Optimal Control for Continuous-time Nonlinear Systems based on a Linear-like Policy Iteration

Adnan Tahirovic[1] and Alessandro Astolfi[2]

***Abstract*— We propose a novel strategy to construct optimal controllers for continuous-time nonlinear systems by means of linear-like techniques, provided that the optimal value function is differentiable and quadratic-like. This assumption covers a wide range of cases and holds locally in general. The proposed strategy avoids solving the Hamilton-Jacobi-Bellman (HJB) equation, that is a nonlinear partial differential equation, which is known to be hard or impossible to solve. Instead, the HJB equation is replaced with an easy-solvable state-dependent Lyapunov matrix equation without introducing any approximation. We achieve this exploiting a linear-factorization of the underlying nonlinear system and a policy-iteration algorithm (PI) to yield a linear-like PI for nonlinear systems. The proposed control strategy solves optimal nonlinear control problems in an exact, yet still linear-like manner. We prove optimality of the resulting solution and illustrate the results via two examples.**

I. INTRODUCTION

The solution of optimal control problems for nonlinear systems is based on the solution of the HJB partial differential equations (PDE), which can be extremely difficult or impossible to solve. Many approximation methods for solving the HJB PDE have been developed, under a variety of assumptions, at the cost of some optimality loss [1].

A first class of techniques is based on the theory of viscosity solutions of the HJB PDE [2]. This solution is proved to be the value function of the underlying optimal control. It is required to be continuous, and not necessarily differentiable, as it is assumed for classical solutions. For this reason, the theory of viscosity solutions also provides a tool for dealing with existence and uniqueness issues for nonlinear PDEs. To get an approximate viscosity solution, finite-difference and finite-element methods have been used: both require a discretization of the state space, hence the computational cost increases exponentially with the dimension of the state space.

A second class of techniques, also relevant to this paper, is based on the PI algorithm, which reduces a nonlinear HJB PDE to a linear PDE [3], [4]. This is used to find the cost associated to an admissible control. The PI algorithm also provides an incremental improvement of the control policy and ensures convergence to the optimal control. In many cases, solving a linear PDE is still not easy. In [5], Galerkin approximations have been used to approximately solve optimal control problems by combining this approximation with the PI algorithm. Some other approaches developed to approximate the solution of the HJB PDE, up to a desired degree of accuracy, have been presented in [6]–[8].

A third class of techniques is based on results obtained for linear systems and for a cost in quadratic form. For such systems the HJB PDE reduces to an algebraic Riccati equation (ARE), which is easy to solve. The methods based on Jacobian linearization of the nonlinear system, feedback linearization [9], [10], dynamic extensions [11], and state-dependent Riccati equations (SDRE) [12]–[14], represent techniques to approximate the optimal control by avoiding solving nonlinear PDEs. The linearization-based approach is feasible only in the vicinity of an equilibrium, while feedback-linearization may cancel "useful" nonlinearities and may not provide a near-to-optimal control law. The dynamic extension-based approach relies on a modified cost to avoid solving the HJB PDE, providing thus a suboptimal control law. It is worth noting that the dynamic extension-based control is capable to extract an upper bound of the modified cost to provide a measure of the sub-optimality level of the solution. The SDRE-based control approach relies upon a *linear-like factorization of the nonlinear system*. Its main disadvantage is the lack of stability guarantee.

We propose a control strategy for input-affine continuous-time nonlinear systems which is based on the PI paradigm combined with the linear-like factorization used in the SDRE approach. We use the PI algorithm to ensure convergence of the policy to the optimal control. Unlike other PI approaches, we use the linear-like factorization of the nonlinear system to avoid solving any PDE, thus replacing the PDE with a state-dependent Lyapunov matrix equation (SDLE). In this way the proposed control strategy solves the optimal nonlinear control problem in an exact, but still linear-like, manner, provided the optimal cost is in a quadratic-like form. If this is not a case, the proposed approach may still ensure a near-optimal solution in the vicinity of an equilibrium, around which it provides a powerful approximation method.

In Section II we define the problem and recall a general form of the PI algorithm. In Section III we recall the SDRE approach with its associated factorization technique and re-define the optimal control problem. In Section IV, we define the linear-like PI which computes the optimal control with a modified cost. Section V introduces the modified linear-like PI to solve the considered nonlinear optimal control problem. Section VI provides an illustration of the results via two examples, while Section VII concludes the paper.

*This work has been partially supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 739551 (KIOS CoE).

[1]Adnan Tahirovic is with Faculty of Electrical Engineering, University of Sarajevo, atahirovic@etf.unsa.ba

[2]Alessandro Astolfi is with Imperial College London, U.K., and also with Dipartimento di Informatica, Sistemi e Produzione, Universita di Roma, Italy. a.astolfi@imperial.ac.uk

II. Control based on policy iteration for continuous-time systems

A. Problem description

Consider a class of continuous-time nonlinear systems described by equations of the form

$$\dot{x} = f(x) + g(x)u, \tag{1}$$

with state $x(t) \in \mathbb{R}^n$, input $u(t) \in \mathbb{R}^m$ and f and g Lipschitz continuous on a compact set $\Omega \subset \mathbb{R}^n$ that contains the origin. Suppose in addition that the system (1) has an equilibrium at the origin for $u = 0$, that is $f(0) = 0$. Finally, assume that the system is controllable in Ω, that is, it is possible to find an input signal $u(t)$ which steers the state of the system to the origin $x_e = 0$ from any initial condition x_0 in Ω.

Consider now the cost function

$$V(x_0, u) = \int_0^\infty (l(x) + \|u\|_R^2) dt, \tag{2}$$

where the state penalty function l is a positive function on Ω, such that $l(0) = 0$, the system (1) with output $y = l(x)$ is zero-state observable, and $R \in \mathbb{R}^{mxm}$ is a symmetric positive definite matrix. Typically, $l(x)$ is quadratic, that is $l(x) = x^T Q x$, where Q is a positive semidefinite matrix.

A feedback control $u = u(x)$ is called an *admissible control*, $u \in \mathscr{A}(\Omega)$, with respect to l on Ω, if u is continuous on Ω, $u(0) = 0$, the zero equilibrium of the closed-loop system is locally asymptotically stable with basin of attraction containing Ω, and the cost (2) is finite for all $x_0 \in \Omega$. The minimal value of the cost function V, obtained for an admissible control $u^*(x)$ (the optimal control), is denoted as the optimal cost $V^*(x)$, $\forall x \in \Omega$. This optimal cost V^*, called the value function, is the solution of the HJB equation

$$\frac{\partial V^*(x)}{\partial x} f(x) - \frac{1}{4} \frac{\partial V^*(x)}{\partial x} g(x) R^{-1} g(x)^T \frac{\partial V^*(x)^T}{\partial x} + l(x) = 0, \tag{3}$$

which is a PDE, provided V is differentiable. Equation (3) is in general hard to solve even in those cases in which a unique solution is known to exist. The PDE makes the optimal control problem virtually impossible to solve in closed-form. If a solution exists, the optimal control is

$$u^* = u^*(x) = -\frac{1}{2} R^{-1} g^T(x) \frac{\partial V^*(x)}{\partial x}^T. \tag{4}$$

B. Policy iteration for nonlinear systems

To compute the value of the cost $\hat{V}_0(x_0)$, for a fixed initial condition x_0 and an admissible control $\hat{u}$, one has to solve (1) with $u = \hat{u}$, which is not always possible, and compute the integral (2) along the corresponding solution. Another way to deal with this problem is to differentiate (2) along the trajectories of the system yielding the linear PDE

$$\frac{\partial \hat{V}(x)}{\partial x} (f(x) + g(x)\hat{u}(x)) + l(x) + \|\hat{u}\|_R^2 = 0, \tag{5}$$

which represents an incremental expression of the cost of the admissible control $\hat{u}$, and it does not depend on the solution trajectories of the system (1). If the optimal control (4) is used, *i.e.* $\hat{u} = u^*$, then (5) transforms into the nonlinear PDE (3), the solution of which directly provides the optimal cost V^* and the optimal control law u^*.

The optimal PI for continuous-time nonlinear systems has been proposed in [4]. The main idea of this iterative algorithm is to choose an arbitrarily initial admissible control $\hat{u}(x) \in \mathscr{A}(\Omega)$ and solve the linear PDE (5) for $\hat{V}$, which should be easier to solve than the nonlinear PDE (3). In order to improve the performance of the arbitrarily selected control $\hat{u}(x)$, one then defines the policy-update

$$\hat{u}^*(x) = \arg\min_u \frac{\partial \hat{V}(x)}{\partial x} (f(x) + g(x)\hat{u}(x)) + l(x) + \|\hat{u}\|_R^2 = -\frac{1}{2} R^{-1} g^T(x) \frac{\partial \hat{V}(x)}{\partial x}^T, \forall x \in \Omega. \tag{6}$$

Having a new and improved control $\hat{u}^*$, one can again solve (5) to obtain the value function $\hat{V}$. By iteratively improving the value function and the control law iterating (5) and (6), the optimal PI algorithm ensures, in principle, the desired convergence, i.e. $\lim_{k\to\infty} \hat{V}_k(x) = V^*(x)$ and $\lim_{k\to\infty} \hat{u}_k(x) = u^*(x)$, $\forall x \in \Omega$, where k is the index of the iteration.

Although equation (5) should be easier to solve for $\hat{V}$ than solving (1) and (2), it is still difficult. For this reason different approaches to approximately deal with equation (5) have been proposed, see, e.g. [4], [5]. The goal of this paper is to show how PI can be exploited to find the optimal control solution without the need to solve any PDE on the basis of a simple linear-like procedure.

C. Policy iteration for linear systems

In this section we consider linear systems, that is system (1) with $f(x) = Ax$, with $A \in \mathbb{R}^{nxn}$, $g(x) = B$, with $B \in \mathbb{R}^{nxm}$ and a quadratic cost, that is $l(x) = x^T Q x$, with $Q = Q^T \geq 0$, in (2). Assume that the pair (A, B) is stabilizable and the pair $(Q^{1/2}, A)$ is detectable.

Assuming that the optimal value function is of the form

$$V^*(x) = x^T P^* x, \tag{7}$$

where $P^* = P^{*T}$ is a positive definite matrix, the HJB equation (3) becomes the ARE

$$A^T P^* + P^* A - P^* B R^{-1} B^T P^* + Q = 0, \tag{8}$$

which is easily solvable and has a unique positive definite solution P^*. The optimal control action can then be computed from (4) yielding

$$u^*(x) = -R^{-1} B^T P^* x = \Pi^* x, \tag{9}$$

where Π^* is the optimal control policy.

Although the solution to the optimal control problem for continuous-time linear systems can be given in the closed-form (9), we recall the optimal PI algorithm to understand how to construct the optimal control in an iterative manner.

In the simplified version of the optimal PI algorithm for linear systems the cost-update equation (5) becomes the Lyapunov Matrix Equation (LME)

$$(A + B\hat{\Pi})^T \hat{P} + \hat{P}(A + B\hat{\Pi}) + Q + \hat{\Pi}^T R \hat{\Pi} = 0, \tag{10}$$

which can be easily solved for a positive definite matrix $\hat{P}$, provided an admissible control $\hat{u} = \hat{\Pi}x$ is given. Additionally, the policy-update equation (6) for linear systems becomes

$$\hat{u}^* = \hat{\Pi}^* x = -R^{-1}B^T\hat{P}x, \quad \hat{\Pi}^* = -R^{-1}B^T\hat{P}. \tag{11}$$

The proof for this linear case is provided in [15], where it has been shown that the PI is actually Kleiman-Newton's method, which ensures convergence to the solution of the ARE whenever the initial control is admissible.

III. Pointwise factorization of the optimal control problem

Under mild regularity assumptions the nonlinear system (1) can be rewritten in the form

$$\dot{x} = A(x)x + g(x)u, \tag{12}$$

where $A(x) : \mathbb{R}^n \to \mathbb{R}^{nxn}$ is a smooth matrix valued function. The main idea behind the factorizations of the function $f(x)$ as $f(x) = A(x)x$ is to represent the nonlinear system (1) as a pointwise linear system by assuming that $A(x)$ and $g(x)$ are constant matrices for each state x along the trajectories of the system (see, e.g. [14]).

In the spirit of the above factorization, similarly to the linear case, we assume a pointwise quadratic form for the optimal value function, namely

$$V^*(x) = x^T P^*(x)x, \tag{13}$$

where $P^* = P^*(x)$ is a state-dependent matrix valued function and it is positive definite for all $x \in \Omega$.

For clarity, we first define the solution to the SDRE [14], which represents the factorized version of the ARE (8).

Definition 1 *A positive definite matrix $\bar{P}$ is the pointwise solution to the SDRE for the state x if*

$$A(x)^T\bar{P} + \bar{P}A(x) - \bar{P}g(x)R^{-1}g(x)^T\bar{P} + Q = 0. \tag{14}$$

As in the case of the ARE, the SDRE is easily solvable for each fixed $x \in \Omega$. By mimicking the linear-like procedure presented in II-C, the control action can be computed by (9) in the pointwise form

$$u^*(x) = -R^{-1}g^T(x)\bar{P}x = \bar{\Pi}x. \tag{15}$$

Equations (14) and (15) form the SDRE-based control method: (14) is solved for each x along the trajectories of the system and the control law is computed as in (15).

Note that the SDRE-based control does not provide the optimal solution to the optimal control problem for the nonlinear system, since (14) has not been derived from the HJB equation (3). Another issue pertains to the matrix $\bar{P}$, for which we do not have a closed form solution, that is $\bar{P} = \bar{P}(x)$, but only the pointwise value for each state x along the trajectories of the system. This prevents $V = x^T\bar{P}x$ from being a Lyapunov function candidate, since its time derivative along the trajectories of the system, namely

$$\dot{V}^*(x) = \dot{x}^T\bar{P}(x)x + x^T\bar{P}(x)\dot{x} + x^T\dot{\bar{P}}(x)x, \tag{16}$$

has the additional term $\dot{\bar{P}}(x)$, which is impossible to obtain analytically and to be used for further analysis.

Lemma 1 [Direct optimal control] *Assume that the optimal value function for the optimal control problem for the nonlinear factorized system (12) is given in the quadratic-like form (13), where $P^*(x)$ is a positive definite matrix for all $x \in \Omega$. Then $P^*(x)$ is the solution of the HJB equation*

$$\begin{aligned} &x^T\{A(x)^TP^* + P^*A(x) - P^*g(x)R^{-1}g(x)^TP^* + Q\}x \\ &+ u_{corr}^T R u_{corr} + x^T\dot{P}^*x = 0, \end{aligned} \tag{17}$$

while the optimal control is given by $u^* = \bar{u} + u_{corr}$, where

$$\bar{u} = -R^{-1}g^T(x)P^*(x)x, \tag{18}$$

$$u_{corr} = -\frac{1}{2}R^{-1}[\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j g^T(x)\frac{\partial p_{i,j}}{\partial x}], \tag{19}$$

and $p_{i,j}$ indicates the $(i,j)^{th}$ element of the matrix $P^*(x)$.

Although Lemma 1 provides the exact solution to the optimal control problem, the HJB equation (17), which is itself a PDE, is as hard to solve for P^* as the initial HJB equation (3). However, equation (17) allows for a separation of the optimal control problem into two simpler problems, one aimed at finding the solution $\bar{u}$, which is a counterpart of (14), and the second one aimed at finding a correction term from the last two terms in (17), which are discussed in Section IV and V, respectively.

IV. An approximate control based on linear-like policy iteration

A. The State-dependent Lyapunov Equation - SDLE

The main idea behind the linear-like PI is to use the PI algorithm for nonlinear systems by avoiding using PDEs, *i.e.* by using only Lypaunov matrix equations as in the linear case presented in Section II-C. To do so, we conduct the PI by omitting the last two terms in (17) to get the Lyapunov equation instead of the PDE at the cost of optimality loss. For clarity, we define the State-dependent Lyapunov Equation (SDLE) which is used as the approximate cost-update equation in the PI algorithm.

Definition 2 (Approximate cost-update) *Consider the admissible control $\hat{u} = \hat{\Pi}x \in \mathscr{A}(\Omega)$. A differentiable function $\hat{V} = x^T\hat{P}(x)x : \Omega \to \mathbb{R}$ ($\hat{V}(0) = 0$), where $\hat{P}(x)$ is a positive definite matrix, is the approximate cost function of $\hat{u}$ if $\hat{P}(x)$ satisfies the SDLE*

$$(A(x) + g(x)\hat{\Pi})^T\hat{P} + \hat{P}(A(x) + g(x)\hat{\Pi}) + Q + \hat{\Pi}^T R\hat{\Pi} = 0. \tag{20}$$

We call (20) the approximate cost-update equation for the nonlinear system and write $\hat{P}(x) = CU_{SDLE}(\hat{\Pi}(x))$, where the index SDLE indicates that one has to solve the state-dependent Lyapunov matrix equation (20) to obtain $\hat{P}(x)$.

Note first that this equation is easy solvable as in the linear case (10). Moreover, unlike the idea behind the SDRE (14), where P is computed pointwise for each single x along the trajectories of the system, the SDLE provides an analytical

form of $\hat{P}(x)$. Having $\hat{P}(x)$ in closed form, it is then possible to compute the time derivative $\dot{\hat{P}}(x)$ along the trajectories of the system, thus circumventing one of the main limitations of the SDRE-based approach.

Note also that the SDLE can be derived from (5), by leting $\hat{u} = \bar{u} + u_{corr}$, where the terms equal to the last two terms in (17) are omitted for simplicity. This would mean that the SDLE can be considered as the cost-update equation when taking $u_{corr}(x) = 0$, for all x, and by omitting the time derivative $\dot{P}(x)$. For this reason we call (20) *the approximate cost-update equation*, and we write $\hat{P}(x) = CU_{SDLE}(\hat{\Pi})$.

B. The linear-like policy iteration based on the SDLE

Along with *Definition* 2, we introduce a new definition and a result to define the linear-like PI based on the SDLE.

Definition 3 [Approximate policy-update] *Consider the differentiable function $\hat{V} = x^T\hat{P}(x)x : \Omega \to \mathbb{R}$ ($\hat{V}(0) = 0$), where for each x, $\hat{P}(x)$ is a positive definite matrix. The control $\hat{u}^*$ is said to update the control $\hat{u}$ (or the policy $\hat{\Pi}^*$ updates the policy $\hat{\Pi}$) in accordance with the approximate policy-update equations for nonlinear systems*

$$\hat{u}^* = -R^{-1}g(x)^T\hat{P}(x)x, \quad \hat{\Pi}^* = -R^{-1}g(x)^T\hat{P}(x), \tag{21}$$

and we write $\hat{\Pi}^* = PU_{SDLE}(\hat{P}(x))$.

Note that (21) includes only the first term (18) of the optimal control given by (18)-(19). For this reason, we also call $\hat{\Pi}^* = PU_{SDLE}(\hat{P}(x))$ *the approximate policy-update equation*.

Theorem 1 [Control based on linear-like policy iteration] *Consider an admissible control $\hat{u}_k = \hat{\Pi}_k x$, which ensures that the matrix $A + g\hat{\Pi}_k$ is stable, and assume that the matrix $Q + \hat{\Pi}_k^T R\hat{\Pi}_k$ is positive definite. Then there is a unique, symmetric and positive definite solution $\hat{P}_k$ to the approximate cost-update in accordance with Definition 2, that is $\hat{P}_k = CU_{SDLE}(\hat{\Pi}_k)$. If now $\hat{\Pi}_{k+1}$ is computed by the approximate policy-update in accordance with Definition 3, that is $\hat{\Pi}_{k+1} = PU_{SDLE}(\hat{P}_k)$, then $\hat{V}_k > \hat{V}_{k+1}$, where $\hat{P}_{k+1} = CU_{SDLE}(\hat{\Pi}_{k+1})$. The pair $(\hat{P}_k, \hat{\Pi}_{k+1})$ represents the k^{th} iteration of the linear-like PI based on the SDLE for nonlinear systems. Moreover, this linear-like PI fully resembles the optimal PI with respect to the modified state cost $l(x) = x^T\bar{Q}x$, where $\bar{Q} = Q - \dot{P}$.*

We call the solution based on this approach the PI-SDLE control. One of the main advantages of the proposed PI-SDLE control is that the linear-like PI can also be computed pointwise using (20), instead of finding a closed form solution. In such a case, one needs to conduct the whole PI algorithm for every single x along the trajectories of the system. Such a procedure is similar to the pointwise computation of the ARE solution when the SDRE-based control is used. Unlike the SDRE-based control, the PI-SDLE based control is proven to be stabilizable in Ω provided the initial control is admissible.

V. Optimal control based on linear-like policy iteration

We now show how it is possible to use the linear-like PI proposed in *Theorem* 1 to obtain the optimal solution for the optimal control problem for continuous-time nonlinear systems.

Let the matrix $\bar{P}^1(x)$ and the control $\bar{u}^1(x)$ be the solutions obtained by the linear-like PI equations (22)-(23), that is

$$(A(x) + g(x)\hat{\Pi}_k^1)^T\hat{P}_k^1 + \hat{P}_k^1(A(x) + g(x)\hat{\Pi}_k^1) + \hat{\Pi}_k^{1^T}R\hat{\Pi}_k^1 + Q = 0, \tag{22}$$

$$\hat{u}_{k+1}^{*1} = -R^{-1}g^T(x)\hat{P}_k^1(x)x. \tag{23}$$

Let also the matrix $\bar{P}^2(x)$ and the control $\bar{u}^2(x)$ be the solutions obtained by the modified linear-like PI equations (24)-(25), that is

$$(A(x) + g(x)\hat{\Pi}_{k,i}^2)^T\hat{P}_{k,i}^2 + \hat{P}_{k,i}^2(A(x) + g(x)\hat{\Pi}_{k,i}^2) + \hat{\Pi}_{k,i}^{2^T}R\hat{\Pi}_{k,i}^2 + Q + \dot{\hat{P}}_{i-1}^2|_{A(x)x+g(x)\bar{u}_{i-1}^2} = 0, \tag{24}$$

$$\hat{u}_{k+1,i}^{*2} = -R^{-1}g^T(x)\hat{P}_{k,i}^2(x)x. \tag{25}$$

The index i indicates one complete i^{th} PI (24)-(25). $\dot{\hat{P}}_{i-1}^2|_{A(x)x+g(x)\bar{u}_{i-1}^2}$ is the time derivative of $\bar{P}_{i-1}^2$ along the trajectories of the system when $\bar{u}_{i-1}^2 = \bar{\Pi}_{i-1}^2 x$ is used. Both $\dot{\hat{P}}_{i-1}^2$ and $\bar{u}_{i-1}^2$ are obtained from the $(i-1)^{th}$ PI (24)-(25) as the respective solutions. This means that $\dot{\hat{P}}_{i-1}^2|_{A(x)x+g(x)\bar{u}_{i-1}^2}$ is a fixed matrix function during the i^{th} PI (24)-(25). The initial admissible control $\bar{u}_0^2$ and the matrix $\dot{\hat{P}}_0^2$ required for the first PI (24)-(25) ($i = 1$), are taken from the solutions of the PI (22)-(23), as $\bar{u}_0^2 = \bar{u}^1$ and $\dot{\hat{P}}_0^2 = \dot{\hat{P}}^1$.

Lemma 2 *The matrix $\dot{\hat{P}}_i^2$ ($\forall i$), is a positive definite matrix in the same region as the matrix $\dot{\hat{P}}^1$, $x^T\dot{\hat{P}}^1x \geq 0$.*

The result from *Lemma* 2 is required to ensure all matrices $\dot{\hat{P}}_{i-1}^2$ in (24) are positive definite when $x^T\dot{\hat{P}}^1x \geq 0$.

Theorem 2 *Assume that the optimal value function is in the form (13). Then the optimal control $u^*(x) \in \mathscr{A}(\Omega)$ is given as*

$$u^*(x) = \begin{cases} \bar{u}^2 & \text{if } x^T\dot{\hat{P}}^1x \geq 0 \\ \bar{u}^1 + \bar{u}_{corr}^1 & \text{if } x^T\dot{\hat{P}}^1x < 0, \end{cases} \tag{26}$$

where $\bar{u}^2 = -R^{-1}g(x)^T\bar{P}^2x$, $\bar{u}^1 = -R^{-1}g(x)^T\bar{P}^1x$, and $\bar{u}_{corr}^1$ is the correction component obtained as solution to the quadratic matrix equation

$$\bar{u}_{corr}^{1^T}R\bar{u}_{corr}^1 + x^T\dot{\hat{P}}^1|_{g(x)\bar{u}_{corr}^1}x + x^T\dot{\hat{P}}^1|_{A(x)x+g(x)\bar{u}^1}x = 0, \tag{27}$$

which preserves continuity in $u^*(x)$.

When $x^T\dot{\hat{P}}^1x \geq 0$, $\dot{\hat{P}}_{i-1}^2$ is positive definite (Lemma 2) implying that (24)-(25) is solvable for a positive definite $\dot{\hat{P}}_i^2$. For this reason, the repetition of the PI (24)-(25) can be interpreted as follows. The first PI ($i = 1$), for which this additional cost is $x^T\dot{\hat{P}}_0^2x = x^T\dot{\hat{P}}^1x$, aims at finding the control $\bar{u}_{i=1}$ (the control after the first PI is completed) for the modified state cost $x^T(Q + \dot{\hat{P}}_0^2)x$. During the first PI (24)-(25) for $i = 1$, the algorithm resembles the optimal control

for the modified cost $x^T(Q+\dot{\bar{P}}_0^2-\dot{\bar{P}}_1^2)x$ (Theorem 1). This cost can be interpreted as an improved cost with respect to the cost considered in the preceding policy itearation. After the i^{th} PI is completed, we obtain the modified state cost $x^T(Q+\dot{\bar{P}}_{i-1}^2-\dot{\bar{P}}_i^2)x$ for which the linear-like PI (24)-(25) finds the optimal solution. In accordance to the the PI algorithm, we have $\lim_{i\to\infty}(\dot{\bar{P}}_{i-1}^2-\dot{\bar{P}}_i^2)=0$. This means that the modified cost, with respect to which the linear-like PI (24)-(25) finds the optimal control, tends to the original state cost, hence the obtained control solution converges to the optimal control.

When $x^T\dot{\bar{P}}^1x<0$, the quadratic equation (27) follows from the last two terms of the HJB equation (17). It is solvable for a real-valued $\bar{u}_{corr}^1$ to provide the correction part of the optimal control.

VI. Illustrative Examples

We provide simulation results by considering two nonlinear systems. For the first system the optimal control and the optimal value function are known, so it is possible to assess the proposed approach against the optimal solution. In the second example we compare our approach against the control based on the Galerkin approximation (GAC) by considering a nonlinear system with an unknown optimal control policy and an unknown optimal value function. This example illustrates the capability of the proposed approach to solve such nonlinear control problems. In all examples, after the linear-like PI (22)-(23) is used, we complete only one modified PI (24)-(25), that is $i=1$, while both PIs have been conducted for three iterations only (up to $k=3$).

A. Optimal control of the Van Der Pol oscillator

Consider the Van Der Pol oscillator

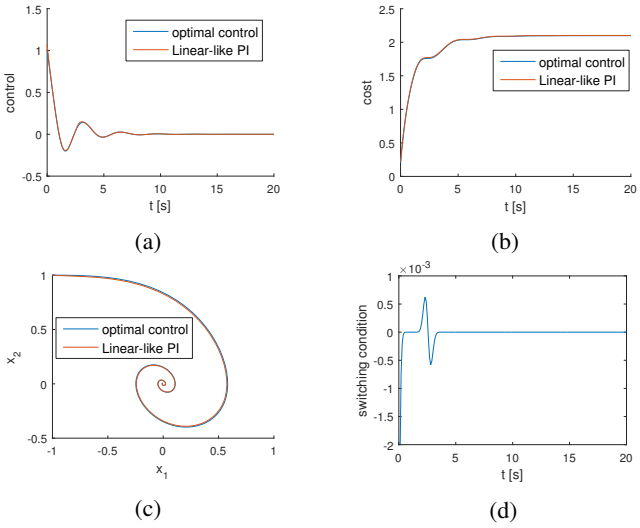
$$\dot{x}_1=x_2,\quad \dot{x}_2=-x_1-\mu(1-x_1^2)x_2+x_1u, \tag{28}$$

with $\mu=0.5$, the state cost $l(x)=x_2^2$ and assume $R=1$. The optimal control is $u^*=-x_1x_2$, with optimal value function $V^*=x_2^1+x_2^2$. The system can be easily factorized with

$$A(x)=\begin{bmatrix}0 & 1\\ -1 & -\frac{1}{2}(1-x_1^2)\end{bmatrix},\quad B(x)=\begin{bmatrix}0\\ x_1\end{bmatrix}. \tag{29}$$

The initial admissible control for the linear-like PI is selected to be the one that cancels out the nonlinearities and stabilizes the system, that is $u=-\frac{1}{2}x_1x_2$.

Fig. 1 provides a comparison between the proposed approach and the optimal control for the system for $x_0=[-1;1]$. From the control signals, cumulative costs and phase portrait, we conclude optimality of the proposed approach. In Fig. 1 one can also see the switching function $x^T\dot{\bar{P}}^1x$, which is used in (26). This function indicates the time intervals when the two different forms of the optimal control (26) have been used. Another interesting observation is that this function becomes zero before the states reach the origin. This phenomenon has not been investigated in this work, and it can be a promising direction for further understanding of the proposed framework. This means that the system trajectory

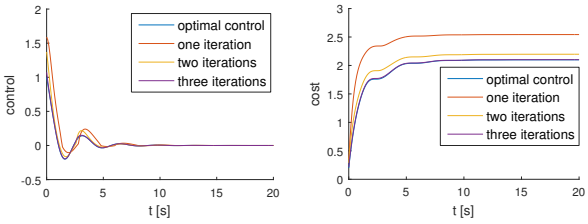


(a) (b) (c) (d)

Fig. 1: Comparison between the optimal and proposed controls in terms of control signals (a), cumulative costs (b), and phase portrait (c) obtained along the trajectory from the initial condition $x=[-1;1]$. Subfigure (d) shows the values of the boundary function used in (26), that is $x^T\dot{\bar{P}}^1x$.

has approached the hyper-surface $x^T\dot{\bar{P}}^1x=0$ (in this example, a curve) and then has moved along this surface towards the equilibrium. Somewhat surprisingly, once the system states are on this hyper-surface, along which the PIs (22)-(23) and (24)-(25) are equivalent, one only needs the linear-like PI (22)-(23) to obtain the remaining part of the optimal control.

Fig. 2 illustrates how the control signal obtained using the proposed approach (a) and its associated cumulative cost (b) converge towards the optimal values depending on the number of iterations used in the linear-like policy-iteration (22)-(23) and in the modified linear-like policy-iteration (24)-(25). One can observe that both cumulative cost and control signal obtained after three iterations (green) are almost identical to the optimal counterparts (blue). We

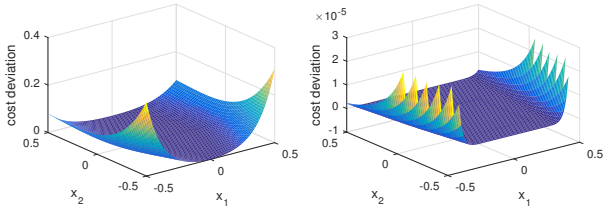


(a) (b)

Fig. 2: Cumulative costs (a) and control signals (b) obtained for the initial condition $x=[-1;1]$ when using a different number of iterations, for both policy-iterations (22)-(23) and (24)-(25), that is $k=\{1\},k=\{1,2\}$ and $k=\{1,2,3\}$.

provide also the final expression of the estimated optimal cost function for the case based on one iteration, due to a very high order of the rational function produced by the case based on three iterations. The estimated optimal function based on

only one iteration is $\hat{V}_1(x) = \dfrac{(x_1^2x_2^2+4)(x_1^2+x_2^2)}{4(x_1^2x_2-x_1^2+1)}$.



(a) (b)

Fig. 3: Cost deviation between the one which is iteratively improved by the linear-like PI approach and the optimal one. The deviations shown in subfigures (a) and (b) are obtained based on one and three iterations, respectively.

Fig. 3 shows the deviation of the estimated optimal cost function obtained from the proposed approach and the actual value function. In Fig. 3a, only one iteration is conducted, that is $k = \{1\}$, while Fig. 3b shows the deviation resulting after three iterations, that is $k = \{1,2,3\}$. From Fig. 3b, one can observe that the convergence is locally achieved.

B. Comparison against control based on Galerkin approximations

Consider the nonlinear system

$$\dot{x} = \begin{bmatrix} -x_1^3 - x_2 \\ x_1 + x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \tag{30}$$

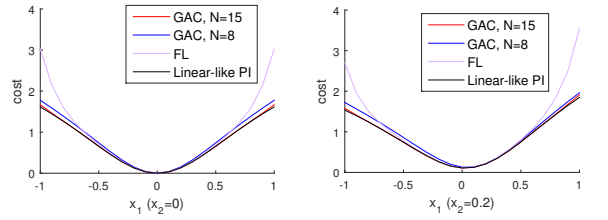
with the state cost $l(x) = x_1^2 + x_2^2$ and assume $R = 1$. The system can be easily factorized with

$$A(x) = \begin{bmatrix} -x_1^2 & -1 \\ 1 & 1 \end{bmatrix}, \quad B(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{31}$$

The initial admissible control for the linear-like PI is selected to be the control based on feedback linearization (FL) which is obtained in the form [5]

$$u(x) = 3x_1^5 + 3x_1^2x_2 - x_2 + 0.4142x_1 - 1.3522(x_1^3 + x_2). \tag{32}$$

The GAC solution has been obtained for different orders of the approximation and those can be found in [5]. In this example, we use two such controls obtained for $N = \{8, 15\}$.



(a) (b)

Fig. 4: The cost values for different initial conditions for x_1, where $x_2 = 0$ (a) and $x_2 = 0.2$ (b).

We illustrate the comparison of the GAC, FL and the proposed approach in terms of their associated costs as in [5]. Fig. 4 shows the costs that have been obtained for different initial conditions in x_1, while x_2 is constant, that is $x_2 = 0$ (a) and $x_2 = 0.2$ (b). One can observe that the proposed linear-like policy-iteration generates the minimal cost, although the GAC with $N = 15$ is similar. However, we stress that the GAC requires a number of preconditions for a valid implementation [5].

VII. CONCLUSION

We have presented several results to develop a method to determine optimal control strategies for continuous-time nonlinear systems. In Definitions 2 and 3 we have defined the approximate linear-like PI based on the SDLE to compute an approximate control law. The potential of such approximate control framework can be seen from the result in Theorem 1, in which it has been shown that the control is optimal with respect to a modified state cost.

In Theorem 2 we have given the main result which pertains to the definition of a novel optimal control approach for continuous-time nonlinear systems. From the results obtained on two case studies one can observe the optimality of the proposed approach and conclude that the proposed approach can be a control choice even in cases in which the optimal value function is not known.

REFERENCES

[1] D. Bertsekas, Dynamic programming and optimal control. Vol. 1. No. 2. Belmont, MA: Athena scientific, 1995.

[2] M.G. Crandall, and P.L. Lions. "Viscosity solutions of Hamilton-Jacobi equations." Transactions of the American mathematical society, 1983.

[3] R.J. Leake, and R.W. Liu. "Construction of suboptimal control sequences." SIAM Journal on Control, 1967.

[4] A. Wernrud, and A. Rantzer. "On approximate policy iteration for continuous-time systems." Proceedings of the 44th IEEE Conference on Decision and Control, 2005.

[5] R.W. Beard, G.N. Saridis, and J. T. Wen. "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation." Automatica, 1997.

[6] A.J. Krener, "The existence of optimal regulators." Proceedings of the 37th IEEE Conference on Decision and Control, 1998.

[7] W.M. McEneaney, "A curse-of-dimensionality-free numerical method for solution of certain HJB PDEs." SIAM journal on Control and Optimization, 2007

[8] A. Wernli, and G. Cook. "Suboptimal control for the nonlinear quadratic regulator problem." Automatica, 1975..

[9] A. Isidori. Nonlinear control systems. Springer Science and Business Media, 2013.

[10] R. Marino. "An example of a nonlinear regulator." IEEE Transactions on Automatic Control, 1984.

[11] M. Sassano, and A. Astolfi. "Dynamic approximate solutions of the HJ inequality and of the HJB equation for input-affine nonlinear systems." IEEE Transactions on Automatic Control, 2012.

[12] J.D. Pearson. "Approximation methods in optimal control I. Sub-optimal control." International Journal of Electronics, 1962.

[13] C.P. Mracek, and J.R. Cloutier. "Control designs for the nonlinear benchmark problem via the state dependent Riccati equation method." International Journal of robust and nonlinear control, 1998.

[14] T. Cimen. "State-dependent Riccati equation (SDRE) control: A survey." IFAC Proceedings Volumes, 2008.

[15] D. Vrabie, et al. "Adaptive optimal control for continuous-time linear systems based on policy iteration." Automatica, 2009.