Time-local formulation and identification of implicit Volterra models by means of diffusive representation

Céline Casenave

Time-Local Formulation and Identification of Implicit Volterra Models by use of Diffusive Representation

Céline Casenave [a,b]

[a] CNRS; LAAS; 7 avenue du colonel Roche, F-31077 Toulouse, France.

[b] Université de Toulouse; UPS, INSA, INP, ISAE; LAAS; F-31077 Toulouse, France.

Abstract

We present a time-continuous identification method for nonlinear dynamic Volterra models of the form $\mathcal{H}X = f(u, X) + v$ with $\mathcal{H}$ a causal convolution operator. It is mainly based on a suitable parameterization of $\mathcal{H}$ deduced from the so-called diffusive representation, which is devoted to state representations of integral operators. Following this approach, the complex dynamic nature of $\mathcal{H}$ can be summarized by a few numerical parameters on which the identification of the dynamic part of the model will focus. The method is validated on a physical numerical example.

Key words: system identification; least-squares method; nonlinear Volterra model; implicit model; nonrational operator; state realization; diffusive representation.

1 Introduction

Numerous methods have been developed for the identification of continuous-time dynamic models of finite dimension [8, 21], such as the ones characterized by rational transfer functions. Beyond the finite dimensional case, classes of specific identification models of infinite dimension, for example for time-delay systems [20], non standard stochastic processes [15] or distributed physical phenomena [3], have also been considered in some works. In general models of low dimension are indeed not sufficient to describe accurately some complex dynamic phenomena, which require more sophisticated operators. For example, when some distributed underlying phenomena are involved, significant nonrational dynamic components are quasi systematically generated.

In this paper, we consider some dynamic models of the form:
$$\mathcal{H}X = f(u, X) + v, \qquad (1)$$
where $u(t)$, $v(t)$, $X(t) \in \mathbb{R}$, $t > 0$, f is a locally integrable nonlinear function defined on $\mathbb{R}^2$, and $\mathcal{H}$ is the causal convolution operator of impulse response [1] h, that is $\mathcal{H} : y \mapsto h * y$. By denoting H the Laplace-transform of function h, the operator $\mathcal{H}$ can be symboli-

cally rewritten [2] $\mathcal{H} = H(\partial_t)$. The operator $H(\partial_t)$ is also assumed to be invertible, so model (1) can be rewritten $X = H(\partial_t)^{-1}(f(u, X) + v)$, that is under the standard Volterra's form:
$$X(t) = \int_0^t k(t - s) \left[f\left(u(s), X(s)\right) + v(s) \right] ds, \qquad (2)$$

with k the impulse response of the operator $H(\partial_t)^{-1}$. In general $H(\partial_t)^{-1}$ will be a nonrational operator, and so it cannot be realized by a finite-dimensional input-output state equation.

In the sequel, (1) will be referred as the *implicit* form of the model, because this equation cannot be solved by simple quadrature of an integral, and so is ill adapted to numerical simulation for example. On the other side, (2) will be called the *explicit* form of the model. Implicit (nonrational) Volterra models of the form (1) are frequently encountered in various domains, as in thermic phenomena [9], electrical engineering [18], control [6, 2], chemical processes [7], electronics [10], etc.

Note that various problems related to identification or estimation and based on some specific Volterra models or Volterra series have been studied in the two last decades; see, for example, [11, 12, 16, 17].

In many concrete situations, the operator $H(\partial_t)$ and/or

---

Email address: `casenave@laas.fr` (Céline Casenave).

[1] Note that h is not a locally integrable function but a singular distribution.

[2] As usual, ∂_t denotes the time derivative operator.

the function f can be ill known. In these cases, the model (1) must be identified, that is some or all of its components must be estimated from physical measures. When the impulse response h is unknown, identifying such models presents several difficulties resulting from both the presence of the convolution operator $H(\partial_t)$ and the dynamic coupling between this operator and the (static) nonlinear function f, via equation (2). The main origin of these difficulties lies in the infinite-dimensionality of the distribution h (or k), which *a priori* needs a great number of coefficients to be accurately described under numerical approximation.

Under the implicit form (1) however, the nonlinearity (represented by function f) and the convolution (by h) are formally decoupled: this property is used in the identification method described in this paper. This method is based on a parameterization of $H(\partial_t)$ deduced from the so-called diffusive representation [5, 13], devoted to time-local state realizations of integral linear operators. Following this approach, the complex dynamic nature of $H(\partial_t)$ can be summarized by a few parameters on which the numerical identification problem will focus. At the same time, the function f is decomposed on a suitable functions' basis, in such a way that we get an equivalent linear-in-parameters model. Under numerical approximation, the identification problem is then of reasonable dimension thanks to the properties of diffusive representation. Thus, standard identification methods can be used, namely least-squares methods.

The paper is organized as follows. In section 2, we recall some basic notions about the diffusive representation, which will be used in the sequel. In section 3, we describe the theoretical principle of the identification method, while we give in section 4 a few indications about the numerical implementation. Then, in section 5, the method is tested on a numerical example for illustration.

2 Preliminaries: diffusive formulation of causal convolution operators

In this section, we present a particular case of a methodology introduced and developed in [13] in a general framework. Details about the proofs can be found in [5, 14].

We denote $\mathcal{L}$ the Laplace transform, $\mathbf{1}_A$ the characteristic function of the set A (that is $\mathbf{1}_A(x) = 1$ if $x \in A$, 0 if $x \notin A$) and ∂_t^{-1} the operator [3] $f \mapsto \int_0^t f(s)\,ds$, whose transfer function is $\frac{1}{p}$. We consider a causal convolution operator defined, on any locally integrable function $v : \mathbb{R}^+ \to \mathbb{R}$, by:

$$v \mapsto \int_0^t q(t-s)\,v(s)\,ds. \qquad (3)$$

We denote $Q = \mathcal{L}q$ the Laplace transform of the locally integrable function q, and $Q(\partial_t)$ the convolution operator defined by (3).

Let $v^t(s) := \mathbf{1}_{]-\infty,t]}(s)\,v(s)$ be the restriction of v to its past and $v_t(s) := v^t(t-s)$ the so-called "history" of v. From causality of $Q(\partial_t)$, we deduce:

$$[Q(\partial_t)(v - v^t)](t) = 0 \text{ for all } t; \qquad (4)$$

then, we have for any continuous function v:

$$[Q(\partial_t)\,v](t) = \left[\mathcal{L}^{-1}\left(Q\,\mathcal{L}v\right)\right](t) = \left[\mathcal{L}^{-1}\left(Q\,\mathcal{L}v^t\right)\right](t). \quad (5)$$

By denoting $\Psi_v(t,p) := \mathrm{e}^{pt}\left(\mathcal{L}v^t\right)(p) = \left(\mathcal{L}v_t\right)(-p)$, by computation of $\partial_t \mathcal{L}v_t$, Laplace inversion and use of (5), we have:

Lemma 1 *1. The function Ψ_v is the solution of the differential equation:*

$$\partial_t \Psi(t,p) = p\,\Psi(t,p) + v, \ \ t > 0, \ \ \Psi(0,p) = 0. \quad (6)$$

2. $[Q(\partial_t)\,v]\,(t) = \frac{1}{2i\pi}\int_{b-i\infty}^{b+i\infty} K(p)\,\Psi_v(t,p)\,dp, \ \forall b > 0.$

We denote Ω the holomorphic domain of Q (after analytic continuation). Let $\gamma : \mathbb{R} \to \mathbb{C}^-$ be a continuous function defining a closed [4] simple arc in $\mathbb{C}^-$, also denoted γ for simplicity. We denote $\Omega_\gamma^- \subset \mathbb{C}^-$ the interior domain defined by γ, and Ω_γ^+ the complementary of $\Omega_\gamma^- \cup \gamma$. By use of standard techniques (Cauchy's theorem, Jordan's lemma), it can be shown:

Lemma 2 *For any $\gamma \subset \Omega$ such that Q is holomorphic in Ω_γ^+, if $Q(p) \to 0$ when $p \to \infty$ in Ω_γ^+, then:*

$$[Q(\partial_t)\,v]\,(t) = \frac{1}{2i\pi}\int_\gamma Q(p)\,\Psi_v(t,p)\,dp. \qquad (7)$$

Under hypotheses of lemma 2 (except $\gamma \subset \Omega$), we have [5, 13, 14]:

Theorem 3 *If the possible singularities of Q on γ are branching points in the neighborhood of which $|Q \circ \gamma|$ is locally integrable in the Lebesgue sense, then, with*

$$\mu = \frac{\gamma'}{2i\pi}\,Q \circ \gamma \ \ and \ \ \psi_v(t,.) = \Psi_v(t,.) \circ \gamma, \qquad (8)$$

$$we \ have: [5] \qquad [Q(\partial_t)\,v]\,(t) = \langle \mu, \psi_v(t,.) \rangle. \qquad (9)$$

Furthermore, $\psi_v(t,\xi)$ is the solution of the following evolution problem on $(t,\xi) \in \mathbb{R}^{*+} \times \mathbb{R}$:

$$\partial_t \psi(t,\xi) = \gamma(\xi)\,\psi(t,\xi) + v(t), \ \psi(0,\xi) = 0. \quad (10)$$

---

[3] Note that ∂_t^{-1} is indeed the inverse of ∂_t in a suitable algebra of causal convolution operators defined on functions with support in $\mathbb{R}^+$; this algebra is not explicitly described here for simplicity.

[4] Possibly at infinity.
[5] We use the notation $\langle \mu, \psi \rangle = \int \mu\,\psi\,d\xi$.

Definition 4 *The function μ is called the γ-symbol of operator $Q(\partial_t)$. The function ψ_v solution of (10) is called the diffusive representation of v.*

Formulation (10,9) can easily be extended to operators of the form $H(\partial_t) := Q(\partial_t) \circ \partial_t$ where $Q(\partial_t)$ admits a γ-symbol. We then have formally:

$$[H(\partial_t)\, v](t) = \langle \mu, \partial_t \psi_v(t,.)\rangle = \langle \mu, \gamma\, \psi_v(t,.) + v(t)\rangle, \quad (11)$$

with μ the γ-symbol of $Q(\partial_t) = H(\partial_t) \circ \partial_t^{-1}$.

One of the great advantages of the formulation (11), is its ability of generating simple, robust and accurate finite dimensional numerical approximations with few parameters (see section 4). It is namely the case when the contour γ satisfies a so-called "sector condition", for example when:

$$\gamma(\xi) = |\xi|\, \mathrm{e}^{\mathrm{i}\,\mathrm{sign}(\xi)(\frac{\pi}{2}+\alpha)} \text{ with } 0 < \alpha \lesssim \frac{\pi}{2}. \quad (12)$$

For the same precision, the numerical cost is then sometimes several orders lower than the one of standard integral quadratures. This cheap numerical cost will be of a great interest for identification problems because models with too many degrees of freedom in general intrinsically suffer from an excessive sensitivity to measurement noises.

Finally, the main advantage of the diffusive representation approach is that it refers to a unified and complete mathematical framework (see [13] for more details) including both rational and nonrational operators (and in particular rational approximations). Indeed, under diffusive representation, rational and nonrational operators have a common state representation (whose state variable ψ is governed by the same diffusive model). The only thing which differs from one formulation to the other is the γ-symbol. Nonrational operators (that is operators with no finite-dimensional state realization) are characterized by a γ-symbol μ with infinite support while, for rational ones, there exists a γ-symbol μ with finite support. As a consequence, efficient numerical approximations are easy to construct. The study of convergences and/or of error estimates is also greatly simplified by use of suitable topological spaces relating to the diffusive representation and γ-symbols spaces (which are linked by a topological duality product). These properties will be used in the following sections.

3   Theoretical principle of the proposed identification method

We consider the problem of identification of implicit Volterra models of the form (1). We suppose the equation (1) is well-posed in the sense of existence, uniqueness and continuous dependence of the solution X with respect to the input (u,v). The goal is to estimate both operator $H(\partial_t)$ and function f from some measures u_m, v_m and X_m of u, v and X in order to get a model sufficiently accurate. The proposed method consists:

- first in parameterizing both operator $H(\partial_t)$ and function f to get an equivalent linear-in-parameters model,
- then in using standard time-continuous system identification methods based on least squares minimization problem to estimate the parameters of the model.

Remark 5 *Given (u,v,X), the solution $(\mathcal{H}, f)$ of (1) is obviously not unique: for any $c \in \mathbb{R}^*$, the model $(\mathcal{H} + cI)X = f(u,X) + cX + v$ is equivalent to (1).*

▶ **Parameterization of** $H(\partial_t)$. Let suppose there exists a separable Hilbert space $\mathcal{E}$ such that $\mu \in \mathcal{E}'$ and $\gamma\, \psi_X(t,.) + X(t) \in \mathcal{E}$ t-a.e.; this will be the case in most of the concrete situations [13]. We then have:

$$H(\partial_t)\, X = \mathcal{A}_X\, \mu, \quad (13)$$

where $\mathcal{A}_X$ is the linear operator defined by:

$$\mathcal{A}_X : \mu \longmapsto \left(t \mapsto \langle \mu, \gamma\, \psi_X(t,.) + X(t)\rangle_{\mathcal{E}',\mathcal{E}}\right). \quad (14)$$

▶ **Parameterization of function** f. We consider a topological basis (that we can suppose orthonormal) $\{\mathbf{g}^p \otimes \mathbf{k}^q\}_{p,q=1:+\infty}$ of a tensorial product $\mathcal{K} = \mathcal{K}_1 \otimes \mathcal{K}_2$ of separable Hilbert spaces, to which f is supposed to belong. We then have:

$$f = \sum_{p,q} a_{pq}\, \mathbf{g}^p \otimes \mathbf{k}^q = a \cdot \mathbf{g} \otimes \mathbf{k},\ a = (a_{pq}) \in \ell^2 \otimes \ell^2. \quad (15)$$

▶ **Optimal estimation of** (μ, a). From the above, model (1) can then be equivalently rewritten under the new following (implicit) time-local form:

$$\begin{cases} \partial_t \psi_X - \gamma\, \psi_X = X \\ \phi_{u,X}\, \theta = v. \end{cases} \quad (16)$$

with $\theta := (\mu, a)$, and $\phi_{u,X}$ the linear operator defined by:

$$\phi_{u,X} : \theta = (\mu, a) \mapsto \mathcal{A}_X\, \mu - [a \cdot (\mathbf{g} \otimes \mathbf{k})] \circ (u,X). \quad (17)$$

This formulation is suitable for identification purpose thanks to the linear dependence with respect to the unknown coefficients μ and a.

With u_m, v_m and X_m some measures of u, v and X, we then get the linear regression equation:

$$v_\mathrm{m} = \phi_{u_\mathrm{m}, X_\mathrm{m}}\, \theta + \varepsilon(\theta), \quad (18)$$

where $\varepsilon(\theta)$ is the so-called equation error. The estimate $\hat{\theta}$ of the unknown parameters $\theta = (\mu, a)$ can then be

defined as the solution of the least squares problem:

$$\min_{\theta \in \mathcal{E}' \times \ell^2 \otimes \ell^2} \|\varepsilon(\theta)\|^2_{\mathcal{F}}, \qquad (19)$$

where $\mathcal{F}$ is a suitable separable Hilbert space. Under the hypothesis that $\phi_{u_{\mathrm{m}}, X_{\mathrm{m}}}(\mathcal{E}' \times \ell^2 \otimes \ell^2)$ is closed in $\mathcal{F}$, the solution of (19) is then formally obtained by orthogonal projection, that is:

$$\hat{\theta} = (\hat{\mu}, \hat{a}) = \phi^{\dagger}_{u_{\mathrm{m}}, X_{\mathrm{m}}} v_{\mathrm{m}}, \qquad (20)$$

where $\phi^{\dagger}_{u_{\mathrm{m}}, X_{\mathrm{m}}}$ is the pseudo-inverse of $\phi_{u_{\mathrm{m}}, X_{\mathrm{m}}}$ [1]. In the sense of the Hilbertian norm of $\mathcal{F}$, this estimation is optimal. When at least one of the two measures X_{m} and u_{m} is noisy, the estimator $\hat{\theta}$ is in general *biased* because $\phi^{\dagger}_{u_{\mathrm{m}}, X_{\mathrm{m}}}$ depends on the measurement noise. To mitigate this problem, it will be interesting to use some classical bias reduction methods.

▶ **Identified model**. The physical system under consideration can then be described by the identified model $\hat{H}(\partial_t) X = \hat{f}(u, X) + v$ where $\hat{H}$ and $\hat{f}$ are deduced from the identified parameters $\hat{\mu}$ and $\hat{a}$, or equivalently by the (implicit) diffusive augmented form deduced from (16):

$$\begin{cases} \partial_t \psi = \gamma \psi + X, \ \psi(0,.) = 0 \\ \langle \hat{\mu}, \gamma \psi + X \rangle = \hat{f}(u, X) + v. \end{cases} \qquad (21)$$

The dynamic model (21) is yet implicit and so ill adapted to numerical simulation. However it is well adapted to some problems as for example motion planning. Indeed, given a trajectory X defined on $t \in [0, T]$, the associated input (u, v) can be deduced by solving the static equation $v + \hat{f}(u, X) = \langle \hat{\mu}, \gamma \psi + X \rangle$ after having integrated (up to numerical approximations) the set of linear differential equations which governs $\psi(., \xi)$, $\xi \in \mathbb{R}$. Note that this equation is trivially solved if f is of the form $\hat{f}(u, X) = \hat{h}(X)$ or $\hat{f}(u, X) = \hat{h}(X) u$. For many other problems however, an explicit formulation would be preferable. When operator $K(\partial_t) = H(\partial_t)^{-1}$ admits a γ-symbol ν, such a formulation can be derived from (2), under the classical (but infinite dimensional) state-space input-output form [6]:

$$\begin{cases} \partial_t \varphi = \gamma \varphi + \hat{f}(u, X) + v, \ \varphi(0,.) = 0, \\ X = \langle \hat{\nu}, \varphi \rangle, \end{cases} \qquad (22)$$

where φ is the γ-representation of $\hat{f}(u, X) + v$ and $\hat{\nu}$ is the γ-symbol associated with $\hat{H}(\partial_t)^{-1}$. This last γ-symbol

---

[6] Note that X appears here as an output of the (explicit) differential dynamic system.

can be computed from $\hat{\mu}$ by (numerical) γ-symbolic inversion techniques [4].

▶ **Regularized problem**. However, note that in practice, the closedness of $\phi_{u_{\mathrm{m}}, X_{\mathrm{m}}}(\mathcal{E}' \times \ell^2 \otimes \ell^2)$ can be a too strong hypothesis for usual spaces $\mathcal{F}$ (such as $L^2(0, T)$ for example). So, it is preferable to consider a regularized problem in place of (19) [19], namely:

$$\min_{\theta \in \mathcal{E}' \times \ell^2 \otimes \ell^2} J_\epsilon(\theta) \text{ with } J_\epsilon(\theta) := \|\varepsilon(\theta)\|^2_{\mathcal{F}} + \epsilon \|\theta\|^2_{\mathcal{E}' \times \ell^2 \otimes \ell^2}, \qquad (23)$$

where $\epsilon > 0$ is a small parameter. We then have the following results:

Theorem 6

1. *If $\phi := \phi_{u_{\mathrm{m}}, X_{\mathrm{m}}} : \mathcal{E}' \times \ell^2 \otimes \ell^2 \to \mathcal{F}$ is continuous and $v_{\mathrm{m}} \in \mathcal{F}$, then there exists a unique $\hat{\theta}_\epsilon \in \mathcal{E}' \times \ell^2 \otimes \ell^2$ such that $J_\epsilon(\hat{\theta}_\epsilon) = \min J_\epsilon$; it is given by:*

    $$\hat{\theta}_\epsilon = (\phi^* \circ \phi + \epsilon I)^{-1} \phi^*(v_{\mathrm{m}}). \qquad (24)$$

    Furthermore, if there exists $\hat{\theta} \in \mathcal{E}' \times \ell^2 \otimes \ell^2$ such that $\phi \hat{\theta} = v_{\mathrm{m}}$, then $\hat{\theta}_\epsilon \underset{\epsilon \to 0}{\to} \hat{\theta}^*$ with $\hat{\theta}^*$ such that $\|\hat{\theta}^*\| = \min\{\|\hat{\theta}\|; \ \phi \hat{\theta} = v_{\mathrm{m}}\}$.

2. *Let us consider a sequence $\mathcal{H}_n$ of finite dimensional vector spaces such that $\mathcal{H}_n \subset \mathcal{H}_{n+1}$ and $\bigcup_n \mathcal{H}_n$ is densely embedded in $\mathcal{E}' \times \ell^2 \otimes \ell^2$; then, with $\hat{\theta}_{\epsilon,n}$ solution of $\min_{\theta \in \mathcal{H}_n} J_\epsilon(\theta)$, we have: $\hat{\theta}_{\epsilon,n} \underset{n \to \infty}{\to} \hat{\theta}_\epsilon$.*

PROOF. The first part is a well known result, see for example [19]. For the second part, let us consider the Hilbert space $\mathcal{G} := \mathcal{F} \times \mathcal{E}' \times \ell^2 \otimes \ell^2$ with norm $\|(v, \theta)\|^2_{\mathcal{G}} := \|v\|^2_{\mathcal{F}} + \epsilon \|\theta\|^2_{\mathcal{E}' \times \ell^2 \otimes \ell^2}$. Because ϕ is continuous, $(\phi \theta_n, \theta_n) \xrightarrow{\mathcal{G}} (y, \theta) \Rightarrow \theta_n \to \theta \Rightarrow \phi \theta_n \to \phi \theta = y$, that is: $\phi(\mathcal{E}' \times \ell^2 \otimes \ell^2) \times \mathcal{E}' \times \ell^2 \otimes \ell^2$ is a closed subspace of $\mathcal{G}$. Furthermore, $J_\epsilon(\theta)$ is clearly the distance between $(v_{\mathrm{m}}, 0)$ and $(\phi \theta, \theta)$, and so, $(\phi \hat{\theta}_\epsilon, \hat{\theta}_\epsilon)$ (resp. $(\phi \hat{\theta}_{\epsilon,n}, \hat{\theta}_{\epsilon,n})$) is the orthogonal projection of $(v_{\mathrm{m}}, 0)$ on $\phi(\mathcal{E}' \times \ell^2 \otimes \ell^2) \times \mathcal{E}' \times \ell^2 \otimes \ell^2$ (resp. $\phi(\mathcal{H}_n) \times \mathcal{H}_n$). It follows that $(\phi \hat{\theta}_{\epsilon,n}, \hat{\theta}_{\epsilon,n})$ is also the orthogonal projection of $(\phi \hat{\theta}_\epsilon, \hat{\theta}_\epsilon)$ on $\phi(\mathcal{H}_n) \times \mathcal{H}_n$ and therefore, the distance between these two points goes to 0 thanks to the property $\overline{\bigcup_n \mathcal{H}_n} = \mathcal{E}' \times \ell^2 \otimes \ell^2$; in particular: $\|\hat{\theta}_{\epsilon,n} - \hat{\theta}_\epsilon\|^2_{\mathcal{E}' \times \ell^2 \otimes \ell^2} \to 0$.

Remark 7 *As said previously (remark 5), the solution of the problem of identification of model (1) is not unique. Nevertheless, as stated in the above theorem, uniqueness is restored by solving the problem (23) with $\epsilon \to 0$, in the sense that the asymptotic solution $(\hat{\mathcal{H}}^*, \hat{f}^*)$ is the one such that $\|\hat{\theta}^*\|$ is minimum.*

So, the spaces $\mathcal{E}$, $\mathcal{F}$, $\mathcal{K}$ must be chosen in such a way that $v_{\mathrm{m}} \in \mathcal{F}$, $f \in \mathcal{K}$, and $\phi_{u_{\mathrm{m}}, X_{\mathrm{m}}}$ is continuous (which, in particular, implies that $\langle \mu, \gamma \psi_{X_{\mathrm{m}}} + X_{\mathrm{m}} \rangle_{\mathcal{E}', \mathcal{E}} \in \mathcal{F}$ and $f(u_{\mathrm{m}}, X_{\mathrm{m}}) \in \mathcal{F}$). In practice, this choice is important. Indeed, thanks to the specific weight matrices induced under finite dimensional approximations, the well-posedness of (23) will insure the stability and convergence of numerical solutions. Such mathematical questions will be deepened in a further paper. Note that in most of the situations, the standard Hilbert space $\mathcal{F} = L^2(0, T)$ is well adapted.

4 Numerical implementation

Let (u, v, X) be a solution of (1). We consider a discrete data set $\{u_{\mathrm{m}}^k, v_{\mathrm{m}}^k, X_{\mathrm{m}}^k\}_{k=0:K}$ where $u_{\mathrm{m}}^k, v_{\mathrm{m}}^k$ and X_{m}^k are the respective measures of u, v and X at time $t_k = k \Delta t$. For convenience, we also denote u_{m}, v_{m} and X_{m} the associated continuous-time trajectories such that $u_{\mathrm{m}}(t_k) = u_{\mathrm{m}}^k$, $v_{\mathrm{m}}(t_k) = v_{\mathrm{m}}^k$ and $X_{\mathrm{m}}(t_k) = X_{\mathrm{m}}^k$.

To numerically solve (23), we have to build an approximate problem of finite dimension depending on the available data only.

▶ **Approximation of f.** It is obtained by truncation at finite order $P \times Q$ of the series (15):

$$f \simeq \sum_{p=1}^{P} \sum_{q=1}^{Q} a_{pq} \, \mathbf{g}^p \otimes \mathbf{k}^q. \qquad (25)$$

The values of P and Q have to be chosen in order to get the best compromise between the errors generated by truncation, and the ones generated by the presence of the noise, or even by structural defect of the model. In practice, such a choice is achieved empirically.

▶ **Approximation of μ.** We consider L-dimensional approximations μ^L of the γ-symbol μ, defined as sums of atomic measures [7] on a suitable mesh $\{\xi_l^L\}_{l=1:L}$:

$$\mu^L = \sum_{l=1}^{L} \mu_l^L \, \delta_{\xi_l^L}, \quad \mu_l^L \in \mathbb{C}. \qquad (26)$$

If $\cup_{L \leqslant \mathbf{L}} \{\xi_l^L\}$ is an increasing sequence, and if $\cup_{L < \infty} \{\xi_l^L\}$ is dense in $\mathbb{R}$, the following (weak) convergence holds with μ_l^L suitably chosen [13]:

$$\forall \psi, \ \langle \mu^L, \psi \rangle \xrightarrow[L \to +\infty]{} \langle \mu, \psi \rangle. \qquad (27)$$

In that sense, we get the approximation of $H(\partial_t)X$:

$$
\begin{aligned}
H^L(\partial_t)X &= \langle \mu^L, \gamma \psi_X + X \rangle \\
&= \sum_{l=1}^{L} \mu_l^L \left(\gamma(\xi_l^L) \psi_X(., \xi_l^L) + X \right) \simeq H(\partial_t)X,
\end{aligned}
$$

---
[7] δ_a denotes the Dirac measure at point a.

associated with the approximate transfer function [14]:

$$H^L(p) = \sum_{l=1}^{L} \frac{\mu_l^L \, p}{p + \gamma(\xi_l^L)} \simeq H(p). \qquad (28)$$

Most of the nonrational operators encountered in practice can be closely approximate with small L (see for example [14]). For the identification of dynamic operators, this is a great advantage because only a few numerical complex parameters μ_l^L have to be identified from experimental data. Moreover, the property (27) ensures the well-posedness and the robustness of the problem as soon as the operator to be identified admits a γ-symbol μ.

▶ **About the choice of γ and ξ_l^L:** to choose the function (and therefore the contour) γ, recall that the symbol $Q(p) = \dfrac{H(p)}{p}$ has to be analytic in Ω_γ^+ (see theorem 3). As a consequence, all the singularities of Q have to be inside the closed domain $\overline{\Omega_\gamma^-}$ delimited by γ. However, as the operator $H(\partial_t)$ is unknown, so is the position of the singularities of Q. So the contour γ will be chosen in such a way that the domain $\overline{\Omega_\gamma^-}$ is sufficiently large. In practice, thanks to physical underlying features, the singularities of Q are in $\mathbb{C}^-$; we take a contour of sector form (12). Thanks to high frequency dissipative properties, such a contour in general allows to accurately approximate $H(\partial_t)$ with very small L, typically around a few tens, with ξ_l^L geometrically spaced. However, note that the more γ is close to the axis $i\mathbb{R}$ (stability limit axis), the finer (and so the more expensive numerically) the discretization in ξ has to be in order to get good approximations of $H(\partial_t)$ and therefore, of the Volterra model under consideration. Empirically, the model can be first identified with a small α and then by increasing α (while decreasing L) as long as the identification results are good. Note that in practice, available information about the operator $H(\partial_t)$ can also help us in the choice of γ.

Some indications about the choice of the discretization points ξ_l^L can be found in [13, 14]. Note that from a discrete set of data $\{u_{\mathrm{m}}^k, v_{\mathrm{m}}^k, X_{\mathrm{m}}^k\}$, the frequency response $H(i\omega)$ can be identified only in a limited frequency band included in $[\frac{1}{t_K}; \frac{1}{2\Delta t}]$. Consequently, the band $[\xi_1^L; \xi_L^L]$ covered by the ξ-discretization will be chosen in such a way that [8] $[\frac{2\pi}{t_K}; \frac{2\pi}{2\Delta t}] \simeq [\min |\gamma(\xi_l^L)|; \max |\gamma(\xi_l^L)|]$.

▶ **Resolution of the numerical problem.** Finally, we consider the following numerical identification model (of finite dimension):

$$v_{\mathrm{m}}^k = \varphi_{\mathrm{m}}^T(t_k) \, \theta + \varepsilon(t_k, \theta), \ \ \theta = (\mu, a)^T, \qquad (29)$$

---
[8] Indeed, the quantity $\frac{|\gamma(\xi)|}{2\pi}$ is a cut-off frequency. In the particular case of (12), we have $|\gamma(\xi)| = |\xi|$.

where $\mu = (\mu_1^L, ..., \mu_L^L)$ and $a = (a_{11}, ..., a_{1Q}, a_{21}, ..., a_{PQ})$, $\varphi_{\mathrm{m}}^T(t_k) = [\Phi_1^k, ..., \Phi_L^k, F_{11}^k, ..., F_{1Q}^k, F_{21}^k, ..., F_{PQ}^k]$ with $F_{pq}^k = -\mathbf{g}^p(u_{\mathrm{m}}^k)\,\mathbf{k}^q(X_{\mathrm{m}}^k)$ and $\Phi_l^k = \gamma(\xi_l^L)\Psi_l^k + X_m^k$, Ψ_l^k being an approximation of $\psi_{X_{\mathrm{m}}}(t_k, \xi_l^L)$ computed from numerical integration of (10) with input X_{m}, that is, for any $k = 1 : K$, $l = 1 : L$:

$$\Psi_l^k = \mathrm{e}^{\gamma(\xi_l^L)\Delta t}\,\Psi_l^{k-1} + \frac{\mathrm{e}^{\gamma(\xi_l^L)\Delta t} - 1}{\gamma(\xi_l^L)}\,X_{\mathrm{m}}^{k-1};\ \Psi_l^0 = 0. \quad (30)$$

With $\mathcal{F} = \mathbb{R}^K$ equipped with the standard Euclidian norm (that is $\mathcal{F}$ is the discrete equivalent of $L^2(0,T)$), the associated least squares estimator $\hat{\theta} = (\hat{\mu}, \hat{a})$ is then given by:

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{C}^L \times \mathbb{R}^{P \times Q}} \left\{ \sum_{k=1}^{K} \varepsilon^2(t_k, \theta)\Delta t + \epsilon \sum_{j=1}^{L+P \times Q} |\theta_j|^2 \right\}$$

$$= \left[\sum_{k=1}^{K} \varphi_{\mathrm{m}}(t_k)\,\varphi_{\mathrm{m}}^T(t_k)\,\Delta t + \epsilon I \right]^{-1} \sum_{k=1}^{K} \varphi_{\mathrm{m}}(t_k)\,v_{\mathrm{m}}^k\,\Delta t.$$

Remark 8 *Up to technical adaptations, the method described above can easily be extended to the case of multiple data sets $\{u_{\mathrm{m}}^{j,k}, v_{\mathrm{m}}^{j,k}, X_{\mathrm{m}}^{j,k}\}$, $j = 1 : J$.*

5 Concrete example

For illustration, the identification method presented above is validated on a model of the electrical behavior of a capacitor realized in cubic Perovskite $CaCu_3Ti_4O_{12}$. This model was proposed in [18] and includes three components as shown in figure 1: a resistor r in serial with a capacitor of impedance Z_c in parallel with a nonlinear resistor of model: $i_G = G(V_c)$. We then have, with $r = 40\,\Omega$:

$$V = Z_c(\partial_t)(i - G(V - ri)) + ri, \quad (31)$$

$$\text{where} \quad Z_c(p) = \frac{\left(1 + 5.0 \times 10^{-4}p\right)^{0.12}}{13 \times 10^{-9}p}, \quad (32)$$

$$\text{and} \quad G(V_c) = c_1 \tan(c_2 V_c) + c_3 V_c, \quad (33)$$

with [9] $c_1 = 2.820 \times 10^{-3}$, $c_2 = 1.692 \times 10^{-2}$ and $c_3 = -4.771 \times 10^{-5}$. This model is of the form (1) with $X = V - ri$, $v = i$, $f(u, X) = -G(X)$, and $H(\partial_t) = Z_c^{-1}(\partial_t)$. As $\frac{H(p)}{p}$ is holomorphic in $\mathbb{C} \backslash \mathbb{R}^-$ and is such that $\frac{H(p)}{p} \to 0$ when $|p| \to \infty$, then $H(\partial_t) \circ \partial_t^{-1}$ admits a γ-symbol [10] which is defined for any $\xi \neq 0$, when γ is given by (12), by $\mu(\xi) = \frac{13 \times 10^{-9}\xi\,\mathrm{e}^{2\mathrm{i}\,\mathrm{sign}(\xi)(\frac{\pi}{2}+\alpha)}}{2\mathrm{i}\pi\left(1 + 5 \times 10^{-4}|\xi|\,\mathrm{e}^{\mathrm{i}\,\mathrm{sign}(\xi)(\frac{\pi}{2}+\alpha)}\right)^{0.12}}$.

---

[9] This function has been chosen to be comparable to the one in figure 1 of [18].

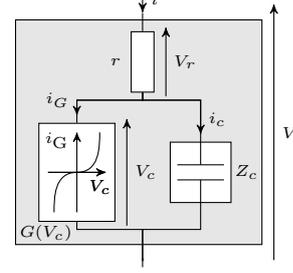[10] Note that μ is a locally integrable function.

Fig. 1. Model structure of the capacitor.

We consider in the sequel the problem of identification of $H(\partial_t)$ and f from $J = 8$ data sets $\{v_{\mathrm{m}}^{j,k}, X_{\mathrm{m}}^{j,k}\}_{j=1:J,k=0:K}$, obtained from accurate simulation (not described here) of (31). The input measures are supposed to be known exactly (no input measurement noise), that is $v_{\mathrm{m}}^{j,k} = v^{j,k} = v^j(k\Delta t)$ where the process v^j is gaussian, with power spectrum constant in the frequency band $[0, 10^6]$ and standard deviation 0.2. The sampled output data set associated with input $v^{j,k}$ is denoted $X^{j,k}$. The output measures $X_{\mathrm{m}}^{j,k}$ are defined by $X_{\mathrm{m}}^{j,k} = X^{j,k} + \eta^{j,k}$ with $\eta^{j,k}$ a Gaussian sequence with cut-off frequency 10^7 and spectral power density equal to σ in the frequency band $[0, 10^7]$, that is with standard deviation $\sigma\sqrt{10^7}$.
The time step is $\Delta t = 10^{-8}$ and we chose $K = 10^6$, so $t_{\max} = K\,\Delta t = 0.01\,\mathrm{s}$.

Because the operator $H(\partial_t) \circ \partial_t^{-1}$ admits a γ-symbol, it can be numerically identified under the form (26). We have taken $L = 60$ and the ξ_l are geometrically spaced between $\xi_1 = 10^2$ and $\xi_{60} = 10^8$, with γ defined by (12) with $\alpha = \frac{\pi}{4}$. The function f is identified under the form (25) with $P = 1$, $Q = 10$, $\mathbf{g}^1 \equiv 1$ and $\mathbf{k}^q$ some classical hat functions.

▶ **Noise free case** ($\sigma = 0$). For validation, the measurement data $X_{\mathrm{m}}^{j,k}$ are first supposed to be noise free (i.e. $\sigma = 0$). The Bode diagram of the identified frequency response $\hat{H}(\mathrm{i}\omega)$ of operator $\hat{H}(\partial_t)$ (deduced from the identified γ-symbol $\hat{\mu}$ by (28)) and the identified function $\hat{f}$ deduced from $\hat{a}$ are given in Fig. 2. The operator $H(\partial_t)$ is well approximated in the band $[10^3, 10^6] \subset [\frac{2\pi}{K\Delta t}, \frac{\pi}{\Delta t}] = [6.28 \times 10^2, 3.14 \times 10^8]$ (covered by the ξ_l). The identification of the nonlinear function f is also accurate.
To quantify the identification efficiency, we introduce the following relative errors:

$$\mathrm{E}^H = \frac{\int_{\xi_1^L}^{\xi_L^L} \left| \hat{H}(\mathrm{i}\omega) - H(\mathrm{i}\omega) \right| \frac{d\omega}{\omega}}{\int_{\xi_1^L}^{\xi_L^L} |H(\mathrm{i}\omega)| \frac{d\omega}{\omega}}, \mathrm{E}^f = \frac{\int_{X_{\mathrm{m}}}^{X_{\mathrm{M}}} \left| \hat{f}(X) - f(X) \right| dX}{\int_{X_{\mathrm{m}}}^{X_{\mathrm{M}}} |f(X)| \, dX},$$

where X_{m} and X_{M} are the minimal and maximal values of X reached by the data sets, that is, in this case,
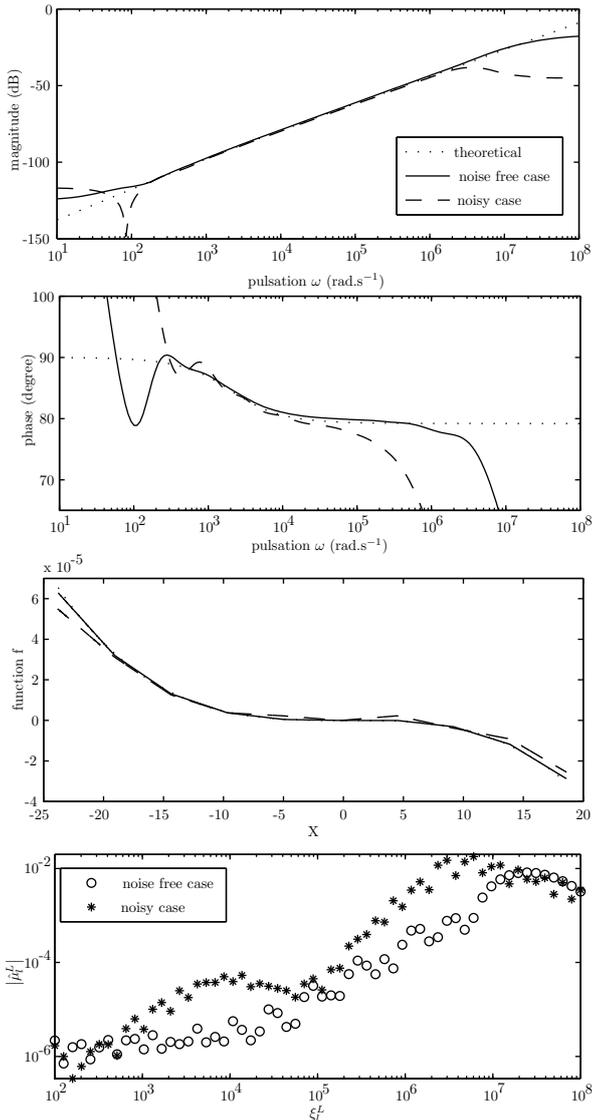
6

Fig. 2. **Identification results with and without measurement noise**. From top to bottom: magnitude and phase of the frequency response of the identified operator $\hat{Z}_c(\partial_t)$, curve of the identified function $\hat{f}$, identified parameters $\hat{\mu}_l^L$.

$X_{\mathrm{m}} = -23.8$ and $X_{\mathrm{M}} = 18.6$. We get:

$$\mathrm{E}^H = 1.72 \times 10^{-11}, \ \mathrm{E}^f = 2.1 \times 10^{-2}. \quad (34)$$

▶ **Noisy case**. The measures $X_{\mathrm{m}}^{j,k}$ are now supposed to be corrupted by additive measurement noise with standard deviation about 0.015V (that is $\sigma = 5 \times 10^{-6} \mathrm{V}/\sqrt{\mathrm{Hz}}$). The obtained results are given in Fig. 2.

The estimate of $H(\partial_t)$ is still good. Note that the frequency domain on which the operator $H(\partial_t)$ can be correctly identified is smaller than in the noise free case. This is a normal consequence of the presence of white noise: indeed, high frequencies are more corrupted than

low ones because, due to continuity, the spectral density of the trajectory X necessarily decreases 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.

Light oscillations appear on the identified $\hat{f}$, but the behavior of the estimate function is still very close to the theoretical one. We have in this case:

$$\mathrm{E}^H = 1.85 \times 10^{-11}, \ \mathrm{E}^f = 12.8 \times 10^{-2}. \quad (35)$$

6 Conclusion

The identification method introduced in this paper is mainly based on the diffusive representation approach and well adapted to a large class of nonlinear Volterra models. Its implementation on a simple numerical example gives good results. However, several problems still need to be studied in order to improve the efficiency of this method. For example, among the most significant ones, the involved Hilbertian norms should be judiciously chosen and adapted to the specific properties of the class of models under consideration. Estimation biases, intrinsically induced by such methods, must also be reduced, particularly when the measurement noise is important.

References

[1] A. Ben-Israel and T.N.E. Greville. *Generalized inverses: Theory and applications.* Springer Verlag, New York, USA, 2003.

[2] P. Bidan. Commande diffusive d'une machine électrique: une introduction. In *European Series on Applied and Industrial Mathematics (ESAIM): Proceedings*, volume 5, pages 55–68, 1998.

[3] P. Bidan, C. Neacsu, and T. Lebey. Pseudo-differential models for propagation and dissipation phenomena in electrical machine windings. In *14th International Symposium of Mathematical Theory of Networks and Systems (MTNS'2000)*, 19-23 June 2000.

[4] C. Casenave. *Représentation diffusive et inversion opératorielle pour l'analyse et la résolution de problèmes dynamiques non locaux.* PhD thesis, Université Paul Sabatier, Toulouse (France), 2009.

[5] C. Casenave and G. Montseny. Introduction to diffusive representation. In *4th IFAC Symposium on System, Structure and Control*, Ancona (Italy), 15-17 September 2010.

[6] F.-A. Devy-Vareta and G. Montseny. Pseudo-invariant H_2 multivariable robust control of a flexible beam. In *Proceedings of the European Control Conference, ECC 2010*, page 2638, Porto (Portugal), 2001.

[7] G. Garcia and J. Bernussou. Identification of the dynamics of a lead acid battery by a diffusive model.

European Series on Applied and Industrial Mathematics (ESAIM): Proceedings, 5:87–98, 1998.

[8] H. Garnier and L. Wang. *Identification of continuous-time models from sampled data.* Springer, 2008.

[9] G. Joulin. Point-source initiation of lean spherical flames of light reactants: an asymptotic theory. *Combustion science and technology*, 43(1-2):99–113, 1985.

[10] F. Lavernhe and J. Solhusvik. Fractional noises: diffusive model for CCD imager band-pass acquisition chain. In *European Series on Applied and Industrial Mathematics (ESAIM): Proceedings*, volume 5, pages 119–130, 1998.

[11] D. Mattera. Identification of polyperiodic Volterra systems by means of input-output noisy measurements. *Signal Processing*, 75(1):41–50, 1999.

[12] A. Monin and G. Salut. IIR Volterra filtering with application to bilinear systems. *IEEE Transactions on Signal Processing*, 44(9):2209–2221, 1996.

[13] G. Montseny. *Représentation diffusive.* Hermes Science Publ., 2005.

[14] G. Montseny. Simple approach to approximation and dynamical realization of pseudodifferential time-operators such as fractional ones. *IEEE Trans on Circuits and System II*, 51(11):613–618, Nov. 2004.

[15] Ph. Mouyon and N. Imbert. Identification of a 2D turbulent wind spectrum. In *Aerospace Science and Technology*, volume 6, pages 3599–605, Kansas City (Missouri USA), 2002.

[16] J.G. Németh, I. Kollár, and J. Schoukens. Identification of Volterra kernels using interpolation. *IEEE Transactions on Instrumentation and Measurement*, 51(4):770–775, 2002.

[17] M.J. Reed and M.O.J. Hawksford. Identification of discrete Volterra series using maximum length sequences. *IEE Proc.-Circuitt Devices Syst. ci*, 143(5), 1996.

[18] A. Rumeau, P. Bidan, T. Lebey, L. Marchin, B. Barbier, and S. Guillemet. Behavior modeling of a $CaCu_3Ti_4O_{12}$ ceramic for capacitor applications. In *IEEE Conference on Electrical Insulation and Dielectric Phenomena*, Kansas City (Missouri USA), 15-18 Oct. 2006.

[19] A.N. Tikhonov, V.Y. Arsenin, and F. John. *Solutions of ill-posed problems.* Vh Winston Washington, DC, 1977.

[20] J. Tuch, A. Feuer, and Z.J. Palmor. Time delay estimation in continuous linear time-invariant systems. *IEEE Transactions on Automatic Control*, 39(4):823, 1994.

[21] H. Unbehauen and G.P. Rao. Continuous-time approaches to system identification–A survey. *Automatica*, 26(1):23–35, 1990.