ON THE OBSERVABILITY AND STATE ESTIMATION IN A CLASS OF GENE-EXPRESSION SYSTEM

RICARDO AGUILAR-LÓPEZ¹ AND JUAN L. MATA-MACHUCA²

¹Department of Biotechnology and Bioengineering CINVESTAV-IPN, IPN 2508, 07360, Mexico City, MEXICO ²Department of Advanced Technologies, National Polytecnic Institute (Instituto Politécnico Nacional) IPN 2580, 07340, Mexico City, MEXICO

ABSTRACT: The failure of available physical sensors for the online measurement of protein concentration in cells is a key problem to understanding the transition in bio-systems. In this paper, a new state observer has been designed to estimate three protein concentrations by using a gene-expression mathematical model. Interestingly, its only input is the concentration of one messenger RNA (mRNA). Similarly by differential-algebraic observability analysis is showed that the gene-expression model is indeed observable. The observer convergence was demonstrated by analysing the estimation error dynamics. *In silico* experiments confirm the satisfactory performance of this new observer.

AMS Subject Classification: 34A34, 92B05, 93B07, 93E10

Key Words: observer design, convergence proof, gene expression model, observability analysis

Received:April 13, 2018;Accepted:June 8, 2018;Published:June 10, 2018doi:10.12732/dsa.v27i3.5Dynamic Publishers, Inc., Acad. Publishers, Ltd.https://acadsol.eu/dsa

1. INTRODUCTION

From the soil bacterial communities to the human immune response cells, all biological systems are well known for their nonlinear behavior. This is due to the intrinsic relationship among their components, and is in turn reflected in the reversible and autocatalytic biochemical reactions, the abundant feedbacks and the circadian rhythms [1, 2]. Currently, the study of biological systems can be done by focusing on the cell and its most basic biochemical units, such as nucleic acids, lipids, carbohydrates and proteins [3].

The quantification of all of the above macromolecules calls for the development of sophisticated mathematical models. These models are tools, which help account for the intracellular behavior with respect to time. Yet, the measurement of macromolecule concentration can be expensive, time consuming and in many cases only possible off line. This is a strong limitation for system manipulation in real time in bioprocess applications. In spite of this, system concentrations can be determined on-line via the design and implementation of state observer algorithms, also named software sensors. They can accurately estimate the vector values, which cannot otherwise be measured by physical sensors [4].

Research on state estimation has developed quickly during the last 50 years [5, 6]. However, there have only been a handful of works on the successful application of the observer algorithms to the estimation of RNA and protein concentrations in metabolically active cells. The work [7] is focused on the reconstruction of the gene expression dynamics when the online measurement of all the transcription network components was not feasible. In their work, they employed a mathematical model of the transcription process, a set of online measurements of the promoter transcriptional activity (obtained by the real-time polymerase chain reaction (RT-PCR) technique), and also the classical extended Luenberger observer. This observer can estimate mRNA and protein concentrations despite non-smooth input perturbations, small parameter uncertainties and low noise measurements. (For a review on current observer design problems, see [8]. Further, the extended Luenberger observer also shows effectiveness in spite of measuring only one target gene. The contribution [9] deals with three different state observers (the extended Luenberger observer, a linear observer and the extended Kalman filter) and a basic model that describes the transcription-translation process. In [9] is aimed to reconstruct mRNA concentration by the sole measurement of total protein concentration. The best performance was achieved with the nonlinear observer, although it exhibited considerable overshoots. In the manuscript [10] is applied the nonlinear Luenberger observer to accurately estimate the fifteen state variables of a gene transcription circuit based on Goodwin's model. And they did so, by measuring an inhibitor from the gene circuit. This observer showed robustness to small model uncertainties and low noise output measurements.

Despite the successful application of this kind of classic observers in estimating the mRNA and proteins concentrations in gene-expression systems, it is imperative to design new nonlinear observer algorithms. These need to significantly improve the performance of the Luenberger observer or the Kalman Filter [6, 11]. And this can be achieved by decreasing the overshoots and the settling time, while increasing the robustness to high noise measurements and large parameter uncertainties. Hence, in this article we propose an alternative observer with dual feedback, comprised of a proportional and a sigmoidal function. This observer is able to estimate the time course of the three protein concentrations in a basic gene-expression model by measuring only one mRNA concentration with the RT-PCR technique. This is demonstrated by the observer convergence proof, which analyzes the estimation error dynamics. Furthermore, numerical simulations corroborate the fine performance of the proposed observer algorithm.

The rest of this work is organized as follows. Section 2 describes the gene expression modeling. Section 3 contains some important concepts such as observability properties and differential-algebraic observability analysis. In Section 4, the state observer is designed. Section 5 explains the obtained results. Finally, we give some concluding remarks in Section 6.

2. GENE EXPRESSION MODELING

The gene expression model was taken from Zhdanov [12], whose work describes the transcription-translation process. In simple terms, transcription is a set of steps that lead to the synthesis of RNA from genomic DNA. On the other hand, translation is the process in which a kind of RNA called mRNA is used as a blueprint to direct protein synthesis.

This model takes into account the relationship between one mRNA and the three proteins synthesized thereafter. Two of these proteins are the precursors of the third one, which is a transcription inhibitor (a feedback regulator). The model can be considered as an analog of a chemical reaction network governed by the mass conservation principle (MCP) [13, 14, 15, 16].

The gene expression model is given by,

$$\frac{dN}{dt} = w \left(\frac{K}{K+n_3}\right)^m - kN \tag{1}$$

$$\frac{dn_1}{dt} = vN - k_{12}n_1 - k_1n_1 \tag{2}$$

$$\frac{dn_2}{dt} = k_{12}n_1 - k_{23}n_2 \tag{3}$$

$$\frac{dn_3}{dt} = k_{23}n_2 - k_3n_3 \tag{4}$$

Where N is the mRNA, n_1 and n_2 are the precursor proteins 1 and 2, respectively, n_3 is the active protein 3, w is the transcription rate at $n_3 \rightarrow 0$, K is the constant

of association of n_3 with a regulatory site, m is the number of regulatory sites, v is the translation rate constant, k_{12} and k_{23} are the conversion rate constants, and k, k_1 and k_3 are the degradation rate constants. All the constants have a strictly positive value.

In order to analyze the dynamic properties of the gene-expression model (1)-(4), let us consider the following condition.

Assumption 1. Let $x = [N, n_1, n_2, n_3]' \in \Omega \subset \mathbb{R}^4_+$ be the state vector of the gene-expression model (1)-(4). Let the set $\Omega \subset \mathbb{R}^4_+$ be the physically realizable domain of the dynamical model if

 $\Omega = \left\{ [N, n_1, n_2, n_3]' \in \mathbb{R}^4_+ \mid 0 \le N \le N_m; \ 0 \le n_i \le n_{im}, \ i = 1, 2, 3 \right\}$

where $N_m < \infty$ and $n_{im} < \infty$, i = 1, 2, 3, are the maximal values for the mRNA, precursor protein 1, precursor protein 2, and active protein, respectively. Thus, we write

$$N_m > \frac{w}{k} \left(\frac{K}{K+n_3}\right)^m; \ n_{1m} > \frac{vN}{k_{12}+k_1}; \ n_{2m} > \frac{k_{12}}{k_{23}}n_1; \ n_{3m} > \frac{k_{23}}{k_3}n_2$$

Now, from the existence and uniqueness theorem [17] it is straightforward to check that if $f(x) = \left(\frac{K}{K+n_3}\right)^m$ is locally Lipschitz on Ω , then the system (1)-(4) with $x(t_0) = x_0$ for any $x_0 \in \Omega$ has a unique solution over $[t_0, t_0 + \delta]$ for some δ .

With respect to the equilibrium points for the gene-expression model (1)-(4), we have the following statements. The equilibrium points are denoted by $\bar{x} = [\bar{N}, \bar{n}_1, \bar{n}_2, \bar{n}_3]' \in \Omega$, and are defined by $\bar{n}_1 = \frac{v\bar{N}}{k_{12}+k_1}, \bar{n}_2 = \frac{k_{12}}{k_{23}}\bar{n}_1$, where \bar{N} and \bar{n}_3 satisfies $k\bar{N} = w \left(\frac{K}{K+n_3}\right)^m$. In the original reference of the gene-expression model [12, 18], it is shown that around one equilibrium $\bar{x} \in \Omega$ there exists an oscillatory attractor. In the following section, we discuss the observability properties of the system (1)-(4).

3. DIFFERENTIAL-ALGEBRAIC OBSERVABILITY ANALYSIS

Observability is a system property, which offers the possibility of determinate if internal states on the basis of input/output data can be estimated [19].

Now, consider a state space representation of the system (1)-(4) as a nonlinear autonomous system which involves an input/output state space, with the following structure:

$$\dot{x} = f(x, u)
y = \Upsilon x$$
(5)

Where $x \in \mathbb{R}^n$ is the state vector; $u \in \mathbb{R}^q$ is the input vector, $y \in \mathbb{R}^p$ is the continuous and differentiable measured vector and Υ is the measurable matrix. In the study of observability conditions for nonlinear models, there exist three main observability methods known as linear, geometric and differential-algebraic approaches [23]. We will consider differential-algebraic methods in this proposal.

The differential-algebraic observability is relatively recently methodology based on differential-algebraic techniques. This observability approach suggests that a system is observable if and only if all the states of the model can be expressed in terms of the measurable outputs and a finite number of its time derivatives. For a preliminary background, see references [20, 21].

Theorem 2. A state x_i is said to be algebraically observable, if it is a function of u (inputs), y (measurable output) and the first $r_1, r_2 \in \mathbb{N}$ time derivatives of u and y [22, 23], that is:

$$x_{i} = F_{i}\left(u, \dot{u}, \ddot{u}, \frac{d^{3}u}{dt^{3}}, \dots, \frac{d^{r_{1}}u}{dt^{r_{1}}}, y, \dot{y}, \ddot{y}, \frac{d^{3}y}{dt^{3}}, \dots, \frac{d^{r_{2}}u}{dt^{r_{2}}}\right) = 0$$
(6)

where the function $F_i : \mathbb{R}^{(r_1+1)q} \times \mathbb{R}^{(r_2+1)p} \to \mathbb{R}$.

To satisfy Theorem 2, it is necessary to express each state variable in terms of the measurable outputs and a finite set of their time derivatives.

If the measured output of the gene model is the mRNA concentration (N), then the following algebraic relationship can be written:

$$N = y \tag{7}$$

Then, considering eq. (1) and substitute it with eq. (7),

$$n_{3} = K \left(\frac{1}{\left(\frac{\dot{y} + ky}{w}\right)^{\frac{1}{m}}} - 1 \right)$$

$$= F_{3}(y, \dot{y})$$
(8)

Replacing eq. (8) and its time derivative in eq. (4) we obtain,

$$n_{2} = -\frac{K}{k_{23}} \left[k_{3} \left(1 - \frac{1}{\left(\frac{\dot{y} + ky}{w}\right)^{\frac{1}{m}}} \right) + \frac{k\dot{y} + \ddot{y}}{mw \left(\frac{\dot{y} + ky}{w}\right)^{\frac{1}{m} + 1}} \right]$$
(9)
= $F_{2}(y, \dot{y}, \ddot{y})$

Finally, substituting eq. (9) and its time derivate in eq. (3),

$$n_{1} = -\frac{K}{k_{12}} \left[k_{3} \left(1 - \frac{1}{\left(\frac{\dot{y} + ky}{w}\right)^{\frac{1}{m}}} \right) + \frac{\left(k\dot{y} + \ddot{y}\right) \left(1 + \frac{k_{3}}{k_{23}}\right)}{mw \left(\frac{\dot{y} + ky}{w}\right)^{\frac{1}{m} + 1}} + \frac{1}{k_{23}mw} \left(\frac{k\ddot{y} + \dddot{y}}{\left(\frac{\dot{y} + ky}{w}\right)^{\frac{1}{m} + 1}} - \frac{k\dot{y} + \dddot{y}}{mw \left(\frac{\dot{y} + ky}{w}\right)^{\frac{1}{m} + 2}} \right) \right]$$

$$= F_{1}(y, \dot{y}, \ddot{y}, \dddot{y})$$

$$(10)$$

Note that eqs. (7)-(10) is a set of differential algebraic equations, where the mRNA and the three proteins mass balances are explicit and are functions of y, \dot{y} , \ddot{y} , \ldots , $\frac{d^{n-1}y}{dt^{n-1}}$. Therefore, from Theorem 2, we conclude that this system is algebraically observable. This result agrees with the observability matrix rank. And thus, a non-linear observer algorithm can be utilized.

4. STATE OBSERVER DESIGN

The nonlinear observer design assumes continuous time measurements. It is based on the mathematical model shown in equations (1)-(4),

$$\begin{aligned} \dot{x} &= f(x, u) \\ y &= \Upsilon x \end{aligned}$$
 (11)

where $x = [N, n_1, n_2, n_3]' \in \Omega \subset \mathbb{R}^4_+$ is the state vector and y = N is the measured output.

Let us take into account the following assumption.

Assumption 3. System (11) satisfies the condition

$$\|f(x,u) - f(\hat{x},u)\| \le L < \infty$$

We suppose the system is bounded in x and uniformly bounded regarding u.

Now, we establish the main result of this work.

Proposition 4. Consider the gene-expression model (1)-(4) on a compact domain $\Omega \subset \mathbb{R}^4_+$, and suppose that the Assumptions 1 and 3 are satisfied. Then the following dynamic system is an exponential observer for the system (1)-(4),

$$\dot{\hat{x}} = f(\hat{x}, u) - d_1(y - \Upsilon \hat{x} - \Upsilon d_2) \tag{12}$$

where the observer gains are given by, $||d_1 \Upsilon d_2|| \cong L$, with $d_1 \in \mathbb{R}^n$, $d_2 \in \mathbb{R}^n$ and the matrix $d_1 \Upsilon$ is stable.

Proof. Let us define the estimation error as,

$$\xi = x - \hat{x}$$

Then, the dynamic equation of the estimation error is provided by,

$$\dot{\xi} = f(x,u) - f(\hat{x},u) + d_1(\Upsilon\xi - \Upsilon d_2)$$
(13)

The correction term $d_1(\Upsilon \xi - d_2)$ is viewed as an exponential measured output feedback for system (13).

Solving the equation (13), we have

$$\xi(t) = \exp(d_1\Upsilon t)\xi_0 + \int_0^t \exp(d_1\Upsilon(t-\tau))(f(x,u) - f(\hat{x},u) - d_1\Upsilon d_2)\,d\tau \qquad (14)$$

where $\xi(0) = \xi_0$.

Applying CauchySchwarz inequality to (15) and using Assumption 3,

$$\|\xi(t)\| \le \|\exp(d_1\Upsilon t)\| \|\xi_0\| + (L - \|d_1\Upsilon d_2\|) \int_0^t \exp(d_1\Upsilon (t - \tau)) d\tau \qquad (15)$$

From this, as above considered it is proposed that the tuning of the observer gains obey the following algebraic relationship,

$$\|d_1 \Upsilon d_2\| \cong L \tag{16}$$

If the above holds, therefore,

$$\|\xi(t)\| \le \|\exp(d_1\Upsilon t)\| \|\xi_0\|$$
 (17)

Since the matrix $d_1 \Upsilon$ is stable, we prove that system (12) is an exponential observer for the system (11).

If $t \to \infty$, we obtain,

$$\lim_{t \to \infty} \|\xi(t)\| = 0$$

Hence, the state estimation is achieved for global initial conditions of the states.

5. RESULTS AND DISCUSSION

Numerical simulations were performed in Matlab[®] with the numerical routine ode113 solver. The performance of the proposed observer was compared with a nonlinear observer (with correction term equal to $d_1 \tanh(\Upsilon \xi)$), which has shown a good performance applied to biological systems [24]. The initial conditions and parameter values for the gene-expression model and both observers used are shown in Table 1 and Table 2, respectively.

Initial conditions	Gene expression model	State Observers
mRNA concentration	500	250
Protein 1 concentration	750	350
Protein 2 concentration	1500	1000
Protein 3 concentration	2250	3000

Table 1: Initial conditions for the gene-expression model and both observers proved.

Table 2: Parameter values for the gene-expression model and both observers proved.

Model parameters	Gene expression model	State Observers
w	$4.16 \times 10^4 \text{ s}^{-1}$	$3.53 \times 10^4 \text{ s}^{-1}$
K_0	500	500
k	$6.66\times 10^{-3}~{\rm s}^{-1}$	$6.66 \times 10^{-3} \text{ s}^{-1}$
v	$0.033 \ {\rm s}^{-1}$	$0.3 \ {\rm s}^{-1}$
k_{12}	$3.33 \times 10^{-3} \text{ s}^{-1}$	$3.83 \times 10^{-3} \ { m s}^{-1}$
k_1	$3.33 \times 10^{-3} \text{ s}^{-1}$	$3.33 \times 10^{-3} \text{ s}^{-1}$
k_{23}	$3.33 \times 10^{-3} \text{ s}^{-1}$	$3.66 \times 10^{-3} \text{ s}^{-1}$
k_3	$3.33 \times 10^{-3} \text{ s}^{-1}$	$3.33 \times 10^{-3} \text{ s}^{-1}$
m	6	6

Figure 1 shows the time series of mRNA, protein 1, protein 2, and protein 3, for the nominal model and both observers. The observers began at the same initial conditions that are different to those used in the nominal model (see Table 1).

Also, the observer gains employed have the same value (see Table 3), where the observer gains $d_1 \in \mathbb{R}^4$ and $d_2 \in \mathbb{R}^4$ are defined as

$$d_1 = [d_{1mRNA}, d_{1Protein 1}, d_{1Protein 2}, d_{1Protein 3}]^{\prime}$$

and

$$d_2 = [d_{2mRNA}, d_{2Protein 1}, d_{2Protein 2}, d_{2Protein 3}]'$$

The estimated dynamic given by the proposed observer rapidly caught up with the real dynamics, in contrast to the hyperbolic tangent observer. This observer took longer to converge to the real dynamics and also showed an overshoot. Additionally, Figure 2 and Figure 3 present a visualization of the real and estimated trajectories as a 3D phase portrait, where the differences between the proposed and the hyperbolic tangent observers are more evident. It is important to stress here the fast and accurate



Figure 1: State estimation.

Gain values	Proposed observer	Hyperbolic tangent observer
d_{2mRNA}	10	—
$d_{2Protein \ 1}$	1	—
$d_{2Protein 2}$	-0.01	—
$d_{2Protein 3}$	-13	—
d_{1mRNA}	10	10
$d_{1Protein \ 1}$	1	1
$d_{1Protein 2}$	-0.01	-0.01
$d_{1Protein \ 3}$	13	13

Table 3: Gain values for both observers proved.

convergence of the proposed observer.

Furthermore, Figure 4 and Figure 5 show the error dynamics of each variable state. As expected, error diminishes quickly and is closer to zero with the proposed observer than with the hyperbolic tangent observer. Finally, the Figure 6 presents the global error dynamics. It is clear that the performance of the new observer is superior to the hyperbolic tangent observer.

6. CONCLUSIONS

In this work, the estimation of protein concentrations was possible in an observablegene expression model, using the mRNA as the only measurable output. The software sensor designed for this task is robust to both uncertainties in the model parame-



Figure 2: Phase portrait of the state estimation.



Figure 3: Phase portrait of the state estimation.

ters and noisy output measurements, which are the current challenges in the state estimation research. Furthermore, the proposed observer has demonstrated a finer performance compared to a hyperbolic tangent observer, a recent nonlinear observer reported in literature.

NOMENCLATURE

- N is the mRNA, mgL⁻¹, the measured output
- n_1 precursor protein 1, mgL⁻¹



Figure 4: Estimation errors.

Figure 5: Estimation errors.

- n_2 precursor protein 2, mgL⁻¹
- n_3 active protein, mgL⁻¹
- w is the transcription rate, mgL⁻¹h⁻¹
- K is the constant of association of n_3 with a regulatory site, mgL⁻¹
- m is the number of regulatory sites
- v is the translation rate constant, mgL⁻¹h⁻¹

 $k_{1,2}, k_{2,3}$ are the conversion rate constants, mgL⁻¹h⁻¹

Figure 6: Global estimation errors.

- k, k_1, k_3 are the degradation rate constants, mgL⁻¹h⁻¹
 - d_1, d_2 observer gains
 - ξ is the estimation error
 - Υ system output vector

REFERENCES

- M. Quach, N. Brunel, F. d'Alche-Buc, Estimating parameters and hidden variables in non-linear state-space models based on odes for biological networks inference, *Bioinformatics*, 23 (2007), 3209.
- [2] A.S. Ribeiro, Stochastic and delayed stochastic models of gene expression and regulation, *Mathematical Biosciences*, **223** (2010), 1.
- [3] R.N. Gutenkunst, J.J. Waterfall, F.P. Casey, K.S. Brown, C.R. Myers, J.P. Sethna, Universally sloppy parameter sensitivities in systems biology models, *Plos. Comput. Biol.*, 3 (2007), 1871.
- [4] G. Goffaux, A. Vande Wouwer, Bioprocess state estimation: Some classical and less classical approaches, *Lect. Notes Contr. Inf.*, **322** (2005), 111.
- [5] S. Bennett, A brief history of automatic control, *IEEE Contr. Syst. Mag.*, 16 (1996), 17.
- [6] G. Besançon, Nonlinear observers and applications, Springer, New York (2007).

- [7] F. Cacace, A. Germani, P. Palumbo, The state observer as a tool for the estimation of gene expression, J. Math. Anal. Appl., 391 (2012), 382.
- [8] H. Nijmeijer, T.I. Fossen, New directions in nonlinear observer design, Springer, New York (1999).
- [9] G. Lillacci, P. Valigi, State observers for the estimation of mRNA and protein dynamics, In: IEEE Life Science Systems and Applications Workshop, 2007, pp. 108.
- [10] V. Ibarra-Junquera, L.A. Torres, H.C. Rosu, G. Arguello, J. Collado-Vides, Nonlinear software sensor for monitoring genetic regulation processes with noise and modeling errors. *Phys. Rev. E*, **72** (2005), 011919 1-9.
- [11] H. Trinh, T. Fernando, Functional observers for dynamical systems, Springer, Berlin (2012).
- [12] V.P. Zhdanov, Periodic perturbation of genetic oscillations, Chaos Solitons Fractals, 45 (2012), 577.
- [13] G.R. Gavalas, Nonlinear differential equations of chemically reacting systems, Springer, Berlin (1968).
- [14] F. Viel, F. Jadot, G. Bastin, Robust feedback stabilization of chemical reactors, *IEEE Trans. Automat. Contr.*, 42 (1997), 473.
- [15] A. Ilchmann, M.F. Weirig, Modelling of general biotechnological processes, Math. Comp. Model. Dyn., 5 (1999), 152.
- [16] J. Alvarez-Ramirez, A. Morales, PI control of continuously stirred tank reactors: Stability and performance, *Chem. Eng. Sci.*, 55 (2000), 5497.
- [17] H.K. Khalil, *Nonlinear systems*, Pearson New International Edition (2013).
- [18] V.P. Zhdanov, Kinetic models of gene expression including non-coding RNAs. *Phys. Rep.*, **500** (2011), 1.
- [19] E.D. Sontag, A concept of local observability, Syst. Control Lett., 5 (1984), 41.
- [20] S. Diop, M. Fliess, Nonlinear observability, identifiability, and persistent trajectories, In: IEEE Conference on Decision and Control, 1991, pp. 714.
- [21] R. Martinez-Guerra, J. De Leon-Morales, Nonlinear estimators: A differential algebraic approach, Appl. Math. Lett., 9 (1996), 21.
- [22] R. Martinez-Guerra, R. Aguilar-Lopez, A. Poznyak, A new robust sliding-mode observer design for monitoring in chemical reactors, J. Dyn. Syst.-T Asme, 126 (2004), 473.
- [23] R. Aguilar-Lopez, J. Mata-Machuca, R. Martinez-Guerra, On the observability for a class of nonlinear (bio)chemical systems, *Int. J. Chem. React. Eng.*, 8 (2010), A3 1-27.

[24] M.I. Neria-Gonzalez, J.C. Figueroa-Estrada, M.R. Cruz-Diaz, R. Aguilar-Lopez, Adaptive smooth observer design for state estimation in Desulfovibrio alaskensis 6SR cultures, *Rev. Mex. Ing. Quim.*, **10** (2011), 137.