

# Optimal Control Formulation of Constrained Least-Square Estimation for Biochemical Pathway Estimation

Cranos Williams<sup>1,\*</sup>, Winser Alexander<sup>1</sup>, William Edmonson<sup>1</sup>

1. Department of Electrical and Computer Engineering,  
North Carolina State University, Raleigh, NC, USA

\*E-mail: cmwilli5@ncsu.edu

## Introduction

In our work, we present an optimal control formulation of a constrained-least squares estimation problem that addresses several issues inherent to biochemical pathway modeling which include limited sparse measurements, nonlinear system characteristics, and biological constraints of pathway components. We verify the functionality of this algorithm on a single-step signal transduction pathway, which serves as a chemical reaction template for larger more complex biochemical pathways. Our results suggests that efforts spents towards further development of these approaches may lead to more successful applications of pathway modeling to problems like drug design, cancer treatment, and biofuel synthesis.

## Methodology

### Single-Step Signal Transduction Pathway

Kutalik et al. present a description of a single-step signal transduction pathway in [3].



The series of reactions in (1) are comprised of four components, the substrate (S), the enzyme (E), the enzyme-substrate complex (ES), and the product (P). The system of difference equations that describe the dynamics of this pathway can be written in discrete state space form

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{p}) \quad (2)$$

where  $\mathbf{x} \in \mathcal{R}^{4 \times 1}$  represents the state vector,  $\mathbf{p} \in \mathcal{R}^{3 \times 1}$  represents the kinetic parameter vector, and  $\mathbf{f}(\cdot) : \mathcal{R}^{4 \times 1} \rightarrow \mathcal{R}^{4 \times 1}$  is a nonlinear vector function. This system has no external inputs, thus the dynamics of the system are completely described by the initial concentrations of the system components and the values of the kinetic rate parameters.

## Theory

Define the following least-squares approach to estimating the initial component concentrations  $\mathbf{x}_0$  and the kinetic parameters  $\mathbf{p}$  of the pathway:

$$\min_{\mathbf{x}_0, \mathbf{p}} J_0(\mathbf{x}_0, \mathbf{p}) = \frac{1}{2} \sum_{i \in \mathcal{I}} (\mathbf{h}(\mathbf{x}_i) - \mathbf{y}_i)^T (\mathbf{h}(\mathbf{x}_i) - \mathbf{y}_i) \quad (3)$$

subject to

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{f}(\mathbf{x}_k, \mathbf{p}), & k &= 0, \dots, N-1 \\ \mathbf{x}_k &\geq 0, & k &= 0, \dots, N-1 \end{aligned} \quad (4)$$

where  $\mathcal{I}$  is the set of time points where measurements are observed,  $\mathbf{h}(\cdot)$  models our ability to measure the system, and  $\mathbf{y}_i$  are the observed measurements. We formulate an optimal control approach to solving this minimization problem, utilizing Lagrange multipliers, the Hamiltonian of the system, and a technique used to handle explicit constraints on the states [4] [5]. This results in a two-point boundary value problem represented by a system of coupled difference equations. These coupled difference equations, along with the associated boundary conditions, were used to implement a conjugate gradient based approach to estimating the initial concentrations and kinetic parameters of the pathway.

## Analysis and Results

Our analysis poses a problem where estimates of  $\mathbf{x}_0$  and  $\mathbf{p}$  are acquired using simulated discrete-time measurements of the substrate and the product. The system was simulated using an initial concentration vector  $\mathbf{x}_0^* = [12, 12, 0, 0]^T$  and a kinetic parameter vector  $\mathbf{p}^* = [0.18, 0.02, 0.23]^T$ . The simulated discrete-time measurements were obtained at 6 uniformly spaced intervals corresponding to 0 min, 2 min, 4 min, 6 min, 8 min, and 10 min. We tested the consistency of the algorithm by executing 50 independent runs at 50 different initial guess where all initial guess were within valid biological ranges.

Table 1 summarizes the average and variance of the estimates over 50 runs. The initial concentrations for  $x_1$  and  $x_4$  are observed and assumed known and constant. The largest absolute error between true value and average estimates are on order  $O(10^{-6})$ . We see a maximum variation of order  $O(10^{-11})$  over the 50 independent runs. These results highlight the algorithm’s ability to consistently provide accurate estimates. Figure 1 shows both the true and estimated state trajectories, illustrating very little difference between the two. The maximum average relative error over all states was on order  $O(10^{-6})$  with the maximum relative error on order  $O(10^{-5})$ . Our results suggests that efforts spent towards further improvement of these estimation approaches may lead to more successful applications of pathway modeling to practical biological and environmental problems.

## References

- [1] H. Kitano, “Systems Biology: A Brief Overview,” *Science*, vol. 295, pp. 1662–1664, 2002.
- [2] —, “Computational systems biology,” *Nature*, vol. 420, pp. 206–210, 2002.

Table 1: True Values & Average Estimates

	$\mathbf{x}_0^*, \mathbf{p}^*$	$\bar{\mathbf{x}}_0, \bar{\mathbf{p}}$	$\text{Var}(\hat{\mathbf{x}}_0, \hat{\mathbf{p}})$
$x_2(0)$	12.0000000	11.9999989	$7.702117e - 11$
$x_3(0)$	$0.0000000e + 00$	$4.3500000e - 08$	$1.4552888e - 11$
$p_1$	0.1800000	0.1800000	$7.928136e - 14$
$p_2$	0.0200000	0.0200000	$5.541936e - 15$
$p_3$	0.2300000	0.2300000	$2.852789e - 15$

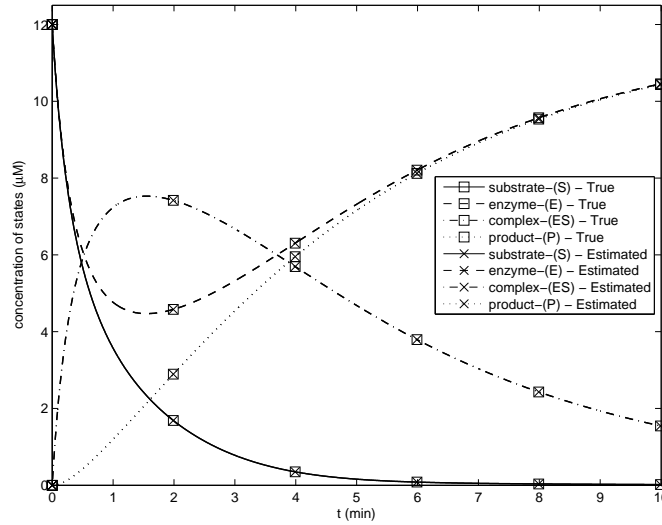


Figure 1: True vs. Estimated State Trajectories

- [3] Z. Kotalik, K. H. Cho, and O. Wolkenhauer, "Optimal sampling time selection for parameter estimation in dynamic pathway modeling," *Bio Systems*, vol. 75, pp. 43–55, 2004.
- [4] F. L. Lewis, *Optimal Control*. John Wiley & Sons, Inc., 1986.
- [5] D. E. Kirk, *Optimal Control Theory: An Introduction*. Dover Publications, Inc., 2004.