Genetic Algorithms for Parameter Estimation in Circadian Model

F.M. Janerio, P. Vo Tan, L. Ton That, and C. Nguyen Thai

Abstract

This paper proposes an algorithm to estimate two parameter values $\boldsymbol{v_s}$, transcription of *frq* gene, and $\boldsymbol{v_d}$, maximum rate of FRQ protein degradation for an existing 3rd order Neurospora model in literature. Details of the algorithm with simulation results are shown in this paper.

Keywords

Biological system • Genetic algorithms • Parameter estimation • Circadian clock • Circadian system

1 Introduction

For almost every living species on Earth, there exist internal periodic oscillations that control their daily biological activities. These oscillations are known as circadian rhythms and are governed by a biological clock, which is placed at suprachiasmatic nucleus (SCN) (for human). SCN has independent rhythms and the ability to re-adjust its own rhythms with the external rhythms [1]. Research results have shown that losing this ability will cause the health to degrade, even leading to dangerous symptom such as cancer [2, 3]. Because of its importance, circadian rhythm is a topic that has attracted researchers over many years. Some results have been published, notably with the discovery of key genes that are parts of circadian rhythms [4, 5], the mathematical models that describe characteristics of circadian rhythms [6, 7], and circadian phase recovery [8, 9].

System identification is one of the inspired research areas for circadian rhythm [10]. Although several mathematical models of circadian rhythms have been created successfully, the measurements made for exact parameter values that can be used for these models are still a challenge. Some results have been achieved in identification of circadian system [9], [11]. In this paper, we propose a method based on Genetic Algorithm to estimate two parameter values of a Neurospora model given in [6]. The parameters that required to estimate are v_s, transcription of *frq* gene, and v_d, maximum rate of FRQ protein degradation. According to the research on sensitivity analysis performed in [5], v_s and v_d are among the parameters that show the most sensitivity to light entrainment, which make them unpredictable.

The search for the correct parameter values is an optimization problem where a suitable cost function needs to be minimized. Although this is simple for the search of one parameter value, it becomes more challenging when multiple parameters are being searched for. If in addition there is a wide search space and the cost function presents local minimums, the traditional search algorithms become unsuitable for the task. In this type of parameter optimization problems, Genetic Algorithms (GA) [12] have been successfully used [13, 14]. They are based on the way biological populations evolve from generation to generation. It will be shown that GAs are capable of estimating both v_s and v_d from the state variables evolution.

F.M. Janerio
Instituto de Telecomunicações, Universidade de Évora, Évora, Portugal

P. Vo Tan (✉) · L. Ton That · C. Nguyen Thai
School of Electrical Engineering, International University-National University HCMC, Ho Chi Minh City, Vietnam
e-mail: vtphuoc@hcmiu.edu.vn

T. Vo Van et al. (eds.), *6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6)*, IFMBE Proceedings 63, https://doi.org/10.1007/978-981-10-4361-1_107

2 Circadian Model

The circadian rhythm of Neurpspora crassa is characterized by the synthesis of the FRQ protein inside the Neurospora cell. The transcription process of *frq* gene produces *frq* mRNA. In the cytoplasm (outside the nucleus), the translation process will synthesize the FRQ proteins. A portion of synthesized FRQ proteins goes back to nucleus to inhibit the transcription process of *frq* gene. The FRQ protein synthesis process acts like a negative feedback mechanism and is described in [6] with the set of differential equations

$$\dot{M} = v_s \frac{K_I^n}{K_I^n + F_N^n} - v_m \frac{M}{K_m + M} \tag{1}$$

$$\dot{F}_C = k_s M - v_d \frac{F_C}{K_d + F_C} - k_1 F_C + k_2 F_N \tag{2}$$

$$\dot{F}_N = k_1 F_C - k_2 F_N \tag{3}$$

where state variables $\{M, F_C, F_N\} = \{r_1, r_2, r_3\}$ are respectively the concentration of *frq* mNRA, the concentration of FRQ protein synthesis and the concentration of nucleus FRQ protein. All parameters involved in (1), (2) and (3) are given in [6] and shown in Table 1.

With initial conditions set as $[10.51.25]^T$, the fluctuation of the state variables $\{M, F_C, F_N\}$ exhibits an oscillation around 21.5 h and is depicted in Fig. 1.

3 Genetic Algorithm

The estimation of the system parameters (shown in Table 1) that result into the periodic evolution shown in Fig. 1 is, in general, a very difficult problem to solve. In this section it will be shown how a Genetic Algorithm can be an effective way to estimate the values of up to 2 parameters in the system of Eqs. (1)–(3). It will be seen that the estimation of a single parameter (v_s) is relatively easy and only requires a sweep in the parameter's domain followed by the application of a traditional search algorithm. However, for the estimation of more than one parameter, the traditional search algorithms are not effective anymore requiring the aid, for example, of a genetic algorithm.

Table 1 Parameter values of the circadian model

Parameter	Value	Unit
v_s	4	nMh^{-1}
K_I	1	nM
n	4	
v_m	0.505	nMh^{-1}
K_m	0.5	nM
k_s	0.5	h^{-1}
v_d	1.4	nMh^{-1}
K_d	0.13	nM
k_1	0.5	h^{-1}
k_2	0.6	h^{-1}

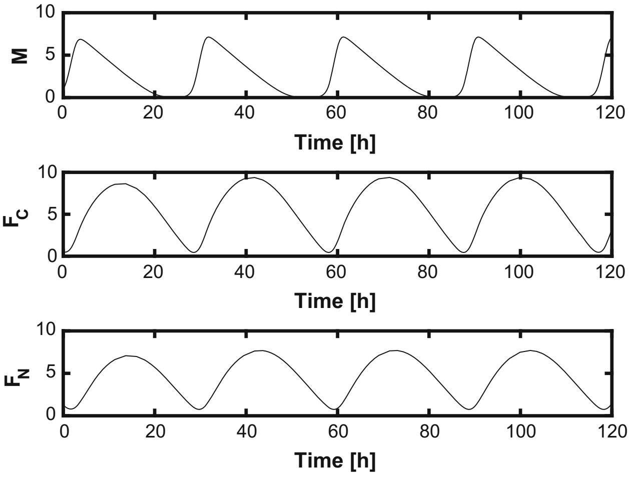

Fig. 1 Circadian oscillations with the parameters shown in Table 1

Assuming that measurements of the 3 state variables are available at times t_k where $k = \{1, 2, \ldots, N_k\}$ and N_k is the number of available measurements for each state variable $r_j^{(target)}$ (with $j = 1, 2, 3$), then the fitness of the each possible solution for the unknown parameters can be assessed by the following cost function

$$\varepsilon = \sqrt{\frac{1}{3N_k} \sum_{j=1}^{3} \sum_{k=1}^{N_k} \left(\frac{r_j^{(\text{estimated})}(t_k) - r_j^{(\text{target})}(t_k)}{r_j^{(\text{target})}(t_k)} \right)^2} \tag{1}$$

which is defined in a least squares sense as the relative difference between the target state variables evolution $r_j^{(target)}$ and the estimated state variables evolution $r_j^{(estimated)}(t_k)$ at the available time stamps t_k. The objective is to minimize the relative difference between the target and estimated evolution of the state variables.

3.1 Estimation of v_s Parameter

When all parameters are known, except v_s, it is possible to search for the value that minimizes the cost function (4). In this case, a sweep in the parameter v_s was performed in the range [0; 10] since the state variables must have a periodic evolution in a circadian cycle. For each candidate v_s, the system of differential Eqs. (1)–(3) was solved resulting in an estimated evolution of the state variables $r_j^{(estimated)}(t_k)$ which

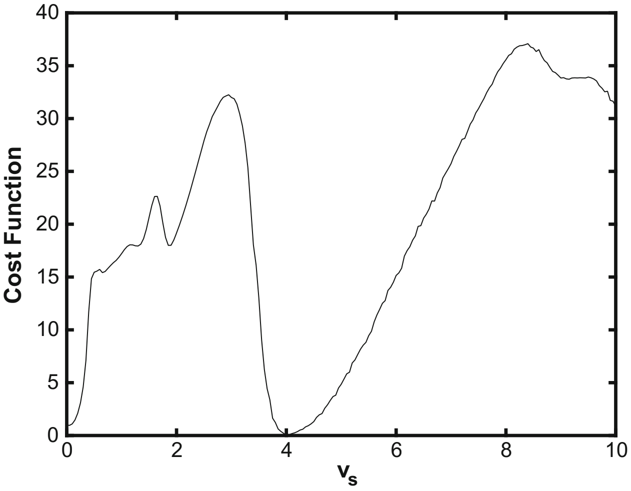

Fig. 2 Cost function as a function of the v_s parameter

is then used in (4) to estimate the suitability of the candidate value. The result of the sweep is shown in Fig. 2 which presents a minimum when the value of v_s is 4, which corresponds to the correct value as shown in Table 1. The actual minimum can be found by applying a traditional search algorithm, such as the Nelder-Mead simplex algorithm, using a v_s in the neighborhood of the absolute minimum as the initial search value.

It can be concluded that, if only one parameter is unknown, it is relatively easy to estimate its value from the evolution of the measured state variables. However, when more than one parameter needs to be estimated, it is not anymore practical to perform sweeps in the parameters to be estimated. A different approach should be used.

3.2 Genetic Algorithm

Genetic algorithms are inspired in the evolution, through the generations, of biological populations. It is based on the survival of the fittest paradigm in the sense that the fittest population members have a greater probability of reproducing, thus passing their good characteristics to the next generations.

The algorithm starts by randomly creating a population of $N = 30$ candidate solutions. Each individual of the population is made up of genes, with each gene encoding the value of an unknown parameter. For example, if the problem consists in estimating 5 parameters, the number of genes in each individual (*i.e.*, candidate solution) would be 5. Each gene would then contain a potential value of each of the parameters to estimate. Once the population is created each individual is tested by solving the system described in Eqs. (1)–(3) to obtain the estimated state variable evolution and compare it with the target evolution through the application of the cost function (4). The population is then sorted according to the fitness of each individual (a fit individual will have a low cost function value). Next, pairs of individuals will be randomly selected for reproduction. The fittest candidates have a higher probability of being chosen to reproduce because a rank-based biased roulette wheel selection scheme is used. Once the individuals are chosen, they have a probability $p_{cross} = 0.8$ of crossing over (thus creating offspring). If they do not reproduce, then they will pass into the next generation unchanged. After crossover, a mutation operation is applied where an individual has a probability $p_{mut} = 0.2$ of suffering a mutation in a randomly chosen gene. At this stage a new generation of potential solutions has been created. Their fitness is evaluated and the process of crossover and mutation repeats itself until a maximum number of generations is reached or the cost function is below a certain predefined threshold (in this paper, the threshold is set to 0.001).

However, it should be noted that a Genetic Algorithm is very efficient in finding the absolute minimum region of the cost function, but has difficulties in fine-tuning the solution onto the absolute minimum. Therefore, when the Genetic Algorithm provides an estimate of the solution, a traditional search method is applied using the GA solution as starting point for the search. This way the absolute minimum of the cost function can be efficiently found.

The crossover and mutation operations are essential to the success of the Genetic Algorithm. Crossover allows the good characteristics of the candidates to pass from generation to generation. Mutation is essential since it provides a way to increase the diversity of the population [12].

3.3 Estimation of v_s and v_d Parameters

This subsection describes the results of the parameter estimation when 2 parameters are unknown. The unknown parameters are the v_s and v_d values. In this case, performing a parameter sweep is not practical due to the two-dimensional nature of the problem. However, Fig. 3 shows the cost function value as a function of the two unknown parameters. While it is still possible to plot this function, for problems with more parameters to estimate, it would not be possible. Additionally, the search space is relatively wide and the cost function presents many grooves and local minimums. Therefore, a traditional search algorithm cannot be applied because it will most likely get trapped in a local minimum. Fortunately, Genetic Algorithms are very well suited to this class of problems which have wide search spaces, local minimums and various parameters to estimate.

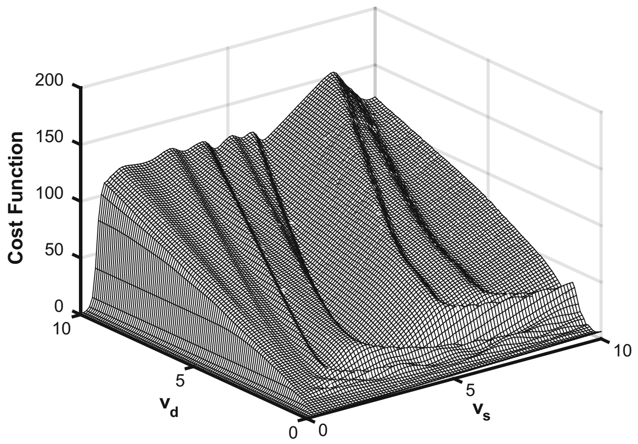

Fig. 3 Cost function as a function of parameters v_s and v_d

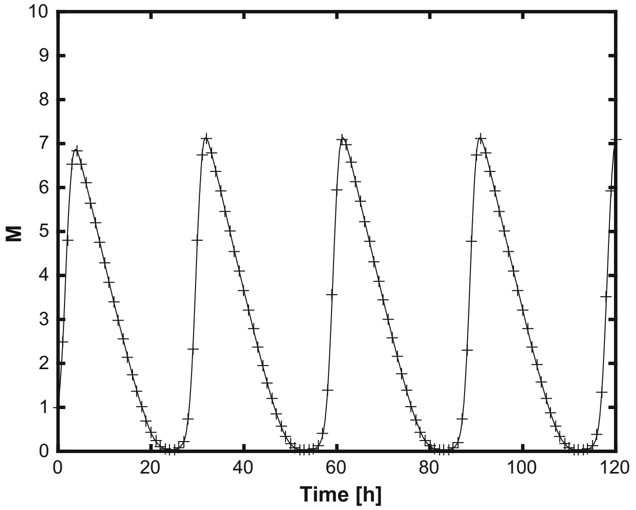

Fig. 4 Evolution of the state variable *M*. The crosses represent the target data available every hour. The continuous line represents the evolution obtained with the v_s and v_d parameters obtained by the Genetic Algorithm

The GA explained in previous section was used to search for the v_s and v_d values that resulted in an evolution of state variables that matched the target results shown in Fig. 1. The search range was [0, 10] for both parameters since the evolution must have a periodic behavior. The target data was sampled to simulate measurements with 1 h sampling time. The target evolution of the 3 state variables is shown with crosses in Figs. 4, 5 and 6. The estimated values by the Genetic Algorithm, followed by the Nelder-Mead simplex algorithm, are shown in Table 2 along with the relative estimation error. The evolution of the state variables obtained with the estimated parameters is also included in Figs. 4, 5 and 6 as continuous lines. The cost function value for this case was $\varepsilon = 7.9x10^{-9}$.

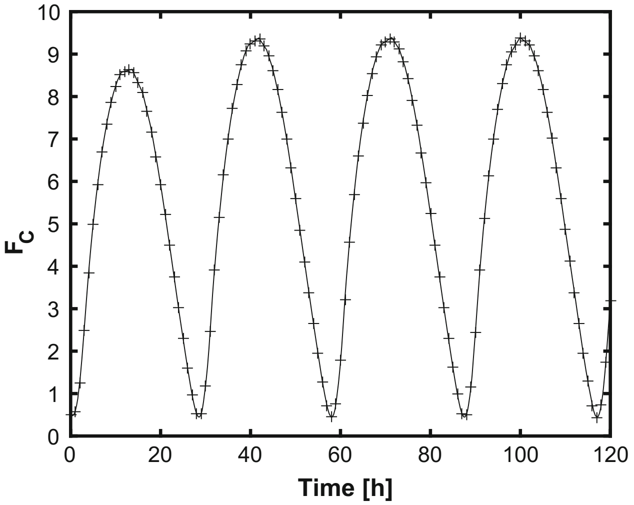

Fig. 5 Evolution of the state variable F_C. The crosses represent the target data available every hour. The continuous line represents the evolution obtained with the parameters v_s and v_d obtained by the Genetic Algorithm

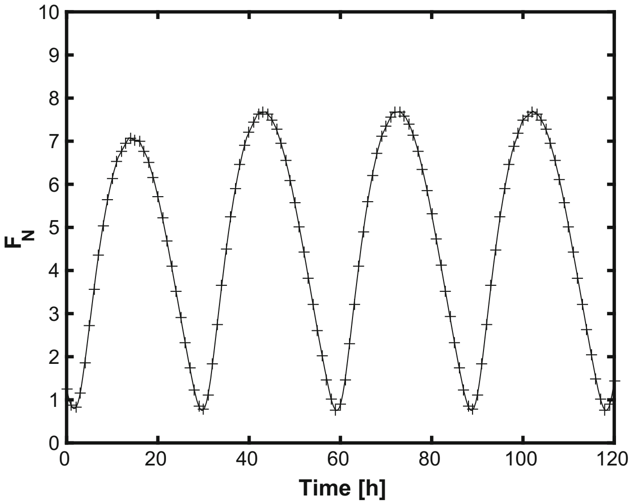

Fig. 6 Evolution of the state variable F_N. The crosses represent the target data available every hour. The continuous line represents the evolution obtained with the parameters v_s and v_d obtained by the Genetic Algorithm

Table 2 Estimated parameters values and relative estimation errors

Parameter	Estimated value	Relative error
v_s	4	1.18×10^{-10}
v_d	1.4	6.63×10^{-10}

4 Conclusions

A Genetic Algorithm has been used to estimate the parameters that govern the circadian cycle of Neurospora. It was shown that when only one model parameter is unknown, a parameter sweep followed by a traditional search algorithm suffices to estimate that parameter. However, when two parameters are unknown, it is not practical to perform a parameter sweep. Therefore, a Genetic Algorithm was successfully used to estimate two parameters of the circadian model of Neurospora.

References

1. Weaver DR (1998) The suprachiasmatic nucleus: a 25 year retrospective. J Biol Rhythms 13(2):100–112
2. Shanmugan V, Wafi A, Al-Taweel N, BusselBerg D (2013) Disruption of circadian rhythms increases the risk of cancer, metabolic syndrome and cardiovascular disease. J Local Glob Health Sci. doi:10.5339/jlghs.2013.3
3. Savvidis C, Koutsilieris M (2012) Circadian rhythm disruption in cancer biology. Mol Med 18(1):1249–1260
4. Kaldi K, Gonzalez BH, Brunner M (2006) Transcriptional regulation of Neurospora circadian clock gene wc-1 affects the phase of circadian output. EMBO Rep 7:199–204
5. Bagheri N, Stelling J, Doyle III FJ (2008) Circadian phase resetting via single and multiple targets. PLOS Comput Biol 4
6. Gonze D, Leloup JC, Goldbeter A (2000) Theoretical models for circadian rhythms in Neurospora and Drosophila. C R l'Acadmie Sci Ser III Sci Vile 323:57–67
7. Scheper T, Klinkenberg D, Pennartz C et al (1999) J Neurosci 19 (1):40–47
8. Vo-Tan P, Ton-That L (2015) Circadian phase control using adaptive synchronization. In: 8th IFAC symposium on robust control design (RODCON) 48(14):68–71
9. Zhang J, Bierman A, Wen J et al (2010) Circadian system modeling and phase control. In: 49th IEEE conference on decision control, Atalanta, GA,USA
10. Sontag ED (2004) Some new directions in control theory inspired by system biology. IEE P Syst Biol 1:9–18
11. Foo M, Somers D, Kim P (2015) System identification of the arabidopsis plant circadian system. J Korea Phys Soc 66(4):700–712
12. Haupt RL, Haupt SE (2004) Practical Genetic Algorithms, 2nd edn. Wiley, New Jersey
13. Janeiro FM, Ramos PM (2009) Impedance measurements using genetic algorithms and multiharmonic signals. IEEE Trans Instrum Meas 58:383–388
14. Janeiro FM, Santos J, Ramos PM (2013) Gene expression programming in sensor characterization: numerical results and experimental validation. IEEE Trans Instrum Meas 62(5)