

# An algebraic approach to challenges on identification problems in systems biology

Mizuka Komatsu

Graduate School of System Informatics, Kobe University

Takaharu Yaguchi

Graduate School of System Informatics, Kobe University

## Abstract

This paper is twofold. A challenge appearing in identification problems, in particular unidentifiability problems in parameter estimations, in the field of systems biology is introduced with an example of PBPK models in the first section. In the second section, an algebraic and algebro-geometric approach to the challenge in a wider context is explained. More precisely, a method to extract a geometric structure that is uniquely determined by observed time series data and unidentifiable state-space models as an algebraic variety is introduced. Our method is based on differential algebra. An application of the proposed method for analysis of viral dynamics during therapy is briefly described.

## 1 Introduction of challenges in identification problems in systems biology

In the field of systems biology, biological phenomena and their interactions are analysed integrately. In such a field, biological systems are often modelled as and investigated through mathematical models. One of common modelling approaches is by ordinary differential equations, ODEs, which are constructed based on biological knowledge. Partial differential equations, PDEs, are also one of the major modelling approaches in such a field. However, considering possible constructions of PDEs based on refining compartment models, a class of ODEs, analytical methods for ODE models are considered to be still useful for analysis of PDEs. Based on this, our study focuses on ODE models appearing in systems biology. In particular, we consider such models with unknown parameters to be estimated using observed time-series data, etc. Parameter estimation problems of such models, which constitute a type of identification problems of ODEs, are the main topic in this study.

In order to illustrate challenges in parameter estimation problems in systems biology, an example based on physiologically-based pharmacokinetic models, so-called PBPK models, is introduced. Roughly speaking, PBPK models describe drug responses in human bodies dynamically [1]. In the field of pharmacokinetics, such models are constructed and practically used for predicting the time course of the concentration of drugs in plasma and other sites. Although PBPK models have their origin in such a field, their applications can be seen in other fields, e.g., [2], so we newly applied the model to immunology. A PBPK model is constructed as compartments, each of which corresponds to an organ or a compound of organs. State variables of a PBPK model correspond to the amounts of the items of interest, for example, drugs in compartments. Some of the model parameters are assigned to the meanings of certain biological functions such as clearance rates at organs. Thus, not only the behaviours of state variables but also parameter values are of importance, and these tend to be unknown. Furthermore, they are considered to be different depending on the subject, for example, by age group. Using relevant time-series data, e.g., drug concentrations in plasma measured at several time points from subjects, such parameters, i.e., biological functions, are quantitatively estimated for the corresponding subjects. In such estimations,

there exists an intrinsic challenge: the parameters tend to be unidentifiable. In other words, parameters cannot be uniquely determined from given data due to their insufficiency [3]. Unidentifiability of parameters causes difficulties in investigation of corresponding biological systems. For example, in the context of immunology, such situation makes detection of immunological abnormalities from estimated parameters difficult in spite of the desire. Besides, estimations conducted without considering the unidentifiable property may overlook possible important considerations for the systems, as we pointed out in [4]. Due to experimental constraints, unidentifiability problems often appear in systems biology, suggesting the importance of approaches to dealing with unidentifiable models.

## 2 Algebraic approaches to unidentifiable state-space models: the parameter variety

Based on the previous section, we consider an approach to unidentifiable models, of which the underlying theory is differential algebra [5]. Roughly speaking, differential algebra is an algebraic framework in which differentiations are allowed as operations. See, e.g., [5, 6] for details and, e.g., [7] for its applications and references therein.

Here, we deal with the following state-space models:

$$\frac{dx}{dt} = f(x, u; a), \quad (1)$$

$$y = g(x, u; a), \quad (2)$$

where  $x(t) \in \mathbb{R}^N$  is the state variable vector,  $u(t) \in \mathbb{R}^M$  is the input vector,  $y(t) \in \mathbb{R}$  is the output, and  $a \in \mathbb{R}^n$  is the unknown parameter vector, where  $N, M, n$  are positive integers.  $u(t)$  and  $y(t)$  are observed. The coefficients of  $f(x, u)$  and  $g(x, u)$  are assumed to be rational functions of  $a$ . Here, we consider the models with polynomials  $f$  and  $g$  of  $x$  and  $u$ . (1) denotes the mathematical model that describes the system under consideration, for example, a PBPK model mentioned in the previous section. (2) denotes an observation model, which describes a data acquisition. In this paper, (1) and (2) are called unidentifiable if for any input  $u(t)$ , there exist two parameters  $a_1 \neq a_2$  in a parameter space such that  $y(u; a_1) = y(u; a_2)$  holds, where  $y(u; a)$  denotes the output of (1) and (2) with  $u$  applied as input given  $a$ . In our method, we focus on the fact that sets of all the possible parameters fitting observed data form algebraic varieties, which we call the parameter varieties, and then propose a method to describe them explicitly.

In control theory, the transfer functions of linear state-space models are derived through methods such as the Laplace transformations. Through these functions, the input-output behaviour of the models can be investigated. This is possible because they do not contain information on the state variables. To investigate the input-output relations of (1) and (2), we apply algebraic techniques, and thus, eliminate the state variables. Technically speaking, the models considered in this paper may not be able to be dealt with by commutative algebra naively since they contain derivatives of variables with respect to time. In fact, in order to derive input-output relationships from such models, derivative operations, which are not allowed in commutative algebra, for the models are required. More precisely, we first consider the differential ideal generated by (1) and (2) in the differential polynomial ring whose field is the rational functions of  $a$  and variables are  $x, u, y$  and their derivatives. Then, we investigate the intersection of the differential ideal and the differential polynomial ring whose field is the rational functions of  $a$  and variables are  $u, y$  and their derivatives, which contains differential polynomials representing the input-output relationships of the model. In this way, our method appears to fall under the umbrella of differential algebra, which is rather difficult compared to the non-differential one. In general, one of the difficulties regarding differential algebra is the fact that the differential polynomial ring is non-Noetherian. This implies that the differential ideal in such a ring, for example, the intersection that we consider, may not be finitely generated. However, thanks to state-space representations of (1) and (2), it is guaranteed that essential polynomials representing the input-output relationships can be derived considering certain truncated differential ideals generated by the models. Such an ideal can be regarded as a non-differential ideal in the non-differential polynomial ring where the derivatives of variables are regarded as other

variables, which is Noetherian. The details of the way of truncation of the differential ideal are described in [4].

Once the truncated ideal for (1) and (2) is considered, the intersection of the ideal and the polynomial ring of which the field is the rational functions of  $a$  and the variables are  $u, y$  and their derivatives up to a finite order is our interest. In the proposed method, such an intersection is described using the Gröbner basis for the truncated ideal at first. A subset of the Gröbner basis that does not contain state variables and their derivatives describes the intersection representing the input-output relations of the model thanks to its elimination property [8]. Then, by introducing observed data into the subset, the sets of parameters, each of which generates the given data, are described as sets of constraints in terms of parameters, i.e., the algebraic varieties. Once the structure, i.e., the variety, is extracted, overlooking the feasible parameters, which may lead to insufficient or inappropriate system considerations, would never occur [4]. See [4] for details of the proposed method.

We applied our method in the analysis of viral dynamics, which reveals an important fact about the efficacy of a drug that was missed in a previous study [9]. In [9], the viral dynamics under a drug therapy, which is described as (3) and illustrated in Figure 1, is investigated.

$$\frac{dx_1}{dt} = \frac{a_1 a_3}{a_4} x_2 - a_1 x_1, \quad \frac{dx_2}{dt} (1 - a_2) a_4 x_2 - a_3 x_3, \quad y = x_2 \quad (3)$$

Figure 2 shows examples of estimated parameter varieties in parameter spaces of an unidentifiable model that appeared in [9] given observed time-series data taken from two different subjects. See [4] for the details of the varieties. As can be seen in Figure 2, our method captures feasible parameters exhaustively, unlike conventional approaches [9], suggesting an applicability of our method to, for example, classifications of subjects in terms of their varieties.

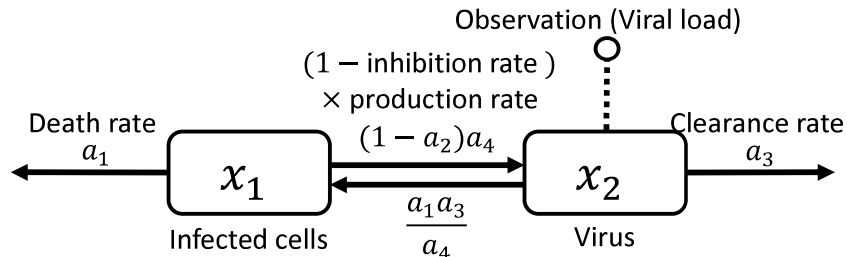


Figure 1: The schematic representation of the viral dynamics under the drug therapy described by (3).  $x_1$  denotes the number of productively infected cells.  $x_2$  denotes the viral load, which is observed. The infected cells die at a constant rate  $a_1$  per cell per day and produce the virus at an average rate  $a_4$ . Virions are cleared at a constant clearance rate  $a_3$ . The drug reduces the production of virions from the infected cells by a fraction  $1 - a_2$ , where  $0 \leq a_2 \leq 1$ . See [9] for details.

## Acknowledgements

This work was supported by JST CREST JPMJCR1914 and JSPS KAKENHI Grant Numbers JP20J21185 and 20K11693.

## References

- [1] S. A. Peters, “Physiologically-based pharmacokinetic(PBPK) modeling and simulations: principles, methods, and applications in the pharmaceutical industry,” *Wiley* (2011).
- [2] M. Komatsu and T. Yaguchi, “Mathematical modeling of kinetics of antigens and antibodies based on experiments toward analysis of the mechanism of allergies,” *Transactions of the Japan Society for Industrial and Applied Mathematics*, Vol. 28, pp. 162–204 (2018).

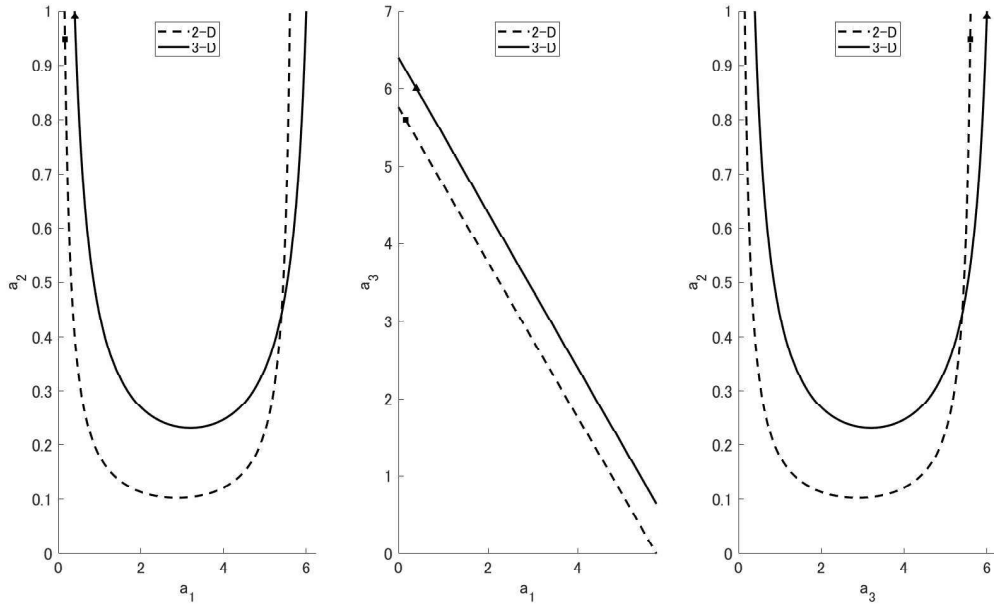


Figure 2: The parameter varieties of an unidentifiable model in two-dimensional parameter spaces that describes viral dynamics [9] given time-series data of viral load.  $2-D$  and  $3-D$  are labels that are assigned to subjects from which temporal measurements are taken.  $a_1, a_2, a_3$  are parameters in the model. The symbols on the varieties correspond to estimated parameters in the previous study [9].

- [3] H. Miao, X. Xia, A. S. Perelson and H. Wu, “On identifiability of nonlinear ODE models and applications in viral dynamics,” *SIAM Rev. Soc. Ind. Appl. Math.*, vol. 53, pp. 3–39 (2011).
- [4] M. Komatsu and T. Yaguchi, “Method for estimating hidden structures determined by unidentifiable state-space models and time-series data based on the Gröbner basis,” *arXiv*.
- [5] J. F. Ritt, “*Differential algebra*,” New York: AMS Colloquium Publications (1950).
- [6] W. Y. Sit, “The Ritt-Kolchin theory for differential polynomials,” *Differential Algebra And Related Topics - Proceedings Of The International Workshop*, pp. 1–70, Singapore: World Scientific (2002).
- [7] M. Fliess and S. T. Glad “An Algebraic Approach to Linear and Nonlinear Control,” *Essays on Control: Perspectives in the Theory and its Applications*, pp. 223–267, Massachusetts: Birkhäuser Boston (1993).
- [8] D. A. Cox, J. Little, J. and D. O’Shea, “Ideals, Varieties, and Algorithms,” New York: *Springer International Publishing* (2015).
- [9] A. U. Neumann, N. P. Lam, H. Dahari, D. R. Gretch, T. E. Wiley, T. J. Layden and A. S. Perelson, “Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon- $\alpha$  therapy,” *Science*, Vol. 282, pp. 103–107 (1998).

Graduate School of System Informatics  
 Kobe University  
 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, JAPAN  
 E-mail address: m-komatsu@stu.kobe-u.ac.jp  
 E-mail address: yaguchi@pearl.kobe-u.ac.jp