Using *a posteriori* error analysis to develop fast and low bias gradient-based ODE inference strategies.

Richard Creswell

2022 May 23

Inference problem for ordinary differential equations (DEs).

$$\frac{dy}{dt} = f(y, t; \theta); \quad y(t = t_0) = y_0$$

- Given an observed time series for y, learn parameters θ.

Inference problem for ordinary differential equations (DEs).

$$\frac{dy}{dt} = f(y, t; \theta); \quad y(t = t_0) = y_0$$

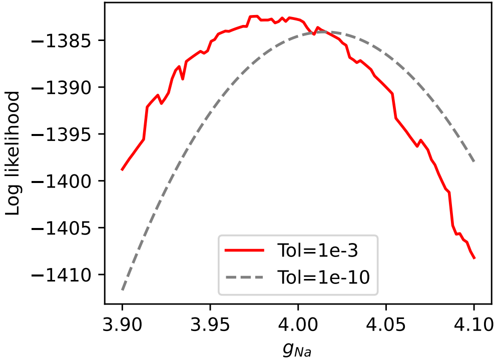
- Given an observed time series for y, learn parameters θ.
- Problems of this type arise throughout mathematical biology (e.g., epidemiology, cardiac electrophysiology, enzyme kinetics, ...).

Inference for DEs requires multiple approximate forward solves.

- Each forward solution requires a numerical approximation.
- Adaptive step size solvers are based on the **local truncation error** without directly controlling the error in the likelihood.

Inference for DEs requires multiple approximate forward solves.

- Each forward solution requires a numerical approximation.
- Adaptive step size solvers are based on the **local truncation error** without directly controlling the error in the likelihood.



Beeler and Reuter: "Reconstruction of the action potential of ventricular myocardial fibres," *Journal of Physiology* (1977).

Some other ideas from probabilistic numerics.

- *Probabilistic numerics* aims to quantify uncertainty in numerical approximations using probability distributions.

Hennig et al.: "Probabilistic numerics and uncertainty in computations," *Proceedings of the Royal Society A* (2015).

Some other ideas from probabilistic numerics.

- *Probabilistic numerics* aims to quantify uncertainty in numerical approximations using probability distributions.
- Place Gaussian process priors on the DE solution, incorporating this uncertainty in the solution in parameter inference.

Hennig et al.: "Probabilistic numerics and uncertainty in computations," *Proceedings of the Royal Society A* (2015).

Chkrebtii et al.: "Bayesian solution uncertainty quantification for differential equations," *Bayesian Analysis* 11.4 (2016)

Some other ideas from probabilistic numerics.

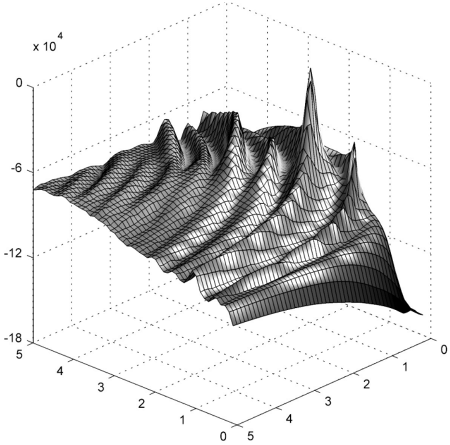
- *Probabilistic numerics* aims to quantify uncertainty in numerical approximations using probability distributions.
- Place Gaussian process priors on the DE solution, incorporating this uncertainty in the solution in parameter inference.
- Randomize the solver by inserting noise (IID Gaussian) at each solver time step, to obtain a distribution over the solutions to the DE.

Hennig et al.: "Probabilistic numerics and uncertainty in computations," *Proceedings of the Royal Society A* (2015).

Chkrebtii et al.: "Bayesian solution uncertainty quantification for differential equations," *Bayesian Analysis* 11.4 (2016)

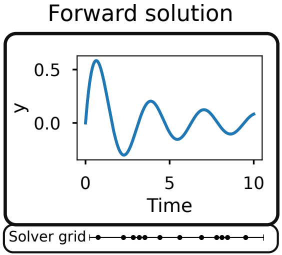
Conrad et al.: "Statistical analysis of differential equations: introducing probability measures on numerical solutions," *Statistics and Computing*, 27.4 (2017).

DE problems often present challenging or high-dimensional posteriors.

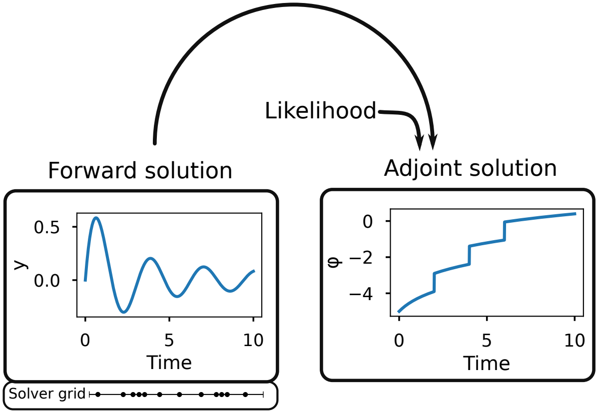


Calderhead and Girolami: "Estimating Bayes factors via thermodynamic integration and population MCMC," *Computational Statistics and Data Analysis*, 53 (12).

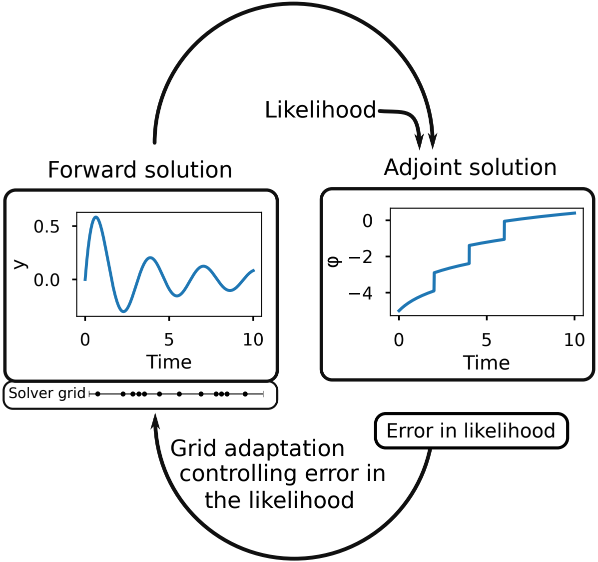
Adjoint methods for ODE inference.



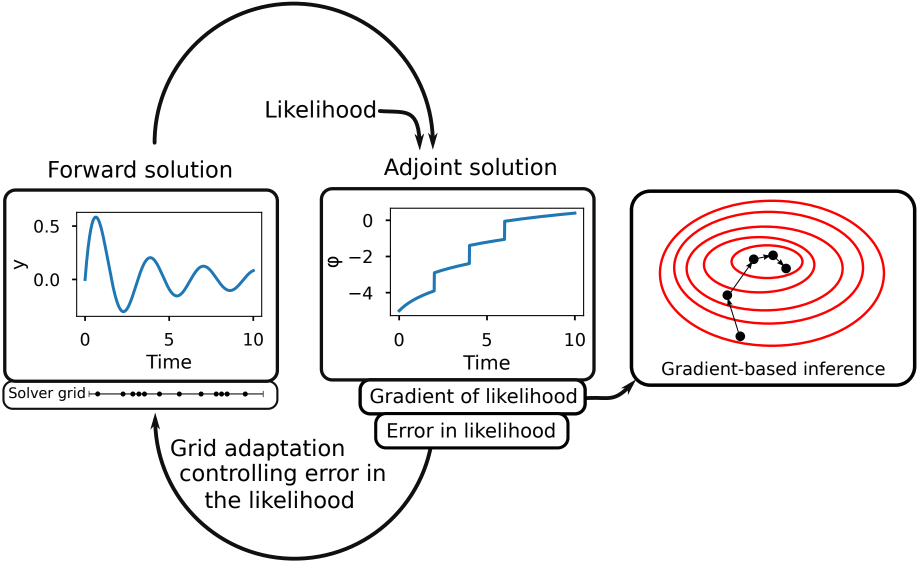
Adjoint methods for ODE inference.



Adjoint methods for ODE inference.



Adjoint methods for ODE inference.



Adjoint-based estimate of error.

$$Q(y) = \int_0^T q(y)dt$$

Q = quantity of interest (e.g., log-likelihood)

y = solution to ODE

Adjoint-based estimate of error.

$$Q(y) = \int_0^T q(y)dt$$

$$q(y) = \sum_{k=1}^{K} \delta(t^{(k)})(\tilde{y}^{(k)} - y(t^{(k)};\theta))^2$$

Q = quantity of interest (e.g., log-likelihood)

y = solution to ODE

$\{t^{(k)}\}$ = time points
$\{\tilde{y}^{(k)}\}$ = data points

Adjoint-based estimate of error.

$$Q(y) = \int_0^T q(y)dt$$

$$q(y) = \sum_{k=1}^{K} \delta(t^{(k)})(\tilde{y}^{(k)} - y(t^{(k)};\theta))^2$$

$$e_Q = Q(y) - Q(\hat{y})$$

Q = quantity of interest (e.g., log-likelihood)

y = solution to ODE

$\{t^{(k)}\}$ = time points
$\{\tilde{y}^{(k)}\}$ = data points

$\hat{y}$ = numerical approximation to ODE solution

Adjoint-based estimate of error.

$$e_Q \approx \int_0^T (y - \hat{y}) \left. \frac{\partial q}{\partial y} \right|_{\hat{y},\theta} dt$$

Q = quantity of interest (e.g., log-likelihood)

y = true solution to ODE

$\hat{y}$ = numerical approx solution

Adjoint-based estimate of error.

$$e_Q \approx \int_0^T (y - \hat{y}) \left. \frac{\partial q}{\partial y} \right|_{\hat{y},\theta} dt$$

$$\begin{aligned} \frac{d}{dt}(y - \hat{y}) &= f(y, t; \theta) - \dot{\hat{y}} \\ &\approx f(\hat{y}, t; \theta) - \dot{\hat{y}} + \frac{\partial f}{\partial y}(y - \hat{y}) \end{aligned}$$

$Q =$ quantity of interest (e.g., log-likelihood)

$y =$ true solution to ODE

$\hat{y} =$ numerical approx solution

Adjoint-based estimate of error.

$$e_Q \approx \int_0^T (y - \hat{y}) \left.\frac{\partial q}{\partial y}\right|_{\hat{y},\theta} dt$$

$$\boxed{= \int_0^T \left(f(\hat{y}, t; \theta) - \dot{\hat{y}}\right) \phi \, dt.}$$

Q = quantity of interest (e.g., log-likelihood)

y = true solution to ODE

$\hat{y}$ = numerical approx solution

ϕ = adjoint state

Adjoint differential equation:

$$\frac{d\phi}{dt} = \phi \left.\frac{\partial f}{\partial y}\right|_{\hat{y},\theta} + \left.\frac{\partial q}{\partial y}\right|_{\hat{y}}; \quad t \in (T, 0]$$

$$\phi(t = T) = 0$$

Adjoint-based estimate of gradient.

Consider the same ODE at parameters $\theta + \Delta\theta$:

$$\frac{d}{dt}(y+z) = f(t, y+z, \theta+\Delta\theta)$$

$$\frac{dQ}{d\theta} \approx \frac{1}{\Delta\theta}\int_0^T z\frac{\partial q}{\partial y}dt$$

Q = quantity of interest (e.g., log-likelihood)

y = true solution to ODE

z = perturbation to ODE solution

Adjoint-based estimate of gradient.

Consider the same ODE at parameters $\theta + \Delta\theta$:

$$\frac{d}{dt}(y + z) = f(t, y + z, \theta + \Delta\theta)$$

$$\frac{dQ}{d\theta} \approx \frac{1}{\Delta\theta} \int_0^T z \frac{\partial q}{\partial y} dt$$

Q = quantity of interest (e.g., log-likelihood)

y = true solution to ODE

z = perturbation to ODE solution

$$\frac{d}{dt}(y + z) \approx f(t, y, \theta) + \frac{\partial f}{\partial y} z + \frac{\partial f}{\partial \theta} \Delta\theta$$

$$\frac{dz}{dt} = \frac{\partial f}{\partial y} z + \frac{\partial f}{\partial \theta} \Delta\theta; \quad z(t = 0) = 0$$

Adjoint-based estimate of gradient.

Consider the same ODE at parameters $\theta + \Delta\theta$:

$$\frac{dQ}{d\theta} \approx \frac{1}{\Delta\theta} \int_0^T z \frac{\partial q}{\partial y} dt$$

$$\boxed{\approx \int_0^T \frac{\partial f}{\partial \theta} \phi \, dt}$$

Q = quantity of interest (e.g., log-likelihood)

y = true solution to ODE

z = perturbation to ODE solution

ϕ = adjoint state

This is the same adjoint state ϕ we used for e_Q.

Summary of adjoint-based method for parameter inference for DEs.

- At parameter values θ proposed by the inference algorithm:
  - Solve the forward problem, to obtain the numerical solution $\hat{y}$.

Summary of adjoint-based method for parameter inference for DEs.

- At parameter values θ proposed by the inference algorithm:
  - Solve the forward problem, to obtain the numerical solution $\hat{y}$.
  - Solve the adjoint problem, to obtain ϕ.

Summary of adjoint-based method for parameter inference for DEs.

- At parameter values θ proposed by the inference algorithm:
  - Solve the forward problem, to obtain the numerical solution $\hat{y}$.
  - Solve the adjoint problem, to obtain ϕ.
  - Evaluate the two previously derived expressions, to obtain:
    - the error in the log-likelihood resulting from numerical approximation
    - and the gradient in the log-likelihood.

Summary of adjoint-based method for parameter inference for DEs.

- At parameter values θ proposed by the inference algorithm:
  - Solve the forward problem, to obtain the numerical solution $\hat{y}$.
  - Solve the adjoint problem, to obtain ϕ.
  - Evaluate the two previously derived expressions, to obtain:
    - the error in the log-likelihood resulting from numerical approximation
    - and the gradient in the log-likelihood.
  - If the error is too high, refine the grid and repeat. Otherwise, use the gradient to drive HMC or NUTS or a similar algorithm to the next parameter values.

How accurate does the log-likelihood need to be?

- For the absolute Bayes factor between the true posterior and the numerical posterior not to exceed $1 + b$, we should have:

$$|e_Q| < b$$

Capistrán et al.: "Error control of the numerical posterior with Bayes factors in Bayesian uncertainty quantification," *Bayesian Analysis* 1.1 (2021).

DE model and inference methods.

$$\ddot{y} + k\dot{y} + cy = F(t)$$

$$y(t=0) = 0; \quad \dot{y}(t=0) = 2$$

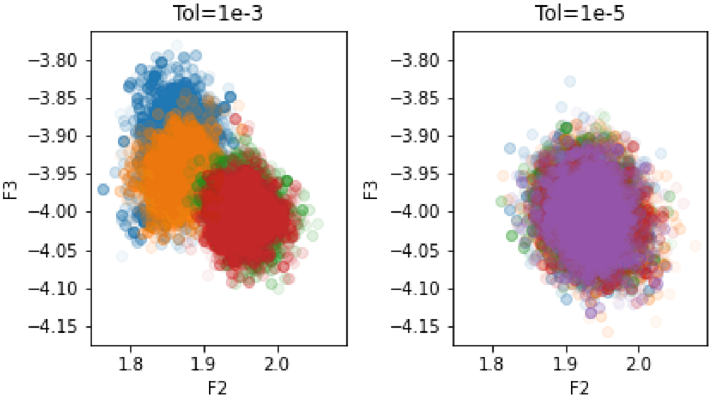
DE model and inference methods.

$$\ddot{y} + k\dot{y} + cy = F(t)$$

$$y(t=0) = 0; \quad \dot{y}(t=0) = 2$$

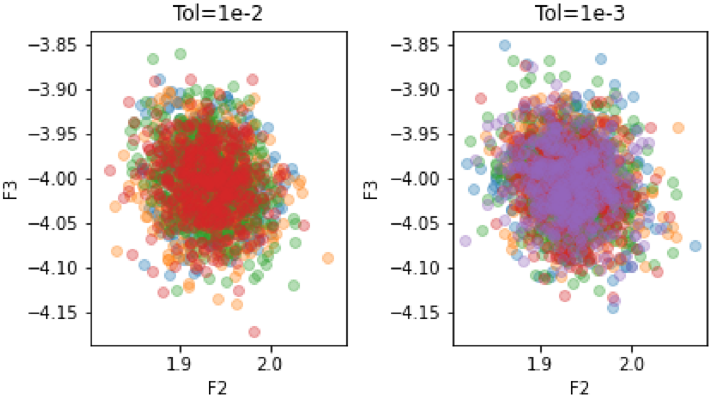
Method	Forward solution adaptation	Inference
Traditional	Local truncation error	Adaptive covariance
Adjoint	**Log-likelihood (adjoint-based)**	**NUTS (adjoint-based gradient)**

Adaptation based on error in the log-likelihood avoids having to tune solver tolerances.



Inference via Traditional method.

Adaptation based on error in the log-likelihood avoids having to tune solver tolerances.

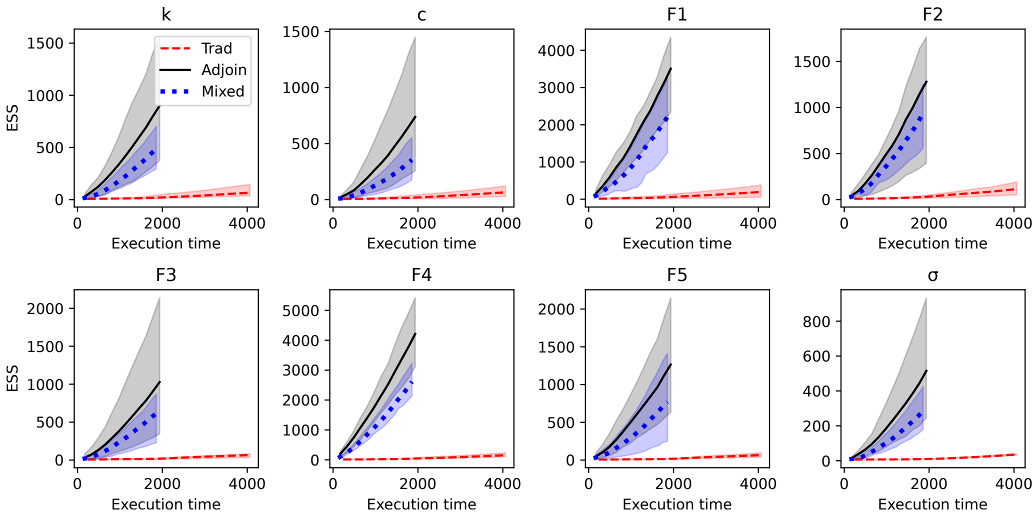


Inference via Adjoint method.

Comparison of inference performance: ESS.

Method	Forward solution adaptation	Inference
Traditional	Local truncation error	Adaptive covariance
Adjoint	**Log-likelihood (adjoint-based)**	**NUTS (adjoint-based gradient)**
Mixed	Local truncation error	**NUTS (adjoint-based gradient)**

Comparison of inference performance: ESS.



Future work

- Extend methods to other solvers.

Future work

- Extend methods to other solvers.
- Improved strategies for grid adaptation.

Acknowledgements.

Collaborative project between:

- Oxford Computer Science (**Richard Creswell**, **Martin Robinson**, **David Gavaghan**)
- University of Exeter (**Ben Lambert**)
- University of Macau (**Chon Lok Lei**)
- Colorado State University (**Simon Tavener**)