Multilevel Markov chain Monte Carlo methods for Uncertainty Quantification

Aretha Teckentrup

School of Mathematics, University of Edinburgh

Joint work with:
Tim Dodwell (Exeter), Chris Ketelsen (UC Boulder), Rob Scheichl (Bath)

SIAM UQ '18 - April 19, 2018

Main ingredient 1: Multilevel Monte Carlo

Set-up

- Suppose we want to compute $\mathbb{E}_\mu[Q]$, for some quantity of interest $Q = Q(\theta)$ depending on parameters $\theta \sim \mu$.
- In many applications, the evaluation of Q involves solving a differential equation. We cannot compute Q exactly, and instead have access to a sequence of approximations $\{Q_\ell\}_{\ell=0}^\infty$ where:
  - Q_0 is the cheapest, but also the least accurate approximation,
  - Q_ℓ is increasing in cost and accuracy as ℓ increases, with $Q_\infty = Q$.

Main ingredient 1: Multilevel Monte Carlo

Set-up

- Suppose we want to compute $\mathbb{E}_\mu[Q]$, for some quantity of interest $Q = Q(\theta)$ depending on parameters $\theta \sim \mu$.
- In many applications, the evaluation of Q involves solving a differential equation. We cannot compute Q exactly, and instead have access to a sequence of approximations $\{Q_\ell\}_{\ell=0}^{\infty}$ where:
  - Q_0 is the cheapest, but also the least accurate approximation,
  - Q_ℓ is increasing in cost and accuracy as ℓ increases, with $Q_\infty = Q$.
- Using the linearity of expectation, we can write

$$\mathbb{E}_\mu[Q] = \mathbb{E}_\mu[Q_0] + \sum_{\ell=1}^{\infty} \mathbb{E}_\mu[Q_\ell - Q_{\ell-1}].$$

- For practical purposes, we need to truncate the series.

Main ingredient 1: Multilevel Monte Carlo

Definition

- Multilevel Monte Carlo [Giles '08]: truncate at fixed $L \in \mathbb{N}$, introducing a bias that decays to 0 as $L \to \infty$:

$$\mathbb{E}_\mu[Q] \approx \mathbb{E}_\mu[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}_\mu[Q_\ell - Q_{\ell-1}].$$

We estimate the $L+1$ terms independently using Monte Carlo:

$$\widehat{Q}^{\mathrm{MLMC}}_{L,\{N_\ell\}} := \frac{1}{N_0}\sum_{i=1}^{N_0} Q_0(\theta_0^{(i)}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{i=1}^{N_0} Q_\ell(\theta_\ell^{(i)}) - Q_{\ell-1}(\theta_\ell^{(i)}),$$

where $\theta_\ell^{(i)} \stackrel{\text{i.i.d.}}{\sim} \mu$.

- Debiasing Monte Carlo [McLeish '11], [Rhee, Glynn '12]: truncate at random $\tilde{L}$, defining an unbiased estimator with finite expected cost.

Main ingredient 1: Multilevel Monte Carlo

Error and Cost

- In practice, we choose L large enough so that the bias is of the same size as the sampling error.

Main ingredient 1: Multilevel Monte Carlo

Definition

- Multilevel Monte Carlo [Giles '08]: truncate at fixed $L \in \mathbb{N}$, introducing a bias that decays to 0 as $L \to \infty$:

$$\mathbb{E}_\mu[Q] \approx \mathbb{E}_\mu[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}_\mu[Q_\ell - Q_{\ell-1}].$$

We estimate the $L+1$ terms independently using Monte Carlo:

$$\widehat{Q}^{\text{MLMC}}_{L,\{N_\ell\}} := \frac{1}{N_0} \sum_{i=1}^{N_0} Q_0(\theta_0^{(i)}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_0} Q_\ell(\theta_\ell^{(i)}) - Q_{\ell-1}(\theta_\ell^{(i)}),$$

where $\theta_\ell^{(i)} \overset{\text{i.i.d.}}{\sim} \mu$.

- Debiasing Monte Carlo [McLeish '11], [Rhee, Glynn '12]: truncate at random $\tilde{L}$, defining an unbiased estimator with finite expected cost.

Main ingredient 1: Multilevel Monte Carlo

Definition

- Multilevel Monte Carlo [Giles '08]: truncate at fixed $L \in \mathbb{N}$, introducing a bias that decays to 0 as $L \to \infty$:

$$\mathbb{E}_\mu[Q] \approx \mathbb{E}_\mu[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}_\mu[Q_\ell - Q_{\ell-1}].$$

  We estimate the $L+1$ terms independently using Monte Carlo:

$$\widehat{Q}^{\mathrm{MLMC}}_{L,\{N_\ell\}} := \frac{1}{N_0}\sum_{i=1}^{N_0} Q_0(\theta_0^{(i)}) + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_0} Q_\ell(\theta_\ell^{(i)}) - Q_{\ell-1}(\theta_\ell^{(i)}),$$

  where $\theta_\ell^{(i)} \overset{\text{i.i.d.}}{\sim} \mu$.

- Debiasing Monte Carlo [McLeish '11], [Rhee, Glynn '12]: truncate at random $\tilde{L}$, defining an unbiased estimator with finite expected cost.

Main ingredient 1: Multilevel Monte Carlo

Definition

- Multilevel Monte Carlo [Giles '08]: truncate at fixed $L \in \mathbb{N}$, introducing a bias that decays to 0 as $L \to \infty$:

$$\mathbb{E}_\mu[Q] \approx \mathbb{E}_\mu[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}_\mu[Q_\ell - Q_{\ell-1}].$$

We estimate the $L+1$ terms independently using Monte Carlo:

$$\widehat{Q}^{\mathrm{MLMC}}_{L,\{N_\ell\}} := \frac{1}{N_0} \sum_{i=1}^{N_0} Q_0(\theta_0^{(i)}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_0} Q_\ell(\theta_\ell^{(i)}) - Q_{\ell-1}(\theta_\ell^{(i)}),$$

where $\theta_\ell^{(i)} \overset{\text{i.i.d.}}{\sim} \mu$.

- Debiasing Monte Carlo [McLeish '11], [Rhee, Glynn '12]: truncate at random $\tilde{L}$, defining an unbiased estimator with finite expected cost.

Main ingredient 1: Multilevel Monte Carlo

Error and Cost

- In practice, we choose L large enough so that the bias is of the same size as the sampling error.

- Why is the multilevel Monte Carlo estimator efficient?

- The sequence $\{N_\ell\}$ is decreasing, since

$$N_0 \propto \mathbb{V}[Q_0], \qquad N_\ell \propto \mathbb{V}[Q_\ell - Q_{\ell-1}].$$

- This means:
  - we take a large number N_0 of cheap samples of Q_0,
  - we take a small number N_ℓ of expensive samples of Q_ℓ, for $\ell \gg 1$, since $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$.

- In contrast, standard Monte Carlo approaches use a large number of expensive samples of Q_L.

Main ingredient 1: Multilevel Monte Carlo

Definition

- Multilevel Monte Carlo [Giles '08]: truncate at fixed $L \in \mathbb{N}$, introducing a bias that decays to 0 as $L \to \infty$:

$$\mathbb{E}_\mu[Q] \approx \mathbb{E}_\mu[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}_\mu[Q_\ell - Q_{\ell-1}].$$

  We estimate the $L+1$ terms independently using Monte Carlo:

$$\widehat{Q}^{\mathrm{MLMC}}_{L,\{N_\ell\}} := \frac{1}{N_0}\sum_{i=1}^{N_0} Q_0(\theta_0^{(i)}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{i=1}^{N_0} Q_\ell(\theta_\ell^{(i)}) - Q_{\ell-1}(\theta_\ell^{(i)}),$$

  where $\theta_\ell^{(i)} \overset{\text{i.i.d.}}{\sim} \mu$.

- Debiasing Monte Carlo [McLeish '11], [Rhee, Glynn '12]: truncate at random $\tilde{L}$, defining an unbiased estimator with finite expected cost.

Main ingredient 1: Multilevel Monte Carlo

Error and Cost

- In practice, we choose L large enough so that the bias is of the same size as the sampling error.

Main ingredient 1: Multilevel Monte Carlo

Definition

- Multilevel Monte Carlo [Giles '08]: truncate at fixed $L \in \mathbb{N}$, introducing a bias that decays to 0 as $L \to \infty$:

$$\mathbb{E}_\mu[Q] \approx \mathbb{E}_\mu[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}_\mu[Q_\ell - Q_{\ell-1}].$$

  We estimate the $L+1$ terms independently using Monte Carlo:

$$\widehat{Q}^{\text{MLMC}}_{L,\{N_\ell\}} := \frac{1}{N_0} \sum_{i=1}^{N_0} Q_0(\theta_0^{(i)}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_0} Q_\ell(\theta_\ell^{(i)}) - Q_{\ell-1}(\theta_\ell^{(i)}),$$

  where $\theta_\ell^{(i)} \overset{\text{i.i.d.}}{\sim} \mu$.

- Debiasing Monte Carlo [McLeish '11], [Rhee, Glynn '12]: truncate at random $\tilde{L}$, defining an unbiased estimator with finite expected cost.

Main ingredient 1: Multilevel Monte Carlo

Error and Cost

- In practice, we choose L large enough so that the bias is of the same size as the sampling error.
- Why is the multilevel Monte Carlo estimator efficient?
- The sequence $\{N_\ell\}$ is decreasing, since

$$N_0 \propto \mathbb{V}[Q_0], \qquad N_\ell \propto \mathbb{V}[Q_\ell - Q_{\ell-1}].$$

- This means:
  - we take a large number N_0 of cheap samples of Q_0,
  - we take a small number N_ℓ of expensive samples of Q_ℓ, for $\ell \gg 1$, since $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$.
- In contrast, standard Monte Carlo approaches use a large number of expensive samples of Q_L.

Main ingredient 2: Bayesian posterior distributions

Set-up [Stuart, '10]

- We are interested in $\mu = \mu^y$ being the posterior distribution in a Bayesian inverse problem (parameter identification problem):

$$\frac{d\mu^y}{d\mu_0}(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}, \qquad \left(\pi^y(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}} \pi_0(\theta)\right).$$

- This arises from
  - incorporating knowledge on θ in a prior distribution μ_0 (with density π_0),
  - observing data $y = F(\theta) + \eta$, with noise $\eta \sim N(0, \Gamma)$,
  - conditioning μ_0 on y, resulting in the posterior distribution μ^y (with density π^y).

Main ingredient 1: Multilevel Monte Carlo

Error and Cost

- In practice, we choose L large enough so that the bias is of the same size as the sampling error.

- Why is the multilevel Monte Carlo estimator efficient?

- The sequence $\{N_\ell\}$ is decreasing, since

$$N_0 \propto \mathbb{V}[Q_0], \qquad N_\ell \propto \mathbb{V}[Q_\ell - Q_{\ell-1}].$$

- This means:
  - we take a large number N_0 of cheap samples of Q_0,
  - we take a small number N_ℓ of expensive samples of Q_ℓ, for $\ell \gg 1$, since $\mathbb{V}[Q_\ell - Q_{\ell-1}] \to 0$ as $\ell \to \infty$.

- In contrast, standard Monte Carlo approaches use a large number of expensive samples of Q_L.

Main ingredient 2: Bayesian posterior distributions

Set-up [Stuart, '10]

- We are interested in $\mu = \mu^y$ being the posterior distribution in a Bayesian inverse problem (parameter identification problem):

$$\frac{d\mu^y}{d\mu_0}(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}, \qquad \left(\pi^y(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}\ \pi_0(\theta)\right).$$

- This arises from
  - incorporating knowledge on θ in a prior distribution μ_0 (with density π_0),
  - observing data $y = F(\theta) + \eta$, with noise $\eta \sim N(0, \Gamma)$,
  - conditioning μ_0 on y, resulting in the posterior distribution μ^y (with density π^y).

Main ingredient 2: Bayesian posterior distributions

Set-up [Stuart, '10]

- We are interested in $\mu = \mu^y$ being the posterior distribution in a Bayesian inverse problem (parameter identification problem):

$$\frac{d\mu^y}{d\mu_0}(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}. \qquad \left(\pi^y(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}} \pi_0(\theta)\right).$$

- This arises from
  - incorporating knowledge on θ in a prior distribution μ_0 (with density π_0),
  - observing data $y = F(\theta) + \eta$, with noise $\eta \sim N(0, \Gamma)$,
  - conditioning μ_0 on y, resulting in the posterior distribution μ^y (with density π^y).

Main ingredient 2: Bayesian posterior distributions

Set-up [Stuart, '10]

- We are interested in $\mu = \mu^y$ being the posterior distribution in a Bayesian inverse problem (parameter identification problem):

$$\frac{d\mu^y}{d\mu_0}(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}, \qquad \left(\pi^y(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}\, \pi_0(\theta)\right).$$

- This arises from
  - incorporating knowledge on θ in a prior distribution μ_0 (with density π_0),
  - observing data $y = F(\theta) + \eta$, with noise $\eta \sim N(0, \Gamma)$,
  - conditioning μ_0 on y, resulting in the posterior distribution μ^y (with density π^y).

Multilevel Metropolis Hastings algorithm

Computational challenges

What are the main challenges?

- The normalising constant of μ^y is unknown, and i.i.d. sampling is not available.
  - We employ Markov chain Monte Carlo samplers.

- The quantity F can typically not be evaluated exactly, but we have access to a sequence of approximations $\{F_\ell\}_{\ell=0}^\infty$, with $F = F_\infty$.
  - This gives rise to a sequence of approximate posterior measures $\{\mu_\ell^y\}_{\ell=0}^\infty$, so how should we define the multilevel estimator?

Main ingredient 2: Bayesian posterior distributions

Set-up [Stuart, '10]

- We are interested in $\mu = \mu^y$ being the posterior distribution in a Bayesian inverse problem (parameter identification problem):

$$\frac{d\mu^y}{d\mu_0}(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}, \qquad \left(\pi^y(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}\,\pi_0(\theta)\right).$$

- This arises from
  - incorporating knowledge on θ in a prior distribution μ_0 (with density π_0),
  - observing data $y = F(\theta) + \eta$, with noise $\eta \sim N(0, \Gamma)$,
  - conditioning μ_0 on y, resulting in the posterior distribution μ^y (with density π^y).

Multilevel Metropolis Hastings algorithm

Computational challenges

What are the main challenges?

- The normalising constant of μ^y is unknown, and i.i.d. sampling is not available.
  - We employ Markov chain Monte Carlo samplers.

Main ingredient 2: Bayesian posterior distributions

Set-up [Stuart, '10]

- We are interested in $\mu = \mu^y$ being the posterior distribution in a Bayesian inverse problem (parameter identification problem):

$$\frac{d\mu^y}{d\mu_0}(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}, \qquad \left(\pi^y(\theta) \propto e^{-\|y-F(\theta)\|^2_{\Gamma^{-1}}}\, \pi_0(\theta)\right).$$

- This arises from
  - incorporating knowledge on θ in a prior distribution μ_0 (with density π_0),
  - observing data $y = F(\theta) + \eta$, with noise $\eta \sim N(0, \Gamma)$,
  - conditioning μ_0 on y, resulting in the posterior distribution μ^y (with density π^y).

Multilevel Metropolis Hastings algorithm

Computational challenges

What are the main challenges?

- The normalising constant of μ^y is unknown, and i.i.d. sampling is not available.
  - We employ Markov chain Monte Carlo samplers.
- The quantity F can typically not be evaluated exactly, but we have access to a sequence of approximations $\{F_\ell\}_{\ell=0}^\infty$, with $F = F_\infty$.
  - This gives rise to a sequence of approximate posterior measures $\{\mu_\ell^y\}_{\ell=0}^\infty$, so how should we define the multilevel estimator?

Multilevel Metropolis Hastings algorithm

Algorithm [Dodwell, Ketelsen, Scheichl, ALT '15]

We need to generate coupled Markov chains $\{\theta_\ell^{(i)}\}_{i\in\mathbb{N}}$ and $\{\Theta_{\ell-1}^{(i)}\}_{i\in\mathbb{N}}$:

- $\{\Theta_{\ell-1}^{(i)}\}$ has marginal distribution $\mu_{\ell-1}^y$,
- $\{\theta_\ell^{(i)}\}$ has marginal distribution μ_ℓ^y,
- $\mathbb{V}[Q_\ell(\theta_\ell^{(i)}) - Q_{\ell-1}(\Theta_{\ell-1}^{(i)})] \to 0$ as $\ell \to \infty$.

The main idea of our algorithm is to:

- use Metropolis-Hastings with $\Theta'_{\ell-1} \sim q(\cdot|\Theta_{\ell-1}^{(i)})$ to generate $\Theta_{\ell-1}^{(i+1)}$,
- use Metropolis-Hastings with $\theta'_\ell = \Theta_{\ell-1}^{(i+1)}$ to generate $\theta_\ell^{(i+1)}$.

The acceptance probability $\alpha^{2\mathrm{L}}$ for $\theta_\ell^{(i)}$ is easy to compute, and we can prove $\mathbb{E}[\alpha^{2L}] \to 1$ as $\ell \to \infty$.

This means $\mathbb{P}\left[\theta_\ell^{(i)} = \Theta_{\ell-1}^{(i)}\right] \to 1$ as $\ell \to \infty$.

Multilevel Metropolis Hastings algorithm

Implementation details [Dodwell, Ketelsen, Scheichl, ALT '15]

- So far, μ_ℓ^y was defined by approximating F by F_ℓ. We also
  - change the number of parameters on level ℓ: $\theta \in \mathbb{R}^{\mathbb{N}} \to \theta_{1:R_\ell} \in \mathbb{R}^{R_\ell}$,
  - change the noise level on level ℓ: $\Gamma \to \Gamma_\ell$, where $\eta \sim \mathrm{N}(0, \Gamma)$.
- In practice, we always start at level 0 when generating samples, and use a sub-sampling rate t_ℓ.
  - This leads to an efficient implementation with small integrated autocorrelation times ($\mathcal{O}(1)$) on levels $\ell \geq 1$ on level ℓ.
- For more details on efficient implementation using DUNE and MUQ, see talk by Linus Seelinger in MS121 at 3.30pm.

Multilevel Metropolis Hastings algorithm

Example in groundwater flow modelling [Dodwell, Ketelsen, Scheichl, ALT '15]

- Unknown permeability of subsurface: $k(x)$, $x \in (0,1)^2$
- Prior distribution: $\log k(x) \sim \mathrm{GP}\big(0, \exp(-2\|x - x'\|_1)\big)$
- Parametrisation of $k(x)$: Karhunen-Loève expansion

$$\log k(x) = \sum_{j=1}^{\infty} \theta_j \psi_j(x) \approx \sum_{j=1}^{R_\ell} \theta_j \psi_j(x), \quad \theta = \{\theta_j\}_{j=1}^{\infty}, \ \psi_j \in L^\infty((0,1)^2)$$

Under the prior distribution, $\theta_j \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,1)$.

- Resulting pressure field $p(x)$: $-\nabla \cdot (k(x)\nabla p(x)) = 1$ (+ b.c.'s)
- Observed data: $y = \{p(x_i) + \eta_i\}_{i=1}^{16}$, with $\eta_i \sim \mathrm{N}(0, 10^{-4})$

- Quantity of interest: outflow over right boundary

$$\mathbb{E}_{\mu^y}[Q] = \mathbb{E}_{\mu^y}\left[-\int_0^1 k \frac{\partial p}{\partial x_1}\Big|_{x_1=1} \mathrm{d}x_2 \right]$$

Multilevel Metropolis Hastings algorithm

Example in groundwater flow modelling [Dodwell, Ketelsen, Scheichl, ALT '15]