Distributed online estimation of biophysical neural networks*

Thiago B. Burghi[1], Timothy O'Leary[1] and Rodolphe Sepulchre[1]

Abstract— In this work, we propose a distributed adaptive observer for a class of nonlinear networked systems inspired by biophysical neural network models. Neural systems learn by adjusting intrinsic and synaptic weights in a distributed fashion, with neuronal membrane voltages carrying information from neighbouring neurons in the network. We show that this learning principle can be used to design an adaptive observer based on a decentralized learning rule that greatly reduces the number of observer states required for exponential convergence of parameter estimates. This novel design is relevant for biological, biomedical and neuromorphic applications.

I. INTRODUCTION

With the improvement of neural recording technology, it may soon be possible to concurrently monitor the membrane potential of hundreds of interconnected neurons in a living brain [1]. This high-resolution data opens up new possibilities for the development of real-time closed-loop interventions aimed at treating disorders of neural excitability such as epilepsy and Parkinson's [2]. The capability to effectively monitor and control spiking systems also impacts the nascent field of neuromorphic engineering [3].

Good closed-loop control design often requires reliable model estimates, and hence any method aimed at controlling neural activity is bound to involve the estimation of neuronal models, which is a nontrivial task. Many techniques have been proposed for *batch-mode* or *offline* estimation of neuronal dynamics, see for instance [4], [5], [6], [7]. However, living brain systems are adaptive [8], and thus online estimation approaches are necessary, especially if real-time applications are involved.

To meet this demand, an adaptive observer-based approach for online estimation of conductance-based neural networks was recently proposed in [9]. The adaptive observers, inspired by [10] and [11], are rooted in the familiar Recursive Least Squares (RLS) algorithm [12], and allow for approximately tracking slowly time-varying parameters. One limitation of RLS-based adaptive observers is the rapid increase in observer states with respect to the number of parameters. More observer states require more computing power, which might become critical when attempting to perform online estimation of large neuronal network models containing thousands of parameters. In this paper, we propose a distributed version of the linear-in-the-parameters

adaptive observer from [9] that results in a scalable algorithm for online parameter estimation of biophysical neural models. The proposed modification, which echoes the diagonal RLS-like update rule of [13], greatly reduces the number of adaptive observer states. We show that for neuronal network models, the proposed adaptive observer becomes distributed over individual neuronal membrane currents and also over neurons in the network. We analyse the adaptive observer using contraction theory [14], and show that a strengthened persistent excitation condition is sufficient for consistent convergence of the parameter estimates.

The paper is organized as follows: in Section II, we pose the problem from an abstract point of view and recall the observer from [9]. In Section III, we introduce and analyse the modified distributed observer. In Section IV, we use the observer to estimate conductance-based biophysical neural networks. In Section V, we discuss the relevance of this work in neuroscience, as well as future research directions.

Notation: We write I_n for the $n \times n$ identity matrix, and I when n is obvious from the context. For two column vectors x and y, we write $\operatorname{col}(x, y) := (x^\mathsf{T}, y^\mathsf{T})^\mathsf{T}$. For a matrix $A \in \mathbb{R}^{n \times n}$, $\lambda_{\max}(A)$ denotes the largest eigenvalue of A. For a vector-valued function $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^m$, we write $\partial_x f(x, y) \in \mathbb{R}^{m \times n_1}$ for the Jacobian of $f(x, y)$ with respect to x. We write $A \succeq B$ ($A \succ B$) if $A - B$ is a positive-semidefinite (positive-definite) matrix.

II. BACKGROUND

We consider nonlinear state-space systems of the form

$$\dot{v} = \sum_{j=1}^{m} \Phi_j^\mathsf{T}(v, w^j, u)\theta^j + a(v, w, u) \tag{1a}$$

$$\dot{w}^j = g_j(v, w^j) \tag{1b}$$

for $j = 1, \ldots, m$. Here, $v \in \mathbb{R}^{n_v}$ is a state vector, which is also the output of the system; $w = \operatorname{col}(w^1, \ldots, w^m)$ is an internal dynamics state vector, with $w^j \in \mathbb{R}^{n_w^j}$; $u \in \mathbb{R}^{n_u}$ a control input vector; and $\theta = \operatorname{col}(\theta^1, \ldots, \theta^m)$ is a parameter vector, with $\theta^j \in \mathbb{R}^{n_\theta^j}$. The matrices $\Phi_j(v, w^j, u) \in \mathbb{R}^{n_\theta^j \times n_v}$ and the vectors $a(v, w, u) \in \mathbb{R}^{n_v}$ and $g_j(v, w^j) \in \mathbb{R}^{n_w^j}$ are assumed to be continuously differentiable in their arguments. We will also use the more compact notation

$$\Phi^\mathsf{T}(v, w, u) := \begin{bmatrix} \Phi_1^\mathsf{T}(v, w^1, u) & \ldots & \Phi_m^\mathsf{T}(v, w^m, u) \end{bmatrix}$$

and

$$g(v, w) := \operatorname{col}(g_1(v, w^1), \ldots, g_m(v, w^m)) .$$

The role of an adaptive observer is to provide an online estimation of the states and the parameters of the system from measurements of the input $u(t)$ and output $v(t)$.

The structure of (1) is motivated by models of neuronal dynamics [15]. The state v represents a vector of membrane voltages in a neural network, while w represents a vector of *gating variables* that dictate ion channel and synaptic dynamics. This specific application is discussed in Section IV.

In our problem formulation, we assume that the trajectories of the system (1) evolve in a compact positively invariant set, and that the internal dynamics of (1) are exponentially contracting [14], uniformly in v:

Assumption 1. There exists a compact set $V \times U$ such that $\{v(t), u(t)\} \in V \times U$ for all $t \geq 0$.

Assumption 2. For each $j = 1, \ldots, m$, there exists a compact convex set W_j which is positively invariant with respect to (1b), uniformly in v on $\mathbb{R}^{n_v}$. Furthermore, there exist a symmetric matrix $M_j(t) = \Theta_j(t)^{\mathsf{T}} \Theta_j(t)$ such that $\underline{\sigma} I \preceq M_j(t) \preceq \overline{\sigma} I$ for some $\underline{\sigma}, \overline{\sigma} > 0$, and a contraction rate $\lambda_j > 0$ such that the generalized Jacobian

$$F_j := (\dot{\Theta}_j + \Theta_j \partial_{w^j} g_j(v, w^j)) \Theta_j^{-1} \qquad (2)$$

satisfies

$$F_j + F_j^{\mathsf{T}} \preceq -\lambda_j I \qquad (3)$$

for all $\{v, w^j\} \in \mathbb{R}^{n_v} \times W_j$ and all $t \geq 0$.

In [9], an adaptive observer-based approach (inspired by the earlier designs of [10] and [11]) was proposed to estimate the parameter vector $\theta = \mathrm{col}(\theta^1, \ldots, \theta^m)$ in real-time. In the present work, our point of departure is the adaptive observer given by

$$\dot{\hat{v}} = \Phi^{\mathsf{T}}(v, \hat{w}, u)\hat{\theta} + a(v, \hat{w}, u) + \gamma(I + \Psi^{\mathsf{T}} P \Psi)(v - \hat{v})$$
$$\dot{\hat{w}} = g(v, \hat{w}) \qquad (4a)$$
$$\dot{\hat{\theta}} = \gamma P \Psi (v - \hat{v})$$

where the matrices P and Ψ evolve according to

$$\dot{\Psi} = -\gamma \Psi + \Phi(v, \hat{w}, u), \qquad (4b)$$
$$\dot{P} = \alpha P - \alpha P \Psi \Psi^{\mathsf{T}} P, \qquad P(0) \succ 0, \qquad (4c)$$

with $\gamma > \alpha > 0$. Under Assumptions 1 and 2, and under the persistent excitation (PE) assumption that

$$\exists T > 0, \ \forall t \geq 0 \ : \ \underline{\delta} I \preceq \int_t^{t+T} \Psi(\tau) \Psi^{\mathsf{T}}(\tau) d\tau \preceq \overline{\delta} I \quad (5)$$

for some $\overline{\delta}, \underline{\delta} > 0$, it can be shown that the adaptive observer state vector $\mathrm{col}(\hat{v}(t), \hat{w}(t), \hat{\theta}(t))$ converges to $\mathrm{col}(v(t), w(t), \theta)$ exponentially fast as $t \to \infty$ (see [9, Theorem 1] and its proof).

The adaptive observer (4) relies on the $n_\theta \times n_\theta$ matrices $P(t)$ and $\Psi \Psi^{\mathsf{T}}$ to update the parameter estimates $\hat{\theta}$. From a computational point of view, when the number $n_\theta = \sum_{j=1}^{m} n_\theta^j$ is large, updating the $(n_\theta)^2$ states of $P(t)$ becomes costly. In this paper, we are interested in redesigning the adaptive observer above so as to decrease the number of required observer states in $P(t)$.

We will explore the simple idea that the matrix $\Psi \Psi^{\mathsf{T}}$ can be approximated by its (block) diagonal elements $\Psi_j \Psi_j^{\mathsf{T}}$.

under suitable assumptions. This will lead to a *decoupled* version of the update rule for each component θ^j of θ which is in addition *distributed* with respect to the internal dynamics states w^j.

III. DISTRIBUTED ADAPTIVE OBSERVER

We now consider the adaptive observer design given by

$$\dot{\hat{v}} = \sum_{j=1}^{m} \Phi_j^{\mathsf{T}}(v, \hat{w}^j, u)\hat{\theta}^j + a(v, \hat{w}, u) \qquad (6a)$$
$$+ \left(\gamma_0 I + \gamma_j \sum_{j=1}^{m} \Psi_j^{\mathsf{T}} P_j \Psi_j\right)(v - \hat{v})$$
$$\dot{\hat{w}}^j = g_j(v, \hat{w}^j) \qquad (6b)$$
$$\dot{\hat{\theta}}^j = \gamma_j P_j \Psi_j (v - \hat{v}) \qquad (6c)$$

where $\gamma_0, \gamma_1, \ldots, \gamma_m > 0$ are constant gains, and the matrices P_j and Ψ_j evolve according to

$$\dot{\Psi}_j = -\gamma_j \Psi_j + \Phi_j(v, \hat{w}^j, u), \qquad \Psi_j(0) = 0, \quad (6d)$$
$$\dot{P}_j = \alpha_j P_j - \alpha_j P_j \Psi_j \Psi_j^{\mathsf{T}} P_j, \qquad P_j(0) \succ 0, \quad (6e)$$

for all $j = 1, \ldots, m$ (there is no loss of generality in the choice of $\Psi_j(0)$ above).

The adaptive observer (6) is obtained from (4) by ignoring the off-diagonal terms in $\Psi \Psi^{\mathsf{T}}$, enforcing a block-diagonal structure in the matrix $P(t)$, and allowing for different learning gains $\gamma_j > 0$ and forgetting rates $\alpha_j > 0$. In the new design, the adaptation law given by (6c)-(6e) is distributed with respect to each "regressor" Φ_j^{T} and its associated internal dynamics (6b). The total number of states in the matrices P_j for $j = 1, \ldots, m$ is now reduced to $m \times \sum_{j=1}^{m}(n_\theta^j)^2 \leq (n_\theta)^2$. In addition, the learning gains and forgetting rates can now be chosen independently, which might be beneficial to account for different timescales in the internal dynamics components (1b). These timescales correspond to the contraction rates λ_j in Assumption 2.

Remark 1. The distributed adaptive observer (6) was designed by modifying the design of [9], which exactly solves a Recursive Least Squares (RLS) problem with exponential forgetting [12, Chapter 2]. In a similar spirit, a modification of the RLS algorithm based on a diagonal update rule was previously proposed by [13]. However, not much is known about the convergence properties of such diagonal update rules, and they do not seem to have been considered in the context of adaptive observers.

Remark 2. Many authors have studied the related problem of *directional forgetting* in RLS estimation, e.g. [16], [17], [18]. The adaptive observer above is not based on directional forgetting methods, however, since such methods still consider off-diagonal terms of $\Psi \Psi^{\mathsf{T}}$ in the update rule. It is also worth mentioning that the diagonal update rule above is reminiscent of the idea of *splitting across features* when solving regression problems with the Alternating Direction Method of Multipliers [19].

Remark 3. While the rather general PE condition (5) is a sufficient condition for the consistent exponential convergence of the adaptive observer (4), the same condition may not be sufficient for exponential convergence of the adaptive observer with the diagonal update rule (6). The reason can be illustrated by considering the simplified linear model

$$y = \gamma \Psi(t)^\mathsf{T} \theta$$

with known $\Psi(t)$. It is well-known that the update rule

$$\dot{\hat{\theta}} = P\Psi(y - \gamma \Psi^\mathsf{T} \hat{\theta}) \tag{7}$$

with P given by (4c) ensures that $\hat{\theta}(t) \to \theta$ exponentially fast as long as the PE condition (5) holds. Indeed, the PE condition ensures that $P(t)$ is uniformly positive definite and bounded above [10], leading to the Lyapunov function candidate $V(t, \tilde{\theta}) = \tilde{\theta}^\mathsf{T} P^{-1}(t) \tilde{\theta}$ where $\tilde{\theta} := \theta - \hat{\theta}$. This function satisfies

$$\dot{V}(t, \tilde{\theta}) = -\tilde{\theta}^\mathsf{T}(\alpha P^{-1} + (2\gamma - \alpha)\Psi \Psi^\mathsf{T})\tilde{\theta},$$

which is uniformly negative definite for $\gamma > \alpha/2$. Suppose now that we replace the update rule (7) by its diagonal counterpart

$$\dot{\hat{\theta}} = \bar{P}\Psi(y - \gamma \Psi^\mathsf{T} \hat{\theta}),$$

with $\bar{P} := \mathrm{diag}\{P_1, \ldots, P_m\}$ and $P_j(t)$ given by (6e), where for simplicity we take $\alpha_j = \alpha$. Persistent excitation with respect to each $\Psi_j \Psi_j^\mathsf{T}$ still ensures that $\bar{P}(t)$ is uniformly positive definite and bounded above, but PE alone is not sufficient to ensure the uniform negative definiteness of the derivative of the Lyapunov function candidate $\bar{V}(t, \tilde{\theta}) = \tilde{\theta}^\mathsf{T} \bar{P}^{-1}(t) \tilde{\theta}$.

The previous remark illustrates that the convergence analysis of the distributed observer requires further investigation. In the same vein as [20], [21], we shall give up on negative definiteness (or semidefiniteness) of the derivative of the Lyapunov function. To state our final assumption, we define the matrices

$$
\begin{aligned}
\bar{P} &:= \mathrm{diag}\{P_1, \ldots, P_m\}, \\
\Gamma &:= \mathrm{diag}\{\gamma_1 I_{n_\theta^1}, \ldots, \gamma_m I_{n_\theta^m}\}, \\
A &:= \mathrm{diag}\{\alpha_1 I_{n_\theta^1}, \ldots, \alpha_m I_{n_\theta^m}\}, \\
D &:= \mathrm{diag}\{\Psi_1 \Psi_1^\mathsf{T}, \ldots, \Psi_m \Psi_m^\mathsf{T}\},
\end{aligned}
\tag{8}
$$

where the P_j above come from (6e).

Assumption 3. For all $v(t), u(t)$, there exists a $T > 0$ such that for all $t \geq 0$, the following hold:
  (i) For each $j = 1, \ldots, m$, we have

$$\int_t^{t+T} \Psi_j(\tau)\Psi_j(\tau)^\mathsf{T} d\tau \succeq \delta_j$$

  for some $\delta_j > 0$.
 (ii) Let $\underline{\alpha} = \min\{\alpha_1, \ldots, \alpha_m\}$. Then there exists a $\beta > 0$ such that

$$\frac{1}{T} \int_t^{t+T} \lambda_{\max}(AD + \gamma_0 \Psi \Psi^\mathsf{T} - \Gamma \Psi \Psi^\mathsf{T} - \Psi \Psi^\mathsf{T} \Gamma) d\tau$$

$$\leq (\underline{\alpha} - \beta) \times \min_j \delta_j \alpha_j e^{-2\alpha_j T}.$$

We can now state our main theoretical result:

Theorem 1. *Consider the true system* (1) *and the adaptive observer* (6). *Under Assumptions 1 to 3, for any* $\hat{v}(0) \in \mathbb{R}^{n_v}$, $\hat{w}(0) \in \prod_j W_j$, *and* $\hat{\theta}(0) \in \mathbb{R}^{n_\theta}$, *we have*

$$\mathrm{col}(\hat{v}(t), \hat{w}(t), \hat{\theta}(t)) \to \mathrm{col}(v(t), w(t), \theta)$$

exponentially fast as $t \to \infty$.

Proof. See Appendix A. □

Remark 4. Assumption 3 part (*ii*) shows that to promote consistent parameter estimation, we can pick $\Gamma = \gamma_0 I$ with $\gamma_0 \gg \alpha_j$ for all j: then $AD + \gamma_0 \Psi \Psi^\mathsf{T} - \Gamma \Psi \Psi^\mathsf{T} - \Psi \Psi^\mathsf{T} \Gamma \approx -\gamma_0 \Psi \Psi^\mathsf{T} \preceq 0$. However, the freedom in choosing different γ_j allows us to be more strategic in terms of how to choose those gains to ensure convergence of the observer, especially when some prior information about the system is available. This is illustrated in Section IV-A.

IV. APPLICATION TO BIOPHYSICAL NEURAL NETWORKS

In the context of biophysical neural network models, each component v_i of the vector $v \in \mathbb{R}^{n_v}$ represents the membrane potential of a single neuronal cell. In each of these cells, the membrane potential evolves according to

$$c_i \dot{v}_i = -\sum_{\mathrm{ion} \in \mathcal{I}} I_i^{\mathrm{ion}} - \sum_{\mathrm{syn} \in \mathcal{S}} \sum_{k \neq i} I_{i,k}^{\mathrm{syn}} - \mu_i^{\mathrm{leak}}(v_i - \nu^{\mathrm{leak}}) + u_i, \tag{9}$$

where $c_i > 0$ is a capacitance,

$$I_i^{\mathrm{ion}} = \mu_i^{\mathrm{ion}}(m_i^{\mathrm{ion}})^{p^{\mathrm{ion}}}(h_i^{\mathrm{ion}})^{q^{\mathrm{ion}}}(v_i - \nu^{\mathrm{ion}}) \tag{10}$$

are *intrinsic ionic currents*,

$$I_{i,k}^{\mathrm{syn}} = \mu_{i,k}^{\mathrm{syn}} s_{i,k}^{\mathrm{syn}}(v_i - \nu^{\mathrm{syn}}) \tag{11}$$

are *synaptic currents*, and $\mu_i^{\mathrm{leak}}(v_i - \nu^{\mathrm{leak}})$ is a leak current. The set $\mathcal{I}$ collects ionic current types, while $\mathcal{S}$ collects synaptic current types. The scalars $\mu_i^{\mathrm{ion}} > 0$ and $\mu_{i,k}^{\mathrm{syn}} > 0$ are *intrinsic* and *synaptic maximal conductances*, and the scalars $\nu^{\mathrm{ion}} \in \mathbb{R}$ and $\nu^{\mathrm{syn}} \in \mathbb{R}$ are *intrinsic* and *synaptic Nernst potentials*, respectively. Finally, the scalars $m_i^{\mathrm{ion}}, h_i^{\mathrm{ion}} \in (0, 1)$ and $s_{i,k}^{\mathrm{syn}} \in (0, 1)$ are *intrinsic* and *synaptic gating variables*, respectively. Those gating variables modulate the intensity of the currents traversing the neuronal membrane, according to the voltage-dependent dynamics

$$\tau_m^{\mathrm{ion}}(v_i) \dot{m}_i^{\mathrm{ion}} = -m_i^{\mathrm{ion}} + \sigma_m^{\mathrm{ion}}(v_i) \tag{12a}$$

$$\tau_h^{\mathrm{ion}}(v_i) \dot{h}_i^{\mathrm{ion}} = -h_i^{\mathrm{ion}} + \sigma_h^{\mathrm{ion}}(v_i) \tag{12b}$$

$$\dot{s}_{i,k}^{\mathrm{syn}} = a_{\mathrm{syn}}\sigma^{\mathrm{syn}}(v_k)(1 - s_{i,k}^{\mathrm{syn}}) - b_{\mathrm{syn}} s_{i,k}^{\mathrm{syn}} \tag{12c}$$

where $\tau(\cdot)$ are bell-shaped[1] functions of the form $\tau(v) = \underline{\tau} + (\overline{\tau} - \underline{\tau}) \exp(-(v - \zeta)^2/\chi^2)$, with $\underline{\tau}, \overline{\tau} > 0$, where $\sigma(\cdot)$ are sigmoids of the form $\sigma(v) = (1 + \exp(-(v - \rho)/\kappa))^{-1}$, and where $a_{\mathrm{syn}}, b_{\mathrm{syn}} > 0$.

For $i, k = 1, \ldots, n_v$, the biophysical neural network model given by (9)-(12) can be put in the parametric form

---

[1]The results easily extend to other forms of time-constant functions and sigmoidal functions.

(1). We illustrate this fact by means of an example in which the components of the parameter vector θ are given by the maximal conductances μ_i^{ion} and $\mu_{i,k}^{\mathrm{syn}}$, which constitute key parameters dictating the behavior of the network (see [9] for more general parametrizations).

Example 1. The Hodgkin-Huxley (HH) biophysical model first introduced by [22] contains a sodium and a potassium intrinsic current, so that $\mathcal{I} = \{\mathrm{Na}, \mathrm{K}\}$. Here we consider two HH neurons interconnected bidirectionally by means of a GABA-type inhibitory synapse (abbreviated by G), so that $\mathcal{S} = \{\mathrm{G}\}$. We parameterize the model according to

$$\theta = \mathrm{col}(\theta^{\mathrm{Na}}, \theta^{\mathrm{K}}, \theta^{\mathrm{G}}),$$

where

$$\theta^{\mathrm{Na}} = (\mu_1^{\mathrm{Na}}, \mu_2^{\mathrm{Na}})^{\mathsf{T}}, \ \theta^{\mathrm{K}} = (\mu_1^{\mathrm{K}}, \mu_2^{\mathrm{K}})^{\mathsf{T}}, \ \theta^{\mathrm{G}} = (\mu_{1,2}^{\mathrm{G}}, \mu_{2,1}^{\mathrm{G}})^{\mathsf{T}}.$$

The voltage dynamics (9) of this model can be expressed as

$$\dot{v} = \Phi_{\mathrm{Na}}(v, w^{\mathrm{Na}})\theta^{\mathrm{Na}} + \Phi_{\mathrm{K}}(v, w^{\mathrm{K}})\theta^{\mathrm{K}} + \Phi_{\mathrm{G}}(v, w^{\mathrm{G}})\theta^{\mathrm{G}} + a(v, u)$$

with the internal states

$$w^{\mathrm{Na}} = (m_1^{\mathrm{Na}}, h_1^{\mathrm{Na}}, m_2^{\mathrm{Na}}, h_2^{\mathrm{Na}})^{\mathsf{T}}, w^{\mathrm{K}} = (m_1^{\mathrm{K}}, m_2^{\mathrm{K}})^{\mathsf{T}}, \text{ and}$$
$$w^{\mathrm{G}} = (s_{1,2}^{\mathrm{G}}, s_{2,1}^{\mathrm{G}})^{\mathsf{T}},$$

the system matrices

$$\Phi_{\mathrm{Na}}(v, w^{\mathrm{Na}}) = -\mathrm{diag}\left((w_1^{\mathrm{Na}})^3 w_2^{\mathrm{Na}}(v_1 - \nu^{\mathrm{Na}}), -(w_3^{\mathrm{Na}})^3 w_4^{\mathrm{Na}}(v_2 - \nu^{\mathrm{Na}})\right),$$
$$\Phi_{\mathrm{K}}(v, w^{\mathrm{K}}) = -\mathrm{diag}\left((w_1^{\mathrm{K}})^4(v_1 - \nu^{\mathrm{K}}), (w_2^{\mathrm{K}})^4(v_2 - \nu^{\mathrm{K}})\right),$$
$$\Phi_{\mathrm{G}}(v, w^{\mathrm{G}}) = -\mathrm{diag}\left(w_1^{\mathrm{G}}(v_1 - \nu^{\mathrm{G}}), w_2^{\mathrm{G}}(v_2 - \nu^{\mathrm{G}})\right),$$

and the known vector

$$a(v, u) = \begin{pmatrix} -\mu_1^{\mathrm{leak}}(v_1 - \nu^{\mathrm{leak}}) + u_1 \\ -\mu_2^{\mathrm{leak}}(v_2 - \nu^{\mathrm{leak}}) + u_2 \end{pmatrix}.$$

With the choice of internal states above, the internal dynamics (12) can be expressed as (1b), and hence the model is in the form (1) (for clarity, we have replaced indexing by $j = 1, \ldots, m$ with indexing by ion $\in \mathcal{I}$ and syn $\in \mathcal{S}$).

It can be shown that any conductance-based model (9)-(12) satisfies Assumptions 1 and 2; in particular the internal dynamics (12) are exponentially contracting (see [9] for rigorous proofs). Hence, (6) can be used to estimate the maximal conductances of any conductance-based model.

Example 1 shows an interesting feature of biophysical neural network models: the matrices $\Phi_j(v, \cdot)$ are diagonal, with the ith diagonal elements depending only on the voltage and internal states of the ith neuron. This feature allows for distributing parameter estimation over the neurons in the network, in addition to distributing it over individual membrane currents. This is illustrated next:

Example 2. Consider the two-neuron HH network of Example 1. With the distributed observer (6), the sodium maximum conductance update rule is given by

$$\dot{\hat{\theta}}^{\mathrm{Na}} = \gamma_{\mathrm{Na}} P_{\mathrm{Na}} \Psi_{\mathrm{Na}}(v - \hat{v})$$
$$\dot{\Psi}_{\mathrm{Na}} = -\gamma_{\mathrm{Na}} \Psi_{\mathrm{Na}} + \Phi_{\mathrm{Na}}(v, \hat{w}^{\mathrm{Na}}) \quad (13)$$
$$\dot{P}_{\mathrm{Na}} = \alpha_{\mathrm{Na}} P_{\mathrm{Na}} - \alpha_{\mathrm{Na}} P_{\mathrm{Na}} \Psi_{\mathrm{Na}} \Psi_{\mathrm{Na}}^{\mathsf{T}} P_{\mathrm{Na}}$$
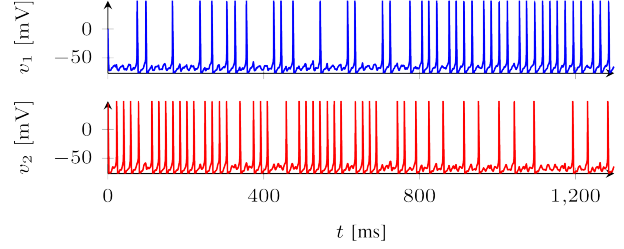

Fig. 1. Voltages of the two-neuron HH network in Section IV-A.

For $\Psi_{\mathrm{Na}}(0) = 0$ and diagonal $P_{\mathrm{Na}}(0) \succ 0$, the matrices $\Psi_{\mathrm{Na}}(t)$, and $P_{\mathrm{Na}}(t)$ are diagonal for all $t \geq 0$, and (13) becomes the network-distributed update rule

$$\dot{\hat{\mu}}_i^{\mathrm{Na}} = \gamma_{\mathrm{Na}} p_i^{\mathrm{Na}} \psi_i^{\mathrm{Na}}(v_i - \hat{v}_i)$$
$$\dot{\psi}_i^{\mathrm{Na}} = -\gamma_{\mathrm{Na}} \psi_i^{\mathrm{Na}} - (\hat{m}_i^{\mathrm{Na}})^3 \hat{h}_i^{\mathrm{Na}}(v_i - \nu^{\mathrm{Na}})$$
$$\dot{p}_i^{\mathrm{Na}} = \alpha_{\mathrm{Na}} p_i^{\mathrm{Na}} - \alpha_{\mathrm{Na}} (p_i^{\mathrm{Na}})^2 (\psi_i^{\mathrm{Na}})^2$$

for $i = 1, 2$. The same simplification applies to the update rules of $\hat{\theta}^{\mathrm{K}}$ and $\hat{\theta}^{\mathrm{G}}$. Hence the observer (6) becomes fully distributed with respect to neurons in the network and individual neuronal membrane currents in each neuron.

A. Numerical simulation

We finish this section by simulating online parameter estimation of the two-neuron HH network of Example 1. By doing so, we also show that the distributed adaptive observer estimates are able to approximately track slowly time-varying parameters. We consider a configuration in which the two neurons in the true system have identical parameters, except for their synaptic maximal conductances and control inputs. The parameters to be estimated are given by $\mu_1^{\mathrm{Na}} = \mu_2^{\mathrm{Na}} = 120$, $\mu_1^{\mathrm{K}} = \mu_2^{\mathrm{K}} = 36$,

$$\mu_{1,2}^{\mathrm{G}} = 0.75 - 0.4(1 + e^{-(t-750)/100})^{-1}, \quad \text{and}$$
$$\mu_{2,1}^{\mathrm{G}} = 0.25 + 0.4(1 + e^{-(t-750)/100})^{-1},$$

while the remaining (non-adaptive) parameters are described in Appendix B (which also describes initial conditions). Figure 1 shows the resulting voltages of the network.

We compare the performance of the distributed observer (6), implemented according to Example 2, with the non-distributed version (4). To check the robustness of the observers to measurement noise, both observers are simulated with the measured $v(t)$ replaced by $v(t) + e(t)$, where $e(t)$ is white Gaussian measurement noise chosen so that the signal-to-noise between $v(t)$ and $e(t)$ is of 40 dB.

The non-distributed observer is simulated with $\alpha = 0.15$ and $\gamma = 2$. The distributed observer is simulated with the two different gain sets below:

Gains	γ_0	γ_{Na}	γ_{K}	γ_{G}	α_{Na}	α_{K}	α_{G}
(A)	2	2	2	2	0.15	0.15	0.15
(B)	2	2	2	0.8	0.15	0.15	0.03

The behaviour of the estimates $\hat{\mu}_i^{\mathrm{Na}}$ and $\hat{\mu}_{i,j}^{\mathrm{G}}$ is shown in Figures 2 and 3 (the behaviour of $\hat{\mu}_i^{\mathrm{K}}$ is qualitatively

similar to that of $\hat{\mu}_i^{\text{Na}}$ and is omitted). The estimates of the non-distributed observer converge very rapidly to a neighborhood of the true parameter values (after only a few spikes, compare with Figure 1). This indicates a rather aggressive choice of gains, which can also be inferred from the perturbation in $\hat{\mu}_{1,2}^{\text{G}}$ seen just before 1200 ms (this perturbation occurs due to a momentary decrease in the excitation provided by v_2, which increases sensitivity to noise). For the distributed observer with gain set (A), which mimics the aggressive gains of the non-distributed observer, it can be seen that $\hat{\mu}_i^{\text{Na}}$ converges with a small smooth transient, but $\hat{\mu}_{i,j}^{\text{G}}$ converges with a large transient with rapid oscillations. A less aggressive choice of gains could mitigate the undesired transients and decrease the sensitivity to noise, at the cost of a slower convergence rate (for a detailed analysis, see [9]). The distributed observer allows being strategic with respect to the choice of gains: the gain set (B) remedies the undesirable transients in $\hat{\mu}_{i,j}^{\text{G}}$ with a less aggressive choice of gains for α_{G} and γ_{G}, while keeping the other gains from the set (A). As a result, the good convergence properties of $\hat{\mu}_i^{\text{Na}}$ (and of $\hat{\mu}_i^{\text{K}}$) are preserved.

V. DISCUSSION

We have shown that the distributed adaptive observer proposed in this paper is well-suited for estimating networks of biophysical neuronal models in real-time. We envision two applications of this work. First, *experiment design in electrophysiology*: neuromodulators are capable of changing the maximal conductances of living neurons, and this process can be studied *in vitro* using a closed-loop technique called dynamic clamp [23]. The adaptive observer provides a means to track neuromodulatory changes in real-time, which allows incorporating this information in experiment design. The second application is the *detection of qualitative changes in excitable regimes* [24]: epileptic seizures are associated to hyperexcitable neuronal states; given intracellular voltage data from a large network of neurons, the distributed adaptive observer provides a tractable means to keep track of the conductances responsible for modulating cellular excitability. This information is valuable for predicting the onset of switches in excitable regime associated to seizures.

In this paper, we have assumed the internal dynamics (1b) is known. To address the case of an uncertain or unknown internal dynamics, a distributed version of the locally convergent nonlinearly parameterized adaptive observer in [9], designed to estimate the internal dynamics, can be proposed. This will be the subject of future work.

APPENDIX

A. Proof of Theorem 1

The proof is inspired by the results of [20], [21] who provided sufficient conditions for exponential stability of nonlinear systems without requiring a Lyapunov function with a negative semidefinite derivative. The proof also uses the idea of *virtual system* from contraction theory, see [14]. The idea is to construct a dynamical system (the virtual system) whose trajectories contain the trajectories of the true system (1) as well as those of the system (6a)-(6c), and then show that the virtual system is exponentially contracting on a positively invariant set. Contraction of the virtual system's trajectories then imply exponential convergence of $(\hat{v}, \hat{w}, \hat{\theta})$ to (v, w, θ). The virtual system is given by

$$
\begin{aligned}
\dot{s} &= g(v, s) \\
\dot{r} &= \tilde{f}(t, s, \eta) + a(v, s, u) + (\gamma_0 I + \Psi^{\mathsf{T}} \bar{P} \Gamma \Psi)(v - r) \\
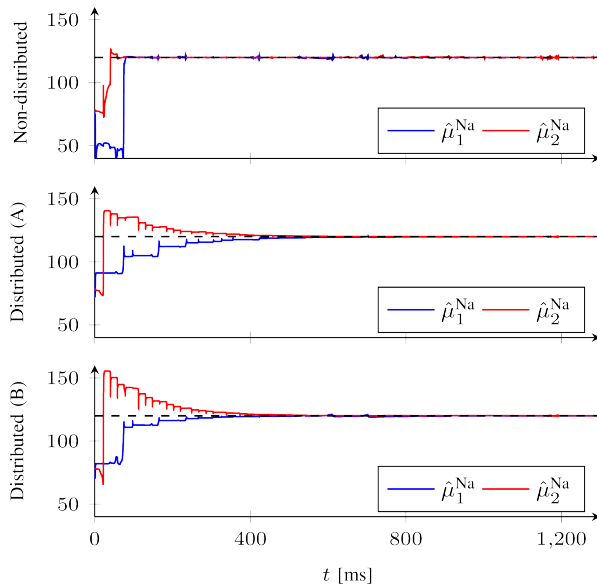\dot{\eta} &= \bar{P} \Gamma \Psi(v - r)
\end{aligned}
$$

(14)

Fig. 2. Estimates of $\mu_1^{\text{Na}} = \mu_2^{\text{Na}}$ (dashed line) for the non-distributed and distributed adaptive observers with gain sets (A) and (B).
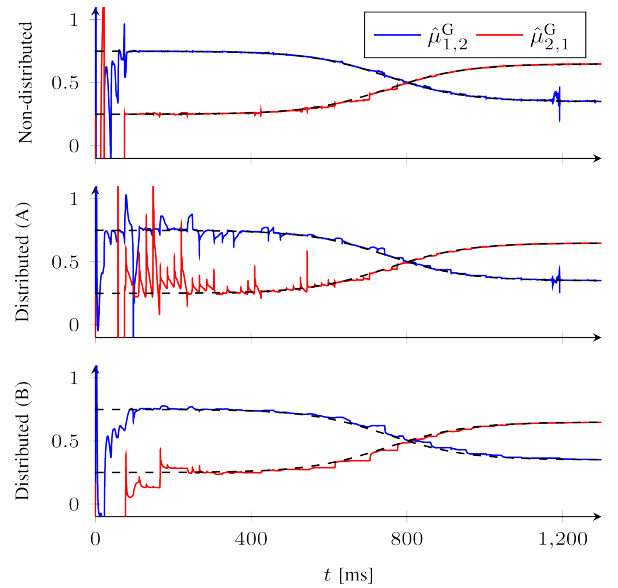

Fig. 3. Estimates of $\mu_{1,2}^{\text{G}}$ and $\mu_{2,1}^{\text{G}}$ (dashed lines) for the non-distributed and distributed adaptive observers with gain sets (A) and (B).

where

$$\tilde{f}(t, s, \eta) = \Phi^{\mathsf{T}}(v, \hat{w}, u)\eta + (\Phi^{\mathsf{T}}(v, s, u) - \Phi^{\mathsf{T}}(v, \hat{w}, u))\theta$$

and where Ψ comes from (6d), which can be rewritten as

$$\dot{\Psi} = -\Gamma\Psi + \Phi(v, \hat{w}, u), \quad \Psi(0) = 0 .$$

Here, s, r and η are the virtual system states, as opposed to $v(t)$, $\hat{w}(t)$, and $u(t)$, which are treated as time-varying signals. Notice that any solutions $\mathrm{col}(w, v, \theta)$ of (1) and $\mathrm{col}(\hat{w}, \hat{v}, \hat{\theta})$ of (6) are particular solutions of the virtual system. The first two equations of the virtual system have been written in a different order with respect to those of the true system and the adaptive observer, to simplify the notation in the remainder of the proof.

To proceed, we derive the *differential* virtual system, given by

$$\begin{pmatrix} \dot{\delta s} \\ \dot{\delta r} \\ \dot{\delta \eta} \end{pmatrix} = \underbrace{\begin{bmatrix} J_{1,1} & 0 \\ J_{2,1} & J_{2,2} \end{bmatrix}}_{J} \begin{pmatrix} \delta s \\ \delta r \\ \delta \eta \end{pmatrix} \tag{15}$$

where the Jacobian J has the components

$$J_{1,1} = \mathrm{diag}\{\partial_{s^1} g_1(v, s^1), \ldots, \partial_{s^m} g_m(v, s^m)\} \tag{16a}$$

$$J_{2,1} = \begin{bmatrix} \partial_s(\Phi^{\mathsf{T}}(v, s, u)\theta + a(v, s, u)) \\ 0 \end{bmatrix}, \tag{16b}$$

$$J_{2,2} = \begin{bmatrix} -(\gamma_0 I + \Psi^{\mathsf{T}}\bar{P}\Gamma\Psi) & \Phi^{\mathsf{T}}(v, \hat{w}, u) \\ -\bar{P}\Gamma\Psi & 0 \end{bmatrix} \tag{16c}$$

Following the ideas of [14], [21] we show that the dynamics (14) are exponentially contracting on the invariant set $W \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_\theta}$ using an infinitesimal coordinate transformation (we write $W := \prod_j W_j$). We define

$$\delta z = \mathrm{col}\left(\mu\bar{\Theta}(t)\delta s, \quad \Theta_0(t) \begin{pmatrix} \delta r \\ \delta \eta \end{pmatrix} \right), \tag{17}$$

where $\mu > 0$ is an arbitrary constant, $\bar{\Theta}$ is given by

$$\bar{\Theta}(t) := \mathrm{diag}\{\Theta_1, \ldots, \Theta_m\},$$

with the $\Theta_j(t)$ from Assumption 2, and $\Theta_0(t)$ is given by

$$\Theta_0(t) = \begin{bmatrix} I & -\Psi^{\mathsf{T}} \\ 0 & \bar{R}^{\frac{1}{2}} \end{bmatrix},$$

with

$$\bar{R}(t) = \mathrm{diag}\{R_1(t), \ldots, R_m(t)\} := \bar{P}(t)^{-1}.$$

Under Assumptions 1 and 2, $\Psi(t)$ is bounded for all $t \geq 0$, which implies $\bar{R}(t)$ is a well-defined upper-bounded inverse of $\bar{P}(t)$, see [9, Lemma 3]. Furthermore, part *(i)* of Assumption 3 ensures that

$$R_j(t) \succeq \delta_j \alpha_j e^{-2\alpha_j T} I, \quad t \geq T \tag{18}$$

and $R_j(t) \succeq e^{-\alpha_j T} R(0) \succ 0$ for $t \in [0, T)$, see [10, Lemma 1]. As a consequence, J in (15) is bounded on $W \times \mathbb{R}^{n_v} \times$

$\mathbb{R}^{n_\theta}$ for all $t \geq 0$, and $\Theta_0(t)$ is uniformly invertible and bounded, with

$$\Theta_0^{-1}(t) = \begin{bmatrix} I & \Psi^{\mathsf{T}}\bar{P}^{\frac{1}{2}} \\ 0 & \bar{P}^{\frac{1}{2}} \end{bmatrix} \tag{19}$$

Following [21], we prove contraction of the virtual system (14) by showing that $\|\delta z(t)\| \to 0$ as $t \to 0$. For that purpose, we now derive the transformed differential system. Taking the derivative of (17) and using the inverse coordinate transformation, we obtain

$$\dot{\delta z} = \underbrace{\begin{bmatrix} \bar{F} & 0 \\ \mu^{-1}J_{2,1}\bar{\Theta}^{-1} & F_0 \end{bmatrix}}_{F(t, s)} \delta z \tag{20}$$

where $\bar{F} = \mathrm{diag}\{F_1, \ldots, F_m\}$, with F_j given by (2), and where F_0 is given by

$$F_0 = (\dot{\Theta}_0 + \Theta_0 J_{2,2})\Theta_0^{-1} \tag{21}$$

(here, we have used $\Theta_0 J_{2,1} = J_{2,1}$).

The dynamics of the distance $\|\delta z\|$ is governed by

$$\frac{d}{dt}\|\delta z(t)\|^2 = \delta z^{\mathsf{T}}(F + F^{\mathsf{T}})\delta z$$

and hence we seek an upper bound for $F + F^{\mathsf{T}}$. To compute $F_0 + F_0^{\mathsf{T}}$, it is worth noticing that since $d(P_j^{-1})/dt = -P_j^{-1}\dot{P}_j P_j^{-1}$, the diagonal elements of $\bar{R}$ obey

$$\dot{R}_j = -\alpha_j R_j + \alpha_j \Psi_j \Psi_j^{\mathsf{T}}$$

from where we derive the relation

$$\tfrac{d}{dt}(\bar{R}^{\frac{1}{2}})\bar{P}^{\frac{1}{2}} + \bar{P}^{\frac{1}{2}}\tfrac{d}{dt}(\bar{R}^{\frac{1}{2}}) = -A + \bar{P}^{\frac{1}{2}}AD\bar{P}^{\frac{1}{2}} \tag{22}$$

It follows from (16c), (19), (21) and (22) that

$$F_0 + F_0^{\mathsf{T}} = \begin{bmatrix} -2\gamma_0 I & -\gamma_0\Psi^{\mathsf{T}}\bar{P}^{\frac{1}{2}} \\ -\gamma_0\bar{P}^{\frac{1}{2}}\Psi & -A + \bar{P}^{\frac{1}{2}}(AD - \Gamma\Psi\Psi^{\mathsf{T}} - \Psi\Psi^{\mathsf{T}}\Gamma)\bar{P}^{\frac{1}{2}} \end{bmatrix}$$

Now, using Assumption 2 and F in (20), we have that

$$F + F^{\mathsf{T}} \leq \varepsilon I + M + N$$

where

$$M = \begin{bmatrix} -\varepsilon I & \mu^{-1}\bar{\Theta}^{-\mathsf{T}}J_{2,1}^{\mathsf{T}} \\ \mu^{-1}J_{2,1}\bar{\Theta}^{-1} & -\varepsilon I + \gamma_0 M_{2,2} \end{bmatrix}$$

$$M_{2,2} = \begin{bmatrix} -I & -\Psi^{\mathsf{T}}\bar{P}^{\frac{1}{2}} \\ -\bar{P}^{\frac{1}{2}}\Psi & -\bar{P}^{\frac{1}{2}}\Psi\Psi^{\mathsf{T}}\bar{P}^{\frac{1}{2}} \end{bmatrix}$$

with $\varepsilon > 0$ an arbitrary constant, and

$$N = \begin{bmatrix} -\min_j\{\lambda_j\}I & 0 & 0 \\ 0 & -\gamma_0 I & 0 \\ 0 & 0 & -A + \bar{P}^{\frac{1}{2}}Q\bar{P}^{\frac{1}{2}} \end{bmatrix} \tag{23}$$

with

$$Q = AD + \gamma_0\Psi\Psi^{\mathsf{T}} - \Gamma\Psi\Psi^{\mathsf{T}} - \Psi\Psi^{\mathsf{T}}\Gamma$$

In what follows, we choose ε such that

$$0 < \varepsilon < \min\{\gamma_0, \lambda_1, \ldots, \lambda_m, \beta\} \tag{24}$$

where $\beta > 0$ is taken from part (*ii*) of Assumption 3.

Using Schur's complement, we see that $M_{2,2} \preceq 0$. Then, since $\mu > 0$ is arbitrary and $\bar{\Theta}$ and $J_{2,1}$ are bounded on on $W \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_\theta}$, uniformly in $t \geq 0$, we can use Schur's complement again to show that $M \preceq 0$ for a sufficiently large choice of $\mu(\varepsilon) > 0$. Hence for that choice it follows that

$$\frac{d}{dt}\|\delta z(t)\|^2 \leq \delta z(t)^\mathsf{T}(N(t) + \varepsilon I)\delta z(t)$$
$$\leq (\lambda_{\max}(N(t)) + \varepsilon)\|\delta z(t)\|^2$$

Solving for this inequality at every initial time $t \geq T$ yields

$$\|\delta z(t+T)\| \leq \|\delta z(t)\| \exp\left(\frac{1}{2}\int_t^{t+T}(\lambda_{\max}(N(\tau)) + \varepsilon)d\tau\right)$$

Now since A, $\bar{P}$ and Q are all symmetric, [25, Theorem 4.5.9] and [25, Corollary 4.3.15] can be used to show that

$$\lambda_{\max}(-A + \bar{P}^{\frac{1}{2}}Q\bar{P}^{\frac{1}{2}}) \leq \lambda_{\max}(-A) + \lambda_{\max}(\bar{P})\lambda_{\max}(Q)$$
$$\leq -\underline{\alpha} + \left(\min_j\left(\frac{\delta_j\alpha_j}{e^{2\alpha_j T}}\right)\right)^{-1}\lambda_{\max}(Q)$$

for all $t \geq T$, where $\underline{\alpha} = \min\{\alpha_1, \ldots, \alpha_m\}$, and where we have used $\lambda_{\max}(\bar{P}) = (\lambda_{\min}(\bar{R}))^{-1}$ as well as (18).

It finally follows from part (*ii*) of Assumption 3, and the form of N in (23), that for all $t \geq T$, there is a T such that

$$\int_t^{t+T}(\lambda_{\max}(N(\tau)) + \varepsilon)d\tau \leq (-\min\{\gamma_0, \lambda_j, \beta\} + \varepsilon)T$$

Given our choice of ε in (24), the right hand side above is strictly negative. Hence, similarly to [21, Corollary 3.1], we conclude that $\|\delta z(t)\| \to 0$ as $t \to 0$ and that the virtual system (14) is exponentially contracting on the invariant set $W \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_\theta}$. As a result, we have $\mathrm{col}(\hat{w}, \hat{v}, \hat{\theta}) \to \mathrm{col}(w, v, \theta)$ as $t \to \infty$ and the result is proven.

B. Simulation parameters

The parameters in $\Phi(v, w)$ and $a(v, u)$ are given by $\mu_1^{\text{leak}} = \mu_2^{\text{leak}} = 0.3$, $\nu^{\text{Na}} = 55$, $\nu^{\text{K}} = -77$, $\nu^{\text{leak}} = -54.4$, and $\nu^{\text{G}} = -80$; the parameters of the intrinsic gating variable dynamics (12a)-(12b) are given in the table in [9, Appendix C.1]; the parameters of the synaptic gating variable dynamics (12c) are given by $a_{\text{syn}} = 2$, $b_{\text{syn}} = 0.1$, $\rho_{\text{syn}} = -45$ and $\kappa_{\text{syn}} = 2$. The two neurons are excited with the control inputs $u_1(t) = 2 + \sin(2\pi t/10) + \sin(2\pi t/7) + \sin(2\pi t/4)$ and $u_2(t) = 1 + 2\sin(2\pi t/9) + \sin(2\pi t/5)$. All simulations are performed using the Euler-Maruyama method with $dt = 10^{-4}$ ms. The initial conditions are given by $v_1(0) = \hat{v}_1(0) = 0$, $v_2(0) = \hat{v}_2(0) = -60$, $w^{\text{Na}}(0) = (0, 0.5, 0, 0.5)^\mathsf{T}$, $w^{\text{K}}(0) = (0, 0.5)^\mathsf{T}$, $w^{\text{G}} = (0, 0.5)^\mathsf{T}$, $\hat{w}^{\text{Na}}(0) = (0.5, 0, 0.5, 0)^\mathsf{T}$, $\hat{w}^{\text{K}}(0) = (0.5, 0)^\mathsf{T}$, $\hat{w}^{\text{G}} = (0.5, 0)^\mathsf{T}$, $\hat{\theta}^{\text{Na}} = \hat{\theta}^{\text{K}} = (78, 78)^\mathsf{T}$, $\hat{\theta}^{\text{G}} = (0, 0)^\mathsf{T}$.

REFERENCES

[1] T. Knöpfel and C. Song, "Optical voltage imaging in neurons: moving from technology development to practical tool," *Nature Reviews Neuroscience*, vol. 20, no. 12, pp. 719–727, Dec. 2019.

[2] E. Tang and D. S. Bassett, "*Colloquium*: Control of dynamics in brain networks," *Reviews of Modern Physics*, vol. 90, no. 3, Aug. 2018.

[3] L. Ribar and R. Sepulchre, "Neuromodulation of Neuromorphic Circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 8, pp. 3028–3040, Aug. 2019.

[4] Q. J. M. Huys, M. B. Ahrens, and L. Paninski, "Efficient Estimation of Detailed Single-Neuron Models," *Journal of Neurophysiology*, vol. 96, no. 2, pp. 872–890, Aug. 2006.

[5] S. Druckmann, Y. Banitt, A. Gidon, F. Schürmann, H. Markram, and I. Segev, "A Novel Multiple Objective Optimization Framework for Constraining Conductance-Based Neuron Models by Experimental Data," *Frontiers in Neuroscience*, vol. 1, no. 1, pp. 7–18, Oct. 2007.

[6] C. D. Meliza, M. Kostuk, H. Huang, A. Nogaret, D. Margoliash, and H. D. I. Abarbanel, "Estimating parameters and predicting membrane voltages with conductance-based neuron models," *Biological Cybernetics*, vol. 108, no. 4, pp. 495–516, Aug. 2014.

[7] T. B. Burghi, M. Schoukens, and R. Sepulchre, "Feedback identification of conductance-based models," *Automatica*, vol. 123, p. 109297, Jan. 2021.

[8] E. Sorrell, M. E. Rule, and T. O'Leary, "Brain–Machine Interfaces: Closed-Loop Control in an Adaptive System," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 167–189, May 2021.

[9] T. B. Burghi and R. Sepulchre, "Adaptive observers for biophysical neuronal circuits," *Submitted to IEEE Transactions on Automatic Control.*, Oct. 2021, under review, preprint available at https://arxiv.org/abs/2111.02176.

[10] Q. Zhang and A. Clavel, "Adaptive observer with exponential forgetting factor for linear time varying systems," in *40th IEEE Conference on Decision and Control*, Orlando, FL, USA, Dec. 2001, pp. 3886–3891.

[11] M. Farza, M. M'Saad, T. Maatoug, and M. Kamoun, "Adaptive observers for nonlinearly parameterized class of nonlinear systems," *Automatica*, vol. 45, no. 10, pp. 2292–2299, Oct. 2009.

[12] K. J. Åström and B. Wittenmark, *Adaptive Control*, 2nd ed. Mineola, NY: Dover Publications, Jan. 2008.

[13] A. Vahidi, A. Stefanopoulou, and H. Peng, "Recursive least squares with forgetting for online estimation of vehicle mass and road grade: theory and experiments," *Vehicle System Dynamics*, vol. 43, no. 1, pp. 31–55, Jan. 2005.

[14] W. Lohmiller and J.-J. E. Slotine, "On Contraction Analysis for Nonlinear Systems," *Automatica*, vol. 34, no. 6, pp. 683–696, Jun. 1998.

[15] E. M. Izhikevich, *Dynamical Systems in Neuroscience*. Cambridge, MA: MIT Press, 2007.

[16] S. Bittanti, P. Bolzern, and M. Campi, "Convergence and exponential convergence of identification algorithms with directional forgetting factor," *Automatica*, vol. 26, no. 5, pp. 929–932, Sep. 1990.

[17] J. E. Parkum, N. K. Poulsen, and J. Holst, "Recursive forgetting algorithms," *International Journal of Control*, vol. 55, no. 1, pp. 109–128, Jan. 1992.

[18] F. Fraccaroli, A. Peruffo, and M. Zorzi, "A new recursive least squares method with multiple forgetting schemes," in *54th IEEE Conference on Decision and Control*, Osaka, Japan, Dec. 2015, pp. 3367–3372.

[19] S. Boyd, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[20] D. Aeyels and J. Peuteman, "On exponential stability of nonlinear time-varying di!erential equations," *Automatica*, vol. 35, no. 6, pp. 1091–1100, Jun. 1999.

[21] J. Jouffroy, "A relaxed criterion for contraction theory: Application to an underwater vehicle observer," in *2003 European Control Conference (ECC)*, Sep. 2003, pp. 2999–3004.

[22] 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*, vol. 117, no. 4, pp. 500–544, 1952.

[23] A. A. Sharp, M. B. O'Neil, L. F. Abbott, and E. Marder, "Dynamic clamp: computer-generated conductances in real neurons," *Journal of Neurophysiology*, vol. 69, no. 3, pp. 992–995, Mar. 1993.

[24] Y. Tang, A. Franci, and R. Postoyan, "On-line detection of qualitative dynamical changes in nonlinear systems: The resting-oscillation case," *Automatica*, vol. 100, pp. 17–28, Feb. 2019.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, UK: Cambridge University Press, 2012.