MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Learning-Based Modular Indirect Adaptive Control for a Class of Nonlinear Systems

Benosman, M.; Farahmand, A.-M.

TR2016-055 July 2016

Abstract

We study in this paper the problem of adaptive trajectory tracking control for a class of nonlinear systems with parametric uncertainties. We propose to use a modular approach: we first design a robust nonlinear state feedback that renders the closed loop inputto-state stable (ISS). Here, the input is considered to be the estimation error of the uncertain parameters, and the state is considered to be the closed-loop output tracking error. Next, we augment this robust ISS controller with a model-free learning algorithm to estimate the model uncertainties. We implement this method with two different learning approaches. The first one is a model-free multi-parametric extremum seeking (MES) method and the second is a Bayesian optimizationbased method called Gaussian Process Upper Confidence Bound (GPUCB). The combination of the ISS feedback and the learning algorithms gives a learning-based modular indirect adaptive controller. We show the efficiency of this approach on a two-link robot manipulator example.

2016 American Control Conference (ACC)

Learning-Based Modular Indirect Adaptive Control for a Class of Nonlinear Systems

Mouhacine Benosman, Amir-massoud Farahmand, Meng Xia

***Abstract*— We study in this paper the problem of adaptive trajectory tracking control for a class of nonlinear systems with parametric uncertainties. We propose to use *a modular approach*: We first design a robust nonlinear state feedback that renders the closed loop input-to-state stable (ISS). Here, the input is considered to be the estimation error of the uncertain parameters, and the state is considered to be the closed-loop output tracking error. Next, we augment this robust ISS controller with a model-free learning algorithm to estimate the model uncertainties. We implement this method with two different learning approaches. The first one is a model-free multi-parametric extremum seeking (MES) method and the second is a Bayesian optimization-based method called Gaussian Process Upper Confidence Bound (GP-UCB). The combination of the ISS feedback and the learning algorithms gives a *learning-based modular indirect adaptive controller*. We show the efficiency of this approach on a two-link robot manipulator example.**

I. INTRODUCTION

Classical adaptive methods can be classified into two main approaches: 'direct' approaches, where the controller is updated to adapt to the process, and 'indirect' approaches, where the model is updated to better reflect the actual process. Many adaptive methods have been proposed over the years for linear and nonlinear systems; we cannot possibly cite them all. Instead we refer the reader to e.g., [1], [2] and the references therein for more detail. Of particular interest to us is the indirect modular approach to adaptive nonlinear control, e.g., [2]. In this approach, first the controller is designed by assuming that all the parameters are known and then an identifier is used to guarantee certain boundedness of the estimation error. The identifier is independent of the designed controller and thus the approach is called 'modular'. For example, a modular approach has been proposed in [3] for adaptive neural control of pure-feedback nonlinear systems, where the input-to-state stability (ISS) modularity of the controller-estimator is achieved and the closed-loop stability is guaranteed by the small-gain theorem (see also [4], [5]).

In this work, we build upon this type of modular adaptive design and provide a framework that combines model-free learning methods and robust model-based nonlinear control. We propose a learning-based modular indirect adaptive controller, in which model-free learning algorithms are used to estimate, in closed-loop, the uncertain parameters of the model. The main difference with the existing model-based indirect adaptive control methods is the fact that we do not use the model to design the uncertainty parameters estimation filters. Indeed, model-based indirect adaptive controllers are based on parameters estimators designed using the system's model, e.g., the X-swapping methods presented in [2], where gradient descent filters obtained using the systems dynamics are designed to estimate the uncertain parameters. We argue that because we do not use the system's dynamics to design uncertainties estimation filters we have less restrictions on the type of uncertainties that we can estimate, e.g., uncertainties appearing nonlinearly can be estimated with the proposed approach, see [6] for some earlier results on a mechatronics application. We also show that with the proposed approach we can estimate at the same time a vector of linearly dependent uncertainties, a case which cannot be straightforwardly solved using model-based filters, e.g., refer to [7] where it is shown that the X-swapping model-based method fails to estimate a vector of linearly dependent model coefficients.

We implement the proposed approach with two different model-free learning algorithms: The first one is a dither-based MES algorithm, and the second one is a Bayesian optimization-based method called GP-UCB. The latter solves the exploration-exploitation problem in the continuous armed bandit problem, which is a non-associative reinforcement learning (RL) setting. MES is a model-free control approach with well-known convergence properties, and has been analyzed in many papers, e.g., [8], [9]. This makes MES a good candidate for the model-free estimation part of our modular adaptive controller, as already shown in some of our preliminary results in [10], [11], [12]. However, one of the main limitations with dither-based MES is the convergence to local minima. To improve this part of the controller, we introduce another model-free learning algorithm in the estimation part of the adaptive controller. We propose in this paper to use a reinforcement learning algorithm based on Bayesian optimization methods, known as GP-UCB [13]. Contrary to the MES algorithm, GP-UCB is guaranteed to reach the global minima under certain mild assumptions.

One point worth mentioning at this stage is that comparing to 'pure' model-free controllers, e.g., pure MES or model-free RL algorithms, the proposed control has a different goal. The available model-free controllers are meant for output or state regulation, i.e., solving a static optimization problem. In contrast, we propose to use model-free learning to complement a model-based nonlinear control to estimate the unknown parameters of the model, which means that the control goal, i.e., state or output trajectory tracking is handled by the model-based controller. The learning algorithm is used to improve the tracking performance of the model-based controller, and once the learning algorithm has converged, one can carry on using the nonlinear model-based feedback controller alone, i.e., without the need of the learning algorithm. Furthermore, due to the fact that we are merging together a model-based control with a model-free learning algorithm, we believe that this type of controller can converge faster to an optimal performance, comparatively to the pure model-free controller, since by 'partly' using a model-based controller, we are taking advantage of the partial information given by the physics of the system, whereas the pure model-free algorithms assume no knowledge about the system, and thus start the search for an optimal control signal from scratch.

Similar ideas of merging model-based control and MES has been proposed in [14], [15], [16], [17], [6], [18], [10], [11], [12]. For instance, extremum seeking is used to complement a model-based controller, under the linearity of the model assumption in [14] (in the direct adaptive control setting, where the controllers gains are estimated), or in the indirect adaptive control setting, under the assumption of linear parametrization of the control in terms of the uncertainties in [15]. The modular design idea of using a model-based controller with ISS guarantee, complemented with an MES-

Mouhacine Benosman (m_benosman@ieee.org) and Amir-massoud Farahmand are with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. Meng Xia is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA (part of this work has been completed during her internship at MERL).

based module can be found in [6], [18], [10], [11], [12], where the MES was used to estimate the model parameters and in [17], [19], where feedback gains were tuned using MES algorithms. The work of this paper falls in this class of ISS-based modular indirect adaptive controllers. The difference with other MES-based adaptive controllers is that, due to the ISS modular design we can use any model-free learning algorithm to estimate the model uncertainties, not necessarily extremum seeking-based. To emphasize this we show here the performance of the controller when using a type of RL-based learning algorithm as well.

The rest of the paper is organized as follows. In Section II, we present some notations, and fundamental definitions that will be needed in the sequel. In Section III, we formulate the problem. The nominal controller design are presented in Section IV. In Section IV-B, a robust controller is designed which guarantees ISS from the estimation error input to the tracking error state. In Section IV-C, the ISS controller is complemented with an MES algorithm to estimate the model parametric uncertainties. In Section IV-D, we introduce the RL GP-UCB algorithm as a model-free learning to complement the ISS controller. Section V is dedicated to an application example and the paper conclusion is given in Section VI.

II. Preliminaries

Throughout the paper, we use $\|\cdot\|$ to denote the Euclidean norm; i.e., for a vector $x \in \mathbb{R}^n$, we have $\|x\| \triangleq \|x\|_2 = \sqrt{x^T x}$, where x^T denotes the transpose of the vector x. We denote by $\mathrm{Card}(S)$ the size of a finite set S. The Frobenius norm of a matrix $A \in \mathbb{R}^{m\times n}$, with elements a_{ij}, is defined as $\|A\|_F \triangleq \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}$. Given $x \in \mathbb{R}^m$, the signum function is defined as $\mathrm{sign}(x) \triangleq [\mathrm{sign}(x_1),\ \mathrm{sign}(x_2),\ \cdots,\ \mathrm{sign}(x_m)]^T$, where sign(.) denotes the classical signum function.

In this paper, we will rely on the concept of ISS for dynamical systems. Due to space restriction, we will not recall the definition and the related Lyapunov direct theorem used here.[1] However, in the following we will use the definition of ISS for time-varying systems, and the associated Lyapunov direct theorem as in [21], [22].

III. Problem Formulation

A. Nonlinear system model

We consider here affine uncertain nonlinear systems of the form

$$\begin{aligned} \dot{x} &= f(x) + \Delta f(t,x) + g(x)u, \\ y &= h(x), \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^p$, $y \in \mathbb{R}^m$ ($p \geq m$), represent the state, the input and the controlled output vectors, respectively. $\Delta f(t,x)$ is a vector field representing additive model uncertainties. The vector fields f, Δf, columns of g and function h satisfy the following assumptions.

Assumption A1 The function $f : \mathbb{R}^n \to \mathbb{R}^n$ and the columns of $g : \mathbb{R}^n \to \mathbb{R}^p$ are $\mathbb{C}^\infty$ vector fields on a bounded set X of $\mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ is a $\mathbb{C}^\infty$ vector on X. The vector field $\Delta f(x)$ is $\mathbb{C}^1$ on X.

Assumption A2 System (1) has a well-defined (vector) relative degree $\{r_1,\ r_2,\ \cdots,\ r_m\}$ at each point $x^0 \in X$, and the system is linearizable, i.e., $\sum_{i=1}^{m} r_i = n$.

[1] Please refer to a longer version posted on arXiv [20], for the complete definitions.

Assumption A3 The desired output trajectories y_{id} ($1 \leq i \leq m$) are smooth functions of time, relating desired initial points $y_{id}(0)$ at $t = 0$ to desired final points $y_{id}(t_f)$ at $t = t_f$.

B. Control objectives

Our objective is to design a state feedback adaptive controller such that the output tracking error is uniformly bounded, whereas the tracking error upper-bound is function of the uncertain parameters estimation error, which can be decreased by the model-free learning. We stress that the goal of learning algorithm is not stabilization but rather performance optimization, i.e., the learning improves the parameters estimation error, which in turn improves the output tracking error. To achieve this control objective, we proceed as follows: First, we design a robust controller which can guarantee input-to-state stability (ISS) of the tracking error dynamics w.r.t the estimation errors input. Then, we combine this controller with a model-free learning algorithm to iteratively estimate the uncertain parameters, by optimizing online a desired learning cost function.

IV. Adaptive Controller Design

A. Nominal Controller

Let us first consider the system under nominal conditions, i.e., when $\Delta f(t,x) = 0$. In this case, it is well know, e.g., [21], that system (1) can be written as

$$y^{(r)}(t) \ = \ b(\xi(t)) + A(\xi(t))u(t), \tag{2}$$

where

$$\begin{aligned} y^{(r)}(t) &= [y_1^{(r_1)}(t),\ y_2^{(r_2)}(t),\ \cdots,\ y_m^{(r_m)}(t)]^T, \\ \xi(t) &= [\xi^1(t),\ \cdots,\ \xi^m(t)]^T, \\ \xi^i(t) &= [y_i(t),\ \cdots,\ y_i^{(r_i-1)}(t)], \quad 1 \leq i \leq m \end{aligned} \tag{3}$$

The functions $b(\xi)$, $A(\xi)$ can be written as functions of f, g and h, and $A(\xi)$ is non-singular in $\tilde{X}$, where $\tilde{X}$ is the image of the set of X by the diffeomorphism $x \mapsto \xi$ between the states of system (1) and the linearized model (2). Now, to deal with the uncertain model, we first need to introduce one more assumption on system (1).

Assumption A4 The additive uncertainties $\Delta f(t,x)$ in (1) appear as additive uncertainties in the input-output linearized model (2)-(3) as follows (see also [23])

$$y^{(r)}(t) \ = \ b(\xi(t)) + A(\xi(t))u(t) + \Delta b(t,\xi(t)), \tag{4}$$

where $\Delta b(t,\xi)$ is $\mathbb{C}^1$ w.r.t. the state vector $\xi \in \tilde{X}$.

It is well known that the nominal model (2) can be easily transformed into a linear input-output mapping. Indeed, we can first define a virtual input vector $v(t)$ as

$$v(t) \ = \ b(\xi(t)) + A(\xi(t))u(t). \tag{5}$$

Combining (2) and (5), we can obtain the following input-output mapping

$$y^{(r)}(t) \ = \ v(t). \tag{6}$$

Based on the linear system (6), it is straightforward to design a stabilizing controller for the nominal system (2) as[2]

$$u_n = A^{-1}(\xi)\left[v_s(t,\xi) - b(\xi)\right], \tag{7}$$

[2] The inverse of A is to be understood in the sense of Moore-Penrose pseudo-inverse which is guaranteed to exist by the relative degree Assumption A2.

where v_s is a $m \times 1$ vector and the i-th ($1 \leq i \leq m$) element v_{si} is given by

$$v_{si} = y_{id}^{(r_i)} - K_{r_i}^i (y_i^{(r_i-1)} - y_{id}^{(r_i-1)}) - \cdots - K_1^i (y_i - y_{id}). \quad (8)$$

If we denote the tracking error as $e_i(t) \triangleq y_i(t) - y_{id}(t)$, we obtain the following tracking error dynamics

$$e_i^{(r_i)}(t) + K_{r_i}^i e^{(r_i-1)}(t) + \cdots + K_1^i e_i(t) = 0, \quad (9)$$

where $i \in \{1,\ 2,\ \cdots,\ m\}$. By properly selecting the gains K_j^i where $i \in \{1,\ 2,\ \cdots,\ m\}$ and $j \in \{1,\ 2,\ \cdots,\ r_i\}$, we can obtain global asymptotic stability of the tracking errors $e_i(t)$. To formalize this condition, we add the following assumption.

Assumption A5 There exists a non-empty set $\mathcal{A}$ where $K_j^i \in \mathcal{A}$ such that the polynomials in (9) are Hurwitz, where $i \in \{1,\ 2,\ \cdots,\ m\}$ and $j \in \{1,\ 2,\ \cdots,\ r_i\}$.

To this end, we define $z = [z^1,\ z^2,\ \cdots,\ z^m]^T$, where $z^i = [e_i,\ \dot{e}_i,\ \cdots,\ e_i^{(r_i-1)}]$ and $i \in \{1,\ 2, \cdots,\ m\}$. Then, from (9), we can obtain

$$\dot{z} \;=\; \tilde{A} z,$$

where $\tilde{A} \in \mathbb{R}^{n \times n}$ is a diagonal block matrix given by

$$\tilde{A} \;=\; \mathrm{diag}\{\tilde{A}_1,\ \tilde{A}_2,\ \cdots,\ \tilde{A}_m\}, \quad (10)$$

and $\tilde{A}_i$ ($1 \leq i \leq m$) is a $r_i \times r_i$ matrix given by

$$\tilde{A}_i \;=\; \begin{bmatrix} 0 & 1 & & & \\ 0 & & 1 & & \\ 0 & & & \ddots & \\ \vdots & & & & 1 \\ -K_1^i & -K_2^i & \cdots & \cdots & -K_{r_i}^i \end{bmatrix}.$$

As discussed above, the gains K_j^i can be chosen such that the matrix $\tilde{A}$ is Hurwitz. Thus, there exists a positive definite matrix $P > 0$ such that (see e.g. [21])

$$\tilde{A}^T P + P \tilde{A} \;=\; -I. \quad (11)$$

In the next section, we build upon the nominal controller (7) to write a robust ISS controller.

B. Lyapunov reconstruction-based ISS Controller

We now consider the uncertain model (1), i.e., when $\Delta f(t,x) \neq 0$. The corresponding exact linearized model is given by (4) where $\Delta b(t,\xi(t)) \neq 0$. The global asymptotic stability of the error dynamics (9) cannot be guaranteed anymore due to the additive uncertainty $\Delta b(t,\xi(t))$. We use Lyapunov reconstruction techniques to design a new controller so that the tracking error is guaranteed to be bounded given that the estimate error of $\Delta b(t,\xi(t))$ is bounded. The new controller for the uncertain model (4) is defined as

$$u_f \;=\; u_n + u_r, \quad (12)$$

where the nominal controller u_n is given by (7) and the robust controller u_r will be given later. By using controller (12), and (4) we obtain

$$\begin{aligned} y^{(r)}(t) \;&=\; b(\xi(t)) + A(\xi(t))u_f + \Delta b(t,\xi(t)), \\ &=\; b(\xi(t)) + A(\xi(t))u_n + A(\xi(t))u_r + \Delta b(t,\xi(t)), \\ &=\; v_s(t,\xi) + A(\xi(t))u_r + \Delta b(t,\xi(t)), \end{aligned} \quad (13)$$

where (13) holds from (7). This leads to the following error dynamics

$$\dot{z} \;=\; \tilde{A} z + \tilde{B}\delta, \quad (14)$$

where $\tilde{A}$ is defined in (10), δ is a $m \times 1$ vector given by

$$\delta \;=\; A(\xi(t))u_r + \Delta b(t,\xi(t)), \quad (15)$$

and the matrix $\tilde{B} \in \mathbb{R}^{n \times m}$ is given by

$$\tilde{B} \;=\; \left[\ \tilde{B}_1^T, \tilde{B}_2^T, \ldots, \tilde{B}_m^T\ \right]^T, \quad (16)$$

where each $\tilde{B}_i$ ($1 \leq i \leq m$) is given by a $r_i \times m$ matrix such that

$$\tilde{B}_i(l,q) \;=\; \begin{cases} 1 & \text{for } l = r_i,\ q = i \\ 0 & \text{otherwise.} \end{cases}$$

If we choose $V(z) = z^T P z$ as a Lyapunov function for the dynamics (14), where P is the solution of the Lyapunov equation (11), we obtain

$$\begin{aligned} \dot{V}(t) \;&=\; \frac{\partial V}{\partial z}\dot{z}, \\ &=\; z^T(\tilde{A}^T P + P\tilde{A})z + 2z^T P\tilde{B}\delta, \\ &=\; -\|z\|^2 + 2z^T P\tilde{B}\delta, \end{aligned} \quad (17)$$

where δ given by (15) depends on the robust controller u_r.

Next, we design the controller u_r based on the form of the uncertainties $\Delta b(t,\xi(t))$. More specifically, we consider the case when $\Delta b(t,\xi(t))$ is of the following form

$$\Delta b(t,\xi(t)) \;=\; E\, Q(\xi,t), \quad (18)$$

where $E \in \mathbb{R}^{m \times m}$ is a matrix of unknown constant parameters, and $Q(\xi,t):\ \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^m$ is a known bounded function of states and time variables. For notational convenience, we denote by $\hat{E}(t)$ the estimate of E, and by $e_E = E - \hat{E}$, the estimate error. We define the unknown parameter vector $\Delta = [E(1,1), ..., E(m,m)]^T \in \mathbb{R}^{m^2}$, i.e., concatenation of all elements of E, its estimate is denoted by $\widehat{\Delta}(t) = [\hat{E}(1,1), ..., \hat{E}(m,m)]^T$, and the estimation error vector is given by $e_\Delta(t) = \Delta - \widehat{\Delta}(t)$.

Next, we propose the following robust controller

$$u_r \;=\; -A^{-1}(\xi)[\tilde{B}^T P z \|Q(\xi,t)\|^2 + \widehat{E}(t)Q(\xi,t)]. \quad (19)$$

The closed-loop error dynamics can be written as

$$\dot{z} \;=\; \tilde{f}(t,z,e_\Delta), \quad (20)$$

where $e_\Delta(t)$ is considered to be an input to the system (20).

Theorem 1: Consider the system (1), under Assumptions A1-A5, where $\Delta b(t,\xi(t))$ satisfies (18). If we apply to (1) the feedback controller (12), where u_n is given by (7) and u_r is given by (19), then the closed-loop system (20) is ISS from the estimation errors input $e_\Delta(t) \in \mathbb{R}^{m^2}$ to the tracking errors state $z(t) \in \mathbb{R}^n$.

Proof: Please refer to a longer version posted on arXiv [20] for the complete proofs, which will also be included in a future longer journal version of this work.

C. MES-based parametric uncertainties estimation

Let us define now the following cost function

$$J(\widehat{\Delta}) \;=\; F(z(\widehat{\Delta})), \quad (21)$$

where $F : \mathbb{R}^n \rightarrow \mathbb{R}$, $F(\mathbf{0}) = 0$, $F(z) > 0$ for $z \in \mathbb{R}^n - \{\mathbf{0}\}$. We need the following assumptions on J.

Assumption A6 The cost function J has a local minimum at $\widehat{\Delta}^* = \Delta$.

Assumption A7 The initial error $e_\Delta(t_0)$ is sufficiently small, i.e., the original parameter estimate vector $\widehat{\Delta}$ are close enough to the actual parameter vector Δ.

Assumption A8 The cost function J is analytic and its variation with respect to the uncertain parameters is bounded in the neighborhood of $\widehat{\Delta}^*$, i.e., $\|\frac{\partial J}{\partial \widehat{\Delta}}(\tilde{\Delta})\| \leq \xi_2$, $\xi_2 > 0$, $\tilde{\Delta} \in \mathcal{V}(\widehat{\Delta}^*)$, where $\mathcal{V}(\widehat{\Delta}^*)$ denotes a compact neighborhood of $\widehat{\Delta}^*$.

We can now present the following result.

Lemma 2: Consider the system (1), under Assumptions A1-A8, where the uncertainty is given by (18). If we apply to (1) the feedback controller (12), where u_n is given by (7), u_r is given by (19), the cost function is given by (21), and $\widehat{\Delta}(t)$ are estimated through the ES algorithm

$$\begin{aligned}\dot{\tilde{x}}_i &= a_i \sin(\omega_i t + \frac{\pi}{2}) J(\widehat{\Delta}),\ a_i > 0, \\ \widehat{\Delta}_i(t) &= \tilde{x}_i + a_i \sin(\omega_i t - \frac{\pi}{2}), \quad i \in \{1, 2, \dots, m^2\} \end{aligned} \tag{22}$$

with $\omega_i \neq \omega_j$, $\omega_i + \omega_j \neq \omega_k$, $i, j, k \in \{1, 2, \dots, m^2\}$, and $\omega_i > \omega^*$, $\forall\ i \in \{1, 2, \dots, m^2\}$, with ω^* large enough. Then, the norm of the error vector $z(t)$ admits the following bound

$$\|z(t)\| \leq \beta(\|z(0)\|, t) + \gamma(\tilde{\beta}(\|e_\Delta(0)\|, t) + \|e_\Delta\|_{\max}),$$

where $\|e_\Delta\|_{\max} = \frac{\xi_1}{\omega_0} + \sqrt{\sum_{i=1}^{m^2} a_i^2}$, $\xi_1 > 0$, $\omega_0 = \max_{i\in\{1,2,\dots,m^2\}} \omega_i$, $\beta \in \mathcal{KL}$, $\tilde{\beta} \in \mathcal{KL}$ and $\gamma \in \mathcal{K}$.

Proof: Please refer to a longer version posted on arXiv [20] for the complete proofs, which will also be included in a future longer journal version of this work.

As we mentioned earlier, the dither-based MES has the problem of local minima. To address this point in the next section we propose to use GP-UCB as the model-free learning algorithm for model uncertainties estimation.

D. GP-UCB based parametric uncertainties estimation

In this section we propose to use Gaussian Process Upper Confidence Bound (GP-UCB) algorithm to find the uncertain parameter vector Δ [13], [24]. GP-UCB is a Bayesian optimization algorithm for stochastic optimization, i.e., the task of finding the global optimum of an unknown function when the evaluations are potentially contaminated with noise. The underlying working assumption for Bayesian optimization algorithms, including GP-UCB, is that the function evaluation is costly, so we would like to minimize the number of evaluations while having as accurate estimate of the minimizer (or maximizer) as possible [25]. For GP-UCB, this goal is guaranteed by having an upper bound on the regret of the algorithm – to be defined precisely later.

One difficulty of stochastic optimization is that since we only observe noisy samples from the function, we cannot really be sure about the exact value of the function at any given point. One may try to query a single point many times in order to have an accurate estimate of the function. This, however, may lead to excessive number of samples, and can be wasteful way of assigning samples when the true value of the function at that point is actually far from optimal. The Upper Confidence Bound (UCB) family of algorithms provides a principled approach to guide the search [26]. These algorithms, which are not necessarily formulated in a Bayesian framework, automatically balance the exploration (i.e., finding the regions of the parameter space that *might* be promising) and the exploration (i.e., focusing on the regions that are known to be the best based on the *current* available knowledge) using the principle of optimism in the face of uncertainty. These algorithms often come with strong theoretical guarantee about their performance. For more information about the UCB class of algorithms, refer to [27], [28], [29]. GP-UCB is a particular UCB algorithms that is suitable to deal with continuous domains. It uses a Gaussian Process (GP) to maintain the mean and confidence information about the unknown function.

We briefly discuss GP-UCB in our context following the discussion of the original papers [13], [24]. Consider the cost function $J : D \to \mathbb{R}$ to be minimized. This function depends on the dynamics of the closed-loop system, which itself depends on the parameters $\widehat{\Delta}$ used in the controller design. So we may consider it as an unknown function of $\widehat{\Delta}$, so $D \subset \mathbb{R}^{m^2}$

For the moment, let us assume that J is a function sampled from a Gaussian Process (GP) [30]. Recall that a GP is a stochastic process indexed by the set D that has the property that for any finite subset of the evaluation points, that is $\{\widehat{\Delta}_1, \widehat{\Delta}_2, \dots, \widehat{\Delta}_t\} \subset D$, the joint distribution of $\left(J(\widehat{\Delta}_i)\right)_{i=1}^{t}$ is a multivariate Gaussian distribution. GP is defined by a mean function $\mu(\widehat{\Delta}) = \mathbb{E}\left[J(\widehat{\Delta})\right]$ and its covariance function (or kernel) $\kappa(\widehat{\Delta}, \widehat{\Delta}') = \mathrm{Cov}(J(\widehat{\Delta}), J(\widehat{\Delta}')) = \mathbb{E}\left[\left(J(\widehat{\Delta}) - \mu(\widehat{\Delta})\right)\left(J(\widehat{\Delta}') - \mu(\widehat{\Delta}')\right)^\top\right]$. The kernel κ of a GP determines the behavior of a typical function sampled from the GP. For instance, if we choose $\kappa(\widehat{\Delta}, \widehat{\Delta}') = \exp\left(-\frac{\|\widehat{\Delta} - \widehat{\Delta}'\|^2}{2l^2}\right)$, the squared exponential kernel with length scale $l > 0$, it implies that the GP is mean square differentiable of all orders.

Let us first briefly describe how we can find the posterior distribution of a GP$(0, \kappa)$, a GP with zero prior mean. Suppose that for $\underline{\widehat{\Delta}}_{t-1} \triangleq \{\widehat{\Delta}_1, \widehat{\Delta}_2, \dots, \widehat{\Delta}_{t-1}\} \subset D$, we have observed the noisy evaluation $y_i = J(\widehat{\Delta}_i) + \eta_i$ with $\eta_i \sim N(0, \sigma^2)$ being i.i.d. Gaussian noise. We can find the posterior mean and variance for a new point $\widehat{\Delta}^* \in D$ as follows: Denote the vector of observed values by $\mathbf{y}_{t-1} = [y_1, \dots, y_{t-1}]^\top \in \mathbb{R}^{t-1}$, and define the Grammian matrix $K \in \mathbb{R}^{t-1\times t-1}$ with $[K]_{i,j} = \kappa(\widehat{\Delta}_i, \widehat{\Delta}_j)$, and the vector $\kappa_* = [\kappa(\underline{\widehat{\Delta}}_1, \widehat{\Delta}^*), \dots, \kappa(\underline{\widehat{\Delta}}_{t-1}, \widehat{\Delta}^*)]$. The expected mean $\mu_t(\widehat{\Delta}^*)$ and the variance $\sigma_t(\widehat{\Delta}^*)$ of the posterior of the GP evaluated at $\widehat{\Delta}^*$ are (cf. Section 2.2 of [30])

$$\begin{aligned}\mu_t(\widehat{\Delta}^*) &= \kappa_* \left[K + \sigma^2 \mathbf{I}\right]^{-1} \mathbf{y}_{t-1}, \\ \sigma_t^2(\widehat{\Delta}^*) &= \kappa(\widehat{\Delta}^*, \widehat{\Delta}^*) - \kappa_*^\top \left[K + \sigma^2 \mathbf{I}\right]^{-1} \kappa_*.\end{aligned}$$

At round t, the GP-UCB algorithm selects the next query point $\widehat{\Delta}_t$ by solving the following optimization problem:

$$\widehat{\Delta}_t \leftarrow \operatorname*{argmin}_{\widehat{\Delta} \in D} \mu_{t-1}(\widehat{\Delta}) - \beta_t^{1/2} \sigma_{t-1}(\widehat{\Delta}). \tag{23}$$

Where β_t depends on the choice of kernel among other parameters of the problem.

The optimization problem (23) is often nonlinear and non-convex. Nonetheless solving it only requires querying the GP, which in general is much faster than querying the original dynamical system. This is important when the dynamical system is a physical system and we would like to minimize the number of interactions with it before finding a $\widehat{\Delta}$ with small $J(\widehat{\Delta})$. One practically easy way to approximately solve (23) is to restrict the search to a finite subset D' of D. The finite subset can be a uniform grid structure over D, or it might consist of randomly selected members of D.

The theoretical guarantee for GP-UCB is in the form of regret upper bound. Let us define $\Delta^* \leftarrow \text{argmin}_{\Delta \in D} J(\Delta)$, the global minimizer of the objective function. The regret at time t is defined by $r_t = J(\widehat{\Delta}_t) - J(\Delta^*)$. This is a measure of sub-optimality of the choice of $\widehat{\Delta}_t$ according the cost function J. The cumulative regret at time T is defined as $R_T = \sum_{t=1}^{T} r_t$. Ideally we would like $\lim_{t\to\infty} \frac{R_T}{T} = 0$.

The behavior of the cumulative regret R_T depends on the set D and the choice of kernel. If D is a compact and convex set of $\mathbb{R}^n$ and we use the squared exponential kernel, for any fixed confidence parameter $\delta > 0$, the asymptotic behavior of R_T is

$$O\left(\sqrt{T[\log^{d+1}(T) + \log(1/\delta)]}\right),$$

with probability at least $1 - \delta$ (cf. Theorems 3 and 5 of [13]). This result does not even require the function J to be a GP. It only requires the function to have a finite norm in the reproducing kernel Hilbert space (RKHS) $\mathcal{H}_\kappa$ defined by the kernel κ.

V. Two-link Manipulator Example

We consider here a two-link robot manipulator, with the following dynamics

$$H(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) \;=\; \tau, \tag{24}$$

where $q \triangleq [q_1, q_2]^T$ denotes the two joint angles and $\tau \triangleq [\tau_1, \tau_2]^T$ denotes the two joint torques. The matrix $H \in \mathbb{R}^{4\times 4}$ is assumed to be non-singular and its elements are given by

$$\begin{aligned} H_{11} &= m_1\ell_{c_1}^2 + I_1 + m_2[\ell_1^2 + \ell_{c_2}^2 + 2\ell_1\ell_{c_2}\cos(q_2)] + I_2,\\ H_{12} &= m_2\ell_1\ell_{c_2}\cos(q_2) + m_2\ell_{c_2}^2 + I_2,\\ H_{21} &= H_{12},\\ H_{22} &= m_2\ell_{c_2}^2 + I_2. \end{aligned} \tag{25}$$

The matrix $C(q, \dot{q})$ is given by

$$C(q,\dot{q}) \;\triangleq\; \begin{bmatrix} -h\dot{q}_2 & -h\dot{q}_1 - h\dot{q}_2 \\ h\dot{q}_1 & 0 \end{bmatrix},$$

where $h = m_2\ell_1\ell_{c_2}\sin(q_2)$. The vector $G = [G_1, G_2]^T$ is given by

$$\begin{aligned} G_1 &= m_1\ell_{c_1} g\cos(q_1) + m_2 g[\ell_2\cos(q_1+q_2) + \ell_1\cos(q_1)],\\ G_2 &= m_2\ell_{c_2} g\cos(q_1+q_2), \end{aligned} \tag{26}$$

where, $\ell_1,\ \ell_2$ are the lengths of the first and second link, respectively, $\ell_{c_1},\ \ell_{c_2}$ are the distances between the rotation center and the center of mass of the first and second link respectively. $m_1,\ m_2$ are the masses of the first and second link, respectively, I_1 is the moment of inertia of the first link and I_2 the moment of inertia of the second link, respectively, and g denotes the earth gravitational constant.

In our simulations, we assume that the parameters take the following values: $I_2 = \frac{5.5}{12}\ kg\cdot m^2$, $m_1 = 10.5\ kg$, $m_2 = 5.5\ kg$, $\ell_1 = 1.1\ m$, $\ell_2 = 1.1\ m$, $\ell_{c_1} = 0.5\ m$, $\ell_{c_2} = 0.5\ m$, $I_1 = \frac{11}{12}\ kg\cdot m^2$, $g = 9.8\ m/s^2$. The system dynamics (24) can be rewritten as

$$\ddot{q} \;=\; H^{-1}(q)\tau - H^{-1}(q)\left[C(q,\dot{q})\dot{q} + G(q)\right]. \tag{27}$$

Thus, the nominal controller is given by

$$\begin{aligned} \tau_n \;=&\; [C(q,\dot{q})\dot{q} + G(q)]\\ &+\; H(q)\left[\ddot{q}_d - K_d(\dot{q} - \dot{q}_d) - K_p(q - q_d)\right], \end{aligned} \tag{28}$$

where $q_d = [q_{1d}, q_{2d}]^T$, denotes the desired trajectory and the diagonal gain matrices $K_p > 0$, $K_d > 0$, are chosen such that the linear error dynamics (as in (9)) are asymptotically stable. We choose as output references the $5th$ order polynomials $q_{1ref}(t) = q_{2ref}(t) = \sum_{i=0}^{5} a_i(t/t_f)^i$, where the a_i's have been computed to satisfy the boundary constraints $q_{iref}(0) = 0, q_{iref}(t_f) = q_f, \dot{q}_{iref}(0) = \dot{q}_{iref}(t_f) = 0, \ddot{q}_{iref}(0) = \ddot{q}_{iref}(t_f) = 0,\ i = 1, 2$, with $t_f = 2\ sec$, $q_f = 1.5\ rad$. In these tests, we assume that the nonlinear model (24) is uncertain. In particular, we assume that there exist additive uncertainties in the model (27), i.e.,

$$\ddot{q} = \; H^{-1}(q)\tau - H^{-1}(q)\left[C(q,\dot{q})\dot{q} + G(q)\right] - E\,G(q), \tag{29}$$

where E is a matrix of constant uncertain parameters. Following (19), the robust-part of the control writes as

$$\tau_r = -H(\tilde{B}^T P z\|G\|^2 - \hat{E}\,G(q)), \tag{30}$$

where

$$\tilde{B}^T = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

P is solution of the Lyapunov equation (11), with

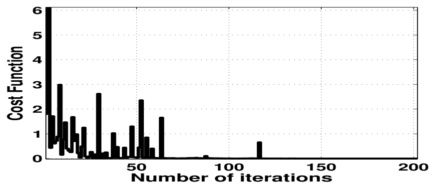
$$\tilde{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -K_p^1 & -K_d^1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -K_p^2 & -K_d^2 \end{bmatrix},$$

$z = [q_1 - q_{1d}, \dot{q}_1 - \dot{q}_{1d}, q_2 - q_{2d}, \dot{q}_2 - \dot{q}_{2d}]^T$, and $\hat{E}$ is the matrix of the parameters' estimates. Eventually, the final feedback controller writes as

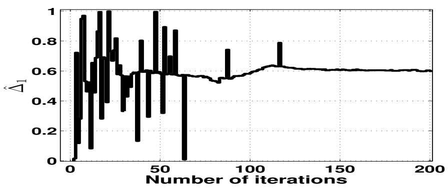
$$\tau = \tau_n + \tau_r. \tag{31}$$

We consider the challenging case where the uncertain parameters are linearly dependent. In this case the uncertainties' 'effect' is not observable from the measured output. Indeed, in the case where the uncertainties enter the model in a linearly dependent function, e.g. when the matrix Δ has only one non-zero line, some of the classical available modular model-based adaptive controllers, like for instance X-swapping controllers, cannot be used to estimate all the uncertain parameters simultaneously. For example, it has been shown in [7], that the model-based gradient descent filters failed to estimate simultaneously multiple parameters in the case of the electromagnetic actuators example. For instance, in comparison with the ES-based indirect adaptive controller of [15], the modular approach does not rely on the parameters mutual exhaustive assumption, i.e., each element of the control vector needs to be linearly dependent on at least one element of the uncertainties vector. More specifically, we consider here the following case: $\Delta(1,1) = 0.3,\ \Delta(1,2) = 0.6$, and $\Delta(2,i) = 0,\ i = 1,2$. In this case, the uncertainties' effect on the acceleration $\ddot{q}_1$ cannot be differentiated, and thus the application of the model-based X-swapping method to estimate the actual values of both uncertainties at the same time is challenging. Similarly, the method of [15], cannot be readily applied because the second control τ_2 is not linearly depend on the uncertainties, which only affects τ_1. However, we show next that, by using the modular ISS-based controller, we manage to estimate the actual values of the uncertainties simultaneously and improve the tracking performance. Due to the limitation of number of pages, we only report the results related to the GP-UCB. For the MES results please refer to [20].

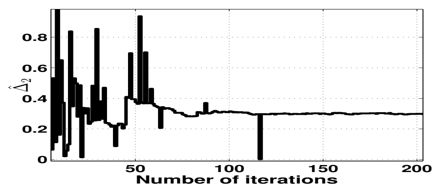
To show that the modular ISS-based controller is independent of the choice of the learning algorithm, we apply the GP-UCB learning algorithm-based estimator to the same two-links manipulator example. We apply the algorithm IV-D, with the following parameters: $\sigma = 0.1$, $l = 0.2$, and $\beta_t = 2\log(\frac{\text{Card}(D')t^2\pi^2}{6\delta})$, with $\delta = 0.05$. We test the GP-UCB algorithm under the uncertainties conditions



(a) Cost function over the learning iterations (GP-UCB)



(b) Estimate of Δ_1 over the learning iterations (GP-UCB)



(c) Estimate of Δ_2 over the learning iterations (GP-UCB)

Fig. 1. Cost function and uncertainties estimates- (GP-UCB) algorithm

stated above. The obtained parameters are reported on figures 1(a), 1(b), 1(c). We can see on these figures that the uncertainties are well estimated. One could argue that they are better estimated with the GP-UCB than with MES algorithm (see [20]) because there is no permanent dither signal, which leads to permanent oscillations in the MES-based learning. The tracking performance improved in this as well due to the precise estimation of the parameters.

VI. Conclusion

We have studied the problem of adaptive control for nonlinear systems which are affine in the control with parametric uncertainties. For this class of systems, we have proposed the following controller: We use a modular approach, where we first design a robust nonlinear controller, designed based on the model (assuming knowledge of the uncertain parameters), and then complement this controller with an estimation module to estimate the actual values of the uncertain parameters. The novelty is that the estimation module that we propose is based on model-free learning algorithms.

References

[1] P. Ioannou and J. Sun, *Robust Adaptive Control*. Dover Publications, 2012.

[2] M. Krstic, I. Kanellakopoulos, and P. Kokotovic, *Nonlinear and Adpative Control Design*. New York: Wiley, 1995.

[3] C. Wang, D. J. Hill, S. S. Ge, and G. Chen, "An ISS-modular approach for adaptive neural control of pure-feedback systems," *Automatica*, vol. 42, no. 5, pp. 723–731, 2006.

[4] E. D. Sontag, "Smooth stabilization implies coprime factorization," *Automatic Control, IEEE Transactions on*, vol. 34, no. 4, pp. 435–443, 1989.

[5] Z.-P. Jiang, A. R. Teel, and L. Praly, "Small-gain theorem for ISS systems and applications," *Mathematics of Control, Signals and Systems*, vol. 7, no. 2, pp. 95–120, 1994.

[6] M. Benosman and G. Atinc, "Nonlinear learning-based adaptive control for electromagnetic actuators," in *European Control Conference*, 2013, pp. 2904–2909.

[7] ——, "Non-linear adaptive control for electromagnetic actuators," *IET Control Theory and Applications*, vol. 9, no. 2, pp. 258–269, 2015.

[8] K. B. Ariyur and M. Krstic, *Real Time Optimization by Extremum Seeking Control*. New York, NY, USA: John Wiley & Sons, Inc., 2003.

[9] M. Rotea, "Analysis of multivariable extremum seeking algorithms," in *American Control Conference, 2000. Proceedings of the 2000*, vol. 1, no. 6, Sep 2000, pp. 433–437.

[10] M. Benosman, "Learning-based adaptive control for nonlinear systems," in *IEEE European Control Conference*, 2014, pp. 920–925.

[11] ——, "Extremum-seeking based adaptive control for nonlinear systems," in *IFAC World Congress*, 2014, pp. 401–406.

[12] M. Xia and M. Benosman, "Extremum seeking-based indirect adaptive control for nonlinear systems with time-varying uncertainties," in *European Control Conference*, 2015, pp. 2780–2785.

[13] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 2010, pp. 1015–1022.

[14] P. Haghi and K. Ariyur, "On the extremum seeking of model reference adaptive control in higher-dimensional systems," in *American Control Conference*, 2011, pp. 1176–1181.

[15] ——, "Adpative feedback linearization of nonlinear MIMO systems using ES-MRAC," in *American Control Conference*, 2013, pp. 1828–1833.

[16] M. Benosman and G. Atinc, "Model-based learning control," Mitsubishi Electric Research Laboratories (MERL), Patent 20130304236, May 2012.

[17] ——, "Multi-parametric extremum seeking-based learning control for electromagnetic actuators," in *American Control Conference*, 2013, pp. 1914–1919.

[18] G. Atinc and M. Benosman, "Nonlinear learning-based adaptive control for electromagnetic actuators with proof of stability," in *IEEE, Conference on Decision and Control*, 2013, pp. 1277–1282.

[19] M. Benosman, *Control Theory: Perspectives, Applications and Developments*, 2015, ch. Extremum Seeking-based Indirect Adaptive Control and Feedback Gains Auto-Tuning for Nonlinear Systems.

[20] M. Benosman, A.-M. Farahmand, and M. Xia, "Learning-based modular indirect adaptive control for a class of nonlinear systems," *arXiv:1509.07860*, 2015.

[21] H. Khalil, *Nonlinear Systems*, 3rd ed. Prentice Hall, 2002.

[22] M. Malisoff and F. Mazenc, "Further remarks on strict input-to-state stable lyapunov functions for time-varying systems," *Automatica*, vol. 41, no. 11, pp. 1973 – 1978, 2005.

[23] M. Benosman, F. Liao, K.-Y. Lum, and J. L. Wang, "Nonlinear control allocation for non-minimum phase systems," *Control Systems Technology, IEEE Transactions on*, vol. 17, no. 2, pp. 394–404, March 2009.

[24] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Information-theoretic regret bounds for gaussian process optimization in the bandit setting," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.

[25] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv:1012.2599*, 2010.

[26] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2-3, pp. 235–256, 2002.

[27] S. Bubeck, R. Munos, G. Stoltz, and Cs. Szepesvári, "X-armed bandits," *Journal of Machine Learning Research (JMLR)*, vol. 12, no. 1655–1695, 2011.

[28] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[29] R. Munos, "From bandits to Monte-Carlo Tree Search: The optimistic principle applied to optimization and planning," *Foundations and Trends in Machine Learning*, vol. 7(1), pp. 1–130, 2014.

[30] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.