

# Finite-Horizon Optimal Control for Linear and Nonlinear Systems Relying on Constant Optimal Costate

Lorenzo Tarantino , Mario Sassano , Senior Member, IEEE, Sergio Galeani , and Alessandro Astolfi , Fellow, IEEE

**Abstract**—A class of finite-horizon optimal control problems, the solution of which relies on a time-varying change of coordinates that incorporates the transition matrix of the system linearized along the current estimate of the optimal process, is studied. The transformed dynamics exhibit a *constant* optimal costate. Differently from existing methods that hinge upon similar tools, the proposed strategy does not require at each step the (numerical) solution of a two-point boundary value problem or of a time-varying Riccati equation, and only the solution of a linear initial value problem is needed. The method is firstly illustrated in the setting of linear dynamics and quadratic cost for which the construction permits the identification of a class of problems in which the solution to the underlying (quadratic) Differential Riccati Equation exhibit a separation between homogeneous and particular contributions.

**Index Terms**—Iterative learning control, nonlinear systems, optimal control.

## I. INTRODUCTION

THE problem of controlling the state of a system in an optimal way from a given initial state to a certain terminal state, within a prescribed time interval, which is fixed *a priori*, has gained a central role in systems and control theory over the past decades [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. In the second half of the last century, two approaches have been introduced for solving such a class of problems, i.e., *Dynamic Programming* (DP) [11], [12] and *Pontryagin's Minimum Principle* (PMP) [13]. The former provides necessary and sufficient conditions of optimality, leading to an optimal

feedback solution characterized via a nonlinear partial differential equation (PDE), i.e. the so-called *Hamilton–Jacobi–Bellman* (HJB) equation. Since the optimal solution, as well as the optimal cost, are characterized for any initial condition in the state space, methods inspired by DP are particularly appealing, whereas the main drawback of this approach is represented by the fact that closed-form solutions to the HJB equation can be seldom obtained. Methods relying on PMP, on the other hand, provide only necessary conditions of optimality, allowing to identify *candidate* optimal solutions, referred to as *extremals*, and this approach essentially leads to open-loop strategies, since the characterization of the optimal control law heavily relies upon the knowledge of the initial condition of the underlying system. Such necessary conditions are characterized in terms of *ordinary differential equations*, instead of PDEs, which must be satisfied by the optimal process, together with an auxiliary variable, the so-called *costate*. It is worth mentioning that, in the case of infinite-horizon problems (see, e.g., [14], [15]), this approach recovers the computational complexity of DP, since determining the correct boundary conditions requires the solution of the HJB PDE. On the contrary, in the finite-horizon setting, a solution to a two-point boundary value problem (TPBVP) is needed. As it has been recognized, the major challenge in solving optimal control problems lies in their intrinsic infinite-dimensional nature, combined with the presence of nonlinearities. To address this challenge, in the past decades a large number of algorithms have been envisioned: These can be essentially classified as *direct* or *indirect* (see [16] and [17] for further discussion on this distinction). A direct method converts the original problem into a sequence of nonlinear optimization problems, for example, by using *Chebyshev polynomials* (see, e.g., [18], [19]). As a matter of fact, a direct method transforms an infinite-dimensional problem into a sequence of finite-dimensional problems, which are in general easier to handle. Indirect methods instead rely essentially on necessary conditions such as those derived from the *Calculus of Variations* [6] and the problem is then handled in the infinite-dimensional context. On the other hand, since Newton method provides particularly efficient iterative strategies to tackle nonlinear optimization tasks, methods based on the linearization of the underlying plant along certain candidate trajectories have emerged with the objective of replicating Newton's arguments in an infinite-dimensional context. This can be

Manuscript received 26 July 2023; accepted 18 October 2023. Date of publication 31 October 2023; date of current version 29 March 2024. This work was supported by the Italian Ministry for Research in the framework of the 2020 Program for Research Projects of National Interest under Grant 2020RTWES4. Recommended by Senior Editor T. Iwasaki. (Corresponding author: Lorenzo Tarantino.)

Lorenzo Tarantino, Mario Sassano, and Sergio Galeani are with the Dipartimento di Ingegneria Civile ed Ingegneria Informatica, Università degli Studi di Roma, 00133 Roma, Italy (e-mail: lorenzo.tarantino@uniroma2.it; mario.sassano@uniroma2.it; sergio.galeani@uniroma2.it).

Alessandro Astolfi is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BX London, U.K., and with the Dipartimento di Ingegneria Civile ed Ingegneria Informatica, Università degli Studi di Roma, 00133 Roma, Italy (e-mail: a.astolfi@ic.ac.uk).

Digital Object Identifier 10.1109/TAC.2023.3328925

accomplished in two different ways (see [20] and [21]). The first one consists in the linearization of the underlying dynamics, together with the expansion of the cost functional in a Taylor series up to second order terms, solving then a *Linear Quadratic Regulator* (LQR) problem at each iteration (see [22], [23], [24], [25], and [26], where the latter refers to the context of *stochastic systems*). The second one relies on imposing some necessary conditions, such as those derived from PMP, and then subsequently linearizing the resulting dynamics (see, e.g., [3] and [1]). Although this latter way has been extensively studied, since there exist efficient numerical techniques for solving TPBVP (see, e.g., [27], [28], [29], [30], [31], [32], [33], [34]), the major drawback of this approach is represented by the fact that the steps involved lead to losing sight on the behavior of the cost functional.

The strategy proposed here follows the first of the two approaches described above to replicate Newton's argument in an infinite-dimensional setting, which in principle relies on solving a time-varying LQR problem. This task could appear discouraging, since the solution of this problem relies on the solution of a time-varying *Riccati equation*, which is a quadratic differential equation. Nonetheless, by borrowing ideas from [1] and [35], in which a class of *zero-sum differential games* is solved by means of an auxiliary variable related to the underlying dynamics via a change of coordinates, it is possible to replace the Riccati equation with a linear, time-varying, differential equation with only initial boundary conditions. This makes the proposed approach computationally viable. Furthermore, this strategy also allows assessing the behavior of the cost functional after each iteration. Therefore, the steps involved lead to global convergence to (possibly local) extremal. Since the resulting iterative scheme is characterized by a gradient-descent nature, it follows that the choice of the optimal step size along the descent direction plays a crucial role in terms of convergence rate of the entire scheme. This issue can be addressed by relying on approximate as well as exact line search methods or by borrowing ideas used in the field of *machine learning* in the training process of neural networks, by relying on heuristic methods, such as the RMSProp scheme [36].

Preliminary results have been presented in [37]. Differently from [37], herein we provide detailed proofs of all the technical results, together with further insights on the structure of the solution. Moreover, the main ideas are firstly presented in the case of the LQR problem, allowing to identify a class of such problems with interesting properties. An extensive comparison is provided between strategies for obtaining the optimal step size at each iteration of the descent algorithm. Finally, the optimal control problem for a mechanical system on a redundant task is solved using the proposed approach.

The rest of the article is organized as follows. The class of problems considered, together with some preliminaries and the discussion of the main ideas by revisiting the LQR problem, leading to a closed-form solution of the *Differential Riccati Equation*, are introduced in Section II. The introduction of the algorithm that allows solving the finite-horizon optimal control problem by exploiting the constant optimal costate is provided in Section III. Further discussion on the proposed algorithm,

together with some implementation aspects are given in Section IV. A discussion on the steps involved by the proposed approach and those of the quasilinearization scheme is provided in Section V. Finally, an example of application on a mechanical system on a redundant task and some concluding remarks are given in Sections VI and VII, respectively.

*Notation:* Given two matrices  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{p \times q}$ , the Kronecker product between  $A$  and  $B$ , denoted as  $A \otimes B \in \mathbb{R}^{pm \times qn}$ , is the block matrix given by

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}$$

where  $a_{ij}$  indicates the  $(i, j)$ th element of the matrix  $A$ . The vectorization of a matrix  $C \in \mathbb{R}^{n \times m}$ , denoted with the operator  $\text{vec}(\cdot)$  consists in stacking the columns of the matrix  $C$ , i.e.,

$$\text{vec}(C) = [c_{11}, \dots, c_{n1}, \dots, c_{1m}, \dots, c_{nm}]^T.$$

A positive definite function is denoted as  $q: \mathbb{R}^n \rightarrow \mathbb{R}_{>0}$ , whereas a positive semidefinite function is denoted as  $q: \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ .

## II. PROBLEM STATEMENT AND PRELIMINARIES

The main purpose of this article is to study a class of finite-horizon optimal control problems described by the cost functional

$$J(u(\cdot)) = q_f(x(t_f)) + \int_{t_0}^{t_f} q(x(t)) + \frac{1}{2} \|u(t)\|_R^2 dt \quad (1)$$

subject to the dynamic constraints

$$\dot{x} = f(x) + g(x)u, \quad x(t_0) = x_0 \quad (2)$$

where  $x(t) \in \mathbb{R}^n$  denotes the state of the system,  $u(t) \in \mathbb{R}^m$  represents the control input, and  $t_f > 0$  is the fixed final time. Moreover, the functions  $q: \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$  and  $q_f: \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$  are assumed to be sufficiently smooth, namely, whenever their (higher-order) derivatives are required, it is assumed that such derivatives are continuous. The matrix  $R \in \mathbb{R}^{m \times m}$ , with  $R = R^T$ , is positive definite. Finally, suppose that the vector field  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  and the mapping  $g: \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$  are sufficiently smooth. The problem under investigation is formally introduced in the following statement.

*Problem 1:* Let  $x_0 \in \mathbb{R}^n$  be given. The *finite-horizon optimal control* problem consists in determining a control law  $u \in C^0(t_0, t_f)$ , that minimizes (1) along the trajectories of (2). ◻

Problem 1 is studied by first considering the case of linear dynamics and quadratic cost functional, which is briefly revisited in the following section. The genuinely nonlinear formulation of Problem 1 is subsequently addressed by first letting, for illustrative purposes (see Theorem 1 below),  $q_f$  be the squared norm of the terminal state  $x_f$  and  $q \equiv 0$ . These structural assumptions are then removed, thus yielding a solution to Problem 1 in the general setting given by (1) and (2) (see Corollaries 1 and 2 below).

### A. Revisiting the Finite Horizon LQR Problem

The main intuition behind the arguments employed to tackle the problem defined by (1), (2) is here discussed in the specially structured case in which the dynamics (2) are linear and time-invariant, namely

$$\dot{x} = Ax + Bu, \quad x(t_0) = x_0 \quad (3)$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times m}$ , and the running and terminal costs in (1) are quadratic functions of the state, namely

$$q(x) = \frac{1}{2} \|x\|_Q^2, \quad q_f(x) = \frac{1}{2} \|x\|_{Q_f}^2 \quad (4)$$

with  $Q \in \mathbb{R}^{n \times n}$ ,  $Q = Q^\top$ ,  $Q \geq 0$  and  $Q_f \in \mathbb{R}^{n \times n}$ ,  $Q_f = Q_f^\top$ ,  $Q_f \geq 0$ . It is well known (see, for instance, [6]) that, in such setting, the optimal solution is described by a time-varying state feedback

$$u^*(t, x) = -R^{-1}B^\top P(t)x \quad (5)$$

where  $P(t) \in \mathbb{R}^{n \times n}$ , with  $P(t) = P(t)^\top \geq 0$ , solves the *Differential Riccati Equation* (DRE)

$$\dot{P}(t) = -A^\top P(t) - P(t)A + P(t)BR^{-1}B^\top P(t) - Q \quad (6a)$$

$$P(t_f) = Q_f. \quad (6b)$$

It is worth mentioning that (5) solves Problem 1 for all  $x_0 \in \mathbb{R}^n$ . Furthermore, the minimal cost attained from a generic time  $t$  and state  $x$  is yielded by the *value function*

$$V(t, x) = \frac{1}{2}x^\top P(t)x$$

for all  $x \in \mathbb{R}^n$  and  $t \in [t_0, t_f]$ . Whenever the state of the system is penalized only at the final time in (1), a particularly insightful structure of the solution, which can be determined in closed form, emerges. The latter then constitutes the main ingredient toward the extension to the nonlinear setting discussed in Section III. To provide a concise statement of the following result, let  $\bar{B}(t) := e^{A(t_f-t)}B$ , for  $t \in [t_0, t_f]$ .

*Proposition 1:* Consider the system (3) and the cost functional (1) with  $q$  and  $q_f$  as in (4). Suppose that  $Q = 0$  and  $Q_f > 0$ . Fix  $x_0 \in \mathbb{R}^n$ . Then, the solution to Problem 1 is provided by

$$u^*(t) = -R^{-1}\bar{B}(t)^\top K(t)^{-1}e^{A(t_f-t)}x(t) \quad (7)$$

with  $K(t)$  defined as  $K(t) := Q_f^{-1} + \int_t^{t_f} S(\tau)d\tau$ , and  $S(t)$  defined as  $S(t) := \bar{B}(t)R^{-1}\bar{B}(t)^\top$ , for  $t \in [t_0, t_f]$ .  $\triangle$

*Proof:* Consider the time-varying change of coordinates defined as

$$z(t) = e^{A(t_f-t)}x(t) \quad (8)$$

for  $t \in [t_0, t_f]$ . In such transformed coordinates, the cost functional (1), (4) with  $Q = 0$  becomes

$$J(u(\cdot)) = \frac{1}{2} \|z(t_f)\|_{Q_f}^2 + \frac{1}{2} \int_{t_0}^{t_f} \|u(t)\|_R^2 dt \quad (9)$$

while the underlying dynamics are

$$\begin{aligned} \dot{z}(t) &= -e^{A(t_f-t)}Ax(t) + e^{A(t_f-t)}\dot{x}(t) = e^{A(t_f-t)}Bu(t) \\ &=: \bar{B}(t)u(t). \end{aligned} \quad (10)$$

It is then straightforward to note that the *Hamiltonian* function associated with the problem in the transformed coordinates reduces to

$$\bar{H}(u, z, \lambda, t) = \frac{1}{2} \|u\|_R^2 + \lambda^\top \bar{B}(t)u \quad (11)$$

which is, in fact, independent of the state  $z$ . Therefore, by relying on standard results (see, e.g., [6]), necessary and sufficient conditions of optimality for the problem described by (9), (10) imply that the optimal process  $(u^*, z^*)$ , satisfies the Hamiltonian dynamics

$$\dot{z} = \frac{\partial \bar{H}}{\partial \lambda}(u, z, \lambda, t) = \bar{B}(t)u \quad (12a)$$

$$\dot{\lambda} = -\frac{\partial \bar{H}^\top}{\partial z}(u, z, \lambda, t) = 0 \quad (12b)$$

together with the boundary conditions

$$z(t_0) = z_0 \quad (13a)$$

$$\lambda(t_f) = Q_f z(t_f). \quad (13b)$$

The control law  $u$  in (12) is instead obtained via the (unconstrained) minimization of  $\bar{H}$ , yielding

$$u(t) = -R^{-1}\bar{B}(t)^\top \lambda(t). \quad (14)$$

By replacing the control law (14) into the dynamics (12a), and noting that, since the optimal costate is constant (see (12b)),  $\lambda(t) = \lambda(t_f) = Q_f z(t_f)$  for all  $t \in [t_0, t_f]$ , one obtains

$$\begin{aligned} z(t) &= z(t_f) + \int_t^{t_f} \bar{B}(\tau)R^{-1}\bar{B}(\tau)^\top d\tau \lambda(t) \\ &= Q_f^{-1}\lambda(t) + \int_t^{t_f} \bar{B}(\tau)R^{-1}\bar{B}(\tau)^\top d\tau \lambda(t) \\ &= \left( Q_f^{-1} + \int_t^{t_f} \bar{B}(\tau)R^{-1}\bar{B}(\tau)^\top d\tau \right) \lambda(t). \end{aligned} \quad (15)$$

From (15) it follows that:

$$\begin{aligned} \lambda(t) &= \left( Q_f^{-1} + \int_t^{t_f} \bar{B}(\tau)R^{-1}\bar{B}(\tau)^\top d\tau \right)^{-1} z(t) \\ &= K(t)^{-1}z(t). \end{aligned} \quad (16)$$

Substitution of (16) into the control law in (14) leads to

$$u(t) = -R^{-1}\bar{B}(t)^\top \left( Q_f^{-1} + \int_t^{t_f} S(\tau)d\tau \right)^{-1} z(t) \quad (17)$$

thus showing the claim.  $\square$

By closely inspecting the structure of (7) in comparison with (5), one immediately realizes that the former implicitly suggests the structure of the matrix  $P$  solving the DRE in closed form in this setting. The above intuition is summarized in the following statement.

*Proposition 2:* Suppose that  $Q_f > 0$  and let

$$P_h(t) = e^{A^\top(t_f-t)} \left( Q_f^{-1} + \int_t^{t_f} S(\tau)d\tau \right)^{-1} e^{A(t_f-t)}. \quad (18)$$

Then  $P_h$  in (18) is a closed-form solution of the DRE (6) with  $Q = 0$ , namely

$$\dot{P}_h(t) = -A^\top P_h(t) - P_h(t)A + P_h(t)BR^{-1}B^\top P_h(t) \quad (19a)$$

$$P_h(t_f) = Q_f. \quad (19b)$$

△

*Proof:* To begin with, observe that  $P_h$  in (18) can be rewritten as  $P_h(t) = e^{A^\top(t_f-t)}K(t)^{-1}e^{A(t_f-t)}$ , by definition of  $K$  as in (7), namely

$$K(t) = Q_f^{-1} + \int_t^{t_f} S(\tau)d\tau. \quad (20)$$

Note now that

$$\begin{aligned} \frac{d}{dt}K^{-1}(t) &= -K^{-1}(t)\left(\frac{d}{dt}K(t)\right)K^{-1}(t) \\ &= K^{-1}(t)\bar{B}(t)R^{-1}\bar{B}(t)^\top K^{-1}(t). \end{aligned}$$

Therefore, replacing the above identity into the time derivative of the candidate solution  $P_h$  in (18), and recalling that  $S(\tau) = e^{A(t_f-\tau)}BR^{-1}B^\top e^{A^\top(t_f-\tau)}$ , yields

$$\begin{aligned} \dot{P}_h(t) &= -A^\top P_h(t) - P_h(t)A \\ &\quad + e^{A(t_f-t)}\frac{d}{dt}\left(Q_f^{-1} + \int_t^{t_f} S(\tau)d\tau\right)^{-1}e^{A^\top(t_f-t)} \\ &= -A^\top P_h(t) - P_h(t)A + P_h(t)BR^{-1}B^\top P_h(t). \end{aligned}$$

Finally, note that the boundary condition is satisfied by construction, hence the proof is complete. □

A few observations about Proposition 2 are in order. First note that, although the DRE (6) is quadratic instead of linear, the statement of the Proposition 2 implicitly entails that  $P_h(t)$  in (18) solves the homogeneous part of (6), obtained by letting the constant terms equal to zero in the vector field of (6). One may then wonder in which circumstances the DRE enjoys a separation into homogeneous and particular solutions, as for linear differential equations. This is precisely illustrated in the result below. To provide a concise statement, let  $B^\perp \in \mathbb{R}^{n \times (n-m)}$  denote a full column rank matrix with the property that  $B^\top B^\perp = 0$ .

*Proposition 3:* Consider the Differential Riccati Equation (6). Suppose that there exists a matrix  $X \in \mathbb{R}^{(n-m) \times (n-m)}$ ,  $X = X^\top$  such that

$$A^\top B^\perp X (B^\perp)^\top + B^\perp X (B^\perp)^\top A + Q = 0. \quad (21)$$

Moreover, suppose that  $Q_f - B^\perp X (B^\perp)^\top$  is positive definite. Then, the solution to the DRE (6) can be decomposed as

$$P(t) = \tilde{P}_h(t) + P_p = \tilde{P}_h(t) + B^\perp X (B^\perp)^\top \quad (22)$$

with

$$\tilde{P}_h(t) = e^{A^\top(t_f-t)}\left((Q_f - P_p)^{-1} + \int_t^{t_f} S(\tau)d\tau\right)^{-1}e^{A(t_f-t)}. \quad (23)$$

△

*Proof:* Since  $P(t_f) = \tilde{P}_h(t_f) + P_p = Q_f$ , the boundary condition is satisfied. Evaluating the time derivative of  $P$  one

has that

$$\begin{aligned} \dot{P}(t) &= \dot{\tilde{P}}_h(t) = -A^\top \tilde{P}_h(t) - \tilde{P}_h(t)A + \tilde{P}_h(t)BR^{-1}B^\top \tilde{P}_h(t) \\ &= -A^\top P(t) - P(t)A + P(t)BR^{-1}B^\top P(t) \\ &\quad + A^\top P_p + P_p A - P_p BR^{-1}B^\top P(t) \\ &\quad - P(t)BR^{-1}B^\top P_p + P_p BR^{-1}B^\top P_p \\ &= -A^\top P(t) - P(t)A + P(t)BR^{-1}B^\top P(t) \\ &\quad + A^\top B^\perp X (B^\perp)^\top + B^\perp X (B^\perp)^\top A. \end{aligned}$$

Hence

$$\dot{P}(t) = -A^\top P(t) - P(t)A + P(t)BR^{-1}B^\top P(t) - Q$$

implying that the  $P$  given in (22) solves the DRE (6a). □

It is worth noting that, since (21) can be equivalently rewritten as

$$(B^\perp \otimes A^\top B^\perp + A^\top B^\perp \otimes B^\perp) \text{vec}(X) = -\text{vec}(Q)$$

there exists a matrix  $X$  with the property that (21) is satisfied if and only if

$$\text{vec}(Q) \in \text{Im}(B^\perp \otimes A^\top B^\perp + A^\top B^\perp \otimes B^\perp). \quad (24)$$

Therefore, if a given LQR problem is such that the condition (24) is met, then the closed-form solution to the Differential Riccati Equation (6) is provided by (22).

*Remark 1:* The results given by Propositions 1–3 remain valid even in the case in which, instead of the positive definiteness of the matrices  $Q_f$  and  $Q_f - B^\perp X (B^\perp)^\top$ , which assures the existence of a unique matrix  $P$  solving the DRE (19) and (6), respectively (see, e.g., [8]), only invertibility of the previous matrices is assumed. However, the invertibility of the matrices (20) and  $(Q_f - P_p)^{-1} + \int_t^{t_f} S(\tau)d\tau$  must be additionally assumed, provided that there exist unique solutions to the DRE (6) and (19). ▲

*Example 1:* As a simple illustration of the previous discussion, suppose that the weighting matrices appearing in (4) are given by

$$Q = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}, \quad Q_f = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

Note that  $Q_f$  is positive definite, whereas  $Q$  is only positive semi-definite. It is also assumed that  $R$  is the identity matrix in the cost functional (1), whereas the dynamic constraint (3) is represented by the matrices

$$A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

In this setting, by following the results discussed in this section, the matrices  $B^\perp$  and  $X$  are given by  $B^\perp = [1 \ 1]^\top$  and  $X = 1$ , respectively. Therefore,  $P(t)$  can be obtained by following the constructions given by Proposition 3, yielding

$$P(t) = -\frac{1}{a(t)} \begin{bmatrix} b(t) & c(t) \\ c(t) & d(t) \end{bmatrix}$$

where

$$\begin{aligned} a(t) &= 6e^8 - 10e^{2(2+t)} \\ b(t) &= 9e^8 + 7e^{4t} - 16e^{2(2+t)} \\ c(t) &= 3e^8 + 7e^{4t} - 14e^{2(2+t)} \\ d(t) &= 9e^8 + 7e^{4t} - 12e^{2(8+t)}. \end{aligned}$$

### III. ITERATIVE OPTIMAL CONTROL OF NONLINEAR SYSTEMS

To provide a concise statement of the main result in the nonlinear case, a few preliminary definitions are introduced. Let a process  $(\hat{u}(t), \hat{x}(t))$ , defined for all  $t \in [t_0, t_f]$ , be given such that  $\hat{x}(t_0) = x_0$  and the time derivative of  $\hat{x}$  satisfies (2) with  $u(\cdot) \equiv \hat{u}(\cdot)$ . Consider then the linearization of (2) along the pair  $(\hat{u}, \hat{x})$ , defined by

$$\begin{aligned} \delta \dot{x} &= \left( \frac{\partial f}{\partial x}(\hat{x}(t)) + \sum_{j=1}^m \frac{\partial g_j}{\partial x}(\hat{x}(t)) \hat{u}^j(t) \right) \delta x + g(\hat{x}(t)) \delta u \\ &=: A(t) \delta x + B(t) \delta u \end{aligned} \quad (26)$$

together with the initial condition  $\delta x(t_0) = 0$ , where  $\delta x = x - \hat{x}$ ,  $\delta u = u - \hat{u}$  describe the incremental variables and  $g_j$ ,  $j = 1, \dots, m$ , are the column vectors of the mapping  $g$ . Finally, the operator  $\beta : C^0(t_0, t_f) \rightarrow \mathbb{R}^{n \times m}$  is defined as

$$\beta(\hat{x}(t)) := \Phi(t_f, t) g(\hat{x}(t)) \quad (27)$$

where  $\Phi(t_f, t_0)$  describes the transition matrix associated with  $A(t)$ , and the *controllability Gramian*  $G(x_0) \in \mathbb{R}^{n \times n}$  associated with the system (26) is given by

$$\begin{aligned} G(x_0) &:= \int_{t_0}^{t_f} \Phi(t_f, t) g(\hat{x}(t)) R^{-1} g(\hat{x}(t))^\top \Phi(t_f, t)^\top dt \\ &= \int_{t_0}^{t_f} \beta(\hat{x}(t)) R^{-1} \beta(\hat{x}(t))^\top dt. \end{aligned} \quad (28)$$

Further, note that the operator  $\beta$  relies upon the knowledge of the state transition matrix of a linear time-varying system. The latter can be seldom obtained in closed form. To overcome this issue, recall (see, e.g., [38]) that for a given linear system  $\dot{x} = A(t)x$ ,  $x(t_0) = x_0$ ,  $\Phi(t_f, t)$  satisfies  $\frac{\partial}{\partial t} \Phi(t_f, t) = -\Phi(t_f, t)A(t)$ , together with the condition  $\Phi(t_f, t_f) = I$ . Thus, defining  $\Theta(t) := \Phi(t_f, t)^\top$ , for all  $t \in [t_0, t_f]$ , it follows that

$$\dot{\Theta} = -A(t)^\top \Theta, \quad \Theta(t_f) = I. \quad (29)$$

It is possible to backward integrate the linear differential equation (29), obtaining the required expression for  $\beta(\hat{x}(t))$ , i.e.,  $\beta(\hat{x}(t)) = \Theta(t)^\top g(\hat{x}(t))$ , for any time. Finally, consider the cost functional

$$J(u(\cdot)) = \|x(t_f)\|_{Q_f}^2 + \frac{1}{2} \int_{t_0}^{t_f} \|u(t)\|_R^2 dt \quad (30)$$

where  $Q_f \in \mathbb{R}^{n \times n}$ ,  $Q_f = Q_f^\top \geq 0$  and  $R \in \mathbb{R}^{m \times m}$ ,  $R = R^\top > 0$ . It is worth anticipating that the specially structured form of this cost functional is motivated by illustrative purposes of the underlying intuition and it is subsequently generalized, therefore

allowing to solve problems involving cost functionals of the form given in (1). By relying on the definitions (26)–(28), it is now possible to state the following theorem, which provides an iterative strategy leading to *candidate* optimal solutions for (30), (2).

*Theorem 1:* Consider the cost functional (30), subject to the constraints (2). Let  $u_0 \in C^0(t_0, t_f)$  be an arbitrary control law defined for all  $t \in [t_0, t_f]$  such that the solution  $x(\cdot)$  to (2) with  $u(\cdot) = u_0(\cdot)$ , exists in  $[t_0, t_f]$ . Consider the iterations described by

$$u_{i+1}(t) = (1 - \alpha_i)u_i(t) + \alpha_i \Gamma_i(t) \quad (31)$$

for  $i = 0, 1, \dots$ , with  $\Gamma_i : [t_0, t_f] \rightarrow \mathbb{R}^m$  given by the dynamic law

$$\begin{aligned} \dot{\Theta}_i &= -A_i(t)^\top \Theta_i \\ \Gamma_i(t) &= R^{-1} \beta_i(t)^\top Q_f \left( (I + G(x_0)Q_f)^{-1} \int_{t_0}^{t_f} \beta_i(t) u_i(t) dt \right. \\ &\quad \left. + (I + G(x_0)Q_f)^{-1} G(x_0) Q_f x_i(t_f) - x_i(t_f) \right) \end{aligned} \quad (32)$$

where  $\Theta_i(t_f) = I$ ,  $A_i$  is defined as in (26) with respect to the process  $(u_i, x_i)$ , and  $\beta_i : \mathbb{R} \rightarrow \mathbb{R}^{n \times m}$  is defined as

$$\beta_i(t) := \beta(x_i(t)) = \Theta_i^\top g(x_i(t)). \quad (33)$$

Then there exists a sequence  $\{\alpha_i\}$ ,  $\alpha_i \in \mathbb{R}_{>0}$ , such that the sequence  $\{u_i\}$  converges pointwise to an extremal of (30), (2). Furthermore, if the problem described by (30), (2) admits an optimal solution  $u^*$  that is the unique extremal, then

$$\lim_{i \rightarrow \infty} \|u_i(t) - u^*(t)\| = 0 \quad (34)$$

for all  $t \in [t_0, t_f]$ .  $\diamond$

*Proof:* Let  $\delta x_i = x_{i+1} - x_i$  and  $\delta u_i = u_{i+1} - u_i$ , and consider the linearization introduced in (26) along the process  $(u_i, x_i)$ . By borrowing ideas similar to those in the constructions introduced in [35] in a different context, consider the auxiliary variable  $\delta z_i$  defined as

$$\delta z_i(t) := \Phi_i(t_f, t) \delta x_i(t) \quad (35)$$

for  $t \in [t_0, t_f]$ . In the transformed coordinates, the problem, once linearized, becomes

$$\begin{aligned} J_L(\delta u_i(\cdot)) &= \frac{1}{2} \|x_i(t_f) + \delta z_i(t_f)\|_{Q_f}^2 \\ &\quad + \frac{1}{2} \int_{t_0}^{t_f} \|u_i(t) + \delta u_i(t)\|_R^2 dt \end{aligned} \quad (36)$$

since, by definition of state transition matrix,  $\delta z_i(t_f) = \delta x_i(t_f)$ , subject to the dynamic constraints

$$\begin{aligned} \delta \dot{z}_i &= \frac{\partial}{\partial t} \Phi_i(t_f, t) \delta x_i + \Phi_i(t_f, t) \delta \dot{x}_i \\ &= -\Phi_i(t_f, t) \left( \left( \frac{\partial f}{\partial x}(x_i(t)) + \sum_{j=1}^m \frac{\partial g_j}{\partial x}(x_i(t)) u_i^j(t) \right) \delta x_i - \delta \dot{x}_i \right) \\ &= \beta_i(t) \delta u_i \end{aligned} \quad (37)$$

together with the initial condition  $\delta z(t_0) = 0$ . The latter is a time-varying system such that the underlying vector field

is independent of the variable  $\delta z_i$ . The Hamiltonian function associated with (36), (37) is given by

$$H(\delta u_i, \delta z_i, \delta \lambda_i, t) = \frac{1}{2} \|u_i(t) + \delta u_i\|_R^2 + \delta \lambda_i^\top \beta_i(t) \delta u_i. \quad (38)$$

By relying on standard results (see, e.g., [6]), necessary conditions of optimality for the problem (36), (37) imply that the optimal process  $(\delta u_i^*, \delta z_i^*)$ , if it exists, satisfies the Hamiltonian dynamics

$$\dot{\delta z}_i = \frac{\partial H_i}{\partial(\delta \lambda_i)}(\delta u_i, \delta z_i, \delta \lambda_i, t) = \beta_i(t) \delta u_i \quad (39a)$$

$$\dot{\delta \lambda}_i = -\frac{\partial H_i^\top}{\partial(\delta z_i)}(\delta u_i, \delta z_i, \delta \lambda_i, t) = 0 \quad (39b)$$

together with the boundary conditions

$$\delta z_i(t_0) = 0 \quad (40a)$$

$$\delta \lambda_i(t_f) = Q_f(\delta z_i(t_f) + x_i(t_f)) \quad (40b)$$

whereas the control law  $\delta u_i$  in (39) is obtained, as in Section II, via the unconstrained optimization of  $H_i$ , yielding

$$\delta u_i(t) = -u_i(t) - R^{-1} \beta_i(t)^\top \delta \lambda_i(t). \quad (41)$$

Moreover, since

$$\frac{\partial^2 H_i}{\partial(\delta u_i)^2} = R > 0 \quad (42)$$

and because of linearity of the underlying dynamics, the conditions provided by (39), (42) are also sufficient conditions for optimality of the process  $(\delta u_i^*, \delta z_i^*)$ . It is worth observing that the dynamics in (39b) imply that the costate variable  $\delta \lambda_i$  is constant in the interval  $[t_0, t_f]$  and equal, by the boundary conditions in (40b), to  $\delta \lambda_i(t_f) = Q_f(\delta z_i(t_f) + x_i(t_f))$ . By replacing the costate into the control law in (41), and, in turn, the latter into (39a), one obtains

$$\delta z_i(t_f) = -\left( \int_{t_0}^{t_f} \beta_i(t) u_i(t) dt + G(x_0) Q_f(\delta z_i(t_f) + x_i(t_f)) \right) \quad (43)$$

which, once solved with respect to  $\delta z_i(t_f)$ , yields

$$\delta z_i(t_f) = -(I + G(x_0) Q_f)^{-1} \left( \int_{t_0}^{t_f} \beta_i(t) u_i(t) dt + G(x_0) Q_f x_i(t_f) \right).$$

The matrix  $I + G(x_0) Q_f$  is nonsingular, since the matrix  $G(x_0) Q_f$ , although not necessarily symmetric, possesses non-negative eigenvalues (see, e.g., [39]). Finally, by replacing the latter into (41), yields

$$\begin{aligned} \delta u_i^*(t) &= -u_i(t) - R^{-1} \beta_i(t)^\top Q_f(\delta z_i(t_f) + x_i(t_f)) \\ &=: -u_i(t) + \Gamma_i(t). \end{aligned} \quad (44)$$

This yields the update rule

$$\begin{aligned} u_{i+1}(t) &= u_i(t) + \alpha_i \delta u_i^*(t) \\ &= (1 - \alpha_i) u_i(t) + \alpha_i \Gamma_i(t) \end{aligned} \quad (45)$$

which describes precisely the iterations given by (31). The first claim follows by noting that the control law  $\delta u_i^*$  is the optimal solution of the linearized problem and it yields a descent direction (see, e.g., [40]) for the nonlinear cost  $J$ .

Finally, the second claim follows immediately from the stated assumptions and the gradient-like nature of the update rule (31).  $\square$

The update rule (45) reveals that moving along the descent direction  $\delta u_i^*$  is equivalent to considering the convex combination of the previous input  $u_i$  and the function  $\Gamma_i$  in (32). It is worth mentioning that, even if the proposed approach iteratively updates the control law in a gradient-like scheme, the properties related to the computational complexity of the proposed algorithm are different from those exhibited by standard gradient descent (see, e.g., [3]) and the conjugate gradient descent methods (see, e.g., [41]). More precisely, the integration of  $2n$  nonlinear differential equations is required for applying the gradient scheme, whereas  $4n$  nonlinear differential equations must be integrated to implement the conjugate gradient scheme. On the other hand, the approach proposed herein relies on the (backward-in-time) integration of a linear, time-varying, matrix equation, namely (29), hence revolving around the solution to  $n^2$  linear time-varying differential equations. As such, the comparison between the number of equations for (conjugate) gradient methods suggest a disadvantage of the proposed method with respect to the former. On the other hand the performance of the proposed approach has been compared, by means of a case study, with those provided by the gradient and conjugate gradient schemes. The numerical simulations have been carried out in the setting described by the cost functional (30), subject to the dynamic constraints given by the following nonlinear system, borrowed from [42, Ch.4], namely

$$\dot{x} = \begin{bmatrix} x_2 - x_3^2 \\ x_3 + 2x_1^2 x_3 \\ x_1^2 \\ x_1 + x_3^2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2x_3 \\ 1 \\ 0 \end{bmatrix} u, \quad x(t_0) = x_0.$$

The initial condition is  $x_0 = (1, -1, 2, -1)$ , whereas the matrix  $Q_f$  in (30) is the identity matrix. Finally, since in this example  $m = 1$ ,  $R$  has been selected as  $R = 1$ . The three methods are implemented by using a prescribed step size, for each iteration, along the descent direction. More precisely, for each scheme, the step size has been numerically selected as the maximum value that ensures monotonic convergence. In particular, in the case of the gradient as well as the conjugate gradient schemes, the maximum step size with the property above is  $\alpha = 5.25 \times 10^{-2}$ . On the other hand, the proposed approach ensures convergence with a unitary step size, i.e., for  $\alpha_i = 1$ . The behavior of the cost functional  $J$  along the iterations, for each of the simulated schemes, is reported in Fig. 1, which shows that, thanks to the possibility of selecting a unitary step size, convergence is achieved faster than the two other gradient schemes by the proposed algorithm.

*Remark 2:* The statement of Theorem 1 entails that the sequence  $\{u_i\}$ , produced by the iterations described by (31), converges in general to an extremal of (30), (2). It is worth

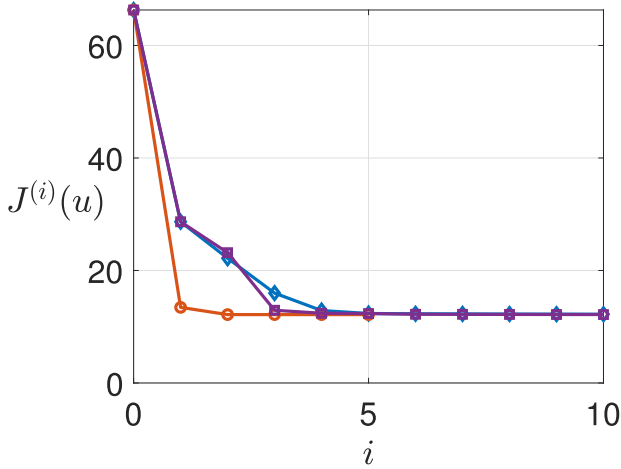


Fig. 1. First ten iterations of the cost functional obtained by using gradient descent scheme (diamond), the conjugate gradient descent scheme (squares) and the proposed approach (circles).

observing that this extremal cannot be a local maximizer of (30), (2). In fact, let  $\{J_i\}$  be the sequence of costs produced by the algorithm, where  $J_i$  is the cost associated with the  $u_i$  obtained at the  $i$ th iteration, and consider the limiting process  $(x^*, u^*)$ . The latter is not a (local) maximizer of (30), (2), since there exists a sequence of admissible processes (namely (31) in reverse order) along which  $J$  monotonically increases. This property is a natural consequence of the gradient descent nature of (31) and it is significantly different from other techniques based on the approximation of the Hamiltonian dynamics (see Section IV).  $\blacktriangle$

The objective of the two following corollaries consists in generalizing the structure of the cost functional (30). This is achieved by modifying first the terminal cost (Corollary 1) and then the running cost (Corollary 2).

*Corollary 1:* Consider the cost functional

$$J(u(\cdot)) = q_f(x(t_f)) + \frac{1}{2} \int_{t_0}^{t_f} \|u(t)\|_R^2 dt \quad (46)$$

subject to the constraints (2), and  $q_f \in C^1(\mathbb{R}^n)$ . Let  $u_0 \in C^0(t_0, t_f)$  be an arbitrary control law defined for all  $t \in [t_0, t_f]$  such that the solution  $x(\cdot)$  of (2) with  $u(\cdot) = u_0(\cdot)$ , exists in  $[t_0, t_f]$ . Let  $\delta z_i^o(t_f)$  be the solution of

$$0 = \delta z_i(t_f) + G(x_0) \frac{\partial q_f}{\partial(\delta z)}(\delta z_i(t_f) + x_i(t_f)) + \int_{t_0}^{t_f} \beta_i(t) u_i(t) dt. \quad (47)$$

Then, letting

$$\Gamma_i(t) = -R^{-1} \beta_i(t)^\top \delta z_i^o(t_f) \quad (48)$$

and considering the iterations (31), there exists a sequence  $\{\alpha_i\}$ ,  $\alpha_i \in \mathbb{R}_{>0}$ , such that the sequence  $\{u_i\}$  converges pointwise to an extremal of (46). Furthermore, if the problem described

by (46), (2) admits an optimal solution  $u^*$  which is the unique extremal, then (34) holds.  $\diamond$

*Proof:* By repeating the steps of the proof of Theorem 1 and noting that, since  $\delta \lambda_i(t) = \delta \lambda_i(t_f) = \frac{\partial q_f}{\partial(\delta z)}(\delta z_i(t_f) + x_i(t_f))$  for all  $t \in [t_0, t_f]$ , the terminal state (43) can be rewritten in the form given by (47), the claim follows by replacing in (48) the solution  $\delta z_i^o(t_f)$ .  $\square$

*Remark 3:* By possibly requiring that  $q_f \in C^2(\mathbb{R}^n)$  it is always possible to find the solution of (47), at least locally, by relying on the *implicit function theorem*. To this end note that, if the difference  $t_f - t_0$  is sufficiently small, the left-hand side of (47) can be interpreted as a function  $F: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $(\delta z_i(t_f), x_i(t_f)) \mapsto F(\delta z_i(t_f), x_i(t_f))$ . Thus, local solvability of (47) is implied by nonsingularity of the matrix

$$\frac{\partial F}{\partial(\delta z)}(0, x_i(t_f)) = I + G(x_0) \frac{\partial^2 q_f}{\partial(\delta z^2)}(x_i(t_f))$$

where  $\delta z_i(t_f)$  has been replaced with zero, since by construction (and its role of describing an increment) it can be reasonably assumed to be sufficiently small.  $\blacktriangle$

*Corollary 2:* Consider the cost functional

$$J(u(\cdot)) = \int_{t_0}^{t_f} \frac{1}{2} \|u(t)\|_R^2 + q(x(t)) dt \quad (49)$$

subject to the constraints (2), with  $q: \mathbb{R}^n \rightarrow \mathbb{R}_{>0}$ . Let  $u_0 \in C^0(t_0, t_f)$  be an arbitrary control law defined for all  $t \in [t_0, t_f]$  such that the solution  $x(\cdot)$  to (2) with  $u(\cdot) = u_0(\cdot)$ , exists in  $[t_0, t_f]$ . Let  $\beta_i: \mathbb{R} \rightarrow \mathbb{R}^{(n+1) \times m}$  be given by (33), where  $\beta$  is given by (27) considering  $F(x) = [f(x)^\top, q(x)]^\top$  and  $G(x) = [g(x)^\top, 0_{m \times 1}]^\top$ , in place of  $f$  and  $g$ , for all  $x \in \mathbb{R}^n$ , respectively. Then

$$\delta z_i(t_f) = - \int_{t_0}^{t_f} \beta_i(t) u_i(t) - \beta_i(t) R^{-1} \beta_i(t)^\top \begin{bmatrix} 0_{n \times 1} \\ 1 \end{bmatrix} dt. \quad (50)$$

Moreover, letting

$$\Gamma_i(t) = -R^{-1} \beta_i(t)^\top \begin{bmatrix} 0_{n \times 1} \\ 1 \end{bmatrix} \quad (51)$$

and considering the iterations described by (31), there exists a sequence  $\{\alpha_i\}$ ,  $\alpha_i \in \mathbb{R}_{>0}$ , such that the sequence  $\{u_i\}$  converges pointwise to an extremal of (49). Furthermore, if the problem described by (49), (2) admits an optimal solution  $u^*$  which is the unique extremal, then (34) holds.  $\diamond$

*Proof:* Note that, by introducing the auxiliary state-variable  $x_v$ , the problem given by (49), (2) can be rewritten as (see, e.g., [6])

$$J(u(\cdot)) = x_v(t_f) + \int_{t_0}^{t_f} \frac{1}{2} \|u(t)\|_R^2 dt \quad (52)$$

subject to

$$\begin{bmatrix} \dot{x} \\ \dot{x}_v \end{bmatrix} = \begin{bmatrix} f(x) \\ q(x) \end{bmatrix} + \begin{bmatrix} g(x) \\ 0_{1 \times m} \end{bmatrix} u, \quad \begin{bmatrix} x(t_0) \\ x_v(t_0) \end{bmatrix} = \begin{bmatrix} x_0 \\ 0 \end{bmatrix} \quad (53)$$

since

$$x_v(t_f) = \int_{t_0}^{t_f} q(x) dt. \quad (54)$$

Therefore, the claim follows by direct application of the previous Corollary 1 noting that, since in this case  $\delta\lambda_i(t) = \delta\lambda_i(t_f) = [0_{1 \times n}, 1]^\top$ , the solution of (47) is given by (50).  $\square$

*Remark 4:* Although the structure exhibited by the term  $\Gamma_i$  in (51) appears simpler once compared to those in the previous statements, it is worth observing that all the information needed to solve the finite-horizon optimal control problem, related in particular to the cost-to-go  $q$ , is essentially encoded, at each iteration, within the (extended)  $\beta_i$  introduced in Corollary 2. The value of the terminal state in (50) appears unused in the construction of the update rule (51) due to the structural absence of a genuine terminal cost in (49). Nonetheless, it would be employed in the descent direction whenever the cost functional (49) contains, in addition to a running cost on the state, a terminal cost similar to those of (30) or (46). More precisely, by combining the results of Corollaries 1 and 2, it is then possible to solve, iteratively, the optimal control problem described by the cost functional (1), subject to the nonlinear dynamic constraints (2).  $\blacktriangle$

*Remark 5:* By inspecting the structure of the update law (31), it appears evident that the main computational complexity in employing expression (32) for the construction of  $\Gamma_i$  is represented by the evaluation of the function  $\beta_i$ , which in turn relies upon the knowledge of a state transition matrix of a linear time-varying system, which can be determined by backward integration of (29). This allows concluding that the proposed scheme solves a sequence of time-varying LQR problems but, instead of the solution of a time-varying Differential Riccati Equation, it requires a solution of a time-varying linear differential equation.  $\blacktriangle$

#### IV. SELECTION OF THE STEP ALONG THE DESCENT DIRECTION

The structure of the update rule (31) entails that the knowledge of the descent direction  $\delta u_i$  in (44) is exploited at each step via a convex combination of the function  $\Gamma_i$  with the control law employed at the previous iteration. Once the descent direction is computed, the implementation of the rule (31) has the property that the cost functional  $J(u_{i+1})$  becomes essentially a function of the step  $\alpha_i$  alone. While, on one hand, the formal statements above ensure the existence of a sufficiently small  $\alpha_i$  such that  $J(u_{i+1}) < J(u_i)$ , the practical selection of such a value remains a crucial design choice with significant consequences on the convergence rate of the updates. A few alternative strategies are discussed and compared in this section. It is worth pointing out that all the strategies discussed in the following section aim to construct the sequence  $\{\alpha_i\}$ , the existence of which is ensured by Theorem 1 and therefore, in practice, some preliminary parameter tuning may be necessary to actually achieve convergence of the iterations.

#### A. Approximate Line Search

The most intuitive approach would be to study the variation of  $J(u_{i+1})$  as a function of  $\alpha_i$ , denoted  $\tilde{J}(\alpha_i)$ , and select  $\alpha_i^* = \arg \min_{\alpha_i \in (0,1)} \tilde{J}(\alpha_i)$ . However, since the analytic expression of  $\tilde{J}$  cannot be obtained in general, the previous strategy may be pursued only in an approximate fashion. This can be achieved in, at least, two different ways, i.e., either approximating (discretizing) the admissible values for  $\alpha_i$  or by approximating directly the expression of  $\tilde{J}(\alpha_i)$ . While the latter strategy is discussed in detail in the following section, the former requires the definition of a set  $\mathcal{T} = \{\alpha_0, \alpha_1, \dots, \alpha_N\} \subset (0, 1)$ , of finite cardinality  $|\mathcal{T}| = N$ , and subsequently obtaining

$$\alpha_i \in \arg \min_{\alpha_i \in \mathcal{T}} \tilde{J}(\alpha_i).$$

Since  $\tilde{J}(\alpha_i)$  should be computed numerically at the values in  $\mathcal{T}$ , the approach is in general particularly time-consuming for large values of  $N$ .

#### B. Optimal Step for Second-Order Expansion

Provided that the descent direction  $\delta u_i(t)$  is known, one can interpret  $x_{i+1}(\cdot)$  as a function of the step  $\alpha_i$  in addition to its canonical dependence on time. Thus, by considering the first-order Taylor series expansion of  $x_{i+1}(t, \alpha_i)$  with respect to  $\alpha_i$  about a nominal value  $\alpha_0$ , it follows that

$$x_{i+1}(t, \alpha_i) = x_{i+1}(t, \alpha_0) + \left. \frac{\partial}{\partial \alpha_i} x_{i+1}(t, \alpha_i) \right|_{\alpha_i = \alpha_0} (\alpha_i - \alpha_0) + r(t, \alpha_i) \quad (55)$$

for all  $t \in [t_0, t_f]$ , where  $r(t, \alpha_i)$  contains higher-order terms of the expansion which decay to zero faster than  $(\alpha_i - \alpha_0)^2$ , whenever  $\alpha_i$  tends to  $\alpha_0$ . Considering the formal claims of Section III concerning a sufficiently small step, a reasonable choice of the nominal value  $\alpha_0$  consists in selecting  $\alpha_0 = 0$ , corresponding to the case in which  $u_{i+1} \equiv u_i$ , implying as a consequence that  $x_{i+1}(t, \alpha_0) = x_{i+1}(t, 0) = x_i(t)$ , for all  $t \in [t_0, t_f]$ . Then, replacing (55), with  $\alpha_0 = 0$ , in the first term of the cost functional (30), one obtains

$$\begin{aligned} \tilde{J}(\alpha_i) &= \frac{1}{2} \|x_i(t_f)\|_{Q_f}^2 \\ &+ \left( x_{i+1}(t_f, \alpha_i)^\top Q_f \frac{\partial}{\partial \alpha_i} x_{i+1}(t, \alpha_i) \right) \Big|_{\alpha_i=0} \alpha_i \\ &+ \frac{1}{2} \left\| \frac{\partial}{\partial \alpha_i} x_{i+1}(t, \alpha_i) \right\|_{\alpha_i=0}^2 \alpha_i^2 \\ &+ \frac{1}{2} \int_{t_0}^{t_f} \|u_i(t) + \alpha_i \delta u_i(t)\|_R^2 dt + \tilde{r}(\alpha_i) \\ &= \tilde{J}_q(\alpha_i) + \tilde{r}(\alpha_i) \end{aligned} \quad (56)$$

where  $\tilde{J}_q(\alpha_i)$  represents the second-order approximation of  $\tilde{J}(\alpha_i)$  about the nominal step value  $\alpha_0 = 0$ . The dependence of  $\tilde{J}_q$  only on  $\alpha_i$  highlights again the property that, once the direction  $\delta u_i(t)$  is given, the optimal control problem reduces to the problem of finding the step  $\alpha_i^*$  that, at least locally, yields



the minimum value of  $\tilde{J}(\alpha_i)$ . The latter is a finite-dimensional optimization task with respect to a scalar variable.

*Proposition 4:* The optimal step that minimizes  $\tilde{J}_q$ , subject to the dynamic constraints

$$\begin{aligned} \dot{x}_{i+1} &= f(x_{i+1}) + g(x_{i+1})u_i + \alpha_i g(x_{i+1})\delta u_i \\ x_{i+1}(t_0) &= x_0 \end{aligned} \quad (57)$$

is given by

$$\alpha_i^* = -\frac{x_i(t_f)^\top Q_f S_{i+1}(t_f) + \int_{t_0}^{t_f} u_i(t) R \delta u_i(t) dt}{\|S_{i+1}(t_f)\|_{Q_f}^2 + \int_{t_0}^{t_f} \|\delta u_i(t)\|_R^2 dt} \quad (58)$$

in which  $S_{i+1}(t) \in \mathbb{R}^n$  is the solution of

$$\begin{aligned} \dot{S}_{i+1}(t) &= \left( \frac{\partial f}{\partial x}(x_i(t)) + \sum_{j=1}^m \frac{\partial g_j}{\partial x}(x_i(t)) u_i^j(t) \right) S_{i+1}(t) \\ &\quad + g(x_i(t)) \delta u_i(t) \end{aligned} \quad (59)$$

from the initial condition  $S_{i+1}(t_0) = 0$ .  $\diamond$

*Proof:* Since finding the optimal step consists in solving a scalar, unconstrained optimization problem, with respect to  $\alpha_i$ , the solution can be found by equating the total derivative of  $\tilde{J}_q(\alpha_i)$ , given by (56), with respect to  $\alpha_i$  to zero, namely

$$\begin{aligned} \frac{d\tilde{J}_q}{d\alpha_i}(\alpha_i) &= \left( x_{i+1}(t_f, \alpha_i)^\top Q_f \frac{\partial}{\partial \alpha_i} x_{i+1}(t, \alpha_i) \right) \Big|_{\alpha_i=0} \\ &\quad + \left\| \frac{\partial}{\partial \alpha_i} x_{i+1}(t, \alpha_i) \right\|_{\alpha_i=0}^2 \alpha_i \\ &\quad + \int_{t_0}^{t_f} (u_i(t) + \alpha_i \delta u_i(t))^\top R \delta u_i(t) dt = 0. \end{aligned} \quad (60)$$

The proof is then concluded by solving the previous equation with respect to  $\alpha_i$ , provided that one is able to evaluate  $\frac{\partial}{\partial \alpha_i} x_{i+1}(t, \alpha_i)|_{\alpha_i=0}$ . This can be accomplished by relying on the theory of *sensitivity functions* (see, e.g., [43]), obtaining the *sensitivity equation* related to the system (57), which is given by (59), namely  $S_{i+1}(t) := \frac{\partial}{\partial \alpha_i} x_{i+1}(t, \alpha_i)|_{\alpha_i=0}$ .  $\square$

*Remark 6:* Whenever the cost functional is in the form given in (46), the first-order Taylor expansion around  $\alpha_0 = 0$  of  $q_f$  yields

$$\begin{aligned} \tilde{J}(\alpha_i) &= q_f(x_i(t_f)) + \alpha_i \frac{\partial q_f(x_{i+1}(t_f, \alpha_i))}{\partial x_{i+1}} S_{i+1}(t_f) \\ &\quad + \frac{1}{2} \int_{t_0}^{t_f} \|u_i(t) + \alpha_i \delta u_i(t)\|_R^2 dt + \bar{r}(\alpha_i) \\ &= \tilde{J}_q(\alpha_i) + \bar{r}(\alpha_i). \end{aligned}$$

Therefore, (60) becomes

$$\begin{aligned} \frac{d\tilde{J}_q}{d\alpha_i}(\alpha_i) &= \frac{\partial q_f(x_{i+1}(t_f, \alpha_i))}{\partial x_{i+1}} S_{i+1}(t_f) \\ &\quad + \int_{t_0}^{t_f} (u_i(t) + \alpha_i \delta u_i(t))^\top R \delta u_i(t) dt = 0. \end{aligned} \quad (61)$$

Provided that one is able to solve (61), then  $\alpha_i^*$  is obtained. As a specially structured case, consider for instance the setting in

which  $q_f(x_i(t_f)) = x_i(t_f)$ , assuming that  $x_i(t_f)$  is a scalar (as in the case of  $x_v(t_f)$  introduced in Corollary 2). Then, the step size  $\alpha_i^*$  that leads to the maximum reduction of  $J_q$  is given by

$$\alpha_i^* = -\frac{S_{i+1}(t_f) + \int_{t_0}^{t_f} u_i(t) R \delta u_i(t) dt}{\int_{t_0}^{t_f} \|\delta u_i(t)\|_R^2 dt}. \quad (62)$$

▲

It is worth observing that the construction of Proposition 4 revolves around the solution of a system of  $n$  linear differential equations (i.e., the sensitivity equations (59)), in order to provide, at least locally, the maximum descent of the cost functional, along the direction  $\delta u_i$ . It is possible, however, to circumvent the need for such a set of equations, for instance, by borrowing the heuristics used in the field of *machine learning* to improve the speed of convergence of the gradient descent algorithm, used in the training process of neural networks, as discussed in the following section.

### C. Heuristic-Based Evaluation of the Step Size

Even if the descent direction of the proposed algorithm (being infinite dimensional) differs, in general, from the one provided by the gradient descent, the idea is that the chosen heuristics are still viable, since the two schemes are essentially based on descent directions. From the vast amount of heuristics available in the literature (see, e.g., [36]), the RMSProp scheme yields an update rule given by  $u_{i+1} = u_i + \Delta u_i$ , where  $\Delta u_i$  is given by

$$\Delta u_i(t) = \theta_i(t) \odot \delta u_i(t) \quad (63)$$

where the symbol  $\odot$  stands for the elementwise product, or *Hadamard product*, between the two vectors  $\theta_i(t)$  and  $\delta u_i(t)$ . The former is given by

$$\theta_i(t) = -\alpha(\delta[1_{m \times 1}] + r_i(t))^{-\frac{1}{2}} \quad (64)$$

where the operation  $((\cdot)^{-\frac{1}{2}})$  is again performed elementwise on the  $m$ -vector  $\delta[1_{m \times 1}] + r_i(t)$ ,  $\alpha \in (0, 1)$  is a fixed step, and  $r_i(t)$  is given by

$$r_i(t) = \rho r_{i-1}(t) + (1 - \rho) \delta u_i(t) \odot \delta u_i(t). \quad (65)$$

The coefficients  $\delta \in \mathbb{R}_{>0}$  and  $\rho \in \mathbb{R}_{>0}$  are left as design parameters, and  $r_0(t) = [0_{m \times 1}]$ . Note that this scheme allows choosing the initial step  $\alpha$ , which is then adjusted along the iterations according to the construction discussed above.

### D. Strategy Comparison on a Continuous Stirred-Tank Chemical Reactor

To compare the recipes introduced in this section, consider the model of a continuous stirred-tank chemical reactor, borrowed from [3, ch. 6], and described by the equations

$$\begin{aligned} \dot{x}_1 &= -2(x_1 + 0.25) + (x_2 + 0.5)e^{\frac{25x_1}{x_1+2}} - (x_1 + 0.25)u \\ \dot{x}_2 &= 0.5 - x_2 - (x_2 + 0.5)e^{\frac{25x_1}{x_1+2}} \end{aligned} \quad (66)$$

where  $x_1$  represents the deviation of the steady-state temperature and  $x_2$  describes the deviation of the steady-state concentration

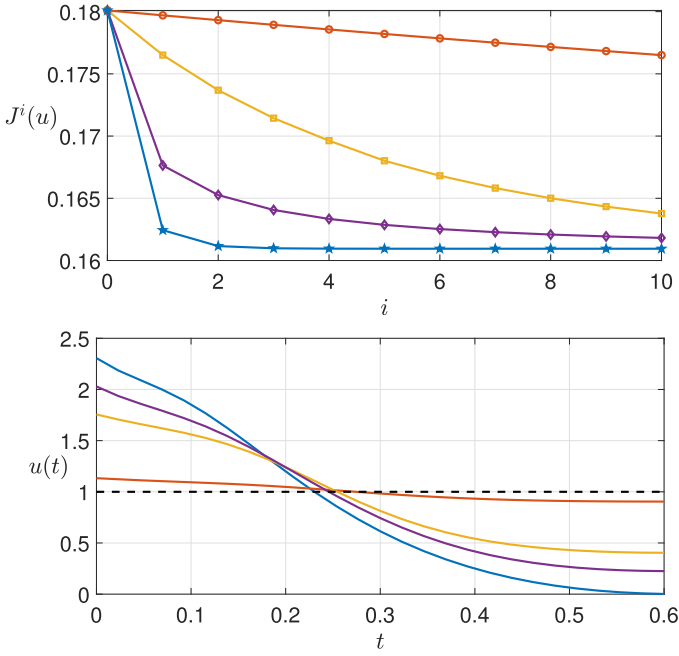


Fig. 2. Top graph: First ten iterations of the cost functional obtained by using a fixed step scheme (circle), the RMSProp scheme (square), the approximation of the optimal step (diamond), the grid search (star). Bottom graph: time histories of the control laws, evaluated on the last (10th) iteration, together with the initial guess  $u_0(t) = 1$  (dashed) for all  $t \in [t_0, t_f]$ .

of the reactor. The cost functional subject to the previous dynamical constraints is

$$J(u(\cdot)) = \int_{t_0}^{t_f} x_1^2(t) + x_2^2(t) + Ru^2(t)dt \quad (67)$$

where  $t_0 = 0$ ,  $t_f = 0.6$ , and  $R = 0.1$ . The structure of the cost functional suggests that the objective is to maintain both the temperature as well as the concentration close to prescribed steady-state values, while minimizing, at the same time, the control effort. The following numerical simulations are carried out with  $x(0) = [0.5 \ -0.5]^\top$ , and the initial guess  $u_0$  on the optimal control, required to initialize the algorithm, is chosen, similarly to [3], as  $u_0(t) = 1$  for all  $t \in [t_0, t_f]$ .

According to Corollary 2, the use of the auxiliary dynamic constraint

$$\dot{x}_v = x_1^2 + x_2^2, \quad x_v(t_0) = 0$$

allows rewriting the cost functional (67) in the form required in (52), which in turn leads to the iterations described by (31), with  $\Gamma_i$  provided by (51). Fig. 2 shows the behavior of the schemes introduced in this section for the selection of the step size  $\alpha$ , namely the grid-search, the RMSProp, the one based on the second-order approximation of  $\tilde{J}(\alpha_i)$ , with  $\alpha_i^*$  provided by (62), and finally the scheme characterized by the choice of a fixed step. Both the RMSProp and the fixed step schemes are initialized with  $\alpha = 0.01$ , whereas the grid search is performed discretizing the interval  $(0, 1)$  in samples uniformly spaced via an offset of  $10^{-4}$ , namely by defining  $\mathcal{T} = \{0.0001, 0.0002, \dots\}$ , with  $|\mathcal{T}| = 9999$ . In particular, the top graph of Fig. 2 depicts the

behavior of the cost functional after the first ten updates, with the lower values achieved by the grid search (starred line). It is worth noting, however, that each iteration of the grid search scheme is computationally much more expensive (time-consuming) with respect to the alternative schemes, requiring approximately 750 s for the simulation to terminate, whereas the other three proposed schemes required less than 5 s to terminate. Therefore, the scheme based on the second-order approximation leads to the best tradeoff between speed of convergence of  $J_i$  to  $J(u^*)$  and time required for the execution of the algorithm, with respect to the alternative strategies. The plot reported in the bottom graph of Fig. 2 provides a comparison between the resulting control laws after ten iterations and confirms the previous discussion. In fact the control law resulting from the second-order approximation of the cost functional is closer than the others to the one obtained by the grid-search strategy.

## V. PERSPECTIVES ON QUASILINEARIZATION

Intuitively, the proposed scheme hinges upon two main *ingredients*, namely *i*) linearization and *ii*) the construction (and solution) of Hamiltonian dynamics. These are also the two main steps on which the strategy of quasilinearization (see [3, ch. 6]) is based, although performed in the opposite order: in fact linearization is carried out after the construction of the nonlinear Hamiltonian dynamics. The following statement shows that these two steps do not commute in general. To avoid burden of notation in the following result, consider the problem (30), (2), in which  $Q_f = I$  for simplicity.

*Fact 1:* Consider the optimal control problem given by (30), (2) with  $Q_f = I$ . The structures of the dynamics obtained by linearizing the nonlinear Hamiltonian dynamics and by constructing the Hamiltonian dynamics for the linearized system differ.  $\diamond$

To show this, consider the Hamiltonian associated with (2), (30) given by

$$H(x, \lambda, u) = \frac{1}{2} \|u\|_R^2 + \lambda^\top (f(x) + g(x)u) \quad (68)$$

and consider the following dynamics, obtained by applying the standard necessary conditions of optimality and by substituting in the resulting Hamiltonian dynamics the candidate optimal control,  $u = -R^{-1}g(x)^\top \lambda$ , namely

$$\dot{x} = f(x) - g(x)R^{-1}g(x)^\top \lambda \quad (69a)$$

$$\dot{\lambda} = -\frac{\partial f^\top}{\partial x} \lambda + \left( \frac{\partial}{\partial x} (g(x)^\top \lambda) \right)^\top R^{-1}g(x)^\top \lambda \quad (69b)$$

together with  $x(t_0) = x_0$  and  $\lambda(t_f) = x(t_f)$ , respectively. Considering then  $\delta x = x_{i+1} - x_i$ ,  $\delta \lambda = \lambda_{i+1} - \lambda_i$ , and linearizing around the pair  $(x_i(t), \lambda_i(t))$ , one obtains

$$\dot{\delta x} = A_{11}(t)\delta x + A_{12}(t)\delta \lambda, \quad \delta x(t_0) = 0$$

$$\dot{\delta \lambda} = A_{21}(t)\delta x + A_{22}(t)\delta \lambda, \quad \delta \lambda_f = \delta x_f + x_i^f - \lambda_i^f \quad (70)$$

where  $\delta\lambda_f = \delta\lambda(t_f)$ ,  $\delta x_f = \delta x(t_f)$ ,  $x_i^f = x_i(t_f)$ , and  $\lambda_i^f = \lambda_i(t_f)$ , and

$$A_{11}(t) = \frac{\partial f}{\partial x}(x_i(t)) - g(x_i(t))R^{-1} \frac{\partial}{\partial x}(g(x_i(t))^\top \lambda_i(t)) - \frac{\partial}{\partial x}(g(x_i(t))R^{-1}g(\xi(t))^\top \lambda_i(t)) \Big|_{\xi(t)=x_i(t)} \quad (71a)$$

$$A_{12}(t) = -g(x_i(t))R^{-1}g(x_i(t))^\top \quad (71b)$$

$$A_{21}(t) = -\sum_{j=1}^n \frac{\partial^2 f_j(x_i(t))}{\partial x^2} \lambda_i^j(t) + \frac{\partial}{\partial x}(\lambda_i(t)^\top g(x_i(t)))R^{-1} \frac{\partial}{\partial x}(g(x_i(t))^\top \lambda_i(t)) + \frac{\partial}{\partial x} \left( \frac{\partial}{\partial x}(\lambda_i(t)^\top g(x_i(t)))R^{-1}(g(\xi(t))^\top \lambda_i(t)) \right) \Big|_{\xi(t)=x_i(t)} \quad (71c)$$

$$A_{22}(t) = -\frac{\partial f}{\partial x}(x_i(t))^\top + \frac{\partial}{\partial x}(\lambda_i(t)^\top g(x_i(t)))R^{-1}g(x_i(t))^\top + \frac{\partial}{\partial \lambda} \left( \frac{\partial}{\partial x}(\lambda_i(t)^\top g(x_i(t)))R^{-1}(g(x_i(t))^\top \zeta(t)) \right) \Big|_{\zeta(t)=\lambda_i(t)} \quad (71d)$$

Consider now the linearization of (2), introduced in Section II and given by (26), along the process  $(u_i, x_i)$ , together with the cost given by (36), in which  $\delta z_i$  is replaced by  $\delta x_i$ , and consider the Hamiltonian function associated with this problem, given by

$$H_\delta(\delta x, \delta \lambda, \delta u, t) = \frac{1}{2} \|\delta u_i(t) + \delta u\|_R^2 + \delta \lambda^\top (A(t)\delta x + B(t)\delta u). \quad (72)$$

By performing the same steps carried out at the beginning of the discussion and by considering the candidate optimal control  $\delta u = -(u_i(t) + R^{-1}B(t)^\top \delta \lambda)$ , one obtains the dynamics

$$\dot{\delta x} = A(t)\delta x - B(t)R^{-1}B(t)^\top \delta \lambda - B(t)u_i(t) \quad (73a)$$

$$\dot{\delta \lambda} = -A(t)^\top \delta \lambda \quad (73b)$$

together with  $\delta x(t_0) = 0$ ,  $\delta \lambda_f = \delta x_f + x_i^f$  and the claim follows by noting that the two dynamics given by (70) and (73a) differ.

To visualize the implications of the previous proposition, consider the cost functional

$$J(u(\cdot)) = \frac{1}{2}x(t_f)^2 + \frac{1}{2} \int_{t_0}^{t_f} u(t)^2 dt$$

subject to the dynamical constraint

$$\dot{x} = -x^2 + xu$$

with initial condition  $x_0 = 2.5$ . By applying, for example, the *Pontryagin's minimum principle*, one obtains the candidate optimal control law  $u(t) = -\lambda(t)x(t)$ , which is then used to initialize the proposed scheme by substituting, respectively,  $\lambda(t)$  and  $x(t)$  with  $x_0(t)$  and  $\lambda_0(t)$ , the latter one obtained by forward integrating the costate equation starting from an arbitrary initial condition  $\lambda(t_0)$ . Note that the pair  $(x_0(t), \lambda_0(t))$  is used also for the initialization of the quasilinearization scheme. Fig. 3

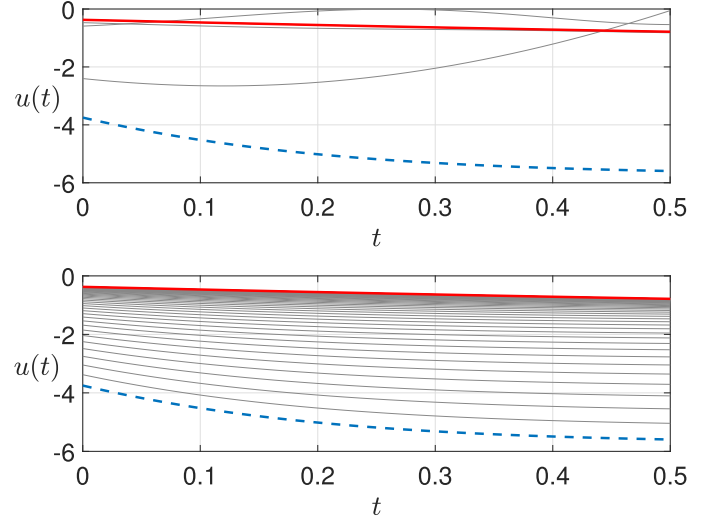


Fig. 3. Successive control update of quasilinearization (top graph) and constant costate (bottom graph) iterations. The dashed blue line indicates the initial guess  $u_0(t) = -\lambda_0(t)x_0(t)$ , whereas the solid red line indicates the candidate optimal control law.

shows how the control is updated by following, respectively, the quasilinearization and the scheme proposed herein, confirming that, although the starting and terminal control laws coincide in both cases, the paths taken by the two schemes, for the update of the optimal control law, differ.

## VI. OPTIMAL CONTROL OF MECHANICAL SYSTEMS ON REDUNDANT TASKS

The effectiveness of the strategies envisioned above is validated by means of an application involving a mechanical system that must accomplish a prescribed redundant task. Toward this end, consider a mechanical system described by the canonical equations

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + \frac{\partial P(q)}{\partial q}^\top = u, \quad q(t_0) = q_0, \quad \dot{q}(t_0) = \dot{q}_0 \quad (74)$$

where  $q(t) \in \mathbb{R}^n$  is the vector of the generalized coordinates,  $\dot{q}(t) \in \mathbb{R}^n$  and  $\ddot{q}(t) \in \mathbb{R}^n$  are the vectors of generalized velocities and accelerations, respectively,  $M(q) \in \mathbb{R}^{n \times n}$  is the inertia matrix, which is symmetric and positive definite. The term  $C(q, \dot{q})\dot{q}$  accounts for the centrifugal and Coriolis forces, whereas the term  $(\partial P(q)/\partial q)^\top$  accounts for the gravity forces, with  $P(q)$  describing the potential energy due to gravity. Finally,  $u$  is the vector of the external joint generalized forces (i.e., the torques). This section is devoted to illustrate the application of the results presented in the previous sections to the case in which the mechanical system (74) is required to perform a desired task with respect to which it is redundant. In particular, *task redundancy* is characterized, for a certain mechanical system, by the presence of a number of controlled degrees of freedom greater than the number of variables needed to describe a certain given task (see, e.g., [44]). A popular solution to this control problem, as briefly recalled below, consists in rewriting the dynamics of the mechanical system with respect to the task; decouple the task itself from the remaining (zero) dynamics and finally by stabilizing the task alone via linear design methods.

However, despite its simplicity, two relevant shortcomings of this strategy can be immediately identified. First, the hidden dynamics may be unstable for certain tasks. Second, somewhat related to the former, the above approach does not permit to fully exploit the intrinsic degrees of redundancy. On the other hand, a direct solution of this problem within the framework of optimal control is in general a discouraging task. In the following we envision a combination of the two approaches above, which permits to retain the advantages of both, in terms of simplicity and robustness. Since a task can be efficiently modeled as a function  $h : \mathbb{R}^n \rightarrow \mathbb{R}^r$ ,  $q \mapsto h(q)$ , mapping the generalized coordinates space into the *operational space*, the objective is to steer  $h(q)$  to a desired function  $h_d(t)$ . Thus, redundancy arises whenever  $r < n$ . Introduce the auxiliary variable  $z = h(q) - h_d$ , for all  $t \geq 0$ . Define  $x_1 = z$ ,  $x_2 = \dot{z}$  and substitute the preliminary control law  $u = (\partial P(q)/\partial q)^\top + C(q, \dot{q})\dot{q} + M(q)\tau$  in (74), obtaining

$$\begin{aligned} \ddot{q} &= \tau \\ \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \nabla h(q)\dot{q} + \dot{q}^\top \nabla^2 h(q)\dot{q} - \ddot{h}_d(t) \end{aligned} \quad (75)$$

where  $\nabla h(q)$  is assumed to be full row rank. Then, replace the first into the last equation and let

$$\tau = -(\nabla h(q))^\sharp (\dot{q}^\top \nabla^2 h(q)\dot{q} - \ddot{h}_d(t) + K_p x_1 + K_d x_2) \quad (76)$$

in which the operator  $(\cdot)^\sharp$  indicates *Moore–Penrose pseudoinversion*, and  $K_p$  and  $K_d$  are such that the dynamic matrix of the linear system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ K_p x_1 + K_d x_2 \end{bmatrix} = \begin{bmatrix} 0 & I \\ K_p & K_d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

possesses all eigenvalues with negative real parts. Thus, one has that the control law given by

$$\begin{aligned} u &= C(q, \dot{q})\dot{q} - M(q)(\nabla h(q))^\sharp (\dot{q}^\top \nabla^2 h(q)\dot{q} - \ddot{h}_d(t)) \\ &\quad - M(q)(\nabla h(q))^\sharp (K_p(h(q) - h_d(t)) + K_d(\dot{h}(q) - \dot{h}_d(t))) \\ &\quad + \frac{\partial P(q)}{\partial q}^\top \end{aligned} \quad (77)$$

steers  $z$  and its successive time-derivatives to zero asymptotically. Consider now instead the problem of applying the strategy of Section III to (74). The cost functional subject to the dynamic constraints (74) is given by

$$\begin{aligned} J(u(\cdot)) &= \frac{1}{2} \int_{t_0}^{t_f} \|h(q(t)) - h_d(t)\|_{W_1}^2 dt \\ &\quad + \frac{1}{2} \int_{t_0}^{t_f} \|q(t)\|_{W_2}^2 + \|\dot{q}(t)\|_{W_3}^2 + \|u(t)\|_R^2 dt \end{aligned} \quad (78)$$

where the matrices  $W_1$ ,  $W_2$ ,  $W_3$ , and  $R$  are of appropriate dimensions. This particular choice of the weights indicates that it is desired that  $h(q(t))$  is as close as possible to  $h_d(t)$  for any time, together with the smallest possible angular velocity, thus settling for a compromise between the task accomplishment and

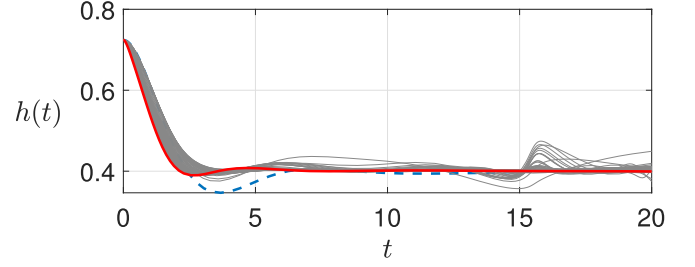


Fig. 4. Update of the function  $h$ , which represents the task.

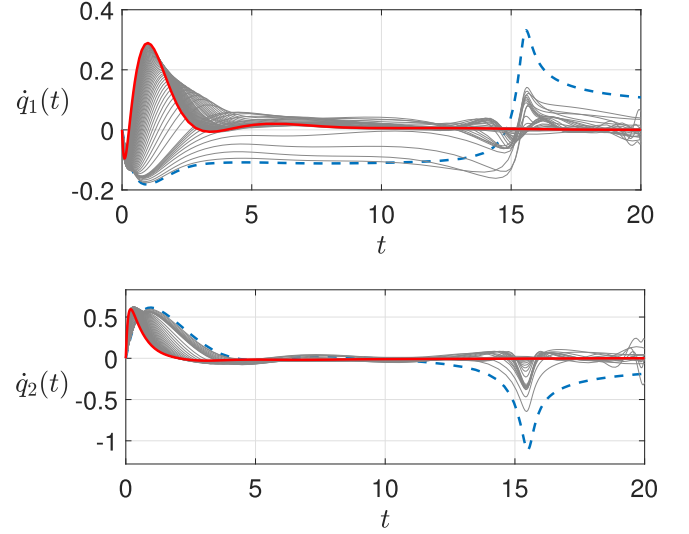


Fig. 5. Updates of the time histories of  $\dot{q}_1$  and  $\dot{q}_2$ . The dashed trajectories are obtained from the application of the initial guess  $u_0$ , whereas the solid lines are obtained by applying the last updated control law  $u^*$ .

the requirement of stability. Since the structure of the problem, given by the cost functional (78) and the constraints (74), belongs to the class of problems addressed by Corollary 2, the update on the control law is given by  $u_{i+1} = (1 - \alpha_i)u_i + \alpha_i\Gamma_i$ , with  $\Gamma_i$  provided by (51). Therefore, the control law (77) can be used for the initialization of the algorithmic scheme proposed in the previous sections. Simulations are carried out in the case of a 2-DOF manipulator evolving on a horizontal plane, for which the task is defined as

$$h(q) := x_e(q) + y_e(q)$$

where  $(x_e, y_e)$  are the coordinates of the origin of the end-effector reference frame. The desired task is then induced by selecting  $h_d = 0.4$ , with  $W_1 = 10^3$ ,  $W_2 = I_{2 \times 2}$ ,  $W_3 = 10^2 \cdot I_{2 \times 2}$ ,  $R = 10^{-3} \cdot I_{2 \times 2}$ ,  $t_0 = 0$  s, and  $t_f = 20$  s. The 2-DOF planar manipulator links lengths are given by  $l_1 = l_2 = 0.3$  m, both having mass  $m_1 = m_2 = 1$  kg. In the considered scenario, the gravitational effects may be neglected, since the robot essentially lays on the horizontal plane. The initial conditions of the manipulator have been chosen as  $q_0 = (\pi/4, \pi/4)$  and  $\dot{q}_0 = (0, 0)$ .

As it can be appreciated from Figs. 4 and 5, the proposed algorithm is such that the robot reaches the desired objective with nearly zero angular velocity at time  $t_f = 20$  s. As a beneficial

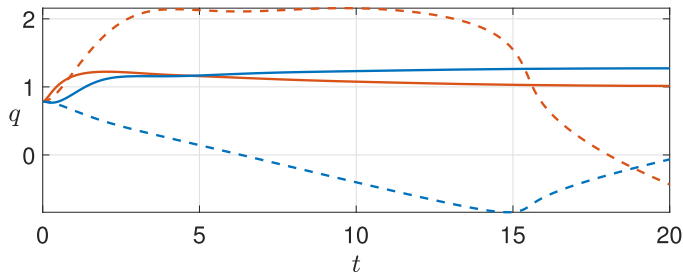


Fig. 6. Time histories of  $q_1$  (red) and  $q_2$  (blue) resulting from the application of the guess  $u_0$  given by (77) (dashed), and from the application of the last update of the control law  $u^*$  (solid).

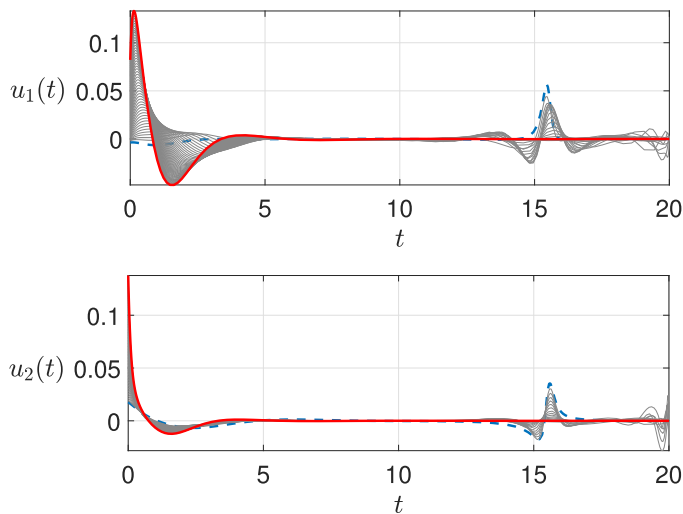


Fig. 7. Time histories of the components of the control law obtained from the proposed algorithm.

consequence, the optimal solution that exploits the redundancy to a full extent has the effect of reducing the undershoot that arises from the application of the control law  $u_0$  given by (77), see Fig. 4. The angular positions at the end of the procedure can be inferred from Fig. 6, which also depicts the same quantities at the beginning of the algorithm, obtained from the application of the control law (77). As graphically illustrated by the dashed lines of Fig. 6, the control law (77) is such that, while accomplishing the prescribed task, the position of the end-effector continues to slide along the task (i.e., the line described by the identity  $x_e(q) + y_e(q) = h_d$ ), inverting the direction of motion each time it reaches the boundary of the feasible workspace.

The updates of the control law are reported in Fig. 7, whereas Fig. 8 shows the behavior of the auxiliary state-variable  $x_v$ , which satisfies, according to Corollary 2, the equation

$$\dot{x}_v = \frac{1}{2} \left( \|h(q(t)) - h_d\|_{W_1}^2 + \|q(t)\|_{W_2}^2 + \|\dot{q}(t)\|_{W_3}^2 \right)$$

together with the initial condition  $x_{v0} = 0$ . It is worth observing that this variable, due to the choice of the weights on the quantities that constitute the cost functional (78), represents the dominant part of the cost functional, since it is of six orders of magnitude greater than the term involving the control variable. Therefore, Fig. 8 depicts, at each iteration of the algorithm, the behavior of the integral of the running cost with respect to

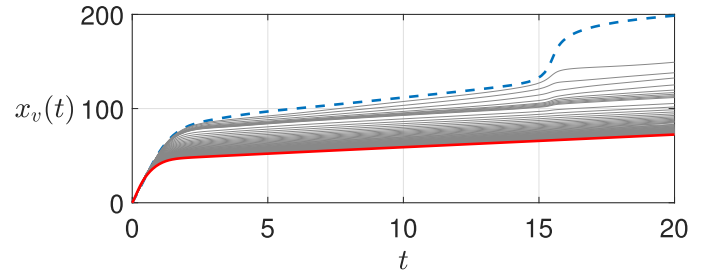


Fig. 8. Time histories of the auxiliary state-variable  $x_v$ , which, in this case, represents the dominant part of the cost functional.

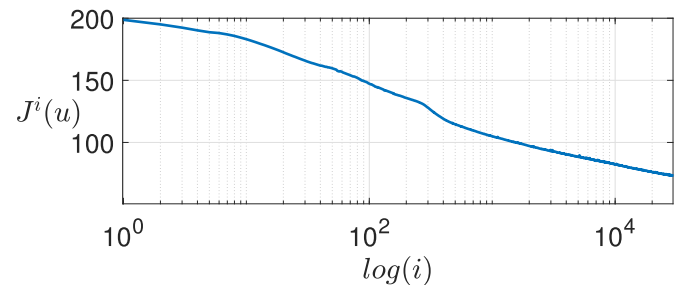


Fig. 9. Cost functional evaluated at each iteration of the proposed algorithm.

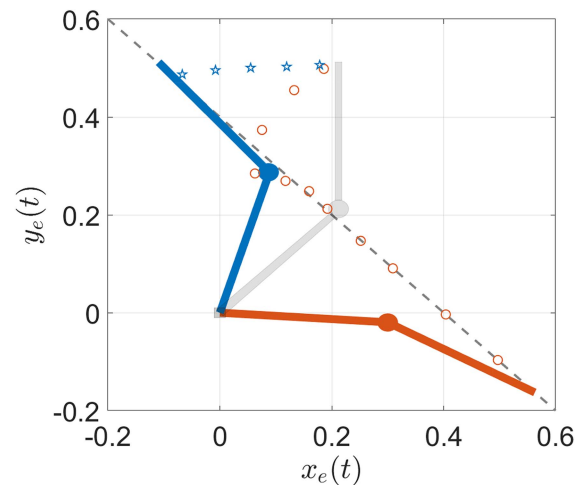


Fig. 10. Trajectories of the 2-DOF planar manipulator, obtained by applying the guess on the optimal control law given by (77) (circles), and from the application of the last updated control law  $u^*$  (stars). The starter links are reported in gray.

time, which is related to the cost functional. The behavior of the aforementioned cost functional, with respect to the iterations, can be visualized in Fig. 9, showing that the proposed scheme improves significantly, with respect to the optimality criteria herein adopted, the performance of the control law (77), since after less than 2500 iterations the cost appears to be reduced to more than the 50% of its initial value. Finally, in Fig. 10, the cartesian trajectory of the end-effector of the planar manipulator is plotted. The gray dashed line corresponds to the equation  $x_e(q) + y_e(q) = h_d$ , whereas the lines in gray represent the initial position of the manipulator. The trajectory individuated by

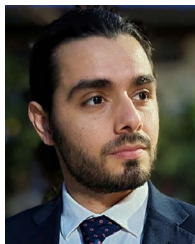
the red-circle points is given by applying the control law (77), with the orange lines indicating the final configuration of the manipulator in this case, while the blue-star points represent the trajectory obtained by applying the last update  $u^*$  of the control law, with the final configuration of the manipulator given by the blue lines. This figure shows, according to the previous figures, that the path traveled by the end-effector of the manipulator in the latter configuration is shorter than the former.

## VII. CONCLUSION AND FURTHER WORK

An algorithm to solve a class of finite-horizon optimal control problems for nonlinear systems has been presented. The key feature of the algorithm is that it is based on an iterative strategy that relies, by means of a suitable time-varying change of coordinates, on the solution of a linear initial value problem, which solves the time-varying LQR problem derived from the linearization of the nonlinear system together with the construction of the corresponding time-varying Hamiltonian dynamics, removing then the necessity of solving a time-varying Riccati equation. It has been shown, for the considered class of problems, that this method yields different convergence properties with respect to those provided by methods that linearize directly the nonlinear Hamiltonian dynamics, such as the quasilinearization method. It has been also shown how using heuristics borrowed from the machine learning field, such as the RMSProp, could improve the speed of convergence of the proposed scheme. Further work will involve the extension of the proposed scheme to *nonlinear differential games*, validation on more complex problems and the solution of problems in which constraints on both the control and the state of the system are present.

## REFERENCES

- [1] A. Bryson and Y. Ho, *Applied Optimal Control: Optimization, Estimation, and Control*. New York, NY, USA: Taylor & Francis, 1965.
- [2] D. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1. Belmont, MA, USA: Athena Scientific, 2012.
- [3] D. E. Kirk, *Optimal Control Theory: An Introduction*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1970.
- [4] L. D. Berkovitz, *Optimal Control Theory*, vol. 12. Berlin, Germany: Springer, 2013.
- [5] R. B. Vinter and R. Vinter, *Optimal Control*. Berlin, Germany: Springer, 2010.
- [6] D. Liberzon, *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton, NJ, USA: Princeton Univ. Press, 2011.
- [7] A. Van der Schaft, *L2-Gain and Passivity Techniques in Nonlinear Control*. Berlin, Germany: Springer, 2000.
- [8] J. Engwerda, *LQ Dynamic Optimization and Differential Games*. Hoboken, NJ, USA: Wiley, 2005.
- [9] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, vol. 1. New York, NY, USA, Wiley, 1972.
- [10] B. D. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*. Chelmsford, MA, USA: Courier Corporation, 2007.
- [11] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.
- [12] R. Bellman, "On the theory of dynamic programming," *Proc. Nat. Acad. Sci.*, vol. 38, no. 8, pp. 716–719, 1952.
- [13] L. S. Pontryagin, *Mathematical Theory of Optimal Processes*. Boca Raton, FL, USA: CRC Press, 1987.
- [14] M. Sassano, T. Mylvaganam, and A. Astolfi, "Infinite-horizon optimal control problems for nonlinear systems," in *Proc. IEEE 60th Conf. Decis. Control*, 2021, pp. 1721–1741.
- [15] M. Sassano and A. Astolfi, "Combining Pontryagin's principle and dynamic programming for linear and nonlinear systems," *IEEE Trans. Autom. Control*, vol. 65, no. 12, pp. 5312–5327, Dec. 2020.
- [16] J. T. Betts, "Survey of numerical methods for trajectory optimization," *J. Guid., Control, Dyn.*, vol. 21, no. 2, pp. 193–207, 1998.
- [17] A. V. Rao, "A survey of numerical methods for optimal control," *Adv. Astronautical Sci.*, vol. 135, no. 1, pp. 497–528, 2009.
- [18] H. Jaddu and E. Shimemura, "Computation of optimal control trajectories using Chebyshev polynomials: Parameterization, and quadratic programming," *Opt. Control Appl. Methods*, vol. 20, no. 1, pp. 21–42, 1999.
- [19] J. Vlassenbroeck and R. Van Dooren, "A Chebyshev technique for solving nonlinear optimal control problems," *IEEE Trans. Autom. Control*, vol. 33, no. 4, pp. 333–340, Apr. 1988.
- [20] R. E. Bellman and R. Kalaba, *Quasilinearization and Nonlinear Boundary-Value Problems*. New York, NY, USA, American Elsevier, 1965.
- [21] H. Wang, D. Wu, and C. Yu, "Solutions of nonlinear optimal control problems using quasilinearization and Fenchel duality," 2022, *arXiv:2201.07483*.
- [22] C. Mracek, J. Cloutier, and C. A. D'Souza, "A new technique for nonlinear estimation," in *Proc. IEEE Int. Conf. Control Appl. Together IEEE Int. Symp. Intell. Control*, 1996, pp. 338–343.
- [23] T. Cimen and S. P. Banks, "Nonlinear optimal tracking control with application to super-tankers for autopilot design," *Automatica*, vol. 40, no. 11, pp. 1845–1863, 2004.
- [24] A. Sideris and J. E. Bobrow, "An efficient sequential linear quadratic algorithm for solving nonlinear optimal control problems," in *Proc. IEEE Amer. Control Conf.*, 2005, pp. 2275–2280.
- [25] J. Van Den Berg, "Iterated LQR smoothing for locally-optimal feedback control of systems with non-linear dynamics and non-quadratic cost," in *Proc. IEEE Amer. Control Conf.*, 2014, pp. 1912–1918.
- [26] E. Todorov and W. Li, "A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *Proc. IEEE Amer. Control Conf.*, 2005, pp. 300–306.
- [27] J. T. Betts, "Sparse Jacobian updates in the collocation method for optimal control problems," *J. Guid., Control, Dyn.*, vol. 13, no. 3, pp. 409–415, 1990.
- [28] G. Elnagar, M. A. Kazemi, and M. Razzaghi, "The pseudospectral Legendre method for discretizing optimal control problems," *IEEE Trans. Autom. Control*, vol. 40, no. 10, pp. 1793–1796, Oct. 1995.
- [29] D. Garg, M. Patterson, W. W. Hager, A. V. Rao, D. A. Benson, and G. T. Huntington, "A unified framework for the numerical solution of optimal control problems using pseudospectral methods," *Automatica*, vol. 46, no. 11, pp. 1843–1851, 2010.
- [30] D. Garg, W. W. Hager, and A. V. Rao, "Pseudospectral methods for solving infinite-horizon optimal control problems," *Automatica*, vol. 47, no. 4, pp. 829–837, 2011.
- [31] C. Chen and Y. Liu, "Solution of two-point boundary-value problems using the differential transformation method," *J. Optim. Theory Appl.*, vol. 99, no. 1, pp. 23–35, 1998.
- [32] P. Roul and K. Thula, "A new high-order numerical method for solving singular two-point boundary value problems," *J. Comput. Appl. Math.*, vol. 343, pp. 556–574, 2018.
- [33] M. Soledad Aronna, J. Frédéric Bonnans, and P. Martinon, "A shooting algorithm for optimal control problems with singular arcs," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 419–459, 2013.
- [34] M. Sassano, "On the connection between costate and the annihilator of the Hamiltonian vector field in optimal control problems," *IEEE Trans. Autom. Control*, 2023, to be published.
- [35] Y. Ho, A. Bryson, and S. Baron, "Differential games and optimal pursuit-evasion strategies," *IEEE Trans. Autom. Control*, vol. AC-10, no. 4, pp. 385–389, Oct. 1965.
- [36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [37] L. Tarantino, M. Sassano, S. Galeani, and A. Astolfi, "Constant costate iterations for finite-horizon optimal control with nonlinear dynamics," in *Proc. IEEE 61st Conf. Decis. Control*, 2022, pp. 3481–3486.
- [38] P. J. Antsaklis and A. N. Michel, *Linear Systems*. Cambridge, MA, USA: Birkhäuser, 2006.
- [39] R. Bhatia, *Matrix Analysis*, vol. 169. Berlin, Germany: Springer, 2013.
- [40] D. G. Luenberger, *Optimization by Vector Space Methods*. Hoboken, NJ, USA: Wiley, 1997.
- [41] L. Lasdon, S. Mitter, and A. Waren, "The conjugate gradient method for optimal control problems," *IEEE Trans. Autom. Control*, vol. AC-12, no. 2, pp. 132–138, Apr. 1967.
- [42] A. Isidori, *Nonlinear Control Systems*. Berlin, Germany: Springer, 2013.
- [43] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
- [44] L. Sciacivico and B. Siciliano, *Modelling and Control of Robot Manipulators*. Berlin, Germany: Springer, 2001.



**Lorenzo Tarantino** received the B.S. degree in computer science engineering and the M.S. degree in automation engineering in 2018 and 2021, respectively, from the University of Rome “Tor Vergata,” Rome, Italy, where he is currently working toward the Ph.D. degree in control, computer science and geoinformation.

His research interests are focused on optimal control and differential games theory, hybrid systems, and nonlinear systems theory.



**Mario Sassano** (Senior Member, IEEE) was born in Rome, Italy, in 1985. He received the B.S. degree in automation systems engineering and the M.S. degree in systems and control engineering from the University of Rome “La Sapienza,” Rome, Italy, in 2006 and 2008, respectively. He received the Ph.D. degree in control theory from Imperial College London, London, U.K. in 2012.

From 2009 to 2012, he had been a Research Assistant with the Department of Electrical and Electronic Engineering, University of Rome La Sapienza. He is an Assistant Professor with the University of Rome “Tor Vergata,” Italy. His research interests are focused on nonlinear observer design, optimal control and differential game theory with applications to mechatronical systems and output regulation for hybrid systems. He is Associate Editor of the IEEE CONTROL SYSTEMS LETTERS, *European Journal of Control*, the IEEE CSS, EUCA Conference Editorial Boards. He is a member of the IFAC Technical Committee on Control Design.



**Sergio Galeani** received the Laurea degree and the Ph.D. degree in computer science and control engineering from the University of Roma Tor Vergata, Roma, Italy, in 1998 and 2002, respectively.

He is currently an Associate Professor with the University of Roma Tor Vergata, where his interests include periodic and multirate control systems, control systems with constraints, robust/adaptive antiwindup techniques, and hybrid, linear, and nonlinear output regulation.

Dr. Galeani is an Associate Editor for IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY and the Conference Editorial Board of the IEEE Control Systems Society and Chair of the IFAC Technical Committee 2.1 (Control Design).



**Alessandro Astolfi** (Fellow, IEEE) was born in Rome, Italy, in 1967. He received the graduate degree in electrical engineering from the University of Rome, Rome, Italy, in 1991. He received the M.Sc. degree in information theory from ETH-Zurich, Zurich, Switzerland, in 1995 and the Ph.D. degree with medal of honor in 1995 with a thesis on discontinuous stabilization of nonholonomic systems. He received the Ph.D. degree in control theory from the University of Rome “La Sapienza,” Rome, Italy, in 1996, for

his work on nonlinear robust control.

Since 1996, he has been with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., where he is currently a Professor of Nonlinear Control Theory. From 1998 to 2003, he was also an Associate Professor with the Department of Electronics and Information, Politecnico di Milano. Since 2005, he has also been a Professor with Dipartimento di Ingegneria Civile e Ingegneria Informatica, University of Rome Tor Vergata. He has been a Visiting Lecturer in “Nonlinear Control” in several universities, including ETH-Zurich from 1995 to 1996; Terza University of Rome in 1996; Rice University, Houston in 1999; Kepler University, Linz, in 2000; SUPELEC, Paris, in 2001, Northeastern University in 2013. He is the author of more than 150 journal papers, 30 book chapters, and over 240 papers in refereed conference proceedings. He is the author (with D. Karagiannis and R. Ortega) of the monograph “Nonlinear and Adaptive Control with Applications” (Springer-Verlag). He served as an Associate Editor for *Automatica*, *Systems and Control Letters*, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *International Journal of Control*, *European Journal of Control*, and *Journal of the Franklin Institute*; as an Area Editor for the *International Journal of Adaptive Control and Signal Processing*; as a Senior Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL; and as an Editor-in-Chief for the *European Journal of Control*.

Dr. Astolfi is the recipient of the IEEE CSS A. Ruberti Young Researcher Prize in 2007, the IEEE RAS Googol Best New Application Paper Award in 2009, the IEEE CSS George S. Axelby Outstanding Paper Award in 2012, and the Automatica Best Paper Award in 2017. He is a “Distinguished Member” of the IEEE CSS and IFAC Fellow. He is currently Editor-in-Chief of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He served as the Chair of the IEEE CSS Conference Editorial Board (2010–2017) and in the IPC of several international conferences. He has been/is a Member of the IEEE Fellow Committee (2016), (2019–2023).