only (a primitive of σ_1). We indeed recover, for $\sigma_2 = 0$, the flat output (6.90) obtained by the previous method.

To conclude this example, let us verify that (6.94) indeed constitutes a flat output.

Differentiating y_1 with respect to time, we get

$$\dot{y}_1 = -\frac{\ddot{x}_3}{\sqrt{1-\dot{x}_3^2}} x_2 + \sigma_1(\dot{x}_3)\ddot{x}_3 = -\frac{\ddot{y}_2}{\sqrt{1-\dot{y}_2^2}} x_2 + \sigma_1(\dot{y}_2)\ddot{y}_2$$

Thus

$$x_2 = -\frac{\sqrt{1-\dot{y}_2^2}}{\ddot{y}_2}\,(\dot{y}_1 - \sigma_1(\dot{y}_2)\ddot{y}_2)$$

and, plugging this expression in $y_1 = x_1 - \frac{\dot{x}_1}{\dot{x}_2}x_2 + \sigma_2(\dot{x}_3)$, we finally get

$$x_1 = y_1 - \arcsin(\dot{y}_2)\frac{\sqrt{1-\dot{y}_2^2}}{\ddot{y}_2}\,(\dot{y}_1 - \sigma_1(\dot{y}_2)\ddot{y}_2) - \sigma_2(\dot{y}_2)$$

which, together with $x_3 = y_2$, achieves to prove that (y_1, y_2) is a flat output.

Chapter 7
Flatness and Motion Planning

Let us consider the nonlinear system $\dot{x} = f(x, u)$.

Given the initial time t_i, the initial conditions

$$x(t_i) = x_i, \quad u(t_i) = u_i, \tag{7.1}$$

the final time t_f and the final conditions

$$x(t_f) = x_f, \quad u(t_f) = u_f, \tag{7.2}$$

the motion planning problem consists in finding a trajectory $t \mapsto (x(t), u(t))$ for $t \in [t_i, t_f]$ that satisfies $\dot{x} = f(x, u)$ and the initial and final conditions (7.1), (7.2)[1]. If we add constraints on the searched trajectory of the type $(x(t), u(t)) \in A(t)$ for $A(t)$ a submanifold of $X \times U$, the problem is called motion planning with constraints.

This problem, in the general case, is quite difficult since it may require an iterative solution by numerical methods to find a control input u such that conditions (7.1), (7.2) are satisfied: we start with an input $t \mapsto u_0(t)$, we integrate the system equations from the initial conditions, evaluate the solution at final time t_f, and then modify the input, say $t \mapsto u_1(t)$, to get closer to the final conditions, assuming that we are given a mechanism that guarantees it, and so on. In this class, a typical method for the determination of u is the Optimal Control approach, e.g. find the control that minimizes the square deviation to an *a priori* given trajectory. For nonlinear systems, it may pose problems that are still open. One can also choose, in some particular cases, the inputs in an *a priori* parameterized class, e.g. classes of sinusoids, for which the solution or an approximation of the solution is sometimes known.

We will see that, in the case of flat systems, this problem is easily solved without approximation and without requiring to integrate the system differential equations.

---

[1] We are not only looking for a feasible trajectory $t \mapsto x(t)$, but also for the open-loop control $t \mapsto u(t)$ that generates it.

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering, DOI 10.1007/978-3-642-00839-9_7,

7.1 Motion Planning Without Constraint

The flatness conditions boil down to the existence of a flat output such that all the system variables can be expressed as functions of this flat output and a finite number of its successive derivatives, this parameterization being such that the system differential equations are identically satisfied.

It results that, if we want to construct a trajectory whose initial and final conditions are specified, it suffices to compute the corresponding flat output trajectory, while integrating the system differential equations is not necessary.

More precisely, let us assume that

$$\begin{aligned} x &= \varphi_0(y, \dot{y}, \ldots, y^{(r)}) \\ u &= \varphi_1(y, \dot{y}, \ldots, y^{(r+1)}). \end{aligned} \tag{7.3}$$

Since the initial and final values of x and u are given, by the surjectivity of (φ_0, φ_1) one can find the initial and final values of $(y, \dot{y}, \ldots, y^{(r_1+1)})$.

It suffices then to find a trajectory $t \mapsto y(t)$ at least $r+1$ times differentiable that satisfies these initial and final conditions since x and u are deduced from y and derivatives up to order $r+1$ by (7.3). Since, moreover, the trajectory $t \mapsto y(t)$ is not required to satisfy any differential equation, one can simply compute it by *polynomial interpolation*, similarly to what we have presented in the linear controllable case (see section 4.1.3).

Before recalling this construction, let us precise that, as a direct consequence of the definition of a flat output, the trajectories $t \mapsto x(t)$ and $t \mapsto u(t)$ obtained by replacing y and derivatives by their values in function of time in (7.3) identically satisfy the system differential equations.

We now detail this construction in the general case first, and then in the important particular case of *rest-to-rest* trajectories, *i.e.* joining two equilibrium points of the system, or, in other words, such that the system is at rest when starting and at the end.

7.1.1 The General Case

Let us thus assume that we are given the following data at time t_i:

$$y_1(t_i), \ldots, y_1^{(r+1)}(t_i), \ldots, y_m(t_i), \ldots, y_m^{(r+1)}(t_i) \tag{7.4}$$

and at time t_f:

$$y_1(t_f), \ldots, y_1^{(r+1)}(t_f), \ldots, y_m(t_f), \ldots, y_m^{(r+1)}(t_f) \tag{7.5}$$

that, together, represent $2(r+2)$ conditions on each of the m components of y.

If we look for $(y_1, \ldots, y_m)$ in the form of m polynomials with respect to time, each component of y must have at least $2(r+2)$ coefficients to satisfy the initial and final conditions, and thus must be at least of degree $2r+3$.

Let us denote by $T = t_f - t_i$, $\tau(t) = \frac{t-t_i}{T}$ and

$$y_j(t) = \sum_{k=0}^{2r+3} a_{j,k} \tau^k(t), \quad j = 1, \ldots, m.$$

Following the same lines as in section 4.1.3, we compute the coefficients $a_{j,k}$ by equating the successive derivatives of y_j at the initial and final times to the data (7.4) and (7.5):

$$y_j^{(k)}(t) = \frac{1}{T^k} \sum_{l=k}^{2r+3} \frac{l!}{(l-k)!} a_{j,l} \tau^{l-k}(t), \; j = 1, \ldots, m$$

or, at $\tau = 0$, which corresponds to $t = t_i$,

$$y_j^{(k)}(t_i) = \frac{k!}{T^k} a_{j,k}, \; k = 0, \ldots, r+1, \; j = 1, \ldots, m, \tag{7.6}$$

and at $\tau = 1$, or $t = t_f$,

$$y_j^{(k)}(t_f) = \frac{1}{T^k} \sum_{l=k}^{2r+3} \frac{l!}{(l-k)!} a_{j,l}, \; k = 0, \ldots, r+1, \; j = 1, \ldots, m \tag{7.7}$$

which makes a total of $2r+4$ linear equations in the $2r+4$ coefficients $a_{j,0}, \ldots, a_{j,2r+3}$, for every $j = 1, \ldots, m$. This system can in fact be reduced to $r+2$ linear equations in the $r+2$ unknown coefficients $a_{j,r+2}, \ldots, a_{j,2r+3}$, since the $r+2$ first equations (7.4) are solved in $a_{j,0}, \ldots, a_{j,r+1}$:

$$a_{j,k} = \frac{T^k}{k!} y_j^{(k)}(t_i) \;, k = 0, \ldots, r+1.$$

The remaining $r+2$ coefficients are given by:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ r+2 & r+3 & & 2r+3 \\ (r+1)(r+2) & (r+2)(r+3) & & (2r+2)(2r+3) \\ \vdots & & & \vdots \\ (r+2)! & \frac{(r+3)!}{2} & \cdots & \frac{(2r+3)!}{(r+2)!} \end{pmatrix} \begin{pmatrix} a_{j,r+2} \\ \vdots \\ a_{j,2r+3} \end{pmatrix}$$

$$= \begin{pmatrix} y_j(t_f) - \sum_{l=0}^{r+1} \frac{T^l}{l!} y_j^{(l)}(t_i) \\ \vdots \\ T^k \left(y_j^{(k)}(t_f) - \sum_{l=k}^{r+1} \frac{T^{l-k}}{(l-k)!} y_j^{(l)}(t_i) \right) \\ \vdots \\ T^{r+1} \left(y_j^{(r+1)}(t_f) - y_j^{(r+1)}(t_i) \right) \end{pmatrix} . \quad (7.8)$$

Remark 7.1. In this construction, the relation giving y in function of x, u and derivatives of u is nowhere needed, so that the interpretation of y in function of the original variables of the system, even though it often gives an interesting insight, is of minor importance in the motion planning problem.

7.1.2 Rest-to-Rest Trajectories

If the starting point $(x(t_i), u(t_i))$ and ending point $(x(t_f), u(t_f))$ are equilibrium points, we have $\dot{x}(t_i) = 0$, $\dot{u}(t_i) = 0$ and $\dot{x}(t_f) = 0$, $\dot{u}(t_f) = 0$. We know, from Theorem 5.2 that $y(t_i)$ and $y(t_f)$ are equilibrium points too for the associated trivial system and, according to (7.3), we have

$$x(t_i) = \varphi_0(y(t_i), 0, \ldots, 0), \quad u(t_i) = \varphi_1(y(t_i), 0, \ldots, 0)$$

$$x(t_f) = \varphi_0(y(t_f), 0, \ldots, 0), \quad u(t_f) = \varphi_1(y(t_f), 0, \ldots, 0).$$

The previous construction is thus easily adapted by replacing all the derivatives of y by 0 at t_i and t_f. We get

Proposition 7.1. *The polynomial rest-to-rest trajectories are of the form*

$$y_j(t) = y_j(t_i) + (y_j(t_f) - y_j(t_i)) \left(\frac{t - t_i}{t_f - t_i} \right)^{r+2} \left(\sum_{k=0}^{r+1} \alpha_{j,k} \left(\frac{t - t_i}{t_f - t_i} \right)^k \right), \quad j = 1, \ldots, m \quad (7.9)$$

with $\alpha_{j,0}, \ldots, \alpha_{j,r+1}$ *solution of*

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ r+2 & r+3 & & 2r+3 \\ (r+1)(r+2) & (r+2)(r+3) & & (2r+2)(2r+3) \\ \vdots & & & \vdots \\ (r+2)! & \frac{(r+3)!}{2} & \dots & \frac{(2r+3)!}{(r+2)!} \end{pmatrix} \begin{pmatrix} \alpha_{j,0} \\ \vdots \\ \alpha_{j,r+1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (7.10)$$

Proof. It suffices to replace the derivatives of y by 0 in (7.8). We thus get $a_{j,0} = y_j(t_i)$, $a_{j,k} = 0$ for $k = 1, \dots, r+1$, and the interpolation polynomials read

$$y_j(t) = y_j(t_i) + \sum_{k=r+2}^{2r+3} a_{j,k}\tau^k(t), \quad j = 1, \dots, m.$$

where the right-hand side of the linear system (7.8) is the vector $\begin{pmatrix} y_j(t_f) - y_j(t_i) \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. Thus, setting $\alpha_{j,k-r-2} = \dfrac{a_{j,k}}{y_j(t_f) - y_j(t_i)}$, we obtain the expression (7.9) and the system (7.10).

Remark 7.2. SInce all the derivatives of y have to be equal to 0 at an equilibrium point, one can add an arbitrary number of initial and final null derivatives of order larger than or equal to $r+1$ without changing the initial and final equilibrium points. In this way, one can increase the regularity of the trajectory and make the start and end "smoother". This may be useful in practice to avoid exciting oscillating or unstable modes at the end point.

7.2 Motion Planning With Constraints

We consider here two kinds of constraints: constraints of geometric type, namely concerning the fact that the flat output must remain inside a given domain, or quantitative limitations on positions, velocities, accelerations, and so on.

7.2.1 Geometric Constraints

Assume that we want to plan rest-to-rest trajectories for a flat system with m inputs and n states, and that we want the flat output trajectory to remain in a domain described by the inequality

$$A(y) \leq 0. \tag{7.11}$$

We assume, for simplicity's sake, that A is onto from $\mathbb{R}^m$ to $\mathbb{R}$ and C^∞ and that the initial point y_i and final point y_f belong to the frontier of this domain, *i.e.* $A(y_i) = A(y_f) = 0$, and to the same connected component, noted A_0, of the submanifold $A = 0$.

We can construct a trajectory $t \mapsto y^*(t)$ satisfying $A(y^*(t)) = 0$ for all t, and thus satisfying (7.11), as follows: since A is onto, by the implicit function Theorem, there exists a mapping Y from a neighborhood of U in $\mathbb{R}^{m-1}$ to $A_0 \subset \mathbb{R}$ such that, possibly up to a permutation of the components of y, $y_m = Y(y_1, \ldots, y_{m-1})$ implies that $A(y_1, \ldots, y_{m-1}, Y(y_1, \ldots, y_{m-1})) = 0$ for all $(y_1, \ldots, y_{m-1}) \in U$. In particular, we have $y_m(t_i) = Y(y_1(t_i), \ldots, y_m(t_i))$ and $y_m(t_f) = Y(y_1(t_f), \ldots, y_m(t_f))$.

It suffices thus to construct $m-1$ curves $t \mapsto y_j(t)$, $j = 1, \ldots, m-1$, satisfying, at the initial time $\dot{y}_j(t_i) = \ldots = y_j^{(r+1)}(t_i) = 0$, for all $j = 1, \ldots, m-1$, and, at the final time, $\dot{y}_j(t_f) = \ldots = y_j^{(r+1)}(t_f) = 0$.

In this case, indeed, we have $y_m(t_i) = Y(y_1(t_i), \ldots, y_m(t_i))$, $\dot{y}_m(t_i) = \sum_{j=1}^{m-1} \frac{\partial Y}{\partial y_j} \dot{y}_j(t_i) = 0$, and, by induction, $y_m^{(r+1)}(t_i) = 0$, so that the initial and final rest-to-rest conditions are necessarily satisfied.

It remains to construct the $m-1$ curves $t \mapsto y_j(t)$, $j = 1, \ldots, m-1$ satisfying the above initial and final conditions. We proceed in the same way as in section 7.1.2:

$$y_j(t) = y_j(t_i) + (y_j(t_f) - y_j(t_i)) \left(\frac{t - t_i}{t_f - t_i} \right)^{r+2} \left(\sum_{k=0}^{r+1} \alpha_{j,k} \left(\frac{t - t_i}{t_f - t_i} \right)^k \right)$$

with the $\alpha_{j,k}$'s given by (7.10) for all $j = 1, \ldots, m-1$.

We achieve this construction by composing these trajectories with Y to obtain the last component of y and hence the desired trajectory y^*. We finally deduce x and u in the usual way.

Example 7.1. We go back to the example of non holonomic vehicle of sections 6.2.4 and 6.4.4.

Assume that we want to parallel park the car along a sidewalk in reverse. Remark that, in reverse, we must change u to $-u$, or equivalently, consider that $u < 0$ in the system equations (6.17).

Let us suppose that the x-axis of the fixed frame is parallel to the boundary of the sidewalk. The initial position, at time t_i, is denoted by (x_i, y_i), the vehicle's axis and its front wheels being parallel to the sidewalk ($\theta(t_i) = 0$,

$\varphi(t_i) = 0$), and the final position, at time t_f, is noted (x_f, y_f), with, again, the vehicle's axis and its front wheels being parallel to the sidewalk ($\theta(t_f) = 0$, $\varphi(t_f) = 0$). The initial speed, in the sidewalk direction, is equal to 0, as well as the final speed.

We assume, for the sake of simplicity, that the geometric constraint is given by $y \geq y_f$, $x \in [x_i, x_f]$, which expresses the fact that the car is not allowed to roll on the sidewalk.

Let us construct a curve $y = Y(x)$ satisfying

$$
\begin{aligned}
y_i &= Y(x_i) \ , \quad 0 = \frac{dY}{dx}(x_i) \ , \quad 0 = \frac{d^2Y}{dx^2}(x_i) \\
y_f &= Y(x_f) \ , \quad 0 = \frac{dY}{dx}(x_f) \ , \quad 0 = \frac{d^2Y}{dx^2}(x_f).
\end{aligned}
$$

in addition to the aforementioned constraint.

The function given by

$$
Y(x) = y_i + (y_f - y_i)\left(\frac{x - x_i}{x_f - x_i}\right)^3 \left(10 - 15\left(\frac{x - x_i}{x_f - x_i}\right) + 6\left(\frac{x - x_i}{x_f - x_i}\right)^2\right)
$$

obtained by polynomial interpolation, satisfies these requirements.

We now have to find a curve $t \mapsto x(t)$ such that $x(t_i) = x_i$, $\dot{x}(t_i) = 0$ and $x(t_f) = x_f$, $\dot{x}(t_f) = 0$.

Using polynomial interpolation, we find

$$
x(t) = x_i + (x_f - x_i)\left(\frac{t - t_i}{t_f - t_i}\right)^2 \left(3 - 2\left(\frac{t - t_i}{t_f - t_i}\right)\right).
$$

It suffices then to compose $y(t) = Y(x(t))$ to obtain the trajectory of the middle of the rear axle. Thus the speed is given by $u = -(\dot{x}^2(t) + \dot{y}^2(t))^{\frac{1}{2}}$ and the other variables, θ, φ, by (6.19), (6.20) for all $t \in]t_i, t_f[$. Moreover, we verify (exercise) that $\lim_{\substack{t \to t_i \\ t \geq t_i}} \tan\theta(t) = 0$, $\lim_{\substack{t \to t_i \\ t \geq t_i}} \tan\varphi(t) = 0$, $\lim_{\substack{t \to t_f \\ t \leq t_f}} \tan\theta(t) = 0$, $\lim_{\substack{t \to t_f \\ t \leq t_f}} \tan\varphi(t) = 0$, which achieves the construction.

It is remarkable that the initial and end points are equilibrium points and that the system is not first-order controllable at these points, which results from the fact that the formulas (6.19), (6.20) are not defined there, because the speed vanishes. The construction in two steps proposed here, may thus be interpreted as a *desingularization*.

Example 7.2. Let us go back to the crane example of section 5.1 and let us construct a rest-to-rest trajectory avoiding an obstacle located approximately in the middle of the load's displacement.

Recall that the position of the load is a flat output and thus that the constraints on the coordinates of the load are directly expressed in terms of the flat output.

Similarly to the previous example, we first construct a geometric path remaining above the obstacle, and then we tune the evolution on this curve to obtain the desired rest-to-rest behavior.

We assume that the initial position is (ξ_i, ζ_i) at time t_i, at rest, that the final position is (ξ_f, ζ_f) at time t_f, at rest, and that the trajectory passes through the point $(\frac{\xi_f+\xi_i}{2}, 2\zeta_f - \zeta_i)$ which must be the maximum of the desired curve between ξ_i and ξ_f.

The desired trajectory $\xi \mapsto \zeta$ must therefore satisfy the four conditions

$$\zeta(\xi_i) = \zeta_i, \quad \zeta(\xi_f) = \zeta_f, \quad \zeta(\frac{\xi_f + \xi_i}{2}) = 2\zeta_f - \zeta_i, \quad \frac{d\zeta}{d\xi}(\frac{\xi_f + \xi_i}{2}) = 0.$$

and the constraint $\frac{d^2\zeta}{d\xi^2}(\frac{\xi_f+\xi_i}{2}) < 0$ to have a local maximum at this point.

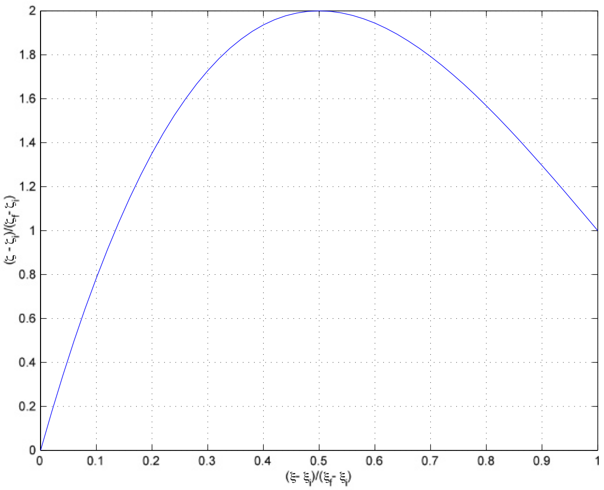

Fig. 7.1 Trajectory of $\frac{\zeta-\zeta_i}{\zeta_f-\zeta_i}$ in function of $\frac{\xi-\xi_i}{\xi_f-\xi_i}$.

The 3rd degree polynomial in ξ:

$$\zeta(\xi) = \zeta_i + (\zeta_f - \zeta_i)\left(\frac{\xi - \xi_i}{\xi_f - \xi_i}\right)\left(9 - 12\left(\frac{\xi - \xi_i}{\xi_f - \xi_i}\right) + 4\left(\frac{\xi - \xi_i}{\xi_f - \xi_i}\right)^2\right)$$

satisfies all the conditions (see Fig 7.1).

It remains to construct a trajectory $t \mapsto \xi(t)$ satisfying

$$\xi(t_i) = \xi_i, \quad \dot{\xi}(t_i) = 0, \quad \ldots, \quad \xi^{(5)}(t_i) = 0$$
$$\xi(t_f) = \xi_f, \quad \dot{\xi}(t_f) = 0, \quad \ldots, \quad \xi^{(5)}(t_f) = 0$$

where we have required that all the derivatives of ξ up to order 5 vanish to guarantee that the derivatives of the force applied to the cart and of the winch torque also vanish. In fact, the successive derivatives up to order 4 must vanish to guarantee that the crane is at rest, and the nullity of 5th order is added to produce a smooth and gradual start and end.

We thus obtain the 11th degree polynomial:

$$\xi(t) = \xi_i + (\xi_f - \xi_i)\,\sigma^6(t) \cdot \big(462 - 1980\sigma(t) + 3465\sigma^2(t) - 3080\sigma^3(t) + 1386\sigma^4(t) - 252\sigma^5(t)\big)$$

with $\sigma(t) = \frac{t-t_i}{t_f-t_i}$.

The trajectories of x, R, θ, T, F and C are deduced from $t \mapsto (\xi(t), \zeta(\xi(t)))$ by (5.8)-(5.9)-(5.10).

7.2.2 Quantitative Constraints

In addition to the geometric constraints of the previous section, it is often required that some system variables, and in particular the inputs, remain bounded.

We may tune the total duration T such that the bounds are respected. Indeed, since the flat output may be expressed in terms of the reduced time

$$\tau(t) = \frac{t - t_i}{t_f - t_i} = \frac{t - t_i}{T}$$

we have

$$\dot{y}(t) = \frac{1}{T}\frac{dy}{d\tau}(\tau(t)),\ \ddot{y}(t) = \frac{1}{T^2}\frac{d^2y}{d\tau^2}(\tau(t)), \ldots,\ y^{(k)}(t) = \frac{1}{T^k}\frac{d^ky}{d\tau^k}(\tau(t)), \ldots$$

and thus

$$\max_{t\in[t_i,t_f]} \|y^{(k)}(t)\| = \frac{1}{T^k} \max_{\tau\in[0,1]} \|\frac{d^ky}{d\tau^k}(\tau)\|, \quad \forall k \geq 1.$$

To guarantee that

$$|\dot{y}\| \leq C_1, \ldots, \|y^{(k)}\| \leq C_k$$

for some given constants $C_1, \ldots, C_k$, it suffices to choose the duration T such that:

$$T = \max\left\{ \frac{1}{C_1} \max_{\tau\in[0,1]} \left\| \frac{dy}{d\tau} \right\|, \ldots, \left(\frac{1}{C_k} \max_{\tau\in[0,1]} \left\| \frac{d^ky}{d\tau^k} \right\| \right)^{\frac{1}{k}} \right\}. \tag{7.12}$$

The constants C_j are then chosen in order to guarantee that the original variables satisfy the requested bounds thanks to the Lie-Bäcklund isomorphism.

We assume for instance that we want $\|u\| \leq C_u$.

Since $u = \varphi_1(y, \dot{y}, \ldots, y^{(r+1)})$, for T large enough, we have

$$\begin{aligned}
\|u\| &= \left\|\varphi_1(y, \dot{y}, \ldots, y^{(r+1)})\right\| = \left\|\varphi_1(y, \frac{1}{T}\frac{dy}{d\tau}, \ldots, \frac{1}{T^{r+1}}\frac{d^{r+1}y}{d\tau^{r+1}})\right\| \\
&\leq \|\varphi_1(y, 0, \ldots, 0)\| + \frac{1}{T}\left\|\frac{\partial\varphi_1}{\partial\dot{y}}\right\| \left\|\frac{dy}{d\tau}\right\| + \ldots + \frac{1}{T^{r+1}}\left\|\frac{\partial\varphi_1}{\partial y^{(r+1)}}\right\| \left\|\frac{d^{r+1}y}{d\tau^{r+1}}\right\| \\
&\leq \|\varphi_1(y, 0, \ldots, 0)\| + \left\|\frac{\partial\varphi_1}{\partial\dot{y}}\right\| C_1 + \ldots + \left\|\frac{\partial\varphi_1}{\partial y^{(r+1)}}\right\| C_{r+1} \\
&\leq C_u
\end{aligned}$$

which allows us to choose the constants C_k, $k = 1, \ldots, r+1$ in function of C_u and then T by (7.12).

To conclude, if there exist trajectories satisfying the given constraints one can obtain them by sufficiently increasing T.

7.3 Application to Predictive Control

The predictive control approach consists in the data of a horizon T, a receding horizon T_0, generally small compared to T, some final conditions and constraints as in the motion planning problem, the main difference being that the perturbations than may deviate the trajectory are taken into account precisely through the deviation with respect to the reference trajectory. To attenuate this deviation after a duration T_0, the system state is measured and a new trajectory relating these new initial conditions to the final ones is computed, and so on.

More precisely, given the initial conditions (x_i, u_i), a reference trajectory relating them to the target (x_f, u_f) at time $t_f = t_i + T$ and satisfying the given constraints is computed. Then, at time $t_i + T_0$, a measurement of $(x(t_i + T_0), u(t_i + T_0))$ is done and, if this new point is not close enough to the reference trajectory, a new trajectory starting from $(x(t_i + T_0), u(t_i + T_0))$ and arriving at (x_f, u_f) at time $t_i + T_0 + T$ [2], again satisfying the given constraints, is recomputed. This approach is iterated until we arrive close enough to the target. If, in the mean time, we cannot find a trajectory satisfying the constraints, one can increase T or try to reach an intermediate point from which the target is reachable under the given constraints.

The previous methods can obviously be applied by computing a reference trajectory of the flat output: at each step, a trajectory of the flat output start-

[2] this is what is called a receding horizon since the time interval in which we are working, namely $[t_i, t_i + T]$, then $[t_i + T_0, t_i + T_0 + T]$, and so on, is progressively shifted

ing from $(y(t_i + kT_0), \ldots, y^{(r+1)}(t_i + kT_0))$ such that $x(t_i + kT_0) = \varphi_0(y(t_i + kT_0), \ldots, y^{(r)}(t_i + kT_0))$, $u(t_i + kT_0) = \varphi_1(y(t_i + kT_0), \ldots, y^{(r+1)}(t_i + kT_0))$, is computed, where $x(t_i + kT_0)$ is the new measured point, ensuring the continuity of the derivatives of the flat output at this time, thus avoiding creating jerks.

Let us remark that the perturbations may correspond to unmodelled phenomena or, in the case where the control is designed to provide an assistance to a system supervised by an operator, to setup modifications required by the operator.

There is, at present, no general result establishing that the objectives are reachable against arbitrary perturbations, and no general evaluation of the duration to reach the target. This method is however very popular in the industry, and particularly in chemical processes. In practice, it gives satisfactory results for systems that are sufficiently stable in open-loop, or unstable with slow enough dynamics compared to the receding horizon T_0, or with small enough perturbations and a precise enough model (see *e.g.* Petit et al. [2001], Devos and Lévine [2006], Devos [2009]). Other approaches may be found in Morari and Lee [1999], Findeisen and Allgöwer [2002], Fliess and Marquez [2000], Delaleau and Hagenmeyer [2006], Hagenmeyer and Delaleau [2008], De Dona et al. [2008].

Chapter 8
Flatness and Tracking

8.1 The Tracking Problem

For the solution of the motion planning problem, all we required was the knowledge of a dynamical model and the time, in other words, anticipative data: the reference trajectory was computed from the present time to some future time according to what we know about the system's evolution. This type of design is called *open-loop*. If the system dynamics is precisely known and if the disturbances (all signals not taken into account in the model that might affect the system's evolution) don't produce significant deviations from the predicted trajectories in the workspace, the open-loop design may sometimes be sufficient. However, if measurements of the system evolution are available, they may be used to compensate such disturbances. More precisely, if disturbances create significant deviations from our predictions, which is more often the case[1], we may *close the loop*, by using the measurements to compute at every time the deviation with respect to our desired trajectory and deduce some correction term in the control to decrease this deviation.

For a flat system, in an open set that doesn't contain singular points (points where the Lie-Bäcklund isomorphism degenerates or is no more defined), and if there are enough sensors to measure all the system state, this trajectory tracking may be designed thanks to Corollary 6.2 establishing the equivalence to a trivial system by endogenous dynamic feedback. Indeed, if y is a flat output of the system whose state is x and input u, assumed to be measured, and if y^* is the reference trajectory of the flat output, let us denote by $e_i = y_i - y_i^*$, $i = 1, \ldots, m$, the components of the error. By Corollary 6.2, we can compute an endogenous dynamic feedback such that the system, up

[1] In fact, in control, as opposed to physics, the art of modelling consists in deciding what is a system variable and what is a disturbance, in order to have a precise enough description of the way the control variables affect the system's evolution, but knowing that the inaccuracies will be compensated on the basis of real-time measurements of some of the system variables.

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_8, © Springer-Verlag Berlin Heidelberg 2009

to a diffeomorphism, reads $y^{(r+1)} = v$. If we set $v^* = (y^*)^{(r+1)}$, the error equation reads

$$e^{(r+1)} = v - v^* + w$$

where w is an unmeasured disturbance term.

It suffices then to set, componentwise:

$$v_i = v_i^* - \sum_{j=0}^{r} k_{i,j} e_i^{(j)} \ , \quad i = 1, \ldots, m \tag{8.1}$$

the gains $k_{i,j}$ being chosen such that the m polynomials $s^{r+1} + \sum_{j=0}^{r} k_{i,j} s^{(j)} = 0$ have their roots with strictly negative real part, $i = 1, \ldots, m$. Thus, if, e.g. $w(t)$ converges to 0 as $t \to \infty$, the error e exponentially converges[2] to 0:

$$e_i^{(r+1)} = -\sum_{j=0}^{r} k_{i,j} e_i^{(j)} + w_i \ , \quad i = 1, \ldots, m, \tag{8.2}$$

and y and all its derivatives up to order $r+1$ converge to their reference y^*, $\ldots$, $(y^*)^{(r+1)}$. Using the differentiability of the Lie-Bäcklund isomorphism

$$x = \varphi_0(y, \ldots, y^{(r)}), \quad u = \varphi_1(y, \ldots, y^{(r+1)})$$

we conclude that the set of variables x and u of the original system locally exponentially converge to their reference.

8.1.1 Pendulum (conclusion)

We go back to the endogenous dynamic feedback computed in section 6.4.3. Using (6.30) and (6.31), the closed-loop system reads, for $w_1 \neq 0$,

$$\begin{cases} \ddot{x} = (w_1 + \varepsilon\dot{\theta}^2)\sin\theta - \dfrac{\varepsilon}{w_1}(v_1\cos\theta - v_2\sin\theta - 2\dot{w}_1\dot{\theta})\cos\theta \\ \ddot{z} = (w_1 + \varepsilon\dot{\theta}^2)\cos\theta + \dfrac{\varepsilon}{w_1}(v_1\cos\theta - v_2\sin\theta - 2\dot{w}_1\dot{\theta})\sin\theta - 1 \\ \varepsilon\ddot{\theta} = \dfrac{\varepsilon}{w_1}(v_1\cos\theta - v_2\sin\theta - 2\dot{w}_1\dot{\theta}) \\ \ddot{w}_1 = v_1\sin\theta + v_2\cos\theta + w_1\dot{\theta}^2. \end{cases}$$

Thus, if v_1^* and v_2^* are the reference inputs generating the reference trajectories y_1^* and y_2^*, and if the whole state $(x, \dot{x}, z, \dot{z}, \theta, \dot{\theta})$ is measured, it suffices to choose

[2] Convergence may be proven under less restrictive assumptions on w. The corresponding results are beyond the scope of this book and are not presented here

$$v_i = v_i^* - \sum_{j=0}^{3} k_{i,j}(y_i^{(j)} - (y_i^*)^{(j)}), \quad i = 1, 2,$$

and to replace the $y_i^{(j)}$'s by their expressions in function of $(x, \dot{x}, z, \dot{z}, \theta, \dot{\theta}, w_1, \dot{w}_1)$, to have the local exponential convergence of all the variables $(x, \dot{x}, z, \dot{z}, \theta, \dot{\theta}, w_1, \dot{w}_1)$ to their respective reference.

8.1.2 Non Holonomic Vehicle (conclusion)

We go back to the endogenous dynamic feedback obtained in section 6.4.4, yielding the closed-loop system:

$$\begin{aligned}
\dot{x} &= u \cos\theta \\
\dot{y} &= u \sin\theta \\
\dot{\theta} &= \frac{1}{u}(-v_1 \sin\theta + v_2 \cos\theta) \\
\dot{u} &= v_1 \cos\theta + v_2 \sin\theta.
\end{aligned}$$

If the whole state (x, y, θ) is measured, and if the speed u doesn't vanish, one can set, as before,

$$v_1 = v_1^* - \sum_{j=0}^{1} k_{1,j}(x^{(j)} - (x^*)^{(j)}) , \quad v_2 = v_2^* - \sum_{j=0}^{1} k_{2,j}(y^{(j)} - (y^*)^{(j)})$$

with suitably chosen gains $k_{i,j}$'s, in order to ensure local exponential convergence of x, y and θ to their respective reference.

This construction can in fact be extended to the case where the speed vanishes, which is indeed desirable for the parallel parking problem, thanks to the following homogeneity property: dividing both sides of the system's equations (6.17) by a function λ of class C^∞ with respect to t, of constant sign on $[t_i, t_f]$, boils down to the change of time

$$\frac{d\tau}{dt} = \lambda(t)$$

and the change of input

$$v = \frac{u}{\lambda}.$$

The system reads:

$$\begin{aligned}
\frac{dx}{d\tau} &= \frac{\dot{x}}{\lambda} = \frac{u}{\lambda}\cos\theta = v\cos\theta \\
\frac{dy}{d\tau} &= \frac{\dot{y}}{\lambda} = \frac{u}{\lambda}\sin\theta = vsin\theta \\
\frac{d\theta}{d\tau} &= \frac{\dot{\theta}}{\lambda} = \frac{u}{\lambda l}\tan\varphi = \frac{v}{l}\tan\varphi.
\end{aligned}$$

We recover a system having the same form as the original one, with respect to the new time

$$\tau = \int_{t_i}^{t} \lambda(s)ds$$

and with the new input v.

Thus, it is possible to desingularize the system by a suitable choice of λ, namely such that $\lim_{t\to t_i} \frac{u(t)}{\lambda(t)} \neq 0$ and $\lim_{t\to t_f} \frac{u(t)}{\lambda(t)} \neq 0$ whereas $u(t_i) = 0$ and $u(t_f) = 0$.

It clearly suffices to choose $\lambda(t) = u^*(t)$, the reference speed realizing the parallel parking maneuver. Indeed, the reference v^* of v is $v^* = \frac{u^*}{u^*} = 1$ and, copying the previous control law in the non vanishing case, but now function of new time τ, we get:

$$\begin{aligned}
\frac{d^2x}{d\tau^2} &= v_1 = v_1^* - k_{1,0}(x - x^*) - k_{1,1}\left(\frac{dx}{d\tau} - \frac{dx^*}{d\tau}\right) \\
\frac{d^2y}{d\tau^2} &= v_2 = v_2^* - k_{2,0}(y - y^*) - k_{2,1}\left(\frac{dy}{d\tau} - \frac{dy^*}{d\tau}\right)
\end{aligned} \tag{8.3}$$

with

$$\begin{aligned}
v_1^* &= -\frac{1}{l}\sin\theta^*\tan\varphi^* \\
v_2^* &= \frac{1}{l}\cos\theta^*\tan\varphi^*.
\end{aligned} \tag{8.4}$$

Let us choose the gains $k_{i,j}$, $i = 1, 2$, $j = 0, 1$, as follows: $k_{1,1} = k_{2,1} = K_1 + K_2$ and $k_{1,0} = k_{2,0} = K_1K_2$, with $K_1, K_2 > 0$ and $K = \min(K_1, K_2)$. We easily verify that the solution of (8.3) satisfies the inequalities:

$$\begin{aligned}
|x(\tau(t)) - x^*(\tau(t))| &\leq \left(C_1\,|x(t_i) - x_i| + C_2\left|\frac{dx}{d\tau}(\tau(t_i))\right|\right) e^{-K\tau(t)} \\
|y(\tau(t)) - y^*(\tau(t))| &\leq \left(C_1'\,|y(t_i) - y_i| + C_2'\left|\frac{dy}{d\tau}(\tau(t_i))\right|\right) e^{-K\tau(t)}
\end{aligned} \tag{8.5}$$

where C_1, C_2, C_1' and C_2' are constants depending only on K_1 and K_2, which proves that the error with respect to the reference trajectory is monotonically decreasing with respect to $\tau(t)$. Since $\tau(t) = \int_{t_i}^{t} u^*(s)ds$ is the integral of the speed modulus, it represents the arc length between the initial point (x_i, y_i) and the point of coordinates $(x(t), y(t))$, so that τ is bounded above by L, the

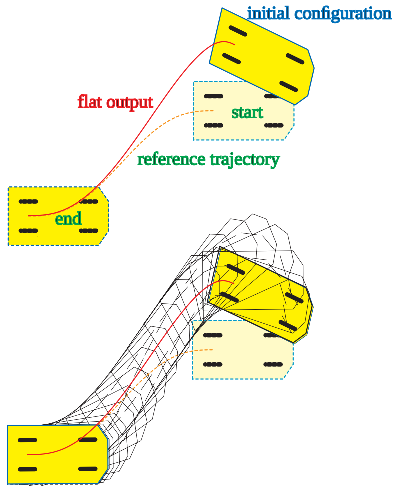

Fig. 8.1 Closed-loop parallel parking maneuver with an initial error on the position and orientation of the car. Top: trajectory of the flat output and its reference. Bottom: stroboscopic representation of the car's motion.

total arc length between (x_i, y_i) and (x_f, y_f). Thus, the error at the end of the maneuver is given by (8.5) for $\tau(t_f) = L$. Analogous inequalities are obtained for $\left|\frac{dx}{d\tau}(\tau(t)) - \frac{dx^*}{d\tau}(\tau(t))\right|$ and $\left|\frac{dy}{d\tau}(\tau(t)) - \frac{dy^*}{d\tau}(\tau(t))\right|$, which, combined with $v = \left(\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2\right)^{\frac{1}{2}}$, yields the monotonic decreasing property of the error $v - v^*$ and thus, if the initial errors are small enough, $v(\tau(t_f)) \approx v^*(t_f) = 1$ which achieves to justify this construction.

8.2 Control of the Clock

For systems requiring to be synchronized, each subsystem being submitted to independent perturbations, if the actuators are not powerful enough to maintain the errors as small as possible, or if the reference trajectories may be unrealistic in presence of some exceptional perturbations, we can use the so-called *control of the clock* method, consisting in slowing down the reference trajectory as long as the error between the measured state and its reference exceeds a given acceptable level.

Assume that we want to track the reference trajectory $t \mapsto y^*(t)$ with the control law (8.1).

Let us consider a function Γ at least continuous from $\mathbb{R}^{m(r+1)}$ to $\mathbb{R}$ such that $\Gamma(0) = 0$. We call $\rho(t)$ the trailing time on the reference trajectory, with

$0 \leq \rho(t) \leq t$ for all t. Since we are using the derivatives of y^* up to the order $r+1$, if we trail this trajectory at the modified time $\rho(t)$, this function ρ must be at least $r+1$ times differentiable with respect to t. It is left as an exercise to check that the $(r+1)$st derivative $\frac{d^{r+1}y^*(\rho(t))}{dt^{r+1}}$ of the composed function $y^*(\rho(t))$ with respect to t, depends on all the derivatives of ρ up to the order $r+1$.

Moreover, if we are only interested in the synchronization, we don't particularly want to recover the original time, lost during the slowing down phase ($\rho(t) - t \leq 0$). However, the synchronization requires that the rate $\frac{d}{dt}(\rho(t) - t) = \dot{\rho} - 1$ converges to 0. We thus set $\varepsilon = \dot{\rho} - 1$ and we want to tune this error dynamics in function of the error between the present state and its reference trajectory via the function Γ:

$$\varepsilon^{(r+1)}(t) = -\sum_{j=0}^{r} \kappa_j \varepsilon^{(j)}(t) - \Gamma(e(t), \dot{e}(t), \ldots, e^{(r+1)}(t)) \tag{8.6}$$

with the gains $\kappa_i > 0$ chosen to ensure the desired behavior.

Since $\varepsilon = \dot{\rho} - 1$ must range between -1 and 0, we replace $\varepsilon(t)$ by -1 as soon as the solution of (8.6) decreases below -1 and the solution of (8.6) is only taken into account if $\varepsilon > -1$. In other words, we replace ε by

$$\tilde{\varepsilon}(t) = \max(\varepsilon(t), -1).$$

We verify that if $\Gamma(e(t), \dot{e}(t), \ldots, e^{(r+1)}(t)) > 0$, when $\varepsilon = 0$, the $(r+1)$st derivative of ε decreases and, integrating $r+1$ times, $\dot{\rho} \leq 1$.

Now, if the the deviation between the real trajectory and its reference and its successive derivatives $\overline{e}$ remain almost constant for some times, the trailing speed stabilizes itself to the equilibrium value given by

$$-\kappa_0 \varepsilon - \Gamma(\overline{e}) = 0$$

which allows to tune κ_0 so that ε remains larger than -1, a value that would stop the system ($\dot{\rho} = 0$ implies $\varepsilon = -1$).

Moreover, one can verify that the closed-loop error equation (8.2) is not modified by this approach since we only modify the reference trajectory. In fact, grouping (8.2) and (8.6):

$$\begin{aligned} e_i^{(r+1)} &= -\sum_{j=0}^{r} k_{i,j} e_i^{(j)}\ , \quad i = 1, \ldots, m \\ \varepsilon^{(r+1)}(t) &= -\sum_{j=0}^{r} \kappa_j \varepsilon^{(j)}(t) - \Gamma(e(t), \dot{e}(t), \ldots, e^{(r+1)}(t)) \end{aligned} \tag{8.7}$$

we see that the error equation with respect to the reference trajectory, and thus its stability, is not affected by the control of the clock, whereas the error to the reference trajectory effectively influences the clock's delay.

A practical example of control of the clock is presented in Chapter 11 and in Lévine [2004].

Chapter 1
Introduction

This book is made of two parts, Theory and Applications.

In the first Part, two major problems of automatic control are addressed: *trajectory generation*, or *motion planning*, and *tracking* of these trajectories.

In order to make this book as self-contained as possible we have included a survey of Differential Geometry and Dynamical System Theory. The viewpoint adopted for these topics has been tailored to prepare the reader to the language and tools of flatness-based control design, that is why we have preferred to place them ahead in Chapters 2 and 3 rather than to release them in an Appendix.

Recalls of linear system theory are also provided in Chapter 4, such as controllability and the corresponding Brunovský canonical form, since they constitute a first solution to the trajectory generation and tracking problems, which are generalized in the next chapters to flat systems, using a different approach, leading to simpler calculations.

The last chapters (from Chapter 5 to Chapter 8), are then devoted to the analysis of Lie-Bäcklund equivalence and flat systems. Note that a large part of Chapter 6 is devoted to the characterization of flat systems. This part, not essential to understand the examples and applications of flatness all along, may be skipped at first reading. However, the reader interested in this essential but difficult theoretical aspect, still full of unsolved questions, may find there a self-contained presentation.

In the second Part, the applications have been selected according to their pedagogical potentials, to illustrate as many control design techniques as possible in various industrial contexts: control of various types of motors, magnetic bearings, cranes and aircraft automatic flight design.

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_1, © Springer-Verlag Berlin Heidelberg 2009

1.1 Trajectory Planning and Tracking

The problems of *trajectory generation*, or *motion planning*, and *tracking* of these trajectories are studied in the context of *finite dimensional nonlinear systems*, namely systems described by a set of nonlinear differential equations, influenced by a finite number of inputs, or control variables.

In practice, a system represents our knowledge of the evolution of some variables with respect to time, and the control variables are often designed as the inputs of the *actuators* driving the system. They may be freely chosen in order to achieve some tasks, or may be subject to constraints resulting from technological restrictions.

Numerous examples of such systems may be found in mechanical systems driven by motors (satellites, aircraft, cars, cranes, machine tools, *etc.*), electric circuits or electronic devices driven by input currents or voltages (converters, electromagnets, motors, *etc.*), thermal machines driven by heat exchangers or resistors, chemical reactors, chemical, biotechnological or food processes driven by input concentrations of some chemical components, or mixtures of these examples.

The notion of trajectory generation, or motion planning, corresponds to what we intuitively mean by preparing a flight plan or a motion plan in advance. More precisely, it consists in the off-line generation of a path, and the associated control actions that generate the path. This path is supposed to relate a prescribed initial point to a prescribed final point, in *open-loop*, *i.e.* based on the knowledge of the system model only, in the ideal case where disturbances are absent, and without taking account of possible measurements of the system state. Such a trajectory is often called *reference* or *nominal trajectory*, and the associated control the *reference* or *nominal control*. This notion is quite natural in the context of controlled mechanical systems such as aircraft, cars, ships, underwater vehicles, cranes, mechatronic systems, machine tools or positioning systems. It is also of interest in many other fields such as chemical, biotechnological or food processes, where we may want to change the concentration of a chemical component from its present value to another one in a fast but smooth way, for energy savings or productivity increase, or some other reason.

The tracking aspect concerns the design of a control law able to follow the reference trajectory even if some unknown disturbances force the system to deviate from it. For this purpose, this control law must take into account additional information, namely on-line measurements, or *observations*, from which the deviations at every time with respect to the reference trajectory can be deduced. In practice, such observations are provided by *sensors*. The class of controls that take into account the system state observations, is generally called *feedback* or *closed-loop* control. Without deviation (*i.e.* without disturbances), the control coincides with its reference, but as soon as a deviation is detected, the closed-loop control law must ensure the convergence of the system to its reference trajectory. The type of convergence (local, global, exponential, polynomial, *etc.*) that can be guaranteed, its rate, sometimes called *time constant* of the closed-loop system, and other robustness properties versus disturbances, modelling errors, *etc.* , will also be addressed in this book.

These two problems are particularly easy to solve for the class of nonlinear systems called *differentially flat*, or shortly *flat*, systems, introduced by M. Fliess, P. Martin, P. Rouchon and the author (Fliess et al. [1992a,b]) and actively developed since then (see e.g. the surveys and books by Martin et al. [1997], Lévine [1999], Rudolph [2003], Rudolph et al. [2003], Sira-Ramirez and Agrawal [2004], Rudolph [2003], Rudolph et al. [2003], Müllhaupt [2009]).

Most of the examples and applications of differential flatness of this book could have been presented using only elementary and intuitive mathematics. Though insufficiently precise for a mathematician, the mathematical ambiguities may be balanced by their physical evidence. However, if the reader wants to acquire a deeper understanding and/or wishes to solve more advanced problems, a precise mathematical background and a rigorous description of flat systems and their properties are required. Unfortunately, the corresponding mathematics are not easy. Their proper background comes from the theory of manifolds of jets of infinite order Krasil'shchik et al. [1986], Zharinov [1992]. Since at present no self-contained presentation of this theory for control systems is available, we have decided to privilege this aspect in this book, while keeping the mathematical level as accessible as possible. Nevertheless, applications also receive a prominent place in this book (Part II) to present flat systems from every angle.

1.2 Equivalence and Flatness

To give an intuitive idea of differential flatness, a flat system is a system whose integral curves (curves that satisfy the system equations) can be mapped in a one-to-one way to ordinary curves (which need not satisfy any differential equation) in a suitable space, whose dimension is possibly different than the one of the original system state space.

This definition can be made rigorous by introducing several notions and tools: we need to work with mappings that are one-to-one between vector spaces or manifolds of different dimension, and infinitely differentiable. According to the well-known *constant rank theorem* (see section 2.3), such mappings don't exist between finite dimensional manifolds. Therefore, it may only become possible if the original manifolds can be embedded in infinite dimensional ones. A classical way to realize this embedding consists in using the natural coordinates together with an infinite sequence of their time derivatives, called *jets of infinite order* (see *e.g.* Krasil'shchik et al. [1986], Zharinov [1992]).

In this framework, if two manifolds of jets of infinite order are mapped in a one-to-one and differentiable way, we say that they are *Lie-Bäcklund equivalent*. More precisely, two systems are said Lie-Bäcklund equivalent if and only if there exists a smooth one-to-one time-preserving mapping between their integral curves (trajectories that are solutions of the system differential equations) which maps tangent vectors to tangent vectors, in order to preserve time differentiation. Going back to our above stated intuitive definition of flatness, a flat system is Lie-Bäcklund equivalent to a system whose integral curves have no differential constraints (ordinary curves), that we call *trivial system*. Thus, finally, a system is flat if and only if it is *Lie-Bäcklund equivalent* to a *trivial system*.

Therefore, it becomes clear that the study of flat systems passes through the study of Lie-Bäcklund equivalence, a notion that plays a central role in this book. In addition, the notion of flatness may be interpreted as a change of coordinates that transforms the system in its "simplest" form, where calculations become elementary since the coordinates and the vector field describing the system are "straightened up". Recall that a transformation *straightens out coordinates, curves, surfaces, vector fields, distributions (families of vector fields), etc.* . if they are changed into lines, planes, constant vector fields, orthonormal frames, *etc.* . In particular, the integration of differential equations or partial differential equations in these coordinates may be done explicitly, as far as the associated straightening out transformations may be obtained.

These considerations indeed strongly suggest that the language of Differential Geometry is particularly well adapted to our context. However, the usual finite dimensional standpoint is too narrow for our purpose and its extension to manifolds of jets of infinite order seems difficult to circumvent. For the sake of completeness, we first introduce the reader to classic finite dimensional tools (Part I, Chapter 2), and then to their extension to jets of infinite order (Part I, Chapter 5).

Other approaches are indeed possible: finite dimensional differential geometric approaches Charlet et al. [1991], Franch [1999], Shadwick [1990], Sluis [1993], differential algebra and related approaches Fliess et al. [1995], Aranda-Bricaire et al. [1995], Jakubczyk [1993], infinite dimensional differential geometry of jets and prolongations Fliess et al. [1999], van Nieuwstadt et al. [1998], Pomet [1993], Pereira da Silva and Filho [2001], Rathinam and Murray [1998].

In the framework of linear finite or infinite dimensional systems, the notions of flatness and *parametrization* coincide as remarked by Pommaret [2001], Pommaret and Quadrat [1999], and in the behavioral approach of Polderman and Willems [1997], flat outputs correspond to *latent variables* of *observable image representations* Trentelman [2004] (see also Fliess [1992] for a module theoretic interpretation of the behavioral approach).

1.3 Equivalence in System Theory

Several equivalence relations have been studied to characterize *system equivalence* by various transformation groups. Traditionally, geometric objects are said to be *intrinsically* defined when their definition is not affected by change of coordinates (diffeomorphism) Boothby [1975], Chern et al. [2000], Choquet-Bruhat [1968], Demazure [2000], Dieudonné [1960], Kobayashi and Nomizu [1996], Olver [1995], Pham [1992]. In other words, two geometric objects are said equivalent if there exists a diffeomorphism mapping the first one into the second and *vice versa*. In the same spirit, *system equivalence by static feedback* has been introduced to deal with the equivalence of systems under static feedback action in an intrinsic way, namely independently of the choice of coordinates where the system and/or the control inputs are expressed. They yield classifications (*i.e.* partition of the set of systems into cosets) and canonical forms ("simplest" system representatives of the cosets) of major interest, such as the ones provided by Brunovský for linear controllable systems Brunovský [1970] (see also Rosenbrock [1970], Wolovich [1974], Tannenbaum [1980], Kailath [1980], Antoulas [1981], Polderman and Willems [1997], Sontag [1998] and, for extensions in the nonlinear case Sommer [1980], Jakubczyk and Respondek [1980], Hunt et al. [1983b], Marino [1986], Charlet et al. [1989, 1991], Gardner and Shadwick [1992], Isidori [1995], Nijmeijer and van der Schaft [1990], Marino and Tomei [1995]). However, equivalence relations which only involve static state feedback appear to be too fine to study flat systems. They are finer than the Lie-Bäcklund one which corresponds to the equivalence under a special class of dynamic feedback called *endogenous dynamic feedback* Martin [1992], Fliess et al. [1995], Martin [1994], van Nieuwstadt et al. [1994], Aranda-Bricaire et al. [1995], Pomet [1993], van Nieuwstadt et al. [1998], Fliess et al. [1999], Lévine [2006], that strictly contains the class of static feedback.

1.4 Equivalence and Stability

In the stability analysis of closed-loop systems, the notion of equivalence, though different than the previously discussed ones and called here *topological equivalence*, is also most important: in the introduction to dynamical system theory (Chapter 3), we emphasize on the equivalence between the behavior (stability or instability) of a nonlinear system around an equilibrium point and the one of its tangent linear approximation.

If the latter tangent linear approximation is *hyperbolic* (if it has no eigenvalue on the imaginary axis of the complex plane), the nonlinear system can be proved to be topologically equivalent to its tangent linear approximation. More precisely (Hartman-Grobman's Theorem), hyperbolic systems can be shown to be equivalent to a linear system made up with two decoupled linear subsystems, the first one being stable and the second one being unstable. These subsystems respectively live in locally defined *invariant manifolds* called *stable* and *unstable*, their respective dimensions corresponding to the number of eigenvalues, counted with their multiplicities, of the tangent linear approximation at the equilibrium point with negative and positive real parts.

In the non hyperbolic case, a nonlinear system may be shown to be topologically equivalent to a system made up with a linear stable subsystem and a linear unstable subsystem, obtained as before from the linear tangent approximation, and completed by a nonlinear neutral one, coupled to the previous linear ones. These subsystems respectively live in locally defined *stable*, *unstable* and *centre manifolds* (Shoshitaishvili's Theorem).

Singularly perturbed systems are introduced in this framework in Section 3.3, which extends the previous approach to control systems. We particularly insist on the links between singularly perturbed systems, multiple time scales and hierarchical control.

1.5 What is a Nonlinear Control System?

1.5.1 Nonlinearity versus Linearity

Before talking about nonlinearity, let us discuss the definition of linearity. First, linearity is a coordinate dependent property since a linear system might look nonlinear after a nonlinear change of coordinates. Take the following elementary example: $\dot{x} = u$ in a sufficiently small neighborhood of the initial condition $x_0 = 0$, and transform x into $\xi = \sqrt{x+1}$ and u into $v = u$. We have $\dot{\xi} = \frac{\dot{x}}{2\sqrt{x+1}} = \frac{u}{2\xi}$. Therefore, the transformed system, namely $\dot{\xi} = \frac{v}{2\xi}$ is no more linear.

Note that in the previous transformation, ξ doesn't depend on u and is invertible in the sense that $x = \xi^2 - 1$, and v is also invertible as a function of u[1]. Clearly, the set of transformations that enjoy these properties forms a group with respect to composition, and the linearity property of the system thus depends on this group[2]. More precisely, a system is said linear if it can be transformed into a linear system by a transformation of this group. The number of linear systems thus depends on the "size" of the group. This is why transformations depending on the input and its successive time derivatives, generating a larger group than the above mentioned one, will be introduced later.

[1] this transformation is actually a local C^∞ diffeomorphism: in addition to its local invertibility, it is of class C^∞ in a neighborhood of $x_0 = 0$, with C^∞ inverse.

[2] indeed, the smoothness of the transformations, which may be C^k for any $k \geq 2$ or analytic, is also part of the group definition.

Linear systems form a distinguished class in the set of nonlinear systems since they enjoy simpler properties as far as controllability, open-loop stability/instability, stabilizability, *etc.* are concerned. Therefore, they should be detected independently of the particular choice of coordinates in which they are expressed.

1.5.2 Uncontrolled versus Controlled Nonlinearity

In order to outline some fundamental differences between linear and nonlinear systems we may start with stability aspects for uncontrolled systems, by considering a linear system perturbed by a small nonlinearity, that significantly modifies the behavior of the original linear system. We next show that, once the system is controlled, what counts is the control efficiency to attenuate or remove the phenomenon created by the open-loop nonlinearity, as presented in the next example.

An introductory Example

This example is presented in three steps. We first start with a linear non controlled system, a spring with linear stiffness (force exerted by the spring proportional to its length variation), with a nonlinear perturbation, that may physically result from a defect of the spring, and modelled as a small nonlinear perturbation of the stiffness coefficient. It turns out that this small defect creates a big change in the system behavior, that doesn't exist in linear systems. At a second step, we connect the system with a passive device, that may be interpreted as a special case of feedback control, and show how the system behavior is locally modified. Finally, the third step consists in replacing the passive device by an active one to globally transform the original nonlinear system behavior into a linear one that may be tuned as we want.

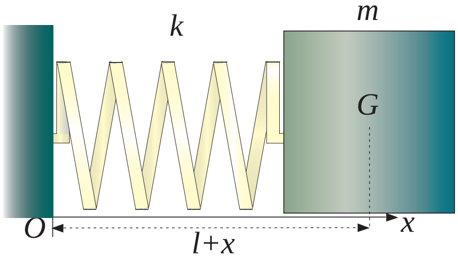

Fig. 1.1 Spring and mass

Uncontrolled nonlinear perturbation

Consider a system made of a mass and an undamped spring of pulsation ω whose position, denoted by x, satisfies :

$$\ddot{x} + \omega^2 x = 0 \tag{1.1}$$

with the spring stiffness k related to the pulsation ω by $\omega = \sqrt{\frac{k}{m}}$, m being the mass of the rigid body attached to the spring.

Setting $\dot{x} = v$, the expression $v^2 + \omega^2 x^2$, proportional to the mechanical energy of the spring, remains constant along any trajectory of (1.1) since $\frac{d}{dt}(v^2 + \omega^2 x^2) = 2\dot{x}(\ddot{x} + \omega^2 x) = 0$. In other words, in the (x, v)-plane (phase plane), these trajectories are the ellipses of equation $v^2 + \omega^2 x^2 = C$, where C is an arbitrary positive constant, and thus are closed curves around the origin, whose focuses are determined by the initial conditions (x_0, v_0). We indeed recover the classical interpretation that once the spring is released from its initial position x_0 with initial velocity v_0, it oscillates forever at the pulsation ω. This motion is neither attenuated nor amplified.

However, if the spring stiffness is not exactly a constant, even very close to it, but if this aspect has been neglected, a very different behavior may be expected.

Assume in fact that the spring stiffness is a linear slowly decreasing function of the length : $\frac{k(x)}{m} = \omega^2 - \varepsilon x$, with $\varepsilon > 0$ small, which means that the pulling force produced by the spring is $k(x)x = \omega^2 x - \varepsilon x^2$. The spring's dynamical equation becomes

$$\ddot{x} + \omega^2 x - \varepsilon x^2 = 0 \tag{1.2}$$

a nonlinear equation because of the x^2 term. Setting as before $v = \dot{x}$, we easily check that the expression (the mechanical energy up to a constant)

$$E_\varepsilon(x, v) = v^2 + x^2(\omega^2 - \frac{2}{3}\varepsilon x) \tag{1.3}$$

is such that $\frac{d}{dt}E_\varepsilon(x, v) = 0$ along the integral curves of (1.2), and thus remains constant with respect to time. The perturbed spring trajectories are therefore described by the curves of equation $E_\varepsilon(x, v) = E_\varepsilon(x_0, v_0)$ shown on Figure 1.2. We see that for a small initial length and velocity, the spring's behavior is not significantly changed with respect to the previous linear one. On the contrary, for larger initial length and velocity, the spring becomes too sluggish and thus unstable.

The differences with respect to the original linear system are thus twofold:

1. the only equilibrium point of the linear system (1.1) is the origin $(0, 0)$ whereas system (1.2) has two equilibria $(0, 0)$ and $(\frac{\omega^2}{\varepsilon}, 0)$;

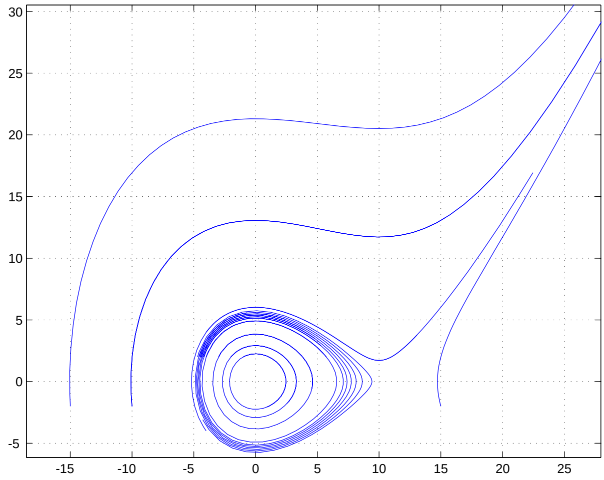

Fig. 1.2 Destabilization of the spring caused by its nonlinear stiffness

2. the linear system behavior is purely oscillatory, whereas the perturbed nonlinear one is oscillatory near the origin but unstable for larger initial conditions.

Adding a damper

This phenomenon is well-known on truck's trailers or on train wagon bogies where it is necessary to add a damper to dissipate the energy excess stored in the spring when released. In fact, the appending of a damper may be interpreted as a feedback: in (1.2), a frictional force Kv, proportional to the velocity, is added, which amounts to consider that the system is controlled by the force $u = Kv$:

$$\ddot{x} + \omega^2 x - \varepsilon x^2 + u = \ddot{x} + \omega^2 x - \varepsilon x^2 + Kv = 0. \tag{1.4}$$

Doing the previous calculation again, $\frac{dE_\varepsilon}{dt}$ along an arbitrary trajectory of (1.4), we find that $\frac{dE_\varepsilon}{dt} = -2Kv^2 < 0$, which proves that the function E_ε is monotonically decreasing along the trajectories of (1.4). It is readily seen that, for $|x| < \frac{\omega^2}{\varepsilon}$, the function E_ε is strictly convex and admits the origin $x = 0, v = 0$ as unique minimum. Consequently, the decreasing rate of E_ε along the trajectories such that $|x| < \frac{\omega^2}{\varepsilon}$ implies that the trajectories all converge to the origin, and thus that the system is stabilized thanks to the damper

Active control

The stability can be improved yet if the damper is replaced by an active hydraulic jack for instance. Indeed, if the force u produced by the damper can be modified at will, it suffices to set

$$u = -\omega^2 x + \varepsilon x^2 + K_1 x + K_2 v$$

with $K_1 > 0$ and $K_2 > 0$, and the equation (1.4) becomes the exponentially stable linear differential equation

$$\ddot{x} = -K_1 x - K_2 \dot{x}.$$

The thread followed in this simple example is quite representative of one of the main orientations of this course: we first analyze the nonlinearities that might influence the non controlled system, and then various feedback loop designs to compensate some or all of the unwanted dynamical responses are studied.

Chapter 9
DC Motor Starting Phase

This chapter is aimed at showing that, even for DC motor control, a quite standard application of control, described by a single input linear system, the so-called flatness-based approach may dramatically improve its performance in a transient phase.

More precisely, we consider a DC motor on which we test two control laws starting at rest to reach a given permanent angular speed in a given duration.

The first control law is a classical PID controller with a step speed reference, and the second one is the same PID controller with a flatness-based reference trajectory in place of the step reference, in order to evaluate the impact of the feedforward design.

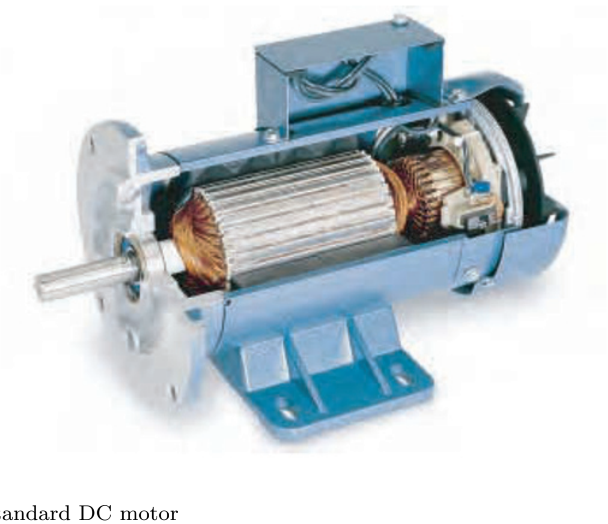

Fig. 9.1 A standard DC motor

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_9,

The motor has inductance L, resistance R, assumed constant, and torque constant K. Its inertia is denoted by J and its coefficient of viscous friction K_v. An unknown resistive torque C_r is applied to the motor. The current through the motor is denoted by I, the angular speed by ω and the input voltage by U, the control variable.

At the initial time t_i we have $\omega(t_i) = \dot{\omega}(t_i) = 0$, and, at the final time t_f (the duration is noted $T = t_f - t_i$), we want the motor to reach the angular speed $\omega(t_f) = \omega_f$ with $\dot{\omega}(t_f) = 0$.

The model is given by

$$\begin{aligned} L\frac{dI}{dt} &= U - RI - K\omega \\ J\frac{d\omega}{dt} &= KI - K_v\omega - C_r. \end{aligned} \tag{9.1}$$

This system is indeed flat with ω as flat output: setting $y = \omega$, all the system variables can be expressed as functions of y and derivatives up to second order:

$$\begin{aligned} \omega &= y \\ I &= \frac{1}{K}(J\dot{y} + K_v y + C_r) \\ U &= L\frac{dI}{dt} + RI + Ky \\ &= \frac{K^2 + RK_v}{K}y + \frac{RJ + LK_v}{K}\dot{y} + \frac{JL}{K}\ddot{y} \\ &\quad + \frac{R}{K}C_r + \frac{L}{K}\dot{C}_r. \end{aligned} \tag{9.2}$$

9.1 Tracking of a Step Speed Reference

A classical approach consists in choosing as reference trajectory of ω the speed step

$$\omega^*(t) = \begin{cases} 0 & \text{si } t_i \le t < t_i + \varepsilon \\ \omega_f & \text{si } t_i + \varepsilon \le t \le t_f \end{cases} \tag{9.3}$$

where ε is an arbitrary real number satisfying $0 \le \varepsilon < T$, generally chosen small compared to T.

We assume that the motor is endowed with an angular speed sensor (tachymeter). To track the step reference (9.3), the following Proportional-Integral-Derivative (PID) output feedback is used:

$$U = U^* - K_P(\omega - \omega^*) - K_D\dot{\omega} - K_I\int_{t_i}^{t}(\omega(\tau) - \omega^*(\tau))d\tau \tag{9.4}$$

where U^* is the steady state voltage corresponding to ω_f and $C_r = 0$ (recall that the resistive torque is unknown), *i.e.* $U^* = RI_f + K\omega_f$, with $KI_f = K_v\omega_f$, and thus $U^* = \left(K + \frac{RK_v}{K}\right)\omega_f$. Recall that the integral part of the controller (9.4) is designed to compensate the asymptotic angular deviation resulting from the ignorance of C_r.

For obvious thermal dissipation and safety reasons, the following limitations on the voltage, current and rate of current are considered:

$$|U| \leq U_{max}, \quad \left|\frac{dI}{dt}\right| \leq \delta, \quad |I| \leq I_{max}. \tag{9.5}$$

All along this section, the motor constants are $L = 3.6.10^{-3}$ H, $R = 1.71\ \Omega$, $K = 0.1$ Nm/s, $J = 6.10^{-5}$ Nms^2/rd and $K_v = 0.3.10^{-5}$ Nms^2/rd, the resistive torque C_r is equal to 0.5 Nm and the constraints are $U_{max} = 25$ V, $\delta = 100$ A/s and $I_{max} = 10$ A.

In the next simulations, the initial angular speed is $\omega(t_i) = 0$, the final one is $\omega_f = 30$ rd/s and the duration $T = 0.1$s. We also make an initial error on the angular speed of 0.087 rd/s ($\approx 5°$/s). The gains of the PID (9.4) are $K_P = 0.056$, $K_I = 7.45$ and $K_D = 10^{-5}$. They correspond to time constants equal to 10^{-2}, $6.6.10^{-3}$ and 4.10^{-3}s.

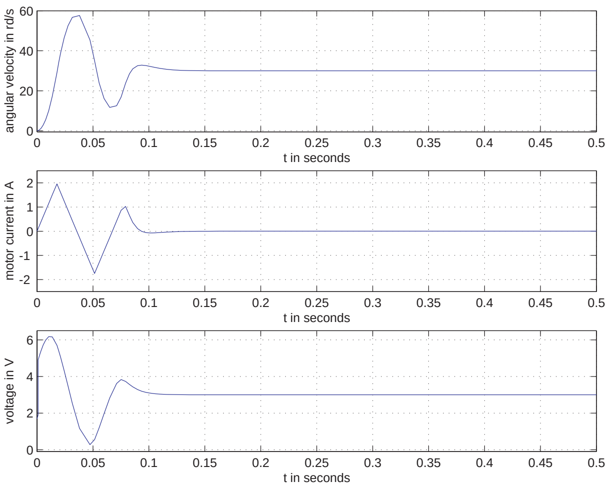

Fig. 9.2 Tracking of the angular speed from 0 to 30 rd/s with a current rate limitation of 100 A/s.

Figure 9.2 depicts a simulation of the closed-loop system using the PID (9.4) with a current rate limitation of 100 A/s. The objective of 30 rd/s is reached after a relatively hectic transient of about 0.1 s. One may remark that the current slope remains saturated during more than 0.07 s during the starting phase. The stabilization objective is reached within the required 0.1 s but with an inacceptable speed overshoot of about twice the desired value ω_f. In fact, the current rate limitation is too strict to let the motor exponentially converge to its reference yielding a slower convergence.

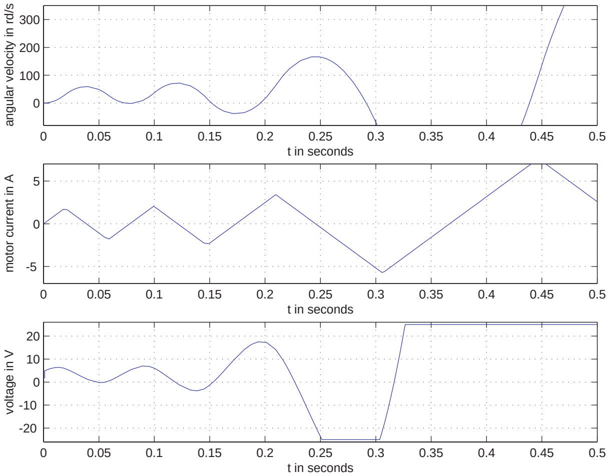

Fig. 9.3 Angular speed step from 0 to 30 rd/s with a current rate limitation of 95 A/s PID tracking.

In Ffigure 9.3, the current rate constraint is slightly stengthened to 95 A/s, all the other constraints remaining unchanged. An unstable response is obtained, which proves that the previous bound of 100 A/s is approximately the lower limit under which the stability is no more guaranteed.

In fact, an increase of this current rate bound up to 500A/s is necessary for this constraint to stop being active, without modifying the bounds of the other constraints. The exponential convergence is thus recovered and effectively corresponds to the pole placement performed by the PID, as shown in Figure 9.4. Note that, even in this case, an overshoot of about 30% of ω_f's value remains.

9.2 Flatness Based Tracking

Another approach consists in replacing the speed step reference (9.3) by a smooth enough trajectory that the motor can effectively follow, deduced from the system flatness. In other words, from a reference trajectory of the angular speed, which is actually a flat output, one can deduce the open-loop voltage

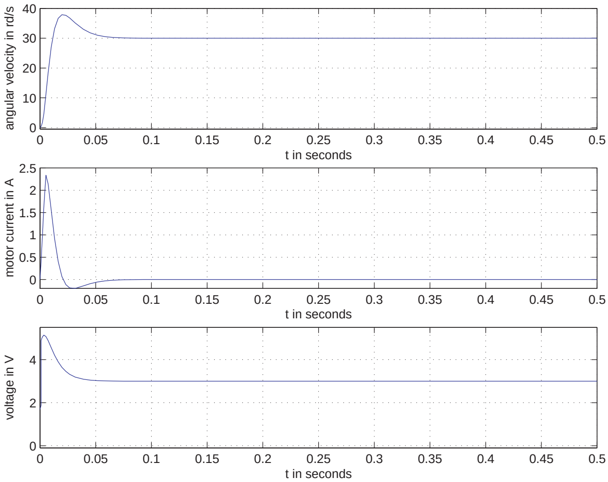

Fig. 9.4 Angular speed step from 0 to 30 rd/s with a current rate limitation of 500 A/s PID tracking.

and current reference trajectories that generate, in the absence of perturbations and modelling errors, the required speed trajectory, assuming that the constraints (9.5) are satisfied. These trajectories are then used in the PID (9.4) to replace the references U^*, ω^* and $\dot{\omega}^*$.

Since the initial and final conditions are $\omega(t_i) = 0$, $\dot{\omega}((t_i)) = 0$ and $\omega(t_f) = \omega_f$, $\dot{\omega}(t_f) = 0$, one can construct an at least twice continuously differentiable reference trajectory ω^{**} of the angular speed ω by polynomial interpolation:

$$\omega^{**} = \omega_f \left(\frac{t - t_i}{T}\right)^2 \left(3 - 2\left(\frac{t - t_i}{T}\right)\right). \tag{9.6}$$

We then deduce the reference current and voltage I^{**} and U^{**} by (9.6) and (9.2), assuming that the unknown resistive torque C_r vanishes:

$$\begin{aligned} I^{**} &= \frac{1}{K}\left(J\frac{d\omega^{**}}{dt} + K_v\omega^{**}\right) \\ U^{**} &= L\frac{dI^{**}}{dt} + RI^{**} + K\omega^{**}. \end{aligned} \tag{9.7}$$

One can easily verify that the trajectory $t \mapsto \omega^{**}(t)$ is non decreasing, and thus that $\omega^{**}(t) \leq \omega_f$ for all $t \leq t_f$ which proves that this reference trajectory has no overshoot.

Furthermore, one easily checks that the maximum current on this trajectory is 0.27 A, that the maximal current rate is 9.8 A/s and that the maximum voltage is 3 V.

Replacing ω^* by the flatness-based reference trajectory ω^{**} in the PID (9.4) we get:

$$U = U^{**} - K_P(\omega - \omega^{**}) - K_D(\dot{\omega} - \dot{\omega}^{**}) \\ -K_I \int_{t_i}^{t} (\omega(\tau) - \omega^{**}(\tau))d\tau. \tag{9.8}$$

The gains K_P, K_I and K_D, as well as the errors and constraints are the same

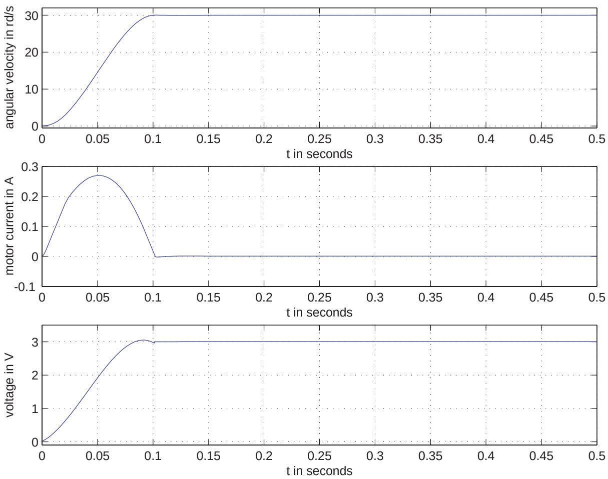

Fig. 9.5 Tracking of the flatness-based angular speed reference trajectory from 0 to 30 rd/s with the PID (9.8).

as in the previous paragraph with the speed step reference. Nevertheless, we obtain an exponential convergence without overshoot, the final angular speed being reached in 0.1 s with a precision of about $\pm$ 2.10^{-3} rd/s ($\approx 0.1°$/s) (see figure 9.5). Let us stress that the maximum current rate limitation is **50 times smaller** than with the speed step for a comparable exponential convergence, thus implying significant energy savings and decreased motor wear for lower power dissipation.

To conclude, with the same PID, the same gains and in presence of the same perturbations, we obtain a significantly different behavior according to the reference trajectory design in the PID. The main advantage of the flatness-based reference trajectory design is here that the reference of ω is such that the motor can track it without saturating the constraints, and that the evolution of the voltage reference U^{**} is at each time compatible with the speed reference ω^{**}. Let us remark in addition that we have kept the same PID for simplicity's sake. However, the flatness-based reference trajectory design clearly allows a significant increase of the PID gains without risk of saturation and we can even expect greater performance and/or robustness improvements.

Chapter 10
Displacements of a Linear Motor With Oscillating Masses

The aim of this chapter[1] is to show, on a barely undamped oscillating system, roughly approximated by a simple single-input linear model, the importance of feedforward design to simultaneously achieve, with the same smooth controller (no switch), high precision positioning, fast displacements, and robustness versus modelling errors, objectives which are, at first sight, antinomic.

We consider a linear motor moving along a rail and, in a first step, related by one flexible rod to an auxiliary mass (see Figure 10.1), and in a second step, by two different flexible rods to two different auxiliary masses (see Figure 10.8). The motor is assumed to be controlled by the force it delivers. All along this study, we assume that the motor position and speed are measured in real time with an arbitrary accuracy, but that the auxiliary mass positions are not measured.

Our aim is to generate fast rest-to-rest displacements of the motor with a motor positioning as accurate as possible. Indeed, when the force is non zero, the motor accelerates, producing oscillations of the auxiliary bodies, and, even if their masses are small compared to the motor's mass, the flexible rods transmit oscillations to the motor, and every controller reaction needed to compensate the corresponding deviation from the motor set point, not only results in a deterioration of the positioning accuracy but in energy expenses and thermal losses. Thus, it is clear that cancelling the oscillations of the auxiliary bodies might be useful. What is less obvious is that, at the same time, the robustness of the controller versus modelling errors is not deteriorated.

---

[1] Work partially done in collaboration with the company Micro-Controle/Newport. Related Patent: US 20020079198A1. Particular thanks are due to Dr. D.V. Nguyen for his help to realize the set-up displayed in Figures 10.1 and 10.8. Videos are available at the URL `http://cas.ensmp.fr/ levine` (Pendules et Systèmes flexibles).

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_10, © Springer-Verlag Berlin Heidelberg 2009

10.1 Single Mass Case

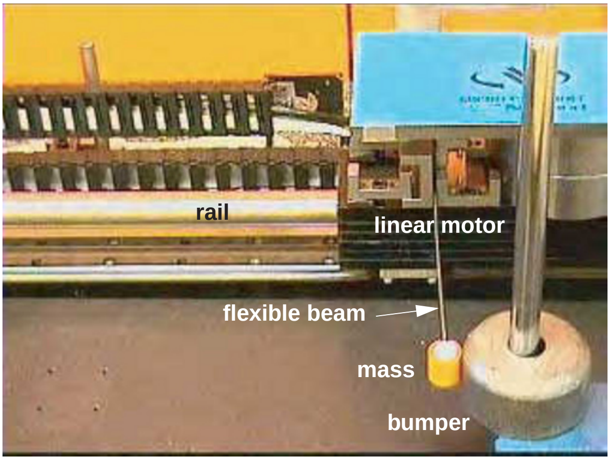

Fig. 10.1 Single mass elastically related to a linear motor.

The total mass of the motor is M. The auxiliary body's mass, including the flexible rod's mass, is m. The flexible rod may be interpreted as a spring and damper, with spring stiffness k and (linear) damping coefficient r. The position of the center of gravity of the motor (resp. of the auxiliary body) is denoted by x (resp. z). At rest, we assume that $x = z$.The motor delivers a force F, considered as the control variable.

In a first approximation[2], the system reads

$$\begin{aligned} M\ddot{x} &= F - k(x-z) - r(\dot{x} - \dot{z}) \\ m\ddot{z} &= k(x-z) + r(\dot{x} - \dot{z}) \end{aligned} \tag{10.1}$$

The most common approach consists in ignoring the influence of the auxiliary mass, thus considered as a perturbation, and in tuning a PID to track rest-to-rest reference trajectories of the motor and obtain the required positioning precision, in the same spirit as in (9.4) of the previous chapter.

This approach will be compared to the one where the auxiliary mass is taken into account and where the reference trajectories are rest-to-rest for both the motor and the auxiliary mass, while the initial and final positions of the motor are the same and with the same displacement duration. Note that such reference trajectories are easily computed thanks to the flatness

[2] In theory, the flexible rod is modelled by a partial differential equation, *i.e.* an infinite dimensional system.The finite dimensional model (10.1) may be interpreted as a modal approximation at the second order.

property of system (10.1) and that, in this approach, the auxiliary mass and the motor stop simultaneously, at least if the characteristics k and r are precisely estimated.

The herebelow presented data are simulation results. The reader can verify that they are fully confirmed by the previously mentioned videos of the Micro-Controle/Newport set-up[3].

In what follows, we use $M = 4$ kg for the mass of the motor, $m = 0.15$ kg for the auxiliary mass, $k = m(4 \cdot 2\pi)^2$ N/m for the stiffness of the rod (corresponding to a resonance frequency of 4 Hz), $r = 2 \cdot 0.01\sqrt{km}$ Ns/m for the damping coefficient of the rod (corresponding to a damping time constant of $\frac{1}{0.01}$ s).

10.1.1 Displacement Without Taking Account of the Auxiliary Mass

If we don't take account of the auxiliary mass the system reads

$$M\ddot{x} = F + p \tag{10.2}$$

the unknown disturbance p including the contribution of the auxiliary mass.

If we neglect the disturbance p, the following reference trajectory

$$\begin{aligned} x_{ref}(t) = x_0 + (x_1 - x_0)\left(\frac{t}{T}\right)^4 \\ \times \left(35 - 84\left(\frac{t}{T}\right) + 70\left(\frac{t}{T}\right)^2 - 20\left(\frac{t}{T}\right)^3\right) \end{aligned} \tag{10.3}$$

realises a rest-to-rest displacement starting from x_0 at rest at time 0 and ending at x_1 at rest at time T, with the force and its derivative equal to 0 at times 0 and T.

One can verify that the reference of the force F_{ref} is given by

$$\begin{aligned} F_{ref} = M\ddot{x}_{ref} = 420\ M\left(\frac{x_1 - x_0}{T^2}\right)\left(\frac{t}{T}\right) \\ \times \left(1 - 4\left(\frac{t}{T}\right) + 5\left(\frac{t}{T}\right)^2 - 2\left(\frac{t}{T}\right)^3\right). \end{aligned}$$

To attenuate the sensitivity of the motor displacement to perturbations, we add, to the reference force, the following PID loop:

---

[3] available at the URL `http://cas.ensmp.fr/ levine` (Pendules et Systèmes flexibles)

$$F = F_{ref} - k_P\left(x - x_{ref}\right) - k_D\left(\dot{x} - \dot{x}_{ref}\right) \\ -k_I \int_0^t \left(x(\tau) - x_{ref}(\tau)\right) d\tau \tag{10.4}$$

whose gains are $k_P = 18260.10^4$, $k_D = 53950$ and $k_I = 13280.10^7$. They approximately correspond to eigenvalues -8.10^3, -4.10^3, -10^3 and $-0,25 \pm 25,13i$ of the closed-loop system (10.1) with (10.4), the two last complex conjugated eigenvalues corresponding to the auxiliary mass dynamics $m\ddot{z} = -kz - r\dot{z}$. The latter eigenvalues cannot be significantly modified by the output feedback (10.4), since the error on z and $\dot{z}$ are not taken into account. The corresponding time constants, except those of the auxiliary mass, are approximately 10^{-3}s, $1,25.10^{-4}$s and $2,5.10^{-4}$s.

10.1.2 Displacements Taking Account of the Auxiliary Mass

To prevent the auxiliary mass from oscillating, without measuring its position, we use the model (10.1) to find rest-to-rest reference trajectories of the motor and the auxiliary mass.

One can express x, z and F in function of a flat output of system (10.1) and its successive derivatives and deduce a rest-to-rest trajectory for both the motor and the auxiliary mass starting from $x_0 = z_0 = 0$ at time $t = 0$ and arriving at $x_1 = z_1 = 0.1$ m at time $t = T$.

Recall from Theorem 6.9 the flat output computation method. We first rewrite (10.1) as:

$$\begin{pmatrix} (M\frac{d^2}{dt^2} + r\frac{d}{dt} + k) & -(r\frac{d}{dt} + k) \\ -(r\frac{d}{dt} + k) & (m\frac{d^2}{dt^2} + r\frac{d}{dt} + k) \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} F$$

and denote by $A(\frac{d}{dt}) = \begin{pmatrix} (M\frac{d^2}{dt^2} + r\frac{d}{dt} + k) & -(r\frac{d}{dt} + k) \\ -(r\frac{d}{dt} + k) & (m\frac{d^2}{dt^2} + r\frac{d}{dt} + k) \end{pmatrix}$ and $B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We immediately see that $C^T = \begin{pmatrix} 0 & 1 \end{pmatrix}$ is such that $C^T B = 0$. Thus, $P(F) = C^T A(\frac{d}{dt}) = \begin{pmatrix} -(r\frac{d}{dt} + k) & (m\frac{d^2}{dt^2} + r\frac{d}{dt} + k) \end{pmatrix}$. We are looking for a polynomial matrix $\Theta = \begin{pmatrix} P_x \\ P_z \end{pmatrix}$ solution of $P(F)\Theta = 0$. The reader may verify that the right-Smith decomposition of $P(F)$ is given by $P(F)U = \begin{pmatrix} 1 & 0 \end{pmatrix}$ with

$$U = \frac{r^2}{mk^2} \begin{pmatrix} \frac{m}{r}\frac{d}{dt} + \left(1 - \frac{mk}{r^2}\right) & m\frac{d^2}{dt^2} + r\frac{d}{dt} + k \\ 1 & r\frac{d}{dt} + k \end{pmatrix}$$

and thus $\Theta = \begin{pmatrix} P_x \\ P_z \end{pmatrix} = \begin{pmatrix} m\frac{d^2}{dt^2} + r\frac{d}{dt} + k \\ r\frac{d}{dt} + k \end{pmatrix} P_0$ where P_0 is an arbitrary non zero real number. Choosing $P_0 = \frac{1}{k}$, we get:

$$\begin{array}{l} x = y + \frac{r}{k}\dot{y} + \frac{m}{k}\ddot{y}, \quad z = y + \frac{r}{k}\dot{y} \\ F = (M+m)\left(\ddot{y} + \frac{r}{k}y^{(3)} + \frac{Mm}{(M+m)k}y^{(4)}\right) \end{array} \tag{10.5}$$

These relations define y in a unique way:

$$y = \frac{r^2}{mk}x + \left(1 - \frac{r^2}{mk}\right)z - \frac{r}{k}\dot{z}. \tag{10.6}$$

Note that the physical meaning of this expression is not obvious. One can however remark that, if one interprets (10.5) as the fact that y is the input of a filter whose output is x and z, formula (10.6) is the inverse of this filter. Furthermore, since $y = z + \frac{r}{k}\left(\frac{r}{m}(x-z) - \dot{z}\right)$, if the rod's stiffness is large enough ($\frac{r}{k} \approx 0$), and if the upper bounds of the deviation $x - z$ and the speed $\dot{z}$ are not too large, we have $y \approx z$, namely the flat output corresponds to the position of the auxiliary mass.

Nevertheless, the expression (10.6) is not necessary to compute the reference trajectories, as will be shown in the next construction where only (10.5) is used.

Let us denote by y_{ref} the required reference trajectory of y. Since we start from $x_0 = z_0$ at time $t = 0$ at rest, thus with null derivatives of x and z up to order 2, and with $F_0 = 0$, and arrive at $x_1 = z_1$ at time T at rest, *i.e.* with null derivatives of x and z up to order 2, with $F_1 = 0$, we deduce from formulas (10.5) that

$$\begin{array}{l} y_{ref}(0) = x_0, \ \dot{y}_{ref}(0) = \ddot{y}_{ref}(0) = y^{(3)}_{ref}(0) = y^{(4)}_{ref}(0) = 0 \\ y_{ref}(T) = x_1, \ \dot{y}_{ref}(T) = \ddot{y}_{ref}(T) = y^{(3)}_{ref}(T) = y^{(4)}_{ref}(T) = 0. \end{array}$$

By polynomial interpolation, we get

$$\begin{aligned} y_{ref}(t) = x_0 + (x_1 - x_0)\left(\frac{t}{T}\right)^5 \\ \times \left(126 - 420\left(\frac{t}{T}\right) + 540\left(\frac{t}{T}\right)^2 \right. \\ \left. -315\left(\frac{t}{T}\right)^3 + 70\left(\frac{t}{T}\right)^4\right). \end{aligned} \tag{10.7}$$

The reference trajectories x_{ref} of x and F_{ref} of F are thus obtained by replacing y by y_{ref} given by (10.7) in (10.5), and then F is obtained by plugging the obtained expressions of x_{ref}, $\dot{x}_{ref}$ and F_{ref} in (10.4). The gains

$k_P = 18260.10^4$, $k_D = 53950$ and $k_I = 13280.10^7$ of the PID are the same as before.

10.1.3 Comparisons

The simulation results are presented for both approaches for T=0.5s, 0.25s or 0.2s, for $k = m(4 \times 2\pi)^2$, +10%, +20%, and for $r = 2 \times 0.01\sqrt{km}$, -10%, and -20%, in Table 10.1 and in Figures 10.3 to 10.7. The notation Δx, denotes the maximal amplitude of the deviation between the motor position and its reference after T, Δz, the maximal amplitude of the oscillations of the auxiliary mass after T, ΔF, the maximal amplitude of the force after T and $|F_{max}|$ the maximal force in absolute value between 0 and T.

		1st approach				**2nd approach**			
k, r	T (s)	Δx (m)	Δz (m)	ΔF (N)	$\|F_{max}\|$ (N)	Δx (m)	Δz (m)	ΔF (N)	$\|F_{max}\|$ (N)
	0,5	2.10^{-10}	8.10^{-3}	$0,8$	13	10^{-12}	6.10^{-4}	0,06	10
exact	0,25	2.10^{-9}	0,1	10	50	$1,2.10^{-12}$	8.10^{-4}	0,06	80
	0,2	2.10^{-9}	0,12	14	80	$1,2.10^{-12}$	4.10^{-4}	0,06	200
	0,5	8.10^{-11}	4.10^{-3}	0,4	13	4.10^{-11}	2.10^{-3}	0,2	10
1,1k, 0,9r	0,25	2.10^{-9}	0,1	10	50	2.10^{-10}	0,01	1	80
	0,2	$2,5.10^{-9}$	0,12	13	80	$2,5.10^{-10}$	0,013	1,4	200
	0,5	2.10^{-11}	10^{-3}	0,1	13	3.10^{-11}	$1,6.10^{-3}$	0,3	10
1,2k, 0,8r	0,25	2.10^{-9}	0,1	10	50	4.10^{-10}	0,02	2	80
	0,2	3.10^{-9}	0,12	14	80	4.10^{-10}	0,02	3	200

Table 10.1 Comparison between the two approaches

We notice that the first approach is relatively robust with respect to errors on k and r. However, the positioning precision dramatically depends on the displacement duration T. For $T = 0,5$s, the precision is significantly better than for $T = 0,25$ or $0,2$s because of the large oscillations of the auxiliary mass: the latter oscillations have an amplitude of approximately 0.1m, the same order of magnitude as the full displacement itself. The positioning precision is thus obtained by spending a big amount of energy, the amplitude of

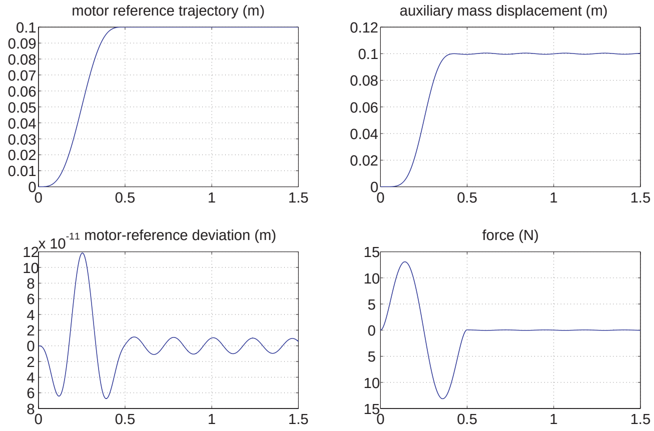

Fig. 10.2 Auxiliary mass taken as a perturbation (1st approach), duration $T = 0,5$s, error of 20% on k and r.

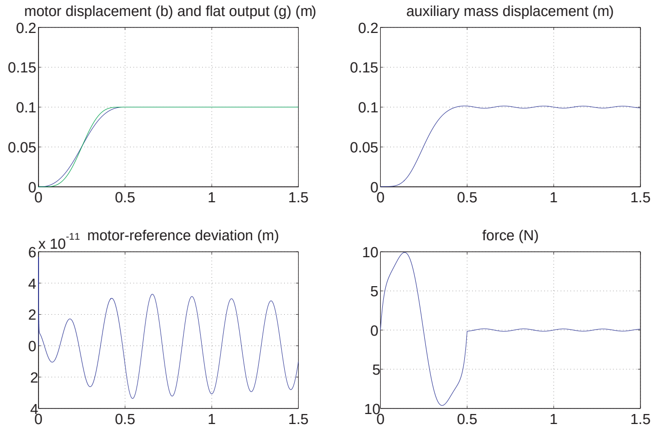

Fig. 10.3 Auxiliary mass taken into account (2nd approach), duration $T = 0,5$s, error of 20% on k and r.

the force necessary to maintain the motor at its final position being of the order of 10N.

This contrasts with the flatness-based approach where the precision is about 10 times finer, with a comparable robustness to errors on k and r, and a relative insensitivity of the precision with respect to the duration T. For $T \leq 0,25$s, even if k and r are erroneous, the amplitude of the auxiliary mass is about 10 times smaller and the force necessary to maintain the motor is also at least 5 times smaller. On the other hand, the maximal force needed to track the reference trajectory is about twice larger than with the first approach.

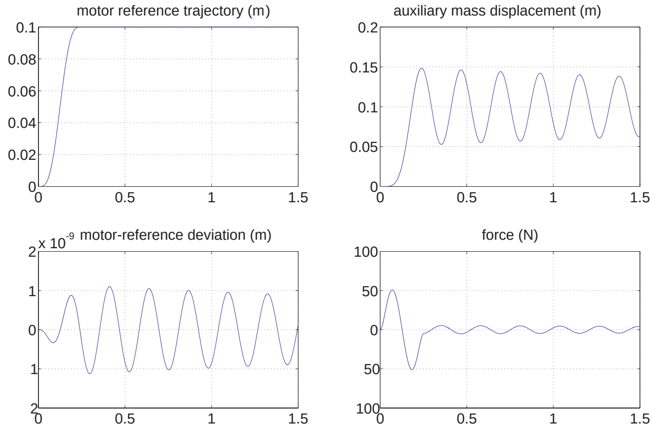

Fig. 10.4 Auxiliary mass taken as a perturbation (1st approach), duration $T = 0,25$s, error of 20% on k and r.

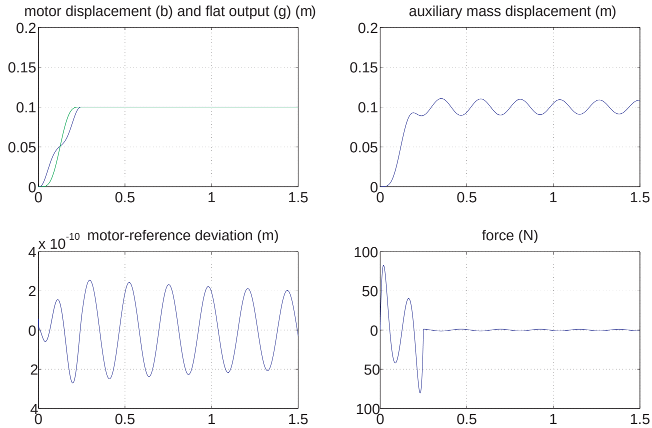

Fig. 10.5 Auxiliary mass taken into account (2nd approach), duration $T = 0,25$s, error of 20% on k and r.

We also remark that for $T = 0,2$s, the motor trajectory is no more a monotone function of time since, once arrived at the middle of the displacement, the motor goes backward before starting again onwards to its final position in order to ensure that the auxiliary mass arrives at rest (see Figure 10.7 top-left).

To conclude, with the same tunings, the second approach provides a significant improvement of the precision without deterioration of the robustness for comparable displacement durations and a smaller energy expense to maintain the precision at the final position, considering that the energy spent is proportional to the integral along the duration T of the squared force. How-

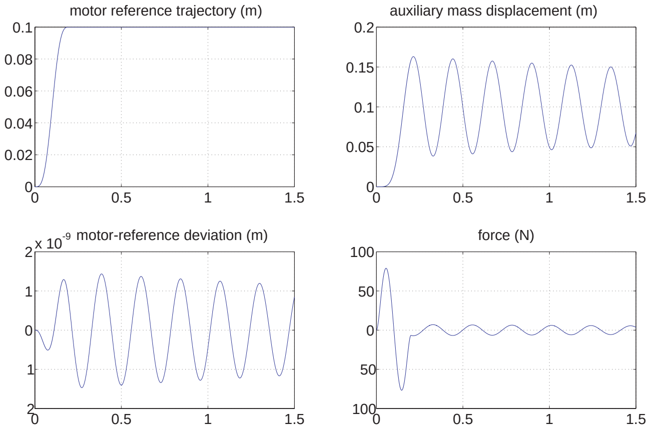

Fig. 10.6 Auxiliary mass taken as a perturbation (1st approach), duration $T = 0,2$s, error of 20% on k and r.

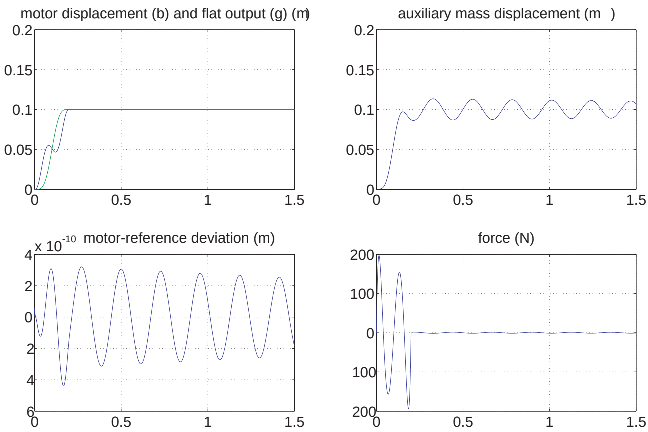

Fig. 10.7 Auxiliary mass taken into account (2nd approach), duration $T = 0,2$s, error of 20% on k and r.

ever, the duration cannot be decreased below the natural oscillation period of the auxiliary mass, whose frequency is 4Hz, *i.e.* a period equal to 0,25s. Below this duration, the reference trajectory may go backwards and have an important overshoot, yielding excessive force peaks, which constitutes a limit to the 2nd approach.

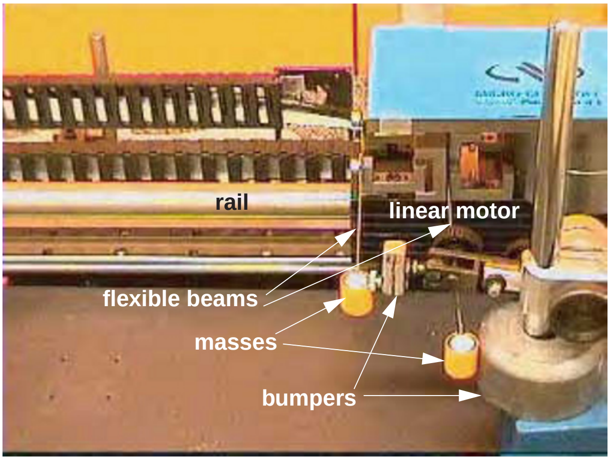

Fig. 10.8 Two masses elastically related to a linear motor.

10.2 Displacements With Two Auxiliary Masses

Let us go on with the previous comparison with a second mass $m' = 0.04$ kg related to the motor via a second flexible rod (see Figure 10.8). The characteristics of the second rod are $k' = m'(7 \cdot 2\pi)^2$ N/m (characteristic frequency of 7 Hz) and $r' = 2 \cdot 0.012\sqrt{k'm'}$ Ns/m (damping time constant of $\frac{1}{0.012}$ s).

The system is given by

$$\begin{aligned} M\ddot{x} &= F - k(x-z) - r(\dot{x}-\dot{z}) - k'(x-z') - r'(\dot{x}-\dot{z}') \\ m\ddot{z} &= k(x-z) + r(\dot{x}-\dot{z}) \\ m'\ddot{z}' &= k'(x-z') + r'(\dot{x}-\dot{z}'). \end{aligned} \tag{10.8}$$

The first approach follows exactly the same lines as in the previous section: the two auxiliary masses being considered as perturbations, the system is given by (10.2) and the control law by (10.4). with the same gains $k_P = 18260.10^4$, $k_D = 53950$ and $k_I = 13280.10^7$.

For the second approach, we first compute a flat output y according to the same principle as in the single mass case, using Theorem 6.9:

$$\begin{aligned} x &= y + \left(\frac{r}{k} + \frac{r'}{k'}\right)\dot{y} + \left(\frac{m}{k} + \frac{m'}{k'} + \frac{rr'}{kk'}\right)\ddot{y} \\ &\qquad + \left(\frac{mr' + rm'}{kk'}\right)y^{(3)} + \frac{mm'}{kk'}y^{(4)} \\ z &= y + \left(\frac{r}{k} + \frac{r'}{k'}\right)\dot{y} + \left(\frac{m'}{k'} + \frac{rr'}{kk'}\right)\ddot{y} + \frac{rm'}{kk'}y^{(3)} \\ z' &= y + \left(\frac{r}{k} + \frac{r'}{k'}\right)\dot{y} + \left(\frac{m}{k} + \frac{rr'}{kk'}\right)\ddot{y} + \frac{mr'}{kk'}y^{(3)} \end{aligned} \tag{10.9}$$

$$
\begin{aligned}
F = \hat{M}\ddot{y} + \hat{M}\left(\frac{r}{k} + \frac{r'}{k'}\right) y^{(3)} \\
+ \left(\frac{m}{k}\bar{M}' + \bar{M}\frac{m'}{k'} + \hat{M}\frac{rr'}{kk'}\right) y^{(4)} \\
+ \left(\frac{mr'}{kk'}\bar{M}' + \bar{M}\frac{rm'}{kk'}\right) y^{(5)} + \frac{Mmm'}{kk'} y^{(6)}
\end{aligned}
\tag{10.10}
$$

with the notations $\hat{M} = (M + m + m')$, $\bar{M} = M + m$ and $\bar{M}' = M + m'$.

As in the single mass case, one can express y as a linear combination of $x, \dot{x}, z, \dot{z}, z', \dot{z}'$. The result being useless for rest-to-rest trajectory computations, we leave it to the reader as an exercise.

To generate rest-to-rest trajectories, formula (10.7) is modified as follows:

$$
\begin{aligned}
y(t) = x_0 + (x_1 - x_0)\left(\frac{t}{T}\right)^7 \\
\times \Bigg(1716 - 9009\left(\frac{t}{T}\right) + 20020\left(\frac{t}{T}\right)^2 \\
-24024\left(\frac{t}{T}\right)^3 16380\left(\frac{t}{T}\right)^4 \\
-6006\left(\frac{t}{T}\right)^5 + 924\left(\frac{t}{T}\right)^6\Bigg).
\end{aligned}
\tag{10.11}
$$

We keep the same PID control law (10.4) with the reference force and trajectory F_{ref} and x_{ref} deduced from (10.9), (10.10), (10.11), still with the gains $k_P = 18260.10^4$, $k_D = 53950$ and $k_I = 13280.10^7$.

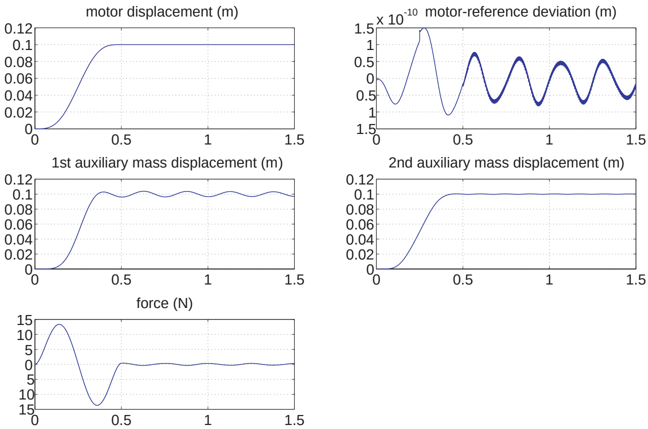

Fig. 10.9 Auxiliary masses taken as perturbations (1st approach), duration $T = 0,5$s, error of 20% on k and r.

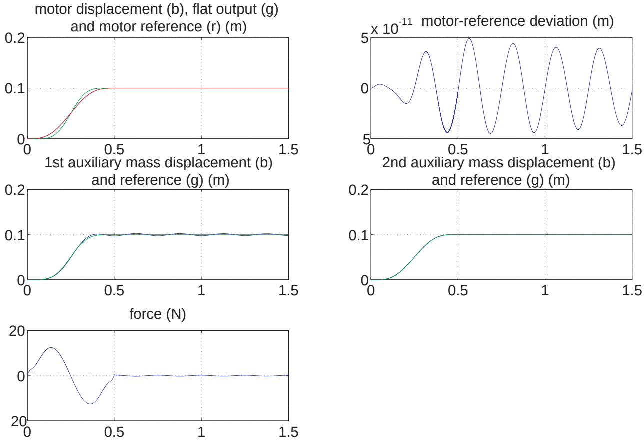

Fig. 10.10 Auxiliary masses taken into account (2nd approach), duration $T = 0,5$s, error of 20% on k and r.

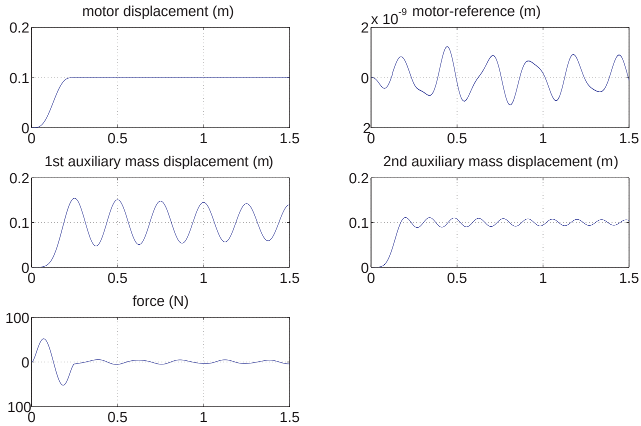

Fig. 10.11 Auxiliary masses taken as perturbations (1st approach), duration $T = 0,25$s, error of 20% on k and r.

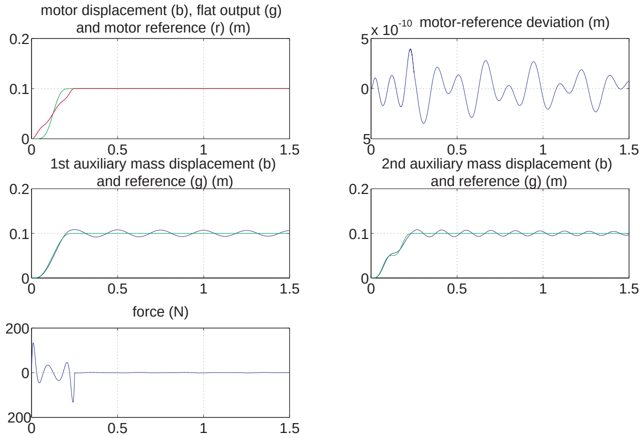

Fig. 10.12 Auxiliary masses taken into account (2nd approach), duration $T = 0,25$s, error of 20% on k and r.

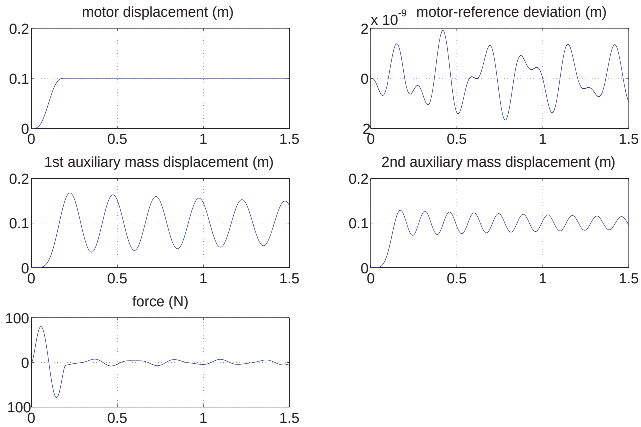

Fig. 10.13 Auxiliary masses taken as perturbations (1st approach), duration $T = 0,2$s, error of 20% on k and r.

The simulation results are presented in Figures 10.9 to 10.14. The conclusions are more or less the same as for a single mass: the flatness-based approach provides a precision of about 10 times higher than with the traditional approach, with comparable displacement durations. A suitable robustness to errors on k and r may also be noticed with this approach, as well as performances which are almost independent of the duration T, at the condition that it remains above 0,25s, where the phenomenon of backward displacement, as in the single mass case, appears (Figure 10.14 top-left). Let us stress that the PID control law is the same in all cases, the only difference being the force and trajectory references F_{ref} and x_{ref}. Let us also remark that the same

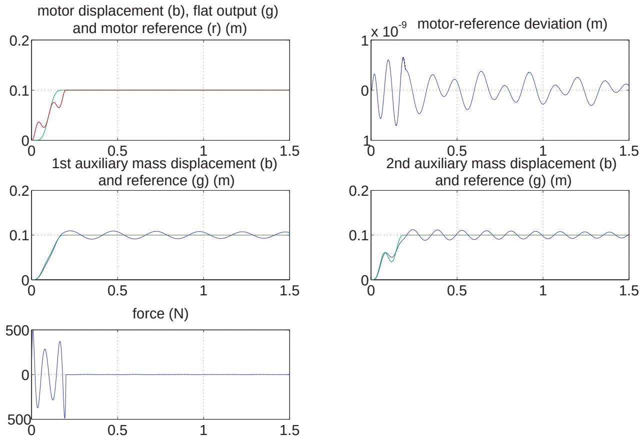

Fig. 10.14 Auxiliary masses taken into account (2nd approach), duration $T = 0,2$s, error of 20% on k and r.

control law could be used to control the rest-to-rest displacements of a motor with an arbitrary number of different auxiliary masses, using the force and trajectory references deduced from the associated flat output.

Chapter 11
Synchronization of a Pair of Independent Windshield Wipers

In this chapter[1], we present an industrial case study whose difficulty lies in the synchronization of two independent systems.

Note that the decoupling problem has received a large number of contributions (see *e.g.* Isidori [1995], Nijmeijer and van der Schaft [1990]), as opposed to the synchronization one, its contrary, especially with decentralized information.

Here, the independent subsystems correspond to two independent windshield wipers that must be synchronized in particular to avoid collisions. They are driven by two independent actuators, each one fed by its own position measurement only. The lack of global information therefore makes the synchronization hard. We propose a decentralized flatness-based control design, where the reference trajectories are specifically designed to avoid collisions and with a trajectory tracking loop, supervised by a clock control loop that regulates the time rate at which both reference trajectories are tracked.

11.1 Introduction

Traditional windshield wipers are actuated by a single constant speed motor related to the wipers by a system of connecting rods, often called the wiper arm (see Fig. 11.1). The wiping movement may be parallel or in opposition and in each case the geometry of the arm is designed such that, first, the wiper movements are generated by almost constant speed rotation of the motor, and second, collisions between the two blades are impossible. There are generally three speeds, fast, slow and intermittent, whose selection is

[1] Work partially done in collaboration with Valeo Electronics. Related registered Patents: US 5757155 and EP0700342

J. Lévine, *Analysis and Control of Nonlinear Systems,* Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_11, © Springer-Verlag Berlin Heidelberg 2009

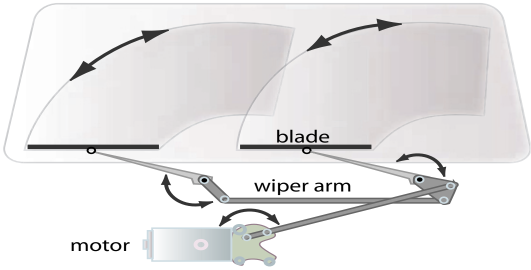

Fig. 11.1 Usual parallel windshield wiper system.

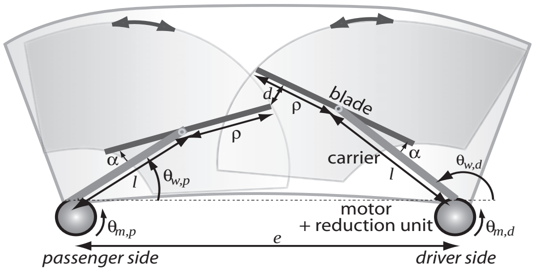

Fig. 11.2 Independent wipers in opposition.

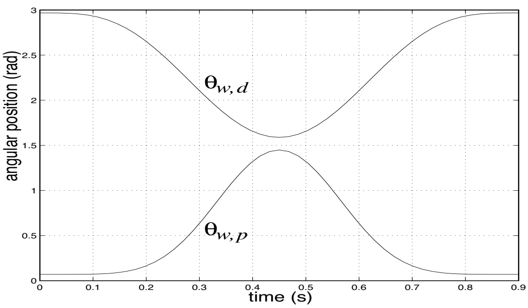

Fig. 11.3 Reference trajectories for the driver's ($\theta_{w,d}$) and passenger's ($\theta_{w,p}$) wipers

made by the driver. In this configuration, the motor control consists only of a PI regulator of the given constant speeds (see e.g. Chin et al. [1991]).

Nevertheless, for a category of vehicles, such as space wagons, with wide bulging windscreens, wiping in opposition is needed to maximize the wiped surface of the screen, and the resulting gearing system is very heavy and bulky, sometimes taking most of the space reserved to the dashboard. In this case, we may need to replace this cumbersome technology by two independent wipers, each one driven by its own motor (see Fig. 11.2). However, the trajectories to be followed by the wipers have to be properly and independently generated by the two motors, which cannot work at constant speed, and the synchronization of the wipers is no longer guaranteed, thus resulting in a complexity increase of the motor controllers.

We consider such a pair of windshield wipers in opposition, driven by two independent DC motors as depicted in Figure 11.2. We indeed want the wipers to follow prescribed smooth and collision-free reference trajectories such as those of Figure 11.3. One difficulty comes from the fact that the wipers are influenced by unmeasured asymmetric torque perturbations that result from the asymmetry of the dust on the screen, of its dampness, of the air flow on the wipers and their contact with the screen. In addition, once the mechanical coupling between the wipers is removed, the elastic contribution of the blades, their carriers and the reduction units between motors and carriers to the overall behavior ceases to be negligible. Further, the measurements of the angular positions and velocities of the wipers are not available, and only the angular positions and velocities of the motors are measured.

Moreover, as it is the case for all automotive equipments, the reliability of this controlled system must be proved in various extreme situations, thus requiring a great amount of robustness. Finally, the cost of this system with two variable speed drives must remain comparable to that of the standard single motor setup, if one wants the former system to effectively replace the latter. Therefore, the motors and the microprocessor driving the system have to be as cheap as possible, resulting in inevitable restrictions of motor power and microprocessor performance. Consequently, a careful trajectory design of the wipers has to be done to be compatible with low powered motors and the controller has to be simple enough to be implemented on a slow microprocessor with small memory.

Our approach is based on the observation that each wiper is naturally decoupled from the other one, the couplings being created by disturbance torques. Therefore, at a first step, the control design can be done wiper by wiper. It is moreover based on *flatness*. Then, the synchronization step is done by *controlling the clock*, i.e. delaying the reference trajectory of the wiper called *follower* when the other one, the *leader*, is behind its reference trajectory, extending ideas previously developed by Dahl and Nielsen Dahl and Nielsen [1990], Dahl [1994].

This chapter is organized as follows: the model of a single wiper system is presented and analyzed in section 11.2. The (open-loop) reference trajectory design problem of the pair of wipers, including open-loop collision avoidance, is studied in section 11.3. Then the tracking aspects, namely the feedback design of a control law that stabilizes the error dynamics of a wiper with respect to its reference trajectory in presence of torque disturbances, are dealt with in section 11.4. We then present our closed-loop synchronization method in section 11.5.

The interested reader may also refer to Bitauld et al. [1997], Lévine [2004].

11.2 The Model of a Single Wiper

Each wiping system is made up with a DC motor, a gear and a wiper. Each motor is equipped, as previously stated, with an angular position sensor, whose precision is enough to assume that the angular velocity of the motor can be satisfactorily estimated. Denoting by $\theta_{m,i}$ the angular position of the motor i ($i = d, p$, the subscript d referring to the driver's side and p to the passenger's side), by $\dot{\theta}_{m,i}$ its angular velocity, and by $I_{m,i}$ the current circulating through it, the electric model of the ith motor is:

$$L\dot{I}_{m,i} + RI_{m,i} + K_e\dot{\theta}_{m,i} = U_{m,i}, \quad i = d, p \tag{11.1}$$

where L is the self-inductance, R the resistance, K_e the counter electromotive force constant of the motors (both motors having the same characteristics) and $U_{m,i}$ the voltage applied to the ith motor, $i = d, p$. $U_{m,d}$ and $U_{m,p}$ are the control variables.

Identifying each wiper to its blade-carrier set, considered as a single rigid body, we denote by $\theta_{w,i}$ the angular position of the ith wiper, $i = d, p$. We assume that the mechanical characteristics of both wipers are the same and that the elastic link between each wiper and its motor's reduction unit axis can be identified to a helicoidal spring of stiffness K_w and damping torque $K_v\dot{\theta}_{w,i}$ (assumed linear w.r.t. the angular velocity $\dot{\theta}_{w,i}$). The reduction ratio being denoted by N, the angular position of each reduction unit is equal to $\frac{\theta_{m,i}}{N}$. Thus, the torque balance, for $i = d, p$, writes

$$\begin{aligned} J_m\ddot{\theta}_{m,i} &= -\frac{K_w}{N}\left(\frac{\theta_{m,i}}{N} - \theta_{w,i}\right) - K_f\dot{\theta}_{m,i} + K_cI_{m,i} \\ J_w\ddot{\theta}_{w,i} &= K_w\left(\frac{\theta_{m,i}}{N} - \theta_{w,i}\right) - K_v\dot{\theta}_{w,i} + \eta_i \end{aligned} \tag{11.2}$$

with J_m and J_w the respective inertias of the motors and wipers with respect to their common axis, K_c the torque constant of the motors and K_f the coefficient of viscous friction of the motor-reduction units. The unknown disturbance η_i corresponds to the resisting torque of the ith wiper on the windscreen. Recall that η_d and η_p may be different.

The dynamics of the current $I_{m,i}$ in (11.1) is assumed faster than the dynamics of $(\theta_{m,i}, \dot{\theta}_{m,i}, \theta_{w,i}, \dot{\theta}_{w,i})$, representing the mechanical contribution. In other words, we assume that the inductance L is negligible so that $I_{m,i}$ reaches its steady-state $\dot{I}_{m,i} = 0$ fast compared to the mechanical dynamics (see e.g. Kokotović et al. [1986] and Section 3.3):

$$I_{m,i} = \frac{1}{R}U_{m,i} - \frac{K_e}{R}\dot{\theta}_{m,i}. \tag{11.3}$$

With the notation $K_r = \frac{K_cK_e}{R} + K_f$, (11.2) rewrites:

$$\begin{aligned} J_m\ddot{\theta}_{m,i} &= -\frac{K_w}{N}(\frac{\theta_{m,i}}{N} - \theta_{w,i}) - K_r\dot{\theta}_{m,i} + \frac{K_c}{R}U_{m,i} \\ J_w\ddot{\theta}_{w,i} &= K_w(\frac{\theta_{m,i}}{N} - \theta_{w,i}) - K_v\dot{\theta}_{w,i} + \eta_i. \end{aligned} \tag{11.4}$$

We now show that (11.4) admits $\theta_{w,d}$ and $\theta_{w,p}$, the driver's and passenger's wiper angular positions, as a *flat output*. Assume for a moment that the perturbation η_i is known and at least twice differentiable with respect to time. Then $\theta_{m,i}$ and $U_{m,i}$, $i = d, p$, are given by:

$$\begin{aligned} \theta_{m,i} &= N\left(\theta_{w,i} + \frac{1}{K_w}\left(K_v\dot{\theta}_{w,i} + J_w\ddot{\theta}_{w,i} - \eta_i\right)\right) \\ U_{m,i} &= \frac{NR}{K_c}\left[\left(K_r + \frac{K_v}{N^2}\right)\dot{\theta}_{w,i}\right. \\ &+ \left(J_m + \frac{J_w}{N^2} + \frac{K_rK_v}{K_w}\right)\ddot{\theta}_{w,i} \\ &\left.+\frac{1}{K_w}\left(J_wK_r + J_mK_v\right)\theta^{(3)}_{w,i} + \frac{J_wJ_m}{K_w}\theta^{(4)}_{w,i}\right] \\ &-\frac{NR}{K_cK_w}\left(\frac{K_w}{N^2}\eta_i + K_r\dot{\eta}_i + J_m\ddot{\eta}_i\right) \end{aligned} \tag{11.5}$$

which proves the assertion. In addition, according to (11.5), system (11.4) reads:

$$\theta^{(4)}_{w,i} = -\sum_{j=0}^{3} S_j\theta^{(j)}_{w,i} + \mu U_{m,i} + \tilde{\eta}_i \tag{11.6}$$

with

$$\begin{aligned} S_0 &= 0, \quad S_1 = \frac{K_w}{J_wJ_m}\left(K_r + \frac{K_v}{N^2}\right), \\ S_2 &= \frac{K_w}{J_wJ_m}\left(J_m + \frac{J_w}{N^2} + \frac{K_rK_v}{K_w}\right), \\ S_3 &= \frac{1}{J_wJ_m}\left(J_wK_r + J_mK_v\right), \\ \mu &= \frac{K_wK_c}{NRJ_wJ_m}, \end{aligned} \tag{11.7}$$

and $\tilde{\eta}_i = \frac{1}{J_wJ_m}\left(\frac{K_w}{N^2}\eta_i + K_r\dot{\eta}_i + J_m\ddot{\eta}_i\right)$.

11.3 Open Loop Synchronization of the Pair of Wipers by Motion Planning

Let us now describe the reference trajectories of the wipers. The total cycle duration T of both wipers is prescribed: $T = 0.9$ s for the fast wiping rate and $T = 1.2$ s for the slow rate. Both wipers must start at rest (0 angular speed and acceleration) at the same time $t = kT$, $k = 0, 1, \ldots$, arrive at their upper positions at the same time $kT + \frac{T}{2}$ at zero angular speed and then return symmetrically to their initial position at the same time $(k+1)T$, $k = 0, 1, \ldots$ Moreover, the driver's wiper must start moving first and remain above the passenger's wiper all along the displacement, with a prescribed safety distance to avoid collisions. Consequently, the passenger's wiper must start at a slower rate than the driver's one and return to its initial position at a faster rate. The driver's wiper initial angular position is $\pi - a_d$ (at time $t = kT$) and its upper position is $\pi - a_d - b_d$. The passenger's wiper initial angular position is a_p and its upper position is $a_p + b_p$ (at time $kT + \frac{T}{2}$). In this particular application, we have chosen $\pi - a_d = 2.967$ rad $= 170°$, $a_p = 0.0698$ rad $= 4°$ and $b_p = b_d = 1.3788$ rad $= 79°$ (full wiped angle).

Additional maximum velocity and acceleration as well as current and voltage constraints are also imposed:

$$\begin{array}{c} |\dot{\theta}_{w,i}| \leq 9\ rad/s,\ \ |\ddot{\theta}_{w,i}| \leq 60\ rad/s^2, \\ |I_{m,i}| \leq 50\ A,\ \ |U_{m,i}| \leq 24\ V,\ \ i = p, d. \end{array} \tag{11.8}$$

Finally, to avoid collision between the blades, the minimal distance separating them, denoted by d (see Figure 11.2), must always exceed 50 mm. An easy calculation, the screen being approximated by a plane, gives the minimal distance by:

$$\begin{array}{c} d = \rho \sin(\theta_{w,p} - \theta_{w,d} - 2\alpha) + e \sin(\theta_{w,d} + \alpha) \\ +l \left(\sin\alpha + \sin(\theta_{w,p} - \theta_{w,d} - \alpha)\right) \end{array} \tag{11.9}$$

where e is the distance between the two motor axles, l the carrier's length, ρ the half-blade's length and α the fixed angle between the blade and the carrier, the same on both sides (see Figure 11.2). Thus the collision avoidance constraint reads

$$d \geq 50 \text{ mm}. \tag{11.10}$$

In our particular application, we have $e = 1260$ mm, $l = 640$ mm and $\rho = 300$ mm.

We now design the wiper position reference trajectories and, according to flatness, use (11.5)-(11.3) to deduce the corresponding motor and input current and voltage reference trajectories. Consider the polynomial:

$$P(\beta) = 2^8 \beta^4 (1 - \beta)^4 \tag{11.11}$$

defined on the interval $[0, 1]$, and set, for t modulo T:

$$\theta_{w,d}(t) = \pi - a_d - b_d P\left(\frac{t}{T}\right) \tag{11.12}$$

i.e., define $\theta_{w,d}(t)$ by (11.12) for $t \in [0,T]$ and $\theta_{w,d}(t+kT) = \theta_{w,d}(t)$ for every $k = 0, 1, \ldots$ Clearly, $\dot{\theta}_{w,d}(kT) = \ddot{\theta}_{w,d}(kT) = 0$ and $\dot{\theta}_{w,d}(kT + \frac{T}{2}) = 0$, for all $k = 0, 1, \ldots$

The reference trajectory of the passenger's wiper is given, in turn, for $t \in [0,T]$, by

$$\theta_{w,p}(t) = a_p + b_p P\left(\frac{t}{T}\right) G_{\frac{1}{2},\sigma}\left(\frac{t}{T}\right) \sin\left(\frac{\pi t}{T}\right) \tag{11.13}$$

where $G_{\frac{1}{2},\sigma}(t)$ is the unnormalized Gaussian function of mean $\frac{1}{2}$ and variance σ^2: $G_{\frac{1}{2},\sigma}(t) = e^{-\frac{1}{2\sigma^2}\left(t-\frac{1}{2}\right)^2}$. Next, we define $\theta_{w,p}(t+kT) = \theta_{w,p}(t)$ for all $k = 0, 1, \ldots$ In other words, (11.13) is valid for t modulo T.

As shown on Figure 11.4, with $2\sigma^2 = 0.1$, the product of the Gaussian and sine functions by P is such that the passenger's wiper starts at a slower rate and arrives at a faster rate. Again, we clearly have $\dot{\theta}_{w,p}(kT) = \ddot{\theta}_{w,p}(kT) = 0$ and $\dot{\theta}_{w,p}(kT + \frac{T}{2}) = 0$, for all $k = 0, 1, \ldots$

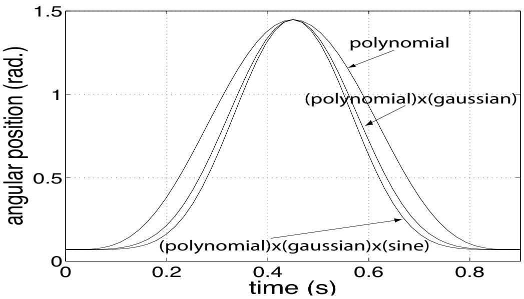

Fig. 11.4 Comparison between the polynomial trajectory $a_p + b_p P(\frac{t}{T})$, $a_p + b_p P\left(\frac{t}{T}\right) G_{\frac{1}{2},\sigma}\left(\frac{t}{T}\right)$ and $\theta_{w,p}(t)$ given by (11.13).

Note that it appeared much simpler to distort the polynomial P by multiplying it by a Gaussian function, and a sine to make it even narrower, than to construct another polynomial of higher degree having the required property.

Figure 11.5 depicts the minimal distance between the wipers along the chosen reference trajectory, showing that the collision avoidance constraint (11.10) is met.

Finally, the reference trajectories for both motor angular positions, voltages and currents are deduced from (11.5)-(11.3) by replacing the perturbation η_i by a constant $\overline{\eta}_i$, chosen, for instance, as the averaged value of η_i, $i = d, p$.

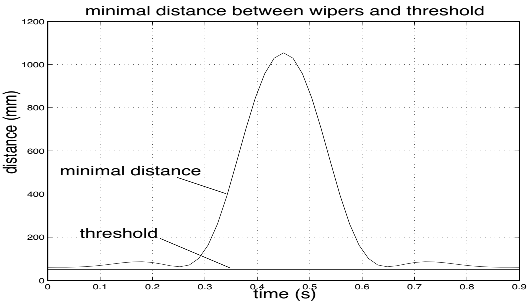

Fig. 11.5 The minimal distance between the wipers along the reference trajectory (11.12)-(11.13) always exceeds the threshold of 50mm.

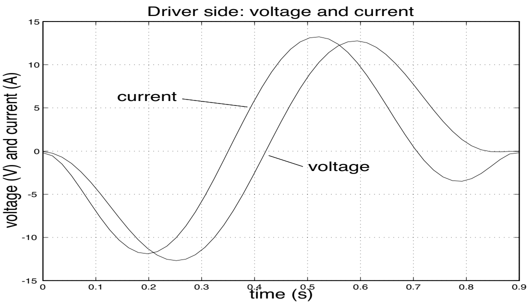

Fig. 11.6 The reference (feedforward) trajectories of voltage and current corresponding to the driver's wiper

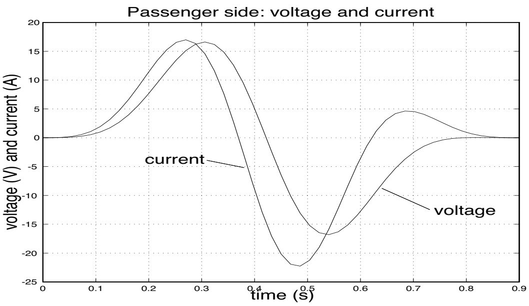

Fig. 11.7 The reference (feedforward) trajectories of voltage and current corresponding to the passenger's wiper

The respective reference voltages and currents of the driver's and passenger's motor, that generate the wiper's angular position trajectories $\theta_{w,d}$ and $\theta_{w,p}$ respectively, are depicted in Figures 11.6 and 11.7. Here, the motor characteristics are $N = 63$ (reduction ratio), $R = 0.29\ \Omega$ (motor resistance), $K_v = 2$ Nms/rad (wiper damping factor), $K_f = 11.10^{-5}$ Nms/rad (motor viscous friction coefficient), $K_w = 600$ Nm/rad (wiper stiffness), $K_c = 0.0276$

Nm/A (torque constant), $K_e = 0.0276$ Vs/rad (counterelectromotive coefficient), $J_w = 0.1$ Nms^2/rad (wiper inertia) and $J_m = 6.10^{-5}$ Nms^2/rad (motor inertia).

The reader can check that $\dot{\theta}_{w,d}$, $\dot{\theta}_{w,p}$, $\ddot{\theta}_{w,d}$, $\ddot{\theta}_{w,p}$, $I_{m,d}$, $I_{m,p}$, $U_{m,d}$ and $U_{m,p}$ satisfy the constraints (11.8).

The same design can be done for slow speed and intermittent modes by simply modifying the value of T, the constraints (11.8) being all the more easy to meet as the period T is large.

11.4 Trajectory Tracking

Given the reference trajectories, noted $\theta^*_{w,i}$, $i = d, p$, described in the previous section, we introduce the angular position error

$$e_{w,i} = \theta_{w,i} - \theta^*_{w,i}, \quad i = d, p. \tag{11.14}$$

and $E_{w,i} = \left(e_{w,i}, \dot{e}_{w,i}, \ddot{e}_{w,i}, e^{(3)}_{w,i}\right)'$ (' denoting transpose). According to (11.6), $E_{w,i}$ satisfies the following error equation, in matrix form, for $i = d, p$:

$$\dot{E}_{w,i} = AE_{w,i} + B(U_{m,i} - U^*_{mi}) + D\delta_i \tag{11.15}$$

with

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -S_0 & -S_1 & -S_2 & -S_3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \mu \end{pmatrix},$$

$$D = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \delta_i = \tilde{\eta}_i - \frac{K_w}{N^2 J_w J_m}\overline{\eta}_i.$$

If we denote by $-\lambda_1, \ldots, -\lambda_4$ the eigenvalues of the matrix A, it is well-known (see e.g. Gantmacher [1966], Kailath [1980], Sontag [1998]) that the relations between the eigenvalues of A and A's entries read: $S_0 = \lambda_1\lambda_2\lambda_3\lambda_4$, $S_1 = \sum_{0\le i<j<k\le 4} \lambda_i\lambda_j\lambda_k$, $S_2 = \sum_{0\le i<j\le 4} \lambda_i\lambda_j$, $S_3 = \sum_{i=1}^{4} \lambda_i$.

Since $S_0 = 0$, and $S_1 > 0$, $S_2 > 0$, $S_3 > 0$, it results that one of the eigenvalues is equal to 0, the remaining ones having negative real part. To the 0 eigenvalue is associated the eigenvector $(e_{w,i}, 0, 0, 0)'$, which means that $e_{w,i}$ is not exponentially stable, whereas all the derivatives of the error up to the 3rd order are.

Thus, since we are concerned with the stabilization of $e_{w,i}$ we want to find a feedback law such that the 0 eigenvalue is replaced by a sufficiently negative one, without modifying the sign of the real parts of the other eigenvalues.

However, $e_{w,i}$ is not measured and only $\theta_{m,i}$ is, the sensors and the actuators being collocated. But since we have, by (11.5), $\theta_{m,i} = N(\theta_{w,i} + \frac{K_v}{K_w}\dot{\theta}_{w,i} + \frac{J_w}{K_w}\ddot{\theta}_{w,i} - \frac{1}{K_w}\eta_i)$, a feedback on $\theta_{m,i}$ will indeed affect $\theta_{w,i}$ and thus $e_{w,i}$. Denoting by $e_{m,i} = \theta_{m,i} - \theta^*_{m,i}$, we therefore introduce the following PID:

$$\begin{aligned} U_{m,i}(t) = U^*_{m,i}(t) - K_P e_{m,i}(t) - K_D \dot{e}_{m,i}(t) \\ -K_I \int_0^t e_{m,i}(s)ds \end{aligned} \tag{11.16}$$

leading to the following error equation

$$e^{(5)}_{w,i} = -\sum_{j=0}^{4} \hat{S}_j e^{(j)}_{w,i} + \hat{\tilde{\eta}}_i \tag{11.17}$$

with

$$\begin{aligned} &\hat{S}_0 = \mu N K_I, \quad \hat{S}_1 = \frac{\mu N}{K_w}(K_w K_P + K_v K_I) \\ &\hat{S}_2 = S_1 + \frac{\mu N}{K_w}(K_v K_P + K_w K_D + J_w K_I) \\ &\hat{S}_3 = S_2 + \frac{\mu N}{K_w}(J_w K_P + K_v K_D) \\ &\hat{S}_4 = S_3 + \frac{\mu N}{K_w} J_w K_D \end{aligned} \tag{11.18}$$

and $\hat{\tilde{\eta}}_i = \dot{\tilde{\eta}}_i + \frac{\mu N}{K_w}(K_I(\eta_i - \overline{\eta}_i) + K_P\dot{\eta}_i + K_D\ddot{\eta}_i)$.

We therefore have to tune the gains K_P, K_D and K_I such that the unperturbed system is stable and such that the coefficients of the perturbative terms are sufficiently small compared to the others.

Choosing $K_I > 0$, if the perturbation converges to a constant, still denoted by η_i, the equilibrium value of the error $\overline{e}_{w,i}$ is given by $\mu N K_I \overline{e}_{w,i} = \frac{\mu N}{K_w} K_I(\eta_i - \overline{\eta}_i)$, or

$$\overline{e}_{w,i} = \frac{1}{K_w}(\eta_i - \overline{\eta}_i) \tag{11.19}$$

which means that the asymptotic bias is independent of the gains and depends only on the stiffness of the wiper which is, however, large enough in our application ($K_w = 600$ Nm/rad), and on the estimate $\overline{\eta}_i$.

Note that the stability is ensured by an arbitrary choice of $K_I, K_P, K_D \geq 0$, with $K_I K_P \neq 0$.

The gains $K_P = 80$, $K_I = 40$ and $K_D = 20$, obtained by elementary root locus and Bode diagram analysis, are such that the lowest frequency almost reaches its maximal value of 0.56 Hz while keeping an acceptable disturbance attenuation and robustness versus perturbations, as depicted by the simulation of Figure 11.8: the torque disturbances have maximum amplitude of 20 Nm and $\overline{\eta}_i = 10$ Nm, $i = p, d$. Figure 11.9 shows the corresponding motor voltages in closed-loop. Indeed, the gains can also be tuned by other techniques such as H_∞ optimization, μ-synthesis or LMI.

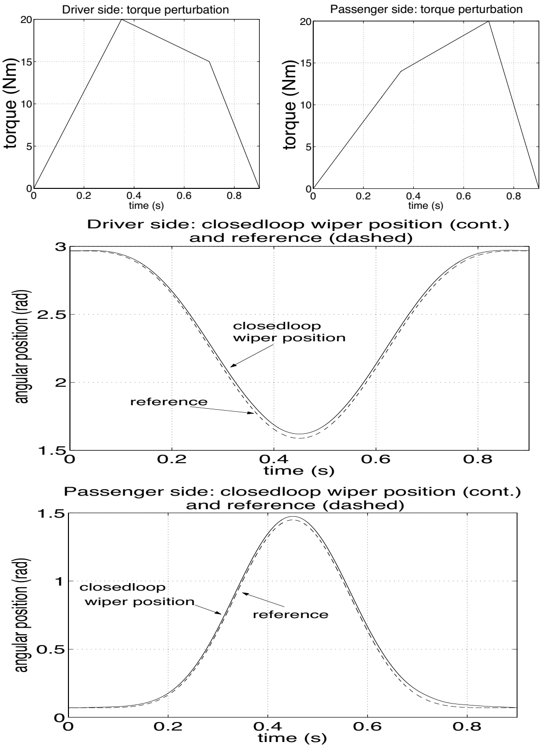

Fig. 11.8 from top to bottom: torque disturbances of driver and passenger wiper; driver's wiper position in closed-loop and its reference; passenger's wiper position in closed-loop and its reference. The gains are $K_P = 80$, $K_I = 40$ and $K_D = 20$ and $\overline{\eta}_i = 10$ Nm, $i = p, d$.

11.5 Synchronization by Clock Control

Since the two wipers are naturally decoupled, there is no standard method to synchronize the wipers if one of them deviates from its reference trajectory. Recall that a deviation of about 5 deg. of the angular position of one of the wipers in the mid-zone of the screen, or more precisely when the passenger's wiper angle lies between 10° and 30° (see Fig. 11.5), might cause a collision with the other wiper's blade and an emergency stop. Keeping in mind that the motors' power is limited and that their voltage must satisfy (11.8),

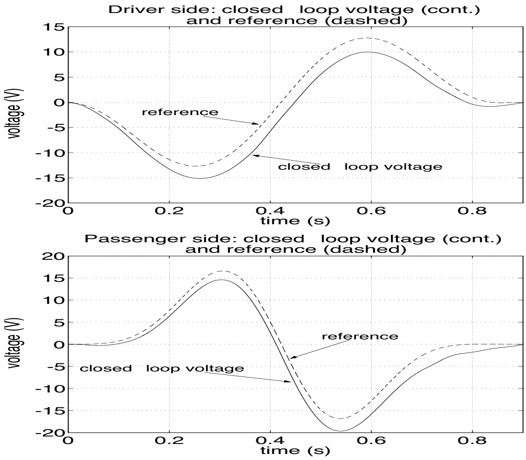

Fig. 11.9 closed-loop voltage applied to the driver's motor and its reference (up) and closed-loop voltage applied to the passenger's motor and its reference (down) with the above disturbances and gains.

the synchronization would not be significantly improved by adding terms depending on the error of the other wiper in the PID controller (11.16): this additional term would increase the motor voltage, thus producing more frequent saturations, and would even deteriorate the individual tracking. In fact, the synchronization is already partly realized through the feedforward reference trajectory design and a natural idea, inspired by Dahl and Nielsen Dahl and Nielsen [1990], Dahl [1994], consists in adjusting the speed at which the reference trajectory of one of the wipers is run in function of the deviation of the other one (see Section 8.2). When the wipers move up, while leaving their parking position, the driver's wiper moves first, and may be considered as the leader, while the passenger's one follows; conversely, when going back to their parking position, the passenger's wiper is the leader and the driver's one the follower. Therefore, to avoid collisions when the leader is behind its reference trajectory, it suffices to design a mechanism to delay the local clock of the follower's reference trajectory.

Let us stress that if a wiper is ahead of its position while the motor accelerates, the feedback might suffice to take it back to its reference since slowing down in this case does not require much extra power. On the contrary, if the voltage of the leading wiper's motor reaches its saturation while this wiper is behind its reference, it is momentarily unable to recover its reference position, and the motor of the follower has no other choice than slowing down to avoid possible collisions. We now develop this approach in detail.

Let us consider the reference trajectories $\theta^*_{w,i}$, $i = d, p$, defined in section 11.3. In what follows, the subscript $\mathcal{L}$ denotes the leader, namely $\mathcal{L} = d$ when $\dot{\theta}^*_{m,d} \leq 0$ and $\mathcal{L} = p$ when $\dot{\theta}^*_{m,p} \leq 0$, and the subscript $\mathcal{F}$ denotes the follower, namely $\mathcal{F} = p$ if $\mathcal{L} = d$ and conversely (it is clear from Figure 11.3

that the reference velocities $\dot{\theta}^*_{m,d}$ and $\dot{\theta}^*_{m,p}$ of the wipers always have opposite signs). We introduce a new time for the follower, denoted by $\tau(t)$, satisfying $0 \leq \tau(t) \leq t$ for all t and at least 4 times continuously differentiable with respect to t so that, by composition, the follower's modified reference trajectory $\theta^*_{w,\mathcal{F}}(\tau(t))$ is also at least 4 times continuously differentiable. A natural way to design the function τ, which we call the *controlled clock*, is to remain as close as possible to the natural time, i.e. $\dot{\tau} = 1$, and to decrease its rate according to the (measured) deviation of the leader's angular position $e_{m,\mathcal{L}}$ and velocity $\dot{e}_{m,\mathcal{L}}$. Let us remark as in Section 8.2, that we are not particularly interested in the convergence of τ to the natural time t after slowing down, the only requirement being the convergence of its rate: $\dot{\tau} \to \dot{t} = 1$.

A possible design is thus to assign its behavior by an at least 4th order differential equation controlled by a scalar input Γ fed back by $e_{m,\mathcal{L}}$ and $\dot{e}_{m,\mathcal{L}}$. Here is a 5th order example (see again Section 8.2):

$$\begin{aligned} \tau^{(5)} = -\gamma_4 \tau^{(4)} - \gamma_3 \tau^{(3)} - \gamma_2 \ddot{\tau} - \gamma_1 (\dot{\tau} - 1) \\ -\Gamma_0(e_{m,\mathcal{L}}(t), \dot{e}_{m,\mathcal{L}}(t)) \end{aligned} \tag{11.20}$$

with initial conditions $\tau(t_0) = t_0$, $\dot{\tau}(t_0) = 1$ and $\tau^{(j)}(t_0) = 0$ for $j = 2, 3, 4$. The gains γ_i, $i = 1, \ldots, 4$ may be chosen arbitrarily and the scalar function Γ, the *clock controller*, is such that $\Gamma(0, 0) = 0$ and is required to be differentiable with respect to its arguments with bounded derivatives. In the sequel, we plan to choose $\Gamma_0((e_{m,\mathcal{L}}(t), \dot{e}_{m,\mathcal{L}}(t))$ proportional to $\dot{e}_{m,\mathcal{L}}(t)$.

Note that $e_{m,\mathcal{L}} = 0$ for all t implies that $\tau(t) = t$ for all t and that, even if $\dot{e}_{m,\mathcal{L}}(t)$ is piecewise continuous, the composed reference trajectory $\theta^*_{w,\mathcal{F}}(\tau(t))$ remains 4 times differentiable for all t. It is then possible to place the poles of the controlled clock system (11.20).

If we replace $\theta^*_{w,\mathcal{F}}(t)$, $\theta^*_{m,\mathcal{F}}(t)$, $U^*_{m,\mathcal{F}}(t)$, $e_{w,\mathcal{F}}(t)$ and $e_{m,\mathcal{F}}(t)$ by $\theta^*_{w,\mathcal{F}}(\tau(t))$, $\theta^*_{m,\mathcal{F}}(\tau(t))$, $U^*_{m,\mathcal{F}}(\tau(t))$, respectively in (11.5) and $e_{w,\mathcal{F}}(t) = \theta_{w,\mathcal{F}}(t) - \theta^*_{w,\mathcal{F}}(\tau(t))$, $e_{m,\mathcal{F}}(t) = \theta_{m,\mathcal{F}}(t) - \theta^*_{m,\mathcal{F}}(\tau(t))$, the PID (11.16) being modified in the same way, the reader may easily check that the corresponding error system is still (11.17), the same stable time-invariant linear system.

However, system (11.20) being a 5th order integrator of Γ, wherever its poles are placed in the left half complex plane, its performance may be poor in case of an emergency such as the locking of the leading wiper, where we require a fast slowing down reaction from the follower, and thus from its clock, e.g. with a time constant less than or equal to 10^{-3} s, thus imposing a high gain design. We are therefore led to replace (11.20) by the following (see Section3.3):

$$\begin{aligned} &\ddot{\tau} = -\tilde{\gamma}_1(\dot{\tau} - 1) - \Gamma(e_{m,\mathcal{L}}(t), \dot{e}_{m,\mathcal{L}}(t)) \\ &\tau(t_0) = t_0, \dot{\tau}(t_0) = 1 \end{aligned} \tag{11.21}$$

and to replace the 3rd and 4th order derivatives of τ by

$$\tau^{(3)}(t) = \tau^{(4)}(t) = 0 \tag{11.22}$$

for all t. System (11.21)-(11.22) may be viewed as a singular perturbation approximation of (11.20) via high gain for a suitable choice of γ_i, $i = 1, \ldots, 4$, Γ_0, $\tilde{\gamma}_1$ and Γ (see e.g. Kokotović et al. [1986] and the proof of Theorem 11.1 below). Note that, as previously remarked, adding a term of the form $-\gamma(\theta - t)$ to force the convergence to the current time t is not needed.

To achieve our controlled clock design, let us specify that the leader's reference trajectory cannot be modified by the follower's local clock and that the current time is reset when the leader and follower exchange roles, namely after each half cycle, when the wipers arrive either at their upper positions or at their parking ones.

The resulting control architecture is presented in Figure 11.10.

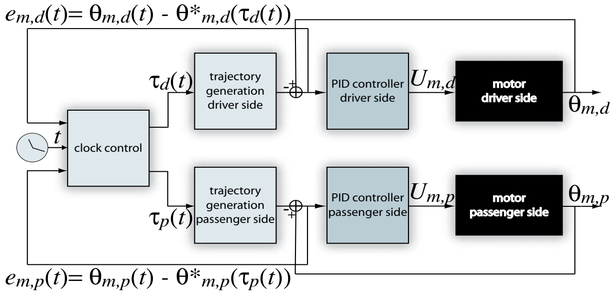

Fig. 11.10 The structure of the controller with clock control

This construction is justified by the fact that modifying the clock *of the follower's reference trajectory only* neither modifies (11.15) nor the feedback gains and thus the closed-loop pole placement. In particular, it does not create an artificial time-dependence of the eigenvalues and eigenvectors of the closed-loop system that could cause instabilities. This is stated in the following result:

Theorem 11.1. *The local stability of the closed-loop system (11.6)-(11.16) is still valid with the synchronization (11.21)-(11.22) with $\tilde{\gamma}_1 > 0$ not too large and Γ such that $\Gamma(0, 0) = 0$, in a domain where τ is strictly increasing.*

Proof. Given a 4 times differentiable non decreasing function τ, we have shown that the error $e_{w,\mathcal{F}} = \theta_{w,\mathcal{F}}(t) - \theta^*_{w,\mathcal{F}}(\tau(t))$ satisfies the time-invariant differential equation (11.17), which is exponentially stable. On the other hand, since $e_{w,\mathcal{L}}$ is not modified by τ, the stability of the pair $(e_{w,\mathcal{L}}, e_{w,\mathcal{F}})$ remains unchanged. It thus remains to prove that (11.20) can be replaced by (11.21)-(11.22). Setting $\tau_1 = \dot{\tau} - 1$, (11.20) reads

$$\tau_1^{(4)} = -\gamma_4 \tau_1^{(3)} - \gamma_3 \ddot{\tau}_1 - \gamma_2 \dot{\tau}_1 - \gamma_1 \tau_1 - \Gamma_0(e_{m,\mathcal{L}}(t), \dot{e}_{m,\mathcal{L}}(t)). \tag{11.23}$$

Let $\varepsilon > 0$ be a sufficiently small real number and let $-\sigma_1, -\frac{1}{\varepsilon}\sigma_2, -\frac{1}{\varepsilon^2}\sigma_3, -\frac{1}{\varepsilon^2}\sigma_4$ be the eigenvalues associated to (11.23) with $\sigma_i > 0$, $i = 1, \ldots, 4$, i.e. $\gamma_1 = \frac{1}{\varepsilon^5}\sigma_1\sigma_2\sigma_3\sigma_4$, $\gamma_2 = \frac{1}{\varepsilon^3}\sigma_1\sigma_2(\sigma_3 + \sigma_4) + \frac{1}{\varepsilon^4}\sigma_1\sigma_3\sigma_4 + \frac{1}{\varepsilon^5}\sigma_2\sigma_3\sigma_4$, $\gamma_3 = \frac{1}{\varepsilon}\sigma_1\sigma_2 + \frac{1}{\varepsilon^2}\sigma_1(\sigma_3 + \sigma_4) + \frac{1}{\varepsilon^3}\sigma_2(\sigma_3 + \sigma_4) + \frac{1}{\varepsilon^4}\sigma_3\sigma_4$ and $\gamma_4 = \sigma_1 + \frac{1}{\varepsilon}\sigma_2 + \frac{1}{\varepsilon^2}(\sigma_3 + \sigma_4)$. Then (11.23), with Γ_0 close to 0, is exponentially stable. Thus, if $\Gamma_0 = 0$, τ_1 converges to 0, or in other words $\dot{\tau}$ converges to 1 from any initial condition $(e_{w,\mathcal{L}}(0), \dot{e}_{w,\mathcal{L}}(0), \ddot{e}_{w,\mathcal{L}}(0), e^{(3)}_{w,\mathcal{L}}(0), e^{(4)}_{w,\mathcal{L}}(0))$, $(e_{w,\mathcal{F}}(0), \dot{e}_{w,\mathcal{F}}(0), \ddot{e}_{w,\mathcal{F}}(0), e^{(3)}_{w,\mathcal{F}}(0), e^{(4)}_{w,\mathcal{F}}(0))$, and $(\tau_1(0), \dot{\tau}_1(0), \ddot{\tau}_1(0), \tau_1^{(3)}(0))$ in an open neighborhood of $0_{14} = (0,0,0,0,0,0,0,0,0,0,0,0,0,0)'$ such that $\dot{\tau} > 0$. In the very fast time-scale defined by $t_2 = \frac{t}{\varepsilon^2}$, denoting by $\nu_1 = \frac{d^2\tau_1}{dt_2^2} + (\sigma_3 + \sigma_4)\frac{d\tau_1}{dt_2} + \sigma_3\sigma_4\tau_1$, we get

$$\frac{d^2\nu_1}{dt_2^2} + (\varepsilon\sigma_2 + \varepsilon^2\sigma_1)\frac{d\nu_1}{dt_2} + \varepsilon^3\sigma_1\sigma_2\nu_1 = \Gamma_0. \tag{11.24}$$

In the fast time-scale defined by $t_1 = \frac{t}{\varepsilon} = \varepsilon t_2$, we note $\nu_2 = \frac{d\nu_1}{dt_1} + \sigma_2\nu_1$. Then we get $\frac{d\nu_2}{dt_1} + \varepsilon\sigma_1\nu_2 = \frac{1}{\varepsilon^2}\Gamma_0$ or, expressed in the original time, the slow dynamics is $\dot{\nu}_2 + \sigma_1\nu_2 = \frac{1}{\varepsilon}\Gamma_0$.

It results that, in the very fast time-scale t_2, for every ν_1, τ_1 converges to $\frac{1}{\sigma_3\sigma_4}\nu_1$ and, in the fast time-scale t_1, for every ν_2, ν_1 converges to $\frac{1}{\sigma_2}\nu_2$. Finally, replacing these expressions in the slow dynamics, and setting $\Gamma = \frac{1}{\varepsilon}\Gamma_0$, we get

$$\dot{\tau}_1 = -\sigma_1\tau_1 + \Gamma \tag{11.25}$$

which is precisely (11.21) with $\tilde{\gamma}_1 = \sigma_1$ and $\Gamma_0 = \varepsilon\Gamma$. The local stability of the system ((11.17), $i = \mathcal{L}, \mathcal{F}$)-(11.21), which is thus an approximation of order $0(\varepsilon)$ of the locally stable system ((11.17), $i = \mathcal{L}, \mathcal{F}$)-(11.20), readily follows for suitable initial conditions such that $\dot{\tau} > 0$ and by choosing ε sufficiently small, using standard singular perturbation arguments Kokotović et al. [1986]. Moreover, since we can choose $\sigma_i > 0$ arbitrarily, the same holds true for $\tilde{\gamma}_1 > 0$.

In the simulations presented in Figures 11.11, 11.12 and 11.13, the torque disturbances have maximum amplitudes 40 Nm, twice as those of Figure 11.8, with both slow and fast transients, while the same value of $\overline{\eta}_i = 10$ Nm and feedback gains, $K_P = 80$, $K_I = 40$ and $K_D = 20$, are used. Without control of the clock, the perturbations produce a collision at $t \approx 0.655$ s as shown by Figure 11.13 (bottom). Using the controlled clock (11.21) with

$$\Gamma(e_{m,\mathcal{L}}, \dot{e}_{m,\mathcal{L}}) = \gamma_0\dot{e}_{m,\mathcal{L}} \tag{11.26}$$

and gains $\gamma_0 = 1.4$, $\gamma_1 = 1$, a very small deviation with respect to the original time (Figures 11.11 and 11.12 bottom) dramatically modifies the tracking

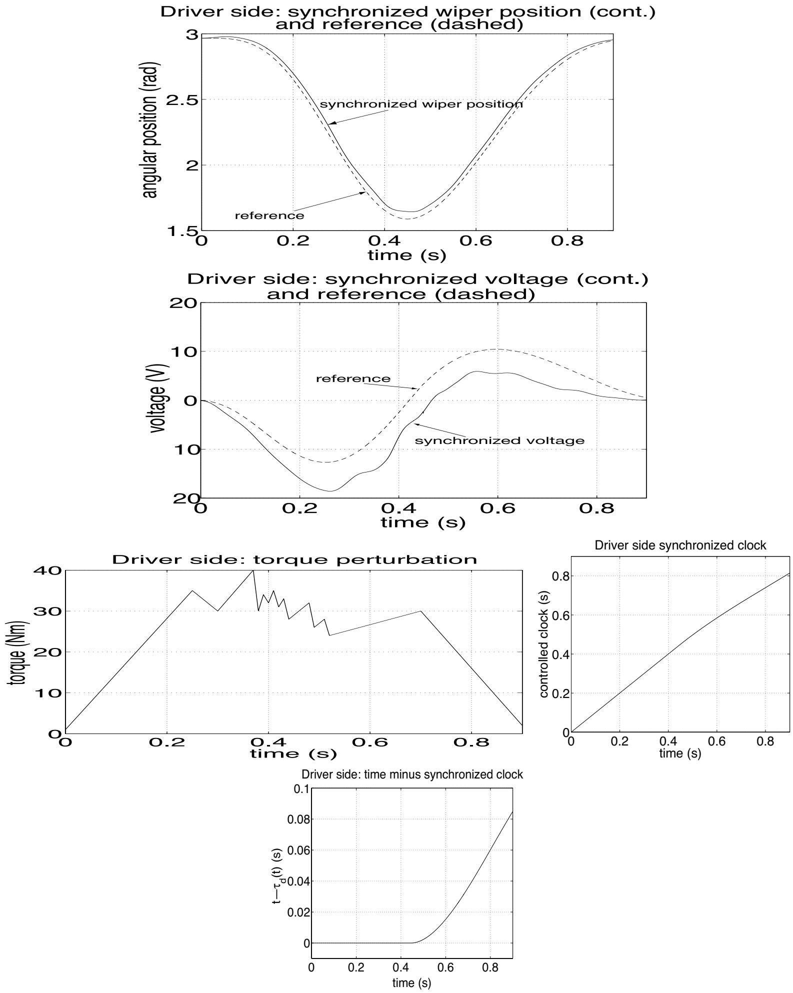

Fig. 11.11 From top to bottom: the driver's synchronized wiper position and its reference; the closed-loop voltage and its reference; the corresponding perturbative torque; the controlled clock and its deviation with respect to time. Saturation voltage: ±24 V.

behavior so that the collision is clearly avoided (figure 11.13 top). A rough global time reset at half period creates the voltage burst of Figure 11.12 (second from top), limited by the voltage saturation at $\pm 24V$. However, contrarily to current bursts, tension bursts present no danger to the motors and the resulting synchronization is acceptable, as seen from Figures 11.11 and 11.12 (top).

Note that the form chosen for the controlled clock (11.21) with (11.26) is such that if the error $\dot{e}_{m,\mathcal{L}}$ on the motor angular velocity of the leader converges to 0, then any equilibrium value of τ is such that

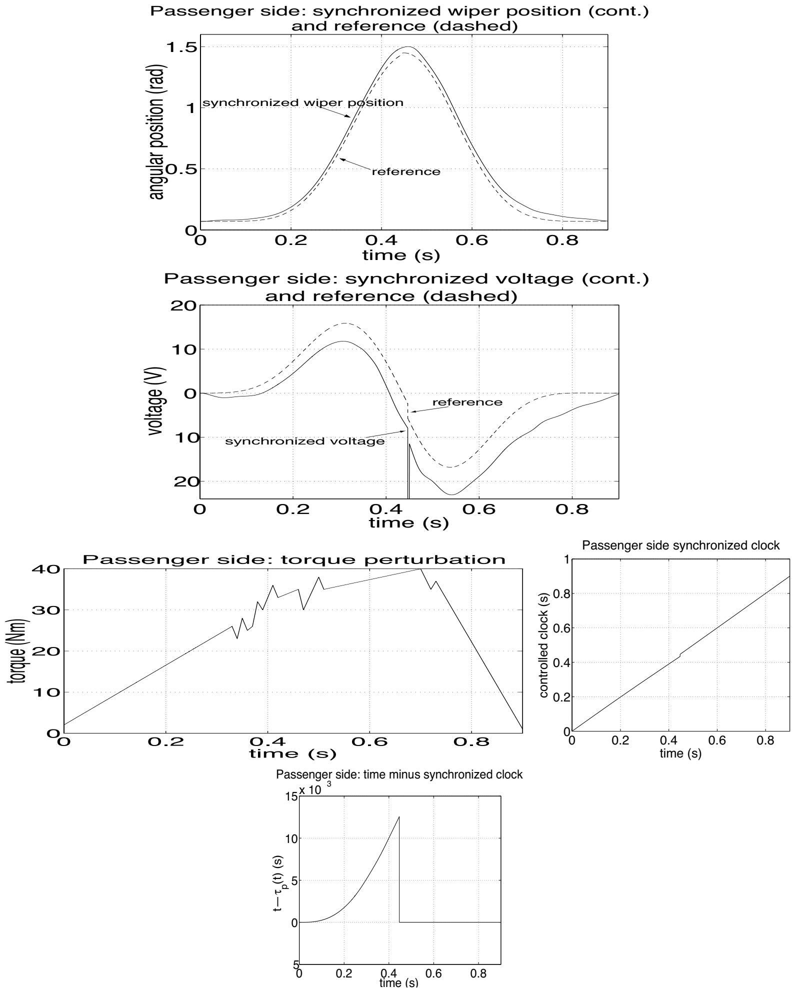

Fig. 11.12 From top to bottom: the passenger's synchronized wiper position and its reference; the closed-loop voltage and its reference; the corresponding perturbative torque; the controlled clock and its deviation with respect to time. Saturation voltage: ± 24 V.

$$0 = -\tilde{\gamma}_1(\dot{\tau} - 1) - \gamma_0 \dot{e}_{m,\mathcal{L}} = -\tilde{\gamma}_1(\dot{\tau} - 1),$$

namely $\dot{\tau} = 1$. Since, moreover, for every $\gamma_0, \tilde{\gamma}_1 > 0$, the controlled clock system is stable, we have proved that $\tau - t$ exponentially converges to a constant, as required. If on the contrary $\dot{e}_{m,\mathcal{L}}$ remains non zero, the clock rate will constantly decrease. This situation corresponds to a system fault since the leader's motor is not able to recover its reference.

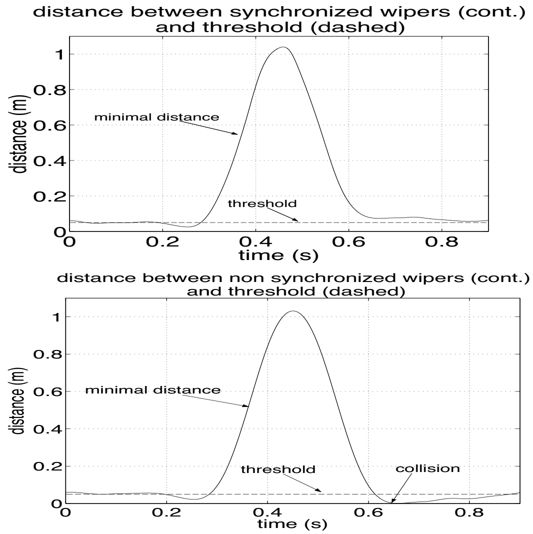

Fig. 11.13 Top: minimal distance between synchronized wipers with $\gamma_0 = 1.4, \gamma_1 = 1$; Bottom: minimal distance without synchronization: a collision occurs at $t \approx 0.655$ s.

Chapter 12
Control of Magnetic Bearings

Introduction

In this chapter[1], we study various aspects of the motion planning and nonlinear feedback design for a rotating shaft actuated by active magnetic bearings. The shaft is here assumed to be rigid and balanced.

The magnetic bearings are made up with pairs of electromagnets, one for each direction to be controlled, possibly completed by a passive thrust. Each electromagnet produces an attractive force, whose intensity depends on the coil current, which explains why they are used by pairs in order to produce forces with both signs.

The modelling and control problems for magnetic bearings has been extensively considered in the open literature, using various approaches. The interested reader may find some references in Lévine et al. [1996], von Löwis et al. [2000].

We propose a simple control design based upon the physical structure of the model, not restricted to linear techniques at an equilibrium point. A major motivation is to improve our mastering of the rotor behavior along given admissible trajectories. This implies being able to do motion planning and trajectory tracking. In this nonlinear perspective, it is also interesting to investigate the role played by premagnetisation currents, generally introduced in many linear approaches of magnetic bearing control to circumvent the singularity at the zero force point. In a more technological perspective, biasing currents are related to iron losses and rotor heating. In some industrial applications such as vacuum pumps, where the rotor heat cannot be easily evacuated, the possibility of avoiding biasing currents without penalizing the dynamical requirements of the bearings is of particular interest.

[1] Joint work with J. Lottin and J.-C. Ponsart. The experimental results presented at the end of ths chapter have been obtained on a bench set-up built at the University of Savoie, Annecy, France.

J. Lévine, *Analysis and Control of Nonlinear Systems,* Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_12, © Springer-Verlag Berlin Heidelberg 2009

Throughout this chapter, the control objectives consist in steering the rotating shaft from a given arbitrary initial position and velocity to another arbitrary position and velocity, with stability, and avoiding as much as possible the use of premagnetisation currents. The electromagnets may be controlled either through their coil currents or their voltages. Therefore, we address the above questions successively in the context of current control, voltage control or hierarchical, or cascaded, control, the latter method consisting of an inner voltage loop and an outer current loop.

Our flatness-based design is most clearly explained when restricting the shaft positioning problem to a one degree of freedom (shortly d.o.f.) system and then generalizing to 4. The 1 d.o.f. problem physically reduces to the positioning of a ball in a one dimensional (e.g. vertical) magnetic field controlled by one pair of electromagnets. In the 4 d.o.f. problem, the positioning of the shaft is operated by 4 pairs of electromagnets.

The nonlinear design of current, voltage and cascaded feedback laws which track sufficiently differentiable reference trajectories are successively presented. The first step consists in planning reference trajectories that can be exactly followed in the absence of perturbations, and then to design a feedback loop that tracks them. We introduce the *current complementarity condition*, which means that, depending on the sign of the required force, only one electromagnet is working at a time. A weaker condition, the *current almost complementarity condition* is also considered. It may be interpreted as the introduction of a small differentiable biasing, vanishing outside a small domain. A current feedback law satisfying the current complementarity condition is constructed. In the voltage case, a linearizing feedback is obtained, which might require infinite voltages at switchings, while the current complementarity condition is only asymptotically satisfied. A cascaded control scheme is then presented in the framework of the current almost complementarity condition. In this case, the regularizing nature of the corresponding biasing greatly simplifies the whole design. This cascaded control strategy is validated by experiments on a 4 d.o.f. active magnetic bearing system, its implementation requiring in particular an observer to reconstruct, from the available measurements, the state used in the feedback loop.

The experiments have been done at the Laboratoire d'Automatique et de Micro-Informatique Industrielle in Annecy (France), on a platform used for pedagogical purposes. However, extensions to practical applications such as pumps, turbines, grinding machines, large gap magnetic suspensions and control of precision motion platforms may be foreseen.

The chapter is organized as follows: the first section concerns the modelling and control of a ball in a vertical magnetic field, controlled by a pair of electromagnets. The motion planning and trajectory tracking are addressed in the context of current, voltage and hierarchical control. The second section extends these results to the general case of a shaft. The third section is devoted to implementation aspects, and in particular to the design of an observer. The experimental results are presented in the last section.

12.1 Analysis and Control of a Ball

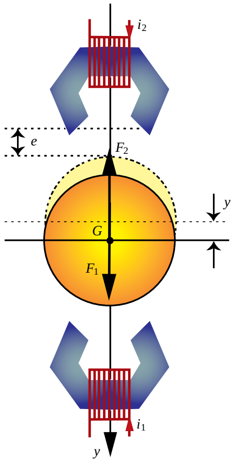

Fig. 12.1 The ball and the pair of electromagnets.

12.1.1 Modelling

The 1 d.o.f. problem reduces to the positioning problem of a ball in a vertical magnetic field created by one pair of electromagnets. Denoting by y the difference between the position of the center of the ball and its nominal position, the y-axis being oriented downwards, e the nominal air gap, F_1 and F_2 the forces created by the two electromagnets, i_1 and i_2 the associated currents, we have, in the open domain $|y| < e$:

$$m\ddot{y} = F_1 - F_2 + mg \tag{12.1}$$

with

$$F_1 = \frac{\lambda_1 (i_1)^2}{(e-y)^2} \ , \quad F_2 = \frac{\lambda_2 (i_2)^2}{(e+y)^2} \tag{12.2}$$

the currents i_1 and i_2 satisfying

$$U_1 = R_1 i_1 + L_1 \frac{di_1}{dt} + i_1 \frac{dL_1}{dt} \ , \quad U_2 = R_2 i_2 + L_2 \frac{di_2}{dt} + i_2 \frac{dL_2}{dt} \tag{12.3}$$

with U_i the input voltage and R_i the resistance of circuit i, $i = 1, 2$, and with the inductances $L_1 = \dfrac{\lambda_1}{(e-y)}$, $L_2 = \dfrac{\lambda_2}{(e+y)}$. The two control variables may be chosen either as the input voltages U_1 and U_2 or, if the electric circuits are assumed much faster than the mechanical dynamics, as the currents i_1 and i_2. This means, in the latter case, that the equations (12.3) are not considered.

12.1.2 Current Control

Recall that the model is given by (12.1) and (12.2), with i_1 and i_2 as control variables.

12.1.2.1 Flatness

A possible choice of flat output is

$$y_1 = y \ , \quad y_2 = \frac{i_2}{e+y}\sqrt{\frac{\lambda_2}{m}} \ , \tag{12.4}$$

Note that y_1 is the position of the center of mass of the ball and that y_2 is the square root of the acceleration induced by the second electromagnet.

Let us prove that the flatness property holds true: since y and the current i_2 are independent variables, the two components y_1 and y_2 of the output are also clearly independent. Moreover, we have:

$$\begin{aligned} y &= y_1, \quad \dot{y} = \dot{y}_1 \\ i_2 &= y_2(e+y_1)\sqrt{\frac{m}{\lambda_2}} \\ (i_1)^2 &= \frac{m}{\lambda_1}(e-y_1)^2(\ddot{y}_1 - g + (y_2)^2) \end{aligned} \tag{12.5}$$

Therefore, all the state and control variables of system (12.1) can be expressed in terms of $(y_1, \dot{y}_1, \ddot{y}_1, y_2)$, which proves our assertion. Note that the expression of i_1 makes sense only if its right-hand side is non negative. We now take this constraint into account in the control design, which indeed begins with path planning considerations.

12.1.2.2 Path Planning With Current Complementarity

By path planning, we mean here the design of a pair of trajectories $t \mapsto (y_1(t), y_2(t))$ starting at time $t = 0$ from $(y_1(0), y_2(0))$, possibly with given

$(\dot{y}_1(0), \ddot{y}_1(0))$, and arriving, at time $t = T$ ($T > 0$ given), at $(y_1(T), y_2(T))$, with possibly given $(\dot{y}_1(T), \ddot{y}_1(T))$.

We want such trajectories to satisfy the requirement that premagnetisation currents are not used, or in other words that one coil current may be zero when the opposite is non zero. Throughout this paper, we will call this property the *current complementarity condition.* It may be expressed as a constraint between the flat outputs: by inspection of (12.5), it is easily seen that $i_2 = 0$ implies $y_2 = 0$ and thus i_1 exists if, and only if, $\ddot{y}_1 - g \geq 0$. On the other hand, $i_1 = 0$ implies that $\ddot{y}_1 - g + (y_2)^2 = 0$ and thus $\ddot{y}_1 - g \leq 0$. Therefore, the current complementarity condition is equivalent to:

$$y_2 = Y(g - \ddot{y}_1) \triangleq \begin{cases} 0 & \text{if } \ddot{y}_1 - g \geq 0 \\ \sqrt{g - \ddot{y}_1} & \text{if } \ddot{y}_1 - g \leq 0\ . \end{cases} \tag{12.6}$$

We call the function Y the *complementarity function.*

Note that this condition does not impose restrictions on the dynamics of the ball or, in other words, that the ball can follow twice differentiable but otherwise arbitrary paths with only one coil energized at a time. We first assign to y_1 an arbitrary time function, at least twice differentiable, and then compute y_2 according to (12.6). The current references are then deduced from $(y_1, \dot{y}_1, \ddot{y}_1, y_2)$ by (12.5).

Nevertheless, for obvious practical reasons, y_2, which is directly related to the current i_2 by (12.4), must remain bounded with a bounded rate of change. Let us construct reference trajectories satisfying these constraints. The boundedness of i_2 is, according to (12.6), a straightforward consequence of the continuity of the complementarity function Y for an at least twice continuously differentiable trajectory y_1. However, since Y is differentiable everywhere but at the zero force point $\ddot{y}_1 = g$, a more careful analysis of the rate of change of y_2 is needed. Whenever $\ddot{y}_1 \neq g$, we have

$$\dot{y}_2 = \begin{cases} 0 & \text{if } \ddot{y}_1 - g > 0 \\ -\dfrac{y_1^{(3)}}{2\sqrt{g - \ddot{y}_1}} & \text{if } \ddot{y}_1 - g < 0\ . \end{cases}$$

All along this paper, we use the notation $y_1^{(j)} = \dfrac{d^j y_1}{dt^j}$ for $j \geq 3$.

We now evaluate the right and left limits $\dot{y}_2(t_0^-) = \lim_{t \to t_0,\ t < t_0} \dot{y}_2(t)$ and $\dot{y}_2(t_0^+) = \lim_{t \to t_0,\ t > t_0} \dot{y}_2(t)$ at the time t_0 when the zero force point is reached, namely when $\ddot{y}_1(t_0) = g$. Using the lemma 12.1 of section 12.1.2.3, we assume that y_1 is analytic in the interval $t_0 - \varepsilon \leq t < t_0$, with $\ddot{y}_1(t) < g$ for $t_0 - \varepsilon \leq t < t_0$. With the notations of lemma 12.1, by posing $\varphi(t) = g - \ddot{y}_1$ and $\psi(t) = \dot{y}_2(t)$, we find that $\dot{y}_2$ has a finite left limit when t tends to t_0 if, and only if, $y_1^{(3)}(t_0^-) = 0$. Moreover, $y_1^{(4)}(t_0^-) \leq 0$, and

$$\dot{y}_2(t_0^-) = -\sqrt{\frac{1}{2}|y_1^{(4)}(t_0^-)|}\ . \tag{12.7}$$

It is straightforward to check that the right limit is $\dot{y}_2(t_0^+) = 0$. We have thus proved that if the reference trajectory y_1 is regular enough for the expression (12.7) to be finite, the rate of change of the currents i_1 and i_2 are finite at switchings.

Let us now outline the construction of such reference trajectories. We assume for simplicity's sake that we want to move from one steady-state to another one. We start, at time $t = 0$, from an arbitrary position $y(0)$, $|y(0)| < e$, with $\dot{y}(0) = 0$, and we want to reach, at time $t = T$, the arbitrary position $y_1(T)$, $|y_1(T)| < e$, with $\dot{y}_1(T) = 0$. The continuity requirements of the initial and final currents and the finiteness of the current rate of changes at the initial and final times, are given by: $\ddot{y}_1(0) = \ddot{y}_1(T) = y_1^{(3)}(0) = y_1^{(3)}(T) = 0$. Finally, at switching, namely at that time when one electromagnet turns off and the opposing electromagnet turns on, the conditions are $y_1^{(3)}(\tau) = 0$, for every $\tau \in [0, T]$ such that $\ddot{y}_1(\tau) = g$.

If we choose a trajectory $y_1(t)$ for $0 \leq t \leq T$ in the class of polynomials in t, namely

$$y_1(t) = \sum_{i=0}^{k} a_i t^i \ , \tag{12.8}$$

its minimal degree must be 7 to satisfy the 8 initial and final constraints. However, a polynomial trajectory is too regular to satisfy the above conditions at switchings. We therefore have to look for y_1 in the class of piecewise polynomials. The construction of a piecewise polynomial trajectory of degree 7 meeting the above requirements is always possible. Details may be found in appendix 12.1.2.4. In figure 12.2, such a trajectory of $y(t)$ with the corresponding currents $(i_1(t), i_2(t))$ are presented for a displacement from $y(0) = -1.10^{-4}$ m to $y(T) = 4.10^{-4}$ m, with $T = 10.10^{-3}$ s. The ball mass is $m = 0.2$ kg, the air gap is $e = 5.10^{-4}$ m and $\lambda = 5.10^{-6}$ ISU.

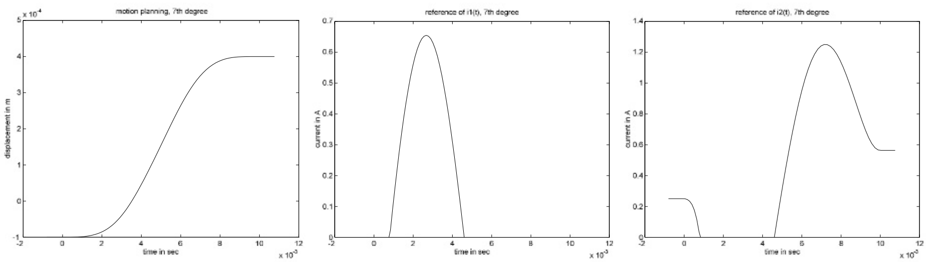

Fig. 12.2 Reference position and currents with 7th degree polynomials and current complementarity

12.1.2.3 A Lemma

Lemma 12.1. *Let φ be a non negative analytic time function on the interval $]t_0-\varepsilon, t_0[$, with left limits $\varphi(t_0^-)$, $\dot\varphi(t_0^-)$ and $\ddot\varphi(t_0^-)$, and with $\varphi(t_0^-)=0$. Let ψ be the function defined by $\psi(t)=\dfrac{\dot\varphi(t)}{2\sqrt{\varphi(t)}}$. The left limit $\psi(t_0^-)$ exists if, and only if, $\dot\varphi(t_0^-)=0$. In this case, we have $\ddot\varphi(t_0^-)\geq 0$ and*

$$\psi(t_0^-) = -\sqrt{\frac{1}{2}\ddot\varphi(t_0^-)}\ . \tag{12.9}$$

Furthermore, if φ is also non negative and analytic on $]t_0, t_0+\varepsilon[$, and has right limits $\varphi(t_0^+)$, $\dot\varphi(t_0^+)$ and $\ddot\varphi(t_0^+)$, with $\varphi(t_0^+)=0$, ψ has a right limit as $t\to t_0$ if, and only if, $\dot\varphi(t_0^+)=0$ and

$$\psi(t_0^+) = \sqrt{\frac{1}{2}\ddot\varphi(t_0^+)}\ . \tag{12.10}$$

Proof. The condition $\dot\varphi(t_0^-)=0$ is clearly necessary. To prove the sufficiency, let us prolong φ analytically in a neighborhood containing t_0 in its interior and expand the prolongated function (still denoted by φ) in Taylor series. Since φ and its first derivative vanish at t_0, we have: $\varphi(t)=\frac{1}{2}(t-t_0)^2\ddot\varphi(t_0^-)+0(|t-t_0|^3)$. Thus, $\varphi(t)\geq 0$ for $t\in]t_0-\varepsilon,t_0[$ implies that $\ddot\varphi(t_0^-)\geq 0$ and $\sqrt{\varphi(t)}=|t-t_0|\sqrt{\frac{1}{2}\ddot\varphi(t_0^-)}+0(|t-t_0|^2)$. Also, differentiating the expression of φ yields $\dot\varphi(t)=(t-t_0)\ddot\varphi(t_0^-)+0(|t-t_0|^2)$. Thus, since $\frac{t-t_0}{|t-t_0|}=-1$,

$$\psi(t)=\frac{(t-t_0)\ddot\varphi(t_0^-)+0(|t-t_0|^2)}{2|t-t_0|\sqrt{\frac{1}{2}\ddot\varphi(t_0^-)}+0(|t-t_0|^2)}=-\frac{\frac{1}{2}\ddot\varphi(t_0^-)}{\sqrt{\frac{1}{2}\ddot\varphi(t_0^-)}}+0(|t-t_0|)$$

which proves that the left limit exists and is given by (12.9). The proof of (12.10) follows the same lines and is omitted.

12.1.2.4 A Polynomial Interpolation Algorithm

We sketch a construction of reference trajectories for y, i_1, i_2, U_1 and U_2 that are well-defined and bounded everywhere, including at switchings. In particular, we want y_1 to have continuous derivatives up to third order. To this aim, we propose the following algorithm:

- we first construct a polynomial curve satisfying the initial and final requirements, using standard polynomial interpolation methods,
- the switching points belonging to this trajectory are then computed,

- the coordinates of the switching points are then used to construct a sequence of polynomials with appropriate degree that join the initial point to the first switching point, two successive switching points, and then the last switching point to the final one. Each polynomial represents a path joining two successive points. Its coefficients are computed by standard methods,
- we check that the new polynomials do not create new switchings, to which case we perturb the coordinates of the previous switching points until no new switchings appear.

For simplicity's sake, we give this construction for 7th degree polynomials. The same kind of construction may be done with higher degrees and possibly additional constraints on the derivatives.

Precisely, let us denote $Y = (y_{0,1}, \ldots, y_{3,1}, y_{0,2}, \ldots, y_{3,2}, T_1, T_2)$ where T_1 (resp. T_2) is the initial (resp. final) time, and where $y_{i,j} = y_1^{(i)}(T_j)$, $i = 0, \ldots, 3$, $j = 1, 2$. We set $P_Y(t) = \sum_{i=0}^{7} a_i(Y)t^i$. P_Y is thus the interpolating polynomial satisfying the initial and final requirements specified in the vector Y. We first construct P_{Y_0} where $Y_0 = (y_1(0), 0, 0, 0, y_1(T), 0, 0, 0, 0, T)$is the vector meaning that we start at $t = 0$ from $y_1(0)$, which is a steady-state, and arrive at the point $y_1(T)$, which is also a steady-state, at time $t = T$. The coefficients $a_i(Y_0)$'s are computed by standard interpolation methods. We then compute the set of switching instants:

$$\{t_1, \ldots, t_q\} \stackrel{\Delta}{=} \{0 < t < T | \ddot{P}_{Y_0}(t) = g\}$$

with $q \leq 5$ since $\ddot{P}_{Y_0}$ is a 5th degree polynomial. We set $t_0 = 0$ and $t_{q+1} = T$. Note that the coordinates of a switching point and its derivatives up to order 3 are, in these notations, $(P_{Y_0}(t_j), \dot{P}_{Y_0}(t_j), g, P_{Y_0}^{(3)}(t_j))$, since $\ddot{P}_{Y_0}(t_j) = g$. To satisfy the switching conditions such that the currents and voltages remain bounded, we have shown in lemma 12.1 that we must replace the 3rd derivative of P_{Y_0} at t_j by 0. Therefore, we are led to construct the new polynomials corresponding to the vectors Z_j of pairs of successive switching points. We set

$$\begin{aligned}
Z_0 &= (P_{Y_0}(t_0), 0, 0, 0, P_{Y_0}(t_1), \dot{P}_{Y_0}(t_1), g, 0, t_0, t_1), \\
Z_j &= (P_{Y_0}(t_j), \dot{P}_{Y_0}(t_j), g, 0, P_{Y_0}(t_{j+1}), \dot{P}_{Y_0}(t_{j+1}), g, 0, t_j, t_{j+1}), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } j = 1, \ldots, q-1, \\
Z_q &= (P_{Y_0}(t_q), \dot{P}_{Y_0}(t_q), g, 0, P_{Y_0}(t_{q+1}), 0, 0, 0, t_q, T).
\end{aligned}$$

Once the 7th degree polynomials P_{Z_j} are obtained for $j = 0, \ldots, q$, we check that the switching points obtained with these new polynomials are the same as those of the previous list, namely that $\ddot{P}_{Z_j}(t) \neq g$ for all $t \in]t_j, t_{j+1}[$, $j = 0, \ldots, q$.

If it is the case, we set

$$y_1(t) = \sum_{j=0}^{q} P_{Z_j}(t) 1_{[t_j, t_{j+1}[}(t)$$

where $1_A(t) = \begin{cases} 1 \text{ if } t \in A \\ 0 \text{ otherwise} \end{cases}$ for a given subset A of the real line. We have constructed a reference for y_1 which is 4 times differentiable everywhere but at the switching instants and with 4th derivative having left and right limits at the switching instants, so that the currents and voltages are well-defined.

If, on the contrary, one of the polynomials satisfies $\ddot{P}_{Z_j}(t) = g$ for at least one $t \in]t_j, t_{j+1}[$, then we perturb the vector Z_j, and, if necessary, its neighbors, until we fall into the previous case.

Note that in the generic case, this last step is not necessary.

12.1.2.5 Trajectory Tracking

Let us now focus attention on the feedback design to track a reference trajectory of the above type, denoted by (y_1^*, y_2^*), and assumed at least twice differentiable. We aim at designing a feedback law such that the tracking error $(y_1 - y_1^*, y_2 - y_2^*)$ asymptotically vanishes. More precisely, we may ask that the first component of the tracking error satisfies

$$\ddot{y} - \ddot{y}_1^* = -k_0(y - y_1^*) - k_1(\dot{y} - \dot{y}_1^*) \tag{12.11}$$

where k_0 and k_1 are real numbers such that the second degree characteristic polynomial $P_2(s) = s^2 + k_1 s + k_0$ is Hurwitz[2].

Let us introduce the following notation:

$$A(y, \dot{y}) = -k_0(y - y_1^*) - k_1(\dot{y} - \dot{y}_1^*) - (g - \ddot{y}_1^*) \; . \tag{12.12}$$

Theorem 12.1. *Assume that k_0 and k_1 are chosen such that the second degree characteristic polynomial P_2 of (12.11) is Hurwitz. The control law*

$$\begin{cases} i_1 = \sqrt{\dfrac{m}{\lambda_1}}(e - y)\sqrt{A(y, \dot{y})} \; , \quad i_2 = 0 \quad \text{if } A(y, \dot{y}) \geq 0 \\ i_1 = 0 \; , \quad i_2 = \sqrt{\dfrac{m}{\lambda_2}}(e + y)\sqrt{-A(y, \dot{y})} \text{ if } A(y, \dot{y}) \leq 0 \end{cases} \tag{12.13}$$

is such that:

[2] Recall that this property means that the two roots of P_2 have negative real parts. In other words, (12.11) is exponentially stable with damping rate given by the smallest (negative) real part of the roots of P_2. It is also useful to interpret this property in terms of damping time constant, the time constant being the inverse of the absolute value of the smallest root's real part

(i) *the current complementarity condition is satisfied,*
(ii) *the first component of the tracking error is exponentially stable and satisfies (12.11),*
(iii) *the second component* $(y_2 - y_2^*)$ *of the tracking error tends to zero as* $t \to \infty$.

Proof. Using the linearizability property of flat systems (see Annex A.3), by differentiating twice the first output y_1, it is straightforward to verify that the system (12.1), (12.2) can be transformed into the linear controllable system:

$$\begin{cases} \ddot{y}_1 = v_1 \\ y_2 = v_2 \end{cases} \tag{12.14}$$

by using the static state feedback

$$v_1 = \frac{\lambda_1 i_1^2}{m(e-y)^2} - \frac{\lambda_2 i_2^2}{m(e+y)^2} + g \ , \quad v_2 = \frac{i_2}{e+y}\sqrt{\frac{\lambda_2}{m}} \ . \tag{12.15}$$

By construction, the reference trajectory $t \mapsto y_1^*(t)$ is at least twice differentiable. Then, if we want the closed loop system to satisfy (12.11) with the gains k_0 and k_1 chosen as above to ensure the exponential stability of $y_1 - y_1^*$, we get

$$v_1 = \ddot{y}_1^* - k_0(y_1 - y_1^*) - k_1(\dot{y}_1 - \dot{y}_1^*) \ . \tag{12.16}$$

Therefore, it is easily seen that the currents

$$\begin{cases} i_1 = \sqrt{\dfrac{m}{\lambda_1}}(e-y)\sqrt{v_1 - g} \ , \quad i_2 = 0 \text{ if } v_1 - g \geq 0 \\ i_1 = 0 \ , \quad i_2 = \sqrt{\dfrac{m}{\lambda_2}}(e+y)\sqrt{g - v_1} \text{ if } v_1 - g \leq 0 \end{cases} \tag{12.17}$$

ensure that (12.11) is satisfied. Combining (12.17) and (12.16), we get (12.13). Furthermore, y_2 satisfies the constraint (12.6) and since the mapping Y is continuous with respect to $\ddot{y}_1$, it results that the exponential convergence of y_1 to y_1^* as t approaches ∞ implies that y_2 converges to y_2^*, which achieves the proof.

Remark 12.1. Note that since Y is not differentiable at $\ddot{y}_1 = g$, the convergence of y_2 is not in general exponential. Remark that, for linear controllable systems, exponential stabilization is always possible. Therefore, the impossibility of designing a feedback law that guarantees an exponential convergence results from the nonlinear structure of the problem, and, more precisely, from the singularity occurring when $\ddot{y} = g$.

Remark 12.2. The above feedback linearization result shows that, though the system is static feedback linearizable, we are not in a classical situation since the second output has no dynamics: $y_2 = v_2$. This is easily handled in the differential flatness approach whereas it falls beyond the static feedback linearizability results of Jakubczyk and Respondek [1980], Hunt et al. [1983a].

Remark 12.3. The feedback law (12.13) may be seen as a force control scheme, followed by a nonlinear inversion of the actuator model.

Remark 12.4. Note that the same feedback applies to the case of a ball rolling on a plane without friction in a horizontal magnetic field, the only difference being that the weight mg disappears. Let us stress that in this case, the linear tangent model at the origin is not controllable whereas this singularity is not present in our nonlinear approach using differential flatness.

12.1.3 Voltage Control

We now consider the model made up with (12.1), (12.2) and (12.3), with U_1 and U_2 as control variables.

12.1.3.1 Flatness

In this case, the same choice of outputs yields, in addition to (12.5), the following relations:

$$\begin{aligned} U_1 &= R_1\sqrt{\frac{m}{\lambda_1}}(e-y_1)\sqrt{\ddot{y}_1 - g + (y_2)^2} + \frac{\sqrt{m\lambda_1}}{2}\frac{y_1^{(3)} + 2y_2\dot{y}_2}{\sqrt{\ddot{y}_1 - g + (y_2)^2}} \\ U_2 &= R_2 y_2(e+y_1)\sqrt{\frac{m}{\lambda_2}} + \dot{y}_2\sqrt{m\lambda_2} \end{aligned} \tag{12.18}$$

which proves that all the system variables may be expressed in terms of the output (y_1, y_2) and a finite number of its derivatives, or, otherwise stated, that the system is still flat, with the same flat output.

Note that the expression of U_1 is only defined, as before, if, and only if, $\ddot{y}_1 - g + (y_2)^2 > 0$. This is verified as long as the force F_1 is different from 0.

12.1.3.2 Path Planning with Current Complementarity

Since we require that the current complementarity condition (12.6) holds, combining (12.18) and (12.6), we get:

$$U_1 = \begin{cases} R_1\sqrt{\dfrac{m}{\lambda_1}}(e-y_1)\sqrt{\ddot{y}_1 - g} + \dfrac{\sqrt{m\lambda_1}}{2}\dfrac{y_1^{(3)}}{\sqrt{\ddot{y}_1 - g}} & \text{if } \ddot{y}_1 - g > 0 \\ 0 \quad \text{if } \ddot{y}_1 - g < 0 \end{cases}$$
$$U_2 = \begin{cases} 0 \quad \text{if } \ddot{y}_1 - g > 0 \\ R_2\sqrt{\dfrac{m}{\lambda_2}}(e+y_1)\sqrt{g - \ddot{y}_1} + \sqrt{m\lambda_2}\dfrac{y_1^{(3)}}{2\sqrt{g - \ddot{y}_1}} & \text{if } \ddot{y}_1 - g < 0 \;. \end{cases} \tag{12.19}$$

For the input variables U_1 and U_2 to remain bounded, according to lemma 12.1, it is necessary and sufficient that $y_1^{(3)}(t_0) = 0$ at every switching time t_0, and we have

$$\begin{cases} U_1(t_0^-) = -\sqrt{\dfrac{m\lambda_1}{2}|y_1^{(4)}(t_0^-)|}\;, \quad U_2(t_0^-) = 0 \text{ if } \ddot{y}_1(t) > g \text{ for } t < t_0 \\ U_1(t_0^-) = 0\;, \quad U_2(t_0^-) = -\sqrt{\dfrac{m\lambda_2}{2}|y_1^{(4)}(t_0^-)|} \text{ if } \ddot{y}_1(t) < g \text{ for } t < t_0 \end{cases} \tag{12.20}$$

The analysis of the switching situation corresponding to the limits when $t \to t_0$, $t > t_0$ follows the same lines.

The references of figure 12.2 have already been computed to ensure that the limits (12.20) at switchings are finite. The corresponding voltages are shown in figure 12.3.

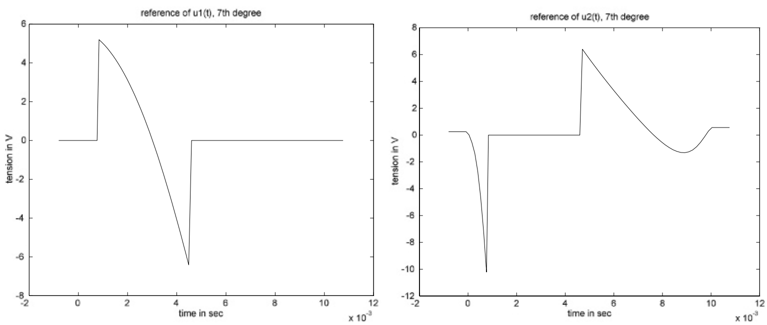

Fig. 12.3 Reference voltages with 7th degree polynomials

12.1.3.3 Trajectory Tracking

Following the same lines as in the feedback construction of theorem 12.1, the system is feedback equivalent to the linear controllable system

$$\begin{cases} y_1^{(3)} = v_1 \\ \dot{y}_2 = v_2 \ . \end{cases} \tag{12.21}$$

Given suitable reference trajectories (y_1^*, y_2^*) satisfying the current complementarity condition, if we want to assign the dynamics of the error $(y_1 - y_1^*, y_2 - y_2^*)$ as follows:

$$\begin{cases} y_1^{(3)} - (y_1^*)^{(3)} = -k_0(y_1 - y_1^*) - k_1(\dot{y}_1 - \dot{y}_1^*) - k_2(\ddot{y}_1 - \ddot{y}_1^*) \\ \dot{y}_2 - \dot{y}_2^* = -k_0'(y_2 - y_2^*) \end{cases} \tag{12.22}$$

then U_1 and U_2 can be computed in feedback form by inverting the following equations, obtained by differentiating (12.1), (12.2), with (12.3):

$$\begin{cases} y_1^{(3)} = \dfrac{2L_1 i_1}{m\lambda_1}(U_1 - R_1 i_1) - \dfrac{2L_2 i_2}{m\lambda_2}(U_2 - R_2 i_2) \\ \dot{y}_2 = \dfrac{1}{\sqrt{m\lambda_2}}(U_2 - R_2 i_2) \ . \end{cases} \tag{12.23}$$

We obtain:

$$\begin{cases} i_1 U_1 = m(e - y_1)(\dfrac{R_1}{\lambda_1}(e - y_1)(\ddot{y}_1 - g + (y_2)^2) + y_2 u_2 + v_1) \\ U_2 = \sqrt{m\lambda_2}(\dfrac{R_2}{\lambda_2} y_2(e + y_1) + v_2) \end{cases} \tag{12.24}$$

with

$$\begin{cases} v_1 = (y_1^*)^{(3)} - k_0(y_1 - y_1^*) - k_1(\dot{y}_1 - \dot{y}_1^*) - k_2(\ddot{y}_1 - \ddot{y}_1^*) \\ v_2 = \dot{y}_2^* - k_0'(y_2 - y_2^*) \end{cases} \tag{12.25}$$

Though U_2 is well-defined, U_1 is not defined when $i_1 = 0$. This is the case in particular at switchings, independently of the fact that the reference trajectory satisfies the current complementarity condition and has bounded current and voltage limits at switchings. However, contrarily to the current control case, the voltage feedback (12.24) does not automatically imply that the current complementarity condition is met: the currents exponentially converge to their references which satisfy it.

In the next section, we propose a different approach, with a simpler design, in terms of hierarchical control.

12.1.4 Hierarchical Control

If we want to control the bearings by the input voltages without being confronted to the above singularity problem, a simple solution consists in starting from the current control scheme of theorem 12.1 and designing a low-level control loop to track the current reference trajectory with stability. To be more specific, we consider the system (12.1) as if it were controlled by the currents i_1 and i_2 and the voltages U_1 and U_2 are designed in such a way that these currents are tracked at a faster rate.

Nevertheless, since the tracked currents are now the real ones, rather than their open-loop references, their regularity at switchings cannot be guaranteed anymore in the presence of perturbations. Therefore, we cannot ensure that $y_1^{(3)}$ vanishes at switchings and the previous approach may fail to produce currents with a bounded rate of change. We now develop a slightly different approach where a specific kind of biasing is allowed to overcome this problem. Note that the main difficulty raised by the current complementarity condition is the lack of differentiability of the complementarity function at the zero force point. This is why we now propose to regularize it.

12.1.4.1 A Regularization Procedure

In place of dealing with the non differentiable complementarity function Y defined by (12.6), we introduce an *almost complementarity function* Y_η where η is a small positive real number. To this aim, we replace Y on the interval $-\eta \leq \xi \leq \eta$, by a non decreasing polynomial $P_\eta(\xi)$, with $P_\eta(\xi) > Y(\xi)$ for $\xi \in]-\eta, +\eta[$, $P_\eta(\pm\eta) = Y(\pm\eta)$, such that the *almost complementarity function* Y_η defined by

$$Y_\eta(\xi) = \begin{cases} \sqrt{\xi} & \text{if } \xi \geq \eta \\ P_\eta(\xi) & \text{if } -\eta \leq \xi \leq \eta \\ 0 & \text{if } \xi \leq -\eta \end{cases} \tag{12.26}$$

is at least twice continuously differentiable.

To ensure that Y_η be twice differentiable, the contact conditions between P_η and Y at $\xi = \pm\eta$ must be:

$$\begin{aligned} &P_\eta(-\eta) = \frac{dP_\eta}{d\xi}(-\eta) = \frac{d^2P_\eta}{d\xi^2}(-\eta) = 0 \\ &P_\eta(\eta) = \sqrt{\eta}\,, \quad \frac{dP_\eta}{d\xi}(\eta) = \frac{1}{2\sqrt{\eta}}\,, \quad \frac{d^2P_\eta}{d\xi^2}(\eta) = -\frac{1}{4\sqrt{\eta^3}}\,. \end{aligned} \tag{12.27}$$

For example P_η may be chosen as:

$$P_\eta(\xi) = \alpha_3(\xi+\eta)^3 + \alpha_4(\xi+\eta)^4 + \alpha_5(\xi+\eta)^5 + \alpha_6(\xi+\eta)^6$$

with $\alpha_3 = \frac{15}{16}\eta^{-\frac{5}{2}}$, $\alpha_4 = -\frac{13}{16}\eta^{-\frac{7}{2}}$, $\alpha_5 = \frac{17}{64}\eta^{-\frac{9}{2}}$ and $\alpha_6 = -\frac{1}{32}\eta^{-\frac{11}{2}}$. The graph of P_η is shown in figure 12.4. We have

$$\sup_{\xi\in\mathbb{R}} |Y_\eta(\xi) - Y(\xi)| = |Y_\eta(0) - Y(0)| = \frac{23}{64}\eta^{\frac{1}{2}}\,.$$

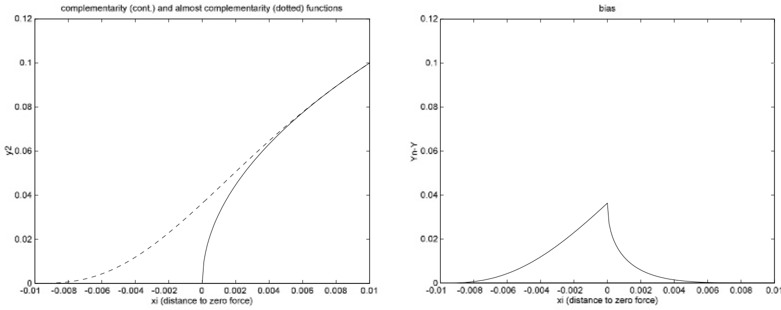

Fig. 12.4 The almost complementarity function and the corresponding non constant bias

This regularization procedure may be interpreted, when $\xi = g - \ddot{y}_1$, as the fact that, before a switching, when the acceleration of the ball is equal to $g + \eta$, a small current is injected in the opposite electromagnet until the acceleration has decreased to $g - \eta$. This current may be interpreted as a small premagnetisation current that works only during a small interval of time, corresponding to the duration needed to achieve the switching. Moreover, this bias is non constant and decreases with the distance to zero force as shown in figure 12.4. Note that outside the domain defined by $-\eta \leq \ddot{y}_1 - g \leq \eta$, the currents i_1 and i_2 exactly satisfy the current complementarity condition. Otherwise stated, if we note $F_{2,\eta} = m(Y_\eta(g - \ddot{y}_1))^2$ and $F_{1,\eta} = m(\ddot{y}_1 - g + (Y_\eta(g - \ddot{y}_1))^2)$, the corresponding forces, and F_1 and F_2 the previously defined forces satisfying the current complementarity condition, it can be easily seen that $F_{i,\eta} = F_i$, $i = 1, 2$, whenever $\ddot{y}_1 - g \leq -\eta$ or $\eta \leq \ddot{y}_1 - g$ and $|F_{i,\eta} - F_i| \leq C\eta$, C being a finite positive real number, for $-\eta \leq \ddot{y}_1 - g \leq \eta$, which shows that there is no biasing outside the arbitrarily small switching domain $-\eta \leq \ddot{y}_1 - g \leq \eta$. This justifies the name *almost complementarity function* given to Y_η. We will also speak of the *current almost complementarity condition* to call the corresponding requirement on the currents.

We now show how to use this approximation in the cascaded design.

12.1.4.2 Path Planning With Almost Current Complementarity

We now roughly sketch the construction of paths that satisfy the current almost complementarity condition. In other words, the second output y_2 must satisfy:

$$y_2 = Y_\eta(g - \ddot{y}_1) = \begin{cases} \sqrt{g - \ddot{y}_1} & \text{if } g - \ddot{y}_1 \geq \eta \\ P_\eta(g - \ddot{y}_1) & \text{if } -\eta \leq g - \ddot{y}_1 \leq \eta \\ 0 & \text{if } g - \ddot{y}_1 \leq -\eta\ . \end{cases} \tag{12.28}$$

As before, we begin our design by choosing a trajectory y_1 to be followed by the ball, and then the currents and voltages are deduced by using (12.28). Combining (12.5), (12.18) and (12.28), we obtain the following formulas:

- if $\ddot{y}_1 - g \geq \eta$

$$\begin{aligned} & y_1 = y \; , \quad y_2 = 0 \\ & i_1 = \sqrt{\frac{m}{\lambda_1}}(e - y_1)\sqrt{\ddot{y}_1 - g} \; , \quad i_2 = 0 \\ & U_1 = R_1\sqrt{\frac{m}{\lambda_1}}(e - y_1)\sqrt{\ddot{y}_1 - g} + \frac{\sqrt{m\lambda_1}}{2}\frac{y_1^{(3)}}{\sqrt{\ddot{y}_1 - g}} \; , \quad U_2 = 0 \end{aligned} \tag{12.29}$$

- if $-\eta \leq \ddot{y}_1 - g \leq \eta$

$$\begin{aligned} & y_1 = y \; , \quad y_2 = P_\eta(g - \ddot{y}_1) \\ & i_1 = \sqrt{\frac{m}{\lambda_1}}(e - y_1)\sqrt{\ddot{y}_1 - g + P_\eta^2(g - \ddot{y}_1)}, \\ & i_2 = \sqrt{\frac{m}{\lambda_2}}(e + y_1)P_\eta(g - \ddot{y}_1) \\ & U_1 = R_1\sqrt{\frac{m}{\lambda_1}}(e - y_1)\sqrt{\ddot{y}_1 - g + P_\eta^2(g - \ddot{y}_1)} \\ & \qquad + \frac{\sqrt{m\lambda_1}y_1^{(3)}\left(1 - 2P_\eta(g - \ddot{y}_1)P'_\eta(g - \ddot{y}_1)\right)}{2\sqrt{\ddot{y}_1 - g + P_\eta^2(g - \ddot{y}_1)}} \\ & U_2 = R_2\sqrt{\frac{m}{\lambda_2}}(e + y_1)P_\eta(g - \ddot{y}_1) - \sqrt{m\lambda_2}y_1^{(3)}P'_\eta(g - \ddot{y}_1) \end{aligned} \tag{12.30}$$

- if $\ddot{y}_1 - g \leq -\eta$

$$\begin{aligned} & y_1 = y \; , \quad y_2 = \sqrt{g - \ddot{y}_1} \\ & i_1 = 0 \; , \quad i_2 = \sqrt{\frac{m}{\lambda_2}}(e + y_1)\sqrt{g - \ddot{y}_1} \\ & U_1 = 0 \; , \quad U_2 = R_2\sqrt{\frac{m}{\lambda_2}}(e + y_1)\sqrt{g - \ddot{y}_1} - \frac{\sqrt{m\lambda_2}}{2}\frac{y_1^{(3)}}{\sqrt{g - \ddot{y}_1}} \end{aligned} \tag{12.31}$$

Note that if y_1 is an at least 4 times continuously differentiable function of time, since the almost complementarity function Y_η is twice continuously differentiable by construction, the obtained references for i_1, i_2 are continuously differentiable and U_1 and U_2 are continuous.

Here, the current references are designed to satisfy the almost complementarity condition:

$$\begin{aligned} & i_1 = \sqrt{\frac{m}{\lambda_1}}(e - y_1)\sqrt{\ddot{y}_1 - g + Y_\eta^2(g - \ddot{y}_1)}, \\ & i_2 = \sqrt{\frac{m}{\lambda_2}}(e + y_1)Y_\eta(g - \ddot{y}_1). \end{aligned} \tag{12.32}$$

12.1.4.3 Hierarchical Trajectory Tracking With Almost Current Complementarity

The high-level controller is given, in place of (12.13), by

$$\begin{aligned} i_1^{**} &= \sqrt{\frac{m}{\lambda_1}}(e-y_1)\sqrt{A(y_1,\dot{y}_1)+Y_\eta^2(-A(y_1,\dot{y}_1))}, \\ i_2^{**} &= \sqrt{\frac{m}{\lambda_2}}(e+y_1)Y_\eta(-A(y_1,\dot{y}_1)) \end{aligned} \tag{12.33}$$

with $A(y_1, \dot{y}_1)$ given by (12.12). Here, in order to distinguish between low and high-level variables, we have renamed i_j by i_j^{**}, $j = 1, 2$.

We want to design the voltage loops such that

- the errors between the currents i_j and their reference i_j^{**}, $j = 1, 2$, locally decrease exponentially fast, the use of high-gains being allowed,
- the high-level tracking error between the ball position and its reference is stabilized,
- without using a precise knowledge of the currents dynamics.

Setting

$$U_j = -K_j(i_j - i_j^{**}) \ , \quad j = 1, 2 \tag{12.34}$$

the error equation satisfies:

$$L_j \frac{d}{dt}(i_j - i_j^{**}) = -(\frac{dL_j}{dt} + R_j + K_j)(i_j - i_j^{**}) - U_j^{**} \ , \quad j = 1, 2 \tag{12.35}$$

where

$$U_j^{**} = R_j i_j^{**} + L_j \frac{di_j^{**}}{dt} + i_j^{**}\frac{dL_j}{dt} \ , \quad j = 1, 2 \ . \tag{12.36}$$

Note that U_j^{**} is not used in the low-level loop (12.34). It depends on the coefficients of the currents dynamics, while the feedback (12.34) does not. Therefore, if we can stabilize (12.35) by (12.34), the result will not be sensitive to modelling inaccuracies in (12.35).

Note also that, according to the above regularization, U_j^{**} remains bounded. This point is crucial for the applicability of the cascaded control design.

Let us introduce ε a small positive real number and let us choose the gain K_j high enough, namely such that

$$\frac{\frac{dL_j}{dt} + R_j - K_j}{L_j} = \frac{\kappa_j}{\varepsilon} \tag{12.37}$$

where κ_j is an arbitrary positive number, $j = 1, 2$. Using standard singular perturbation arguments (see Theorems 3.8 and 3.9. See also Kokotović et al. [1986]), we obtain:

Theorem 12.2. *For every ε small enough, the high-gain feedback (12.34) is such that*

(i) the current error of subsystem (12.35) exponentially converges to a limit of order $O(\varepsilon)$ as $t \to +\infty$ and its closed-loop dynamics is fast compared to the dynamics of $y_1 - y_1^*$,
(ii) the ball position error $y_1(t) - y_1^*(t)$ exponentially converges to a limit of order $O(\varepsilon)$ as $t \to +\infty$.

Proof. We have

$$\begin{aligned} \ddot{y}_1 - \ddot{y}_1^* &= \frac{\lambda_1}{m}\left(\frac{i_1}{e-y_1}\right)^2 - \frac{\lambda_2}{m}\left(\frac{i_2}{e+y_1}\right)^2 + g - \ddot{y}_1^* \\ &= \frac{\lambda_1}{m}\left(\left(\frac{i_1}{e-y_1}\right)^2 - \left(\frac{i_1^{**}}{e-y_1}\right)^2\right) \\ &\qquad - \frac{\lambda_2}{m}\left(\left(\frac{i_2}{e+y_1}\right)^2 - \left(\frac{i_2^{**}}{e+y_1}\right)^2\right) \\ &\qquad\quad + \frac{\lambda_1}{m}\left(\frac{i_1^{**}}{e-y_1}\right)^2 - \frac{\lambda_2}{m}\left(\frac{i_2^{**}}{e+y_1}\right)^2 + g - \ddot{y}_1^* \,. \end{aligned} \tag{12.38}$$

According to (12.33), we have

$$\begin{aligned} \frac{\lambda_1}{m}\left(\frac{i_1^{**}}{e-y_1}\right)^2 - \frac{\lambda_2}{m}\left(\frac{i_2^{**}}{e+y_1}\right)^2 + g - \ddot{y}_1^* &= A(y_1, \dot{y}_1) + g - \ddot{y}_1^* \\ &= -k_0(y_1 - y_1^*) - k_1(\dot{y}_1 - \dot{y}_1^*) \end{aligned}$$

and

$$\begin{aligned} &\frac{\lambda_j}{m}\left(\left(\frac{i_j}{e+(-1)^j y_1}\right)^2 - \left(\frac{i_j^{**}}{e+(-1)^j y_1}\right)^2\right) \\ &\qquad = \frac{\lambda_j}{m}\left(\frac{i_j + i_j^{**}}{(e+(-1)^j y_1)^2}\right)(i_j - i_j^{**}), \quad j = 1, 2. \end{aligned}$$

Therefore we have

$$\begin{aligned} \ddot{y}_1 - \ddot{y}_1^* = -k_0(y_1 - y_1^*) &- k_1(\dot{y}_1 - \dot{y}_1^*) \\ &+ \frac{\lambda_1}{m}\left(\frac{i_1 + i_1^{**}}{(e-y_1)^2}\right)(i_1 - i_1^{**}) \\ &- \frac{\lambda_2}{m}\left(\frac{i_2 + i_2^{**}}{(e+y_1)^2}\right)(i_2 - i_2^{**}). \end{aligned} \tag{12.39}$$

On the other hand, let us combine (12.35) with (12.37). We get:

$$\frac{d}{dt}(i_j - i_j^{**}) = -\frac{\kappa}{\varepsilon}(i_j - i_j^{**}) - \frac{U_j^{**}}{L_j} \ , \quad j = 1, 2 \tag{12.40}$$

The error equations are thus given by (12.39) and (12.40), and we may recognize the standard form of singular perturbations (see Section 3.3 and, e.g., Kokotović et al. [1986]). Clearly, (12.40) is fast and stable by construction and the corresponding steady-states satisfy $i_j - i_j^{**} = \frac{\varepsilon}{\kappa}\frac{U_j^{**}}{L_j}$, $j = 1, 2$. Therefore, since $\frac{U_j^{**}}{L_j}$ is bounded, the fast transients of $i_j - i_j^{**}$ converge fast to a limit of order $O(\varepsilon)$ and, by Tikhonov's theorem Tikhonov et al. [1980] (see also Theorems 3.8 and 3.9 and Kokotović et al. [1986]), the right-hand side of the slow dynamics (12.39) may be approximated by replacing $i_j - i_j^{**}$ by its limit, which proves that it is stable and that $y_1(t) - y_1^*(t)$ exponentially converges to a limit of order $O(\varepsilon)$ as $t \to +\infty$, which achieves to prove the theorem.

Remark 12.5. The above result would not hold true anymore if the references U_j^{**} were not bounded. Note that the boundedness of the voltage references for every current reference is a consequence of our regularization procedure and therefore of the corresponding biasing.

Remark 12.6. The singularity problem at switchings for the trajectory tracking by voltage control is here circumvented since we do not use the 3rd order derivative of $y_1 - y_1^*$ to stabilize it.

Remark 12.7. This cascaded control technique replaces the stabilization problem of (12.21), made up with a third order subsystem and a first order one, by the stabilization of a second order subsystem and two first order ones. It can be interpreted as a way to "shorten the integrators" (see Krstić et al. [1995] for a backstepping approach to this problem)

Remark 12.8. Lower bounds for the gains K_j can be easily computed by assuming that we restrict the velocity $\dot{y}_1$ to a bounded domain and that a given tracking precision on the currents and positions is desired. Note that, in practice, the gains may not be chosen too large to avoid an artificial amplification of external disturbances.

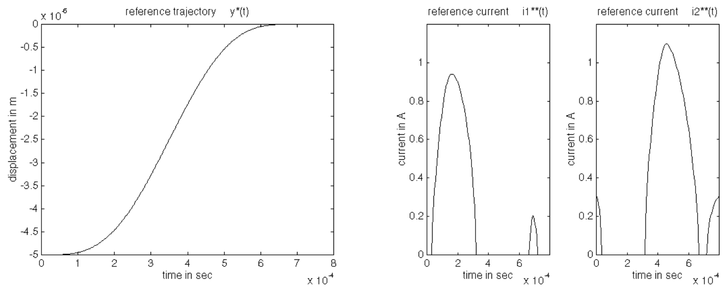

Fig. 12.5 Hierarchical control for the ball: reference position and control inputs

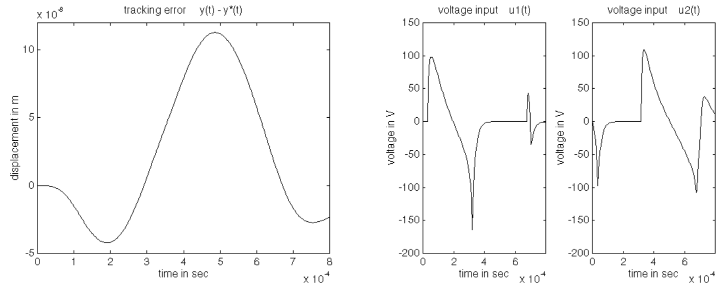

Fig. 12.6 Hierarchical control: tracking error and voltage inputs

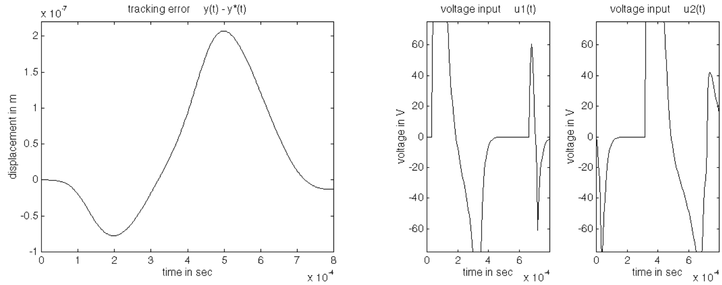

Fig. 12.7 Hierarchical control: tracking error and saturated voltage inputs

The figures displayed in figure 12.5 show the evolution of $y^*(t)$ and the corresponding reference currents i_1^{**} and i_2^{**}, the high-level loop inputs. Figure 12.6 gives the tracking error and the actual voltage inputs. Because of the excessive amplitude of low-level control inputs, another test has been carried out with a saturation on the low-level voltage inputs. The corresponding curves are given in figure 12.7. The tracking error is not too much affected by this saturation.

12.1.4.4 Some Comparisons

To conclude this section, let us briefly sketch a comparison between this approach and a more classical one based on tangent linearization techniques with biasing currents. To make the comparison more interesting, we assume that there is no gravity, which means that we set $g = 0$ in what precedes, so that the zero force point is an equilibrium point. We assume that the mass of the ball is $m = 0.2$ kg, the air gap is $e = 0.5$ mm. The coils are identical and we set $\lambda_1 = \lambda_2 = \lambda = 5.10^{-6}$ Nm^2/A^2, and $R_1 = R_2 = R = 1\ \Omega$. Moreover, the currents are saturated at 3 A and the voltages at 20 V. The initial position of the ball is $y(0) = -0.4$ mm with null currents $i_1(0) = i_2(0) = 0$ and we want to reach the final position $y(T) = 0$, at $T = 10$ ms, with null currents

$i_1(T) = i_2(T) = 0$. After $t = 13$ ms and during 4 ms, the ball is submitted to a perturbative force, which is triangular with maximal amplitude equal to 140 N (recall that the mass of the ball is 0.2 kg).

We compare four approaches:

- the nonlinear design based upon the current complementarity condition and the voltage linearizing feedback (recall that the voltages and currents are saturated);
- the nonlinear cascaded control method, using the almost complementarity condition, whose maximal biasing level corresponds to 0.035 A, with $\eta = 1$ m/s^2;
- a linear approach, based on the tangent linearization at the equilibrium point $y = 0$, with the reference trajectory computed as in the previous case and with two levels of biasing currents: $I = 0.035$ A and $I = 1.5$ A $(= \frac{i_{max}}{2})$;
- a linear approach, also based on the tangent linearization at the equilibrium point $y = 0$, but with a reference trajectory replaced by the set point $y = 0$. The two above levels of biasing are considered.

In the four cases, the poles of the linear dynamics at $y = 0$ are all equal to -4000.

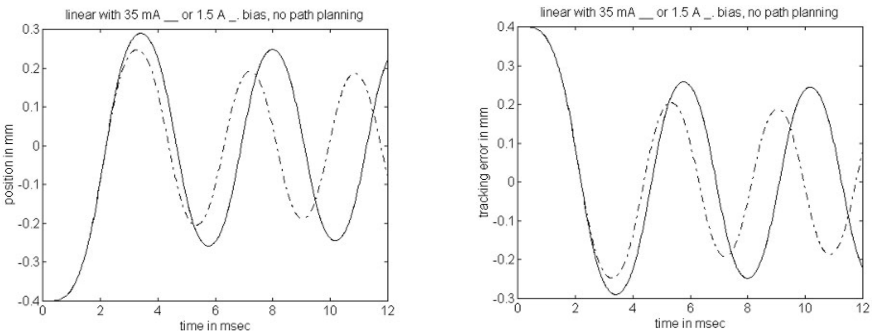

Fig. 12.8 Linear control without path planning: position of the ball (left) and tracking error (right).

The simulation results are displayed in figures 12.9, 12.10 and 12.8. The differences between the methods based upon the complementary condition and the almost one are so tiny that we have only shown the last one (figure 12.9). We may first observe that motion planning is useful since the linear approach without trajectory planning, and for a tracking dynamics comparable to the nonlinear one, does not yield a satisfactorily convergent behavior, whatever the biasing current level (figure 12.8). With motion planning, the linear tracking is quite good with a small biasing current, while the sensitivity

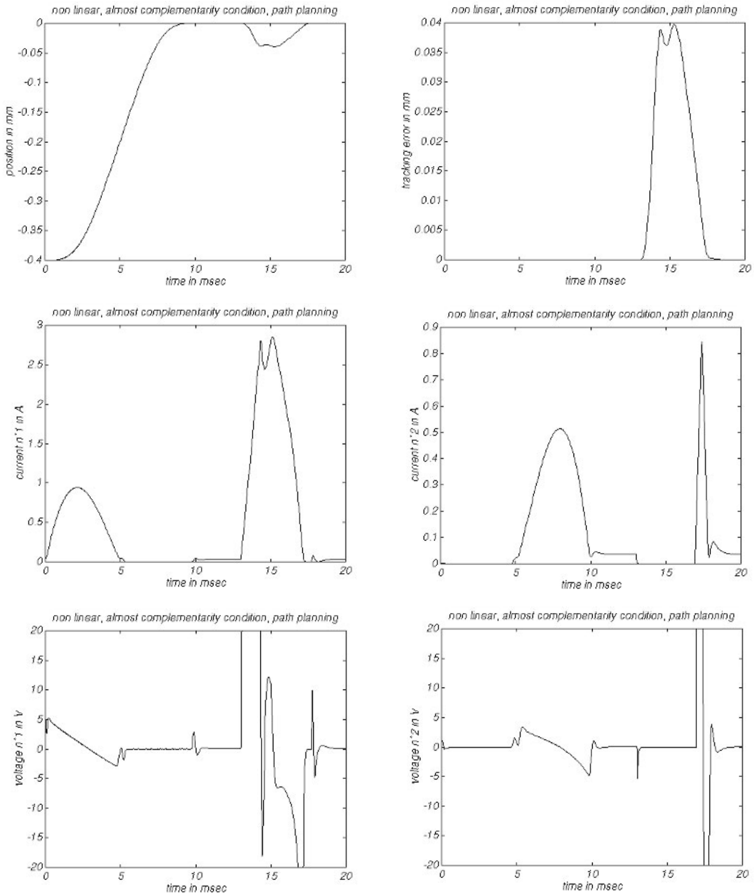

Fig. 12.9 The nonlinear cascaded control: position of the ball (upleft), error between the ball position and its reference (upright), currents 1 and 2 (middle) and voltages 1 and 2 (below).

to perturbations remains too high (see the continuous curves of figure 12.10 after 13 ms). If we increase the bias, the robustness to perturbations is improved but the system's answers to large tracking errors are slightly deteriorated and produce saturating currents and voltages (see the dotted curves of figure 12.10 between 0 and 5 ms). Finally, both nonlinear approaches ensure a good tracking and an acceptable robustness to perturbations (see the responses to the force disturbance on figure 12.9). It is interesting to note that the nonlinear controller requires larger currents and voltages than the linear controller with large bias (figure 12.10, dotted curves), to attenuate the perturbation. This is caused by the fact that, at low currents, a large current variation is required to achieve a force variation of 1 unit, whereas with a large bias, a smaller current variation achieves the same force variation.

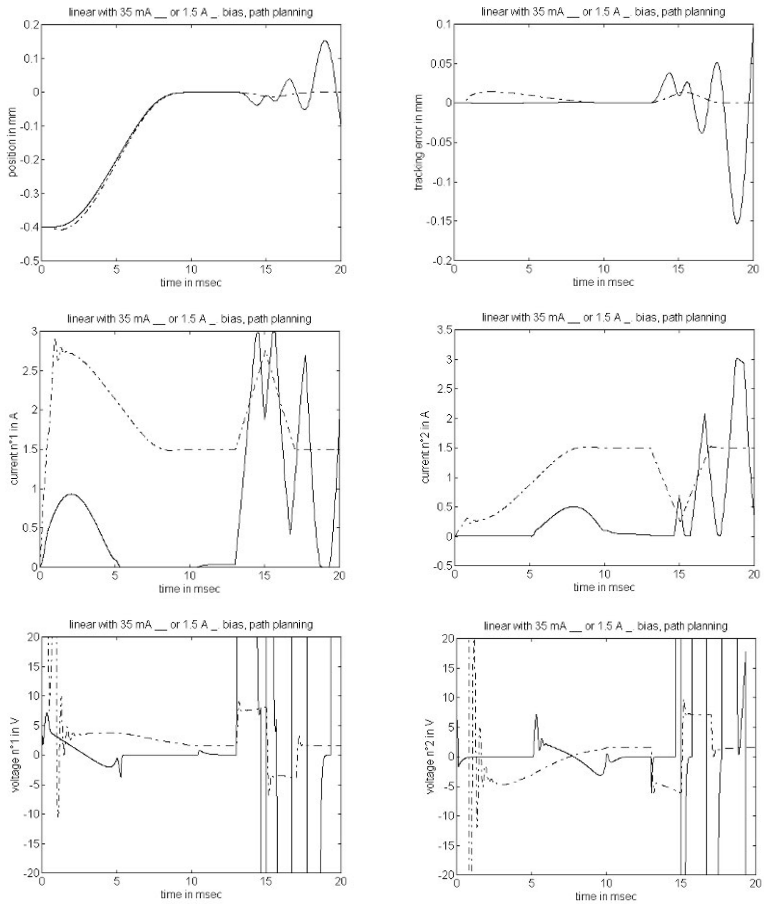

Fig. 12.10 Linear control with path planning: position of the ball (upleft), error between the ball position and its reference (upright), currents 1 and 2 (middle) and voltages 1 and 2 (below).

Indeed, the same considerations extend to the more general 4 d.o.f. problem of the next section.

12.2 The General Shaft

We now turn to the case of a rotating shaft with two active magnetic bearings. The bearings control four degrees of freedom and a passive thrust stabilizes the projection along the main axis of the movements of the shaft. The rotor is assumed to be a rigid body. We consider the moving frame attached to the center of mass G in the fixed coordinates of the stator. We denote by θ and

ψ the angles corresponding to a small rotation around the axis Gy and Gz respectively. The lengths and forces are as displayed in Figure 12.11.

Due to the presence of a passive thrust, the dynamics of the shaft along the Gx axis are neglected. The following model is obtained by expressing the balance of forces and torques at the center of mass G. The action of the passive thrust in the directions of Gy and Gz is supposed to be equivalent to the one of a spring with stiffness k_b in both directions. The corresponding radial forces are denoted by F_{b_1} and F_{b_2}.

12.2.1 Modelling

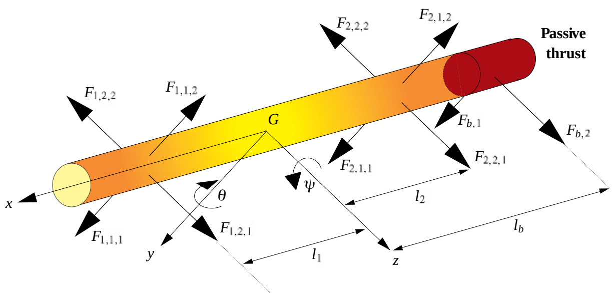

Fig. 12.11 The shaft and the four pairs of electromagnets.

In the xGy plane:

$$\begin{aligned} m\ddot{y} &= F_{1,1,1} - F_{1,1,2} + F_{2,1,1} - F_{2,1,2} + F_{b_1} + mg_1 \\ J\ddot{\psi} &= (F_{1,1,1} - F_{1,1,2})l_1 - (F_{2,1,1} - F_{2,1,2})l_2 - F_{b_1} l_b + \omega J_x \dot{\theta} \end{aligned} \tag{12.41}$$

with $F_{b_1} = -k_b(y - l_b \sin\psi)$, and in the xGz plane:

$$\begin{aligned} m\ddot{z} &= F_{1,2,1} - F_{1,2,2} + F_{2,2,1} - F_{2,2,2} + F_{b_2} + mg_2 \\ J\ddot{\theta} &= -(F_{1,2,1} - F_{1,2,2})l_1 + (F_{2,2,1} - F_{2,2,2})l_2 + F_{b_2} l_b - \omega J_x \dot{\psi} \end{aligned} \tag{12.42}$$

where $F_{b_2} = -k_b(z + l_b \sin\theta)$, where ω is the angular velocity around the Gx axis and g_1 and g_2 are the projections of the gravitational field along Gy and Gz respectively. According to the geometry of the shaft, we assume that the radial inertia moments are equal: $J_y = J_z = J$.

Let us note :

$$x_1 = y \ , \quad x_2 = z \ , \quad x_3 = \psi \ , \quad x_4 = \theta \ . \tag{12.43}$$

The electromagnetic forces can be expressed as follows:

$$F_{j,k,l} = \frac{\lambda_{j,k,l}(i_{j,k,l})^2}{(e_j + (-1)^l(x_k + (-1)^{j+k} l_j \sin(x_{k+2})))^2} \ , \quad j,k,l = 1,2 \ , \quad (12.44)$$

where $\lambda_{j,k,l}$ and e_j are the characteristic constants of the bearing number j, $j,k,l = 1,2$.

Note that the 8 forces are, as before, functions of the 8 corresponding coil currents which may be, in a first approach, considered as the 8 input variables. In a second stage, we will adjoin the electrical equations relating the currents $i_{j,k,l}$ to the voltages $U_{j,k,l}$, $j,k,l = 1,2$, given by:

$$U_{j,k,l} = R_{j,k,l} i_{j,k,l} + L_{j,k,l} \frac{di_{j,k,l}}{dt} + i_{j,k,l} \frac{dL_{j,k,l}}{dt} \ , \quad j,k,l = 1,2 \qquad (12.45)$$

with

$$L_{j,k,l} = \frac{\lambda_{j,k,l}}{e_j + (-1)^l(x_k + (-1)^{j+k} l_j \sin(x_{k+2}))} \ , \quad j,k,l = 1,2 \ . \qquad (12.46)$$

In this latter case, the 8 voltages are to be considered as control inputs.

Most of the ideas developed in the case of the ball easily extend to the shaft. Indeed, the difficulties also extend. Since we have shown that a direct approach of the voltage control problem is not completely satisfactory, we only insist now on the hierarchical control problem. The current control problem being a necessary preparation to it, we will just sketch the new aspects. In the same spirit, the current complementarity condition could be used in this case, though we only present the results related to the current almost complementarity condition, needed for the hierarchical control solution.

12.2.2 Current Control

12.2.2.1 Flatness

A possible choice of flat output is

$$\begin{aligned} y_1 &= y \ \ y_5 = \frac{i_{1,1,2}\sqrt{\lambda_{1,1,2}}}{e_1 + y + l_1 \sin\psi} \\ y_2 &= z \ \ y_6 = \frac{i_{1,2,2}\sqrt{\lambda_{1,2,2}}}{e_1 + z - l_1 \sin\theta} \\ y_3 &= \psi \ \ y_7 = \frac{i_{2,1,2}\sqrt{\lambda_{2,1,2}}}{e_2 + y - l_2 \sin\psi} \\ y_4 &= \theta \ \ y_8 = \frac{i_{2,2,2}\sqrt{\lambda_{2,2,2}}}{e_2 + z + l_2 \sin\theta} \ . \end{aligned} \qquad (12.47)$$

To verify the flatness property, it suffices to express all the state and input variables as functions of the y_i's and a finite number of their derivatives:

$$\begin{aligned}
x_j &= y_j \ , \quad j = 1, \ldots, 4 \\
\dot{x}_j &= \dot{y}_j \ , \quad j = 1, \ldots, 4 \\
i_{j,k,2} &= \frac{1}{\sqrt{\lambda_{j,k,2}}} y_{2j+k+2}(e_j + y_k + (-1)^{j+k} l_j \sin(y_{k+2})) \\
i_{j,k,1}^2 &= \frac{(e_j - y_k + (-1)^{j+k} l_j \sin y_{k+2})^2}{\lambda_{j,k,1}} \big\{ y_{2j+k+2}^2 \\
&\quad + \frac{1}{(l_1 + l_2)} \big[m l_{3-j}(\ddot{y}_k - g_k) + (-1)^{j+k} J \ddot{y}_{k+2} \\
&\quad + k_b(l_{3-j} + (-1)^j l_b)(y_k + (-1)^k l_b \sin y_{k+2}) \\
&\quad + (-1)^j \omega J_x \dot{y}_{5-k} \big] \big\}
\end{aligned} \tag{12.48}$$

12.2.2.2 Path Planning With Almost Current Complementarity

As before, we choose twice continuously differentiable functions of time $y_1^*, \ldots, y_4^*$ that play the role of the references for y, z, ψ, θ. To construct the currents in the 4 pairs of electromagnets, we use the current almost complementarity condition for each pair of electromagnets, and the associated almost complementarity function Y_η, which lead to the formulas, expressed for a pair (j,k), $j,k = 1,2$, of electromagnets:

$$\begin{aligned}
i_{j,k,1} &= \frac{(e_j - y_k + (-1)^{j+k} l_j \sin y_{k+2})}{\sqrt{\lambda_{j,k,1}}} \sqrt{B_{j,k} + Y_\eta^2(-B_{j,k})} \\
i_{j,k,2} &= \frac{(e_j + y_k + (-1)^{j+k} l_j \sin y_{k+2})}{\sqrt{\lambda_{j,k,2}}} Y_\eta(-B_{j,k})
\end{aligned} \tag{12.49}$$

where we have noted, for $j,k = 1,2$,

$$\begin{aligned}
B_{j,k} = \frac{1}{(l_1 + l_2)} &(m l_{3-j}(\ddot{y}_k^* - g_k) + (-1)^{j+k} J \ddot{y}_{k+2}^* \\
&+ k_b(l_{3-j} + (-1)^j l_b)(y_k^* + (-1)^k l_b \sin y_{k+2}^*) + (-1)^j \omega J_x \dot{y}_{5-k}^*)).
\end{aligned} \tag{12.50}$$

12.2.2.3 Trajectory Tracking

The feedback construction follows the same lines as in theorem 12.1 and consists in linearizing the system by feedback. The computations are mostly uninteresting and do not bring new insight. They are omitted. The feedback form obtained for the currents is the same as in (12.49), the only difference being that the function $B_{j,k}$ must be replaced by $C_{j,k}$:

$$C_{j,k} = \frac{1}{(l_1+l_2)}(ml_{3-j}(A_k - g_k) + (-1)^{j+k} J A_{k+2}$$
$$k_b(l_{3-j} + (-1)^j l_b)(y_k + (-1)^k l_b \sin y_{k+2}) + (-1)^j \omega J_x \dot{y}_{5-k})) \tag{12.51}$$

with

$$A_j = -k_{0,j}(y_j - y_j^*) - k_{1,j}(\dot{y}_j - \dot{y}_j^*) + \ddot{y}_j^* \ , j = 1, \ldots, 4 \ . \tag{12.52}$$

The same result as in theorem 12.1 can be stated for the above designed feedback: let us introduce the tracking behavior for the 4 first components of the output:

$$\ddot{y}_j - \ddot{y}_j^* = -k_{0,j}(y_j - y_j^*) - k_{1,j}(\dot{y}_j - \dot{y}_j^*) \ , \quad j = 1, \ldots, 4 \tag{12.53}$$

and assume that the polynomials $P_j(s) = s^2 + k_{1,j}s + k_{0,j}$ are Hurwitz.

Theorem 12.3. *The control law*

$$\begin{aligned} i_{j,k,1} &= \frac{(e_j - y_k + (-1)^{j+k} l_j \sin y_{k+2})}{\sqrt{\lambda_{j,k,1}}} \sqrt{C_{j,k} + Y_\eta^2(-C_{j,k})} \\ i_{j,k,2} &= \frac{(e_j + y_k + (-1)^{j+k} l_j \sin y_{k+2})}{\sqrt{\lambda_{j,k,2}}} Y_\eta(-C_{j,k}) \end{aligned} \tag{12.54}$$

for $j, k = 1, 2$*, is such that:*

(i) the current almost complementarity condition is satisfied for each pair of electromagnets,
(ii) the tracking errors $y_j - y_j^*$*,* $j = 1, \ldots, 4$*, are exponentially stable and satisfy (12.53),*
(iii) the tracking errors $y_j - y_j^*$*,* $j = 5, \ldots, 8$*, tend to 0 as* t *tends to* ∞*.*

The proof follows the same lines as the one of theorem 12.1 and is omitted.

12.2.3 Hierarchical Control

Once the high-level loop designed as in theorem 12.3, the low-level loop is exactly the same as in the case of the ball since the low-level (voltage) stabilization is realized by each electromagnet independently. Therefore, the stabilization result of theorem 12.2 holds true here, the only difference being that the high-level loop is designed as above.

12.3 Implementation

This section is devoted to the implementation aspects of the above hierarchical control design in the case of the shaft. Two important problems remain to be addressed. At first, the knowledge of the position and velocity of the shaft as well as the currents in the electromagnets are necessary to compute the feedback: since the position and currents only are measured, an observer is needed to reconstruct the velocity from the measurements. Secondly, the digital implementation requires a discretized version of the above control laws.

12.3.1 Observer Design

The informations given by the sensors concern the position of the rotor relatively to the stator. We call ζ_j, $j = 1, \ldots, 4$, the corresponding measurements, related to the state variables of the shaft as follows:

$$\begin{array}{ll} \zeta_1 = x_1 + l_{c_1} \sin x_3 & \zeta_2 = x_1 - l_{c_2} \sin x_3 \\ \zeta_3 = x_2 - l_{c_1} \sin x_4 & \zeta_4 = x_2 + l_{c_2} \sin x_4 \end{array} \tag{12.55}$$

where l_{c_j}, $j = 1, 2$, is the distance between the jth plane of the two corresponding sensors and the center of mass.

To the list of measured variables, we may add the currents $i_{j,k,l}$, $j, k, l = 1, 2$. We denote by $\zeta = (\zeta_1, \ldots, \zeta_4)$ and i the vector made up with the 8 currents.

Clearly, we can directly recover the positions x_j, $j = 1, \ldots, 4$ from (12.55):

$$\begin{array}{ll} x_1 = \dfrac{l_{c_1}\zeta_2 + l_{c_2}\zeta_1}{l_{c_1} + l_{c_2}} & x_2 = \dfrac{l_{c_1}\zeta_4 + l_{c_2}\zeta_3}{l_{c_1} + l_{c_2}} \\ x_3 = \arcsin(\dfrac{\zeta_1 - \zeta_2}{l_{c_1} + l_{c_2}}) & x_4 = \arcsin(\dfrac{\zeta_4 - \zeta_3}{l_{c_1} + l_{c_2}}) \end{array} \tag{12.56}$$

The aim of this section is to reconstruct, by means of an observer, the velocities $\dot{x}_j$, $j = 1, \ldots, 4$. In fact, with the same approach, we can also reconstruct the biases of the forces and torques that result from modelling errors. Note that, since the shaft rotates around its main axis at an angular speed ω, we aim at designing the observer such that its behavior is independent of ω. This additional requirement is mainly motivated by the avoidance of critical speeds.

Let us denote $\beta = (\beta_1, \ldots, \beta_4)$ the force and torque bias, supposed constant or slowly varying, and $F_j(\zeta, i)$ the force defined by

$$\begin{array}{l} F_j(\zeta, i) = F_{1,j,1} - F_{1,j,2} + F_{2,j,1} - F_{2,j,2} + F_{b_j} + mg_j \text{ , for } j = 1, 2 \\ F_j(\zeta, i) = (-1)^{j-1}(F_{1,j-2,1} - F_{1,j-2,2})l_1 \\ \qquad + (-1)^j(F_{2,j-2,1} - F_{2,j-2,2})l_2 + (-1)^j F_{b_{j-2}} l_b \text{ , for } j = 3, 4 \ . \end{array} \tag{12.57}$$

The model (12.41), (12.42) may be rewritten as follows:

$$\begin{array}{ll} m\ddot{x}_1 = F_1(\zeta,i) + \beta_1 & \dot{\beta}_1 = 0 \\ m\ddot{x}_2 = F_2(\zeta,i) + \beta_2 & \dot{\beta}_2 = 0 \\ J\ddot{x}_3 = -\omega J_x \dot{x}_4 + F_3(\zeta,i) + \beta_3 & \dot{\beta}_3 = 0 \\ J\ddot{x}_4 = \omega J_x \dot{x}_3 + F_4(\zeta,i) + \beta_4 & \dot{\beta}_4 = 0 \end{array} \tag{12.58}$$

Note that these forces (12.57) are functions of the available measurements only and do not contain the variables we want to estimate. The system (12.58) belongs to the class of linear systems up to output injection (see e.g. Isidori [1995], Krener and Isidori [1983], Krener and Respondek [1985], Bestle and Zeitz [1981], Xia and Gao [1989]). Moreover, if we use the inverted relations (12.56) to observe directly x_j, $j = 1, \dots, 4$, we obtain 3 independent subsystems, two subsystems of dimension 2 for the translations along y and z and one subsystem of dimension 4 for the rotations:

- for the translations along y and z:

$$\begin{cases} m\ddot{x}_1 = \beta_1 + F_1(\zeta,i) & \dot{\beta}_1 = 0 \\ x_1 = \dfrac{l_{c_1}\zeta_2 + l_{c_2}\zeta_1}{l_{c_1} + l_{c_2}} \end{cases} \tag{12.59}$$

  and

$$\begin{cases} m\ddot{x}_2 = \beta_2 + F_2(\zeta,i) & \dot{\beta}_2 = 0 \\ x_2 = \dfrac{l_{c_1}\zeta_4 + l_{c_2}\zeta_3}{l_{c_1} + l_{c_2}} \end{cases} \tag{12.60}$$

- for the rotations:

$$\begin{cases} J\ddot{x}_3 = -\omega J_x \dot{x}_4 + \beta_3 + F_3(\zeta,i) & \dot{\beta}_3 = 0 \\ J\ddot{x}_4 = \omega J_x \dot{x}_3 + \beta_4 + F_4(\zeta,i) & \dot{\beta}_4 = 0 \\ x_3 = \arcsin(\dfrac{\zeta_1 - \zeta_2}{l_{c_1} + l_{c_2}}) & x_4 = \arcsin(\dfrac{\zeta_4 - \zeta_3}{l_{c_1} + l_{c_2}}) \ . \end{cases} \tag{12.61}$$

The construction of an observer for the velocities reduces thus to the construction of three independent observers for linear systems with output injection of the form

$$\begin{cases} \dot{\mu}_j = A_j \mu_j + b_j(\nu_j) \\ \nu_j = C_j \mu_j \ . \end{cases} \quad , \quad j = 1,2,3 \tag{12.62}$$

with the pairs (A_j, C_j) observable for $j = 1, 2, 3$. For the translation subsystems, we have

$$A_1 = A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & \frac{1}{m} \\ 0 & 0 & 0 \end{pmatrix} \quad , \quad C_1 = C_2 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

and for the rotation subsystem:

$$A_3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{J} & 0 & -\omega\frac{J_x}{J} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & \omega\frac{J_x}{J} & 0 & 0 & 0 & \frac{1}{J} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} , \quad C_3 = \begin{pmatrix} 1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0 \end{pmatrix} .$$

Since these pairs are observable, there exist gain matrices K_j, $j = 1, 2, 3$,

$$K_j = \begin{pmatrix} \gamma_{1,j} \\ \gamma_{2,j} \\ \gamma_{3,j} \end{pmatrix} , \quad j = 1, 2$$

and K_3 a matrix 6×2 with entries noted $\gamma_{i,j,3}$, $i = 1, \ldots, 6$, $j = 1, 2$, such that $A_j + K_j C_j$ is stable, $j = 1, 2, 3$. We thus construct observers of the form

$$\dot{\hat{\mu}}_j = A_j \hat{\mu}_j + b_j(\nu_j) - K_j(\nu_j - C_j \hat{\mu}_j) \tag{12.63}$$

where $\hat{\mu}_j$ is an estimate of μ_j. The estimation error $e_j = \mu_j - \hat{\mu}_j$ satisfies:

$$\dot{e}_j = A_j e_j + K_j C_j e_j = (A_j + K_j C_j) e_j \ , \quad j = 1, 2, 3, \tag{12.64}$$

which proves the stability of the error equations or, in other words, e_j tends to 0 exponentially fast, the exponential rate of decay depending on the choice of the eigenvalues of $A_j + K_j C_j$ which may be placed arbitrarily by means of the entries of K_j, $j = 1, 2, 3$.

It results that, if the measurements are accurate enough, the above observers are able to produce convergent estimates of the velocities and of the force and torque biases. Indeed, since the observer is used to synthesize the feedback, its convergence rate must be chosen significantly higher than the one of the controller.

Note in addition that A_3 depends on the angular velocity ω, which is measured, and that a proper choice of the entries of K_3 may be done such that the eigenvalues of $A_3 + K_3 C_3$ are independent of ω. Assume for example that the eigenvalues of $A_3 + K_3 C_3$ are solution of the Hurwitz polynomial

$$(s^3 + a_{1,3}s^2 + a_{1,2}s + a_{1,1})(s^3 + a_{2,3}s^2 + a_{2,2}s + a_{2,1}) = 0$$

where the coefficients $a_{j,k}$ are conveniently chosen, independent of ω. A possible choice of K_3 is given in this case by:

$$K_3 = \begin{pmatrix} a_{1,3} & \omega \\ a_{1,2} - \omega^2 & a_{2,3}\omega \\ a_{1,1} & 0 \\ -\omega & a_{2,3} \\ -a_{1,3}\omega & a_{2,2} - \omega^2 \\ 0 & a_{2,1} \end{pmatrix} .$$

12.3.2 Digital Control

According to the fact that the model of the shaft is flat, or more precisely equivalent to a linear system, we can take profit of the linearity to obtain a discrete-time model by exact discretization. The original variables are then obtained by the inverse transformations. This technique is indeed exact only if constant controls are used, which is not the case here. However, when the discretization period is small enough, the errors induced by the discretization scheme can be seen as perturbations of the system, which are naturally dampen by the stability of the closed-loop system.

Since the discretization aspects are not crucial in this implementation, and since the formulas have little interest in themselves, they are withdrawn.

12.4 Experimental Results

12.4.1 Platform Description

The experimental bench is composed by a magnetic suspension unit, a personal computer with a control interface, two transputers and a power supply.

The magnetic suspension is made of a rotor and a stator. The rotor turns around its main axis by means of an electric motor and its position with respect to the stator is controlled by four magnetic bearings. Four electromagnetic sensors give a differential position measurement from which the displacement is computed by a devoted circuit.

Parameters	Shaft	Units
l_1	0.0504	m
l_2	0.0606	m
l_b	0.1000	m
l_{c_1}	0.088	m
l_{c_2}	0.091	m
m	1.165	kg
J_x	0.149E-03	kg.m^2
J_y	4.72E-03	kg.m^2
λ_l	2.3E-06	N.m^2/A^2
λ_r	2.0E-06	N.m^2/A^2
k_r	-4000	N/m

Table 12.1 Characteristics of the Shaft

The main characteristics of the shaft are displayed in the table 12.1. The notations are those of the previous sections.

Concerning the computing units, two transputers are used, one for the analog/digital converters and the other for the numerical treatment. The numerical program is presently written in C, without optimization with respect to real time aspects. In particular, a lower level programming language might increase the computational speed in real time.

12.4.2 Experiments

Two types of tests have been done on the experimental bench. They concern the transitions to steady-states when the stator is in horizontal or vertical position. In the first case, we start from the landing position (all the bearings switched off) to a centered equilibrium position. In the second (vertical) case, we move around the vertical centered position corresponding to currents in the electromagnets remaining always close to 0.

12.4.2.1 The Horizontal Case

The shaft moves up in a plane to reach its centered position. The reference trajectories correspond to the formulas given in the motion planning section with a 5th degree polynomial for y and ψ.

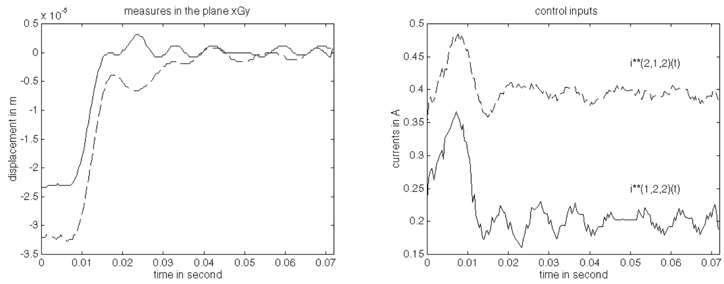

Fig. 12.12 Horizontal shaft : measured relative positions of the rotor w.r.t. the stator and currents

The curves displayed in figure 12.12, 12.13 show the evolution of y, ψ and the currents i^{**}, the high-level loop inputs. Note that though the dynamical change due to the take off of the shaft is not taken into account in our model, the behavior at starting is not significantly affected.

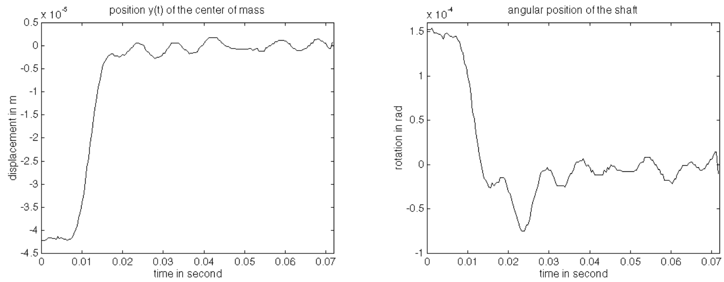

Fig. 12.13 Horizontal shaft: position of the center of mass and angular position of the shaft

12.4.2.2 The Vertical Case

The same kind of tests have been done when the stator is in vertical position. The model in this case is modified to properly take into account the gravity but the control synthesis remain unchanged. The interest of this position relies on the fact that the coil currents vanish at the steady-state. As in the previous case, the control references are computed from the references of y and ψ which are chosen as 5th degree polynomials. The curves presented in figures 12.14 to 12.16 are those of the evolution of y, ψ and the corresponding input currents i^{**}.

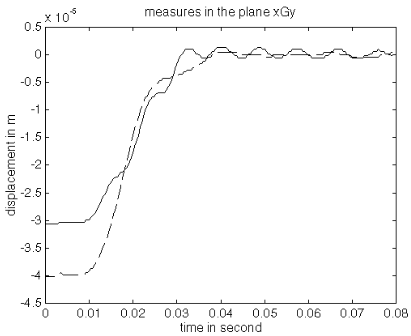

Fig. 12.14 Vertical shaft: measured relative positions of the rotor w.r.t. the stator

Note that in both the vertical and the horizontal cases, residual oscillations, of amplitude less than 5.10^{-6}m, are present. In the vertical case, they produce numerous switchings since the currents remain close to zero. They apparently result from external vibrations transmitted by the stator, which

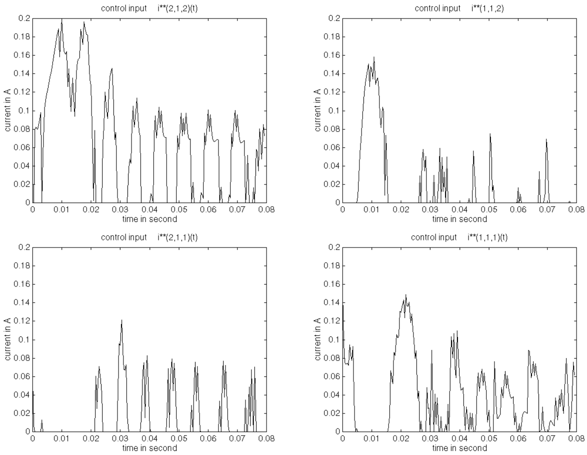

Fig. 12.15 Vertical shaft: measured current inputs

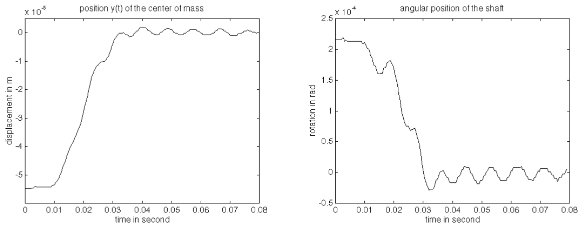

Fig. 12.16 Vertical shaft: position of the center of mass and angular position of the shaft

is not vibration isolated, and perturb both the system dynamics and the measurements. They do not concern a presumably oscillatory behavior of the controller.

12.4.2.3 A Rough Experimental Comparison

We now compare the above nonlinear design in the vertical case to a classical linear one, using a biasing current equal to 1.5 A. For technological reasons, the current path planning could not be implemented with the linear controller. Therefore, and, contrarily to the simulated comparisons of section 12.1.4.4, the linear tracking dynamics around the equilibrium point has been significantly slowed down to avoid oscillations. The experimental results are shown in figure 12.17. The linear controller achieves the transfer from $y = -0.4$ mm to $y = 0$ in approximatively 1400 ms with a peak current of about 2A. The nonlinear tracking achieves the same transfer in 30 ms (more than 45 times faster!) with peak currents smaller than 1A. The current variations all along the transfer are comparable in both cases. Clearly, the dynamical response of the linear controller might be improved by using motion planning. Further experiments are required to better evaluate the nonlinear feedback performances.

Note that we have more insisted on the starting aspects and that little have been said about the rotating performances.The present nonlinear controller has been satisfactorily tested up to frequencies of 600 Hz with a positioning precision of about 2μm.

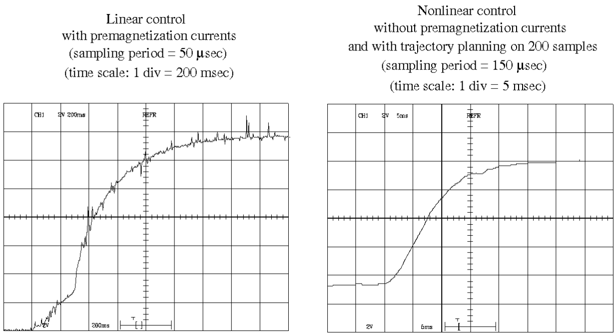

Fig. 12.17 An experimental comparison of positioning of the vertical shaft: linear control using bias currents without path planning (left) and nonlinear cascaded control (right).

Chapter 13
Crane Control

The present chapter constitutes a continuation of the introductory example of section 5.1, chapter 5, and of example 7.2, chapter 7. We show how simple measurement feedbacks may be designed to track different types of rest-to-rest trajectories and compare them.

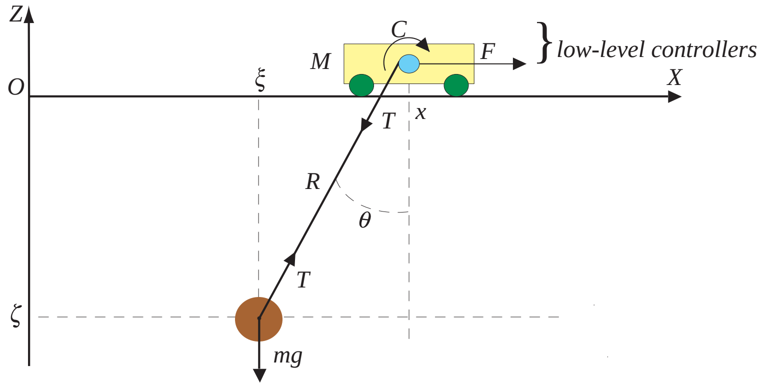

Fig. 13.1 Two-dimensional overhead crane.

13.1 Orientation

Our aim is to control an overhead crane that carries a load from a given configuration to another one, without swing at the end point. We are only interested here in *rest-to-rest displacements*, namely starting from a given point at rest and arriving to another given point at rest, with a duration as

J. Lévine, *Analysis and Control of Nonlinear Systems*, Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_13, © Springer-Verlag Berlin Heidelberg 2009

short as possible. We therefore assume that the motors are powerful enough to deliver the required force and torque. We also assume that there are technological limitations on the motors and therefore that the force and torque are subject to constraints: the absolute value of the force and torque are limited to 100000N and 10000kgm^2 respectively. These constraints will not be taken into account explicitly in the control design but they are present in the simulation software where the force and the winch torque are replaced by their saturated value each time they exceed it.

One of the distinctive features of this application is that the position and velocity of the load, as well as the angle of the cable with respect to the vertical and its velocity are not measured. Only the position and velocity of the cart on the one hand and the angular position and velocity of the winch (which is equivalent to measuring the length of the cable and its velocity) on the other hand are measured. Implementing a full state feedback is therefore impossible. More precisely, a feedback linearizing design, as in (5.11), where the load is virtually controlled by a linear controller:

$$\xi^{(4)} = v_1, \quad \zeta^{(4)} = v_2$$

(see the corresponding full state dynamic feedback design of Andréa-Novel and Lévine [1990]), requires measurements of $(x, \dot{x}, R, \dot{R}, \theta, \dot{\theta})$ in real time, that are not available.

This is why we consider here a much simpler design that doesn't require measurements of θ and $\dot{\theta}$. The presented material is a particular case of the results obtained for a class of weight-handling systems in Kiss [2001], Kiss et al. [2000a,c]. The interested reader may also refer to Delaleau and Rudolph [1998] for a different construction using the notion of quasi-static feedback.

Recall from section 5.1 that the mass of the cart is denoted by M, its position by x, and the force applied to it by F. The winch has radius ρ, and is submitted to the torque C. The length of the cable is R, the angle of the cable with the vertical is θ and its tension is T. The load's mass is m and its position in the fixed frame (O, OX, OZ) is denoted by (ξ, ζ). We indeed assume that $R \leq R_0$, where R_0 corresponds to the height of the cart with respect to the ground, and that $T \geq 0$, which means that the cable cannot push the load.

Recall that the dynamical equations are given by:

$$\begin{aligned} m\ddot{\xi} &= -T\sin\theta \\ m\ddot{\zeta} &= T\cos\theta - mg \\ M\ddot{x} &= -\gamma_1(\dot{x}) + F + T\sin\theta \\ \frac{J}{\rho}\ddot{R} &= -\gamma_2(\dot{R}) - C + T\rho \end{aligned} \tag{13.1}$$

and that the geometric constraints are:

$$\begin{aligned}\xi &= x + R\sin\theta \\ \zeta &= -R\cos\theta.\end{aligned} \tag{13.2}$$

To simulate the system, the following explicit form (see again section 5.1) will be needed:

$$\begin{pmatrix} (\frac{M}{m}+\sin^2\theta) & \sin\theta & 0 \\ \sin\theta & (\frac{J}{m\rho^2}+1) & 0 \\ \cos\theta & 0 & R \end{pmatrix} \begin{pmatrix} \ddot{x} \\ \ddot{R} \\ \ddot{\theta} \end{pmatrix} = \begin{pmatrix} (R\dot{\theta}^2 + g\cos\theta)\sin\theta - \frac{1}{m}\gamma_1(\dot{x}) + \frac{F}{m} \\ R\dot{\theta}^2 + g\cos\theta - \frac{1}{m\rho}\gamma_2(\dot{R}) - \frac{C}{m\rho} \\ -2\dot{R}\dot{\theta} - g\sin\theta \end{pmatrix}. \tag{13.3}$$

We set

$$J_1 = \frac{J}{m\rho^2}, \quad J_2 = J_1 + 1, \quad M_1 = \frac{M}{m},$$

$$\mu(\theta) = J_2M_1 + J_1\sin^2\theta, \quad \alpha(R,\theta,\dot{\theta}) = R\dot{\theta}^2 + g\cos\theta,$$

and assume the frictions linear with respect to the corresponding velocities:

$$\gamma_1(\dot{x}) = \Gamma_1\dot{x}, \quad \gamma_2(\dot{R}) = \Gamma_2\dot{R}.$$

After inversion of the left-hand side matrix of (13.3), we get:

$$\begin{aligned}
\ddot{x} &= \frac{1}{\mu(\theta)}\left[J_1\alpha(R,\theta,\dot{\theta})\sin\theta + \frac{\sin\theta}{m\rho}\left(C + \Gamma_2\dot{R}\right) + \frac{J_2}{m}(F - \Gamma_1\dot{x})\right] \\
\ddot{R} &= \frac{1}{\mu(\theta)}\left[M_1\alpha(R,\theta,\dot{\theta}) - \frac{(M_1+\sin^2\theta)}{m\rho}\left(C + \Gamma_2\dot{R}\right)\right. \\
&\qquad\left. - \frac{\sin\theta}{m}(F - \Gamma_1\dot{x})\right] \\
\ddot{\theta} &= -\frac{\cos\theta}{R\mu(\theta)}\left[J_1\alpha(R,\theta,\dot{\theta})\sin\theta + \frac{\sin\theta}{m\rho}\left(C + \Gamma_2\dot{R}\right) + \frac{J_2}{m}(F - \Gamma_1\dot{x})\right] \\
&\qquad - \frac{1}{R}\left(2\dot{R}\dot{\theta} + g\sin\theta\right).
\end{aligned} \tag{13.4}$$

In the simulations, we have chosen the following values:

$$\begin{aligned}&m = 500\ kg, \quad M = 5000\ kg, \quad g = 10\ m\cdot s^{-2}, \quad J = 50 kg\cdot m^2, \\ &\rho = 0.4\ m, \quad \Gamma_1 = 20\ kg\cdot s^{-1}, \quad \Gamma_2 = 20\ kg\cdot m\cdot s^{-1}\end{aligned}$$

The friction coefficients Γ_1 and Γ_2, the cart and load masses and the winch inertia are generally inaccurately known and, in the controller, they are replaced by erroneous values.

The curves presented below have been obtained by simulating (13.4) with errors of 50% on these coefficients, as well as errors on the initial position of 1%. We'll see that the presented controllers are robust with respect to the masses, inertia and friction coefficients, but require precise measurements of the cart position and cable length, which explains why the initial error is so small compared to the other ones.

13.2 Straight Line Displacement

13.2.1 Approximate Tracking of Straight Line by Hierarchical PID Control

In this section, we assume that the angle θ remains sufficiently small, as well as its angular speed $\dot{\theta}$, so that, according to (13.2), we have $x \approx \xi$ and $z \approx -R$.

We want to follow a straight line starting from the initial position

$$\xi_i = x_i \ , \quad \zeta_i = -R_i$$

at time t_i, with

$$\theta(t_i) = \dot{\theta}(t_i) = 0 \ , \quad \dot{\xi}(t_i) = \dot{x}(t_i) = 0 \ , \quad \dot{R}(t_i) = -\dot{\zeta}(t_i) = 0$$

and arriving at the final position

$$\xi_f = x_f \ , \quad \zeta_f = -R_f$$

at time t_f, with

$$\theta(t_f) = \dot{\theta}(t_f) = 0 \ , \quad \dot{\xi}(t_f) = \dot{x}(t_f) = 0 \ , \quad \dot{R}(t_f) = -\dot{\zeta}(t_f) = 0$$

(see figure 13.2), with velocities $\dot{x}$ and $\dot{R}$ small enough (quasi-static displacement).

Denoting by $T = t_f - t_f$ the transfer duration, the reference trajectory of the load is given by:

$$\begin{aligned}
&P(t) = \left(\tfrac{t}{T}\right)^5 \left(126 - 420\left(\tfrac{t}{T}\right) + 540\left(\tfrac{t}{T}\right)^2 - 315\left(\tfrac{t}{T}\right)^3 + 70\left(\tfrac{t}{T}\right)^4\right) \\
&\xi_{ref}(t) = x_i + (x_f - x_i)P(t) \\
&\zeta_{ref}(t) = -R_i - (R_f - R_i)\left(\frac{\xi_{ref}(t) - x_i}{x_f - x_i}\right),
\end{aligned} \tag{13.5}$$

The reference trajectory of the cart and cable length are thus, according to the small angle assumption, $x^*(t) = \xi_{ref}(t)$ and $R^*(t) = -\zeta_{ref}(t)$.

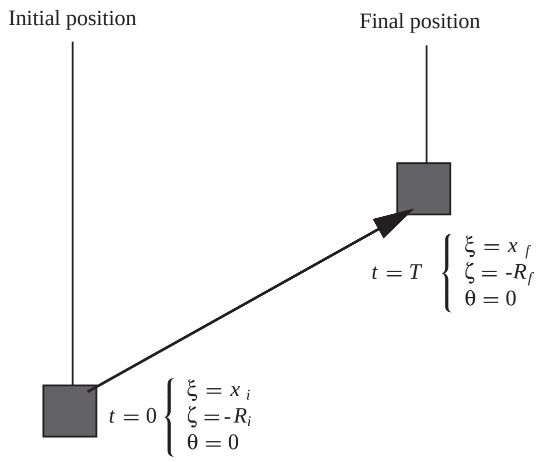

Fig. 13.2 Load displacement in straight line

The duration T may be tuned to make the deviations $x - \xi$ and $\zeta + R$ sufficiently small.

In the reminder of this section, the end point and final point are set to:

$$R_i = 5\text{m}, \quad R_f = 3\text{m}, \quad x_i = 0\text{m}, \quad x_f = 15\text{m}.$$

To track the trajectory (13.5) we introduce the following high-gain PID design for F and C:

$$\begin{aligned} F &= -\frac{M}{\varepsilon_x}\left((\dot{x} - \dot{x}^*) + K_1(x - x^*)\right), \\ C &= mg\rho + \frac{\frac{J}{\rho} + m\rho}{\varepsilon_R}\Big((\dot{R} - \dot{R}^*) + K_2(R - R^*)\Big) \end{aligned} \tag{13.6}$$

where K_1 and K_2 are tunable gains and ε_x and ε_R are tunable time constants, but otherwise assumed to be small enough. This choice is motivated by the fact that, in a neighborhood of the equilibrium point $x = x_{fin}$, $R = R_{fin}$ and $\theta = \theta_{fin} = 0$, the x- and R-dynamics are almost decoupled: combining (13.4) and (13.6), with $\varepsilon_x = \varepsilon_R = \varepsilon$ for simplicity's sake, and denoting by $\delta_x = x - x^*$, $\delta_R = R - R^*$, we get the closed-loop system:

$$\begin{aligned}
\varepsilon\ddot{x} &= \frac{\varepsilon}{\mu(\theta)}\left[J_1\alpha(R,\theta,\dot{\theta})\sin\theta + \sin\theta\left(g + \frac{1}{m\rho}\Gamma_2\dot{R}\right) - \frac{J_2}{m}\Gamma_1\dot{x}\right] \\
&\quad + \frac{J_2\sin\theta}{\mu(\theta)}\left[\dot{\delta}_R + K_2\delta_R\right] - \frac{J_2M_1}{\mu(\theta)}\left[\dot{\delta}_x + K_1\delta_x\right] \\
\varepsilon\ddot{R} &= \frac{\varepsilon}{\mu(\theta)}\left[M_1\alpha(R,\theta,\dot{\theta}) - (M_1+\sin^2\theta)\left(g + \frac{1}{m\rho}\Gamma_2\dot{R}\right) + \frac{\sin\theta}{m}\Gamma_1\dot{x}\right] \\
&\quad -(M_1+\sin^2\theta)J_2\left[\dot{\delta}_R + K_2\delta_R\right] + M_1\sin\theta\left[\dot{\delta}_x + K_1\delta_x)\right] \\
\varepsilon\ddot{\theta} &= -\frac{\varepsilon\cos\theta}{R\mu(\theta)}\left[J_1\alpha(R,\theta,\dot{\theta})\sin\theta + \sin\theta\left(g + \frac{1}{m\rho}\Gamma_2\dot{R}\right) - \frac{J_2}{m}\Gamma_1\dot{x}\right] \\
&\quad -\frac{\varepsilon}{R}\left(2\dot{R}\dot{\theta} + g\sin\theta\right) \\
&\quad -\frac{J_2M_1\sin\theta\cos\theta}{R\mu(\theta)}\left[\dot{\delta}_R + K_2\delta_R\right] + \frac{J_2M_1\cos\theta}{R\mu(\theta)}\left[\dot{\delta}_x + K_1\delta_x)\right]
\end{aligned}$$

Therefore, for ε small enough, in a neighborhood of $\theta = 0$, an approximation of the slow dynamics of this closed-loop system at the order 0 in ε is given by

$$\dot{\delta}_x + K_1\delta_x = 0, \quad \dot{\delta}_R + K_2\delta_R = 0$$

(see section 3.3).

Moreover, K_1 and K_2 are chosen positive and such that the tangent approximation of the closed-loop system at the equilibrium point $x = x_{fin}$, $R = R_{fin}$, $\theta = 0$, is exponentially stable. For instance, if we take $\varepsilon = 0.01$ and $K_1 = K_2 = 20$, the corresponding eigenvalues are

$$\begin{array}{l}
-27.64 \\
-72.37 \\
-1.710^{-7} + 0.58i \\
-1.710^{-7} - 0.58i \\
-27.60 \\
-72.46
\end{array}$$

which proves that, if θ can be kept small enough during the displacement, the end point is a locally exponentially stable equilibrium point, though poorly damped in view of the pair of conjugated eigenvalues with real part of order 10^{-7}.

This is shown in simulation on Figure 13.3 for a transfer duration of 10s. Note that the force and torque controls are saturated at the initial time and that the maximum value of θ is 0.2745rd ($\approx 16°$).

A simulation in the same conditions but with $T = 8$s is shown in Figure 13.4. The force and torque controls are again saturated at the initial time and the maximum value of θ is now 0.6958rd ($\approx 40°$). Though the motion is still stable, it is significantly deteriorated and the damping is too slow.

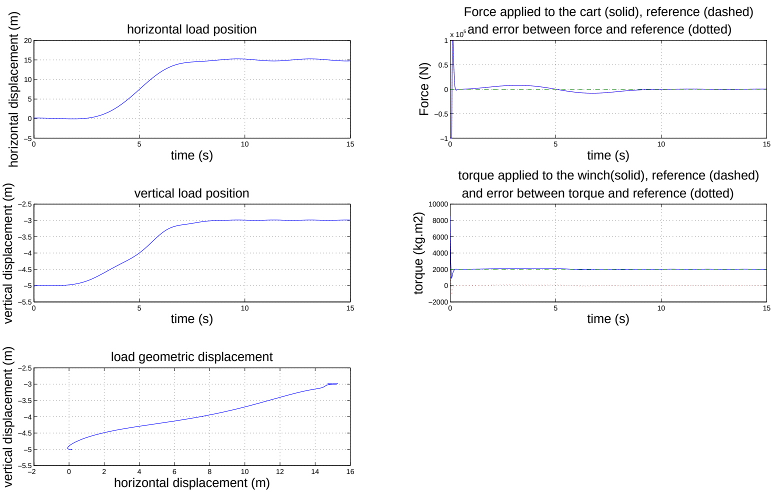

Fig. 13.3 Straight line following with small angle approximation and $T = 10$s.

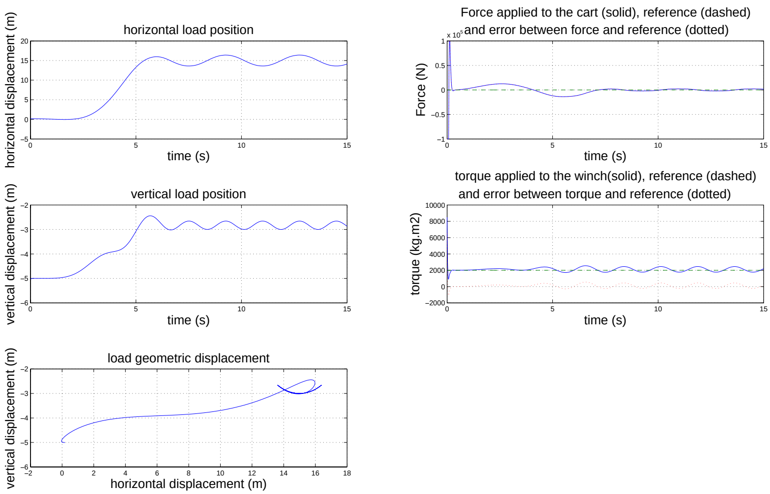

Fig. 13.4 Straight line following with small angle approximation and $T = 8$s.

We finally present a simulation, still in the same conditions, with $T = 6$s in Figure 13.5. The motion is no more acceptable since the maximum of θ is now $\approx 90°$.

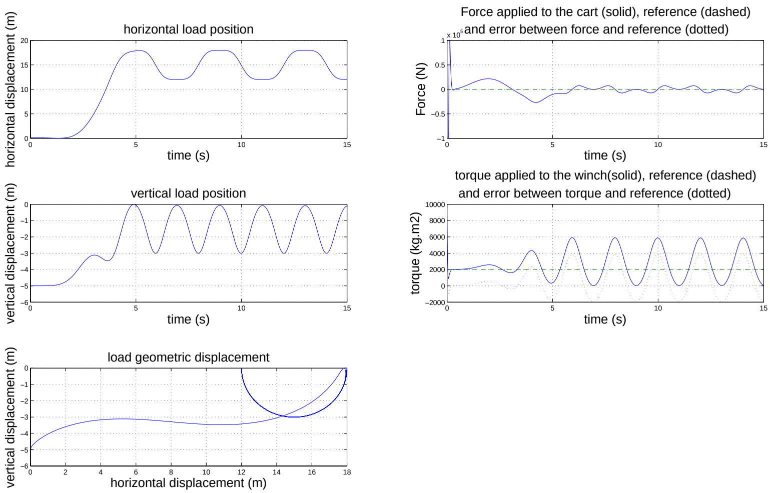

Fig. 13.5 Straight line following with small angle approximation and $T = 6$s.

13.2.2 Straight Line Tracking Without Small Angle Approximation

If, in place of the previous approximation, we compute the exact reference trajectory of x and R deduced from (5.8):

$$\begin{aligned} x^{**}(t) &= \xi_{ref}(t) - \frac{\ddot{\xi}_{ref}(t)\zeta_{ref}(t)}{\ddot{\zeta}_{ref}(t) + g} \\ R^{**}(t) &= \sqrt{\zeta_{ref}^2(t) + \left(\frac{\ddot{\xi}_{ref}(t)\zeta_{ref}(t)}{\ddot{\zeta}_{ref}(t) + g}\right)^2} \end{aligned} \tag{13.7}$$

where ξ_{ref} and ζ_{ref} are still given by (13.5), and we replace the controller (13.6) by the following one

$$\begin{aligned} F &= F_{ref} - \frac{M}{\varepsilon_x}\left((\dot{x} - \dot{x}^{**}) + K_1(x - x^{**})\right), \\ C &= C_{ref} + \frac{\frac{J}{\rho} + m\rho}{\varepsilon_R}\left((\dot{R} - \dot{R}^{**}) + K_2(R - R^{**})\right) \end{aligned} \tag{13.8}$$

We keep the same values of ε_x, ε_R, K_1 and K_2 as in the previous section. We now prove the local stability of the closed-loop system around the endpoint, following Kiss et al. [2000a].

Theorem 13.1. *The end point is a semi-globally*[1] *stable equilibrium of the closed-loop system.*

Proof. Let us consider the following function :

$$\begin{aligned} V(t,X) = &\frac{1}{2}M\dot{x}^2 + \frac{1}{2}\frac{J}{\rho^2}\dot{R}^2 + \frac{1}{2}m\left(\dot{\xi}^2 + \dot{\zeta}^2\right) + mgR(1-\cos\theta) \\ &+\frac{1}{2}\left(\frac{MK_1}{\varepsilon_x}\right)(x - x^{**}(t))^2 + \frac{1}{2}\left(\frac{J+m\rho}{\rho^2\varepsilon_R}\right)(R - R^{**}(t))^2 \end{aligned} \tag{13.9}$$

where we have denoted by $X = (x, \dot{x}, R, \dot{R}, \theta, \dot{\theta})$.

We have $V(t,X) \geq 0$ for all t, X since the sum of squares is non negative and the only non quadratic term $mgR(1-\cos\theta)$ is also non negative since $R \geq 0$.

The level set $S(t,C) = \{X \in \mathbb{R}^2 \times ([0,R_0] \times \mathbb{R}) \times ([0,2\pi] \times \mathbb{R})\,|V(t,X) \leq C\}$, for all $t \geq t_i$ and $0 < C < \infty$, is bounded. Clearly, if $S(t,C)$ was unbounded, at least one of the components of X could be infinite. But it is impossible that one of the components $x, \dot{x}, R, \dot{R}$ be infinite without implying that the corresponding squared term in V is infinite and thus V itself too, which is absurd since $V \leq C$. Moreover $\theta \in [0, 2\pi]$ is bounded by construction. It remains to prove that $\dot{\theta}$ is bounded. Following the same argument as before, we easily deduce that $\dot{\xi}$ and $\dot{\zeta}$ are bounded. But $\dot{\zeta} = -\dot{R}\cos\theta + R\dot{\theta}sin\theta$ with x, $R, \dot{R}$ and θ bounded, and $R \geq 0$, immediately yields the boundedness of $\dot{\theta}$, hence the result.

An easy but lengthy computation shows that

$$\begin{aligned} \dot{V}(t,X) = -&\left[\Gamma_1\dot{x}^2 + \frac{\Gamma_2}{\rho}\dot{R}^2 + \frac{M}{\varepsilon_x}(\dot{x}-\dot{x}^{**})^2 + \frac{J+m\rho}{\rho^2\varepsilon_R}\left(\dot{R}-\dot{R}^{**}\right)^2\right] \\ &-\frac{M}{\varepsilon_x}\dot{x}^{**}\left(K_1(x-x^{**}) + (\dot{x}-\dot{x}^{**})\right) \\ &-\frac{J+m\rho}{\rho^2\varepsilon_R}\dot{R}^{**}\left(K_2(R-R^{**}) + (\dot{R}-\dot{R}^{**})\right) \end{aligned} \tag{13.10}$$

Therefore, since we have $\dot{x}^{**}(t) = \dot{R}^{**}(t) = 0$ for all $t \geq t_f$, we get

$$\begin{aligned} \dot{V}(t,X) = -&\left[\left(\Gamma_1 + \frac{M}{\varepsilon_x}\right)\dot{x}^2 + \left(\frac{\Gamma_2}{\rho} + \frac{J+m\rho}{\rho^2\varepsilon_R}\right)\dot{R}^2\right] \\ &\leq 0\,, \quad \text{for all } \; t \geq t_f. \end{aligned} \tag{13.11}$$

We have thus proven that we can apply LaSalle's invariance principle to V (see Theorem 3.3 of section 3.2.4): for every initial condition in $S(t,C)$ with $t \geq t_f$ and arbitrary $C > 0$, the state of the closed-loop system (13.4) with (13.8) converges to the largest invariant subset W contained in $\dot{V} = 0$,

[1] this means that every compact neighborhood contained in $\mathbb{R}^2 \times ([0,R_0] \times \mathbb{R}) \times ([0,2\pi] \times \mathbb{R})$ is positively invariant and that the state starting from this subset asymptotically converges to the equilibrium end point

which, as a consequence of (13.11), is contained in the set $\dot{x} = \dot{R} = 0$ for all $t \geq t_f$.

Since $\dot{x} = \dot{R} = 0$ for all $t \geq t_f$, x and R must remain constant. Let us denote by $\bar{x}$ and $\bar{R}$ these constants, and by $\bar{F}$ and $\bar{C}$ the corresponding limit force and torque, namely

$$\bar{F} = -\frac{MK_1}{\varepsilon_x}(\bar{x} - x^{**}), \quad \bar{C} = mg\rho + \frac{(J + m\rho)K_2}{\rho\varepsilon_R}(\bar{R} - R^{**})$$

But, according to the dynamics, using (13.1) with $\dot{x} = \ddot{x} = \dot{R} = \ddot{R} = 0$, we get

$$g\cos\theta\sin\theta = -\frac{\bar{F}}{m}, \quad g\cos\theta = \frac{\bar{C}}{m\rho}$$

We immediately deduce that since $\bar{F}$ and $\bar{C}$ are constant, θ must be equal to a constant too, say $\bar{\theta}$, and that

$$g\cos\bar{\theta}\sin\bar{\theta} = \frac{\bar{C}}{m\rho}\sin\bar{\theta} = -\frac{\bar{F}}{m} = -\frac{MK_1}{\varepsilon_x}(\bar{x} - x^{**}). \tag{13.12}$$

The fact that $\theta = \bar{\theta}$ implies that $\dot{\theta} = \ddot{\theta} = 0$ and, using the last equation of (13.1), we get $-g\sin\bar{\theta} = 0$, or $\bar{\theta} = 0 \mod \pi$. From (13.12), we immediately get that $\bar{x} = x^{**}$ and that $g\cos\bar{\theta} = g = \frac{\bar{C}}{m\rho}$ which also proves that $\bar{R} = R^{**}$, which achieves to prove that the largest invariant subset contained in $\dot{V} = 0$ is reduced to the equilibrium point $x = x^{**}$, $R = R^{**}$ and $\theta = 0$.

Let us now illustrate this result by the next simulations.

If we apply the controller (13.8) in place of (13.6) for the same duration $T = 6$s, we obtain a motion which is smoother that the one obtained in the previous section for $T = 10$s, with oscillation amplitude at the end divided by 2, though almost twice as fast. Remark that the force and torque are saturated at start, but don't significantly perturb the motion. Note also that the error on the friction coefficients explains the torque deviation of Figure 13.6, bottom right, automatically compensated in closed-loop.

The same simulation is presented for $T = 4$s in Figure 13.7. Though the motion is accelerated by a factor of 33%, no significant deterioration is noticed. However, since the horizontal and vertical accelerations are larger at start, the force and torque are saturated on longer periods at start.

If the total duration is decreased to $T = 3.5$s (Figure 13.8), the accelerations become so large that they are saturated half of the time at start and destabilize the motion.

Nevertheless, the above simulations show that using the flatness-based reference trajectories (13.7) allows to significantly accelerate the motion without amplifying the oscillations and that this design is robust (50% error on the

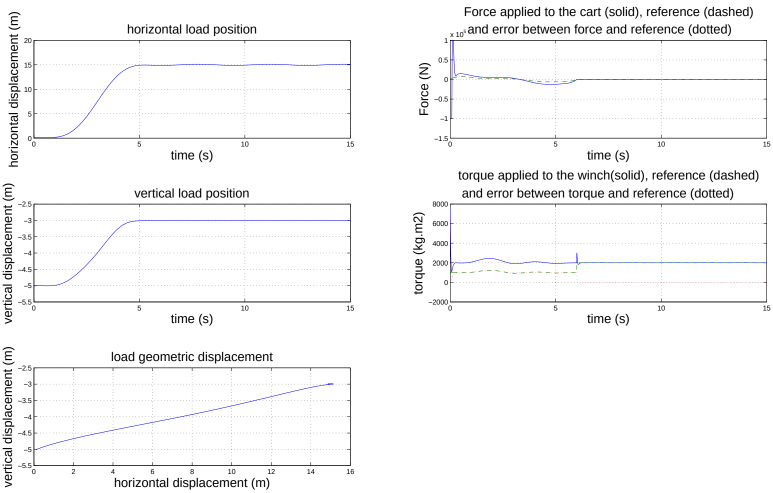

Fig. 13.6 Straight line following without small angle approximation and $T = 6$s.

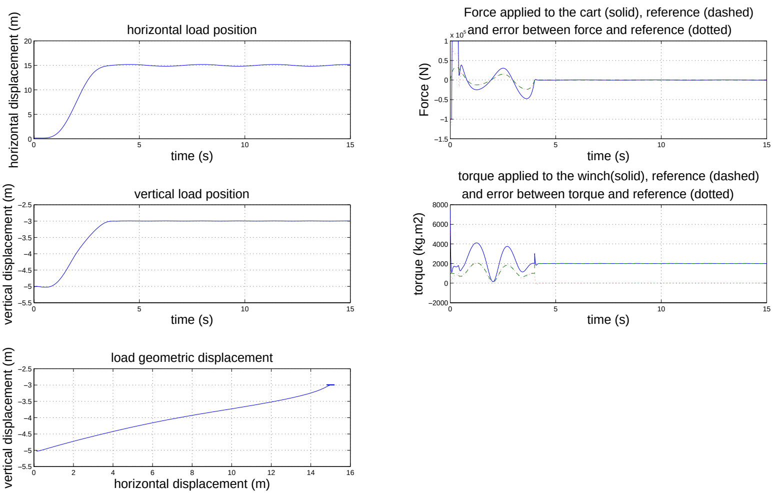

Fig. 13.7 Straight line following without small angle approximation and $T = 4$s.

masses, inertia and friction coefficients) if we except the position and velocity errors which must remain very small.

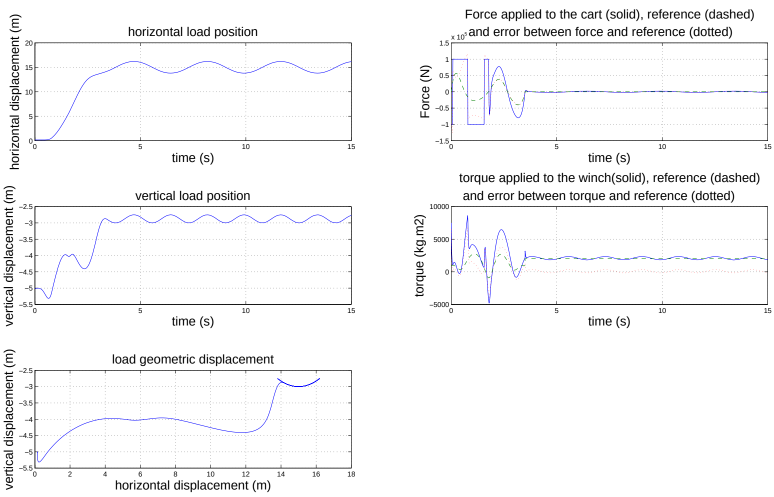

Fig. 13.8 Straight line following without small angle approximation and $T = 3.5$s.

13.3 Obstacle Avoidance

We now want to avoid an obstacle located in the middle of the load trajectory as shown in Figure 13.9. More precisely, we want the load to follow a polynomial trajectory whose maximum is $(\frac{\xi_f+\xi_i}{2}, 2\zeta_f - \zeta_i)$.

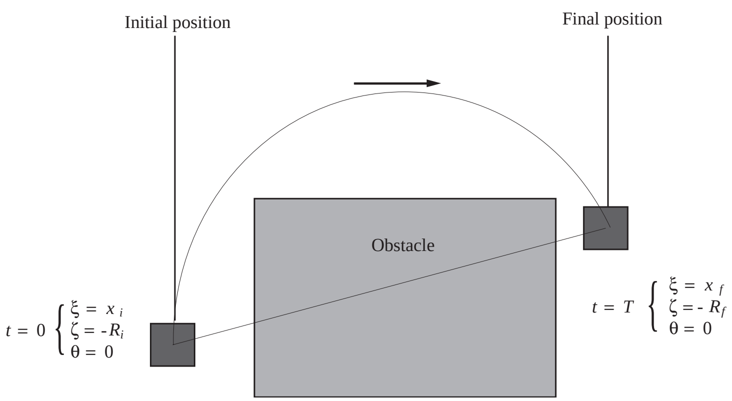

Fig. 13.9 Polynomial trajectory for obstacle avoidance.

One can verify that the following polynomial trajectory:

$$\zeta_{ref}(t) = \zeta_i + (\zeta_f - \zeta_i)\left(\frac{\xi_{ref}(t) - \xi_i}{\xi_f - \xi_i}\right) \cdot \\ \cdot \left(9 - 12\left(\frac{\xi_{ref}(t) - \xi_i}{\xi_f - \xi_i}\right) + 4\left(\frac{\xi_{ref}(t) - \xi_i}{\xi_f - \xi_i}\right)^2\right) \tag{13.13}$$

with ξ_{ref} given by (13.5), satisfies these requirements.

This new reference trajectory will be used in this section instead of the straight line (13.5).

13.3.1 Tracking With Small Angle Approximation

As in section 13.2.1, assuming that the angle θ remains small, the reference trajectories of x and R are given by

$$x^*(t) = \xi_{ref}(t), \quad R^*(t) = -\zeta_{ref}(t)$$

with ξ_{ref} given by (13.5) and ζ_{ref} by (13.13). The feedback is again given by (13.6) with the same gains as before. Note that the local stability of the endpoint remains unchanged.

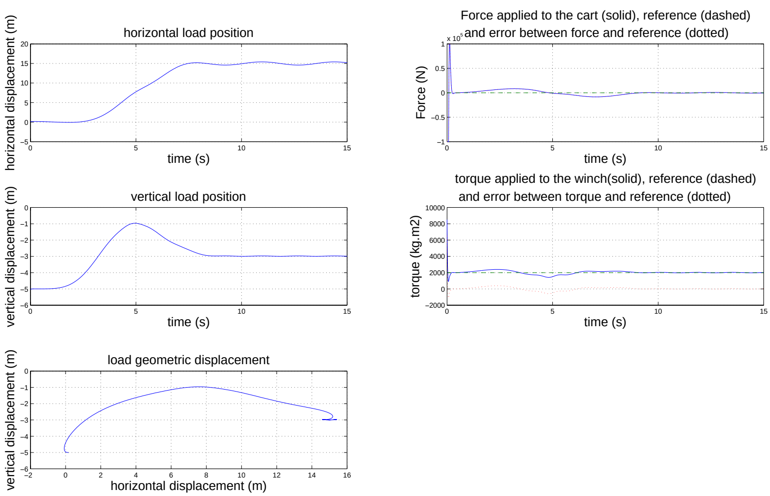

Fig. 13.10 Obstacle avoidance with small angle approximation and $T = 10$s.

The motion corresponding to a duration $T = 10$s is presented in Figure 13.10. The same remarks on too slow damping and saturation at start as in section 13.2.1 apply, as well as for $T = 8\text{s}$ (Figure 13.11). Again, for $T = 6\text{s}$, the motion becomes unstable (Figure 13.12).

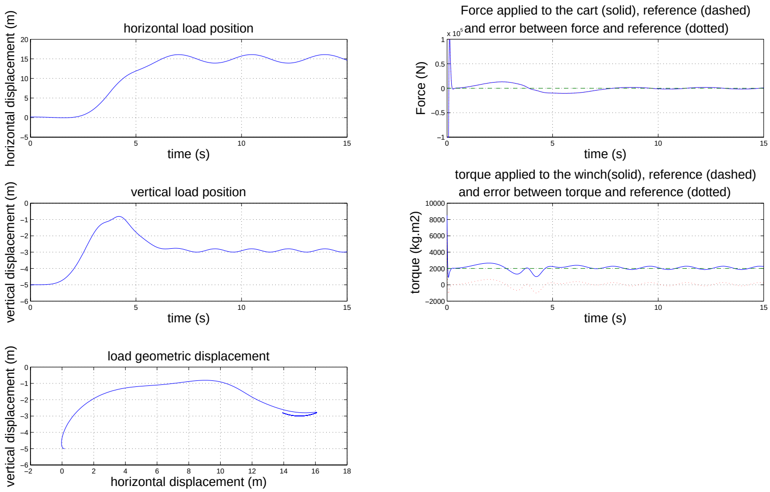

Fig. 13.11 Obstacle avoidance with small angle approximation and $T = 8\text{s}$.

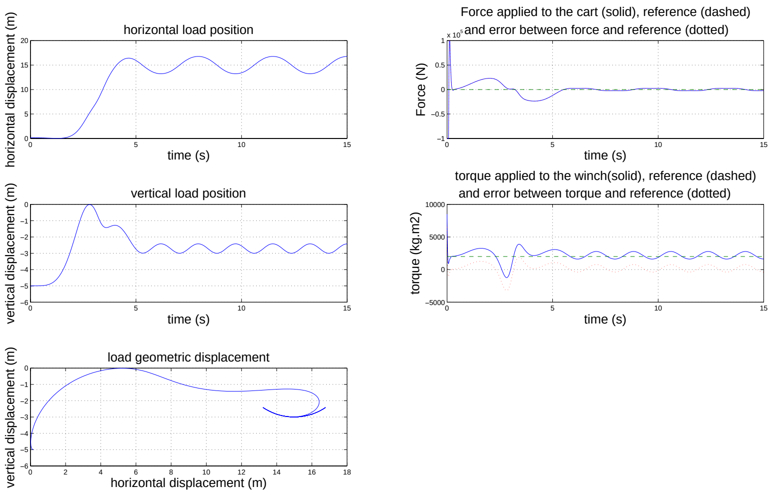

Fig. 13.12 Obstacle avoidance with small angle approximation and $T = 6\text{s}$.

13.3.2 Tracking Without Small Angle Approximation

We now use the reference trajectory ξ_{ref} given by (13.5) and ζ_{ref} by (13.13), x^{**} and z^{**} being deduced from (13.7). The feedback is again (13.8) with these references. Note, as before, that Theorem stabcrane-thm is still valid since it doesn't depend on the reference trajectory, the only needed property being that it arrives at rest. As in the case of straight line tracking, the quality of tracking is now considerably improved: for $T = 6$s (Figure 13.13), the swing amplitude at the end is at least twice as small as the one with the previous controller for $T = 10$s.

The duration T can be decreased to 5s without noticeable drawback(Figure 13.14).

Only when $T \leq 4.3$s, the intitial accelerations are too large compared to their saturation values and the motion becomes inacceptable (Figure 13.15).

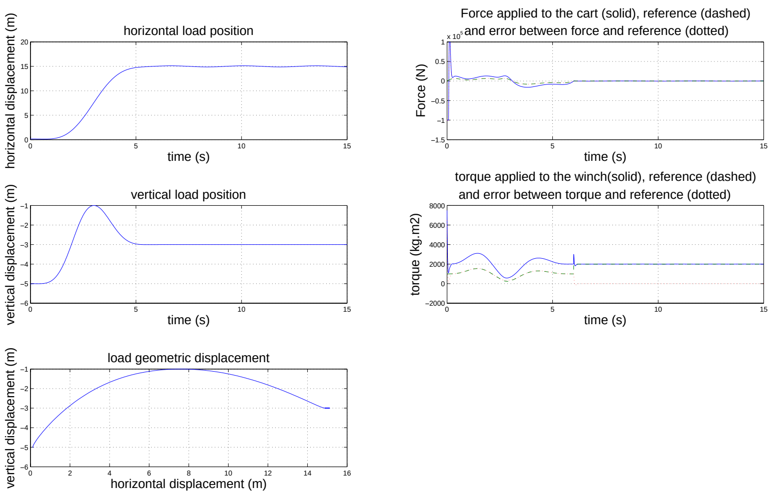

Fig. 13.13 Obstacle avoidance without small angle approximation and $T = 6$s.

We therefore conclude, as in the straight line section, that using the flatness-based reference trajectories makes possible a significant decrease of the transfer duration without amplifying the oscillations and in a robust way, except for the position and velocity errorsl. This shows that particular efforts must be made on the construction of filters or observers to reduce the observation errors in order to improve the robustness.

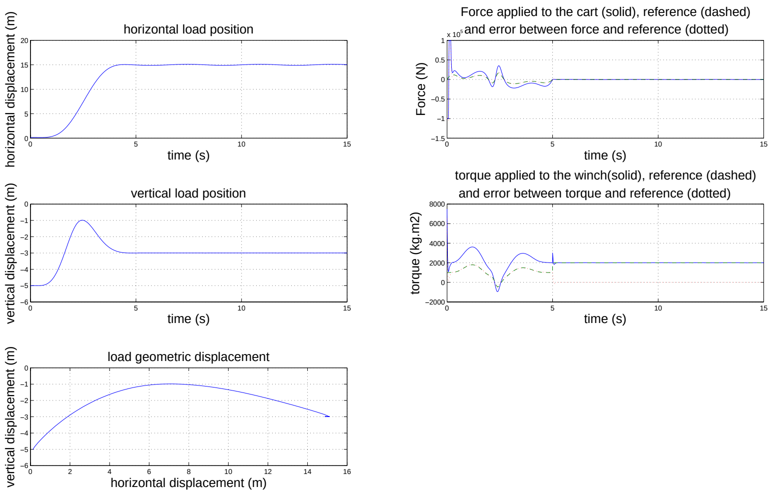

Fig. 13.14 Obstacle avoidance without small angle approximation and $T = 5$s.

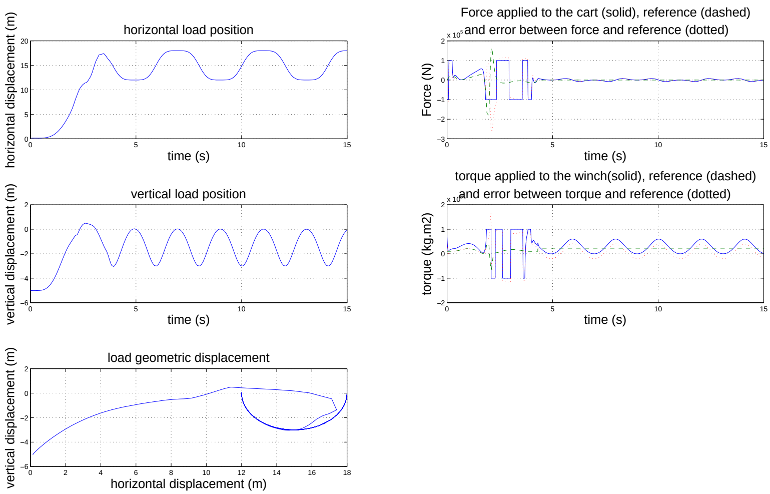

Fig. 13.15 Obstacle avoidance without small angle approximation and $T = 4.3$s.

Chapter 14
Automatic Flight Control Systems

The present chapter[1] is devoted to the flatness-based control design for one of the major control applications of the 20th century, namely flight control.

Automatic Flight Control Systems (AFCS) have been, since the 60's, a major field of application of control methods. The main drawback of most current AFCSs is that they follow the principle of *one function–one controller*. In each case, pitch attitude control, wing leveler, side-slip suppressor, landing, terrain-following, target tracking, etc., a different controller is associated. These controllers may be based on different models, i.e., linear approximation along the trajectory to be followed, linear or nonlinear input-output description, the output of which is chosen such that its regulation corresponds to the control objective (see e.g. Etkin [1982], McLean [1990], McRuer et al. [1973], Vukobratović and Stojić [1988], Wanner [1984]).

On advanced autopilots, the number of possible combinations of functions and reference trajectories is large. Moreover, strong limitations must be imposed on both the maneuvers and their sequencing rate, which must be kept slow enough to avoid instabilities. Indeed, as far as linear techniques are concerned, one has to restrict to sequences of trajectories for which the tangent approximation is nearly time invariant (pieces of straight lines, circles, or helices with slowly varying altitude, etc.). On the other hand, in most nonlinear approaches, the control objective is translated into an output variable whose reference trajectory, generally constant, is assumed to satisfactorily describe the maneuver. In some cases, however, the stability of the overall closed-loop dynamics causes difficulties. In most cases, designing the switches from one function to another requires a lot of attention. Furthermore, since the highest level loop is the pilot himself, the switching policy may lead to complex tradeoffs in order to take into account the pilot's online queries.

New complex tasks such as collision avoidance, or, more specifically in the case of military or remotely operated aircraft, terrain following or orientation control are needed to complement the classical ones. Hence, a simpler control

[1] Work done by Ph. Martin and the author, in a collaboration with Sextant Avionique

J. Lévine, *Analysis and Control of Nonlinear Systems,* Mathematical Engineering,
DOI 10.1007/978-3-642-00839-9_14, © Springer-Verlag Berlin Heidelberg 2009

architecture is required to both improve reliability and decrease development costs. To this aim, we propose a single controller comprised of a universal reference trajectory generator and a fixed tracking loop. In addition, general tracking tasks, for which linearization about operating points is insufficient, require that the main nonlinearities of the system are properly taken into account. We show that such a controller can be designed thanks to the (almost) flatness property of generic aircraft models. We briefly describe its construction, performances and some perspectives.

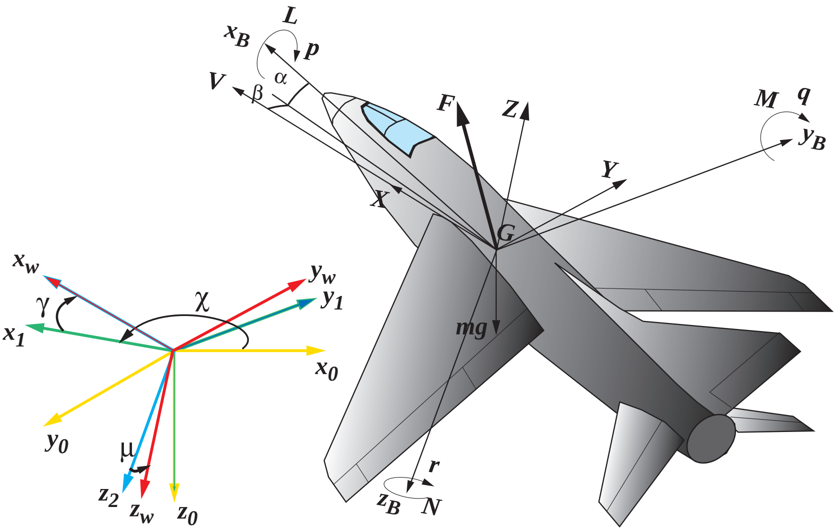

Fig. 14.1 Aircraft coordinate frames, forces and moments

14.0.3 Generic Aircraft Model

The aircraft we consider is *generic* in the sense that, as in most civil and military ones, it is actuated through its thrust and deflection surface positions, namely the elevator, ailerons and rudder, in the 3 directions.

A complete discussion of the model can be found in Martin [1992, 1996]. We partly follow Charlet et al. [1991], Vukobratović and Stojić [1988], Wanner [1984] (see also Etkin [1982], McLean [1990], McRuer et al. [1973]).

Aircraft dynamics are generally described by a set of 12 variables: the 3 components of the position of the center of mass, 3 variables for the velocity vector, 3 angular positions describing the aircraft attitude, and 3 variables for the corresponding angular velocities.

There are many possible choices, depending on the coordinate frames where positions, velocities, forces and moments are expressed. Here, we are using (see Figure 14.1):

- x, y, z, components of the center of mass in the Earth axes,
- V, α, β, velocity, angle of attack and side-slip angle in the body axes,
- χ, γ, μ, orientation of the wind axes with respect to the Earth axes,
- p, q, r, components of the angular velocity in the body axes.

The aircraft is conventionally actuated by four independent controls: the thrust F and the positions $(\delta_l, \delta_m, \delta_n)$ of the deflection surfaces. We may also want to include the servo model that produce the required deflection positions, the inputs of which are denoted $(\tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n)$.

The resultant of the external forces, expressed in the wind axes, is denoted by (X, Y, Z). It is equal to the sum of the aerodynamic, gravitational (weight) and propulsive (thrust) forces:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \underbrace{\frac{1}{2}\rho S V^2 \begin{pmatrix} -C_x \\ C_y \\ -C_z \end{pmatrix}}_{\text{Aerodynamic forces}} + \underbrace{F \begin{pmatrix} \cos(\alpha+\varepsilon)\cos\beta \\ \cos(\alpha+\varepsilon)\sin\beta \\ -\sin(\alpha+\varepsilon) \end{pmatrix}}_{\text{Thrust}} + \underbrace{mg \begin{pmatrix} -\sin\gamma \\ \cos\gamma\sin\mu \\ \cos\gamma\cos\mu \end{pmatrix}}_{\text{Weight}}. \tag{14.1}$$

The dimensionless so-called *aerodynamic coefficients* C_x, C_y, C_z (which are in fact functions of all the configuration variables, as detailed below) are experimentally determined in a wind tunnel (see Vukobratović and Stojić [1988] for an example). They are naturally expressed in the wind axes, in which the measurements are carried out, which justifies our (quite unusual) choice of coordinate frame. The air density is denoted by ρ. The remaining constants are the reference surface S and the angle ε between the axis of the propeller and the axis Gx_B (see figure 14.1) in the symmetry plane of the aircraft.

Similarly, let (L, M, N) denote the components, in the body axes, of the sum of the external torques, generated by the external forces about the center of mass:

$$\begin{pmatrix} L \\ M \\ N \end{pmatrix} = \frac{1}{2}\rho S V^2 \begin{pmatrix} aC_l \\ bC_m \\ cC_n \end{pmatrix} + F \begin{pmatrix} 0 \\ d \\ 0 \end{pmatrix}, \tag{14.2}$$

where C_l, C_m, C_n are dimensionless *aerodynamic coefficients* (same remark as above) and a, b, c, d are constant (reference lengths).

It is commonly assumed that $C_x, C_y, C_z, C_l, C_m, C_n$ depend on the translational velocity (i.e., V, α, β) and acceleration $(\dot{V}, \dot{\alpha}, \dot{\beta})$, angular velocity (p, q, r), position of the deflection surfaces $(\delta_l, \delta_m, \delta_n)$, and Mach number (i.e., V/c, where c is the velocity of sound in air). They vary a lot with the Mach number during transsonic flight. A crucial remark for the design of our control law will be that, for nearly every aircraft, the dependence of C_x, C_y, C_z on $\dot{V}, \dot{\alpha}, \dot{\beta}$, p, q, r and $\delta_l, \delta_m, \delta_n$ is **weak**. In the same way, though it is not as important, the dependence of C_l, C_m, C_n on $\dot{V}, \dot{\alpha}, \dot{\beta}, p, q, r$ is also **weak**. Of course, on the contrary, C_l, C_m, C_n strongly depend on the deflection surface positions $\delta_l, \delta_m, \delta_n$. Thanks to this dependence, the moments, and hence the aircraft attitude, can be controlled.

Applying Newton's Second Law, we get

$$
\begin{aligned}
&\dot{x} = V \cos\chi \cos\gamma \\
&\dot{y} = V \sin\chi \cos\gamma \\
&\dot{z} = -V \sin\gamma \\
&\dot{V} = \frac{X}{m} \\
&\dot{\alpha}\cos\beta = -p\cos\alpha\sin\beta + q\cos\beta - r\sin\alpha\sin\beta + \frac{Z}{mV} \\
&\dot{\beta} = p\sin\alpha - r\cos\alpha + \frac{Y}{mV} \\
&\dot{\gamma} = -\frac{Y\sin\mu}{mV} - \frac{Z\cos\mu}{mV} \\
&\dot{\chi}\cos\gamma = \frac{Y\cos\mu}{mV} - \frac{Z\sin\mu}{mV} \\
&\dot{\mu}\cos\beta = p\cos\alpha + r\sin\alpha + \frac{Y\cos\mu\tan\gamma\cos\beta}{mV} \\
&\qquad\qquad - \frac{Z(\sin\mu\tan\gamma\cos\beta + \sin\beta)}{mV} \\
&I\begin{pmatrix}\dot{p}\\ \dot{q}\\ \dot{r}\end{pmatrix} = \begin{pmatrix}L\\ M\\ N\end{pmatrix} - \begin{pmatrix}p\\ q\\ r\end{pmatrix} \wedge I\begin{pmatrix}p\\ q\\ r\end{pmatrix}
\end{aligned}
\tag{14.3}
$$

where $\wedge$ denotes the wedge product of 3-vectors, with the exterior forces X, Y, Z given by (14.1) and the torques L, M, N given by (14.2).

These equations must be completed by the servoactuator dynamics, whose slowest part, though much faster than (14.3), may be represented by:

$$
\varepsilon^2 \begin{pmatrix}\dot{\delta}_l\\ \dot{\delta}_m\\ \dot{\delta}_n\end{pmatrix} = A\begin{pmatrix}\delta_l\\ \delta_m\\ \delta_n\end{pmatrix} + \begin{pmatrix}\tilde{\delta}_l\\ \tilde{\delta}_m\\ \tilde{\delta}_n\end{pmatrix}
\tag{14.4}
$$

where ε represents the order of magnitude between the fastest time constants of (14.3) and these of (14.4), and where $\tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n$ are the actual controls. The matrix A is stable. It may be noted, as will be detailed below, that the time constants corresponding to the angular acceleration subsystem of (14.3) (last equation of (14.3)) are of order ε compared to these of the remaining part of (14.3), and that those of (14.4) are in turn of order ε compared to these of the angular acceleration subsystem. This explains why in (14.4) the scaling ε^2 is needed.

For the sake of simplicity, we do not consider here the dynamical model relating the thrust F and the throttle.

14.0.4 Flatness Based Autopilot Design

Assume that the aerodynamic coefficients C_x, C_y, C_z depend only on (z, V, α, β), the altitude, Mach number, angle of attack, and side-slip angle respectively. Assume furthermore that C_l, C_m, C_n depend only on $(z, V, \alpha, \beta, \delta_l, \delta_m, \delta_n)$, namely on the positions of the deflection surfaces in addition to the aforementioned variables. This means in particular that we have neglected the naturally small contribution of the variables $(\dot{V}, \dot{\alpha}, \dot{\beta}, p, q, r, \delta_l, \delta_m, \delta_n)$ in C_x, C_y, C_z, and of $(\dot{V}, \dot{\alpha}, \dot{\beta}, p, q, r)$ in C_l, C_m, C_n. Following Martin [1996], we call these approximated coefficients and the corresponding model (14.3), (14.4) *ideal*.

We first sketch the construction of a trajectory generator for the ideal aircraft and then of a unique controller for the real aircraft. This controller will be independent of the trajectories to be tracked.

14.0.4.1 Flatness of the Ideal Aircraft Model

Let us show that all the trajectories of all the variables of the ideal aircraft model are parametrized by (x, y, z, β), and a finite number of their time derivatives. In other words, given a smooth (sufficiently continuously differentiable) arbitrary trajectory for the quadruple (x, y, z, β), one can uniquely deduce the trajectory of each variable of the system, including the control inputs, without integrating the system differential equations. It means that the ideal aircraft is flat with (x, y, z, β) as a flat output (see Fliess et al. [1995, 1999], Martin [1992, 1996]).

More precisely, let us show that all the system variables $(x, y, z, V, \gamma, \chi, \mu, F, \alpha, \beta, p, q, r, \delta_l, \delta_m, \delta_n, \tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n)$ can be expressed as functions of $(x, \ldots, x^{(5)}, y, \ldots, y^{(5)}, z, \ldots, z^{(5)}, \beta, \ldots, \beta^{(3)})$, where the superscript $^{(i)}$ stands for the ith order time derivative, $i \in \mathbb{N}$, of the corresponding function.

Inverting the 3 first equations of (14.3), we get V, γ, χ as functions of $x, \dot{x}, y, \dot{y}, z, \dot{z}$:

$$V = \left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right)^{\frac{1}{2}}, \quad \gamma = \arctan\left(\frac{\dot{z}}{\left(\dot{x}^2 + \dot{y}^2\right)^{\frac{1}{2}}}\right), \quad \chi = \arctan\left(\frac{\dot{y}}{\dot{x}}\right).$$

Differentiating them, and inverting the equations of $\dot{V}, \dot{\chi}, \dot{\gamma}$, we obtain, using formulas (14.1) and (14.2) with the ideal coefficients, that F, α, μ are functions of $x, \dot{x}, \ddot{x}, y, \dot{y}, \ddot{y}, z, \dot{z}, \ddot{z}, \beta$ (note that these functions cannot be computed explicitly in general and may be obtained by numerical approximation). Differentiating again, we find that p, q, r are functions of $x, \dot{x}, \ddot{x}, x^{(3)}, y, \dot{y}, \ddot{y}, y^{(3)}, z, \dot{z}, \ddot{z}, z^{(3)}, \beta, \dot{\beta}$ and that $\delta_l, \delta_m, \delta_n$ are functions of $x, \dot{x}, \ddot{x}, x^{(3)}, x^{(4)}, y, \dot{y}, \ddot{y}, y^{(3)}, y^{(4)}, z, \dot{z}, \ddot{z}, z^{(3)}, z^{(4)}, \beta, \dot{\beta}, \ddot{\beta}$. Finally, by the same technique, it results that $\tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n$ are functions of $x, \dot{x}, \ddot{x}, x^{(3)}, x^{(4)}, x^{(5)}, y, \dot{y}, \ddot{y}, y^{(3)}, y^{(4)}, y^{(5)}, z, \dot{z}, \ddot{z}, z^{(3)}, z^{(4)}, z^{(5)}, \beta, \dot{\beta}, \ddot{\beta}, \beta^{(3)}$.

This construction is summarized in the following diagram, where the left side column contains the flat output components, the middle column the state variables, and the right side column the input variables:

$$\begin{array}{ccccccc}
(x, y, z) & \xrightarrow{\frac{d}{dt}} & (V, \gamma, \chi) & & \\
& & \downarrow \frac{d}{dt} & & \\
\beta & \longrightarrow & (\alpha, \mu) & \longrightarrow & F \\
& & \downarrow \frac{d}{dt} & & \\
& & (p, q, r) & & \\
& & \downarrow \frac{d}{dt} & & \\
& & (\delta_l, \delta_m, \delta_n) & \xrightarrow{\frac{d}{dt}} & (\tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n).
\end{array} \qquad (14.5)$$

14.0.4.2 Trajectory Generation

As a consequence of flatness, any trajectory of the ideal aircraft will be easily generated via the construction of a smooth curve $t \mapsto (x(t), y(t), z(t), \beta(t))$: given suitable initial and final conditions (positions, velocities, accelerations, jerks,...) and possibly saturation constraints on the path curvature, velocity, acceleration, etc., it may be constructed by interpolation. Then the state and input variables are recovered according to the above calculations, without need to integrate the system.

For instance, in the case where we want the side-slip angle β to remain equal to 0, the trajectories for x, y, and z may be represented by three 7th degree polynomials with respect to time, with possible simplifications for simple maneuvers as those considered in classical flight control systems. More details are given in Martin [1992, 1996].

14.0.4.3 Theoretical Comparison Between Classical AFCS and Flatness-based Designs

Let us sketch the differences between our approach and the classical ones on the simple example of a steady turn. A steady turn consists in setting $\dot{V} = \dot{\alpha} = \dot{\beta} = 0$, $\dot{p} = \dot{q} = 0$ and $\dot{r}$ equal to a constant. The corresponding forces and torques are immediately deduced from these conditions. The reference trajectory is then computed by connecting the corresponding circle to the current aircraft position by drawing a straight line along the current velocity vector (assuming that it is constant or almost constant), that touches tangentially the circle (see Fig. 14.2). However, at the junction point, the acceleration may jump, thus creating a perturbation to be attenuated by the controller. It can also be filtered to avoid a too shaky behavior.

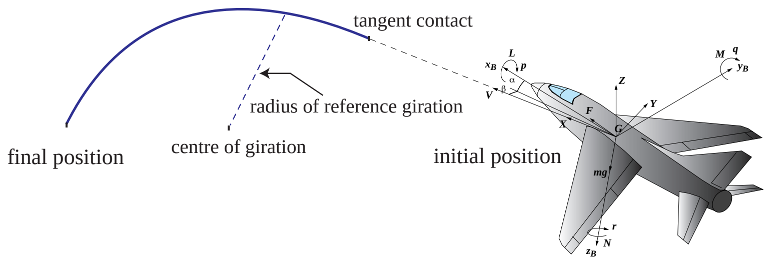

Fig. 14.2 Steady turn with tangent junction

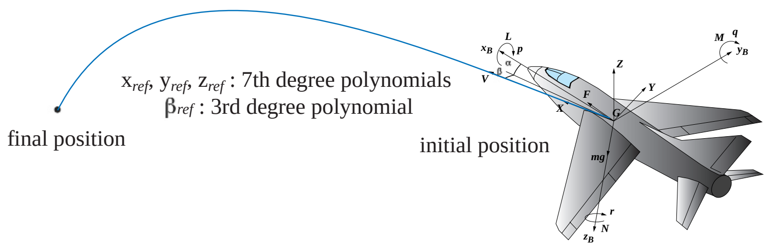

Fig. 14.3 Flatness based reference trajectory

In our approach, we only need specifying the initial and final state, velocity, acceleration and jerk (see Fig. 14.3), and translate them into initial and final conditions for the flat output and its derivatives. We generate a flat output reference $(x_{ref}, y_{ref}, z_{ref}, \beta_{ref})$ by polynomial interpolation with respect to time, more precisely, by interpolating this set of initial and final conditions, which involve the derivatives of x, y and z up to the order 3 and those of β up to the order 1. We thus end up with a 7th degree polynomial for x_{ref}, y_{ref} and z_{ref} and a 3rd degree one for β_{ref}. The resulting state and input variable reference trajectories satisfy the ideal system, or, in other words, exactly satisfy the ideal aircraft model differential equations, with an initial error which is 0 by construction, if the initial state is precisely measured. If the ideal system is not too far from reality, which we can reasonably expect, the perturbations created by the error between the ideal model and reality might be smaller than with the previous method. Therefore, we decrease the risk of saturation of the actuators, have smoother and faster responses, generating less wear, and so on.

Note that this generator is not constrained to sequences of simple geometric paths, corresponding to steady dynamics and, moreover, that constraints on the velocity, acceleration, etc. can be taken into account by tuning the maneuver duration (see section 7.2.2) or, if necessary, the clock rate at which the reference trajectory is tracked (see section 8.2). Therefore, applications e.g. to terrain following or orientation control are possible in this approach.

14.0.4.4 Controller Design

The flatness property is not only useful for motion planning and trajectory generation, but also for the design of a single universal controller. With this term, we mean a controller able to make the aircraft track any trajectory without reconfiguration of the variables on which we close the loop and without modification of the gains and other parameters involved in the feedback loop. Moreover, it can be synthesized as a three level cascaded PID, which is particularly surprising compared to the apparent complexity of the model[2].

Recalling the hierarchical structure of (14.5), it can be interpreted in the following way: the ideal system is composed of

1. a first part consisting of the string of integrators

$$(p, q, r, F) \longrightarrow (\alpha, \beta, \mu) \longrightarrow (V, \gamma, \chi) \longrightarrow (x, y, z) \tag{14.6}$$

   where (F, p, q, r) may be considered as the control variables and where (x, y, z, β) is a flat output; note that the thrust F corresponds to a real control variable, whereas (p, q, r), the angular velocity vector, is a fictitious input, serving as an output for the subsystems described below, and indirectly controlled by their corresponding (lower level) inputs;
2. a second part, corresponding to the angular velocity dynamics, faster than the previous one, consisting of the 3 dimensional subsystem

[2] Since flatness implies dynamic feedback linearization, full linearization of the system (14.3)-(14.4) can be achieved by dynamic feedback (see Charlet et al. [1991]). However, the presence of different time scales makes the corresponding feedback ill-conditioned and the above cascaded controller is more adapted to this particular structure.

$$(\delta_l, \delta_m, \delta_n) \longrightarrow (p, q, r) \tag{14.7}$$

where the positions of the deflection surfaces $(\delta_l, \delta_m, \delta_n)$ play the role of control variables that can be used in order to realize the control requirements for (p, q, r);

3. finally a third part, corresponding to the servoactuator loop, much faster than the two previous ones, constituted by the 3 dimensional local loop

$$(\tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n) \longrightarrow (\delta_l, \delta_m, \delta_n) \tag{14.8}$$

where $(\tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n)$ are the real control variables, that can be used, indeed, to realize the requirements on the fictitious control variables $(\delta_l, \delta_m, \delta_n)$ of the intermediate subsystem (14.7).

More precisely, denoting by ξ the state vector $(x, y, z, V, \gamma, \chi, \alpha, \beta, \mu, F)$ of the subsystem described by (14.6) and by ζ the vector $(x, \dot{x}, \ddot{x}, y, \dot{y}, \ddot{y}, z, \dot{z}, \ddot{z}, \beta)$, it results from flatness and (14.6) that ξ and ζ are diffeomorphic and that the first nine equations of the ideal system, with the dynamic extension $\dot{F} = F_1$, are of the form:

$$\begin{aligned} x^{(3)} &= X_0(\zeta) + X_1(\zeta) w \\ y^{(3)} &= Y_0(\zeta) + Y_1(\zeta) w \\ z^{(3)} &= Z_0(\zeta) + Z_1(\zeta) w \\ \dot{\beta} &= B_0(\zeta) + B_1(\zeta) w \end{aligned} \tag{14.9}$$

where $w = (p, q, r, F_1)'$ (the prime denoting transpose) and with the matrix $\Xi_1 = \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ B_1 \end{pmatrix}$ invertible (X_1, Y_1, Z_1 and B_1 are independent line vectors). Note that in $w = (\omega, F_1)'$, the derivative of the thrust F_1, is directly controlled, and $\omega = (p, q, r)'$, the angular velocity, is a fictitious control variable.

One can choose w in order to track a given sufficiently differentiable reference trajectory $(x_{ref}, y_{ref}, z_{ref}, \beta_{ref})$ with stability by feedback linearization:

$$\begin{aligned} X_0(\zeta) + X_1(\zeta) w^* &= x_{ref}^{(3)} - K_{x,0}(x - x_{ref}) - K_{x,1}(\dot{x} - \dot{x}_{ref}) \\ &\quad - K_{x,2}(\ddot{x} - \ddot{x}_{ref}) \\ Y_0(\zeta) + Y_1(\zeta) w^* &= y_{ref}^{(3)} - K_{y,0}(y - y_{ref}) - K_{y,1}(\dot{y} - \dot{y}_{ref}) \\ &\quad - K_{y,2}(\ddot{y} - \ddot{y}_{ref}) \\ Z_0(\zeta) + Z_1(\zeta) w^* &= z_{ref}^{(3)} - K_{z,0}(z - z_{ref}) - K_{z,1}(\dot{z} - \dot{z}_{ref}) \\ &\quad - K_{z,2}(\ddot{z} - \ddot{z}_{ref}) \\ B_0(\zeta) + B_1(\zeta) w^* &= \dot{\beta}_{ref} - K_\beta(\beta - \beta_{ref}) \end{aligned} \tag{14.10}$$

with the gains $K_{i,j}$, $i = x, y, z$, $j = 0, 1, 2$, such that the polynomials $s^3 + K_{i,2}s^2 + K_{i,1}s + K_{i,0}$ are Hurwitz, for $i = x, y, z$, and with $K_\beta > 0$.

Note that the positions, velocities and accelerations $(x, \dot{x}, \ddot{x}, y, \dot{y}, \ddot{y}, z, \dot{z}, \ddot{z})$ needed to close this loop can be obtained from gyroscopic measurements. Thus, if w^*, satisfying (14.10), is applied to the system, the closed loop system is

$$\begin{aligned} &x^{(3)} - x_{ref}^{(3)} = -K_{x,0}(x - x_{ref}) - K_{x,1}(\dot{x} - \dot{x}_{ref}) - K_{x,2}(\ddot{x} - \ddot{x}_{ref}) \\ &y^{(3)} - y_{ref}^{(3)} = -K_{y,0}(y - y_{ref}) - K_{y,1}(\dot{y} - \dot{y}_{ref}) - K_{y,2}(\ddot{y} - \ddot{y}_{ref}) \\ &z^{(3)} - z_{ref}^{(3)} = -K_{z,0}(z - z_{ref}) - K_{z,1}(\dot{z} - \dot{z}_{ref}) - K_{z,2}(\ddot{z} - \ddot{z}_{ref}) \\ &\dot{\beta} - \dot{\beta}_{ref} = -K_\beta(\beta - \beta_{ref}) \end{aligned} \tag{14.11}$$

and is exponentially stable.

In order to realize the desired angular velocity vector ω^* (recall that $w^* = (\omega^*, F_1^*)$), we have to compute $(\delta_l, \delta_m, \delta_n)$, and then $(\tilde{\delta}_l, \tilde{\delta}_m, \tilde{\delta}_n)$, to make ω fastly converge to ω^*.

We can rewrite the subsystem (14.7) as follows

$$\varepsilon\dot{\omega} = \Omega_0(\zeta, \omega, F) + \Omega_1(\zeta, \omega, F)\delta \tag{14.12}$$

where ε is the time scaling between the slow part (14.9) and the faster part corresponding to (14.7), where $\delta = (\delta_l, \delta_m, \delta_n)'$, and where Ω_1 is an invertible matrix for every (ζ, ω, F) in the flight domain. It suffices then to choose

$$\delta^* = \Omega_1^{-1}(\zeta, \omega, F)(\Lambda(\omega - \omega^*) - \Omega_0(\zeta, \omega, F)) \tag{14.13}$$

with Λ an arbitrary but otherwise stable matrix, to transform (14.12) into $\varepsilon\dot{\omega} = \Lambda(\omega - \omega^*)$, ensuring the desired fast convergence of ω to ω^*. Note that (14.13) is not a high-gain feedback. The fast rate of convergence results only from the time scaling, preserved by this feedback.

Finally, it remains to compute the real servoactuator input vector $\tilde{\delta}^* = (\tilde{\delta}_l^*, \tilde{\delta}_m^*, \tilde{\delta}_n^*)'$ that realizes the required deflection position δ^*. Rewriting (14.4) as

$$\varepsilon^2\dot{\delta} = A\delta + \tilde{\delta} \tag{14.14}$$

one can set

$$\tilde{\delta}^* = -A\delta^* \tag{14.15}$$

which completes our feedback design. This construction is justified in Martin [1992], using singular perturbation techniques (see also Theorems 3.8 and 3.9, and Kokotović et al. [1986], Lévine and Rouchon [1994], Tikhonov et al. [1980]).

Summing up, we have obtained

$$\tilde{\delta}^* = -A\Omega_1^{-1}(\zeta, \omega, F)\big(\Lambda(\omega - \omega^*) - \Omega_0(\zeta, \omega, F)\big) \tag{14.16}$$

with ω^* given by (14.10).

Finally, *the error between the real system*, with non ideal aerodynamic coefficients, *and the ideal one* may be seen as a *perturbation* that the above controller may naturally attenuate, as a consequence of the exponential stability.

Moreover, the control law defined by (14.10), (14.16) shows a good robustness versus modelling errors and perturbations if the time scalings are preserved by the choice of (not too high) gains. In particular, the feedback formulas involve several first order partial derivatives of the aerodynamic coefficients, which are not too accurately known, for instance in transsonic flight, but without noticeable consequences on the overall performance.

Let us insist on the fact that this cascaded controller, defined by (14.10), (14.13) and (14.16), and whose structure is displayed in figure 14.4, is defined for all reference trajectories that remain in the flight domain, and that its gains do not depend on the particular reference trajectory to follow.

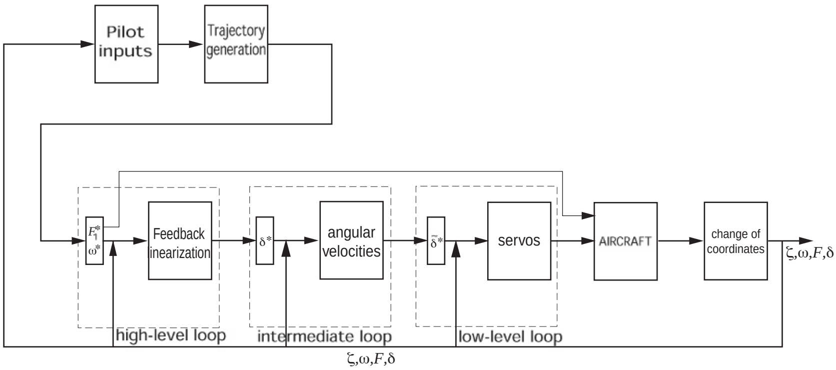

Fig. 14.4 The cascaded control architecture

14.0.4.5 The Pilot in the Loop

In most control systems, wether civil or military ones, the pilot prescribes the aircraft reference trajectory to follow. The pilot enters, for instance through the instrument panel, the direction towards which the aircraft must be pointing, its velocity, and various angles describing the aircraft's attitude, in particular the angle of attack $\alpha$and the side-slip angle β, as well as several rates describing the desired path. These indications are used to compute the next piece of trajectory to follow during a given interval of time.

The authority allocation between the pilot and the automatic controller remains one of the most difficult questions: *when and in which circumstances can the pilot enforce the autopilot to modify the current reference trajectory?* Clearly, this external loop might sometimes have destabilizing effects since if the pilot updates the trajectory at a faster rate than the aircraft's response, the servoactuators will have more and more difficulties to follow their reference and, the error increasing, they may soon be saturated, sometimes switching the controller to an emergency mode.

The structure of the above controller is well adapted to design this external loop for at least two reasons:

1. the pilot inputs interact with the internal loops only through the trajectory generator;
2. the time constants of the internal loops being fixed independently of the maneuver, the rate at which the pilot inputs are updated may be chosen slower than the slowest closed-loop time constant, to prevent the controller from injecting too fast dynamics in the slow ones, but otherwise fast enough compared to the pilot's reactions.

The reference trajectory is thus refreshed by computing at each step a new piece of reference trajectory, according to the pilot's indications, starting from the current aircraft flight conditions, and without interfering with the inner loops. For obvious reasons, the same technique may also be used in remote operation mode.

To conclude this section, let us note that such a controller might also provide important simplifications concerning failure detection aspects.

Comparable approaches may be used for other types of aircraft such as Vertical Take-Off and Landing aircraft Fliess et al. [1999] and helicopters (see Müllhaupt et al. [1997, 1999, 2008] for example).

References

R. Abraham and J. Marsden. *Foundations of Mechanics.* Benjamin/Cummings Pub. Co., Reading Mass., 2nd edition, 1978.

R. L. Anderson and N. H. Ibragimov. *Lie-Bäcklund Transformations in Applications.* SIAM, Philadelphia, 1979.

B. D' Andréa-Novel and J. Lévine. Modelling and nonlinear control of an overhead crane. In M.A. Kashoek, J.H. van Schuppen, and A.C.M. Rand, editors, *Robust Control of Linear and Nonlinear Systems, MTNS'89*, volume II, pages 523–529. Birkhäuser, Boston, 1990.

D.V. Anosov and V.I. Arnold. *Dynamical Systems, I.*, volume 1 of *Encyclopaedia of Mathematical Sciences.* Springer-Verlag, New York, 1980.

A.C. Antoulas. On canonical forms for linear constant systems. *Int. J. Control*, 33(1):95–122, 1981.

E. Aranda-Bricaire, C.H. Moog, and J.-B. Pomet. A linear algebraic framework for dynamic feedback linearization. *IEEE Trans. Automat. Contr.*, 40(1):127–132, 1995.

V.I. Arnold. *Équations Différentielles Ordinaires.* MIR, Moscou, 1974.

V.I. Arnold. *Chapitres Supplémentaires de la Théorie des Équations Différentielles Ordinaires.* MIR, Moscou, 1980.

D. Avanessoff and J.-B. Pomet. Flatness and Monge parameterization of two-input systems, control-affine with 4 states or general with 3 states. *ESAIM. Contrôle, Optimisation et Calcul des Variations*, 13(2):237–264, 2007.

L. Bers. On Hilbert's 22nd problem. In Browder F., editor, *Mathematical Developments Arising From Hilbert Problems, Proceedings of Symposia in Pure Mathematics*, pages 559–609. American Mathematical Society, Providence, Rhode Island, 1976.

D. Bestle and M. Zeitz. Canonical form design for nonlinear observers with linearizable error dynamics. *Int. J. of Control*, 23:419–431, 1981.

L. Bitauld, M. Fliess, and J. Lévine. A flatness based control synthesis of linear systems and application to windshield wipers. In *Proc. ECC'97, Brussels*, July 1997.

W. Boothby. *An introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press, New York, 1975.

P. Brunovský. A classification of linear controllable systems. *Kybernetica*, 6: 176–178, 1970.

J. Carr. *Application of Center Manifold Theory.* Springer, 1981.

É. Cartan. Sur l'équivalence absolue de certains systèmes d'équations différentielles et sur certaines familles de courbes. *Bull. Soc. Math. France*, 42:12–48, 1914. reédité in Oeuvres Complètes, part II, vol 2, pp. 1133–1168, CNRS, Paris, 1984.

H. Cartan. *Calcul Différentiel.* Hermann, Paris, 1967.

B. Charlet, J. Lévine, and R. Marino. On dynamic feedback linearization. *Systems & Control Letters*, 13:143–151, 1989.

B. Charlet, J. Lévine, and R. Marino. Sufficient conditions for dynamic state feedback linearization. *SIAM J. Control Optimization*, 29(1):38–57, 1991.

S.S. Chern, W.H. Chen, and K.S. Lam. *Lectures on Differential Geometry*, volume 1 of *Series on University Mathematics*. World Scientific, 2000.

V.N. Chetverikov. New flatness conditions for control systems. In *Proceedings of NOLCOS'01, St. Petersburg*, pages 168–173, 2001.

C. Chevalley. *Theory of Lie Groups*. Princeton University Press, 1946.

Y.K. Chin, A. Kade, J. Kowalik, and D. Graham. Electronic windshield wiper system i: modelling and validation, ii: control and sensitivity study. *Int. J. of Vehicle design*, 12(2):175–196, 1991.

Y. Choquet-Bruhat. *Géométrie différentielle et systèmes extérieurs*. Dunod, Paris, 1968.

P.M. Cohn. *Free Rings and Their Relations*. Academic Press, London, 1985.

O. Dahl. Path-constrained robot control with limited torques–experimental evaluation. *IEEE Trans. on Robotics and Automation*, 10(5):658–669, 1994.

O. Dahl and L. Nielsen. Torque-limited path following by on-line trajectory time scaling. *IEEE Trans. on Robotics and Automation*, 6(5):554–561, 1990.

J. De Dona, F. Suryawan, M. Seron, and J. Lévine. A flatness-based iterative method for reference trajectory generation in constrained NMPC. In D. Raimondo L. Magni and F. Allgöwer, editors, *Proc. of the International Workshop on Assessment and Future Directions of Nonlinear Model Predictive Control, NMPC'08*, Pavia, Italy, 2008.

E. Delaleau and V. Hagenmeyer. Commande prédictive non linéaire fondée sur la platitude différentielle. In D. Dumur, editor, *La commande prédictive : Avancées et perspectives*. Hermès, 2006.

E. Delaleau and J. Rudolph. Control of flat systems by quasi-static feedback of generalized states. *Int. J. Control*, 71(5):745–765, 1998.

M. Demazure. *Bifurcations and Catastrophes, Geometry of Solutions to Nonlinear Problems*. Springer, Universitext, 2000.

Th. Devos. *Étude et comparaison de plusieurs lois de commande de grues*. PhD thesis, Université Paris-Sud, 2009.

Th. Devos and J. Lévine. A flatness-based nonlinear predictive approach for crane control. In M. Alamir and F. Allgöwer, editors, *Proc. of the IFAC Conf. on Nonlinear Model Predictive Control for Fast Systems, NMPC'06*, Grenoble, France, 2006.

J. Dieudonné. *Fondements de l'Analyse Moderne*. Gauthier-Villars, Paris, 1960.

B. Etkin. *Dynamics of flight-Stability and control*. John Wiley, 2nd edition, 1982.

N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *J. Diff. Equations*, 31:53–98, 1979.

A.F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Kluwer Academic Publishers, Dordrecht, Boston, London, 1988.

R. Findeisen and F. Allgöwer. An introduction to nonlinear model predictive control. In *Proc. 21st Benelux Meeting on Systems and Control*, Veldhoven, The Netherlands, 2002.

M. Fliess. Some basic structural properties of generalized linear systems. *Systems & Control Letters*, 15:391–396, 1990.

M. Fliess. A remark on Willems' trajectory characterization of linear controllability. *Systems & Control Letters*, 19:43–45, 1992.

M. Fliess and R. Marquez. Continuous-time linear predictive control and flatness : a module-theoretic setting with examples. *International Journal of Control*, 73:606–623, 2000.

M. Fliess, J. Lévine, and P. Rouchon. A simplified approach of crane control via a generalized state-space model. In *Proc. 30th IEEE Control Decision Conf., Brighton*, pages 736–741, 1991.

M. Fliess, J. Lévine, Ph. Martin, and P. Rouchon. Sur les systèmes non linéaires différentiellement plats. *C.R. Acad. Sci. Paris*, I–315:619–624, 1992a.

M. Fliess, J. Lévine, Ph. Martin, and P. Rouchon. On differentially flat nonlinear systems. In *Proc. IFAC-Symposium NOLCOS'92, Bordeaux*, pages 408–412, 1992b.

M. Fliess, J. Lévine, and P. Rouchon. A generalized state variable representation for a simplified crane description. *Int. J. Control*, 58:277–283, 1993.

M. Fliess, J. Lévine, Ph. Martin, and P. Rouchon. Flatness and defect of nonlinear systems: introductory theory and examples. *Int. J. Control*, 61 (6):1327–1361, 1995.

M. Fliess, J. Lévine, Ph. Martin, and P. Rouchon. A Lie-Bäcklund approach to equivalence and flatness of nonlinear systems. *IEEE Trans. Automat. Control*, 44(5):922–937, 1999.

E. Fossas and J. Franch. Linearization by prolongations: a new bound for three input systems. In *Proceedings of the 14th IFAC World Congress, Beijing*, 1999.

J. Franch. *Flatness, Tangent Systems and Flat Outputs.* PhD thesis, Universitat Politècnica de Catalunya Jordi Girona, 1999.

F.R. Gantmacher. *Théorie des Matrices, t. I, II.* Dunod, Paris, 1966.

R.B. Gardner and W.F. Shadwick. The GS algorithm for exact linearization to Brunovsky normal form. *IEEE. Trans. Automat. Contr.*, 37:224–230, 1992.

C. Godbillon. *Géométrie Différentielle et Mécanique Analytique.* Hermann, 1969.

J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* Springer-Verlag, New York, 1983.

V. Hagenmeyer and E. Delaleau. Continuous-time non-linear flatness-based predictive control: an exact feedforward linearisation setting with an induction drive example. *International Journal of Control*, 81(10):1645—1663, 2008.

D. Hilbert. Über den Begriff der Klasse von Differentialgleichungen. *Math. Ann.*, 73:95–108, 1912. Also in Gesammelte Abhandlungen, Vol. 3, pages 81–93, Chelsea, New York, 1965.

D. Hilbert. Mathematische probleme. *Archiv für Mathematik und Physik*, 1: 44–63 and 213–237, 1901. Also in Gesammelte Abhandlungen, Vol.3, pages 290–329, Chelsea, New York, 1965.

M.W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems and Linear Algebra.* Acamedic Press, New-York, 1974.

M.W. Hirsch, C.C. Pugh, and M. Shub. *Invariant Manifolds.* Lecture Notes in Mathematics. Springer, 1977.

L.R. Hunt, R. Su, and G. Meyer. Design for multi-input nonlinear systems. In R.W. Brockett, R.S. Millman, and H.J. Sussmann, editors, *Differential Geometric Control Theory*, pages 268–298. Birkhäuser, Boston, 1983a.

L.R. Hunt, R. Su, and G. Meyer. Global transformations of nonlinear systems. *IEEE Trans. Automat. Control*, 28:24–31, 1983b.

N. H. Ibragimov. *Transformation Groups Applied to Mathematical Physics.* Reidel, Boston, 1985.

A. Isidori. *Nonlinear Control Systems.* Springer, New York, 3rd edition, 1995.

B. Jakubczyk. Invariants of dynamic feedback and free systems. In *Proc. ECC'93, Groningen*, pages 1510–1513, 1993.

B. Jakubczyk and W. Respondek. On linearization of control systems. *Bull. Acad. Pol. Sci. Ser. Sci. Math.*, 28(9–10):517–522, 1980.

T. Kailath. *Linear Systems.* Prentice Hall, Englewood Cliffs, NJ, 1980.

H. K. Khalil. *Nonlinear Systems.* Prentice Hall, Englewood Cliffs, NJ, 1996.

B. Kiss. *Planification de trajectoires et commande d'une classe de systèmes mécaniques plats et liouvilliens.* PhD thesis, École des Mines de Paris, 2001.

B. Kiss, J. Lévine, and Ph. Müllhaupt. Modelling, flatness and simulation of a class of cranes. *Periodica Polytechnica*, 43(3):215–225, 1999.

B. Kiss, J. Lévine, and Ph. Müllhaupt. A simple output feedback PD controller for nonlinear cranes. In *39th IEEE Conference on Decision and Control, Sydney, Australia*, 2000a.

B. Kiss, J. Lévine, and Ph. Müllhaupt. Modelling and motion planning for a class of weight handling equipment. *J. Systems Science*, 26(4):79–92, 2000b.

B. Kiss, J. Lévine, and Ph. Müllhaupt. Control of a reduced size model of US Navy crane using only motor position sensors. In A. Isidori, F. Lamnabhi-Lagarrigue, and W. Respondek, editors, *Nonlinear Control in the Year 2000*, volume 2, pages 1–12. Springer, 2000c.

S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry*, volume I. John Wiley & Sons, 1996.

P.V. Kokotović, H.K. Khalil, and J. O'Reilly. *Singular Perturbation Methods in Control: Analysis and Design.* Academic Press, London, 1986.

M.V. Kondratieva, A.V. Mikhalev, and E.V. Pankratiev. On Jacobi's bound for systems of differential polynomials. *Algebra, Moscow Univ. Press, Moscow*, pages 79–85, 1982. in russian.

A.I. Kostrikin and I.R. Shafarevich. *Algebra, I.*, volume 11 of *Encyclopaedia of Mathematical Sciences.* Springer-Verlag, New York, 1980.

I. S. Krasil'shchik, V. V. Lychagin, and A. M. Vinogradov. *Geometry of Jet Spaces and Nonlinear Partial Differential Equations.* Gordon and Breach, New York, 1986.

A.J. Krener and A. Isidori. Linearization by output injection and nonlinear observers. *Systems & Control Lett.*, 3:47–52, 1983.

A.J. Krener and W. Respondek. Nonlinear observers with linear error dynamics. *SIAM J. Control Optimiz.*, 23:197–216, 1985.

M. Krstić, I. Kanellakopoulos, and P. Kokotović. *Nonlinear and Adaptive Control Design.* Wiley, New York, 1995.

J.P. LaSalle and S. Lefschetz. *Stability by Lyapunov's Direct Method With Application.* Academic Press, New York, 1961.

J. Lévine. Are there new industrial perspectives in the control of mechanical systems? In P.M. Frank, editor, *Advances in Control, Highlights of ECC'99*, pages 197–226. Springer, London, 1999.

J. Lévine. Static and dynamic state feedback linearization. In A. Fossard and D. Normand-Cyrot, editors, *Nonlinear Systems*, volume 3, pages 93–126. Chapman & Hall, 1997.

J. Lévine. On necessary and sufficient conditions for differential flatness. *http://www.arxiv.org*, arXiv:math.OC/0605405, 2006.

J. Lévine. On the synchronization of a pair of independent windshield wipers. *IEEE Trans. Control Systems Technology*, 12(5):787–795, 2004.

J. Lévine. On necessary and sufficient conditions for differential flatness. In *Proc. of IFAC NOLCOS 2004 Conference*, Stuttgart, 2004.

J. Lévine and D.V. Nguyen. Flat output characterization for linear systems using polynomial matrices. *Systems & Control Letters*, 48:69–75, 2003.

J. Lévine and P. Rouchon. Quality control of binary distillation columns based on nonlinear aggregated models. *Automatica*, 27(3):463–480, 1991.

J. Lévine and P. Rouchon. An invariant manifold approach for robust control design and applications. In U. Helmke, R. Mennicken, and J. Saurer, editors, *Systems and Networks: Mathematical Theory and Applications, Proc. MTNS-93, Regensburg, Germany*, volume 2, pages 309–312. Akademie Verlag, Berlin, 1994.

J. Lévine, J. Lottin, and J. C. Ponsart. A nonlinear approach to the control of magnetic bearings. *IEEE Trans. Control Systems Technology*, 4(5):524–544, 1996.

J. Lévine, P. Rouchon, G. Yuan, C. Grebogi, B.R. Hunt, E. Kostelich, E. Ott, and J. Yorke. On the control of US navy cranes. In *Proc. ECC'97, Brussels, paper N. 717*, 1997.

A. Liapounoff. *Problème Général de la Stabilité du Mouvement.* Annales de la Faculté des Sciences de Toulouse, deuxième série. Edouard Privat, 1907. reprint J. Gabay, Paris, 1988.

R. Marino. On the largest feedback linearizable subsystem. *Systems & Control Letters*, 6:345–351, 1986.

R. Marino and P. Tomei. *Nonlinear Control Design.* Prentice Hall, Englewood Cliffs, NJ, 1995.

Ph. Martin. Aircraft control using flatness. In *Proc. CESA Multiconference, Lille*, pages 194–199, 1996.

Ph. Martin. Endogenous feedbacks and equivalence. In U. Helmke, R. Mennicken, and J. Saurer, editors, *Systems and Networks: Mathematical Theory and Applications, Proc. MTNS-93, Regensburg, Germany*, volume 2, pages 343–346. Akademie Verlag, Berlin, 1994.

Ph. Martin. *Contribution à l'étude des systèmes diffèrentiellement plats.* PhD thesis, École des Mines de Paris, 1992.

Ph. Martin and P. Rouchon. Systems without drift and flatness. In *Proc. MTNS 93, Regensburg, Germany*, August 1993.

Ph. Martin, R.M. Murray, and P. Rouchon. Flat systems. In G. Bastin and M. Gevers, editors, *Plenary Lectures and Minicourses, Proc. ECC 97, Brussels*, pages 211–264, 1997.

D. McLean. *Automatic Flight Control Systems.* Prentice Hall, 1990.

D.T. McRuer, I.L. Ashkenas, and D.C. Graham. *Aicraft Dynamics and Automatic Control.* Princeton University Press, 1973.

J. Milnor. *Topology from the Differential Viewpoint.* University Press of Virginia, Charlottesville, 1978.

M. Morari and J.H. Lee. Model predictive control: past, present and future. *Computers & Chemical Engineering*, 23:667–682, 1999.

P. Müllhaupt. *Introduction à l'Analyse et à la Commande des Systèmes Non Linéaires.* Presses Polytechniques et Universitaires Romandes, Lausanne, Switzerland, 2009.

P. Müllhaupt, B. Srinivasan, J. Lévine, and D. Bonvin. A toy more difficult to control than the real thing. In *Proc. ECC'97, Brussels*, July 1997.

P. Müllhaupt, B. Srinivasan, J. Lévine, and D. Bonvin. Cascade control of the toycopter. In *Proc. ECC'99, Karslruhe*, 1999.

P. Müllhaupt, B. Srinivasan, J. Lévine, and D. Bonvin. Control of the toycopter using a flat approximation. *IEEE Trans. Control Systems Technology*, 16(5):882–896, 2008.

Ju.I. Neĭmark and N.A. Fufaev. *Dynamics of Nonholonomic Systems.* American Mathematical Society, Providence, Rhode Island, 1972.

H. Nijmeijer and A.J. van der Schaft. *Nonlinear Dynamical Control Systems.* Springer, New York, 1990.

F. Ollivier. Standard bases of differential ideals. In *Proceedings of AAECC8*, volume 508 of *Lecture Notes In Computer Science*, pages 304–321. Springer, 1990.

F. Ollivier and S. Brahim. La borne de Jacobi pour une diffiété définie par un système quasi régulier (Jacobi's bound for a diffiety defined by a quasi-regular system). *Comptes rendus Mathematique*, 345(3):139–144, 2007.

P.J. Olver. *Equivalence, Invariants, and Symmetry.* Cambridge University Press, 1995.

P. S. Pereira da Silva. Flatness of nonlinear control systems : a Cartan-Kähler approach. In *Proc. Mathematical Theory of Networks and Systems (MTNS'2000), Perpignan*, pages 1–10, 2000.

P.S. Pereira da Silva and C. Corrêa Filho. Relative flatness and flatness of implicit systems. *SIAM J. Control and Optimization*, 39(6):1929–1951, 2001.

N. Petit, Y. Creff, L. Lemaire, and P. Rouchon. Minimum time constrained control of acid strength on a sulfuric acid alkylation unit. *Chemical Engineering Science*, 56:2767–2774, 2001.

F. Pham. *Géométrie et Calcul Différentiel sur les Variétés.* InterEditions, Paris, 1992.

H. Poincaré. Sur l'uniformisation des fonctions analytiques. *Acta Mathematica*, 31:1–63, 1907. Also in Œuvres de Henri Poincaré, t. 4, pages 70–139, Gauthier-Villars, Paris, 1950.

J.W. Polderman and J.C. Willems. *Introduction to Mathematical System Theory: a Behavioral Approach.* Springer-Verlag, Berlin, 1997.

J.-B. Pomet. A differential geometric setting for dynamic equivalence and dynamic linearization. In B. Jakubczyk, W. Respondek, and T. Rzežuchowski, editors, *Geometry in Nonlinear Control and Differential Inclusions*, pages 319–339. Banach Center Publications, Warsaw, 1993.

J.-B. Pomet. On dynamic feedback linearization of four-dimensional affine control systems with two inputs. *ESAIM-COCV*, 1997. http://www.emath.fr/Maths/Cocv/Articles/articleEng.html.

J.B. Pomet, C. Moog, and E. Aranda. A non-exact Brunovsky form and dynamic feedback linearization. In *Proc. 31st. IEEE Conf. Decision Cont.*, pages 2012–2017, 1992.

J.F. Pommaret. *Partial Differential Control Theory.* Kluwer academic publishers, 2001.

J.F. Pommaret and A. Quadrat. Localization and parametrization of linear multidimensional control systems. *Systems & Control Letters*, 37:247–269, 1999.

L. Pontriaguine. *Équations différentielles Ordinaires.* MIR, Moscou, 1975.

M. Rathinam and R.M. Murray. Configuration flatness of Lagrangian systems underactuated by one control. *SIAM J. Control and Optimization*, 36(1): 164–179, 1998.

J. F. Ritt. Jacobi's problem on the order of a system of differential equations. *Annals of Mathematics*, 36:303–312, 1935.

H.H. Rosenbrock. *Multivariable and State-Space Theory.* Wiley, New York, 1970.

P. Rouchon. Necessary condition and genericity of dynamic feedback linearization. *J. Math. Systems Estim. Control*, 4(2):257–260, 1994.

J. Rudolph. *Flatness based control of distributed parameter systems.* Shaker Verlag, Aachen, 2003.

J. Rudolph, J. Winkler, and F. Woittenek. *Flatness based control of distributed parameter systems: examples and computer exercises from various technological domains.* Shaker Verlag, Aachen, 2003.

D. Ruelle. *Elements of Differentiable Dynamics and Bifurcation Theory.* Academic Press, 1989.

S. Sastry. *Nonlinear Systems, Analysis, Stability and Control.* Springer, New York, 1999.

J.A. Saunders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems.* Springer, 1987.

K. Schlacher and M. Schöberl. Construction of flat outputs by reduction and elimination. In *Proc. 7th IFAC Symposium on Nonlinear Control Systems, Pretoria, South Africa*, pages 666–671, August 2007.

W.F. Shadwick. Absolute equivalence and dynamic feedback linearization. *Systems Control Letters*, 15:35–39, 1990.

H. Sira-Ramirez and S.K. Agrawal. *Differentially flat systems.* Marcel Dekker, New York, 2004.

J.J.E. Slotine and W. Li. *Applied Nonlinear Control.* Prentice Hall, Englewood Cliffs, NJ, 1991.

W.M. Sluis. A necessary condition for dynamic feedback linearization. *Systems Control Letters*, 21:277–283, 1993.

W.M. Sluis and D.M. Tilbury. A bound on the number of integrators needed to linearize a control system. *Systems & Control Letters*, 29(1):43–50, 1996.

R. Sommer. Control design for multivariable non-linear time-varying systems. *Int. J. Control*, 31:883–891, 1980.

E.D. Sontag. *Mathematical Control Theory, Deterministic Finite Dimensional Systems.* Springer, New York, 2nd edition, 1998.

H.J. Sussmann. A general theorem on local controllability. *SIAM J. Control and Optimiz.*, 25:158–194, 1987.

A. Tannenbaum. *Invariance and System Theory: Algebraic and Geometric Aspects.* Springer, New York, 1980.

R Thom. *Stabilité Structurelle et Morphogénèse.* InterEdition, Paris, 2nd edition, 1977.

A. Tikhonov, A. Vasil'eva, and A. Sveshnikov. *Differential Equations.* Springer, New York, 1980.

H.L. Trentelman. On flat systems, behaviors and observable image representations. *Systems & Control Letters*, 21:51–55, 2004.

M. van Nieuwstadt, M. Rathinam, and R.M. Murray. Differential flatness and absolute equivalence. In *Proc. 33th IEEE Conf. Decision Control, Lake Buena Vista, Fl.*, pages 326–332, 1994.

M. van Nieuwstadt, M. Rathinam, and R.M. Murray. Differential flatness and absolute equivalence of nonlinear control systems. *SIAM J. Control Optim.*, 36(4):1225–1239, 1998.

M. Vidyasagar. *Nonlinear Systems Analysis.* Prentice Hall, Englewood Cliffs, NJ, 2nd edition, 1993.

J. von Löwis, J. Rudolph, J. Thiele, and F. Urban. Flatness-based trajectory tracking control of a rotating shaft. In *7th International Symposium on Magnetic Bearings, Zürich*, pages 299–304, 2000.

M. Vukobratović and R. Stojić. *Modern Aircraft Flight Control.* Springer-Verlag, 1988.

J.C. Wanner. *Dynamique du vol et pilotage des avions.* Technical report, École Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, 1984.

E.T. Whittaker. *A Treatise on the Analytical Dynamics of Particules and Rigid Bodies.* Cambridge University Press, Cambridge, 4th edition, 1937.

W.A. Wolovich. *Linear Multivariable Systems*, volume 11 of *Series in Applied Mathematical Sciences.* Springer, New York, 1974.

W.M. Wonham. *Linear Multivariable Control: a Geometric Approach.* Springer, 1974.

X.H. Xia and W.B. Gao. Nonlinear observer design by observer error linearization. *SIAM J. Control Optimiz.*, 27:199–216, 1989.

V.V. Zharinov. *Geometrical Aspect of Partial Differential Equations.* World Scientific, Singapore, 1992.

Index