



# HHS Public Access

Author manuscript

*IEEE Life Sci Lett.* Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

*IEEE Life Sci Lett.* 2016 September ; 2(3): 31–34. doi:10.1109/LLS.2016.2646498.

## Regularization Techniques to Overcome Overparameterization of Complex Biochemical Reaction Networks

**Daniel P. Howsmon [Member, IEEE]** and

Biomedical Engineering Department at Rensselaer Polytechnic Institute, Troy, NY 12180 USA  
(hahnj@rpi.edu)

**Juergen Hahn [Senior Member, IEEE]**

Chemical and Biological Engineering Department at Rensselaer Polytechnic Institute, Troy, NY 12180 USA (howsmj@rpi.edu)

### Abstract

Models of biochemical reaction networks commonly contain a large number of parameters while at the same time there is only a limited amount of (noisy) data available for their estimation. As such, the values of many parameters are not well known as nominal parameter values have to be determined from the open scientific literature and a significant number of the values may have been derived in different cell types or organisms than that which is modeled. There clearly is a need to estimate at least some of the parameter values from experimental data, however, the small amount of available data and the large number of parameters commonly found in these types of models, require the use of regularization techniques to avoid over fitting. A tutorial of regularization techniques, including parameter set selection, precedes a case study of estimating parameters in a signal transduction network. Cross validation rather than fitting results are presented to further emphasize the need for models that generalize well to new data instead of simply fitting the current data.

### Index Terms

nonlinear dynamical systems; parameter estimation; computational systems biology

### I. Introduction

MODELS of biochemical reaction networks comprised of ordinary differential equations (ODEs) have been successfully employed for a variety of purposes, including summarizing experimental evidence [1], distinguishing between possible mechanisms [2], and reducing experimental burden [3]. Herein, the term “biochemical reaction networks” is used as a general term to incorporate all ODE models of biological systems, such as pharmacokinetic/dynamic models, signal transduction pathways, metabolic networks, and quantitative systems pharmacology models. These models typically contain a large number of uncertain parameters to adequately represent the underlying biology. When models such as these 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. This represents a clear

need to estimate some or all of the parameters of these models to ensure that the model can fit and predict experimental data. While it would be desirable to have as much high quality data as possible, there are several challenges to this: measuring protein concentrations and other states inside of living cell is challenging, as the measurements can be noisy, and the cost and sometimes the destructive nature (or required offline analysis) of certain measurements leads to sparse data collection. Therefore, regularization techniques are typically employed to avoid over fitting the model to the experimental data.

The most common regularization techniques employed for the parameter estimation problem in biochemical reaction networks are parameter set selection, Tikhonov regularization, and  $\mathcal{L}_1$  regularization. Herein, these three estimation techniques are presented together to allow the user to readily compare the objectives of each technique. A case study involving a representative example of a biochemical reaction network compares results obtained by each technique.

## II. Problem Formulation

From here on, consider nonlinear dynamic models with  $n$  states,  $m$  inputs,  $\ell$  parameters, and  $k$  outputs of the form

$$\dot{\hat{x}}=f(t, \hat{x}, u, p), \quad \hat{x}(t_0)=\hat{x}_0 \quad \hat{y}=g(t, \hat{x}, u, p) \quad (1)$$

where  $\hat{x} \in \mathbb{R}^n$  is the state vector,  $u \in \mathbb{R}^m$  is the input vector,  $p \in \mathbb{R}^\ell$  is the parameter vector, and  $\hat{y} \in \mathbb{R}^k$  is the output vector. In models of biochemical reaction networks, each output is often a single model state or proportionally related to a state. If the initial conditions are not precisely known, they may be included as additional parameters for estimation. Also, additional algebraic relations can be incorporated to form a systems of differential algebraic equations (DAEs) rather than ODEs.

Without regularization, the parameter estimation problem is usually posed as a least-squares problem and can be written as

$$\min_p \sum_{i=1}^j (y(t_i) - \hat{y}(t_i))^2 \quad (2)$$

$$\text{s.t.} \quad \dot{\hat{x}}=f(t, \hat{x}, u, p), \quad \hat{x}(t_0)=\hat{x}_0$$

$$\hat{y}=g(t, \hat{x}, u, p)$$

where  $y$  signifies the experimental measurements collected at various times. Systems of DAEs can be handled in the same way with additional equality constraints on the state variables. This objective can be modified (e.g. weights for different outputs, log scaling, etc.)

to suit the needs of the modeler, but should be chosen to reduce the estimation error by manipulating the parameters. For additional parameter estimation techniques, the interested reader is referred to [4], [5].

Various solution strategies to solve ODE-constrained optimization problems have been developed (e.g. [6]–[8]) and are generally categorized into sequential or simultaneous approaches [9]. The sequential approach solves the system of ODEs at one level and an optimization algorithm updates the decision variables which requires multiple solutions of the ODEs at each iteration. While this approach is conceptually straightforward, a large number of decision variables will greatly increase the computational expense of the problem. The simultaneous approach instead discretizes the ODEs by some collocation method and includes the collocation equations as constraints in one large optimization problem.

When regularization is added, the objective function of the optimization problem becomes

$$\min_p \sum_{i=1}^j (y(t_i) - \hat{y}(t_i))^2 + h(p, p_0) \quad (3)$$

where the regularization function  $h(p, p_0)$  is a function of both the estimated parameter values  $p$  and the nominal parameter values  $p_0$ . Path and terminal constraints can be added to provide regularization for the state and output trajectories, but they will be ignored here for simplicity.

### A. 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. A variety of methods for parameter set selection have been proposed in the literature (e.g. [10]–[12]), but all rely on local sensitivity analysis in some form to determine parameter effects on the outputs. Some, but not all of the available methods consider the correlations among parameters.

When the nominal values of most parameters are close to their true values, e.g., parameter estimation only results in small changes of the parameter values, then local sensitivity analysis at the nominal parameter values provides an acceptable approximation to the sensitivity of the system over the range of interest. However, this can only be determined after estimation has been performed. Also, as the uncertainty in the parameter values increases, a more detailed treatment of the systems sensitivity over the entire region of feasible parameters may be warranted. Uncertainty, as measured by changes of the sensitivity vectors, can be incorporated at the parameter set selection stage through computation of sensitivity cones which envelop all sensitivity vectors derived for different nominal values of the parameters [9].

After a subset  $q \in p$  of parameters is chosen for estimation, the objective of the parameter estimation problem can be written in the form of 3 with

$$h(p, p_0) = \|p - p_0\|_W^2 \quad (4)$$

where the weight matrix is diagonal with

$$W_{ii} = \begin{cases} \infty & i \cap q = \emptyset \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

This formulation does not constrain parameters selected for estimation while it constrains the parameters not selected to such a degree that they remain at their nominal values. The choice of the norm is arbitrary. Although this formulation would never be numerically implemented as stated here, i.e., one would select a set of parameters to estimate using one of the existing parameter set selection approaches and then only estimate this subset of parameters, this formulation highlights that parameter set selection is one of the regularization techniques that can be used to deal with over-parameterized problems.

## B. Tikhonov Regularization

Tikhonov regularization, also referred to as ridge regression, can be written in the form of 3 with the regularization function given by

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

where  $L$  is chosen to represent the relative scaling of the parameters and  $\lambda$  is a positive constant that balances the estimation error with the distance from the nominal parameter values. Tikhonov regularization is widely popular since it results in objective functions that are continuously differentiable, a requirement for many commonly available optimization packages.

## C. $\mathcal{L}_1$ Regularization

Also known as the LASSO operator, the regularization function for penalizing the  $\mathcal{L}_1$  norm can be written as

$$h(p, p_0) = \lambda \|L(p - p_0)\|_1 \quad (7)$$

where  $L$  and  $\lambda$  have the same purpose as in Tikhonov regularization. Due to the sharp increase of the regularization function as  $p$  moves away from  $p_0$ ,  $\mathcal{L}_1$  regularization promotes sparsity in the parameter estimates, retaining more parameters at their nominal values than the Tikhonov approach.

## D. Combination of Regularization Techniques

The above approaches can be combined for more sophisticated estimation strategies. Parameters set selection can remove unimportant parameters and directions from the full parameter set prior to, e.g., Tikhonov regularization. Tikhonov and  $\mathcal{L}_1$  regularization can be combined to form an elastic net [13], a common methodology in empirical modeling.

It is important to point out that it is the authors belief that there is no one “best” regularization technique for all parameter estimation problems, but rather that (a) some form of regularization is almost always advantageous over no regularization when estimating parameters in complex biochemical reaction networks, (b) that each method can be tuned to achieve good outcomes, and (c) that each technique has its advantages and its own set of drawbacks.

## III. Case Study: IL-6 Signaling

A model of interleukin-6 (IL-6) signaling developed previously [1] is used to illustrate the regularization approaches described above. The signal transduction model is described by 75 states and 128 parameters and measurements correspond to the transcription factor concentration STAT3 taken at different points in time.

### A. Parameter Set Selection

The method described in [12] is used to determine the parameters for estimation. Briefly, sensitivity vectors are obtained through local sensitivity analysis and scaled by the nominal parameter values. If a normalized sensitivity vector is small, the parameter can be removed from further analysis. Then the cosine distance between every pair of remaining sensitivity vectors is used to cluster parameters via agglomerative hierarchical clustering. The normalized sensitivity vectors are presented in a dendrogram as shown in Figure 1. In this example, a threshold of 0.1 is placed on the normalized sensitivity vector magnitude to remove parameters with small sensitivity from further consideration. This reduces the number of parameters from 128 to 33.

The dendrogram is used to view correlations between each pair of sensitivity vectors and each cluster of parameters below a user-defined cosine similarity threshold is said to be indistinguishable from other parameters in the same cluster. For example, a cosine similarity of 0.7 would divide the dendrogram into two clusters, namely  $\{p_{40}\}$  in one cluster and all other parameters in a separate cluster.

### B. Tikhonov and $\mathcal{L}_1$ Regularization

Before Tikhonov and  $\mathcal{L}_1$  regularization can be applied to this problem, an appropriate  $L$  matrix must be defined. If information on the uncertainty in the parameters is known, the covariance matrix can be used to normalize the parameters. In the absence of information about the covariance of the parameters,  $L$  can be chosen to simply normalize the parameter differences. Therefore, the  $L$  matrix is constructed as  $(1/\sqrt{p_0})$  for Tikhonov regularization and  $L = \text{diag}(1/p_0)$  for  $\mathcal{L}_1$  regularization.

### C. Estimation Results

All parameter estimation problems were solved with a simultaneous optimization strategy with Legendre-Gauss-Radau collocation on finite elements. A variant step collocation method could be used since the system stiffness varies over time; however, this work uses a fixed-step, three-point Radau collocation.

Since regularization techniques aim to reduce over fitting of the model to data, cross validation is used to evaluate the effects of regularization on the parameter estimation problem. Five data sets of nuclear STAT3 fluorescence were collected following step changes in IL-6 administration. Table I presents cross validation results for different parameter sets obtained through parameter set selection as well as Tikhonov and  $\mathcal{L}_1$  results at different values of  $\lambda$ .

Parameter set selection indicates that estimating only a few parameters can achieve similar cross validation MSE as estimating all parameters with Tikhonov or  $\mathcal{L}_1$  regularization. If parameter correlations are ignored,  $\{p_{14}, p_{48}\}$  would be selected for estimation. While this set is a decent choice in this example, it is not the best set as shown in Table I. The best set of parameters selected for estimation for this particular example is the set  $\{p_{40}, p_{48}\}$ ;  $p_{40}$  is in a cluster by itself while  $p_{48}$  is the representative parameter from the remaining three clusters and in fact this parameter has one of the largest sensitivity values. The quality of choosing this parameter set for estimation can be judged by it resulting in the lowest cross validation mean-squared error (MSE) as shown in Table I. Furthermore, it can be seen from the results that including more than two parameters for estimation does not improve prediction accuracy as the lowest MSE for any set of three parameters is larger than the MSE achieved by estimating the parameter set  $\{p_{40}, p_{48}\}$ .

The cross validation results indicate that estimating only two parameters using parameter set selection gives similar performance to estimating all parameters with either Tikhonov or  $\mathcal{L}_1$  regularization. Moreover, when the 10 largest parameters are estimated with only box constraints on the parameters to constrain the estimates to positive values, the cross validation error increases and it is clear that too many parameters were estimated in the model. The modest changes in mean-squared error seen in this example are a reflection of the good choice of initial nominal parameter values; more drastic results are seen when the nominal parameter values do not allow the model to predict the data as well.

### IV. CONCLUSION

Due to the large number of uncertain parameters and small amounts of available data which, furthermore, tend to be noisy, the parameter estimation problem in biochemical reaction networks requires some form of regularization to avoid over fitting the model to the data. The choice of regularization depends on the modeler's preference and the desired qualities of the estimated parameters. The three techniques presented here work equally well to regularize the parameter estimation problem presented in this work; however, while this is often the case based upon the authors' experience on a number of biochemical reaction network, there is no guarantee that this will always hold and each particular example may lead to slightly different conclusions. Although parameter estimation explicitly minimizes

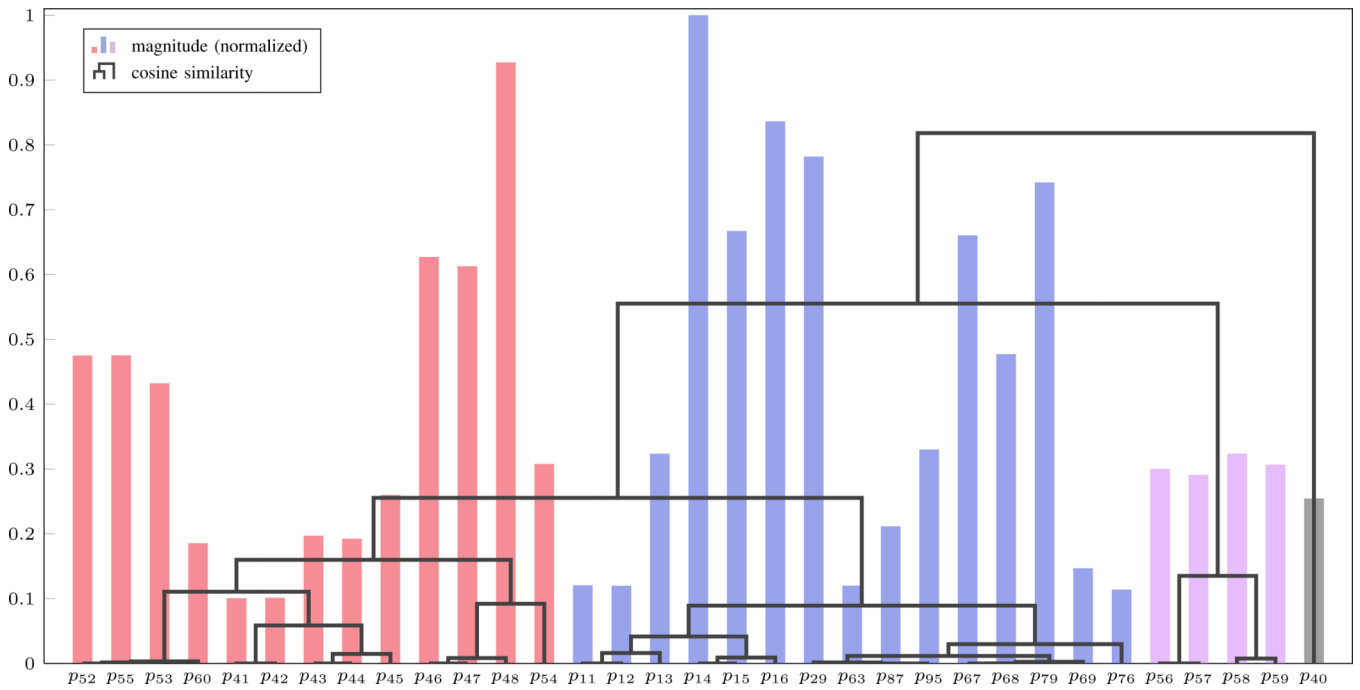
the difference between the model outputs and experimental data, cross validation should be employed to ensure that the model can predict new data well rather than simply reflect the current data.

## Acknowledgments

This work was supported by the National Institute of Health under Grant 1R01AI110642.

## References

1. Moya C, Huang Z, Cheng P, Jayaraman A, Hahn J. Investigation of IL-6 and IL-10 signalling via mathematical modelling. *IET Systems Biology*. 2011; 5(1):15–26. [PubMed: 21261398]
2. Sneyd J, Tsaneva-Atanasova K, Reznikov V, Bai Y, Sanderson MJ, Yule DI. A method for determining the dependence of calcium oscillations on inositol trisphosphate oscillations. *PNAS*. 2006; 103(6):1675–1680. [PubMed: 16446452]
3. Dalla Man C, Micheletto F, Lv D, Breton M, Kovatchev B, Cobelli C. The UVA/PADOVA Type 1 Diabetes Simulator New Features. *Journal of Diabetes Science and Technology*. 2014; 8(1):26–34. [PubMed: 24876534]
4. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology: parameter estimation for biochemical models: Parameter estimation in systems biology. *FEBS Journal*. 2009; 276(4):886–902. [PubMed: 19215296]
5. Vanlier J, Tiemann CA, Hilbers PAJ, van Riel NAW. Parameter uncertainty in biochemical models described by ordinary differential equations. *Mathematical Biosciences*. 2013; 246(2):305–314. [PubMed: 23535194]
6. Tjoa IB, Biegler LT. Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems. *Industrial & Engineering Chemistry Research*. 1991; 30(2): 376–385.
7. Poyton AA, Varziri MS, McAuley KB, McLellan PJ, Ramsay JO. Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & Chemical Engineering*. 2006; 30(4):698–708.
8. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(5):741–796.
9. Dai W, Bansal L, Hahn J, Word D. Parameter set selection for dynamic systems under uncertainty via dynamic optimization and hierarchical clustering. *AIChE Journal*. 2014; 60(1):181–192.
10. Brun R, Reichert P, Knsch HR. Practical identifiability analysis of large environmental simulation models. *Water Resources Research*. 2001; 37(4):1015–1030.
11. Lund BF, Foss BA. Parameter ranking by orthogonalization-Applied to nonlinear mechanistic models. *Automatica*. 2008; 44(1):278–281.
12. Chu Y, Hahn J. Parameter Set Selection via Clustering of Parameters into Pairwise Indistinguishable Groups of Parameters. *Industrial & Engineering Chemistry Research*. 2009; 48(13):6000–6009.
13. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. Apr; 2005 67(2):301–320.



**Fig. 1.** Sensitivity vector magnitude (normalized) and agglomerative hierarchical clustering based on the cosine similarity between sensitivity vectors. For a cosine similarity threshold of 0.2, the dendrogram shows separation of the parameters into four clusters, indicated by the four colors in the bar graph.



Cross-Validation Results

TABLE I

	MSE	Value of $\lambda$	MSE
$\{P_{14}\}$ *	70.4	0.001	74.5
$\{P_{48}\}$	70.7	0.01	72.7
$\{P_{58}\}$	72.6	0.1	70.9
$\{P_{40}\}$	70.4	1	69.9
$\{P_{14}, P_{40}\}$ *	69.6	10	69.2
<b>Tikhonov Regularization</b>			
<b>Parameter Set Selection</b>			
$\{P_{14}, P_{48}\}$	70.0	100 <sup>†</sup>	69.0
$\{P_{48}, P_{40}\}$	69.5	1000	69.0
$\{P_{14}, P_{58}, P_{40}\}$ *	69.8	10000	71.2
$\{P_{14}, P_{48}, P_{40}\}$	69.9	100000	78.1
<hr/>			
$\{P_{48}, P_{58}, P_{40}\}$	70.7	0.001	74.2
10 largest	75.2	0.01	72.6
<hr/>			
<b><math>\mathcal{R}_\lambda</math> Regularization</b>			
		0.1	71.0
		1	69.5
		10	69.3
		100	69.2
		1000 <sup>†</sup>	69.2
		10000	74.6
		100000	78.1

\* Parameter sets of differing sizes obtained from the parameter set selection procedure.

<sup>†</sup> Optimal value of  $\lambda$