Multilevel Delayed Acceptance MCMC with an Adaptive Error Model in PyMC3

Mikkel B. Lykkegaard
Grigorios Mingas
Robert Scheichl
Colin Fox
Tim J. Dodwell

Intro

- We have developed a **multilevel** version of the popular **Delayed Acceptance** MCMC algorithm.
- The algorithm is implemented in the development version of the **probabilistic programming** framework **PyMC3**.
- If coupled with a state-independent **Adaptive Error Model** (AEM), the **Effective Sample Size (ESS)** can be drastically **increased**.

Methods

- The classic **Delayed Acceptance** (DA) MCMC algorithm is **extended** to support a **hierarchy** of coarse models, with each coarse level running finite length **subchains** (see Algorithm 2).

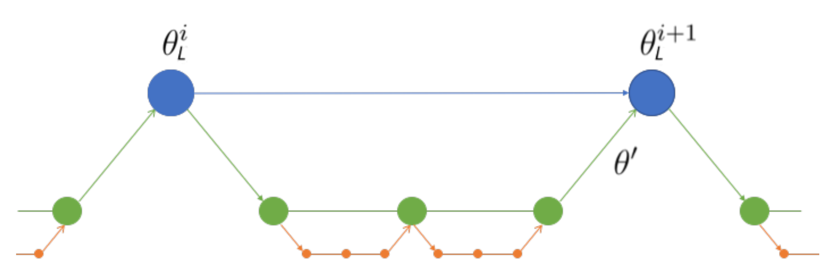

Fig. 1: Diagram of the MLDA algorithm. The finest level (blue) receives proposals from the coarse level below (green), which again receives proposals from the coarsest level below (orange). Each coarse level is running a subchain before passing a proposal.

- Since coarse models are **approximations** of the fine, they will embody additional approximation errors, e.g. **discretisation errors** in case of solving a PDE on a coarser grid.
- These approximation errors can be **corrected** using our multilevel extension of the DA **Adaptive Error Model** (AEM), which effectively offsets, scales and rotates the coarse **likelihood functions**.

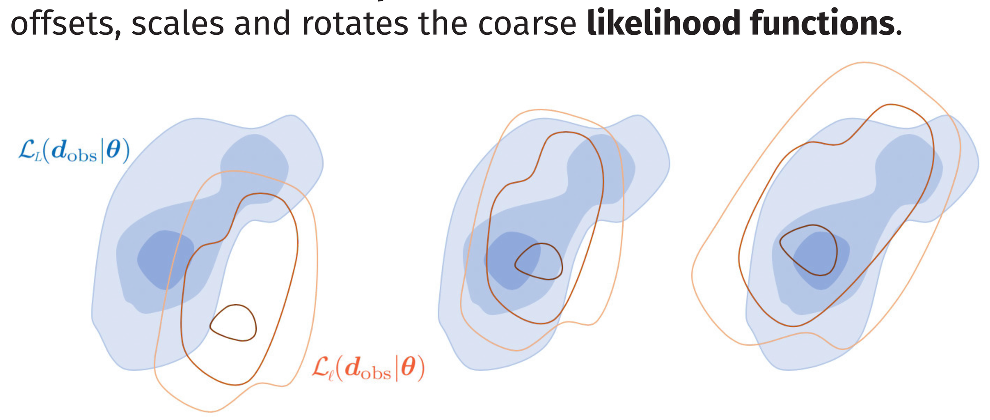

Fig. 2: Diagram of the AEM. Before correcting (left), the coarse likelihood (red/orange isolines) approximates the fine (blue contours) poorly. By offsetting (middle), scaling and rotating (right) the coarse likelihood function, it approaches the fine.

Results

- We tested the method on a steady state **subsurface flow problem,** which involves solving an **elliptic partial differential equation,** when evaluating the **likelihood function.**
- Using an **extremely coarse** model (see Fig. 4) on the lowest level and **no AEM,** the sampler did **not converge.** Utilising the **AEM,** the sampler **converged** and we achieved very **high ESS.**

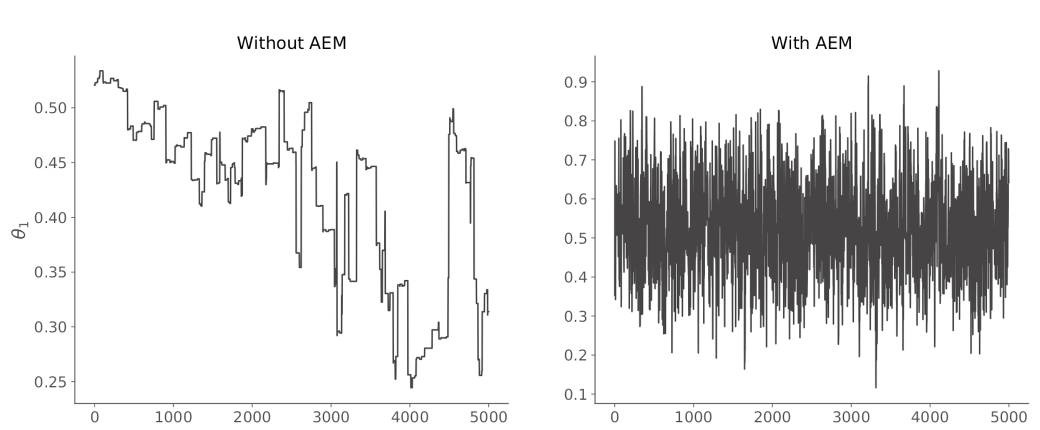

Fig. 3: Traceplot of a model parameter, with (right) and without (left) the AEM.

Using an **Adaptive Error Model** with Multilevel Delayed Acceptance MCMC can drastically **increase** the **Effective Sample Size**

Supplementary Material

Multilevel Delayed Acceptance MCMC

Algorithm 2 (Multilevel Delayed Acceptance MCMC):
Choose θ^0 and set the states of all subchains $\theta_0^0 = \ldots = \theta_{L-1}^0 = \theta^0$. Then, for $j = 0, \ldots, J-1$:

1. Given θ^j and $\theta_\ell^{j_\ell}$ such that $j_\ell < J_\ell$ for all $1 \leq \ell < L$, generate a subchain of length J_0 with Alg. 1 on level 0, starting from $\theta_0^0 = \theta_1^{j_1}$ and using the transition kernel $q(\theta_0' | \theta_0^{j_1})$.
2. Let $\ell = 1$ and $\theta_1' = \theta_0^{J_0}$.
3. If $\ell = L$ go to Step 7. Otherwise compute the delayed acceptance probability on level ℓ, i.e.,
$$\alpha_\ell = \min\left\{1, \frac{\mathcal{L}_\ell(d|\theta_\ell')\, \mathcal{L}_{\ell-1}(d|\theta_\ell^{j_\ell})}{\mathcal{L}_\ell(d|\theta_\ell^{j_\ell})\mathcal{L}_{\ell-1}(d|\theta_\ell')}\right\}.$$
4. Set $\theta_\ell^{j_\ell+1} = \theta_\ell'$ with probability α_ℓ and $\theta_\ell^{j_\ell+1} = \theta_\ell^{j_\ell}$ otherwise. Increment $j_\ell \to j_\ell + 1$.
5. If $j_\ell = J_\ell$ set $\theta_{\ell+1}' = \theta_\ell^{J_\ell}$, increment $\ell \to \ell + 1$ and return to Step 3.
6. Otherwise set $j_k = 0$ and $\theta_k^0 = \theta_\ell^{j_\ell}$, for all $0 \leq k < \ell$, and return to Step 1.
7. Compute the delayed acceptance probability on level L, i.e.,
$$\alpha_L = \min\left\{1, \frac{\mathcal{L}_\ell(d|\theta_L')\mathcal{L}_{\ell-1}(d|\theta^j)}{\mathcal{L}_\ell(d|\theta^j)\, \mathcal{L}_{\ell-1}(d|\theta_L')}\right\}.$$
Set $\theta^{j+1} = \theta_L'$ with probability α_L and $\theta^{j+1} = \theta^j$ otherwise. Increment $j \to j+1$.
8. Set $j_\ell = 0$ and $\theta_\ell^0 = \theta^j$, for all $0 \leq \ell < L$, and return to Step 1.

Adaptive Error Model

We extend the model definition for the coarse levels with a telescopic sum of the differences between model levels to account for approximation errors:

$$d = \mathcal{F}_L(\theta) + \epsilon = \mathcal{F}_\ell(\theta) + \mathcal{B}_\ell(\theta) + \epsilon \quad \text{with} \quad \mathcal{B}_\ell(\theta) := \sum_{k=\ell}^{L-1} \underbrace{\mathcal{F}_{k+1}(\theta) - \mathcal{F}_k(\theta)}_{:=B_k(\theta)}$$

Assuming that $\mathcal{B}_\ell^* \sim \mathcal{N}(\mu_{\mathcal{B},\ell}, \Sigma_{\mathcal{B},\ell})$, the likelihood function for the coarse levels can then be written as:

$$\mathcal{L}_\ell^*(d|\theta) \propto \exp\left(-\frac{1}{2}\left(d - \mathcal{F}_\ell(\theta) - \mu_\epsilon - \mu_{\mathcal{B},\ell}\right)^T \left(\Sigma_\epsilon + \Sigma_{\mathcal{B},\ell}\right)^{-1} \left(d - \mathcal{F}_\ell(\theta) - \mu_\epsilon - \mu_{\mathcal{B},\ell}\right)\right)$$

Forward Model

The forward model for the example was the steady state groundwater flow equation with random coefficients:

$$-\nabla \cdot k(x,\omega) \nabla p(x,\omega) = f(x)$$

The subsurface permeability was modelled as a log-Gaussian random field with a squared exponential covariance function, and parametrised using a truncated Karhunen-Loève expansion:

$$\log k(x,\omega) = \sum_{i=1}^{R} \sqrt{\mu_i} \phi_i(x) \theta_i(\omega)$$

We used three model levels with an extremely coarse model on the lowest level (Fig. 4).

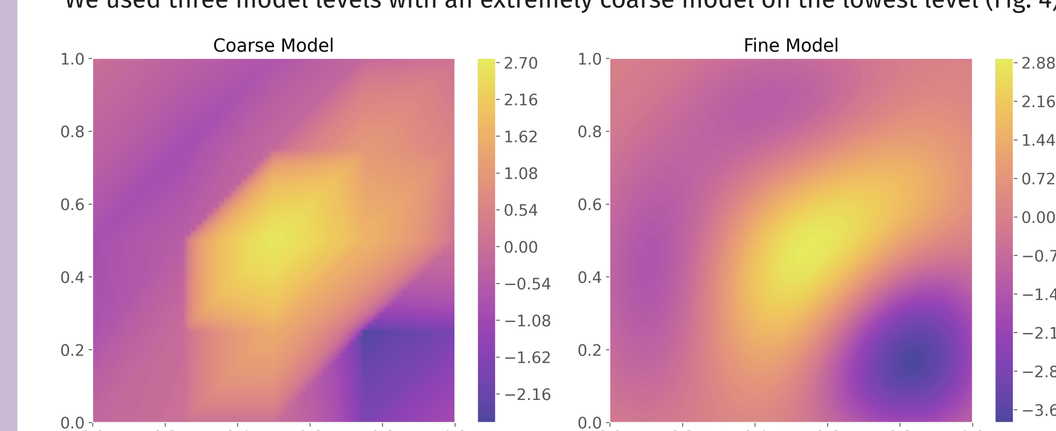

Fig. 4: A realisation of the permeability on the lowest (left) and highest level (right).

Acknowledgements

The work was funded by a Turing AI fellowship (2TAFFP\100007) and the Water Informatics Science and Engineering Centre for Doctoral Training (WISE CDT) under a grant from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/L016214/1.

References

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. In: The Journal of Chemical Physics, 21(6)

Christen, J. A., & Fox, C. (2005). Markov chain Monte Carlo Using an Approximation. In: Journal of Computational and Graphical Statistics, 14(4), 795–810.

Cui, T., Fox, C., & O'Sullivan, M. J. (2011). Adaptive Error Modelling in MCMC Sampling for Large Scale Inverse Problems, Technical report

This poster was inspired by the #betterposter design:
https://www.youtube.com/watch?v=1RwJbhkCA58