Multi-level Markov Chain Monte Carlo Methods for Bayesian Inverse Problems

Juan Pablo Madrigal Cianci

May 4th, 2022

ESI WS1: Multilevel and multifidelity sampling methods in UQ for PDEs

Chair of Scientific Computing and Uncertainty Quantification (CSQI), EPF Lausanne
Collaborators: Prof. Fabio Nobile, Prof. Raúl Tempone.

Introduction: Bayesian Inverse Problems

Given measured data $y \in Y$ modeled as

$$y = F(u) + e, \quad e \sim \mu_{\text{noise}}, \quad u \in X,$$

- u: Parameter $u \sim \mu_{\text{pr}}$.
- $F : X \to Y$.
- e: Random noise.

Introduction: Bayesian Inverse Problems

Given measured data $y \in \mathsf{Y}$ modeled as

$$y = F(u) + e, \quad e \sim \mu_{\text{noise}}, \quad u \in \mathsf{X},$$

- u: Parameter $u \sim \mu_{\text{pr}}$.
- $F : \mathsf{X} \to \mathsf{Y}$.
- e: Random noise.

Goal: sample u **conditioned on** y from μ^y, with

$$\underbrace{\mu^y(u)}_{\text{Posterior}} := \mu(u|y) = \frac{1}{Z} \underbrace{\mu_{\text{noise}}(y - F(u))}_{\text{likelihood}} \underbrace{\mu_{\text{pr}}(u)}_{\text{prior}}$$

Given measured data $y \in \mathrm{Y}$ modeled as

$$y = F(u) + e, \quad e \sim \mu_{\text{noise}}, \quad u \in \mathrm{X},$$

- u: Parameter $u \sim \mu_{\text{pr}}$.
- $F : \mathrm{X} \to \mathrm{Y}$.
- e: Random noise.

Goal: sample u **conditioned on** y from μ^y, with

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_{\text{pr}}}(u) := \frac{1}{Z} e^{-\Phi(u;y)}, \quad \Phi(u;y) = \|y - F(u)\|_{\mathrm{Y}}^2,$$

$$Z = \int_{\mathrm{X}} e^{-\Phi(u;y)} \mu_{\text{pr}}(\mathrm{d}u),$$

Introduction: Bayesian Inverse Problems

Given measured data $y \in \mathsf{Y}$ modeled as

$$y = F(u) + e, \quad e \sim \mu_{\text{noise}}, \quad u \in \mathsf{X},$$

- u: Parameter $u \sim \mu_{\text{pr}}$.
- $F : \mathsf{X} \to \mathsf{Y}$.
- e: Random noise.

Goal: sample u **conditioned on** y from μ^y, with

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_{\text{pr}}}(u) := \frac{1}{Z} e^{-\Phi(u;y)}, \quad \Phi(u;y) = \|y - F(u)\|_{\mathsf{Y}}^2,$$

$$Z = \int_{\mathsf{X}} e^{-\Phi(u;y)} \mu_{\text{pr}}(\mathrm{d}u),$$

Caveat: F usually gets evaluated numerically with discr. parameter L of high accuracy/cost. $\implies$ Induces F_{L}

Introduction: Bayesian Inverse Problems

Given measured data $y \in \mathrm{Y}$ modeled as

$$y = F_{\mathrm{L}}(u) + e, \quad e \sim \mu_{\text{noise}}, \quad u \in \mathrm{X},$$

- u: Parameter $u \sim \mu_{\mathrm{pr}}$.
- $F_{\mathrm{L}} : \mathrm{X} \to \mathrm{Y}$.
- e: Random noise.

Goal: sample u **conditioned on** y from μ_{L}^{y}, where

$$\frac{\mathrm{d}\mu_{\mathrm{L}}^{y}}{\mathrm{d}\mu_{\mathrm{pr}}}(u) = \frac{1}{Z_{\mathrm{L}}} e^{-\Phi_{\mathrm{L}}(u;y)}, \qquad \Phi_{\mathrm{L}}(u;y) = \|y - F_{\mathrm{L}}(u)\|_{\mathrm{Y}}^{2},$$

$$Z_{\mathrm{L}} = \int_{\mathrm{X}} e^{-\Phi_{\mathrm{L}}(u;y)} \mu_{\mathrm{pr}}(\mathrm{d}u),$$

with $\quad \mu_{\mathrm{L}}^{y} \to \mu^{y} \quad$ as $\mathrm{L} \to \infty$.

Introduction: MCMC- Metropolis Hastings

Constructs Markov chain $\{u^n_{\mathrm{L}}\}_{n=0}^N$ with invariant distribution μ^y_{L}

procedure MH(μ^y_{L}, Q,N,λ_0).

Sample $u^0 \sim \lambda_0$

for $n = 0, \dots, N-1$ **do**

Sample $z \sim Q(u^n, \cdot)$.

Set $u^{n+1} = z$ w/prob. α_{L}:

$$\alpha_{\mathrm{L}}(u^n, z) = \min\left[1, \frac{\mu^y_{\mathrm{L}}(z)Q(z, u^n)}{\mu^y_{\mathrm{L}}(u^n)Q(u^n, z)}\right],$$

set $u^{n+1} = u^n$ otherwise.

end for

Output $\{u^n\}_{n=0}^N$

end procedure

Constructs Markov chain $\{u_L^n\}_{n=0}^N$ with invariant distribution μ_L^y

procedure MH(μ_L^y, Q,N,λ_0).
  Sample $u^0 \sim \lambda_0$
  for $n = 0, \dots, N-1$ **do**
    Sample $z \sim Q(u^n, \cdot)$.
    Set $u^{n+1} = z$ w/prob. α_L:

$$\alpha_L(u^n, z) = \min\left[1, \frac{\mu_L^y(z)Q(z, u^n)}{\mu_L^y(u^n)Q(u^n, z)}\right],$$

    set $u^{n+1} = u^n$ otherwise.
  end for
  Output $\{u^n\}_{n=0}^N$
end procedure

In short: $u^{n+1} \sim p(u^n, \cdot)$.

Introduction: MCMC- Metropolis Hastings

Constructs Markov chain $\{u_L^n\}_{n=0}^N$ with invariant distribution μ_L^y

procedure MH(μ_L^y, Q,N,λ_0).
 Sample $u^0 \sim \lambda_0$
 for $n = 0, \dots, N-1$ **do**
  Sample $z \sim Q(u^n, \cdot)$.
  Set $u^{n+1} = z$ w/prob. α_L:

$$\alpha_L(u^n, z) = \min\left[1, \frac{\mu_L^y(z)Q(z,u^n)}{\mu_L^y(u^n)Q(u^n,z)}\right],$$

  set $u^{n+1} = u^n$ otherwise.
 end for
 Output $\{u^n\}_{n=0}^N$
end procedure

In short: $u^{n+1} \sim p(u^n, \cdot)$. Once $\{u^n\}_{n=0}^N$ is obtained, we can estimate

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \mathbb{E}_{\mu_L^y}[\mathrm{QoI}_L]$$

$$\approx \widehat{\mathrm{QoI}_L} = \frac{1}{N}\sum_{n=0}^{N} \mathrm{QoI}_L(u^n).$$

Introduction: MCMC- Metropolis Hastings

Constructs Markov chain $\{u_L^n\}_{n=0}^N$ with invariant distribution μ_L^y

procedure MH(μ_L^y, Q, N, λ_0).

Sample $u^0 \sim \lambda_0$

for $n = 0, \ldots, N-1$ **do**

Sample $z \sim Q(u^n, \cdot)$.

Set $u^{n+1} = z$ w/prob. α_L:

$$\alpha_L(u^n, z) = \min\left[1, \frac{\mu_L^y(z)Q(z,u^n)}{\mu_L^y(u^n)Q(u^n,z)}\right],$$

set $u^{n+1} = u^n$ otherwise.

end for

Output $\{u^n\}_{n=0}^N$

end procedure

In short: $u^{n+1} \sim p(u^n, \cdot)$. Once $\{u^n\}_{n=0}^N$ is obtained, we can estimate

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \mathbb{E}_{\mu_L^y}[\mathrm{QoI}_L]$$
$$\approx \widehat{\mathrm{QoI}}_L = \frac{1}{N}\sum_{n=0}^{N} \mathrm{QoI}_L(u^n).$$

Accuracy is controlled by N and L.

Issue: Large N + Expensive F_L.

Introduction: MCMC- Metropolis Hastings

Constructs Markov chain $\{u_L^n\}_{n=0}^N$ with invariant distribution μ_L^y

procedure MH(μ_L^y, Q,N,λ_0).
 Sample $u^0 \sim \lambda_0$
 for $n = 0, \ldots, N-1$ **do**
  Sample $z \sim Q(u^n, \cdot)$.
  Set $u^{n+1} = z$ w/prob. α_L:

$$\alpha_L(u^n, z) = \min\left[1, \frac{\mu_L^y(z)Q(z,u^n)}{\mu_L^y(u^n)Q(u^n,z)}\right],$$

  set $u^{n+1} = u^n$ otherwise.
 end for
 Output $\{u^n\}_{n=0}^N$
end procedure

In short: $u^{n+1} \sim p(u^n, \cdot)$. Once $\{u^n\}_{n=0}^N$ is obtained, we can estimate

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \mathbb{E}_{\mu_L^y}[\mathrm{QoI}_L]$$

$$\approx \widehat{\mathrm{QoI}}_L = \frac{1}{N}\sum_{n=0}^{N} \mathrm{QoI}_L(u^n).$$

Accuracy is controlled by N and L.

Issue: Large N + Expensive F_L.

Solution: ML-MCMC

ML-MCMC: motivation

Introduce $\{F_\ell\}_{\ell=0}^{L}$ of increasing accuracy/cost; $F_\ell \to F$ as $\ell \to \infty$.

(a) $\ell = 0$, F_0 (b) $\ell = 1$, F_1 (c) $\ell = 2$, F_2 (d) $\ell = 3$, F_3

Induces μ_ℓ^y, $\ell = 0, 1, \ldots, L$, with

$$\frac{d\mu_\ell^y}{d\mu_{pr}}(u) = \frac{1}{Z_\ell} \exp\left(-\|y - F_\ell(u)\|_Y^2\right)$$

Introduce $\{F_\ell\}_{\ell=0}^{L}$ of increasing accuracy/cost; $F_\ell \to F$ as $\ell \to \infty$.

(a) $\ell = 0$, F_0 (b) $\ell = 1$, F_1 (c) $\ell = 2$, F_2 (d) $\ell = 3$, F_3

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \simeq \mathbb{E}_{\mu^y_L}[\mathrm{QoI}_L] = \mathbb{E}_{\mu^y_0}[\mathrm{QoI}_0] + \sum_{\ell=1}^{L}\left(\mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y_{\ell-1}}[\mathrm{QoI}_{\ell-1}]\right)$$

$$\approx \widehat{\mathrm{QoI}}_{L,\{N_\ell\}_{\ell=0}^{L}} := \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u^n_{0,0})] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell({u^m}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}({u^m}_{\ell,\ell-1}))}_{:= Y^m_\ell},$$

ML-MCMC: motivation

Introduce $\{F_\ell\}_{\ell=0}^{L}$ of increasing accuracy/cost; $F_\ell \to F$ as $\ell \to \infty$.

(a) $\ell = 0$, F_0 (b) $\ell = 1$, F_1 (c) $\ell = 2$, F_2 (d) $\ell = 3$, F_3

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \simeq \mathbb{E}_{\mu^y_L}[\mathrm{QoI}_L] = \mathbb{E}_{\mu^y_0}[\mathrm{QoI}_0] + \sum_{\ell=1}^{L}\left(\mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y_{\ell-1}}[\mathrm{QoI}_{\ell-1}]\right)$$

$$\approx \widehat{\mathrm{QoI}}_{L,\{N_\ell\}_{\ell=0}^{L}} := \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u^n_{0,0})] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell({u^m}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}({u^m}_{\ell,\ell-1}))}_{:= Y^m_\ell},$$

$u_{\star,\ell} \sim \mu^y_\ell$, $\quad u_{\star,\ell-1} \sim \mu^y_{\ell-1}$.

ML-MCMC: motivation

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u^n_{0,0})] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell({u^m}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}({u^m}_{\ell,\ell-1}))}_{:= Y_\ell^m}.$$

Fact: Choosing $N_\ell < N_{\ell-1}$ very carefully gives a much cheaper estimator with same accuracy as SL.

ML-MCMC: motivation

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u_{0,0}^n)] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell(u^m{}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}(u^m{}_{\ell,\ell-1}))}_{:= Y_\ell^m}.$$

Theorem [DKST15] (Informal) : If $\exists\ \alpha_w, \beta, \gamma > 0$:

- T1. $|\mathbb{E}_{\mu_\ell^y}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w \ell}$,
- T2. $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}$.
- T3. $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell]$.
- T4. $\mathcal{C}_\ell(\mathrm{QoI}_\ell) \leq C_\gamma M^{\gamma\ell}$.

[DKST15]: Dodwell, et Al. (2015).

ML-MCMC: motivation

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u^n_{0,0})] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell({u^m}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}({u^m}_{\ell,\ell-1}))}_{:=Y_\ell^m}.$$

Theorem [DKST15] (Informal) : If $\exists\ \alpha_w, \beta, \gamma > 0$:

T1. $|\mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w \ell}$,

T2. $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}$.

T3. $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell]$.

T4. $\mathcal{C}_\ell(\mathrm{QoI}_\ell) \leq C_\gamma M^{\gamma\ell}$.

$$\mathrm{Cost}\left(e_{\mathrm{ML}}, \mathrm{tol}^2\right) \leq C_{\mathrm{ML}} \begin{cases} \mathrm{tol}^{-2}|\log \mathrm{tol}| & \text{if } \beta > \gamma, \\ \mathrm{tol}^{-2}|\log \mathrm{tol}|^3, & \text{if } \beta = \gamma, \\ \mathrm{tol}^{-2+(\gamma-\beta)/\alpha_w}|\log \mathrm{tol}|, & \text{if } \beta < \gamma. \end{cases}$$

$$\mathrm{Cost}\left(e_{\mathrm{SL}}, \mathrm{tol}^2\right) \leq C_{\mathrm{SL}} \mathrm{tol}^{-2-\gamma/\alpha_w}$$

ML-MCMC: motivation

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u^n_{0,0})] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell({u^m}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}({u^m}_{\ell,\ell-1}))}_{:=\, Y^m_\ell}.$$

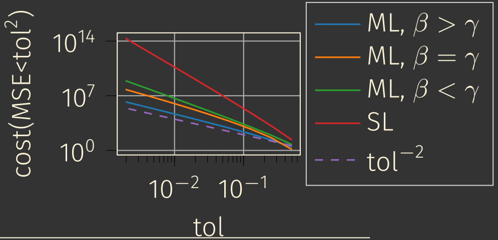
Theorem [DKST15] (Informal) : If $\exists\ \alpha_w, \beta, \gamma > 0$:

T1. $|\mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w \ell}$,

T2. $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}$.

T3. $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell]$.

T4. $\mathcal{C}_\ell(\mathrm{QoI}_\ell) \leq C_\gamma M^{\gamma\ell}$.



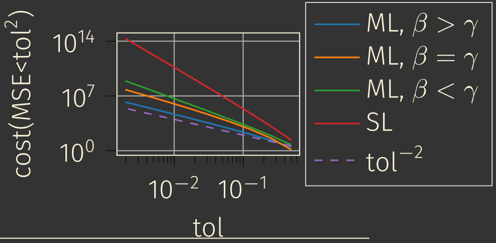
[DKST15]: Dodwell, et Al. (2015).

ML-MCMC: motivation

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u^n_{0,0})] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell({u^m}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}({u^m}_{\ell,\ell-1}))}_{:=Y^m_\ell}.$$

Theorem [DKST15] (Informal) : If $\exists\ \alpha_w, \beta, \gamma > 0$:

- T1. $|\mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \le C_w M^{-\alpha_w \ell}$,
- T2. $\mathbb{V}_{\nu_\ell}[Y_\ell] \le C_v M^{-\beta\ell}$.
- T3. $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \le N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell]$.
- T4. $\mathcal{C}_\ell(\mathrm{QoI}_\ell) \le C_\gamma M^{\gamma\ell}$.



T2: $Y^m_\ell \xrightarrow{\ell\to\infty} 0 \implies \mathbb{V}_{\nu_\ell}[Y_\ell] \xrightarrow{\ell\to\infty} 0$

(Not trivial)

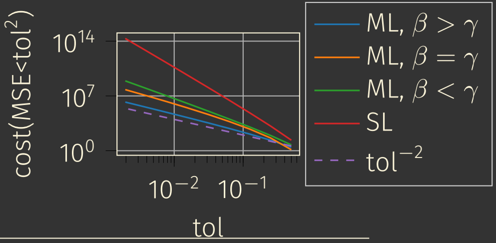
[DKST15]: Dodwell, et Al. (2015).

ML-MCMC: motivation

$$\mathbb{E}_{\mu^y}[\mathrm{QoI}] \approx \frac{1}{N_0}\sum_{n=0}^{N_0}[\mathrm{QoI}_0(u^n_{0,0})] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{m=0}^{N_\ell}\underbrace{(\mathrm{QoI}_\ell({u^m}_{\ell,\ell}) - \mathrm{QoI}_{\ell-1}({u^m}_{\ell,\ell-1}))}_{:= Y_\ell^m}.$$

Theorem [DKST15] (Informal) : If $\exists\ \alpha_w, \beta, \gamma > 0$:

- T1. $|\mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w \ell}$,
- T2. $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}$.
- T3. $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell]$.
- T4. $\mathcal{C}_\ell(\mathrm{QoI}_\ell) \leq C_\gamma M^{\gamma\ell}$.



T2: $Y_\ell^m \xrightarrow{\ell\to\infty} 0 \implies \mathbb{V}_{\nu_\ell}[Y_\ell] \xrightarrow{\ell\to\infty} 0$

(Not trivial)

Key: Construct coupled chains by coupling proposals.

[DKST15]: Dodwell, et Al. (2015).

ML-MCMC: generating couplings

Γ is a **coupling** of $Q(u^n_{\ell,\ell},\cdot)$ and $R(u^n_{\ell,\ell-1},\cdot)$, if:

A. $(u'_{\ell,\ell-1}, u'_{\ell,\ell}) \sim \Gamma, \implies u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1},\cdot)$ and $u_{\ell,\ell} \sim Q(u^n_{\ell,\ell},\cdot)$.

ML-MCMC: generating couplings

Γ is a **coupling** of $Q(u^n_{\ell,\ell}, \cdot)$ and $R(u^n_{\ell,\ell-1}, \cdot)$, if:

A. $(u'_{\ell,\ell-1}, u'_{\ell,\ell}) \sim \Gamma, \implies u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1}, \cdot)$ and $u_{\ell,\ell} \sim Q(u^n_{\ell,\ell}, \cdot)$.

For any coupling:

B. $\mathbb{P}_\Gamma(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) \geq \left\| Q(u^n_{\ell,\ell}, \cdot) - R(u^n_{\ell,\ell-1}, \cdot) \right\|_{\mathrm{TV}}.$

We say it is a maximal coupling if B. holds with equality.

ML-MCMC: generating couplings

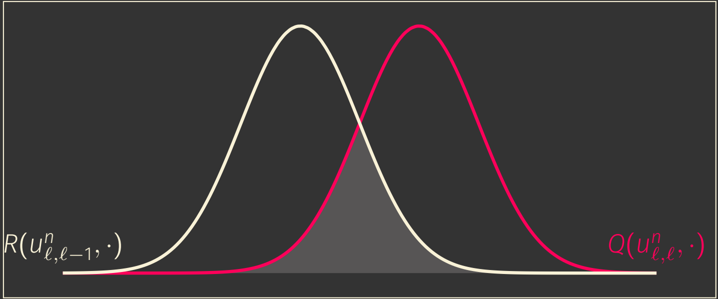
Γ is a **coupling** of $Q(u^n_{\ell,\ell},\cdot)$ and $R(u^n_{\ell,\ell-1},\cdot)$, if:

A. $(u'_{\ell,\ell-1}, u'_{\ell,\ell}) \sim \Gamma, \implies u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1},\cdot)$ and $u_{\ell,\ell} \sim Q(u^n_{\ell,\ell},\cdot)$.

For any coupling:

B. $\mathbb{P}_\Gamma(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) \geq \left\| Q(u^n_{\ell,\ell},\cdot) - R(u^n_{\ell,\ell-1},\cdot) \right\|_{\mathrm{TV}}.$

We say it is a maximal coupling if B. holds with equality.



ML-MCMC: generating couplings

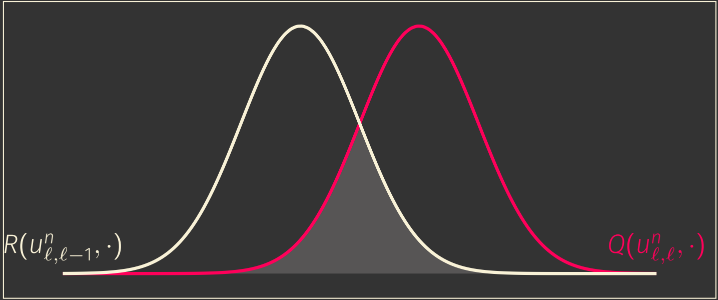
Γ is a **coupling** of $Q(u^n_{\ell,\ell},\cdot)$ and $R(u^n_{\ell,\ell-1},\cdot)$, if:

A. $(u'_{\ell,\ell-1}, u'_{\ell,\ell}) \sim \Gamma, \implies u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1},\cdot)$ and $u_{\ell,\ell} \sim Q(u^n_{\ell,\ell},\cdot)$.

For any coupling:

B. $\mathbb{P}_\Gamma(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) \geq \left\| Q(u^n_{\ell,\ell},\cdot) - R(u^n_{\ell,\ell-1},\cdot) \right\|_{\mathrm{TV}}.$

We say it is a maximal coupling if B. holds with equality.



1. IMH $Q(u^n_{\ell,\ell},\cdot) = R(u^n_{\ell,\ell-1},\cdot) = Q(\cdot)$.
2. MxC: $Q(u^n_{\ell,\ell},\cdot)$, $R(u^n_{\ell,\ell-1},\cdot)$.

   (g)pCN, RWM, HMC,..

ML-MCMC: coupling via IMH

IMH: Couples chains by **always** proposing the same state (independently).

[DKST15]: Dodwell, et Al. (2015). [MCNT21]: MC, Nobile, Tempone. (2021.) [MC22]: MC. (2022.)

ML-MCMC: coupling via IMH

IMH: Couples chains by **always** proposing the same state (independently).

[DKST15]: Dodwell, et Al. (2015). [MCNT21]: MC, Nobile, Tempone. (2021.) [MC22]: MC. (2022.)

ML-MCMC: coupling via IMH

IMH: Couples chains by **always** proposing the same state (independently).

procedure IMH($\{\mu^y_{\ell-1}, \mu^y_{\ell}\}, \{u^n_{\ell,\ell-1}, u^n_{\ell,\ell}\}, Q_\ell$)

Sample $z \sim Q_\ell$ and $w \sim \mathcal{U}(0,1)$.

for $j = \ell - 1, \ell$ **do**

Set $u^{n+1}_{\ell,j} = z$ if $w < \alpha_j(u^n_{\ell,j}, z)$, where

$$\alpha_j(u^n_{\ell,j}, z) := \min\left[1, \frac{\mu^y_j(z) Q_\ell(u^n_{\ell,j})}{\mu^y_j(u^n_{\ell,j}) Q_\ell(z)}\right].$$

Set $u^{n+1}_{\ell,j} = u^n_{\ell,j}$ otherwise.

end for

Output $\{u^{n+1}_{\ell,\ell-1}, u^{n+1}_{\ell,\ell}\}$.

end procedure

[DKST15]: Dodwell, et Al. (2015). [MCNT21]: MC, Nobile, Tempone. (2021.) [MC22]: MC. (2022.)

ML-MCMC: coupling via IMH

IMH: Couples chains by **always** proposing the same state (independently).

procedure IMH($\{\mu^y_{\ell-1}, \mu^y_\ell\}, \{u^n_{\ell,\ell-1}, u^n_{\ell,\ell}\}, Q_\ell$)
  Sample $z \sim Q_\ell$ and $w \sim \mathcal{U}(0,1)$.
  for $j = \ell-1, \ell$ **do**
    Set $u^{n+1}_{\ell,j} = z$ if $w < \alpha_j(u^n_{\ell,j}, z)$, where

$$\alpha_j(u^n_{\ell,j}, z) := \min\left[1, \frac{\mu^y_j(z) Q_\ell(u^n_{\ell,j})}{\mu^y_j(u^n_{\ell,j}) Q_\ell(z)}\right].$$

    Set $u^{n+1}_{\ell,j} = u^n_{\ell,j}$ otherwise.
  end for
  Output $\{u^{n+1}_{\ell,\ell-1}, u^{n+1}_{\ell,\ell}\}$.
end procedure

- Induces joint kernel $\mathbf{p}_{\ell,IMH}$.
- Choosing Q_ℓ is critical
- [DKST15] proposes $Q_\ell(\cdot) = \frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu^y_{\ell-1}$, however, **can give biased results**.
- [MCNT21,MC22] extend to a wider class of ind. proposals Q_ℓ w/ certain conditions (prior, KDE, Laplace approximation) . Difficult to find in some situations.

[DKST15]: Dodwell, et Al. (2015). [MCNT21]: MC, Nobile, Tempone. (2021.) [MC22]: MC. (2022.)

ML-MCMC: coupling via MxC

MxC: Γ^n of state-dependent proposals[MCN22][HJLT21] $Q^n = Q(u^n_{\ell,\ell}, \cdot)$, $R^n = R(u^n_{\ell,\ell-1}, \cdot)$

[MCN22]: MC, Nobile. (2022) . [HJLT21]: Heng, et. al. (2021).

ML-MCMC: coupling via MxC

MxC: Γ^n of state-dependent proposals[MCN22][HJLT21] $Q^n = Q(u^n_{\ell,\ell}, \cdot)$, $R^n = R(u^n_{\ell,\ell-1}, \cdot)$

procedure MxC($\mu^y_\ell, \mu^y_{\ell-1}, u^n_{\ell,\ell-1}, u^n_{\ell,\ell}, Q^n, R^n$)
    Sample $(u'_{\ell,\ell-1}, u'_{\ell,\ell}) \sim \Gamma^n$, $w \sim \mathcal{U}([0,1])$
    for $j = \ell-1, \ell$ **do**:
        if $w \leq \alpha_j(u^n_{\ell,j}, u'_{\ell,j})$ **then**
            Set $u^{n+1}_{\ell,j} = u'_{\ell,j}$.
        else
            Set $u^{n+1}_{\ell,j} = u^n_{\ell,j}$.
        end if
    end for
    Output $(u^{n+1}_{\ell,\ell-1}, u^{n+1}_{\ell,\ell})$.
end procedure

[MCN22]: MC, Nobile. (2022) . [HJLT21]: Heng, et. al. (2021).

ML-MCMC: coupling via MxC

MxC: Γ^n of state-dependent proposals[MCN22][HJLT21] $Q^n = Q(u^n_{\ell,\ell}, \cdot)$, $R^n = R(u^n_{\ell,\ell-1}, \cdot)$

procedure MxC($\mu^y_\ell, \mu^y_{\ell-1}, u^n_{\ell,\ell-1}, u^n_{\ell,\ell}, Q^n, R^n$)
  Sample $(u'_{\ell,\ell-1}, u'_{\ell,\ell}) \sim \Gamma^n$, $w \sim \mathcal{U}([0,1])$
  for $j = \ell-1, \ell$ **do**:
    if $w \leq \alpha_j(u^n_{\ell,j}, u'_{\ell,j})$ **then**
      Set $u^{n+1}_{\ell,j} = u'_{\ell,j}$.
    else
      Set $u^{n+1}_{\ell,j} = u^n_{\ell,j}$.
    end if
  end for
  Output $(u^{n+1}_{\ell,\ell-1}, u^{n+1}_{\ell,\ell})$.
end procedure

- Induces joint kernel $\mathbf{p}_{\ell,MxC}$.
- ML-MCMC w/ state-dependent proposals.
- Proposes same state w.p. $1 - \left\| Q(u^n_{\ell,\ell}, \cdot) - R(u^n_{\ell,\ell-1}, \cdot) \right\|_{TV}$, small if $u^n_{\ell,\ell-1}, u^n_{\ell,\ell}$ are too far apart

[MCN22]: MC, Nobile. (2022) . [HJLT21]: Heng, et. al. (2021).

ML-MCMC: IMH convergence

Theorem 1: Convergence of joint IMH chain[MCNT21]**.**

A.1 $\exists\, c \in (0,1)$ independent of ℓ s.t. $\forall \ell \geq 1$:

$$\operatorname*{ess\,inf}_{z\in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0, j = \ell-1, \ell, \ \mu_{\text{pr}} - \text{a.s.} \tag{1}$$

[MCNT21]: MC, Nobile, Tempone. (2021)

Theorem 1: Convergence of joint IMH chain[MCNT21].

A.1 $\exists\, c \in (0,1)$ independent of ℓ s.t. $\forall \ell \geq 1$:

$$\operatorname*{ess\,inf}_{z\in X}\left\{\frac{Q_\ell(z)}{\mu_j^y(z)}\right\} \geq c > 0, j = \ell-1, \ell,\ \mu_{\mathrm{pr}} - \text{a.s.} \tag{1}$$

Then, $\forall \ell \geq 1\ \exists!$ a invariant ν_ℓ induced by $\mathbf{P}_{\ell,\mathrm{imh}}$. Furthermore,

$$\sup_{\|f\|_{L_\infty(X^2,\mu_{\mathrm{pr}}^2)}} \left| \int_{X^2} f(\mathbf{u}_\ell')\mathbf{p}_\ell^n(\mathbf{u}_\ell, \mathrm{d}\mathbf{u}_\ell') - \int_{X^2} f(\mathbf{u}_\ell)\nu_\ell(\mathrm{d}\mathbf{u}_\ell)\right| \leq 2(1-\rho_\ell)^n, \quad \forall \mathbf{u}_\ell \in X^2, n \in \mathbb{N},$$

$\rho_\ell := c \int_X \min\{\mu^y(z), \mu_{\ell-1}^y(z)\}\mu_{\mathrm{pr}}(\mathrm{d}z) = c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)]$.

[MCNT21]: MC, Nobile, Tempone. (2021)

Theorem 1: Convergence of joint IMH chain[MCNT21]**.**

A.1 $\exists\, c \in (0,1)$ independent of ℓ s.t. $\forall \ell \geq 1$:

$$\operatorname*{ess\,inf}_{z\in X}\left\{\frac{Q_\ell(z)}{\mu_j^y(z)}\right\} \geq c > 0, j = \ell-1, \ell,\ \mu_{\text{pr}} - \text{a.s.} \tag{1}$$

Then, $\forall \ell \geq 1\ \exists!$ a invariant ν_ℓ induced by $\mathbf{P}_{\ell,\text{imh}}$. Furthermore,

$$\sup_{\|f\|_{L_\infty(X^2,\mu_{\text{pr}}^2)}}\left|\int_{X^2} f(\mathbf{u}_\ell')\mathbf{p}_\ell^n(\mathbf{u}_\ell, \mathrm{d}\mathbf{u}_\ell') - \int_{X^2} f(\mathbf{u}_\ell)\nu_\ell(\mathrm{d}\mathbf{u}_\ell)\right| \leq 2(1-\rho_\ell)^n, \quad \forall \mathbf{u}_\ell \in X^2, n \in \mathbb{N},$$

$\rho_\ell := c \int_X \min\{\mu^y(z), \mu_{\ell-1}^y(z)\}\mu_{\text{pr}}(\mathrm{d}z) = c[1 - d_{\text{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)]$.
Conversely, if (1) doesn't hold $\implies$ no geometric ergodicity.

[MCNT21]: MC, Nobile, Tempone. (2021)

ML-MCMC: IMH convergence

Theorem 1: Convergence of joint IMH chain[MCNT21]**.**

A.1 $\exists\, c \in (0,1)$ independent of ℓ s.t. $\forall \ell \geq 1$:

$$\operatorname*{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0, j = \ell - 1, \ell, \ \mu_{\mathrm{pr}} - \text{a.s.} \tag{1}$$

Then, $\forall \ell \geq 1\ \exists!$ a invariant ν_ℓ induced by $\mathbf{P}_{\ell,\mathrm{imh}}$. Furthermore,

$$\sup_{\|f\|_{L_\infty(X^2,\mu_{\mathrm{pr}}^2)}} \left| \int_{X^2} f(\mathbf{u}_\ell')\mathbf{p}_\ell^n(\mathbf{u}_\ell, \mathrm{d}\mathbf{u}_\ell') - \int_{X^2} f(\mathbf{u}_\ell)\nu_\ell(\mathrm{d}\mathbf{u}_\ell) \right| \leq 2(1-\rho_\ell)^n, \quad \forall \mathbf{u}_\ell \in X^2, n \in \mathbb{N},$$

$\rho_\ell := c \int_X \min\{\mu^y(z), \mu_{\ell-1}^y(z)\}\mu_{\mathrm{pr}}(\mathrm{d}z) = c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)]$.
Conversely, if (1) doesn't hold $\implies$ no geometric ergodicity.

Importance: first result of its kind + guideline on implementation + $\nu_\ell = \nu_\ell(Q_\ell)$.

[MCNT21]: MC, Nobile, Tempone. (2021)

Theorem 2: Convergence of joint MxC chain[MCN22]

A.2.1 $\Phi_\ell(u;y) > 0, \ \forall u \in X, y \in Y.$

A.2.2 $\mu_{\text{pr}} = \mathcal{N}(0, \mathcal{C})$

A.2.3 $Q(u^n_{\ell,\ell}, \cdot) = \mathcal{N}(\sqrt{1-\rho^2}(u^n_{\ell,\ell}, \rho^2\mathcal{C}), \quad R(u^n_{\ell,\ell-1}, \cdot) = \mathcal{N}(\sqrt{1-\rho^2}u^n_{\ell,\ell-1}, \rho^2\mathcal{C})$

+ some technical conditions on the conv. of marginal pCN.

[MCN22]: MC, Nobile. (2022)

ML-MCMC: MxC convergence

Theorem 2: Convergence of joint MxC chain[MCN22]

A.2.1 $\Phi_\ell(u;y) > 0, \ \forall u \in X, y \in Y.$

A.2.2 $\mu_{\text{pr}} = \mathcal{N}(0, \mathcal{C})$

A.2.3 $Q(u^n_{\ell,\ell}, \cdot) = \mathcal{N}(\sqrt{1-\rho^2}(u^n_{\ell,\ell}, \rho^2\mathcal{C}), \quad R(u^n_{\ell,\ell-1}, \cdot) = \mathcal{N}(\sqrt{1-\rho^2}u^n_{\ell,\ell-1}, \rho^2\mathcal{C})$

+ some technical conditions on the conv. of marginal pCN.

Then, $\forall \ell \geq 1 \ \exists!$ invariant ν_ℓ induced by $\mathbf{P}_{\ell,\text{mc}}$. Furthermore, $\exists \varrho \in (0,1), M < +\infty$, $V_\ell : X^2 \to [1, \infty)$ s.t.

$$\sup_{\|f\|_{L_\infty(X^2, \mu^2_{\text{pr}})}} \left| \int_{X^2} f(\mathbf{u}'_\ell)\mathbf{p}^n_\ell(\mathbf{u}_\ell, d\mathbf{u}'_\ell) - \int_{X^2} f(\mathbf{u}_\ell)\nu_\ell(d\mathbf{u}_\ell) \right| \leq M V_\ell(\mathbf{u}_\ell)\varrho^n, \quad \forall \mathbf{u}_\ell \in X^2, n \in \mathbb{N}.$$

[MCN22]: MC, Nobile. (2022)

ML-MCMC: MSE bound

Theorem 3: MSE Bound[MCNT21] (informal). $\mu p = \mu$, p **non-nec. reversible**, $\mu^0 \neq \mu$, and p mixes sufficiently fast. Then $\forall f \in L_2(\mathsf{X}, \mu)$:

[MCNT21]: MC, Nobile, Tempone. (2021)

Theorem 3: MSE Bound[MCNT21] (informal). $\mu p = \mu$, p **non-nec. reversible**, $\mu^0 \neq \mu$, and p mixes sufficiently fast. Then $\forall f \in L_2(\mathsf{X}, \mu)$:

$$\mathrm{E}_{\mu^0, p}\left[\left|\frac{1}{N}\sum_{n=0}^{N} f(u^n) - \mathbb{E}_\mu[f]\right|^2\right] \leq C_{\gamma_{\mathrm{ps}}} \frac{\mathbb{V}_\mu[f]}{N}.$$

[MCNT21]: MC, Nobile, Tempone. (2021)

ML-MCMC: MSE bound

Theorem 3: MSE Bound[MCNT21] (informal). $\mu p = \mu$, p **non-nec. reversible**, $\mu^0 \neq \mu$, and p mixes sufficiently fast. Then $\forall f \in L_2(\mathsf{X}, \mu)$:

$$\mathsf{E}_{\mu^0, p}\left[\left|\frac{1}{N}\sum_{n=0}^{N} f(u^n) - \mathbb{E}_\mu[f]\right|^2\right] \leq C_{\gamma_{\mathrm{ps}}} \frac{\mathbb{V}_\mu[f]}{N}.$$

Importance:

1. useful beyond ML-MCMC.
2. Needed for complexity result $(\mathrm{MSE}(\hat{Y}_{\ell, N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell])$.
3. Similar results in literature[Rud12] **but** difficult to verify assumption if non-reversible (as ML-MCMC).

[MCNT21]: MC, Nobile, Tempone. (2021) , [Rud12]: Rudolf. (2012).

ML-MCMC: IMH complexity

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u)s^{-\alpha\ell}, \ \forall u \in X.$

[MCNT21]: MC, Nobile, Tempone. (2021).

ML-MCMC: IMH complexity

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u) s^{-\alpha \ell}, \ \forall u \in X.$

A.3.2 $|\mathrm{QoI}_\ell(u) - \mathrm{QoI}(u)| \leq C_q(u) s^{-\alpha_q \ell}, \ \forall u \in X.$

[MCNT21]: MC, Nobile, Tempone. (2021).

ML-MCMC: IMH complexity

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u)s^{-\alpha\ell}, \ \forall u \in X.$

A.3.2 $|\mathrm{QoI}_\ell(u) - \mathrm{QoI}(u)| \leq C_q(u)s^{-\alpha_q\ell}, \ \forall u \in X.$

A.3.3 $\int_X |\mathrm{QoI}_\ell^m(u)|\mu_{\mathrm{pr}}(du) \leq C_{\mathrm{m}}$, for some $m > 2$

+ some μ_{pr}-integrability conditions on constants and Q,

[MCNT21]: MC, Nobile, Tempone. (2021).

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u) s^{-\alpha\ell}, \ \forall u \in X.$

A.3.2 $|\mathrm{QoI}_\ell(u) - \mathrm{QoI}(u)| \leq C_q(u) s^{-\alpha_q \ell}, \ \forall u \in X.$

A.3.3 $\int_X |\mathrm{QoI}_\ell^m(u)| \mu_{\mathrm{pr}}(\mathrm{d}u) \leq C_{\mathrm{m}}$, for some $m > 2$

+ some μ_{pr}-integrability conditions on constants and Q, then,

T.1 $|\mathbb{E}_{\mu_\ell^y}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w \ell},$

T.2 $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}.$

T.3 $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell].$

Here, $\alpha_w = \min\{a_q, a\}, \ \beta = \min\{2\, a_q, \alpha(1 - 2/m)\}$

[MCNT21]: MC, Nobile, Tempone. (2021).

ML-MCMC: IMH complexity

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u) s^{-\alpha \ell}, \ \forall u \in X.$

A.3.2 $|\mathrm{QoI}_\ell(u) - \mathrm{QoI}(u)| \leq C_q(u) s^{-\alpha_q \ell}, \ \forall u \in X.$

A.3.3 $\int_X |\mathrm{QoI}_\ell^m(u)| \mu_{\mathrm{pr}}(du) \leq C_{\mathrm{m}}$, for some $m > 2$

+ some μ_{pr}-integrability conditions on constants and Q, then,

T.1 $|\mathbb{E}_{\mu_\ell^y}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w \ell},$

T.2 $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta \ell}.$

T.3 $\mathrm{MSE}(\hat{Y}_{\ell, N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell].$

Here, $\alpha_w = \min\{a_q, a\}, \ \beta = \min\{2\, a_q, \alpha(1 - 2/m)\} \implies$ Dodwell's complexity result holds

[MCNT21]: MC, Nobile, Tempone. (2021).

ML-MCMC: IMH complexity

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u) s^{-\alpha\ell}, \ \forall u \in X.$

A.3.2 $|\mathrm{QoI}_\ell(u) - \mathrm{QoI}(u)| \leq C_q(u) s^{-\alpha_q \ell}, \ \forall u \in X.$

A.3.3 $\int_X |\mathrm{QoI}_\ell^m(u)| \mu_{\mathrm{pr}}(du) \leq C_{\mathrm{m}}$, for some $m > 2$

+ some μ_{pr}-integrability conditions on constants and Q, then,

T.1 $|\mathbb{E}_{\mu_\ell^y}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w \ell},$

T.2 $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}.$

T.3 $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell].$

Here, $\alpha_w = \min\{a_q, a\}, \ \beta = \min\{2\, a_q, \alpha(1-2/m)\} \implies$ Dodwell's complexity result holds

Similar result for MxC if $\exists c_\ell \not\to 0$ s.t. $\int_{\Delta^c} \mathrm{p}_\ell(\mathbf{u}_\ell, \Delta)\nu_\ell(d\mathbf{u}_\ell) > c_\ell$ (strong)

[MCNT21]: MC, Nobile, Tempone. (2021).

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u)s^{-\alpha\ell}, \ \forall u \in X.$

A.3.2 $|\mathrm{QoI}_\ell(u) - \mathrm{QoI}(u)| \leq C_q(u)s^{-\alpha_q\ell}, \ \forall u \in X.$

A.3.3 $\int_X |\mathrm{QoI}_\ell^m(u)|\mu_{\mathrm{pr}}(du) \leq C_{\mathrm{m}}$, for some $m > 2$

+ some μ_{pr}-integrability conditions on constants and Q, then,

T.1 $|\mathbb{E}_{\mu_\ell^y}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w\ell},$

T.2 $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}.$

T.3 $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell].$

Here, $\alpha_w = \min\{a_q, a\}, \ \beta = \min\{2\, a_q, \alpha(1 - 2/m)\} \implies$ Dodwell's complexity result holds

[MCNT21]: MC, Nobile, Tempone. (2021).

ML-MCMC: IMH complexity

Theorem 4: Complexity of ML-IMH algorithm [MCNT21] (informal) Suppose A.1 holds and:

A.3.1 $|\Phi_\ell(u;y) - \Phi(u;y)| \leq C_\Phi(u)s^{-\alpha\ell}, \ \forall u \in X.$

A.3.2 $|\mathrm{QoI}_\ell(u) - \mathrm{QoI}(u)| \leq C_q(u)s^{-\alpha_q\ell}, \ \forall u \in X.$

A.3.3 $\int_X |\mathrm{QoI}_\ell^m(u)|\mu_{\mathrm{pr}}(du) \leq C_{\mathrm{m}}$, for some $m > 2$

+ some μ_{pr}-integrability conditions on constants and Q, then,

T.1 $|\mathbb{E}_{\mu_\ell^y}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \leq C_w M^{-\alpha_w\ell},$

T.2 $\mathbb{V}_{\nu_\ell}[Y_\ell] \leq C_v M^{-\beta\ell}.$

T.3 $\mathrm{MSE}(\hat{Y}_{\ell,N_\ell}) \leq N_\ell^{-1} C_{\mathrm{mse}} \mathbb{V}_{\nu_\ell}[Y_\ell].$

Here, $\alpha_w = \min\{a_q, a\}, \ \beta = \min\{2\,a_q, \alpha(1-2/m)\} \implies$ Dodwell's complexity result holds

importance: Extends Dodwell's result beyond $Q_\ell = \mu_{\ell-1}^y$.

[MCNT21]: MC, Nobile, Tempone. (2021).

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname*{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell \to \infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell \to \infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u_{\ell-1,\ell-1}^i}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell \to \infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
    2.1 Moderate dim. using KDE, prior, measure transport.

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname*{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell\to\infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
   2.1 Moderate dim. using KDE, prior, measure transport.
   2.2 Higher dim. if sufficiently good Laplace approx. can be constructed (small noise, many obs, μ_{pr} Gaussian...)

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0$.

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \stackrel{\ell\to\infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
   2.1 Moderate dim. using KDE, prior, measure transport.
   2.2 Higher dim. if sufficiently good Laplace approx. can be constructed (small noise, many obs, μ_{pr} Gaussian...)
3. Otherwise, use MxC:

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname*{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell \to \infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
   2.1 Moderate dim. using KDE, prior, measure transport.
   2.2 Higher dim. if sufficiently good Laplace approx. can be constructed (small noise, many obs, μ_{pr} Gaussian...)
3. Otherwise, use MxC:
   - Easy to implement proposals (e.g., pCN) coupling algo.

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell \to \infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
   - 2.1 Moderate dim. using KDE, prior, measure transport.
   - 2.2 Higher dim. if sufficiently good Laplace approx. can be constructed (small noise, many obs, μ_{pr} Gaussian...)
3. Otherwise, use MxC:
   - Easy to implement proposals (e.g., pCN) coupling algo.
   - Less efficient than a good IMH.

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname*{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0$.

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell \to \infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
   - 2.1 Moderate dim. using KDE, prior, measure transport.
   - 2.2 Higher dim. if sufficiently good Laplace approx. can be constructed (small noise, many obs, μ_{pr} Gaussian...)
3. Otherwise, use MxC:
   - Easy to implement proposals (e.g., pCN) coupling algo.
   - Less efficient than a good IMH.
   - Depends on the coupling algorithm; dimensionality, multi-modality.

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell\to\infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
   2.1 Moderate dim. using KDE, prior, measure transport.
   2.2 Higher dim. if sufficiently good Laplace approx. can be constructed (small noise, many obs, μ_{pr} Gaussian...)
3. Otherwise, use MxC:
   - Easy to implement proposals (e.g., pCN) coupling algo.
   - Less efficient than a good IMH.
   - Depends on the coupling algorithm; dimensionality, multi-modality.
   - If $\left\|Q(u^n_{\ell,\ell}, \cdot) - R(u^n_{\ell,\ell-1}, \cdot)\right\|_{\mathrm{TV}} \approx 1 \implies \mathbb{P}(u_{\ell,\ell-1} \neq u_{\ell,\ell}) \approx 1.$

ML-MCMC: Some key points on the choice of method

Want marginals to stay close.

A1: $\operatorname*{ess\,inf}_{z \in X} \left\{ \frac{Q_\ell(z)}{\mu_j^y(z)} \right\} \geq c > 0.$

1. If $\mu_{\ell-1}^y$ available AND A1 $\implies$ use sub-sampling (IMH with $Q_\ell = \mu_{\ell-1}^y$). Indeed rate $1 - c[1 - d_{\mathrm{TV}}(\mu_\ell^y, \mu_{\ell-1}^y)] \overset{\ell\to\infty}{\to} 0$. (Tricky in practice since $\frac{1}{N_{\ell-1}} \sum_{i=1}^{N_{\ell-1}} \delta_{u^i_{\ell-1,\ell-1}}(\cdot) \approx \mu_{\ell-1}^y$).
2. If one has $Q_\ell \approx \mu_{\ell-1}^y$ AND A1 $\implies$ use IMH.
   - 2.1 Moderate dim. using KDE, prior, measure transport.
   - 2.2 Higher dim. if sufficiently good Laplace approx. can be constructed (small noise, many obs, μ_{pr} Gaussian...)
3. Otherwise, use MxC:
   - Easy to implement proposals (e.g., pCN) coupling algo.
   - Less efficient than a good IMH.
   - Depends on the coupling algorithm; dimensionality, multi-modality.
   - If $\left\|Q(u^n_{\ell,\ell}, \cdot) - R(u^n_{\ell,\ell-1}, \cdot)\right\|_{\mathrm{TV}} \approx 1 \implies \mathbb{P}(u_{\ell,\ell-1} \neq u_{\ell,\ell}) \approx 1$.
   - A workaround: $\mathbf{p}_{c,\ell} = (1-\omega)\mathbf{p}_{\ell,MxC} + \omega \mathbf{p}_{\ell,IMH}$, $\omega \in (0,1)$.

ML-MCMC Example: Subsurface flow

Given

$$\begin{cases} -\nabla_x \cdot (e^{u(x)} \nabla_x h(x,u)) = 1, \quad x \in D, \; u \in \mathrm{X}, \\ +\text{Suitable BCs} \end{cases}, \tag{2}$$

ML-MCMC Example: Subsurface flow

Given

$$\begin{cases} -\nabla_x \cdot (e^{u(x)} \nabla_x h(x,u)) = 1, \quad x \in D, \ u \in \mathrm{X}, \\ +\text{Suitable BCs} \end{cases}, \tag{2}$$

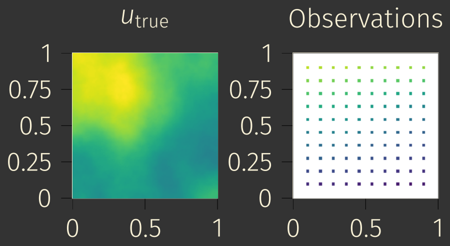
Goal: Find μ^y of permeability $\kappa(x,u) = e^{u(x)}$ in domain D given noisy measurements of pressure head h on D and Estimate $\mathbb{E}_{\mu^y}[\mathrm{QoI}]$, $\mathrm{QoI}(u) := \ln\left(\int_{\Gamma_{D_1}} e^{u(x)} \nabla h(x,u) \cdot \mathbf{n} \, ds\right)$.

Given

$$\begin{cases} -\nabla_x \cdot (e^{u(x)} \nabla_x h(x,u)) = 1, \quad x \in D,\ u \in X, \\ +\text{Suitable BCs} \end{cases}, \tag{2}$$

Goal: Find μ^y of permeability $\kappa(x,u) = e^{u(x)}$ in domain D given noisy measurements of pressure head h on D and Estimate $\mathbb{E}_{\mu^y}[\mathrm{QoI}]$, $\mathrm{QoI}(u) := \ln\left(\int_{\Gamma_{D_1}} e^{u(x)} \nabla h(x,u) \cdot \mathbf{n}\, ds\right)$.

- Introduce a FE mesh of $K^* = (16 \times 2^4 + 1)^2$ elements.
- Data y generated by sampling $u_{\mathrm{true}} \sim \mu_{\mathrm{pr}}$, solving (2), obs. at 100 points in $[0, 0.9]^2$ and polluting with 1% Gaussian noise.
- $u \sim \mathcal{N}(0, \mathcal{A}^{-2}) =: \mu_{\mathrm{pr}}$, $\mathcal{A} : \Delta + \frac{1}{2} I$



ML-MCMC Example: Subsurface flow

Introduce $\{F_\ell\}_{\ell=0}^{3}$ by discr. of (2) on F.E mesh $K_\ell := (16 \times 2^\ell + 1)^2$ elements $u = \sum_j^{K_\ell} \mathrm{u}_j \phi_j$

$\implies \quad u_{\ell,\ell} = (\mathrm{u}_1, \ldots, \mathrm{u}_{K_\ell})$

ML-MCMC Example: Subsurface flow

Introduce $\{F_\ell\}_{\ell=0}^{3}$ by discr. of (2) on F.E mesh $K_\ell := (16 \times 2^\ell + 1)^2$ elements $u = \sum_j^{K_\ell} \mathrm{u}_j \phi_j$
$\implies \quad u_{\ell,\ell} = (\mathrm{u}_1, \ldots, \mathrm{u}_{K_\ell})$ Induces

$$\frac{\mathrm{d}\mu_\ell^y}{\mathrm{d}\mu_{\mathrm{pr}}}(u_{\ell,\ell}) = \frac{1}{Z_\ell} \exp\left(-\frac{1}{2}\left\|y - F_\ell(u_{\ell,\ell})\right\|_{\mathrm{Y}}^2\right), \quad \dim(\mathrm{X}_\ell) = (16 \times 2^\ell + 1)^2, \quad \ell = 0, 1, 2, 3.$$

ML-MCMC Example: Subsurface flow

Introduce $\{F_\ell\}_{\ell=0}^{3}$ by discr. of (2) on F.E mesh $K_\ell := (16 \times 2^\ell + 1)^2$ elements $u = \sum_j^{K_\ell} \mathrm{u}_j \phi_j$
$\implies \quad u_{\ell,\ell} = (\mathrm{u}_1, \ldots, \mathrm{u}_{K_\ell})$ Induces

$$\frac{d\mu_\ell^y}{d\mu_{\mathrm{pr}}}(u_{\ell,\ell}) = \frac{1}{Z_\ell} \exp\left(-\frac{1}{2}\|y - F_\ell(u_{\ell,\ell})\|_Y^2\right), \quad \dim(X_\ell) = (16 \times 2^\ell + 1)^2, \quad \ell = 0, 1, 2, 3.$$

ML-MCMC Example: Subsurface flow

Introduce $\{F_\ell\}_{\ell=0}^{3}$ by discr. of (2) on F.E mesh $K_\ell := (16 \times 2^\ell + 1)^2$ elements $u = \sum_j^{K_\ell} \mathrm{u}_j \phi_j$
$\implies \quad u_{\ell,\ell} = (\mathrm{u}_1, \ldots, \mathrm{u}_{K_\ell})$ Induces

$$\frac{\mathrm{d}\mu_\ell^y}{\mathrm{d}\mu_{\mathrm{pr}}}(u_{\ell,\ell}) = \frac{1}{Z_\ell} \exp\left(-\frac{1}{2}\|y - F_\ell(u_{\ell,\ell})\|_{\mathrm{Y}}^2\right), \quad \dim(\mathrm{X}_\ell) = (16 \times 2^\ell + 1)^2, \quad \ell = 0, 1, 2, 3.$$

IMH:

$$Q_\ell(\cdot) = \mathcal{N}(\mathrm{map}_\ell, \tilde{\mathcal{C}}_\ell)$$

MxC:

$$Q_\ell(u_{\ell,\ell}, \cdot) = \mathcal{N}(\mathrm{map}_\ell + \sqrt{1-\rho^2}(u_{\ell,\ell} - \mathrm{map}_\ell), \rho^2 \tilde{\mathcal{C}}_\ell),$$
$$R_\ell(u_{\ell,\ell-1}, \cdot) = \mathcal{N}(\mathrm{map}_\ell + \sqrt{1-\rho^2}(u_{\ell,\ell-1} - \mathrm{map}_\ell), \rho^2 \tilde{\mathcal{C}}_\ell)$$

ML-MCMC Example: Subsurface flow

Introduce $\{F_\ell\}_{\ell=0}^3$ by discr. of (2) on F.E mesh $K_\ell := (16 \times 2^\ell + 1)^2$ elements $u = \sum_j^{K_\ell} \mathrm{u}_j \phi_j$
$\implies \quad u_{\ell,\ell} = (\mathrm{u}_1, \ldots, \mathrm{u}_{K_\ell})$ Induces

$$\frac{\mathrm{d}\mu_\ell^y}{\mathrm{d}\mu_{\mathrm{pr}}}(u_{\ell,\ell}) = \frac{1}{Z_\ell} \exp\left(-\frac{1}{2}\|y - F_\ell(u_{\ell,\ell})\|_{\mathrm{Y}}^2\right), \quad \dim(\mathrm{X}_\ell) = (16 \times 2^\ell + 1)^2, \quad \ell = 0, 1, 2, 3.$$

IMH:

$$Q_\ell(\cdot) = \mathcal{N}(\mathrm{map}_\ell, \tilde{\mathcal{C}}_\ell)$$

MxC:

$$Q_\ell(u_{\ell,\ell}, \cdot) = \mathcal{N}(\mathrm{map}_\ell + \sqrt{1-\rho^2}(u_{\ell,\ell} - \mathrm{map}_\ell), \rho^2\tilde{\mathcal{C}}_\ell),$$
$$R_\ell(u_{\ell,\ell-1}, \cdot) = \mathcal{N}(\mathrm{map}_\ell + \sqrt{1-\rho^2}(u_{\ell,\ell-1} - \mathrm{map}_\ell), \rho^2\tilde{\mathcal{C}}_\ell)$$

ML-MCMC Example: Subsurface flow

Introduce $\{F_\ell\}_{\ell=0}^3$ by discr. of (2) on F.E mesh $K_\ell := (16 \times 2^\ell + 1)^2$ elements $u = \sum_j^{K_\ell} \mathrm{u}_j \phi_j$

$\implies$ $u_{\ell,\ell} = (\mathrm{u}_1, \ldots, \mathrm{u}_{K_\ell})$ Induces

$$\frac{d\mu_\ell^y}{d\mu_{\mathrm{pr}}}(u_{\ell,\ell}) = \frac{1}{Z_\ell} \exp\left(-\frac{1}{2}\|y - F_\ell(u_{\ell,\ell})\|_Y^2\right), \quad \dim(X_\ell) = (16 \times 2^\ell + 1)^2, \quad \ell = 0, 1, 2, 3.$$

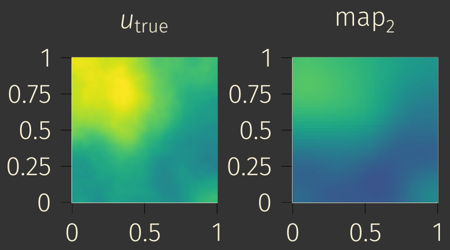
IMH:

$$Q_\ell(\cdot) = \mathcal{N}(\mathrm{map}_\ell, \tilde{\mathcal{C}}_\ell)$$

MxC:

$$Q_\ell(u_{\ell,\ell}, \cdot) = \mathcal{N}(\mathrm{map}_\ell + \sqrt{1-\rho^2}(u_{\ell,\ell} - \mathrm{map}_\ell), \rho^2 \tilde{\mathcal{C}}_\ell),$$

$$R_\ell(u_{\ell,\ell-1}, \cdot) = \mathcal{N}(\mathrm{map}_\ell + \sqrt{1-\rho^2}(u_{\ell,\ell-1} - \mathrm{map}_\ell), \rho^2 \tilde{\mathcal{C}}_\ell)$$

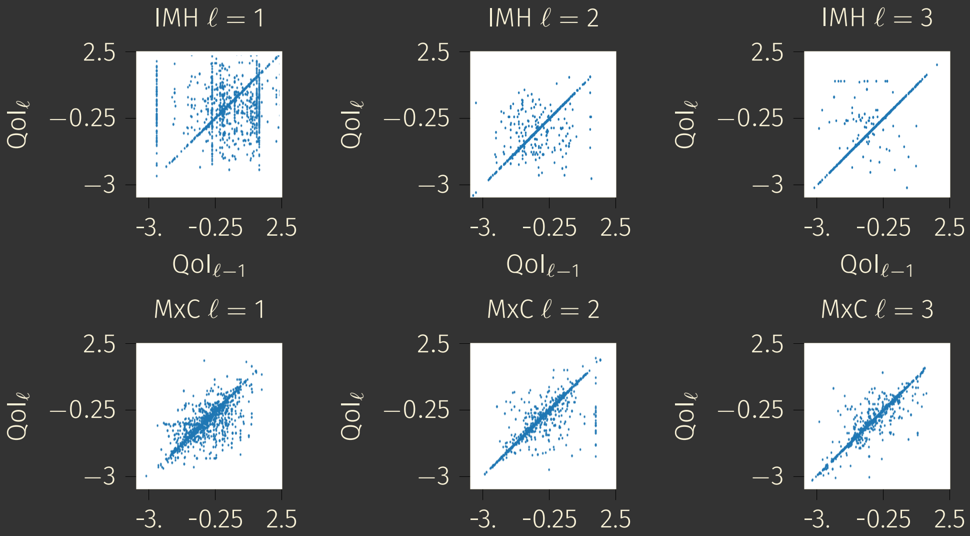


ML-MCMC Example: Subsurface flow –concentration

Verification: run ML-MCMC algs. 50 ind. times w/ $N_\ell = [5000, 5000, 5000, 2000]$ each.

ML-MCMC Example: Subsurface flow –concentration

Verification: run ML-MCMC algs. 50 ind. times w/ $N_\ell = [5000, 5000, 5000, 2000]$ each.

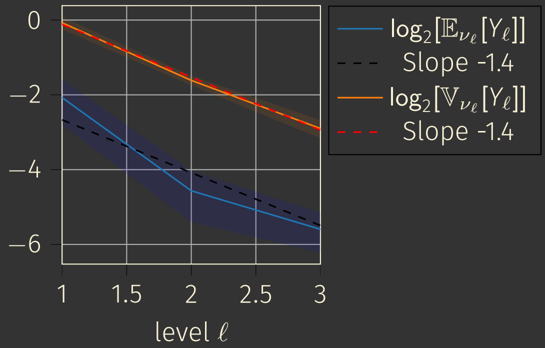


ML-MCMC Example: Subsurface flow –rates

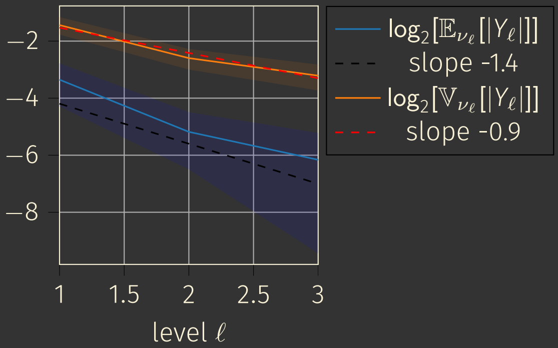
Verification: run ML-MCMC algs. 50 ind. times w/ $N_\ell = [5000, 5000, 5000, 2000]$ each.

Verification: run ML-MCMC algs. 50 ind. times w/ $N_\ell = [5000, 5000, 5000, 2000]$ each.

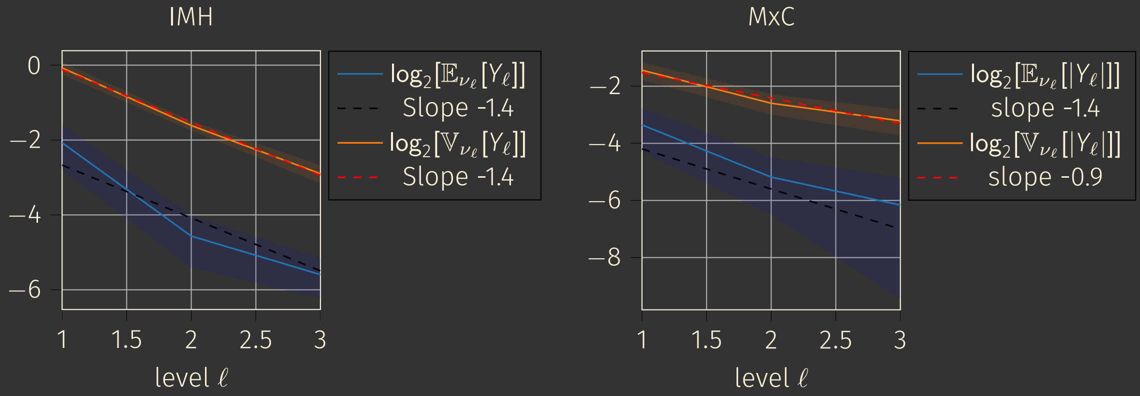
IMH



MxC



Verification: run ML-MCMC algs. 50 ind. times w/ $N_\ell = [5000, 5000, 5000, 2000]$ each.



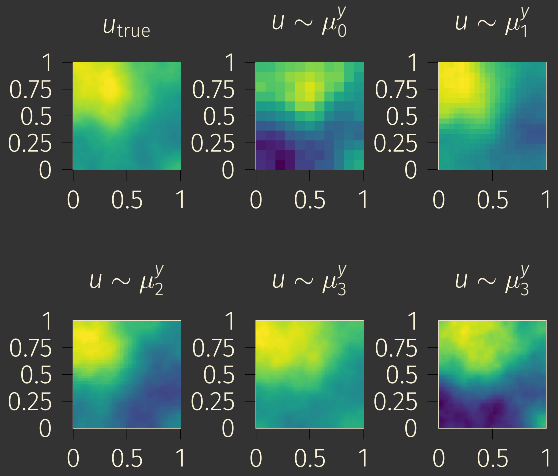
pay a price for applicability of MxC

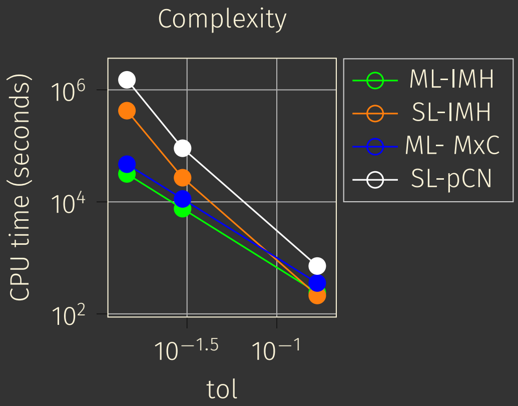
ML-MCMC Example: Subsurface flow –costs

By estimating rates we can estimate complexity:

ML-MCMC Example: Subsurface flow –costs

By estimating rates we can estimate complexity:





ML-MCMC

1. Proposed methods can be understood as generalizations of current ML-MCMC techniques.
   1.1 IMH: First extension of SS approach to allow wider class of proposals.
   1.2 MxC: ML-MCMC allowing state-dependent proposals.
2. Theoretical analysis for convergence and cost of both proposed methods, MxC can be refined.
3. Identification of cases on when should each method be used.
4. MxC easy to implement but lower perfomance than SS or good IMH (when applicable).

Perspectives/open problems

- Do results extend (trivially) to other, more recent methods? (c.f. Dodwell's talk)

Perspectives/open problems

- Do results extend (trivially) to other, more recent methods? (c.f. Dodwell's talk)
- Normalizing flows/Transport maps + ML-MCMC: If $\mu_\ell^y = T_\sharp \mu_{\ell-1}^y$ set $u_{\ell,\ell-1} \sim \mu_{\ell-1}^y, u_{\ell,\ell} = T_\sharp(u_{\ell,\ell-1})$.

Perspectives/open problems

- Do results extend (trivially) to other, more recent methods? (c.f. Dodwell's talk)
- Normalizing flows/Transport maps + ML-MCMC: If $\mu_\ell^y = T_\sharp \mu_{\ell-1}^y$ set $u_{\ell,\ell-1} \sim \mu_{\ell-1}^y, u_{\ell,\ell} = T_\sharp(u_{\ell,\ell-1})$.
- MF/MI-MCMC?

Perspectives/open problems

- Do results extend (trivially) to other, more recent methods? (c.f. Dodwell's talk)
- Normalizing flows/Transport maps + ML-MCMC: If $\mu_\ell^y = T_\sharp \mu_{\ell-1}^y$ set $u_{\ell,\ell-1} \sim \mu_{\ell-1}^y, u_{\ell,\ell} = T_\sharp(u_{\ell,\ell-1})$.
- MF/MI-MCMC?
- More efficient maximal couplings: HMC, max-coup of kernels. Could one generate a max-coup of $\mu_{\ell-1}^y, \mu_\ell^y$? Max-couplings in other norms? rejoining chains?

Perspectives/open problems

- Do results extend (trivially) to other, more recent methods? (c.f. Dodwell's talk)
- Normalizing flows/Transport maps + ML-MCMC: If $\mu_\ell^y = T_\sharp \mu_{\ell-1}^y$ set $u_{\ell,\ell-1} \sim \mu_{\ell-1}^y, u_{\ell,\ell} = T_\sharp(u_{\ell,\ell-1})$.
- MF/MI-MCMC?
- More efficient maximal couplings: HMC, max-coup of kernels. Could one generate a max-coup of $\mu_{\ell-1}^y, \mu_\ell^y$? Max-couplings in other norms? rejoining chains?
- combine ML-MCMC with tempering for multi-modal problems.

References

MCNT21. Madrigal-Cianci, Nobile and Tempone
Analysis of a class of Multi-Level Markov Chain Monte Carlo algorithms based on Independent Metropolis-Hastings
arXiv preprint arXiv:2105.02035, 2021. (in review).

MC22. Madrigal-Cianci
Hierarchical Markov Chain Monte Carlo Methods for Bayesian Inverse Problems
Doctoral Thesis, EPFL, 2022.

HJLT21. Heng, Jasra, Law, and Tarakanov
On unbiased estimation for discretized models.
arXiv preprint arXiv:2102.12230 (2021).

DKST15. Dodwell, Ketelsen, Scheichl and Teckentrup
A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow
SIAM/ASA Journal on Uncertainty Quantification, 2015.

References

MCN22. Madrigal-Cianci and Nobile
Multi-Level Markov Chain Monte Carlo based on maximally coupled proposals
In preparation, 2022.

JOA19. Jacob, O'Leary and Atchadé
Unbiased Markov chain Monte Carlo methods with couplings
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2019.

Rud12. Rudolf, Daniel
Explicit error bounds for Markov chain Monte Carlo
Dissertationes Mathematicae 485 (2012), 1-93.

Thanks!

Sanity check: Gaussians

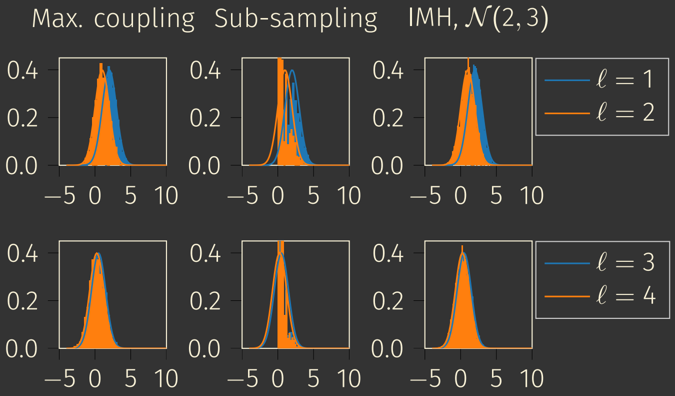
$\mu_\ell^y = \mathcal{N}\left(2^{-\ell+2}, 1\right), \ell = 0, 1, \ldots, \mathsf{L}, \mathsf{QoI}_\ell(u_{\ell,\ell}) = u_{\ell,\ell}.$

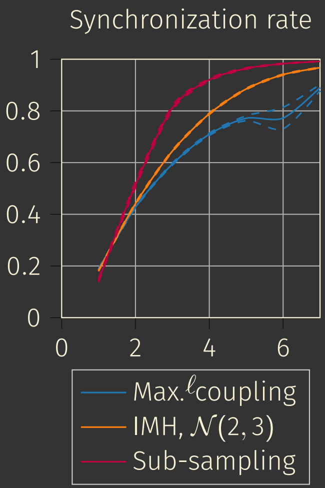
Sanity check: Gaussians

$\mu_\ell^y = \mathcal{N}\left(2^{-\ell+2}, 1\right)$, $\ell = 0, 1, \ldots, \mathrm{L}$, $\mathrm{QoI}_\ell(u_{\ell,\ell}) = u_{\ell,\ell}$.100 ind. runs, $N_\ell = 50,000$, $\ell = 0, \ldots, 7$.

Sanity check: Gaussians

$\mu_\ell^y = \mathcal{N}\left(2^{-\ell+2}, 1\right)$, $\ell = 0, 1, \ldots, \mathrm{L}$, $\mathrm{QoI}_\ell(u_{\ell,\ell}) = u_{\ell,\ell}$.100 ind. runs, $N_\ell = 50,000$, $\ell = 0, \ldots, 7$.



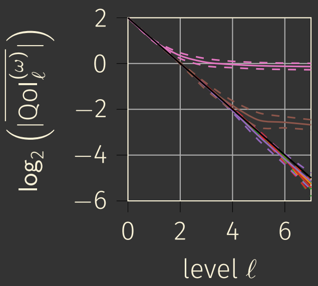


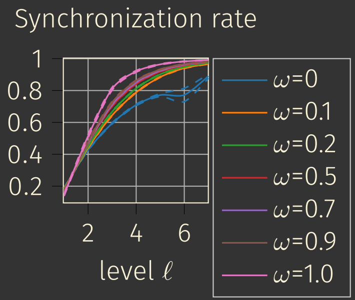
Sanity check: Gaussians

$\mathrm{QoI}_\ell = u_\ell \implies \mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] = 2^{-\ell+2}$

$\mathbf{p}_{c,\ell} = (1-\omega)\mathbf{p}_{\ell,MxC} + \omega\mathbf{p}_{\ell,IMH}$, $\omega \in (0,1)$

Using the mixed kernel with the SS of [1], we obtain for different values of $\omega \in (0,1)$





1: Dodwell, et Al. 2015.

Examples: Subsurface flow

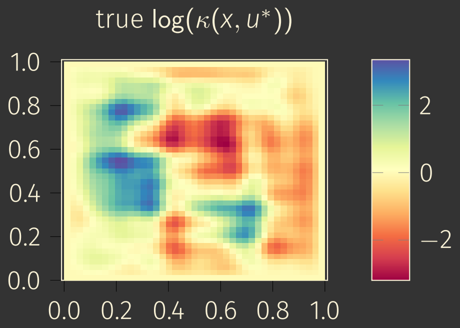
Let $\bar{D} = [0,1]^2$, $\partial D = \Gamma_N \cup \Gamma_D$, $\mathring{\Gamma}_N \cap \mathring{\Gamma}_D = \emptyset$, $\Gamma_D := \{(x_1, x_2) \in \partial D,\ \text{s.t. } x_1 = \{0,1\}\}$.

$$\begin{cases} -\nabla_x \cdot (\kappa(x,u)\nabla_x u(x,u)) = 1, & x \in D,\ u \in X, \\ u(x,u) = 0 & x \in \Gamma_D,\ u \in X,\ , \\ \partial_n u(x,u) = 0 & x \in \Gamma_N,\ u \in X, \end{cases} \tag{3}$$

$$\log(\kappa(x,u)) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \sqrt{\lambda_{m,n}}\phi_{m,n}(x)u_{m,n}, \quad u_{n,m} \sim \mathcal{N}(0,1)$$

$\lambda_{m,n} = \frac{1}{\pi m^2}\frac{1}{\pi n^2}$,
$\phi_{m,n}(x) = \sin(m\pi x_1)\sin(n\pi x_2),\ m,n \in \mathbb{N}$.

$$\mathrm{QoI}(u) := \ln\left(-\int_{\Gamma_{D_1}} \kappa(x,u)\nabla u(x,u)\cdot \mathbf{n}\ \mathrm{d}s\right),$$



Goal: Estimate $\mathbb{E}_{\mu^y}[\mathrm{QoI}]$.

Examples: Subsurface flow: discretization and setup

1. Reorder $\log(\kappa(x,u))$ in terms of a single index j s.t $\lambda_{m,n} =: \lambda_j > \lambda_{j+1}$, and truncate at K:

$$\log(\kappa(x,u)) \approx \sum_{j=1}^{K} \sqrt{\lambda_j}\phi_j(x)u_j, \quad u_j \sim N(0,1),\ j = 1,\dots,K. \tag{4}$$

Examples: Subsurface flow: discretization and setup

1. Reorder $\log(\kappa(x,u))$ in terms of a single index j s.t $\lambda_{m,n} =: \lambda_j > \lambda_{j+1}$, and truncate at K:

$$\log(\kappa(x,u)) \approx \sum_{j=1}^{K} \sqrt{\lambda_j}\phi_j(x)u_j, \quad u_j \sim N(0,1),\ j = 1, \dots, K. \tag{4}$$

2. We introduce $\ell = \{0,1,2,3,4\}$, For each ℓ, the solution to Equation (3) is numerically approximated using FE on a mesh of $2^\ell \cdot 22 \times 2^\ell \cdot 22$ piece-wise linear elements.

Examples: Subsurface flow: discretization and setup

1. Reorder $\log(\kappa(x,u))$ in terms of a single index j s.t $\lambda_{m,n} =: \lambda_j > \lambda_{j+1}$, and truncate at K:

$$\log(\kappa(x,u)) \approx \sum_{j=1}^{K} \sqrt{\lambda_j}\phi_j(x)u_j, \quad u_j \sim N(0,1),\ j = 1,\ldots,K. \tag{4}$$

2. We introduce $\ell = \{0,1,2,3,4\}$, For each ℓ, the solution to Equation (3) is numerically approximated using FE on a mesh of $2^\ell \cdot 22 \times 2^\ell \cdot 22$ piece-wise linear elements.
3. Data y is obtained by solving PDE using a discretization level $\mathrm{L}^* = 6$, with $u_k \sim \mathcal{N}(0,1), k = 1,2,\ldots,K^* = 150$. $\eta \sim \mathcal{N}(0, \sigma^2_{\mathrm{noise}} I_{16\times 16})$

Examples: Subsurface flow: discretization and setup

1. Reorder $\log(\kappa(x,u))$ in terms of a single index j s.t $\lambda_{m,n} =: \lambda_j > \lambda_{j+1}$, and truncate at K:

$$\log(\kappa(x,u)) \approx \sum_{j=1}^{K} \sqrt{\lambda_j}\phi_j(x)u_j, \quad u_j \sim N(0,1),\ j = 1,\dots,K. \tag{4}$$

2. We introduce $\ell = \{0,1,2,3,4\}$, For each ℓ, the solution to Equation (3) is numerically approximated using FE on a mesh of $2^\ell \cdot 22 \times 2^\ell \cdot 22$ piece-wise linear elements.
3. Data y is obtained by solving PDE using a discretization level $\mathrm{L}^* = 6$, with $u_k \sim \mathcal{N}(0,1), k = 1,2,\dots,K^* = 150.$ $\eta \sim \mathcal{N}(0, \sigma^2_{\text{noise}} I_{16\times 16})$
4. Set $K = 50$, $\ell \in \{0,1,2,3,4\}$, $\mathrm{L} = 4$, $u \mapsto F_\ell(u)$: sol PDE at level ℓ, using a log-permeability field modeled by (4), and observed at a grid of 4×4 points inside D $\Phi_\ell(u,y) = \frac{1}{2\sigma^2_{\text{noise}}} \|y - F_\ell(u)\|^2$, 1% noise.

Examples: Subsurface flow: discretization and setup

1. Reorder $\log(\kappa(x,u))$ in terms of a single index j s.t $\lambda_{m,n} =: \lambda_j > \lambda_{j+1}$, and truncate at K:

$$\log(\kappa(x,u)) \approx \sum_{j=1}^{K} \sqrt{\lambda_j}\phi_j(x)u_j, \quad u_j \sim N(0,1),\ j = 1,\dots,K. \tag{4}$$

2. We introduce $\ell = \{0,1,2,3,4\}$, For each ℓ, the solution to Equation (3) is numerically approximated using FE on a mesh of $2^\ell \cdot 22 \times 2^\ell \cdot 22$ piece-wise linear elements.
3. Data y is obtained by solving PDE using a discretization level $\mathrm{L}^* = 6$, with $u_k \sim \mathcal{N}(0,1)$, $k = 1,2,\dots,K^* = 150$. $\eta \sim \mathcal{N}(0, \sigma^2_{\mathrm{noise}} I_{16\times 16})$
4. Set $K = 50$, $\ell \in \{0,1,2,3,4\}$, $\mathrm{L} = 4$, $u \mapsto F_\ell(u)$: sol PDE at level ℓ, using a log-permeability field modeled by (4), and observed at a grid of 4×4 points inside D $\Phi_\ell(u,y) = \frac{1}{2\sigma^2_{\mathrm{noise}}} \|y - F_\ell(u)\|^2$, 1% noise.
5. $\mu_{\mathrm{pr}} = \bigotimes_{i=1}^{K} \mathcal{N}(0,1)$,

Examples: subsurface

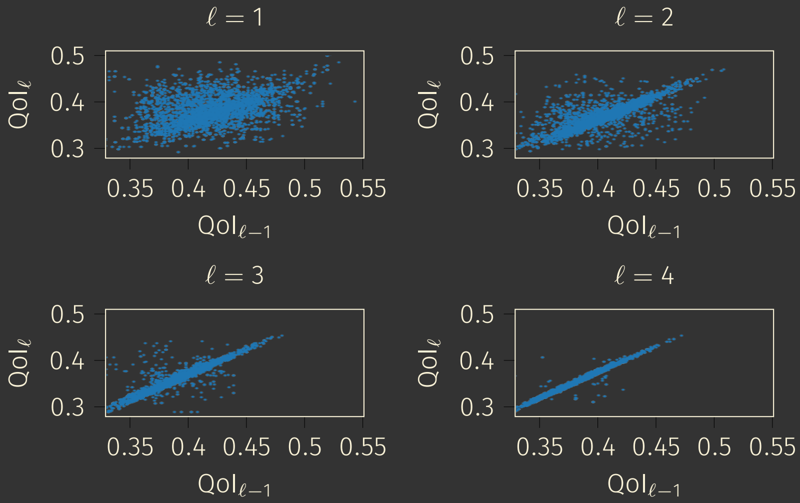
We implement our Maximal Coupling method with ω_ℓ given by $\omega_1 = 0.1$, and $\omega_2 = \omega_3 = \omega_4 = 0.5$.

As a verification of our method, we run our ML-MCMC algorithm 50 independent times, with $N_\ell = [5000, 5000, 5000, 2000, 1000]$ samples per level, per run, for $\ell = 0, 1, \ldots, 4$.

We want to (i) check that chains stay correlated, (ii) verify assumptions form complexity Theorem.

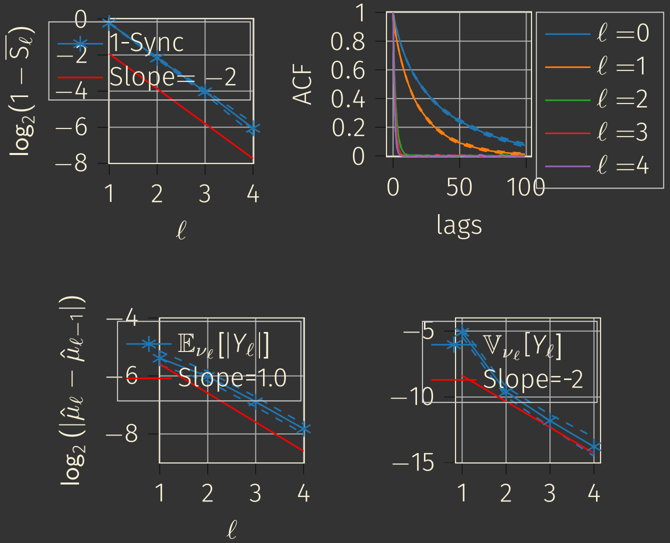
Examples: subsurface: results

Verification that chains stay increasingly more sync'd $\rightarrow$ concentrate around a diagonal



Verification of rates in cost Theorem:

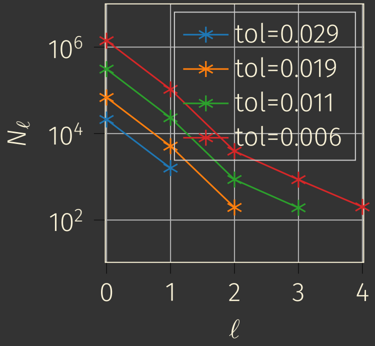
$$|\mathbb{E}_{\mu^y_\ell}[\mathrm{QoI}_\ell] - \mathbb{E}_{\mu^y}[\mathrm{QoI}]| \le C_w M^{-\alpha_w \ell}, \quad \mathbb{V}_{\nu_\ell}[Y_\ell] \le C_v M^{-\beta\ell}.$$

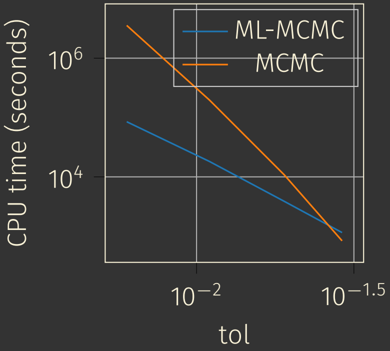


Examples: subsurface: results

$N_\ell = \left\lceil 2\text{tol}^{-2}\sqrt{V_\ell/C_\ell}\left(\sum_{k=0}^{L}\sqrt{V_k C_k}\right)\right\rceil$, Here: $V_\ell = \mathbb{V}_{\nu_\ell}(Y_\ell)$, C_ℓ cost of 1 sample at ℓ.

We can assume that $C_\ell \lessapprox 2^{\gamma\ell}$, $\gamma = 2$

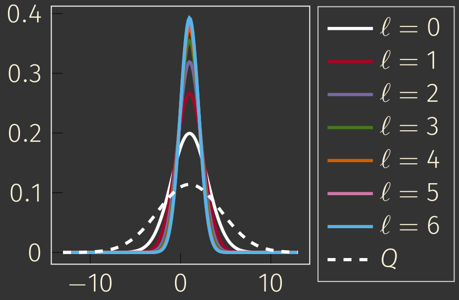




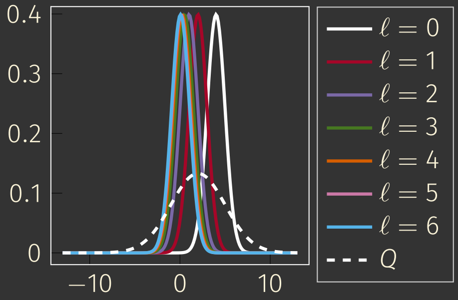
In words: In a single-level we would need $O(10^6)$ PDE solves at the finest level. **With the ML-MCMC approach we need just** $O(10^2)$

Coupling via independent sampler: illustrative example.

Dodwell et. Al. 2015, proposes to sub-sample from the previous chain targeting $\mu^y_{\ell-1} \to \{u^n_{\ell,\ell}\}_{n=0}^{N_\ell} \subset \{u_{0,0}\}_{n=0}^{N_0}$. however this can only work if $\mathrm{supp}(\mu^y_\ell) \subset \mathrm{supp}(\mu^y_{\ell-1})$.



(a) ✓

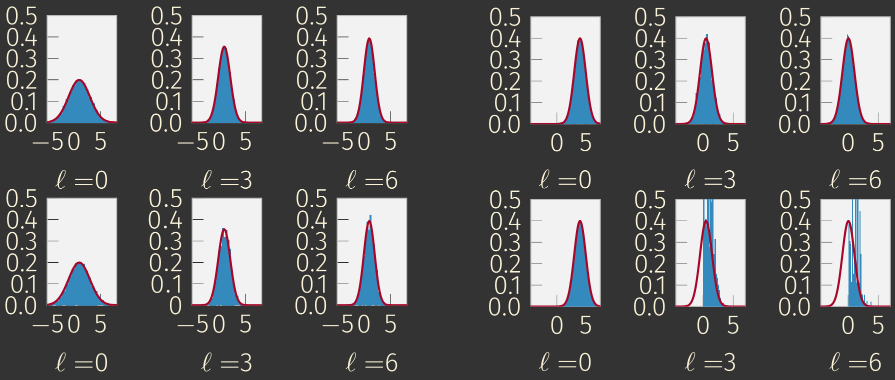


(b) ✗

Coupling via independent sampler: illustrative example.

Dodwell et. Al. 2015, proposes to sub-sample from the previous chain targeting $\mu^y_{\ell-1} \rightarrow \{u^n_{\ell,\ell}\}_{n=0}^{N_\ell} \subset \{u_{0,0}\}_{n=0}^{N_0}$. however this can only work if $\mathrm{supp}(\mu^y_\ell) \subset \mathrm{supp}(\mu^y_{\ell-1})$.

Top: ours. **Bottom:** Dodwell et Al. 2015

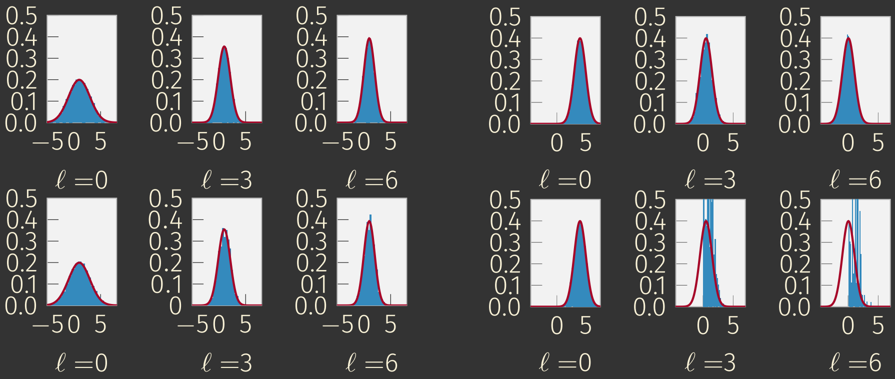


(a) ✓ Both methods work (b) ✗ Only our method works

Coupling via independent sampler: illustrative example.

Dodwell et. Al. 2015, proposes to sub-sample from the previous chain targeting $\mu^y_{\ell-1} \rightarrow \{u^n_{\ell,\ell}\}_{n=0}^{N_\ell} \subset \{u_{0,0}\}_{n=0}^{N_0}$. however this can only work if $\mathrm{supp}(\mu^y_\ell) \subset \mathrm{supp}(\mu^y_{\ell-1})$.

Difficulty finding Q (IMH) suitable for realistic problems $\rightarrow$ Try localized proposals



(a) ✓ Both methods work

(b) ✗Only our method works

Maximal coupling: An example

Sample $u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1}, \cdot)$, and $u'_{\ell,\ell} \sim Q(u^n_{\ell,\ell}, \cdot)$ s.t. $\mathbb{P}(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) = \|Q - R\|_{\mathrm{TV}}$

4: Jacob et Al. 2019.

Maximal coupling: An example

Sample $u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1}, \cdot)$, and $u'_{\ell,\ell} \sim Q(u^n_{\ell,\ell}, \cdot)$ s.t. $\mathbb{P}(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) = \|Q - R\|_{\mathrm{TV}}$

Mixture approach [4]:

1. Case 1: With probability $a = \int \min\{Q, R\} dz$ sample

$$u'_{\ell,\ell} \sim \frac{\min\{Q, R\}}{a}, \quad u'_{\ell,\ell} = u'_{\ell,\ell-1}.$$

2. Case 2: Otherwise

$$u'_{\ell,\ell} \sim \frac{Q - \min\{Q, R\}}{1 - a}, \quad u'_{\ell,\ell-1} \sim \frac{R - \min\{Q, R\}}{1 - a} \text{ independently.}$$

4: Jacob et Al. 2019.

Maximal coupling: An example

Sample $u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1}, \cdot)$, and $u'_{\ell,\ell} \sim Q(u^n_{\ell,\ell}, \cdot)$ s.t. $\mathbb{P}(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) = \|Q - R\|_{\mathrm{TV}}$

Mixture approach [4]:

1. Case 1: With probability $a = \int \min\{Q, R\} \mathrm{d}z$ sample
$$u'_{\ell,\ell} \sim \frac{\min\{Q, R\}}{a}, \quad u'_{\ell,\ell} = u'_{\ell,\ell-1}.$$
2. Case 2: Otherwise
$$u'_{\ell,\ell} \sim \frac{Q - \min\{Q, R\}}{1-a}, \quad u'_{\ell,\ell-1} \sim \frac{R - \min\{Q, R\}}{1-a} \text{ independently.}$$

$u_{\ell,\ell} \sim Q$

$$\begin{aligned}\mathbb{P}(u_{\ell,\ell} \in A) &= a \int_A \frac{\min\{Q, R\}}{a} + (1-a) \int_A \frac{Q - \min\{Q, R\}}{1-a} \\ &= \int_A Q \mathrm{d}z\end{aligned}$$

4: Jacob et Al. 2019.

Maximal coupling: An example

Sample $u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1}, \cdot)$, and $u'_{\ell,\ell} \sim Q(u^n_{\ell,\ell}, \cdot)$ s.t. $\mathbb{P}(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) = \|Q - R\|_{\mathrm{TV}}$

Mixture approach [4]:

1. Case 1: With probability $a = \int \min\{Q, R\}\mathrm{d}z$ sample
$$u'_{\ell,\ell} \sim \frac{\min\{Q, R\}}{a}, \quad u'_{\ell,\ell} = u'_{\ell,\ell-1}.$$
2. Case 2: Otherwise
$$u'_{\ell,\ell} \sim \frac{Q - \min\{Q, R\}}{1 - a}, \quad u'_{\ell,\ell-1} \sim \frac{R - \min\{Q, R\}}{1 - a} \text{ independently.}$$

$u_{\ell,\ell-1} \sim R$

$$\begin{aligned}\mathbb{P}(u_{\ell,\ell-1} \in A) &= a \int_A \frac{\min\{Q, R\}}{a} + (1 - a) \int_A \frac{R - \min\{Q, R\}}{1 - a} \\ &= \int_A R\mathrm{d}z\end{aligned}$$

4: Jacob et Al. 2019.

Maximal coupling: An example

Sample $u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1}, \cdot)$, and $u'_{\ell,\ell} \sim Q(u^n_{\ell,\ell}, \cdot)$ s.t. $\mathbb{P}(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) = \|Q - R\|_{\mathrm{TV}}$

Mixture approach [4]:

1. Case 1: With probability $a = \int \min\{Q, R\}\mathrm{d}z$ sample
$$u'_{\ell,\ell} \sim \frac{\min\{Q, R\}}{a}, \quad u'_{\ell,\ell} = u'_{\ell,\ell-1}.$$
2. Case 2: Otherwise
$$u'_{\ell,\ell} \sim \frac{Q - \min\{Q, R\}}{1-a}, \quad u'_{\ell,\ell-1} \sim \frac{R - \min\{Q, R\}}{1-a} \text{ independently.}$$

$\mathbb{P}(u_{\ell,\ell-1} = u_{\ell,\ell}) = 1 - \|Q - R\|_{\mathrm{TV}}$

$$\begin{aligned}\mathbb{P}(u_{\ell,\ell-1} = u_{\ell,\ell}) &= \int \min\{Q, R\}\mathrm{d}z = \frac{1}{2}\int Q\mathrm{d}z + \frac{1}{2}\int R\mathrm{d}z - \int \frac{1}{2}|Q - R|\mathrm{d}z \\ &= 1 - \frac{1}{2}\int |Q - R|\mathrm{d}z = 1 - \|Q - R\|_{\mathrm{TV}}.\end{aligned}$$

4: Jacob et Al. 2019.

Maximal coupling: An example

Sample $u'_{\ell,\ell-1} \sim R(u^n_{\ell,\ell-1}, \cdot)$, and $u'_{\ell,\ell} \sim Q(u^n_{\ell,\ell}, \cdot)$ s.t. $\mathbb{P}(u'_{\ell,\ell-1} \neq u'_{\ell,\ell}) = \|Q - R\|_{\mathrm{TV}}$

Mixture approach [4]:

1. Case 1: With probability $a = \int \min\{Q, R\}\mathrm{d}z$ sample

$$u'_{\ell,\ell} \sim \frac{\min\{Q, R\}}{a}, \quad u'_{\ell,\ell} = u'_{\ell,\ell-1}.$$

2. Case 2: Otherwise

$$u'_{\ell,\ell} \sim \frac{Q - \min\{Q, R\}}{1 - a}, \quad u'_{\ell,\ell-1} \sim \frac{R - \min\{Q, R\}}{1 - a} \text{ independently.}$$

Basis for more complicated algorithms. Again: Max coupling is not unique.

4: Jacob et Al. 2019.

Reflection Maximal coupling

Create coupling from $\mathcal{N}(m(u^n_{\ell,\ell}),\tilde{\mathcal{C}})$ and $\mathcal{N}(m(u^n_{\ell,\ell-1}),\tilde{\mathcal{C}})$. $\varphi_0 = \mathcal{N}(0,I)$

procedure REFLECTION-COUPLING$(m(u^n_{\ell,\ell}), m(u^n_{\ell,\ell}))$

Set $z^n = m(u^n_{\ell,\ell}) - m(u^n_{\ell,\ell-1})$ and set $e^n = z^n / \|z^n\|_X$.

Sample $\xi \sim \varphi_0$, and $w \sim \mathcal{U}([0,1])$.

if $w \leq \min\left\{1, \frac{\varphi_0(\xi+z^n)}{\varphi_0(\xi)}\right\}$ **then**

Set $\zeta = \xi + z^n$. ▷ case I

else

Set $\zeta = \xi - 2\langle e^n, \xi\rangle_X e^n$. ▷ case II

end if

Set $u'_{\ell,\ell} = m(u^n_{\ell,\ell}) + \tilde{\mathcal{C}}^{1/2}\xi$ and $u'_{\ell,\ell-1} = m(u^n_{\ell,\ell-1}) + \tilde{\mathcal{C}}^{1/2}\zeta$

Output $(u'_{\ell,\ell-1}, u'_{\ell,\ell})$.

end procedure