MITIGATING OVERPARAMETERIZATION OF BIOCHEMICAL REACTION NETWORKS VIA REGULARIZATION TECHNIQUES

Juergen Hahn* and Daniel Howsmon
Rensselaer Polytechnic Institute
Troy, NY 12180

Abstract

Models of biochemical reaction networks, such as signaling or metabolic networks, contain a large number of components and an even larger number of parameters associated with reaction and transport processes. While the nominal parameter values are usually sourced from the literature, they may represent kinetics in a different cell type or organism than that which is modeled. As such, parameters are often estimated from the available experimental data. However, because of the limited amount of data available and the large number of parameters, regularization is needed to avoid over fitting. A tutorial of regularization techniques including parameter set selection precedes a discussion of selected state-of-the-art procedures for estimating parameters in complex biochemical networks.

Keywords

biochemical reaction networks, regularization, parameter set selection, parameter estimation

Introduction

Biochemical reaction network models have been successfully employed for a variety of purposes, including summarizing existing experimental evidence (Moya et al. 2011), distinguishing between possible mechanisms (Sneyd et al. 2006), and reducing experimental efforts (Man et al. 2014). When models are first developed to describe a given data set, parameters are often sourced from the available literature. In practice, it is common to source parameters from different cell types/organisms or similar reaction networks when more relevant data are unavailable. Coupled with the inherent experimental error, this mismatch between the system under study and the system(s) described by the parameters usually requires parameters to be estimated from experimental data to better describe the system under study. However, both the cost and destructive nature of some measurements (or required offline analysis) leads to sparse data collection. Furthermore, the models developed are typically composed of a large number of uncertain parameters. Therefore, regularization techniques are typically employed to avoid over fitting models to experimental data.

Problem Formulation

Without loss of generality, consider nonlinear dynamic models with n states, m inputs, ℓ parameters, and k outputs of the form

$$\begin{aligned} \dot{x} &= f(x,u,p), \qquad x(t_0) = x_0 \\ y &= g(x,u,p) \end{aligned} \tag{1}$$

where $x \in R^n$ is the state vector, $u \in R^m$ is the input vector, $p \in R^\ell$ is the parameter vector, and $y \in R^k$. In models of biochemical reaction networks, each output is often proportional to a single model state. Without

* To whom all correspondence should be addressed

regularization, the parameter estimation problem is usually posed as a least-squares problem and can be written as

$$\begin{aligned} min_p \quad & \|y^{data} - y^{model}\|_2^2 \\ s.t. \quad & \dot{x} = f(x,u,p), \qquad x(t_0) = x_0 \\ & y = g(x,u,p) \end{aligned} \tag{2}$$

The objective function can be modified to suit the needs of the modeler, but should be chosen to minimize the fitting error by manipulating the parameter values. This optimization problem can be solved by either a sequential or simultaneous approach (Dai et al. 2014).

With regularization, the objective function of the optimization problem becomes

$$min_p \quad \|y^{data} - y^{model}\|_2^2 + h(p,p_0) \tag{3}$$

where the regularization function $h(p,p_0)$ is a function of both the estimated parameter values p and the nominal parameter values p_0.

Parameter Set Selection

Parameter set selection procedures regularize the parameter estimation problem by choosing a subset of parameters to estimate and setting the remaining parameters to their nominal values. After a subset $\ell_p \in \ell$ of parameters is chosen, the objective of the parameter estimation problem can be put into the form of Eq. (3) with

$$h(p,p_0) = \|p - p_0\|_W^2 \tag{4}$$

where the weight matrix W is diagonal with

$$W_{ii} = \begin{cases} \infty & i \cap \ell_p = \emptyset \\ 0 & otherwise \end{cases} \tag{5}$$

This formulation does not constrain parameters selected for estimation, however, parameters not selected for estimation will remain at their nominal values. The choice of norm here is arbitrary, but the $\mathcal{L}_2$ norm provides a useful comparison with Tikhonov regularization.

Tikhonov Regularization

Tikhonov regularization is another very popular regularization technique. The regularization function is identical to Eq. (4) but is usually cast as

$$h(p,p_0) = \lambda\|L(p - p_0)\|_2^2 \tag{6}$$

where L is usually either chosen to represent the "size" of the parameters or an identity matrix and λ is a positive constant chosen to balance the estimation error and the distance from the nominal parameter values.

$\mathcal{L}_1$ Regularization

Penalizing the $\mathcal{L}_1$ norm of the parameter vector is another common regularization function. The regularization function can be formalized as

$$h(p,p_0) = \lambda\|L(p - p_0)\|_1 \tag{7}$$

Where λ and L have the same purposes as in Tikhonov regularization. $\mathcal{L}_1$ regularization is used when sparsity is desired. However, this approach fundamentally differs from the parameter set selection approaches where sparsity is induced by setting specific elements of the $(p - p_0)$ vector to 0 before the estimation begins. Furthermore, pure parameter set selection approaches do not penalize changes in parameters selected for estimation, whereas $\mathcal{L}_1$ regularization still penalizes these changes.

Case Study

A model of IL-6 signaling has been developed previously (Huang et al. 2010) and will serve as a case study for understanding the effects of parameter set selection, Tikhonov regularization, and $\mathcal{L}_1$ regularization. The model is described by 13 states and 19 parameters. Furthermore, we will show how some of these techniques can be combined to obtain other regularized solutions.

Conclusions

Ill-posed inverse problems such as the parameter estimation problem in nonlinear biochemical reaction network models require some form of regularization to avoid over fitting of the model to data. The choice of regularization depends on the analyzed model and the desired qualities of the estimated parameters.

References

Dai W, Bansal, L, Hahn J. (2014). Parameter Set Selection for Dynamic Systems under Uncertainty via Dynamic Optimization and Hierarchical Clustering. *AIChE Journal*. 60(1):181-92.

Huang Z, Chu Y, Hahn J. (2010). Model Simplification Procedure for Signal Transduction Pathway Models: An Application to IL-6 Signaling. *Chem Eng Sci*. 65:1964-75.

Man CD, Micheletto F, Lv D, Breton M, Kovatchev B, Cobelli C. (2014). The UVA/PADOVA Type 1 Diabetes Simulator New Features. *J Diabetes Sci Technol.* 8(1):26–34.

Moya C, Huang Z, Cheng P, Jayaraman A, Hahn J. (2011). Investigation of IL-6 and IL-10 Signaling via Mathematical Modelling. *IET Syst Biol*. 5(1):15–26.

Sneyd J, Tsaneva-Atanasova K, Reznikov V, Bai Y, Sanderson MJ, Yule DI. (2006). A Method for Determining the

Dependence of Calcium Oscillations on Inositol Trisphosphate Oscillations. *Proc Natl Acad Sci*. 103(6):1675–80.