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

Gaussian Processes-based Parametric Identification for Dynamical Systems

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

TR2017-089 July 2017

Abstract

In this paper we present some results on parametric identification for dynamical systems. More specifically, we consider the general case of dynamics described by partial differential equations (PDEs), which includes the special case of ordinary differential equations (ODEs). We follow a stochastic approach and formulate the identification problem as a Gaussian process optimization with respect to the unknown parameters of the PDE. We use proper orthogonal decomposition (POD) model reduction theory together with a data-driven Gaussian Process Upper Confidence Bound (GP-UCB), to solve the identification problem. The proposed approach is validated on the coupled Burgers' equation benchmark.

World Congress of the International Federation of Automatic Control (IFAC)

Gaussian Processes-based Parametric Identification for Dynamical 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: In this paper we present some results on parametric identification for dynamical systems. More specifically, we consider the general case of dynamics described by partial differential equations (PDEs), which includes the special case of ordinary differential equations (ODEs). We follow a stochastic approach and formulate the identification problem as a Gaussian process optimization with respect to the unknown parameters of the PDE. We use proper orthogonal decomposition (POD) model reduction theory together with a data-driven Gaussian Process Upper Confidence Bound (GP-UCB), to solve the identification problem. The proposed approach is validated on the coupled Burgers' equation benchmark.

1. INTRODUCTION

The aim of system identification (id) is to estimate the best possible model of a system, given a set of noisy experimental measurements. System identification can be classified as system identification for linear vs. nonlinear models, time-domain based vs. frequency-domain based, open-loop vs. closed-loop identification, stochastic vs. deterministic identification algorithms, identification for control vs. identification for simulation and prediction, goal-oriented based identification or co-design identification. Due to this large ramification of system id, we cannot present here an exhaustive review of the field, instead, we refer the reader to some outstanding surveys, e.g., Astrom and Eykhoff [1971],Ljung and Vicino [2005],Gevers [2006],Ljung [2010],Pillonetto et al. [2014].

We will focus here on one specific subarea of system id, namely, identification for dynamical systems described by partial differential equations (PDEs). In this subarea of system identification, we will present some results on a stochastic approach for open-loop parametric identification in the time domain.

Indeed, PDEs are important mathematical models, which are used to model complex dynamic systems in applied sciences. For instance, PDEs used to model flexible beams and ropes Montseny et al. [1997], Barkana [2014], crowd dynamics Huges [2003], Colombo and Rosini [2005], or fluid dynamics Rowley [2005], Li et al. [2013], MacKunis et al. [2011], Cordier et al. [2013], Balajewicz et al. [2013], etc. However, PDEs are infinite dimensional systems, which makes them hard to solve in closed-form, and computationally demanding to solve numerically. For instance, when using FEM discretization methods, one may end-up with a large discretization space, which implies large computation time. Because of this complexity, it is often hard to use PDEs to analyze, predict or control systems in real-time. Instead, one approach that is often used in real applications, is to first reduce the PDE model to an ordinary differential equation (ODE) model, which has a finite dimension and then use this ODE to control the system. This step of converting a PDE to a reduced order model (ROM) ODE, while maintaining a small error between the solutions of both models, is known as model reduction. However, if the original PDE has unknown physical parameters, then the associated ODE will inherit the same unknown physical parameters, which need to be identified based on real-time measurements.

Many results have been proposed for PDEs identification. For example in Xun et al. [2013], two methods have been proposed to estimate parameters of PDE models. Indeed, the first method is based on a parameter cascading approach, and the second method is a Bayesian approach based on Markov chain Monte Carlo (MCMC) techniques. Both methods rely on decomposing the PDE solutions in a linear basis function and then solving an optimization problem in the coefficients of the basis function, as well as, the PDE parameters to be identified. In Muller and Timmer [2004], two approaches have been investigated, one classified as a regression-based method, where all the terms of the PDE are computed based on measured data, and then the unknown coefficients of the PDE are obtained by solving an algebraic optimization problem, equaling both sides of the PDE equation. The second method is in the form of a dynamical approach, in the sense that the unknown parameters of the PDE are obtained by solving an optimization problem which minimizes the distance between the measured data and the solutions of the PDE solutions over time. Many other work on PDE identification fall into one of these two categories, e.g., refer to Parlitz and C.Merkwirth [2000], Voss et al. [1999] for some regression-based identification techniques, and Baake et al. [1992], Muller and Timmer [2002], Benosman [2016] for a dynamical approach for PDEs identification.

In this paper, we propose an alternative stochastic method, which can be classified as a dynamical approach. Indeed, we follow here the stochastic identification formulation, and use GP-UCB optimization methods to solve the PDE id problem. We use POD model reduction theory together with a data-driven stochastic optimization approach to solve the identification problem. We formulate the identification problem as a minimization of performance cost function modeled as a Gaussian Process (GP), and use the

GP-UCB to solve the optimization problem online, leading to a simple real-time solution for open-loop parametric identification for PDEs.

One of the main advantages of using GP-UCB to solve the identification (stochastic) optimal problem, is that it is a data-driven optimization algorithm which needs a small amount of measurements to direct the search for the optimal parameter. Moreover, GP-UCB algorithms are well known to be robust to measurement noise, which makes them a good candidate for solving identification problems, where measurements are often contaminated with noise, e.g., Ljung and Vicino [2005]. Another, non-negligible, advantage of GP-UCB algorithms, is that they are proven to converge to the global optimum, of a possibly non-convex function, in a compact search space, e.g., Rasmussen and Williams [2006].

This paper is organized as follows: We first introduce some notations and definitions in Section 2. Section 3 is dedicated to the problem formulation and the presentation of the proposed solution. The case of the coupled Burgers' equation is studied in Section 4. Finally, a conclusion is presented in Section 5.

2. BASIC NOTATIONS AND DEFINITIONS

Throughout the paper we will use $\|.\|$ to denote the Euclidean vector norm; i.e., for $x \in \mathbb{R}^n$ we have $\|x\| = \sqrt{x^T x}$. The Kronecker delta function is defined as: $\delta_{ij} = 0$, for $i \neq j$ and $\delta_{ii} = 1$. We will use $\dot{f}$ for the short notation of time derivative of f, and x^T for the transpose of a vector x. A function is said to be analytic in a given set, if it admits a convergent Taylor series approximation in some neighborhood of every point of the set. We consider the Hilbert space $\mathcal{Z} = L^2([0,1])$, which is the space of Lebesgue square integrable functions, i.e., $f \in \mathcal{Z}$, iff $\int_0^1 |f(x)|^2 dx < \infty$. We define on $\mathcal{Z}$ the inner product $\langle \cdot, \cdot \rangle_{\mathcal{Z}}$ and the associated norm $\|.\|_{\mathcal{Z}}$, as $\|f\|_{\mathcal{Z}}^2 = \int_0^1 |f(x)|^2 dx$, and $\langle f, g \rangle_{\mathcal{Z}} = \int_0^1 f(x)g(x)dx$, for $f, g \in \mathcal{Z}$. A function $\omega(t,x)$ is in $L^2([0,T];\mathcal{Z})$ if for each $0 \leq t \leq T$, $\omega(t,\cdot) \in \mathcal{Z}$, and $\int_0^T \|\omega(t,\cdot)\|_{\mathcal{Z}}^2 dt \leq \infty$.

Definition 1. (Haddad and Chellaboina [2008]).
An autonomous system $\dot{x} = f(t,x)$ is said to be *Lagrange stable* if for every initial condition x_0 associated with the time instant t_0, there exists $\epsilon(x_0)$, such that $\|x(t)\| < \epsilon$, $\forall t \geq t_0 \geq 0$.

3. IDENTIFICATION OF PDE MODELS BY GAUSSIAN PROCESS OPTIMIZATION

3.1 GP-UCB-based parameters identification

Consider a stable dynamical system modelled by a nonlinear PDE of the form

$$\dot{z} = \mathcal{F}(z,p) \in \mathcal{Z}, \tag{1}$$

where $\mathcal{Z}$ is an infinite-dimension Hilbert space, and $p \in \mathcal{P} \subset \mathbb{R}^m$ represents the vector of physical parameters to be identified, in the compact domain $\mathcal{P}$. While solutions to this PDE can be obtained through numerical discretization, e.g., finite elements, finite volumes, finite differences etc., these computations are often very expensive and not suitable for online applications. However, solutions of the original PDE often exhibit low rank representations in an 'optimal' basis, which is exploited to reduce the PDE to a finite dimension ODE.

The general idea is as follows: One first finds a set of 'optimal' (spatial) basis vectors $\phi_i \in \mathbb{R}^n$ (the dimension n is generally very large and comes form a 'brute-force' discretization of the PDE, e.g., finite element discretization), and then approximates the PDE solution as

$$z(t) \approx \Phi z_r(t) = \sum_{i=1}^{r} z_{ri}(t)\phi_i, \tag{2}$$

where Φ is a $n \times r$ matrix containing the basis vectors ϕ_i as column vectors. Next, the PDE equation is projected into the finite r-dimensional space via classical nonlinear model reduction techniques, e.g., Galerkin projection, to obtain a ROM of the form

$$\dot{z}_r(t) = F(z_r(t), p) \in \mathbb{R}^r, \tag{3}$$

where $F : \mathbb{R}^r \to \mathbb{R}^r$ is obtained from the original PDE structure, through the model reduction technique, e.g., the Galerkin projection. Clearly, the problem lies in the selection of this 'optimal' basis matrix Φ. There are many model reduction methods to find the projection basis functions for nonlinear systems. For example proper orthogonal decomposition (POD) Kunisch and Volkwein [2007], Gunzburger et al. [2007], dynamic mode decomposition (DMD) Williams et al. [2015], and reduced basis (RB) Haasdonk et al. [2008], are some of the most used methods. We will recall hereafter the POD method, however, the GP-UCB-based identification results are independent of the type of model reduction approach, and the id method presented in this paper applies regardless of the selected model reduction approach.

POD model reduction We give here a brief recall of POD basis functions computation, the interested reader can refer to Kunisch and Volkwein [2007], Gunzburger et al. [2007] for a more complete presentation about POD theory.

We consider here the case where the POD basis functions are computed from (exact) solutions' snapshots of the PDE obtained by a discretization method, e.g. FEM. The general idea behind POD is to select a set of basis functions that capture an optimal amount of energy of the original PDE. The POD basis functions are obtained from a collection of snapshots over a finite time support of the PDE solutions. In the context of this work, these snapshots are obtained by solving an approximation (discretization) of the PDE equation, e.g., using finite element method (FEM). The POD basis functions computation steps are presented below in more details.

First, the original PDE is discretized using any finite element basis functions, e.g., piecewise linear functions or spline functions, e.g., Sordalen [1997], Fletcher [1983]. Let us denote the associated PDE solutions approximation by $z_{fem}(t,x)$, where t stands for the scalar time variable, and x stands for the space variable. We consider here (for simplicity of the notations) the case of one dimension, where x is a scalar in a finite interval, which we consider, without loss of generality, to be [0, 1]. Next, we compute a set of s snapshots of approximate solutions as

$$S_z = \{z_{fem}(t_1,.), ..., z_{fem}(t_s,.)\} \subset \mathbb{R}^N, \tag{4}$$

where N is the selected number of FEM basis functions. Now we define the so called correlation matrix K^z elements as

$$K^z{}_{ij} = \frac{1}{s} < z_{fem}(t_i,.), z_{fem}(t_j,.) >, \; i,j = 1, ..., s. \tag{5}$$

We then compute the normalized eigenvalues and eigenvectors of K^z, denoted as λ^z, and v^z. Eventually, the i^{th} POD basis function is given by

$$\phi_i^{pod}(x) = \frac{1}{\sqrt{s}\sqrt{\lambda_i^z}} \sum_{j=1}^{j=s} v_i^z(j) z^{fem}(t_j, x),\ i = 1, ..., N_{pod}, \tag{6}$$

where $N_{pod} \leq s$ is the number of retained POD basis functions, which depends on the application.
One of the main properties of the POD basis functions is orthonormality, which means that the basis function satisfy the following equalities

$$< \phi_i^{pod}, \phi_j^{pod} >= \int_0^1 \phi_i^{pod}(x)\phi_j^{pod}(x)dx = \delta_{ij}, \tag{7}$$

where δ_{ij} denotes the Kronecker delta function. The solution of the PDE (1) can then be approximated as

$$z^{pod}(t, x) = \sum_{i=1}^{i=N_{pod}} \phi_i^{pod}(x) q_i^{pod}(t), \tag{8}$$

where q_i^{pod}, $i = 1, ..., N_{pod}$ are the POD projection coefficients (which play the role of the z_r in the ROM (3)). Finally, the PDE (1) is projected on the reduced dimension POD space using a Galerkin projection, i.e., both sides of equation (1) are multiplied by the POD basis functions, where z is substituted by z^{pod}, and then both sides are integrated over the space interval [0, 1], which using the orthonormality constraints (7) and the boundary constraints of the original PDE, leads to an ODE of the form

$$\dot{q}^{pod}(t) = F(q^{pod}(t), p) \in \mathbb{R}^{N_{pod}}, \tag{9}$$

where the structure (in terms of nonlinearities) of the vector field F is related to the structure of the original PDE, and where $p \in \mathbb{R}^m$ represents the vector of parametric uncertainties to be identified. Note that due to (9), and (8), z^{pod} is implicitly function of p, as well.

We can now proceed with the GP-based identification of the parametric uncertainties.

GP-UCB-based PDEs open-loop parameters estimation
We will use here a GP-UCB algorithm to estimate the PDE's parametric uncertainties, based on its reduced order model; the POD ROM. First, we need to introduce some basic stability assumptions.

Assumption A1 The solutions of the original PDE model (1) are assumed to be in $L^2([0, \infty); \mathcal{Z})$, and the associated POD reduced order model (8), (9) is Lagrange stable, $\forall\ p \in \mathcal{P} \subset \mathbb{R}^m$.

Remark 1. Assumption A1 is needed to be able to perform open-loop identification of the system.

Now, to be able to use the GP-UCB framework to identify the parameters vector p, we define an identification cost function as

$$Q(\hat{p}) = H(e_z(t, \hat{p})),\ t \in [0,\ T], \tag{10}$$

where $\hat{p}$ denotes the estimate of p, H is a positive definite function of e_z, and e_z represents the error between the ROM model (8),(9) and the system's measurements z_m, defined as

$$e_z(t, \hat{p}) = z^{pod}(t, x_m, \hat{p}) - z_m(t, x_m, \eta), \tag{11}$$

x_m being the points in space where the measurements are obtained, and η represents additive (Gaussian) measurement noise. At this point, we need to formalize the assumption that the true values of the parameters are optimal for the identification cost function Q.

Assumption A2 The cost function Q is continuous and has a minimum at $\hat{p}^* = p$.

Let us first recall that a Gaussian Process (GP) is a stochastic process indexed by the set $D \subseteq \mathcal{P}$ that has the property that for any finite subset of the evaluation points, that is $\{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_t\} \subset D$, the joint distribution of $(Q(\hat{p}_i))_{i=1}^t$ is a multivariate Gaussian distribution. GP is defined by a mean function $\mu(\hat{p}) = \mathbb{E}\left[Q(\hat{p})\right]$ and its covariance function (or kernel) $\kappa(\hat{p}, \hat{p}') = \mathrm{Cov}(Q(\hat{p}), Q(\hat{p}')) = \mathbb{E}\left[\left(Q(\hat{p}) - \mu(\hat{p})\right)\left(Q(\hat{p}') - \mu(\hat{p}')\right)^T\right]$. The kernel κ of a GP determines the behavior of a typical function sampled from the GP Rasmussen and Williams [2006]. For instance, if we choose

$$\kappa(\hat{p}, \hat{p}') = \exp\left(-\frac{\|\hat{p} - \hat{p}'\|^2}{2l^2}\right), \tag{12}$$

the squared exponential kernel with length scale $l > 0$, it implies that the GP is mean square differentiable of all orders.

Let us now briefly describe how we can find the posterior distribution of a $\mathrm{GP}(0, \kappa)$; a GP with zero prior mean. Suppose that for $\underline{\hat{p}}_{t-1} \triangleq \{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_{t-1}\} \subset D$, we have observed the noisy evaluation $y_i = Q(\hat{p}_i) + \eta_i$ 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 $\hat{p}^* \in D$ as follows: Denote the vector of observed values by $\mathbf{y}_{t-1} = [y_1, \ldots, y_{t-1}]^\top \in \mathbb{R}^{t-1}$, and define the Grammian matrix $K \in \mathbb{R}^{t-1 \times t-1}$ with $[K]_{i,j} = \kappa(\hat{p}_i, \hat{p}_j)$, and the vector $\kappa_* = [\kappa(\underline{\hat{p}}_1, \hat{p}^*), \ldots, \kappa(\underline{\hat{p}}_{t-1}, \hat{p}^*)]$. The expected mean $\mu_t(\hat{p}^*)$ and the variance $\sigma_t(\hat{p}^*)$ of the posterior of the GP evaluated at $\hat{p}^*$ are (cf. Section 2.2 of Rasmussen and Williams [2006])

$$\begin{aligned} \mu_t(\hat{p}^*) &= \kappa_* \left[K + \sigma^2 \mathbf{I}\right]^{-1} \mathbf{y}_{t-1}, \\ \sigma_t^2(\hat{p}^*) &= \kappa(\hat{p}^*, \hat{p}^*) - \kappa_*^\top \left[K + \sigma^2 \mathbf{I}\right]^{-1} \kappa_*. \end{aligned} \tag{13}$$

At time t, the GP-UCB algorithm selects the next query point $\hat{p}_t$ by solving the following optimization problem:

$$\hat{p}_t \leftarrow \operatorname*{argmin}_{\hat{p} \in D} \mu_{t-1}(\hat{p}) - \beta_t^{1/2} \sigma_{t-1}(\hat{p}), \tag{14}$$

where we choose β_t as [1]

$$\beta_t = 2\,\|Q\|_{\mathcal{H}_\kappa} + 300\gamma_t \log^3(t/\delta), \tag{15}$$

where $\delta \in (0, 1)$, represents the confidence parameter, and $\gamma_t = \log(t)^c,\ c > 0$.

Remark 2. The optimization problem (14) 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 real system and we would like to minimize the number of interactions with it before finding a $\hat{p}$ with small $Q(\hat{p})$. One practically way to approximately solve (14) 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.

Based on the above assumptions, we can summarize the open-loop identification result in the following Lemma.

Lemma 1. Consider the system (1), then under Assumptions A1, A2, the uncertain parameters vector p, can be estimated online using the algorithm

$$\hat{p}(t) = p_{nom} + \Delta p(t), \tag{16}$$

[1] $\|.\|_{\mathcal{H}_\kappa}$ denotes the norm associated with the reproducing kernel Hilbert space (RKHS), e.g. Rasmussen and Williams [2006].

where p_{nom} is the nominal value of p, $\Delta p = [\delta p_1, ..., \delta p_m]^T$ is computed using the GP-UCB algorithm (10), (11), (13), (14), and (15). Then, the parameters vector p is estimated, over $[0, T]$, with a cumulative regret $R_T = \sum_{t=0}^{T} r_t,\ r_t = Q(\hat{p}_t) - Q(p)$, which admits the following upper-bound

$$R_T \leq \sqrt{\frac{8T\beta_T\gamma_T}{1+\log(1+\sigma^{-2})}},$$

with probability at least $1-\delta$, where $\delta > 0$ is the confidence parameter.

Proof: Based on Assumption A1, we can ensure the existence and boundedness of the cost function given by (10), and (11). Next, to provide an upper bound on the regret of the method, we follow (Theorem 3, in Srinivas et al. [2010]), and fix the confidence parameter $\delta > 0$. If the reproducing kernel Hilbert space (RKHS) $\mathcal{H}_\kappa$ defined by the kernel κ is such that $\|Q\|_{\mathcal{H}_\kappa} < \infty$, we can choose

$$\beta_t = 2\|Q\|_{\mathcal{H}_\kappa} + 300\gamma_t \log^3(t/\delta),$$

in which γ_t should depend on kernel κ. For the exponential kernel and $D \subset \mathbb{R}^d$, we have $\gamma_t = O((\log t)^{d+1})$.

We define $\hat{p}* \leftarrow \text{argmin}_{\hat{p}\in D} Q(\hat{p})$, the global minimizer of the objective function. We define the regret at time t by $r_t = Q(\hat{p}_t) - Q(p)$, which is a measure of sub-optimality of the choice of $\hat{p}_t$ according the cost function Q. The cumulative regret over $[0, T]$ is defined as $R_T = \sum_{t=1}^{T} r_t$. If we choose $\hat{p}_t$ according to (14) with the aforementioned parameters, we can write the cumulative regret's upper-bound Srinivas et al. [2010]

$$R_T \leq \sqrt{\frac{8T\beta_T\gamma_T}{1+\log(1+\sigma^{-2})}},$$

with probability at least $1-\delta$.

4. THE COUPLED BURGERS' PDE EQUATION

We consider here the case of the coupled Burgers' equation, e.g., Kramer [2011]

$$\begin{cases} \dfrac{\partial\omega(t,x)}{\partial t} + \omega(t,x)\dfrac{\partial\omega(t,x)}{\partial x} = \mu\dfrac{\partial^2\omega(t,x)}{\partial x^2} - \kappa T(t,x), \\ \dfrac{\partial T(t,x)}{\partial t} + \omega(t,x)\dfrac{\partial T(t,x)}{\partial x} = c\dfrac{\partial^2 T(t,x)}{\partial x^2} + f(t,x), \end{cases} \tag{17}$$

where T represents the temperature, and ω represents the velocity field. κ is the coefficient of the thermal expansion, c the heat diffusion coefficient, μ the viscosity coefficient (inverse of the Reynolds number R_e), x is the one dimensional space variable $x \in [0, 1]$, $t > 0$, and f is the external forcing term such that $f \in L^2((0,\infty), X),\ X = L^2([0, 1])$. The previous equation is associated with the following boundary conditions

$$\begin{aligned} \omega(t,0) = \delta_1,\ \frac{\partial\omega(t,1)}{\partial x} = \delta_2, \\ T(t,0) = T_1,\ T(t,1) = T_2, \end{aligned} \tag{18}$$

where $\delta_1,\ \delta_2,\ T_1,\ T_2 \in \mathbb{R}_{\geq 0}$.

We consider here the following general initial conditions

$$\begin{aligned} \omega(0,x) = \omega_0(x) \in L^2([0, 1]), \\ T(0,x) = T_0(x) \in L^2([0, 1]). \end{aligned} \tag{19}$$

Following a Galerkin-type projection into POD basis functions, e.g., Kramer [2011], the coupled Burgers equation is reduced to a POD ROM with the following structure

$$\begin{aligned} \begin{pmatrix} \dot{q}_\omega^{pod} \\ \dot{q}_T^{pod} \end{pmatrix} &= B_1 + \mu B_2 + \mu\ D\ q^{pod} + \tilde{D}q^{pod} + Cq^{pod}q^{podT}, \\ \omega_{ROM}(x,t) &= \omega_0(x) + \sum_{i=1}^{i=N_{pod\omega}} \phi(x)_{\omega i}^{pod} q_{\omega i}^{pod}(t), \\ T_{ROM}(x,t) &= T_0(x) + \sum_{i=1}^{i=N_{podT}} \phi(x)_{Ti}^{pod} q_{Ti}^{pod}(t), \end{aligned} \tag{20}$$

where matrix B_1 is due to the projection of the forcing term f, matrix B_2 is due to the projection of the boundary conditions, matrix D is due to the projection of the viscosity damping term $\mu\frac{\partial^2\omega(t,x)}{\partial x^2}$, matrix $\tilde{D}$ is due to the projection of the thermal coupling and the heat diffusion terms $-\kappa T(t,x),\ c\frac{\partial^2 T(t,x)}{\partial x^2}$, and the matrix C is due to the projection of the gradient-based terms $\omega\frac{\omega(t,x)}{\partial x}$, and $\omega\frac{\partial T(t,x)}{\partial x}$. The notations $\phi_{\omega i}^{pod}(x),\ q_{\omega i}^{pod}(t)\ (i = 1, ..., N_{pod\omega})$, $\phi_{Ti}^{pod}(x),\ q_{Ti}^{pod}(t)\ (i = 1, ..., N_{podT})$, stand for the space basis functions and the time projection coordinates, for the velocity and the temperature, respectively. $\omega_0(x),\ T_0(x)$ represent the mean values (over time) of ω and T, respectively.

4.1 Burgers' equation parameters estimation

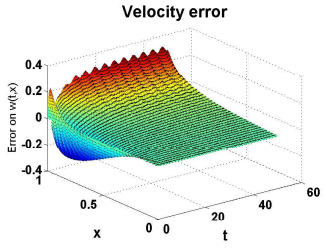
- *Test 1:* First, we report the case with an uncertainty on the Reynolds number R_e. We consider the coupled Burgers equation (17), with the parameters $R_e = 500,\ \kappa = -1,\ c = 0.01$, the boundary conditions $\delta_1 = 0,\ \delta_2 = 5,\ T_1 = 0,\ T_2 = 0.1sin(0.5\pi t)$, the initial conditions $\omega_0(x) = 2(x^2(0.5 - x)^2)$, $T_0(x) = 0.5sin(\pi x)^5$, and a zero forcing term f. We assume large uncertainties on both R_e, and κ. We consider that their assumed values are $R_{e-nom} = 50$, and $\kappa - nom = -0.5$. We apply the GP-UCB-based estimation algorithm of Lemma 1, to estimate both R_e, and κ. We choose the identification cost function as

$$Q = \sum_{t=0}^{t=T=50} Q_1\|e_T(t)\|^2 + \sum_{t=0}^{t=T=50} Q_2\|e_\omega(t)\|^2, \tag{21}$$

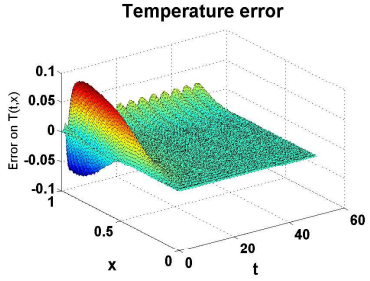
with $Q_1,\ Q_2 > 0$, $e_T(t) = T(t) - T_{ROM}(t)$, $e_\omega(t) = \omega(t) - \omega_{ROM}(t)$ define the errors at instant t, between the measurements and the POD ROM solution for temperature and velocity, respectively. We assume that the measurements are corrupted with additive (Gaussian) white noise with standard deviation $\sigma = 10^{-2}$. We applied the GP-UCB id algorithm (13), (14), and (15), with $\delta = 0.05,\ l = 10,\ \gamma_t = \log(t)^3,\ Q_1 = Q_2 = 1$. We consider the case of limited number of measurements, where we assume that we only have 10 sensors for the velocity and 10 for the temperature, uniformly located over the space $[0, 1]$.

We first show in Figure 1, the plots of error between the true solutions (obtained by solving the Burgers' PDE with finite elements method, with a uniform grid of 100 elements in time and space [2]), and the velocity and temperature profiles, obtained by the nominal POD ROM with 10 POD modes for the velocity and 10 modes for the temperature, considering the initial (incorrect) parameters' values $R_e = 50,\ \kappa = -0.5$. Now, we show the GP-UCB-based identification of the uncertain parameters $R_e,\ \kappa$. We first report in Figure 2(a), the id cost function over the identification iterations. We notice that the GP

[2] We thank here Dr. Boris Krämer @MIT, for sharing his codes to solve the Burgers' equation.



(a) Error between the true velocity and the nominal POD ROM velocity profile



(b) Error between the true temperature and the nominal POD ROM temperature profile

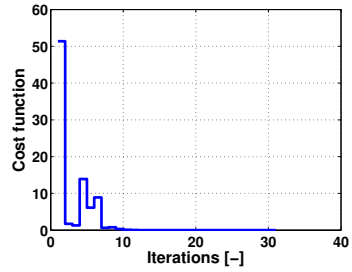
Fig. 1. Errors between the nominal POD ROM and the true solutions- Test 1

process exploration leads quickly, within 12 iteration, to the neighborhood of the true value of R_e, κ, as seen in Figures 2(b), 2(c). The error between the POD ROM after identification and the true solutions are depicted in Figure 3. By comparing Figure 1 and Figure 3, we can see that the error between the POD ROM solutions and the true solutions have been largely reduced with the identification of the actual values of the parameters.

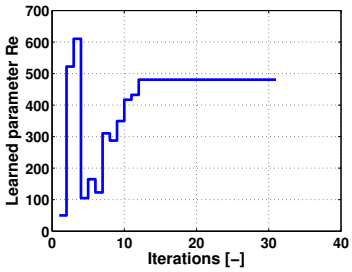
In a second set of tests, we compared the the performance of the GP-UCB-based identification algorithm, to the deterministic extremum seeking-based identification algorithm, introduced in Benosman [2016]. The associated numerical results could not be reported here to due to the limitation on the number of pages. The tests showed that the deterministic MES-based identification algorithm, managed to converge to a small neighborhood of the true parameters' values. However, in comparison with the GP-UCB stochastic id algorithm, the extremum-seeking-based identification of the parameters was less precise, which is due to the fact that the deterministic MES-based identification do not handle measurement noises as well as the stochastic GP-based id algorithm.

5. CONCLUSION

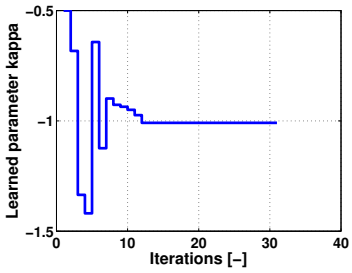
In this work we have studied the problem of dynamical systems parametric identification. More specifically, we have considered the case of dynamical systems described by PDEs (the case of ODEs being a sub-class which can be deduced from this formulation). We have formulated the problem as a stochastic optimization with respect the unknown parameters, and have proposed to use data-driven Gaussian processes GP-UCB theory to search for the PDE parameters. We believe that one of the main advantages of using GP-UCB theory for parametric identification is the fact that GP-UCB algorithms require a minimal number of cost function probing, i.e., real-time measurements. Furthermore, GP-UCB algorithms are well known to converge to the global optimum within a compact search set. Finally, GP-UCB algorithms, due to their stochastic nature, are robust to measurement noise. In this context, we have proposed to merge together POD model reduction theory and GP-UCB theory to propose a solution for PDEs parametric identification. The proposed approach showed a good performance when tested on the Burger's equation, however, its performance needs to be validated with more complex nonlinear PDEs, which will be the focus of our future research in this direction.



(a) id cost function vs. number of iterations



(b) Identified parameter $\hat{R}_e$ vs. number of iterations



(c) Identified parameter $\hat{\kappa}$ vs. number of iterations

Fig. 2. Identified parameters and id cost function- Test 1

REFERENCES

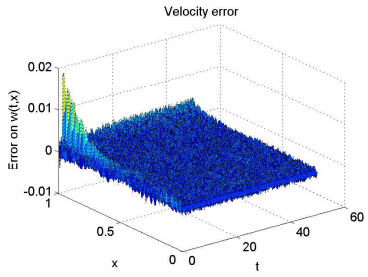
K.J. Astrom and P. Eykhoff. System identification– a survey. *Automatica*, 7:123–162, 1971.

E. Baake, M. Baake, H. Bock, and K. Briggs. Fitting ordinary differential equations to chaotic data. *Phys. Rev. A*, 45:5524–5529, 1992.

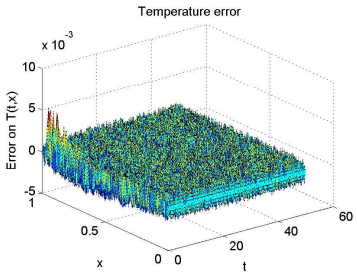
M.J. Balajewicz, E.H. Dowell, and B.R. Noack. Low-dimensional modelling of high–Reynolds–number shear flows incorporating constraints from the Navier-Stokes equation. *Journal of Fluid Mechanics*, 729(1):285–308, 2013.

I. Barkana. Simple adaptive control– a stable direct model reference adaptive control methodology–brief survey. *Int. Journal of Adaptive Control and Signal Processing*, 28:567–603, 2014.

M. Benosman. Extremum seeking-based parametric identification for partial differential equations. In *2nd IFAC*



(a) Error between the true velocity and the id-based POD ROM velocity profile



(b) Error between the true temperature and the id-based POD ROM temperature profile

Fig. 3. Errors between the id-based POD ROM and the true solutions- Test 1

Workshop on Control of Systems Governed by Partial Differential Equations, pages 19–24, 2016.

Rinaldo M. Colombo and Massimiliano D. Rosini. Pedestrian flows and non-classical shocks. *Mathematical Methods in the Applied Sciences*, 28(13):1553–1567, September 2005.

L. Cordier, B.R. Noack, G. Tissot, G. Lehnasch, J. Delville, M. Balajewicz, G. Daviller, and R. K. Niven. Identification strategies for model-based control. *Experiments in Fluids*, 54(1580):1–21, 2013.

C. A. J. Fletcher. The group finite element formulation. *Computer Methods in Applied Mechanics and Engineering*, 37:225–244, 1983.

M. Gevers. A personal view of the development of system identification a 30–year journey through an exciting field. *IEEE Control Systems Magazine*, pages 93–105, 2006.

M.D. Gunzburger, J. S. Peterson, and J.N. Shadid. Reduced–order modeling of time–dependent PDEs with mulptiple parameters in the boundary data. *Computer Methods in Applied Mechanics and Engineering*, 196(4–6):10301047, 2007.

B. Haasdonk, M. Ohlberger, and G. Rozza. A reduced basis method for evolution schemes with parameter-dependent explicit operators. *Electronic Transactions on Numerical Analysis*, 32:145–161, 2008.

W.M. Haddad and V. S. Chellaboina. *Nonlinear dynamical systems and control: a Lyapunov–based approach*. Princeton University Press, 2008.

R.L. Huges. The flow of human crowds. *Annual Review of Fluid Mechanics*, 35:169–182, 2003.

B. Kramer. Model reduction of the coupled Burgers equation in conservation form. Masters of science in mathematics, Virginia Polytechnic Institute and State University, 2011.

K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical Analysis*, 40(2): 492–515, 2007.

Kangji Li, Hongye Su, Jian Chu, and Chao Xu. A fast-POD model for simulation and control of indoor thermal environment of buildings. (60):150–157, 2013.

L. Ljung. Perspectives on system identification. *Annual Review in Control*, 34(1):112, April 2010.

L. Ljung and A. Vicino. Special issue on identification. *IEEE, Transactions on Automatic Control*, 50(10), Oct 2005.

W. MacKunis, S.V. Drakunov, M. Reyhanoglu, and L. Ukeiley. Nonlinear estimation of fluid velocity fields. In *IEEE, Conference on Decision and Control*, pages 6931–6935, 2011.

G. Montseny, J. Audounet, and D. Matignon. Fractional integro-differential boundary control of the Euler–Bernoulli beam. In *IEEE, Conference on Decision and Control*, pages 4973–4978, San Diego, California, 1997.

T. Muller and J. Timmer. Fitting parameters in partial differential equations from partially observed noisy data. *Physica D*, 171:1–7, 2002.

T. Muller and J. Timmer. Parameter identification techniques for partial differential equations. *Int. J. Bifurcation Chaos*, 14(06), 2004.

U. Parlitz and C.Merkwirth. Prediction of spatiotemporal time series based on reconstructed local states. *Phys. Rev. Lett.*, 84:1890–1893, 2000.

G. Pillonetto, F. Dinuzzo, T. Chenc, G. De Nicolao, and L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, (50):112, April 2014.

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

C.W. Rowley. Model reduction for fluids using balanced proper orthogonal decomposition. *INT. J. on Bifurcation and Chaos*, 2005.

O.J. Sordalen. Optimal thrust allocation for marine vessels. *Control Engineering Practice*, 15(4):1223–1231, 1997.

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.

H. U. Voss, P. Kolodner, M. Abel, and J. Kurths. Amplitude equations from spatiotemporal binary–fuid convection data. *Phys. Rev. Lett.*, 83:3422–3425, 1999.

M. Williams, I. Kevrekidis, and C. Rowley. A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, 2015. ISSN 1432-1467.

Xiaolei Xun, Jiguo Cao, Bani Mallick, Arnab Maity, and Raymond J. Carroll. Parameter estimation of partial differential equation models. *Journal of the American Statistical Association*, 108(503):1009–1020, 2013.