Convergence radius and guaranteed error bound for the Volterra series expansion of finite dimensional quadratic systems

THOMAS HELIE  
CNRS UMR9912, Ircam - Centre G. Pompidou  
Analysis/Synthesis Team  
1, place Igor Stravinsky, 75004 Paris  
FRANCE

BEATRICE LAROCHE  
Univ. Paris-Sud, CNRS, Supelec  
UMR 8506  
3, rue Joliot Curie, 91190 Gif-Sur-Yvette  
FRANCE

Abstract: In this paper, the Volterra series decomposition of a class of quadratic, time invariant single-input finite dimensional systems is considered. These systems are represented using Volterra series. The convergence of the series towards a weak solution is proven. An explicit and computable lower bound of the radius of convergence is obtained. Moreover, guaranteed error bounds in $L^\infty$ are obtained. In section 4, numerical simulations are performed on academic examples. This illustrates how easily the truncated Volterra series can be implemented.

Key–Words: Volterra series, Convergence, Quadratic SIMO-Systems, Simulation.

1 Introduction

Volterra series were introduced by the Italian mathematician Vito Volterra [Vol59]. They can be viewed as the generalization of the transfer function of a linear system. These functional series expansions are convenient tools for on-line simulation or system identification [DPO02], but it is often difficult to obtain convergence results and bounds for the series.

In this paper, such convergence results and bounds are obtained in the case of finite dimensional ODE quadratic systems.

The paper is organized as follows. In section 2, some recalls on Volterra series are given. In section 3, the class of systems under consideration is defined (sec. 3.1) and a standard recursive formula for the associated Volterra kernels is derived (sec. 3.2). Section 4 is devoted to the main point of the paper: first, the convergence of the Volterra series is proven and an explicit and computable lower bound for the radius of convergence is obtained; second, guaranteed error bounds in $L^\infty$ are given for the truncated series. Finally, in section 5, numerical simulations are performed on academic examples. This illustrates how easily the truncated Volterra series can be implemented.

Detailed proofs of the theoretical results presented in section 4 as well as their extensions to MIMO systems can be found in [HL07].

2 Volterra series

2.1 Volterra series of time-variant systems

Following [LL94, p.113], the Volterra series of a time-variant system can be defined as follows.

Definition 1 A causal SISO-system can be described by a Volterra series $\{h_m\}_{m \in \mathbb{N}}$ if there exists functions $h_m : \mathbb{R}^{m+1}_+ \to \mathbb{R}$, for $m \in \mathbb{N}$ which are locally bounded, piecewise continuous and such that, for all $T > 0$, there exists $\epsilon(T) > 0$ such that for all piecewise continuous function $u$ satisfying $|u(t)| \leq \epsilon$, $\forall t \in [0, T]$ the series

$$y(t) = h_0(t) + \sum_{m \in \mathbb{N}} \int_{[0,t]^m} h_m(t, \tau_{1,m}) \prod_{j=1}^{m} u(\tau_j) d\tau_{1,m}$$

is normally convergent, using the concise notations, for $1 \leq p \leq q$,

$$\tau_{p,q} := (\tau_p, \tau_{p+1}, \ldots, \tau_q),$$

$$d\tau_{p,q} := \prod_{j=p}^{q} d\tau_j.$$  

(2)

Nevertheless, natural extensions to more general settings can be defined. For example, taking $h_m$ in $L^1_{loc}(\mathbb{R}^{m+1}_+)$ or $L^\infty(\mathbb{R}_+, L^1_{loc}(\mathbb{R}^m_+))$ still yields well-posed definitions. In this paper, more specific spaces will be introduced in section 2.2.2.
2.2 Volterra series of time-invariant systems

We refer to [Rug81, Boy85, Has99] for developments in this section.

2.2.1 Time domain and Laplace domain

For a time-invariant system, the kernels are such that, for \( m \in \mathbb{N}^* \), it exists \( \tilde{h}_m \) such that

\[
h_m(t, \tau_1, \ldots, t - \tau_m) = \tilde{h}_m(t - \tau_1, \ldots, t - \tau_m).
\]

(3)

Moreover, the zero-input response of the system \( h_0 \) can be omitted considering the difference output \( \tilde{y}(t) = y(t) - h_0(t) \). Then, using the change of variables \( t_j = t - \tau_j \), equation (1) reduces to a sum of standard multi-convolutions given by

\[
\tilde{y}(t) = \sum_{m \in \mathbb{N}^*} \int_{[0, \xi]^m} \tilde{h}_m(t_1, \ldots, t_m) \prod_{j=1}^m u(t - t_j) \, dt_{1:m}.
\]

(4)

For sake of legibility, the tilde of \( \tilde{h}_m(t_1, \ldots, t_m) \) will be omitted (without ambiguity since the number of independent variables in \( h_m \) makes the time-variant and time-invariant versions distinguishable with \( m + 1 \) and \( m \) variables, respectively).

The mono-lateral Laplace transform of the time-invariant kernels is denoted with capital letters and defined by \( \forall m \in \mathbb{N}^* \), \( \forall (s_1, m) \in \mathcal{D}_{hm} \subset \mathbb{C}^m \),

\[
H_m(s_1, \ldots, s_m) = \int_{\mathbb{R}_+^m} h_m(t_1, \ldots, t_m) e^{-(s_1 t_1 + \cdots + s_m t_m)} \, dt_{1:m},
\]

(5)

where \( \mathcal{D}_{hm} \) denotes the domain of convergence of the Laplace transform. For stable systems, \( H_m \) is analytic in \( \mathcal{D}_{hm} \subset (\mathbb{C}_0^+)^m \) where \( \mathbb{C}_0^+ = \{ s \in \mathbb{C} \mid \Re(s) > 0 \} \).

2.2.2 Functional spaces, characteristic function and a BIBO-convergence theorem

Definition 2 (Functional spaces) Let \((m, n) \in \mathbb{N}^* \times \mathbb{N}^* \) and \( p \in [1, \infty[ \). The spaces \( \mathcal{V}_p^{m,n} \) and \( \mathcal{B}_p^n \) are defined by

\[
\mathcal{V}_p^{m,n} = L^1(\mathbb{R}_+^m, \mathbb{R}^n_p)
\]

(6)

\[
\mathcal{B}_p^n = L^\infty(\mathbb{R}_+, \mathbb{R}^n_p)
\]

(7)

where \( \mathbb{R}^n_p \) is the euclidean space of dimension \( n \) endowed with the standard \( p \)-norm defined by \( \| x \|_p = (|x_1|^p + \cdots + |x_N|^p)^{1/p} \) for \( p \in [1, \infty[ \) and by \( \| x \|_\infty = \max(|x_1|, \ldots, |x_N|) \) for \( p = \infty \). When \( n = 1 \), all the \( p \)-norms are identical so that \( p \) is omitted in this case.

Definition 3 (Characteristic function) Let \( \{ h_m \}_{m \in \mathbb{N}^*} \) be the Volterra series of a time-invariant SISO-system, such that \( \forall m \in \mathbb{N}^* \), \( \| h_m \|_{\mathcal{V}_p^{m,n}} = \int_{\mathbb{R}_+^m} \| h_m(t_1, \ldots, t_m) \|_1 \, dt_{1:m} \) is bounded. The characteristic function \( \varphi_h \) of \( \{ h_m \}_{m \in \mathbb{N}^*} \) is defined by the power series

\[
\varphi_h(z) = \sum_{m \in \mathbb{N}^*} \| h_m \|_{\mathcal{V}_p^{m,n}} z^m, \quad \forall |z| < \rho,
\]

(8)

where \( \rho \) is the radius of convergence of the power series.

Theorem 4 Let \( \{ h_m \}_{m \in \mathbb{N}^*} \) be the Volterra series of a time-invariant SISO-system such that the characteristic function \( \varphi_h \) has a radius of convergence \( \rho > 0 \). The Volterra series is convergent in \( \mathbb{B}^1 \) for inputs such that \( \| u \|_{\mathbb{B}^1} < \rho \). In this case, the output \( y \) is bounded and satisfies

\[
\| y \|_{\mathbb{B}^1} \leq \varphi_h(\| u \|_{\mathbb{B}^1}).
\]

(9)

This result is quite interesting for system analysis since it is non-local in time. Nevertheless, it requires the determination of the radius of convergence \( \rho \) and bounding \( \| h_m \|_{\mathcal{V}_p^{m,n}} \) is not straightforward. This paper copes with this practical problem and establishes practicable BIBO-results.

2.2.3 Interconnection laws

Let \( \{ f_m \}_{m \in \mathbb{N}^*} \) and \( \{ g_m \}_{m \in \mathbb{N}^*} \) be the Volterra kernels of two systems, associated to the characteristic functions \( \varphi_f \) and \( \varphi_g \) with radius of convergence \( \rho_f \) and \( \rho_g \), respectively.

\[
\begin{align*}
(a) \quad & \begin{array}{c}
\{ f_m \} \\
\{ g_m \}
\end{array} \\
\begin{array}{c}
y(t)
\end{array} & \begin{array}{c}
\oplus
\end{array} & \begin{array}{c}
u(t)
\end{array}
\end{align*}
\]

(10)

\[
\begin{align*}
(b) \quad & \begin{array}{c}
\{ f_m \} \\
\{ g_m \}
\end{array} \\
\begin{array}{c}
y(t)
\end{array} & \begin{array}{c}
\otimes
\end{array} & \begin{array}{c}
u(t)
\end{array}
\end{align*}
\]

(11)

\[
\begin{align*}
(c) \quad & \begin{array}{c}
\{ f_m \}
\end{array} \\
\begin{array}{c}
y_1(t)
\end{array} & \begin{array}{c}
\rightarrow
\end{array} & \begin{array}{c}
y(t)
\end{array}
\end{align*}
\]

(12)

Figure 1: Sum (a), product (b), and cascade (c) of two systems

Connecting these systems through a sum of outputs, a product of outputs or a cascade (Fig. 1a, b, c) still defines a Volterra series [Has99, p. 34, 35] with kernels \( \{ h_m \}_{m \in \mathbb{N}^*} \) such that for \( m \in \mathbb{N}^* \),
Sum (Fig. 1a) : For \((t_1,m) \in (\mathbb{R}_+)^{m}, (s_1,m) \in D_{f_m} \cap D_{g_m}, z \in [0, min(\rho_f, \rho_g)], \)

\[ h_m(t_1,m) = f_m(t_1,m) + g_m(t_1,m), \]

\[ H_m(s_1,m) = F_m(s_1,m) + G_m(s_1,m), \]

\[ \varphi_h(z) \leq \varphi_f(z) + \varphi_g(z). \]

Product (Fig. 1b) : For \((t_1,m) \in (\mathbb{R}_+)^{m}, (s_1,m) \in 1 \leq p \leq m-1 (D_{f_k} \times D_{g_{m-k}}), z \in [0, min(\rho_f, \rho_g)], \)

\[ h_m(t_1,m) = \sum_{k=1}^{m-1} f_k(t_1,k) g_{m-k}(t_{k+1,m}), \]

\[ H_m(s_1,m) = \sum_{k=1}^{m-1} F_k(s_1,k) G_{m-k}(s_{k+1,m}), \]

\[ \varphi_h(z) \leq \varphi_f(z) \varphi_g(z). \]

Cascade with a linear system (Fig. 1c) : For \((t_1,m) \in (\mathbb{R}_+)^{m}, (s_1,m) \in \{(s_1,m) \in D_{f_m} | \bar{s}_1,m \in D_{g_1}\}, z \in [0, \rho_f], \)

\[ h_m(t_1,m) = \int_{[0, \min(t_1,m)]} g_1(\theta_1) f_m(t_1,m- \theta_1)d\theta_1, \]

\[ H_m(s_1,m) = G_1(\bar{s}_1,m) F_m(s_1,m), \]

\[ \varphi_h(z) \leq \|g_1\|_{\mathcal{Y}_1} \|\varphi_f(z)\|. \]

where \(\bar{s}_1,m\) denotes the sum of the Laplace variables

\[ \bar{s}_1,m = s_1 + \cdots + s_m. \]

3 Quadratic SIMO systems

3.1 System under consideration

Let the quadratic ODE system be defined by

\[
\dot{x} = Ax + \begin{bmatrix} x^T E_1 x \\ \vdots \\ x^T E_N x \end{bmatrix} + Bu, \tag{20}
\]

\[ y = Cx, \tag{21}\]

for \(t \in \mathbb{R}^+\) with \(x(0) = 0\), where \(u(t) \in \mathbb{R}, x(t) \in \mathbb{R}^N\) and \(y(t) \in \mathbb{R}^Q\) are the input, state and output of the system, respectively. All matrices are real and \(A\) is \(N \times N\), \(B\) is \(N \times 1\), \(C\) is \(\times N\), and \(E_n\) \((n = 1, \ldots, N)\) are \(N \times N\). This system can be viewed as a second order approximation of a nonlinear system of the form \(\dot{x} = f(x) + Bu, y = Cx\) around the initial state \(x(0) = 0\).

Definition 5 (Strong and weak solutions) Let \(C^1(\mathbb{R}_+; \mathbb{R}^N)\) denote the set of all \(C^1, \mathbb{R}^N\)-valued functions with compact support in \(\mathbb{R}_+\). \((x,y)\) is said to be a weak solution of (20-21) in \(\mathcal{B}_p^N \times \mathcal{B}_q^Q\) with \(p \in [1, \infty]\) iff \(\forall w \in C^1(\mathbb{R}_+, \mathbb{R}^N), \)

\[
\int_{\mathbb{R}_+} w^T \dot{x} dt + \int_{\mathbb{R}_+} w^T Ax dt \\
+ \int_{\mathbb{R}_+} w^T \begin{bmatrix} x^T E_1 x \\ \vdots \\ x^T E_N x \end{bmatrix} dt + \int_{\mathbb{R}_+} w^T Bu dt = 0, \tag{22}\]

and \(y\) satisfies (21). Moreover, \((x,y)\) is said to be a strong solution, if it is a weak solution and \(x\) is \(C^1(\mathbb{R}_+, \mathbb{R}^N)\).

3.2 Derivation of the Volterra kernels

The formal computation of Volterra kernels is well-known and standard (see [Flü81, LL94, Ls95]). Here, a method relying on a cancelling system is used (see e.g. [Has99, HH04]), which also proves to be practicable for PDE’s systems.

3.2.1 Output and state kernels

As the nonlinearity of the system is embedded in the state equation (20), it is quite convenient to consider the Volterra series which maps the input \(u\) to each state coordinate \(x_n\). Thus, let \(\{g_m\}_{m \in \mathbb{N}^*}\) denote the Volterra series of the SIMO-system \(S_{u \rightarrow y}\) with input \(u\) and output \(y\). Let \(\{h_m^n\}_{m \in \mathbb{N}^*}\) for \(1 \leq n \leq N\) denote the Volterra series of the SISO-system \(S_{u \rightarrow x_n}\) with input \(u\) and output \(x_n\). Let \(\{h_m\}_{m \in \mathbb{N}^*}\) denote the Volterra series of the SIMO-system \(S_{u \rightarrow x}\) with input \(u\) and output \(x\) so that \(h_m = [h_m^1, \ldots, h_m^N]^T\).

From (10-11) and (21), \(\{g_m\}_{m \in \mathbb{N}^*}\) are related to \(\{h_m\}_{m \in \mathbb{N}^*}\) through the equations, for \(m \in \mathbb{N}^*\),

\[ g_m(t_1,m) = C h_m(t_1,m) \]

\[ G_m(s_1,m) = C H_m(s_1,m) \]

in the time and the Laplace domains, respectively.

3.2.2 Cancelling system

A convenient way to derive a set of equations satisfied by the kernels is to build the system described in Fig. 2. From (20), it is the null-system so that all its Volterra kernels are zero. Writing these zero-kernels from the interconnection laws (10-17) yields the equations satisfied by kernels \(\{h_m\}\) which are linear differential in the time domain and algebraic in the Laplace domain.
3.2.3 Kernels in the Laplace domain

In the upper branch of the Laplace domain (see Fig. 2), the Laplace transform of the linear operator is

\[ s \mapsto I_N s - A. \]

The kernels of the system \( S_{u - \cdot - Ax} \) associated to this upper branch can be derived, for each coordinate \( \hat{x} - Ax \), from (11) and (17). They are given by \( [s I_N - A]H_m(s_{1,m}) \), for \( m \in \mathbb{N}^* \). The middle branch corresponds to the quadratic nonlinearities \( Q_{E_n} \). For each sub-system \( S_{u - \cdot - x}E_n \), the kernels are deduced from (14). They are given by

\[ m \in \mathbb{N}^*, \text{ which is zero for } m = 1. \]

The kernels for the bottom branch are zero kernels associated to the constant kernel \( s_1 - B \) for \( m = 1 \) and to zero kernels \( s_{1,m} - 0 \) for \( m \geq 2 \). Now, writing from (11) that the sum of these kernels is zero yields the recursive algebraic equations, for all \( m \in \mathbb{N}^* \),

\[
H_m(s_{1,m}) = [s_{1,m}I_N - A]^{-1}F_m(s_{1,m}), \quad (25)
\]

\[
F_1(s_1) = B \quad (26)
\]

\[
F_m(s_{1,m}) = \sum_{k=1}^{m-1} \left[ (H_k(s_{1,k}))^T E_n H_{m-k}(s_{k+1,m}) \right]
\]

if \( m \geq 2. \) \( (27) \)

3.2.4 Kernels in the time domain

Using the notation of the time-variant systems \( h_m(t, \tau_1, \tau_2) \) rather than the time-invariant version \( h_m(t_{1,m}) \) with \( t_i = t - \tau_i \), the time domain versions of (25-27) are, for all \( m \in \mathbb{N}^* \),

\[
[I_N \partial_t - A]h_m(t, \tau_1, \tau_2) = f_m(t, \tau_1, \tau_2), \quad (28)
\]

\[
f_1(t, \tau_1) = B\delta(t - \tau_1), \quad (29)
\]

\[
f_m(t, \tau_1, \tau_2) = \sum_{k=1}^{m-1} \left[ (h_k(t, \tau_1, \tau_2))^T E_n h_{m-k}(t, \tau_k, \tau_{k+1}) \right]
\]

if \( m \geq 2. \) \( (30) \)

The solution is

\[
h_1(t, \tau_1) = e^{A(t - \tau_1)}B 1_{\mathbb{R}^+}(t - \tau_1), \quad (31)
\]

\[
h_m(t, \tau_1, \tau_2) = \int_{\tau_2}^{t} e^{A(t - \theta)}f_m(\theta, \tau_1, \tau_2) \, d\theta \cdot 1_{\mathbb{R}^+}(t - \max(\tau_1, \tau_2)), \text{ if } m \geq 2, \quad (32)
\]

where \( 1_{\mathbb{R}^+} \) denotes the Heaviside function.

4 Convergence and guaranteed error bounds

In this section, standard \( p \)-norms of vectors \( x \), matrices \( M \) and bilinear forms \( E \) are considered for a fixed \( p \in [1, \infty] \) and given by

\[
\|x\|_p = (|x_1|^p + \ldots + |x_N|^p)^{\frac{1}{p}}, \quad \text{if } p \in [1, \infty[ \], \quad (33)
\]

\[
\|x\|_p = \max(|x_1, \ldots, |x_N|), \quad \text{if } p = \infty, \quad (34)
\]

\[
\|M\|_p = \sup_{\|x\|_p = 1} \|Mx\|_p, \quad (35)
\]

\[
\|E\|_{\mathcal{Q}_p} = \sup_{\|x\|_p = 1, \|y\|_p = 1} |y^T Ex|. \quad (36)
\]

Theorem 6 Consider system (20) with \( \max(\Re(\text{Spec}(A))) < 0 \). Let \( \{h_m \}_{m \in \mathbb{N}^*} \) be the Volterra kernels defined by (31-32). Then, for all \( p \in [1, +\infty) \),

\[
\|h_m\|_{\mathcal{V}_p, N} \leq \Phi_m(\epsilon_p \alpha_p)^{m-1} (\|h_1\|_{\mathcal{V}_p, N})^m, \quad (37)
\]

with \( \epsilon_p = \left\| \left[ \|E_1\|_{\mathcal{Q}_p} \ldots \|E_N\|_{\mathcal{Q}_p} \right]^T \right\|_p, \quad (38) \]

\[
\alpha_p = \int_{\mathbb{R}^+} \|e^{A\xi}\|_p d\xi < \infty, \quad (39)
\]

\[
\Phi_m = C_{m-1} = \frac{(2(m - 1))!}{m!(m - 1)!}. \quad (40)
\]

Note that \( \|h_1\|_{\mathcal{V}_p, N} \leq \alpha_p \|B\|_p < \infty, \quad (41) \)

and that \( \Phi_m = \Phi_{m+1} \) for \( m \in \mathbb{N} \) are the Catalan numbers (see e.g. [FS07]).

Proof: The main steps of the proof (detailed in [HL07]) are sketched below.

Equation (41) is straightforward and, (37) is satisfied for \( m = 1 \) with equality by defining \( \Phi_1 = 1. \)

Then, (37) is proven with \( \Phi_1 = \sum_{k=1}^{m-1} \Phi_k \Phi_{m-k} \), by induction, considering that (37) is satisfied for any \( m' \) with \( 1 \leq m' \leq m - 1 \) and making use of, for \( m \geq 2, \)

\[
\|h_m\|_{\mathcal{V}_p, N} \leq \epsilon_p \alpha_p \sum_{k=1}^{m-1} \|h_k\|_{\mathcal{V}_p, N} \|h_{m-k}\|_{\mathcal{V}_p, N} \|h_{m-k}\|_{\mathcal{V}_p, N}, \quad (42)
\]
with the notations defined in the theorem. This recursive relationship on \( \Phi_m \) defines the Catalan numbers \( C_{m-1} \) recalled in (40).

For \( m \leq 2 \), the inequality (42) can be derived by proving the successive following inequalities. From (3) and (32) choosing \( t = 0 \),

\[
||h_m||_{\mathbb{V}_p} \leq \int_{[\xi, +\infty]^m} ||e^{A_1}d\|_{\mathbb{P}} \left( \int ||f_m(\xi, -t_1, 1)||_{\mathbb{P}} d\xi \right) d\xi.
\]

From (30),

\[
||f_m(t, \tau_1, \tau_1, m)||_{\mathbb{P}} \leq \epsilon_p \sum_{k=1}^{m-1} \||h_k(t, \tau_1, m)||_{\mathbb{P}} \||h_m-k(t, \tau_1, m)||_{\mathbb{P}}.
\]

Then,

\[
\int_{[\xi, +\infty]^m} ||f_m(\xi, -t_1, 1)||_{\mathbb{P}} d\xi \leq \epsilon_p \sum_{k=1}^{m-1} \||h_k||_{\mathbb{V}_p} ||h_m-k||_{\mathbb{V}_p}.
\]

so that (42) is satisfied.

**Theorem 7** Let \( p \in [1, \infty] \). Let the system (20-21) be such that \( \max \left( \text{Re(Spec (A))o} \right) < 0 \) so that \( ||h_1||_{\mathbb{V}_p} < \infty \). Then, the Volterra series expansions of the state and output of the system (20-21) converge in \( \mathbb{B}_p^\infty \) and \( \mathbb{B}_1 \), respectively, for all input \( u \in \mathbb{B}_1 \) such that

\[
Z_p(u) < 1/4,
\]

where \( Z_p : \mathbb{B}_1 \to \mathbb{R}_+ \) is defined by

\[
Z_p(u) = \epsilon_p \alpha_p ||h_1||_{\mathbb{V}_p} ||u||_{\mathbb{B}_1},
\]

so that the radius of convergence \( \rho_h \) satisfies

\[
\rho_h \geq \rho_h^* = \frac{4\epsilon_p \alpha_p ||h_1||_{\mathbb{V}_p}^{-1}}{1}.
\]

In this case, the following results hold:

(i) The sum of the series is a weak solution of the system. If \( u \) is in \( C^0(\mathbb{R}_+, \mathbb{R}) \), this solution is also a strong one.

(ii) The output \( y \) and the state \( x \) are bounded as:

\[
||y||_{\mathbb{B}_p^\infty} \leq ||C||_p ||x||_{\mathbb{B}_p^\infty},
\]

\[
||x||_{\mathbb{B}_p^\infty} \leq \varphi_{h, p}(||u||_{\mathbb{B}_1}) \leq \frac{\Phi(Z_p(u))}{\epsilon_p \alpha_p}.
\]

(iii) Errors due to the truncation of order \( M \in \mathbb{N}^* \) have guaranteed bounds:

\[
||x - V_M x||_{\mathbb{B}_p^\infty} \leq \frac{F_M(Z_p(u))}{\epsilon_p \alpha_p},
\]

\[
||y - V_M y||_{\mathbb{B}_p^\infty} \leq ||C||_p \frac{F_M(Z_p(u))}{\epsilon_p \alpha_p},
\]

where

\[
|F_M(z)| \leq \frac{\Phi_{M+1} z^{M+1}}{1 - 4z}
\]

\[
\leq \frac{1}{2 \pi (M+1)(2M+1)} \frac{(4z)^{M+1}}{1 - 4z}.
\]

These results involve the following definitions:

\[
\varphi_{h, p}(z) = \sum_{m=1}^{\infty} ||h_m||_{\mathbb{V}_p^m, N} z^m,
\]

\[
\Phi(z) = (1 - \sqrt{1 - 4z})/2,
\]

\[
V_M x(t) = \sum_{m=1}^{M} \int_{t}^{m+1} h_m(\tau_1, m) u(t - \tau_1) d\tau_1,
\]

\[
V_M y(t) = C V_M x(t),
\]

\[
F_M(z) = \sum_{m=M+1}^{\infty} \Phi_m z^m,
\]

and \( \epsilon_p, \alpha_p, \Phi_m \) are given in theorem 6.

Note that \( \varphi_{h, p} \) generalizes definition 3 to SIMO-systems.

**Proof:** The sketch of proof of (43) and (ii,iii) is as follows (for a detailed version of (43), (ii,iii), see [HL07]).

From (21), \( ||y||_{\mathbb{B}_p^\infty} \leq ||C||_p ||x||_{\mathbb{B}_p^\infty} \). Moreover, from theorem 6, \( ||x||_{\mathbb{B}_p^\infty} \leq \varphi_{h, p}(||u||_{\mathbb{B}_1}) \leq \Phi(Z_p(u))/\epsilon_p \alpha_p \) where \( \Phi(z) = \sum_{m \in \mathbb{N}} \Phi_m z^m = \sum_{m \in \mathbb{N}} \Phi_m z^m = \frac{1 - \sqrt{1 - 4z}}{2} \) is absolutely convergent for \( z < 1/4 \). Similarly, for the remainder \( R_M x = x - V_M x \), we prove that

\[
||R_M x||_{\mathbb{B}_p^\infty} \leq \sum_{m=M+1}^{\infty} ||h_m||_{\mathbb{V}_p^m, N} \left( ||u||_{\mathbb{B}_1} \right)^m
\]

\[
\leq \frac{1}{\epsilon_p \alpha_p} \sum_{m=M+1}^{\infty} \Phi_m(Z_p(u))^m,
\]

which converges for \( Z_p(u) \leq 1/4 \). Now, rewriting the Catalan numbers as \( \Phi_m = \frac{4m-1}{m} \left( \frac{m+1/2}{\Gamma(m+1)} \right)^m \) and using Wallis’ formula \( \frac{1}{\sqrt{m}} (1 - \frac{1}{8m}) < \frac{1}{\Gamma(m+1/2)} < \frac{1}{\sqrt{m}} \), yield

\[
1 - \frac{1}{8m} \xi_m < \Phi_m < \xi_m = \frac{4m-1}{\sqrt{\pi} \sqrt{m} (m-1)},
\]

from which (48-50) are deduced using the superior bound \( \xi_m \).

**5 Simulation**

To illustrate the previous theoretical results, computation and simulation are proposed on two academic examples. The optimality of the bounds computed is investigated. In each case, guaranteed convergence radii
and guaranteed error bounds are explicitly computed for three \( p \)-norms \((p = 1, 2, \infty)\). A low-cost numerical simulation of the truncated series is proposed.

5.1 First example

Consider the system (20-21) with \( N = Q = 2 \) and

\[
A = -\mu I_2 \tag{56}
\]

\[
B = \begin{pmatrix} 1 \\ \gamma \end{pmatrix} \tag{57}
\]

\[
C = I_2, \tag{58}
\]

\[
E_n = \frac{\beta_n}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ for } n \in \{1, 2\}, \tag{59}
\]

where \( \mu, \gamma, \beta_1 \) and \( \beta_2 \) are strictly positive parameters. Such a model describes phenomena involving two entities with the same decay rate \( \mu \), a growth speed \( \beta_{1,2} x_1 x_2 \) and an external input term proportional to \( u \). The quadratic factor \( x_1 x_2 \) of the growth speed models the probability for \( x_1 \) to meet \( x_2 \) and, for \( E_n \), to produce more \( x_n \) through the coefficient \( \beta_n \). For instance, these equations could model some catalytic processes, or auto-activating gene network, or the death/birth process in an animal population structured according to gender, etc. Note that this system is positive: \( x_{1,2}(0) \geq 0 \) and \( u \geq 0 \) implies \( x_{1,2} \geq 0 \) on \( \mathbb{R}^+ \) (\( x_{1,2} \) represent the (positive) quantities of the entities “1” and “2”).

For numerical applications, parameters are chosen as follows: \( \mu = 0.3, \gamma = 0.1, \beta_1 = 0.04 \), and \( \beta_2 = 0.02 \).

5.1.1 Volterra series and guaranteed results

We start by investigating the existence and local stability of a positive equilibrium state for a constant positive input \( u = a \). An easy computation shows that the system has two positive equilibrium states if and only if \( 0 \leq a < a^* = \frac{135 - 45 \sqrt{5}}{16} \approx 2.15 \), one of them being locally stable and the other one unstable. Numerical simulation confirms that if \( a < a^* \) the systems state is bounded, and on the contrary, if \( a > a^* \), then the state is unbounded.

As a consequence, the convergence radius for the Volterra series is expected to be less than \( a^* \).

For \( p = 1, p = \infty \) and \( p = 2 \), calculations lead to

\[
\alpha_1 = \alpha_2 = \alpha_\infty = \frac{1}{\mu} \approx 3.33,
\]

\[
\epsilon_\infty = \beta_1 = 0.04, \epsilon_1 = \frac{\beta_1 + \beta_2}{2} = 0.03
\]

\[
\epsilon_2 = \frac{\beta_1^2 + \beta_2^2}{2} \approx 0.022,
\]

\[
\|h_1\|_{V_{1,2}} = (1 + \gamma)/\mu \approx 3.667,
\]

\[
\|h_1\|_{V_{2,1,2}} = \sqrt{1 + \gamma^2}/\mu \approx 3.35,
\]

\[
\|h_1\|_{V_{1,1,10}} = 1/\mu \approx 3.33.
\]

From theorem 7, the corresponding guaranteed convergence radii are \( \rho_p^* = [4 \epsilon_p \alpha_p \|h_1\|_{V_{p,2}}]^{-1}, \) namely,

\[
\rho_1^* = \mu^2/(4 \epsilon_1 (1 + \gamma)) \approx 0.68,
\]

\[
\rho_2^* = \mu^2/(4 \epsilon_2 \sqrt{1 + \gamma^2}) \approx 1.008,
\]

\[
\rho_\infty^* = \mu^2/(4 \epsilon_\infty) = 0.56.
\]

As all \( p \)-norms are equivalent in finite dimensional spaces, the series is convergent for any \( p \)-norm although the criterion is not necessarily met. Here, the best convergence radius among \( \rho_1^*, \rho_2^* \) and \( \rho_\infty^* \) is then \( \rho^* = \rho_2^* \) which satisfies as expected: \( \rho^*/\alpha^* \approx 0.5 < 1 \).

Hence, for \( U = \|u\|_{B^1} < \rho^* \), the output is guaranteed to be bounded and the truncated series (at order \( M \)) yields an error less than:

\[
\mathcal{E} = \frac{\Phi_{M+1}(\epsilon_2 \alpha_2 \|h_1\|_{V_{2,1,2}} U)^{M+1}}{\epsilon_2 \alpha_2 \|h_1\|_{V_{2,1,2}} U} \approx \frac{\Phi_{M+1}(0.248 U)^{M+1}}{0.074 - 0.992 U}.
\]

Criterion (43) in theorem 7 therefore provides a lower bound for the convergence radius of the Volterra series that might seem conservative when restricting the system to the positive input/positive state situation (or equivalently here to the negative input/negative state situation because of symmetries). This problem is related to the use of matrix norms that cannot take into account the signature of the quadratic forms involved in the quadratic part of the system, as we shall see on the second example.

5.1.2 Numerical simulation

The simulation of this system is performed below for \( M = 3 \) (so that \( \Phi_{M+1} = 5 \)). Inputs \( u \) such that
\(\|u\|_{B^1} \leq U^* = 0.6 < \rho^*\) are considered. The error is guaranteed to be less than 
\[ \mathcal{E} \approx 0.08.\]

Real time implementation of Volterra decomposition is done as follows: the first order contribution denoted as \(u_1\) is given by the linear part of the system. In the Laplace domain this contribution is
\[ W_1(s_1) = (s_1 I - A)^{-1} B U(s_1) = H_1(s_1) U(s_1). \]

Then \(m^{th}\) order contribution \(w_m\) is computed as the sum of \(m - 1\) terms \(w_{mk}\), with \(1 \leq k \leq m - 1\), as shown in figure 3. In the time domain, a realization for

\[ \text{Figure 3: Implementation of the Volterra series decomposition at order 3} \]

each \(w_{mk}\) is obtained by computing the \(N\) dot products \((w_k(t))^T E_i w_{m-k}(t)\), for \(1 \leq i \leq N\), and filtering the resulting \(N\) dimensional vector with the filter with \(N\) inputs, \(N\) outputs whose impulse response is \(e^{A_1 t} 1_{\mathbb{R}^+}(t)\), as shown on figure 4. This realization shall be denoted as \(T\). On figure 6, truncated Volterra

\[ \text{Figure 4: Realization of } w_{mk} \]

series of order 3 for a constant input \(u(t) = 0.6\) is plotted with the associated guaranteed error bounds. The truncated series matches almost exactly the true solution of the system, so that, as pointed out before, the convergence bound as well as the truncation error bound are conservative in this situation.

\[ \text{Figure 5: Simulation of the Volterra approximation at order 3 (--), guaranteed intervals (---) using error bounds, and exact solution (-).} \]

5.2 Second example

In order to investigate the influence of the signature of the quadratic forms involved in system (20-21) we consider the same system as above, the only change being the expression of matrices \(E_n\):
\[ E_n = \frac{\beta_n}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{for } n \in \{1, 2\}, \]

(60)

Parameters \(\mu, \gamma, \beta_1,\) and \(\beta_2\) have the same values as before, and this system is still positive.

5.2.1 Volterra series and guaranteed results

As before, we first investigate the existence and local stability of a positive equilibrium state for a constant positive input \(u = a\). We easily find that the system has two positive equilibrium states if and only if \(0 < a < \frac{945 + 45 \sqrt{605}}{64} \approx 1.04\), one of them being locally stable and the other one unstable.

The parameters involved in the computation of the convergence radius and the guaranteed error bound of the Volterra series of the system are the same as in the first example: the only change is in the quadratic forms matrices \(E_1\) and \(E_2\) but the resulting values for parameters \(e_1, e_2\) and \(e_\infty\) do not change. This time we find that the best convergence radius is \(\rho^* = \rho_2^*\) which satisfies: \(\rho^*/a^* \approx 1\), so that our bound is very good for this example.

5.2.2 Numerical simulation

On figure 6, truncated Volterra series of order 3 for a constant input \(u(t) = 0.6\) is plotted with the associated guaranteed error bounds. We see that the computed error bound gives a sensible value in this case.
Figure 6: Simulation of the Volterra approximation at order 3 (- - -), guaranteed intervals (⋯⋯) using error bounds, and exact solution (−).

6 Conclusion and perspectives

An algorithm to build the kernels for the Volterra series decomposition for a stable system with quadratic state nonlinearity in $L^\infty(\mathbb{R}_+, \mathbb{R}^N)$, as well as a bound on the input and on the truncation error have been obtained. The resulting truncated system is easy to implement and simulate.

Further works will now consist in improving the quality of the guaranteed radius of convergence and error bounds (1), and extending these results to nonlinear systems more general than quadratic ordinary differential ones (2):

- More precisely, the first point (1) will be dedicated to refine the results of theorems 6 and 7 in order to have best guarantees on particular situations, e.g. to take into account the signature of the quadratic forms involved and the positivity of the system (if any).
- The second point (2) will consist in establishing conditions under which the extension of these results to a $n^{th}$ order polynomial state nonlinearity is possible.

As a next step, this analysis can also be extended (with greater technical difficulties to be overcome) to some families of infinite dimensional systems such as nonlinear propagation (see e.g. [HH04, Hélo6]) or diffusion equations with polynomial in the state diffusion coefficients.

References:


