REVIEW

Experimental design and model reduction in systems biology

Jenny E. Jeong[1], Qinwei Zhuang[2], Mark K. Transtrum[3], Enlu Zhou[4] and Peng Qiu[5,*]

[1] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA
[2] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30318, USA
[3] Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602, USA
[4] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA
[5] Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA; Emory University, Atlanta, GA 30322, USA
* Correspondence: peng.qiu@bme.gatech.edu

Background: In systems biology, the dynamics of biological networks are often modeled with ordinary differential equations (ODEs) that encode interacting components in the systems, resulting in highly complex models. In contrast, the amount of experimentally available data is almost always limited, and insufficient to constrain the parameters. In this situation, parameter estimation is a very challenging problem. To address this challenge, two intuitive approaches are to perform experimental design to generate more data, and to perform model reduction to simplify the model. Experimental design and model reduction have been traditionally viewed as two distinct areas, and an extensive literature and excellent reviews exist on each of the two areas. Intriguingly, however, the intrinsic connections between the two areas have not been recognized.

Results: Experimental design and model reduction are deeply related, and can be considered as one unified framework. There are two recent methods that can tackle both areas, one based on model manifold and the other based on profile likelihood. We use a simple sum-of-two-exponentials example to discuss the concepts and algorithmic details of both methods, and provide Matlab-based code and implementation which are useful resources for the dissemination and adoption of experimental design and model reduction in the biology community.

Conclusions: From a geometric perspective, we consider the experimental data as a point in a high-dimensional data space and the mathematical model as a manifold living in this space. Parameter estimation can be viewed as a projection of the data point onto the manifold. By examining the singularity around the projected point on the manifold, we can perform both experimental design and model reduction. Experimental design identifies new experiments that expand the manifold and remove the singularity, whereas model reduction identifies the nearest boundary, which is the nearest singularity that suggests an appropriate form of a reduced model. This geometric interpretation represents one step toward the convergence of experimental design and model reduction as a unified framework.

Keywords: experimental design; model reduction; model manifold; profile likelihood

Author summary: In systems biology, a common challenge is that models are often highly complex while data is almost always insufficient. Two intuitive strategies to address this challenge are experimental design (obtain more data to improve parameter estimation) and model reduction (simplify the model to reveal key mechanism). In the literature, those two have been viewed as distinct areas. We present a geometric framework to connect the two areas. We consider a model as a manifold, and explore its geometry to perform experimental design and model reduction. This framework is interesting because of both its mathematical beauty (unifying two seemingly distinct areas) and its potential impact to biology (helping biologists to design experiments and find important mechanisms).

INTRODUCTION

Mathematical modeling is undoubtedly an important tool for understanding complex biological systems, and for making predictions about system behavior far away from its nominal operating conditions [1,2]. These models can take the forms of ordinary differential equations [3], stochastic differential equations [4], Boolean networks [5], Bayesian networks [6], Petri net [7], and other frameworks [8–10] according to properties of the biological processes they intend to describe. In systems biology, a popular modeling tool for the dynamics of biological networks is ordinary differential equations (ODEs), describing changes of abundance or concentration of interacting components over time. ODE-based systems biology models are often constructed by including prior knowledge of interactions among individual genes and proteins in the complex system, resulting in highly complex models with many unknown parameters, such as reaction rates, binding affinity, hill coefficient, *etc* [11–15]. In general, these unknown model parameters can be estimated based on the experimentally observed data. However, compared to the complexity of the models, the amount of experimentally available data is almost always limited and not enough to constrain the parameters. As a result, it is possible that drastically different sets of parameters can fit the data equally well, which is a manifestation of an information gap between the model complexity and the data [16]. Parameter estimation in this situation is an ill-posed problem, and thus very challenging.

To bridge this gap between high model complexity and limited available experimental data, one strategy is to develop better optimization algorithms for parameter estimation, and investigate sensitivity and identifiability to evaluate which parameters are accurately estimated and which ones have large uncertainty. Literature along this line is voluminous, and most progress took place in the fields of statistics, machine learning, systems and control theory [17–21]. An alternative strategy is to make the problem better conditioned by obtaining more data or simplifying the model, with methods that are referred to as experimental design [22–24] and model reduction [25–28] in the literature. This paper focuses on discussing the second strategy, from both the experimental design and the model reduction perspectives.

Experimental design is an intuitive approach to address the information gap between complex models and limited experimental data, by performing new experiments to obtain more data. To design a new experiment, one typically needs to decide what perturbation (activation or inhibition) is to be applied to which components (genes and proteins) in the system, as well as which components

should be measured at which time points. These choices together constitute the design of a new experiment. The experimental design question is: given an ODE model and existing experimental data, what additional experiment is expected to be maximally informative in improving parameter estimation and reducing uncertainty?

Model reduction is another intuitive approach, which aims to simplify the complex model to an extent that is compatible to the available experimental data. For example, if additional experimental data cannot be obtained for some practical reasons (*e.g.*, biological samples are unavailable or experiments are too costly), the available experimental data is fixed. In this situation, a highly complex model may be unnecessary to describe the available data. Model reduction can be applied to derive simpler models that can describe the data with fewer parameters. Furthermore, model reduction can lead to a minimal model that cannot be further reduced without losing its ability to explain the data, and the minimal model may elucidate the key controlling mechanisms that give rise to the data. The challenge is to identify the appropriate reduction among a huge number of possible ways to write down reduced models (*e.g.*, remove or combine parameters or variables). The model reduction question is: given an ODE model and limited experimental data, how to systematically derive a sequence of reduced models, each with one fewer degree of freedom, such that the reduced models retain the ability to fit the data?

Both these questions have been studied in the literature. For example, experimental design algorithms have been developed based on Bayesian posterior sampling [29–32], information theory [33,34] and sensitivity analysis [22,35,36]. Model reduction methods have been applied to complex biological systems by exploiting system properties, such as time scales [27,28,37–40], modularity [41–44] and sensitivity [14,45–49]. Although experimental design and model reduction have been largely considered as distinct problems, these two problems share deep connections and there exist methods that can tackle both problems, such as the profile likelihood [50,51] and the model manifold analysis [52–54]. In the following, we use a simple dynamical system as an example to illustrate in details the processes of experimental design and model reduction using the profile likelihood and model manifold methods.

EXPERIMENTAL DESIGN

Existing experimental design methods

Given an ODE model and limited amount of experimental data, experimental design aims to identify the experiment

that is expected to be maximally informative in improving parameter estimation. Different existing experimental design methods differ by how "maximally informative" is defined. For example, methods based on Bayesian posterior sampling aim to identify experiments which optimizes the expected value of some objective functions associated to candidate experiments [29,30,32,50]. Information theoretic methods use entropy and mutual information to quantify the additional amount of information contained in candidate experiments [33,34]. Methods based on sensitivity analysis aim to find experiments which can maximally reduce the variance and uncertainty of the estimated parameters [22,35,52, 55–57].

The Approximate Bayesian Computation method based on Sequential Monte Carlo (ABC-SMC) is a powerful approach for parameter estimation and model selection [58]. The ABC-SMC algorithm can be applied to sample the parameter space from a posterior distribution defined based on the fit between parameter values and experimental data. As an output of this sampling process, an empirical distribution of the parameters is obtained, which describes how well the parameters are constrained by the experimental data. Compared to many parameter optimization algorithms that provide a point estimate, the empirical distribution contains richer information, which can be used for experimental design. In a Bayesian active learning method [32], this empirical distribution was used to compute, for each candidate experiment, the expected value of a loss function defined by the error of the model predictions, and the expected loss was the criterion for selecting the optimal experiment to perform next. In another Bayesian method [30], the empirical distribution of parameters was used to compute the expected variance of estimated parameters if the data is augmented by each candidate experiment, and the expected variance served as the criterion for experimental design. In the profile likelihood method [50], a variation of this empirical distribution was used to calculate the range of model predictions for each candidate experiment, and suggested that the optimal experiment should be the one with the widest spread of predictions.

Information theoretic experimental design approaches incorporate entropy and mutual information measures into the Bayesian methods. The typical criterion for optimal experiment is to maximize the mutual information between parameters and candidate experiments, or more precisely, the Kullback-Leibler divergence between the prior distribution of the parameters and the posterior distribution of parameters given data from candidate experiments [33,34]. The prior distribution of the parameters encodes either the prior knowledge of the parameters if no experimental data is available, or the posterior distribution of parameters given available data

from previously performed experiments, allowing iterative procedures between the computational analysis of experimental design and the biological efforts of carrying out the experiments.

Sensitivity analysis examines the derivatives of model predictions with respect to the parameters, which form the Jacobian and the Fish Information Matrix (FIM) [59]. For each candidate experiment, a separate FIM can be constructed by considering the derivatives of model predictions associated to the available experiments and the candidate experiment. When the FIM is evaluated at a point estimate of the parameters given the currently available data, the resulting matrix represents a linear approximation of the model, and the properties of the FIM provide quantification of parameter uncertainty if the candidate experiment is performed, which can be used as criteria for experimental design [60]. One of these criteria is A-optimality, which minimizes the trace of the inverse of the FIM, and hence minimizes the variance of the estimated parameters [52,56]. Another popular criterion is the D-optimality, which maximizes the determinant of the FIM [22,35,55,57]. Although the linear approximation seems insufficient in handling complex nonlinear dynamical models, it has been shown to be an effective method for experimental design in many systems biology studies, and is computationally much more efficient compared to Bayesian approaches.

An example model: sum of two exponentials

To discuss experimental design in details, we introduce a simple toy example, the sum of two exponentials. Assume we have a dynamical system that contains two exponentially decaying variables with unknown decay rates. Assume the two variables cannot be measured separately, and we can measure the sum of them at time points of our choosing. For simplicity, we further assume that at $t = 0$, the initial values of both variables are 1. This system can be modeled by the following ODEs in Equation (1),

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = -\theta_1 x_1 \\[2mm] \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = -\theta_2 x_2 \\[2mm] x_3(t) = x_1(t) + x_2(t) \\[2mm] x_1(t=0) = 1 \\[2mm] x_2(t=0) = 1 \end{cases} \tag{1}$$

where x_1 and x_2 represent the two dynamical variables, which exponentially decay at rates θ_1 and θ_2, respectively. x_3 represents the sum that can be measured. The

simplifying assumption is modeled by the initial conditions. Since this ODE system is simple and linear, we can obtain its analytical solution in Equation (2). However, this is special to the simple example. For complex systems biology models, such an explicit analytical solution is typically unavailable,

$$\begin{cases} x_1(t) = e^{-\theta_1 t} \\ x_2(t) = e^{-\theta_2 t} \\ x_3(t) = e^{-\theta_1 t} + e^{-\theta_2 t} \end{cases} . \qquad (2)$$

To estimate the decay rates, we design an initial experiment, where we measure the sum x_3 at time points $t = 1$ and $t = 3$. The mathematical description of the experiment is in Equation (3),

$$\begin{cases} obs_1 = x_3(t = 1) \\ obs_2 = x_3(t = 3) \end{cases} . \qquad (3)$$

Assume that after we carry out this experiment, the observed measurements are 1.10×10^{-5} and 1.04×10^{-15} (corresponding to noise-free simulation using true parameter values $\theta_1 = 14$, $\theta_2 = 11.5$).

Using this example, we introduce a few terminologies. Since the system has two unknown parameters θ_1 and θ_2, the parameter space is 2-dimensional. A 45 degree line is drawn in the parameter space because of the symmetry in the model. The experiment makes two measurements obs_1 and obs_2, and therefore the experimental data is a point living the a 2-dimensional data space. The mathematical descriptions of both the system (1) and the data (3) together constitute the model, which defines a mapping from the parameter space to the data space. If we consider all parameter settings in the parameter space and map them to the data space, we will obtain a collection of points in the data space that are achievable by the model

(shaded area in Figure 1), which is what we call model manifold. Typically, the model manifold does not occupy the entire data space, because of the constraints imposed by the mathematical structure of the model.

Albeit the simplicity of this example model, the fact of the experimental measurements being 1.10×10^{-5} and 1.04×10^{-15} represents a situation where the data does not contain enough information compared to the complexity of the model. There are two ways to explain this. (i) The measured variable x_3 is the sum of two exponential decays that will eventually go down to 0. By the time the two measurements are taken, x_3 is already extremely close to 0. Because the measurements are taken too late, there is little information in the observed data to infer the decay rates. (ii) The observed data corresponds to one point (star) in the data space, which sits in the bottom left corner of the model manifold, as shown in Figure 2B. If we formulate a least squares problem to estimate the parameters by an optimization algorithm (e.g., gradient descent, interior point [17]), and set a small error tolerance threshold of 10^{-5}, the optimization algorithm will stop when it reaches a solution in the red region of the model manifold in Figure 2B, which is tiny and only visible in the zoomed-in view in Figure 2C. However, this tiny red region of the manifold corresponds to a large region in the parameter space shown in Figure 2A. Therefore, the estimated parameter can land anywhere in the red region of the parameter space, far away from the true parameters indicated by the star, which means large uncertainty in the estimated parameters. In this example, parameter estimation is difficult because the available experiment is performed at an incorrect time scale. Intuitively, we should measure earlier time points. In the following, we use two experimental design methods to identify the appropriate time points we should measure.
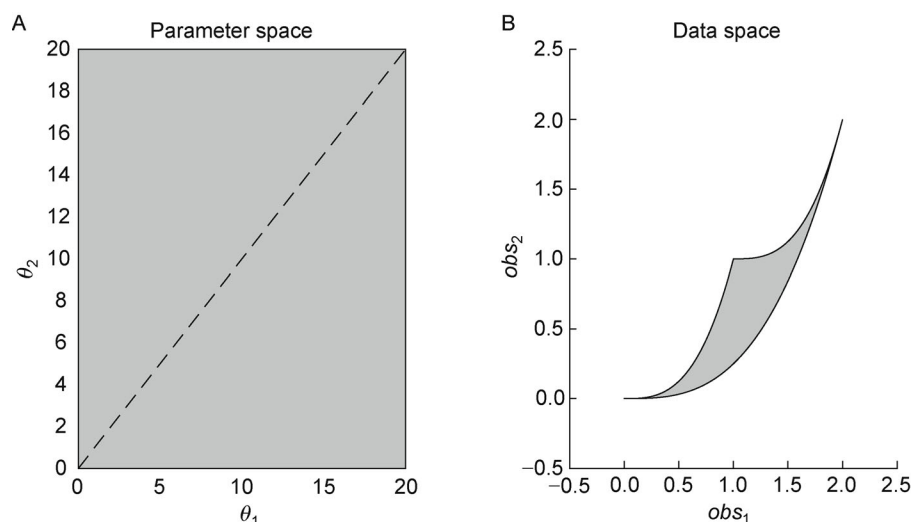

Figure 1. The parameter space and the model manifold in the data space for the sum-of-two-exponentials example.
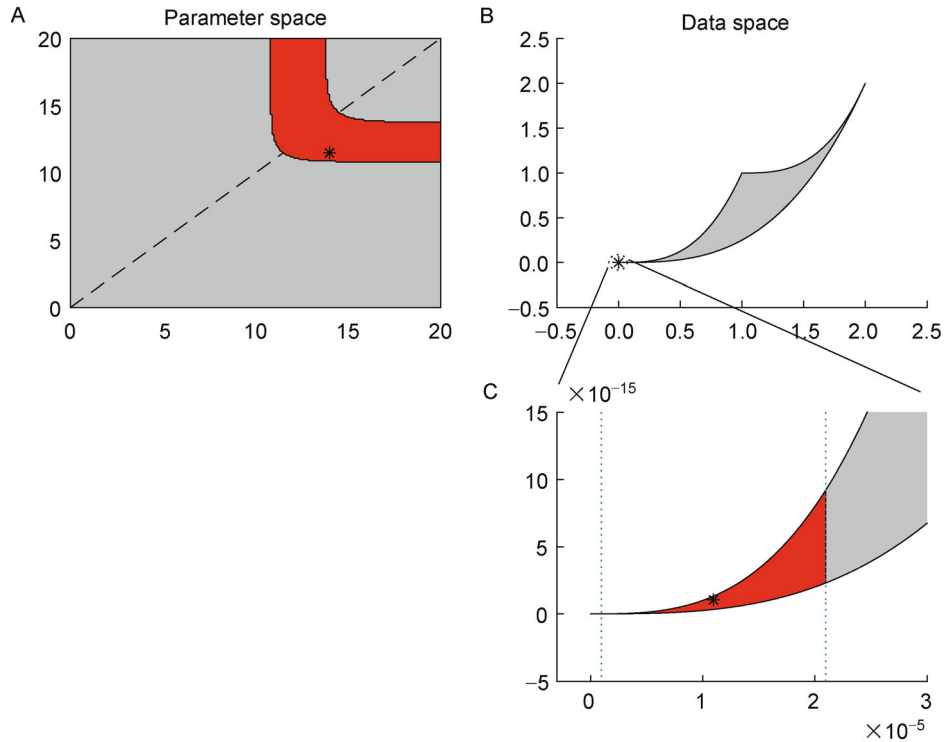
Figure 2. Small region of the manifold can map to large region in the parameter space.

Experimental design algorithm based on uncertainty reduction

We previously developed an experimental design algorithm based on sensitivity analysis and uncertainty quantification [52]. The key idea behind this algorithm is to identify the experiment that maximally reduces the uncertainty of the estimated parameter. The algorithm is as follows:

1. Set up a least squares cost function for parameter estimation $c = 0.5 \times \sum_i (obs_i - pred_i(\theta))^2$. Perform parameter estimation and obtain the best estimate based on currently available data.

2. Compute the Jacobian matrix $J_{i,j} = \partial pred_i(\theta) / \partial \theta_j$ at the best estimate. Compute the Fisher Information matrix $I = J^T J$, which is an approximation of the Hessian of the cost function at the current best estimate. Compute the parameter uncertainty: $D = trace(I^{-1})$.

3. For each candidate experiment, compute the extended Jacobian, which is the Jacobian in step 2 appended by the partial derivatives of the model predictions of the candidate experiment with respect to the parameters. Compute parameter uncertainty based on the extended Jacobian. Suggest the experiment with smallest uncertainty as the next new experiment.

4. Carry out the suggested experiment to obtain new data.

5. Iterate through steps 1–4 to suggest new experiments

until uncertainty cannot be reduced.

Given the data from the initial experiment $x_3(t = 1) = 1.10 \times 10^{-5}$ and $x_3(t = 3) = 1.04 \times 10^{-15}$, we formulate the least squares cost function in step 1, and optimize it using the Levenberg-Marquardt algorithm implemented in Matlab. The estimated parameter values are [10.70, 10.70]. Although the initial experimental data corresponds to noise-free simulation of the true parameter, the optimization algorithm stops early before reaching the true parameter due to numerical imprecision. However, the optimization actually performs well and achieves a very small value of 1.16×10^{-9} for the least squares cost. This disconnection between low cost and incorrect parameter is a manifestation of the fact that the data is limited compared to the complexity of the model.

In step 2, we compute the Jacobian matrix which is composed of derivatives of experimental observations with respect to the parameters, evaluated at the estimated parameter. In this example, these derivatives can be written analytically in Equation (4),

$$J = \begin{bmatrix} \dfrac{\partial x_3(t=1)}{\partial \theta_1} & \dfrac{\partial x_3(t=1)}{\partial \theta_2} \\ \dfrac{\partial x_3(t=3)}{\partial \theta_1} & \dfrac{\partial x_3(t=3)}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} -e^{-\theta_1} & -e^{-\theta_2} \\ -3e^{-3\theta_1} & -3e^{-3\theta_2} \end{bmatrix}.$$

(4)

After evaluating the Jacobian at the current estimated parameter [10.70, 10.70] and computing the Fisher Information Matrix and its trace, we can see that the uncertainty of the estimated parameter is huge 4.93×10^{44}. In complex models where the analytical solution of the Jacobian is not available, the derivatives in the Jacobian can be approximated by finite differences or the sensitivity equations [61].

At this moment, although we notice the huge uncertainty associated with the estimated parameter, since the least squares objective is very small, we do not have evidence against the estimated parameter. In other words, we do not have reasons to suggest the estimated parameter is incorrect, and we just do not have much confidence in it. Therefore, we assume the estimated parameter is correct, and design a new experiment that maximally increase our confidence by reducing the uncertainty. When the new experiment is carried out, we may realize that our assumption is wrong, and we can perform parameter estimation again based on the current and additional data to update the estimated parameter.

In this example, the collection of all possible new experiments is defined as measuring x_3 at another time point. For each possible new experiment (a candidate time point), we compute the Jacobian associated to the existing experiments and the new experiment, evaluate it at the current estimated parameter, and compute the trace of the Fisher Information Matrix to quantify the parameter uncertainty if the new experiment is performed. Figure 3 shows the resulting uncertainty at each candidate time point. The peak at $t = 1$ shows that repeating a previous measurement does not efficiently reduce the uncertainty. The flat region after $t = 3$ indicates that measuring any

time points after $t = 3$ does not reduce the uncertainty at all. Because $t = 3$ is already too late with observed $x_3 (t = 3)$ extremely close to 0, any time point after that is even closer to 0, which does not provide any additional information about the underlying parameter. The lowest uncertainty is achieved at $t = 0.0008$, which is the new experiment that our algorithm suggests.

After the suggested experiment is performed, the mathematical description of the experiment becomes $obs_1 = x_3 (t = 1)$, $obs_2 = x_3 (t = 3)$, and $obs_3 = x_3 (t = 0.0008)$, and the experimental data becomes $[1.10 \times 10^{-5}, 1.04 \times 10^{-15}, 1.9797]$. Repeating the parameter estimation and uncertainty quantification in steps 1 and 2, the estimated parameter becomes (13.85, 11.65), which is closer to the true parameter and has smaller uncertainty 3.30×10^{10}.

Repeating step 3 leads to the design of the next experiment. In Figure 4A, we can see that the uncertainty landscape changes. Very early time points are no longer very useful, because we already have a measurement taken at $t = 0.0008$. The uncertainty score after $t = 1$ is flat, indicating that a new time point between 1 and 3 is not useful any more. The maximal uncertainty reduction occurs at $t = 0.46$, which is in a gap between the available time points. After the suggested experiment is carried out, the entire experimental design process can iterate to suggest new time points as shown in Figure 4B–4D.

Figure 5 shows a summary of five iterations of the experimental design algorithm based on uncertainty reduction. Figure 5A shows the suggested time point at each iteration. Figure 5B shows that the estimated parameter is almost identical to the true parameter after the second experimental design iteration. Interestingly in Figure 5C, after the second iteration is completed,
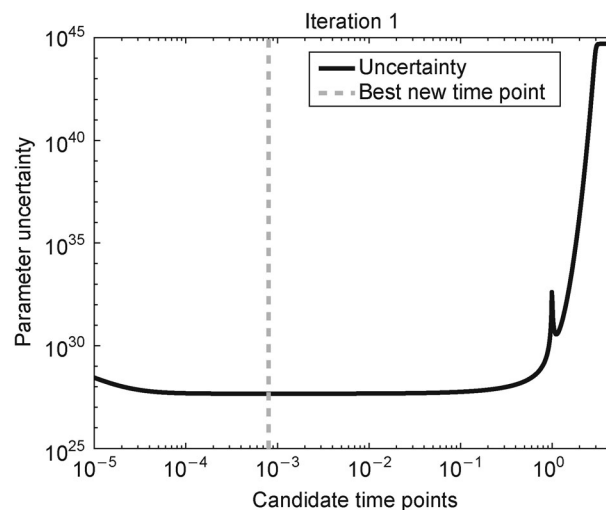

Figure 3. Uncertainty profile for experimental design. In the first experimental design iteration, parameter uncertainty at each candidate time point. The time point $t = 0.0008$ is chosen for the next experiment because it leads to lowest uncertainty.
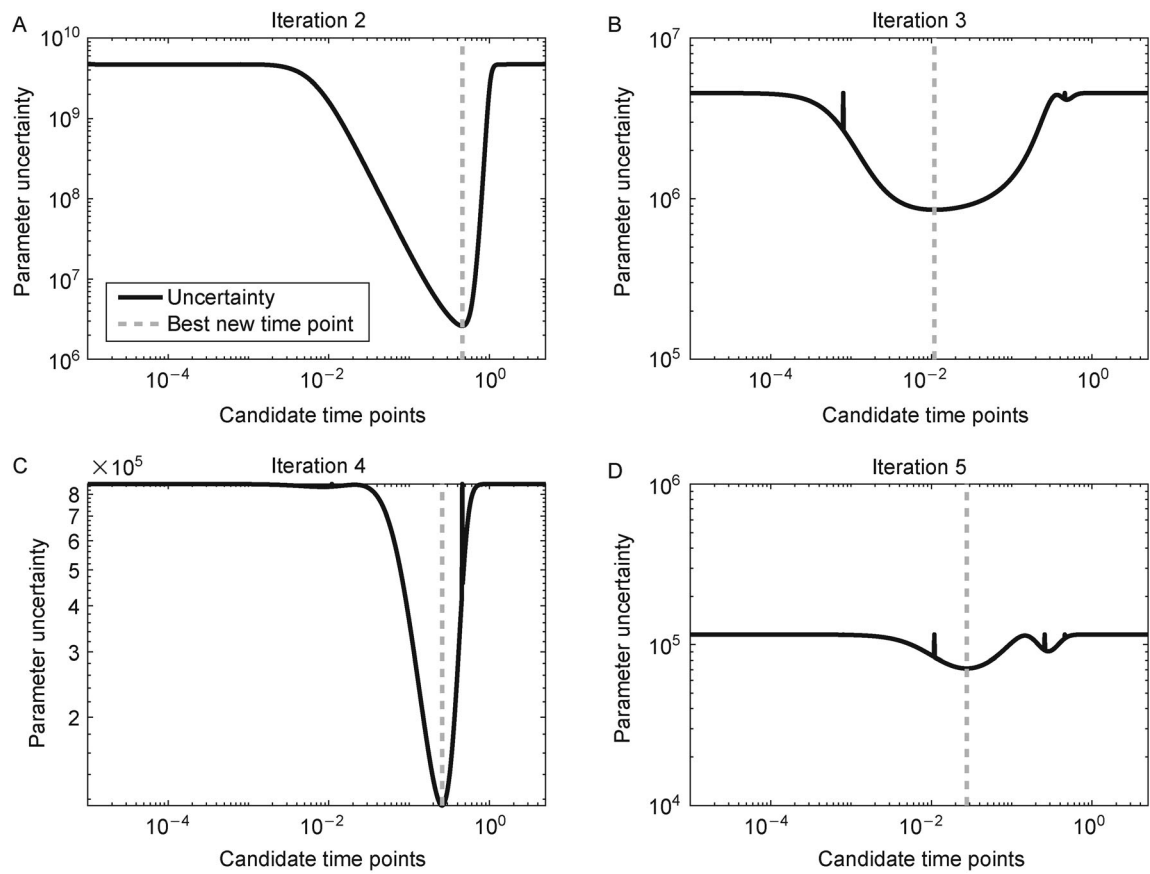
Figure 4. Parameter uncertainty at each candidate time point in iterations 2–5. The selected time points are 0.4589, 0.011309, 0.2576 and 0.0277, respectively.
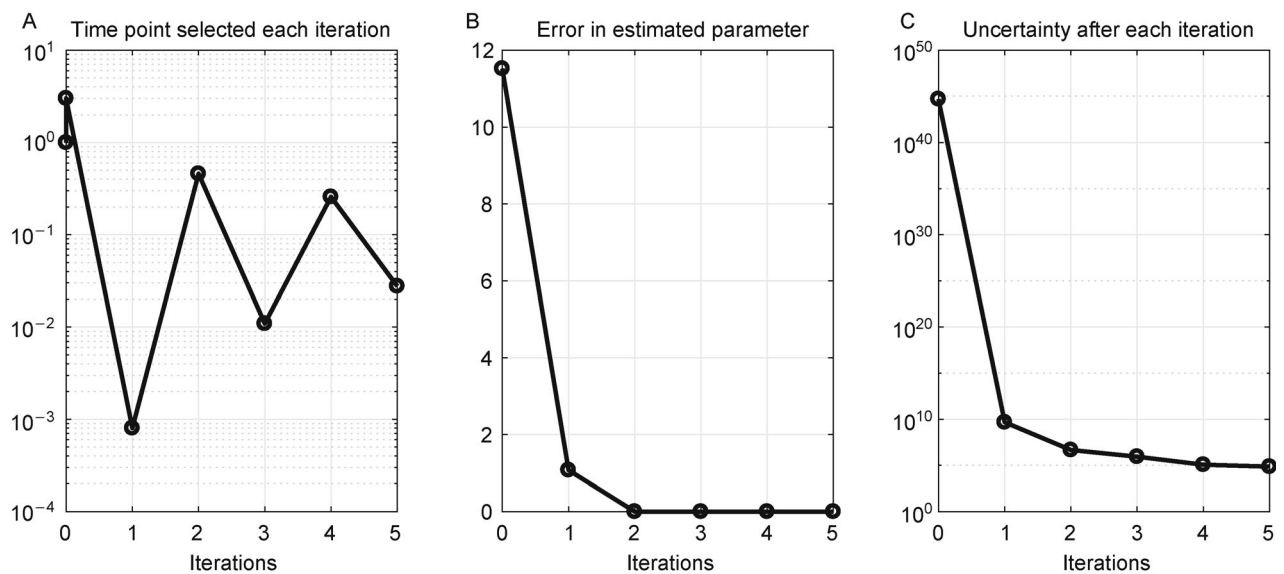

Figure 5. Summary of five iterations of experimental design based on uncertainty reduction.

subsequent iterations only lead to moderate amounts of reduction in parameter uncertainty, suggesting that the experimental data after the second iteration becomes sufficient compared to the complexity of the model, and the subsequent iterations are unnecessary.

In this example and the subsequent examples, no experimental noise is included in the simulations, although noise is an important aspect that affects uncertainty quantification and experimental design. This is mainly because operationally, with or without experimental noise, the algorithm is exactly the same: use the experimental data to perform parameter estimation, use the estimated parameter to quantify uncertainty, evaluate all possible new experiments to identify the one that leads to the most amount of uncertainty reduction, and select that one as the new experiment to perform next. Therefore, noise is not necessary for the purpose of illustrating how the algorithm works. Experimental noise in the data affects parameter estimation, which subsequently affects uncertainty quantification at the estimated parameter and the selected experiment. In an extreme case where the noise is so large that the estimated parameter is completely wrong, it is possible that the selected new experiment is not useful in reducing the parameter uncertainty. However, data from the new experiment, together with the previous data, will be able to help us to perform parameter estimation better. With the increased amount of data, the newly estimate parameter will move toward the true parameter, making the next iteration of uncertainty quantification and experimental design more meaningful. To investigate the impact of noise on a specific experimental design problem, we only need to include one more operation in this example, adding experimental noise right after the noise-free experimental data is simulated.

Experimental design algorithm based on profile likelihood

Experimental design of the sum-of-two-exponentials model can be performed using another algorithm based on the profile likelihood. In Ref. [50], the likelihood is defined with a Gaussian assumption,

$$Likelihood(obs|\theta) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(obs_i - pred_i(\theta))^2}{2\sigma_i^2}}, \quad (5)$$

where σ_i represents measurement noise. If we assume all measurements are equally accurate, the negative loglikelihood can be simplified as the following:

$$NegativeLogLikelihood(obs|\theta) = 0.5$$
$$\times \sum_i (obs_i - pred_i(\theta))^2, \quad (6)$$

which is the same as the cost function used in the experimental design algorithm based on uncertainty reduction. The likelihood function in Equation (5) or the negative loglikelihood cost function in Equation (6) can be used to define the profile likelihood.

The profile likelihood for each parameter is a function of the parameter, defined by solving many optimization problems. Take parameter θ_1 as an example. If we fix θ_1 to be a certain value (*e.g.,* 0.123), we can optimize the negative log-likelihood cost function with respect to the remaining parameters (θ_2, ..., θ_d if there are d unknown parameters), which identifies the best cost function value and the estimated values for the remaining parameters. If we vary the value of θ_1 across its feasible range and perform the optimization for each possible value of θ_1, we obtain two curves: one is the optimal cost as a scalar function of θ_1, and the other is the optimal values for the remaining parameters as a vector function of θ_1. The first of the two curves is the profile likelihood of θ_1. The profile likelihood of θ_1 characterizes the optimal cost function values when θ_1 is fixed. The same analysis can be performed for each parameter, obtaining a total of d profile likelihoods, where d is the number of parameters.

For each profile likelihood, for example the one of θ_1, we can identify the minimal, which indicates the best value to fix θ_1. We can also define a confidence interval around the best value of θ_1. For example, a generous definition would be a threshold at the 90 percentile of the optimized cost function values, which exclude 10 percent of possible θ_1 values with optimized cost function larger than the threshold. The accepted θ_1 values and the corresponding estimated values for the remaining parameters form a collection of acceptable parameter settings derived from the profile likelihood of θ_1. If we perform the same analysis for the profile likelihood of all the parameters, we will obtain a larger collection of acceptable parameter settings.

These acceptable parameter settings can then be used for experimental design. For one candidate experiment, we can use these acceptable parameter settings to simulate the experiment, and examine the range of the simulated data. If the range is small, all these acceptable parameters lead to similar predictions of the data that will be generated by the candidate experiment, which means this candidate experiment does not have the ability to differentiate the acceptable parameters. However, if the range is large, this candidate experiment is able to differentiate acceptable parameters and further constrain the parameters. Here is an algorithm for performing experimental design based on the profile likelihood:

1. Compute profile likelihood for each parameter. Basically, pick one parameter and vary its value in the feasible region. For each feasible value of the picked parameter, perform optimization to estimate the other

parameters. The optimization objective is a least squares cost function $c = 0.5 \times \sum_i (obs_i - pred_i(\theta))^2$.

2. Define thresholds to obtain the collection of acceptable parameter settings from the profile likelihood of each parameter.

3. For each candidate experiment, simulate the model using all the acceptable parameter settings, and compute the range of the simulated predictions. Suggest the experiment with the largest range as the next new experiment.

4. Carry out the suggested experiment to obtain new data.

5. Iterate through steps 1−4 to suggest new experiments until some stopping criterion (the profiles become very sharp, or the range of simulated predictions becomes very tiny).

In the sum-of-two-exponentials example, given the data from the initial experiment $x_3 (t = 1) = 1.10 \times 10^{-5}$ and $x_3 (t = 3) = 1.04 \times 10^{-15}$, we formulate the negative loglikelihood cost function in step 1 and compute the profile likelihood for each parameter using the Levenberg-Marquardt optimization algorithm implemented in Matlab. We first fix θ_1 at different values, optimize the cost function with respect to θ_2, and generate the profile likelihood for θ_1 shown in Figure 6A. In Figure 6A, when θ_1 is fixed at small values, the optimized cost function is large because it is impossible to tune θ_2 alone to achieve small cost function value. Correspondingly in Figure 6C, the optimized θ_2 exhibits a strange oscillating pattern when θ_1 is fixed at small values, which is likely caused by numerical issues of the optimization algorithm. Since the two parameters are symmetric, the profile likelihood of the two parameters are identical, and therefore Figure 6B and 6D are identical to Figure 6A and 6C.

To define acceptable parameter settings in step 2, we need to choose a threshold. In this example, the threshold is defined as 90% of the best likelihood in Equation (5). The corresponding threshold on the profile likelihood is shown by the horizontal line in Figure 6A and 6B. This is a generous threshold, accepting many parameter settings along the profile likelihood. The acceptable parameter settings are shown in black in Figure 6C and 6D, whereas gray indicates the parameter settings that are not accepted according to the threshold.

In step 3, the model is simulated using the acceptable parameter settings to obtain the range of predicted data for all candidate experiments (time points). Figure 6E shows the range of x_3's behaviors across the acceptable parameter settings. The time point $t = 0.0135$ is suggested as the new experiment, because it corresponds to the largest range, and hence has the most discriminating power among the acceptable parameter settings.

After the suggested experiment is performed, the experimental data becomes $[obs_1, obs_2, obs_3] = [x_3 (t = 1),$ $x_3 (t = 3), x_3 (t = 0.0135)] = [1.10 \times 10^{-5}, 1.04 \times 10^{-15}, 1.6851]$. With the new experimental data, if we perform parameter estimation again, the estimated parameter becomes [14.3431, 11.1697], closer to the ground truth. We can perform a second iteration of the experimental design algorithm to suggest the next time point. In the second iteration, the updated profile likelihood is shown in Figure 7A and 7C. The sharp peaks indicate that the current data is already able to constrain the parameters close to the true parameter values. In Figure 7E, we can see that the simulated behaviors of x_3 still exhibit large range at late time points, and the algorithm suggests $t = 0.2899$ as the next time point to measure.

A summary of the first five iterations of the profile likelihood based experimental design is shown in Figure 8. Figure 8A shows the suggested time points in each iteration. After the second iteration, the algorithm stops exploring the new time points. This coincides with Figure 8B, showing that the error in the estimated parameters approaches 0 after the second iteration. Although profile likelihood does not consider parameter uncertainty, the suggested time points do lead to great reduction of parameter uncertainty along the process, as shown in Figure 8C.

Geometric interpretation of experimental design

In each iteration of the experimental design process, a new experiment is suggested and performed. More available data means that the dimension of the data space is increased. Since the number of parameters stays the same, the intrinsic dimension of the model manifold does not change. Therefore, increasing the data by new experiments will expand the data space by adding new dimensions, and deform the model manifold into the new dimensions, but will not change the dimension of the model manifold itself.

Figure 9 illustrates how experimental design changes the geometry of the model manifold using the sum-of-two-exponentials example. With the initial experimental design $obs = [x_3 (t = 1), x_3 (t = 3)]$, the model manifold is a 2D object (2 parameters) that lives in the 2D data space, shown in Figure 9A. The manifold is colored by the sum of the two parameters. The upper right corner (blue) corresponds to both parameters being small. The bottom left corner (red) corresponds to both parameters being large. The data of the initial experiment sits in the bottom left corner of the manifold. After the first iteration of experimental design based on uncertainty reduction, a third observation $x_3 (t = 0.0008)$ is made, and the model manifold becomes a 2D object in the 3D data space shown in Figure 9B, colored in the same way. If we look at the model manifold in Figure 9B from top to down, we will see the exact same figure as the 2D version in Figure 9A.
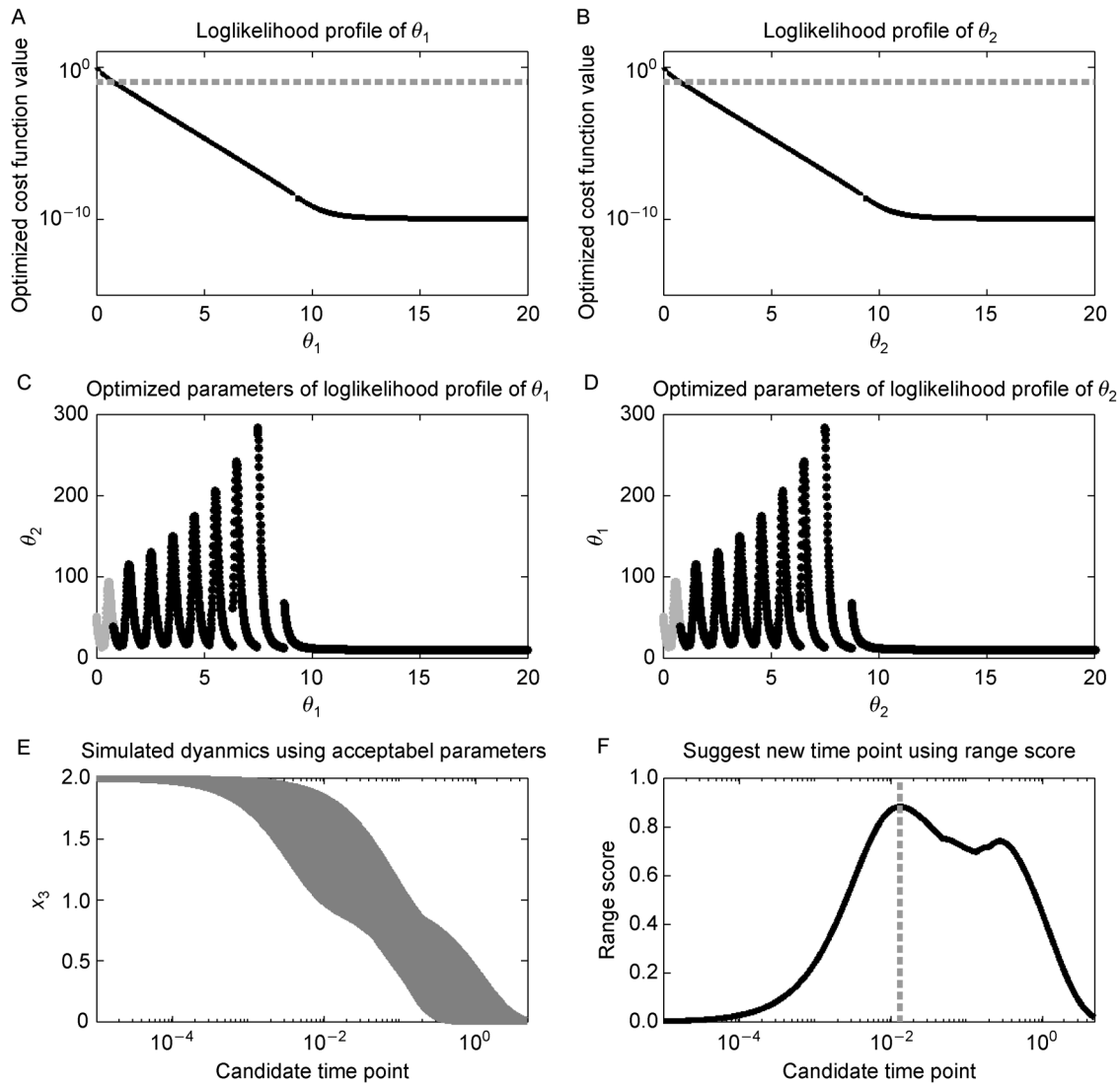
Figure 6. The first iteration of profile likelihood based experimental design.

From the color, we can see that the new experiment expanded the bottom left corner (red region) of the model manifold and stretched that region downwards, while the rest of the model manifold (blue region) is less affected. This is because the new experiment is designed to improve parameter estimation of the data that sits at the bottom left corner. Figure 9C shows the geometry of the model manifold after the first iteration of the experimental design algorithm based on profile likelihood, which is qualitatively the same as Figure 9B. As the experimental design algorithm iterates, the model manifold is expanded to higher dimension. The manifold region around where the experimental data sits is expanded more than the rest of the manifold.

When applied to the sum-of-two-exponentials example, the above experimental design algorithms show similar performance. In both cases, the first two suggested time points are sufficient in constraining the parameters, and therefore, the subsequent iterations are actually unnecessary. The two algorithms are comparable in terms of the error and the uncertainty in the estimated parameters. We implement the experimental design process of the sum-of-two-exponentials example in Matlab, and provide the code and documentations at http://pengqiu.gatech.edu/software/model_manifold/html/publish_model_manifold.html.

In this example, we assume the collection of all possible new experiments is defined by measuring x_3 at any time point. This is certainly unrealistic, because getting data very frequently at highly precise time points is very difficult experimentally. In reality, the collection of all possible and realistic new experiments is typically a
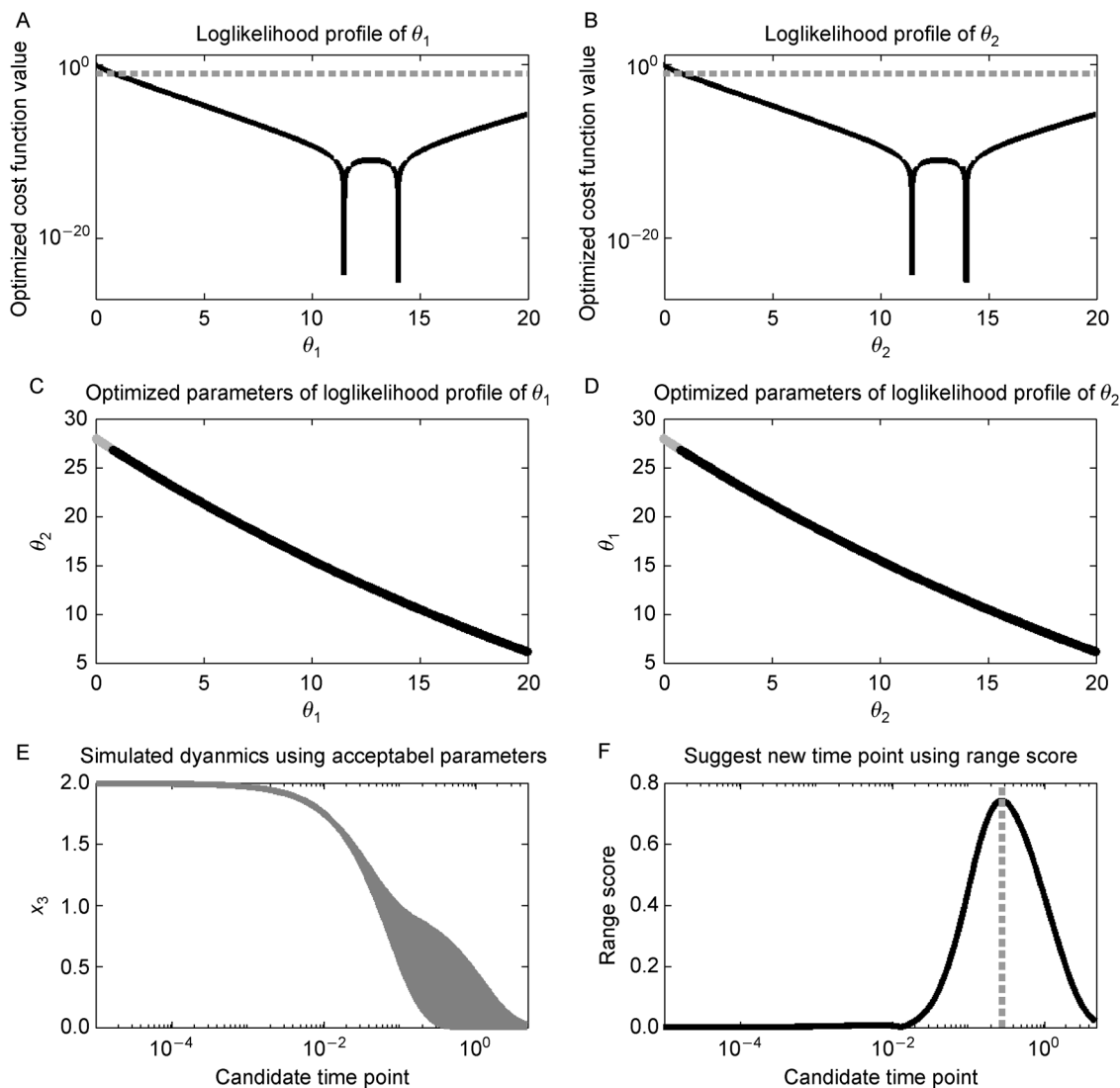
Figure 7. **The second iteration of profile likelihood based experimental design.**

smaller set due to experimental and technical constraints, and the uncertainty quantification for experimental design should only consider the realistic experiments.

MODEL REDUCTION

Existing model reduction methods

Given an ODE model and a limited amount of experimental data, model reduction aims to derive reduced models that can fit the data with fewer parameters, and identify the minimum model to elucidate the key mechanisms that give rise to the experimental observations. The challenge is to identify the appropriate reduction among a huge number of possible ways to write down reduced models (*e.g.*, remove or combine para-

meters or variables). Most existing model reduction methods follow a two-step process: first identify which part of the model should be simplified by exploiting special properties and sensitivities of the model, and then write down the mathematical form of the simplified model using ad hoc rules or biological insights. Useful properties for model reduction include separation of time scales [27,28,37–40,62], clustering and lumping of variables into modules [42–45], and insensitive parameters or variables [14,45–49]. Methods exploiting these properties are summarized in a recent survey on model reduction [63].

Biological systems often contain processes that occur in different time scales [64]. Compared to the experimentally observed behavior, processes that occur much faster can be assumed to be in the steady state and
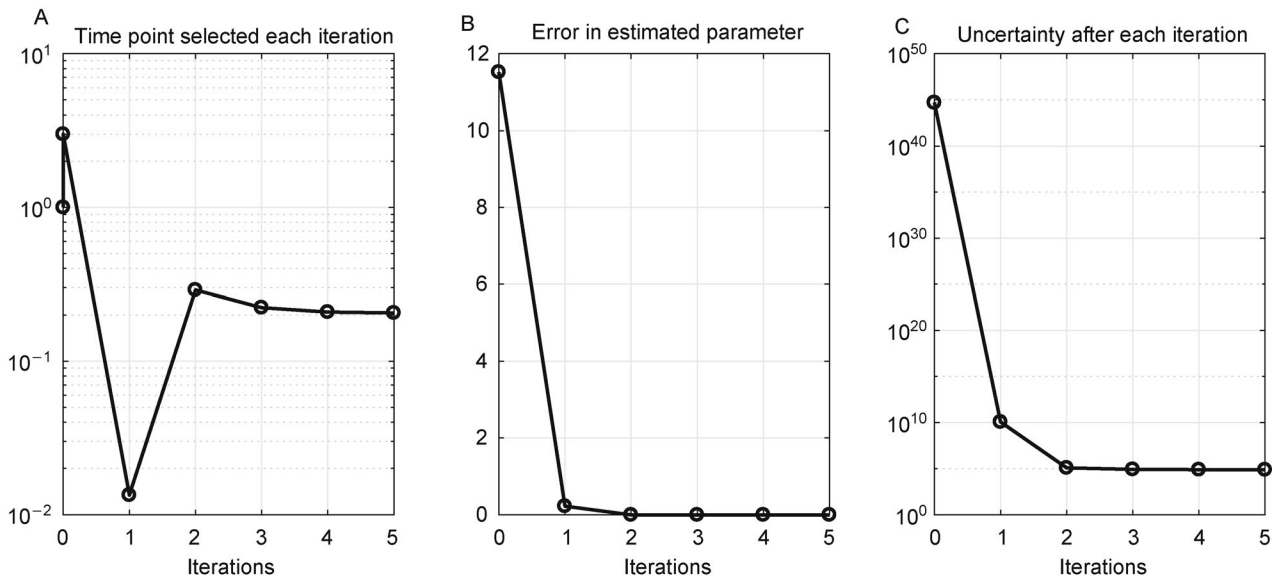
Figure 8. Summary of five iterations of experimental design based on profile likelihood.

processes that occur much slower can be approximated as constants [65]. In an analysis of the *Wnt/β*-Catenin signaling pathway [37], the experimental observations are made on a time scale of hours. Since separate experiments observed no detectable degradation of several proteins in the pathway in several hours, the concentration of these proteins are assumed to be constants, so that the corresponding differential equations are converted into algebraic equations, reducing the number of dynamical variables. In addition, a few reversible binding reactions are known to occur much faster, which lead to quasi-equilibrium approximations that turn reaction rates into ratios, reducing the number of parameters [37]. For complex models, computational singular perturbation analysis can be applied to systematically identify processes or transformations associated to fast and slow time scales, which lead to model reduction [27,38–40,62].

Lumping methods for model reduction aims to remove dynamical variables from the system and replace them by new lumped dynamical variables that represent affine combinations of the removed variables [41,42]. Variables to be lumped can be identified by either biological intuition of the structural properties of the model [26,43,66–68], or systematic algorithms based on principle component analysis [69,70], greedy iterative forward selection [45] and decomposition algorithms [13]. One special case of lumping methods is called proper lumping, where each variable in the system contributes to only one of the new lumped variables. This constraint allows the interpretation of the reduced model to be clearly connected to the interpretation of the original variables [13,43,44].

Sensitivity analysis is another popular approach for model reduction. In complex models with limited experimentally observed data, there typically are parameters or parameter combinations that are not important for the observed data. In sensitivity analysis, the insensitive model parameters that affect the dynamics the least are eliminated [45–47]. The insensitive parameters can be identified by principal component analysis of the Jacobian [14,49] and flux analysis of the stoichiometry [48].

The sum-of-two-exponentials example for model reduction

Similar to the discussion of experimental design, we use the sum-of-two-exponentials model to illustrate model reduction algorithms. The model reduction problem starts with the mathematical model in Equations (1) and (3). Similar as before, we assume the true underlying parameter values are $\theta_1 = 14$, $\theta_2 = 11.5$, and the experimental data are $obs_1 = x_3\,(t = 1) = 1.10 \times 10^{-5}$ and $obs_2 = x_3\,(t = 3) = 1.04 \times 10^{-15}$. Therefore, the model and data together represent a situation where the data is not enough to constrain the model parameters, making this problem amendable to model reduction. In the following, we apply two model reduction methods to derive reduced models for this example.

Model reduction based on manifold boundary approximation method (MBAM)

The manifold boundary approximation method (MBAM) [53] performs the model reduction by manifold boundaries. The basic idea is to view the experimental data as a
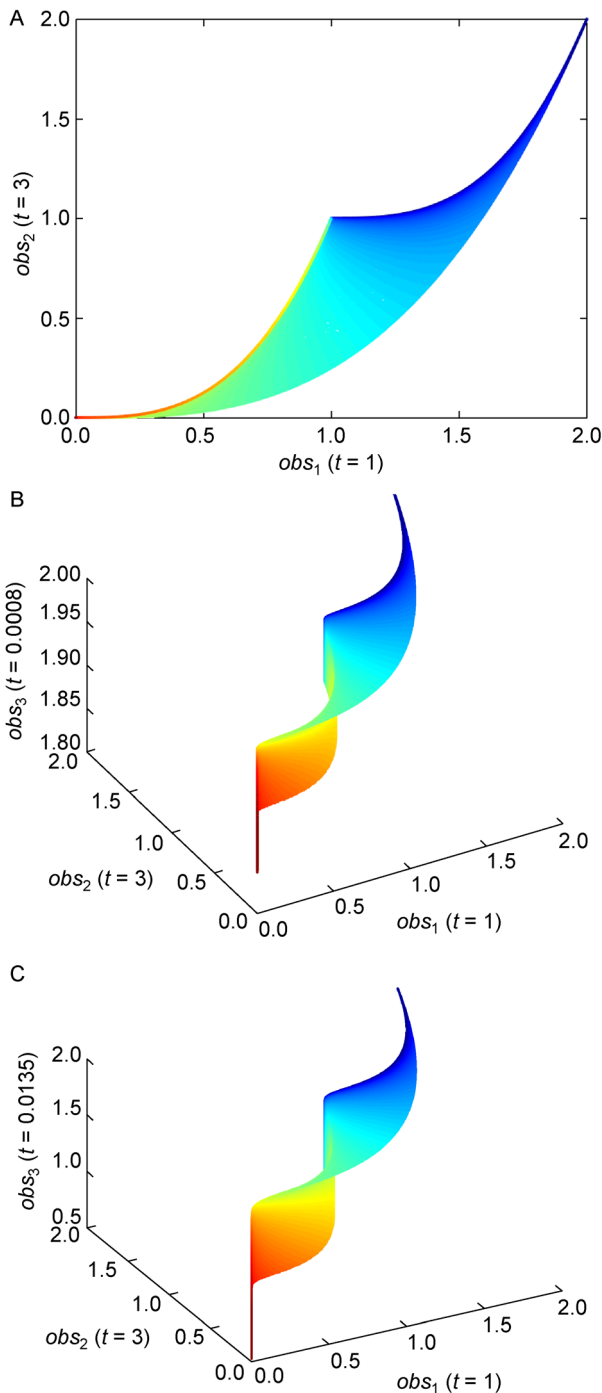
Figure 9. Geometry of model manifold is changed by the experimental design process. (A) Initial experimental design. (B) Third time point at 0.0008. (C) Third time point at 0.0135.

point in the data space, project it on the model manifold, and identify the nearest manifold boundary. This manifold boundary corresponds to a reduced model with one fewer degree of freedom, and is able to better fit the data than other reduced models corresponding to other boundaries

of the manifold.

In the sum-of-two-exponentials model, the model manifold has three boundaries, which are highlighted in Figure 10. Starting from the original model in Equations (1) and (2), if θ_1 is set to infinity, the solution of the model becomes $x_3(t) = e^{-\theta_2 t}$, as the exponential term corresponding to infinite θ_1 decays to 0 instantaneously. Given the same initial experiment that makes two measurement, the model manifold for the reduced model becomes a 1D object in the 2D data space, which is the blue boundary of the original model manifold in Figure 10. Setting θ_1 is close to 0 leads to another reduced model $x_3(t) = e^{-\theta_2 t} + 1$, whose corresponding model manifold is the green boundary in Figure 10. Lastly, if the two decay rates are equal $\theta_1 = \theta_2$, the original model reduces to $x_3(t) = 2e^{-\theta_2 t}$, and the model manifold reduces to the red boundary. This example shows the boundaries of the original model manifold correspond to reduced models that can be derived by physically meaningful limits of the parameters.
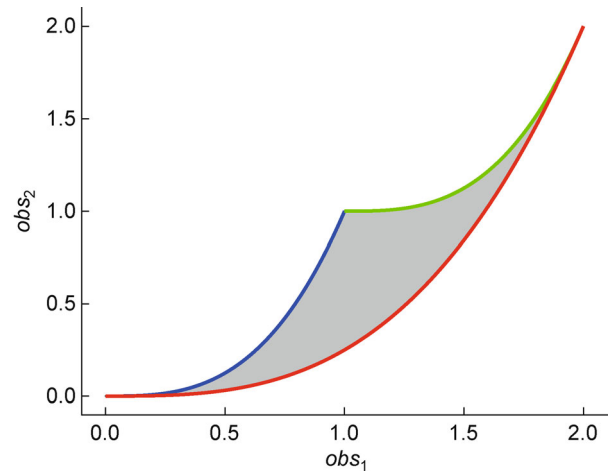

Figure 10. Model manifold boundaries.

Motivated by the correspondence between manifold boundaries and reduced models, we developed the MBAM algorithm for model reduction [53]. The MBAM algorithm is as follows:

1. Based on the current model, set up a least squares cost function for parameter estimation $c = 0.5 \times \sum_i (obs_i - pred_i(\theta))^2$. Perform parameter estimation and obtain the best estimated parameter based on the data, which corresponds to the projection of the experimental data onto the model manifold.

2. Identify the manifold boundary nearest to the experimental data by numerical integration of the geodesic equation, using the estimated parameter as the 0-th order initial condition and the least sensitive

parameter directions from the Fisher Information Matrix as the first order initial condition. The geodesic equation is a set of N nonlinear second-order differential equations: $\frac{d^2\theta^\mu}{d\tau^2} + \sum_{\alpha,\beta}\Gamma^\mu_{\alpha,\beta}\frac{d\theta^\alpha}{d\tau}\frac{d\theta^\beta}{d\tau} = 0$, where μ runs from 1 to N, indexing the parameters. $\Gamma^\mu_{\alpha,\beta}$ is known as the connection coefficient. $\Gamma^\mu_{\alpha,\beta} = \sum_{v,m}\left(g^{\mu v}\frac{\partial pred_m}{\partial\theta^v}\right.$ $\left.\frac{\partial^2 pred_m}{\partial\theta^\alpha\partial\theta^\beta}\right)$ where, $g^{\mu v} = \left(\sum_m\frac{\partial pred_m}{\partial\theta^\mu}\frac{\partial pred_m}{\partial\theta^v}\right)^{-1}$ is the (μ,v) element of the inverse of the Fisher Information Matrix. In the summations, α, β, v run from 1 to the number of parameters, and m runs from 1 to the number of experimental observations. If we set $\theta\,(\tau = 0)$ to be the estimated parameter and set $\theta'(\tau = 0)$ to be the eigenvector of the smallest eigenvalue of the Fisher Information Matrix (the least sensitive parameter direction), solution to the geodesic equation is a nonlinear curve in the parameter space which maps to a "straight" line on the model manifold, leading to the boundary closest to where the projection of the experimental data sits.

3. As the geodesic path approaches a boundary, the Fisher Information Matrix becomes singular, and certain parameters approach to limiting values (such as $0, \infty$). We evaluate the limits and manually write down the mathematical form of a reduced model.

4. With the reduced model, go through steps 1–3 to derive further model reductions, until the reduced model cannot fit the experiment data well.

The most challenging step of MBAM is the numerical integration of the geodesic equation, which is a set of second-order differential equations. To use existing differential equation solvers to perform numerical integration, we turn the geodesic equation into the following first-order form:

$$\frac{d}{d\tau}(\theta^\mu) = \left(\frac{d\theta^\mu}{d\tau}\right), \mu = 1, 2, ..., N,$$

$$\frac{d}{d\tau}\left(\frac{d\theta^\mu}{d\tau}\right) = -\sum_{\alpha,\beta}\Gamma^\mu_{\alpha,\beta}\frac{d\theta^\alpha}{d\tau}\frac{d\theta^\beta}{d\tau}, \mu = 1, 2, ..., N, \qquad (7)$$

Denote $\frac{d\theta^\mu}{d\tau}$ as new variables $[d\theta^\mu]$, Equation (7) becomes clearer,

$$[\dot\theta^\mu] = [d\theta^\mu],$$

$$[d\dot\theta^\mu] = -\sum_{\alpha,\beta}\Gamma^\mu_{\alpha,\beta}[d\theta^\alpha][d\theta^\beta]. \qquad (8)$$

where each "[...]" represents one variable in the first-order form of the geodesic equation, and the dot above represents the first order derivative with respect to τ. Numerical integration of Equation (8) requires the connection coefficients $\Gamma^\mu_{\alpha,\beta}$, which are functions of the

parameters θ. As mentioned in step 2 of MBAM, calculating the connection coefficients involves the first- and second-order partial derivatives of the model predictions of the experimental data with respect to the parameters, which can be computed by finite differences or the sensitivity equations [61]. Overall, the process of integrating the geodesic equation is the following: first specify an initial parameter and initial velocity, numerically compute the connection coefficients at the initial parameter, inch forward both θ^μ and $d\theta^\mu$ using Euler formula, then recompute the connection coefficient at the new θ^μ, inch forward again, and keep repeating this process to obtain the geodesic path in the parameter space.

Since MBAM uses manifold boundaries to derive reduced models, it requires the model manifold to be bounded and requires the model to be differentiable. In systems biology, models are usually constructed with biological assumptions and constraints such as bounded production, exponential decay and mass conservation, which often make the model manifold bounded. Therefore, MBAM is generally applicable to models in systems biology. The manifold for the sum-of-two-exponentials example is obviously bounded, and hence can be analyzed using MBAM. In the sum-of-two-exponentials example, the decay parameters are non-negative. Many models in systems biology also involve non-negative constraints on the parameters. For such constraints, one neat trick is to modify the model to work with log-parameters, which are unconstrained because the exponentiation automatically takes care of the non-negative constraints. The modified model in log-parameters is shown in Equation (9):

$$\begin{cases} x_3(t) = e^{-te^{\log\theta_1}} + e^{-te^{\log\theta_2}} \\ obs_1 = x_3(t = 1) \\ obs_2 = x_3(t = 3) \end{cases}, \qquad (9)$$

where the two parameters are $\log\theta_1$ and $\log\theta_2$.

In this example, we assume the parameter estimation in step 1 is perfectly accurate, and use the true parameter as the initial parameter for integrating the geodesic equation, $\log\theta^\mu = [\log(14), \log(11.5)]' = [2.6391, 2.4423]'$. The initial velocity $d\log\theta^\mu$ is the least sensitive parameter direction. To obtain the initial velocity, we compute the Jacobian matrix J at the initial parameter, compute the Fisher Information Matrix $J'J$, and find its eigenvalues and eigenvectors. The eigenvector corresponding to the smallest eigenvalue is the initial velocity direction. Equivalently, the initial velocity direction can be defined by the right singular vector of J that corresponds to J's smallest singular value. The eigenvector (or the singular vector) is not the initial velocity yet, there is still ambiguity about the direction, because the opposite direction of an eigenvector is also an eigenvector corresponding to the same eigenvalue. To determine the

direction, we compute the right hand side of the $[\text{dlog}\theta^{\mu}]$ equation in Equation (8), which defines the acceleration. If the inner product of the eigenvector and the acceleration is positive, we define the initial velocity to be the eigenvector itself. If the inner product is negative, we define the initial velocity to be the negative of the eigenvector. In this case, the initial velocity is $[-0.9950, 0.0994]'$.

After determining the initial conditions of the geodesic Equation (8), we numerically integrate it to identify the nearest boundary. Figure 11A shows the trajectory obtained by integrating the geodesic equation, the log-parameter values as functions of τ which parameterizes the geodesic path. We can see that the two parameters gradually approach each other and become almost the same. In Figure 11C, we show the geodesic path in the original parameter space, overlaid with the least squares cost surface. We can see that the geodesic path starts from the initial parameter we specified, and moves along the canyon of the cost surface until some limit is achieved. As shown in Figure 11D, the accelerations of both parameters grow very large. This is an indication that the identified boundary corresponds to a limit that involve both parameters. Since the two parameters gradually approach each other along the geodesic path, the limit involving both parameters corresponds to the symmetry of the model, where the two parameters are equal. Evaluating this limit leads to a reduced model $x_3(t) = 2e^{-te^{\log\theta_2}}$, corresponding to the bottom-right boundary (red) of the original model manifold shown in Figure 10.

After presenting one geodesic path in the parameter space, natural next questions are how about its corresponding image in the data space on the model manifold, and how does the geodesic path look like given different starting points? To answer these questions, we compute initial velocities corresponding to various initial log-parameters that map to different points on the model manifold, and map the initial velocities of the log-
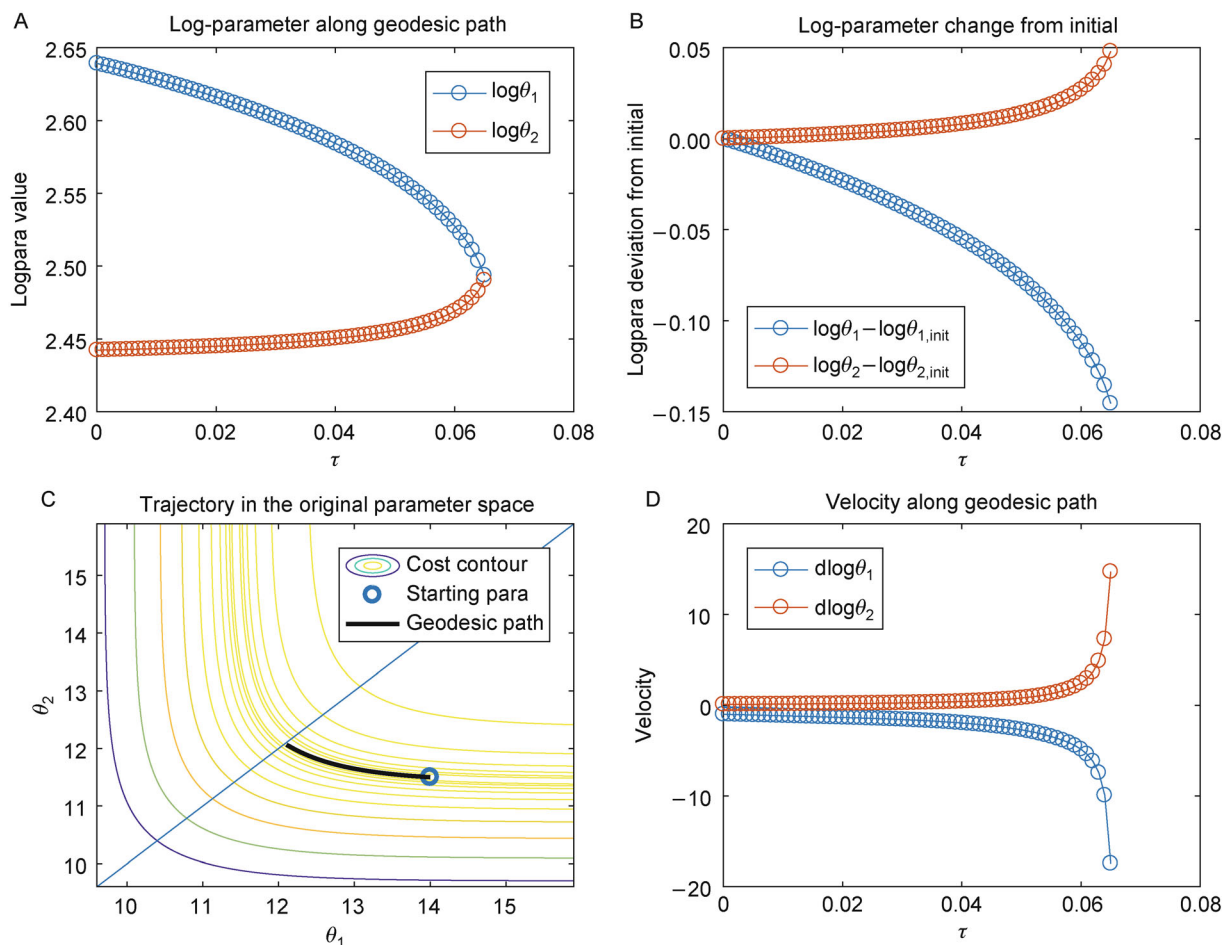

Figure 11. Geodesic path obtained in the first iteration of MBAM model reduction for the sum-of-two-exponentials model. (A) The log-parameter values along the geodesic path. (B) The deviation of log-parameter from the initial point. (C) The geodesic path in the original parameter space overlaid with the cost surface contours. (D) The velocity of the log-parameters along the geodesic path.

parameters to directions on the model manifold. The mapped velocities on the manifold are visualized as a vector field on the model manifold in Figure 12. From this vector field, it is easy to imagine the trajectory of the geodesic paths on this model manifold, and map out regions on the manifold that lead to each boundary.
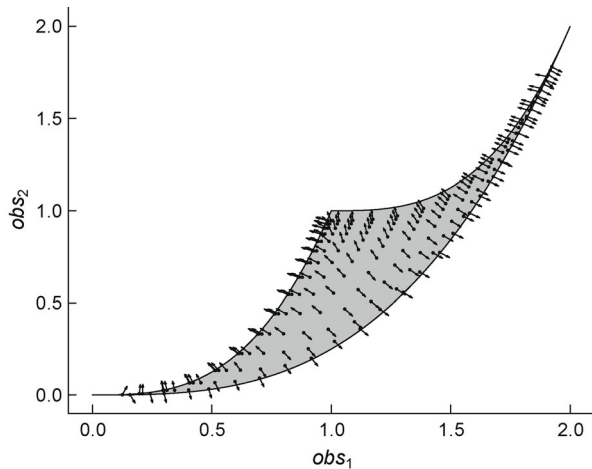

Figure 12. The initial velocities in the data space at the starting points of geodesic paths corresponding to initial parameters that map to various regions on the model manifold.

After the first iteration of MBAM, the model reduces to $x_3(t) = 2e^{-te^{\log\theta_2}}$ with only one parameter. If we perform a second iteration of MBAM with the initial log-parameter $\log\theta_2 = \log(11.5) = 2.4423$, the initial velocity is $+1$, and the geodesic path leads to the limit of $\log\theta_2 \to \infty$. Evaluating this limit further reduces the model to a constant of $x_3 = 0$, which corresponds to the bottom-left corner of the original model manifold.

Model reduction based on the profile likelihood method

The profile likelihood approach for experimental design can also be applied to perform model reduction [51]. The idea is related to the manifold boundary approximation method, but has notable differences. As mentioned in the discussions of experimental design, the profile likelihood of one parameter is a function of the parameter and is defined by solving many optimization problems with the parameter fixed at various values. The optimized values for other parameters along the profile likelihood form a trajectory in the parameter space, along a canyon defined by cross-sections of the cost surface at various values of the fixed parameter. The profile likelihood approach examines these canyons associated with the profile likelihood of all the parameters, and suggest appropriate

limits to derive reduced models. Here is a simplified procedure for model reduction based on profile likelihood:

1. Compute profile likelihood for each parameter. Basically, pick one parameter and vary its value in the feasible region, and for each value perform optimization to estimate the other parameters. The optimization objective is a least squares cost function $c = 0.5 \times \sum_i (obs_i - pred_i(\theta))^2$. Perform the same analysis for each parameter.

2. Define thresholds for acceptable likelihoods.

3. Examine the shapes of the profiles with respect to the threshold and decide on the limits to simplify the model. If both ends of a profile exceed the threshold, the corresponding parameter is considered to be identifiable and thus cannot be reduced. If neither ends of a profile exceed the threshold, the corresponding parameter is unidentifiable and can be fixed to an arbitrary value. If one end exceeds the threshold and the other stays below it, the corresponding parameter can be taken to the limit associated to the end that stays below the threshold.

The profile likelihood of the sum-of-two-exponential model is already presented in Figure 6A and 6B, along with thresholds for acceptable likelihoods. For both parameters, the profile likelihood exceeds the threshold on the left end, and stays below the threshold on the right end, indicating that both parameters can be taken to $+\infty$ without making the profile likelihood unacceptable. Since the model is symmetric with respect to the two parameters, an appropriate reduction should take either one parameter to $+\infty$, or both parameters to $+\infty$. The corresponding reduced model is either $x_3(t) = e^{-\theta_2 t}$, or a constant model $x_3(t) = 0$, corresponding to the bottom-left boundary (blue) or the bottom-left corner of the original model manifold in Figure 10. Both are reasonable reduced models for this example because the observed data sits close to the bottom-left corner of the original model manifold.

Remarks on manifold boundary and profile likelihood for model reduction

As shown in this example, both manifold boundary and profile likelihood methods perform model reduction by generating parameter trajectories that follow canyons on the cost surface, and both algorithms generate reasonable reduced models, which correspond to sub-manifolds (boundaries and corners) close to where the experimental data sits. We implement the model reduction process of this sum-of-two-exponentials example in Matlab, and provide the code and documentations. In addition, we provide model reduction analysis of enzyme catalyzed reaction dynamics, which is a larger model containing

four variables and three parameters. The Matlab code, documentations and additional example are available at http://pengqiu.gatech.edu/software/model_manifold/html/publish_ model_manifold.html.

One key difference between the two algorithms is the computational complexity. Each iteration of the manifold boundary approximation method involves one run of parameter optimization, and computation of the connection coefficient along the geodesic path, which means calculation or approximation of the first-and second-order derivatives of the model predictions with respect to the parameters along the geodesic path. The profile likelihood method is more computationally expensive. Each iteration of the profile likelihood method involves computation of N profiles. Each profile is computed by many runs of parameter optimization, with one parameter fixed at various values. Since parameter optimization is typically the most time consuming operation in these analyses, the difference in the number of parameter optimization runs suggests that the manifold boundary method is more computationally efficient compared to profile likelihood. On the other hand, the parameter estimations required in profile likelihood can be achieved by relatively standard optimization procedures, whereas the numerical integration of geodesics in the model manifold approach requires more sophisticated mathematical machinery and higher numerical precision. For very complex models, it is possible that the profile likelihood approach works but the model manifold approach fails because of numerical issues with the geodesic integration.

Another key difference between the two algorithms is whether parameter symmetry is considered. The profile likelihood can identify limits that involve taking individual parameters to their limits (0 or $\pm\infty$), and can also identify combinations of parameters that go to their limits together. However, profile likelihood is not able to identify limits related to symmetry in the system, for example the limit of $\theta_1 \rightarrow \theta_2$ in the sum-of-two-exponentials example. In contrast, the manifold boundary approximation method is able to identify limits associated to parameter symmetry and derive the corresponding reduced models. This kind of symmetry is actually quite common in systems biology and also in other engineering practices. For example, nearly all machine learning models in regression analysis and neural networks have lots of internal symmetries.

CONCLUSIONS

Mathematical modeling is a crucial tool for studying complex biological processes. In systems biology, mathematical modeling often faces the situation of highly complex models and insufficient data, making it challenging to perform parameter estimation and obtain insights

into the underlying biological mechanisms that give rise to the data. Two intuitive approaches to address this challenge are experimental design and model reduction. Given a complex model and limited data, these two approaches aim to answer what additional experiment is maximally informative, and what is the simplest mechanism to explain experimentally observed behaviors. These are problems that biologists consider on a daily basis. Although experimental design and model reduction have been largely considered as distinct problems in the literature, these two problems share deep connections that can unified them into a common framework. Here, we focus on the model manifold method and the profile likelihood method, which can tackle both problems by exploring their connections.

From the model manifold perspective, we consider a mathematical model as a manifold living in a data space, and consider the observed experimental data as a point in the data space. Parameter estimation can be viewed as projecting the data point onto the manifold. By examining the singularity around the projected point on the manifold, we can perform both experimental design and model reduction. Experimental design is to identify new experiments that expand the manifold and remove the singularity to reduce parameter uncertainty. Model reduction is to identify the nearest boundary, which is the nearest singularity that suggests an appropriate form of a reduced model.

From the profile likelihood perspective, we consider a mathematical model and observed experimental data together as an optimization problem. We can use parameter estimation techniques and sampling techniques to obtain a collection of acceptable parameters, all of which fit to the data decently well. By examining this collection of acceptable parameters, we can perform both experimental design and model reduction. Experimental design examines model predictions of new experiments based on the acceptable parameters, and identifies the new experiment with largest variations in the model predictions. Model reduction examines the range of the acceptable parameters, and identifies which parameters can be taken to the limits and hence removed, while still maintaining a decent fit.

To demonstrate how these methods work, we introduce a sum-of-two-exponentials model, and present details of both methods applied to both problems in lengthy details, along with Matlab code and documentations at a supplemental website http://pengqiu.gatech.edu/software/model_manifold/html/publish_model_manifold.html. The sum-of-two-exponentials model is simple enough to be tractable but complicated enough to be interesting. The detailed presentations are helpful for the dissemination of experimental design and model reduction ideas in the biology community, and are resources for

interested researchers to generalize these approaches to other systems.

ACKNOWLEDGMENTS

COMPLIANCE WITH ETHICS GUIDELINES

The authors Jenny E. Jeong, Qinwei Zhuang, Mark K. Transtrum, Enlu Zhou and Peng Qiu declare that they have no conflict of interests.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Lander, A. D. (2004) A calculus of purpose. PLoS Biol., 2, e164

2. Sobie, E. A., Lee, Y. S., Jenkins, S. L. and Iyengar, R. (2011) Systems biology—biomedical modeling. Sci. Signal., 4, tr2

3. Fages, F., Gay, S. and Soliman, S. (2015) Inferring reaction systems from ordinary differential equations. Theor. Comput. Sci., 599, 64–78

4. Jha, S. K. and Langmead, C. J. (2012) Exploring behaviors of stochastic differential equation models of biological systems using change of measures. BMC Bioinformatics, 13, S8

5. Kauffman, S. A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol., 22, 437–467

6. Sachs, K., Gifford, D., Jaakkola, T., Sorger, P. and Lauffenburger, D. A. (2002) Bayesian network approach to cell signaling pathway modeling. Sci. STKE, 2002, pe38

7. Koch, I. (2015) Petri nets in systems biology. Soft. Syst. Model., 14, 703–710

8. Materi, W. and Wishart, D. S. (2007) Computational systems biology in drug discovery and development: methods and applications. Drug Discov. Today, 12, 295–303

9. Machado, D., Costa, R. S., Rocha, M., Ferreira, E. C., Tidor, B. and Rocha, I. (2011) Modeling formalisms in systems biology. AMB Express, 1, 45

10. Bartocci, E. and Lió, P. (2016) Computational modeling, formal analysis, and tools for systems biology. PLoS Comput. Biol., 12, e1004591

11. Kitano, H. (2002) Computational systems biology. Nature, 420, 206–210

12. Aldridge, B. B., Burke, J. M., Lauffenburger, D. A. and Sorger, P. K. (2006) Physicochemical modelling of cell signalling pathways. Nat. Cell Biol., 8, 1195–1203

13. Anderson, J., Chang, Y. C. and Papachristodoulou, A. (2011) Model decomposition and reduction tools for large-scale networks in systems biology. Automatica, 47, 1165–1174

14. Quaiser, T., Dittrich, A., Schaper, F. and Mönnigmann, M. (2011) A simple work flow for biologically inspired model reduction —application to early JAK-STAT signaling. BMC Syst. Biol., 5, 30

15. Villaverde, A. F., Henriques, D., Smallbone, K., Bongard, S.,

16. Schmid, J., Cicin-Sain, D., Crombach, A., Saez-Rodriguez, J., Mauch, K., Balsa-Canto, E., et al. (2015) BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. BMC Syst. Biol., 9, 8

16. Machta, B. B., Chachra, R., Transtrum, M. K. and Sethna, J. P. (2013) Parameter space compression underlies emergent theories and predictive models. Science, 342, 604–607

17. Boyd, S. and Vandenberghe, L. (2004) Convex Optimization. New York: Cambridge University Press

18. Moles, C. G., Mendes, P. and Banga, J. R. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res., 13, 2467–2474

19. Ramsay, J. O., Hooker, G., Campbell, D. and Cao, J. (2007) Parameter estimation for differential equations: a generalized smoothing approach. J. R. Stat. Soc. Series B Stat. Methodol., 69, 741–796.

20. Zenker, S., Rubin, J. and Clermont, G. (2007) From inverse problems in mathematical physiology to quantitative differential diagnoses. PLoS Comput. Biol., 3, e204

21. Campbell, D. A. and Chkrebtii, O. (2013) Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates. Math. Biosci., 246, 283–292

22. Banga, J. R. and Balsa-Canto, E. (2008) Parameter estimation and optimal experimental design. Essays Biochem., 45, 195–210

23. Kreutz, C. and Timmer, J. (2009) Systems biology: experimental design. FEBS J., 276, 923–942

24. Meyer, P., Cokelaer, T., Chandran, D., Kim, K. H., Loh, P. R., Tucker, G., Lipson, M., Berger, B., Kreutz, C., Raue, A. (2014) Network topology and parameter estimation: from experimental design methods to gene regulatory network kinetics using a community based approach. BMC Syst. Biol., 8, 13

25. Apri, M., de Gee, M. and Molenaar, J. (2012) Complexity reduction preserving dynamical behavior of biochemical networks. J. Theor. Biol., 304, 16–26

26. Danø, S., Madsen, M. F., Schmidt, H. and Cedersund, G. (2006) Reduction of a biochemical model with preservation of its basic dynamic properties. FEBS J., 273, 4862–4877

27. Kourdis, P. D., Palasantza, A. G. and Goussis, D. A. (2013) Algorithmic asymptotic analysis of the NF-κB signaling system. Comput. Math. Appl., 65, 1516–1534

28. Radulescu, O., Gorban, A. N., Zinovyev, A. and Noel, V. (2012) Reduction of dynamical biochemical reactions networks in computational biology. Front. Genet., 3, 131

29. Vanlier, J., Tiemann, C. A., Hilbers, P. A. J. and van Riel, N. A. W. (2012) An integrated strategy for prediction uncertainty analysis. Bioinformatics, 28, 1130–1135

30. Vanlier, J., Tiemann, C. A., Hilbers, P. A. J. and van Riel, N. A. W. (2012) A Bayesian approach to targeted experiment design. Bioinformatics, 28, 1136–1142

31. Huan, X. and Marzouk, Y. M. (2013) Simulation-based optimal Bayesian experimental design for nonlinear systems. J. Comput. Phys., 232, 288–317

32. Pauwels, E., Lajaunie, C. and Vert, J. P. (2014) A Bayesian active

learning strategy for sequential experimental design in systems biology. BMC Syst. Biol., 8, 102

33. Liepe, J., Filippi, S., Komorowski, M. and Stumpf, M. P. H. (2013) Maximizing the information content of experiments in systems biology. PLoS Comput. Biol., 9, e1002888

34. Busetto, A. G., Hauser, A., Krummenacher, G., Sunnåker, M., Dimopoulos, S., Ong, C. S., Stelling, J. and Buhmann, J. M. (2013) Near-optimal experimental design for model selection in systems biology. Bioinformatics, 29, 2625–2632

35. Faller, D., Klingmüller, U. and Timmer, J. (2003) Simulation methods for optimal experimental design in systems biology. Simulation, 79, 717–725

36. Casey, F. P., Baird, D., Feng, Q., Gutenkunst, R. N., Waterfall, J. J., Myers, C. R., Brown, K. S., Cerione, R. A. and Sethna, J. P. (2007) Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. IET Syst. Biol., 1, 190–202

37. Krüger, R. and Heinrich, R. (2004) Model reduction and analysis of robustness for the *Wnt*/β-Catenin signal transduction pathway. Genome Inform., 15, 138–148

38. Gerdtzen, Z. P., Daoutidis, P. and Hu, W. S. (2004) Non-linear reduction for kinetic models of metabolic reaction networks. Metab. Eng., 6, 140–154

39. Vora, N. and Daoutidis, P. (2001) Nonlinear model reduction of chemical reaction systems. Aiche J., 47, 2320–2332

40. Lam, S. H. (2013) Model reductions with special CSP data. Combust. Flame, 160, 2707–2711

41. Kuo, J. C. W. and Wei, J. (1969) Lumping analysis in monomolecular reaction systems. analysis of approximately lumpable system. Ind. Eng. Chem. Fundam., 8, 124–133

42. Liao, J. C. and Lightfoot, E. N. Jr. (1988) Lumping analysis of biochemical reaction systems with time scale separation. Biotechnol. Bioeng., 31, 869–879

43. Brochot, C., Tóth, J. and Bois, F. Y. (2005) Lumping in pharmacokinetics. J. Pharmacokinet. Pharmacodyn., 32, 719–736

44. Dokoumetzidis A, Aarons L (2009) Proper lumping in systems biology models. IET Syst. Biol., 3, 40–51

45. Seigneur, C., Stephanopoulos, G. and Carr Jr., R. W. (1982) Dynamic sensitivity analysis of chemical reaction systems: a variational method. Chem. Eng. Sci., 37, 845–853

46. Turányi, T., Bérces, T. and Vajda, S. (1989) Reaction rate analysis of complex kinetic systems. Int. J. Chem. Kinet., 21, 83–99

47. Petzold, L. and Zhu, W. (1999) Model reduction for chemical kinetics: an optimization approach. Aiche J., 45, 869–886

48. Liu, G., Swihart, M. T. and Neelamegham, S. (2005) Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling. Bioinformatics, 21, 1194–1202

49. Schmidt, H., Madsen, M. F., Danø, S. and Cedersund, G. (2008) Complexity reduction of biochemical rate expressions. Bioinformatics, 24, 848–854

50. Steiert, B., Raue, A., Timmer, J. and Kreutz, C. (2012) Experimental design for parameter estimation of gene regulatory networks. PLoS One, 7, e40052

51. Maiwald, T., Hass, H., Steiert, B., Vanlier, J., Engesser, R., Raue, A., Kipkeew, F., Bock, H. H., Kaschek, D., Kreutz, C., *et al.* (2016) Driving the model to its limit: profile likelihood based model reduction. PLoS One, 11, e0162366

52. Transtrum, M. K. and Qiu, P. (2012) Optimal experiment selection for parameter estimation in biological differential equation models. BMC Bioinformatics, 13, 181

53. Transtrum, M. K. and Qiu, P. (2014) Model reduction by manifold boundaries. Phys. Rev. Lett., 113, 098701

54. Transtrum, M. K. and Qiu, P. (2016) Bridging mechanistic and phenomenological models of complex biological systems. PLoS Comput. Biol., 12, e1004915

55. Kutalik, Z., Cho, K. H. and Wolkenhauer, O. (2004) Optimal sampling time selection for parameter estimation in dynamic pathway modeling. Biosystems, 75, 43–55

56. Bandara, S., Schlöder, J. P., Eils, R., Bock, H. G. and Meyer, T. (2009) Optimal experimental design for parameter estimation of a cell signaling model. PLoS Comput. Biol., 5, e1000558

57. Hagen, D. R., White, J. K. and Tidor, B. (2013) Convergence in parameters and predictions using computational experimental design. Interface Focus, 3, 20130008

58. Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M. P. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface, 6, 187–202

59. Frieden, B.R. (2000) Physics from fisher information: a unification. Am. J. Phys., 68, 1064–1065

60. Transtrum, M. K., Machta, B. B. and Sethna, J. P. (2011) Geometry of nonlinear least squares with applications to sloppy models and optimization. Phys. Rev. E Stat. Nonlin. Soft Matter Phys., 83, 036701

61. Leis, J. R. and Kramer, M. A. (1988) The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. ACM Trans. Math. Softw., 14, 45–60

62. Kumar, A., Christofides, P. D. and Daoutidis, P. (1998) Singular perturbation modeling of nonlinear processes with nonexplicit time-scale multiplicity. Chem. Eng. Sci., 53, 1491–1504

63. Snowden, T. J., van der Graaf, P. H. and Tindall, M. J. (2017) Methods of model reduction for large-scale biological systems: a survey of current methods and trends. Bull. Math. Biol., 79, 1449–1486

64. Heinrich, R. and Schuster, S. (1996) The Regulation of Cellular Systems. Springer: New York

65. Voit, E. (2012) A First Course in Systems Biology. 1st ed., Garland Science: New York

66. Okino, M. S. and Mavrovouniotis, M. L. (1998) Simplification of mathematical models of chemical reaction systems. Chem. Rev., 98, 391–408

67. Wolf, J. and Heinrich, R. (2000) Effect of cellular interaction on glycolytic oscillations in yeast: a theoretical investigation. Biochem. J., 345, 321–334

68. Sauter, T., Gilles, E. D., Allgöwer, F., Saez-Rodriguez, J., Conzelmann, H. and Bullinger, E. (2004) Reduction of mathema-

tical models of signal transduction networks: simulation-based approach applied to EGF receptor signalling. Syst. Biol. (Stevenage), 1, 159–169

69. Liebermeister, W., Baur, U. and Klipp, E. (2005) Biochemical network models simplified by balanced truncation. FEBS J., 272, 4034–4043

70. Maertens, J., Donckels, B., Lequeux, G. and Vanrolleghem, P. (2005) Metabolic model reduction by metabolite pooling on the basis of dynamic phase planes and metabolite correlation analysis. In Proceedings of the Conference on Modeling and Simulation in Biology, Medicine and Biomedical Engineering. Linkping, Sweden