Documentation for the RMHMC Source

Andrei Kramer

December 11, 2013

This text provides some documentation and mathematical considerations and definitions for the implementation of the Riemannian Manifold Hamiltonian Monte Carlo (RMHMC) algorithm.

1 Model Specifications

We consider a deterministic ODE model and a stochastic measurement model. System states are captured by the state variables $x(t) \in \mathbb{R}^n$. The model parameters $\theta \in \mathbb{R}^m$ describe interactions between the model state variables and are unknown. A second set of parameters $u \in \mathbb{R}^l$ describes the conditions of an experimental setup. These parameters are considered inputs and we assume that they can be set by the experimenter. They can be external parameters, e.g. the temperature, or describe modifications to the system, e.g. inhibitions to some of the interactions. The known time derivative of $x(t)$ defines the model:

$$\dot{x} = f(t, x; \theta, u)\,, \tag{1}$$

$$y(t_j; \theta, u) = C(\zeta)x(t_j; \theta, u) + \epsilon_i(t_j)\,, \tag{2}$$

where $y(t; \theta, u) \in \mathbb{R}^k$ is the output of the measurement process, which is recorded at T time points. The linear transformation $C \in \mathbb{R}^{k\times n}$ models the capabilities of the measurement setup. The measurements are obscured by the noise process $\epsilon_i(t_j) \sim \mathcal{N}(0, \sigma_{ij}^2)$ $(i = 1, \ldots, n; j = 1, \ldots, T)$. In addition, the observations might be done in unknown units, such that a reference Experiment might be needed to interpret any numerical values of $y(t_j)$. To deal with this problem, C may contain unknown scaling parameters ζ, to be estimated via MCMC. Another approach, which instead eliminates the scaling parameters, is to take the ratios:

$$\tilde{y}_i(t_j; \theta, u_b) = \frac{y_i(t_j, \theta, u_b)}{y_i(t_j; \theta, u_0)}\,, \tag{3}$$

where $u_b \in \mathbb{R}^l (b = 1, \ldots, n_{\mathrm{E}})$ is any particular experimental setup (n_{E} is the number of experiments) and u_0 is the reference experiment setup. In consequence we have to recalculate the standard deviation of $\tilde{y}_i(t_j;\theta,u_b)$ from the standard deviations[1] of $y_i(t_j,\theta,u_b)$ and $y_i(t_j;\theta,u_0)$:

$$\tilde{\sigma}_{i,j} \approx \left|\frac{\partial\tilde{y}_i(t_j;\theta,u_b)}{\partial y_i(t_j;\theta,u_b)}\right|\sigma_{ij,b} + \left|\frac{\partial\tilde{y}_i(t_j;\theta,u_b)}{\partial y_i(t_j;\theta,u_0)}\right|\sigma_{ij,0}\,, \tag{4}$$

and in this case:

$$\tilde{\sigma}_{i,j} \approx \left|\frac{\tilde{y}_i(t_j;\theta,u_b)}{y_i(t_j;\theta,u_b)}\right|\sigma_{ij,b} + \left|\frac{\tilde{y}_i(t_j;\theta,u_b)}{y_i(t_j;\theta,u_0)}\right|\sigma_{ij,0}\,. \tag{5}$$

2 Sensitivities

The sensitivity of $\tilde{y}(t)$ in terms of the (known) $y(t;\theta,u)$ sensitivities $S(t;\theta,u)$ is:

$$\begin{aligned}\partial_{\theta_j}\tilde{y}_i(t;\theta,u_b) &= \frac{S_i^{\,j}(t;\theta,u_b)y_i(t;\theta,u_0) - y_i(t;\theta,u_b)S_i^{\,j}(t;\theta,u_0)}{y_i(t;\theta,u_0)^2}\\ &= \frac{S_i^{\,j}(t;\theta,u_b) - \tilde{y}_i(t;\theta,u_b)S_i^{\,j}(t;\theta,u_0)}{y_i(t;\theta,u_0)}\,.\end{aligned} \tag{6}$$

The code for this operation is located in the function `Likelihood` and is organized such, that if the data is absolute and does not require normalization, then $S_i^{\,j}(t;\theta,u_0)$ is set to 0 for all i,j and the reference $y_i(t;\theta,u_0) = 1$ for all i.

3 Sensitivity Gradient

We take the derivative of (6) for any $u_b \in \{u_1,\dots,u_{n_\mathrm{E}}\}$:

$$\begin{aligned}\partial_{\theta_k}\tilde{S}_i^{\,j}(t;\theta,u_b) = \Bigg(&\frac{\partial S_i^{\,j}(t;\theta,u_b)}{\partial\theta_k} - \tilde{S}_i^{\,k}(t;\theta,u_b)S_i^{\,j}(t;\theta,u_0) -\\ &\tilde{y}_i(t;\theta,u_b)\frac{\partial S_i^{\,j}(t;\theta,u_0)}{\partial\theta_k}\Bigg)\frac{1}{y_i(t;\theta,u_0)}\\ &+\Big(\tilde{y}_i(t;\theta,u_b)S_i^{\,j}(t;\theta,u_0) - S_i^{\,j}(t;\theta,u_b)\Big)\frac{S_i^{\,k}(t;\theta,u_0)}{(y_i(t;\theta,u_0))^2}\,.\end{aligned} \tag{7}$$

4 Sampling in logarithmic Space

Ode models are often unstable for negative parameters, which is definitely the case for biological models. Sampling in logarithmic space θ and passing $\rho = \exp(\theta)$ to the ode system fixes the problem. But it also means that we have to modify the above

[1] we append the input indeces $b,0$ to the standard deviation symbol for distinction

expressions. Because `vfgen` and `cvodes` provide sensitivity analysis with respect to the nominal model parameters ρ:

$$
\begin{aligned}
\dot{x} &= f(t,x;\rho,w)\,, & \rho_j &= \exp(\theta_j)\,,\\
y(t;\rho,w) &= Cx(t;\rho,w)\,, & b &= 1,\dots,l\,,\\
\tilde{y}_i(t;\rho,w) &= \frac{y_i(t;\rho,w)}{y_i(t,\rho,u_0)}\,,\\
\tilde{S}_i^{\,j}(t;\rho,w) &= \partial_{\rho_j}\tilde{y}_i(t;\rho,w)\,,\\
\partial_{\theta_j}\tilde{y}(t;\rho,w) &= \frac{\partial\tilde{y}_i(t;\rho,w)}{\partial\rho_j}\frac{\partial\rho_j}{\partial\theta_j}\\
&= \tilde{S}_i^{\,j}(t;\rho,w)\rho_j\,.
\end{aligned}
\tag{8}
$$

for any input w and consequently:

$$
\begin{aligned}
\partial_{\theta_j}\tilde{y}_i(t;\rho,w) &= \tilde{S}_i^{\,j}(t;\rho,w)\rho_j\,,\\
\partial_{\theta_k}\tilde{S}_i^{\,j}(t;\rho,w)\rho_j &= \frac{\partial\tilde{S}_i^{\,j}(t;\rho,w)}{\partial\rho_l}\overbrace{\frac{\partial\rho_l}{\partial\theta_k}}^{\rho_l\delta_{lk}}\rho_j + \tilde{S}_i^{\,j}(t;\rho,w)\frac{\partial\rho_j}{\partial\theta_k}\\
&= \frac{\partial\tilde{S}_i^{\,j}(t;\rho,w)}{\partial\rho_k}\rho_k\rho_j + \tilde{S}_i^{\,j}(t;\rho,w)\rho_k\delta_{jk}\,.
\end{aligned}
\tag{9}
$$

Note that it is possible to use `Expressions` in vfgen models:

```
<Parameter Name="\theta_1" DefaultValue="0">
<Expression Name="rho_1" Formula="exp(theta_1)"/>
```

to convert the parameters from logspace, then use these expressions to define fluxes. But, since `Expressions` need to be recalculated at every step this is very wasteful. On the other hand, VFGEN will then compute correct sensitivities ($dx_i/d\theta_j$). But, to save calls to the exp function, the software does this conversion before calling the solver and instead transforms the sensitivities as described here.