Introduction to State-Space Probabilistic ODE Solvers

Oksana Chkrebtii
Department of Statistics
The Ohio State University

#GPSS22
September 15, 2022

The Bayesian paradigm

Bayesian inference is concerned with modeling and updating degree of belief about an unknown quantity via probability statements.

For example, we may not know η but we may have some prior knowledge about, e.g., its range, most probable values, even before any data is collected,

$$\eta \sim \pi(\eta)$$

We seek to update our prior knowledge by conditioning on any new information, d, e.g., field data, model evaluations, via Bayes' Rule,

$$\pi(\eta \mid d) = \frac{p(d \mid \eta)\,\pi(\eta)}{\int p(d \mid \eta)\,\pi(\eta)\,d\eta} \propto p(d \mid \eta)\,\pi(\eta)$$

The inverse problem

We wish to estimate the unknown parameters, $\theta \in \Theta$, from observations,

$$y(x_i) = A\{u(x_i, \theta)\} + \varepsilon(x_i), \quad x_i \in \mathcal{X}, \;\; i = 1, \ldots, T,$$

of the deterministic state $u(x_i, \theta)$ transformed via an observation process A, and contaminated with stochastic noise ε.

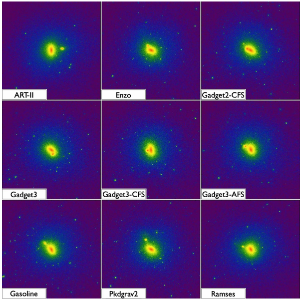
Likelihood function involves the unknown explicit solution, $u(x_i, \theta)$. This is termed the forward model.

The classical approach constructs a surrogate model by replacing u with an approximate numerical solution, u^N, discretized over a grid/mesh of size N. We wish to characterize the uncertainty in this approximation.

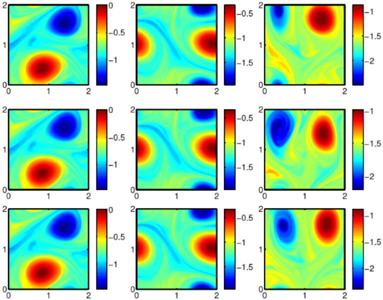
Motivation: the need for numerical uncertainty quantification in the forward problem

Motivation

When the solution to the system equations is not known in closed form we may wish to replace numerical approximation with a stochastic process reflecting probable trajectories for the solution.



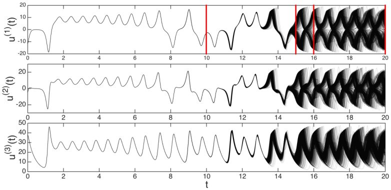
Density-weighted numerical solutions from nine different numerical solvers for a model of squared dark matter density



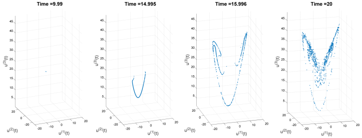
Time evolution of vorticity from the discretized Navier-Stokes equations on a torus with periodic boundary conditions

Motivation

Do numerical error bounds correctly characterize uncertainty when the solution is geometrically constrained?



1000 draws for Lorenz63 system at four fixed time points
(fixed initial states and model parameters in the chaotic regime).



1000 draws for Lorenz63 system at four fixed time points
(fixed initial states and model parameters in the chaotic regime).

Ultimate goal is inference on θ

We can incorporate this uncertainty about the exact solution within a Bayesian Hierarchical Model

$$
\begin{aligned}
[y \mid u, \theta] &\propto \rho\{y - A(u)\} \\
[u \mid \theta] &= \text{a probability model representing uncertainty} \\
&\quad\ \text{in the solution given discretization of size } N \\
[\theta] &= \pi(\theta).
\end{aligned}
$$

Probabilistic numerical methods provide a model for the uncertainty arising from discretization of a fixed but unknown solution.

Probability Modeling for Discretization Uncertainty

The unknown ODE solution - what we know

For fixed θ, consider the ODE initial value problem,

$$\begin{cases} Du &= f\left(t, u\right), & t \in (0, L], \\ u &= u_1, & t = 0, \end{cases}$$

- u_1 is a vector of initial states,
- D is a linear differential operator,
- $f : [0, L] \times \mathbb{R}^p \to \mathbb{R}^p$ is Lipschitz continuous in the second argument.

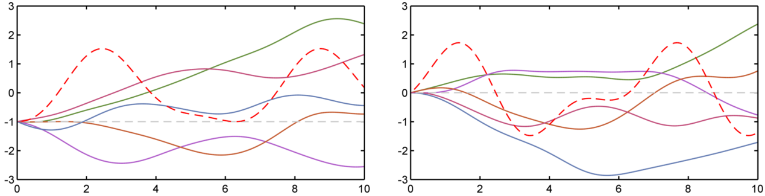
What we know a priori about the unique solution:

- Boundary constraints, precision (e.g. stable, stiff, chaotic), smoothness of $u(t)$
- For $t_1 < t_2$, the solution $u(t_2)$ is a function of $u(t_1)$ that does not depend on $u(\tau), \tau \in [0, t_1)$

Prior model over ODE solution [Skilling 1991]

Gaussian process (GP) prior for the solution and its time derivative given fixed hyperparameters $(m_t^0, m^0, \alpha, \lambda)$ is,

$$\begin{pmatrix} Du(t_k) \\ u(t_\ell) \end{pmatrix} \sim \mathcal{GP}\left\{ \begin{pmatrix} Dm^0(t_k) \\ m^0(t_\ell) \end{pmatrix}, \begin{pmatrix} DC^0(t_k, t_k)D^* & DC^0(t_k, t_\ell) \\ C^0(t_\ell, t_k)D^* & C^0(t_\ell, t_\ell) \end{pmatrix} \right\},$$



Five samples from the prior process: state (left) and first derivative (right)

$$\begin{cases} \frac{d}{dt^2} u^2 = \sin(2t) - u, & t \in [0, 10], \\ \frac{d}{dt} u(0) = 0, \ u(0) = -1, \end{cases}$$

Bayesian(ish) ODE solvers

[Cockayne, Oates, Sullivan, Girolami, 2017]

- *Exact* Bayesian collocation-based approach
- Pros: conditioning on exact solution evaluations (or arbitrarily close in practice)
- Cons: speed, direct sampling infeasible except when the ODE admits a solvable Lie algebra [Wang, Cockayne, Oates, 2018]

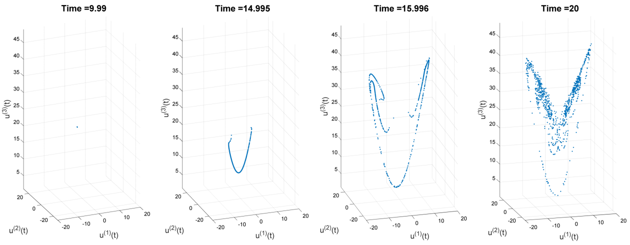
[Skilling, 1991], [Hennig and Hauberg, 2014]

- Extrapolate, condition on predictive mean to update GP (*approximate*)
- Pros: posterior is a GP - fast, simple, intuitive
- Cons: posterior is a GP - cannot easily be restricted to a manifold

Bayesian(ish) ODE solvers

[Chkrebtii, 2014], [Chkrebtii, Campbell, Calderhead, Girolami 2016]

- State-space based approach is also *approximate*
- Cons: requires eliciting an error model and hyperparameters
- Pros: admits uncertainty estimates that are non-Gaussian
- Pros: Cost proportional to numerical solver & fully parallelizable



1000 draws for Lorenz63 system at four fixed time points

(fixed initial states and model parameters in the chaotic regime).

State-space probabilistic ODE solver

[Chkrebtii, 2014], [Chkrebtii, Campbell, Calderhead, Girolami 2016]

Discretize time domain by an ordered grid $\{s_i\}_{i=1,\ldots,N}$. ODE is interrogated sequentially at these grid points by generating auxiliary pairs of state and derivative evaluations:

$$A_i = \left\{\, u_i^{i-1} := u^{i-1}(s_i),\; f_i := f(s_i, u_i^{i-1}) \,\right\}, \quad i = 1, \ldots, N.$$

Posterior density over the state u evaluated at time $s \in [0, L]$ is,

$$\pi(u \mid f(\cdot), u_1, \theta, N) = \int \pi(u, A_{1:N} \mid f(\cdot), u_1, \theta, N)\, dA_{1:N}$$

$$\propto \int p\left(u \mid f_{1:N}, u_1\right) \prod_{i=1}^{N} \left\{\, p\left(f_i \mid u_i^{i-1}\right)\; p\left(u_i^{i-1} \mid f_{1:i-1}, u_1\right) \,\right\} dA_{1:N}.$$

State-space probabilistic ODE solvers

Probabilistic analogue of linearization is the error model,

$$f_i := f(s_i, u^{i-1}(s_i)) = Du(s_i) + \xi(s_i), \quad i = 1, \ldots, N,$$

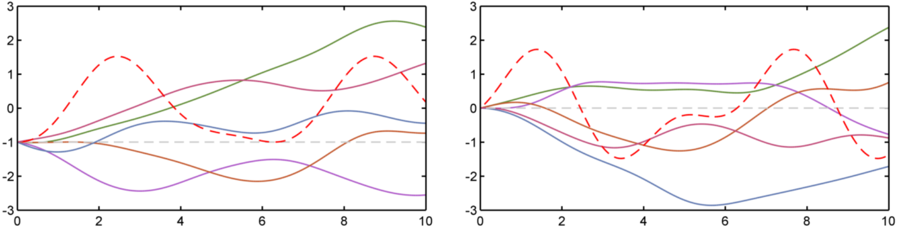
the term $\xi \sim N(0, Q(s_i, s_i))$ represents solution uncertainty. Updates are:

$$\left(\begin{matrix} Du(t_k) \\ u(t_\ell) \end{matrix} \,\middle|\, f_{1:i} \right) \sim \mathcal{GP} \left\{ \begin{pmatrix} Dm^i(t_k) \\ m^i(t_\ell) \end{pmatrix}, \begin{pmatrix} DC^i(t_k, t_k)D^* & DC^i(t_k, t_\ell) \\ C^i(t_\ell, t_k)D^* & C^i(t_\ell, t_\ell) \end{pmatrix} \right\},$$

where means and covariances can be defined recursively as,

$$\begin{aligned} m^i(t) &= m^{i-1}(t) + K^i(t, s_i) \left\{ f_i - Dm^{i-1}(s_i) \right\}, \\ C^i(t_k, t_\ell) &= C^{i-1}(t_k, t_\ell) - K^i(t_k, s_i) DC^{i-1}(s_i, t_\ell), \\ K^i(t, s_i) &= C^{i-1}(t, s_i) D^* \left(Q(s_i, s_i) + DC^{i-1}(s_i, s_i) D^* \right)^{-1}. \end{aligned}$$

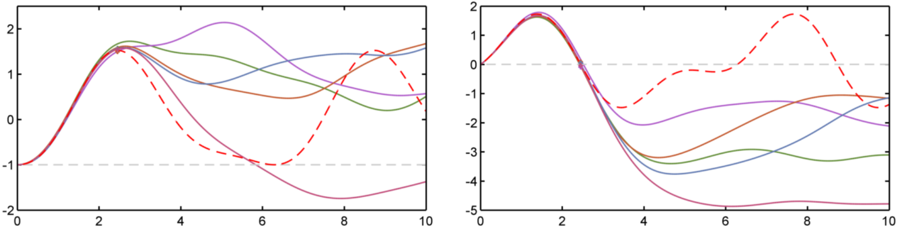
Example - simple ODE initial value problem



Five draws from the updated process for the state (left) and first derivative (right)

$$
\begin{cases}
\frac{d}{dt^2} u^2 = \sin(2t) - u, \quad t \in [0, 10], \\
\frac{d}{dt} u(0) = 0, \; u(0) = -1,
\end{cases}
$$

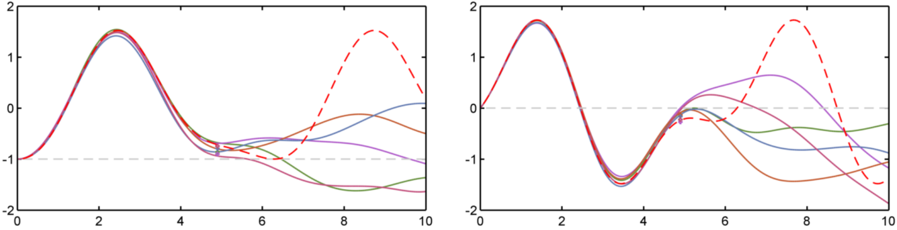
Example - simple ODE initial value problem



Five draws from the updated process for the state (left) and first derivative (right)

$$\begin{cases} \frac{d}{dt^2}u^2 = \sin(2t) - u, \quad t \in [0, 10], \\ \frac{d}{dt}u(0) = 0, \ u(0) = -1, \end{cases}$$

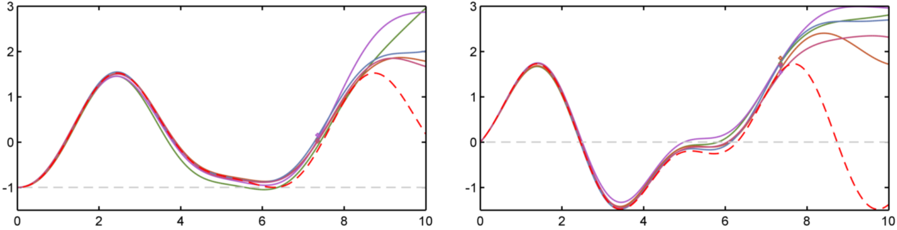
Example - simple ODE initial value problem



Five draws from the updated process for the state (left) and first derivative (right)

$$\begin{cases} \frac{d}{dt^2}u^2 = \sin(2t) - u, \quad t \in [0, 10], \\ \frac{d}{dt}u(0) = 0, \ u(0) = -1, \end{cases}$$

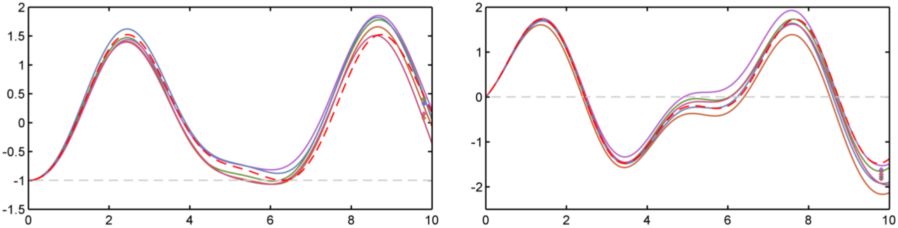
Example - simple ODE initial value problem



Five draws from the updated process for the state (left) and first derivative (right)

$$\begin{cases} \frac{d}{dt^2}u^2 = \sin(2t) - u, \quad t \in [0, 10], \\ \frac{d}{dt}u(0) = 0, \ u(0) = -1, \end{cases}$$

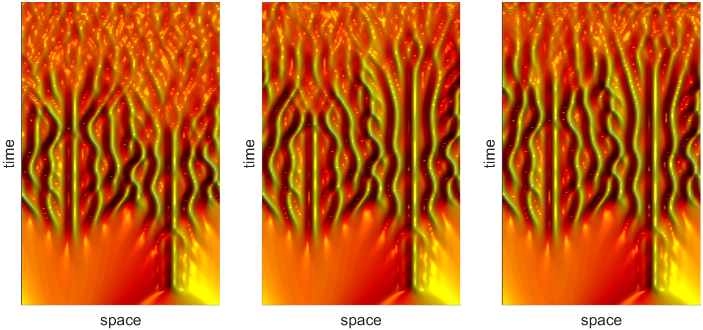
Example - simple ODE initial value problem



Five draws from the updated process for the state (left) and first derivative (right)

$$\begin{cases} \frac{d}{dt^2}u^2 = \sin(2t) - u, \quad t \in [0, 10], \\ \frac{d}{dt}u(0) = 0, \; u(0) = -1, \end{cases}$$

Uncertainty decreases with grid size



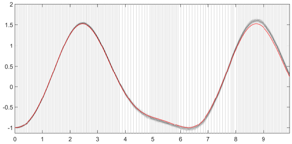
Kuramoto-Sivashinsky model of a reaction-diffusion system discretized with 1000, 2000, and 3000 equally spaced points respectively,

$$\begin{cases} \frac{\partial}{\partial t} u & = -u \frac{\partial}{\partial x} u - \frac{\partial^2}{\partial x^2} u - \frac{\partial^4}{\partial x^4} u, \qquad x \in [0, 32\pi],\ t \in (0, 150] \\ u & = \cos\left(\frac{x}{16}\right) \{1 + \sin\left(\frac{x}{16}\right)\}, \qquad x \in [0, 32\pi],\ t = 0. \end{cases}$$

Some related work

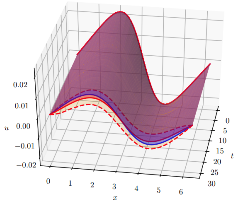
[Chkrebtii, Campbell, 2019]

Propose to adaptively select the discretization grid for the state-space probabilistic method by solving a design problem.



[Wang, Cockayne, Chkrebtii, Sullivan, Oates, 2021]

Develop a forward-in-time, continuous-in-space (FTCS) approach to solving nonlinear PDEs.



Selected references

1. Skilling, J. *Bayesian Solution of Ordinary Differential Equations* (Kluwer Academic Publishers, Seattle), 1991.
2. Cockayne, J., Oates, C.J. , Sullivan, T.J., Girolami, M., *Bayesian Probabilistic Numerical Methods*, SIAM Review, 2019.
3. Wang, J., Cockayne, J., and Oates, C. J., *A Role for Symmetry in the Bayesian Solution of Differential Equations*, Bayesian Analysis, 2020.
4. Chkrebtii, O.A., Campbell, D.A., Calderhead, B., Girolami, M. , *Bayesian Solution Uncertainty Quantification for Differential Equations*, Bayesian Analysis, 2016.
5. Hennig, P., and Hauberg, S., *Probabilistic Solutions to Differential Equations and their Application to Riemannian Statistics*, In Proc. of the 17th int. Conf. on Artificial Intelligence and Statistics (AISTATS), 2014.
6. Chkrebtii, O.A., Campbell, D.A. *Adaptive Grid Designs for State-Space Probabilistic ODE Solvers*, Statistics and Computing, 2019.
7. Wang, J., Cockayne, J., Chkrebtii, O.A., Sullivan, T.J., Oates, C.J., *Bayesian Numerical Methods for Nonlinear Partial Differential Equations*, Statistics and Computing, 2021.