

## INVERSION IN INDIRECT OPTIMAL CONTROL OF MULTIVARIABLE SYSTEMS

FRANÇOIS CHAPLAIS<sup>1</sup> AND NICOLAS PETIT<sup>2</sup>

**Abstract.** This paper presents the role of vector relative degree in the formulation of stationarity conditions of optimal control problems for affine control systems. After translating the dynamics into a normal form, we study the Hamiltonian structure. Stationarity conditions are rewritten with a limited number of variables. The approach is demonstrated on two and three inputs systems, then, we prove a formal result in the general case. A mechanical system example serves as illustration.

**Mathematics Subject Classification.** 34C20, 34H05, 49K15, 93C10, 93C35.

Received February 2, 2006. Revised September 15, 2006.  
Published online October 13, 2007.

### 1. INTRODUCTION

Geometric tools of nonlinear control theory [22,31] have long been used for feedback linearization of control-affine systems. The induced changes of variables readily solve the inverse problems of computing inputs corresponding to a prescribed behavior of outputs. In the context of these inverse problems, trajectory optimization is often important, especially in applications. For that purpose, two families of numerical techniques are commonly used (see [42]). The direct methods imply a discretization of the optimal control problem, yielding a nonlinear program (NLP). On the other hand, indirect methods (a.k.a. adjoint methods) are based on the solution of necessary conditions for optimality, as derived by the calculus of variations. While direct methods have been the workhorse of control engineers [6,7,20,21], indirect methods are usually reported to produce higher accuracy solutions, although being relatively instable. Both approaches can be cascaded to take advantages of these properties (see [11,37,38,42]).

Inversion has lately been used in direct methods of numerical optimal control. The numerical impact of the relative degree (as defined in [22]) of the output chosen to cast the optimal control problem into a NLP was emphasized in [25,33]. Given the system dynamics and an optimal cost, it was shown how to take advantage of the geometric structure of the dynamics to reduce the dimensionality of a numerical collocation scheme. In general collocation methods, coefficients are used to approximate with basis functions both states and inputs [21]. While it was known since [36] that it is numerically efficient to eliminate the control, it was emphasized in [25,33] that it is possible to reduce the problem further. Choosing outputs with maximum relative degrees is the key to

---

*Keywords and phrases.* Optimal control, inversion, adjoint states, normal form.

<sup>1</sup> Centre Automatique et Systèmes, École Nationale Supérieure des Mines de Paris, 35 rue Saint-Honoré, 77305 Fontainebleau Cedex, France; [chaplais@cas.ensmp.fr](mailto:chaplais@cas.ensmp.fr)

<sup>2</sup> Centre Automatique et Systèmes, École Nationale Supérieure des Mines de Paris, 60 bd Saint-Michel, 75272 Paris Cedex 06, France; [petit@cas.ensmp.fr](mailto:petit@cas.ensmp.fr)

efficient variable elimination that lowers the number of required coefficients (see for example [30]). In differential equations, constraints, and cost functions, unnecessary variables are substituted with successive derivatives of the chosen outputs. For that reason, it is a smart choice to represent these outputs with basis functions that can be easily differentiated. A prime example are B-splines functions as in the software package NTG [26]. Other possibilities, such as Legendre pseudospectral differentiation, or Chebyshev approximations of the derivatives are presented in [35] and [15,16] respectively. When combined to a NLP solver (such as NPSOL [20] for instance), this can induce drastic speed-ups in numerical solving [2,26,29,40]. The best-case scenario is full feedback linearization (as implied by flatness [18,19]) considered theoretically, numerically in [17,32,34,39,41], and in practice in [27].

In this paper, we focus on indirect methods. In this framework, we show how to use the geometric structure of the dynamics. In [14], we addressed the case of single-input single-output (SISO) systems with a  $n$ -dimensional state. We emphasized that  $r$  the relative degree of the primal system also plays a role in the adjoint (dual) dynamics. The two-point boundary value problem (TPBVP) can be rewritten by eliminating many variables. Only  $n - r$  variables are required. In the case of full feedback linearisability, the primal and adjoint dynamics take the form of a  $2n$ -degree differential equation in a single variable: the linearizing output. The adjoint variables are computed and eliminated. In this paper, we address the general case of multi-inputs multi-outputs (MIMO) systems. We note  $n$  the dimension of the state,  $m$  the number of inputs, and  $r$  the total relative degree (see Def. 1). The proposed results encompass the SISO case but, not surprisingly, requires more in-depth investigations of stationarity conditions. The system under consideration may only be partly feedback linearizable (*i.e.* may have a zero dynamics). The main contribution of the paper is the derivation of a  $2n$  dimensional necessary state space form equation for the primal and adjoint dynamics using a reduced number of variables ( $m + 2(n - r)$ ). Adjoint states corresponding to the linearizable part of the dynamics are explicitly computed and eliminated from stationarity conditions.

The article is organized as follows. In Section 2, we present an introductory example to stress main noticeable points and motivate our approach. The classic forced van der Pol system is considered. Adjoint variables are eliminated from the TPBVP. After a numerical resolution with a standard software package, the adjoint variables are analytically recovered. This yields a straightforward computation of neighboring extremals (as defined in [9]) and provides answers to post optimal analysis. Interestingly, these results do not really depend on the numerical method used to compute primal variables optimal trajectories: both direct and indirect methods can be used. In any case, the adjoint variables can be derived through stationarity conditions. The main method of adjoint variables recovery and elimination is illustrated by this simple example: the key is to iteratively differentiate the stationarity relation of the Hamiltonian with respect to the control variables. Numerical experiments are reported to illustrate the relevance of our approach. More generally, in Section 3, we define a Lagrange optimal control problem for which we aim at proving a general result. Recalling the Weak Minimum Principle, we detail stationarity conditions involving high derivative orders of linearizing outputs. We use a normal form obtained by feedback linearization. Elimination of adjoint variables corresponding to the linear part of the normal form is explained in Lemma 1. Substitutions in stationarity conditions of the Hamiltonian yield high order necessary differential equations for the linearizing outputs. These give Theorem 2. Further, it is possible to lower orders of this set of necessary differential equations by more in-depth investigations. Sequentially, elimination of variables is performed in Section 4. Eventually, this procedure yields the desired  $2n$  dimensional state space form system in  $m + 2(n - r)$  variables. To provide a direct reading of the proposed approach, two cases of practical interest are detailed (two and three inputs systems respectively). We address the general case in Theorem 3. In Section 5, we illustrate the proposed approach with a mechanical system example. Finally, we give conclusions and future directions of our work in Section 6.

## 2. INTRODUCTORY EXAMPLE AND MOTIVATION

### 2.1. Three approaches to an example from the literature

In this section, we want to stress some noticeable points in optimal control problems that can be rewritten under a normal form. We consider the classic forced van der Pol Problem that served as a benchmark problem in [15,26,34]. The dynamics of this single input cascade system is

$$\dot{x}_1 = x_2 \tag{1}$$

$$\dot{x}_2 = -x_1 + (1 - x_1^2)x_2 + u. \tag{2}$$

The optimal control problem we consider is the following

**Problem 1.** Minimize the quadratic cost function  $J = \frac{1}{2} \int_0^5 (x_1^2(s) + x_2^2(s) + u^2(s)) ds$  subject to the dynamics (1,2) and the endpoint constraints  $x_1(0) = 1, x_2(0) = 0, x_2(5) - x_1(5) = 1$ .

The dynamics is flat, and feedback linearizable by static feedback. Indeed,  $z_1 = x_1$  is a linearizing output. This means that the state variables and the control write in terms of  $x_1$  and its derivatives. Here, we have  $x_2 = \dot{x}_1$ , and  $u = \ddot{x}_1 + x_1 - (1 - x_1^2)\dot{x}_1$ . There is a one-to-one relationship between the trajectories of the system  $t \mapsto (x_1, x_2, u)(t)$  and the trajectories of  $t \mapsto x_1(t)$  through these last relations. A first possibility to solve problem 1 by taking advantage of this trajectory correspondance is to follow the approach presented in [26]: cast the optimal control problem into a high order problem in the  $x_1$  variable. This leads to the following solution method.

**Solution method 1** [26]. Use a collocation method to minimize the cost function  $J = \frac{1}{2} \int_0^5 (x_1^2(s) + (\dot{x}_1)^2(s) + (\ddot{x}_1 + x_1 - (1 - x_1^2)\dot{x}_1)^2(s)) ds$  subject to the endpoint constraints  $x_1(0) = 1, \dot{x}_1(0) = 0, \dot{x}_1(5) - x_1(5) = 1$ . Finally, recover  $u = \ddot{x}_1 + x_1 - (1 - x_1^2)\dot{x}_1$  once the optimal trajectory  $x_1$  is found.

This solution method can be implemented in a very efficient algorithm, because only a single variable is used. The trajectories of  $x_1$  are chosen in a set of B-splines functions. The derivatives are analytically computed and the cost function is approximated by quadrature formulas. This leads to a nonlinear programming problem that can be solved by a NLP software package (such as NPSOL [20]).

We can also use an indirect approach and derive a two point boundary value problem (TPBVP). For that purpose, we note the Hamiltonian

$$H = \frac{1}{2}(x_1^2 + x_2^2 + u^2) + \lambda_1 x_2 + \lambda_2(-x_1 + (1 - x_1^2)x_2 + u)$$

and derive the adjoint equations

$$\dot{\lambda}_1 = -\frac{\partial H}{\partial x_1} = -(x_1 - \lambda_2 - 2\lambda_2 x_1 x_2) \tag{3}$$

$$\dot{\lambda}_2 = -\frac{\partial H}{\partial x_2} = -(x_2 + \lambda_1 + \lambda_2(1 - x_1^2)). \tag{4}$$

This approach yields the following solution

**Solution method 2.** Solve the four dimensional TPBVP (1,2,3,4) with boundary conditions  $x_1(0) = 1, x_2(0) = 0, x_2(5) - x_1(5) = 1, \lambda_1(5) + \lambda_2(5) = 0$ . Recover  $u$  from the relation  $\frac{\partial H}{\partial u} = 0$ , i.e.  $u = -\lambda_2$ , once the optimal trajectory is found.

Here, numerous numerical approaches can be used among which are collocation (as in the Matlab routine `bvp4c`), and shooting techniques (see [10] for instance). In that solution method, we have 4 unknowns. Yet,

it is possible to derive another TPBVP involving only a single variable. Solving the two stationarity equations  $\frac{\partial H}{\partial u} = 0$  and  $\frac{d}{dt} \frac{\partial H}{\partial u} = 0$ , we get analytic expressions for the adjoint variables in terms of  $x_1$  and its derivatives

$$\lambda_1 = x_1^{(3)} + 2x_1\dot{x}_1^2 + (x_1 - (1 - x_1^2)\dot{x}_1)(1 - x_1^2) \tag{5}$$

$$\lambda_2 = -x_1^{(2)} - x_1 + (1 - x_1^2)\dot{x}_1. \tag{6}$$

Then, one can substitute the expression for  $\lambda_1$  and  $\lambda_2$  into the differential equation (3) to get a fourth order differential equation to be satisfied by the linearizing output. Eventually, rewriting the boundary conditions of Problem 1 yields the following approach.

**Solution method 3.** Solve the fourth order TPBVP

$$x_1^{(4)} = -2\dot{x}_1^3 - 6x_1\dot{x}_1\ddot{x}_1 - (2x_1\dot{x}_1^2 - (1 - x_1^2)\ddot{x}_1)(1 - x_1^2) - 2x_1 - \ddot{x}_1$$

with boundary conditions

$$\begin{aligned} x_1(0) = 1, \quad \dot{x}_1(0) = 0, \quad (\dot{x}_1 - x_1)(5) = 1 \\ (x_1^{(3)} - x_1^3 + 2x_1\dot{x}_1^2 + (1 - x_1^2)x_1^2\dot{x}_1 - \ddot{x}_1)(5) = 0. \end{aligned}$$

Then, recover  $u = \ddot{x}_1 + x_1 - (1 - x_1^2)\dot{x}_1$  once the solution is found.

So far, we have reduced the number of unknown variables to 1. Instead of  $x_1, x_2, \lambda_1, \lambda_2$  and  $u$ , only  $x_1$  needs to be considered both in the differential equations to be satisfied by optimal trajectories, and in the boundary conditions. Solving this last TPBVP can be done with many numerical solvers, *e.g.* the above mentioned Matlab routine `bvp4c`. Results are comparable to those obtained with solution method 1 in [26] and [34]. The obtained cost is 1.68568.

### 2.2. Post-optimal analysis

Now, it is interesting to notice that no matter which solution technique (1 or 2 or 3) is used to compute optimal trajectories, we can recover the optimal adjoint variables histories through (5,6) without any kind of integration or differential equation solving. The reason for this is that the adjoint states  $\lambda_1$  and  $\lambda_2$  write in terms of the linearizing output through (5) and (6).

From here, post-optimal analysis and neighboring extremal computations can be performed by considering the following time-varying matrices

$$\begin{aligned} f_x = \begin{pmatrix} 0 & 1 \\ -1 - 2x_1x_2 & 1 - x_1^2 \end{pmatrix}, \quad f_u = (0 \ 1)^T, \quad H_{uu} = 1, \quad H_{ux} = 0, \quad H_{xx} = \begin{pmatrix} 1 - 2\lambda_2x_2 & -2x_1\lambda_2 \\ -2x_1\lambda_2 & 1 \end{pmatrix}, \\ A = f_x, \quad B = f_u H_{uu}^{-1} f_u^T, \quad C = H_{xx}. \end{aligned}$$

Consider a perturbation in the state  $\delta x$  and a revision in the terminal condition  $\delta \psi$ . The perturbed TPBVP dynamic is

$$\begin{aligned} \dot{\delta x} &= A(t)\delta x - B(t)\delta \lambda \\ \dot{\delta \lambda} &= -C(t)\delta x - A^T(t)\delta \lambda. \end{aligned}$$

This can be readily solved by a backward sweep method (see [9]) for which we must consider

$$\dot{S} = SBS - SA - A^T S - C, \quad \dot{R} = -A^T R + SBR, \quad \dot{Q} = R^T BR$$

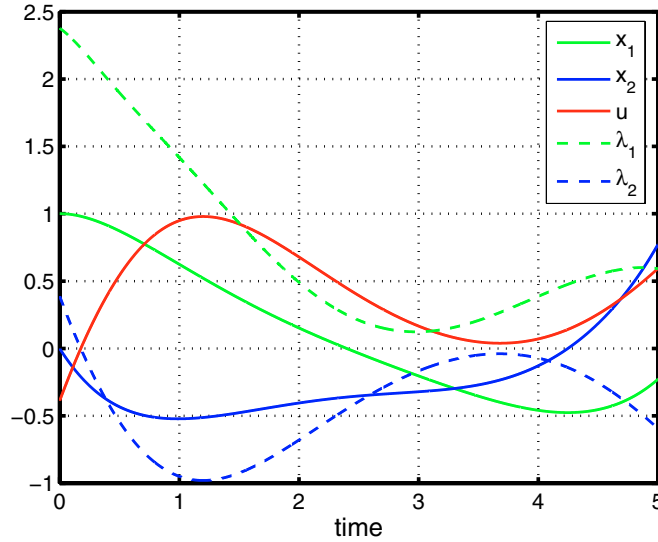


FIGURE 1. Solution to an optimal control problem for the van der Pol system. Optimal state, control and adjoint variables are computed from  $x_1$  and its derivatives.

with boundary conditions  $S(5) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ ,  $R(5) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ ,  $Q(5) = 0$ . Solving these equations backward in time gives the perturbation feedback control law

$$\delta u = -H_{uu} (H_{ux} + f_u^T (S - RQ^{-1}R^T)) \delta x - H_{uu}^{-1} f_u^T RQ^{-1} \delta \psi.$$

For example, one can easily compute  $\delta u(0)/\delta x(0) = -0.4336$ , and the partial derivative of the optimal cost value with respect to a change of initial condition is  $(2.377 \ 0.388)$ .

Of course, we derived the preceding results in a straightforward manner. This was for sake of motivation. Our conclusion here is that the optimal history of the linearizing output  $x_1$  actually carries a lot of information: histories of adjoint variables and, consequently, information about neighboring extremals, closed loop approximate optimal control, and post optimal analysis. One can wonder how general this property is. In Sections 3 and 4, we actually prove similar results in general multivariable cases.

### 2.3. Increase in accuracy

#### 2.3.1. Theoretical aspects

We would like to mention that some significant impact on convergence of numerical solvers dedicated to the approach we advocate can be expected. In [5], Section 5.6, numerical schemes for solving boundary value problems for high order differential equations are studied. A collocation scheme is proposed along with various implementations. A first convergence result for linear boundary value problems is proven. We note  $p$  the regularity of the coefficients of the linear differential system, and  $m$  its order. Approximate solutions are sought after among piecewise polynomials of degree  $k + m$ . There are  $k$  collocation points, and  $h$  corresponds to the mesh size. Under an orthogonality condition on the collocation points, the following error estimates are derived in [5], Theorem 5.140. At the mesh points  $x_i$

$$|u^{(j)}(x_i) - u_{\pi}^{(j)}(x_i)| = O(h^p), \quad 0 \leq j \leq m - 1 \tag{7}$$

where  $u$  is the exact solution of the BVP problem and  $u_\pi$  is the approximate solution obtained through the collocation scheme for the high order system. Outside the mesh points, we have

$$|u^{(j)}(x) - u_\pi^{(j)}(x)| = O(h^{k+m-j}) + O(h^p), \quad 0 \leq j \leq m - 1. \tag{8}$$

Interestingly, if we choose to use the proposed collocation method on an equivalent state space form, (7) remains unchanged, but (8) is replaced by

$$|y(x) - y_\pi(x)| = O(h^{k+1}) + O(h^p) \tag{9}$$

where  $y$  (resp.  $y_\pi$ ) is the exact (resp. approximate collocation) solution of the equivalent state space form BVP ( $y$  is the concatenation of the derivatives of  $u$  from order 0 to  $m - 1$  see [5], pp. 220–222). In terms of convergence, the upper bound of (8) is better than (9). If  $p$  is large enough, we see that the collocation method for the high order system is more accurate than the collocation method for the state space form at points outside the mesh.

These approaches are then extended to the nonlinear case using quasi-linearization and a Newton method to solve the nonlinear problem. Roundoff errors depend on which functions basis is used for collocation. This is beyond the scope of this remark; interested readers can refer to [5], Section 5.6.4.

2.3.2. Numerical investigations

For sake of illustration, we present an optimal control example which possesses an analytical solution. In this problem of optimal investment (see [24] for original problem formulation), it is desired to optimize the following cost

$$J = \int_0^T \exp(-\beta t) 2\sqrt{u} dt$$

under the dynamical constraint

$$\dot{x} = \alpha x - u$$

with fixed boundary conditions

$$x(0) = S, \quad x(T) = 0 \tag{10}$$

where  $T > 0, S > 0, \alpha > \beta > \alpha/2$ . The analytic solution to this problem is

$$x(t) = S \exp(2(\alpha - \beta)t) \frac{1 - \exp((\alpha - 2\beta)(T - t))}{1 - \exp((\alpha - 2\beta)T)}$$

Following solution method 3, we determine the higher order differential equation to be satisfied by optimal solutions

$$\ddot{x} = \alpha \dot{x} - 2(\beta - \alpha)(\alpha x - \dot{x}) \tag{11}$$

while the control and the adjoint state can be recovered as

$$u = \alpha x - \dot{x}, \quad \lambda = \frac{\exp(-\beta t)}{\sqrt{\alpha x - \dot{x}}}$$

The two point boundary value problem (11), (10) can be numerically treated with the Scilab implementation `bvode` of the Fortran package `colnew` (see [3,5]) for the numerical solution of multi-point boundary value problems for mixed order systems of ordinary differential equations. This routine has the possibility to directly address higher order differential equation. It can also deal with a set of first order equations. To illustrate the improvement in accuracy obtained when using higher order equations, we decide to solve (11), (10) under the form of the derived second order equation (11), and, separately, under the form of two first order equations.

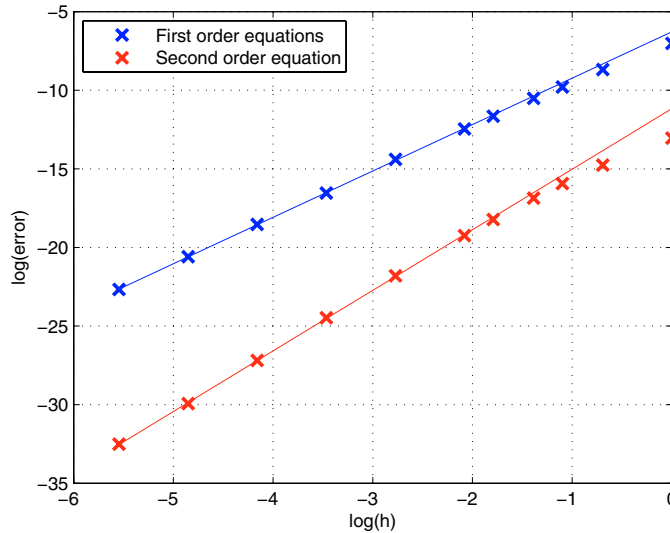


FIGURE 2. Testing the accuracy of a first order method versus a higher order approach (optimal investment problem).

Results are reported in Figure 2. Mesh is automatically refined by the solver to satisfy constraints (including collocation constraints) within a user-specified tolerance. The final mesh size is reported on the  $x$ -axis ( $\log(h)$ ) of Figure 2. Afterwards, we evaluate the difference between the analytic solution and the obtained result over a very fine grid (much finer than the solver mesh). This error is reported on the  $y$ -axis ( $\log(\text{error})$ ) of Figure 2. The observed results are in great accordance with equations (8) and (9). Here,  $\alpha = 2$ ,  $\beta = 4/3$ ,  $S = 1$ ,  $T = 1$ ,  $k = 2$ ,  $m = 2$ ,  $p = \infty$ . Numerical fits of  $\log(\text{error})$  as an affine function of  $\log(h)$  provide slopes of 2.9452 and 3.8305, for first order and second order methods respectively. Other values for  $k$  provide consistent results<sup>1</sup>. These are very close to the theoretical values of 3 and 4 that are given by equations (9) and (8). In theory and in practice on this example as well, accuracy is improved by using higher order equations instead of first order equations.

#### 2.4. Numerical comparisons with other approaches

Finally, we propose here to investigate the computational impact solving the two boundary value problem under the proposed higher order form. For that purpose, we consider the above presented forced van der Pol problem. Separately, we test solution technique 1 with the NTG software package [25], and solution technique 3 with the Scilab `bvode` routine (presented in Sect. 2.3.2). Two cases are considered. Successively, the `bvode` routine uses a formulation of the two point boundary value problem (presented in solution method 3) under the form of 4 first order differential equations or under the form of the single fourth order differential equation. In summary, we use three different approaches to the same problem.

Results are reported in Tables 1–3, respectively. For each computed solution, a numerical integration using a Runge-Kutta scheme is carried out using a control evaluated over a fine grid. The purpose of this is to come with a fair (and method independent) comparison of obtained cost values and constraints violations. The quality of a method is summarized by these recomputed values. Tests were conducted on a 2 GHz Pentium M computer.

<sup>1</sup> It shall be noted that at very high values, such as  $k = 7$ , round off errors become dominant for mesh size under  $1/4$ .

TABLE 1. Numerical results obtained with solution technique 1.

|                                   |           |           |
|-----------------------------------|-----------|-----------|
| Found optimal cost                | 1.6857621 | 1.6857371 |
| Re-computed cost value            | 1.6857452 | 1.6857107 |
| CPU time (ms)                     | 9.1       | 120       |
| Memory usage (floats)             | 137 000   | 310 000   |
| Number of intervals               | 10        | 30        |
| Re-computed constraints violation | 2.70 E-06 | 2.89 E-07 |

TABLE 2. Numerical results obtained with solution technique 3 using 4 first order differential equations.

|                                   |            |            |            |            |
|-----------------------------------|------------|------------|------------|------------|
| Found optimal cost                | 1.6856685  | 1.6856832  | 1.6856832  | 1.6856832  |
| Re-computed cost value            | 1.6856825  | 1.6856834  | 1.6856829  | 1.6856832  |
| CPU time (ms)                     | 15.9       | 19.7       | 27.1       | 41.7       |
| Memory usage (floats)             | 3465       | 5775       | 10 395     | 19 635     |
| Number of intervals               | 2          | 4          | 8          | 16         |
| Requested tolerance               | 1.00 E-04  | 1.00 E-05  | 1.00 E-07  | 1.00 E-09  |
| Re-computed constraints violation | -1.86 E-06 | -2.53 E-07 | -2.78 E-07 | -2.02 E-07 |

TABLE 3. Numerical results obtained with solution technique 3 using the single 4th order differential equation.

|                                   |            |            |            |            |
|-----------------------------------|------------|------------|------------|------------|
| Found optimal cost                | 1.6856842  | 1.6856841  | 1.6856841  | 1.6856841  |
| Re-computed cost value            | 1.6856834  | 1.6856829  | 1.6856832  | 1.6856832  |
| CPU time (ms)                     | 6.2        | 9.0        | 16.6       | 26.8       |
| Memory usage (floats)             | 630        | 1050       | 1890       | 3570       |
| Number of intervals               | 2          | 4          | 8          | 16         |
| Requested tolerance               | 1.00 E-03  | 1.00 E-04  | 1.00 E-06  | 1.00 E-07  |
| Re-computed constraints violation | -9.54 E-08 | -6.37 E-07 | -7.00 E-08 | -2.15 E-07 |

While, roughly speaking, all the methods converge to similar solutions and comparable cost values, several points should be noted. The higher order method outperforms the other two in terms of memory usage, CPU time and accuracy.

With similar or better values for the obtained cost and the constraints violation, the reduction in memory usage is due to the fact that only a single variable need to be considered in the higher order approach. The relatively high memory space required by the NTG software package is due to the call to the external NPSOL solver used for solving the SQP problem which is comparatively more memory demanding.

The reduction in CPU time is important. The method using the high order TPBVP can provide accurate solutions faster than NTG<sup>2</sup>. Besides, the higher order method is constantly faster than the corresponding first order method as accuracy requirements become more stringent.

Finally, the higher order method appears as the most accurate of the three methods. Comparisons is particularly relevant with the first order method. Constraints violations is lower while requested tolerance can be relaxed.

<sup>2</sup> It must be noticed that this method (as-is) can not address interval state constraints, while NTG can.



### 3. NORMAL FORM AND HIGH ORDER STATIONARITY CONDITIONS

#### 3.1. Problem statement and feedback linearization

Consider a multivariable control-affine system

$$\dot{x} = f(x) + \sum_{i=1}^m g_i(x)u_i \tag{12}$$

where  $f(x), g_1(x), \dots, g_m(x)$  are smooth vector fields  $\mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $m \leq n$ . We note  $g(x) = [g_1(x) \dots g_m(x)]$ . First, we are interested in putting it into a normal form. For that reason, we consider  $m$  smooth functions  $\mathbb{R}^n \rightarrow \mathbb{R}$

$$\xi_1 = h_1(x), \dots, \xi_m = h_m(x) \tag{13}$$

and investigate the vector relative degree defined as follows (where  $L$  is the Lie derivative)

**Definition 1** [13]. A system of the form (12,13) is said to have *vector relative degree*  $\{r_1, r_2, \dots, r_m\}$  at  $x^0$  if

$$L_{g_j} h_i(x) = \dots = L_{g_j} L_f^{r_i-2} h_i(x) = 0$$

for all  $1 \leq i, j \leq m$  and all  $x$  in a neighborhood of  $x^0$ , and the matrix

$$A(x) = \{a_{ij}(x)\} = \{L_{g_j} L_f^{r_i-1} h_i(x)\}$$

is nonsingular at  $x = x^0$ .

Further, this system is said to have *uniform relative degree*  $\{r_1, r_2, \dots, r_m\}$  if it has vector relative degree  $\{r_1, r_2, \dots, r_m\}$  at all  $x \in \mathbb{R}^n$ . The sum  $r = r_1 + \dots + r_m$  is called *total relative degree*.

We assume having such uniform relative degree. We note, for all  $1 \leq j \leq r_i, 1 \leq i \leq m$

$$\xi_i^j(x) = L_f^{j-1} h_i(x)$$

and the mapping  $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^r$  defined by

$$\mathcal{H}(x) = \text{col} (\xi_1^1(x) \dots \xi_1^{r_1}(x) \dots \xi_m^1(x) \dots \xi_m^{r_m}(x)). \tag{14}$$

Set

$$\begin{aligned} \tilde{f} &= f - gA^{-1}(x) \begin{pmatrix} L_f^{r_1} h_1(x) \\ \vdots \\ L_f^{r_m} h_m(x) \end{pmatrix} \\ \tilde{g} &= gA^{-1}(x) \triangleq [\tilde{g}_1(x) \dots \tilde{g}_m(x)]. \end{aligned}$$

The following result directly arises from [23], pp. 109–118.

**Theorem 1.** *Suppose system (12,13) has uniform relative degree  $\{r_1, \dots, r_m\}$  and set  $r = r_1 + \dots + r_m$ . Suppose  $Z^* = \mathcal{H}^{-1}(0)$  is non empty (where  $\mathcal{H}$  is defined by (14)). Suppose the vector fields*

$$Y_j^k(x) = \text{ad}_{\tilde{f}}^{k-1} \tilde{g}_j, \quad 1 \leq j \leq m, \quad 1 \leq k \leq r_j$$

*are complete. Then  $Z^*$  is connected and there is a globally defined diffeomorphism  $\Upsilon : \mathbb{R}^n \rightarrow Z^* \times \mathbb{R}^r$  which changes (12) into the following normal form (15,16), with  $X \triangleq (\xi_1^1, \dots, \xi_1^{r_1}, \xi_2^1, \dots, \xi_2^{r_2}, \dots, \xi_m^1, \dots, \xi_m^{r_m}, \eta)^T \in \mathbb{R}^n$ ,*

$\eta \triangleq (\eta_1, \dots, \eta_q)^T \in \mathbb{R}^q$  with  $q = n - r$ , and  $v \triangleq (v_1, \dots, v_m) \in \mathbb{R}^m$

$$\frac{d\xi_i^1}{dt} = \xi_i^2, \dots, \frac{d\xi_i^{r_i}}{dt} = L_f^{r_i} h_i + \sum_{j=1}^m a_{ij} u_j \triangleq v_i \tag{15}$$

$$\frac{d\eta_j}{dt} = \alpha^j(X) + \sum_{i=1}^m \beta_i^j(X) v_i = \mathcal{B}_j(X, v) \tag{16}$$

where  $\alpha^j$  and  $\beta_i^j$  are smooth  $\mathbb{R}^n \rightarrow \mathbb{R}$  mappings, for all  $j = 1, \dots, q$ , and  $i = 1, \dots, m$ .

In the previous normal form, we performed a feedback linearization  $v_i = L_f^{r_i} h_i + \sum_{j=1}^m a_{ij} u_j$ . This substitution is invertible since  $A$  is nonsingular by assumption.

It is now assumed we want to solve a Lagrange optimal control problem, originally of the form

$$\int_0^T L(x(t), u(t)) dt$$

where  $x$  and  $u$  satisfy (12). In this expression,  $[0, T]$  is a fixed time interval and  $L$  is a smooth real valued function whose Hessian with respect to  $u$  is definite positive. Without loss of generality, infinite horizon, terminal cost, initial or final general constraints could be considered as well but remain out of the scope of the paper.

We now cast this problem into the newly defined variables  $X$ , and  $v$  (see Th. 1). The following proposition details a set of necessary conditions to be satisfied by optimal controls. We use them in the rest of the paper.

**Proposition 1** (Weak Minimum Principle [8]). *Consider the system  $\dot{X} = f_0(X(t), v(t))$  ( $f_0$  being a  $C^\infty$  mapping) where  $v(\cdot) \in V = L^\infty([0, T])$  and the minimisation problem:  $\min_{v(\cdot) \in V} \int_0^T \mathcal{L}(X(t), v(t)) dt$  where  $T$  and the extremities  $X_0, X_1$  are fixed ( $\mathcal{L}$  being  $C^1$ ). If  $v^*$  and its corresponding trajectory  $X^*$  are optimal and correspond to a normal case in the calculus of variations (i.e.  $v(\cdot)$  is regular for the system  $\dot{X}(t) = f_0(X(t), v(t))$ , (see [8], p. 43, for details<sup>3</sup>), then there exists  $p^*(t) \in \mathbb{R}^n$  such that  $(X^*, p^*, v^*)$  satisfies*

$$\dot{X}^*(t) = \frac{\partial H}{\partial p}(X^*(t), p^*(t), v^*(t)) \tag{17}$$

$$\dot{p}^*(t) = -\frac{\partial H}{\partial X}(X^*(t), p^*(t), v^*(t)) \tag{18}$$

$$\frac{\partial H}{\partial v}(X^*(t), p^*(t), v^*(t)) = 0 \tag{19}$$

where  $H = \mathcal{L}(X, v) + p^T f_0(X, v)$ .

These results and their included hypothesis lead us to consider the following optimal control problem used throughout the rest of the paper.

**Definition 2** (optimal control problem definition). Let the multivariable control system under normal form (15,16)  $\dot{X} = f_0(X(t), v(t))$  ( $f_0$  being a  $C^\infty$  mapping) where  $v(\cdot) \in V = L^\infty([0, T])$ . It is desired to solve the optimal control problem:  $\min_{v(\cdot) \in V} \int_0^T \mathcal{L}(X(t), v(t)) dt$  where  $T$  and the extremities  $X_0, X_1$  are fixed, and

---

<sup>3</sup> We restrict ourselves to these normal extremals. Computations of abnormal extremals is known to be a very difficult task and those may even be not optimal. One can refer to [1,28] for further details and discussions.

$\mathcal{L} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$  is a  $C^2$  mapping whose Hessian is assumed to be positive definite<sup>4</sup> with respect to  $v$ :

$$\left( \frac{\partial^2 \mathcal{L}}{\partial v^2} \right) > 0. \quad (20)$$

Because the control system is in normal form, the Hamiltonian  $H$  has a very particular expression

$$H = \mathcal{L}(X, v) + \sum_{i=1}^m \sum_{j=1}^{r_i-1} \lambda_i^j \xi_i^{j+1} + \sum_{i=1}^m \lambda_i^{r_i} v_i + \sum_{j=1}^q \mu_j \mathcal{B}_j(X, v)$$

where  $\lambda \triangleq (\lambda_1^1, \dots, \lambda_1^{r_1}, \dots, \lambda_m^1, \dots, \lambda_m^{r_m})^T$  represents the adjoint states related to the cascade dynamics (15) and  $\mu \triangleq (\mu_1, \dots, \mu_q)^T$  corresponds to the zero dynamics (16). Thus, the stationarity adjoint equations (18) write

$$\frac{d\lambda_i^1}{dt} = -\frac{\partial \mathcal{L}}{\partial \xi_i^1} - \sum_{j=1}^q \mu_j \frac{\partial \mathcal{B}_j}{\partial \xi_i^1}, \quad i = 1 \dots p \quad (21)$$

$$\frac{d\lambda_i^j}{dt} = -\lambda_i^{j-1} - \frac{\partial \mathcal{L}}{\partial \xi_i^j} - \sum_{k=1}^q \mu_k \frac{\partial \mathcal{B}_k}{\partial \xi_i^j}, \quad i = 1 \dots p, \quad j = 2 \dots r_i \quad (22)$$

$$\frac{d\mu_j}{dt} = -\frac{\partial \mathcal{L}}{\partial \eta_j} - \sum_{k=1}^q \mu_k \frac{\partial \mathcal{B}_k}{\partial \eta_j}, \quad j = 1 \dots q. \quad (23)$$

**Stationarity of the Hamiltonian.** Using (16), stationarity equations (19) of the Hamiltonian with respect to the control variables yield, for  $i = 1 \dots m$

$$\frac{\partial H}{\partial v_i} = \frac{\partial \mathcal{L}}{\partial v_i} + \lambda_i^{r_i} + \sum_{j=1}^q \mu_j \beta_i^j(X) = 0 \quad (24)$$

This equation can be differentiated to get a high order system of differential equations implying only the variables  $\xi_i^1$ ,  $i = 1 \dots m$ ,  $\eta$  and  $\mu$ . For sake of simplicity, we now note  $\xi_i \triangleq \xi_i^1$ . In particular, we have  $\xi_i^{(j-1)} = \xi_i^j$  and  $\xi_i^{(r_i)} = v_i$ .

### 3.2. Obtaining a higher order differential system

**Lemma 1.** *Let  $i \in \{1 \dots m\}$  such that  $r_i \geq 2$ . For  $j = 1 \dots r_i - 1$ , there exists a function  $G_i^j$  from  $\mathbb{R}^{2n-r+mj}$  to  $\mathbb{R}$  such that*

$$\frac{d^j}{dt^j} \frac{\partial H}{\partial v_i} = \sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+j)} + (-1)^j \lambda_i^{r_i-j} + G_i^j(\eta, \mu, \dots, \xi_l \dots \xi_l^{(r_l+j-1)}, \dots) \quad (25)$$

where  $l$  is a running index ranging from 1 to  $m$ .

*Proof.* We start the proof with  $j = 1$ . To that end, we differentiate (24). First, we have

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial v_i} = \sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+1)} + \sum_{k=1}^m \sum_{j=1}^{r_k} \frac{\partial^2 \mathcal{L}}{\partial v_i \partial \xi_k^j} \xi_k^{(j)} + \sum_{l=1}^q \frac{\partial^2 \mathcal{L}}{\partial v_i \partial \eta_l} \mathcal{B}_l(X, \dots, \xi_k^{r_k}, \dots)$$

<sup>4</sup> Interestingly, the reader may notice that this assumption holds if in the coordinates  $(x, u)$  the cost to be minimized is  $\int_0^T L(x, u) dt$  with  $L$  having a positive definite Hessian with respect to  $u$ . Indeed, the linearizing affine change of coordinates  $v_i = L_f^{r_i} h_i + \sum_{j=1}^m a_{ij} u_j$  yields the matrix equality  $\left( \frac{\partial^2 \mathcal{L}}{\partial v^2} \right) = (A^{-1})^T \left( \frac{\partial^2 L}{\partial u^2} \right) A^{-1} > 0$ .

which is of the form

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial v_i} = \sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+1)} + \gamma_i^1(\eta, \dots, \xi_l \dots \xi_l^{(r_l)}, \dots) \tag{26}$$

where  $\gamma_i^1$  is a function from  $\mathbb{R}^{n+m}$  to  $\mathbb{R}$ . From (22) with  $j = r_i$ , we get

$$\frac{d\lambda_i^{r_i}}{dt} = -\lambda_i^{r_i-1} - \frac{\partial \mathcal{L}}{\partial \xi_i^{r_i}} - \sum_{k=1}^q \mu_k \frac{\partial \mathcal{B}_k}{\partial \xi_i^{r_i}}$$

which, with  $k$  a running index ranging from 1 to  $m$ , writes under the form

$$\frac{d\lambda_i^{r_i}}{dt} = -\lambda_i^{r_i-1} + \bar{\gamma}_i^1(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k)}, \dots) \tag{27}$$

where  $\bar{\gamma}_i^1$  is a function from  $\mathbb{R}^{2n-r+m}$  to  $\mathbb{R}$ . The right hand side of (23) involves only  $\mu$ ,  $X$  and  $v$ . In other words, it is some function of the variables  $(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k)}, \dots)$ . Similarly, the total derivative of  $\beta_i^j(X)$  expresses in terms of  $X$ , and  $v$ . Finally,  $\frac{d}{dt} \sum_{j=1}^{j=q} \mu_j \beta_i^j(X)$  is of the form

$$\frac{d}{dt} \sum_{j=1}^q \mu_j \beta_i^j(X) = \bar{\gamma}_i^j(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k)}, \dots) \tag{28}$$

where  $\bar{\gamma}_i^j$  is a function from  $\mathbb{R}^{2n-r+m}$  to  $\mathbb{R}$ . According to (24), summing up (26), (27), and (28) provides us with an expression of  $\frac{d}{dt} \frac{\partial H}{\partial v_i}$  and defines the sought after  $G_i^j$ . This proves (25) for  $j = 1$ . The induction goes pretty much along the same lines. Assume that (25) holds for some  $j < r_i - 1$ . The differentiation involves three terms. The derivative of the first term is of the form

$$\frac{d}{dt} \sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+j)} = \sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+j+1)} + \gamma_i^{j+1}(\eta, \dots, \xi_l \dots \xi_l^{(r_l+j)}, \dots) \tag{29}$$

where  $\gamma_i^{j+1}$  is a function from  $\mathbb{R}^{n+m(j+1)}$  to  $\mathbb{R}$ . The second term is given by the adjoint equation (22) which, since  $r_i - j \geq 2$ , is

$$\frac{d}{dt} \lambda_i^{r_i-j} = -\lambda_i^{r_i-j-1} - \frac{\partial \mathcal{L}}{\partial \xi_i^{r_i-j}} - \sum_{k=1}^q \mu_k \frac{\partial \mathcal{B}_k}{\partial \xi_i^{r_i-j}}$$

This term is of the form

$$\frac{d}{dt} \lambda_i^{r_i-j} = -\lambda_i^{r_i-(j+1)} + \bar{\gamma}_i^{j+1}(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k+j)}, \dots) \tag{30}$$

where  $\bar{\gamma}_i^{j+1}$  is a function from  $\mathbb{R}^{2n-r+m(j+1)}$  to  $\mathbb{R}$ . Finally, the last term  $\frac{d}{dt} G_i^j$  is trivially of the form

$$\frac{d}{dt} G_i^j(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k+j-1)}, \dots) = \bar{\gamma}_i^{j+1}(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k+j)}, \dots) \tag{31}$$

where  $\bar{\gamma}_i^{j+1}$  is a function from  $\mathbb{R}^{2n-r+m(j+1)}$  to  $\mathbb{R}$ . Putting together (29), (30), and (31) gives the expression of  $G_i^{j+1}$ . This proves the induction and concludes the proof.  $\square$

We now prove that  $\xi_1 \dots \xi_m$  satisfy a set of differential equations in which none of the components of  $\lambda$  appears.

**Theorem 2.** For  $i = 1 \dots m$ , there exists a function  $G_i$  from  $\mathbb{R}^{2n-r+mr_i}$  to  $\mathbb{R}$

$$G_i(\eta, \mu, \dots, \xi_l \dots \xi_l^{(r_l+r_i-1)}, \dots)$$

where  $l$  is a running index ranging from 1 to  $m$ , such that

$$\sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+r_i)} + G_i(\eta, \mu, \dots, \xi_l \dots \xi_l^{(r_l+r_i-1)}, \dots) = 0. \tag{32}$$

Together with (16) and (23), these equations are a set of differential equations on  $\xi_i$ ,  $i = 1, \dots, m$ ,  $\eta$  and  $\mu$  from which the  $\lambda_i^j$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, r_m$ , have been eliminated.

*Proof.* We start with the particular case  $r_i = 1$ . Stationarity of the Hamiltonian reads

$$\frac{\partial H}{\partial v_i} = \frac{\partial \mathcal{L}}{\partial v_i} + \lambda_i^1 + \sum_{j=1}^q \mu_j \beta_i^j(X) = 0. \tag{33}$$

Differentiating this expression with respect to time gives rise to three groups of terms. The first term is

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial v_i} = \sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+1)} + \frac{\partial^2 \mathcal{L}}{\partial v_i \partial X} \frac{dX}{dt}(X, v)$$

which is of the form

$$\sum_{k=1}^m \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+r_i)} + \gamma_i(\eta, \mu, \dots, \xi_l \dots \xi_l^{(r_l+r_i-1)}, \dots) \tag{34}$$

where  $\gamma_i$  is from  $\mathbb{R}^{2n-r+mr_i}$  to  $\mathbb{R}$ . The second term is

$$\frac{d\lambda_i^1}{dt} = -\frac{\partial \mathcal{L}}{\partial \xi_i^1} - \sum_{j=1}^q \mu_j \frac{\partial \mathcal{B}_j}{\partial \xi_i^1}$$

which comes from the adjoint equation (21) and is of the form

$$\bar{\gamma}_i(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k)}, \dots) \tag{35}$$

where  $\bar{\gamma}_i$  is from  $\mathbb{R}^{2n-r+m}$  to  $\mathbb{R}$ . Finally, differentiation of  $\mu_j \beta_i^j(X)$  yields a term of the form

$$\bar{\bar{\gamma}}_i(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k+r_i-1)}, \dots) \tag{36}$$

where  $\bar{\bar{\gamma}}_i$  is from  $\mathbb{R}^{2n-r+mr_i}$  to  $\mathbb{R}$ . Summing up expressions (34), (35) and (36) defines  $G_i$  and gives the desired expression (32).

Now, let us turn to the case  $r_i \geq 2$ . From Lemma 1 with  $j = r_i - 1$ , we know that there exists a function  $G_i^j$  from  $\mathbb{R}^{2n-r+m(r_i-1)}$  to  $\mathbb{R}$  such that

$$\sum_{k=1}^{k=m} \frac{\partial^2 \mathcal{L}}{\partial v_i \partial v_k} \xi_k^{(r_k+r_i-1)} + (-1)^j \lambda_i^1 + G_i^j(\eta, \mu, \dots, \xi_k \dots \xi_k^{(r_k+r_i-2)}, \dots) = 0.$$

Following the computations presented for the case  $r_i = 1$ , we can easily figure out that the differentiation of  $\lambda_i^1$  makes the adjoint state  $\lambda$  disappear. Eventually, we get equation (32). □

#### 4. HIGH ORDER STATIONARITY CONDITIONS UNDER STATE SPACE FORM

Obtaining a differential system where the highest order derivatives of the variables are explicit functions of their lower derivatives, *i.e.* a state space form, is important for two reasons. First (and obviously), state space methods can be used to solve the boundary value problem. Moreover, the methods used to solve higher order differential systems (such as those presented in [4,5]) require such an explicit formulation. In this section, we prove that the differential equations (32, 16, 23) of Theorem 2 can be used to define a state space form set of differential equations involving a reduced number of variables, namely  $(\xi_1, \dots, \xi_m)$ ,  $\eta$  and  $\mu$ . This result is demonstrated on two cases of particular interest ( $m = 2$  and  $m = 3$ ), and, eventually, proven in the general case  $m > 1$ . In the case  $m=1$ , (32) already has a state space form. Specific comments can be found in [14].

##### 4.1. Two inputs systems

If  $r_1 = r_2 = r$ , noting that the Hessian (20) is invertible, we obtain a state space form by solving system (32) which is linear with respect to  $\xi_1^{(2r)}$  and  $\xi_2^{(2r)}$ . More generally, we now assume  $r_1 > r_2$ . Under this assumption, system (32) is not under state space form. Indeed, from Theorem 2, we have

$$\frac{\partial^2 \mathcal{L}}{\partial v_1^2} \xi_1^{(2r_1)} + \frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_2^{(r_1+r_2)} + G_1(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(r_1+r_2-1)}) = 0 \quad (37)$$

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2)} + G_2(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}) = 0 \quad (38)$$

where  $G_1$  is from  $\mathbb{R}^{2n-r+2r_1}$  to  $\mathbb{R}$  and  $G_2$  is from  $\mathbb{R}^{2n-r+2r_2}$  to  $\mathbb{R}$ . In the previous equations, the highest orders of differentiation of  $\xi_1$  and  $\xi_2$  appear only in equation (37), because  $r_1 > r_2$ . This does not provide a direct extraction of these highest orders of differentiation terms from (37) and (38). In fact, the highest order of differentiation in  $\xi_2$  in equation (37) can be eliminated thanks to (38) and its time derivatives. As a result (see following proof and its development which leads to Prop. 2), one can obtain a state space form of order  $2r_1$  with respect to  $\xi_1$  and  $2r_2$  with respect to  $\xi_2$ .

**Lemma 2.** For  $0 \leq i \leq r_1 - r_2$ , there exists  $\bar{G}_2^i$

$$\bar{G}_2^i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i-1)}, \xi_2 \dots \xi_2^{(2r_2-1)})$$

from  $\mathbb{R}^{2n-r+2r_2+i}$  to  $\mathbb{R}$ , such that

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2+i)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2+i)} + \bar{G}_2^i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}) = 0. \quad (39)$$

*Proof.* The equation holds for  $i = 0$ . Let us proceed by induction. We assume that (39) holds for  $i$  and differentiate it. We get an equation of the form

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2+i+1)} + g_1(X, \xi_1^{(r_1)}, \xi_2^{(r_2)}) \xi_1^{(r_1+r_2+i)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2+i+1)} \\ + g_2(X, \xi_1^{(r_1)}, \xi_2^{(r_2)}) \xi_2^{(2r_2+i)} + \tilde{G}_2^i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2)}) = 0. \end{aligned}$$

We use (39) to replace  $\xi_2^{(2r_2+i)}$  by a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2-1)}$ . We also use (38) to replace  $\xi_2^{2r_2}$  by a function of  $\eta, \mu, \xi_1 \dots \xi_1^{r_1+r_2}, \xi_2 \dots \xi_2^{(2r_2-1)}$ . We thus obtain an equation of the form

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2+i+1)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2+i+1)} + \tilde{G}_2^i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2-1)}) = 0$$

with  $\tilde{G}_2^i$  from  $\mathbb{R}^{2n-r+2r_2+i+1}$  to  $\mathbb{R}$ , which proves the induction and concludes the proof. □

For  $i = r_1 - r_2$ , equation (39) writes

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(2r_1)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(r_1+r_2)} + \bar{G}_2^{r_1-r_2}(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}) = 0$$

where  $\bar{G}_2^{r_1-r_2}$  is from  $\mathbb{R}^{2n}$  to  $\mathbb{R}$ . Together with equation (37), and because the Hessian (20) is invertible, solving this system of two equations with respect to the highest derivatives of  $\xi_1$  and  $\xi_2$  leads to an expression of  $\xi_1^{(2r_1)}$  as a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(r_1+r_2-1)}$ . From equation (39), and noticing that the positive definiteness of the Hessian (20) implies  $\left(\frac{\partial^2 \mathcal{L}}{\partial v_2^2}\right) > 0$ , we see that  $\xi_2^{(2r_2)}$  is a function of  $\eta, \mu, \xi_1, \dots, \xi_1^{(r_1+r_2)}, \xi_2, \dots, \xi_2^{(2r_2-1)}$ . Differentiating this equation further shows that that  $\xi_2^{(2r_2+i)}$  is a function of  $\eta, \mu, \xi_1, \dots, \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2-1)}$ . Hence, there exists a function  $G_3$  from  $\mathbb{R}^{2n}$  to  $\mathbb{R}$  such that

$$\xi_1^{(2r_1)} = G_3(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}) \tag{40}$$

which, together with equation (38) gives the following partial state space model

$$\frac{d\xi_1^{(2r_1-1)}}{dt} = G_3(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, t) \tag{41}$$

$$\frac{d\xi_2^{(2r_2-1)}}{dt} = \left(\frac{\partial^2 \mathcal{L}}{\partial v_2^2}\right)^{-1} \left(\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2)} + G_2(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, t)\right). \tag{42}$$

Observe that  $\xi_1^{(r_1+r_2-1)}$  is a derivative of  $\xi_1$  which is of an order lower than  $2r_1 - 1$  because  $r_2 < r_1$  and that, as a vector, the right hand side of equations (41,42) is a function from  $\mathbb{R}^{2n}$  to  $\mathbb{R}^2$ . Thus, the following proposition holds.

**Proposition 2.** *Equations (16), (23), (41) and (42) are a  $2n$  dimensional state space form set of equations to be satisfied by optimal solutions.*

### 4.2. Three inputs systems

Here, we consider systems with three control variables. This gives rise to three chains of integrators in the normal form. The associated lengths are sorted (without loss of generality) and noted  $r_1 > r_2 > r_3$ . This assumption is made because, when two or three chains have equal lengths, a state space form can be derived using the same method as in the case of two or one input systems respectively. These particular cases are also addressed by the general result presented in Section 4.3. We detail them to introduce and demonstrate this last result.

From Theorem 2, we have to consider three differential equations

$$\frac{\partial^2 \mathcal{L}}{\partial v_1^2} \xi_1^{(2r_1)} + \frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_2^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_3} \xi_3^{(r_1+r_3)} + G_1(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(r_1+r_2-1)}, \xi_3 \dots \xi_3^{(r_1+r_3-1)}) = 0 \tag{43}$$

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_3^{(r_2+r_3)} + G_2(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(r_2+r_3-1)}) = 0 \tag{44}$$

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_3} \xi_1^{(r_1+r_3)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_2^{(r_2+r_3)} + \frac{\partial^2 \mathcal{L}}{\partial v_3^2} \xi_3^{(2r_3)} + G_3(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_3-1)}, \xi_2 \dots \xi_2^{(r_2+r_3-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0. \tag{45}$$

As noted in Section 4.1, equations (43), (44), and (45) are not in a state space form. The simple elimination method used for two chains in the previous section fails here because we have to eliminate a larger number of higher orders in equations (43, 44, 45). However, the same technique of orders of differentiation raising can be considered, provided it is improved upon. First, we differentiate (45)  $r_2 - r_3$  times as in the case  $m = 2$  to get

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_3} \xi_1^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_2^{(2r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_3^2} \xi_3^{(r_2+r_3)} + G_4(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0 \tag{46}$$

where  $G_4$  is from  $\mathbb{R}^{2n+r_2-r_1}$  to  $\mathbb{R}$ . As before, we know that, for  $i < r_2 - r_3$ ,  $\xi_3^{2r_3+i}$  as a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_3+i)}, \xi_2 \dots \xi_2^{(r_2+r_3+i)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ . Hence, we can replace (44) by

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_3^{(r_2+r_3)} + \tilde{G}_2(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0 \tag{47}$$

where  $\tilde{G}_2$  is from  $\mathbb{R}^{2n+r_2-r_1}$  to  $\mathbb{R}$ . The matrix

$$\begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial v_2^2} & \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \\ \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} & \frac{\partial^2 \mathcal{L}}{\partial v_3^2} \end{bmatrix}$$

is invertible because it is diagonally extracted from the Hessian (20) which is positive definite. Therefore, we can draw  $\xi_2^{(2r_2)}$  and  $\xi_3^{(r_2+r_3)}$  from (46) and (47) as a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ . Now, let us show recursively that there exist  $\tilde{G}_i$  from  $\mathbb{R}^{n+r_2-r_1+i}$  to  $\mathbb{R}$  and  $\check{G}_i$  from  $\mathbb{R}^{n+r_2-r_1+i}$  to  $\mathbb{R}$  such that

$$\begin{aligned} &\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_3} \xi_1^{(r_1+r_2+i)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_2^{(2r_2+i)} + \frac{\partial^2 \mathcal{L}}{\partial v_3^2} \xi_3^{(r_2+r_3+i)} \\ &+ \tilde{G}_i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0 \end{aligned} \tag{48}$$

$$\begin{aligned} &\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2+i)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2+i)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_3^{(r_2+r_3+i)} \\ &+ \check{G}_i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0 \end{aligned} \tag{49}$$

and that  $\xi_2^{(2r_2+i)}$  and  $\xi_3^{(r_2+r_3+i)}$  is a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ . From (46) and (47), this holds for  $i = 0$ . Differentiating (48) and (49) gives

$$\begin{aligned} &\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_3} \xi_1^{(r_1+r_2+i+1)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_2^{(2r_2+i+1)} + \frac{\partial^2 \mathcal{L}}{\partial v_3^2} \xi_3^{(r_2+r_3+i+1)} \\ &+ \tilde{G}_i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2)}, \xi_3 \dots \xi_3^{(2r_3)}) = 0 \\ &\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(r_1+r_2+i+1)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(2r_2+i+1)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_3^{(r_2+r_3+i+1)} \\ &+ \check{G}_i(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2)}, \xi_3 \dots \xi_3^{(2r_3)}) = 0. \end{aligned}$$

But, we already know, from the resolution of (46) and (47), that  $\xi_2^{(2r_2)}$  is a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ , and, from (45), that  $\xi_3^{(2r_3)}$  is a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_3)}, \xi_2 \dots \xi_2^{(r_2+r_3)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ . This proves the induction for equation (48), and (49). Solving (48), and (49) at  $i + 1$  with respect to



$\xi_2^{(2r_2+i+1)}$  and  $\xi_3^{(r_2+r_3+i+1)}$  recursively proves that these are functions of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i+1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ . Now, let us take  $i = r_1 - r_2$ . Equations (48) and (49) becomes

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_3} \xi_1^{(2r_1)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_2^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_3^2} \xi_3^{(r_1+r_3)} + \tilde{G}_i(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0 \quad (50)$$

$$\frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_1^{(2r_1)} + \frac{\partial^2 \mathcal{L}}{\partial v_2^2} \xi_2^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_2 \partial v_3} \xi_3^{(r_1+r_3)} + \check{G}_i(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0 \quad (51)$$

where  $\tilde{G}_i$  and  $\check{G}_i$  are from  $\mathbb{R}^{2n}$  to  $\mathbb{R}$ . In (43), we can replace  $\xi_2^{(2r_2+i)}$  by a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ , while  $\xi_3^{(2r_3+i)}$  can be replaced, for  $0 \leq i < r_2 - r_3$ , by a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_3+i)}, \xi_2 \dots \xi_2^{(r_2+r_3+i)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ , and, for  $i \geq 0$ ,  $\xi_3^{(r_2+r_3+i)}$  can be substituted with a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2+i)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ . Thus, (43) yields

$$\frac{\partial^2 \mathcal{L}}{\partial v_1^2} \xi_1^{(2r_1)} + \frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_2} \xi_2^{(r_1+r_2)} + \frac{\partial^2 \mathcal{L}}{\partial v_1 \partial v_3} \xi_3^{(r_1+r_3)} + G_5(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) = 0 \quad (52)$$

where  $G_5$  is also from  $\mathbb{R}^{2n}$  to  $\mathbb{R}$ . Solving (50), (51), and (52) with respect to  $\xi_1^{(2r_1)}, \xi_2^{(r_1+r_2)}$  and  $\xi_3^{(r_1+r_3)}$  (a linear system whose matrix is the Hessian (20) which is non singular) gives  $\xi_1^{(2r_1)}$  as a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}, t$ . This, together with  $\xi_2^{(2r_2)}$  being a function of  $\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}$ , and with (45) yields a state space form for equations (43), (44), and (45) of the following form

$$\frac{d\xi_1^{(2r_1-1)}}{dt} = G_6(\eta, \mu, \xi_1 \dots \xi_1^{(2r_1-1)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) \quad (53)$$

$$\frac{d\xi_2^{(2r_2-1)}}{dt} = G_7(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_2)}, \xi_2 \dots \xi_2^{(2r_2-1)}, \xi_3 \dots \xi_3^{(2r_3-1)}) \quad (54)$$

$$\frac{d\xi_3^{(2r_3-1)}}{dt} = G_8(\eta, \mu, \xi_1 \dots \xi_1^{(r_1+r_3)}, \xi_2 \dots \xi_2^{(r_2+r_3)}, \xi_3 \dots \xi_3^{(2r_3-1)}). \quad (55)$$

Observe that  $r_2 + r_3 \leq 2r_2 - 1$ , and that both  $r_1 + r_2$  and  $r_1 + r_3$  are smaller than  $2r_1 - 1$ . As a vector right hand side,  $G_6, G_7$  and  $G_8$  are globally from  $\mathbb{R}^{2n}$  to  $\mathbb{R}$ .

**Proposition 3.** *Equations (16), (23), (53), (54) and (55) are a  $2n$  dimensional state space form set of equations to be satisfied by optimal solutions.*

### 4.3. General case

Among the  $m$  integrators chains (15), several may share the same length. Thus, we note  $p$  the number of distinct chain lengths. For  $1 \leq i \leq p$ , there are  $n_i$  chains of length  $r_i$ , and we have

$$\sum_{i=1}^p n_i r_i \leq n, \quad \sum_{i=1}^p n_i = m.$$

The chain lengths  $r_i$  are sorted in decreasing order. For  $1 \leq i \leq p$ ,  $1 \leq k \leq n_i$ ,  $\xi_{i,k}$  denotes the primal state which starts the  $k^{th}$  chain of differentiation of length  $r_i$ . We note  $\xi_{i,k}^{[l]}$  the concatenation of the variables  $\xi_{i,k} \dots \xi_{i,k}^{(n)} \dots \xi_{i,k}^{(l)}$ . Similarly,  $v_{i,k}$  denotes the control variable associated to this same chain. This reordering gives a rewriting of equations (32) in Theorem 2 as follows: for  $1 \leq i \leq p$ , and  $1 \leq j \leq n_i$

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{i,j} \partial v_{k,l}} \xi_{k,l}^{(r_i+r_k)} + G_{i,j}(\eta, \mu, \dots \xi_{q,r}^{[r_i+r_q-1]} \dots) = 0 \quad (56)$$

where  $q$  ranges from 1 to  $p$  and  $r$  ranges from 1 to  $n_q$ . As in the cases  $m = 2$  or  $m = 3$  addressed in Sections 4.1 and 4.2 respectively, the highest differentiation order of  $\xi_{i,k}$  (for any  $i, k$ ) appears only in the first group of equations (i.e. with  $i = r_1$  in (56)). This prevents us from solving a system of equations with respect to these high order derivatives to obtain a state space form. In fact, many of these high order derivatives can be expressed as functions of lower order derivatives. In this section, we obtain a state space form with derivatives of order  $2r_i$  for  $\xi_{i,k}$ . The total dimension of this state space form is  $2n$ .

We proceed along the following constructive proof. Sequentially, we differentiate the equations in (56) with respect to time. We start with the equations for  $i = p$  until the order  $r_{p-1} + r_p$  (in the factor of the second derivative of  $\mathcal{L}$ ) is replaced by  $2r_{p-1}$ . Together with the equations for  $i = p - 1$ , we solve the obtained system with respect to  $\xi_{p-1,l}^{(2r_{p-1})}$  and  $\xi_{p,l}^{(2r_p)}$ ,  $l$  ranging from 1 to  $n_{p-1}$ , and from 1 to  $n_p$ , respectively. This stresses that these are functions of the lower derivatives of  $\xi_{p-1,l}$  and  $\xi_{p,l}$ . Then, by induction, we differentiate sets of equations combining contiguous  $i$  (these sets are of increasing size) to obtain a system in the  $\xi_{i,l}^{2r_i}$ . At each step, the linear system has a matrix diagonally extracted from the Hessian (20) (modified by the presented reordering). We solve this system and, recursively, show that  $\xi_{i,j}^{2r_i}$  is a function  $G_{i,j}$  of  $(\eta, \mu, \dots, \xi_{k,l}^{[r_i+r_k-1]}, \dots, \xi_{i,1}^{[2r_i-1]}, \dots, \xi_{i,n_i}^{[2r_i-1]}, \dots, \xi_{p,n_p}^{[2r_p-1]})$ . This gives us the desired state space form as detailed in Theorem 3.

**Theorem 3.** For  $1 \leq i \leq p$  the following proposition (H1) holds:

(H1) For  $1 \leq j \leq n_i$ ,  $\xi_{i,j}^{2r_i}$  is a function  $G_{i,j}$  of  $(\eta, \mu, \dots, \xi_{k,q}^{[r_i+r_k-1]} \dots \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]})$ ,  $k$  and  $q$  being running indexes with  $1 \leq k \leq i - 1$  and  $1 \leq q \leq n_k$ .

Together with (16) and (23), this proposition, for  $1 \leq i \leq p$  (and  $1 \leq j \leq n_i$ ), gives the following state form which is satisfied by the  $\xi_{i,j}$ ,  $\eta$  and  $\mu$

$$\begin{aligned} \frac{d\xi_{i,j}}{dt} &= \xi_{i,j}^1 \\ &\vdots \\ \frac{d\xi_{i,j}^{2r_i-2}}{dt} &= \xi_{i,j}^{2r_i-1} \\ \frac{d\xi_{i,j}^{2r_i-1}}{dt} &= G_{i,j}(\eta, \mu, \dots, \xi_{k,q}^{r_i+r_k-1} \dots \xi_{i,1}^{2r_i-1} \dots \xi_{i,n_i}^{2r_i-1} \dots \xi_{p,n_p}^{2r_p-1}). \end{aligned}$$

Observing that, for  $k \leq i$  we have  $r_i + r_k - 1 \leq 2r_k - 1$ , the previous differential system is a state space form of dimension  $2n = 2n - 2r + 2 \sum r_i$ .

*Proof.* (H1) is verified for  $i = p$  (this is (56) with  $i = p, 1 \leq j \leq n_p$ , solved with respect to  $\xi_{p,l}^{2r_p}$ , observing that, for  $k < p, r_p + r_k \leq 2r_k - 1$ ). We now assume that (H1) holds for  $i \leq \tilde{i} \leq p$ ,  $\tilde{i}$  being the running index labelled by  $i$  in (H1).

To prove the induction on (H1), we show the more general inductions on (H2) and (H3) on  $i$  and  $s$ : for  $1 \leq i \leq p$  and  $r_i \leq s < r_{i-1}$  we note

(H2): For  $j \geq i, 1 \leq l \leq n_j, \xi_{j,l}^{(r_j+s)}$  is a function of  $\eta, \mu, \dots, \xi_{q,r}^{[r_q+s-1]} \dots \xi_{i,1}^{[2r_i-1]} \dots, \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}$ ,  $q$  being a running index with  $1 \leq q < i$ .

Observe that (H1) is verified if (H2) holds for  $j = i$  and  $s = r_i$ . We start the induction by observing that (H2) holds for  $i = p$  and  $s = r_p$ .

For  $1 \leq i \leq p$  and  $r_i \leq s < r_{i-1}$  we note

(H3): For  $j \geq i, 1 \leq h \leq n_j$ , there exists a function  $\tilde{G}_{j,h,s}$  such that

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{j,h} \partial v_{k,l}} \xi_{k,l}^{(r_k+s)} + \tilde{G}_{j,h,s}(\eta, \mu, \dots, \xi_{q,r}^{[r_q+s-1]} \dots \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}) = 0. \tag{57}$$

(H3) holds for  $i = p$  and  $s = r_p$ . We now assume that it is verified for the same  $i$  as the index used in the induction assumption on (H1). We make the same assumption on (H2). The induction proceed as follows: we prove that if (H3) holds for  $s$ , it is also valid for  $s + 1$ ; by solving the system with respect to the appropriate  $\xi_{k,l}^{(r_k+s+1)}$ , we prove that (H2) holds for  $s + 1$ , and hence (H2) and (H3) hold for  $r_i \leq s < r_{i-1}$ . To show that they are valid for the next  $i$ , *i.e.*  $i - 1$  and  $s = r_{i-1}$ , we increment  $s$  in (57) and add an extra set of equations. This augmented set is solved with respect to the appropriate  $\xi_{k,l}^{(r_{i-1}+r_k)}$  to prove that (H1) holds for  $i - 1$ . By carefully observing that the orders of differentiation in the  $\xi_{\tilde{i},l}$  for  $\tilde{i} \geq i$ , we prove that (H2) and (H3) hold for  $i - 1$  and  $s = r_{i-1}$ .

Now let us proceed. Differentiating (57) together with assumption (H1) for  $i \leq \tilde{i} \leq p$  (which allows us to eliminate  $\xi_{\tilde{i},l}^{[2r_{\tilde{i}}]}$ ) yields

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{j,h} \partial v_{k,l}} \xi_{k,l}^{(r_k+s+1)} + \tilde{G}_{j,h,s+1}(\eta, \mu, \dots, \xi_{q,r}^{[r_q+s]} \dots \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}) = 0 \quad (58)$$

which gives the induction on (H3) with respect to  $s$ . This system of  $n_i + n_{i+1} \dots + n_p$  equations is solved with respect to  $\xi_{k,l}^{(r_k+s+1)}$  for  $p \geq k \geq i$  and  $1 \leq l \leq n_k$  using the matrix  $\left[ \frac{\partial^2 \mathcal{L}}{\partial v_{j,h} \partial v_{k,l}} \right]$  for  $j \geq i$ ,  $1 \leq h \leq n_j$ ,  $p \geq k \geq i$  and  $1 \leq l \leq n_k$ , which is diagonally extracted from the Hessian (20) (modified by the presented reordering). This shows that, for  $p \geq k \geq i$ ,  $\xi_{k,l}^{(r_k+s+1)}$  is a function of  $\eta, \mu, \dots, \xi_{q,r}^{[r_q+s]} \dots, \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}$  which proves the induction of (H2) with respect to  $s$ . Hence, (H2) and (H3) hold for  $r_i \leq s < r_{i-1}$ . We now proceed by induction on the index  $i$  by adding an extra block of equations. We prove for  $j \geq i$  that both (H2) and (H3) hold for  $i - 1$  and  $s = r_{i-1}$ . To do so, we differentiate (57) once again to obtain  $s = r_{i-1}$ . This gives

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{j,h} \partial v_{k,l}} \xi_{k,l}^{(r_k+r_{i-1})} + \tilde{G}_{j,h,s}(\eta, \mu, \dots, \xi_{q,r}^{[r_q+r_{i-1}-1]} \dots \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}) = 0. \quad (59)$$

For  $q = i - 1$ ,  $\xi_{q,r}^{[r_q+r_{i-1}-1]} = \xi_{i-1,r}^{[2r_{i-1}-1]}$ . Hence, (59) can be written under the form

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{j,h} \partial v_{k,l}} \xi_{k,l}^{(r_k+r_{i-1})} + \tilde{G}_{j,h,s}(\eta, \mu, \dots, \xi_{q,r}^{[r_q+r_{i-1}-1]} \dots \xi_{i-1,1}^{[2r_{i-1}-1]} \dots \xi_{i-1,n_{i-1}}^{[2r_{i-1}-1]} \dots \xi_{p,n_p}^{[2r_p-1]}) = 0 \quad (60)$$

where  $q$  is a running index with  $1 \leq q < i - 1$ . On the other hand, let us consider (56) for  $i$  set to  $i - 1$  and  $1 \leq j \leq n_{i-1}$ . Renaming  $j$  as  $h$  to be consistent with (59) yields

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{i-1,h} \partial v_{k,l}} \xi_{k,l}^{(r_{i-1}+r_k)} + G_{i-1,h}(\eta, \mu, \dots, \xi_{q,r}^{[r_{i-1}+r_q-1]} \dots) = 0. \quad (61)$$

For  $q = i - 1$ , we have  $\xi_{q,r}^{[r_{i-1}+r_q-1]} = \xi_{i-1,r}^{[2r_{i-1}-1]}$ . For  $q \geq i$ ,  $r_{i-1} + r_q - 1 = r_q + s$  with  $s = r_{i-1} - 1$ . But (H2) assumes that for  $i \leq q \leq p$  and  $r_i \leq s \leq r_{i-1} - 1$ ,  $\xi_{q,r}^{(r_q+s)}$  is a function of  $(\eta, \mu, \dots, \xi_{k,l}^{[r_k+s-1]} \dots \xi_{i,1}^{[2r_i-1]}, \dots, \xi_{i,n_i}^{[2r_i-1]}, \dots, \xi_{p,n_p}^{[2r_p-1]})$ . In particular, for  $s = r_{i-1} - 1$ , we see that  $\xi_{q,r}^{[r_{i-1}+r_q-1]}$  is a function of  $(\eta, \mu, \dots, \xi_{k,l}^{[r_k+r_{i-1}-2]} \dots \xi_{i,1}^{[2r_i-1]}, \dots, \xi_{i,n_i}^{[2r_i-1]}, \dots, \xi_{p,n_p}^{[2r_p-1]})$  as long as  $q \geq i$ . Hence, (61) can be rewritten under the form

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{i-1,h} \partial v_{k,l}} \xi_{k,l}^{(r_{i-1}+r_k)} + \tilde{G}_{i-1,h}(\eta, \mu, \dots, \xi_{k,l}^{[r_k+r_{i-1}-1]} \dots \xi_{i-1,1}^{[2r_{i-1}-1]} \dots \xi_{i-1,n_{i-1}}^{[2r_{i-1}-1]} \dots \xi_{p,n_p}^{[2r_p-1]}) = 0. \quad (62)$$

Gathering (60) and (62) is equivalent to having (59) with  $i - 1 \leq j \leq p$  (a set of  $n_{i-1} + n_i \dots + n_p$  equations with coefficients diagonally extracted from the Hessian). We solve this system with respect to the  $n_{i-1} + n_i \dots + n_p$  unknowns  $\xi_{k,l}^{(r_k+r_{i-1})}$ ,  $i - 1 \leq k \leq p$  and  $1 \leq l \leq n_k$ . In particular, for  $k = i - 1$ , we see that  $\xi_{i-1,l}^{(2r_{i-1})}$  is a function of  $(\eta, \mu, \dots, \xi_{q,r}^{[r_q+r_{i-1}-1]}, \dots, \xi_{i-1,1}^{[2r_{i-1}-1]} \dots \xi_{i-1,n_{i-1}}^{[2r_{i-1}-1]} \dots \xi_{p,n_p}^{[2r_p-1]})$ , with  $1 \leq q < i - 1$ , and  $1 \leq r \leq n_q$ . This proves the induction for (H1) from index  $i$  to  $i - 1$ . Furthermore, we can initialize the inductions on (H2) and (H3) at  $i - 1$  and  $s = r_{i-1}$ . To do so, we need to extend (H2) and (H3) to  $i \leq j \leq p$ , with the index  $i$  now set to  $i - 1$ .

We observe that, for  $i_0 \leq j$  with  $i \leq i_0$ , and, for  $r_{i_0} \leq s \leq r_{i_0-1}$ ,  $\xi_{j,l}^{(r_j+s)}$  is a function of  $\eta, \mu, \dots, \xi_{q,r}^{[r_q+s-1]} \dots \xi_{i_0,1}^{[2r_{i_0}-1]} \dots \xi_{i_0,n_{i_0}}^{[2r_{i_0}-1]} \dots \xi_{p,n_p}^{[2r_p-1]}$ . Since  $i \leq i_0$ , it is also a function of  $\eta, \mu, \dots, \xi_{q,r}^{[r_q+s-1]} \dots \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}$ , because, for  $i_1 \leq i_0$ , the order  $2r_{i_1} - 1$  is greater than  $r_{i_0} + r_{i_1} - 1$ . Hence, all derivatives of  $\xi_{j,l}$  (with  $j \geq i$ ) from order  $2r_j$  up to  $r_j + r_{i-1} - 1$  are functions of  $\eta, \mu, \dots, \xi_{m,n}^{[r_m+r_{i-1}-1]} \dots \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}$ . Therefore, equation (58) can be rewritten under the form

$$\sum_{k=1}^p \sum_{l=1}^{n_k} \frac{\partial^2 \mathcal{L}}{\partial v_{i-1,h} \partial v_{k,l}} \xi_{k,l}^{(r_k+r_{i-1})} + \tilde{G}_{i-1,h,s}(\eta, \mu, \dots, \xi_{q,r}^{[r_q+r_{i-1}-1]} \dots \xi_{i,1}^{[2r_i-1]} \dots \xi_{i,n_i}^{[2r_i-1]} \dots \xi_{p,n_p}^{[2r_p-1]}, t) = 0.$$

Together with system (57), it yields a system as required by (H3) but with  $i$  set to  $i - 1$  and  $s = r_{i-1}$ . Solving it with respect to the variables  $\xi_{k,l}^{(r_k+r_{i-1})}$ , for  $k \geq i - 1$  and  $1 \leq l \leq n_k$ , we prove that (H2) holds for  $i - 1$  and  $s = r_{i-1}$ . This proves the inductions on (H1), (H2) and (H3). Observing that (H1) holds for  $1 \leq i \leq p$  we deduce the desired state space form which concludes the proof.  $\square$

As a final remark, we now show that the adjoint states are functions of the states reported in Theorem 3. First, we rewrite equation (25) by grouping equations of equal length  $r_i$ . We rewrite  $\lambda_{i,l}^{r_i-j}$  as  $\lambda_{i,l}^{r_i-j}$ , where  $r_i$  is a unique length of chain, and  $l$  indexes which chain is considered. We also group the  $\xi_k$  as  $\xi_{k,l}$ . These notations are similar to those used in Theorem 3. Equation (25) involves derivatives of  $\xi_{k,l}$  at the order  $r_k + j$ , with  $j$  ranging from 1 to  $r_i - 1$ . In particular,  $\xi_{k,l}^{(r_k+r_{i-1})}$  may be a derivative with a rank which is greater than the order  $2r_k - 1$ , which is the maximum order of differentiation of  $\xi_{k,l}$  making it a state variable in Theorem 3. Yet, Theorem 3 provides us with a state space form. Starting with the longest chain (*i.e.* the  $\xi_{k,l}$  with the highest order derivative in the state space form), this shows that, for a given  $\xi_{k,l}$ , the derivatives with an order greater than the order of the state space can be expressed as a function of the state variables. This implies that

**Corollary 1.** For  $1 \leq i \leq p$ ,  $0 \leq j \leq r_i - 1$ ,  $1 \leq l \leq n_p$ ,  $\lambda_{i,l}^{r_i-j}$  is a function of the state variables used in the differential system of Theorem 3.

*Proof.* For  $j = 1 \dots r_i - 1$ ,  $\lambda_{i,l}^{r_i-j}$ , one uses equation (25) and the fact that derivatives of  $\xi_{q,r}$  of order greater or equal to  $2r_q$  are actually functions of the state variable. This is obtained by differentiating enough times the last equation of Theorem 3, starting with the longest chain of integrators, *i.e.*, the smallest  $i$ . For  $j = 0$ , one uses equation (24), and the fact that  $X$  is a function of the state of Theorem 3.  $\square$

**Remark.** As pointed out in [4,5], there exist collocation methods for high order differential systems that are more accurate than collocation methods for first order systems (see discussion in Sect. 2). If we use the functions mentioned in Corollary 1 and compute numerical estimates of the adjoint variables from the obtained solutions, and if, additionally, these functions satisfy Lipschitz inequalities, then it is reasonable to expect that the numerical estimates for the adjoint variables obtained from such high order collocation methods will be more accurate than the estimates obtained from the equivalent state space collocation method. As a consequence, the numerical accuracy of adjoint-based methods for solving optimal control problems could benefit from this improved accuracy. This is a challenging point for further studies.

## 5. EXAMPLE: ROBOTIC LEG PROBLEM

We now propose an example to illustrate the proposed approach. The dynamics we consider corresponds to a mechanical system with second order dynamics studied in [12]. It is defined by the following under-actuated dynamics

$$m\ddot{r} - mr\dot{\psi}^2 = u_1 \quad (63)$$

$$J\ddot{\theta} = u_2 \quad (64)$$

$$mr^2\ddot{\psi} + 2mr\dot{r}\dot{\psi} = -u_2 \quad (65)$$

In the following, we note  $X = (r, \dot{r}, \theta, \dot{\theta}, \psi, \dot{\psi})^T$ ,  $u = (u_1, u_2)^T$ ,  $\dot{x} = f(x, u)$ . The cost function we consider is

$$\frac{1}{2} \int_0^T (X^T X + u^T u) dt.$$

Using  $\dot{r} = v_1$  and  $\dot{\theta} = v_2$ , the Hamiltonian is given as

$$H = \frac{1}{2} \left( X^T X + m^2(v_1 - r\dot{\psi}^2)^2 + J^2 v_2^2 \right) + \lambda_1 \dot{r} + \lambda_2 v_1 + \lambda_3 \dot{\theta} + \lambda_4 v_2 + \mu_1 \dot{\psi} - \mu_2 \left( \frac{Jv_2}{mr^2} + 2\frac{\dot{r}}{r}\dot{\psi} \right).$$

The adjoint dynamics is

$$\dot{\lambda}_1 = -H_r = -r + \left( m^2\dot{r} - m^2r\dot{\psi}^2 \right) \dot{\psi}^2 - 2\mu_2 \left( \frac{\dot{r}}{r^2}\dot{\psi} + \frac{Jv_2}{mr^3} \right) \quad (66)$$

$$\dot{\lambda}_2 = -H_{\dot{r}} = -\dot{r} - \lambda_1 + \mu_2 \frac{2}{r}\dot{\psi} \quad (67)$$

$$\dot{\lambda}_3 = -H_{\dot{\theta}} = -\dot{\theta} \quad (68)$$

$$\dot{\lambda}_4 = -H_{v_2} = -\dot{\theta} - \lambda_3 \quad (69)$$

$$\dot{\mu}_1 = -H_{\dot{\psi}} = -\dot{\psi} \quad (70)$$

$$\dot{\mu}_2 = -H_{\psi} = -\dot{\psi} - \mu_1 + 2 \left( m^2\dot{r} - m^2r\dot{\psi}^2 \right) \dot{\psi}r + 2\mu_2 \frac{\dot{r}}{r}. \quad (71)$$

Our claim, according to Theorem 3 (with  $p = 1$ ,  $n_1 = 2$  and  $r_1 = 2$ ) and Corollary 1, is twofold. First, all the original 12 variables from the stationarity conditions (6 primal and 6 adjoint variables) can be rewritten using only 6 variables. Further, the 12 first order optimality conditions can be rewritten as 2 fourth order equations and 4 first order equations (as in Prop. 2 for the case where the two chain lengths are equal).

In a first step, we recover the  $\lambda_2$  and  $\lambda_4$  adjoint states from the  $H_{v_1} = 0$  and  $H_{v_2} = 0$  equations. These give

$$\lambda_2 = -m^2\dot{r} + m^2r\dot{\psi}^2 \quad (72)$$

$$\lambda_4 = -J^2\dot{\theta} + J \frac{\mu_2}{mr^2} \quad (73)$$

which corresponds to equation (24). Next, the first time derivatives of  $H_{v_1} = 0$  and  $H_{v_2} = 0$  give

$$\dot{\lambda}_2 = -m^2r^{(3)} + 2m^2r\dot{\psi}\ddot{\psi} + m^2\dot{r}\dot{\psi}^2$$

$$\dot{\lambda}_4 = -J^2\theta^{(3)} + J \frac{\dot{\mu}_2}{mr^2} - 2\frac{J}{mr^3}\mu_2.$$

After substitution in (67) and (69) respectively, one directly gets

$$\lambda_1 = -\dot{r} + \mu_2 \frac{2}{r} \dot{\psi} + m^2 r^{(3)} - 2m^2 r \dot{\psi} \ddot{\psi} - m^2 \dot{r} \dot{\psi}^2 \quad (74)$$

$$\lambda_3 = -\dot{\theta} + J^2 \theta^{(3)} - J \frac{\dot{\mu}_2}{mr^2} + 2J \frac{\mu_2 \dot{r}}{mr^3} = -\dot{\theta} + J^2 \theta^{(3)} + \frac{J}{mr^2} (\dot{\psi} + \mu_1 + 2\lambda_2 \dot{\psi} r) \quad (75)$$

which corresponds to equation (25). The linearizable part of the dynamics and its adjoint have been fully explored through the previous equations. One can readily get the two fourth order equations corresponding to these last necessary conditions by differentiating (74) and (75) with respect to time. This gives

$$r^{(4)} = 4 \ddot{\psi} \dot{\psi} \dot{r} + 2 r \ddot{\psi}^2 - 2 \ddot{r} \dot{\psi}^2 + \frac{\ddot{r} - r}{m^2} + 3 r \dot{\psi}^4 + 2 \frac{\dot{\psi}^2 + \dot{\psi} \mu_1}{m^2 r} + 2 r \dot{\psi} \psi^{(3)} \quad (76)$$

$$\theta^{(4)} = 2 \frac{\dot{r} \dot{\psi} + \dot{r} \mu_1}{mr^3 J} - 2 \frac{m \dot{\psi} \dot{r} \ddot{r}}{r^2 J} + \frac{\ddot{\theta} - \theta}{J^2} + \frac{\psi - \ddot{\psi}}{mr^2 J} + 2 \frac{m \ddot{\psi} \ddot{r} + m \dot{\psi} r^{(3)}}{r J} - 6 \frac{m \dot{\psi}^2 \ddot{\psi}}{J}. \quad (77)$$

Yet, it is possible to simplify these expressions further to obtain an equation which is similar to (32). The  $\ddot{\psi}$  and  $\psi^{(3)}$  terms are not needed. As in the proofs of Lemma 1 and Theorem 2, they can be derived from the following equation and its time derivative (arising from (64) and (65))

$$\ddot{\psi} = -\frac{J\ddot{\theta}}{mr^2} - 2\frac{\dot{r}}{r}\dot{\psi} \quad (78)$$

which gives an expression  $\psi^{(3)} = p(r, \dot{r}, \ddot{r}, \ddot{\theta}, \theta^{(3)}, \dot{\psi})$ . After substitution in (76) and (77), we can gather the following necessary conditions

$$\left. \begin{aligned} r^{(4)} &= g_1(r, \dot{r}, \ddot{r}, \ddot{\theta}, \theta^{(3)}, \dot{\psi}, \mu_1) \\ \theta^{(4)} &= g_2(r, \dot{r}, \ddot{r}, r^{(3)}, \theta, \ddot{\theta}, \dot{\psi}, \mu_1) \\ \ddot{\psi} &= -\frac{J\ddot{\theta}}{mr^2} - 2\frac{\dot{r}}{r}\dot{\psi} \\ \dot{\mu}_1 &= -\psi \\ \dot{\mu}_2 &= -\dot{\psi} - \mu_1 + 2 \left( m^2 \ddot{r} - m^2 r \dot{\psi}^2 \right) \dot{\psi} r + 2\mu_2 \frac{\dot{r}}{r} \end{aligned} \right\} \quad (79)$$

where  $g_1$  and  $g_2$  are given by the following

$$\begin{aligned} r^{(4)} &= -4 \frac{(J\ddot{\theta} + 2\dot{r}\dot{\psi}mr)}{mr^2} \dot{\psi} \dot{r} + 2 \frac{(J\ddot{\theta} + 2\dot{r}\dot{\psi}mr)^2}{r^3 m^2} - 2 \ddot{r} \dot{\psi}^2 + \frac{\ddot{r} - r}{m^2} \\ &\quad + 3 r \dot{\psi}^4 + 2 \frac{\dot{\psi}^2 + \dot{\psi} \mu_1}{m^2 r} + 2 r \dot{\psi} \left( \frac{-J\theta^{(3)} r + 4 \dot{r} J\ddot{\theta} + 6 \dot{r}^2 \dot{\psi} m r - 2 \dot{\psi} m \ddot{r} r^2}{mr^3} \right) \\ \theta^{(4)} &= \frac{2\dot{r}(2\dot{\psi} + \mu_1)}{Jmr^3} - 6 \frac{m\dot{r}\dot{\psi}\ddot{r}}{Jr^2} + \frac{2m\dot{\psi}(r^{(3)} + 6\dot{r}\dot{\psi}^2)}{Jr} + \frac{\ddot{\theta} - \theta}{J^2} + \frac{\ddot{\theta}}{m^2 r^4} - 2 \frac{\ddot{r}\ddot{\theta}}{r^3} + 6 \frac{\dot{\psi}^2 \ddot{\theta}}{r^2} + \frac{\psi}{Jmr^2}. \end{aligned}$$

System (79) is a set of necessary conditions for the primal and adjoint equations corresponding to the original stationarity conditions. This corresponds to Theorem 2, or Theorem 3 with  $p = 1$ . As a remark, one can notice that we do not have to differentiate further  $r$  or  $\theta$  to eliminate derivatives of order greater than 4 (as in the proof of Th. 3) because we have only one length of chain ( $p = 1$ ).

## 6. CONCLUSIONS AND PERSPECTIVES

In this paper, we give explicit ways to derive a state space form set of stationarity conditions that involve a reduced number of variables. We stress that numerous adjoint variables can be computed using zero dynamics states, corresponding adjoints, and, most importantly, successive derivatives of the linearizing outputs. Thus, in many optimal control problems, numerous variables (primal and adjoint variables) can be recovered through such relations and need not be considered. Without loss of generality, constraints (*e.g.* endpoints constraints) can be addressed similarly but, in their general formulation, are out of the scope of the work presented here. This is a direction for further developments. On the numerical side, this number of variables reduction proved efficient in direct methods of trajectory optimization. Future directions of this research should imply development of specific numerical tools, *e.g.* collocation methods for higher order systems. As suggested in [4,5] and demonstrated in Section 2, efficiency can be gained by applying those directly to higher order differential equations rather than corresponding first-order systems.

*Acknowledgements.* The authors would like to thank Emmanuel Trélat and Laurent Praly for fruitful discussions and constructive comments.

## REFERENCES

- [1] A.A. Agrachev and A.V. Sarychev, On abnormal extremals for Lagrange variational problems. *J. Math. Systems Estim. Control* **1** (1998) 87–118.
- [2] S.K. Agrawal and N. Faiz, A new efficient method for optimization of a class of nonlinear systems without Lagrange multipliers. *J. Optim. Theor. Appl.* **97** (1998) 11–28.
- [3] U.M. Ascher, J. Christiansen and R.D. Russel, Collocation software for boundary-value ODE's. *ACM Trans. Math. Software* **7** (1981) 209–222.
- [4] U.M. Ascher, R.M.M. Mattheij and R.D. Russell, *Numerical solution of boundary value problems for ordinary differential equations*. Prentice Hall Series in Computational Mathematics Prentice Hall, Inc., Englewood Cliffs, NJ (1988).
- [5] U.M. Ascher, R.M.M. Mattheij and R.D. Russell, *Numerical solution of boundary value problems for ordinary differential equations*, *Classics in Applied Mathematics* **13**. Society for Industrial and Applied Mathematics (SIAM) (1995).
- [6] J.T. Betts, Survey of numerical methods for trajectory optimization. *J. Guid. Control Dyn.* **21** (1998) 193–207.
- [7] J.T. Betts, *Practical methods for optimal control using nonlinear programming*, *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2001).
- [8] B. Bonnard and M. Chyba, *Singular trajectories and their role in control theory*, *Mathématiques & applications* **40**. Springer-Verlag-Berlin-Heidelberg-New York (2003).
- [9] A.E. Bryson and Y.C. Ho, *Applied Optimal Control*. Ginn and Company (1969).
- [10] R. Bulirsch, F. Montrone and H.J. Pesch, Abort landing in the presence of windshear as a minimax optimal control problem, part 2: Multiple shooting and homotopy. *J. Optim. Theor. Appl.* **70** (1991) 223–254.
- [11] R. Bulirsch, E. Nerz, H.J. Pesch and O. von Stryk, Combining direct and indirect methods in optimal control: Range maximization of a hang glider, in *Optimal Control*, R. Bulirsch, A. Miele, J. Stoer and K.H. Well Eds., *International Series of Numerical Mathematics*, Birkhäuser **111** (1993).
- [12] F. Bullo and A.D. Lewis, *Geometric Control of Mechanical Systems, Modeling, Analysis, and Design for Simple Mechanical Control Systems*, *Texts in Applied Mathematics* **49**. Springer-Verlag (2004).
- [13] C.I. Byrnes and A. Isidori, Asymptotic stabilization of minimum phase nonlinear systems. *IEEE Trans. Automat. Control* **36** (1991) 1122–1137.
- [14] F. Chaplais and N. Petit, Inversion in indirect optimal control, in *Proc. of the 7th European Control Conf.* (2003).
- [15] M. El-Kady, A Chebyshev finite difference method for solving a class of optimal control problems. *Int. J. Comput. Math.* **80** (2003) 883–895.
- [16] F. Fahroo and I.M. Ross, Direct trajectory optimization by a Chebyshev pseudo-spectral method. *J. Guid. Control Dyn.* **25** (2002) 160–166.
- [17] N. Faiz, S.K. Agrawal and R.M. Murray, Differentially flat systems with inequality constraints: An approach to real-time feasible trajectory generation. *J. Guid. Control Dyn.* **24** (2001) 219–227.
- [18] M. Fliess, J. Lévine, P. Martin and P. Rouchon, Flatness and defect of nonlinear systems: introductory theory and examples. *Int. J. Control* **61** (1995) 1327–1361.
- [19] M. Fliess, J. Lévine, P. Martin and P. Rouchon, A Lie-Bäcklund approach to equivalence and flatness of nonlinear systems. *IEEE Trans. Automat. Control* **44** (1999) 922–937.

- [20] P.E. Gill, W. Murray, M.A. Saunders and M.A. Wright, *User's Guide for NPSOL 5.0: A Fortran Package for Nonlinear Programming*. Systems Optimization Laboratory, Stanford University, Stanford, CA 94305 (1998).
- [21] C. Hargraves and S. Paris, Direct trajectory optimization using nonlinear programming and collocation. *AIAA J. Guid. Control* **10** (1987) 338–342.
- [22] A. Isidori, *Nonlinear Control Systems*. Springer, New York, 2nd edn. (1989).
- [23] A. Isidori, *Nonlinear Control Systems II*. Springer, London-Berlin-Heidelberg (1999).
- [24] D.G. Luenberger, *Optimization by vector spaces methods*. Wiley-Interscience (1997).
- [25] M.B. Milam, *Real-time optimal trajectory generation for constrained systems*. Ph.D. thesis, California Institute of Technology (2003).
- [26] M.B. Milam, K. Mushambi and R.M. Murray, A new computational approach to real-time trajectory generation for constrained mechanical systems, in *IEEE Conference on Decision and Control* (2000).
- [27] M.B. Milam, R. Franz and R.M. Murray, Real-time constrained trajectory generation applied to a flight control experiment, in *Proc. of the IFAC World Congress* (2002).
- [28] R. Montgomery, Abnormal minimizers. *SIAM J. Control Optim.* **32** (1994) 1605–1620.
- [29] R.M. Murray, J. Hauser, A. Jadbabaie, M.B. Milam, N. Petit, W.B. Dunbar and R. Franz, Online control customization via optimization-based control, in *Software-Enabled Control, Information technology for dynamical systems*, T. Samad and G. Balas Eds., Wiley-Interscience (2003) 149–174.
- [30] T. Neckel, C. Talbot and N. Petit, Collocation and inversion for a reentry optimal control problem, in *Proc. of the 5th Intern. Conference on Launcher Technology* (2003).
- [31] H. Nijmeijer and A.J. van der Schaft, *Nonlinear Dynamical Control Systems*. Springer-Verlag (1990).
- [32] J. Oldenburg and W. Marquardt, Flatness and higher order differential model representations in dynamic optimization. *Comput. Chem. Eng.* **26** (2002) 385–400.
- [33] N. Petit, M.B. Milam and R.M. Murray, Inversion based constrained trajectory optimization, in *5th IFAC Symposium on Nonlinear Control Systems* (2001).
- [34] I.M. Ross and F. Fahroo, Pseudospectral methods for optimal motion planning of differentially flat systems, in *Proc. of the 41th IEEE Conf. on Decision and Control* (2002).
- [35] I.M. Ross, J. Rea and F. Fahroo, Exploiting higher-order derivatives in computational optimal control, in *Proc. of the 2002 IEEE Mediterranean Conference* (2002).
- [36] H. Seywald, Trajectory optimization based on differential inclusion. *J. Guid. Control Dyn.* **17** (1994) 480–487.
- [37] H. Seywald and R.R. Kumar, Method for automatic costate calculation. *J. Guid. Control Dyn.* **19** (1996) 1252–1261.
- [38] H. Shen and P. Tsiotras, Time-optimal control of axi-symmetric rigid spacecraft using two controls. *J. Guid. Control Dyn.* **22** (1999) 682–694.
- [39] H. Sira-Ramirez and S.K. Agrawal, *Differentially Flat Systems*. Control Engineering Series, Marcel Dekker (2004).
- [40] M.C. Steinbach, *Optimal motion design using inverse dynamics*. Technical report, Konrad-Zuse-Zentrum für Informationstechnik Berlin (1997).
- [41] M.J. van Nieuwstadt. *Trajectory generation for nonlinear control systems*. Ph.D. thesis, California Institute of Technology (1996).
- [42] O. von Stryk and R. Bulirsch, Direct and indirect methods for trajectory optimization. *Ann. Oper. Res.* **37** (1992) 357–373.