



HAL
open science

System identification by operatorial cancellation of nonlinear terms and application to a class of Volterra models

Céline Casenave, Emmanuel Montseny, Gérard Montseny

► **To cite this version:**

Céline Casenave, Emmanuel Montseny, Gérard Montseny. System identification by operatorial cancellation of nonlinear terms and application to a class of Volterra models. *International Journal of Robust and Nonlinear Control*, 2017, 27 (8), pp.1211-1241. 10.1002/rnc.3622 . hal-01606015

HAL Id: hal-01606015

<https://hal.science/hal-01606015v1>

Submitted on 23 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

System Identification by Operatorial Cancellation of Nonlinear Terms and Application to a Class of Volterra Models

C. Casenave^{1,*}, E. Montseny^{2,3} and G. Montseny^{2,4}

¹ INRA, UMR MISTEA, 2 place Pierre Viala, 34060 Montpellier.

² CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France.

³ Univ de Toulouse, UPS, LAAS, F-31400 Toulouse, France.

⁴ Univ de Toulouse, LAAS, F-31400 Toulouse, France.

Abstract

In this paper, a method is proposed for the identification of some SISO nonlinear models with two ill-known components of different nature: a linear (possibly dynamic) part and a static nonlinear one. This method is well adapted when no a priori information is available about the nonlinear component to be identified. It is based on a difference operator which enables to cancel the nonlinear term when applied to the model. Only the ill-known linear part remains in the transformed model; it can therefore be identified independently of the nonlinear term. Based on the identified linear component, we have access to a pseudo-graph of the nonlinear term, whose shape can give precious information for the parameterization of the unknown nonlinear part and its identification. The identification model under consideration is defined in an abstract framework, with very weak hypotheses, so that the proposed approach has a large scope. To highlight the method, a class of dynamic Volterra models including some hybrid models such as dynamic inclusions is considered for application.

KEY WORDS: System identification; Operatorial cancellation; Nonlinear system; Volterra model; Diffusive representation

1 Introduction

System identification methods depend on the class of identification models under consideration. Although discrete-time (DT) models have been mainly studied [1, 2], numerous methods are now available to identify continuous-time (CT) systems [3, 4]. Two types of CT-based identification methods have been developed: the “direct” and the “indirect” ones. Whereas the model is directly identified under its CT form when using a “direct” approach, an “indirect” CT-based identification method consists first in identifying a DT model and then in transforming it into a CT model. In the present paper, we propose a direct CT-based identification method.

Identification methods also depend on the structure of the model and the nature of the unknown components to be identified. Numerous methods have been developed for the identification of linear dynamic models (see for example [3]), but nonlinear models are now more and more considered [5, 6] as they are frequently involved in Physics and Biology. The identification methods for nonlinear systems can be classified into three categories, depending on the form of the considered identification models: (1) the block structured models (e.g. Hammerstein and Wiener models [7, 8]); (2) the kernel or nonparametric models (e.g. Volterra series models [9, 10]); (3) and the parametric models (e.g.

*Correspondence to: Email: celine.casenave@supagro.inra.fr

NARMAX models [11]). Among the block-structured approaches, some methods only focus on the identification of a nonlinear static component, whereas other ones consider models in which both nonlinear static and linear dynamic components have to be identified (e.g. the Hammerstein models [12], or other structured models [13]).

Concerning the dynamic part of the model, a lot of methods (modulating functions, integral equation, linear filters approaches, etc. [3]) have been developed for the identification of linear time-invariant dynamic operators which can be characterized by means of a rational transfer function. In that case, the goal is to identify the coefficients of the transfer function, whose structure (number of poles and zeros) is given. The main advantage of such systems is that they admit a state representation in which the parameters to be identified appear linearly. However, such models cannot be used to represent all types of linear dynamic operators. Indeed, some complex dynamic phenomena need more sophisticated models, not necessarily of finite dimension. Several papers consider some specific classes of infinite dimensional systems such as time-delay systems [14], non standard stochastic processes [15] or some distributed physical phenomena [16]. Distributed delay models (a particular case of time integro-differential equations which can be characterized by a general transfer function, non necessarily rational) have also been considered in some works [17, 18].

In the present paper, we propose an identification method for a generic class of SISO nonlinear systems with two ill-known components of different nature: the first one is nonlinear static and the second one is linear (possibly dynamic and infinite dimensional). The novelty of the proposed approach mainly comes from the principle of the identification method itself. It consists in decoupling the identification of the dynamic part of the model from the one of its nonlinear static part by use of an original “difference” operator which takes advantage of the difference of nature of the two unknown components. When applied to the original system, this operator enables to cancel the nonlinear term of the dynamic model, in such a way that the identification of the remaining ill-known linear dynamic part can first be done independently. Then, based on the so-obtained estimation of the linear component, we have access to a pseudo-graph of the unknown nonlinear component whose shape can give precious information for the parameterization of the nonlinear component and its identification. For example, it can reveal the presence of a discontinuity or a singularity. Note also that the method has a large scope; the considered class of models includes the Hammerstein models for which specific identification methods have been developed (see for example [12, 19]). It also includes the class of nonlinear integral Volterra models already considered in [20], on which the method described in this paper is applied for illustration in section 6.

The idea of decoupling the linear and nonlinear parts of the model to identify them is not new: relaxation approaches (such as [8]) and other less common methods [21] are based on this attractive idea. However, relaxation approaches require the use of an iterative algorithm which can have problems of convergence [22, 23]. As for the method proposed in [21], it is only suitable for Hammerstein models, and is not based on the same idea than the one presented in this paper. In [21] the decoupling between the linear and nonlinear parts of the model is indeed obtained for a particular input signal (Pseudo-Random Binary Signal) whereas in the present approach, the decoupling is based on the properties of static functions and therefore does not depend on the input.

To present the method, the paper has been organized as follows. In section 2, the model and the identification problem are introduced and some concrete examples are given. Then, the method is presented in details in section 3. In section 4, we propose an analysis of the method previously described. Then, a simple example is considered in section 5 to show how to implement the method in concrete situations. In section 6, the method is applied to a class of nonlinear Volterra models and finally implemented on a concrete example to highlight the relevance of the approach. On this last example, the proposed method is compared to a classical relaxation approach [24, 8].

Note that section 4 is rather technical but can be skipped without loss of information for the understanding of sections 5 and 6 or for the implementation of the method. For the sake of clarity, the notations used in the paper are defined in Appendix A. Only the technical details which are specific to the proposed method are given in Appendix B.

2 Presentation of the identification problem

2.1 Identification model under consideration

Let \mathbf{X} , \mathbf{Z} be some Banach spaces, F a separable Hilbert space and Ω a non-empty compact subset of \mathbb{R}^n . Let Ω^i , $i = 1 : I$ be I non-empty compact subsets of \mathbb{R}^n and \mathbf{X}_0 a closed bounded subset of \mathbf{X} ; most of the time, Ω^i will be time intervals of the form $[t_0^i, t_f^i] \subset \mathbb{R}$.

Given “admissible” subsets $\mathcal{X} \subset \prod_{i=1}^I C_{\mathbf{X}_0}^0(\Omega^i)$ and $\mathcal{Z} \subset \prod_{i=1}^I L_{\mathbf{Z}}^\infty(\Omega^i)$, we consider the abstract model:

$$g \circ x + Ay + z = 0, \quad (1)$$

with $g : \mathbf{X}_0 \rightarrow \mathbf{Z}$ a Borel measurable function, $x = (x^i) \in \mathcal{X}$, $y \in F$, $z = (z^i) \in \mathcal{Z}$, and A a linear operator:

$$A = (A^i) \in \prod_{i=1}^I \mathcal{L}(F, L_{\mathbf{Z}}^\infty(\Omega^i)). \quad (2)$$

Because $z^i \in L_{\mathbf{Z}}^\infty(\Omega^i)$, equation (1) must be understood in the almost everywhere sense:

$$\forall i, g(x^i(t)) + (A^i y)(t) + z^i(t) = 0 \quad t\text{-a.e. in } \Omega^i. \quad (3)$$

We suppose in addition that g is “almost everywhere continuous” in the following sense:

$$\forall x \in \mathcal{X}, \forall i, x^{i-1}(W) \text{ is Lebesgue negligible, where } W = \{\xi \in \mathbf{X}_0 \text{ s.t. } g \text{ is discontinuous at } \xi\}. \quad (4)$$

In practice, the condition (4) (as well as the choice of admissible subsets \mathcal{X} , \mathcal{Z}), can be analyzed by taking into account available information on the real system under study (see in particular remarks 14). From a pragmatic point of view, this hypothesis can also be postulated *a priori* and then confirmed or refuted *a posteriori*, according to the obtained identification results.

2.2 Formulation of the identification problem

The problem under consideration in this paper is introduced here-after:

Problem 1 (Identification problem) *Given model (1), consider some measurement trajectories x_m^i, z_m^i, A_m^i , $i = 1 : I$, and denote $x_m := (x_m^i)$, $z_m := (z_m^i)$ and $A_m := (A_m^i)$ (A_m is an operatorial trajectory).*

The problem under consideration is to estimate both y and g , supposed to be unknown, from the measurement trajectories x_m^i, z_m^i, A_m^i , $i = 1 : I$, in such a way that $(x_m, z_m, A_m y)$ is close to a solution (x, z, Ay) of (1).

Remark 2 • *In practice, the trajectories x_m^i, z_m^i, A_m^i can be deduced from both physical measurements (with possible noise) and a priori knowledge, with possible numerical computations (see section 6 for an example).*

- *Thanks to the abstract form of model (1), the quantity y to be identified can be of various type: for example, it can be the kernel h of a convolution operator (or another characteristic functional object, see section 6), but also an unknown input of the system (e.g. input, perturbations, etc.) as in section 5.*

2.3 Examples of application

The identification model (1) under consideration in this paper is defined in an abstract framework and under weak hypotheses, so that it enables to identify a wide class of dynamic models. In

particular, it can be used for the identification of a generic class of nonlinear Volterra models of the form:

$$H(\partial_t)X = f(X) + u, \quad (5)$$

where u , X are some real valued trajectories, f is a nonlinear function, and $H(\partial_t)$ denotes the linear causal convolution operator associated with the transfer function H and the impulse response h , that is the operator defined by:

$$\forall X \text{ such that } X|_{\mathbb{R}^{*-}} = 0, [H(\partial_t)X](t) = (h * X)(t) := \int_0^t h(t-s)X(s)ds. \quad (6)$$

Model (5) is of the form (1) with, for example:

$$x = X, y = h, z = -u, g = -f, A : y \mapsto y * X. \quad (7)$$

Other definitions of A and y can be considered, depending on the quantities to be identified (see examples of sections 5 and 6).

Two concrete examples of such models are given here-after; many others can be found in the literature (e.g. in [25, 26, 27, 28]). This class of models is considered for identification in section 6 where the method is numerically implemented on the concrete example 4.

Example 3 In [29], a Volterra model has been proposed to describe, in suitable thermodynamic conditions, the evolution of the radius r of a spherical flame initiated by a source e at point 0 in a mixture of reactive species. The model is written:

$$r \partial_t^{1/2} r = 2r \ln(r) + 2e, \quad (8)$$

where $\partial_t^{1/2}$ is the non rational operator with transfer function $p^{1/2}$. This model is of the form (5) with $X = r$, $u = 2 \frac{e}{r}$, $f = 2 \ln$ and $H(\partial_t) = \partial_t^{1/2}$.

Example 4 In [30], a dynamic model has been proposed for the electrical behavior of a capacitor realized in cubic Perovskite $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$. The system includes three components as shown in Figure 1: a resistor r in series with a capacitor of impedance Z_c in parallel with a nonlinear conductance: $I_G = G(V_c) \times V_c$.

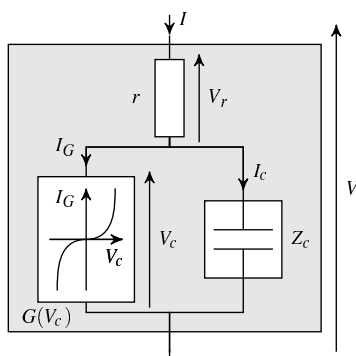


Figure 1: **System of example 4:** structure of the capacitor realized in cubic Perovskite $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$.

The model is written as follows:

$$V = Z_c(\partial_t)(I - G(V - rI) \times (V - rI)) + rI, \quad (9)$$

$$\Leftrightarrow Z_c(\partial_t)^{-1}(V - rI) = -G(V - rI) \times (V - rI) + I, \quad (10)$$

where V and I are the voltage and the current, G is the non linear conductance function, and $Z_c(\partial_t)$ is the impedance operator whose (non rational) transfer function is of the form $Z_c(p) = \frac{(1+\tau p)^\beta}{C p}$ with $C, \tau > 0$ and $0 < \beta < 1$. This model is of the form (5) with $X = V - rI$, $u = I$, $f : X \mapsto -G(X)X$ and $H(\partial_t) = Z_c(\partial_t)^{-1}$.

Finally, one can note that the identification model (1) (and thus the identification method proposed in this paper) can be used to identify nonlinear and dynamic linear unknown components of some dynamic inclusions models like [31, 32, 33]. Indeed, considering a function g discontinuous at some points $\xi \in \mathbf{X}_0$, if we denote \tilde{g} a set-valued map defined by: $\tilde{g}(\xi) \subset \mathbf{Z}$ for all $\xi \in \mathbf{X}_0$ and $\tilde{g}(\xi) = \{g(\xi)\}$ if g is continuous at point ξ , then models such as:

$$\tilde{g}(x^i(t)) \ni -(A^i y)(t) - z^i(t) \quad (11)$$

can straightforwardly be identified using the formulation (1), assuming that hypothesis (4) on measured trajectories is satisfied. This remark applies also for the particular case of Volterra models (5) which can be considered as dynamic inclusions when f is a discontinuous function; a concrete example of identification of such a ‘‘pseudo-differential inclusion’’ is treated in section 6.4.5.

2.4 Technical comments about model (1)

Some comments about model (1) are given here-after.

- First, the assumption that x^i has values in a bounded set $\mathbf{X}_0 \subset \mathbf{X}$ is justified by the fact that necessarily, for a given x (namely the one which is used, via its measurement x_m , for identification), the set $\text{Im}(x) := \bigcup_{i=1}^I x^i(\Omega^i)$ is compact in \mathbf{X} . Indeed, $x^i(\Omega^i)$ is compact because x^i is a continuous function on the compact set Ω^i . In addition, because function g must be identified on \mathbf{X}_0 , this set must be chosen such that $\text{Im}(x)$ is close to \mathbf{X}_0 in the sense:

$$\sup_{\eta \in \mathbf{X}_0} \inf_{\xi \in \text{Im}(x)} \|\xi - \eta\|_{\mathbf{X}} \leq \alpha, \quad (12)$$

with $\alpha > 0$ small enough.

In practice, the only available data will be the measurement x_m . Assume that the measurement noise belongs to $\prod_i C_{\mathbf{X}}^0(\Omega^i)$. As usual, some pre-filtering of the data can be required to satisfy this property, namely when measurement noises are white. Under this assumption, and if we suppose that $\|x_m - x\| \leq \frac{\alpha}{2}$, then (12) can be achieved by taking \mathbf{X}_0 as follows:

$$\mathbf{X}_0 = \bigcup_{\xi \in \text{Im}(x_m)} B_{\mathbf{X}}(\xi, \beta), \quad \beta > \frac{\alpha}{2}, \quad (13)$$

where $B_{\mathbf{X}}(\xi, \beta)$ denotes the closed ball in \mathbf{X} , with center ξ and radius β .

In addition, to make possible a significant identification of g , it can be expected that \mathbf{X}_0 contains some subset $\mathbf{C} \subset \mathbf{X}$ given *a priori*. This will be insured if x_m is such that:

$$\sup_{\eta \in \mathbf{C}} \inf_{\xi \in \text{Im}(x_m)} \|\xi - \eta\|_{\mathbf{X}} \leq \frac{\alpha}{2}. \quad (14)$$

Roughly speaking, this simply requires the obvious property that the trajectories x^i are ‘‘rich enough’’ to well fill the set \mathbf{C} .

- The aim is to propose, in an abstract framework, an identification method which enables to find a couple (y, g) such that model (1) is a good representation of the system under consideration. So, uniqueness of the couple (y, g) with respect to given (x, z, A) is not required at this stage. In fact, without any additional hypothesis, the solution is not always unique (see example of section 5 and remark 22). In concrete situations however, this property is in general quite desirable and it will be necessary to study it specifically (as well, possibly, as the well-posedness of the identified model).

3 Description of the identification method

In this section, we present the identification method. Only the concrete technical steps of the method are detailed here-after; the mathematical analysis of the method is done in section 4.

3.1 Principle of the approach

When no *a priori* information about function g is available, the problem of identifying g and y simultaneously can present serious difficulties, in particular if g is highly nonlinear (for example non differentiable, discontinuous or even weakly singular). In such cases indeed, linear parameterizations of g (whose advantage is to transform (1) into a linear problem) of the form $g = \sum_k a_k \mathbf{g}_k$ with (\mathbf{g}_k) a topological basis chosen *a priori*, may not be accurate enough except if the number of a_k is quite large; but on the other hand, if this number is too large, the estimation is generally bad due to excessive sensitivity to the measurement noises.

However, one can note that thanks to the linearity of operator A , if the function g (and then $g \circ x$) was known then y could be estimated more easily: indeed the identification problem could then be written in the sense of a classical least-squares problem, that is, with $b_m = g(x_m) + z_m$ and G a suitable Hilbert space of \mathbf{Z} -valued trajectories:

$$\min_{y \in F} \|A_m y + b_m\|_G^2, \quad (15)$$

the solution of which is formally given by the pseudo-inverse of operator A_m [34].

The method proposed in this paper takes advantage of this remark by cancelling the nonlinear term $g \circ x$ by means of a suitable operatorial transformation of model (1), allowing to estimate y alone. Under some weak hypotheses, frequently satisfied in practice, the so-obtained new model indeed reveals itself sufficient for a good estimation of y .

Finally, once an estimate \hat{y} of y has been computed, it becomes easy to estimate *a posteriori* the function g using standard methods applied to the available data, including the new ones deduced from the knowledge of \hat{y} , namely $A_m \hat{y}$.

The principle of the operatorial cancellation of $g \circ x$ is given here after. Assume that the measured data are noise-free, and consider two trajectories x^i and x^j with $i, j \in \{1 : I\}$. If we find two instants τ and t such that:

$$x^i(t) = x^j(\tau), \quad (16)$$

then, as g is a static function, it follows that $g(x^i(t)) = g(x^j(\tau))$. By subtraction, we so get from the i^{th} and j^{th} equations of (1) the following relationship:

$$(Ay^i)(t) - (Ay^j)(\tau) + z^i(t) - z^j(\tau) = 0. \quad (17)$$

If we write it for all couples (t, τ) such that $x^i(t) = x^j(\tau)$ with $i, j \in \{1 : I\}$, then we get a set of equations in which the static function g has been cancelled. If this set is moreover rich enough, then we can use it to identify y easily as the equations are linear with respect to y . Classical identification methods such as least-squares ones can then be used.

In practice, the method can be decomposed in three steps:

1. Step 1: Construction of sets of couples (t, τ) such that $g(x^i(t)) = g(x^j(\tau))$ (or, more precisely, such that $g(x^i(t)) - g(x^j(\tau))$ is negligible) for all $i, j \in \{1 : I\}$.
2. Step 2: Identification of y from the set of equations (17) which are independent of the unknown function g and linear with respect to y .
3. Step 3: Identification of g from model (1), using the identified value of y .

The details of each step is given here after. In Figure 2, a graphical abstract gives an overview of the identification method and summarizes its different steps.

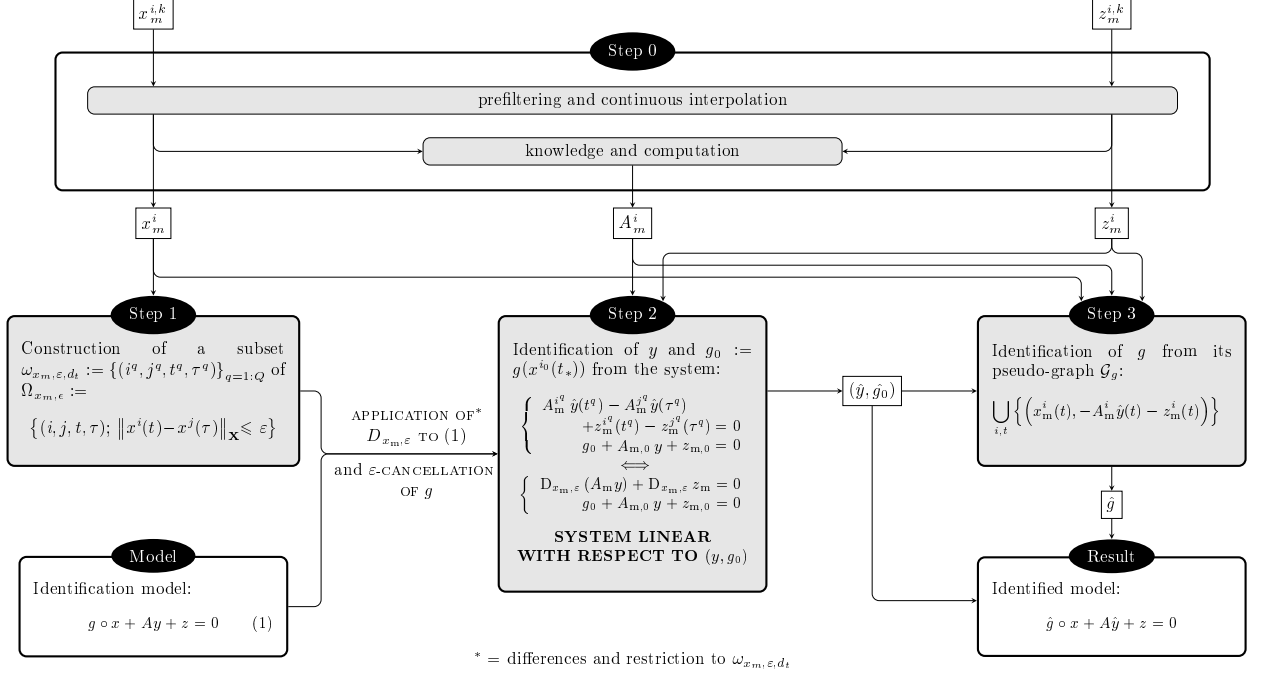


Figure 2: Graphical abstract of the identification method.

3.2 Step 1: Construction of the set $\Omega_{x,\varepsilon}$

As explained in the introduction of section 3, the idea of the method is to consider all $(i, j) \in \{1 : I\}^2$ and all couples (t, τ) such that $x^i(t) = x^j(\tau)$, in order to cancel the term $g(x^i(t)) - g(x^j(\tau))$ in a set of equations which can then be used for the independent identification of y . To this end, we consider $\Omega_{x,0}$ the set of indexes (i, j) and times (t, τ) such that the above mentioned cancellation of the nonlinear term $g(x)$ occurs; such a set is defined by:

$$\Omega_{x,0} := \{(i, j, t, \tau) \in \{1 : I\}^2 \times \Omega^{i,j}; x^i(t) = x^j(\tau)\} \text{ where } \Omega^{i,j} := \Omega^i \times \Omega^j. \quad (18)$$

However, requiring the perfect equality between $x^i(t)$ and $x^j(\tau)$ is not relevant in practice, for several reasons. First, as the measurement trajectories are noisy, we are not able to test the equality between two values of some trajectories, but only between two measurements of these trajectories. Moreover, when the dimension of the space \mathbf{Z} is greater than 2, the property $x^i(t) = x^j(\tau)$ in general becomes quite rare (or even with null probability), except at the trivial and not significant points defined by $i = j, \tau = t$. As a consequence, to overcome these problems, it is more relevant to consider the sets of couples (t, τ) such that $x^i(t)$ is “close to” $x^j(\tau)$. Considering $\varepsilon \geq 0$, such sets are defined by:

$$\Omega_{x,\varepsilon}^{i,j} := \{(t, \tau) \in \Omega^{i,j}; \|x^i(t) - x^j(\tau)\|_{\mathbf{X}} \leq \varepsilon\}. \quad (19)$$

Roughly speaking, $\Omega_{x,\varepsilon}^{i,j}$ contains all couples (t, τ) for which the nonlinear term g is “almost-cancelled” after subtraction of the i^{th} equation of (1) from the j^{th} one at respective times t and τ : in the sequel, we will call this operation “ ε -cancellation”. For convenience, and following the notation introduced in (18), we denote in the sequel:

$$\Omega_{x,\varepsilon} := \bigcup_{(i,j) \in \mathcal{I}_{x,\varepsilon}} (\{(i, j)\} \times \Omega_{x,\varepsilon}^{i,j}), \quad (20)$$

with $\mathcal{I}_{x,\varepsilon} := \{(i, j) \in \{1 : I\}^2; \Omega_{x,\varepsilon}^{i,j} \neq \emptyset\}$.

As an illustration, one can find in Figure 3 an example of some couples $(t, \tau) \in \Omega_{x,\varepsilon}^{i,j}$; the construction of the whole set $\Omega_{x,\varepsilon}$ is obtained when the variable x_0 of Figure 3 takes all the values in

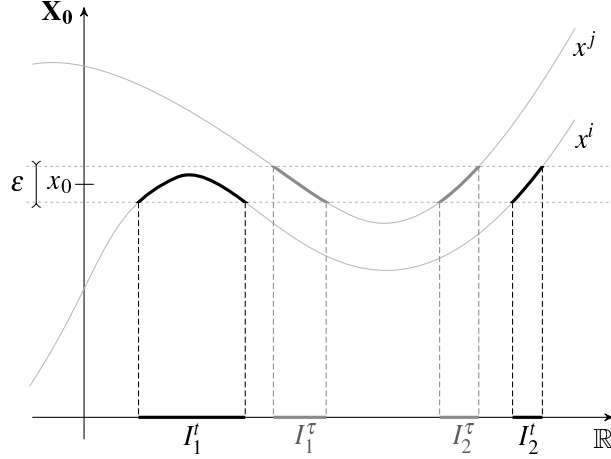


Figure 3: **Example of couples** $(t, \tau) \in \Omega_{x,\varepsilon}^{i,j}$. The admissible couples (t, τ) are obtained by taking t and τ in the union of **black sets** $I_1^t \cup I_2^t$ and the union of **gray sets** $I_1^\tau \cup I_2^\tau$ respectively.

\mathbf{X}_0 . This construction process can be used to build some subsets of $\Omega_{x,\varepsilon}^{i,j}$ in practice (see Appendix B).

In the case $\mathbf{X}_0 = \mathbb{R}$, we can note that another expression of the sets $\Omega_{x,\varepsilon}^{i,j}$ can be given. Indeed, the union of black sets (I_1^t and I_2^t for time t) and the union of gray sets (I_1^τ and I_2^τ for time τ) highlighted in Figure 3 can respectively be expressed by $x^{i-1}([x_0 - \frac{\varepsilon}{2}, x_0 + \frac{\varepsilon}{2}])$ and $x^{j-1}([x_0 - \frac{\varepsilon}{2}, x_0 + \frac{\varepsilon}{2}])$. Thus, the set of all couples of instants (t, τ) of Figure 3 are given by the cartesian product of these two sets, which leads to the following expression of $\Omega_{x,\varepsilon}^{i,j}$:

$$\Omega_{x,\varepsilon}^{i,j} = \bigcup_{x_0 \in \mathbf{X}_0} x^{i-1} \left([x_0 - \frac{\varepsilon}{2}, x_0 + \frac{\varepsilon}{2}] \right) \times x^{j-1} \left([x_0 - \frac{\varepsilon}{2}, x_0 + \frac{\varepsilon}{2}] \right). \quad (21)$$

Finally, note that the sets $\Omega_{x,\varepsilon}^{i,i}$ are never empty because they contain at least all couples of the form (t, t) . Furthermore, if $\varepsilon' \leq \varepsilon$, then $\Omega_{x,\varepsilon'}^{i,j} \subset \Omega_{x,\varepsilon}^{i,j}$. We also have the following result:

Proposition 5 *For any i, j , the set $\Omega_{x,\varepsilon}^{i,j}$ is compact. If $\varepsilon > 0$ and $\Omega_{x,\varepsilon}^{i,j} \neq \emptyset$, then $\Omega_{x,\varepsilon}^{i,j}$ has a non-empty interior (and therefore has non-null Lebesgue measure). Furthermore, $\bigcap_{\varepsilon > 0} \Omega_{x,\varepsilon}^{i,j} = \Omega_{x,0}^{i,j}$.*

Proof. Thanks to the continuity of $\|D^{i,j}x\| : \Omega^{i,j} \rightarrow \mathbb{R}$ and the two properties: $\Omega_{x,\varepsilon'}^{i,j} \subset \mathbb{R}^{2n}$ and $\Omega_{x,\varepsilon'}^{i,j} \subset \Omega_{x,\varepsilon}^{i,j}$ if $\varepsilon' \leq \varepsilon$. \square

3.3 Step 2: Identification of the linear component y

From the definition of the set $\Omega_{x,\varepsilon}$, by subtracting the i^{th} equation of (1) evaluated at time t from the j^{th} equation of (1) evaluated at time τ and by neglecting the term $g(x^i(t)) - g(x^j(\tau))$ for all $(i, j, t, \tau) \in \Omega_{x,\varepsilon}$, we get the set of equations:

$$A^i y(t) - A^j y(\tau) + z^i(t) - z^j(\tau) = 0, \quad \forall (i, j, t, \tau) \in \Omega_{x,\varepsilon}. \quad (22)$$

To get a more synthetic expression, for any $(i, j) \in \mathcal{I}_{x,\varepsilon}$ we denote by $D_{x,\varepsilon}^{i,j}$ the operator defined on any function $v \in \prod_i L_V^\infty(\Omega^i)$ by:

$$D_{x,\varepsilon}^{i,j} v : \begin{array}{l} \Omega_{x,\varepsilon}^{i,j} \rightarrow \mathbf{V} \\ (t, \tau) \mapsto v^i(t) - v^j(\tau); \end{array} \quad (23)$$

thus, each equation of the form (22) can simply be written $D_{x,\varepsilon}^{i,j}(Ay) + D_{x,\varepsilon}^{i,j}z = 0$. By denoting $D_{x,\varepsilon} := \left(D_{x,\varepsilon}^{i,j} \right)_{(i,j) \in \mathcal{I}_{x,\varepsilon}}$ we finally get the model:

$$D_{x,\varepsilon}(Ay) + D_{x,\varepsilon}z = 0, \quad (24)$$

which presents two advantages: first it does not involve the nonlinear function g any longer, and second it is linear with respect to the remaining unknown y .

However, as the equations (22) were obtained by subtractions, the constant terms of the model have also been cancelled, even those who are related to the unknown y . To mitigate this problem, we need to add at least one equation to (24). More explanation about this additive equation are given in section 4. This additional equation will be of the form:

$$g(x^{i_0}(t_*)) + (A^{i_0}y)(t_*) + z^{i_0}(t_*) = 0, \quad (25)$$

with $t_* \in \Omega^{i_0}$, and $i_0 \in \{1 : I\}$ arbitrarily chosen. For simplicity, we will denote in the sequel $g_0 = g(x^{i_0}(t_*)) \in \mathbf{Z}$, $A_0y = (A^{i_0}y)(t_*) \in \mathbf{Z}$, and $z_0 = z^{i_0}(t_*) \in \mathbf{Z}$.

Finally, we consider the following system built from (1):

$$\begin{cases} D_{x,\varepsilon}(Ay) + D_{x,\varepsilon}z = 0 \\ g_0 + A_0y + z_0 = 0, \end{cases} \quad (26)$$

which is independent from the nonlinear function g and linear with respect to the unknowns y and g_0 . The identification of y and g_0 from measurements x_m , z_m and A_m is thus simplified. We simply rewrite the system (26) under the form:

$$\phi_{x,A}^\varepsilon \cdot \theta = b_{x,z}^\varepsilon, \quad (27)$$

with:

$$\theta := (y, g_0), \quad b_{x,z}^\varepsilon := (-D_{x,\varepsilon}z, -z_0), \quad (28)$$

and $\phi_{x,A}^\varepsilon$ the linear continuous operator defined by:

$$\phi_{x,A}^\varepsilon : \theta = (y, g_0) \mapsto (D_{x,\varepsilon}(Ay), g_0 + A_0y). \quad (29)$$

Then, assuming that x_m , A_m and z_m are sufficiently close to x , A and z (see section 4), we consider the following equation of linear regression form:

$$b_{x_m, z_m}^\varepsilon = \phi_{x_m, A_m}^\varepsilon \cdot \theta + e^\varepsilon(\theta), \quad (30)$$

where $e^\varepsilon(\theta)$ is the so-called equation error. An estimate $\hat{\theta}$ of the unknown parameter $\theta = (y, g_0)$ can finally be obtained by classical error-in-equation identification methods [4, 35] which take advantage of the linearity of the model with respect to the unknown parameters.

For example, if \mathbf{Z} is a separable Hilbert space, we can consider the least-squares estimator, solution of:

$$\hat{\theta} = \arg \min_{\theta \in F \times \mathbf{Z}} \|e^\varepsilon(\theta)\|_{G^\varepsilon \times \mathbf{Z}}^2, \quad (31)$$

with G^ε a suitable separable Hilbert space, such as:

$$G^\varepsilon = \prod_{(i,j) \in \mathcal{I}_{x_m, \varepsilon}} L_{\mathbf{Z}}^2(\Omega_{x_m, \varepsilon}^{i,j}), \quad (32)$$

which is such that $L_{\mathbf{Z}}^\infty(\Omega_{x_m, \varepsilon}^{i,j}) \hookrightarrow L_{\mathbf{Z}}^2(\Omega_{x_m, \varepsilon}^{i,j})$. This estimator is formally expressed by:

$$\hat{\theta} = (\hat{y}, \hat{g}_0) = (\phi_{x_m, A_m}^\varepsilon)^\dagger b_{x_m, z_m}^\varepsilon, \quad (33)$$

where $(\phi_{x_m, A_m}^\varepsilon)^\dagger$ is the pseudo-inverse of $\phi_{x_m, A_m}^\varepsilon$ [34].

Remark 6 As usual, when the problem (31) is ill conditioned or when the solution is not unique, some “regularized” problems can advantageously be considered, as the following one:

$$\hat{\theta} = \arg \min_{\theta \in F \times \mathbf{Z}} \{ \|e^\varepsilon(\theta)\|_{G^\varepsilon \times \mathbf{Z}}^2 + \eta \|\theta\|_{F \times \mathbf{Z}}^2 \}, \quad (34)$$

where $\eta > 0$ is a small parameter.

Note that, as the linear operator $\phi_{x_m, A_m}^\varepsilon$ depends on the noise, some of the classical estimators (as the least-squares one mentioned previously) can be biased [1]. Some bias reduction methods [35, 36, 37] can be used to mitigate this problem.

Remark 7 *This identification method can easily be extended, up to simple technical adaptation, to the case where function g is different from one trajectory to the other, that is $g \circ x := (g^i \circ x^i)_{i=1:I}$. In such cases, the cancellation of the function g is only possible for $j = i$; thus only the sets $\Omega_{x, \varepsilon}^{i,i}$ will be used which implies that the difference operator $D_{x, \varepsilon}$ will be given by $(D_{x, \varepsilon}^{i,i})_{i=1:I}$.*

3.4 Step 3: Identification of the nonlinear function g

Under the *a priori* hypothesis that the data x_m , A_m , z_m allow to get a good identification of y (that is $\|\hat{y} - y\|_F$ small enough), then, from continuity properties, we have for any $(i, j, t, \tau) \in \Omega_{x_m, \varepsilon}$:

$$A_m^i \hat{y}(t) - A_m^j \hat{y}(\tau) + z_m^i(t) - z_m^j(\tau) \simeq 0. \quad (35)$$

Moreover, although g is yet unknown at this stage (except at point $x_0 = x^{i_0}(t_*)$ for which we have an estimate \hat{g}_0 of its value), we can expect to have, for any $i \in \{1 : I\}$ and any $t \in \Omega^i$, $A_m^i \hat{y}(t) + g(x_m^i(t)) + z_m^i(t) \simeq 0$, property which can be written under the form:

$$g(x_m^i(t)) \simeq -A_m^i \hat{y}(t) - z_m^i(t), \forall t \in \Omega^i, \forall i = 1 : I. \quad (36)$$

From this last equation, identification of the nonlinear function g can be achieved with classical regression methods implemented on the following “pseudo-graph” of g , deduced from the available data:

$$\mathcal{G}_g = \bigcup_{i=\{1:I\}, t \in \Omega^i} \{(x_m^i(t), -A_m^i \hat{y}(t) - z_m^i(t))\} \subset \mathbf{X} \times \mathbf{Z}. \quad (37)$$

This will be highlighted on numerical examples in sections 5 and 6.

3.5 Advantages and drawbacks of the method

The first advantage of the method is that it allows to have access (after the identification of the linear unknown y) to a pseudo-graph of the unknown function g before its parameterization for the identification. Thus, the parameterization of the function g can be adapted to the shape of the pseudo-graph which is a great advantage especially when the function g is highly nonlinear, non differentiable, discontinuous or even weakly singular.

The identification model (1) and then the identification method presented in this paper can also easily be extended to the case where the function g varies from one trajectory to the other simply by setting $g \circ x := (g^i \circ x^i)_{i=1:I}$ in model (1) and by simple technical adaptation of the method (see remark 7 for more details). It can be useful if the measured trajectories have been obtained under different conditions which only impact the nonlinear term of the system (experiments made at different temperatures for example).

Compared to relaxation approaches (such as [8]), the decoupling between the identification of the linear component y and the nonlinear function g does not imply the use of an iterative algorithm and the associated problems of convergence [22, 23]. The method can also be applied regardless of the type of the input signal (which is not the case for other methods [21]).

However, the method has also some drawbacks. It requires the construction of the set (or more precisely of a finite subset of) $\Omega_{x_m, \varepsilon}$ which is not trivial. In particular, when the measured trajectories are noisy, the set $\Omega_{x_m, \varepsilon}$ can become very large compared to the noise-free case, and the pre-filtering of the data then becomes necessary. A lot of questions remain therefore open, especially about the best way (in terms of computation time and quality of the results) to do the pre-filtering (see for example [38]), and to determine the couples of instants (t, τ) that will be used in the identification process.

Finally, the numerical implementation of the method in the case of systems of high dimension (that is when the spaces \mathbf{X} and \mathbf{Z} are of high dimensions) will need extensive studies to deal with the numerical complexity of the problem.

4 Analysis of the method

In this section, we analyze the identification method previously introduced. In a first part, we discuss in details all the steps leading to model (26), starting from (1), and give some results to justify or explain them. Then, we study the impact of the measurement noise on the method. Finally, we give some results useful for the practical application of the method.

Note that the reader who is not interested in the mathematical analysis of the method, which is rather technical, can skip this section without missing any information necessary to understand the sections 5 and 6, and to implement the method.

4.1 Establishment of model (26)

The transformation process leading to model (26), starting from (1), is composed of three steps:

(i) subtraction of the j^{th} equation of (1) evaluated at time τ from the i^{th} equation of (1) evaluated at time t for all $(i, j, t, \tau) \in \{1 : I\}^2 \times \Omega^{i,j}$;

(ii) restriction of the so-obtained “difference model” to subsets $\Omega_{x,0}^{i,j}$ of $\Omega^{i,j}$ (cancellation of g) and then to subsets $\Omega_{x,\varepsilon}^{i,j}$;

(iii) approximation by ε -cancellation of g .

4.1.1 Difference model and equivalence with model (1).

We first introduce an operator that we call *difference operator*, which is denoted D and defined here after.

Definition 8 For any given Banach space \mathbf{V} , we consider the operator D defined by:

$$D : \prod_{i=1}^I L_{\mathbf{V}}^{\infty}(\Omega^i) \longrightarrow \prod_{i,j=1}^I L_{\mathbf{V}}^{\infty}(\Omega^{i,j}) \quad (38)$$

$$v = (v^i) \longmapsto Dv = (D^{i,j}v) \quad \text{with } D^{i,j}v(t, \tau) := v^i(t) - v^j(\tau).$$

The operator D is obviously linear; thanks to the continuity of the $+$ operation in Banach spaces, we have:

$$D \in \mathcal{L} \left(\prod_i L_{\mathbf{V}}^{\infty}(\Omega^i), \prod_{i,j} L_{\mathbf{V}}^{\infty}(\Omega^{i,j}) \right), \quad (39)$$

and, with D_{C^0} denoting the restriction of D to the subspace $\prod_i C_{\mathbf{V}}^0(\Omega^i)$:

$$D_{C^0} \in \mathcal{L} \left(\prod_i C_{\mathbf{V}}^0(\Omega^i), \prod_{i,j} C_{\mathbf{V}}^0(\Omega^{i,j}) \right). \quad (40)$$

In the sequel, both of these operators will be denoted D for simplicity.

By applying operator D to both members of (1), we get I^2 difference equations that can be summarized by the following equation in $\prod_{i,j} L_{\mathbf{Z}}^{\infty}(\Omega^{i,j})$:

$$D(g \circ x) + D(Ay) + Dz = 0. \quad (41)$$

At this stage, the equivalence between models (41) and (1) is lost, because operator D is mainly based on differences; the equivalence is recovered by considering an additional equation of the form (25), as stated in the following result:

Proposition 9 Model (1) is equivalent to the following system in $\prod_{i,j} L_{\mathbf{Z}}^{\infty}(\Omega^{i,j}) \times \mathbf{Z}$:

$$\begin{cases} D(g \circ x) + D(Ay) + Dz = 0 \\ g_0 + A_0y + z_0 = 0, \end{cases} \quad (42)$$

with $g_0 := g(x^{i_0}(t_*)) \in \mathbf{Z}$, $A_0y := (A^{i_0}y)(t_*) \in \mathbf{Z}$, $z_0 := z^{i_0}(t_*) \in \mathbf{Z}$, where $i_0 \in \{1 : I\}$ and $t_* \in \Omega^{i_0}$ are arbitrarily chosen.

Proof. The implication (1) \Rightarrow (42) is obvious. Let us show that (42) \Rightarrow (1). We have for any $(i, j) \in \{1 : I\}^2$ and any $(t, \tau) \in \Omega^{i,j}$:

$$g(x^i(t)) - g(x^j(\tau)) + A^i y(t) - A^j y(\tau) + z^i(t) - z^j(\tau) = 0. \quad (43)$$

In particular, for $j = i_0$ and $\tau = t_*$, we get:

$$\forall t \in \Omega^i, \quad g(x^i(t)) - g(x^{i_0}(t_*)) + A^i y(t) - A^{i_0} y(t_*) + z^i(t) - z^{i_0}(t_*) = 0. \quad (44)$$

As $(g \circ x^{i_0})(t_*) + A^{i_0} y(t_*) + z^{i_0}(t_*) = 0$, it follows that $\forall t \in \Omega^i, \quad g \circ x^i(t) + A^i y(t) + z^i(t) = 0$. \square

Thus, we have built a new model equivalent to (1) which, restricted to some suitable subsets of $\Omega^{i,j}$, can provide a set of equations independent from g and so well adapted to the identification of y alone. The restriction operations to the subsets $\Omega_{x,\varepsilon}^{i,j}$ of $\Omega^{i,j}$ are studied here-after.

4.1.2 Restriction to subsets $\Omega_{x,\varepsilon}^{i,j}$.

We introduce the following “restriction operators” to subsets $\Omega_{x,\varepsilon}^{i,j} \subset \Omega^{i,j}$:

$$\begin{aligned} R_{\Omega_{x,\varepsilon}^{i,j}} : L_{\mathbf{V}}^\infty(\Omega^{i,j}) &\longrightarrow L_{\mathbf{V}}^\infty(\Omega_{x,\varepsilon}^{i,j}) \\ w &\longmapsto w|_{\Omega_{x,\varepsilon}^{i,j}}. \end{aligned} \quad (45)$$

By composition of operator $D^{i,j}$ with $R_{\Omega_{x,\varepsilon}^{i,j}}$, we get the operator $D_{x,\varepsilon}^{i,j}$, introduced in section 3, and defined by:

Definition 10 *The operator $D_{x,\varepsilon}^{i,j}$ is defined, for any $(i, j) \in \{1 : I\}^2$ such that $\Omega_{x,\varepsilon}^{i,j} \neq \emptyset$, by:*

$$D_{x,\varepsilon}^{i,j} = R_{\Omega_{x,\varepsilon}^{i,j}} \circ D^{i,j} = D^{i,j}(\cdot)|_{\Omega_{x,\varepsilon}^{i,j}} \quad (46)$$

By definition of the norm in L^∞ spaces, the operator $R_{\Omega_{x,\varepsilon}^{i,j}}$ is linear continuous; so, similarly to $D^{i,j}$, we have, if $\Omega_{x,\varepsilon}^{i,j} \neq \emptyset$:

$$D_{x,\varepsilon}^{i,j} \in \mathcal{L} \left(\prod_l L_{\mathbf{V}}^\infty(\Omega^l), L_{\mathbf{V}}^\infty(\Omega_{x,\varepsilon}^{i,j}) \right), \text{ and } D_{x,\varepsilon}^{i,j} \in \mathcal{L} \left(\prod_l C_{\mathbf{V}}^0(\Omega^l), C_{\mathbf{V}}^0(\Omega_{x,\varepsilon}^{i,j}) \right). \quad (47)$$

For convenience, we denote (with $\mathcal{I}_{x,\varepsilon}$ defined in section 3.2):

$$D_{x,\varepsilon} := (D_{x,\varepsilon}^{i,j}) \in \prod_{(i,j) \in \mathcal{I}_{x,\varepsilon}} \mathcal{L} \left(\prod_l L_{\mathbf{V}}^\infty(\Omega^l), L_{\mathbf{V}}^\infty(\Omega_{x,\varepsilon}^{i,j}) \right). \quad (48)$$

The application of operator $D_{x,0}$ to equation (1) leads to the restriction of model (42) to the subsets $\Omega_{x,0}^{i,j}$, whose interest comes from the following property of operator $D_{x,0}$:

Proposition 11 *Let \mathbf{X}_0 be a closed subset of \mathbf{X} . For any $x \in \prod_{i=1}^I C_{\mathbf{X}_0}^0(\Omega^i)$ and any function $g : \mathbf{X}_0 \rightarrow \mathbf{Z}$, we have:*

$$D_{x,0}(g \circ x) = 0 \quad (\text{and in particular: } D_{x,0} x = 0). \quad (49)$$

Proof. From definition (19) of sets $\Omega_{x,\varepsilon}^{i,j}$, we trivially get:

$$\forall i, j, \forall (t, \tau) \in \Omega_{x,0}^{i,j}, \quad \|(D^{i,j} x)(t, \tau)\|_{\mathbf{X}} \leq 0, \quad (50)$$

and then $(D^{i,j} x)(t, \tau) = 0$; this last result can also be written $x^i(t) = x^j(\tau)$, from which we finally deduce $g(x^i(t)) = g(x^j(\tau)) \Leftrightarrow (g(x^i))(t) = (g(x^j))(\tau)$, that is:

$$\forall i, j, \forall (t, \tau) \in \Omega_{x,0}^{i,j}, \quad (D^{i,j}(g \circ x))(t, \tau) = 0, \quad (51)$$

or equivalently $D_{x,0}(g \circ x) = 0$. \square

Thanks to this property, the application of operator $D_{x,0}$ to (42) leads to the following model, in $\prod_{\mathcal{I}_{x,0}} L_{\mathbf{Z}}^{\infty}(\Omega_{x,0}^{i,j}) \times \mathbf{Z}$:

$$\begin{cases} D_{x,0}(Ay) + D_{x,0}z = 0 \\ g_0 + A_0y + z_0 = 0. \end{cases} \quad (52)$$

The advantage of this new model is that the nonlinear function g has been *cancelled*, while the unknowns $y \in F$ and $g_0 \in \mathbf{Z}$ are *linearly* involved. Classical identification methods such as least-squares ones can therefore be used to identify y and g_0 .

However, as emphasized in paragraph 3.2, the set $\Omega_{x,0}$ cannot be built in practice because of the measurement noise. Indeed, we have:

$$\|x_m^i(t) - x_m^j(\tau)\| = 0 \not\Rightarrow \|x^i(t) - x^j(\tau)\| = 0. \quad (53)$$

On the other hand, it can be shown (see section 4.2) that by using only the measurement x_m , we can have access to some sets $\Omega_{x,\varepsilon}$ where $\varepsilon > 0$ depends on the noise magnitude. For these reasons, the restriction is done on the subset $\Omega_{x,\varepsilon}$ (with $\varepsilon > 0$) instead of $\Omega_{x,0}$, which leads to the following model:

$$\begin{cases} D_{x,\varepsilon}(g \circ x) + D_{x,\varepsilon}(Ay) + D_{x,\varepsilon}z = 0 \\ g_0 + A_0y + z_0 = 0. \end{cases} \quad (54)$$

By construction of $\Omega_{x,\varepsilon}$, we expect the term $D_{x,\varepsilon}(g \circ x)$ to be negligible (see paragraph 3.2), so that we can get a model linear with respect to y and g_0 on which the identification of y can be achieved easily.

The study of this ε -cancellation of $D_{x,\varepsilon}(g \circ x)$ is done in the next paragraph.

4.1.3 ε -cancellation of g .

The following proposition gives conditions under which the term $D_{x,\varepsilon}(g \circ x)$ can be considered as negligible.

Proposition 12 *Let \mathbf{X}_0 be a closed subset of \mathbf{X} . For any $x \in \prod_{i=1}^I C_{\mathbf{X}_0}^0(\Omega^i)$ and any function $g : \mathbf{X}_0 \rightarrow \mathbf{Z}$, the following properties hold:*

(i) $\|D_{x,\varepsilon}x\| \leq \varepsilon;$

(ii) *if g is continuous at points $\xi, \eta \in X_0$, then $D_{x,\varepsilon}^{i,j}(g \circ x)$ is continuous at any point of the set $[x^{i-1}(\{\xi\}) \times x^{j-1}(\{\eta\})] \cap \Omega_{x,\varepsilon}^{i,j}$; in particular, $D_{x,\varepsilon}^{i,j}(g \circ x)$ vanishes at any point of the set $x^{i-1}(\{\xi\}) \times x^{j-1}(\{\eta\}) \subset \Omega_{x,0}^{i,j} \subset \Omega_{x,\varepsilon}^{i,j}$, in the sense:*

$$\left[(t_\varepsilon, \tau_\varepsilon) \in \Omega_{x,\varepsilon}^{i,j} \xrightarrow{\varepsilon \rightarrow 0^+} (t, \tau) \in x^{i-1}(\{\xi\}) \times x^{j-1}(\{\eta\}) \right] \Rightarrow \left[(D_{x,\varepsilon}^{i,j}(g \circ x))(t_\varepsilon, \tau_\varepsilon) \xrightarrow{\varepsilon \rightarrow 0^+} 0 \right]; \quad (55)$$

(iii) *if g is k -Lipschitz, then $\|D_{x,\varepsilon}(g \circ x)\| \leq k\varepsilon$.*

Proof. We have:

(i) $\|D_{x,\varepsilon}(x)\| = \sup_{i,j} \sup_{(t,\tau) \in \Omega_{x,\varepsilon}^{i,j}} \|(D^{i,j}x)(t,\tau)\|_{\mathbf{X}} \leq \varepsilon$ by definition of $\Omega_{x,\varepsilon}^{i,j}$.

(ii) for any $(\tau, t) \in [x^{i-1}(\{\xi\}) \times x^{j-1}(\{\eta\})] \cap \Omega_{x,\varepsilon}^{i,j}$, the function g is continuous at points $x^i(t)$, $x^j(\tau)$; so, for any sequence $(t_n, \tau_n) \subset \Omega_{x,\varepsilon}^{i,j}$ such that $(t_n, \tau_n) \rightarrow (t, \tau)$, thanks to the continuity of x :

$$g(x^i(t_n)) \rightarrow g(x^i(t)), \quad g(x^j(\tau_n)) \rightarrow g(x^j(\tau)), \quad (56)$$

and so: $g(x^i(t_n)) - g(x^j(\tau_n)) \rightarrow g(x^i(t)) - g(x^j(\tau))$.

If $(t, \tau) \in x^{i-1}(\{\xi\}) \times x^{j-1}(\{\eta\})$, then necessarily $x^i(t) = x^j(\tau) = \xi$ and so, $(t, \tau) \in \Omega_{x,0}^{i,j}$; it follows from proposition 11 that:

$$(D_{x,\varepsilon}^{i,j}(g \circ x))(t_\varepsilon, \tau_\varepsilon) \rightarrow 0. \quad (57)$$

$$\begin{aligned}
\text{(iii)} \quad \|D_{x,\varepsilon}(g \circ x)\| &= \sup_{i,j} \sup_{\Omega_{x,\varepsilon}^{i,j}} \|(D^{i,j}(g \circ x))(t, \tau)\|_{\mathbf{Z}} = \sup_{i,j} \sup_{\Omega_{x,\varepsilon}^{i,j}} \|g(x^i(t)) - g(x^j(\tau))\|_{\mathbf{Z}} \\
&\leq \sup_{i,j} \sup_{\Omega_{x,\varepsilon}^{i,j}} k \|x^i(t) - x^j(\tau)\|_{\mathbf{X}} = k \sup_{i,j} \sup_{\Omega_{x,\varepsilon}^{i,j}} \|(D^{i,j}x)(t, \tau)\|_{\mathbf{X}} \leq k\varepsilon \text{ by definition of } \Omega_{x,\varepsilon}^{i,j}.
\end{aligned}$$

□

The continuity with respect to ε expressed in the property (ii) of proposition 12 ensures that the term $D_{x,\varepsilon}(g \circ x)$ will remain small enough, in such a way that (54) can be supposed “quasi equivalent” to the following equation in $\prod_{\mathcal{I}_{x,\varepsilon}} L_{\mathbf{Z}}^{\infty}(\Omega_{x,\varepsilon}^{i,j}) \times \mathbf{Z}$:

$$\begin{cases} D_{x,\varepsilon}(Ay) + D_{x,\varepsilon}z = 0 \\ g_0 + A_0y + z_0 = 0. \end{cases} \quad (58)$$

Quantitative questions about such a quasi equivalence will be deepened in a further work. In the present paper, we suppose that the approximate model (58) will be sufficient, at least in some situations, to get good identification results. In other words, we suppose that the identification bias induced by the above approximation will remain negligible. The legitimacy of this hypothesis will be highlighted by numerical examples presented in sections 5 and 6.4.

Remark 13 *The convergence deduced from (iii) is uniform, while we have in (ii) a weaker topology of simple convergence. For this reason, it can be expected that when g is not Lipschitz-continuous on \mathbf{X}_0 , numerical treatments can be more delicate. However, in many concrete situations, function g is almost everywhere continuous and Lipschitz-continuous on some subsets $\mathbf{X}_{\eta} \subset \mathbf{X}_0$, in such a way that the property $\|D_{x,\varepsilon}(\tilde{g}_{\eta} \circ x)\| \leq k_{\eta}\varepsilon$ holds for k_{η} -Lipschitz-continuous function \tilde{g}_{η} defined on \mathbf{X}_0 and such that $\tilde{g}_{\eta}|_{\mathbf{X}_{\eta}} = g|_{\mathbf{X}_{\eta}}$. We will see in section 6.4 that such a property (in addition to property (ii)) can be sufficient to get good identification results.*

Remark 14 *In proposition 12, the function g is not assumed to be continuous on the whole set \mathbf{X}_0 . Let $W = \{\xi \in \mathbf{X}_0 \text{ s.t. } g \text{ is discontinuous at } \xi\}$; if for any i , the (Lebesgue-measurable) set $x^{i-1}(W)$ is negligible in Ω^i (that is a null set in the sense of the Lebesgue measure), then the set of all the points of $\Omega_{x,\varepsilon}^{i,j}$ at which $D_{x,\varepsilon}^{i,j}(g \circ x)$ is possibly discontinuous, which can be written:*

$$V_{x,\varepsilon}^{i,j} = \{(t, \tau) \in \Omega_{x,\varepsilon}^{i,j}; x^i(t) \in W \text{ or } x^j(\tau) \in W\} \quad (59)$$

$$= \left[\left(x^{i-1}(W) \times \Omega^j \right) \cup \left(x^{j-1}(W) \times \Omega^i \right) \right] \cap \Omega_{x,\varepsilon}^{i,j}, \quad (60)$$

is obviously negligible in $\Omega_{x,\varepsilon}^{i,j}$ as soon as $\varepsilon > 0$ and $\Omega_{x,\varepsilon}^{i,j} \neq \emptyset$. Furthermore, $V_{x,0}^{i,j}$ rewrites:

$$V_{x,0}^{i,j} = \{(t, \tau) \in \Omega_{x,0}^{i,j}; x^i(\tau) = x^j(t) \in W\} = \left(x^{i-1}(W) \times x^{j-1}(W) \right) \cap \Omega_{x,0}^{i,j}. \quad (61)$$

If the set $V_{x,0}^{i,j}$ is negligible in $\Omega_{x,0}^{i,j}$, then (55) holds for almost any $(t, \tau) \in \Omega_{x,0}^{i,j}$. Such properties can be obtained for example when g is continuous except on a discrete subset of \mathbf{X}_0 . It is the case for example for the function $g(x) = \text{sign}(x)$: although the subset on which g is discontinuous is negligible (in the Lebesgue sense), the function g is really discontinuous.

4.2 Impact of the measurement noise

Let now discuss about the impact of the measurement noise on the efficiency of the identification method. When x_m is noisy, that is $x_m \neq x$, the (theoretical) set $\Omega_{x,\varepsilon}^{i,j}$ and the set $\Omega_{x_m,\varepsilon}^{i,j}$ (and so, the operators $D_{x,\varepsilon}$ and $D_{x_m,\varepsilon}$) are in general different, because, due to the measurement noise, we have:

$$\|x^i(\tau) - x^j(t)\| \leq \varepsilon \not\Rightarrow \|x_m^i(\tau) - x_m^j(t)\| \leq \varepsilon. \quad (62)$$

However, we will see in the sequel that, in spite of the presence of some measurement noise, the transformation of model (1) by the (linear continuous) operator $D_{x_m,\varepsilon}$ can be used to get an identification model based on measurement trajectories. This is what is shown in the following results.

Proposition 15 Let $x, \tilde{x} \in \prod_{l=1}^l C_{\mathbf{X}}^0(\Omega^l)$; if $\|\tilde{x} - x\| \leq \varepsilon$, then for any i, j :

$$\Omega_{x, \varepsilon/3}^{i,j} \subset \Omega_{\tilde{x}, \varepsilon}^{i,j} \subset \Omega_{x, 3\varepsilon}^{i,j} \subset \Omega_{\tilde{x}, 9\varepsilon}^{i,j}. \quad (63)$$

As a consequence, if $\varepsilon' \leq \frac{\varepsilon}{3}$ then we have:

$$\Omega_{\tilde{x}, \varepsilon'}^{i,j} \subset \Omega_{x, \varepsilon}^{i,j}; \quad (64)$$

furthermore the operator:

$$\begin{aligned} \prod_l C_{\mathbf{X}}^0(\Omega^l) &\longrightarrow \mathcal{L}\left(\prod_l L_{\mathbf{V}}^\infty(\Omega^l), L_{\mathbf{V}}^\infty(\Omega_{x, \varepsilon'}^{i,j})\right) \\ x &\longmapsto R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}} \circ D_{x, \varepsilon}^{i,j} \end{aligned} \quad (65)$$

is well defined on the ball with center \tilde{x} and radius ε and equal to $D_{\tilde{x}, \varepsilon'}^{i,j}$.

N.B. We recall that $R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}}$ is simply a restriction operation to subset $\Omega_{\tilde{x}, \varepsilon'}^{i,j}$ (see section 4.1.2).

Proof. From $\|x - \tilde{x}\| = \sup_i \sup_{\Omega^i} \|x^i(\tau) - \tilde{x}^i(\tau)\|_{\mathbf{X}}$, we deduce that for any i, j, τ, t :

$$\|x^i(\tau) - \tilde{x}^i(\tau)\|_{\mathbf{X}} \leq \varepsilon \text{ and } \|\tilde{x}^j(t) - x^j(t)\|_{\mathbf{X}} \leq \varepsilon; \quad (66)$$

moreover, from the definition of $\Omega_{\tilde{x}, \varepsilon}^{i,j}$, if $(\tau, t) \in \Omega_{\tilde{x}, \varepsilon}^{i,j}$, then $\|\tilde{x}^i(\tau) - \tilde{x}^j(t)\|_{\mathbf{X}} \leq \varepsilon$. We deduce that:

$$\|x^i(\tau) - x^j(t)\|_{\mathbf{X}} \leq \|x^i(\tau) - \tilde{x}^i(\tau)\|_{\mathbf{X}} + \|\tilde{x}^i(\tau) - \tilde{x}^j(t)\|_{\mathbf{X}} + \|\tilde{x}^j(t) - x^j(t)\|_{\mathbf{X}} \leq 3\varepsilon, \quad (67)$$

which means that $(\tau, t) \in \Omega_{x, 3\varepsilon}^{i,j}$. The proofs of $\Omega_{x, \varepsilon/3}^{i,j} \subset \Omega_{\tilde{x}, \varepsilon}^{i,j}$ and $\Omega_{x, 3\varepsilon}^{i,j} \subset \Omega_{\tilde{x}, 9\varepsilon}^{i,j}$ are similar.

The inclusion (64) is directly deduced from (63). Consequently, the function $x \mapsto R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}} \circ D_{x, \varepsilon}^{i,j}$ is well defined on $B(\tilde{x}, \varepsilon)$, the ball with center \tilde{x} and radius ε . We then deduce from (64) that $R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}} \circ R_{\Omega_{x, \varepsilon}^{i,j}} = R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}}$, and from definition of $D_{x, \varepsilon}^{i,j}$, we get, for any $x \in B(\tilde{x}, \varepsilon)$:

$$R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}} \circ D_{x, \varepsilon}^{i,j} = R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}} \circ R_{\Omega_{x, \varepsilon}^{i,j}} \circ D_{x, \varepsilon}^{i,j} = R_{\Omega_{\tilde{x}, \varepsilon'}^{i,j}} \circ D_{x, \varepsilon}^{i,j} = D_{\tilde{x}, \varepsilon'}^{i,j}. \quad (68)$$

□

From proposition 15, if the noise trajectories are continuous (which excludes *white* measurement noises; in practice, a suitable pre-filtering of the measured data can be required) and small enough in \mathbf{X} , we so have the nice property $\Omega_{x_m, \varepsilon}^{i,j} \subset \Omega_{x, 3\varepsilon}^{i,j}$. From proposition 5, we also know that $\Omega_{x_m, \varepsilon}^{i,j}$ is closed, with non-empty interior, if $\varepsilon > 0$. We then easily deduce the following important result.

Corollary 16 If $\|x_m - x\| \leq \varepsilon$, then for any $v \in \prod_l L_{\mathbf{V}}^\infty(\Omega^l)$, we have:

$$D_{x_m, \varepsilon/3}^{i,j} v = D_{x, \varepsilon}^{i,j} v|_{\Omega_{x_m, \varepsilon/3}^{i,j}}, \quad (69)$$

with $\Omega_{x_m, \varepsilon/3}^{i,j} \subset \Omega_{x, \varepsilon}^{i,j}$.

Proof. By application of proposition 15 for $\tilde{x} = x_m$ and $\varepsilon' = \frac{\varepsilon}{3}$. □

From the practical point of view, this last result means that in spite of the measurement noise $e_m := x_m - x$, if this noise is such that $\|e_m\| \leq \varepsilon$, then the (unknown and not directly accessible) operator $D_{x, \varepsilon}^{i,j}$ can be *exactly* determined from the only measurement x_m by applying the operator $D_{x_m, \varepsilon/3}^{i,j}$. In other words, by applying $D_{x_m, \varepsilon/3}^{i,j}$, we get the same “result” than if we apply $D_{x, \varepsilon}^{i,j}$ (in particular the ε -cancellation of g leading to model (26), which is used for the identification of y ; see sections 3.3 and 4.1), but restricted to a smaller set (namely $\Omega_{x_m, \varepsilon/3}^{i,j}$ instead of $\Omega_{x, \varepsilon}^{i,j}$), which means that we get less equations to identify y . Thus, we can state that the consequence of the measurement noise is only a loss of precision in the identification process of y .

Remark 17 Note that, in practice, it can seem more natural to denote ε the size of the quasi-cancellation band on measured trajectories x_m instead of the one on theoretical trajectories x , that is to build sets $\Omega_{x_m, \varepsilon}^{i,j}$ instead of sets $\Omega_{x_m, \varepsilon/3}^{i,j}$. In fact, this has no importance at all as the value of ε can be arbitrarily chosen, and all the above results can be applied by replacing ε by 3ε , as it is done in section 3.3.

Remark 18 There is no restriction about the probabilistic nature of the noise measurement: it is only assumed that the noise trajectories e_m are continuous with norm bounded by some $\varepsilon > 0$. In particular, Gaussian as well as non-Gaussian measurement noise are admissible. In the case of white measurement noise, a pre-filtering of the measured data would be required in order to get continuous noise trajectories; the cut-off frequency should be chosen as small as possible (in order to get ε as small as possible) but far away the frequency band of the dynamic system under consideration, in such a way that the disturbance on the measured trajectories remains negligible. Note that in practical situations, such a pre-filtering is anyway required for the construction of a finite subset of $\Omega_{x_m, \varepsilon}$ (see Appendix B).

4.3 Properties of difference operators D and $D_{x, \varepsilon}$

We give in this paragraph an important result used in section 6, about operators D and $D_{x, \varepsilon}$. We first introduce the following notation: if g is a function defined on \mathbf{V} and $v = (v^i)$ a vector of \mathbf{V} -valued functions v^i , then, by convention, we denote $g \circ v$ the vector-function with components defined by:

$$(g \circ v)^i := g \circ v^i. \quad (70)$$

With this convention, if $v \in \prod_i L_{\mathbf{V}}^\infty(\Omega^i)$ (with Ω^i , $i = 1 : I$ some non-empty compact subsets of \mathbb{R}^n) and if g is measurable with values in a Banach space \mathbf{Z} , then $g \circ v$ is a vector of \mathbf{Z} -valued measurable functions, namely: $g \circ v \in \prod_i \mathcal{M}(\Omega^i; \mathbf{Z})$.

Proposition 19 Let $B \in \mathcal{L}(\mathbf{V}, \mathbf{Z})$; then the linear operator:

$$\mathbf{B} : \prod_{\mathcal{I}} L_{\mathbf{V}}^\infty(\tilde{\Omega}) \xrightarrow{\quad} \prod_{\mathcal{I}} L_{\mathbf{Z}}^\infty(\tilde{\Omega}) \\ v \quad \longmapsto \quad B \circ v \quad (71)$$

is continuous for $\tilde{\Omega} = \Omega^j$ and $\mathcal{I} = \{1 : I\}$ or $\tilde{\Omega} = \Omega^{i,j}$ and $\mathcal{I} = \{1 : I\}^2$. Furthermore:

$$\mathbf{D} \circ \mathbf{B} = \mathbf{B} \circ \mathbf{D} \in \mathcal{L} \left(\prod_j L_{\mathbf{V}}^\infty(\Omega^j), \prod_{i,j} L_{\mathbf{Z}}^\infty(\Omega^{i,j}) \right), \quad (72)$$

$$D_{x, \varepsilon} \circ \mathbf{B} = \mathbf{B} \circ D_{x, \varepsilon} \in \mathcal{L} \left(\prod_j L_{\mathbf{V}}^\infty(\Omega^j), \prod_{\mathcal{I}, \varepsilon} L_{\mathbf{Z}}^\infty(\Omega_{x, \varepsilon}^{i,j}) \right). \quad (73)$$

Proof. The continuity of \mathbf{B} is deduced from:

$$\|\mathbf{B}v\| = \sup_{k \in \mathcal{I}} \sup_{s \in \tilde{\Omega}} \left\| B(v^k(s)) \right\|_{\mathbf{Z}} \leq \sup_{k \in \mathcal{I}} \sup_{s \in \tilde{\Omega}} \|B\| \|v^k(s)\| \leq \|B\| \|v\|. \quad (74)$$

Then we have:

$$\begin{aligned} [(\mathbf{D} \circ \mathbf{B})v]^{i,j}(\tau, t) &= [\mathbf{D}(B \circ v)]^{i,j}(\tau, t) = (B \circ v)^i(\tau) - (B \circ v)^j(t) \\ &= B v^i(\tau) - B v^j(t) = B(v^i(\tau) - v^j(t)) = B[(D^{i,j}v)(\tau, t)] \\ &= [\mathbf{B} \circ (\mathbf{D}v)]^{i,j}(\tau, t) = [(\mathbf{B} \circ \mathbf{D})v]^{i,j}(\tau, t). \end{aligned} \quad (75)$$

(73) is finally proved by restriction to $\Omega_{x, \varepsilon}$. \square

Remark 20 Due to the different natures of the involved vector spaces, operators \mathbf{B} and \mathbf{D} in expression $\mathbf{B} \circ \mathbf{D}$ are in fact different from the ones in expression $\mathbf{D} \circ \mathbf{B}$: they are either defined with $\tilde{\Omega} = \Omega^j$ and $\mathcal{I} = \{1 : I\}$ or $\tilde{\Omega} = \Omega^{i,j}$ and $\mathcal{I} = \{1 : I\}^2$. For simplicity, the same notation is used for both cases.

5 A simple academic example

In this section, we present an academic example to apply the method on a simple and noise-free case where the data x_m , the quantity z and the sets $\Omega_{x,\epsilon}^{i,j}$ can be analytically computed. The aim of this example is essentially to illustrate the steps of the identification method and to show how it works. A more complex and relevant example is treated in section 6.

We consider the dynamic model:

$$\forall t > 0, \partial_t X(t) = f(X(t)) + u(t), \quad X(0) = 0, \quad (76)$$

where X and u are real-valued trajectories, and f is a function defined on \mathbb{R} . Given one measured trajectory X_m (N.B.: we take here $I = 1$), the problem under consideration is to identify both the function f and the “input” trajectory u . Note that without any additional hypothesis, the solution to this problem is not unique. Indeed, for any f chosen a priori and for $X_m = X$, a trajectory u is deduced directly from (76), namely: $u(t) = (\partial_t X)(t) - f(X(t))$.

The model (76) can be identified under the form (1) with $\Omega^1 = [0, T]$, $\mathbf{X} = \mathbf{Z} = \mathbb{R}$, $F \subset L^\infty(\Omega^1)$ (and then $F \subset L^2(\Omega^1)$) a L -dimensional subspace of bounded functions and:

$$x = X, \quad y = u, \quad z = -\partial_t X, \quad g = f, \quad A : y \mapsto y. \quad (77)$$

Note that in spite of its trivial definition, the operator $A : F \rightarrow L^\infty(\Omega^1)$ is not the identity operator but must be understood as the canonical injection: $F \hookrightarrow L^\infty(\Omega^1)$.

Assume that a (noise-free) measured solution of (76) is given by:

$$x_m(t) = X(t) = \sin(t^2), \quad t \in \Omega^1 = [0, \sqrt{\pi}]. \quad (78)$$

Then, we have: $z(t) = -\partial_t X(t) = -2t \cos(t^2)$. Following the identification method, we build the set (ε can be chosen equal to 0 because the measurements are noise-free and $\mathbf{Z} = \mathbb{R}$):

$$\Omega_{x_m,0}^{1,1} = \left\{ (t, t), \left(t, \sqrt{\pi - t^2} \right), t \in [0, \sqrt{\pi}] \right\}. \quad (79)$$

One can see on Figure 4(a) the subset of $\Omega_{x_m,0}^{1,1}$ obtained in a concrete situation for a uniform discretization $\{t_k\}_{k=1:40}$ of $[0, \sqrt{\pi}]$.

We know by construction that for any $(t, \tau) \in \Omega_{x,0}^{1,1}$ we have:

$$u(t) - u(\tau) - (\partial_t X)(t) + (\partial_t X)(\tau) = 0. \quad (80)$$

As explained in Step 2 (cf. 3.3), these equations can be written under the synthetic form (27) with:

$$b_{x,z}^0 : (t, \tau) \in \Omega_{x,0}^{1,1} \mapsto ((\partial_t x)(t) - (\partial_t x)(\tau), (\partial_t x)(t_*)), \quad (81)$$

$$\phi_{x,A}^0(u, g_0) : (t, \tau) \in \Omega_{x,0}^{1,1} \mapsto (u(t) - u(\tau), g_0 + u(t_*)), \quad (82)$$

where $t_* \in [0, \sqrt{\pi}]$ is arbitrarily chosen, and $g_0 = f(x(t_*))$. From there the unknowns $\theta = (u, g_0)$ have been identified by using the least-squares estimator given by (34) with:

$$t_* = 0, \quad \varepsilon = 0, \quad G^0 := L^2(\Omega_{x_m,0}^{1,1}) \ni D_{x_m,0} y, \quad (83)$$

and a conditioning parameter $\eta = 10^{-4}$ which ensures the uniqueness of the solution of this least-square problem. The identified input \hat{u} obtained with this identification process (for the above-mentioned discretization) can be seen on Figure 4(b).

Finally, following the Step 3 (cf. 3.4), we have made an *a posteriori* estimation of f using a classical regression of its pseudo-graph (using the estimation \hat{u} previously identified). The result of this identification is given in Figure 4(c).

To highlight the accuracy of the identified model, we give in Figure 4(d) both the measured trajectory $X_m = X$ and the reconstructed trajectory \hat{X} numerically computed from the identified model:

$$\partial_t \hat{X} = \hat{f}(\hat{X}) + \hat{u}, \quad \hat{X}(0) = X_m(0) = 0. \quad (84)$$

As we can see, the simulated trajectory perfectly fits the measured data.

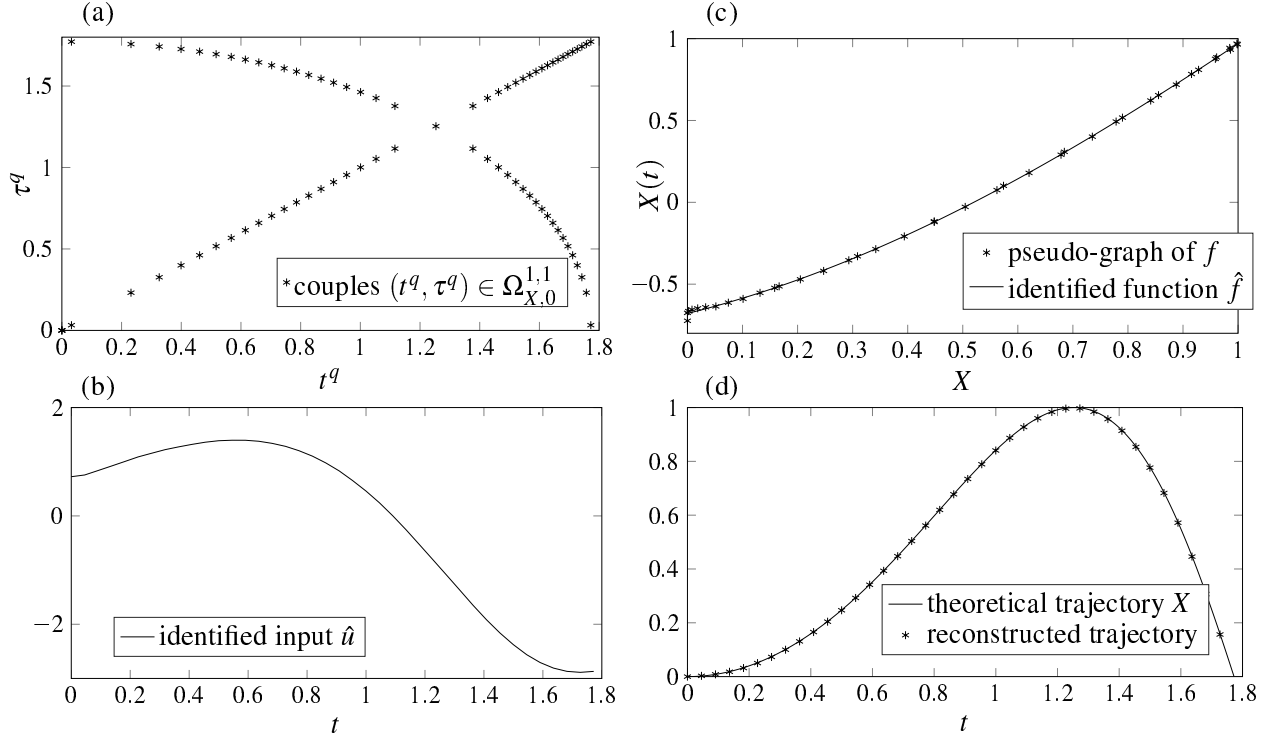


Figure 4: **Identification example of section 5:** (a) set $\Omega_{X,0}^{1,1} = \Omega_{x_m,0}^{1,1}$; (b) identified input \hat{u} ; (c) identified function \hat{f} ; (d) theoretical trajectory X used for the identification and reconstructed trajectory after identification.

Remark 21 *The method can still be used for the identification of model (76) with additional constraints, for example $u(0) = u_0$ or u a T_0 -periodic function (with $T_0 < T$). Indeed, those problems can still be expressed under the form (1) with $x = X$, $y = u$, $f = \begin{pmatrix} f \\ 0 \end{pmatrix}$, and respectively:*

- $z = \begin{pmatrix} -\partial_t X \\ -u_0 \end{pmatrix}$, $A = \begin{pmatrix} I \\ \langle \delta, \cdot \rangle \end{pmatrix}$ for the case $u(0) = u_0$;
- $z = \begin{pmatrix} -\partial_t X \\ 0 \end{pmatrix}$, $A = \begin{pmatrix} I \\ e^{-T_0 p} - I \end{pmatrix}$ for the case where u is a T_0 -periodic function ($e^{-T_0 p}$ being the symbolic notation for the delayed operator $v \mapsto v(\cdot - T_0)$).

This is one of the benefits of the generic formulation (1) chosen to develop the identification method of this paper.

6 Application to a class of Volterra models

In this section, the identification method is applied to a class of Volterra models in which a nonlinear function and a linear dynamic operator with non rational transfer function have to be identified.

6.1 Model under consideration and identification problem

Consider the multi-trajectory model of the form:

$$H(\partial_t) X^i = f(X^i) + u^i, \quad i = 1 : I, \quad (85)$$

where:

- $u^i \in L^\infty(\mathbb{R}_t^+)$, $i = 1 : I$, are the inputs (supposed to be known),

- $X^i \in C^0(\mathbb{R}_t^+)$, $i = 1 : I$, are the outputs (which are measured),
- f is a nonlinear function,
- and $H(\partial_t)$ is a linear operator for which there exists a function $\mu \in L^2(\mathbb{R}_\xi; \mathbb{C})$ such that $H(\partial_t)$ can be realized by the infinite dimensional state formulation:

$$\begin{cases} \partial_t \psi^i(t, \xi) = \gamma(\xi) \psi^i(t, \xi) + X^i(t), \quad \xi \in \mathbb{R}, \quad \psi(0, \cdot) = 0 \\ H(\partial_t) X^i(t) = \int_{\mathbb{R}} \mu(\xi) [\gamma(\xi) \psi^i(t, \xi) + X^i(t)] d\xi, \end{cases} \quad (86)$$

with $\gamma : \mathbb{R} \rightarrow \mathbb{C}^-$ a given continuous function defining an infinite arc in \mathbb{C}^- .

The operators $H(\partial_t)$, as defined above, are the so-called diffusive operators, associated with the diffusive representation theory. A complete statement of this theory, devoted to state realizations of integral operators, can be found in [39, 40]. Note that the problem of identification of convolution operators by means of diffusive representation is studied in [17, 20].

Models of the form (85) are frequently encountered in various domains (see examples 4 and 3). With the convenient notation: $\langle \mu, \eta \rangle := \int \mu(\xi) \eta(\xi) d\xi$, model (85) can be rewritten under the equivalent form:

$$\begin{cases} \langle \mu, \gamma \psi^i + X^i \rangle = f(X^i) + u^i, \quad i = 1 : I, \\ \partial_t \psi^i = \gamma \psi^i + X^i, \quad \psi^i(0, \cdot) = 0, \quad i = 1 : I. \end{cases} \quad (87)$$

The problem under consideration is to estimate, with the method introduced in section 3, both the operator $H(\partial_t)$ (characterized by the function μ) and the function f , from measurements X_m^i of X^i and the associated inputs u^i .

Remark 22 *Without additional hypothesis, the solution $(H(\partial_t), f)$ is of course not unique, namely because equation $(H(\partial_t) + cI) X = f(X) + cX + u$ is equivalent to $H(\partial_t) X = f(X) + u$. The uniqueness of the identification problem is restored when, for example, the static gain of the operator $H(\partial_t)$, that is the value $H(p)|_{p=0}$ or the value of function f at a given point, is known (see paragraph 6.4).*

Remark 23 *The transfer function $H(p)$ of the causal convolution operator $H(\partial_t)$ can be expressed by:*

$$H(p) = \left\langle \mu, \frac{p}{p - \gamma(\cdot)} \right\rangle. \quad (88)$$

Indeed, by denoting \mathcal{L} the Laplace transform with respect to the time variable t and p the Laplace symbolic variable, we have:

$$\mathcal{L}(H(\partial_t) X^i) = \mathcal{L} \int_{\mathbb{R}} (\gamma \psi^j + X^i) \mu(\xi) d\xi = \int \mathcal{L}(\gamma \psi^j + X^i) \mu d\xi = \int \mathcal{L}(\partial_t \psi^j) \mu d\xi \quad (89)$$

$$= \int p \mathcal{L} \psi^j \mu d\xi = \int \frac{p \mu}{p - \gamma} \mathcal{L} X^i d\xi \quad (\text{from (86)}) = H(p) \mathcal{L} X^i \quad \text{with } H(p) = \int \frac{p \mu}{p - \gamma} d\xi. \quad (90)$$

This property will be used in the example of section 6.4.

6.2 Application of the method

Let define $\Psi_{X^i}(t) := \gamma(\cdot) \psi^i(t, \cdot) + X^i(t)$. It can be shown (see for example [39]) that if $X^i \in L_{\text{loc}}^\infty(\mathbb{R}_t^+)$, then $\Psi_{X^i}(t) \in L^2(\mathbb{R}_\xi; \mathbb{C}) \cap C^0(\mathbb{R}_\xi; \mathbb{C})$ for any $t > 0$ and that $t \mapsto \|\Psi_{X^i}(t)\|_{L^2}$ is bounded on any $[0, T]$. We then denote, by convenience, $A_{X^i} := \langle \cdot, \Psi_{X^i} \rangle$, that is:

$$\begin{aligned} A_{X^i} : L^2(\mathbb{R}_\xi; \mathbb{C}) &\rightarrow L_{\text{loc}}^2(\mathbb{R}_t^+; \mathbb{C}) \\ \mu &\mapsto \langle \mu, \Psi_{X^i}(\cdot) \rangle. \end{aligned} \quad (91)$$

Using these notations, for any X^i and μ we have $A_{X^i} \mu = H(\partial_t) X^i$ and then model (87) can be written under the form (1) with:

$$A^i = A_{X^i}, \quad g = -f, \quad y = \mu \in L^2(\mathbb{R}; \mathbb{C}), \quad x^i = X^i \in C^0([t_0^i, t_f^i]), \quad z^i = -u^i \in L^\infty([t_0^i, t_f^i]). \quad (92)$$

In accordance with section 3, after application of the difference operator $D_{x_m, \varepsilon}$ and ε -cancellation of the nonlinear term, we get model (27), whose unknowns to be identified are $\theta = (\mu, g_0) \in L^2(\mathbb{R}_\xi; \mathbb{C}) \times \mathbb{R}$. To make explicit the operator $\phi_{x_m, A_m}^\varepsilon$ necessary for the identification of θ (see formula (30)), we use the proposition 19 with $\mathbf{B} : \Psi_{x^i}(t) \in L^2(\mathbb{R}_\xi; \mathbb{C}) \mapsto \langle \mu, \Psi_{x^i}(t) \rangle \in \mathbb{C}$ from which we get:

$$D_{x_m, \varepsilon}(A_m \mu) = D_{x_m, \varepsilon} \circ \mathbf{B} \Psi_{x_m}(\cdot) = \mathbf{B} \circ D_{x_m, \varepsilon} \Psi_{x_m}(\cdot) = \langle \mu, D_{x_m, \varepsilon} \Psi_{x_m}(\cdot) \rangle, \quad (93)$$

with $\Psi_{x_m}(t) = (\Psi_{x_m^i}(t))$ and $A_m = (A_{x_m^i})$. Such a property is interesting because it makes the linearity of μ appear explicitly in the expression of operator $\phi_{x_m, A_m}^\varepsilon$ which is indeed given by:

$$\phi_{x_m, A_m}^\varepsilon : (\mu, g_0) \mapsto \left(\langle \mu, D_{x_m, \varepsilon} \Psi_{x_m}(\cdot) \rangle, g_0 + \left\langle \mu, \Psi_{x_m^{i_0}}(t_*) \right\rangle \right), \quad (94)$$

with $t_* \in [t_0^i, t_f^i]$, and $i_0 \in \{1 : I\}$ arbitrarily chosen. With in addition $b_{x_m, z_m}^\varepsilon = (D_{x_m, \varepsilon} u, u(t_*))$, θ can be identified by means of the estimator (33). After the identification of θ , the function g can be identified following the method presented in paragraph 3.4.

Remark 24 *We recall that, following the notations introduced in this paragraph, $\Psi_{x_m^i} := \gamma \psi_{x_m^i} + x_m^i$ with $\psi_{x_m^i}$ the solution of the Cauchy problem:*

$$\partial_t \psi(t, \xi) = \gamma(\xi) \psi(t, \xi) + x_m^i(t), \quad \psi(0, \cdot) = 0. \quad (95)$$

Remark 25 *Note that, even if the operator $H(\partial_t)$ is rational, the use of the diffusive representation is quite useful, namely because the function μ (to be identified), which completely defines the operator $H(\partial_t)$, always appears linearly in the model, which would not be the case when using the coefficients (numerator and denominator) of a rational transfer function.*

6.3 Numerical implementation

From the numerical point of view, thanks to the continuity of Ψ_{x_m} , a finite dimensional approximate identification model can be obtained simply by considering L -dimensional approximations μ^L of μ such as atomic measures (δ_b denotes the Dirac measure at point b) defined on a suitable sequence of mesh $\{\xi_l^L\}_{l=1:L}$ [39]; such an approximation takes the form:

$$\mu^L = \sum_{l=1}^L \mu_l^L \delta_{\xi_l^L}, \quad \mu_l^L \in \mathbb{C}. \quad (96)$$

If $\bigcup_{L \leq \mathbf{L}} \{\xi_l^L\}$ is an increasing sequence and $\bigcup_{L < +\infty} \{\xi_l^L\}$ is dense in \mathbb{R} , the following (weak) convergence holds with μ_l^L suitably chosen [39]:

$$\forall \eta \in C^0(\mathbb{R}), \quad \langle \mu^L, \eta \rangle \xrightarrow{L \rightarrow +\infty} \langle \mu, \eta \rangle. \quad (97)$$

In that sense, we get the following L -dimensional approximation of $(A_m \mu)(t) = \langle \mu, \Psi_{x_m}(t) \rangle$:

$$(A_m \mu)(t) \simeq \langle \mu^L, \Psi_{x_m}(t) \rangle = \sum_{l=1}^L \mu_l^L (\Psi_{x_m}(t))(\xi_l^L). \quad (98)$$

Many of (non rational or rational) operators $H(\partial_t)$ encountered in practice can be closely approximated with small L (see for example [41]).

With such an approximation, we can consider the least-squares estimator of $\theta = (\mu, g_0)$ given by (34). The nonlinear function g is then identified after decomposition on a suitable topological basis. The choice of the basis functions will depend on the form of the pseudo-graph of g obtained after identification of θ (see the following paragraph).

6.4 A concrete application example

For the numerical application, we consider the problem of identification of the electrical behavior of a capacitor realized in cubic Perovskite $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$. The identification model considered is the one described in example 4 (and in [30]). For an other concrete example of application, see [42] in which the identification method has been applied to the model of the spherical flame described in example 3.

6.4.1 Model of the system.

In the example 4, the model of the system is of the form (85). The operator $H(\partial_t)$ to be identified is the inverse of the impedance operator of the capacitor, that is $Z_c^{-1}(\partial_t)$, whose transfer function is given by [30]:

$$H(p) = \frac{13 \times 10^{-9} p}{(1 + 5.0 \times 10^{-4} p)^{0.12}} \quad (99)$$

This operator can be realized by an infinite dimensional state realization of the form (86) with [40, 39]:

$$\gamma(\xi) = |\xi| e^{i \text{sign}(\xi)(\frac{\pi}{2} + \alpha)}, \text{ with } \alpha \in \left] 0, \frac{\pi}{2} \right[, \quad (100)$$

$$\text{and } \mu(\xi) = \frac{13 \times 10^{-9} \xi e^{2i \text{sign}(\xi)(\frac{\pi}{2} + \alpha)}}{2i\pi \left(1 + 5 \times 10^{-4} |\xi| e^{i \text{sign}(\xi)(\frac{\pi}{2} + \alpha)} \right)^{0.12}}, \forall \xi \in \mathbb{R}. \quad (101)$$

The nonlinear unknown function f is defined by $f : X \mapsto -G(X)X$ where G is the conductance nonlinear function of the system. The shape of the function $-f$ is given in Fig. 1 in [30]. For the simulation of the data, a comparable but analytical function has been considered under the form:

$$f : X \mapsto -p_1 \tan(p_2 X) - p_3 X, \quad (102)$$

with $p_1 = 2.820 \times 10^{-3}$, $p_2 = 1.692 \times 10^{-2}$ and $p_3 = -4.771 \times 10^{-5}$ (see Figure 7 in 1(b)).

6.4.2 Available data sets and identification parameters.

For the identification, we consider $I = 4$ sets of data $\{u_m^{i,k}, X_m^{i,k}\}$, $i = 1 : I$ (k being the time discretization index), which were obtained from accurate simulation of (85) with the above-mentioned operator $H(\partial_t)$ and function f (see [40, 39] for details about such a simulation). The sampling frequency is equal to $\frac{1}{\Delta t} = 10^6 \text{ Hz}$ and $K = 6 \times 10^6$, so that $t_{\max} = K \Delta t = 6 \text{ s}$.

The input measurements $\{u_m^{i,k}\}$ are assumed noise-free, that is $u_m^{i,k} = u^{i,k}$. For any i , the considered input $u^{i,\cdot}$ is a colored noise obtained by filtering a stationary white Gaussian process with standard deviation 1 with a first order filter of transfer function $\frac{a}{p+a}$, $a = \frac{100}{2\pi}$.

The sampled output data set associated with input $u^{i,k}$ is denoted $X^{i,k}$. The output measurements $X_m^{i,k}$ are defined by $X_m^{i,k} = X^{i,k} + \eta^{i,k}$ with $\eta^{i,\cdot}$ a wide-band Gaussian sequence with standard deviation $\sigma_{\Delta t}$. From the continuous-time point of view, this measurement noise is a wide-band stationary Gaussian process with spectral power density $\sigma = \sqrt{\Delta t} \sigma_{\Delta t}$.

Remark 26 *The band of the measurement noise is limited: recall indeed that its trajectories must be continuous functions (otherwise, the sets $\Omega_{x,\varepsilon}^{i,j}$ would not be well defined). In practical situations, a pre-filtering of the measured data may be required.*

According to the Shannon Theorem, the continuous-time information contained in the available discrete data sets $\{u_m^{i,k}, X_m^{i,k}\}$, $i = 1 : I$ is limited to the angular frequency band $\left[\frac{2\pi}{K\Delta t}, \frac{\pi}{\Delta t} \right] = [1.05, 3.14 \times 10^6]$. In this example, the objective is to identify $H(\partial_t)$ on the specific angular frequency band $[2 \times 10^0, 10^6] \subset \left[\frac{2\pi}{K\Delta t}, \frac{\pi}{\Delta t} \right]$ and the nonlinear function f on the domain $[-50, 50]$ covered by the data.

For that, the function $g = -f$ is identified under the form $g \simeq \sum_{p=1}^P c_p \mathbf{g}^p$, $c_p \in \mathbb{R}$ with $P = 60$, and \mathbf{g}^p some classical hat functions. We denote $\{\hat{c}_p\}_{p=1:P}$ the coefficients computed by the identification method. To get the uniqueness of the solution (μ, f) of (87) (for given X and u), we suppose that the value $f(20)$ (this value has been arbitrarily chosen in our case) is a known data that we denote f_{20} :

$$f(20) = -p_1 \tan(20p_2) - 20p_3 = f_{20}. \quad (103)$$

The identified function \hat{f} will be therefore given by:

$$\hat{f}(X) = \frac{f_{20} + \sum_{p=1}^P \hat{c}_p \mathbf{g}^p(20)}{20} X - \sum_{p=1}^P \hat{c}_p \mathbf{g}^p(X). \quad (104)$$

Then we consider the approximation (98) of $A_m \mu$ in which γ is defined by (100) with $\alpha = \frac{70}{180} \pi$, and with $L = 120$ discretization points ξ_l geometrically spaced between $\xi_1 = 10^0$ and $\xi_{120} = 10^6$ (10 points per frequency decade). Some indications about the choice of γ and the discretization $\{\xi_l\}_{l=1:L}$ can be found in [17, 20]. We denote $\{\hat{\mu}_l\}_{l=1:L}$ the coefficients computed by the identification method. From (96) and (88), and to satisfy constraint (103), the identified transfer function \hat{H} is given by:

$$\hat{H}(p) = p \sum_{l=1}^L \frac{\hat{\mu}_l}{p - \gamma(\xi_l)} + \frac{f_{20} + \sum_{p=1}^P \hat{c}_p \mathbf{g}^p(20)}{20}. \quad (105)$$

6.4.3 Identification results and influence of noise.

In this paragraph the results obtained with the proposed identification method are presented in noise-free and noisy cases. To facilitate the reading, the proposed identification method will be called ‘‘Cancellation method’’ in the sequel.

- **Noise-free case** ($\sigma = 0$)

The measured data $X_m^{j,k}$ are first supposed to be noise-free (i.e. $\sigma_{\Delta t} = 0$).

For illustration, a part of the graph of $X_m^1(t)$ and the associated subset of $\Omega_{X_m,0}^{1,1}$ (in fact $\Omega_{X_m,0}^{1,1} \cap [1.2; 1.3]^2$) are shown in Figure 5.

In Figure 6, we can see the Bode diagram of the frequency response $\hat{H}(i\omega)$ (deduced from $\hat{\mu}$ through formula (105)), associated with the identified operator $\hat{H}(\partial_t)$ (in 1(a) and 1(b)). The operator $H(\partial_t)$ is well approximated on the expected angular frequency band $[2 \times 10^0, 10^6]$ covered by the ξ_l .

In Figure 7, in 1(b), the graph of the identified function \hat{f} (see formula (104)) is compared to the theoretical function f . The function is correctly reconstructed on the whole domain $[-50, 50]$ with an absolute error smaller than 1.5×10^{-6} (see Figure 7, 1(c)). As an example, one of the resulting pseudo-graphs of f , namely:

$$\mathcal{G}_f^i = \bigcup_k \left\{ \left(X_m^{i,k}, \sum_{l=1}^L \hat{\mu}_l \left(\gamma(\xi_l^L) \psi_l(i, t_k^i) + X_m^{i,k} \right) - u_m^{i,k} \right) \right\}, \quad (106)$$

where $\psi_l(i, t)$ is an approximation of $\psi_{X_m^i}(t, \xi_l^L)$ computed from numerical integration of (95) with input X_m^i , is also given in Figure 7, in 1(a).

- **Noisy case** ($\sigma \neq 0$)

The measured data $X_m^{i,k}$ are now corrupted by additive wide-band measurement noises, with $\sigma_{\Delta t} = 4$. The obtained results are given in Figures 6 and 7 (right columns). Note that before applying the identification method, a pre-filtering of the data has been performed with a first order filter of transfer function $\frac{b}{p+b}$ with $b = 2 \times 10^4$.

In spite of the measurement noise, the identified transfer function $\hat{H}(i\omega)$ is still close to $H(i\omega)$, but on a smaller frequency band than in the noise-free case (see Figure 6, in 2(a) and 2(b)): this is a normal consequence of the presence of measurement noise. Indeed, high frequencies are more

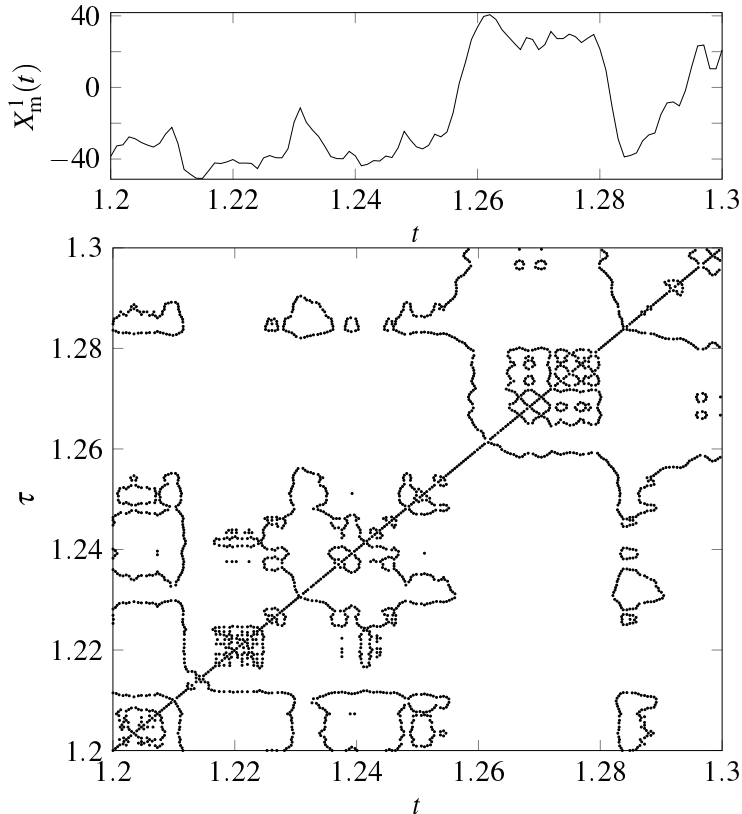


Figure 5: Part of the graph of $X_m^1(t)$ and the associated subset of $\Omega_{X_m,0}^{1,1}$.

corrupted than low ones because the trajectories X^i are continuous, with spectral density decreasing at high frequencies. This highlights the intuitive feature that accurate identification at high frequencies needs more and more measurement data as the noise magnitude increases.

The *a posteriori* identified function \hat{f} is also still very close to the theoretical one (see Figure 7 in 2(b)), the absolute error being smaller than 2×10^{-5} . Note that in spite of the presence of noise, the pseudo-graph of f (shown in Figure 7 in 2(a)) clearly exhibits significant informations, useful for the *a posteriori* choice of the basis functions \mathbf{g}^p .

6.4.4 Comparison with an iterative identification method.

To highlight its efficiency, the cancellation method is now compared to an other identification approach. We have chosen to consider an iterative equation error approach based on a relaxation algorithm and dedicated to the identification of Hammerstein models. This approach is described in several papers (see e.g. [8, 24]). It has the advantage to be a block-structured approach (like the method presented in this paper) which is able to identify non rational dynamic operators. It is not the case of all the identification methods for Hammerstein models, which often assume a rational expression for the dynamic operators.

The considered iterative method can be summarized as follows. The dynamic linear term $[H(\partial_t)X](t)$ is identified under the form:

$$\sum_{s=0}^S h_s X(t - s\Delta t), \quad (107)$$

which is a discrete approximation of the convolution integral $h * X$, assuming that the support of the kernel h of $H(\partial_t)$ is included in $[0, S\Delta t]$.

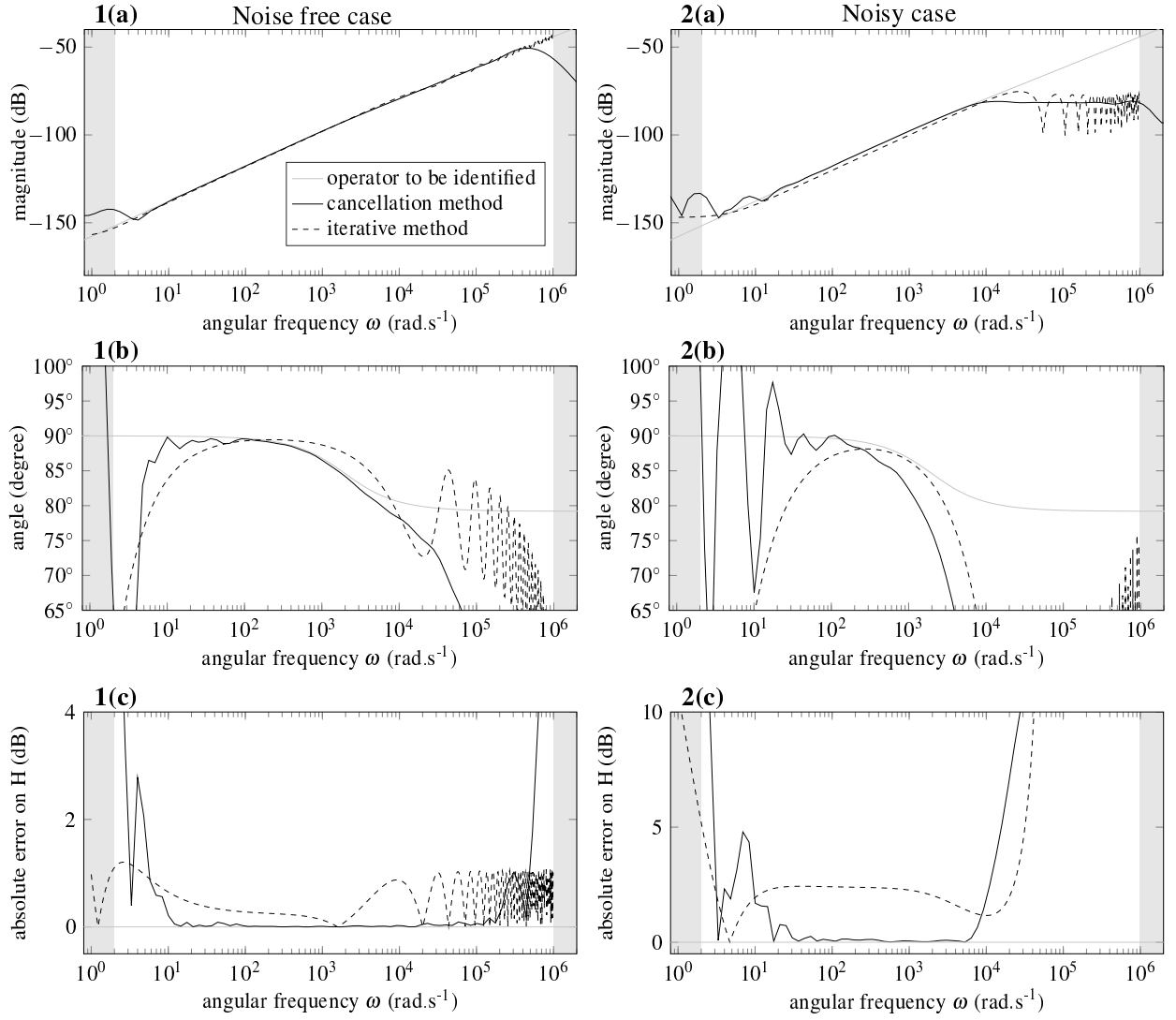


Figure 6: **Identification of the dynamic operator $H(\partial_t)$ of a Volterra model (section 6.4)** in the noise-free case (case 1 - left column) and in the noisy case (case 2 - right column): (a) and (b) comparison of the Bode diagram of the transfer function \hat{H} of the identified operator $\hat{H}(\partial_t)$ with the theoretical one; (c) absolute error between the magnitude of $\hat{H}(i\omega)$ and $H(i\omega)$. The results obtained with the cancellation method are plotted in solid black line, the ones obtained with the iterative method are plotted in dashed black line, and the theoretical function is plotted in gray solid line. The gray bands correspond to the angular frequencies which are outside the band on which the identification has been preformed.

The function $g = -f$ is identified under the same form than for the cancellation method, that is: $\sum_{p=1}^P c_p \mathbf{g}^p$, $c_p \in \mathbb{R}$ with \mathbf{g}^p some hat functions.

The iterative method consists in minimizing the function:

$$V(\{h_s\}_s, \{c_p\}_p) := \sum_{k=1}^K \sum_{j=1}^J \left\| \sum_{s=0}^S h_s X_m^{j,k-s} + \sum_{p=1}^P c_p \mathbf{g}^p(X_m^{j,k}) - u_m^{j,k} \right\|^2 \quad (108)$$

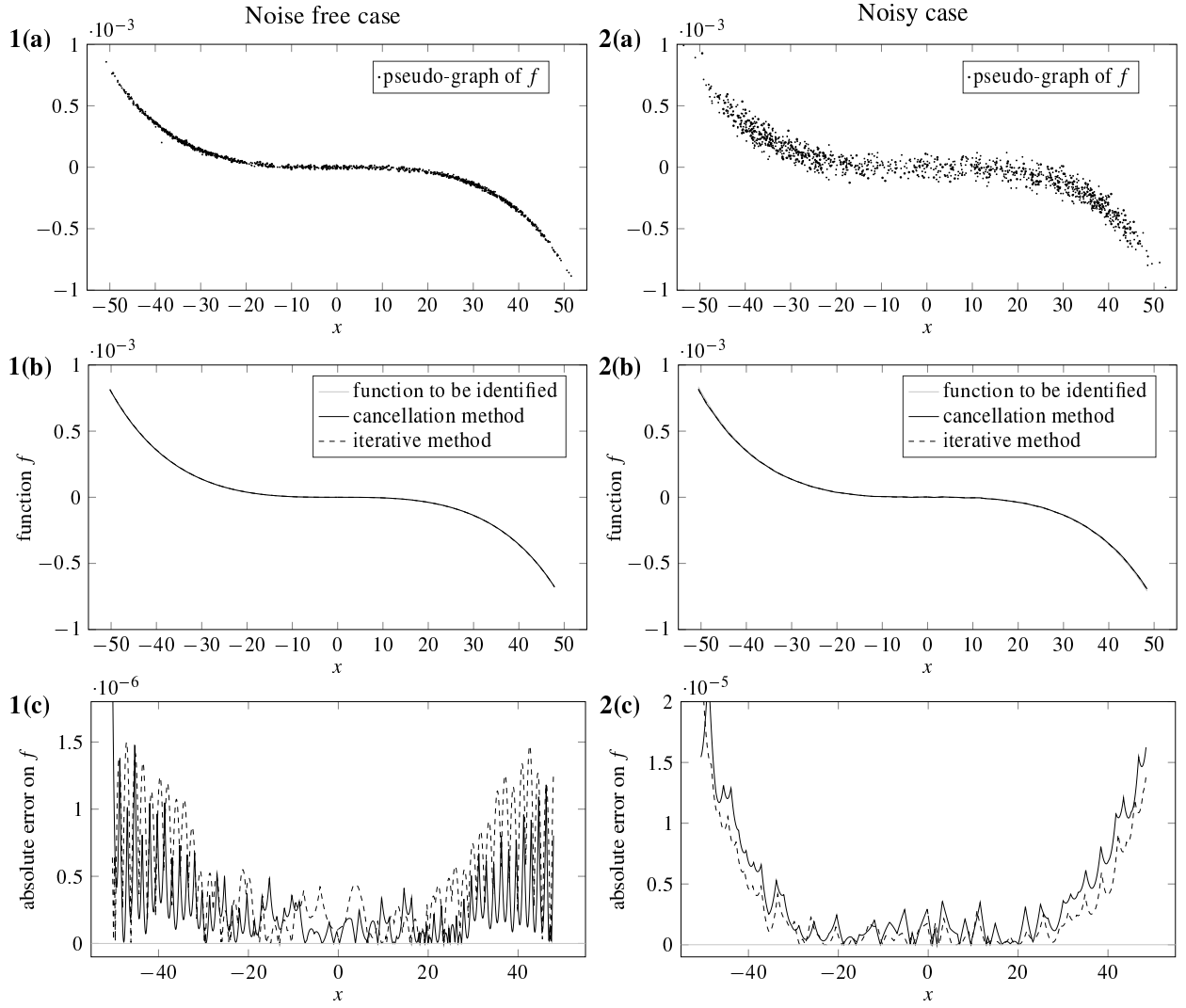


Figure 7: **Identification of the nonlinear function f of a Volterra model (section 6.4)** in the noise-free case (case 1 - left column) and in the noisy case (case 2 - right column): (a) pseudo-graph of f ; (b) comparison between the identified function \hat{f} and the theoretical one; (c) absolute error between f and \hat{f} . The results obtained with the cancellation method are plotted in solid black line, the ones obtained with the iterative method are plotted in dashed black line, and the theoretical function is plotted in gray solid line.

with respect to the parameters $\{h_s\}_{s=0:S}$ and $\{c_p\}_{p=1:P}$, with the following relaxation algorithm:

for $n = 1, 2, \dots$ (loop)

$$\text{Step 1: } \{\hat{h}_s^n\}_{s=0:S} = \arg \min_{\{h_s\}_s} V(\{h_s\}_s, \{\hat{c}_p^{n-1}\}_p) \quad (109)$$

$$\text{Step 2: } \{\hat{c}_p^n\}_{p=1:P} = \arg \min_{\{c_p\}_p} V(\{\hat{h}_s^n\}_s, \{c_p\}_p) \quad (110)$$

end for

where $\{\hat{c}_p^0\}_p$ is given.

For the Step 1 of the iterative method (identification of $\{h_s\}_s$ given $\{\hat{c}_p^{n-1}\}_p$), the algorithm of Bojanczyk, Brent & de Hoog (BBH, [43]) is used. For the Step 2 (identification of $\{c_p\}_p$ given $\{\hat{h}_s^n\}_s$), the classical least-squares method is applied.

The identification results obtained with the iterative method are given in Figures 6 and 7. To compare the cancellation and the iterative methods, we have used the same number of parameters

than for the cancellation method, that is $P = 60$ parameters for the nonlinear function f and $K + 1 = 120$ parameters for the dynamic operator $H(\partial_t)$. For all the noise-free and noisy cases, the iterative method needed less than 10 iterations to converge.

In Figure 6, we can compare the identified transfer functions of the dynamic operator $H(\partial_t)$ obtained with the cancellation method and the iterative method in both noise-free and noisy cases. For the iterative method, the transfer function has been deduced from $\{h_s\}_s$ by a Fast Fourier Transform (FFT) algorithm. In the noise-free case, the cancellation method gives results slightly better than the iterative method. In particular, the transfer function obtained with the iterative method exhibits some oscillations at high frequencies which are not present with the cancellation method. In the noisy case, an identification bias appears with the iterative method which degrades the quality of the identification of $H(\partial_t)$ in comparison with the cancellation method.

In Figure 7, the functions f identified with both methods can be compared. In both noise-free and noisy cases, the identification of f is accurate and the results quality of both methods are similar.

The good quality of the identification results obtained by application of the cancellation method on this concrete example is a proof of concept which demonstrates the feasibility of the method. From the comparison between the cancellation and the iterative methods, we moreover see that the cancellation method can be sometimes better than some other well-known methods. However, as explained in section 3.5, further study is required to improve and optimize the cancellation method and to determine the cases where it can have a substantial advantage in comparison with other methods. The case where f is discontinuous can be one of them.

6.4.5 Case where f is discontinuous.

To highlight one of the advantage of the method, we now consider the case where f is discontinuous. The system is the same as previously except for the function f which is now discontinuous at $X = 0$. The function f to be identified is now given by:

$$f : X \mapsto -p_1 \tan(p_2 X) - p_3 X - p_4 \text{sign}(X), \quad (111)$$

with $p_1 = 2.820 \times 10^{-3}$, $p_2 = 1.692 \times 10^{-2}$, $p_3 = -4.771 \times 10^{-5}$ and $p_4 = 10^{-3}$.

Remark 27 Because function f is discontinuous at point $X = 0$, the dynamic model to be identified is in fact the “pseudo-differential inclusion” $H(\partial_t) X - u \in \tilde{f}(X)$, or equivalently:

$$\begin{cases} \int \mu (\gamma \psi + X) d\xi - u \in \tilde{f}(X) \\ \partial_t \psi = \gamma \psi + X, \quad \psi(0, \cdot) = 0 \end{cases} \quad \text{with } \tilde{f}(x) = \begin{cases} \{f(x)\} & \text{if } x \neq 0 \\ [-c_2, c_2] & \text{if } x = 0. \end{cases} \quad (112)$$

Note that in this case, the sets $X_m^{i-1}(\{0\})$ must be negligible in \mathbb{R} to fit hypothesis (4), as written in paragraph 2.4. Such dynamic inclusions are frequently encountered in some fields of mechanics, for example when dry frictions are involved [44].

The function $g = -f$ is still identified under the form $g \simeq \sum_{p=1}^P c_p \mathbf{g}^p$, $c_p \in \mathbb{R}$ with $P = 60$, and \mathbf{g}^p some classical hat functions, except at point 0 where the basis function is taken discontinuous, to get accurate approximations of f with only a small number of parameters. Note that in real situations, i.e. when f is entirely unknown, such an information can be easily deduced from a detailed analysis of the pseudo-graph \mathcal{G}_g (see Figure 8, 2(a)). The same identification parameters than in the previous case (with f continuous) have been considered for the identification of $H(\partial_t)$. The considered data sets have also the same characteristics than before: number, sampling frequency, maximal time, input, noise level. The results obtained with the two identification methods in the noisy case are given in Figure 8.

For both methods, the identification of the operator $H(\partial_t)$ gives results similar to the case where f is continuous, on an angular frequency band slightly smaller. The results of the iterative method

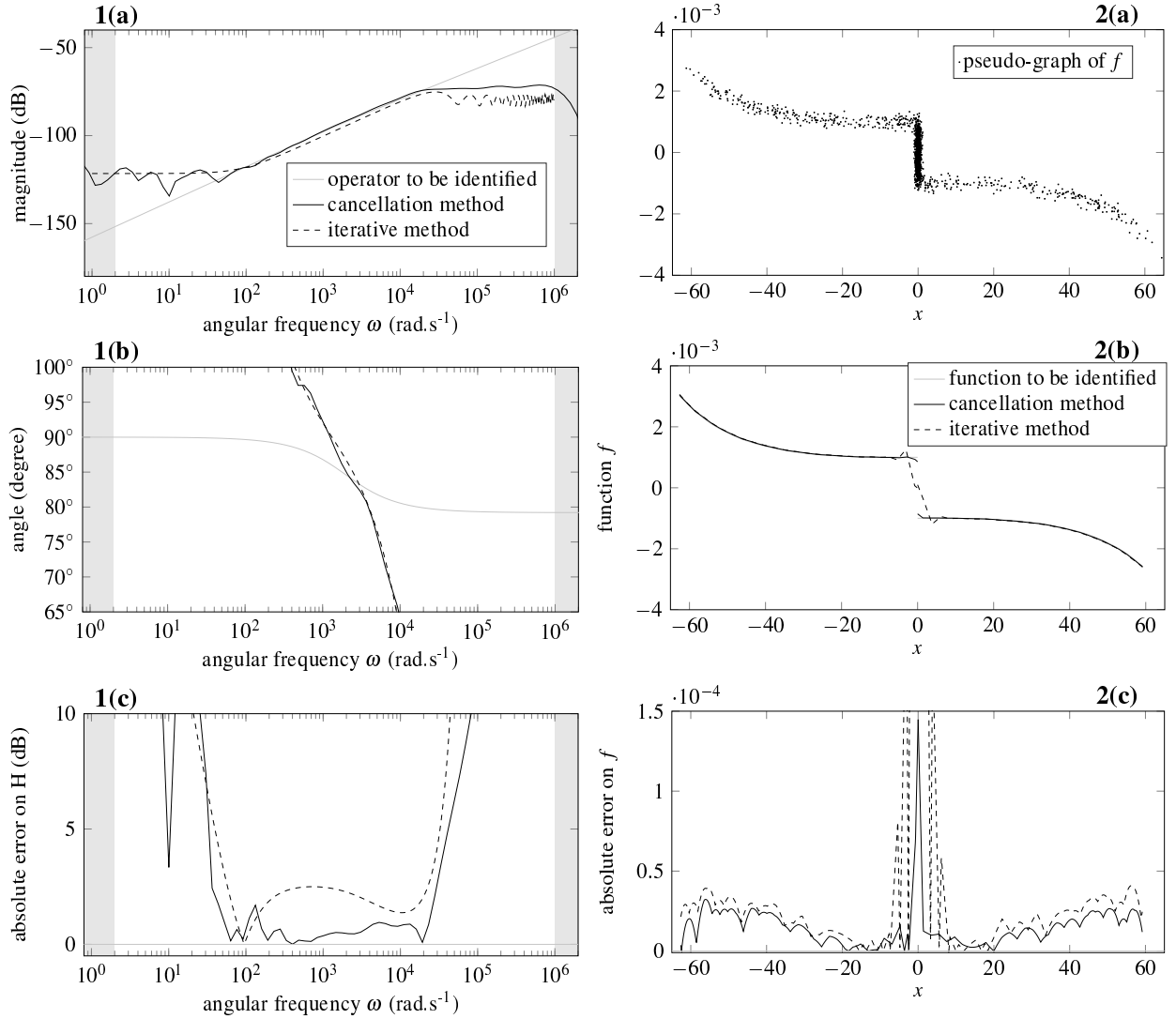


Figure 8: **Identification of a Volterra model with a discontinuous function f (section 6.4).** Left column - identification of the operator $H(\partial_t)$: (a) magnitude of $H(i\omega)$; (a) angle of $H(i\omega)$; (a) absolute error on the magnitude of $H(i\omega)$. Right column - identification of the function f : (a) pseudo-graph of f ; (b) comparison between the identified function \hat{f} and the theoretical one; (c) absolute error between f and \hat{f} . The results obtained with the cancellation method are plotted in solid black line, the ones obtained with the iterative method are plotted in dashed black line, and the theoretical function is plotted in grey solid line.

are still less accurate than the ones of the cancellation method, and still subject to an identification bias. The function f , as well as its set-valued extension \tilde{f} (see (112)), are correctly reconstructed with the cancellation method, even around the discontinuity. Although the function f is not Lipschitz-continuous around $x = 0$, property (iii) of proposition 12 remains satisfied around this point. In comparison, the iterative method fails to well-identify the function f around the discontinuity, despite the fact that the same basis functions have been used for both identification methods. It comes from the fact that, with the cancellation method, a pseudo-graph of f is available before beginning the identification of the nonlinear static term, which enables to adapt the identification method. In this particular example, the discontinuity of f at $X = 0$, which clearly appears on the numerical pseudo-graph, can be taken into account.

7 Conclusion

In the present paper, it has been shown that, by use of a particular difference operator, accurate identification of nonlinear (possibly singular or discontinuous) models is possible, at least in some cases, even if no information is available about the nonlinear terms to be identified.

This difference operator is mainly based on the construction of a set of couples of instants at which the measured trajectories take the same value. Thus it allows to transform the problem in such a way that the nonlinear term of the model is (quasi) cancelled. Then the identification of the residual linear unknown becomes a classical problem with straightforward solution. Once the unknown linear term of the model is identified, the identification of the nonlinear term is greatly simplified.

Some questions will of course need more in-depth investigations, in particular from the numerical point of view when the system is of high dimension. In such cases indeed, numerical complexity has to be correctly controlled.

Some extensions of this approach are currently under study, namely when the nonlinear term is no more an ordinary function but a set-valued function associated with some hysteresis operators.

A Notations used in the paper

The mathematical notations used in the paper are given hereafter.

- Given \mathbf{V} a Banach space and Ω a non-empty compact subset of \mathbb{R}^n , we denote by $C^0(\Omega; \mathbf{V})$ (resp. $L^\infty(\Omega; \mathbf{V})$) the Banach space of continuous functions defined on Ω (resp. essentially bounded measurable functions defined on Ω) with values in \mathbf{V} . The space $C^0(\Omega; \mathbf{V})$ identifies with a closed subspace of $L^\infty(\Omega; \mathbf{V})$, the topology of which is the one of the uniform convergence, that is:

$$\|v\| = \sup_{t \in \Omega} \|v(t)\|_{\mathbf{V}}. \quad (113)$$

For brevity, we denote:

$$L^\infty_{\mathbf{V}}(\Omega) := L^\infty(\Omega; \mathbf{V}), L^2_{\mathbf{V}}(\Omega) := L^2(\Omega; \mathbf{V}), C^0_{\mathbf{V}}(\Omega) := C^0(\Omega; \mathbf{V}) \quad (114)$$

and, as usual: $L^\infty(\Omega) = L^\infty_{\mathbb{R}}(\Omega)$, $L^2(\Omega) = L^2_{\mathbb{R}}(\Omega)$, $C^0(\Omega) = C^0_{\mathbb{R}}(\Omega)$.

- Given \mathbf{V}_0 a closed subset of \mathbf{V} , we denote:

$$C^0_{\mathbf{V}_0}(\Omega) := \{v \in C^0_{\mathbf{V}}(\Omega); \forall t \in \Omega, v(t) \in \mathbf{V}_0\}; \quad (115)$$

the set $C^0_{\mathbf{V}_0}(\Omega)$ is a closed subset of $C^0_{\mathbf{V}}(\Omega)$.

- Let \mathbf{V}^i be Banach spaces; then the product space $\prod_{i=1}^I \mathbf{V}^i$ is a Banach space when equipped with the norm:

$$\|v\| = \sup_{i=1:I} \|v^i\|_{\mathbf{V}^i}. \quad (116)$$

- With \mathbf{V} and \mathbf{W} two Banach spaces, we denote by $\mathcal{L}(\mathbf{V}, \mathbf{W})$ the vector space of all the linear continuous operators from \mathbf{V} into \mathbf{W} ; $\mathcal{L}(\mathbf{V}, \mathbf{W})$ is a Banach space when equipped with the norm:

$$\|f\| = \sup_{\|v\|_{\mathbf{V}} \leq 1} \|f \cdot v\|_{\mathbf{W}}. \quad (117)$$

B About the numerical implementation of Step 1 (case $\Omega^j \subset \mathbb{R}$)

In this appendix, we give some details about the practical implementation of the Step 1 (see section 3.2) of the method in the case where $\Omega^j \subset \mathbb{R}$. We present one method allowing to construct a finite subset $\omega_{x_m, \varepsilon, d_t}$ of $\Omega_{x_m, \varepsilon}$ from a measurement x_m of x , in the case where $\Omega^i \subset \mathbb{R}$ and $\mathbf{Z} = \mathbb{R}$. This simple method has been used for the numerical examples of section 6. The case of spaces of higher dimension requires some technical and/or numerical adaptations.

Consider I trajectories (x^i, A^i, z^i) , $i = 1 : I$, solution of (1) for some given g and $y \in F$ to be identified. Consider also I sets of discrete data $x_m^{i,k}, z_m^{i,k}$ defined by:

$$x_m^{i,k} = x^i(t_k^i) + \varepsilon_x^{i,k}, \quad z_m^{i,k} = z^i(t_k^i) + \varepsilon_z^{i,k}, \quad (118)$$

with $\{t_k^i\}_{k=1:K_i} \subset \Omega^i$, and $\{\varepsilon_x^{i,k}\}_{k=1:K_i}$ and $\{\varepsilon_z^{i,k}\}_{k=1:K_i}$ some zero mean noises with finite (and sufficiently small) standard deviations.

We denote $x_m := (x_m^i)$, $z_m := (z_m^i)$ where x_m^i and z_m^i are the *continuous* trajectories obtained from a continuous *interpolation* process (linear, splines [45], etc.) and a possible pre-filtering of the data $x_m^{i,k}$ and $z_m^{i,k}$. Let $\{x_0^l\}_{l=1:L}$ be a mesh of the (bounded) set $\bigcup_i x_m^i(\Omega^i)$.

Given $\varepsilon > 0$, we proceed in three steps to construct a finite subset $\omega_{x_m, \varepsilon, d_t}$ of $\Omega_{x_m, \varepsilon}$:

1. First, for any i and any l , we determine (by analytic computation of real intervals) the open sets $\omega_{x_m, \varepsilon, l}^i := x_m^{i-1}(|x_0^l - \frac{\varepsilon}{2}|, x_0^l + \frac{\varepsilon}{2}|)$, that is the sets of t such that $|x_m^i(t) - x_0^l| < \frac{\varepsilon}{2}$.

2. Then, from the sets $\omega_{x_m, \varepsilon, l}^i$, we build the subsets $\omega_{x_m, \varepsilon}^{i,j}$ defined by:

$$\omega_{x_m, \varepsilon}^{i,j} := \bigcup_{l=1:L} \omega_{x_m, \varepsilon, l}^i \times \omega_{x_m, \varepsilon, l}^j, \quad (119)$$

that are such that: $\omega_{x_m, \varepsilon}^{i,j} \subset \Omega_{x_m, \varepsilon}^{i,j}$. Indeed:

$$(t, \tau) \in \omega_{x_m, \varepsilon}^{i,j} \Leftrightarrow \exists l, \left\{ \begin{array}{l} |x_m^i(t) - x_0^l| < \frac{\varepsilon}{2} \\ |x_m^j(\tau) - x_0^l| < \frac{\varepsilon}{2} \end{array} \right\} \Rightarrow |x_m^i(t) - x_m^j(\tau)| \leq \varepsilon \Leftrightarrow (t, \tau) \in \Omega_{x_m, \varepsilon}^{i,j}. \quad (120)$$

Although in general $\omega_{x_m, \varepsilon}^{i,j} \subsetneq \Omega_{x_m, \varepsilon}^{i,j}$, these subsets can remain rich enough in practice to ensure an accurate identification.

3. Finally, by choosing a time step d_t small enough, we can discretize the 2D-sets $\omega_{x_m, \varepsilon}^{i,j}$ to get some finite subsets $\omega_{x_m, \varepsilon, d_t}^{i,j} \subset \omega_{x_m, \varepsilon}^{i,j}$. The finite set:

$$\omega_{x_m, \varepsilon, d_t} := \bigcup_{i,j} \left(\{(i, j)\} \times \omega_{x_m, \varepsilon, d_t}^{i,j} \right) \subset \Omega_{x_m, \varepsilon}, \quad (121)$$

can then be used for numerical purposes. By denoting $Q := \text{card}(\omega_{x_m, \varepsilon, \Delta t}) = \sum_{(i,j) \in \mathcal{I}_{x, \varepsilon}} \text{card}(\omega_{x_m, \varepsilon, d_t}^{i,j})$, the set $\omega_{x_m, \varepsilon, d_t}$ can then be simply denoted by $\{(i^q, j^q, \tau^q, t^q)\}_{q=1:Q}$.

References

- [1] Ljung L. *System identification: theory for the user*. Prentice-Hall Englewood Cliffs, NJ, 1987.
- [2] Söderström T, Stoica P, Johansson R. System identification. *Journal of Dynamic Systems, Measurement, and Control* 1993; **115**:739.
- [3] Rao G, Unbehauen H. Identification of continuous-time systems. *IEE Proceedings - Control Theory and Applications*, vol. 153, IET, 2006; 185–220.
- [4] Garnier H, Wang L. *Identification of continuous-time models from sampled data*. Springer, 2008.
- [5] Ljung L. Approaches to identification of nonlinear systems. *Proceedings of the 29th Chinese Control Conference (CCC 2010)*, Beijing, China, 2010.
- [6] Alanqar A, Durand H, Christofides PD. On identification of well-conditioned nonlinear systems: Application to economic model predictive control of nonlinear processes. *AIChE Journal* 2015; **61**(10):3353–3373.

- [7] Venkatesh S, Dahleh M. On learning the input-output behaviour of nonlinear fading memory systems from finite data. *International Journal of Robust and Nonlinear Control* 2000; **10**:931–959.
- [8] Hunter IW, Korenberg MJ. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological cybernetics* 1986; **55**(2):135–144.
- [9] Rugh WJ. *Nonlinear system theory - The Volterra/Wiener Approach*. Johns Hopkins University Press Baltimore, 1981.
- [10] Gruber JK, Ramirez DR, Limon D, Alamo T. A convex approach for NMPC based on second order Volterra series models. *International Journal of Robust and Nonlinear Control* 2015; **25**(18):3546–3571.
- [11] Billings SA. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, 2013.
- [12] Greblicki W. Continuous-time Hammerstein system identification. *IEEE Transactions on Automatic Control* 2000; **45**(6):1232–1236.
- [13] Billings S, Fakhouri S. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica* 1982; **18**(1):15–26.
- [14] Tuch J, Feuer A, Palmor Z. Time delay estimation in continuous linear time-invariant systems. *IEEE Transactions on Automatic Control* 1994; **39**(4):823.
- [15] Mouyon P, Imbert N. Identification of a 2D turbulent wind spectrum. *Aerospace Science and Technology* 2002; **6**(8):3599–605.
- [16] Bidan P, Neacsu C, Lebey T. Pseudo-differential models for propagation and dissipation phenomena in electrical machine windings. *14th International Symposium of Mathematical Theory of Networks and Systems (MTNS'2000)*, Perpignan (France), 2000.
- [17] Casenave C, Montseny G. Identification and state realization of non-rational convolution models by means of diffusive representation. *IET Control Theory and Applications* 2011; **5**(7):934–942.
- [18] Favini A, Lorenzi A. Identification problems for singular integro-differential equations of parabolic type II. *Nonlinear Analysis: Theory, Methods & Applications* 2004; **56**(6):879–904.
- [19] Ding F, Chen T. Identification of Hammerstein nonlinear ARMAX systems. *Automatica* 2005; **41**(9):1479–1489.
- [20] Casenave C. Time-local formulation and identification of implicit Volterra models by use of diffusive representation. *Automatica* 2011; **47**(10):2276–2278.
- [21] Bai EW. Decoupling the linear and nonlinear parts in Hammerstein model identification. *Automatica* 2004; **40**(4):671–676.
- [22] Stoica P. On the convergence of an iterative algorithm used for Hammerstein system identification. *IEEE Transactions on Automatic Control* 1981; **26**(4):967–969.
- [23] Li G, Wen C. Convergence of fixed-point iteration for the identification of Hammerstein and Wiener systems. *International Journal of Robust and Nonlinear Control* 2013; **23**(13):1510–1523.
- [24] Haber R, Keviczky L. The identification of the discrete-time Hammerstein model. *Periodica Polytechnica Electrical Engineering* 1974; **18**(1):71–84.

- [25] Müller S, Kästner M, Brummund J, Ulbricht V. A nonlinear fractional viscoelastic material model for polymers. *Computational Materials Science* 2011; **50**(10):2938–2949.
- [26] Das S, Gupta P, Rajeev. A Fractional Predator-Prey Model and its Solution. *International Journal of Nonlinear Sciences and Numerical Simulation* 2009; **10**(7).
- [27] Tabares PCC, Ferreira JD, Rao V. Weak Allee effect in a predator-prey system involving distributed delays. *Computational & Applied Mathematics* 2011; **30**(3):675–699.
- [28] Wolkowicz GSK, Xia H, Ruan S. Competition in the chemostat: a distributed delay model and its global asymptotic behavior. *SIAM Journal on Applied Mathematics* 1997; **57**(5):1281–1310.
- [29] Joulin G. Point-source initiation of lean spherical flames of light reactants: an asymptotic theory. *Combustion Science and Technology* 1985; **43**(1-2):99–113.
- [30] Rumeau A, Bidan P, Lebey T, Marchin L, Barbier B, Guillemet S. Behavior modeling of a $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$ ceramic for capacitor applications. *IEEE Conference on Electrical Insulation and Dielectric Phenomena*, Kansas City (Missouri USA), 2006.
- [31] Deimling K. *Multivalued differential equations*. Walter de Gruyter, 1992.
- [32] Liberzon D. *Switching in systems and control*. Springer, 2003.
- [33] Prieur C, Goebel R, Teel AR. Hybrid feedback control and robust stabilization of nonlinear systems. *IEEE Transaction on Automatic Control* 2007; **52**(11):2103–2117.
- [34] Ben-Israel A, Greville T. *Generalized inverses: Theory and applications*. Springer Verlag: New York, USA, 2003.
- [35] Mahata K, Garnier H. Direct identification of continuous-time errors-in-variables models. *Proceedings of the 16th Triennial IFAC World Congress on Automatic Control*, Prague (Czech Republic), 2005; 3–8.
- [36] Welsh J, Goodwin G, Garnier H. A simple method for bias reduction in time domain least squares parameter estimation. *3rd IFAC Symposium on System, Structure and Control, SSSC'07*, Foz do Iguçu(Brasil), 2007.
- [37] Casenave C, Montseny E, Camon H. Identification of nonlinear dynamic models of electrostatically actuated MEMS. *Control Engineering Practice* 2010; **18**(8):954 – 969.
- [38] Stojanovic V, Nedic N. Robust Kalman filtering for nonlinear multivariable stochastic systems in the presence of non-Gaussian noise. *International Journal of Robust and Nonlinear Control* 2016; **26**(3):445–460.
- [39] Montseny G. *Représentation diffusive*. Hermès-science: Paris, 2005.
- [40] Casenave C, Montseny G. Introduction to Diffusive Representation. *4th IFAC Symposium on System, Structure and Control, SSSC 2010*, Ancona, Italy, 2010.
- [41] Montseny G. Simple approach to approximation and dynamical realization of pseudodifferential time-operators such as fractional ones. *IEEE Transaction on Circuits and System II* 2004; **51**(11):613–618.
- [42] Casenave C, Montseny G. Diffusive Identification of Volterra Models by Cancellation of the Nonlinear Term. *15th IFAC Symposium on System Identification, SYSID 2009*, Saint-Malo, France, 2009.
- [43] Bojanczyk A, Brent R, Hoog F. QR factorization of Toeplitz matrices. *Numerische Mathematik* 1986; **41**(1):81–94.

- [44] Visintin A. *Differential models of hysteresis*. Springer: Berlin, 1994.
- [45] Schumaker L. *Spline functions: Basic Theory*. Cambridge Mathematical Library, 2007.