# Nonparametric Collocation ODE Parameter Estimation: Application in Biochemical Pathway Modelling*

**Martin. Brown[1], Fei. He[1], Choujun. Zhan[2], Lam. Fat. Yeung[2]**

[1]*School of Electrical and Electronic Engineering, The University of Manchester, Manchester M60 1QD, UK (Tel: +44(0)1613064672; e-mail: martin.brown@machester.ac.uk; fei.he@postgrad.manchester.ac.uk).*
[2] *Electronic Engineering Department, City University of Hong Kong, Hong Kong SAR, China (e-mail: zchoujun2@student.cityu.edu.hk; eelyeung@cityu.edu.hk)*

**Abstract:** Parameter estimation of non-linear differential equations has long been an active and challenge research area. Conventionally methods are computationally intensive and often poorly conditioned. In the context of biochemical pathway modeling, a new method focused on this paper is the so-called "collocation" method, which is a nonparametric data smoothing based approach. The statistical property of a sort of linear smoothing spline based collocation methods is explicitly analyzed. It is concluded that this approach is computational efficient, but leads a non-zero estimation bias and it changes the independence assumption in the additive noise.

## 1. INTRODUCTION

Ordinary differential equations (ODEs) are widely used for the modelling of dynamic processes in physics, engineering, chemistry, molecular biology, etc. Of particular interest in this paper is their use for describing biochemical signalling, metabolic, or cell regulatory networks. This is based on two important assumptions: the pathway system is assumed to be a homogeneous "well-stirred" reaction system and spatial effects are not considered, secondly, the number of molecules of each reaction species is sufficiently large. Otherwise, partial differential equations (PDEs) or discrete stochastic models would be more appropriate.

The parameter estimation problem associated with "inverse modelling" such dynamic processes is an interesting and challenge research area because models are often both high-dimensional and non-linear, so the parameter estimation process is a non-convex, badly conditioned optimization problem. Traditionally, this inverse problem is solved by *initial valve problem* (IVP) approaches (Press *et al*. 1992; Cheney and Kincaid 1994) based on maximum likelihood estimation (MLE) criterion. These are based on iterative approaches using some gradient-based recursive algorithm to improve fitting. This method has several drawbacks (Bock 1982) and can especially easily fall into local minimum in the optimization process. A more reliable approach to overcome these problems is the multiple shooting method (MSM) (van Domselaar and Hemker 1975; Bock 1981 and 1982), which involves solving a set of IVP over a series of subintervals and enforcing continuity across the boundaries of intervals. More state/data information is used in MSM and therefore it is less likely to get trapped with local minimums. However, both IVP and MSM are computational intensive since a numerical approximation to a possible complex process is required for each update of parameter estimation, and the stability and reliability of both approaches are largely determined by the initial condition of parameters and state variables.

Another family of parameter estimation strategy discussed in this paper is known as *collocation* method which is a trajectory approximation/smoothing based approaches (Varah 1982; Ramsay, *et al*., 2007; Brunel, *et al*., 2007). It is less well known but has been extensively applied in recent systems biology literature (Madar, et. al, 2003; Voit and Almedia 2004; Lall and Voit 2005; Han *et al*., 2007). The main advantage of this approach is that instead of solving a complex non-convex optimization problem as in IVP/MSM, it reduces the problem into a simpler two-stage convex optimization process, thus computationally more efficient. This method is especially attractive when the states' derivatives are linear function of the parameters (but nonlinear in the state) which is common in biochemical pathway modelling problems, as the parameter estimation process can be transformed into a least squares problem. Additionally, no prior knowledge of parameters is needed.

However, the statistical properties of this sort of parameter estimation approach have not been explicitly studied in the current literature. This paper focuses on biochemical pathway models and especially the linear in the parameters ODEs systems. A smoothing spline based collocation method is taken as a representative for the statistical analysis of this sort of parameter estimation method. It is concluded that the smoothing spline based approaches have a non-zero estimation bias and it changes the independence assumption in the additive noise. This is demonstrated in the simulation study using a Michaelis-Menten signalling pathway.

## 2. ODE PARAMETER ESTIMATION

In this section, both the biochemical pathway models and the different parameter estimation methods are introduced.

* This work was supported by City University of Hong Kong, grant number: CityU/122305E

*2.1 State Space Model Representation*

For a biochemical pathway system with *n* states (species concentrations) and *m* (kinetic) parameters, let $x = [x_1 \quad x_2 \quad \cdots \quad x_n]^T$ and $\theta = [k_1 \quad k_2 \quad \cdots \quad k_m]^T$ represent the state and parameter vectors respectively. The autonomous biochemical pathway model can be represented as,

$$\dot{x}(t) = f(x(t), \theta), \quad x(t_0) = x_0 \tag{1}$$

For some pathway systems, e.g. Michaelis-Menten (Cho *et al.*, 2003a), ERK (Cho *et al.*, 2003b), NF-κB pathways (Hoffmann *et al.*, 2002), the first-order derivatives of state variables are linearly dependent on the parameters, which can be represented by the linear in the parameter ODE model:

$$\dot{x}(t) = f(x(t), \theta) = \Phi(x(t))\theta \tag{2}$$

Considering a discrete sampling set of length *T*, $\dot{x}$ would be a $nT \times 1$ column vector, $\Phi(x(t))$ will be a $nT \times m$ matrix. The solution of (2) is denoted $\hat{x}(t, \theta)$, and the output measurements are then:

$$y = x^* + N(0, \sigma^2) \tag{3}$$

where $x^*$ is the true, unknown signal (solution to the ODE) and the additive random Gaussian noise $N(0, \sigma^2)$ is assumed independent and identically distributed (i.i.d).

*2.2 ODE Parameter Estimation Review*

The conventional ODE parameter estimation methods are based on maximum likelihood estimation (MLE), which maximize the log-likelihood:

$$\hat{\theta} = \arg\max_{\theta} \sum_{l=1}^{T} \log g(y(t_l) - \hat{x}(t_l, \theta)) \tag{4}$$

where *g* denotes the probability density function (PDF) of the residual. MLE is statistically efficient, but $\hat{x}$ is unknown, and parameter space is big with unknown initial conditions, therefore it is generally a non-convex, badly conditioned optimization problem and computationally consuming.

*2.2.1 Least Squares and Initial Value Problems*

The corresponding nonlinear least squares (NLS) cost function with respect to the MLE (4) is given as,

$$J(\theta) = \| y - \hat{x}(t, \theta) \|_2^2 = \| r \|_2^2 \tag{5}$$

The least squares framework is based on the assumption that additive measurement noise is zero mean, Gaussian, and i.i.d. Generally, for a nonlinear ODE system, (5) is non-quadratic. Given system (1) together with the initial value of $x(t_0) = x_0$ the initial value problem is to find trajectories $x(t)$ and parameters $\theta$ that closely match the given measurement data at discrete times. This initial value approach (IVA) is an iterative process. An initial guess is updated again and again when some convergence criterion is met. The update step is usually based on the gradient or the Hessian computed from (5), leading to different optimization methods. The gradient of least squares loss function $J(\theta)$ with respect to the *j*th parameter can be expressed as

$$g = \frac{\partial J}{\partial \theta_j} = \frac{1}{2} \sum_i \sum_l \frac{\partial (y_i(t_l) - \hat{x}_i(t_l, \theta))^2}{\partial \theta_j} = -\sum_i \sum_l r_i(t_l) \hat{s}_{i,j}(t_l) \tag{6}$$

where $r_i(t_l)$ is the output residual for the *i*th state at time *l* and $\hat{s}_{i,j}(t_l)$ is the model's local sensitivity matrix. The Hessian matrix (curvature) of $J(\theta)$ can also be found by calculating the second order derivatives

$$H = \frac{\partial^2 J}{\partial \theta_j \partial \theta_k} = \sum_i \sum_l \hat{s}_{i,j}(t_l) \hat{s}_{i,k}(t_l) - \sum_i \sum_l r_i(t_l) \frac{\partial \hat{s}_{i,j}(t_l)}{\partial \theta_k} \tag{7}$$

Typically, the second term is neglected when the residuals are small, and the Hessian matrix *H* is approximated by:

$$H \approx \sum_i \sum_l \hat{s}_{i,j}(t_l) \hat{s}_{i,l}(t_l) = \hat{S}^T \hat{S} \tag{8}$$

Thus both gradient and Hessian can be expressed with respect to sensitivity matrix. They can be used in any gradient-based, first, second or quasi-second order NLS optimization algorithms for parameter estimation (solving initial value problem), for instance, the steepest decent method (Ljung 1999), the conjugate gradient method (Fletcher and Reeves 1964), the Levenberg-Marquardt algorithm (Marquardt 1963), and Gauss-Newton algorithm (Ljung 1999), etc.

However, due to least squares estimation aims at minimizing the estimated parameters' variance, estimation results can be easily affected by large outliers. Additionally, the gradient based NLS estimation process can be easily trapped with local minimums or when the optimization cost function is badly conditioned. Here, a simple example is given to show the bad conditioning cases. Consider a 2 parameter ODE system, which is subject to an external step signal and therefore has the corresponding solution is given by:

$$x(t) = \theta_1 (1 - e^{-\theta_2 t}) \tag{9}$$

As can be seen from Fig. 1, the basic optimization problem does not contain any local minima, and locally convex and symmetric around the optimal values, however, globally it is extremely badly condition for parameter values far from the true estimates. Thus $\theta_1$ could be estimated but $\theta_2$ would be difficult when the initial parameter values are far from the true ones.

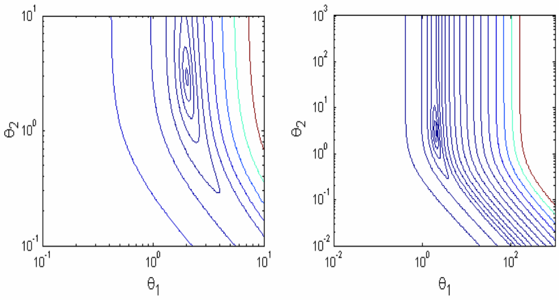

Fig. 1. Cost function contour plot of exemplar ODE system in parameter space. Locally (left) cost function is close to convex; more globally (right) it is still convex, but badly conditioned.

### 2.2.2 Robust Parameter Estimation

In order to overcome the drawbacks of NLS approaches, and especially to solve the problem caused by measurement outliers, robust criteria are often used in the parameter estimation process. Several measures of robustness are discussed in statistical literature, and most common ones are the *influence function* (Hampel 1986) and the *breakdown point* (Huber 1981; Rousseeuw and Leroy 1987), correspondingly the two main robust estimation techniques are M-estimators (Huber 1981) and least median of squares (LMS) (Rousseeuw 1984). M-estimators are generalizations of MLEs and least squares, and try to reduce the effect of outliers by replacing the squared residuals $r^2$ in (5) by another robust cost function of residuals $\rho(r)$, thus

$$\hat{\theta} = \arg\min_{\theta} \sum_i \rho\left(r_i(x_i,\theta)\right) \qquad (10)$$

where $\rho$ is usually a selected symmetric, positive-definite function with a unique minimum at zero. The minimization in (10) is achieved by finding a set $\theta$ such that

$$\sum_i \psi(r_i)\frac{\partial r_i}{\partial \theta_j} = 0\,, \quad for\ i = 1,\ldots,n,\ j = 1,\ldots m \qquad (11)$$

where the derivative $\psi(r) = d\rho(r)/dr$ is known as the *influence function*, which measures the influence of a data set on the value of the parameter estimate. For instance, the least squares with $\rho(r) = r^2/2$, the influence function is $\psi(r) = r$, which implies the influence of a data set on the estimate increases linearly with the size of its error, that accounts for why least squares estimation is non-robust. A robust estimator requires that the influence of any observation data to the parameter estimation is limited and insufficient to yield significant offset, which in term requires the influence function $\psi$ is bounded and the objective function $\rho$ have a unique minimum. Some commonly used M-estimators include $L_1$ (1-norm), $L_1 - L_2$, Huber (Huber 1981), Cauchy (Holland and Welsch 1977), Welsch (Holland and Welsch 1977), etc. LMS introduces robustness using a median function operator over the least squares cost function

$$\hat{\theta} = \arg\min_{\theta} \operatorname{median} \sum_i r_i^2(x_i,\theta) \qquad (12)$$

Since the median function is not differentiable, special search techniques (Edelsbrunner and Souvaine 1990) and efficient sampling techniques (Rousseeuw 1984) are required. LMS could handle large fractions of outliers, up to the theoretical limit of 50% the data, but has low statistical efficiency. M-estimators have higher statistical efficiency but tolerate much lower percentages of outliers unless properly initialized.

### 2.2.3 Parameter Estimation using Collocation

MLE based estimation methods are theoretical satisfactory, however, the computational cost of the optimization process would be extremely large with the rise of model nonlinearity and dimensionality, since the approximation of ODE solution and the exploration of high-dimensional parameter space. Accordingly, instead of optimizing $\theta$ directly as a parametric problem, an alternative parameter estimation strategy is based on functional estimation to derive a smooth approximation of the solution without solving the ODE, which is known as the *collocation* method (Varah 1982) and to be focused in this paper. This approach is based on a two-step procedure in which ODE state output and derivative are first estimated by nonparametric data smoothing in a model-free fashion. Then the system parameters are estimated in a static fashion by minimizing the (least squares) distance between the estimated and predicted state derivatives. Specifically, the first step is based on nonparametric estimation of the states $x$ ($\hat{x}$) and their derivatives $\dot{x}$ ($\hat{\dot{x}} = \dot{\hat{x}}$), and the parameter estimation the parameter estimation optimization problem is given by:

$$\hat{\theta} = \arg\min_{\theta} \left\|\dot{\hat{x}} - f(\hat{x},\theta)\right\|_p^p \qquad (13)$$

The state estimates, $\hat{x}$, can be obtained by polynomials or smoothing splines, neural networks, wavelets, etc. The detailed analysis of using a smoothing spline is given in Section 3. In the second step, $p$ is often selected as 1 or 2. When $p$ = 2, it is least squares estimation; when $p$ = 1, it is a $L_1$ median estimator. In this paper, we focus on the $p$ = 2 case, so the optimization in (13) becomes a nonlinear least squares regression problem.

Assuming $\hat{x}$ and $\dot{\hat{x}}$ are known or have been approximated. When considering a 2-norm SSE cost function:

$$J = \left\|\dot{\hat{x}} - f(\hat{x},\theta)\right\|_2^2 = \frac{1}{nT}\left\|\dot{\hat{x}} - \Phi(\hat{x})\theta\right\|_2^2 \qquad (14)$$

when functions $f$ are linear in the parameters, the least squares estimates of parameters can always be obtained analytically:

$$\hat{\theta} = (\Phi(\hat{x})^T\Phi(\hat{x}))^{-1}\Phi(\hat{x})^T\dot{\hat{x}} \qquad (15)$$

However, for nonlinear functions $f$, an iterative optimization procedure has to be employed to solve the corresponding static, nonlinear regression process. It is worth noting that collocation based approach transforms the dynamic, non-convex ODE estimation problem into a static and convex (assuming linear in the parameters) optimization procedure. The consistent and asymptotic property of this method was well studied by (Brunel 2007). In recent systems biology literature, a cubic spline based smoothing approach with 1-norm regularization was applied for regulatory network estimation (Han *et al.* 2007); an artificial neural network smoothing based collocation method has been applied for biochemical S-system models identification (Voit and Almedia 2004)

## 3. STATE DERIVATIVE ESTIMATION VIA SMOOTHING

In section 2.2.3, the general framework of the two-step collocation parameter estimation approach was introduced. In this section, a smoothing spline is used to estimate both the states and their derivatives and the properties of such an approach is analysed.

Consider a linear cubic polynomial smoothing spline which minimises a penalized residual sum of squares (RSS):

$$RSS(x,\lambda)=\sum_{i=1}^{n}\left(y_i-\hat{x}(t_i)\right)^2+\lambda\int_{t_1}^{t_n}\hat{x}''(t)^2\,dt \tag{16}$$

where smoothing parameter $\lambda$ is a measure of the flexibility of the slope of a fit. If $\lambda=\infty$, (16) produces a constant slope linear regression fit while $\lambda=0$ produces an interploatory fit with completely flexible slopes. The choice of $\lambda$ can usually be obtained from the generalized cross-validation (GCV) criterion (Craven and Wahba 1979) or robustified cross-validation (RCV) criterion. $\hat{x}(\cdot)$ is the smoother fitting function with continuous second derivatives, e.g. for a cubic spline with *N-1* knots, the solution is given as,

$$\hat{x}(t)=\sum_{j=1}^{N}h_j(t)\beta_j=H\hat{\beta} \tag{17}$$

where $H$ is the basis function matrix, and spline parameter $\beta$ can be estimated using penalized least squares (20), therefore the first order derivative of ODE solution can also be approximated by

$$\hat{\dot{x}}=\dot{\hat{x}}=\sum_{j=1}^{N}h_j'(t)\hat{\beta}_j=H^1\hat{\beta} \tag{18}$$

where is $H^1$ the first derivative of $H$. Thus the ODE parameters can be further estimated using standard least squares (15) or M-estimators. If robust parameter estimation is still required, a proper smoothing parameter $\lambda$ should be selected as a result of smoothing the noise. Take the cubic spline as an example, the parameter estimation property of using this smoothing based approaches are analyzed by following theorems.

**Theorem 1**

When a deterministic signal is corrupted by additive independent and identically distributed Gaussian noise $\aleph(0,\sigma^2)$, the smoothing spline parameters are described by $\hat{\beta}\sim\aleph(S(\lambda)H\beta,\sigma^2S(\lambda)S(\lambda)^T)$, with $S(\lambda)$ given in (21).

Proof:

Assuming that the measurements are

$$y=x^*+N(0,\sigma^2) \tag{19}$$

where $x^*=H\beta^*$ is the true, unknown signal (solution to the ODE). Then the optimal least squares estimator is given by:

$$\hat{\beta}=(H^TH+\lambda\Omega)^{-1}H^Ty \tag{20}$$

and the corresponding estimated model output is:

$$\begin{aligned}\hat{x}&=H\hat{\beta}=H(H^TH+\lambda\Omega)^{-1}H^Ty\\&=HS(\lambda)y\end{aligned} \tag{21}$$

where $S(\lambda)=(H^TH+\lambda\Omega)^{-1}H^T$ is often called the smoother matrix, and $\Omega=(H^2)^TH^2$, *where* $\{\Omega\}_{jk}=\int h''_j(t)\,h''_k(t)dt$. As the relationship between the estimated parameters and the noisy measurements is linear, $\hat{\beta}$ is normally distributed. The mean is given by:

$$\begin{aligned}E(\hat{\beta})&=E(S(\lambda)y)=S(\lambda)E(y)\\&=S(\lambda)H\beta\end{aligned} \tag{22}$$

Therefore the bias associated with the spline parameter estimation process is:

$$\beta-S(\lambda)H\beta \tag{23}$$

Obviously, this is zero when $\lambda=0$ or when $\hat{x}$ is a linear function, otherwise the bias associated with the parameter estimates is non-zero. The variance/covariance of the estimated parameters is:

$$\begin{aligned}\Sigma_{\hat{\beta}}&=E((E(\hat{\beta})-\hat{\beta})(E(\hat{\beta})-\hat{\beta})^T)\\&=E((S(\lambda)H\beta-\hat{\beta})(S(\lambda)H\beta-\hat{\beta})^T)\\&=E(S(\lambda)ee^TS(\lambda)^T)\\&=S(\lambda)E(ee^T)S(\lambda)^T\\&=\sigma^2S(\lambda)S(\lambda)^T\end{aligned} \tag{24}$$

where when $\lambda=0$, $e=H(\beta-\hat{\beta})$ and $\Sigma_{\hat{\beta}}=\sigma^2(H^TH)^{-1}$. Here, we assume that the errors are i.i.d.

**Theorem 2**

For any linear transformation $H'$ of the optimal smoothing spline's output (e.g. derivative or integral), the output $\hat{x}'$ is described by $\aleph(H'S(\lambda)H\beta,\sigma^2H'S(\lambda)S(\lambda)^TH'^T)$.

Proof:

As the smoother's output (or derivative or integral) is a linear transformation of the normally distributed parameter estimates as shown in (17) and (18), so, for instance, the smoother's derivative (i.e. $\dot{\hat{x}}=H'\hat{\beta}=H^1\hat{\beta}$) is also normally distributed with mean and variance given by:

$$\dot{\hat{x}}\sim N(H^1S(\lambda)H\beta,\sigma^2H^1S(\lambda)S(\lambda)^TH^{1T}) \tag{25}$$

Q.E.D

It is important to notice that the smoothing spline, its derivative and it's integral will all have, in general, a non-zero bias and also the errors are neither identically distributed nor independent. Therefore, when the systems of linear equations which are constructed via the differential or integral to algebraic transformations:

$$\int_0^t \Phi(\hat{x}(t))d\theta = \hat{x}(t) - \hat{x}(0)$$

$$\int_{t-1}^t \Phi(\hat{x}(t))d\theta = \hat{x}(t) - \hat{x}(t-1) \quad (26)$$

$$\Phi(\hat{x}(t))\theta = \dot{\hat{x}}(t)$$

are solved, the parameter estimates will be biased and the output errors can no longer be assumed to be independent. In addition, "input errors" which are due to the reconstructed smoothed states on the left hand side of the algebraic will also be biased.

**Theorem 3**

The pathway parameter estimates resulting from the least squares derivative (or integral) algebraic equations have a non-zero bias and variances:

$\hat{\theta} \sim N(ZH^1S(\lambda)H\beta, \sigma^2 ZH^1S(\lambda)S(\lambda)^T H^{1T} Z^T)$ , assuming input errors are negligible. Where $Z$ is as given in (28).

Proof:

Assuming the input errors on the left hand side of (26) are negligible, pathway parameters are related to the outputs as

$$\hat{\theta} = \hat{Z}\dot{\hat{x}} \quad (27)$$

where $\hat{Z} = (\Phi^T(\hat{x})\Phi(\hat{x}))^{-1}\Phi^T(\hat{x})$ , $\dot{\hat{x}} = H^1\hat{\beta}$ . Since for most pathway systems, the nonlinearities in the states described by $\Phi$ are simply bilinear terms between different states, and by assuming different state variables are independent, then

$$\begin{aligned} Z = E(\hat{Z}) &= (E(\Phi^T(x))E(\Phi(x)))^{-1}E(\Phi^T(x)) \\ &= (\Phi^T(E(x))\Phi^T(E(x)))^{-1}\Phi^T(E(x)) \\ &= (\Phi^T(HS(\lambda)H\beta)\Phi^T(HS(\lambda)H\beta))^{-1}\Phi^T(HS(\lambda)H\beta) \end{aligned} \quad (28)$$

Therefore, the estimated parameters are distributed as

$$\hat{\theta} \sim N(ZH^1S(\lambda)H\beta, \sigma^2 ZH^1S(\lambda)S(\lambda)^T (H^1)^T Z^T) \quad (29)$$

Q.E.D

Obviously, this may cause a significant bias and increase in the parametric uncertainty, depending on which scheme is used and the quality of the smoother, whereas Theorem 3 is applicable to all linear smoother or estimator. The bias of parameter estimation can be decreased as the number of smoothing spline basis increases, whereas this non-zero estimation bias always exists. It is expected too difficult to obtain analytic results about which approach is better and how much worse it is than performing standard least squares analysis. However, comparative results can be obtained from simulation studies. Based on Theorem 3, we can quantify the bias/covariance induced by the smoother for typical pathway models and assess the effect of compensating for the covariance by multiplying errors by the inverse covariance matrix. However, it's difficult to compensate the bias, since it is a fundamental property of this type approaches.

## 4. SIMULATION STUDY

In this section, Michaelis-Menten signalling pathway is employed for simulation study to verify the theoretical analysis results in Section 3. The model ODEs is given as,

$$\begin{aligned} \dot{x}_1 &= -k_1x_1x_2 + k_2x_3 \\ \dot{x}_2 &= -k_1x_1x_2 + (k_2+k_3)x_3 \\ \dot{x}_3 &= k_1x_1x_2 - (k_2+k_3)x_3 \\ \dot{x}_4 &= k_3x_3 \end{aligned} \quad (30)$$

The noise training data is generated from (30) with additive i.i.d Gaussian noise ( $N \sim (0, 0.25)$ ). A cubic smoothing spline ( $\hat{x} = H\hat{\beta}$ ) is then used to fit the training data, and corresponding first order derivative of prediction can also be calculated ( $\dot{\hat{x}} = H^1\hat{\beta}$ ). Here, simulation time is set to be [0:0.1:8], and the nominal true parameters are $\theta = [k_1 \;\; k_2 \;\; k_3]^T = [0.18 \;\; 0.2 \;\; 0.23]^T$ .

Firstly, we set $\lambda = 0$ , the smoothing spline parameters can be estimated from (20) and model output and derivative from (21) and (18), the model parameters can therefore be estimated using least squares estimation based on (15). This process is repeated 1000 times as a random sampling of measurement noise to get a distribution of estimated spline parameter values as well as spline model predictions. Therefore, the statistical properties of experimental results are then compared with theoretical analysis results presented in this paper, to further verify theorems 1-3. For two different cubic smoothing splines with different number of bases/knots, the simulated parameter estimation results are compared with theoretical analysis results based on (29) in Table 1. For the second state variable, the cubic spline fitting to both state output and its derivative are given in Fig. 1 and 2.

Table 1. Parameter estimation results using cubic smoothing spline based nonparametric method

| | True parameter | Cubic spline fitting with 3 knots | | | | Cubic spline fitting with 6 knots | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Experimental | | Theoretical | | Experimental | | Theoretical | |
| | | mean | variance | mean | variance | mean | variance | mean | variance |
| k1 | 0.18 | 0.1098 | 1e-3 0.0222 | 0.1019 | 1e-3 0.0313 | 0.1454 | 0.0001 | 0.1451 | 0.0002 |
| k2 | 0.2 | 0.0228 | 1e-3 0.3562 | -0.0053 | 1e-3 0.2026 | 0.1124 | 0.0011 | 0.1111 | 0.0012 |
| k3 | 0.23 | 0.2287 | 1e-3 0.0940 | 0.2288 | 1e-3 0.0753 | 0.2292 | 0.0001 | 0.2297 | 0.0001 |

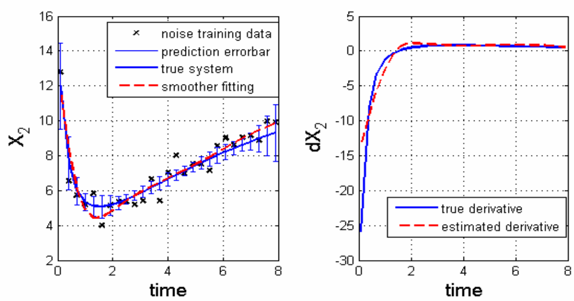


Fig. 1. Cubic spline (with 3 knots) fitting to the output and its derivative of the 2nd state. Error bars in the figures represent the standard deviations of spline fitting.

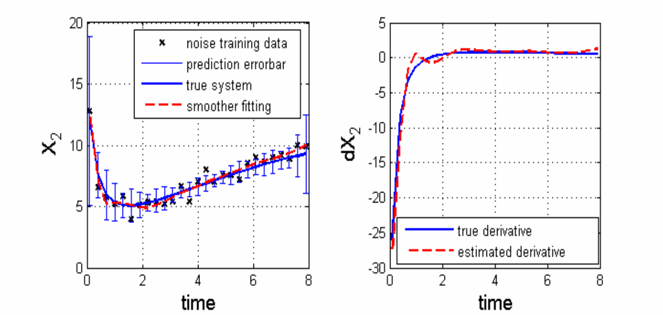


Fig. 2. Cubic spline (with 6 knots) fitting to the output and its derivative of the 2[nd] state.

It is explicit to see for both cases, parameter k3 can be well estimated, but parameters k1 and k2 are not as good as k3. However, for all three parameters, their mean values of experimental result and theoretical analysis are considerably accordant. This verifies that smoothing spline based parameter estimation results are theoretically biased. When the smoothing spline bases increase from 3 to 6, the parameter estimation results of all three parameters are improved, however, the bias with respect to parameter estimation still exists, and the variance of parameter estimation increases a bit.

## 5. CONCLUSIONS

In the context of biochemical pathway modelling, robust ODE parameter estimation problem is discussed in this paper. Particular focus is on the nonparametric collocation methods, due to its specific functional approximation property and extensive study in recent systems biology literature. The statistic property of a sort of smoothing spline based collocation approach is analyzed in detail. We conclude that this smoothing spline based approaches have a non-zero estimation bias and it changes the independence assumption in the additive noise.

## REFERENCES


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

Bock, H.G., (1982) Recent advances in parameter identification techniques for O.D.E., in *Numerical Treatment of Inverse Problems in Differential and Integral equations*, Deuflhard, P., Hairer, E., Eds., Boston: Birkhauser, chap. 7, pp. 95-112.

Brunel, N. (2007), Parameter estimation of ODE's via nonparametric estimators, *Electronic Journal of Statistics*, vol. **0**, pp. 1-24.

Cheney, W., Kincaid, D., *Numerical Mathematics & Computing*, 3rd edition, Pacific Grove, CA: Brooks/Cole, 1994.

Cho, K-H., Shin, S-Y., Lee, H-W., and Wolkenhauer, O., (2003a) Investigations in the Analysis and Modelling of the TNFalpha Mediated NF-kappaB Signaling Pathway, *Genome Research*, **13**, pp. 2413-2422.

Cho, K-H., Shin, S-Y., Kim, H-W., Wolkenhauer, O., McFerran, B., and Kolch, W., (2003b) Mathematical modeling of the influence of RKIP on the ERK signaling pathway, *Computational Methods in Systems Biology* (CMSB'03). vol. 2602 of *Lecture Notes in Computer Science*., Springer-Verlag.

Craven, P., and Wahba, G. (1979). Smoothing noisy day with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation, *Numer. Math.*, **31**, 377-403.

Edelsbrunner, H., and Souvaine, D.L., (1990) Computing least median of squares regression lines and guided topological sweep, *J. Amer. Statist. Assoc.*, **85**, 115-119.

Fletcher, R., Reeves, C.M., (1964) Function minimization by conjugate gradients, *Computer J.* **7**, 149-154.

Hampel, F.R., Rousseeuw, P.J., Ronchetti, E., and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*, John Wiley, New York.

Han, S., Yoon, Y., Cho, K-H., (2007) Inferring biomolecular interaction networks based on convex optimization, *Computational Biology and Chemistry*, **31** pp. 347-354.

Hoffmann, A., et al., (2002) The IkB-NF-kB signaling module: temporal control and selective gene activation. *Science*, **298**, pp. 1241-1245.

Holland P.W. and. Welsch, R.E., (1977) Robust regression using iteratively reweighted leastsquares, *Comm. Statist. Theory Methods*, A6, pp. 813-827.

Huber, P. J., (1981) *Robust Statistics*, John Wiley, New York.

Lall, R., Voit, E.O., (2005) Parameter estimation in modulated, unbranched reaction chains within biochemical systems, *Computational Biology and Chemistry*, **29** (5): 309-318.

Ljung, L., (1999) *System Identification: Theory for the user*, 2nd Edn, Prentice Hall PTR.

Madar, J., Abonyi, J., Roubos, H., and Szeifert, F., (2003) Incorporating prior knowledge in a cubic spline approximation – application to the identification of reaction kinetic models, *Ind. Eng. Chem. Res*, **42**: 4043-4049.

Marquardt, D.W., (1963) An algorithm for least squares estimation of nonlinear parameters. *J. SIAM*, **11**:431-441.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., *Numerical Recipes*, Cambridge University, Cambridge, 1992.

Ramsay, J.O., Hooker, G. Campbell, D. and Cao J., (2007) Parameter estimation for differential equations: a generalized smoothing approach, *J.R. Statist. Soc.* B. **69** (5): 741-796.

Rousseeuw, P.J., (1984) Least median of squares regression, *J. Amer. Statist. Assoc.*, **79**, 871-880.

Rousseeuw, P.J., and Leroy, A.M., (1987) *Robust Regression and Outlier Detection*, John Wiley, New York.

van Domselaar, B., Hemker, P., (1975) Nonlinear parameter estimation in initial value problems, Technical Report NW 18/75, Mathematical Centre Amsterdam.

Varah, J.M., (1982) A spline least squares method for numerical parameter estimation in differential equations. *SIAM J. Scient. Computn*, **3**, 28-46.

Voit, E.O., Almedia, J., (2004) Decoupling dynamical systems for pathway identification from metabolic profiles, *Bioinformatics*, **20**(11), pp. 1670-1681.