Multilevel Monte Carlo methods for inverse problems constrained by partial differential equations

Aretha Teckentrup

School of Mathematics, University of Edinburgh

Joint work with:

Tim Dodwell (Exeter), Chris Ketelsen (DigitalGlobe),
Rob Scheichl (Heidelberg), Andrew Stuart (Caltech)

43rd Woudschoten Conference - October 5, 2018

THE UNIVERSITY of EDINBURGH
School of Mathematics

Outline

Bayesian Inverse Problems

Definition and Applications

- An inverse problem is concerned with determining causal factors from observed results.
- In mathematical terms, we want to determine model inputs based on data comprised of observable model outputs.

Bayesian Inverse Problems

Definition and Applications

- An inverse problem is concerned with determining causal factors from observed results.
- In mathematical terms, we want to determine model inputs based on data comprised of observable model outputs.
- Inverse problems appear in many application areas, including
  - the determination of the geological structure of the sub-surface from measurements of flow rates at wells, $\Rightarrow -\nabla \cdot (k(x)\nabla p(x)) = g(x)$
  - the detection of flaws or cracks within a concrete structure from acoustic or electromagnetic measurements at its surface.
- The model is frequently given by a partial differential equation, where initial conditions, boundary conditions and/or coefficients are inputs and observables of the solution are the output.

Bayesian Inverse Problems

Mathematical Formulation

- Suppose we are given a mathematical model of a process, represented by a (typically non-linear) map $\mathcal{G}$.

Bayesian Inverse Problems

Mathematical Formulation

- Suppose we are given a mathematical model of a process, represented by a (typically non-linear) map $\mathcal{G}$.
- We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U \subseteq \mathbb{R}^{d_u}$ such that

$$y = \mathcal{G}(u) + \eta,$$

  where η represents observational noise, due to for example measurement error.

  In the context of yesterday's talk, we could have $u = \{\xi_n\}_{n=1}^{d_u}$.

Bayesian Inverse Problems

Mathematical Formulation

- Suppose we are given a mathematical model of a process, represented by a (typically non-linear) map $\mathcal{G}$.
- We are interested in the following inverse problem: given observational data $y \in \mathbb{R}^{d_y}$, determine model parameters $u \in U \subseteq \mathbb{R}^{d_u}$ such that

$$y = \mathcal{G}(u) + \eta,$$

  where η represents observational noise, due to for example measurement error.
  In the context of yesterday's talk, we could have $u = \{\xi_n\}_{n=1}^{d_u}$.
- Simply "inverting $\mathcal{G}$" is not possible, since
  - we do not know the value of η, and
  - typically the problem is ill-posed and/or ill-conditioned.

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10] [Kaipio, Somersalo '04]

- The Bayesian approach treats u, y and η as random variables/fields.

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10] [Kaipio, Somersalo '04]

- The Bayesian approach treats u, y and η as random variables/fields.
- We choose a prior measure μ_0 on u with density π_0.

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10] [Kaipio, Somersalo '04]

- The Bayesian approach treats u, y and η as random variables/fields.
- We choose a prior measure μ_0 on u with density π_0.
- Under the measurement model $y = \mathcal{G}(u) + \eta$ with $\eta \sim N(0, \Gamma)$, we have $y|u \sim N(\mathcal{G}(u), \Gamma)$, and the likelihood of the data y is

$$\mathcal{P}(y|u) \eqsim \exp\Big(- \frac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Gamma^{-1}} \Big) =: \exp\big(- \Phi(u)\big).$$

Bayesian Inverse Problems

Mathematical Formulation [Stuart '10] [Kaipio, Somersalo '04]

- The Bayesian approach treats u, y and η as random variables/fields.
- We choose a prior measure μ_0 on u with density π_0.
- Under the measurement model $y = \mathcal{G}(u) + \eta$ with $\eta \sim N(0, \Gamma)$, we have $y|u \sim N(\mathcal{G}(u), \Gamma)$, and the likelihood of the data y is

$$\mathcal{P}(y|u) \eqsim \exp\Big(-\frac{1}{2}\|y - \mathcal{G}(u)\|^2_{\Gamma^{-1}}\Big) =: \exp\big(-\Phi(u)\big).$$

- Using Bayes' Theorem, we obtain the posterior measure μ^y of $u|y$ with density π^y, given by

$$\pi^y(u) = \frac{1}{Z}\exp\big(-\Phi(u)\big)\pi_0(u), \qquad \Big(\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\exp\big(-\Phi(u)\big)\Big)$$

where $Z = \mathbb{E}_{\pi_0}[\exp\big(-\Phi\big)]$.

Bayesian Inverse Problems

Computational Challenges

- For the remainder of this talk, suppose are interested in computing $\mathbb{E}_{\pi^y}[\phi]$, for some quantity of interest $\phi(u)$.

Bayesian Inverse Problems

Computational Challenges

- For the remainder of this talk, suppose are interested in computing $\mathbb{E}_{\pi^y}[\phi]$, for some quantity of interest $\phi(u)$.
- In most cases, we do not have a closed form expression for the posterior density π^y, since the normalising constant Z is not known explicitly.
  (Exception: parameter-to-observation map $\mathcal{G}$ linear and prior π_0 Gaussian $\Rightarrow$ posterior π^y also Gaussian.)
- We will present two efficient approaches to computing $\mathbb{E}_{\pi^y}[\phi]$:
  - (multilevel) ratio estimators,
  - (multilevel) Metropolis Hastings estimators.

(Multilevel) Ratio Estimators

Computing Expectations

Suppose are interested in computing $\mathbb{E}_{\pi^y}[\phi]$. Using Bayes' Theorem, we can write this as

$$\mathbb{E}_{\pi^y}[\phi] = \mathbb{E}_{\pi_0}[\frac{1}{Z}\exp[-\Phi]\,\phi] = \frac{\mathbb{E}_{\pi_0}[\phi\exp[-\Phi]]}{\mathbb{E}_{\pi_0}[\exp[-\Phi]]}.$$

We have rewritten the posterior expectation as a **ratio of two prior expectations**.

(Multilevel) Ratio Estimators

Computing Expectations

Suppose are interested in computing $\mathbb{E}_{\pi^y}[\phi]$. Using Bayes' Theorem, we can write this as

$$\mathbb{E}_{\pi^y}[\phi] = \mathbb{E}_{\pi_0}[\frac{1}{Z}\exp[-\Phi]\,\phi] = \frac{\mathbb{E}_{\pi_0}[\phi\exp[-\Phi]]}{\mathbb{E}_{\pi_0}[\exp[-\Phi]]}.$$

We have rewritten the posterior expectation as a **ratio of two prior expectations**. We can now approximate

$$\mathbb{E}_{\pi^y}[\phi] = \frac{Q}{Z} =\approx \frac{\widehat{Q}}{\widehat{Z}},$$

where $\widehat{Q}$ is an estimator of $Q := \mathbb{E}_{\pi_0}[\phi\exp(-\Phi)]$ and $\widehat{Z}$ is an estimator of $Z = \mathbb{E}_{\pi_0}[\exp\big(-\Phi\big)]$.

The prior distribution is known in closed form, and furthermore often has a simple structure (e.g. i.i.d. Gaussian or uniform).

(Multilevel) Ratio Estimators

Monte Carlo and Multilevel Monte Carlo Estimators

The standard Monte Carlo (MC) estimator of $Z = \mathbb{E}_{\pi_0}[\exp(-\Phi)]$,

$$\widehat{Z}_{h,N}^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} \exp(-\Phi_h(u^{(i)})),$$

is an equal weighted average of N i.i.d samples $\exp(-\Phi_h(u^{(i)}))$, where Φ_h denotes an approximation of Φ using mesh width h.

(Multilevel) Ratio Estimators

Monte Carlo and Multilevel Monte Carlo Estimators

The standard Monte Carlo (MC) estimator of $Z = \mathbb{E}_{\pi_0}[\exp(-\Phi)]$,

$$\widehat{Z}^{\mathrm{MC}}_{h,N} = \frac{1}{N}\sum_{i=1}^{N}\exp(-\Phi_h(u^{(i)})),$$

is an equal weighted average of N i.i.d samples $\exp(-\Phi_h(u^{(i)}))$, where Φ_h denotes an approximation of Φ using mesh width h.

The multilevel Monte Carlo (MLMC) estimator

$$\widehat{Z}^{\mathrm{ML}}_{\{M_\ell,N_\ell\}} = \frac{1}{N_0}\sum_{i=1}^{N_0} \mathrm{e}^{-\Phi_{h_0}(u^{(i,0)})} + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\mathrm{e}^{-\Phi_{h_\ell}(u^{(i,\ell)})} - \mathrm{e}^{-\Phi_{h_{\ell-1}}(u^{(i,\ell)})},$$

is a sum of $L+1$ independent MC estimators. The sequence $\{N_\ell\}$ is decreasing, which means a significant portion of the computational effort is shifted onto the coarse grids.

(Multilevel) Ratio Estimators

Convergence of Mean Square Error [Stuart, Scheichl, T. '17]

- We want to bound the mean square error (MSE): $\mathbb{E}\Big[\Big(\frac{Q}{Z} - \frac{\widehat{Q}}{\widehat{Z}}\Big)^2\Big]$.

- Rearranging the MSE and applying the triangle inequality, we have

$$\mathbb{E}\Big[\Big(\frac{Q}{Z} - \frac{\widehat{Q}}{\widehat{Z}}\Big)^2\Big] \leq \frac{2}{Z^2}\Big(\mathbb{E}\big[(Q - \widehat{Q})^2\big] + \mathbb{E}\big[(\widehat{Q}/\widehat{Z})^2(Z - \widehat{Z})^2\big]\Big).$$

(Multilevel) Ratio Estimators

Convergence of Mean Square Error [Stuart, Scheichl, T. '17]

- We want to bound the mean square error (MSE): $\mathbb{E}\Big[\Big(\frac{Q}{Z} - \frac{\widehat{Q}}{\widehat{Z}}\Big)^2\Big]$.

- Rearranging the MSE and applying the triangle inequality, we have

$$\mathbb{E}\Big[\Big(\frac{Q}{Z} - \frac{\widehat{Q}}{\widehat{Z}}\Big)^2\Big] \leq \frac{2}{Z^2}\Big(\mathbb{E}\big[(Q - \widehat{Q})^2\big] + \mathbb{E}\big[(\widehat{Q}/\widehat{Z})^2(Z - \widehat{Z})^2\big]\Big).$$

- If $(\widehat{Q}/\widehat{Z})^2 \in L^\infty$, then the MSE of $\widehat{Q}/\widehat{Z}$ can be bounded in terms of the individual MSEs of $\widehat{Q}$ and $\widehat{Z}$.

- If $(\widehat{Q}/\widehat{Z})^2 \in L^r$, for some $2 < r < \infty$, then the MSE of $\widehat{Q}/\widehat{Z}$ can be bounded in terms of the MSE of $\widehat{Q}$ and a higher order error of $\widehat{Z}$, $\mathbb{E}\big[(Z - \widehat{Z})^p\big]$ for some $p > 2$.

(Multilevel) Ratio Estimators

Convergence of Mean Square Error [Stuart, Scheichl, T. '17]

- Consider the diffusion problem

$$-\nabla \cdot (k(x)\nabla p(x)) = g(x), \text{ in } D \subset \mathbb{R}^d, \qquad \mathcal{G}(u) = \mathcal{O}(p),$$

with prior distribution

- a log-normal distribution $k(x) = \exp\left(\sum_{n=1}^{d_u} u_n \phi_n(x)\right)$, with $\{\phi_n\}_{n=1}^{d_u}$ given functions in $L^\infty(D)$ and $u_n \sim N(0,1) \Rightarrow \widehat{Q}/\widehat{Z} \in L^r, r < \infty$
- a uniform distribution, i.e. $k(x) = m(x) + \sum_{n=1}^{d_u} u_n \phi_n(x)$, with $\{\phi_n\}_{n=1}^{d_u}$ given functions in $L^\infty(D)$ and $u_n \sim U[-1,1] \Rightarrow \widehat{Q}/\widehat{Z} \in L^\infty$

(Multilevel) Ratio Estimators

Convergence of Mean Square Error [Stuart, Scheichl, T. '17]

- Consider the diffusion problem

$$-\nabla \cdot (k(x)\nabla p(x)) = g(x), \text{ in } D \subset \mathbb{R}^d, \qquad \mathcal{G}(u) = \mathcal{O}(p),$$

  with prior distribution

  - a log-normal distribution $k(x) = \exp\left(\sum_{n=1}^{d_u} u_n \phi_n(x)\right)$, with $\{\phi_n\}_{n=1}^{d_u}$ given functions in $L^\infty(D)$ and $u_n \sim N(0,1) \Rightarrow \widehat{Q}/\widehat{Z} \in L^r, r < \infty$
  - a uniform distribution, i.e. $k(x) = m(x) + \sum_{n=1}^{d_u} u_n \phi_n(x)$, with $\{\phi_n\}_{n=1}^{d_u}$ given functions in $L^\infty(D)$ and $u_n \sim U[-1,1] \Rightarrow \widehat{Q}/\widehat{Z} \in L^\infty$

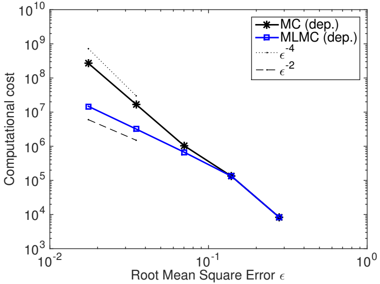
- For the multilevel Monte Carlo estimator $\widehat{Z}^{\text{ML}}_{\{M_\ell, N_\ell\}}$ to be positive, we need to choose h_0 small enough.

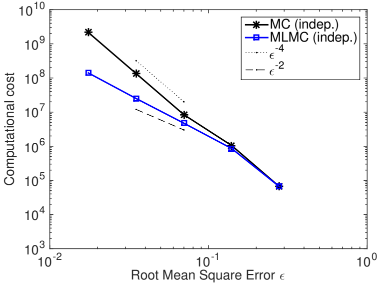
- The samples in $\widehat{Q}$ and $\widehat{Z}$ can be the same or independent.

(Multilevel) Ratio Estimators

Numerical Results [Stuart, Scheichl, T. '17]

- Flow cell model problem on $(0,1)^2$
- u log-normal random field
- Observed data: $y = \{p(x_i) + \eta_i\}_{i=1}^{9}$, with $\eta_i \sim \mathrm{N}(0, 0.09)$
- QoI ϕ is outflow over right boundary





(Multilevel) Ratio Estimators

Related works

A number of works have recently considered the ratio estimator approach in the context of PDE constrained inverse problems, as it allows to reuse machinery developed for π_0:

- [Schillings, Schwab '13]: dimension-adaptive sparse grids
- [Stuart, Scheichl, T. '17]: Quasi-Monte Carlo
- [Dick, Gantner, Le Gia, Schwab '17]: (multilevel) higher order Quasi-Monte Carlo
- [Gantner, Peters '18]: higher order Quasi-Monte Carlo for PDEs on random domains

(Multilevel) Ratio Estimators

Related works

A number of works have recently considered the ratio estimator approach in the context of PDE constrained inverse problems, as it allows to reuse machinery developed for π_0:

- [Schillings, Schwab '13]: dimension-adaptive sparse grids
- [Stuart, Scheichl, T. '17]: Quasi-Monte Carlo
- [Dick, Gantner, Le Gia, Schwab '17]: (multilevel) higher order Quasi-Monte Carlo
- [Gantner, Peters '18]: higher order Quasi-Monte Carlo for PDEs on random domains

When $\|\Gamma\| \ll 1$ or $d_y \gg 1$, the posterior density π^y may concentrate, and the prior evaluations become difficult to evaluate accurately.

- [Schillings, Schwab '16]: rescaling of parameter space around (unique) MAP point

(Multilevel) Metropolis Hastings Estimators

Introduction

- The second approach we are going to discuss is the (multilevel) Metropolis-Hastings algorithm, a type of Markov chain Monte Carlo (MCMC) algorithm.

(Multilevel) Metropolis Hastings Estimators

Introduction

- The second approach we are going to discuss is the (multilevel) Metropolis-Hastings algorithm, a type of Markov chain Monte Carlo (MCMC) algorithm.
- The estimator of $\mathbb{E}_{\pi^y}[\phi]$ takes the form

$$\widehat{E}^{\mathrm{MH}}_{h,N} = \frac{1}{N} \sum_{i=1}^{N} \phi_h(u^{(i)}),$$

where $u^{(i)} \sim \pi_h^y$, for $1 \le i \le N$, and $\pi_h^y(u) = \frac{1}{Z_h} \exp(-\Phi_h(u))\pi_0(u)$.

- Since π_h^y is not known in closed form, it is generally not possible to generate i.i.d. samples distributed according to π_h^y.

(Multilevel) Metropolis Hastings Estimators

Standard Metropolis-Hastings [Robert, Casella '99]

ALGORITHM 1. (Standard MH-MCMC)

- Choose $u^{(1)}$ with $\pi_h^y(u^{(1)}) > 0$.
- At state $u^{(i)}$, sample a proposal u' from density $q(u' \,|\, u^{(i)})$.
- Accept sample u' with probability

$$\alpha(u' \,|\, u^{(i)}) = \min\left(1, \frac{\pi_h^y(u')\, q(u^{(i)} \,|\, u')}{\pi_h^y(u^{(i)})\, q(u' \,|\, u^{(i)})}\right),$$

i.e. $u^{(i+1)} = u'$ with probability $\alpha(u' \,|\, u^{(i)})$; otherwise stay at $u^{(i+1)} = u^{(i)}$.

(Multilevel) Metropolis Hastings Estimators

Standard Metropolis-Hastings [Robert, Casella '99]

ALGORITHM 1. (Standard MH-MCMC)

- Choose $u^{(1)}$ with $\pi_h^y(u^{(1)}) > 0$.
- At state $u^{(i)}$, sample a proposal u' from density $q(u' \,|\, u^{(i)})$.
- Accept sample u' with probability

$$\alpha(u' \,|\, u^{(i)}) = \min\left(1, \frac{\pi_h^y(u')\, q(u^{(i)} \,|\, u')}{\pi_h^y(u^{(i)})\, q(u' \,|\, u^{(i)})}\right),$$

i.e. $u^{(i+1)} = u'$ with probability $\alpha(u' \,|\, u^{(i)})$; otherwise stay at $u^{(i+1)} = u^{(i)}$.

- The proposal density q is chosen to be easy to sample from.

(Multilevel) Metropolis Hastings Estimators

Standard Metropolis-Hastings [Robert, Casella '99]

ALGORITHM 1. (Standard MH-MCMC)

- Choose $u^{(1)}$ with $\pi_h^y(u^{(1)}) > 0$.
- At state $u^{(i)}$, sample a proposal u' from density $q(u' \,|\, u^{(i)})$.
- Accept sample u' with probability

$$\alpha(u' \,|\, u^{(i)}) = \min\left(1, \frac{\pi_h^y(u')\, q(u^{(i)} \,|\, u')}{\pi_h^y(u^{(i)})\, q(u' \,|\, u^{(i)})}\right),$$

i.e. $u^{(i+1)} = u'$ with probability $\alpha(u' \,|\, u^{(i)})$; otherwise stay at $u^{(i+1)} = u^{(i)}$.

- The proposal density q is chosen to be easy to sample from.
- The accept/reject step is added in order to obtain samples from π_h^y.

(Multilevel) Metropolis Hastings Estimators

Standard Metropolis-Hastings [Robert, Casella '99]

ALGORITHM 1. (Standard MH-MCMC)

- Choose $u^{(1)}$ with $\pi_h^y(u^{(1)}) > 0$.
- At state $u^{(i)}$, sample a proposal u' from density $q(u' \,|\, u^{(i)})$.
- Accept sample u' with probability

$$\alpha(u' \,|\, u^{(i)}) = \min\left(1, \frac{\pi_h^y(u')\, q(u^{(i)} \,|\, u')}{\pi_h^y(u^{(i)})\, q(u' \,|\, u^{(i)})}\right),$$

i.e. $u^{(i+1)} = u'$ with probability $\alpha(u' \,|\, u^{(i)})$; otherwise stay at $u^{(i+1)} = u^{(i)}$.

- The proposal density q is chosen to be easy to sample from.
- The accept/reject step is added in order to obtain samples from π_h^y.
- Knowledge of the normalising constant Z_h of π_h^y is not required.

(Multilevel) Metropolis Hastings Estimators

Convergence of Standard Metropolis-Hastings

The mean square error of $\widehat{E}^{\mathrm{MH}}_{h,N}$ satisfies

$$
\begin{aligned}
e(\widehat{E}^{\mathrm{MH}}_{h,N})^2 &= \mathbb{E}[(\widehat{E}^{\mathrm{MH}}_{h,N} - \mathbb{E}_{\pi^y}[\phi])^2] \\
&\leq \underbrace{\mathbb{V}[\widehat{E}^{\mathrm{MH}}_{h,N}] + 2\big(\mathbb{E}\big[\widehat{E}^{\mathrm{MH}}_{h,N}\big] - \mathbb{E}_{\pi^y_h}[\widehat{E}^{\mathrm{MH}}_{h,N}]\big)^2}_{\text{sampling error}} + \underbrace{2\big(\mathbb{E}_{\pi^y_h}\big[\phi_h\big] - \mathbb{E}_{\pi^y}[\phi]\big)^2}_{\text{numerical error}}.
\end{aligned}
$$

(Multilevel) Metropolis Hastings Estimators

Convergence of Standard Metropolis-Hastings

The mean square error of $\widehat{E}^{\mathrm{MH}}_{h,N}$ satisfies

$$e(\widehat{E}^{\mathrm{MH}}_{h,N})^2 = \mathbb{E}[(\widehat{E}^{\mathrm{MH}}_{h,N} - \mathbb{E}_{\pi^y}[\phi])^2]$$

$$\leq \underbrace{\mathbb{V}[\widehat{E}^{\mathrm{MH}}_{h,N}] + 2\Big(\mathbb{E}\big[\widehat{E}^{\mathrm{MH}}_{h,N}\big] - \mathbb{E}_{\pi^y_h}[\widehat{E}^{\mathrm{MH}}_{h,N}]\Big)^2}_{\text{sampling error}} + \underbrace{2\Big(\mathbb{E}_{\pi^y_h}\big[\phi_h\big] - \mathbb{E}_{\pi^y}[\phi]\Big)^2}_{\text{numerical error}}.$$

Theorem (Non-asymptotic bound on sampling error)

Under certain regularity assumptions on ϕ given in [Rudolf, '11], *we have*

$$\mathbb{V}[\widehat{E}^{\mathrm{MH}}_{h,N}] + 2\Big(\mathbb{E}\big[\widehat{E}^{\mathrm{MH}}_{h,N}\big] - \mathbb{E}_{\pi^y_h}[\widehat{E}^{\mathrm{MH}}_{h,N}]\Big)^2 \ \leq \ C\ \frac{\mathbb{E}[\phi_h^2]}{N},$$

for a constant C independent of N and h.

(Multilevel) Metropolis Hastings Estimators

Convergence of Standard Metropolis-Hastings

The mean square error of $\widehat{E}^{\text{MH}}_{h,N}$ satisfies

$$e(\widehat{E}^{\text{MH}}_{h,N})^2 = \mathbb{E}[(\widehat{E}^{\text{MH}}_{h,N} - \mathbb{E}_{\pi^y}[\phi])^2]$$

$$\leq \underbrace{\mathbb{V}[\widehat{E}^{\text{MH}}_{h,N}] + 2\left(\mathbb{E}\left[\widehat{E}^{\text{MH}}_{h,N}\right] - \mathbb{E}_{\pi^y_h}[\widehat{E}^{\text{MH}}_{h,N}]\right)^2}_{\text{sampling error}} + \underbrace{2\left(\mathbb{E}_{\pi^y_h}\left[\phi_h\right] - \mathbb{E}_{\pi^y}[\phi]\right)^2}_{\text{numerical error}}.$$

Theorem (Non-asymptotic bound on sampling error)

Under certain regularity assumptions on ϕ given in [Rudolf, '11], *we have*

$$\mathbb{V}[\widehat{E}^{\text{MH}}_{h,N}] + 2\left(\mathbb{E}\left[\widehat{E}^{\text{MH}}_{h,N}\right] - \mathbb{E}_{\pi^y_h}[\widehat{E}^{\text{MH}}_{h,N}]\right)^2 \;\leq\; C\,\frac{\mathbb{E}[\phi_h^2]}{N},$$

for a constant C independent of N and h.

$\Rightarrow$ need to choose N large and h small!

(Multilevel) Metropolis Hastings Estimators

Definition of Multilevel Metropolis-Hastings [Dodwell, Ketelsen, Scheichl, T. '15]

Key ingredients in multilevel method:

- **telescoping sum:** $\mathbb{E}\left[\phi_{h_L}\right] = \mathbb{E}\left[\phi_{h_0}\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\phi_{h_\ell}\right] - \mathbb{E}\left[\phi_{h_{\ell-1}}\right]$,
- number of samples per level N_ℓ is **decreasing**.

(Multilevel) Metropolis Hastings Estimators

Definition of Multilevel Metropolis-Hastings [Dodwell, Ketelsen, Scheichl, T. '15]

Key ingredients in multilevel method:

- **telescoping sum:** $\mathbb{E}\left[\phi_{h_L}\right] = \mathbb{E}\left[\phi_{h_0}\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\phi_{h_\ell}\right] - \mathbb{E}\left[\phi_{h_{\ell-1}}\right]$,
- number of samples per level N_ℓ is **decreasing**.

The target distribution $\pi^y_{h_\ell}$ depends on h_ℓ, so need to define multilevel estimator carefully!

$$\mathbb{E}_{\pi^y_{h_L}}\left[\phi_{h_L}\right] = \mathbb{E}_{\pi^y_{h_0}}\left[\phi_{h_0}\right] + \sum_{\ell=1}^{L} \mathbb{E}_{\pi^y_{h_\ell}}\left[\phi_{h_\ell}\right] - \mathbb{E}_{\pi^y_{h_{\ell-1}}}\left[\phi_{h_{\ell-1}}\right]$$

(Multilevel) Metropolis Hastings Estimators

Definition of Multilevel Metropolis-Hastings [Dodwell, Ketelsen, Scheichl, T. '15]

Key ingredients in multilevel method:

- **telescoping sum:** $\mathbb{E}\left[\phi_{h_L}\right] = \mathbb{E}\left[\phi_{h_0}\right] + \sum_{\ell=1}^{L} \mathbb{E}\left[\phi_{h_\ell}\right] - \mathbb{E}\left[\phi_{h_{\ell-1}}\right]$,
- number of samples per level N_ℓ is **decreasing**.

The target distribution $\pi^y_{h_\ell}$ depends on h_ℓ, so need to define multilevel estimator carefully!

$$\mathbb{E}_{\pi^y_{h_L}}\left[\phi_{h_L}\right] = \underbrace{\mathbb{E}_{\pi^y_{h_0}}\left[\phi_{h_0}\right]}_{\text{standard MCMC}} + \sum_{\ell=1}^{L} \underbrace{\mathbb{E}_{\pi^y_{h_\ell}}\left[\phi_{h_\ell}\right] - \mathbb{E}_{\pi^y_{h_{\ell-1}}}\left[\phi_{h_{\ell-1}}\right]}_{\text{2 level MCMC (NEW)}}$$

$$\widehat{E}^{\mathrm{MLMH}}_{\{h_\ell, N_\ell\}} := \frac{1}{N_0} \sum_{i=1}^{N_0} \phi_{h_0}(u_0^{(i)}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \Big(\phi_{h_\ell}(u_\ell^{(i)}) - \phi_{h_{\ell-1}}(U_{\ell-1}^{(i)}) \Big)$$

(Multilevel) Metropolis Hastings Estimators

Definition of Multilevel Metropolis-Hastings [Dodwell, Ketelsen, Scheichl, T. '15]

Key ingredients in multilevel method:

- **telescoping sum:** $\mathbb{E}[\phi_{h_L}] = \mathbb{E}[\phi_{h_0}] + \sum_{\ell=1}^{L} \mathbb{E}[\phi_{h_\ell}] - \mathbb{E}\left[\phi_{h_{\ell-1}}\right]$,
- number of samples per level N_ℓ is **decreasing**.

The target distribution $\pi^y_{h_\ell}$ depends on h_ℓ, so need to define multilevel estimator carefully!

$$\mathbb{E}_{\pi^y_{h_L}}[\phi_{h_L}] = \underbrace{\mathbb{E}_{\pi^y_{h_0}}[\phi_{h_0}]}_{\text{standard MCMC}} + \sum_{\ell=1}^{L} \underbrace{\mathbb{E}_{\pi^y_{h_\ell}}[\phi_{h_\ell}] - \mathbb{E}_{\pi^y_{h_{\ell-1}}}\left[\phi_{h_{\ell-1}}\right]}_{\text{2 level MCMC (NEW)}}$$

$$\widehat{E}^{\mathrm{MLMH}}_{\{h_\ell, N_\ell\}} := \frac{1}{N_0} \sum_{i=1}^{N_0} \phi_{h_0}(u_0^{(i)}) + \sum_{\ell=1}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(\phi_{h_\ell}(u_\ell^{(i)}) - \phi_{h_{\ell-1}}(U_{\ell-1}^{(i)}) \right)$$

We need to be able to choose $N_\ell \to 1$ as $\ell \to \infty$!

(Multilevel) Metropolis Hastings Estimators

Definition of Multilevel Metropolis-Hastings [Dodwell, Ketelsen, Scheichl, T. '15]

We need to generate (coupled) Markov chains $\{u_\ell^{(i)}\}$ and $\{U_{\ell-1}^{(i)}\}$, such that

- $\{U_{\ell-1}^{(i)}\}$ has marginal distribution $\pi^y_{h_{\ell-1}}$,
- $\{u_\ell^{(i)}\}$ has marginal distribution $\pi^y_{h_\ell}$,
- $\mathbb{V}[\phi_{h_\ell}(u_\ell^{(i)}) - \phi_{h_{\ell-1}}(U_{\ell-1}^{(i)})] \to 0$ as $\ell \to \infty$.

(Multilevel) Metropolis Hastings Estimators

Definition of Multilevel Metropolis-Hastings [Dodwell, Ketelsen, Scheichl, T. '15]

We need to generate (coupled) Markov chains $\{u_\ell^{(i)}\}$ and $\{U_{\ell-1}^{(i)}\}$, such that

- $\{U_{\ell-1}^{(i)}\}$ has marginal distribution $\pi^y_{h_{\ell-1}}$,
- $\{u_\ell^{(i)}\}$ has marginal distribution $\pi^y_{h_\ell}$,
- $\mathbb{V}[\phi_{h_\ell}(u_\ell^{(i)}) - \phi_{h_{\ell-1}}(U_{\ell-1}^{(i)})] \to 0$ as $\ell \to \infty$.

The main idea of our algorithm is to:

- use Metropolis-Hastings with $U'_{\ell-1} \sim q(\cdot|U_{\ell-1}^{(i)})$ to generate $U_{\ell-1}^{(i+1)}$,
- use Metropolis-Hastings with $u'_\ell = U_{\ell-1}^{(i+1)}$ to generate $u_\ell^{(i+1)}$.

(Multilevel) Metropolis Hastings Estimators

Definition of Multilevel Metropolis-Hastings [Dodwell, Ketelsen, Scheichl, T. '15]

We need to generate (coupled) Markov chains $\{u_\ell^{(i)}\}$ and $\{U_{\ell-1}^{(i)}\}$, such that

- $\{U_{\ell-1}^{(i)}\}$ has marginal distribution $\pi_{h_{\ell-1}}^y$,
- $\{u_\ell^{(i)}\}$ has marginal distribution $\pi_{h_\ell}^y$,
- $\mathbb{V}[\phi_{h_\ell}(u_\ell^{(i)}) - \phi_{h_{\ell-1}}(U_{\ell-1}^{(i)})] \to 0$ as $\ell \to \infty$.

The main idea of our algorithm is to:

- use Metropolis-Hastings with $U'_{\ell-1} \sim q(\cdot|U_{\ell-1}^{(i)})$ to generate $U_{\ell-1}^{(i+1)}$,
- use Metropolis-Hastings with $u'_\ell = U_{\ell-1}^{(i+1)}$ to generate $u_\ell^{(i+1)}$.

In practice, we use t_ℓ steps of Metropolis-Hastings to generate $U_{\ell-1}^{(i+1)}$, such that $U_{\ell-1}^{(i)}$ and $U_{\ell-1}^{(i+1)}$ are essentially uncorrelated. This significantly simplifies the computation of the acceptance probability for $u_\ell^{(i+1)}$.

(Multilevel) Metropolis Hastings Estimators

Implementation details [Dodwell, Ketelsen, Scheichl, T. '15]

- The acceptance probability $\alpha^{2\mathrm{L}}$ for $u_\ell^{(i)}$ is easy to compute, and we can prove $\mathbb{E}[\alpha^{2L}] \to 1$ as $\ell \to \infty$. This means $\mathbb{P}\left[u_\ell^{(i)} = U_{\ell-1}^{(i)}\right] \to 1$ as $\ell \to \infty$.

(Multilevel) Metropolis Hastings Estimators

Implementation details [Dodwell, Ketelsen, Scheichl, T. '15]

- The acceptance probability $\alpha^{2\mathrm{L}}$ for $u_\ell^{(i)}$ is easy to compute, and we can prove $\mathbb{E}[\alpha^{2L}] \to 1$ as $\ell \to \infty$. This means $\mathbb{P}\left[u_\ell^{(i)} = U_{\ell-1}^{(i)}\right] \to 1$ as $\ell \to \infty$.

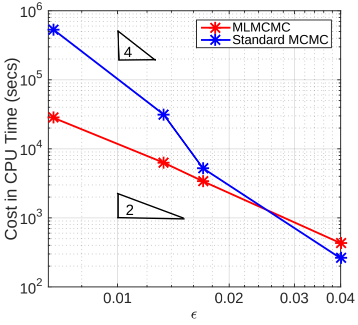
- To define the level ℓ approximation $\pi_{h_\ell}^y$, we can
  - change the number of parameters on level ℓ: $u \in \mathbb{R}^{\mathbb{N}} \to u_{1:R_\ell} \in \mathbb{R}^{R_\ell}$,
  - change the noise level on level ℓ: $\Gamma \to \Gamma_\ell$, where $\eta \sim \mathrm{N}(0, \Gamma)$.

- 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.

(Multilevel) Metropolis Hastings Estimators

Numerical Example [Dodwell, Ketelsen, Scheichl, T. '15]

- Flow cell model problem on $(0,1)^2$
- u log-normal random field
- Observed data: $y = \{p(x_i) + \eta_i\}_{i=1}^{16}$, with $\eta_i \sim \mathrm{N}(0, 10^{-4})$
- QoI ϕ is outflow over right boundary



(Multilevel) Metropolis Hastings Estimators

Related works

- Earlier related work using coarse models to "pre-screen" proposals for fine models:
  - [Liu, '01] : surrogate transition method
  - [Christen, Fox '05] : delayed acceptance
  - [Efendiev, Hou, Lou '06] : application to PDEs
- There are several other works discussing multilevel methods in the context of inverse problems, see Raul Tempone's second talk and the MLMC community webpage.

Conclusions

- Standard sampling methods quickly become computationally infeasible for practical applications.
- Multilevel methods are powerful tools for reducing the computational burden in practical applications, leading to a cost reduction of orders of magnitude.
- In the context of Bayesian inverse problems, multilevel methods include
  - multilevel ratio estimators (importance sampling),
  - multilevel Metropolis-Hastings estimators.
- Multilevel methods are generally applicable!

References I

Mulilevel Community Webpage.
https://people.maths.ox.ac.uk/gilesm/mlmc_community.html.

J. A. Christen and C. Fox, *Markov chain Monte Carlo using an approximation*, Journal of Computational and Graphical statistics, 14 (2005), pp. 795–810.

J. Dick, R. N. Gantner, Q. T. Le Gia, and C. Schwab, *Multilevel higher-order quasi-Monte Carlo Bayesian estimation*, Mathematical Models and Methods in Applied Sciences, 27 (2017), pp. 953–995.

T. Dodwell, C. Ketelsen, R. Scheichl, and A. Teckentrup, *A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2017), pp. 1075–1108.

Y. Efendiev, T. Hou, and W. Luo, *Preconditioning Markov chain Monte Carlo simulations using coarse-scale models*, SIAM Journal on Scientific Computing, 28 (2006), pp. 776–803.

References II

R. N. Gantner and M. D. Peters, *Higher order quasi-Monte Carlo for Bayesian shape inversion*, Tech. Rep. 2016-42, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2016.

J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, Springer, 2004.

J. S. Liu, *Monte Carlo strategies in scientific computing*, Springer, 2001.

C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 1999.

D. Rudolf, *Explicit error bounds for Markov chain Monte Carlo*, arXiv preprint arXiv:1108.3201, (2011).

R. Scheichl, A. Stuart, and A. Teckentrup, *Quasi-Monte Carlo and Multilevel Monte Carlo Methods for Computing Posterior Expectations in Elliptic Inverse Problems*, SIAM/ASA Journal on Uncertainty Quantification, 5 (2017), pp. 493–518.

C. Schillings and C. Schwab, *Sparse, adaptive Smolyak quadratures for Bayesian inverse problems*, Inverse Problems, 29 (2013), p. 065011.

References III

——, *Scaling limits in computational Bayesian inversion*, ESAIM: Mathematical Modelling and Numerical Analysis, 50 (2016), pp. 1825–1856.

A. M. Stuart, *Inverse Problems: A Bayesian Perspective*, Acta Numerica, 19 (2010), pp. 451–559.