Parameter Estimation for Ordinary Differential Equations

Jonathan Calver (supervised by Wayne Enright)

Department of Computer Science
University of Toronto

CAIMS
Whistler, B.C
June 11, 2019

Outline

Parameter Estimation for ODEs

Two Stage Algorithm

Numerical Results

Discussion

Motivation

- Mathematical models describing natural phenomena often take the form of systems of ordinary differential equations (ODEs).
- These models usually contain unknown parameters, $\boldsymbol{p}$.
- The goal is to estimate the parameters, $\hat{\boldsymbol{p}}$, that best fit the observed data ($\hat{\boldsymbol{y}}(t_i)\ \boldsymbol{y}(t_i) + N(0, \sigma^2),\ i = 1, \ldots, n_o$).
- This is a non-linear least squares problem.

$$\hat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}} \sum_{i=1}^{n_o} \frac{\|\hat{\boldsymbol{y}}(t_i) - \boldsymbol{y}(t_i, \boldsymbol{p})\|^2}{2\sigma^2}$$

We will consider using the **simple (single) shooting approach** (Bard, 1974).

Definitions
ODEs

We consider the initial value problem (IVP),

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{f}(t, \boldsymbol{y}(t), \boldsymbol{p})$$
$$\boldsymbol{y}(0) = \boldsymbol{y}_0$$
$$t \in [0, T]$$

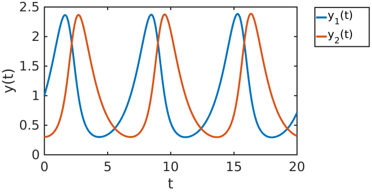
- Numerous numerical methods exist for approximating $\boldsymbol{y}(t, p)$ that satisfy this IVP, for a given value of $\boldsymbol{p}$
- Most methods simulate the **trajectory**, $\boldsymbol{y}(t, p)$, starting from $\boldsymbol{y}(0)$. (e.g. continuous Runge-Kutta with defect control Enright and Yan (2010))

Examples I

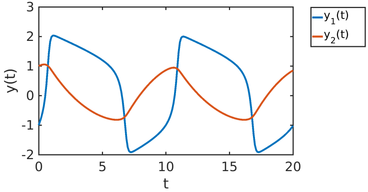
- Systems biology - enzyme kinetics, population dynamics, predator-prey models
- Epidemiology - spread of disease models such as SIR or SEIR
- Chemistry - rate equations, chemical kinetics
- Many more

Examples II

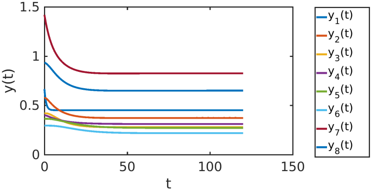
Predator-Prey model



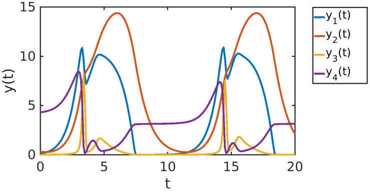
model of squid neuron spikes



a biochemical model



Calcium Ion model



Motivation

This non-linear least squares problem may be challenging if:

1. $\boldsymbol{y}(t, \boldsymbol{p})$ needs to be approximated (using a numerical method)
2. a good initial guess for $\hat{\boldsymbol{p}}$ is unavailable (only know broad ranges that parameters can take)
3. observations are only available for a subset of the components of $\boldsymbol{y}$ or the observations are **noisy** or **sparse** (potential identifiability issues)

Two Stage ODE Parameter Estimation

First Stage: find a suitable initial guess, $\boldsymbol{p}_o$

- use a heuristic global search
- solve an alternative optimization problem
- This first stage is similar to the ideas of proposal generation in computer vision or a warm start when training neural networks.

Second Stage: apply a local, fast converging optimization to obtain $\hat{\boldsymbol{p}}$

- Levenberg-Marquardt (Quasi-Newton)
- Full Newton (accelerated least squares (ACCEL) method (Dattner and Gugushvili, 2015))

The combination of a global heuristic search and a local optimizer is sometimes referred to as a **hybrid optimizer** (Rodriguez-Fernandez et al., 2006b)

Methods for obtaining a suitable p_0

- Smooth and match estimator (SME) (Gugushvili et al., 2012)
  - Numerical Discretization based estimation (DBE) is similar (Wu et al., 2012)
  - Gradient Matching (Macdonald and Husmeier, 2015) in the ML literature
  - Varah (1982); Bellman and Roth (1971) and others recognized that if one uses the observed values of $\boldsymbol{y}(t)$ to approximate $\boldsymbol{y}'(t)$, then one can formulate a related least squares problem,

$$\boldsymbol{p}_{\text{sme}} = \arg\min_{\boldsymbol{p}} \int_0^T \left\|(\tilde{\boldsymbol{y}}'(t) - \boldsymbol{f}(t, \tilde{\boldsymbol{y}}(t), \boldsymbol{p}))\right\|^2 dt.$$

Integral smooth and match (INT-SME)

- A benefit of SME is that the numerical derivative approximations can be significantly faster than a simulation of an ODE IVP trajectory.
- Furthermore, an ODE IVP trajectory simulation may fail for certain parameter values - not a problem for SME
- SME can't estimate intial conditions (could use nearby data points)
- Gugushvili et al. (2012) proposed using the integral form of the ODE IVP instead:

$$\min_{\boldsymbol{y}_0,\boldsymbol{p}} \int_0^T \left\| \tilde{y}(t) - \left(\boldsymbol{y}_0 + \int_0^t f(s, \tilde{y}(s), \boldsymbol{p})\, ds\right) \right\|^2 dt, \qquad (1)$$

where $\boldsymbol{y}_0$ are the initial conditions to be estimated. Note, this formulation is linear in $\boldsymbol{y}_0$. Unlike SME, this does not require us to approximate $\boldsymbol{y}'(t)$.

Exploiting ODE Structure - Linear Parameters

- Some parameters may appear linearly in $f(t, \boldsymbol{y}(t), \boldsymbol{p})$.
- $\boldsymbol{p} = [\boldsymbol{q}\,,\, \boldsymbol{r}]$

$$f(t, \boldsymbol{y}(t), \boldsymbol{p}) = G(t, \boldsymbol{y}(t), \boldsymbol{q})\boldsymbol{r} + g(t, \boldsymbol{y}(t), \boldsymbol{q})$$

- SME becomes:
$$\min_{\boldsymbol{q}} \int_0^T \left\| (\tilde{\boldsymbol{y}}'(t) - G(t, \tilde{\boldsymbol{y}}(t), \boldsymbol{q})\boldsymbol{r}(\boldsymbol{q}) - g(t, \tilde{\boldsymbol{y}}(t), \boldsymbol{q}) \right\|^2 \, dt.$$
- Since $\boldsymbol{r}$ appears linearly in SME, we can view this as a minimization over only the nonlinear parameters, $\boldsymbol{q}$.

Exploiting ODE Structure - Decoupling

- the system of ODEs can sometimes be decoupled
- subsets of the parameters may appear in f, such that each subset can be independently determined (Kotte and Heinemann, 2009)

Example: Goodwin Model

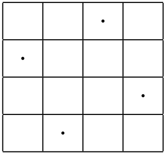
$$f(t, y(t), p) = \begin{pmatrix} \boxed{\frac{1}{q_1 + y_3(t)^{q_2}} \quad -y_1(t)} & 0 & 0 & 0 & 0 \\ 0 \quad 0 & \boxed{y_1(t) \quad -y_2(t)} & 0 & 0 \\ 0 \quad 0 & 0 \quad 0 & \boxed{y_2(t) \quad -y_3(t)} \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_6 \end{pmatrix}$$

Subproblem #	$\lvert\boldsymbol{q}\rvert$	$\lvert\boldsymbol{r}\rvert$
1	2	2
2	0	2
3	0	2
total	2	6

- 3 independent subproblems
- reduces the size of the search space

Latin Hypercube Sampling (LHS)

- For our application, in Stage 1 it is often recommended to sample from the log of the parameter space ((Raue et al., 2013; Maiwald and Timmer, 2008))
- $q \in [10^a, 10^b]^{n_q}$
- $\log q \in [a, b]^{n_q}$
- the scale is usually more important than the exact value
- we used lhsdesign in MATLAB - attempts to distribute points more appropriately when generating the LHS samples
- 2-D example:



Full Algorithm

Stage I: **foreach** *subproblem* S_k, $k = 1, \ldots, K$ **do**

- obtain samples of q^k from the log of the parameter space using LHS
- calculate the subproblem objective function for each sample and sort them
- run local optimizer from the N_{best} samples of q^k
- let P_k be the set of unique minimizers found for subproblem S_k

end

let $P_0 = P_1 \times P_2 \times ... \times P_K$ (Cartesian product)

compute the objective function for each $p_0 \in \{P_0\}$ and sort them

Stage II:

run local optimizer on p_0's from $\{P_0\}$ until satisfied that $\hat{\boldsymbol{p}}$ has been found

Algorithm 1: Two Stage Parameter Estimation Procedure

Choice of Smoother

We have considered:

- cubic splines (Varah, 1982)
- smoothed splines (e.g. splinefit in MATLAB)
- local polynomial estimation (lpe) (Dattner, 2015)

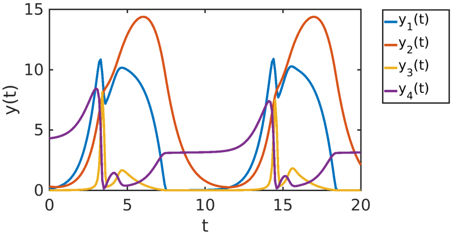
Other approaches:

- neural net or any other function approximator
- for SME, can use the Koopman operator to approximate $\boldsymbol{y}'(t)$ (Mauroy and Goncalves, 2017)

Calcium Ion Test Problem [Kummer et al. (2000) ?]

$$
\begin{aligned}
{G^*_\alpha}' &= k_1 + k_2 G^*_\alpha - k_3 PLC^* \frac{G^*_\alpha}{G^*_\alpha + Km_1} - k_4 Ca_{cyt} \frac{G^*_\alpha}{G^*_\alpha + Km_2}, \\
PLC^{*\prime} &= k_5 G^*_\alpha - k_6 \frac{PLC^*}{PLC^* + Km_3}, \\
{Ca_{cyt}}' &= k_7 PLC^* Ca_{cyt} \frac{Ca_{er}}{Ca_{er} + Km_4} + k_8 PLC^* + k_9 G^*_\alpha - k_{10} \frac{Ca_{cyt}}{Ca_{cyt} + Km_5} - k_{11} \frac{Ca_{cyt}}{Ca_{cyt} + Km_6}, \\
{Ca_{er}}' &= -k_7 PLC^* Ca_{cyt} \frac{Ca_{er}}{Ca_{er} + Km_4} + k_{11} \frac{Ca_{cyt}}{Ca_{cyt} + Km_6},
\end{aligned}
$$

- 11 linear parameters (k)
- 6 non-linear parameters (Km)
- Km_2 kept fixed for identifiability.
- 6.5% relative noise added to true trajectory sampled at 200 uniformly spaced times



Mendes Problem

- test problem that was originally posed in Moles et al. (2003)
- subsequently studied in Rodriguez-Fernandez et al. (2006b,a); Balsa-Canto et al. (2008); Gábor and Banga (2015).

$$f(t, y(t), p) = \begin{bmatrix} G_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & G_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & G_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & G_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & G_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & G_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & G_{78} \end{bmatrix} \begin{bmatrix} r_1 \\ \\ \\ \vdots \\ \\ \\ r_{15} \end{bmatrix}$$

$$G_1 = \left[\frac{1}{1+\left(\frac{P}{q_1}\right)^{q_2}+\left(\frac{q_3}{S}\right)^{q_4}} \quad -y_1 \right]$$

$$G_2 = \left[\frac{1}{1+\left(\frac{P}{q_5}\right)^{q_6}+\left(\frac{q_7}{y_7}\right)^{q_8}} \quad -y_2 \right]$$

$$G_3 = \left[\frac{1}{1+\left(\frac{P}{q_9}\right)^{q_{10}}+\left(\frac{q_{11}}{y_8}\right)^{q_{12}}} \quad -y_3 \right]$$

$$G_4 = \left[\frac{y_1}{y_1+q_{13}} \quad -y_4 \right]$$

$$G_5 = \left[\frac{y_2}{y_2+q_{14}} \quad -y_5 \right]$$

$$G_6 = \left[\frac{y_3}{y_3+q_{15}} \quad -y_6 \right]$$

$$G_{78} = \begin{bmatrix} \frac{y_4(\frac{1}{q_{16}})(S-y_7)}{1+(\frac{S}{q_{16}})+(\frac{y_7}{q_{17}})} & -\frac{y_5(\frac{1}{q_{18}})(y_7-y_8)}{1+(\frac{y_7}{q_{18}})+(\frac{y_8}{q_{19}})} & 0 \\ 0 & \frac{y_5(\frac{1}{q_{18}})(y_7-y_8)}{1+(\frac{y_7}{q_{18}})+(\frac{y_8}{q_{19}})} & -\frac{y_6(\frac{1}{q_{20}})(y_8-P)}{1+(\frac{y_8}{q_{20}})+(\frac{P}{q_{21}})} \end{bmatrix}$$

- 36 parameters to be estimated (15 linear, 21 non-linear)
- Hill coefficients $(q_2, q_4, q_6, q_8, q_{10}, q_{12})$ are assumed to be in $[0.1, 10]$.
- Other parameters are assumed to be in $[10^{-12}, 10^6]$.
- true values are all close to 1
- The initial conditions, as well as S and P, are assumed to be known.
- The data for this test problem consists of 21 uniformly spaced observations of the 8 state variables (over the interval $[0, 120]$), for each of 16 pairs of values for P and S.
- Either 3% or 5% relative noise added to simulated data.

Calcium Ion Results

initial	smoother	Average Time			success
		initial	opt	total	
SME	spline	0.62	7.57	8.19	98 / 100
SME	lpe	0.50	6.82	7.32	81 / 100
INT-SME	spline	0.93	5.52	6.45	77 / 100
INT-SME	lpe	0.79	5.49	6.28	70 / 100

SME more likely to succeed than INT-SME, but slightly slower

- SME produces more initial guesses to try than INT-SME

spline seems to be better than lpe

- smoothing can dampen the peaks in the noisy data (worse p_o)

Mendes Results

initial	smoother	noise (%)	Average Time			success
			initial	opt	total	
SME	cubic spline	3	5.0	2.6	7.5	77 / 100
SME	cubic spline	5	5.0	3.0	8.1	80 / 100
SME	lpe	3	5.1	3.4	8.6	73 / 100
SME	lpe	5	5.1	3.2	8.4	84 / 100
SME	splinefit	3	5.1	2.7	7.8	84 / 100
SME	splinefit	5	5.0	2.4	7.4	83 / 100
INT-SME	cubic spline	3	3.0	0.6	3.6	100 / 100
INT-SME	cubic spline	5	3.0	0.7	3.7	100 / 100
INT-SME	lpe	3	3.3	0.6	3.8	100 / 100
INT-SME	lpe	5	3.3	0.6	3.9	100 / 100
INT-SME	splinefit	3	3.0	0.6	3.5	100 / 100
INT-SME	splinefit	5	3.0	0.7	3.7	100 / 100

Mendes Results

- 10^4 seconds, global-local hybrid optimizer (Rodriguez-Fernandez et al., 2006b).
- $\approx$ 300 seconds, global scatter search heuristic (Rodriguez-Fernandez et al., 2006a).
- $\approx$ 30 seconds, scatter search with additional heuristics (Gábor and Banga, 2015)
- Our approach is competitive with this. ($\approx$ 4 seconds with INT-SME)

Difficulties

- numerical gradients are not likely to be accurate if default tolerances used in MATLAB
  - choose appropriate tolerances for finite differences to provide accurate gradients
  - simulate the variational equations to approximate the required gradients
  - use a different (more reliable or faster) ODE solver (e.g. DDEM Zivari-Piran and Enright (2012) / CRK56 Enright and Yan (2010))

Numerical Uncertainty

What about the fact that $\boldsymbol{y}(t, p)$ is a numerical approximation to the solution of the ODE IVP?

- add the numerical error, $\sigma_{num}(t)$, to the error model (usually try to choose TOL such that $\sigma_{num} << \sigma$)
- can use an estimate of the global error to approximate $\sigma_{num}(t)$ (Rive and Pasciutti, 1975; Constantinescu, 2015).

$$\hat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}} \sum_{i=1}^{n_o} \frac{\|\hat{y}(t_i) - y(t_i, \boldsymbol{p})\|^2}{2(\sigma_{num}^2(t_i) + \sigma^2)}$$

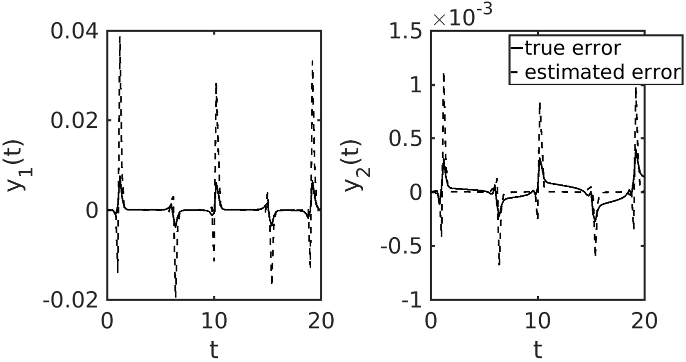
- uncertainty quantification for ODEs / probabilistic numerical methods (PNMs) (Teymur et al., 2018; Tronarp et al., 2018; Schober et al., 2014)

Example - FitzHugh-Nagumo Model

A model of squid neuron electrical potentials (FitzHugh, 1961).

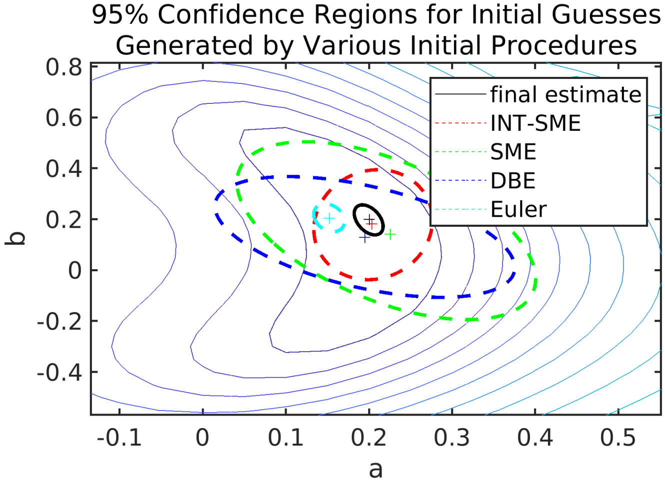
$$V'(t) = {y_1}'(t) = c\left(y_1(t) - \frac{y_1(t)^3}{3} + y_2(t)\right),$$

$$R'(t) = {y_2}'(t) = \frac{-\left(y_1(t) - a + by_2(t)\right)}{c},$$



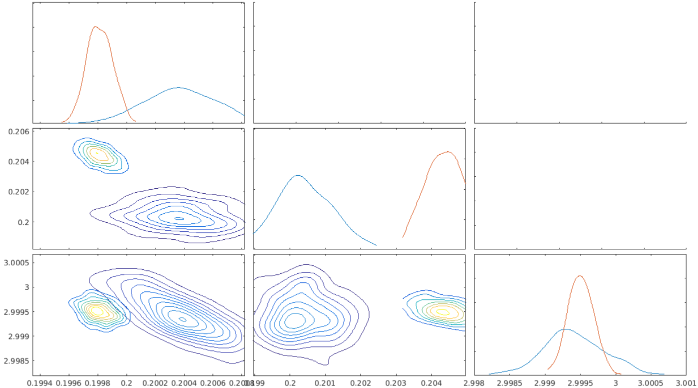
Numerical error varies with t.

Example - FitzHugh-Nagumo Model



The error introduced by using a fixed step size Euler ODE solver results in **biased** estimates.

Example - FitzHugh-Nagumo Model
Distribution of MAP estimates



Inclusion of the numerical error increases the uncertainty and may change the estimates.

Summary

- Presented a two stage procedure for the ODE IVP parameter estimation problem
  1. obtain a set of candidate $\boldsymbol{p}_o$'s
  2. use Levenberg-Marquardt to find $\hat{\boldsymbol{p}}$
- Demonstrated the performance of the procedure on two test problems from the literature
- Discussed how numerical error can impact $\hat{\boldsymbol{p}}$.

Joint work with Wayne Enright and Jienan Yao

Thanks for Listening

Questions?

E. Balsa-Canto, M. Peifer, J. Banga, J. Timmer, and C. Fleck. Hybrid optimization method with general switching strategy for parameter estimation. *BMC Systems Biology*, 2(1), 2008.

Y. Bard. *Nonlinear parameter estimation*. Academic press, 1974.

R. Bellman and R.S. Roth. The use of splines with unknown end points in the identification of systems. *Journal of Mathematical Analysis and Applications*, 34(1):26–33, 1971.

Emil Constantinescu. Estimating global errors in time stepping. *arXiv preprint arXiv:1503.05166*, 2015.

I. Dattner. A model-based initial guess for estimating parameters in systems of ordinary differential equations. *Biometrics*, 71(4): 1176, 2015.

I. Dattner and S. Gugushvili. Accelerated least squares estimation for systems of ordinary differential equations. *arXiv preprint arXiv:1503.07973*, 2015.

W.H. Enright and L Yan. The reliability/cost trade-off for a class of ode solvers. *Numerical Algorithms*, 53(2):239–260, 2010.

R. FitzHugh. Impulses and Physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445 – 466, 1961.

A. Gábor and J.R Banga. Robust and efficient parameter estimation in dynamic models of biological systems. *BMC systems biology*, 9(1):74, 2015.

S. Gugushvili, C.A.J. Klaassen, et al. $\sqrt{n}$-consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing. *Bernoulli*, 18(3): 1061–1098, 2012.

O. Kotte and M. Heinemann. A divide-and-conquer approach to analyze underdetermined biochemical models. *Bioinformatics*, 25 (4):519–525, 2009.

U. Kummer, L. F. Olsen, C. J. Dixon, A. K. Green, E. Bornber-Bauer, and G. Baier. Switching from simple to complex oscillations in calcium signaling. *Biophys. J.*, 79(3): 1188–1195, 2000.

B. Macdonald and D. Husmeier. Gradient matching methods for computational inference in mechanistic models for systems biology: a review and comparative analysis. *Frontiers in bioengineering and biotechnology*, 3:180, 2015.

T. Maiwald and J. Timmer. Dynamical modeling and multi-experiment fitting with potterswheel. *Bioinformatics*, 24 (18):2037–2043, 2008.

Alexandre Mauroy and Jorge Goncalves. Koopman-based lifting techniques for nonlinear systems identification. *arXiv preprint arXiv:1709.02003*, 2017.

C.G. Moles, P. Mendes, and J.R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome research*, 13(11):2467–2474, 2003.

A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B.D. Harms, F.J. Theis, et al. Lessons learned from quantitative dynamical modeling in systems biology. *PloS one*, 8(9):e74335, 2013.

L. Dalle Rive and F. Pasciutti. Runge-kutta methods with global error estimatesâĂă. *IMA Journal of Applied Mathematics*, 16 (3):381–386, 1975. doi: 10.1093/imamat/16.3.381. URL http://dx.doi.org/10.1093/imamat/16.3.381.

M. Rodriguez-Fernandez, J.A. Egea, and J.R. Banga. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics*, 7(1):1–18, 2006a.

M. Rodriguez-Fernandez, P. Mendes, and J.R. Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, 83(2):248–265, 2006b.

Michael Schober, David K Duvenaud, and Philipp Hennig. Probabilistic ode solvers with runge-kutta means. In *Advances in neural information processing systems*, pages 739–747, 2014.

Onur Teymur, Han Cheng Lie, Tim Sullivan, and Ben Calderhead. Implicit probabilistic integrators for odes. *arXiv preprint arXiv:1805.07970*, 2018.

Filip Tronarp, Hans Kersting, Simo Särkkä, and Philipp Hennig. Probabilistic solutions to ordinary differential equations as non-linear bayesian filtering: A new perspective. *arXiv preprint arXiv:1810.03440*, 2018.

J.M. Varah. A spline least squares method for numerical parameter estimation in differential equations. *SIAM J. of Sci. and Stat. Comput.*, 3(1):28–46, 1982.

H. Wu, H. Xue, and A. Kumar. Numerical discretization-based estimation methods for ordinary differential equation models via penalized spline smoothing with applications in biomedical research. *Biometrics*, 68(2):344–352, 2012.

H. Zivari-Piran and W.H. Enright. Accurate first-order sensitivity analysis for delay differential equations. *SIAM Journal on Scientific Computing*, 34(5):A2704–A2717, 2012.