Optimal sampling time selection for parameter estimation in dynamic pathway modeling

Zoltán Kutalik\textsuperscript{a}, Kwang-Hyun Cho\textsuperscript{b}, Olaf Wolkenhauer\textsuperscript{c,}\textsuperscript{*}

\textsuperscript{a} School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK
\textsuperscript{b} School of Electrical Engineering, University of Ulsan, Ulsan 680-749, South Korea
\textsuperscript{c} Department of Computer Science, University of Rostock, Albert Einstein Str. 21, Rostock 18059, Germany

Abstract

Systems Biology is an emerging research area, which considers mathematical representations of inter- and intra-cellular dynamics. Among the many research problems that have been addressed, dynamic modeling of signal transduction pathways has received increasing attention. The usual approach to represent intra-cellular dynamics are nonlinear, usually ordinary, differential equations. The purpose of the models is to test and generate hypothesis of specific pathways and it is therefore required to estimate model parameters from experimental data. The experiments to generate data are complex and expensive, as a consequence of which the time series available are usually rather short, with few if any replicates. Almost certainly, not all variables one would like to include in a model can be measured. Parameter estimation is therefore an important research problem in Systems Biology and the focus of this paper. In particular, we are interested in optimizing the sampling time selection in order to minimize the variance of the parameter estimation error. With few sampling time points feasible, their selection is of practical importance in experimental design. Finally, the theoretical results are supported with an application.

© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: Systems Biology; Dynamic pathway modeling; Experimental design; Multiple shooting; Parameter estimation

1. Introduction

Systems Biology investigates how genes, proteins, metabolites and cells interact and react to changes in their environment (Wolkenhauer et al., 2003a,b). Modeling the dynamics of signal transduction pathways has received increasing attention in the last 3 years. Cell signaling, or "signal transduction," is the study of the mechanisms by which this transfer of biological information comes about. Signaling impinges on all aspects of biology, from development to disease.

Many diseases, such as cancer, involve malfunction of signal transduction pathways (Downward, 2001). In Asthagiri and Lauffenburger (2001) feedback effects on signal dynamics in a MAPK pathway model are investigated. In Schoeberl et al. (2002) the focus is on surface and internalized EGF receptors. The simulation of signal cascades induced by EGF showed good agreement with experimental data despite the need to complement experimental data for parameter estimation with information from the literature. The work described in Swameye et al. (2003) is another example where mathematical modeling and parameter estimation have been successful in studying signal transduction pathways. Like the models of the TNF\textsuperscript{a} mediated NF-\textsuperscript{a}B pathway described in Cho et al. (2003a,b), most approaches to this day, use nonlinear ordinary
differential equations (ODEs) to describe cellular dynamics. Alternative representations that have been proposed are process algebras such as π-calculus (Regev et al., 2001). Although the estimation of parameters for nonlinear, delayed differential equations is in itself an active research area, the multiple-shooting method (MSM) (Bock, 1981, 1983) has been shown to work well in various applications (Timmer, 1998; Müller, 2002; Müller et al., 2002) and therefore forms the basis of this paper. The principal aim of this paper is to gain insight into the hidden relationship between the selection of sampling time points and the variance of the parameter estimation error induced by the random noise inherent in the data. From this investigation, an approach to choose sampling time to optimally reduce the variation of parameter estimates is developed. Because the experiments to generate data for parameter estimation are time consuming and expensive, an optimal choice of sampling times has practical value in experimental design.1

Here we consider ordinary differential equations as the basis for modeling bio-reactions. The inverse problem is to identify the parameters of a particular system from experimental time series data. A major challenge to dynamic pathway modeling is that for data on only few time points are available. Replicate experiments (which could help us to establish a noise model) are often the subject of hard fought negotiations between dry-lab and wet-lab collaborators. One approach to improve parameter estimates is to increase the accuracy of the numerical methods used and to handle the noise in measurements more efficiently. Multiple shooting is one approach to parameter estimation which has been successfully applied to a range of applications. The idea is derived from the numerical solution for the boundary value problem with known parameters as described in Stoer and Bulirsch (2002) and has been extended to the case when as well the parameters of the system are unknown (Müller, 2002; Bock, 1981, 1983). Here we develop a concept of an improved experimental design to enhance parameter estimation. In Baltes et al. (1994) the authors introduced a control function into the system to reduce the variance of the parameter estimates. In contrast, we focus on the proper choice of sampling times to optimally reduce the variation of the parameter estimates. It has to be emphasized that the term optimal experimental design refers to a substantial enhancement of the parameter estimation error by finding a suboptimal sampling time selection. The proposed method is illustrated with a simple set of bio-kinetic reactions.

The outline of this paper is as follows. In Section 2, we provide a short summary of parameter estimation for modeling of signal transduction pathways. Then Section 3 introduces the proposed approach to experimental design. The summary of the results and discussion are in Sections 4 and 5, respectively. All algorithms were implemented in Matlab and are available at http://www.sbi.uni-rostock.de/.

2. Signal transduction pathway modeling and parameter estimation

To develop the idea of multiple shooting in general setting, we consider a dynamic system modeled by ODEs. Given the function $f : \mathbb{R}^{n+1} \to \mathbb{R}^n$ let us assume that function $x : \mathbb{R} \to \mathbb{R}^n$ satisfies the following system of ODEs:

$$x(t) = f(t, x(t), k_0), \quad x(t_0) = x_0,$$

where $k_0 \in \mathbb{R}^m$ are the parameters of the differential equation and $x_0 \in \mathbb{R}^n$ are the initial values of the system. For simplicity, we assume that the system $x(t, p_0)$ is measured directly at certain time points, $t_1, \ldots, t_N \in [T_0, T_1]$. Allowing for measurement noise, we have:

$$x(t_i) = x_i(t_i) + \epsilon_i, \quad i = 1, \ldots, N,$$

where

$$x_i \sim \mathcal{N}(0, \sigma^2).$$

In this paper, optimal experimental design refers to a substantial enhancement of the parameter estimation error by finding a suboptimal sampling time selection.

In the present paper we will use the following notation: $x, \dot{x}, \ddot{x}$ refer to the first, second and third time derivatives of function $x$, respectively. Vectors will be denoted by bold fonts.
In Müller (2002) and Baltes et al. (1994) the authors describe the more general situation in which the states are not directly observable, introducing an observation function $g$, for which only the $g(x(t))$ values are accessible. In the present paper the noise, $\epsilon$, is assumed to be normally distributed with common variance, $\sigma^2$. We thereby simplify the original approach (Müller, 2002; Swameye et al., 2003) to learn more about the information carried in the choice of time points of the measurement, rather than complicating the case with different variance of noise occurring in the experiment. Our goal is to identify the true parameter values, rather than complicating the case with different variance of noise occurring in the experiment. Therefore neither is a perfect measurement, for which only the $x_j(t_i)$ values are acquired from noisy data, and therefore to provide an optimal set of time points to reduce the absolute value of the relative error of the estimate $|I[p_0 - p_0]|/p_0$. To this end one has to to approximate the covariance matrix of the estimated parameters.

2.1. Variance of the estimated parameters

Maximum likelihood estimation is one of the most robust statistical approaches to parameter estimation. When the noise is assumed to be normally distributed the maximization problem reduces to the minimization of the following function:

$$\chi^2(p) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{n} \left( \frac{x_i(t_j) - g(x_j(t_i), p)}{\sigma} \right)^2.$$  \hspace{1cm} (4)

Rather than minimizing (4) directly, we linearize function $x_j(t_i)$ around $p_0$, and minimize this simpler function. Using the Taylor series expansion (Rudin, 1976) of $x_j(t_i, p)$ around $p_0$:

$$x_j(t_i, p) = x_j(t_i, p_0) + \nabla_p x_j(t_i, p_0) \Delta p + o(\Delta p),$$  \hspace{1cm} (5)

where $\Delta p = p - p_0$ and $\nabla$ is the so-called nabla differential operator, i.e.,

$$\nabla_p = \left( \frac{\partial}{\partial p_1}, \ldots, \frac{\partial}{\partial p_n} \right)^\top.$$

Substituting (5) to (4) yields

$$\chi^2(p) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{n} \left( \frac{\sigma^2}{\sigma} - 2 \sigma \nabla_p x_j(t_i, p_0)^\top \Delta p + \Delta p^\top \nabla_p x_j(t_i, p_0) \nabla_p x_j(t_i, p_0)^\top \Delta p + o(\Delta p) \right).$$  \hspace{1cm} (6)

To measure the accuracy of the estimate we like to summarize the information about the variability in the covariance matrix into a single number. Here we use the determinant of the matrix as the function that transforms a matrix into a scalar. This is quite informative as in fact it relates to the volume of the multidimensional simplex defined by the column/row vectors of the matrix. In Baltes et al. (1994) it was suggested to use the ratio of the largest and the smallest eigenvalue of the matrix. As the determinant is the product of the eigenvalues, these two measures bear common roots. Since the ratio of eigenvalues does not change if we multiply the matrix with a scalar, this functional is unable to reflect an approximately uniform increase in all the parameter noises. On the other hand, the determinant can nearly vanish even if only one of the noises is close to zero, which may hide the information about the other noises. Therefore neither is a perfect summary of the Fischer information matrix. Further comparison of these functionals (det and $\lambda_{max}/\lambda_{min}$) lies outside the scope of this paper. As a compromise, we used the determinant and at each step additionally $\lambda_{max}/\lambda_{min}$ was controlled to not exceed a certain limit.

$$\lim_{h \to 0} \frac{w(h)}{h^2} = 0.$$
To put the theory developed to this point into practice, a robust solution to the minimization problem (4) is summarized in the next section.

2.2. Numerical methods for parameter estimation

Here we are concentrating on the MSM (Stoer and Bulirsch, 2002; Bock, 1981, 1983). However, since extended multiple shooting has been shown to outperform the usual multiple shooting (Müller, 2002), our focus is primarily not on the numerical part of parameter estimation, but on the optimal sampling time selection.

What follows is a brief introduction to the MSM. Our aim is to minimize the function \( \chi^2(p) \) defined in (4). We treat the given data as \( N-1 \) boundary value problems. Starting off with initial estimates for the unknown parameters, \( k_0 \) and for the initial estimates for the system \( \{x_0(t_i)\}_{i=1}^N \), in the \( i \)th iteration we compute the following quantities:

\[
\begin{align*}
\gamma^{(i)}(t_i) = & \begin{cases}
\gamma^{(i)}(t_i) = x^{(i)}(t_i), \quad i = 1, \\
\gamma^{(i)}(t_{i-1}) + \int_{t_{i-1}}^{t_i} f(t, x^{(i)}(t), k^{(i)}_0) \, dt, \quad i = 2, \ldots, N,
\end{cases}
\end{align*}
\]

For the given data as \( N-1 \) boundary value problems. Starting off with initial estimates for the unknown parameters, \( k_0 \) and for the initial estimates for the system \( \{x_0(t_i)\}_{i=1}^N \), in the \( i \)th iteration we compute the following quantities:

\[
\begin{align*}
F_{i}(k_0, x^{(i)}(t_1), \ldots, x^{(i)}(t_N)) &= y^{(i)}(t_i) - x^{(i)}(t_i), \quad i = 1, \ldots, N, \quad (11) \\
G_{i}(k_0, x^{(i)}(t_1), \ldots, x^{(i)}(t_N)) &= y^{(i)}(t_i) - x^{(i)}(t_i), \quad i = 2, \ldots, N, \quad (12)
\end{align*}
\]

where \( F : \mathbb{R}^{nN+m} \rightarrow \mathbb{R}^{n(N-1)} \) and \( G : \mathbb{R}^{nN+m} \rightarrow \mathbb{R}^{nN+m} \). The integration in (10) is carried out numerically. In our implementation the Runge–Kutta method was used. Having computed these functions one faces the following constrained minimization problem

\[
\min_{v} \|F(v)\|_2 \quad \text{subject to} \quad \|G(v)\|_2 = 0. \quad (13)
\]

Function \( G \) describes the continuity constraints which guarantee that the curve is continuous. The idea not to treat this condition as strictly as the single shooting method, but keeping in mind that the curve has to be close to the measurement points, as controlled by the minimization of function \( F \). Various methods have been designed to perform this optimization in (13) (Stoer and Bulirsch, 2002; Gill and Murray, 1974; Bertsekas, 1982; Bock, 1981).

Fig. 1. The convergence of multiple shooting algorithm for noisy Lotka–Volterra differential equations. The thin dark line is the real solution, dark dots represent the measurements used for parameter estimation, while thick grey line is the fitted curve in each step of the algorithm.
followed (Stoer and Bulirsch, 2002; Bock, 1981) by locally linearizing the functions $F$ and $G$. For the linearly constrained least squares optimization we used the *lsqlin* function in Matlab’s optimization toolbox. The optimization of (13) yields new values for the parameters

$$\{x_1^{u+1}(t_i)\}_{i=1}^N$$ and $$k_1^{u+1},$$

which are then used for the initialization of the $(u + 1)$th iteration step. Iterations continue until the termination criterion $\|G\|_2 < \delta$ is met. The initial, the fourth, and the twelfth steps of the algorithm are illustrated in Fig. 1. One can see how both continuity and the accuracy (in a least squares sense) of the data improve as the algorithm converges.

In parameter estimation the most dominant error arises from the uncertainty associated with noise in the observation of a system. This error can be reduced or at least accounted for through a statistical model. The design of experiments plays then an important role, especially if the costs and logistics of experiments in signal transduction studies prevent us from measuring at a large number of sampling points. In the following section we therefore focus on experimental design, and specifically investigate the possibility to choose sampling instances in a relatively optimal way.

### 3. Optimization of sampling time selection

Time course experiments in molecular biology are rarely producing large and accurate data sets. These experiments are time consuming and expensive. Despite the availability of increasingly sophisticated technology to identify protein interactions, data usually provide an indirect reflection of the true intra-cellular processes. The design of experiments is subsequently an important issue and covers various aspects. Unfortunately, we cannot provide an optimal solution and in fact require for our approach that a simple data set is already available before a new design of experiments is considered. It also has to be emphasized that the term *optimal experimental design* refers to a substantial enhancement of the parameter estimation error by finding a suboptimal sampling time selection.

For signal transduction studies a relatively small number of time points are used for sampling. The sampling intervals are usually not evenly spaced and based on heuristics. Here we are to investigate an approach to guide the process of selecting time points in a relatively optimal way to minimize the variance of parameter estimates. In our study we first define a time interval from which we are going to select a fixed number of time points. Naturally, for sampling points that are very close, replicate experiments should be considered.

As can be seen in (9) the standard deviation of the parameters is a linear function of the standard deviation of the noise $\sigma$. The information transferred from the experiment will be expressed by $\text{det}(F)$, where matrix $F$ was introduced in (8). The larger the determinant of $F$, the more information can be extracted about the parameters. Hereafter the term *information*—by agreement—refers to $\text{det}(F)$. Since there is no complete ordering among the Fischer information matrices, the multi-objective minimization (i.e. to minimize the variances of the all parameters) is turned into the maximization of $\text{det}(F)$.

Let us denote by $z(\cdot)$ the determinant of $F$ as a function of the sampling time points:

$$z(t_1, \ldots, t_N) = \text{det}(F(t_1, \ldots, t_N)).$$

We then have the following maximization problem to solve

$$\text{max}(z(t_1, \ldots, t_N))$$ subject to $\{t_j\}_{j=1}^N \in [T_0, T_1]^N \subseteq \mathbb{R}^N$. (14)

Without the constraints in (14) one would use differential calculus to find the maximum. Let $m \in \{1, \ldots, N\}$ and compute the partial derivatives of function $z$:

$$\frac{\partial}{\partial t_m} z(t_1, \ldots, t_N) = \frac{\partial}{\partial t_m} \text{det}

= \text{det} \left[ \sum_{j=1}^{N} \sum_{l=1}^{N} \frac{\partial^2}{\partial t_m \partial t_l} \text{det} (F(t_1, \ldots, t_N)) \right]

= \text{det} \left[ \sum_{j=1}^{N} \sum_{l=1}^{N} \frac{\partial^2}{\partial t_m \partial t_l} \text{det} (F(t_1, \ldots, t_N)) + \sum_{j=1}^{N} \frac{\partial^2}{\partial t_m \partial t_l} \text{det} (F(t_1, \ldots, t_N)) \right].$$

[... content continues ...]
Let us denote

\[ A = \sum_{i \in \mathbb{N}, j = 1}^{n} \nabla p x_j(t_i, p_0) \nabla p x_j(t_i, p_0)^\top. \]

Therefore

\[ \frac{\partial}{\partial x}(z(t)) = 2 \cdot \det(A) \sum_{j = 1}^{n} \nabla p f_j(t_0, x(t_0, p_0))^\top \]

\[ \times A^{-1} \nabla p x_j(t_0, p_0), \]

which is proven in Appendix A. Let \( H(t_1, \ldots, t_n) \) denote the Hessian of \( z \), i.e., \( H(t_0, t_1, \ldots, t_n) = \left( \frac{\partial^2}{\partial x \partial \theta} z(t_1, \ldots, t_n) \right) \). By differentiation of (15) the diagonal of the Hessian is

\[ \frac{\partial^2}{\partial x \partial \theta} z(t) = 2 \cdot \det(A) \sum_{j = 1}^{n} \nabla p f_j(t_0, x(t_0, p_0))^\top \]

\[ \times A^{-1} \nabla p x_j(t_0, p_0) + 2 \cdot \det(A) \sum_{j = 1}^{n} \nabla p f_j(t_0, x(t_0, p_0))^\top A^{-1} \nabla p f_j(t_0, x(t_0, p_0)), \]

where

\[ \nabla p f_j(t_0, x(t_0, p_0)) = \frac{\partial}{\partial \theta} f_j(t_0, x(t_0, p_0)) \]

\[ + \frac{\partial}{\partial x} f_j(t_0, x(t_0, p_0)) \cdot f_j(t_0, x(t_0, p_0)) \]

and the non-diagonal elements are \( \{ j \neq m \} \)

\[ \frac{\partial^2}{\partial x \partial \theta} z(t) = 2 \cdot \det(B) \left( \sum_{j = 1}^{n} \nabla p f_j(t_1, x(t_1, p_0))^\top B^{-1} \nabla p x_j(t_1, p_0) \right) \]

\[ \times \left[ \sum_{j = 1}^{m} \nabla p f_j(t_0, x(t_0, p_0))^\top A^{-1} \nabla x_j(t_0, p_0) \right] \]

\[ + 2 \cdot \det(A) \sum_{j = 1}^{n} \nabla p f_j(t_0, x(t_0, p_0))^\top A^{-1} \nabla x_j(t_0, p_0) \]

\[ \times \frac{\partial}{\partial x} z(t), \]

where

\[ \frac{\partial}{\partial x} A^{-1} = \sum_{j = 1}^{n} \nabla p f_j(t_0, x(t_0, p_0))^\top A^{-1} \]

\[ + \nabla p x_j(t_0, p_0) \nabla p f_j(t_0, x(t_0, p_0))^\top A^{-1} \]

and

\[ B = \sum_{i \in \mathbb{N}, j = 1}^{n} \nabla p x_j(t_1, p_0) \nabla p x_j(t_0, p_0)^\top. \]

Having these quantities one can use Newton’s method to find the maximum. This iterative method is carried out as follows. An initial guess for the maximum is denoted by \( t' \) and the next point is obtained by:

\[ t'' = t' - \lambda_s \cdot \nabla', \]

where

\[ \nabla' = H(t')^{-1} \nabla z(t') \]

and \( \lambda_s \) is defined as suggested in Stoer and Bulirsch (2002) and Bock (1981). This computation is rather cumbersome and the constraints make it even more difficult to handle.

Alternatively, we can use Powell’s quadratically convergent method (Press et al., 1987). Although the method here yields only a suboptimum due to the constraints on the search space, this sub optimum can easily coincide with the optimum. A possible reason for this is that the optimum point—according to our trials—usually occurred on the boundary of the multidimensional rectangle \([T_0, T_1]^T\). Fig. 2 illustrates this in three dimensions. Originally all 16 time points were picked equidistantly. Looking at the figure, we start with the 15th measurement fixed at 9.39 min and begin to decrease the location of the 10th time point,
beginning at 6 min. The determinant of the inverse covariance matrix increases nearly 1.5 times (in other words, the determinant of the covariance matrix reduces to two-third). The thick black line shows how Powell’s algorithm evolves to find the maximum. The maximization can radically increase the determinant of the inverse covariance matrix, which will be demonstrated in Section 4.

One could argue that in practice we do not know the true parameters, without which how can we construct the covariance matrix? Here we assume an experiment was carried out before with a few (equidistant) time points. From this we obtain an estimate for the parameter values by using MSM described in the last part of Section 2. Using these estimates to reconstruct the covariance matrix provides similarly good results, since the reconstructed function $x$ used in the covariance matrix closely agrees with the exact solution even if the parameter estimates are relatively poor. This is guaranteed by the nature of the multiple shooting method, which does not allow neither the objective (minimization) function, nor the constraint equation to dominate the convergence. To support it with formula, since $x$ is differentiable with respect to the parameters and the determinant function is differentiable too, $\det(F(\tilde{p}))$ is differentiable. Consequently there exists a constant $M$ such that

$$\left| \det(F(t, \tilde{p})) - \det(F(t, p_0)) \right| < |\Delta p| \cdot M$$

for all $t \in [T_0, T_1]^N$, (20)

where $\tilde{p}$ is the estimate for the parameter vector $p_0$, and $\Delta p = \tilde{p} - p_0$. Let $K$ denote

$$K = \det(F(t^*, \tilde{p})) - \det(F(t^0, \tilde{p}))$$

(21)

where $t^*$ is the suboptimal sampling time vector given by our method, while $t^0$ is the original sampling time vector. Hence our method is successfully applicable when the inequality $K > |\Delta p| \cdot M$ holds. Typically $K$ is greater than $|\Delta p| \cdot M$ by orders of magnitude.

Therefore, after an initial trial, one can elaborate the main experimental design for more time points. The efficiency of this approach is discussed in the following section.
4. Simulation studies for a single step signal transduction module

Our approach can be used in two ways, both of which will be illustrated by a simple example. Either to determine that for a given number of measurements how many time points and at each time point how many replicates we need such that the number of measurements totals the previously specified; or given a tolerance level for the variance of the parameter estimates, it finds the minimal number of measurements and its precise layout.

To provide a simple example for our investigation, we are going to model a single step in a signal transduction pathway cascade and represent this module in analogy to an enzyme kinetic reaction,

\[ S + E \xrightarrow{k_1} ES \xrightarrow{k_2} E + P. \]

Suppose that a given enzyme combines with a substrate to form an enzyme-substrate complex with a rate constant \( k_1 \). The complex holds two possible outcomes in the next step. It can become dissociated with a rate constant \( k_2 \), or it can further proceed to form a product with a rate constant \( k_3 \). It is assumed that none of the products reverts to the initial substrate. It is required to express the relations between the rate of catalysis and the change of concentration for the substrate, the enzyme, the complex, and the product.

In Cho et al. (2003a,b) and Wolkenhauer et al. (2003a,b) we established complex pathway models that are composed of modules, each of which is represented by equations akin to a basic enzyme-kinetic reaction. The presented example is therefore to illustrate the concept with one of those modules, from which more complex pathway structures can be covered.

Based on these reaction kinetics (Robert and Tom, 2001; Cho et al., 2003a,b), illustrated in Fig. 3, we obtain the following set of ODEs

\[
\begin{align*}
\dot{x}_1 &= -k_1 x_1 x_3 + k_2 x_3 \\
\dot{x}_2 &= -k_1 x_1 x_2 + (k_2 + k_3) x_3 \\
\dot{x}_3 &= k_1 x_1 x_2 - (k_2 + k_3) x_3 \\
\dot{x}_4 &= k_3 x_3,
\end{align*}
\]

Here \( x_1 \) refers to the “enzyme” concentration, \( x_2 \) to that of the “substrate,” \( x_3 \) to the “enzyme-substrate” complex, while \( x_4 \) denotes the concentration of the “product.” For simulation studies, we set the parameter values as follows: \( T_0 = 0, T_1 = 10, x_1(0) = 12, x_2(0) = 12, x_3(0) = 0, x_4(0) = 0, k_1 = 0.18, k_2 = 0.02, k_3 = 0.23 \). First, the concentration curves were simulated with a normally distributed noise, standard deviation 0.2, added to the process. As discussed earlier, the magnitude of the noise is not important, since the relationship between the variance of the parameter estimates and the variance of the noise is linear. We first sampled the concentration at 16 equidistant time points in the domain \([0, 10]\), then we applied the MSM. This gave the estimates for the parameters \( k_1 = 0.1768, k_2 = 0.0155, k_3 = 0.2322 \). The reconstruction of the concentration profiles for each protein can be seen in Fig. 4. The covariance matrix was computed based on the estimated parameter values. As described in Section 3, the determinant of the covariance matrix was optimized by Powell’s quadratically convergent minimization method to yield the suboptimal set of time points to minimize the variance of the discrepancy between the real and the estimated parameter values of the system. The subsequently selected sampling time points for experimental design are summarized in Table 1 and its efficiency is shown in Fig. 5.

![Fig. 3. A template pathway modeling block for single-step signal transduction pathway. The pathway model can be constructed from basic reaction modules like this enzyme kinetic reaction for which a set of four ODEs is required.](image_url)

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Enhanced experimental design for sampling time points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time points (min)</td>
<td>0</td>
</tr>
<tr>
<td>Number of replicates</td>
<td>5</td>
</tr>
</tbody>
</table>
In general the interpretation of the enhanced sampling time selection depends on the nature of the experiment. Since $\det(F)$ is a continuous function of the parameters, slight modifications in the suggested sampling times do not influence the performance substantially. When subsequent measurements can follow each other practically without any break, there is no constraint on the minimal time elapsed between adjacent time points. In this case, replicated measurements (suggested by our method) may be substituted by measurements carried out in rapid succession. Very often it is not the case and more measurements at the same time points should be interpreted as the replication of the experiment. The strong point of our approach is that it can decide in every situation whether replicated measurements, or the introduction of new time points would be more beneficial to the accuracy of the estimates.

The set of time points are shown vertically at each iteration step. As can be seen at the initial iteration, the time points were scattered equidistantly between zero and ten, while in the last iteration step they converged to the (sub)optimum detailed in Table 1. The continuous curve is the determinant corresponding to the actual set of sampling time points at the specified iteration step. Since the condition number of the covariance matrix did not tend to infinity as the number of measurements increased, the parameters are locally identifiable (Müller et al., 2002).

In the second step the original (noisy) process was then sampled at the new set of sampling time points (summarized in Table 1). MSM was applied for these measurements and produced new estimates for the parameters of the system: $\hat{k}_1 = 0.1779, \hat{k}_2 = 0.0231, \hat{k}_3 = 0.2296$. The relative error, defined in Section 2 compared to the one obtained from the original design, was improved considerably, as can be seen in Table 2. Here we considered only one particular realization of an experiment where the optimization led to a better estimate for the parameters. In general, we can also characterize its expected performance, i.e. the reduction of variance on average. As seen in Fig. 5, the

![Fig. 4. Reconstructed concentration curves for the enzyme kinetic reaction. Stars, triangles, dots and plus signs refer to 'measured' data.](image-url)
As has been mentioned in Section 2.1, the approach can be extended to heterogeneous noise, i.e. where \( \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{ij}) \). The Fisher information matrix changes accordingly:

$$ F = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{\sigma_{ij}^2} \nabla^2_{p} \ln \mathcal{L}(\theta, \mathbf{P}) \nabla_p \theta_i | \theta_0, \mathbf{P} \right]^T. $$ (26)

To this end we ran a simulation where the standard deviation of the measurement noise was proportional to the measured value. Since we would like to avoid zero variance, slightly changed the initial parameter values: \( x_3(0) = 1, x_4(0) = 1 \), the rest remained unchanged. Then the concentration curves were simulated with a normally distributed noise, standard deviation 15% of the true values, added to the process. Having sampled the concentration at 16 equidistant time points in the time domain \([0, 10]\), multiple shooting was applied. The estimates for the parameters \( \hat{k}_1 = 0.2025, \hat{k}_2 = 0.0377, \hat{k}_3 = 0.2168 \) are

\begin{table}[h]
\centering
\caption{Comparison of the parameter estimates according to the sampling time points}
\begin{tabular}{|l|c|c|c|}
\hline
 & \( k_1 \) & \( k_2 \) & \( k_3 \) \\
\hline
Real values & 0.1800 & 0.2200 & 0.2500 \\
Estimate: & & & \\
Equidistant & 0.1768 & 0.2155 & 0.2322 \\
Enhanced & 0.1779 & 0.231 & 0.2296 \\
Relative error & & & \\
Equidistant & -0.0177 & -0.0250 & 0.0095 \\
Enhanced & -0.0117 & 0.135 & -0.0017 \\
\hline
\end{tabular}
\end{table}
obtained along with the reconstructed concentration profiles for each protein. The profiles with the simulated observations and their “confidence intervals” (value ± estimated standard deviation) can be seen in Fig. 6. The covariance matrix was computed based on the estimated parameter values. The determinant of the covariance matrix was optimized to yield the suboptimal set of time points which (locally) minimizes the variance of the estimated parameter values. The subsequently selected sampling time points for experimental design are summarized in Table 3. To provide a more complete picture, instead of a single realization of the experiment we now compute the exact standard deviation of the parameters $k_1$, $k_2$, $k_3$ in case of equidistant and enhanced sampling time points. These values can be obtained from the diagonal of the inverse Fischer information matrix. The average discrepancies from the real values can be read in Table 4. From the comparison of the standard deviations of the parameter estimates, our approach was able to reduce the noise of the estimates to 20% on average. This means that carrying out the initial experiment with equidistant time points, then repeating the experiment with the enhanced time points results in more than double the accuracy in parameter estimation than simply repeating the initial experiment with the equidistant sampling time points twice over. In other form, 36 repetitions of 16 equidistant measurements produce the same accu-

Table 3
Enhanced experimental design for sampling time points in case of heterogeneously distributed noise

<table>
<thead>
<tr>
<th>Time points (min)</th>
<th>0.00</th>
<th>0.23</th>
<th>0.82</th>
<th>0.93</th>
<th>2.30</th>
<th>2.53</th>
<th>6.30</th>
<th>8.50</th>
<th>9.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of replicates</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 6. Reconstructed concentration curves for the enzyme kinetic reaction with heterogeneous noise. Dots refer to “measured” data, and for each dot the standard deviation of the “measurement” noise is also shown.
racy as 16 equidistant and 16 optimal ones together. Moreover, with our enhanced time point selection the volume of the 95% confidence ellipsoid of the parameter triplet \((k_1, k_2, k_3)\) has shrunk to 6.4% compared to that in the replicated equidistant case. These aforementioned facts excellently demonstrate that our approach substantially reduces the experimental costs and significantly improves the estimation accuracy at the same time.

5. Conclusions

Modeling cellular dynamics based on experimental data is at the heart of Systems Biology. If the dynamics can be represented by ODEs, the first step in modeling is to decide upon the structure of the equations. The most important next step is to identify the parameters of the system from experimental data. Here we investigated multiple shooting both for parameter estimation and application in experimental design. Considering the various applications that exist for the MSM and from our study of these ideas, multiple-shooting based methods are a very good solution for parameter estimation in dynamic pathway modeling. Based on these ideas we developed a concept for time point selection in measurements.

As our study has shown, the impact of optimization is very strong on the accuracy of parameter estimates. In spite of all inaccuracies, such as turning the Fischer information matrix into \(\det(F)\), using only a reconstructed function \(x\), or finding only a sub-optimum of \(\det(F)\), it still provides a significantly better sampling time selection, where the quality of the parameter estimates are by far more precise. We illustrated the significance of our approach through applications with homogeneous and heterogeneous noise; with mean performance and with single realization. This study yielded that time point selection plays a crucial role in parameter estimation. It also has to be emphasized that with our approach—even including the initial experiment—altogether much less measurements are needed to achieve a certain level of accuracy (in parameter estimation) than with equidistant, or any ad hoc sampling. Furthermore, the advantage of our approach becomes more significant as the number of allowed measurements increases.

Although the approach to select optimal sampling time points relies on an existing data set, we claim to provide a solution that will substantially reduce the costs of experiments, and enhance the information extracted from the measurements. Moreover, at the beginning of interdisciplinary projects it is often the case that some data exist, based on which one designs further experiments.

Acknowledgements

This work was funded as a part of the DEFRA project ‘Application of post-genomics to \(M. bovis\).’ The authors would like to thank Dr. Jens Timmer for his helpful comments on the paper. Kwang-Hyun Cho acknowledges the support received by a grant from the Korea Ministry of Science and Technology (Korean Systems Biology Research Grant, M10309000006-03B5000-00211) and also by grant No. (R05-2004-000-10549-0) from the Korea Ministry of Science & Technology.

Appendix A. Determinant properties

Suppose that \(A \in \mathbb{R}^{r \times r}\) and invertible, then

\[
\det(A) = \prod_{i=1}^{r} \lambda_i
\]

(A.1)

holds. Where as usual \(\{\lambda_i\}_{i=1}^{r}\) are the eigenvectors of matrix \(A\). Furthermore it is known that if \(A, B \in \mathbb{R}^{r \times r}\) both non-singular, then \(\det(AB) = \det(A)\det(B)\). Therefore

\[
\det(A + xx^\top) = \det(A)\det(I + A^{-1}xx^\top).
\]

(A.2)
Theorem 1.

\[
\det(I + A^{-1}xx^\top) = 1 + x^\top A^{-1}x.
\]  

(A.3)

Proof. Let us collect the eigenvectors of \(I + A^{-1}xx^\top\). There are \(r - 1\) orthogonal vectors to \(x, x^\bot_i\}_{i=1}^{r-1}\). Thus \(I + A^{-1}xx^\top x^\bot_i = x^\bot_i\). Their corresponding eigenvalues are all one. The final eigenvector is \(A^{-1}x\). To verify this:

\[
(I + A^{-1}xx^\top)(A^{-1}x) = (1 + x^\top A^{-1}x)(A^{-1}x) \quad \Box
\]  

(A.4)

Finally, from (A.2) and (A.3) we have that:

\[
\det(A + xx^\top) = \det(A)(1 + x^\top A^{-1}x).
\]

References


