Symbolic-Numeric Estimation of Kinetic Parameters in Biochemical Pathways by Quantifier Elimination

Shigeo Orii1
orii@strad.ssg.fujitsu.com

Hirokazu Anai2
anai@jp.fujitsu.com

Katsuhisa Horimoto3
khorimot@ims.u-tokyo.ac.jp

1 Science Solutions Group, FUJITSU LTD., Nakase 1-9-3, Mihama-ku, Chiba 261-8588, Japan
2 IT Core Laboratories, FUJITSU LABORATORIES LTD./JST, CREST., Kamikodanaka 4-1-1, Nakahara-ku, Kawasaki 211-8588, Japan
3 Laboratory of Biostatistics, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

Keywords: computer algebra, parameter estimation, quantifier elimination

1 Introduction

The sequencing of complete genomes allows analyses of the interactions between various biological molecules on a genomic scale, which prompted us to simulate the global behaviors of biological phenomena on the molecular level. One of the basic mathematical problems in the simulation is the parameter optimization in the kinetic model for complex dynamics, and many optimization methods have been designed. Here, we introduce a new approach to optimize the parameters in biological kinetic models by quantifier elimination (QE), in combination with numerical simulation methods [3]. The optimization method was applied to a model for the inhibition kinetics of HIV proteinase with ten parameters and nine variables [1, 2].

2 Materials and Methods

2.1 Mathematical Framework

We consider the following fitting problem. The biological kinetic model analyzed here is of the form:

\[ \dot{x}_i = v_i(X, K) \quad (i=1, \ldots, r) \] (1)

where \( X = \{x_1, \ldots, x_n\} \) is a set of variables, and \( K = \{k_1, \ldots, k_m\} \) is a set of parameters. The problem is to fit the parameters \( K \) of the model to the observed data \( \tilde{X}_{obs} = \{\tilde{x}_1^{(t)}, \ldots, \tilde{x}_n^{(t)}\} \) for \( t = 0, 1, \ldots, N \) and \( n_d \leq n \), under the following additional conditions:

(i) Conservation laws: \( h_q(X) = 0 \quad (q=1, \ldots, s) \),

(ii) Variable ranges: \( x_j \in D_j, \) where \( D_j = [a_j, b_j] \), \( a_j, b_j \in \mathbb{R} \cup \{\infty\} \) \( \quad (j=1, \ldots, n) \).

Here, we set up the leading formula of this study. As mentioned above, we have the following constraints \( \psi \) with error variables \( e_i \) from kinetic models:

\[ \psi = \bigwedge_{i} \psi_i \quad (i=1, \ldots, r), \]

where

\[ \psi_i = \dot{x}_i - v_i(X, K) + e_i = 0. \]

We remark that \( \psi \) is in fact a formula over the real algebraic field, which is a discretized version of the differential equations in (1). In \( \psi \), we regard \( \dot{x}_i \) as new variables that are not in \( X \); \( \dot{x}_i \) are substituted by the corresponding numerical values obtained by observation or numerical simulation. Hereafter, we denote the set of the variables \( \dot{x}_i \) by \( \dot{X} \). As for the error variables \( e_i \), we introduce a new variable, \( e_{\text{max}} \), which
refers to the magnitude of the error variables: $|e_i| \leq e_{\text{max}}$. Moreover, for the variables that have observed data, we consider the following objective conditions:
\[
x_l - \bar{x}_l^{(i)} = 0 \quad (l=1,\ldots,n_d),
\]
to achieve fitting. Then the “basic formula” is given as
\[
F(\mathbf{x}, \mathbf{X}, K, e_{\text{max}}, e_i) = \left( \Psi \land \bigwedge_q \left( l_q (X) = 0 \right) \land \bigwedge_l \left( x_l \in D_l \right) \land \bigwedge_i \left( e_i \leq e_{\text{max}} \right) \land \bigwedge_l \left( x_l - \bar{x}_l^{(i)} = 0 \right) \right).
\]
We apply our symbolic-numeric approach to formulas derived by slightly modifying the basic formula according to various purposes.

2.2 Procedure
The procedure of symbolic-numeric optimization consists of the following six parts: Numerical simulation, Formulation, Computation of offset by QE, Estimation of $e_{\text{max}}$ and $K_4$ by QE, Computation of sum of squares ($SSq$), and Termination [3]. The details of the procedure will be described at the spot.

2.3 Kinetic Model
We analyzed a model for the inhibition kinetics of HIV proteinase [1]; there are ten parameters and nine variables. According to the previous studies [1, 2], five parameters ($k_{11}, k_{12}, k_{21}, k_{41}, k_{51}$) were given, and the remaining five unknown parameters ($k_{22}, k_3, k_{42}, k_{52}, k_6$), two initial values ($E_{\text{init}}, S_{\text{init}}$) and the offset of the fluorimeter were estimated by the present method. The experimental data of the product $[P]$ were downloaded from a web site [4].

3 Results and Discussion
The optimized parameters with the six sets of observed data are listed in Table 1, together with the iteration number, the offset, the $e_{\text{max}}$, the initial values of $[E_{\text{init}}]$ and $[S_{\text{init}}]$, and the goodness of fit measured by $SSq$. One of the remarkable features of the present fitting is that only one or two points of the observed data were sufficient to fit 300 data points with an $SSq$ value of less than 0.01. Another feature is that the values of the parameters agreed well with those in the previous studies [1, 2]. The merits and pitfalls of the present method will be discussed at the spot.

Table 1: Goodness of fit with optimized parameters by the symbolic-numeric method.

<table>
<thead>
<tr>
<th>Time</th>
<th>Iteration</th>
<th>Offset</th>
<th>$e_{\text{max}}$</th>
<th>$E_{\text{init}}$</th>
<th>$S_{\text{init}}$</th>
<th>$k_{22}$</th>
<th>$K_3$</th>
<th>$k_{42}$</th>
<th>$k_{52}$</th>
<th>$k_6$</th>
<th>$SSq$</th>
</tr>
</thead>
<tbody>
<tr>
<td>336</td>
<td>1</td>
<td>-0.02737</td>
<td>1.436E-10</td>
<td>0.00495</td>
<td>27.5</td>
<td>250.9</td>
<td>9.776</td>
<td>1296</td>
<td>0.103</td>
<td>0.0969</td>
<td>0.00895</td>
</tr>
<tr>
<td>984</td>
<td>1</td>
<td>-0.03937</td>
<td>1.069E-11</td>
<td>0.00470</td>
<td>28.0</td>
<td>162.8</td>
<td>9.980</td>
<td>1134</td>
<td>0.102</td>
<td>0.0982</td>
<td>0.00795</td>
</tr>
<tr>
<td>1848</td>
<td>1</td>
<td>-0.06337</td>
<td>1.113E-12</td>
<td>0.00390</td>
<td>29.0</td>
<td>38.67</td>
<td>9.990</td>
<td>1342</td>
<td>0.101</td>
<td>0.0986</td>
<td>0.04250</td>
</tr>
<tr>
<td>2</td>
<td>-0.03937</td>
<td>4.705E-10</td>
<td>0.00465</td>
<td>28.0</td>
<td>149.5</td>
<td>9.805</td>
<td>1241</td>
<td>0.110</td>
<td>0.0970</td>
<td>0.00835</td>
<td></td>
</tr>
<tr>
<td>336, 984</td>
<td>1</td>
<td>-0.02737</td>
<td>1.692E-02</td>
<td>0.00485</td>
<td>27.5</td>
<td>226.2</td>
<td>9.868</td>
<td>1324</td>
<td>0.100</td>
<td>0.1000</td>
<td>0.00728</td>
</tr>
<tr>
<td>336, 1848</td>
<td>1</td>
<td>-0.02737</td>
<td>4.732E-02</td>
<td>0.00475</td>
<td>27.5</td>
<td>191.8</td>
<td>9.990</td>
<td>1208</td>
<td>0.100</td>
<td>0.1000</td>
<td>0.00743</td>
</tr>
<tr>
<td>984, 1848</td>
<td>1</td>
<td>-0.03937</td>
<td>2.783E-02</td>
<td>0.00455</td>
<td>28.0</td>
<td>120.1</td>
<td>9.980</td>
<td>996.2</td>
<td>0.100</td>
<td>0.1000</td>
<td>0.000944</td>
</tr>
<tr>
<td>Mendes &amp; Kell</td>
<td>-</td>
<td>-0.008962</td>
<td>-</td>
<td>0.00547</td>
<td>26.79</td>
<td>201.1</td>
<td>7.352</td>
<td>1171</td>
<td>1.31E+04</td>
<td>3.00E+04</td>
<td>0.00513</td>
</tr>
</tbody>
</table>

For reference, the values related to the present optimization are also cited from previous studies [1, 2].

References