

System identification of biophysical neuronal models*

Thiago B. Burghi¹, Maarten Schoukens² and Rodolphe Sepulchre¹

Abstract—After sixty years of quantitative biophysical modeling of neurons, the identification of neuronal dynamics from input-output data remains a challenging problem, primarily due to the inherently nonlinear nature of excitable behaviors. By reformulating the problem in terms of the identification of an operator with fading memory, we explore a simple approach based on a parametrization given by a series interconnection of Generalized Orthonormal Basis Functions (GOBFs) and static Artificial Neural Networks. We show that GOBFs are particularly well-suited to tackle the identification problem, and provide a heuristic for selecting GOBF poles which addresses the ultra-sensitivity of neuronal behaviors. The method is illustrated on the identification of a bursting model from the crab stomatogastric ganglion.

I. INTRODUCTION

This paper explores the potential of a simple parametrization for the identification of a nonlinear system that can be represented as the feedback interconnection between an integrator and an operator with fading memory [3], as shown in Figure 1. Our interest in this particular structure is that it encompasses most biophysical models of neuronal circuits [4], [5]. In such models, the integrator represents the neuronal membrane model, whereas the operator with fading memory represents the input-output mean-field relationship between voltage and the internal currents arising from the opening and closing of myriad ion channels. Feedback between these two components destroys the fading memory property, and indeed this property is ruled out by observed neuronal behaviors such as excitability, autonomous oscillations, and chaos.

The motivation to preserve the biophysical decomposition of Figure 1 in a system identification setting is obvious: the identification of nonlinear systems with fading memory is a mature topic [16], [26], [25], whereas the identification of feedback systems lacking this property is challenging. In the latter case, it is difficult to obtain any guarantees on the behavior of the identified system. While the availability of noiseless state measurements allows some results to be established [19], internal neuronal states cannot be measured. This limitation poses further issues concerning model identifiability [28] and optimization tractability [1]. Furthermore, the lack of fading memory prohibits the use of contraction constraints recently advocated to improve the trainability of

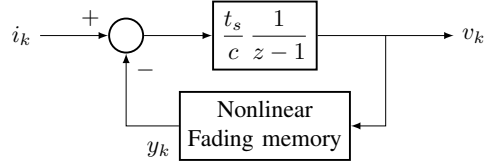


Fig. 1. The general model structure of a neuronal system.

recurrent models [23]. To avoid some of these difficulties, we explore directly identifying the nonlinear fading memory component using a feedforward parametrization with universal approximation properties: the series interconnection of Generalized Orthonormal Basis Functions (GOBFs) with a static Artificial Neural Network (ANN) nonlinearity.

The contributions of the present paper are the following: we show how the problem of neuronal system identification can be cast in terms of a problem of identifying a fading memory operator; we present a proof that the series interconnection of GOBFs with ANNs implements a universal approximator; and we provide a heuristic for choosing GOBF poles to identify neuronal systems, based on the input-output property of ultrasensitivity [6], [9].

The paper is structured in the following way: In Section II, we formulate the identification problem and argue that the model class of Figure 1 encompasses most neuronal systems. In Section III we prove that the proposed model structure is a universal approximator of fading memory operators and discuss how to choose the poles of the GOBFs for neuronal systems. In Section IV, we illustrate how the model structure can be used to identify a bursting neuronal model.

II. PROBLEM STATEMENT

In this section, we introduce the general problem of neuronal system identification. We restrict the treatment in this paper to single-input single-output systems. We write $\mathbb{Z}_+ = \{0, 1, \dots\}$ and $\mathbb{N} = \{1, 2, \dots\}$. The set $\mathbb{R}(\mathbb{Z}_+)$ is the set of real-valued sequences defined on \mathbb{Z}_+ , and $\ell_\infty(\mathbb{Z}_+)$ is the set of all real-valued bounded sequences defined on \mathbb{Z}_+ .

A. Neuronal systems

A neuronal system dictates the evolution of the cellular membrane potential of a neuron, denoted by $v \in \mathbb{R}$. The voltage v changes in time due to the flow of ionic currents through the membrane, which can be reasonably represented by a capacitor. Thus any discrete-time single-compartment biophysical neuronal system has the form

$$c \frac{v_{k+1} - v_k}{t_s} = -y_k + i_k \quad (1)$$

*The research leading to these results has received funding from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Brasil (Finance Code 001) and the European Research Council (under the Advanced ERC Grant Agreement Switchlet n.670645).

¹Thiago B. Burghi (tbb29@cam.ac.uk) and Rodolphe Sepulchre (r.sepulchre@eng.cam.ac.uk) are with the Department of Engineering, Control Group, University of Cambridge, CB2 1PZ, UK.

²Maarten Schoukens (m.schoukens@tue.nl) is with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands.

where $c > 0$ is the membrane capacitance, $t_s > 0$ is a sampling period, $i_k \in \mathbb{R}$ is an external applied current, and

$$y_k = (F_{\text{int}}v)_k \quad (2)$$

represents the aggregate internal ionic currents traversing the neuronal membrane (we assume forward-Euler discretization of the physical system). The time-invariant operator $F_{\text{int}} : \ell_\infty(\mathbb{Z}_+) \rightarrow \mathbb{R}(\mathbb{Z}_+)$ represents the dynamics of ion channels embedded in the membrane, and it is the accurate modeling of this operator which allows neuronal behavior to be successfully reproduced in neurocomputational studies.

B. The internal current operator has fading memory

The operator F_{int} is a mean-field model of complex molecular events dictating the opening and closing of ion channels. Ever since the pioneering work of Hodgkin and Huxley [11], ion channel kinetics have been primarily modeled using step response experiments implemented by means of the so-called *voltage-clamp* technique. It is the design of such experiments that led to the foundation of biophysical modeling in neurophysiology. In an experimental setting, the voltage output v_k is measured, and the experimenter has full control over the current input i_k . The voltage-clamp experiment implements a high-gain output feedback law

$$i_k = \gamma(v_k - r_k) \quad (3)$$

where r_k is a reference signal chosen by the experimenter, and $\gamma > 0$ is the feedback gain. When the voltage-clamp experiment is carried out in a biological neuron, it is observed that for any sufficiently high gain $\gamma > 0$, the measured v_k closely tracks the reference r_k ; furthermore, the tracking error decreases as γ is increased. This is in stark contrast with the open-loop behavior of the neuron, as illustrated in Figure 2. The behavior observed experimentally confirms the basic capacitive modeling assumption behind (1): it is the behavior of a system with relative degree one, whose inverse (the system with input v_k and output i_k) can be obtained at the limit of high gain $\gamma \rightarrow \infty$.

Furthermore, the voltage-clamp experiment indicates that the system inverse, and thus also F_{int} , have fading memory, in the following sense:

Definition 1 ([22], [3]). For arbitrary $\beta > 0$, consider the class $\mathcal{U}_\beta \subset \ell_\infty(\mathbb{Z}_+)$ of sequences u such that $\sup_{k \in \mathbb{Z}_+} |u_k| < \beta$. The system operator $F : \ell_\infty(\mathbb{Z}_+) \rightarrow \mathbb{R}(\mathbb{Z}_+)$ is said to have fading memory on \mathcal{U}_β if there exists a decreasing sequence $w : \mathbb{Z}_+ \rightarrow (0, 1]$ with $\lim_{k \rightarrow \infty} w_k = 0$ such that given an $\epsilon > 0$, there is a $\delta > 0$ such that $\max_{m \in \{0, 1, \dots, k\}} \|u_m - \tilde{u}_m\|_\infty w_{k-m} < \delta$ implies $|(Fu)_k - (F\tilde{u})_k| < \epsilon$ for every $u, \tilde{u} \in \mathcal{U}_\beta$ and every $k \in \mathbb{Z}_+$.

In words, a system operator has fading memory if “two inputs which are close in the recent past (but not necessarily close in the remote past) yield outputs which are close in the present” [3]. In particular, fading memory operators share many properties with stable linear systems. An example of this behavior can be seen in Figure 2 (left column): the response of F_{int} to a step input is reminiscent of that of

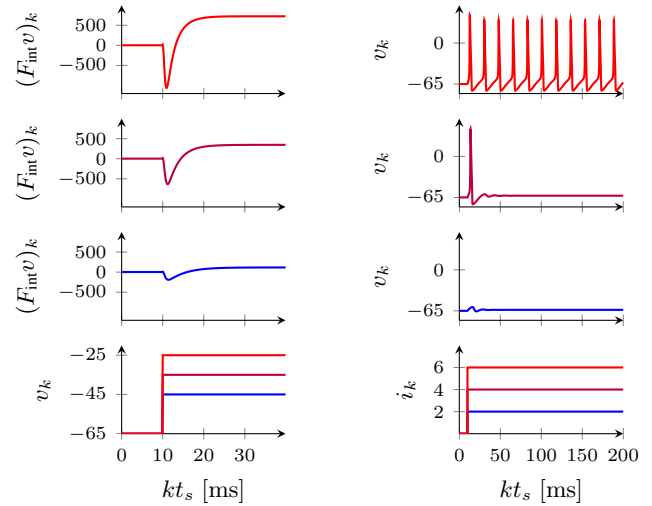


Fig. 2. Left column: Typical aggregate ionic current behavior in a biological neuron under voltage-clamp. Three different experiments are shown in which the reference r_k is stepped to -45 , -35 and -25 mV at 10 ms. Right column: Typical behavior of a neuron (without voltage-clamp) subject to steps of different amplitudes at the input i_k . Voltages are in [mV], and currents are in [$\mu\text{A}/\text{cm}^2$].

a linear system, and its output tends to a unique steady-state value whenever its input tends to a constant as $k \rightarrow \infty$.

While the fading memory of F_{int} can be assumed from empirical evidence, it can also be derived from continuous-time biophysical conductance-based neuronal models. In [5], it is shown that all such models have an exponentially contracting internal dynamics (as defined by [17]), and that choosing a sufficiently small t_s preserves this property in the discretized system. Ultimately, this implies that in conductance-based models, the operator F_{int} has fading memory with an exponentially fast fading rate.

C. Identification problem

We pose the following problem statement: *Identify the input-output behavior of the system (1)-(2), where the operator F_{int} has fading memory on \mathcal{U}_β .* In other words, we search for a model that accurately reproduces the set of admissible input-output trajectories $(i, v) \in \mathbb{R}(\mathbb{Z}_+)^2$ of the system being identified.

In this paper, we focus on the question of model structure selection, and ignore the issue of noise. Noise plays an important role in neuronal dynamics, and input current (process) noise is particularly relevant [10]. However, as long as measurement noise is kept low, the methods discussed in the next section should not be significantly affected (see [5] for related results where input noise is taken into account).

III. IDENTIFICATION METHOD

System operators that have fading memory can be uniformly approximated by simple classes of nonlinear systems, called universal approximators. The parametrization of fading memory operators has a long history going back to Volterra series with notable contributions by Boyd and Chua in [3] and a specific link to neuroscience (though in a

different context from the present paper) by Maas and Sontag [18]. To tackle the problem stated in Section II-C, we propose to use the “direct approach” of closed-loop identification [7] and directly identify F_{int} with a model structure based on the series interconnection of Generalized Orthonormal Basis Functions (GOBFs) [21] and static Artificial Neural Networks (ANNs). This structure was first proposed by [26]. In this section, we first define this structure and show that it has universal approximation power; then, we discuss how it can be tailored to identify neuronal models.

A. A model structure for identifying F_{int}

We define an ANN with logistic nonlinearities as follows:

Definition 2. Let $L \in \mathbb{N}$ be the number of layers in the network. For $\ell = 1, \dots, L$, let $M^{(\ell)} \in \mathbb{N}$ be the number of activation functions in the ℓ^{th} hidden layer, and $M^{(0)}$ and $M^{(L+1)}$ the number of inputs and outputs of the network, respectively. Given the weight matrices $W^{(\ell)} \in \mathbb{R}^{M^{(\ell+1)} \times M^{(\ell)}}$ and the bias vectors $b^{(\ell)} \in \mathbb{R}^{M^{(\ell+1)}}$, let

$$y^{(\ell)} = W^{(\ell)}u^{(\ell)} + b^{(\ell)} \quad (4)$$

where $\ell = 0, 1, \dots, L$, and

$$u_a^{(\ell)} = \left(1 + \exp\left(-y_a^{(\ell-1)}\right)\right)^{-1} \quad (5)$$

where $\ell = 1, \dots, L$ and $a = 1, \dots, M^{(\ell)}$. We define an artificial neural network nonlinearity $\psi : u \mapsto y$ by setting $u^{(0)} := u$ and $y := y^{(L)}$, so that that $M^{(0)} = \dim(u)$ and $M^{(L+1)} = \dim(y)$.

Sandberg et al. showed that single-layer time-delay neural networks constitute a universal approximator for fading memory operators [24, Theorem 1] and [22, Proposition 1]:

Lemma 1. Consider a time-invariant causal operator $F : \ell_\infty(\mathbb{Z}_+) \rightarrow \mathbb{R}(\mathbb{Z}_+)$ with fading memory on \mathcal{U}_β . For $n \in \mathbb{N}$, let $H^{(n)} : \ell_\infty(\mathbb{Z}_+) \rightarrow \mathbb{R}(\mathbb{Z}_+)^{(n+1)}$ be given by

$$(H^{(n)}v)_k = [v_k, v_{k-1}, \dots, v_{k-n}]^\top, \quad (6)$$

and let ψ be a neural network with $L = 1$, $M^{(0)} = n+1$ and $M^{(2)} = 1$. Then, there are $n, M^{(1)} > 0$ and real parameters $W^{(0)}, W^{(1)}, b^{(0)}, b^{(1)}$ such that for any $\epsilon > 0$,

$$|(Fv)_k - \psi((H^{(n)}v)_k)| < \epsilon, \quad k \in \mathbb{Z}_+$$

for all $v \in \mathcal{U}_\beta$.

To improve on the time-delay model structure above, we follow [26] and replace (6) by a set of GOBFs:

Definition 3. (see [21]) Let $\{\xi_0, \xi_1, \xi_2, \dots\}$ be a sequence of (possibly complex) poles such that $|\xi_i| < 1$ for all $i \in \mathbb{Z}_+$. The set of Generalized Orthonormal Basis Functions is defined by the causal transfer functions

$$\begin{aligned} G_0(z) &= z^d \frac{\sqrt{1 - |\xi_0|^2}}{z - \xi_0} \\ G_i(z) &= z^d \frac{\sqrt{1 - |\xi_i|^2}}{z - \xi_i} \prod_{j=0}^{i-1} \frac{1 - \bar{\xi}_j z}{z - \xi_j}, \quad i = 1, 2, \dots \end{aligned} \quad (7)$$

with $d = 0$ or $d = 1$.

Note that when $d = 1$ and all the poles in $\{\xi_i\}_{i \in \mathbb{Z}_+}$ are equal to zero, we recover the set of time-delay basis functions. The utility of GOBFs comes from the fact that poles are allowed to be distinct, and thus the choice of poles may be adapted to the system which is being approximated. The set of GOBFs may form a basis for various system spaces. In particular, a simple condition ensures that it is fundamental¹ in $\ell_1(\mathbb{Z}_+)$, the space of causal discrete-time LTI systems whose impulse responses are absolutely integrable [2, Corollary 10]:

Lemma 2. The set (7), with $d = 1$, is fundamental in $\ell_1(\mathbb{Z}_+)$ if $k^{-\alpha} = O(1 - |\xi_k|)$ for some $0 < \alpha < 1$.

The next result, which, as far as we know, has not been previously published, generalizes Lemma 1:

Theorem 1. Consider a time-invariant causal operator $F : \ell_\infty(\mathbb{Z}_+) \rightarrow \mathbb{R}(\mathbb{Z}_+)$ with fading memory on \mathcal{U}_β . Let $\{\xi_i\}_{i \in \mathbb{Z}_+}$ be a sequence of poles satisfying the condition stated in Lemma 2. For $m \in \mathbb{N}$, let $G^{(m)} : \ell_\infty(\mathbb{Z}_+) \rightarrow \mathbb{R}(\mathbb{Z}_+)^{(m+1)}$ be defined by

$$(G^{(m)}v)_k = [(G_0v)_k, (G_1v)_k, \dots, (G_mv)_k]^\top \quad (8)$$

with the G_i being the operators associated to (7) with $d = 1$. Let ψ be a neural network with $L \geq 1$, $M^{(0)} = m + 1$ and $M^{(L+1)} = 1$. Then, there are $m, M^{(\ell)} > 0$ and real parameters $W^{(\ell)}, b^{(\ell)}$ ($\ell = 1, \dots, L$) such that for any $\epsilon > 0$,

$$|(Fv)_k - \psi((G^{(m)}v)_k)| < \epsilon, \quad k \in \mathbb{Z}_+ \quad (9)$$

for all $v \in \mathcal{U}_\beta$.

Proof. See Appendix A. \square

To obtain a model structure that satisfies Lemma 2 and allows for a direct term from input to output, we fix a finite sequence $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, a number $n_{\text{rep}} \in \mathbb{N}$, and set $\xi_0 = 0$ and $\xi_{i+jn} = \lambda_i$, for $i = 1, \dots, n$ and $j = 0, \dots, n_{\text{rep}} - 1$. Let,

$$G(z) = [G_0(z), G_1(z), \dots, G_{nn_{\text{rep}}}(z)]^\top$$

where the SISO transfer functions $G_i(z)$ are given by (7), with $d = 1$. Let $\psi(\cdot; \theta)$ be an ANN specified by L and M , with the vector of parameters θ encompassing the elements of the matrices $W^{(\ell)}$ and the elements of the bias vectors $b^{(\ell)}$. Then the model structure is given by

$$\hat{y}_k(\theta) = \psi(G(q)v_k; \theta) \quad (10)$$

where q is the forward shift operator.

B. GOBF pole selection for neuronal identification

Since the choice of GOBF poles $\{\xi_i\}$ in the model structure is arbitrary, we pose the question of how to choose them in order to minimize the number of basis functions required to approximate (2) to a given tolerance. If the operator F_{int}

¹A fundamental set in a normed space X is a subset $M \subset X$ whose span is dense in X [14].

were linear and finite dimensional, with a transfer function $F_{\text{int}}(s)$, then the obvious choice would be to place $\{\xi_i\}$ as close as possible to the poles of $F_{\text{int}}(s)$. However, since F_{int} is nonlinear and, in general, infinite-dimensional, it is not immediately clear how to choose $\{\xi_i\}$. With that in mind, we now consider a heuristic for choosing the GOBF poles, keeping the discussion at the conceptual level.

The heuristic involves choosing the GOBF poles so as to ensure that the linearized model structure

$$\delta \hat{y}_k(\theta) = \left. \frac{\partial \psi(u; \theta)}{\partial u} \right|_{u=G(1)\bar{v}} G(q) \delta v_k$$

(whose poles are $\{\xi_i\}$) is able to approximate well the small-signal behavior of F_{int} around a particular setpoint $\bar{v} \in \mathbb{R}$. This is the behavior of F_{int} subject to an input $v_k = \bar{v} + \tilde{v}_k$, $k \geq 0$, where $|\tilde{v}_k|$ has a small amplitude. To identify excitable neuronal systems, the choice of the setpoint \bar{v} can be informed by the closed-loop nature of (1)-(2).

We argue that this setpoint \bar{v} should be a point of *ultra-sensitivity* [9] of the system (1)-(2). To explain this feature of neuronal behaviors, we consider a case in which the fading memory operator F_{int} has a smooth finite-dimensional realization. Since F_{int} has fading memory, when it is subject to a constant input $v_k = \bar{v}$, $k \geq 0$, its output converges to a unique constant value $i_\infty \in \mathbb{R}$, that is, $(F_{\text{int}}v)_k \rightarrow i_\infty$ as $k \rightarrow \infty$ (see, e.g., [3]). Let $dF_{\text{int}}(z; \bar{v})$ be the transfer function associated to the linearized realization of F_{int} at \bar{v} . Because \bar{v} is also an equilibrium of the closed-loop system (1)-(2) for $i_k = i_\infty$, we can write down the transfer function of the linearized closed-loop system at \bar{v} and i_∞ as

$$T(e^{j\omega t_s}; \bar{v}) = \frac{t_s}{c(e^{j\omega t_s} - 1) + t_s dF_{\text{int}}(e^{j\omega t_s}; \bar{v})} \quad (11)$$

where $\omega \geq 0$ denotes the frequency variable. Neuronal systems are characterized by the existence of at least one point $(\bar{\omega}, \bar{v})$ where (11) is singular. At such a point, $T(z; \bar{v})$ has a pole on the unit circle, which, in terms of state-space dynamics, corresponds to an equilibrium bifurcation at \bar{v} (see [13] for a state-space perspective). For system identification, the most relevant bifurcation occurs at the equilibrium voltage $\bar{v} = v^*$ where the neuronal behavior transitions from a constant-steady state to an oscillatory or chaotic motion. We call v^* a point of ultra-sensitivity [9].

Importantly, the point of ultra-sensitivity can be characterized experimentally; this can be done, for instance, by stepping the input i_k to different constant values, as illustrated in Figure 2. The critical role that v^* plays in the dynamics of (1)-(2) suggests that we should place the GOBF poles as close as possible to the (stable) poles of $dF_{\text{int}}(z; v^*)$, since those are the poles that shape the closed-loop frequency-response (11) at a point of ultra-sensitivity.

This heuristic requires that the poles of $dF_{\text{int}}(z; v^*)$ be identified prior to identifying F_{int} . But since small changes in the current i_k may cause large voltage deviations away from v^* , it is not immediately clear how to achieve that. The answer lies in *voltage-clamp*: the output feedback law (3) can be used to stabilize the system at the voltage v^* , suppressing

any oscillations and allowing the use of a low-amplitude input r_k to probe the small-signal behavior of F_{int} around v^* . If signal amplitudes are kept low, linear system identification methods can be applied to the measurements obtained under the output feedback law (3), leading to estimates for the poles of $dF_{\text{int}}(z; v^*)$; these are in turn used to define the GOBFs.

This heuristic leads to a systematic choice for the GOBF poles that takes into account the ultra-sensitivity of neuronal systems. Notice that, in principle, a finite-dimensional F_{int} is not required. The heuristic is in line with the idea that a pre-processing linear identification step should be implemented, if possible, in order to make an informed choice on GOBF poles for nonlinear system identification [27].

C. Parameter estimation and validation

In practice, given measurements of v_k and the input i_k , the system (1)-(2) can be identified by solving

$$\min_{\eta, \theta} \frac{1}{N} \sum_{k=0}^N \left(\frac{v_{k+1} - v_k}{t_s} - (\hat{y}_k(\theta) + \eta i_k) \right)^2 \quad (12)$$

where η is used to obtain an estimate of $1/c$. Because $\hat{y}_k(\theta)$ is a feed-forward ANN, the above problem can be efficiently solved with backpropagation algorithms. Once this is achieved, we can define

$$\hat{v}_{k+1} = \hat{v}_k + t_s \eta (-\psi(G(q)\hat{v}_k; \theta) + i_k) \quad (13)$$

to obtain a closed-loop identified neuronal model.

Due to the frequent bifurcations happening in a neuronal model, different metrics are used for validating the subthreshold and the superthreshold (i.e., spiking) dynamics of the model [10, Section 10.3]. Here, we will focus on validating the latter. This can be done by comparing the spike timings of the identified model, denoted by \hat{k}_s , with the spike timings of the validation dataset, denoted by k_s (we assume the timing of a spike is given by the location of its maximum). Spike timings define impulse trains (called spike trains) given by $s_k = \sum_{k_s} \delta_{kk_s}$ and $\hat{s}_k = \sum_{\hat{k}_s} \delta_{k\hat{k}_s}$, where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ otherwise. A measure of spike coincidence can be defined using the inner product of smoothed spike trains,

$$(s, \hat{s}) = \sum_{k=0}^N (w * s)(k) (w * \hat{s})(k),$$

where w_k is a smoothing kernel. Here, we will use a Gaussian kernel $w_k = \exp(-k^2/(2\rho^2))/\sqrt{2\pi\rho^2}$. The angular separation between the smoothed spike trains, measured by

$$\Delta_\rho = \frac{(s, \hat{s})}{\sqrt{(s, s)}\sqrt{(\hat{s}, \hat{s})}}, \quad (14)$$

is a good measure of spike coincidence: Δ_ρ approaches 1 for very similar spike trains, and 0 for very different ones.

IV. EXAMPLE

The purpose of this section is to illustrate the impact of the choice of basis functions (7) on the identification of a single-compartment neuron. For this illustration, we use a neuronal model from the crab stomatogastric ganglion

(STG), a system responsible for producing rhythmic muscular activity in the crab’s stomach [20]. Neurons in the STG are capable of bursting autonomously, and, to our knowledge, they have never been successfully identified using a generic model structure. To generate data for parameter estimation, we use a forward-Euler discretization of the conductance-based STG neuron model described in [8, Figure 1.A.a]; see also [15, pp. 2318-2319] for the model’s ion channel kinetics. The model in question is a state-space model with 12 states representing the membrane voltage and six different types of ionic currents. It contains over a hundred parameters, which were originally determined by first using ad-hoc methods for fitting individual ionic currents, and then hand tuning the parameters to match the observed neuronal behavior.

We compare the validation performance of the closed-loop model (13) in two cases: when the basis functions used to define $G(z)$ are given by time-delays, and when they are given by GOBFs with poles chosen in accordance with Section III-B. We denote the basis function vector in each of these cases by $G^{\text{td}}(z)$ and $G^{\text{gobf}}(z)$, respectively. In both cases, we use the same two-layer ANN parametrization, with $M^{(1)} = 15$ and $M^{(2)} = 12$. Letting n_{td} be a maximum time delay, we set $G^{\text{td}}(z) = [1, z^{-1}, \dots, z^{-(n_{\text{td}}-1)}]^\top$. To choose the GOBF poles used to define $G^{\text{gobf}}(z)$, for brevity, we assume knowledge of the local dynamics of the continuous-time STG model². It is known that the STG model starts bursting due to a saddle-node bifurcation close to $i_\infty = -0.25$ and $\bar{v} = -49$ mV. At this equilibrium, the linearized F_{int} has poles given by $\{\xi_i^{\text{ct}}\}_{i=1,\dots,11} = \{-7.716, -0.560, -0.179, -0.156, -0.112, -0.074, -0.050, -0.037, -0.019, -0.018, -0.004\}$. We define twelve GOBFs ($d = 1$) by setting discrete-time poles at $\xi_0 = 0$ and $\xi_i = 1 + t_s \xi_i^{\text{ct}}$, with t_s the sampling time (we use $n_{\text{rep}} = 1$).

Using $t_s = 0.0075$, we identify the excitability of the discretized STG model based on its response v_k to the input

$$\hat{i}_k = -0.5 + \varepsilon_k \quad (15)$$

where ε_k is a Gaussian white-noise signal such that $\sigma[\varepsilon_k] = 20 \mu\text{A}/\text{cm}^2$. We obtain a training and a validation dataset, both of length $Nt_s \approx 10$ s, using two different realizations of (15). After training³ each of the models defined with G^{td} and G^{gobf} , we simulated the resulting system (13) with the validation input, obtaining \hat{v}_k^{td} and \hat{v}_k^{gobf} , respectively.

Figure 3 shows the validation voltage v_k and the simulated voltages \hat{v}_k^{td} and \hat{v}_k^{gobf} during; we have used $n_{\text{td}} = 12$ so that both models have the same number of basis functions. The spike coincidence metric (14), computed with $\rho = 3$ using the full validation dataset⁴, was $\Delta_3 = 0.42$ for the model with time-delays, and $\Delta_3 = 0.73$ for the model with GOBFs.

²For numerical simulations where the poles of $dF_{\text{int}}(z; \bar{v})$ are identified using output feedback and linear identification methods, see [4, Section 4.6].

³We trained the ANNs of the two model structures by solving (12) with the Levenberg-Marquadt backpropagation algorithm. The models were trained with ten different sets of randomly selected initial parameters, and the best fitted models were used in the results. To eliminate transient effects, we discarded 1 s of data from the training dataset.

⁴The choice of $\rho = 3$ ms for the smoothing kernel standard deviation was based on the mean width of the observed spikes, which was of 6 ms.

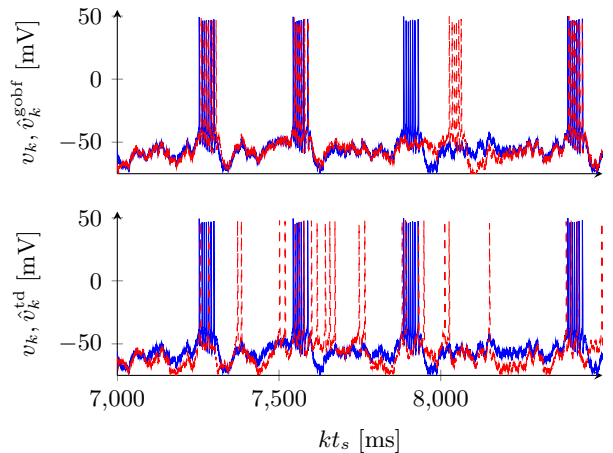


Fig. 3. Top: validation v_k of the true STG model (blue, solid) and output \hat{v}_k^{gobf} of the model defined with twelve GOBFs (red, dashed). Bottom: validation v_k of the true STG model (blue, solid) and output \hat{v}_k^{td} of the model defined with $n_{\text{td}} = 12$ time delays (red, dashed).

Thus it can be said regarding the model structure (10) that using GOBFs with judiciously chosen poles leads to better results than using time delay basis functions. In particular, the model obtained with a maximum delay of $n_{\text{td}} = 12$ samples can *spike*, but it cannot *burst*. This highlights the importance of long timescales in bursting models.

V. CONCLUSION

In this paper, we have shown that GOBFs and static ANNs can be used to efficiently solve the problem of identifying biophysical neuronal systems. One of the main advantages of the method is its simplicity: the estimation of model parameters can be done with off-the-shelf backpropagation algorithms. In addition, the method does not rely on measurements of internal states of the neuron – something which is, in practice, impossible to obtain.

The method also dispenses entirely with the conductance-based formalism, in which optimization of all model parameters may become an intractable problem. This comes, of course, at the cost of biophysical interpretability. In fact, further work on this topic should aim to understand how much the model structure discussed here can be improved so as to provide more interpretability while retaining its capabilities as a universal approximator.

Finally, we point out that while the use of feedforward ANNs for the identification of nonlinear systems is classical, such networks can only be guaranteed to approximate fading memory operators, de facto excluding the excitable nature of neuronal systems. Our approach shows that by retaining the biophysical feedback structure of the model in the identification step, such a limitation can be overcome.

APPENDIX

A. Proof of Theorem 1

We first prove the result for a single-layer ANN ($L = 1$) by modifying the input weights of the ANN obtained from Lemma 1. Let $\epsilon_0 > 0$ be given. By Lemma 1, there is a

number $n = n(\epsilon_0) > 0$, and parameters $M^{(1)} > 0$ and $W_{\text{fir}}^{(0)} \in \mathbb{R}^{M^{(1)} \times (n+1)}$, defining an artificial neural network $\psi : \mathbb{R}^{(n+1)} \rightarrow \mathbb{R}$ such that

$$|(Fv)_k - \psi((H^{(n)}v)_k)| < \epsilon_0, \quad k \in \mathbb{Z}_+$$

for all $v \in \mathcal{U}_\beta$. Let $W_{\text{obf}}^{(0)} \in \mathbb{R}^{M^{(1)} \times (m+1)}$, and consider the vector of transfer functions

$$Q^{(n,m)}(z) = W_{\text{fir}}^{(0)} H^{(n)}(z) - W_{\text{obf}}^{(0)} G^{(m)}(z)$$

By Lemma 2, for every n and every $\epsilon_1 > 0$, there exists an $m = m(n, \epsilon_1) > 0$, and a certain $W_{\text{obf}}^{(0)}$ such that

$$\|q_i^{(n,m)}\|_{\ell_1} = \sum_{k=0}^{\infty} |(q_i^{(n,m)})_k| \leq \epsilon_1, \quad i = 1, \dots, M^{(1)},$$

where $q_i^{(n,m)}$ is the impulse response of the i^{th} element of $Q^{(n,m)}(z)$. Now, let $\tilde{\psi} : \mathbb{R}^{M^{(1)}} \rightarrow \mathbb{R}$ be defined by $\tilde{\psi}(W_{\text{fir}}^{(0)} u) = \psi(u)$. Since the activation functions (5) are globally Lipschitz, there exists a constant l_0 such that

$$\begin{aligned} & |\tilde{\psi}(W_{\text{fir}}^{(0)}(H^{(n)}v)_k) - \tilde{\psi}(W_{\text{obf}}^{(0)}(G^{(m)}v)_k)| \\ & \leq l_0 \|W_{\text{fir}}^{(0)}(H^{(n)}v)_k - W_{\text{obf}}^{(0)}(G^{(m)}v)_k\|_\infty \\ & = l_0 \|(Q^{(n,m)}v)_k\|_\infty \\ & \leq l_0 \max_i \|q_i^{(n,m)}\|_{\ell_1} \|v\|_{\ell_\infty} \leq l_0 \epsilon_1 \beta \end{aligned}$$

for all $k \in \mathbb{Z}_+$. Here, we have used the fact that, for an arbitrary LTI system $Q : \ell_\infty \rightarrow \ell_\infty^M$, $\max_{i=1, \dots, M} \|q_i\|_{\ell_1}$ is the induced system gain [29], and that the elements of \mathcal{U}_β are bounded by β . From the above inequalities, we obtain

$$\begin{aligned} |(Fv)_k - \tilde{\psi}(W_{\text{obf}}^{(0)}(G^{(m)}v)_k)| & \leq |(Fv)_k - \psi((H^{(n)}v)_k)| \\ & \quad + |\psi((H^{(n)}v)_k) - \tilde{\psi}(W_{\text{obf}}^{(0)}(G^{(m)}v)_k)| \\ & \leq \epsilon_0 + l_0 \epsilon_1 \beta \end{aligned}$$

for all $k \in \mathbb{Z}_+$ and all $v \in \mathcal{U}_\beta$. Thus given $\epsilon > 0$, we can choose $\epsilon_0 < \epsilon/2$ and $\epsilon_1 < \epsilon/(2l_0\beta)$, and there exists a large enough m such that the ANN $\tilde{\psi}(W_{\text{obf}}^{(0)} \cdot)$ satisfies (9).

This proves the result for $L = 1$. The fact that the result also holds for a multi-layer ANN ($L > 1$) follows from the fact that multi-layer ANNs with continuous activation functions are capable of arbitrarily accurate approximation (in the uniform norm) to any continuous function over a compact set [12, Theorem 2.1]. A multi-layer ANN can thus be used to approximate the single-layer network obtained above to arbitrary precision, concluding the proof.

REFERENCES

- [1] H. Abarbanel, D. Creveling, R. Farsian, and M. Kostuk. Dynamical State and Parameter Estimation. *SIAM Journal on Applied Dynamical Systems*, 8(4):1341–1381, January 2009.
- [2] Hüseyin Akçay and Brett Ninness. Rational Basis Functions for Robust Identification from Frequency and Time-Domain Measurements. *Automatica*, 34(9):1101–1117, September 1998.
- [3] S. Boyd and L. Chua. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, 32(11):1150–1161, November 1985.
- [4] Thiago B. Burghi. *Feedback for neuronal system identification*. PhD Thesis, University of Cambridge, Cambridge, UK, 2020.
- [5] Thiago B. Burghi, Maarten Schoukens, and Rodolphe Sepulchre. Feedback Identification of conductance-based models. *Automatica*, 2020. In press; available at: <https://arxiv.org/abs/2002.09626>.
- [6] G. Drion, T. O’Leary, J. Dethier, A. Franci, and R. Sepulchre. Neuronal behaviors: A control perspective. In *54th IEEE Conference on Decision and Control*, pages 1923–1944, December 2015.
- [7] Urban Forssell and Lennart Ljung. Closed-loop identification revisited. *Automatica*, 35(7):1215–1241, July 1999.
- [8] Alessio Franci, Guillaume Drion, and Rodolphe Sepulchre. Robust and tunable bursting requires slow positive feedback. *Journal of Neurophysiology*, 119(3):1222–1234, December 2017.
- [9] Alessio Franci, Guillaume Drion, and Rodolphe Sepulchre. The sensitivity function of excitable feedback systems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4723–4728, December 2019. ISSN: 2576-2370.
- [10] Wulfram Gerstner, Werner M. Kistler, Richard Naud, and Liam Paninski. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, Cambridge, UK, 2014.
- [11] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, August 1952.
- [12] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, January 1989.
- [13] Eugene M. Izhikevich. *Dynamical Systems in Neuroscience*. MIT Press, Cambridge, MA, 2007.
- [14] Erwin Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons, Hoboken, NJ, 1989.
- [15] Zheng Liu, Jorge Golowasch, Eve Marder, and L. F. Abbott. A Model Neuron with Activity-Dependent Conductances Regulated by Multiple Calcium Sensors. *Journal of Neuroscience*, 18(7):2309–2320, April 1998.
- [16] L. Ljung. Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control*, 23(5):770–783, October 1978.
- [17] Winfried Lohmiller and Jean-Jacques E. Slotine. On Contraction Analysis for Non-linear Systems. *Automatica*, 34(6):683–696, June 1998.
- [18] Wolfgang Maass and Eduardo D. Sontag. Neural Systems as Nonlinear Filters. *Neural Computation*, 12(8):1743–1772, August 2000.
- [19] I. R. Manchester, M. M. Tobenkin, and J. Wang. Identification of nonlinear systems with stable oscillations. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 5792–5797, Orlando, FL, December 2011.
- [20] Eve Marder and Dirk Bucher. Understanding Circuit Dynamics Using the Stomatogastric Nervous System of Lobsters and Crabs. *Annual Review of Physiology*, 69(1):291–316, 2007.
- [21] B. Ninness and F. Gustafsson. A unifying construction of orthonormal bases for system identification. *IEEE Transactions on Automatic Control*, 42(4):515–521, April 1997.
- [22] J. Park and I. W. Sandberg. Criteria for the approximation of nonlinear systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 39(8):673–676, August 1992.
- [23] Max Revay and Ian Manchester. Contracting Implicit Recurrent Neural Networks: Stable Models with Improved Trainability. In *Proceedings of Machine Learning Research*, volume 120, pages 1–11, The cloud, June 2020.
- [24] I. W. Sandberg. Approximation theorems for discrete-time systems. *IEEE Transactions on Circuits and Systems*, 38(5):564–566, May 1991.
- [25] Maarten Schoukens and Koen Tiels. Identification of block-oriented nonlinear systems starting from linear approximations: A survey. *Automatica*, 85:272–292, November 2017.
- [26] G. Schram, M.H.G. Verhaegen, and A.J. Krijgsman. System Identification with Orthogonal Basis Functions and Neural Networks. *IFAC Proceedings Volumes*, 29(1):4150–4155, June 1996.
- [27] Koen Tiels and Johan Schoukens. Wiener system identification with generalized orthonormal basis functions. *Automatica*, 50(12):3147–3154, December 2014.
- [28] T. Wigren. Nonlinear identification of neuron models. In *2015 IEEE Conference on Control Applications (CCA)*, pages 1340–1346, September 2015.
- [29] Kemin Zhou, John Comstock Doyle, and Keith Glover. *Robust and Optimal Control*. Prentice Hall, Upper Saddle River, NJ, 1996.