Journal of Computational and Applied Mathematics 258 (2014) 135–150

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Proper orthogonal decomposition for parameter estimation in oscillating biological networks

Amanda M. Rehm [b], Elizabeth Y. Scribner [b], Hassan M. Fathallah-Shaykh [a,b,c,d,e,f,g,*]

[a] Department of Neurology, The University of Alabama at Birmingham, Birmingham, AL, United States
[b] Department of Mathematics, The University of Alabama at Birmingham, Birmingham, AL, United States
[c] Department of Cell, Developmental and Integrative Biology, The University of Alabama at Birmingham, Birmingham, AL, United States
[d] Department of Biomedical Engineering, The University of Alabama at Birmingham, Birmingham, AL, United States
[e] Department of Mechanical Engineering, The University of Alabama at Birmingham, Birmingham, AL, United States
[f] The UAB Comprehensive Neuroscience Center, Birmingham, AL, United States
[g] The UAB Comprehensive Cancer Center, Birmingham, AL, United States

ARTICLE INFO

Article history:
Received 20 September 2012
Received in revised form 25 August 2013

Keywords:
Proper orthogonal decomposition
Ordinary differential equations
Biological networks
Oscillations
Parameters

ABSTRACT

Proper orthogonal decomposition (POD) is frequently applied to estimate parameters of partial differential equations. This study examines the application of the POD method in estimating the parameters of an Ordinary Differential Equation (ODE) model of stable oscillating biological networks. The mathematical model used to simulate molecular interactions in these oscillating networks is related to the Gause–Lotka–Volterra equations. The findings reveal that POD generates accurate estimates of the parameters even in the presence of experimental noise; furthermore, extrapolating biologically measured data points to a number of oscillations improves the curve fits, C^1 approximations, and parameter estimations.

1. Introduction

The Proper Orthogonal Decomposition method was initially presented by Kari Karhunen and Michel Loéve independently of one another in the 1940s, and therefore, it is also known as the Karhunen–Loéve Expansion (KLE), [1]. Lawrence Sirovich later developed the method of "snapshots", or observations, to the KLE, which reduces the number of eigenvectors by excluding eigenvectors with eigenvalues less than a certain value [2–4]. The use of POD in mathematical modeling has found applications in several disciplines of study, including signal analysis, pattern recognition, control theory, fluid dynamics, and more recently cardiac electrophysiology [1,5–11]. The goal of the POD method is to produce a reduced-order model of a system, which can then be used to solve an inverse problem, like parameter estimation in a set of differential equations. It is most commonly used to estimate parameters in a system of partial differential equations; while several authors have applied the POD method in parameter estimation problems [5,8,10,11], to our knowledge, none have applied the technique in estimating parameters for stable oscillating biological networks.

* Correspondence to: The University of Alabama at Birmingham, FOT 1020, 150 3rd Avenue S, Birmingham, Al 35295-2410, United States. Fax: +1 205 934 1432.
E-mail address: hfathall@uab.edu (H.M. Fathallah-Shaykh).

0377-0427/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cam.2013.09.004

In nature, many biological systems have periodic tendencies. Some examples in the human body are neural activity in the central nervous system, hormone levels within the menstrual cycle, and the proteins and genes regulating the circadian clock. Mathematical models are important because they allow scientists and researchers to better understand the properties of each molecular network, such as the effects of small perturbations or mutations on the stability of the oscillating system [12].

Ordinary Differential Equations (ODEs) are often used to model these oscillating biological networks, yet they require parameter values that accurately reflect specific molecular interactions in the network-namely, the magnitude of each species' activation or suppression rate on another species. One advantage of working with stable oscillating networks is that if a certain number of molecular concentrations are known in a single oscillation from biological assays, these values can be extrapolated to any number of time-points in subsequent oscillations given the periodic tendency of the system. Then, having molecular concentrations from an optimal number of oscillations, one is faced with an inverse problem of estimating the parameters of an ODE model of the system. This parameter discovery can be computationally intense, especially in networks involving a large number of interacting elements. Hence, we hypothesize that we can apply the POD method to describe the interaction of elements in a stable oscillating network.

Until recently, the most common approach to parameter estimation for non-linear systems has been through non-linear least-squares fitting [13]. However, the POD method presents many advantages over the least-squares approach. First, the POD method produces a reduced order model that is a linear combination of an optimal set of basis elements, computed specifically to capture the main characteristics of the system [5]. In fact, our specific investigation will show that the POD method more accurately models the known biological data than other data-fitting methods, such as splines. Finally, as will be explained in subsequent sections, the method is less computationally intense because it takes an optimally large set of parameter values and then narrows the set to a few estimates to be tested.

The paper is organized as follows. After a brief overview of the types of differential equations used to model oscillating biological networks in Section 2, we present a specific stable oscillating network and its corresponding set of Ordinary Differential Equations in Section 3. Here also, we outline the nature of the inverse problem, which is to use the reduced-order POD model to estimate the unknown parameters in the system. In Section 4, we present the main theory behind the POD method as well as a step-by-step guide to its application in first modeling a stable oscillating system and then solving for the system's parameters. Finally, in Section 5, we apply this guide to a specific optimization problem, where we solve for two of the parameters in the proposed system from Section 3. In particular, we investigate the optimal number of parameter values needed to compute the POD basis elements, and we also optimize the number of oscillations to which we extrapolate our known set of experimental data points. Section 6 details our results, followed by a conclusion in Section 7.

2. System of ordinary differential equations

Assuming that genes/proteins $j \in \{1, \ldots, n\}$ regulate the production of gene/protein i, we use the following type of differential equations as a general model:

$$\frac{dx_i(t)}{dt} = u_i(t)x_i(t)(s_i - x_i(t)), \quad 1 \le i \le n, \tag{1}$$

$$u_i = B_i + \sum_{j=1}^{n} \lambda_{j,i} x_j(t), \quad B_i \ge 0$$

where x_i is the state vector representing the concentration of molecule i at its site of action. The initial conditions are taken such that $x_i(t) \in [0, s_i]$, where s_i is the saturation level of molecule i. The real parameters $\lambda_{j,i}$ are regulatory weights that encode the effects of molecule j on the production rate of molecule i. The sign of $\lambda_{j,i}$ indicates the effect of j on i, where positive and negative signs denote activation and suppression, respectively. The magnitude of $\lambda_{j,i}$ reflects the strength of stimulation or repression. The model incorporates a logistic term $[(x_i)(s_i - x_i)]$. Note that the term $\lambda_{i,i}$ describes the sum of actions of a molecule on itself, which could reflect degradation and possible self activation. The term $B_i \ge 0$ is a constant term representing basal synthesis. The term $u_i(t)$, which controls the sign of the derivative $\frac{dx_i(t)}{dt}$, reflects the linear sum of the regulatory forces acting on molecule i. Note that when $u_i(t) > 0$, the sum of the negative regulatory influences (NRI) (degradation and repression) is less than the sum of the positive regulatory influences (PRI) (activation). Likewise, $u_i(t) < 0$ when the sum of PRI is less than the sum of NRI. Notice that $\dot{x}_i = 0$ implies that $u_i(t) = 0$ (NRI = PRI) since the logistic term ensures that $x_i(t)$ is never identically zero or at full saturation s_i for any finite time (as long as the initial concentration of x_i is not identically zero or s_i, although it can be arbitrarily close).

The overall structure of Eq. (1) is similar to that of a logistic equation, $\frac{dx(t)}{dt} = rx(t)(s - x(t))$, where r is the growth or decline rate of population x, and s is the carrying capacity, or saturation level. Observe that the logistic term generates s-shaped curves that mimic biological data. However, in Eq. (1), the constant rate r has been replaced by a varying rate of change, $u_i(t)$, which is dependent on the summation of a molecule's regulatory signals at a given time t. At any given time, the population (or concentration of a molecular species) is growing or decaying at a rate $u_i(t)$, which fluctuates between the maximal and the minimum rates of formation of i. Thus, the closer $u_i(t)$ is to the maximal (minimal) rate of formation, the faster the molecule i will approach its saturation level s_i (zero).