IFAC

Black-box identification and simulation of continuous-time nonlinear systems with random processes

Sylvain Vinet and Emmanuel Vazquez

Supélec, 91192 Gif-sur-Yvette, France

Abstract This paper proposes a methodology for black-box identification and simulation of continuous-time nonlinear dynamical systems based on random process modeling and kriging. It is assumed that the (finite-dimensional) state vector is observed with noise at regularly or irregularly spaced instants. The proposed identification method consists of two steps. The first step is the estimation of the time derivatives of the state vector. The second step consists in the approximation of the controlled vector field. For the simulation of the system, a new integration scheme is proposed. This integration scheme makes it possible to deal consistently with the error of approximation of the vector field.

Keywords: Nonlinear system identification; Continuous time system; Nonparametric methods

1. INTRODUCTION

This paper proposes a random process approach for black-box identification and simulation of continuous-time nonlinear dynamical systems defined by an ordinary differential equation. Let us consider an ODE

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}), \qquad (1)$$

with $\mathbf{x} \in \mathbb{R}^d$ (the state of the system), $\mathbf{u} \in \mathbb{R}^q$ (the control input), and $\mathbf{f} : \mathbb{R}^d \times \mathbb{R}^q \mapsto \mathbb{R}^d$ a C^1 nonlinear map (the vector field). Assume that $\mathbf{f}$ is unknown and that $\mathbf{x}$ is observed at a finite number of instants $t_1 < \cdots < t_n$. A first objective of this paper is to estimate $\mathbf{f}$ from $\mathbf{x}(t_1), \ldots, \mathbf{x}(t_n)$. A second task is to simulate the future trajectory of the system, i.e. to predict the state of the system at any instant $t > t_n$, given $\{\mathbf{u}(s) ; s \leq t\}$ and the set of observations.

A classical approach for black-box identification of continuous dynamical systems is to use *delay embedding*. More precisely, assume regularly sampled observations, and denote by $\mathbf{x}^k$ the vector $(\mathbf{x}_{k-\tau}, \ldots, \mathbf{x}_k)$, with $\mathbf{x}_i := \mathbf{x}(t_i)$, and by $\mathbf{u}^k$ the vector $(\mathbf{u}_{k-\tau}, \ldots, \mathbf{u}_k)$, with $\mathbf{u}_i := \mathbf{u}(t_i)$. Then, the nonlinear system (1) may be approximated by a recurrence equation written as

$$\mathbf{x}_k = \mathbf{g}(\mathbf{x}^{k-1}, \mathbf{u}^{k-1}), \qquad (2)$$

where the function $\mathbf{g}$ can be of two different types. *Parametric modeling* refers to the case where $\mathbf{g}$ is a parametric function, with a relatively small number of parameters (see, e.g., Walter and Pronzato, 1997). *Nonparametric modeling* generally means that $\mathbf{g}$ belongs to a space of infinite dimension, typically a space of splines or more generally a reproducing kernel Hilbert space (see, e.g., Sjöberg et al., 1995 ; Girard et al., 2003).

In many cases though, it would be useful to preserve the continuous-time representation of the system, i.e. to estimate $\mathbf{f}$ directly from the observations. A number of parametric approaches have already been proposed: state-dependent parameter estimation (Young et al., 2003), estimation based on differential algebra techniques (Fliess et al., 2006), the method of Hartley modulating functions (see, e.g., Rao and Unbehauen, 2006), etc. To the best of our knowledge, *nonparametric* approaches have not been addressed yet. In this paper, we propose a nonparametric approach to approximate $\mathbf{f}$ based on a framework of *random processes* and *Kriging*. The identification procedure can be parted into two steps. Since $\mathbf{f}$ maps the state to its derivative, the first step of the procedure is to estimate the derivative of the state from the set of observations. In the second step, an approximation of $\mathbf{f}$ is built based on the observations and the estimated derivatives. As mentioned above, a random process (rp) framework is used: a first rp, indexed by t, is used to model the state of the system as a function of time and a second rp, indexed by $\mathbf{x}$ and $\mathbf{u}$, models the unknown function $\mathbf{f}$. Using random processes makes it possible to deal comfortably and consistently with the errors of estimation involved at each step. Moreover, this framework yields a natural method of simulation of the system after the estimation of $\mathbf{f}$ has been carried out.

This paper is organized as follows. Section 2 recalls fundamental notions on the theory of linear prediction of random processes. Section 3 presents our two-step identification procedure. Section 4 deals with the simulation of the estimated system. Finally, Section 5 provides an example based on the Lotka-Volterra non-linear system.

2. LINEAR PREDICTION OF A RANDOM PROCESS

In this section, we shall recall some main results about *kriging* and *intrinsic kriging* (Matheron, 1973). These methods originate from geostatistics (see, e.g., Chilès and Delfiner, 1999) and they are used in this paper to build approximations of non-linear functions. Kriging and intrinsic kriging are primarily statistical methods but they can also be understood from the point of view of reproducing kernel Hilbert space methods (see, e.g., Wahba, 1990). Let us consider a function $z : \mathbb{T} \to \mathbb{R}$, where $\mathbb{T}$ is a set of parameters ($\mathbb{T} \subseteq \mathbb{R}^d$, for instance). We wish to approximate z based on a finite set of observations $z(t_i)$, $i \in \{1, \ldots, n\}$.

A classical idea is to model $z(t)$ by a second-order rp. An approximation of z is then obtained by considering the *average* of all sample paths of the rp that interpolate the observations. The theory of kriging and intrinsic kriging is concerned with the computation of this approximation based on the second-order moments of the rp. Kriging and intrinsic kriging are also known as the *best linear unbiased predictors* (BLUP). More specifically, kriging is used when the mean of ξ is assumed to be known, and intrinsic kriging is used otherwise.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and denote by $\xi : \Omega \times \mathbb{T} \to \mathbb{R}$ a second-order rp, i.e. a set of random variables indexed by elements of $\mathbb{T}$. Thus, for all $t \in \mathbb{T}$, $\xi(t) := \xi(\cdot, t)$ is an element of $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Let $m(t) := \mathrm{E}[\xi(t)]$ be the mean of $\xi(t)$ (the average of $\xi(t)$ on all sample paths) and denote by $k(t, s) := \mathrm{Cov}(\xi(t), \xi(s))$, $t, s \in \mathbb{T}$, its covariance function. It is assumed that z is a sample path of ξ and thus, each observation $z(t_i)$ corresponds to a particular outcome of the random variable $\xi(t_i)$. Although we do not need to assume specifically the Gaussianity of ξ throughout this paper, note that sample paths of Gaussian random processes can already define a very large class of functions when the covariance function and the mean vary (see e.g. Theorem 4 of Ghosal and Roy (2006)). Many properties of the sample paths follow from the characteristics of the covariance, especially in the Gaussian case. In particular, it is essential to choose k consistently with the regularity, the differentiability, the spectral properties, etc. of z. For the sake of brevity, we shall not talk about this issue, which is discussed extensively in the statistical literature. In practice, the covariance is chosen under the form of a parameterized function, the parameters of which are adapted to the observations using a goodness-of-fit criterion, such as *maximum likelihood* (Stein, 1999).

2.1 Kriging

Zero-mean random processes play an important role because their covariance function then correspond to a scalar product. Indeed, if $\forall t \in \mathbb{T}$, $m(t) = 0$, then $\forall t, s \in \mathbb{T}$, $(\xi(t), \xi(s))_{L^2} = \mathrm{E}[\xi(t)\xi(s)] = k(t, s)$. Let $\mathcal{H}$ denote the Hilbert space generated by a zero-mean process ξ, i.e. the completion of the vector space whose elements are finite linear combinations of random variables $\xi(t)$, $t \in \mathbb{T}$, endowed with the scalar product $(\xi(t), \xi(s))_{L^2} = k(t, s)$.

Let ξ_0 be a random variable in $\mathcal{H}$. Assume that we observe a finite set of random variables $\xi_i \in \mathcal{H}$, $i = 1, \ldots, n$. The *kriging predictor* of ξ_0 from $\xi_1, \ldots, \xi_n$ corresponds to the L^2-norm closest approximation of ξ_0 by a random variable $\widehat{\xi_0}$ in $\mathcal{H}_S = \mathrm{span}\{\xi_1, \ldots, \xi_n\}$. The orthogonal projection theorem states that the L^2-norm $\mathrm{E}[(\xi_0 - \widehat{\xi_0})^2]$ (the mean square error between the random variable and its predictor) is minimized *iff* $\widehat{\xi_0}$ is the orthogonal projection of ξ_0 onto $\mathcal{H}_S$. For the moment, we shall omit details since we will be more specific in the following section.

2.2 Intrinsic kriging

In this paper, we use *intrinsic Kriging* (IK) to obtain a linear predictor when the mean of $\xi(t)$ is unknown. We recall here the main results (Matheron, 1973).

Any rp ξ can always be rewritten as the sum of a zero-mean process and a deterministic function (the mean of ξ). In this paper, we shall consider polynomial mean functions only, so that ξ can be written as

$$\xi(t) = \sum_{i=1}^{m} \alpha_i p_i(t) + \eta(t), \qquad (3)$$

where $\eta(t)$ is a zero-mean rp, and where the p_is form a basis of low-degree d-variate polynomials. Let $\mathcal{P}$ be the m-dimensional vector space spanned by the functions $\{p_i\}_{i=1,\ldots,m}$ and $\mathcal{H}_\eta$ be the Hilbert space generated by η. When $m(t)$ is known, the framework of the previous section can be used by considering $\xi(t) - m(t)$. The difficulty to extend linear prediction (kriging) when $m(t)$ is unknown is that the spaces $\mathcal{H}_\eta$ and $\mathcal{P}$ are of different nature, and therefore, it is difficult to find a natural scalar product in the space generated by ξ, to apply the orthogonal projection theorem again.

To circumvent this difficulty and to deal comfortably with random processes that possess an unknown mean in $\mathcal{P}$, Matheron (1973) introduces a notion of *generalized random processes*, which extends that of random processes. Let $\tilde{\Lambda}$ be the vector space of all finite-support measures, *i.e.* the space of linear combinations $\sum_{i=1}^{n} \lambda_i \delta_{t_i}$, where δ_t stands for the Dirac measure, such that for any $B \subset \mathbb{T}$, $\delta_t(B)$ equals one if $t \in B$ and zero otherwise. Let $\tilde{\Lambda}_{\mathcal{P}\perp}$ be the subset of the elements of $\tilde{\Lambda}$ that vanish on $\mathcal{P}$. Thus, $\lambda \in \tilde{\Lambda}_{\mathcal{P}\perp}$ implies

$$\langle \lambda, z \rangle := \sum_{i=1}^{n} \lambda_i z(t_i) = 0, \quad \forall z \in \mathcal{P}.$$

Remark $\Lambda_{\mathcal{P}\perp}$ can be viewed as a set of finite-difference (increment) operators. For example, the condition for $\lambda = \sum_{i=1}^{n} \lambda_i \delta_{t_i}$ to be orthogonal to constant functions can be expressed as $\sum_{i=1}^{n} \lambda_i = 0$. Thus, $\lambda = \sum_{i=1}^{n} \lambda_i(\delta_{t_i} - \delta_{t_1})$, so λ is a linear combination of increment measures $\delta_{t_i} - \delta_{t_1}$.

If $\xi(t)$, $t \in \mathbb{T}$, is a second-order rp with mean $m(t)$ in $\mathcal{P}$ and covariance $k(t, s)$, the linear map

$$\xi : \tilde{\Lambda} \qquad\qquad \to \mathcal{H} = \mathrm{span}\{\xi(t) \,; t \in \mathbb{T}\}$$
$$\lambda = \sum_{i=1}^{n} \lambda_i \delta_{t_i} \qquad \mapsto \xi(\lambda) = \sum_{i=1}^{n} \lambda_i \xi(t_i),$$

extends $\xi(t)$ on $\tilde{\Lambda}$. $(\lambda, \mu)_{\tilde{\Lambda}} := (\xi(\lambda), \xi(\mu))_{\mathcal{H}}$ defines an inner product on $\tilde{\Lambda}$. Let Λ be the completion of $\tilde{\Lambda}$ under this inner product and extend $\xi(\lambda)$ on Λ by continuity. A *generalized rp* is then obtained. Note that for all $\lambda \in \Lambda$, $\mathrm{E}[\xi(\lambda)] = \langle \lambda, m(\cdot) \rangle$, and for all $\lambda = \sum_i \lambda_i \delta_{t_i} \in \tilde{\Lambda}$ and $\mu = \sum_j \mu_j \delta_{s_j} \in \tilde{\Lambda}$,

$$k(\lambda, \mu) := \mathrm{Cov}[\xi(\lambda), \xi(\mu)] = \sum_{i,j} \lambda_i \mu_j k(t_i, s_j). \qquad (4)$$

$k(\lambda, \mu)$ can be extended on Λ by continuity of the covariance operator. Denote also by $\Lambda_{\mathcal{P}\perp}$ the completion of $\tilde{\Lambda}_{\mathcal{P}\perp}$ under the inner product $(\cdot, \cdot)_\Lambda$. Since $\lambda \in \Lambda_{\mathcal{P}\perp}$ filters out any function of $\mathcal{P}$ and the mean of $\xi(t)$ is in $\mathcal{P}$, $\forall \lambda \in \Lambda_{\mathcal{P}\perp}$, $\xi(\lambda)$ is a zero-mean random variable. Remark also that $\forall \lambda, \mu \in \Lambda_{\mathcal{P}\perp}$, $k(\lambda, \mu) = (\lambda, \mu)_\Lambda$.

We can now recall the main result of IK. Let ξ be a rp with mean $m(t) \in \mathcal{P}$ and covariance $k(t, s)$. Let $\xi_0, \xi_1, \ldots, \xi_n$ be random variables in $\mathcal{H} = \mathrm{span}\{\xi(t) \,; \, t \in \mathbb{T}\}$. For all $i \in \{0, 1, \ldots, n\}$, there exists an element $\lambda_i \in \Lambda$ such that $\xi_i = \xi(\lambda_i)$. Assume n observations be sample values of the random variables $\xi_i^{\mathrm{obs}} = \xi_i + \varepsilon_i$, $i = 1, \ldots, n$, where the

ε_is are zero-mean random variables independent of $\xi(t)$, with covariance matrix $\mathbf{K}_\varepsilon$.

The *intrinsic kriging predictor* of ξ_0 based on the observations, is the linear projection

$$\widehat{\xi_0} = \sum_{i=1}^n a_{0,i}\xi_i^{\text{obs}} = \xi(\widehat{\lambda_0}) + \sum_{i=1}^n a_{0,i}\varepsilon_i \quad \left(\widehat{\lambda_0} = \sum_{i=1}^n a_{0,i}\lambda_i\right) \tag{5}$$

of ξ_0 onto $\mathcal{H}_S = \text{span}\{\xi_i^{\text{obs}}, i = 1, \ldots, n\}$, such that the variance of the prediction error $\xi_0 - \widehat{\xi_0}$ is minimized under the constraint

$$\lambda_0 - \widehat{\lambda_0} = \lambda_0 - \sum_{i=1}^n a_{0,i}\lambda_i \in \Lambda_{\mathcal{P}\perp}. \tag{6}$$

The coefficients $a_{0,i}$, $i = 1, \ldots, n$, are solutions of a system of linear equations (Matheron, 1973), which can be written in matrix form as

$$\begin{pmatrix} \mathbf{K} + \mathbf{K}_\varepsilon & \mathbf{P}^\mathsf{T} \\ \mathbf{P} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{a}_0 \\ \mu_0 \end{pmatrix} = \begin{pmatrix} \mathbf{k}_0 \\ \mathbf{p}_0 \end{pmatrix}, \tag{7}$$

where $\mathbf{K}$ is the $n \times n$ matrix with entries $k(\lambda_i, \lambda_j)$, $i, j = 1, \ldots, n$, $\mathbf{P}$ is a $m \times n$ matrix with entries $\langle \lambda_j, p_i \rangle$ for $j = 1, \ldots, n$ and $i = 1, \ldots, m$, μ_0 is a vector of Lagrange coefficients, $\mathbf{k}_0$ is a vector of size n with entries $k(\lambda_i, \lambda_0)$ and $\mathbf{p}_0$ is a vector of size m with entries $\langle \lambda_0, p_i \rangle$, $i = 1, \ldots, m$.

The variance of the prediction error, which accounts for the uncertainty of the prediction, is given by

$$\begin{aligned} \widehat{\sigma_0}^2 &:= \text{Var}[\xi_0 - \widehat{\xi_0}] \\ &= k(\lambda_0, \lambda_0) - 2\mathbf{a}_0^\mathsf{T}\mathbf{k}_0 + \mathbf{a}_0^\mathsf{T}(\mathbf{K} + \mathbf{K}_\varepsilon)\mathbf{a}_0 \\ &= k(\lambda_0, \lambda_0) - \mathbf{a}_0^\mathsf{T}\mathbf{k}_0 - \mathbf{p}_0^\mathsf{T}\mu_0. \end{aligned} \tag{8}$$

This variance, also called the *kriging variance*, makes it possible to assert confidence intervals for the predictor. In the following paragraphs, we shall use the notations

$$\begin{cases} \mathcal{K}(\xi_0 \mid \xi_1^{\text{obs}}, \ldots, \xi_n^{\text{obs}}) := \mathbf{a_0}^\mathsf{T}, \\ \mathcal{V}(\xi_0 \mid \xi_1^{\text{obs}}, \ldots, \xi_n^{\text{obs}}) := \text{Var}[\xi_0 - \widehat{\xi_0}]. \end{cases} \tag{9}$$

2.3 Prediction of derivatives

In this section, we recall how to use IK to estimate the derivatives of a rp from point-wise observations (Vazquez and Walter, 2005b). To simplify, suppose $\mathbb{T} = \mathbb{R}$. Recall that a zero-mean second-order rp $\xi(t)$ with covariance function $k(t, s)$ is mean-square differentiable at t if

$$\xi_h(t) = \frac{1}{h}(\xi(t + h) - \xi(t)) \tag{10}$$

converges in mean square when $h \to 0$. The limit exists if and only if $\partial^2 k(u, s)/\partial u \partial s$ exists at (t, t). If $\xi(t)$ is mean-square differentiable for all t, the limit process is called the *derivative process* and is denoted by $\dot{\xi}$. Higher-order derivatives are obtained by iteration and are denoted by $\xi^{(r)}$. It is straightforward to check that

$$\text{Cov}[\xi^{(q)}(t), \xi^{(r)}(s)] = \frac{\partial^{q+r}}{\partial t^q \partial s^r} k(t, s). \tag{11}$$

We now deal with the differentiability of generalized random processes. Let $\tau_h : \Lambda \to \Lambda$ be the translation operator such that for $\lambda = \sum_i \lambda_i \delta_{t_i} \in \tilde{\Lambda}$, $\tau_h \lambda = \sum_i \lambda_i \delta_{t_i+h}$. Then, define

$$\lambda_h := \frac{1}{h}(\tau_h \lambda - \lambda).$$

A generalized rp $\{\xi(\lambda) \; ; \; \lambda \in \Lambda\}$ is said to be *mean-square differentiable* at $\lambda \in \Lambda$ if $\xi(\lambda_h)$ converges in mean square as $h \to 0$. When the limit exists, it is denoted by $\dot{\xi}(\lambda)$. If $\partial^2 k(u, s)/\partial u \partial s$ exists at (t, t) for all $t \in \mathbb{R}$, then $\xi(\lambda)$ is mean-square differentiable for all λ (see Vazquez and Walter, 2005b). Remark that if $\dot{\xi}(\lambda)$ exists, there also exists $\dot{\lambda} \in \Lambda$ such that $\dot{\xi}(\lambda) = \xi(\dot{\lambda})$.

Derivatives of order r are denoted by $\xi^{(r)}(\lambda)$. Denote also by $\lambda^{(r)}$ the elements of Λ such that $\xi^{(r)}(\lambda) = \xi(\lambda^{(r)})$. Given $\lambda = \sum_i \lambda_i \delta_{t_i}^{(q_i)}$ and $\mu = \sum_j \mu_j \delta_{s_j}^{(r_j)}$ in Λ, it is easy to check that

$$\text{Cov}[\xi(\lambda), \xi(\mu)] = \sum_{i,j} (-1)^{r_j} \lambda_i \mu_j \frac{\partial^{q_i+r_j}}{\partial t_i^{q_i} \partial s_j^{r_j}} k(t_i, s_j).$$

Example   In this paragraph, we intent to give a practical example. Consider the estimation of the derivative of a function $z : [0, 1] \to \mathbb{R}$ from noisy observations of this function at a number of irregularly spaced points $t_i \in [0, 1]$, $i = 1, \ldots, n$. Assume z is a sample path of a rp ξ with constant mean (therefore, $\mathcal{P} = \text{span}\{1\}$) and stationary covariance $k(t, s) = k_s(t - s)$. The noisy observations of z are supposed to be sample values of the random variables $\xi_i^{\text{obs}} = \xi(t_i) + \varepsilon_i$, $i = 1, \ldots, n$, where the ε_is are i.i.d. zero-mean Gaussian random variables that model the noise of observation. To estimate the derivative of z, we compute the IK predictor of $\dot{\xi}(t)$ for all $t \in [0, 1]$. Thus, for all t, we search for a linear combination $\widehat{\dot{\xi}}(t) = \sum_{i=1}^n a_{t,i}\xi_i^{\text{obs}}$ such that $\text{Var}[\dot{\xi}(t) - \widehat{\dot{\xi}}(t)]$ is minimized under the constraint

$$\dot{\delta}_t - \sum_{i=1}^n a_{t,i}\delta_{t_i} \in \Lambda_{\mathcal{P}\perp} \Leftrightarrow \sum_{i=1}^n a_{t,i} = 0. \tag{12}$$

The solution can be obtained using (7), which reads

$$\begin{pmatrix} & & & 1 \\ \mathbf{K} + \sigma_\varepsilon^2 \mathbf{I}_n & & \vdots \\ & & & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{a}_t \\ \mu_t \end{pmatrix} = \begin{pmatrix} \mathbf{k}_t' \\ 0 \end{pmatrix}, \tag{13}$$

where $\mathbf{I}_n$ denotes the identity matrix and $\mathbf{k}_t'$ corresponds to the vector with elements $k_s'(t - t_i)$. More examples (with illustrations) can be found in Vazquez and Walter (2005b).

3. TWO-STEP IDENTIFICATION PROCEDURE

In this section, we present a procedure for black-box identification of a continuous-time nonlinear dynamical system from a finite number of (possibly noisy) observations of the state vector. In principle, the proposed procedure could also be applied if the state vector were only partially observed, provided that the non-observed components can be recovered from the observed components through linear operations (such as differentiation or integration, for instance). Indeed, the framework of kriging makes it possible to predict the result of any linear operator acting on a rp. For the sake of brevity, we shall not deal with partially observed state vectors in this paper. Besides, to avoid obfuscation of notations, the presentation will be further simplified by supposing the state vector has dimension one ($d = 1$). The generalization to higher dimensions is straightforward (as will be seen in the example of Section 5). As mentioned in the introduction, the procedure consists of two steps. The objective

of the first step is to estimate the derivative of x. In the second step, an approximation $\widehat{f}$ of f is computed.

3.1 Estimation of the derivative of the state

Let $\{x_i^{\mathrm{obs}},\ i = 1\ldots n\}$ be a set of noisy observations of the state at the instants t_i, which may not be evenly spaced. Our objective is to approximate $\dot{x}(t)$, $t \in \mathbb{R}$, from the noisy observations. Consider a mean-square differentiable rp $X \in \mathbb{R}$, indexed by $\mathbb{R}$, with constant but unknown mean m, and (at least) twice-differentiable covariance function k_X. Assume moreover that x is a sample path of X, and that the noise is modeled by i.i.d. random variables V_i, $i = 1, \ldots, n$, with zero-mean and known variance σ_V^2. Thus, for all i, x_i^{obs} is a sample value of the random variable

$$X_i^{\mathrm{obs}} := X(t_i) + V_i. \tag{14}$$

Using the results of Section 2.3, the derivative of x at t_i, $i = 1\ldots n$, can be estimated by computing the IK predictor $\widehat{\dot{X}}(t_i)$ of $\dot{X}(t_i)$ based on X_j^{obs}, $j = 1, \ldots, n$. Then, for all $i = 1, \ldots, n$, $\dot{x}(t_i)$ may be estimated by

$$\widehat{\dot{x}}(t_i) = \mathcal{K}_X\big(\dot{X}(t_i)|\mathbf{X}^{\mathrm{obs}}\big)\, \mathbf{x}^{\mathrm{obs}}, \tag{15}$$

with

$$\mathbf{X}^{\mathrm{obs}} = (X_1^{\mathrm{obs}}, \ldots, X_n^{\mathrm{obs}})^{\mathsf{T}} \text{ and}$$
$$\mathbf{x}^{\mathrm{obs}} = (x_1^{\mathrm{obs}}, \ldots, x_n^{\mathrm{obs}})^{\mathsf{T}}.$$

The notation $\mathcal{K}_X$ indicates that the covariance k_X is used to compute the kriging coefficients. The estimation error has a variance given by $\mathcal{V}_X(\dot{X}(t_i)|\mathbf{X}^{\mathrm{obs}})$.

3.2 System approximation

The next step of the identification procedure is to approximate the function $f : \mathbb{R} \to \mathbb{R}$ using the estimated derivatives (15). (Remember that we assumed $d = 1$. If $d > 1$, each component f_i, $i = 1, \ldots, d$, of the vector-valued function $\mathbf{f}$ should be approximated separately, cf Section 5.) Suppose that f is a sample path of a rp denoted by F, indexed by x and $\mathbf{u}$, with mean $m(x, \mathbf{u})$ and covariance $k_F\{(x, \mathbf{u}), (x', \mathbf{u}')\}$, where $(x, \mathbf{u}), (x', \mathbf{u}') \in \mathbb{R} \times \mathbb{R}^q$. As mentioned in Section 2.2, $m(x, \mathbf{u})$ can be an unknown constant or a low-degree polynomial in x and $\mathbf{u}$, which makes it possible to incorporate prior knowledge on f (for instance, one can specify that F has a linear trend along x or $\mathbf{u}$ (Vazquez and Walter, 2005a)). For the approximation of f, two cases are to be considered depending on whether or not the observation noise is assumed low or strong.

When the noise is low, we assume that the observation error (the difference between x_i^{obs} and $x(t_i)$) can be neglected. Then, for all i, the estimated derivative (15) can be viewed as a sample value of F at $(x_i^{\mathrm{obs}}, \mathbf{u}(t_i)) \approx (x(t_i), \mathbf{u}(t_i))$. The estimation error of the derivative of the state is taken into account by introducing an independent Gaussian random variable W_i with zero-mean and variance equal to $\mathcal{V}_X(\dot{X}(t_i)|\mathbf{X}^{\mathrm{obs}})$. Thus, for all i, define the random variable $F_i^{\mathrm{obs}} := F(x_i^{\mathrm{obs}}, \mathbf{u}(t_i)) + W_i$. For all x and $\mathbf{u}$, the IK predictor $\widehat{F}(x, \mathbf{u})$ of $F(x, \mathbf{u})$ based on the random variables F_i^{obs}, $i = 1, \ldots, n$, can be used to obtain an approximation of $f(x, \mathbf{u})$, written as

$$\widehat{f}(x, \mathbf{u}) = \mathcal{K}_F\big(F(x, \mathbf{u}) \mid \mathbf{F}^{\mathrm{obs}}\big)\, \widehat{\mathbf{x}} \tag{16}$$

with

$$\mathbf{F}^{\mathrm{obs}} = (F_1^{\mathrm{obs}}, \ldots, F_n^{\mathrm{obs}})^{\mathsf{T}} \quad \text{and}$$
$$\widehat{\mathbf{x}} = (\widehat{x}(t_1), \ldots, \widehat{x}(t_n))^{\mathsf{T}}.$$

The notation $\mathcal{K}_F$ indicates that the covariance k_F is used to compute the kriging coefficients.

In the strong noise case, the positioning error of the observation of F has to be taken in account. The observation model can now be written as $F_i^{\mathrm{obs}} := F(x(t_i) + V_i, \mathbf{u}(t_i)) + W_i$, with V_i the observation noise. Of course, the covariance between F_i^{obs} and $F(x', \mathbf{u}')$ is not $k_F\{(x(t_i), \mathbf{u}(t_i)), (x', \mathbf{u}')\}$. However, the correct covariance can be derived quite easily and a linear predictor similar to (16) can be obtained again. In the literature of kriging, the modification of the covariance function due to some positioning uncertainty is a classical issue (see, e.g., Chilès and Delfiner, 1999, p. 74–80). Although of practical importance, the case of strong noise we will not be developed in this paper due to the lack of space.

After the second step of the identification procedure, an approximation $\widehat{f}$ of f is obtained. One can then use a standard ODE solver to simulate the approximate system $\dot{x} = \widehat{f}(x, \mathbf{u})$. However, we believe that the error of approximation of f should be taken into account during the simulation. In the next section, we propose an integration method that uses the rp point of view presented above.

4. SIMULATION

In this section, we propose a numerical integration method of the ODE (1) based on linear prediction of the state vector. As in Section 3, we shall assume that $d = 1$ to simplify the presentation. Again, the extension to $d > 1$ is straightforward, since each dimension can be considered separately. In essence, the proposed integration scheme is a multi-step integration method, such as for instance the fixed-step Adams method (see, e.g., Butcher, 2003). A multi-step predictor of the state may be written as

$$x_{n+1} = \sum_{i=0}^{p} \alpha_i x_{n-i} + \sum_{j=0}^{q} \beta_j f(x_{n-j}, \mathbf{u}((n-j)h)),\ n \geq 0, \tag{17}$$

where h is the integration step size and x_n denotes the prediction of $x(nh)$ given the initial conditions $x_0 := x(0), x_{-1} := x(-h), x_{-2} := x(-2h), \ldots$ In classical integration methods, the coefficients α_i and β_j are obtained by minimizing the prediction error under a polynomial approximation of x. In the proposed simulation method, the coefficients are obtained as the result of the best linear prediction of the rp X.

First, assume that f is known exactly. To predict x_{n+1} from $x_n, \ldots, x_{n-q}$, $q \geq 0$, consider the IK linear predictor $\widehat{X}((n+1)h)$ of $X((n+1)h)$ based on $X(nh)$ and $\dot{X}((n-q)h), \ldots, \dot{X}(nh)$. Then, a one-step ahead prediction of the state can be written as

$$x_{n+1} = \mathcal{K}_X\big(X((n+1)h) \mid$$
$$\qquad X(nh), \dot{X}((n-q)h), \ldots, \dot{X}(nh)\big)\, \mathbf{x}_n^{\sharp}, \tag{18}$$

with $\mathbf{x}_n^{\sharp} := (x_n, \dot{x}_{n-q}, \ldots, \dot{x}_n)^{\mathsf{T}}$ and, for all i,

$$\dot{x}_i := f(x_i, \mathbf{u}(ih)). \tag{19}$$

Remark 1 The predictor (18) is a particular case of (17) with $p = 0$ (see Figure 1). When $p = 0$, the stability of the integration scheme (17) is ensured if $\alpha_0 \leq 1$ (Butcher, 2003). Due to the unbiasedness condition (6), we have in fact here $\alpha_0 = 1$. This explains our choice to consider a linear prediction without $X((n-1)h), X((n-2)h)$, etc.

Remark 2 Assume that the covariance is stationary so that $k(t,t') = k_{\text{iso}}(h)$, with $h = |t - t'|$. Then the proposed integration method is consistent. Moreover if $k_{\text{iso}}(h)$ is s-times differentiable at the origin, then $\text{E}\big[\big(\widehat{X}((n + 1)h) - X((n + 1)h)\big)^2\big] = O(h^s)$. We do not provide the proof of this result in this paper due to the lack of space.

Remark 3 A variable step-size procedure could also be proposed by adapting the procedure above.

When f is approximated, it is possible to account for the error of approximation of f in the proposed integration scheme. In this case, consider the IK linear predictor $\widehat{X}((n + 1)h)$ of $X((n + 1)h)$ based on $X(nh)$ and $\dot{X}((n - i)h) + \tilde{W}_i$, $i = 0, \ldots, q$, where the $\tilde{W}_i$s are independent zero-mean random variables with variance equal to

$$\mathcal{V}_F(F(x_{n-i}, \mathbf{u}((n - i)h)) \mid F^{\text{obs}}_j, j = 1, \ldots, n).$$

Here, the $\tilde{W}_i$s carry the uncertainty on $\dot{x}((n - i)h)$ due to the error of approximation of f. Then, f has to be replaced by $\widehat{f}$ in (19).
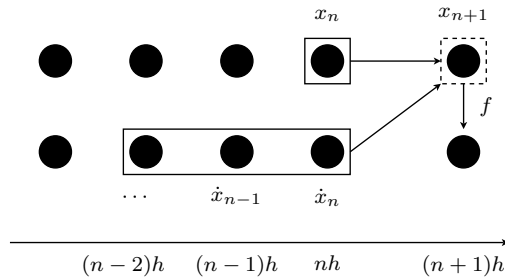

Figure 1. Illustration of the proposed integration scheme

5. EXPERIMENT

In this section, we present some experimental results to illustrate the proposed black-box identification and simulation procedures. We consider a driven Lotka-Volterra nonlinear ODE

$$\begin{cases} \dot{x}_1(t) = f_1(x_1, x_2, u_1) = \alpha x_1 - \beta x_1 x_2 + u_1\,, \\ \dot{x}_2(t) = f_2(x_1, x_2, u_2) = -\gamma x_2 - \delta x_1 x_2 + u_2\,, \end{cases} \quad (20)$$

with $\alpha = \gamma = 1$, $\beta = 0.2$, $\delta = 0.5$ and

$$u_1(t) = \cos^2(t/\exp(1))\cos^2(t)\,,$$
$$u_2(t) = \cos^2(t/\exp(1))\sin^2(t)\,.$$

The observation set consists of $n = 40$ samples of the state vector at evenly spaced instants $0, h, \ldots, (n - 1)h$, with $h = 0.5$, simulated using a standard ODE solver and corrupted with a Gaussian $\mathcal{N}(0, 10^{-4})$ white noise. The components of the state vector are modeled by two independent rps X_1 and X_2 with constant but unknown mean and a Wendland covariance function (see, e.g., Wendland, 2005). The class of Wendland functions provides stationary covariances with compact support. Here, the same C^6 Wendland function

$$k_X(t,s) = \sigma^2\big(1 - h/\rho\big)^8_+ \big(32(h/\rho)^3 + 25(h/\rho)^2 + 8h/\rho + 1\big)\,,$$

$h = |t - s|$, was chosen for the covariance functions of X_1 and X_2. The parameters σ^2 and ρ were estimated by maximum likelihood (Stein, 1999). Figure 2 shows the state vector components and their approximate first derivatives, as estimated in the first step of identification procedure by IK (Section 3.1).
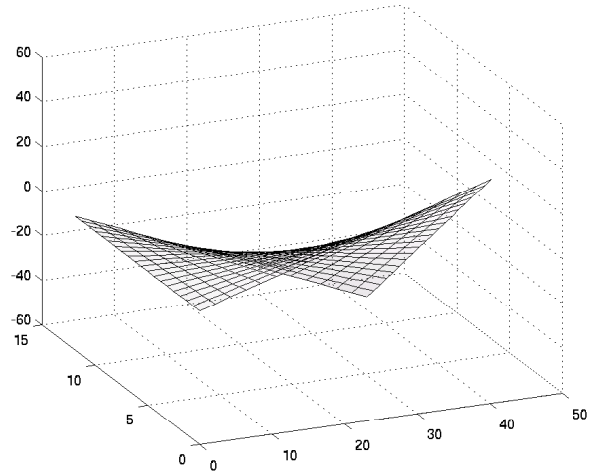

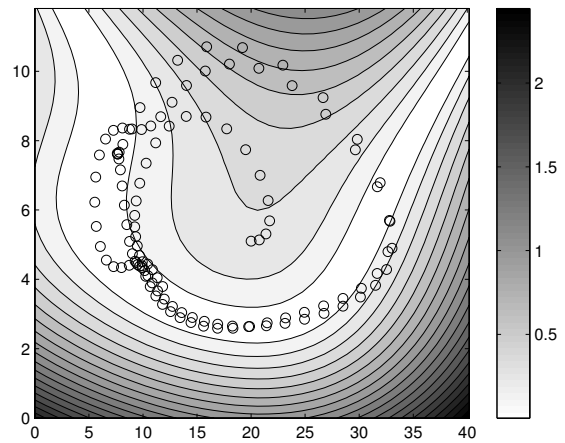Figure 3. Representation of $f_1(x_1, x_2, 0) = x_1 - 0.2x_1 x_2$.

Figure 4. Contour plot in the plane (x_1, x_2) of the absolute difference between $f_1(x_1, x_2, 0)$ and its approximation $\widehat{f}_1(x_1, x_2, 0)$ obtained by our identification procedure. The circles indicate the position of the observations in the plane (x_1, x_2).

The next part of the identification procedure consists in the approximation of $\mathbf{f}$. Here, the two components of $\mathbf{f}$ are modeled by two independent rp F_1 and F_2, with an unknown mean of the form $\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{1,1} x_1 x_2 + \alpha_{2,0} x_1^2 + \alpha_{0,2} x_2^2$ and a generalized covariance written as $k_F(\mathbf{x}, \mathbf{x}) = a_1 h - a_3 h^3 + a_5 h^5$, with $h = \|\mathbf{x} - \mathbf{x}'\|$. The parameters $a_i \geq 0$ are estimated by maximum likelihood. Figures 3 and 4 illustrate the approximation of the first component of $\mathbf{f}$ as obtained in the second step of our identification procedure.

Once $\mathbf{f}$ has been approximated, we want to predict its future trajectory. In Figure 2, we show the result of the simulation as obtained by the procedure described in Section 4 against the true trajectory. The prediction error remains small and does not grow with time.
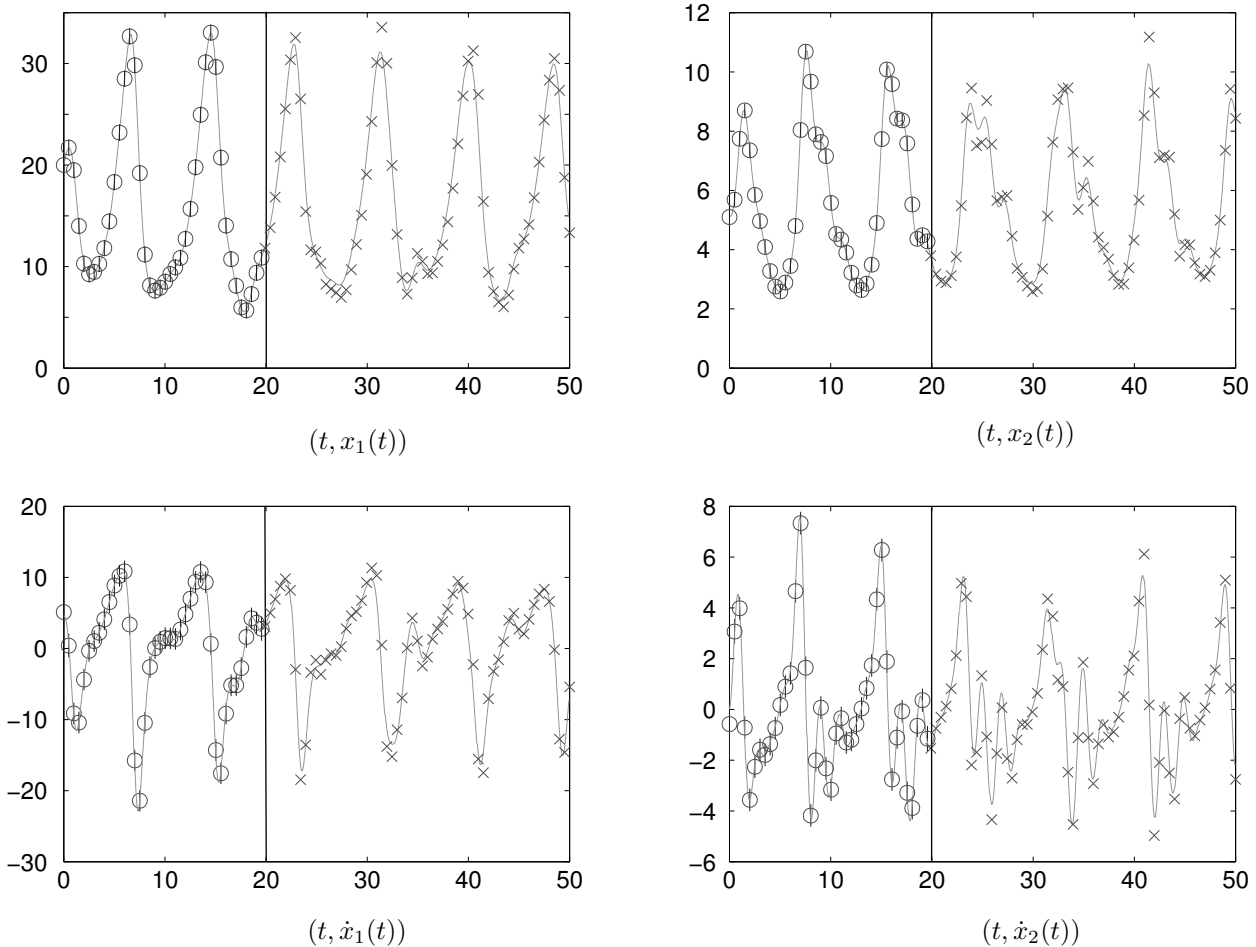
14395

Figure 2. Simulation of the system (20). Solid line: $x_1(t)$, $x_2(t)$, $\dot{x}_1(t)$ and $\dot{x}_2(t)$ computed by the ODE solver. Circles: the observations and the estimated derivatives in the identification procedure. Crosses: components of the state vector and their derivative obtained with the proposed simulation method. The vertical line at $t = 20$ marks the starting time of the simulation.

REFERENCES

J.C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2003.

J.-P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 1999.

M. Fliess, C. Join, and H. Sira-Ramirez. Complex continuous nonlinear systems: Their black box identification and their control. In *Proc. SYSID'2006*, 2006.

S. Ghosal and A. Roy. Posterior consistency of gaussian process prior for nonparametric binary regression. *Ann. Statist.*, 34: 2413–2429, 2006.

A. Girard, C. E. Rasmussen, J. Quiñonero Candela, and R. Murray-Smith. Multiple-step ahead prediction for non linear dynamic systems – a Gaussian Process treatment with propagation of uncertainty. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15, pages 529–536. MIT press, 2003.

G. Matheron. The intrinsic random functions, and their applications. *Adv. Appl. Prob.*, 5:439–468, 1973.

G.P. Rao and H. Unbehauen. Identification of continuous-time systems. *IEE Proceedings - Control Theory and Applications*, 153(2):185–220, 2006.

J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.

M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.

E. Vazquez and E. Walter. Intrinsic kriging and prior information. *Appl. Stoch. Models Bus. Ind.*, 2005a.

E. Vazquez and E. Walter. Estimating derivatives and integrals with kriging. In *IEEE 44th Conference on Decision and Control*, 2005b.

G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1990.

E. Walter and L. Pronzato. *Identification of Parametric Models from Experimental Data*. Communication and Control Engineering. Springer, London, 1997.

H. Wendland. *Scattered Data Approximation*. Cambridge University Press, New York, 2005.

P. Young, H. Garnier, and A. Jarvis. The identification of continuous-time linear and nonlinear models: a tutorial with environmental applications. In *In Proc. SYSID'2003*, 2003.