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

Bayesian Optimization-based Modular Indirect Adaptive Control for a Class of Nonlinear Systems

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

TR2016-061 June 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 adaptive approach, where we first design a robust nonlinear state feedback which renders the closed loop input-to-state stable (ISS). 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. We augment this robust ISS controller with a model-free learning algorithm to estimate the model uncertainties. We implement this method with 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 test the efficiency of this approach on a two-link robot manipulator example, under noisy measurements conditions.

IFAC International Workshop on Adaptation and Learning in Control and Signal Processing

Bayesian Optimization-based Modular Indirect Adaptive Control for a Class of Nonlinear Systems

Mouhacine Benosman *

* *Mouhacine Benosman (m_benosman@ieee.org) is with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA.*

Amir-massoud Farahmand **

** *Amir-massoud Farahmand (farahmand@merl.com) is with MERL.*

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 adaptive approach*, where we first design a robust nonlinear state feedback which renders the closed loop input-to-state stable (ISS). 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. We augment this robust ISS controller with a model-free learning algorithm to estimate the model uncertainties. We implement this method with 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 test the efficiency of this approach on a two-link robot manipulator example, under noisy measurements conditions.

1. INTRODUCTION

Many adaptive methods have been proposed over the years for linear and nonlinear systems, e.g., Krstic et al. [1995]. In this work we focus on a specific type of adaptive control, namely, the indirect modular approach to adaptive nonlinear control, e.g., Krstic et al. [1995], Wang et al. [2006], Benosman and Atinc [2013], Atinc and Benosman [2013], Benosman [2014b,a], Xia and Benosman [2015], Lavretsky [2009], Haghi and Ariyur [2011]. In the direct approach, first a controller is designed by assuming that all the parameters are known (certainly equivalence principle), and then an identifier is used to estimate the unknown parameters online. The identifier might be independent of the designed controller, in which case the approach is called 'modular'. A modular approach has been proposed in Wang et al. [2006] 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, e.g., Sontag [1989].

In this work, we present a modular adaptive design which combines model-free learning methods and robust model-based nonlinear control to propose a learning-based modular indirect adaptive controller. Here, a model-free learning algorithm is 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 parameters estimation filters. Indeed, model-based indirect adaptive controllers are based on parameters' estimators designed using the model of the system, e.g., the X-swapping methods presented in Krstic et al. [1995]. Here, because we do not use the system dynamics to design the estimation filters we can deal with a more general class of uncertainties, e.g., nonlinear uncertainties can be estimated with the proposed approach, see Atinc and Benosman [2013] for some preliminary results. Furthermore, with the proposed approach we can estimate a vector of linearly dependent uncertainties, a case which cannot be solved using model-based filters, e.g., in Benosman and Atinc [2015] it is shown that the X-swapping model-based method fails to estimate a vector of linearly dependent parameters.

In this work, we implement the proposed approach with 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.

We want to underline here that compared to 'pure' model-free controllers, e.g., pure RL algorithms, the proposed control has a different goal. The available model-free controllers are meant for output or state regulation. In the contrary, here we propose to use model-free learning to complement a model-based nonlinear control to estimate the unknown parameters of the model. Here 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. Once the learning algorithm has converged, one can carry on using the nonlinear model-based feedback controller alone, without the need of the learning algorithm. Moreover, we believe that this type of controller can converge faster to an optimal performance, comparatively to the pure model-free controller. The reason is that the model-free algorithms assume no knowledge about the system, and thus start the search for an optimal control signal from scratch. On the other hand, by 'partly' using a model-based controller we are taking advantage of the partial information given by the physics of the system.

A modular design merging model-based control and an extremum seeker has been proposed in Haghi and Ariyur [2011, 2013], Benosman and Atinc [2012, 2013], Atinc and Benosman [2013], Benosman [2014b,a], Xia and Benosman

[2015]. In Haghi and Ariyur [2011, 2013], extremum seeking is used to complement a model-based controller, under linearity of the model assumption in Haghi and Ariyur [2011], or under the assumption of linear parametrization of the control in terms of the uncertainties in Haghi and Ariyur [2013]. The modular design idea of using a model-based controller with ISS guarantee, complemented with an ES-based module can be found in Atinc and Benosman [2013], Benosman [2014b,a], Xia and Benosman [2015], where the ES was used to estimate the model parameters, and in Benosman and Atinc [2013], Benosman [2015] where feedback gains were tuned using ES algorithms. The work of this paper falls in this class of ISS-based modular indirect adaptive controllers. The difference with other ES-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.

The rest of the paper is organized as follows. In Section 2, we formulate the problem. The nominal controller design are presented in Section 3. In Section 3.2, a robust controller is designed which guarantees ISS from the estimation error input to the tracking error state. In section 3.3, we introduce the RL GP-UCB algorithm as a model-free learning to complement the ISS controller. Section 4 is dedicated to an application example and the paper conclusion is given in Section 5.

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 $\mathrm{sign}(.)$ denotes the classical signum function.

2. PROBLEM FORMULATION

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., $\sum_{i=1}^m r_i = n$.

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

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 here 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.

3. ADAPTIVE CONTROLLER DESIGN

3.1 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., Khalil [2002], 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 $x \to \xi$ 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:

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

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

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

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

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

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

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

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

$$\tilde{A} \ = \ \mathrm{diag}\{\tilde{A}_1,\ \tilde{A}_2,\ \cdots,\ \tilde{A}_m\}, \tag{10}$$

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. Khalil [2002])

$$\tilde{A}^T P + P\tilde{A} \ = \ -I. \tag{11}$$

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

3.2 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

$$u_f \ = \ u_n + u_r, \tag{12}$$

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

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

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

$$\dot{z} \ = \ \tilde{A}z + \tilde{B}\delta, \tag{14}$$

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

$$\delta \ = \ A(\xi(t))u_r + \Delta b(t, \xi(t)), \tag{15}$$

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

$$\tilde{B} \ = \ \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ \vdots \\ \tilde{B}_m \end{bmatrix}, \tag{16}$$

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

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

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 here the case when $\Delta b(t, \xi(t))$ is of the following form

$$\Delta b(t, \xi(t)) \ = \ E\, Q(\xi, t), \tag{18}$$

where $E \in \mathbb{R}^{m\times m}$ is a matrix of unknown constant parameters, and $Q(\xi, t) : \ \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$ is a known bounded function of sate 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

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

The closed-loop error dynamics can be written as

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

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 Xia and Benosman [2015] for a similar proof.

3.3 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 Δ vector Srinivas et al. [2010], Srinivas et al. [2012]. 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 Brochu et al. [2010]. 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 Auer et al. [2002]. These algorithms, which are not necessarily formulated in a Bayesian framework, automatically balance the exploration (i.e., finding 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 Bubeck et al. [2011], Bubeck and Cesa-Bianchi [2012], Munos [2014]. 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 Srinivas et al. [2010, 2012]. 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 $\widehat{\Delta}$.

For the moment, let us assume that J is a function sampled from a Gaussian Process (GP) Rasmussen and Williams [2006]. 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) $\kappa(\widehat{\Delta}, \widehat{\Delta}') = \mathrm{Cov}(J(\widehat{\Delta}), J(\widehat{\Delta}')) = \mathbb{E}\left[\left(J(\widehat{\Delta}) - \mu(\widehat{\Delta})\right)\left(J(\widehat{\Delta}') - \mu(\widehat{\Delta}')\right)^T\right]$. 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. We write $J \sim \mathrm{GP}(\mu, \kappa)$.

Let us first briefly describe how we can find the posterior distribution of a $\mathrm{GP}(0, \kappa)$, 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 Rasmussen and Williams [2006])

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

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

$$\widehat{\Delta}_t \leftarrow \operatorname*{argmin}_{\widehat{\Delta}\in D} \mu_{t-1}(\widehat{\Delta}) - \beta_t^{1/2} \sigma_{t-1}(\widehat{\Delta}). \tag{21}$$

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

The optimization problem (21) 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 (21) 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.

Remark 1. It is well know in classical adaptive control that a persistent excitation (PE) condition is needed for the controller to recover the true values of the estimated parameters. In the context of the work presented here, the PE condition is satisfied by the exploration part of the GP-UCB algorithm. Indeed, the GP-UCB estimator is based on regular (possibly random) selection of exploration areas within the search set, formulated in the choice of the function β_t. This function satisfies a PE condition.

4. TWO-LINK MANIPULATOR EXAMPLE

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

$$H(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) \;=\; \tau, \tag{22}$$

where $q \triangleq [q_1, q_2]^T$ denotes the joint angles, and $\tau \triangleq [\tau_1, \tau_2]^T$ denotes the joint torques. The symmetric matrix $H \in \mathbb{R}^{4\times 4}$ is assumed to be non-singular and its elements are given by: $H_{11} = m_1\ell_{c_1}^2 + I_1 + m_2[\ell_1^2 + \ell_{c_2}^2 + 2\ell_1\ell_{c_2}\cos(q_2)] + I_2$, $H_{12} = m_2\ell_1\ell_{c_2}\cos(q_2) + m_2\ell_{c_2}^2 + I_2$, $H_{22} = m_2\ell_{c_2}^2 + I_2$. The matrix $C(q, \dot{q})$ is given by: $C_{11} = -h\dot{q}_2$, $C_{12} = -h\dot{q}_1 - h\dot{q}_2$, $C_{21} = h\dot{q}_1$, $C_{44} = 0$, where $h = m_2\ell_1\ell_{c_2}\sin(q_2)$. The vector $G = [G_1, G_2]^T$ is given by: $G_1 = m_1\ell_{c1} g\cos(q_1) + m_2 g[\ell_2\cos(q_1 + q_2) + \ell_1\cos(q_1)]$, $G_2 = m_2\ell_{c_2} g\cos(q_1 + q_2)$, where, ℓ_1, ℓ_2 are the lengths of the first and second link, respectively, ℓ_{c_1}, ℓ_{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.

Here, 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 (22) can be rewritten as

$$\ddot{q} \;=\; H^{-1}(q)\tau - H^{-1}(q)\left[C(q, \dot{q})\dot{q} + G(q)\right]. \tag{23}$$

The nominal controller is given by

$$\begin{aligned}\tau_n \;=\;& [C(q, \dot{q})\dot{q} + G(q)] \\ &+\; H(q)\left[\ddot{q}_d - K_d(\dot{q} - \dot{q}_d) - K_p(q - q_d)\right],\end{aligned} \tag{24}$$

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 (22) is uncertain. In particular, we assume that there exist additive uncertainties in the model (23), i.e.,

$$\ddot{q} = \ H^{-1}(q)\tau - H^{-1}(q)\left[C(q,\dot{q})\dot{q} + G(q)\right] - E\ G(q). \tag{25}$$

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

$$\tau_r = -H(\tilde{B}^T P z \|G\|^2 - \hat{E}\ G(q)), \tag{26}$$

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

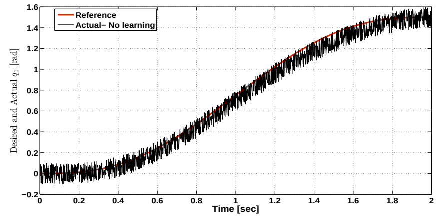
$$\tilde{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -K_p^1 & -K_d^1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -K_p^2 & -K_d^2 \end{bmatrix},$$

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

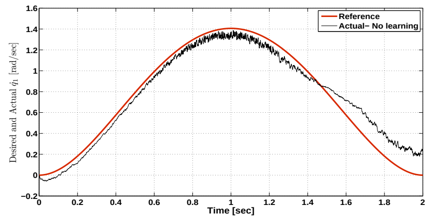
$$\tau = \tau_n + \tau_r. \tag{27}$$

We consider here the challenging case where the uncertain parameters are linearly dependent, and where the measurements are corrupted with non-negligible additive noise. Due to the linearity-dependence, the uncertainties might 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 Benosman and Atinc [2015], that the model-based gradient descent filters failed to estimate simultaneously multiple parameters in the case of the electromagnetic actuators example. In comparison with the ES-based indirect adaptive controller of Haghi and Ariyur [2013], 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: $E(1,1) = 1,\ E(1,2) = 0.5$, and $E(2,i) = 0,\ i = 1,2$. 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 Haghi and Ariyur [2013], 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 that by using the modular ISS-based controller we manage to estimate the actual values of the uncertainties simultaneously, and improve the tracking performance. First, we see the effect of the uncertainties in figure 1, where the ISS controller is applied without the parameters learning. We can see a clear degradation of the tracking performance. Next, we apply the GP-UCB learning algorithm 3.3, with the following parameters: $\sigma = 0.1$, $l = 0.2$, and $\beta_t = 2\log(\frac{Card(D')t^2\pi^2}{6\delta})$,



(a) Obtained vs. desired first angular trajectory (No-learning)



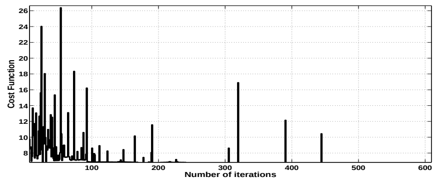
(b) Obtained vs. desired first angular velocity trajectory (No-learning)

Fig. 1. Obtained vs. desired trajectories (No-learning)

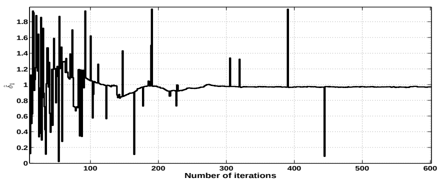
with $\delta = 0.05$. We test the GP-UCB algorithm under noisy measurements conditions, where we assume uniformly distributed additive noises on the angles measurements, with maximal excursion of $0.1\ rad$. The obtained parameters and tracking results are reported on figures 2(a), 2(b), 2(c), 3(a), 3(b). We can see on these figures that the uncertainties are well estimated despite the high measurements noise, as observed on the angular trajectory in figure 3(a). Furthermore, due to the ISS guarantee, the tracking performance is clearly improved, as seen in figures 3(a), 3(b).

5. CONCLUSION

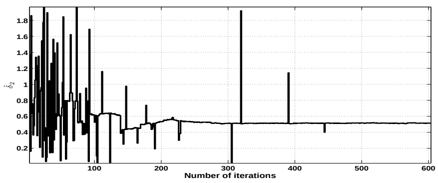
We have studied the problem of adaptive control for nonlinear systems which are affine in the control with parametric uncertainties. For this class of systems, we have proposed the following controller: We use a modular approach, where we first design a robust nonlinear controller based on the model (assuming knowledge of the uncertain parameters), and then complement this controller with an estimation module to estimate the actual values of the uncertain parameters. We propose to use a GP-UCB algorithm to learn in realtime the uncertainties of the model. We have guaranteed the stability (while learning) of the proposed approach by ensuring that the model-based robust controller leads to an ISS results, which guarantees boundedness of the states of the closed-loop system even during the learning phase. The ISS result together with a convergent GP-UCB learning-algorithm eventually leads to a bounded output tracking-error, which decreases with the decrease of the estimation error. We believe that one of the main advantages of the proposed controller, comparatively to the existing model-based adaptive controllers, is that we can learn (estimate) multiple uncertainties at the same time even if they appear in the model equation in a challenging structure, e.g., linearly dependent uncertainties affecting only one output, or uncertainties appearing in a nonlinear term of the model, which are well known limitations of the model-based estimation approaches. Another advantage of the proposed approach, is that due to its modular design, one could easily change the learning algorithm without having to change the model-based part of the controller.



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

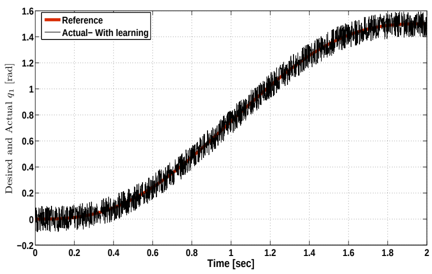


(b) Estimate of Δ_1 over the learning iterations (GP-UCB)

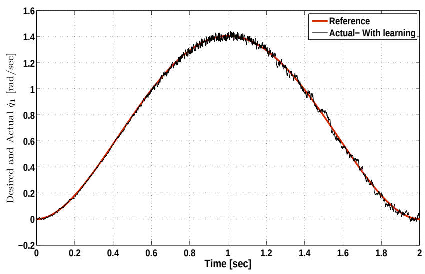


(c) Estimate of Δ_2 over the learning iterations (GP-UCB)

Fig. 2. Cost function and uncertainties estimates- (GP-UCB) algorithm



(a) Obtained vs. desired first angular trajectory (GP-UCB)



(b) Obtained vs. desired first angular velocity trajectory (GP-UCB)

Fig. 3. Obtained vs. desired trajectories (GP-UCB)

REFERENCES

G. Atinc and M. Benosman. Nonlinear learning-based adaptive control for electromagnetic actuators with proof of stability. In *IEEE, Conference on Decision and Control*, pages 1277–1282, 2013.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

M. Benosman. Extremum-seeking based adaptive control for nonlinear systems. In *IFAC World Congress*, pages 401–406, 2014a.

M. Benosman. Learning-based adaptive control for nonlinear systems. In *IEEE European Control Conference*, pages 920–925, 2014b.

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

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

M. Benosman and G. Atinc. Multi-parametric extremum seeking-based learning control for electromagnetic actuators. In *American Control Conference*, pages 1914–1919, 2013.

M. Benosman and G.M. Atinc. Non-linear adaptive control for electromagnetic actuators. *IET Control Theory and Applications*, 9(2):258–269, 2015.

Eric Brochu, Vlad M. Cora, and Nando 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.

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research (JMLR)*, 12(1655–1695), 2011.

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

P. Haghi and K. Ariyur. Adpative feedback linearization of nonlinear MIMO systems using ES-MRAC. In *American Control Conference*, pages 1828–1833, 2013.

Z-P Jiang, Andrew R Teel, and Laurent Praly. Small-gain theorem for ISS systems and applications. *Mathematics of Control, Signals and Systems*, 7(2):95–120, 1994.

H. Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2002.

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

E. Lavretsky. Combined/composit model reference adaptive control. *IEEE, Transactions on Automatic Control*, 54(11):2692–2697, 2009.

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

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Eduardo D Sontag. Smooth stabilization implies coprime factorization. *Automatic Control, IEEE Transactions on*, 34(4):435–443, 1989.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pages 1015–1022, 2010.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, 2012.

Cong Wang, David J Hill, Shuzhi Sam Ge, and Guanrong Chen. An ISS-modular approach for adaptive neural control of pure-feedback systems. *Automatica*, 42(5):723–731, 2006.

M. Xia and M. Benosman. Extremum seeking-based indirect adaptive control for nonlinear systems with time-varying uncertainties. In *European Control Conference*, pages 2780–2785, 2015.