Optimal Sampling Time Selection for Parameter Estimation in Dynamic Pathway Modelling

Zoltan Kutalik[†], Kwang-Hyun Cho[‡], Steve V. Gordon[§] and Olaf Wolkenhauer[¶*]

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 modelling 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. We are first going to investigate numerical aspects of the multiple-shooting approach to parameter estimation. In particular a discussion on forward and backward integration is provided. The numerical analysis of the algorithm suggests an approach for optimal sampling time selection. With few sampling points feasible, their selection is of practical importance in experimental design.

Keywords: systems biology, dynamic pathway modelling, experimental design, multiple shooting, nonlinear ordinary differential equations, parameter estimation

1 Introduction

Systems Biology investigates how genes, proteins, metabolites and cells interact and react to changes in their environment (Wolkenhauer, Kitano and Cho 2003, Wolkenhauer, Kolch and Cho 2003). Modelling the dynamics of signal transduction pathways has received increasing attention in the last three 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

[†] Control Systems Centre, Department of Electrical Engineering and Electronics, UMIST, Manchester, UK.

[‡]School of Electrical Engineering, University of Ulsan, Ulsan, 680-749, South Korea

[§]Veterinary Laboratories Agency, Woodham Lane, New Haw, Addlestone, Surrey, KT 15 3NB, UK.

[¶] Department of Biomolecular Sciences and Department of Electrical Engineering and Electronics, UMIST, Manchester, UK.

[*] *Author for correspondence.* Address: Control Systems Centre, P.O. Box 88, Manchester M60 1QD, UK. E-mail: wolkenhauer@umist.ac.uk, Tel./Fax: +44-(0)161-200-4672.

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 modelling and parameter estimation have been successful in studying signal transduction pathways. Like the models of the TNFα mediated NF-κB pathway described in (Cho, Shin, Lee and Wolkenhauer 2003, Cho, Shin, Kim, Wolkenhauer, McFerran and Kolch 2003), 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 (Regeve 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 1983, Bock 1981) 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 numerical properties of the multiple shooting approach by providing a comparison of forward and backward integration. From this investigation, an approach to chose 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.

Ordinary differential equations are the basis for modelling 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 modelling is that data for only few time points are available. Replicate experiments (which could help us to establish a noise model) are 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 un-

known (Müller 2002, Bock 1983, Bock 1981). Here we are to investigate the numerical performance of the MSM by comparing forward and backward numerical integration. Moreover, we develop a concept of an improved experimental design to enhance parameter estimation. Baltes et al. (1994) 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. 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 modelling of signal transduction pathways. This is followed by a theoretical comparison of forward and backward integration in multiple shooting methods in Section 3. Section 4 introduces the proposed approach to experimental design. The summary of the results and discussion are in Section 5 and Section 6 respectively. All algorithms were implemented in Matlab and are available upon request from the authors.

2 Signal transduction pathway modeling and parameter estimation

To develop the idea of multiple shooting in general setting, we consider a dynamic system modelled 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*

$$\dot{x}(t) = f(t, x(t), \mathbf{k}_0), \quad x(T_0) = \mathbf{x}_0 \ , \quad \mathbf{p}_0 = \begin{pmatrix} \mathbf{x}_0 \\ \mathbf{k}_0 \end{pmatrix} \ , \quad t \in [T_0, T_1] \ , \tag{1}$$

where $\mathbf{k}_0 \in \mathbb{R}^m$ are the parameters of the differential equation and $\mathbf{x}_0 \in \mathbb{R}^n$ are the initial values of the system. For simplicity, we assume that the system $x(t, \mathbf{p}_0)$ is measured directly at certain time points, $t_1, \ldots, t_N \in [T_0, T_1]$. Allowing for measurement noise, we have

$$x_j^O(t_i) = x_j(t_i) + \varepsilon_{ij}, \quad i = 1, \ldots, N \quad j = 1, \ldots, n \tag{2}$$

where

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2). \tag{3}$$

Müller (2002) and Baltes et al. (1994) 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 were

*In the present paper we will use the following notation: $\dot{x}, \ddot{x}, x^{(3)}$ refer to the first, second and third time derivatives of function x, respectively. On the other hand f' is used for the derivative (matrix) of vector-valued functions of several variable. The components of the vector-valued function f are $(f_1, \ldots, f_n)^\top$. The regular partial derivatives are defined as usual $\partial_1 f = \frac{\partial}{\partial t} f(t, x)$ and $\partial_2 f = \frac{\partial}{\partial x} f(t, x)$. Vectors will be denoted by bold fonts, while vector valued functions are in normal type face.

accessible. In the present paper the noise, ε_{ij} is assumed to be normally distributed with common variance, σ^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, most importantly $\mathbf{k}_0$, 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 $|(\hat{\mathbf{p}}_\mathbf{0} - \mathbf{p}_\mathbf{0})/\mathbf{p}_\mathbf{0}|$. 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

$$\chi^2(\mathbf{p}) = \sum_{i=1}^{N}\sum_{j=1}^{n}\left(\frac{x_j^O(t_i) - x_j(t_i,\mathbf{p})}{\sigma}\right)^2. \tag{4}$$

Rather than minimizing (4) directly, we linearize function $x_j(t_i,\cdot)$ around $\mathbf{p}_0$, and minimize this simpler function. Using the Taylor series expansion (Rudin 1976) of $x_j(t_i,\mathbf{p})$ around $\mathbf{p}_0$

$$x_j(t_i,\mathbf{p}) = x_j(t_i,\mathbf{p}_0) + \nabla_p x_j(t_i,\mathbf{p}_0)^\top \mathbf{\Delta p} + o(\mathbf{\Delta p}), \tag{5}$$

where $\mathbf{\Delta p} = \mathbf{p} - \mathbf{p}_0$ and ∇ is the so-called *nabla* differential operator, i.e.,

$$\nabla_p = \left(\frac{\partial}{\partial p_1}, \ldots, \frac{\partial}{\partial p_{n+m}}\right)^\top.$$

$o(h^k)$ refers to a family of functions $w(h)$ for which

$$\lim_{h\to 0}\frac{w(h)}{h^k} = 0.$$

Substituting (5) to (4) yields

$$\chi^2(\mathbf{p}) = \frac{1}{\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n}\left(\varepsilon_{ij}^2 - 2\varepsilon_{ij}\nabla_p x_j(t_i,\mathbf{p}_0)^\top \mathbf{\Delta p} + \mathbf{\Delta p}^\top\left(\nabla_p x_j(t_i,\mathbf{p}_0)\nabla_p x_j(t_i,\mathbf{p}_0)^\top\right)\mathbf{\Delta p}\right) + o(\mathbf{\Delta p}). \tag{6}$$

The $o(\mathbf{\Delta p})$ component is assumed to be negligible and the minimization of (6) can thus be solved explicitly for $\mathbf{\Delta p}$, yielding

$$\mathbf{\Delta p} = F^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n}\frac{\varepsilon_{ij}}{\sigma^2}\nabla_p x_j(t_i,\mathbf{p}_0) \tag{7}$$

to obtain the minimum of (4). In this equation

$$F = \left[\frac{1}{\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{n} \nabla_p x_j(t_i, \mathbf{p}_0) \nabla_p x_j(t_i, \mathbf{p}_0)^\top \right] \tag{8}$$

is the so-called Fischer information matrix (Walter and Pronzato 1997, Baltes et al. 1994). From the solution one can compute the covariance matrix of the parameter vector as

$$\Sigma = E[\mathbf{\Delta p \Delta p}^\top] = F^{-1}. \tag{9}$$

To measure the accuracy of the estimate we like to summarise 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. Baltes et al. (1994) 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. Comparing these functionals (det and $\lambda_{max}/\lambda_{min}$) lies outside the scope of this paper.

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 1983, Bock 1981). However, since extended multiple shooting has been shown 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 $\chi^2(\mathbf{p})$ defined in (4). We treat the given data as $N-1$ boundary value problems. Starting off with initial estimates for the unknown parameters, $\mathbf{k}_0^0$ and for the initial estimates for the system $\{x^0(t_i)\}_{i=1}^N$, in the u-th iteration we compute the following quantities

$$y^u(t_i) = \begin{cases} y^u(t_i) = x^u(t_i) & : \quad i = 1, \\ x^u(t_{i-1}) + \int\limits_{t_{i-1}}^{t_i} f(t, x^u(t), \mathbf{k}_0^u) dt & : \quad i = 2, \ldots, N, \end{cases} \tag{10}$$

$$F_i\big(\mathbf{k}_0^u, x^u(t_1), \ldots, x^u(t_N)\big) = y^u(t_i) - x^0(t_i) \quad i = 1, \ldots, N, \tag{11}$$

$$G_i\big(\mathbf{k}_0^u, x^u(t_1), \ldots, x^u(t_N)\big) = y^u(t_i) - x^u(t_i) \quad i = 2, \ldots, N, \tag{12}$$

where $F : \mathbb{R}^{nN+m} \to \mathbb{R}^{nN}$ and $G : \mathbb{R}^{nN+m} \to \mathbb{R}^{n(N-1)}$. The integration in (10) is carried out numerically. In our implementation the Runge-Kutta method was used. A noteworthy remark on

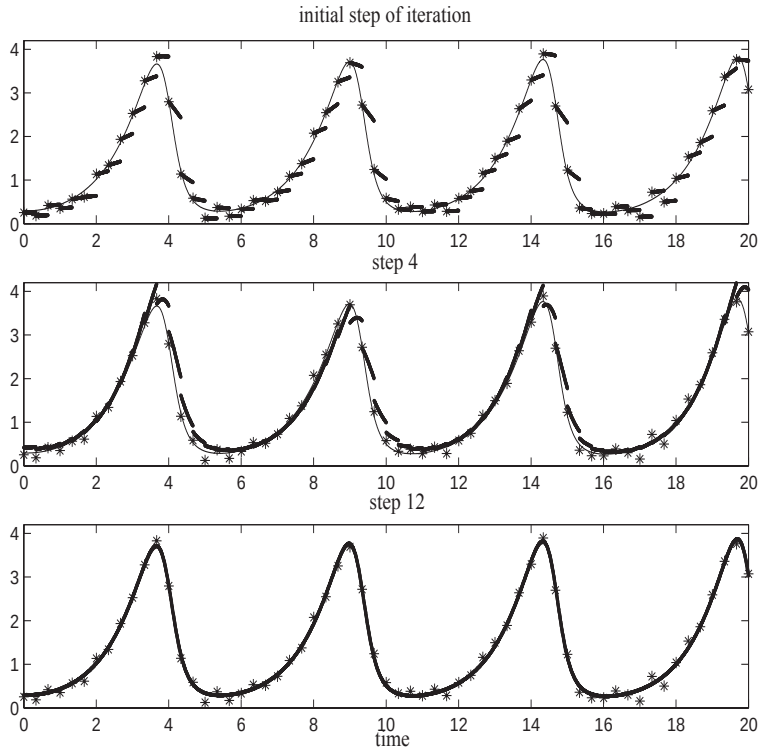

Figure 1: The convergence of multiple shooting algorithm for noisy Lotka-Volterra differential equations. The thin dark line is the real solution, dark dots represent the measurements used for parameter estimation, while thick grey line is the fitted curve in each step of the algorithm.

the numerical integration is to be mentioned in Section 3. Having computed these functions one faces the following constrained minimization problem

$$\min_{\mathbf{v}\in\mathbb{R}^{nN+m}} \|F(\mathbf{v})\|_2 \quad \text{subject to} \quad \|G(\mathbf{v})\|_2 = 0. \tag{13}$$

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). We 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

$$\{x^{u+1}(t_i)\}_{i=1}^{N} \qquad \text{and} \qquad \mathbf{k}_0^{u+1},$$

which are then used for the initialization of the $(u+1)$-th iteration step. Iterations continue until the termination criterion $\|\mathbf{G}\|_2 < \delta$ is met. The initial, the fourth, and the twelfth steps of the

algorithm are illustrated in Figure 1. One can see how both continuity and the accuracy (in a least squares sense) of the data improve as the algorithm converges.

As mentioned earlier, a very important problem for the development of parameter estimation methods is to improve the numerical recipe used. In the following section a comparative analysis will be carried out between the use of forward and backward integration as a subroutine in the MSM. This is going to provide us with a better understanding of the parameter estimation problem and is to lead us to a novel approach for sampling time selection in experimental design.

3 Numerical integration: Forward vs. backward

The inaccuracy of the parameter estimation can be attributed to at least two sources. One is the noise inherent in the experiment (measurement) and the other is the inaccuracy of the numerical integration. In this section we discuss the accuracy of parameter estimates related to numerical integration of the differential equations, which represent the system under consideration.

In what follows, we are to derive two expressions for forward and backward integration. We then compare the errors and describe cases for which either method is preferable. Suppose that the dynamic system under consideration satisfies the following equations, which are a special case of (1)

$$\dot{x}(t) = f\big(x(t)\big), \quad x(T_0) = \mathbf{x}_0, \quad t \in [T_0, T_1], \tag{14}$$

where $x : \mathbb{R} \to \mathbb{R}^n, f : \mathbb{R}^n \to \mathbb{R}^n$. Then $x(T_1)$ is to be computed by

$$x(T_1) = \mathbf{x}_0 + \int_{T_0}^{T_1} f\big(x(t)\big)\ dt. \tag{15}$$

The first order, forward, one-step numerical integration is given by

$$\begin{aligned} h &= (T_1 - T_0)/L, \\ \hat{x}(T_0 + h) &= x(T_0) + h \cdot f(x(T_0)) \\ \hat{x}(T_0 + (i+1)h) &= \hat{x}(T_0 + i \cdot h) + h \cdot f\big(\hat{x}(T_0 + i \cdot h)\big), \qquad i = 0, \ldots, L-1 \end{aligned}$$

where L is a fixed constant which describes the accuracy of the integration. The larger value is assigned to L the better the accuracy achieved. Denote the error of the approximation by

$$\mathbf{\Delta}_i = x(T_0 + i \cdot h) - \hat{x}(T_0 + i \cdot h) \tag{16}$$

and therefore compute the error by second order Taylor expansion (Rudin 1976) iteratively starting off with

$$x(T_0+h) = x(T_0) + h\dot{x}(T_0) + \frac{h^2}{2}\ddot{x}(\xi^0), \quad \text{where} \qquad \ddot{x}(\xi^0) = \big(\ddot{x}_1(\xi_1^0), \ldots, \ddot{x}_n(\xi_n^0)\big)^\top$$
$$(\xi_1^0, \ldots, \xi_n^0)^\top \in [T_0, T_0+h]^n$$

thus

$$\boldsymbol{\Delta}_0 = \frac{h^2}{2}\ddot{x}(\xi^0)$$

and in step i

$$x\big(T_0+(i+1)h\big) \;=\; [\hat{x}(T_0+i\cdot h) + \boldsymbol{\Delta}_i] + [h\cdot f(\hat{x}(T_0+i\cdot h)) + h\cdot f'(\zeta^i)\boldsymbol{\Delta}_i] + \frac{h^2}{2}\ddot{x}(\xi^i).$$

Here

$$\begin{aligned}
f'(\zeta^i) &= \big(\nabla f_1(\zeta_1^i), \ldots, \nabla f_n(\zeta_n^i)\big)^\top, \\
\ddot{x}(\xi^i) &= (\ddot{x}_1(\xi_1^i), \ldots, \ddot{x}_n(\xi_n^i))^\top, \\
\xi_j^i &\in [T_0 + i\cdot h, T_0 + (i+1)h], \qquad i = 0, \ldots, L-1 \qquad j = 1, \ldots, n \\
\zeta_j^i &\in \left[\min\big(\hat{x}_j(T_0+i\cdot h), x_j(T_0+i\cdot h)\big), \max(\hat{x}_j(T_0+i\cdot h), x_j(T_0+i\cdot h))\big)\right], \\
&\qquad\qquad i = 1, \ldots, L-1 \qquad j = 1, \ldots, n.
\end{aligned}$$

Therefore the recursion for $\boldsymbol{\Delta}_i$ is

$$\begin{aligned}
\boldsymbol{\Delta}_0 &= \frac{h^2}{2}\ddot{x}(\xi^0), \\
\boldsymbol{\Delta}_{i+1} &= \boldsymbol{\Delta}_i + \frac{h^2}{2}\ddot{x}(\xi^{i+1}) + hf'(\zeta^{i+1})\boldsymbol{\Delta}_i \qquad i = 0, \ldots, L-2,
\end{aligned}$$

leading to

$$\boldsymbol{\Delta}_{L-1}^{\mathrm{FWD}} \;=\; x(T_1) - \hat{x}(T_1) = \frac{h^2}{2}\sum_{i=0}^{L-1}\left[\prod_{j=1}^{i}\big(I + h\cdot f'(\zeta^{L-j})\big)\right]\ddot{x}(\xi^{L-1-i}) \tag{17}$$

here $I \in \mathbb{R}^{n\times n}$ is the identity matrix. Next, we derive a similar expression for backward integration before we compare the two cases. Let

$$\dot{x}(t) = f(x(t)), \quad x(T_1) = \mathbf{x_1}, \quad t \in [T_0, T_1], \tag{18}$$

and $x(T_0)$ is to be determined. The iterative numerical solution is computed as follows

$$\begin{aligned}
h &= (T_1 - T_0)/L, \\
\hat{x}(T_1 - h) &= x(T_1) - h\cdot f(x(T_1)) \\
\hat{x}(T_1 - (i+1)h) &= \hat{x}(T_1 - i\cdot h) - h\cdot f(\hat{x}(T_1 - i\cdot h)), \qquad i = 0, \ldots, L-1.
\end{aligned}$$

Similarly the same lines one can obtain the error formula

$$\boldsymbol{\Delta}_{L-1}^{\mathrm{BWD}} = x(T_0) - \hat{x}(T_0) = \frac{h^2}{2} \sum_{i=0}^{L-1} \left[\prod_{j=1}^{i} \left(I - h \cdot f'(\vartheta^{L-j}) \right) \right] \ddot{x}(\eta^{L-1-i}) \tag{19}$$

where $f'(\vartheta^i)$, $\ddot{x}(\eta^i)$, ϑ^i and η^i are defined as

$$\begin{aligned}
f'(\vartheta^i) &= (\nabla f_1(\vartheta_1^i), \ldots, \nabla f_n(\vartheta_n^i))^\top \\
\ddot{x}(\eta^i) &= (\ddot{x}_1(\eta_1^i), \ldots, \ddot{x}_n(\eta_n^i))^\top \\
\eta_j^i &\in [T_1 - i \cdot h, T_1 - (i+1)h], \qquad i = 0, \ldots, L-1 \qquad j = 1, \ldots, n \\
\vartheta_j^i &\in \left[\min\left(\hat{x}_j(T_1 - i \cdot h), x_j(T_1 - i \cdot h) \right), \max(\hat{x}_j(T_1 - i \cdot h), x_j(T_1 - i \cdot h)) \right], \\
&\qquad\qquad i = 1, \ldots, L-1 \qquad j = 1, \ldots, n.
\end{aligned}$$

Now pick the interval $[T_0, T_1]$ such that none of the following functions changes its sign: $\left\{ \ddot{x}_j \right\}_{j=1}^n$, $\left\{ \frac{\partial f_j}{\partial_i} \right\}_{i,j=1}^n$. From formulae (17), (19) it is obvious that in case h is small enough the components of $\boldsymbol{\Delta}_i$ will not change its sign, i.e. the sign depends only on the convexity of functions $x_1, \ldots, x_n$. To compute the difference between the forward and the backward formulae, i.e. (17) and (19), we split the difference into two major parts

$$\begin{aligned}
\boldsymbol{\Delta}^{\mathrm{FWD}} - \boldsymbol{\Delta}^{\mathrm{BWD}} &= \frac{h^2}{2} \sum_{i=0}^{L-1} \left[\prod_{j=1}^{i} \left(I + h \cdot f'(\zeta^{L-j}) \right) \right] \ddot{x}(\xi^{L-1-i}) - \\
&- \frac{h^2}{2} \sum_{i=0}^{L-1} \left[\prod_{j=1}^{i} \left(I - h \cdot f'(\vartheta^{L-j}) \right) \right] \ddot{x}(\eta^{L-1-i}) \\
&= \frac{h^2}{2} \sum_{i=0}^{L-1} \left[\prod_{j=1}^{i} \left(I + h \cdot f'(\zeta^{L-j}) \right) \right] \left(\ddot{x}(\xi^{L-1-i}) - \ddot{x}(\eta^i) \right) + \\
&+ \frac{h^2}{2} \sum_{i=0}^{L-1} \left\{ \left[\prod_{j=1}^{i} \left(I + h \cdot f'(\zeta^{L-j}) \right) \right] - \left[\prod_{j=1}^{i} \left(I - h \cdot f'(\vartheta^j) \right) \right] \right\} \ddot{x}(\eta^i).
\end{aligned} \tag{20}$$

In Appendix A we will show that the first term in (20) is $o(h)$ in magnitude, thus

$$\begin{aligned}
\boldsymbol{\Delta}^{\mathrm{FWD}} - \boldsymbol{\Delta}^{\mathrm{BWD}} &= \frac{h^2}{2} \sum_{i=0}^{L-1} \left\{ \left[\prod_{j=1}^{i} \left(I + h \cdot f'(\zeta^{L-j}) \right) \right] - \left[\prod_{j=1}^{i} \left(I - h \cdot f'(\vartheta^j) \right) \right] \right\} \ddot{x}(\eta^i) + o(h) \\
&= h \frac{1}{2} \sum_{i=0}^{L-1} h \left\{ \left[\exp\left(h \cdot \sum_{j=1}^{i} f'(\vartheta^j) \right) - \exp\left(-h \cdot \sum_{j=1}^{i} f'(\zeta^{L-j}) \right) \right] \ddot{x}(\eta^i) \right\} + o(h)
\end{aligned}$$

where $\exp(A)$ for matrix $A \in \mathbb{R}^{n \times n}$ is defined as

$$\exp(A) = \sum_{i=0}^{\infty} \frac{A^i}{i!}.$$

Denote therefore the constant* vector $\mathbf{c} \in \mathbb{R}^n$

$$\mathbf{c} = \frac{1}{2}\sum_{i=0}^{L-1} h \left\{ \left[\exp\left(h \cdot \sum_{j=1}^{i} f'(\vartheta^j) \right) - \exp\left(-h \cdot \sum_{j=1}^{i} f'(\zeta^{L-j}) \right) \right] \ddot{x}(\eta^i) \right\}, \tag{21}$$

leading to the difference between the two integration methods described by

$$\boldsymbol{\Delta}^{\mathrm{FWD}} - \boldsymbol{\Delta}^{\mathrm{BWD}} = h\mathbf{c} + o(h) \ . \tag{22}$$

Monotonicity	**Convexity**	**Direction**
increasing	convex	BWD
decreasing	convex	FWD
increasing	concave	FWD
decreasing	concave	BWD

Table 1: This table summarizes which direction of numerical integration is more accurate depending on the properties of function x in case of one dimension (i.e. $n = 1$). FWD refers to forward, while BWD to backward integration.

Therefore when h is chosen small enough from (22) we find that the sign of the components of $\mathbf{c}$ will decide which direction of the numerical integration is more accurate in the corresponding component of function x. Namely, when the $\mathbf{c}_l$ is positive the backward numerical integration will boast with more accuracy at the component x_l, while in the case $\mathbf{c}_l$ is negative the forward numerical integration will lead to better results at the component x_l. However, in general the determination of the sign of $\mathbf{c}$ is computationally expensive but in one dimension, i.e. when $n = 1$ (21) simplifies to

$$sgn(c) = sgn(f') \cdot sgn(\ddot{x}) \tag{23}$$

where sgn is the signum function, i.e.

$$sgn(q) = \begin{cases} 1 & : \quad q > 0, \\ 0 & : \quad q = 0, \\ -1 & : \quad q < 0. \end{cases}$$

The consequences of (23) are summarized in Table 1. This shows for example that for monotone increasing, convex solutions ($\dot{x} > 0, \ddot{x} > 0$), the backward integration improves the accuracy compared to the commonly used forward integration. More concisely, as $\ddot{x} = f'(x)\dot{x}$, when f' is positive it is suggested to use backward and when negative, forward integration. Furthermore, the integration accuracy for both directions of integration is $o(h)$, whilst the difference between the two types of integration is $o(h)$ as well. To visualize the difference we refer to Figure 2. For the

*$\mathbf{c}$ is constant in the sense that $\lim_{h\to 0} \mathbf{c}(h) = \mathbf{c}_0 \in \mathbb{R}$, i.e. $\mathbf{c} = \mathbf{c}_0 + o(1)$. $o(1)$ is defined as $o(h^0)$. This guarantees that the sign of $\mathbf{c}$ will be independent of h in case h is small enough.

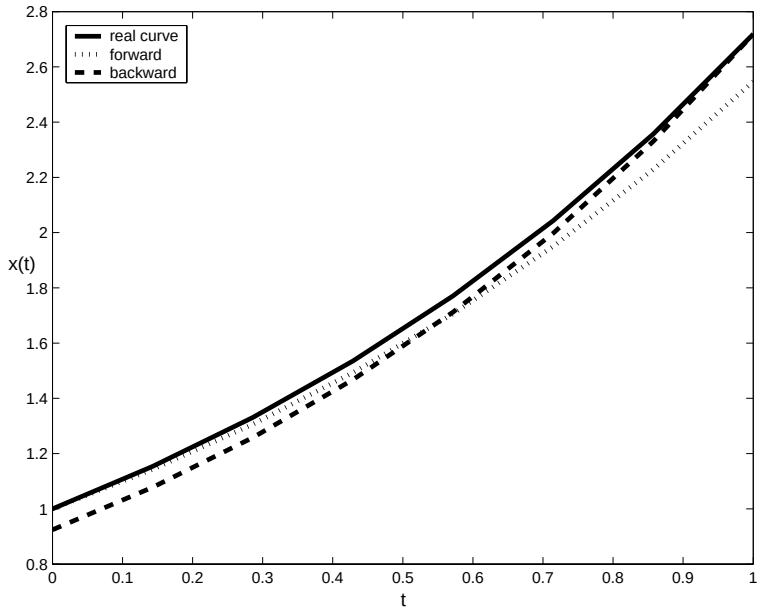

Figure 2: The difference between forward and backward integration.

ODE $\dot{x}(t) = x(t)$, $T_0 = 0$, $T_1 = 1$, with solution $x(t) = e^t$, we carried out forward and backward integration. In case of MS the difference between the two methods can lead to significantly different solutions for the parameters, as will be shown in Section 5.

In this section we considered the inaccuracies of parameter estimates as a result of numerical optimization. The aim was to gain insight into the numerical algorithms and MSM. In parameter estimation a more important 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 an optimal way.

4 Optimal experimental design

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 entirely 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.

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 an 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 σ. The information transferred from the experiment will be expressed by $\det(F)$, where matrix F was introduced in (8). The larger the determinant of F, the more information can be extracted about the parameters. Let us denote by $z(\cdot)$ the determinant of F as a function of the sampling time points:

$$z(t_1, \ldots, t_N) = \det\big(F(t_1, \ldots, t_N)\big).$$

We then have the following maximization problem to solve

$$\max\big(z(t_1, \ldots, t_N)\big) \qquad \text{subject to} \quad \{t_i\}_{i=1}^N \in [T_0, T_1]^N \subseteq \mathbb{R}^N. \tag{24}$$

Without the constraints in (24) one would use differential calculus to find the maximum. Let $m \in \{1, \ldots, N\}$ and compute the partial derivatives of function z

$$\begin{aligned}
\frac{\partial}{\partial t_m} z(t_1, \ldots, t_N) &= \frac{\partial}{\partial t_m} \det\left[\sum_{i=1}^{N}\sum_{j=1}^{n} \nabla_p x_j(t_i, \mathbf{p}_0) \nabla_p x_j(t_i, \mathbf{p}_0)^\top\right] \\
&= \frac{\partial}{\partial t_m} \det\left[\sum_{i\neq m}\sum_{j=1}^{n} \nabla_p x_j(t_i, \mathbf{p}_0) \nabla_p x_j(t_i, \mathbf{p}_0)^\top + \sum_{j=1}^{n} \nabla_p x_j(t_m, \mathbf{p}_0) \nabla_p x_j(t_m, \mathbf{p}_0)^\top\right].
\end{aligned}$$

Let us denote

$$A = \sum_{i\neq m}\sum_{j=1}^{n} \nabla_p x_j(t_i, \mathbf{p}_0) \nabla_p x_j(t_i, \mathbf{p}_0)^\top.$$

Therefore

$$\frac{\partial}{\partial t_m} z(\mathbf{t}) = 2 \cdot \det(A) \sum_{j=1}^{n} \nabla_p f_j(t_m, x(t_m, \mathbf{p}_0))^\top A^{-1} \nabla_p x_j(t_m, \mathbf{p}_0) \tag{25}$$

where f_j is the j-th component of function f, introduced in (1). To obtain this we used the following property of the determinant of sums of special matrices

$$\det(A + \mathbf{x}\mathbf{x}^\top) = \det(A)(1 + \mathbf{x}^\top A^{-1}\mathbf{x}),$$

which is proven in Appendix B. Let $H(t_1,\ldots,t_N)$ denote the Hessian of z, i.e., $H_{(l,m)}(t_1,\ldots,t_N) = \frac{\partial^2}{\partial t_l t_m} z(t_1,\ldots,t_N)$. By differentiation of (25) the diagonal of the Hessian is

$$\begin{aligned}\frac{\partial^2}{\partial t_m t_m} z(\mathbf{t}) &= 2 \cdot \det(A) \sum_{j=1}^{n} \nabla_p \frac{\partial}{\partial t_m} f_j(t_m, x(t_m,\mathbf{p}_0))^\top A^{-1} \nabla_p x_j(t_m,\mathbf{p}_0) + \\ &+ 2 \cdot \det(A) \sum_{j=1}^{n} \nabla_p f_j(t_m, x(t_m,\mathbf{p}_0))^\top A^{-1} \nabla_p f_j(t_m, x(t_m,\mathbf{p}_0)), \end{aligned} \tag{26}$$

where

$$\frac{\partial}{\partial t_m} f_j(t_m, x(t_m,\mathbf{p}_0)) = \partial_1 f_j(t_m, x(t_m,\mathbf{p}_0)) + \partial_2 f_j(t_m, x(t_m,\mathbf{p}_0)) \cdot f(t_m, x(t_m,\mathbf{p}_0))$$

and the non-diagonal elements are ($l \neq m$)

$$\begin{aligned}&\frac{\partial^2}{\partial t_l t_m} z(\mathbf{t}) = \\ &2 \cdot \det(B) \left[\sum_{j=1}^{n} \nabla_p f_j(t_l, x(t_l,\mathbf{p}_0))^\top B^{-1} \nabla_p x_j(t_l,\mathbf{p}_0)\right] \left[\sum_{j=1}^{n} \nabla_p f_j(t_m, x(t_m,\mathbf{p}_0))^\top A^{-1} \nabla_p x_j(t_m,\mathbf{p}_0)\right] \\ &+ 2 \cdot \det(A) \sum_{j=1}^{n} \nabla_p f_j(t_m, x(t_m,\mathbf{p}_0))^\top \left[\frac{\partial}{\partial t_l} A^{-1}\right] \nabla_p x_j(t_m,\mathbf{p}_0) \end{aligned} \tag{27}$$

where

$$\frac{\partial}{\partial t_l} A^{-1} = A^{-1} \left[\sum_{j=1}^{n} \nabla_p f_j(t_l, x(t_l,\mathbf{p}_0)) \nabla_p x_j(t_l,\mathbf{p}_0)^\top + \nabla_p x_j(t_l,\mathbf{p}_0) \nabla_p f_j(t_l, x(t_l,\mathbf{p}_0))^\top\right] A^{-1}$$

and

$$B = \sum_{i \neq m,l} \sum_{j=1}^{n} \nabla_p x_j(t_i,\mathbf{p}_0) \nabla_p x_j(t_i,\mathbf{p}_0)^\top.$$

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 $\mathbf{t^0}$ and the next point is obtained by

$$\mathbf{t}^{s+1} = \mathbf{t}^s - \lambda_s \mathbf{v}^s \tag{28}$$

where

$$\mathbf{v}^s = \mathbf{H}(\mathbf{t}^s)^{-1} \nabla_t z(\mathbf{t}^s) \tag{29}$$

and λ_s is defined as suggested in (Stoer and Bulirsch 2002, 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 this method yields only a suboptimum due to the reduction in the dimensionality of the

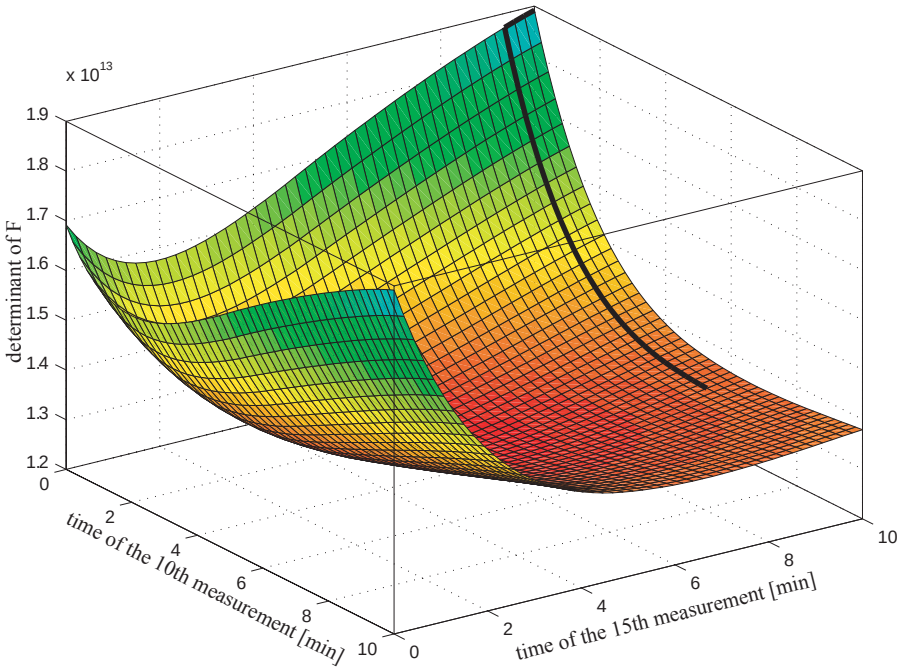

Figure 3: A three-dimensional section of the (17-dimensional) surface to be maximized. The thick black line shows how Powell's algorithm approaches the maximum point.

search space, this suboptimum commonly coincides with the optimum. The reason for this is that the optimum point usually occurs on the boundary of the multidimensional rectangle $[T_0, T_1]^N$. Figure 3 illustrates this in three dimensions. Originally all 16 time points were picked equidistantly. Looking at the figure, we start with the 15-th measurement fixed at 9.39min and begin to decrease the location of the 10th time point, beginning at 6 minutes. 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 5.

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. 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.

5 Simulation studies for a single step signal transduction module

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,

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P.$$

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.

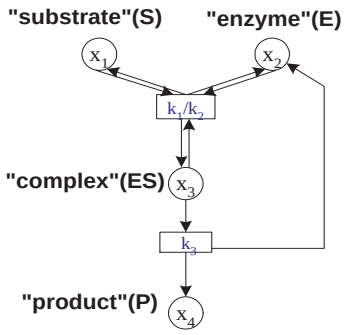

Figure 4: A template pathway modeling block for single-step signal transduction pathway. The pathway model can be constructed from basic reaction modules like this enzyme kinetic reaction for which a set of four ODEs is required.

Based on these reaction kinetics (Robert and Tom 2001, Cho, Shin, Lee and Wolkenhauer 2003, Cho, Shin, Kim, Wolkenhauer, McFerran and Kolch 2003), illustrated in Figure 4, we obtain the following set of ODEs

$$\dot{x}_1 = -k_1 x_1 x_2 + k_2 x_3 \tag{30}$$

$$\dot{x}_2 = -k_1 x_1 x_2 + (k_2 + k_3) x_3 \tag{31}$$

$$\dot{x}_3 = k_1 x_1 x_2 - (k_2 + k_3) x_3 \tag{32}$$

$$\dot{x}_4 = k_3 x_3. \tag{33}$$

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.

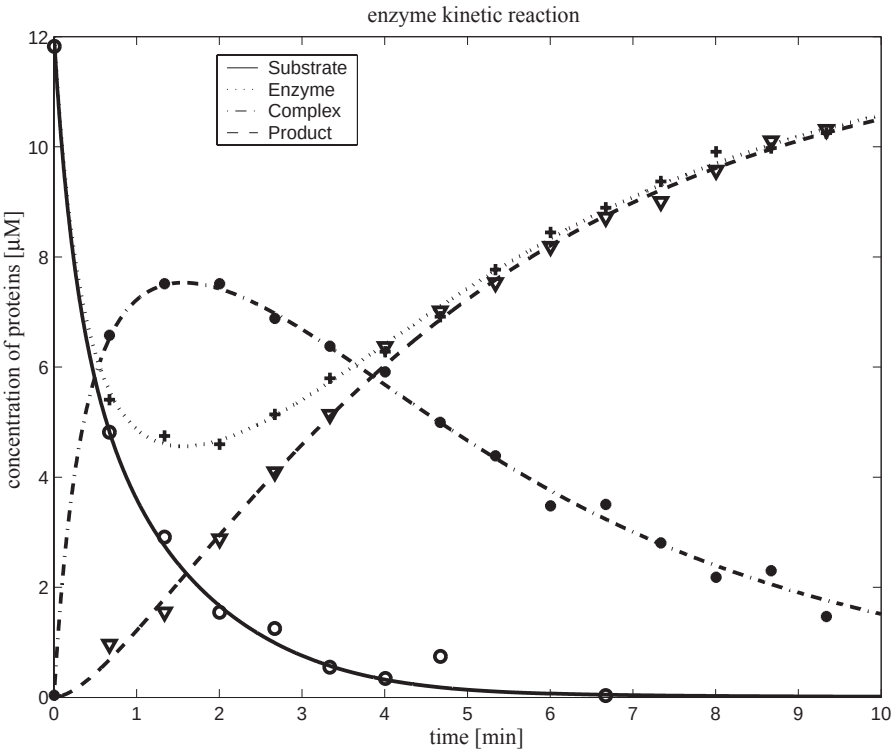

Figure 5: Reconstructed concentration curves for the enzyme kinetic reaction. Stars, triangles, dots and plus signs refer to 'measured' data.

First we 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 $\hat{k}_1 = 0.1768$, $\hat{k}_2 = 0.0155$, $\hat{k}_3 = 0.2322$. The reconstruction of the concentration profiles for each protein can be seen in Figure 5. The covariance matrix was computed based on the estimated parameter values. As described in Section 4, the determinant of the covariance matrix was optimized by Powell's quadratically convergent minimization method to yield the optimal set of time points to minimize the variance of the discrepancy between the real and the estimated parameter values of the system. The selected optimal sampling time points for experimental design are summarized in Table 2 and its efficiency is shown in Figure 6.

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

time points (min)	0	0.36	0.43	3.38	3.86	4.65
number of replicates	5	2	3	2	3	1

Table 2: Optimal experimental design for sampling time points.

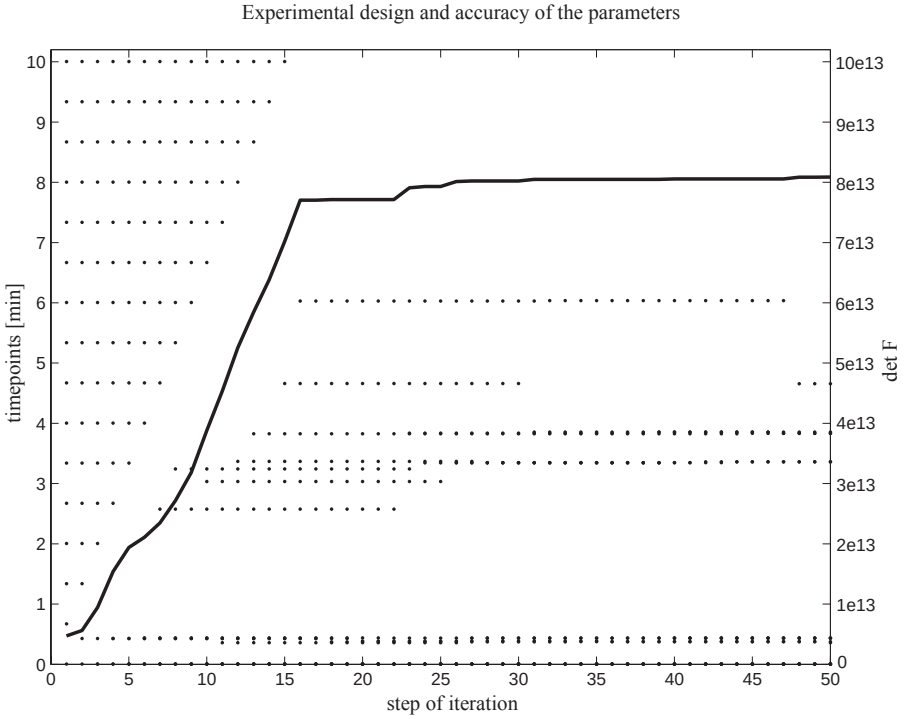

Figure 6: Figure illustrating how the accuracy increases as the design of time point collection changes during the minimization process. Each column of points refers to a set of time points.

last iteration step they converged to the optimal detailed in Table 2. 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 2). 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 3.

To demonstrate how significant the difference can be between forward and backward numerical integration we implemented the backward multiple shooting method as well. The backward

		k_1	k_2	k_3
real values		0.1800	0.0200	0.2300
forward estimate	equidistant	0.1768	0.0155	0.2322
	optimal	0.1779	0.0231	0.2296
relative error	equidistant	−0.0177	−0.2250	0.0095
	optimal	−0.0117	0.155	−0.0017

Table 3: Comparison of the parameter estimates according to the sampling time points.

algorithm agrees with the usual MSM except (10) which has to be changed in the following way:

$$y^u(t_i) = \begin{cases} y^u(t_i) = x^u(t_i) & : \quad i = N, \\ x^u(t_{i+1}) - \int\limits_{t_i}^{t_{i+1}} f(t, x^u(t), \mathbf{k_0^u}) dt & : \quad i = 1, \ldots, N-1. \end{cases} \tag{34}$$

This leads us back to the problem analyzed in Section 3. We solved it using the aforementioned backward integration formula (19). Table 4 compares the results of backward with the usual forward MSM.

		k_1	k_2	k_3
real values		0.1800	0.0200	0.2300
estimate	forward	0.1768	0.0155	0.2322
	backward	0.1782	0.0174	0.2324
relative error	forward	−0.0177	−0.2250	0.0095
	backward	−0.0100	−0.1300	0.0104

Table 4: Comparison of the parameter estimates according to the forward and backward MSM.

In this particular case the backward MSM outperformed the forward MSM, although this is dependent on the particular experiment. As discussed in Section 3, the performance basically depends on the shape of the curve (Table 1), but the magnitude of the noise at specific sampling time points plays naturally a crucial role in the performance.

Here we considered only one particular realization of an experiment where the optimization led to a considerably 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 Figure 6, the change of the determinant was approximately 20-fold between the original and the optimal design. This means that the change in the determinant of the covariance matrix is around 1/20. Since the matrix here is diagonally dominant, we can say that the determinant is roughly the product of the diagonal elements ($\sigma^2(\hat{k}_1)$, $\sigma^2(\hat{k}_2)$, $\sigma^2(\hat{k}_3)$, ...). The variance of the parameters have thus been diminished to approximately $\sqrt[7]{1/20} = 65\%$.

6 Conclusions

Modelling 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 modelling 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 for parameter estimation and experimental design, trying to gain insights into the numerical properties of the approach and to consider an 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 modelling. Based on these ideas we develop a concept for time point selection in measurements. Since the approach to select optimal sampling time points relies on an existing data set we cannot claim to provide a solution that will necessarily reduce the costs of experiments, but will 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*'.

References

Asthagiri, A. and Lauffenburger, D.: 2001, A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model, *Biotechnol. Prog.* **17**, 227–239.

Baltes, M., Scheider, R., Sturm, C. and Reuss, M.: 1994, Optimal experimental design for parameter estimation in unstructured growth models, *Biotechnol. Prog.* **10**, 480–488.

Bertsekas, D.: 1982, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press: London.

Bock, H.: 1981, Numerical treatment of inverse problems in chemical reaction kinetics, *in* K. Ebert, P. Deuflhard and W. Jager (eds), *Modeling of Chemical Reaction Systems*, Vol. 18, Springer: NY, pp. 102–125.

Bock, H.: 1983, Recent advances in parameter-identification for ordinary differential equations, *in* P. Deuflhard and E. Hairer (eds), *Progress in Scientific Computing*, Vol. 2, Birkhäuser: Boston, pp. 95–121.

Cho, K.-H., Shin, S.-Y., Kim, H.-W., Wolkenhauer, O., McFerran, B. and Kolch, W.: 2003, Mathematical modeling of the influence of RKIP on the ERK signaling pathway, *in* C. Priami (ed.), *Computational Methods in Systems Biology*, number 2602 in *Lecture Notes in Computer Science (LNCS)*, Springer-Verlag: Berlin, pp. 127–141.

Cho, K.-H., Shin, S.-Y., Lee, H.-W. and Wolkenhauer, O.: 2003, Investigations into the analysis and modeling of the TNF α mediated NF-κ B signaling pathway, *Genome Research* . Accepted for publication.

Downward, J.: 2001, The ins and outs of signaling, *Nature* **411**(14), 759–762.

Gill, P. and Murray, W.: 1974, *Numerical Methods for Constrained Optimization*, Academic Press: London.

Müller, T.: 2002, *Modeling complex systems with differential equations*, PhD thesis, Albert-Ludwigs Universität Freiburg. http://www.freidok.umi-freiburg.de/volltexte/556.

Müller, T., Noykova, N., Gyllenberg, M. and Timmer, J.: 2002, Parameter identification in dynamical models of anaerobic wastewater treatment, *Mathematical Biosciences* **177-178**, 147–160.

Press, W., Flannery, B., Teukolsky, S. and Vetterling, W.: 1987, *Numerical Recipes*, Cambridge University Press: Cambridge.

Regeve, A., Silverman, W. and Shapiro, E.: 2001, Representation and simulation of biochemical processes using π-calculus process algebra, *Pacific Symposium of Biocomputing PSB2001*, pp. 459–470.

Robert, D. and Tom, M.: 2001, Kinetic modeling approaches to *in vivo* imaging, *Nature Reviews* **2**, 898–907.

Rudin, W.: 1976, *Principles of Mathematical Analysis*, McGraw-Hill.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. and Müller, G.: 2002, Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors, *Nature Biotechnology* **20**, 370–375.

Stoer, J. and Bulirsch, R.: 2002, *Introduction to Numerical Analysis*, Texts in Applied Mathematics, 3rd edn, Springer-Verlag: NY.

Swameye, I., Müller, T., Timmer, J., Sandra, O. and Klingmüller, U.: 2003, Identification of nucleocytopasmic cycling as a remote sensor in cellular signaling by databased modeling, *Proc. Natl. Acad. Sci.* **100**, 1028–1033.

Timmer, J.: 1998, Modeling noisy time series: physiological tremor, *International Journal of Bifurcation and Chaos* **8**(7), 1505–1516.

Walter, E. and Pronzato, L.: 1997, *Identification of Parametric Models from Experimental Data*, Springer.

Wolkenhauer, O., Kitano, H. and Cho, K.-H.: 2003, Systems biology: Opportunities and challenges in the postgenomic era of the bio-medical sciences, *IEEE control Systems Magazine* . Accepted for publication.

Wolkenhauer, O., Kolch, W. and Cho, K.-H.: 2003, Mathematical systems biology: Genomic cybernetics, *in* R. Paton (ed.), *Computations in Cells and Tissues: Perspectives and Tools of Thought*, Springer Series on Natural Computation, Springer-Verlag, London. In press.

Appendix

A Order of magnitudes

Given the introduced notations and definitions in Section 3 we will show that in (20) the magnitude of the first term is that of $o(h)$.

Theorem 1.

$$\frac{h^2}{2}\sum_{i=0}^{L-1}\left[\prod_{j=1}^{i}\left(I+h\cdot f'(\zeta^{L-j})\right)\right]\left(\ddot{x}(\xi^{L-1-i})-\ddot{x}(\eta^i)\right)=o(h)$$

Proof. In case h is small enough there exist a constant $d\in\mathbb{R}$ such that

$$\left|\frac{h^2}{2}\sum_{i=0}^{L-1}\left[\prod_{j=1}^{i}\left(I+h\cdot f'(\zeta^{L-j})\right)\right]\left(\ddot{x}(\xi^{L-1-i})-\ddot{x}(\eta^i)\right)\right|<\left|\frac{h^2}{2}d\sum_{i=0}^{L-1}\left(\ddot{x}(\xi^{L-1-i})-\ddot{x}(\eta^i)\right)\right|$$

Using again Taylor's remainder theorem (Rudin 1976), there exists a vector κ^i such that

$$\kappa_j^i\in[\min\left(\xi_j^{L-1-i},\eta_j^i\right),\max\left(\xi_j^{L-1-i},\eta_j^i\right)]$$

for which

$$\left|\frac{h^2}{2}d\sum_{i=0}^{L-1}\left(\ddot{x}(\xi^{L-1-i})-\ddot{x}(\eta^i)\right)\right| = \left|\frac{h^2}{2}d\sum_{i=0}^{L-1}x^{(3)}(\kappa^i)\cdot h\right| = \left|\frac{h^2}{2}d\int_{T_0}^{T_1}x^{(3)}(t)dt\right| + o(h^2) = o(h)$$

holds. As usual $x^{(3)}$ denotes the third derivative of function x. □

B Determinant properties

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

$$\det(A) = \prod_{i=1}^{r}\lambda_i \tag{35}$$

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

$$\det(A + \mathbf{x}\mathbf{x}^\top) = \det(A)\ \det(I + A^{-1}\mathbf{x}\mathbf{x}^\top). \tag{36}$$

Theorem 2.

$$\det(I + A^{-1}\mathbf{x}\mathbf{x}^\top) = 1 + \mathbf{x}^\top A^{-1}\mathbf{x}. \tag{37}$$

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

$$(I + A^{-1}\mathbf{x}\mathbf{x}^\top)(A^{-1}\mathbf{x}) = (1 + \mathbf{x}^\top A^{-1}\mathbf{x})(A^{-1}\mathbf{x}). \tag{38}$$

□

Finally, from (36) and (37) we have that

$$\det\ (A + \mathbf{x}\mathbf{x}^\top) = \det\ (A)(1 + \mathbf{x}^\top A^{-1}\mathbf{x}).$$