Solving ODEs in a Bayesian context: challenges and opportunities

Charles C. Margossian[1], Lu Zhang[1], Sebastian Weber[2] and Andrew Gelman[1] [1]Department of Statistics, Columbia University, USA; [2]Novartis Pharma, Switzerland.

mc-stan.org

Bayesian inference with Hamiltonian Monte Carlo

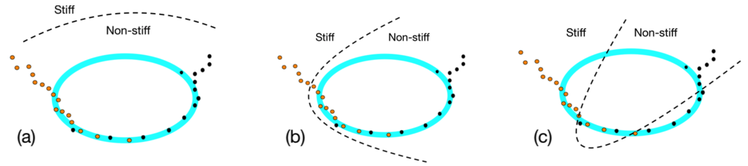

Figure 1. *As the Markov chains (orange and black dots) move across the parameter space, the behavior of the ODE may change. The elliptical blue band represents the sampling region, i.e. the region where the posterior probability mass concentrates. Three hypothetical cases for how the ODE may behave.*

Hamiltonian Monte Carlo. Given observations, y, and parameters, θ, our goal is to characterize the posterior distribution $p(\theta \mid y)$. We focus on pharmacometrics models, for which the likelihood $p(y \mid \theta)$ uses an ODE. *Dynamic Hamiltonian Monte Carlo* (HMC) is a state-of-the-art Markov chains Monte Carlo method which draws approximate samples from $p(\theta \mid y)$ [1]. Effective sampling crucially depends on properly tuning HMC's sampling parameters during a *warmup phase*.

Warmup phases of dynamic HMC (e.g. for 500 iterations) **and prototype Path Finder.**

	Adaptive HMC	**Prototype Path finder**
I (75 iter)	• starting at initialization, θ_0, early exploration with highly varying θ. • convergence to sampling region. • initial tuning of the sampling parameters.	• Construct optimization path between θ_0 and the mode of $p(\theta \mid y)$. • Find sampling region along path and initialize HMC Phase I.
II (375 iter)	• semi-stable exploration of sampling region. • more extensive tuning of sampling parameters, as we learn more about $p(\theta \mid y)$.	• Use variational approx. of $p(\theta \mid y)$ to tune sampling parameters. • OR use Phase II of HMC.
III (50 iter)	• continued exploration of the sampling region. • final tuning of sampling parameters.	• Use HMC's Phase III.

Challenge. As θ changes, so can the behavior of the ODE in our model (Figure 1).

Pathfinder [3]. The pathfinder offers alternative options for warming up HMC by (i) improving initialization and (ii) estimating certain sampling parameters. (Figure 2)

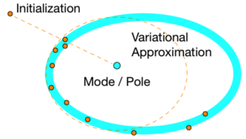

Figure 2.

Behavior of ODE-based models during HMC sampling

Michaelis-Menten PK model. Consider a simple non-linear PK model:

$$y_0' = -k_a y_0$$
$$y_1' = k_a y_0 - \frac{V_m C}{K_m + C},$$

where $C = y_1/V$ is the drug concentration in a central compartment. The patient is administered a single dose and the drug concentration c_{obs} is measured over time.

The full model, with parameters $\theta = \{k_a, V, V_m, K_m, \sigma\}$, is

$$k_a \sim \text{logNormal}(\log(2.5), 3); \quad V \sim \text{logNormal}(\log(35), 0.5); \quad V_m \sim \text{logNormal}(\log(10), 0.5);$$
$$K_m \sim \text{logNormal}(\log(2.5), 3); \quad \sigma \sim \text{Normal}^+(0, 1); \quad c_{\text{obs}} \sim \text{Normal}(C, \sigma).$$

Which ODE integrator should we use?

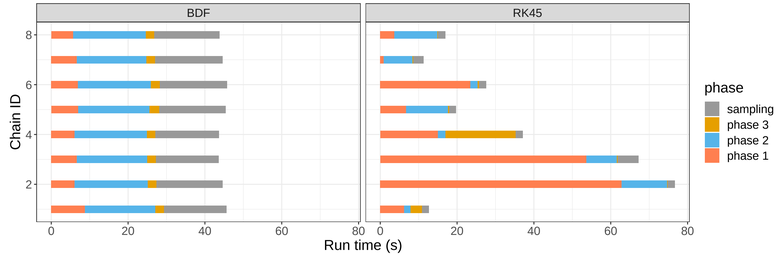

Figure 2. *Model runtimes using a stiff backward differentiation (BDF) integrator and a non-stiff Runge-Kutta $4^{th}/5^{th}$ (RK45) integrator. For each integrator we run 8 HMC chains (500 warmup + 500 sampling iterations) using Stan [2].*

Combining different integrators. A robust integrator may be necessary during the warmup, but overkill when sampling. Applying this heuristic, we propose the following schemes:

	Pathfinding	**Phase I**	**Phase II**	**Phase III**	**Sampling**
HMC warmup, early switch	NA	BDF	RK45	RK45	RK45
HMC warmup, late switch	NA	BDF	BDF	RK45	RK45
Pathfinder, RK45	BDF	RK45	RK45	RK45	RK45
Pathfinder, late switch	BDF	BDF	BDF	RK45	RK45
Pathfinder, approx. tuning	BDF	NA	NA	RK45	RK45

Performance Study

We run all five proposed sampling schemes, as well as HMC using only RK45 or BDF on the Michaelis-Menten model and a population version of it. The inference they produce are all in agreement.

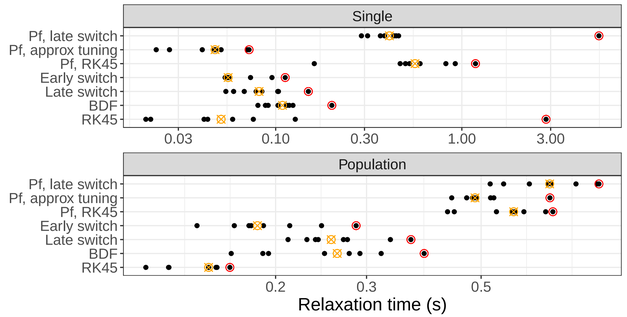

Figure 3. *Relaxation time, i.e. time to increase the effective sample size by 1, measured for $\log p(\theta, y)$. The orange crossed dot is the median time, and the red circled dot the worst time. For method which run 8 or more chains in parallel, we may prefer consistency to good median performance.*

Model details. The population model uses 3 patients, with partial pooling. We tried a centered and non-centered parameterization, and reported the former which for this particular problem produced more effective sampling. We also found adapting a dense mass matrix for HMC worked better than using the default diagonal mass matrix. The population uses 1,000 warmup and 1,000 sampling iterations.

Results. • For the single patient model, using BDF during the warmup phase improves the stability of the sampler. Improved initialization do not warrant the additional cost of running the pathfinder; however estimating tuning parameters based on the variational approximation produces the most efficient sampler. Improved implementation can further increase the benefits of the pathfinder.
• For the population model, we suspect the additional data stabilizes the posterior distribution, making uniform RK45 the best option. For this model, running the pathfinder is expansive and the estimated tuning parameter is suboptimal (we can diagnose this with the Pareto smooth importance sampling [3]).

References

[1] M. Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434v1*, 2018.

[2] Stan development team. *Stan reference manual.* 2021.

[3] L. Zhang, B. Carpenter, A. Vehtari, and A. Gelman. Pathfinder: Parallel quasi-newton variational inference. *arXiv:2108.03782*, 2021.

Population Approach Group in Europe, September 2021