# A Symbolic Approach to Modeling Cellular Behavior

Bhubaneswar Mishra

[1] Courant Institute, New York University,
251 Mercer Street,
New York, NY 10012, USA,
mishra@nyu.edu,
WWW home page: http://www.cs.nyu.edu/cs/faculty/mishra/
[2] Watson School of Biological Sciences, Cold Spring Harbor Laboratory,
Demerec Building, 1 Bungtown Road,
Cold Spring Harbor, NY 11724, USA

**Abstract.** The author examines the connection between classical differential algebra of Ritt and Kolchin and differential algebraic models of biochemical systems–in particular, the models generated by S-system of Savageau. Several open problems of both biological and mathematical significance are proposed.

## 1 Introduction

Unprecedented advances in genomics have made it possible for the first time for a biologist to access enormous amounts of information at the genomic level for a number of organisms, including human, mouse, arabidopsis, fruit fly, yeast and *E. coli*. These developments are at the heart of the many renewed ambitious attempts by the biologists to understand the functional roles of a group of genes using powerful computational algorithms and high-throughput microbiological protocols. The freshly emerging field of systems biology and its sister field of bioinformatics focuses on creating a finely detailed picture of biology at the cellular level by combining the *part-lists* (e.g., genes, regulatory sequences, and other objects from an annotated genome), with the observations of transcriptional states of a cell (using Microarrays) and translational states of the cell (using new proteomic tools). In the process it has become self-evident that the mathematical foundations of these systems need to be explored exhaustively and accurately. In this paper, we describe the basic structure of the underlying differential-algebraic system and the mathematical and computational problem they naturally lead to.

### 1.1 Outline

1. **S-systems:** Section two gives a short biological introduction and then describes Savageau-Voit approach to model bio-chemical reactions based on S-systems. (See [11, 10].)

2. **Canonical Forms:** Section three provides a canonical description of an S-system in terms of a system of differential binomial equations and set of linear equality constraints. This formulation suggests that a biological system can always be described as a differential system evolving on a linear subspace of a high-dimensional embedding space.
3. **Differential Algebra:** Section four describes the elimination theoretic approaches from Ritt-Kolchin differential algebra that can be used in this context to understand the input-output behavior of a bio-chemical system.
4. **Open questions:** The paper concludes with a short description of the open questions.

## 2 S-Systems

The genome of an organism is a collection of its genes, encoded by four chemical *bases* in its DNA (Deoxyribo Nucleic Acid), and forms the genetic core of a cell. The genes ultimately encode for the protein (a chain of amino acids) and in turn, the genes are regulated by transcription factors and other operons, many of which are proteins. The sequence of amino acids, specified by the DNA through transcription and translation processes, determine the three dimensional structure and biochemical properties of the proteins as well as the nature of their interactions. Furthermore, mRNA stability, protein degradation, post-translational modifications and many other bio-chemical processes tightly regulate the time-constants involved in the resulting bio-chemical machinery. Proteins also associate in complexes to form *dimers* (pair of proteins), *trimers* (triplets) and *multimers*. An *isoform* of a protein is a slightly different protein with closely related sequence, and often share similar functional properties, e.g., enzymatic reactions, but are regulated differently.

An *enzyme*, $E$, is a protein which can enhance the activity of a chemical reaction by attaching to a *substrate*, $A$, and making the formation of the *product*, $P$, energetically easier.

$$E + A \leftrightarrows EA \to E + P.$$

In general, equations of these kind take the form

$$A + B \rightleftharpoons^{K_+}_{K_-} C + D,$$

and the rate of change of $A$'s concentration is given by the difference of the "synthesis rate" ($K_-[C][D]$) and the "degradation rate" ($K_+[A][B]$).

$$\frac{d[A]}{dt} = K_-[C][D] - K_+[A][B]. \tag{1}$$

Using a system of first order differential equations (in explicit form), one can construct a general model of a rather complex biochemical reaction involving many genes and proteins. One such model is Savageau-Voit S-system, whose ingredients are $n$ dependent variables, denoted $X_1, \ldots, X_n$ and $m$ independent variables $X_{n+1}, \ldots, X_m$ with $D_1, \ldots, D_{n+m}$ being the domains where these $n +$

$m$ variables take value. In addition the differential equations may need to be constrained by algebraic equations corresponding to stoichiometric constraints, or conserved rates for concentrations.

The basic differential equations of the system are of the form:

$$\dot{X}_i(t) = V_i^+(X_1(t), \ldots, X_{n+m}(t)) - V_i^-(X_1(t), \ldots, X_{n+m}(t)), \tag{2}$$

for each dependent variable $X_i$ (see [11]). The functions $V^+$ and $V^-$ are arbitrary rational functions over $\mathbb{R}$. The set of algebraic constraints take the form

$$\{C_j(X_1(t), \ldots, X_{n+m}(t)) = 0\} \tag{3}$$

## 3 Canonical Forms

However, one can rewrite (recast) the system of equations as the one shown above in a much more simpler manner. We show that every such system admits a canonical form involving first order ordinary differential equations with binomial terms and linear constraints.

**Theorem 1** *Every bio-chemical system arising from an S-system model can be expressed in a canonical form involving $r > n + m$ variables $Z_1$, $Z_2$, …, $Z_r$:*

$$\begin{bmatrix} \dot{Z}_1 \\ \dot{Z}_2 \\ \vdots \\ \dot{Z}_r \end{bmatrix} = \begin{bmatrix} m_1^+(\mathbf{Z}) - m_1^-(\mathbf{Z}) \\ m_2^+(\mathbf{Z}) - m_2^-(\mathbf{Z}) \\ \vdots \\ m_r^+(\mathbf{Z}) - m_r^-(\mathbf{Z}) \end{bmatrix}, \tag{4}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{sr} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{5}$$

*where $m_i^+$'s and $m_i^-$'s are ratios of monomials and $a_{ij}$'s are constants in $\mathbb{R}[Z_1$, …, $Z_r]$ with positive coefficients.*

PROOF:

Starting from the original description, one can derive a description in the canonical form by repeated applications of the following rules:

1. Assume that an equation is given as

$$\dot{X}(t) = \frac{p(X(t))}{q(X(t))},$$

where the right hand side of the explicit form is a rational function.

$$\begin{aligned} p &= \alpha_1 m_1^+ + \cdots + \alpha_k m_k^+ - \beta_1 m_1^- + \cdots + \beta_l m_l^- \\ q &= \alpha_1' m_1'^+ + \cdots + \alpha_k' m_{k'}'^+ - \beta_1' m_1'^- + \cdots + \beta_l' m_{l'}'^-, \end{aligned}$$

where $m^+$'s and $m^-$'s are power-products with arbitrary powers and positive valued coefficients $\alpha$'s and $\beta$'s.
Replace the above equation by the following system:

$$\begin{aligned} \dot{X} &= p(X(t))y(t)^{-1} \\ \dot{c_1} &= q(X(t)) - y(t)^{-1} \\ c_1 &= 0. \end{aligned}$$

2. An algebraic constraint of the form

$$r(X(t)) = \gamma_1 m_1 + \cdots + \gamma_k m_k = 0,$$

is replaced by

$$\begin{aligned} \dot{c_2} &= r(X(t)) \\ c_2 &= 0. \end{aligned}$$

3. Finally, an equation of the form

$$\begin{aligned} \dot{X}(t) &= \alpha_1 m_1^+ + \cdots + \alpha_k m_k^+ - \beta_1 m_1^- - \cdots - \beta_l m_l^+ \\ &= [\alpha_1 m_1^+ - (1/k)W(t)] + \cdots + [\alpha_k m_k^+ - (1/k)W(t)] \\ &\quad - [\beta_1 m_1^- - (1/l)W(t)] - \cdots - [\beta_l m_l^+ - (1/l)W(t)] \end{aligned}$$

is replaced by

$$\begin{aligned} \dot{\Gamma}_i(t) &= \alpha_i m_i^+ - (1/k)W(t), \qquad 1 \le i \le k, \\ \dot{\Gamma}_i(t) &= \beta_i m_i^- - (1/l)W(t), \qquad k+1 \le i \le k+l, \\ X(t) - \Gamma_1(t) - \cdots - \Gamma_k(t) + \Gamma_{k+1}(t) + \cdots + \Gamma_{k+l}(t) &= 0. \end{aligned}$$

Repeated applications of these three rules to any S-system of equations not in the canonical form terminates after finitely many steps and results in the desired final canonical form. $\Xi$

## 4 Differential Algebra

The semantics for a bio-chemical reaction then can be given by the evolution equations in the explicit form, or more geometrically, by the *trajectory* semantics where all possible evolution paths of the system are explicitly represented. A more compact geometric picture can be given in terms of the *distributions*, e.g., the *classical phase portraits* represented as a vector field. For instance, a simple model of a circadian clock can be represented in terms of the mRNA level of *per*, $M$, and corresponding protein levels of PER at various degrees of phosphorylation, $P_0$, $P_1$ and $P_2$, and in terms of its location inside the nucleus

$P_N$ or cytoplasm:

$$\begin{aligned}
\dot{M} &= v_s \frac{K_1^n}{(K_1^n + P_N^n)} - v_m \frac{M}{(K_m + M)} \\
\dot{P_0} &= k_s M - V_1 \frac{P_0}{(K_1 + P_0)} + V_2 \frac{P_1}{(K_2 + P_1)} \\
\dot{P_1} &= V_1 \frac{P_0}{(K_1 + P_0)} - V_2 \frac{P_1}{(K_2 + P_1)} - V_3 \frac{P_1}{(K_3 + P_1)} + V_4 \frac{P_2}{(K_4 + P_2)} \\
\dot{P_2} &= V_3 \frac{P_1}{(K_3 + P_1)} - V_4 \frac{P_2}{(K_4 + P_2)} - k_1 P_2 + k_2 P_N - v_d \frac{P_2}{(k_d + P_2)} \\
\dot{P_N} &= k_1 P_2 - k_2 P_N \\
P_t &= P_0 + P_1 + P_2 + P_N
\end{aligned}$$

Its phase portrait then can be analyzed to determine if the system has a stable and robust limit-cycle (e.g., by applying "Bendixon criteria," etc.).

Another approach is to describe the system in terms of an automaton, whose state can be represented as a finite-dimensional vector $S(t)$ and its transition from $S(t)$ and $S(t + \Delta t)$ can be determined by following the trajectory starting at state $S(t)$:

$$\int_t^{t+\Delta t} F(S(\tau))\, d\tau,$$

subject to the constraints on the system. Wherever an appropriate numerical integrator is available, such an automata can be numerically described by the "traces" of the numerical integrator. In order to keep the complexity of such an automaton simple one can obtain "approximate versions" of the automaton by discretization and collapse operations that hide all or some of the "internal states."

An ultimate example of collapsing involves hiding all the internal state variables and just describing the evolution of outputs in terms of its input. Here, one describes the system in terms of its *input-output relation* that describes only the relation between the control inputs and the output variables starting from a redundant state-space description. From an algebraic point of view, this is exactly the problem of *variable elimination* and comes under the subject of *elimination theory*. Thus all the theories related to *standard bases*, *characteristic sets* and *differential-algebraic resultants* play important roles.

Assume that the system (SISO) is described as shown below:

$$\begin{aligned}
\dot{x}_1 &= p_1(X, u, \dot{u}, \ldots, u^{(k)}) \\
&\vdots \\
\dot{x}_r &= p_r(X, u, \dot{u}, \ldots, u^{(k)}) \\
0 &= q_1(X, u) \\
&\vdots
\end{aligned}$$

$$0 = q_s(X, u)$$
$$y = h(X, u)$$

Consider the following differential ideal $I$ in the differential ring $\mathbb{R}\{X, u, y\}$:

$$I = [\dot{x}_1 - p_1, \ldots, \dot{x}_r - p_r, q_1, \ldots, q_s, y - h].$$

The input-output relation is then obtained by finding the contraction $I^c$ of the ideal $I$ to the ring $\mathbb{R}\{u, y\}$. The generators of $I^c = I \cap \mathbb{R}\{u, y\}$ give the differential polynomials involving $u$ and $y$. However, the underlying algorithmic questions for differential algebraic elimination remain largely unsolved.

**Example** Consider the following system (adapted from Forsman [4]):

$$A \to B,$$

with the following kinetic equations:

$$\dot{[B]} = [A]^{0.5} - [B]^{0.5}.$$

The input $u$ controls the concentration $[A]$ as follows:

$$\dot{[A]} = u[A]^{-2} - [A]^{-1.5},$$

and the output $y$ is simply $[B]$:

$$y = [B].$$

We can simplify the above system to a polynomial system by following transformations:

$$x_1^2 = [A] \quad \text{and} \quad x_2^2 = [B].$$

Thus,

$$I = [2x_1^5\dot{x}_1 + x_1 - u, 2x_2\dot{x}_2 + x_2 - x_1, x_2^2 - y].$$

After eliminating $x_1$ and $x_2$, we obtain the following input-output relation:

$$\begin{aligned}
&(20\dot{y}^8y^2 - 4\dot{y}^{10}y - 40\dot{y}^6y^3 + 40\dot{y}^4y^4 - 20\dot{y}^2y^5 + 4y^6)\ddot{y}^2 \\
&+ (4u\dot{y}^5y - 4\dot{y}^6y - 20\dot{y}^4y^2 + 40u\dot{y}^3y^2 + 20\dot{y}^2y^3 + 20u\dot{y}y^3 + 4y^4)\ddot{y} \\
&- \dot{y}^2y^5 + 5\dot{y}^4y^4 - 10\dot{y}^6y^3 + 20u\dot{y}^3y^2 + 10\dot{y}^8y^2 + y^2 - 8\dot{y}^6y + 10u\dot{y}^5y \\
&- u^2y + 2u\dot{y}y - \dot{y}^2y - 5\dot{y}^{10}y + \dot{y}^{12} + 8\dot{y}^2y^3 + 2u\dot{y}y^3 = 0. \quad \Xi
\end{aligned}$$

## 5 Open Questions

Several interesting questions remain to be further explored.

1. **Reactions Models:** We have primarily focused on a simple ODE model (Differential Algebraic Equations, DAE) and narrowed this even further to a model based on S-systems. Does this imply that there is a significant deviation from reality? How can a stochastic model representing small number of molecules interacting pair-wise and randomly be incorporated?

2. **Hybrid Systems:** Certain interactions are purely discrete and after each such interaction, the system dynamics may change. Such a hybrid model implies that the underlying automaton must be modified for each such mode. How do these enhancements modify the basic symbolic model?
3. **Spatial Models:** The cellular interactions are highly specific to their spatial locations within the cell. How can these be modeled with symbolic cellular-automata? How can we account for dynamics due to changes to the cell volume? The time constants associated with the diffusion may vary from location to location; how can that be modeled?
4. **State Space (Product Space):** A number of interacting cells can be modeled by product automata. In addition to the classical "state-explosion problem" we also need to pay attention to the variable structure due to i) Cell division, ii) Apoptosis and iii) Differentiation.
5. **Communication:** How do we model the communication among the cells mediated by the interactions among the extra-cellular factor and external receptor pairs?
6. **Hierarchical Models:** Finally, as we go to more and more complex cellular processes, a clear understanding can only be obtained through modularized hierarchical models. What are the ideal hierarchical models? How do we model a population of cells with related statistics?
7. **Simulation :** If a biologist wishes to obtain a visualization based on numerical simulation, how can we take advantage of the underlying symbolic description?
8. **Symbolic Verification:** If a biologist wishes to reason about the system with logical queries in an appropriate query language (e.g., temporal logic), what are the best query languages? What are the best algorithms that take advantage of the symbolic structures? What are the correct way to solve problems associated with i) Model Equivalence, ii) Experimental Analysis, and iii) Reachability Analysis?

## References

[1] Brockett, R.: Nonlinear Systems and Differential Geometry. Proceedings of the IEEE, **64** (1976): 61–72.

[2] Carrá Ferro, G.: Gröbner Bases and Differential Ideals. Proceedings of AAECC-5, Lecture Notes in Computer Science, Springer-Verlag, (1987): 129–140.

[3] Diop, S.: Elimination in Control Theory. Math. Control Signals Systems, **4** (1991): 17–32.

[4] Forsman, K.: *Constructive Commutative Algebra in Nonlinear Control Theory*, Linköping Studies in Science and Technology, Dissertation, No. 261, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 1992.

[5] Gallo, G., Mishra, B., and Ollivier, F.: Some Constructions in Rings of Differential Polynomials. Proceedings of AAECC-9, Lecture Notes in Computer Science, Springer-Verlag, **539** (1991): 171–182.

[6] Kolchin, E.R.: On the Basis Theorem for Differential Systems. Transactions of the AMS, **52** (1942): 115–127.

[7] Mishra, B.: Computational Differential Algebra. Geometrical Foundations of Robotics, (Ed. Jon Selig), World-Scientific, Singapore, **Lecture 8** (2000): 111-145.

[8] Ritt, J.F.: *Differential Equations from the Algebraic Standpoint*, AMS Coloq. Publ. 14, New York 1932.

[9] Seidenberg, A.: An Elimination Theory for Differential Algebra. University of California, Berkeley, Publications in Mathematics, **3** (1956): 31–65.

[10] Savageau, M.A.: *Biochemical System Analysis: A Study of Function and Design in Molecular Biology* . Addison-Wesley, 1976.

[11] Voit, E.O.: *Computational Analysis of Biochemical Systems*. Cambridge, 2000.