The First International Symposium on Optimization and Systems Biology (OSB'07)
Beijing, China, August 8–10, 2007
, pp. 259–266

Robust Measurement Selection for Biochemical Pathway Experimental Design*

Martin Brown[1] Fei He[1,2] Lam Fat Yeung[2]

[1] Control Systems Centre, School of Electrical and Electronic Engineering
The University of Manchester, M60 1QD, Manchester, UK
[2] Dept. of Electronic Engineering, City University of Hong Kong, Hong Kong

1 Introduction

In molecular systems biology study, the dynamical properties of metabolic and signal transduction pathways are usually modelled by a set of nonlinear ordinary differential equations (ODEs) following mass balance principles and under the "well stirred" assumption [1]. During the modelling process, both the model structure and the parameters must be determined and even though the structure of some pathways are reliably established; parameter estimation still remains the limiting step due to a lack of quantitative measurement data sets covering all state variables involved. As performing time-series experiments to obtain rich data is usually expensive and time-consuming, how to select a subset of overall possible measurement data in order to estimate unknown parameters with the best statistical quality is particular important. For this purpose, experimental design [2] techniques should be incorporated with the parameter estimation procedure.

Experimental design is the process of designing experimental procedure in order to maximise the information gathered about the quantity under investigation. There are many aspects can be considered for biochemical pathway experimental design, such as initial molecular concentrations, external cellular signals, sampling time, measurement set, etc. In this paper we primarily concerned with the problem of selecting the most informative set of states to measure. As for many systems biology pathway problems, only a small subset of all states can be measured. Experimental design is typically represented as an optimization problem [3], relative to the model hypothesis space. In literature traditional optimal design criterion is generally a function of the model's Jacobian/sensitivity derivative which in turn depends on the estimated parameter values for non-linear models. However, in practice as these are only approximately known information about model parameter priori, some form of sequential or robust [4] process must be used.

*This work is supported by Hong Kong Research Grant Council, Grant No. 9041041, and partially supported by City University of Hong Kong, Grant No. 7001821, and 122305.

Sequential experimental design involves iterating between the experiment design and parameter identification processes. Whereas this iterates usually towards a locally optimal solution, and costs associated with actually performing the experiments often limit the number of iterations. Robust experimental design is concerned with producing a design that is optimal for all parameters within a specified range around the nominal value. The region of uncertainty is hypothesized to contain the optimal values, so the calculated design will be sufficiently identifiable for the true, unknown parameter values. Min/max approaches have mainly been used where the aim is to calculate a design which bounds the least identifiable parameters within the specified uncertainty region.

Aiming at the problem of output measurement selection, this paper mainly investigates and compares two robust experimental design methods. One is regularization based design approach, which incorporates additive uncertainty with parametric sensitivity matrix, and transfers the uncertainty based design into a robust semidefinite programming problem with the uncertainty bound as a regularization parameter. Another proposed robust design strategy is the Taguchi robust design method, in this approach uncertainties are directly considered subject to model parameters to form a two or three-level design, incorporating with Taguchi's design technique, an orthogonal factorial type design strategy is formed. Finally, these two robust design approaches are comparatively studied based on a simplified $I\kappa B\alpha - NF - \kappa B$ signal transduction pathway model.

2 Biochemical Pathway Modelling

In this section, dynamic pathway modelling, parameter estimation and parametric sensitivity analysis are discussed. The local sensitivity derivatives produce a lower bound for the parametric uncertainty matrix, which is also the centre information matrix to be minimized in experimental design schemes.

2.1 Model Representation

In this paper, pathway dynamics can be modelled by the following ODEs:

$$\begin{aligned} \dot{x}(t) &= f(x(t),u(t),\theta),\ \ x(t_0) = x_0 \\ y(t) &= g(x(t)) + w(t) \end{aligned} \tag{1}$$

where $x \in \mathrm{R}^m$, $u \in \mathrm{R}^p$, and $\theta \in \mathrm{R}^n$ are the state, input and parameter column vectors, andx_0 is the initial states vector. From biochemical modeling viewpoint, x, u, and θ represent molecular concentrations, external cellular signals and reaction rates. $f(\cdot)$ is a set of nonlinear transition functions describing pathway dynamics. $g(\cdot)$ here is the measurement function which determines which states can be measured and $w(t)$ is a zero mean, Gaussian measurement noise term. This generalised representation includes many of the biochemical pathway models that have been developed in the literature. For instance, the Michaelis-Menten enzyme kinetics, JAK-STAT, RKIP-ERK, $TNF\alpha - NF - \kappa B$ and $I\kappa B - NF - \kappa B$ pathway, etc, and most of them are

bilinear or nonlinear in the states. It should be noted that using ODEs to model biochemical reactions assumes that the system is well-stirred in a homogeneous medium, spatial effects are irrelevant [1].

2.2 Parameter Estimation and Local Sensitivity Analysis

To estimate the parameters of the state space model (1) requires performing experiments where the discrete time exemplar data, $\{u(t_k), y(t_k)\}$, is collected. Then parameter estimation is achieved by minimizing the sum of residual square between measurement data and model output$\hat{y}_i$:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{m} \sum_{k=1}^{l} \left(y_i(t_k) - \hat{y}_i(t_k, \theta)\right)^2 \tag{2}$$

This is, in general, a non-quadratic function of the parameters θ so iterative schemes must be used to estimate the optimal parameter values, $\hat{\theta}$. The summation operator is being performed over both the l discrete time and m measurable states. As the cost function (2) is not quadratic, optimization of parameters is actually a nonlinear programming (NLP) problem. In the literature, parameter estimation of pathway ODEs is usually reduced into solving nonlinear boundary value problem by using multiple shooting method.

Dynamic sensitivity analysis plays an important role in parameter estimation, selection and uncertainty analysis for system identification procedure [1]. The first-order local sensitivity coefficient $S_{i,j}$ is defined as the partial derivative of i^{th} state to j^{th} parameter: $S_{i,j}(t) = \partial x_i(t) \big/ \partial \theta_j$. On the other hand, the Fisher Information Matrix (FIM) is a key measure of the estimated parameters' quality and is given by:

$$\mathbf{F} = \sum_i \left(\frac{d\hat{y}_i}{d\theta}\right)^T \left(\frac{d\hat{y}_i}{d\theta}\right) = \sum_i S_i^T S_i = S^T S \tag{3}$$

where $S = dy/d\theta$ is the $m \times n \times l$ local sensitivity matrix. The inverse FIM is a lower bound on the parameter covariance matrix. This is a key measure of parametric identifiability which determines how easily the parameter values can be reliably estimated from the data, or alternatively, how many experiments would need to be performed in order to estimate the parameters to a pre-defined level of confidence. The model local sensitivity derivative can be directly computed by differentiating model ODEs (1) together with parameter sensitivity equations:

$$\dot{S} = JS + P \tag{4}$$

where $J = df/dx$ and $P = df/d\theta$ are the Jacobian and parameter Jacobian matrix.

3 Pathway Experimental Design

The aim of the experimental design process is to select a subset of the m states which are the most informative for system identification process. Parametric covariance matrix is used as the measure of identifiability and the design aim is therefore

to minimize this matrix, then state selection process can be represented as:

$$\begin{array}{l} \min \sum = \sigma^2 (\sum_{i=0}^{m} \lambda_i S_i^T S_i)^{-1} \\ s.t. \quad \lambda_i \in \{0,1\}; \quad \mathbf{1}^T \lambda = m \end{array} \tag{5}$$

where $\mathbf{1}$ is a column vector of ones. As each state is related to one weighting termλ_i, the state measurement selection problem is then to select λ_i that with the value of 1. This discrete programming problem is generally transformed into a continuous form:

$$\begin{array}{l} \min \ (\sum_{i=0}^{m} \lambda_i S_i^T S_i)^{-1} \\ s.t. \quad \lambda_i \succ 0, \ \forall i; \quad \mathbf{1}^T \lambda = 1 \end{array} \tag{6}$$

Treating λ as a continuous variable means that the optimal solution will be a lower bound for the original discrete problem. In this paper, the aim is to produce a single, robust design, hence the weighting parameters with the largest values are selected as being the most identifiable sub-set of states to measure.

3.1 Optimal Experimental Design

The vector optimization problem described in (6) does not provide a total ordering of the solutions, since it is not always possible to compare two experimental designs by stating that $\sum_1 - \sum_2 < 0$. Therefore, most experimental design procedures consider a scalarized criterion such as: A-optimal design (min $Tr(\sum) = E(\|\mathbf{e}\|_2^2)$), D-optimal design (min $\log\det(\sum)$), E-optimal design (min $\lambda_{\max}(\sum)$), and modified E-optimal design (min $\lambda_{\max}(\sum)/\lambda_{\min}(\sum)$). The relationship between these various criteria has been well studied [2]. As noted in [3], all four scalarization criteria result in convex optimization problems when the FIM is an appropriate function of the experimental design parameters. For instance, the A-optimal design, which minimizes the trace of $\sum$ which in turn minimizes the expected parametric uncertainty, can be cast as a semi-definite programme (SDP):

$$\begin{array}{l} \min \ \mathbf{1}^T t \\ s.t. \quad \left[\begin{array}{cc} \sum_{i=1}^{m} \lambda_i S_i^T S_i & e_k \\ e_k^T & t_k \end{array} \right] \succ 0, \ \ k = 1, \ldots, n \\ \qquad \lambda_i \succ 0, \ \forall i; \quad \mathbf{1}^T \lambda = 1 \end{array} \tag{7}$$

where $t \in \mathbf{R}^n, \lambda \in \mathbf{R}^m$, e_k denotes the k^{th} column of identity matrix I_n. A-optimal design minimizes the overall dimensions of joint parametric confidence region. Similarly, E-optimal design, can also be cast as a SDP:

$$\begin{array}{ll} \min & -t \\ s.t & \sum_{i=1}^{m} \lambda_i S_i^T S_i \succ t\mathbf{I} \\ & \lambda_i \succ 0, \ \forall i; \quad \mathbf{1}^T \lambda = 1 \end{array} \tag{8}$$

As minimises the largest eigenvalue of covariance matrix, it thus minimises the size of the major axis of the joint parametric confidence region.

3.2 Robust Experimental Design

Optimal experimental design criteria in Sec 3.1 make the explicit assumption that both model structure and parameters are known. However, in practice model parameters can only be known to some extent or subject to external perturbations. Thus, a key part of robust experimental design is to generate a design which is appropriate for a range of models by taking account of uncertainty in parameters.

3.2.1 Regularization Based Robust Experimental Design

As sensitivity coefficients is a function of model parameters, taking account of model uncertainty in parameters is equivalent to consider uncertainty with respect to sensitivity coefficients. When each of the sensitivity matrices S_i is subject to a random, additive uncertainty u_i, the parameter covariance matrix is given by:

$$\sum \propto \left(\sum_{i=1}^{m} \lambda_i (S_i^T S_i + u_i^T u_i)\right)^{-1} \tag{9}$$

subject to the simplex constraints on λ. It will be assumed that the additive uncertainty is common so that $u_i^T u_i = mu = U_i$, where U_i is a $n \times n$ matrix for $i = 1, \ldots, m$. The magnitude of uncertainty is bounded by: $\|blkdiag(U_1, \ldots, U_m)\| \leq \rho$. This uncertainty representation can be combined with optimal design criteria (7)-(8) and recaset as the SDPs. For E-optimal design, when considering uncertainty the optimal design problem can be transferred into a minimax problem:

$$\begin{array}{ll} \min & \max \ \|U\| \leq \rho^{-t} \\ s.t & \sum_{i=1}^{m} \lambda_i (S_i^T S_i + U_i) \succ t\mathbf{I} \\ & U = blkdiag(U_1, \ldots, U_m) \\ & \lambda_i \succ 0, \ \forall i; \quad \mathbf{1}^T \lambda = 1 \end{array} \tag{10}$$

Then by employing linear fractional representation and assuming $U_1 = \cdots = U_m$. We can transfer the minimax SDP (10) into a regularized optimization problem [4]:

$$\begin{array}{ll} \min & -t \\ s.t. & \sum_{i=1}^{m} \lambda_i S_i^T S_i - \rho\sqrt{m}\,\|\lambda\|_2 \succ t\mathbf{I}_n \\ & \lambda \succ \mathbf{0}; \quad \mathbf{1}^T \lambda = 1 \end{array} \tag{11}$$

Similarly, a robust A-optimal design SDP can be developed based on criteria (7),

$$\begin{array}{ll} \min & \mathbf{1}^T t \\ s.t. & \sum_{i=1}^{m} \lambda_i S_i^T S_i - \rho\sqrt{m}\,\|\lambda\|_2 \succ \frac{1}{t_k} e_k e_k^T, \ \ k = 1, \ldots, n \\ & \lambda \succ \mathbf{0}; \quad \mathbf{1}^T \lambda = 1 \end{array} \tag{12}$$

Thus, design of experiments with uncertainty can be deduced into the regularized optimization problem, with uncertainty bound ρ as regularization parameter.

3.2.2 Taguchi Robust Experimental Design

Instead of considering the additive uncertainty with respect to sensitivity matrix, an alternative way for robust experimental design can be based on directly considering uncertainty to parameters. For each parameter, the additive /multiplicative uncertainties can be regarded as two (+/-) or three-level (+, 0, -) design. Thus a factorial

type design can be considered. For each combination of parameter to their levels, it can be regarded as one design model. Then the aim of robust experimental design is to select most informative measurement set which is appropriate for all these models. However, a full factorial type design would be computational costly if a large number of parameters involved. In order to reduce the number of factorial design test, Taguchi design strategy [6] should be considered which employs orthogonal arrays to select a subset of full factorial design. Thus the number of design experiments can be largely reduced, while the key experimental information can still be kept. The proposed Taguchi robust measurement selection algorithm can be implemented as follows:

1. Define parameter uncertainty. In a factorial design frame, uncertainties can be added (subtracted) or multiplied to the nominal parameter values to form a two-level or three-level design. Taking additive three-level design for instance, for the j^{th} parameter θ_j, Level 1 is $\theta_j - \Delta_j$, Level 2 is θ_j, and Level 3 is $\theta_j + \Delta_j$. Here, $\Delta_j$denotes the uncertainty with respect to the j^{th} parameter.
2. According the number of parameters, construct the corresponding two or three-level orthogonal array table. The orthogonal array tables can be constructed or found in literature [6]. Considering a simple case with three parameters and three-level design, the orthogonal array can be build up as in Table I. If using full factorial design, there would be $3 \times 3 \times 3 = 27$ combinations of parameters to their levels to be tested; here orthogonal design only requires 9 trials.

Table 1: Orthogonal array, L_9

Trial	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}
P1	Level 1	Level 1	Level 1	Level 2	Level 2	Level 2	Level 3	Level 3	Level 3
P2	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
P3	Level 1	Level 2	Level 3	Level 2	Level 3	Level 1	Level 3	Level 1	Level 2
Result	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9

3. Calculate the result which is the sensitivity matrix for each trial/model. Each combination of parameters is one design model.
4. Construct the total sensitivity matrix with uncertainty information. After all the sensitivity matrices with respect to all the combination of parameters are calculated in Step 3, a compromised overall sensitivity matrix can be deduced. An easy way can be the mean or sum of all the sensitivity matrices calculated.
5. Based on the overall sensitivity matrix developed in Step 4, optimal design criteria (7)-(8) can be implemented to realize measurement selection.

Compared with regularization based robust design, instead of transforming parametric uncertainty into a single regularized parameter in optimization process, proposed Taguchi design method incorporates uncertainty into a series of orthogonal factorial design and further into a summarized sensitivity matrix.

4 Simulation Results

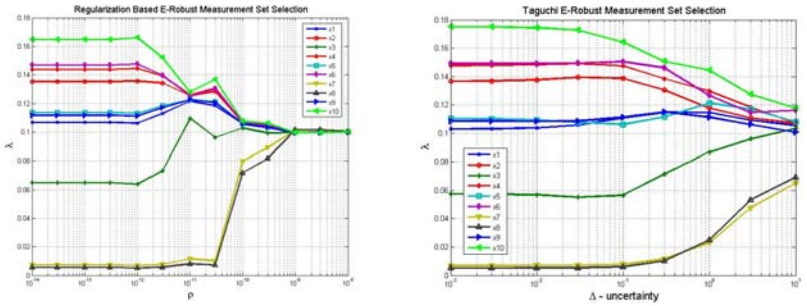
In this section, we demonstrate two robust experimental design strategies discussed in Sec 3.2 based on a simplified $I\kappa B\alpha$ activated $NF-\kappa B$ signal transduction pathway model. This reduced model contains 10 reaction species which participate 24 reactions. The model reaction mechanisms, model ODEs, parameter values and other model information is discussed in [5] in detail. The objective of experimental design is then selecting the best N state variables from the 10 state candidates for time-series measurement. For demonstration, we firstly only investigate E-optimal design. By solving the SDP (8) using SeDuMi 1.1, the E-optimal state measurement selection result can be get:

Table 2: E-optimal Measurement Set Selection

weight	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
value	0.1027	0.1364	0.0577	0.1477	0.1104	0.1492	0.0069	0.0052	0.1086	0.1752

The values of weighing terms show that x_{10}, x_6, x_4, x_2 are the top four state measurement sets should be selected according to E-optimal criteria.

For the regularization based robust design process, when considering magnitude change of perturbations, it is equivalent to change regularization parameter ρ, Here, E-robust design criteria (11) is applied for the state measurement selection.



(a) (b)

Figure 1: (a) Regularization based E-robust design, (b) Taguchi E-Robust design

Figure 1a shows that when ρ is small, experimental weights are close to the E-optimal design result in Table II. As ρ increase, the design weights converge to a uniform value 0.1. This indicates as parametric uncertainties increase, the contribution of different state measurements to modelling design tends to be the same and experimental design is most robust to parameter uncertainty.

For Taguchi based robust design, we firstly define parametric uncertainty is multiplicative to parameters with 2 levels. For simplicity, we take $\Delta_1 = \cdots = \Delta_{24} = \Delta$. According to Taguchi's book [6], a P_{24}, L_{28} 2-Level orthogonal array can then be constructed for our case. By implementing Taguchi design algorithm discussed in Sec

3.2.2, and employing E-optimal criteria, we can get the robust design results shown in Fig. 1b. As uncertainty increase, design weighting term λ tend to converge to an average value 0.1 although not as equally distributed as shown in Fig. 1a. This result clear indicates that when parametric uncertainty increase to a comparatively large scale, for both Taguchi and regularization based robust design, all the state measurement set are inclined to be equally selected.

5 Conclusion

In this work, we study the time-series experimental design problem of biochemical pathway modelling, and especially focus on robust state measurement set selection when parametric uncertainty is considered. Based on the extension of traditional optimal experimental design criteria, two robust experimental design methods are developed and comparatively studied. The regularization based robust design approach is more computational efficient, however, it needs to assume additive uncertainties with respect to all the parameters are identical. The Taguchi based design strategy is more flexible and practice tractable. Whereas, it would be computational costly, as corresponding parametric orthogonal array would be large in scale and difficult to construct for a complex biochemical pathway system. By implementing these two design approaches to a simplified $I\kappa B\alpha - NF - \kappa B$ signalling pathway system, their design advantages and drawbacks are clearly specified. When large parametric uncertainty presents, two robust design methods all tend to provide a similar uniform design result.

References

[1] Z. Szallasi, J. Stelling, and V. Periwal, Systems modelling in cellular biology: from concept to nuts and bolts, MIT Press, Cambridge, MA, Edition edn., 2006

[2] A.C. Atkinson, and A.N. Donev, Optimum Experimental Design, Oxford Science Publications, 1996.

[3] S. Boyd, and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[4] P. Flaherty, M.I. Jordan and A.P. Arkin, Robust Design of Biological Experiments, NIPS 2005.

[5] Y. Jin, H. Yue, M. Brown, Y. Liang and D. B. Kell, Improving data fitting of a signal transduction model by global sensitivity analysis, *American Control Conference 2007*, New York, USA.

[6] G. Taguchi, System of experimental design: engineering methods to optimize quality and minimize costs, White Plains, NY: UNIPUB/Krause, 1987.