Black-box model reduction of the C. Elegans nervous system*

Ruxandra Bărbolescu¹ and L. Miguel Silveira¹

Abstract—In recent years, modeling neurons and neuronal collections with high accuracy have become central issues of neuroscience. The development of efficient algorithms for their simulation as well as the increase in computational power and parallelization need to keep up with the quantity and complexity of novel recordings and reconstructions reported by the experimental neuroscientists. The extraction of low-order equivalents that capture the essential aspects of the high-accuracy models is an essential part of the simulation process. The complexity of these models require the use of black-box data-oriented reduction approaches. We create a detailed model of the nervous system of a very known organism, C. Elegans, and show that it can be reduced using a modified data-driven model reduction method up to the order of 4 with very little loss in accuracy. The reduced model is able to predict the behaviour of the original for time ranges beyond the data used for the reduction.

I. INTRODUCTION

Caenorhabditis Elegans (C. Elegans) is a transparent nematode of about 1 mm in length. Extensive research has been done for a deep understanding of the behavioral and structural biology of this worm. This almost unprecedented scientific interest in C. Elegans is built upon the fact that the animal is able to solve basic problems such as feeding, mate-finding and predator avoidance with a rather small nervous system, of less than 1000 cells across all sexes and around 15000 synapses [14]. Another reason is that the nervous system of nematodes is almost invariant across individuals (every individual possesses the same number of neurons, they occupy fixed positions in the organism and the synaptic connection structure is preserved). The idea is to extend the knowledge, methods and instruments developed for C. Elegans to more complex nervous systems.

The pursuit of a deep, thorough understanding of C. Elegans biology has resulted in comprehensive databases of genetics and genomics [8], [1], electron micrographs and associated data [16], online books [9] and atlases [15].

The connectome (also called "wiring diagram" or "connectivity graph") is a map of the neural connections in the brain, described as a neuronal network or graph (where the nodes are the neurons and the edges represent the synapses). For C. Elegans there have been published complete connectomes for the adult hermaphodite containing 302 neurons [27], partial connectomes for the adult male – the posterior nervous system with 144 neurons [19] and recently the complete connectomes for the two adult sexes – having 385 neurons for the male [14], but for the latter the respective 3D reconstructions are not yet published [17].

While reconstructing the wiring diagrams of the nervous system is an important step in its exploration, solely knowing the static connections in the brain is not enough to connect it to its function. The inherent complexity of the biological mechanisms calls for a multi-scale and multi-algorithm approach, based on digitally reconstructed neurons in 3D, which are available in open-source databases [4], [6], [5]. Collaborative solutions such as [2] will provide robust and flexible environments for the simulation of generic complex biological systems. OpenWorm is trying to build a computational model of C. Elegans and has made available the full connectome of C. Elegans hermaphrodite in NeuroML format. NeuroML is an XML-based specification language, therefore platform-independent; the program that is traditionally associated with NeuroML files is neuroConstruct. The open-source nature of the software and its integration with dedicated neuronal simulators opens the door for different specializations to investigate the dynamics of the C. Elegans nervous system.

A second-layer containing the biophysical processes of neural responses and interactions needs to be added to the static first-layer (the connectome) to make it into a dynamical model (the interactome) [22]. The interactome in [22] is an interactive network, but the underlying neurons are compact models with no geometrical distribution whereas the neural responses are modeled with graded potentials.

With an interest in computational aspects of modeling and simulation and model order reduction, our approach is slightly different. We start from a very detailed model, both geometrically and physically, and we are interested in determining to which extent the model description can be compressed without loss of accuracy. Model reduction techniques are used in neuroscience for large-scale networks to extract coherent features which are not apparent at the level of individual neurons [20]. The intrinsic nonlinearity of the nervous system requires the development of new model reduction approaches, more specifically black-box techniques, which don’t require prior knowledge of the original system’s structure and equations. These are data-driven methods that only look at the system’s behaviour towards a particular input signal and attempt to preserve the input-output behaviour of the original model. Different versions of Dynamic mode decomposition (DMD) [11], Proper Orthogonal Decomposition (POD) [21] and Discrete Empirical Interpolation (DEIM) [24] have been applied to neuronal networks with various levels of morphologic accuracy. Our goal is to use the detailed C. Elegans model as a benchmark.

*This work was not supported by any organization
¹Ruxandra Bărbolescu and Luis Miguel Silveira are with INESC-ID/IST Tecnico Lisboa, Universidade de Lisboa, Rua Alves Redol 9, 1000 Lisboa, Portugal, ruxi@inesc-id.pt, lms@inesc-id.pt

978-1-7281-1178-0/21/$31.00 ©2021 IEEE 4174
for testing and developing data-driven reduction methods for morphologically realistic models and to possibly infer the findings for larger-scale networks. We use a version of POD to reduce the original model in a black-box setting, from snapshots of the states’ evolution over time and we approximate the low-order model with a linear/quadratic mathematical system. The reduced model of order 4 extracted piecewise is able to accurately predict the behaviour of the original system for time spans beyond the range used for the reduction. The ultimate goal is to extract the connectivity, coefficients and/or parameters of a smaller network that corresponds to the reduced model extracted from the snapshots of the original model.

II. THE NETWORK

A. Model Definition

We define a complete model of the C. elegans hermaphrodite, built upon detailed 3D reconstructions of neurons. The model is described in Python and NEURON [13], one of the traditional neural simulators that achieves high efficiency in simulating large-scale networks while having support for biologically realistic multicompartmental models of neurons due to its built-in neuronal-oriented algorithms and methods.

The network contains all of the 302 neurons of the C. elegans hermaphrodite, the biophysical properties and the connectivity data from [27]. A 3D reproduction is extracted from NEURON in Fig. 1. With the NeuroML reconstructions from OpenWorm, we attempted to import the descriptions into NEURON with pyNeuroML. PyNeuroML is a Python package unifying scripts for reading, writing, simulating and analysing NeuroML / LEMS models. LEMS (Low Entropy Model Specification language) [3] is an XML-based language used for specifying generic models of hybrid dynamical systems. We start from the NeuroML files associated with the neurons and network structure of C. elegans, representing the connectome and add biophysical properties thus creating the associated interactome. The NeuroML files are further wrapped inside LEMS files, so that they can be manipulated by pyNeuroML.

However, pyNeuroML only handles LEMS files describing single-compartment models of cells [18], [28]. We created our workaround for this issue: we use pyNeuroML to generate the Python file using the LEMS definitions for cells with only one segment, we then generate the NEURON .hoc files directly from the NeuroML files with pyNeuroML and finally we modify the Python file to use the new .hoc files, which contain all segments. The Python file is then executed in NEURON and the outputs are exported automatically and processed in an external software (Matlab). In this setting, the network can be easily regenerated with modified neuronal and/or synaptic parameters.

Despite the fact that our approach has its own starting point and conception, it turned out being relatively similar with the one used in the c302 modeling framework for C. elegans, presented in [18]. One of the features of the c302 framework is that it allows the creation of the model for the C. elegans nervous system using LEMS and the pyNeuroML package [7]. However, the two approaches differ from a programming standpoint and were independently developed.

B. Validation against Scenarios

We validate the correct definition of the C. elegans network against four scenarios, described in [22].

1) Forward Crawling Motion

We apply constant stimulus on the touch sensitive sensory neurons and the interneurons known to be part of the forward movement circuit (1.4 nA for PLM neurons, 2.3 nA for AVB interneurons) and we check the activity of the motor neurons associated with forward locomotion.

We observe activity in 170 (56%) of the total number of neurons in the network, out of which 81 are motor neurons. The most responsive motor neurons are VB+ and DB+ (Ventralic and Dorsal type B) and AS07. The interneurons with the largest activity are AVB, DVA, LUA, PVC and PVR, which is in accordance with the results reported in the literature. As in [22], we also obtained strong responses from LUA and PVR neurons.

2) Ablation of AVB interneurons + Forward Crawling Motion

It is known that removing the AVB interneurons impedes forward locomotion. In this scenario we remove AVB from the network and repeat the Forward Crawling Motion (stimulus only on the PLM neurons). We should see much less activity in most neurons.

We indeed confirm that the number of neurons with any activity dropped from 170 (56%) to 33 (10%). We observe strong responses in only 16 neurons, out of which only 3 are motor neurons.

3) Ablation of AVA interneurons + Forward Crawling Motion

The elimination of the AVA interneurons does not impact forward motion. We therefore repeat the Forward Crawling Motion simulation with the AVA neurons previously removed from the network. We should see no significant differences in the neurons activity compared to the first scenario.

The comparison shows that the Forward Crawling
III. MODEL REDUCTION

We simulated in NEURON the model in the scenario corresponding to the Forward Crawling Motion for 200 ms, with $n = 16$ outputs and $m = 2001$ time moments, generating a matrix of snapshots $A \in \mathbb{R}^{16 \times 2001}$. The POD method is based on the decomposition of the snapshots matrix into modes (a set of basis functions that spans the collection of samples) and ranking them according to their energy content, in order to preserve only the most important ones, which bring new information to the model [29]. One of the optimal methods to find the POD modes is Singular Value Decomposition (SVD). The singular values give information on the linearly independent character of the samples matrix. SVD allows the identification and elimination of the "almost singular" part of the matrix, the lines that are almost linearly dependent.

The samples matrix $A \in \mathbb{R}^{n \times m} = \{a_k, k = 1, m\}$ is decomposed into singular values:

$$A = U \Sigma V^T,$$

where $\Sigma = \text{diag}(\sigma_i)$ is a rectangular diagonal matrix with the singular values on the diagonal and the columns of $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ are the corresponding left and right singular vectors. The POD modes $\phi_j$ are the columns of $U$ and the relative energy of the POD mode $i$ is given by $\sigma_i^2 / \sum_{j=1}^{n} \sigma_j^2$. The decomposition is then truncated by keeping only the $r < n < m$ first most important POD modes, so that $\phi_{\text{red}} \in \mathbb{R}^{n \times r}$. Note that in contrast to many industry applications where $n \gg m$, in neuroscience many recordings can be extracted during a simulation [20], so also in our case the number of snapshots is much greater than the number of states.

The snapshots can be represented as a linear combination of the truncated POD modes $\phi_j, j = 1, r$ with their corresponding POD coefficients $\alpha_j$. These are computed by projecting the snapshots matrix onto the POD modes

$$\alpha_j = \phi_j^T a_k, \text{ in matrix form } \alpha = \phi_{\text{red}}^T A. \quad (1)$$

The $r$-order approximation $a_{\text{red},k}$ of $a_k$ is given by

$$a_{\text{red},k} = \sum_{j=1}^{r} \alpha_{j,k} \phi_j, \text{ in matrix form } A_{\text{red}} = \phi_{\text{red}} \alpha. \quad (2)$$

However, the reduced model should be able to predict the behaviour for a time range beyond the one used to compute the POD subspace. To go beyond this time span we need to make some assumptions about the reduced model [12], observe strong responses only in RME, RMG, RMH.

The overall behaviour of the model against the four scenarios proves its correctness. The network confirms the patterns reported in the literature for neurons correlated to particular dynamics, such as the touch response (the first three scenarios) and nictation (the fourth scenario). Note that the validation is only partially quantitative, as we are not interested at this point in the exact peak amplitude of certain neurons, but in the overall relative activity in the network.

Motion simulations with and without the AVA interneurons produce comparable results in terms of neurons activity overall, as the latter shows very strong responses in 61 neurons and average activity in 150 neurons.

4) Nictation (the worm stands on its tail and waves its head in three dimensions) The nictation behaviour is regulated in C. Elegans by a specific set of neurons, the IL2 cells [23]. By applying stimuli of 4.8 nA for IL2DL/IL2DR, IL2L/IL2R and IL2VL/IL2VR neurons, we should be able to observe activity mostly in the neurons associated with the head muscles: RMG, RMH, RME (RMEL, RMER, RMED) and SMD motor neurons.

In our case, as expected the most responsive neurons do not include the ones associated with forward/backward locomotion, but from the motor neurons that are known to inervate head muscles we
Fig. 4. The coefficients of the first four POD modes, \( \alpha_i, i = 1,4 \), vs. time. The blue dotted line represents the modal coefficients computed with (1) from the original snapshots matrix. The red markers show the approximate modal coefficients obtained by integrating (5) for the piecewise linear reduced models.

[10] and to extract a mathematical description. With the POD modes and their coefficients already computed from the snapshots matrix, we proceed with fitting this low-order model with an approximation with a known mathematical form.

We follow the main idea in [10], hence we only give here the information relevant for our problem. We consider a linear-quadratic approximation for the low-order model, whose evolution is thus described by a linear system of equations

\[
\dot{x} = a + Bx
\]

(3)

or by a quadratic one

\[
\dot{x} = a + Bx + x^T Cx,
\]

(4)

where \( a \in \mathbb{R}^{r\times1} \), \( B \in \mathbb{R}^{r\times r} \) and \( C \in \mathbb{R}^{r\times r\times r} \).

We project (3) and (4) on the subspace spanned by the POD modes to obtain a low-order model that lies in the POD basis coordinate system, whose mathematical form is now, for a fixed \( i = 1, r \):

\[
\frac{d\alpha_i}{dt} = a_i + \sum_{j=1}^{r} b_{ij} \alpha_j
\]

(5)

or in the quadratic case

\[
\frac{d\alpha_i}{dt} = a_i + \sum_{j=1}^{r} b_{ij} \alpha_j + \sum_{k=1}^{r} \sum_{j=1}^{r} c_{ijk} \alpha_j \alpha_k,
\]

(6)

so its dynamics is described by the evolution of the modal coefficients \( \alpha \). In these expressions \( \alpha \) – computed with (1) and their derivatives – computed with finite differences from (1) are known. To find the mathematical expression of the low-order model we solve an inverse problem, where the unknowns are the coefficients \( a_i, b_{ij} \) and \( c_{ijk} \).

Solution of (5) and (6) can be determined numerically using appropriate time discretization methods leading to a system of algebraic equations at each of the \( m \) time points. For a sufficiently large \( m \) the systems (5) and (6) are overdetermined, with \((m-1) \times r\) equations and \( r^3 + r \) unknowns \((r^3 + r^2 + r\) for the quadratic case). As in [10], we solve the systems in a least-squares sense to obtain the coefficients of the low-order model. We then use \( a_i, b_{ij} \) and \( c_{ijk} \) as coefficients of the reduced model, whose evolution is represented by the solution of the differential linear/quadratic system of equations (5) or (6) with now known coefficients.

Our goal is to compare the response of the two systems for a time range beyond the one used to obtain the reduction.

IV. RESULTS AND CONCLUSIONS

The relative energies of the POD basis functions are shown in Fig. 2. Note that these are identical with the energies of the eigenvalues of the covariance matrix \( \mathbf{R} = \frac{1}{m} \mathbf{A}^T \mathbf{A} \). The relative energies of the POD modes are in the order of \( 10^{-6} \) already from the 5th mode, indicating that their contributions
are less significant. This is also proved by the very small deviations, of 0.2\%, between the algebraic computation of $A_{\text{red}}$ for the time span considered. All the relative errors in this study are computed as $\max_{i=1}^N \left( \frac{||a_i - a_{\text{red}}||_2}{\max(a_i) \sqrt{m}} \right)$, where $i$ is a row in $A$ or $A_{\text{red}}$.

We show in this section the evolution of one state over time, the sensory neuron PHCL (the other states show relatively similar variations from the reduction point of view). Due to the distinctive shape of the neuronal signals, the linear approximation (5) is defective, as shown in Fig. 3 for $r = 4$. Moreover, the singularity at $t = 100$ ms impedes the quadratic time integration (6). Even for the time range ending before the singularity and for a larger number of POD modes, the approximation shows oscillations (Fig. 3).

To obtain an adequate fitting we remove the singularities by partitioning the time domain and approximate by parts with a linear model. This sectioning approach has previously been applied with success in building piecewise reduced models in other contexts [26], [25].

The singularities are identified as the points where the second derivative of the snapshots has very large absolute values (the second derivative is easily approximated using the snapshot information itself). Fig. 4 shows the coefficients of the four POD modes over time, where the original snapshots are separated in two parts by the removal of the singularity at $t = 100$ ms and each section is approximated by a linear low-order model. The blue dotted line represents the modal coefficients computed with (1) from the original snapshots matrix. The red markers show the approximate modal coefficients obtained by integrating (5) for the piecewise linear reduced models, with the coefficients $a_i$ and $b_{ij}$ previously computed in a least-squares sense for each section. The modal coefficients are good approximates for their original counterparts suggesting good accuracy for the states’ variations as well.

Moreover, the low-order model extracted this way has good prediction capabilities, as it accurately approximates the snapshots for a time range beyond the initial span. This is apparent in Fig. 5 where the snapshots for the first 50 ms are used to extract the approximate low-order model and then the latter is evaluated for an interval of 100 ms. In this case there was no need for sectioning, as the response has no singularities within the given time span. The case where the original snapshots contain a singularity is shown in Fig. 6. Here, the reduction procedure makes a piecewise approximation of 150 ms and the low-order model is used to predict the response for the next 50 ms. In both cases, the low-order model accurately predicts the behaviour of the original beyond the given time range.

The prediction of periodically repeated input is shown in Fig. 7 where the low-order subspace is obtained from the snapshots for 200 ms and the computed coefficients are
subsequently used to simulate a time range of 400 ms. The relative error at order $r = 4$ is 2% for the initial time span and 5.2% for the prediction beyond the time range of the initial data. There is no indication that this error would increase for an extended time span. The particular shape of the neuronal signal generates a singularity at every spike resulting most probably in two sections. Then the piecewise sub-models extracted from the initial time span are used to estimate the behaviour for the rest of the time domain. Therefore, extending the time domain should not increase the error nor add much to the overall cost, since it only contributes to the online phase, of evaluations of the low-order model. Even if a larger number of sections is needed, the identification of the singularities in the online phase does not boost the overall cost since the sections are automatically determined at constant cost with the local estimation of the derivatives.

With a sectioning approach such as the one proposed, one can efficiently simulate with precision the behaviour of the C. Elegans network with different internal parameters, to a variety of different stimuli and under various scenarios, using piecewise linear models and the accurate identification of the dominant time constants in each section.

The next step will be to create a small network of neurons with states and parameters now corresponding to the reduced model extracted and implement it in NEURON. This entails an analysis of the correct formulation of the inverse problem corresponding to extracting the topology, coefficients and internal parameters of a sub-circuit from given snapshots of an unknown nonlinear system. This endeavour can bring insight into the neuronal pattern redundancy in empirically observed behaviours and the identification and prediction of low-dimensional subspaces upon which the brain functions.

ACKNOWLEDGMENT

This work was partially supported by Portuguese national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020 as well as project PTDC/EEI-EEE/31140/2017.

REFERENCES