Practical unbiased Monte Carlo for Uncertainty Quantification

Sergios Agapiou

Department of Statistics, University of Warwick

MiR@W day: Uncertainty in Complex Computer Models,
2nd February 2015, University of Warwick

Enabling Quantification of
EQUIP
Uncertainty for Inverse Problems

http://www.sergiosagapiou.com/

- S. Agapiou, G. O. Roberts and S. J. Vollmer, *Unbiased Monte Carlo: posterior estimation for intractable/infinite dimensional models*, http://arxiv.org/abs/1411.7713

- C. H. Rhee, *Unbiased estimation with biased samples*, PhD thesis, Stanford University, 2013 (supervisor P. W. Glynn).

Outline

Problem

Want to estimate expectations of functions f wrt an intractable measure μ,

$$\mathbb{E}_\mu[f] := \mathbb{E}_\mu[f(\cdot)].$$

- Would like to use Monte Carlo estimator: for $X^{(m)} \overset{iid}{\sim} \mu$ let

$$Y_M := \frac{1}{M}\sum_{m=1}^{M} f(X^{(m)}).$$

For all M

$$\mathbb{E}[Y_M] = \mathbb{E}_\mu[f] \qquad (Y_M \text{ unbiased})$$

and

$$Y_M \xrightarrow{M} \mathbb{E}_\mu[f], \text{ almost surely} \qquad (Y_m \text{ consistent})$$

- Intractability of μ forces the use of approximations μ_i introducing bias.

Debiasing idea - John von Neumann, Stanislaw Ulam

- We study unbiased estimation of $\mathbb{E}_\mu[f]$ using biased samples.
- Assume $\mathbb{E}_{\mu_i}[f] \xrightarrow{i} \mathbb{E}_\mu[f]$.
- Let $X_i \sim \mu_i$ and define $\Delta_i := f(X_i) - f(X_{i-1})$.
- If Fubini applies

$$\mathbb{E}_\mu[f] = \sum_{i=1}^{\infty}(\mathbb{E}_{\mu_i}[f] - \mathbb{E}_{\mu_{i-1}}[f]) = \sum_{i=1}^{\infty}\mathbb{E}\Delta_i \stackrel{?}{=} \mathbb{E}\sum_{i=1}^{\infty}\Delta_i.$$

- $\sum_{i=1}^{\infty}\Delta_i$ is unbiased but requires infinite computing time.

Debiasing idea - John von Neumann, Stanislaw Ulam

$$Z := \sum_{i=0}^{N} \frac{\Delta_i}{\mathbb{P}(N \geq i)},$$

N integer-valued r.v. independent of Δ_i, s.t. $\mathbb{P}(N \geq i) > 0, \forall i$.

- If Fubini applies then Z unbiased

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i=0}^{\infty} \frac{\mathbb{1}_{\{N \geq i\}} \Delta_i}{\mathbb{P}(N \geq i)}\right] \stackrel{?}{=} \sum_{i=0}^{\infty} \frac{\mathbb{E}[\mathbb{1}_{\{N \geq i\}} \Delta_i]}{\mathbb{P}(N \geq i)} = \sum_{i=0}^{\infty} \mathbb{E}\Delta_i = \mathbb{E}_\mu[f].$$

- To be practical, Z needs to have finite variance and finite expected computing time.

Unbiasing theory of Glynn and Rhee

Proposition (GR13)

Assume

$$\sum_{i\le\ell} \frac{||\Delta_i||_2 ||\Delta_\ell||_2}{\mathbb{P}(N \ge i)} < \infty.$$

Then $Z := \sum_{i=0}^{N} \frac{\Delta_i}{\mathbb{P}(N\ge i)}$ is an unbiased estimator for $\mathbb{E}_\mu[f]$ with finite variance.

Can use $\tilde{\Delta}_i$ copy of Δ_i s.t. $\{\tilde{\Delta}_i\}$ mutually independent.

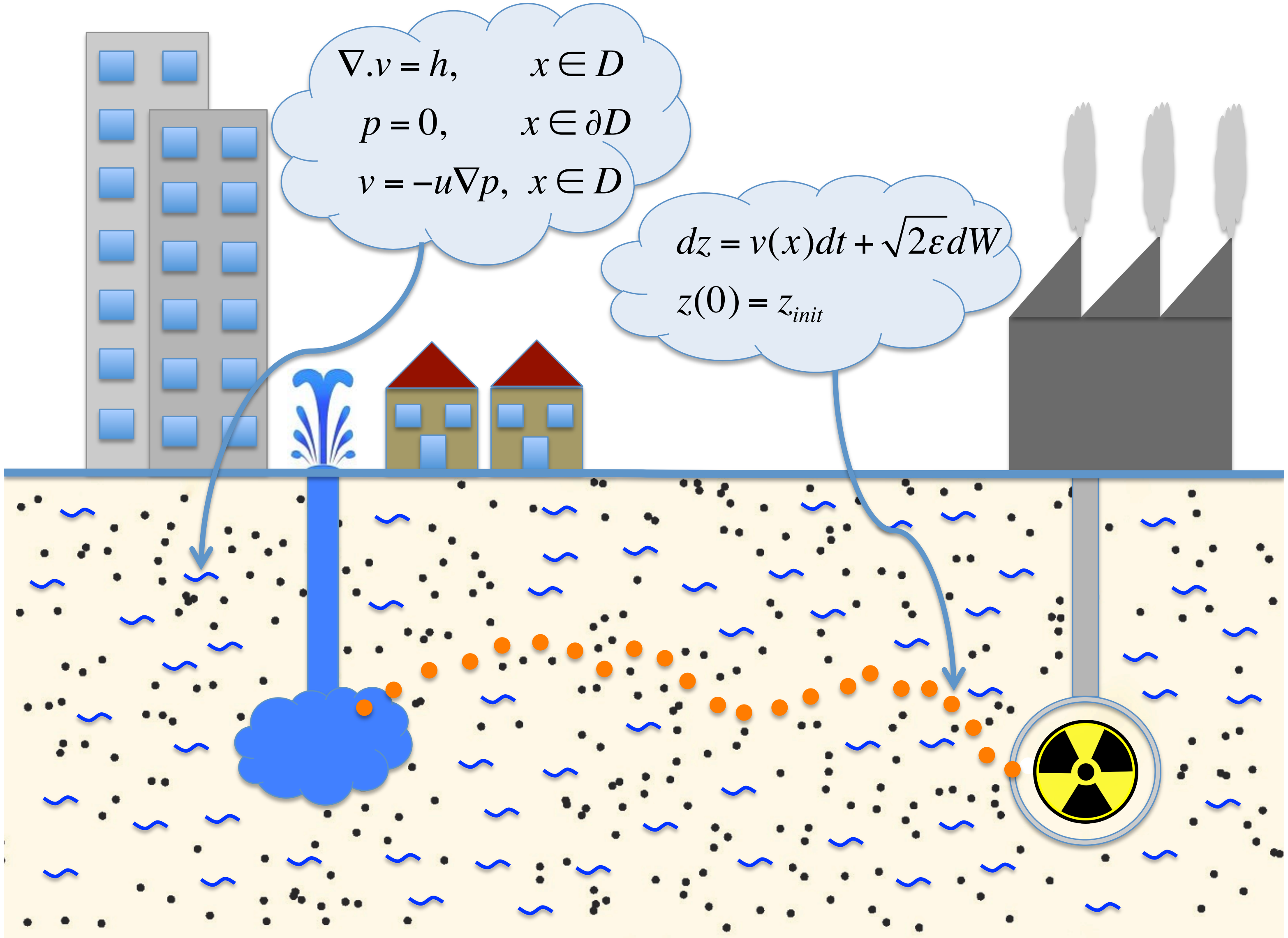
- t_i expected cost of generating Δ_i. Expected computing time of Z

$$\mathbb{E}(\tau) = \mathbb{E}\sum_{i=0}^{N} t_i = \sum_{i=0}^{\infty} t_i \mathbb{P}(N \ge i).$$

- To be possible to choose $\mathbb{P}(N \ge i)$ s.t. Z practical, suffices to generate Δ_i's with correct expectation s.t. $||\Delta_i||_2^2$ decays sufficiently faster than t_i blows-up.

Example - Contamination scenario

- u permeability field
- p pressure
- v Darcy velocity
- $p = G(u)$



Quantity of interest: $f(u) = \mathbb{E}[\inf_{t \geq 0}\{|z(t)| > R\}]$

Example - UQ in contamination scenario

Permeability field u unknown, have prior information $u \sim \mu_0$.

- **Vanilla-UQ**: probe $\mu_0 \circ f^{-1}$, e.g. estimate $\mathbb{E}_{\mu_0}[f(u)]$.
- Have noisy indirect measurements of pressure: data model in $\mathbb{R}^J$

$$y = \mathcal{G}(u) + \eta, \ \eta \sim N(0, \Gamma).$$

Formulate Bayesian inverse problem (see DS13), μ^y posterior on $u|y$

$$\frac{d\mu^y}{d\mu_0}(u; y) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}(u)\|^2\right).$$

BIP-UQ: probe $\mu^y \circ f^{-1}$, e.g. estimate $\mathbb{E}_{\mu^y}[f(u)]$.

- μ_0 is ∞-dim, needs to be approximated by $\mu_{0,i}$ in $\mathbb{R}^i$ introducing discretization bias (ARV14).
- cannot sample μ^y directly, construct Markov chain targeting μ^y, use finite-time distributions $\mu^{y,k}$ burn-in time issues (GR13, ARV14).
- to implement in computer construct Markov chain targeting approximation μ_i^y in $\mathbb{R}^i$, use finite-time distributions $\mu_i^{y,k}$ introducing discretization bias and burn-in time issues (ARV14).

Removing discretization bias

- $\mathcal{X} = L^2[0,1]$, $\{\varphi_\ell\}$ complete orthonormal basis.
- μ Gaussian measure in $\mathcal{X}$ given via the Karhunen-Loeve expansion:

$$\mu = \mathcal{L}\left(\sum_{\ell=1}^{\infty} \ell^{-a}\xi_\ell\varphi_\ell\right), \qquad \xi_\ell \overset{iid}{\sim} N(0,1), \quad a > \frac{1}{2}.$$

- To estimate $\mathbb{E}_\mu[f]$, need to truncate introducing discretization bias in MC estimators. (Vanilla-UQ example)

> Aim: unbiasedly estimate $\mathbb{E}_\mu[f]$ in finite time.

- Approximations $\mu_i = \mathcal{L}\left(\sum_{\ell=1}^{j_i} \ell^{-a}\xi_\ell\varphi_\ell\right)$, $\{j_i\}$ increasing.
- $\Delta_i = f(u_i) - f(u_{i-1})$, $u_i \sim \mu_i$.

Removing discretization bias

Theorem 1 (ARV14)

Assume $a > 1$ and f Lipschitz. Then $\exists$ choices j_i and $\mathbb{P}(N \geq i)$, s.t. $Z = \sum_{i=1}^{N} \frac{\Delta_i}{\mathbb{P}(N \geq i)}$ is unbiased estimator of $\mathbb{E}_\mu[f]$ with finite variance and finite expected computing time.

Proof.

- Consider $j_i = 2^i$. Use Proposition.
- Cost of Δ_i, $t_i = \mathcal{O}(j_i) = \mathcal{O}(2^i)$ $\quad$ ($\#$ $N(0,1)$ draws).
- Bound

$$\|\Delta_i\|_2^2 = \mathbb{E}(|f(u_i) - f(u_{i-1})|^2) \leq \|f'\|_\infty^2 \mathbb{E}(\|u_i - u_{i-1}\|^2) = \mathcal{O}(2^{i(1-2a)}).$$

- $\|\Delta_i\|_2^2$ decays sufficiently faster than t_i blows-up.
- Can choose $\mathbb{P}(N \geq i)$ s.t. $\mathbb{E}(\tau), \mathrm{Var}(Z) < \infty$.

□

Removing burn-in time bias

- $\mathcal{X}$ general state space, d distance in $\mathcal{X}$, f d-Lipschitz.
- Measure μ intractable, cannot be sampled directly but can construct $\mathbb{X} = (X_n)_{n\in\mathbb{N}}$ Markov chain with stationary distribution μ.
- $\{a_i\}$ increasing sequence of positive integers.
- To estimate $\mathbb{E}_\mu[f]$, use finite-time distributions $\mu_i = \mathcal{L}(X_{a_i})$ introducing burn-in issues.

Aim: unbiasedly estimate $\mathbb{E}_\mu[f]$ in finite time.

Removing burn-in time bias

- Weak convergence of μ_i not enough to get convergence of Δ_i.
- Contracting coupling assumption: we can simultaneously generate chains started at different states s.t. they come together in d geometrically quickly.
- Use top level chain $\mathcal{T}_\cdot^i$ running for a_i steps and bottom level chain $\mathcal{B}_\cdot^i$ running for a_{i-1} steps, coupled as follows:

$$
\begin{array}{cccccccccc}
 & & x_0 = & \mathcal{B}^i_{-a_{i-1}} & \dots & \mathcal{B}^i_{-a_0} & \dots & \mathcal{B}^i_0 & & \\
 & & & | & | & | & | & | & \}\Delta_i = f(\mathcal{T}^i_0) - f(\mathcal{B}^i_0) & \\
x_0 = & \mathcal{T}^i_{-a_i} & \dots & \mathcal{T}^i_{-a_{i-1}} & \dots & & \dots & \mathcal{T}^i_0 & &
\end{array}
$$

Removing burn-in time bias

Theorem 2 (ARV14)

$\exists$ choices a_i and $\mathbb{P}(N \geq i)$, s.t. $Z = \sum_{i=1}^{N} \frac{\Delta_i}{\mathbb{P}(N \geq i)}$ is unbiased estimator of $\mathbb{E}_\mu[f]$ with finite variance and finite expected computing time.

Proof.

- Use Proposition.
- Using assumptions, can show $\|\Delta_i\|_2^2 \leq \|f'\|_\infty^2 \mathbb{E} d^2 \left(\mathcal{T}_0^i, \mathcal{B}_0^i\right) \leq c r^{a_i}$.
- Cost of Δ_i, $t_i = \mathcal{O}(a_i)$ (# steps).
- $\|\Delta_i\|_2^2$ decays sufficiently faster than t_i blows-up.
- Can choose $\mathbb{P}(N \geq i)$ s.t. $\mathbb{E}(\tau), \text{Var}(Z) < \infty$.

$\square$

UE for BIP-UQ in function space

- Combining can perform UE of $\mathbb{E}_\mu[f]$ for μ both ∞-dim and only accessible in the limit of a Markov chain (BIP-UQ example).
- Approximation using finite-time distributions and discretizing space: top chain $\mathcal{T}^i_\cdot$ more steps and higher discretization level than bottom chain $\mathcal{B}^i_\cdot$

$$
\begin{array}{lcccccccl}
j_{i-1}: & & x_0 = \mathcal{B}^i_{-a_{i-1}} & \dots & \mathcal{B}^i_{-a_0} & \dots & \mathcal{B}^i_0 & & \\
 & & | & | & | & | & | & \}\Delta_i = f(\mathcal{T}^i_0) - f(\mathcal{B}^i_0) & \\
j_i: \; x_0 = & \mathcal{T}^i_{-a_i} \;\dots & \mathcal{T}^i_{-a_{i-1}} & \dots & & \dots & \mathcal{T}^i_0 & &
\end{array}
$$

- In ARV14, achieve this:
  1. in non-linear BIPs (e.g. groundwater flow example) with uniform priors, using independence sampler;
  2. for targets μ which have Lipschitz log-density wrt Gaussian, using pCN algorithm.
     (MH with proposal $X_{k+1} = \lambda X_k + \sqrt{1-\lambda^2}\xi$)

Comparison of Unbiased Estimator vs Ergodic Average

- 1d Gaussian autoregression

$$X_{n+1} = \rho X_n + \sqrt{1-\rho^2}\, \xi_{n+1},$$

  $\rho \in (0,1)$, ξ_n i.i.d. $N(0,1)$.

- Ergodic with invariant distribution $\mu = N(0,1)$. Estimate $\mathbb{E}_\mu[\mathrm{Id}] = 0$.
- Compare MSE-work product of Monte Carlo estimator based on UE vs EA.
- For EA

$$\lim_{n\to\infty} \text{MSE-work} = \frac{1+\rho}{1-\rho} T_{\text{step}}.$$

- For UE have non-asymptotic expression for MSE-work product, depending on a_i and $\mathbb{P}(N \geq i)$. Optimize by minimizing wrt a_i and $\mathbb{P}(N \geq i)$: hard!

Comparison of Unbiased Estimator vs Ergodic Average

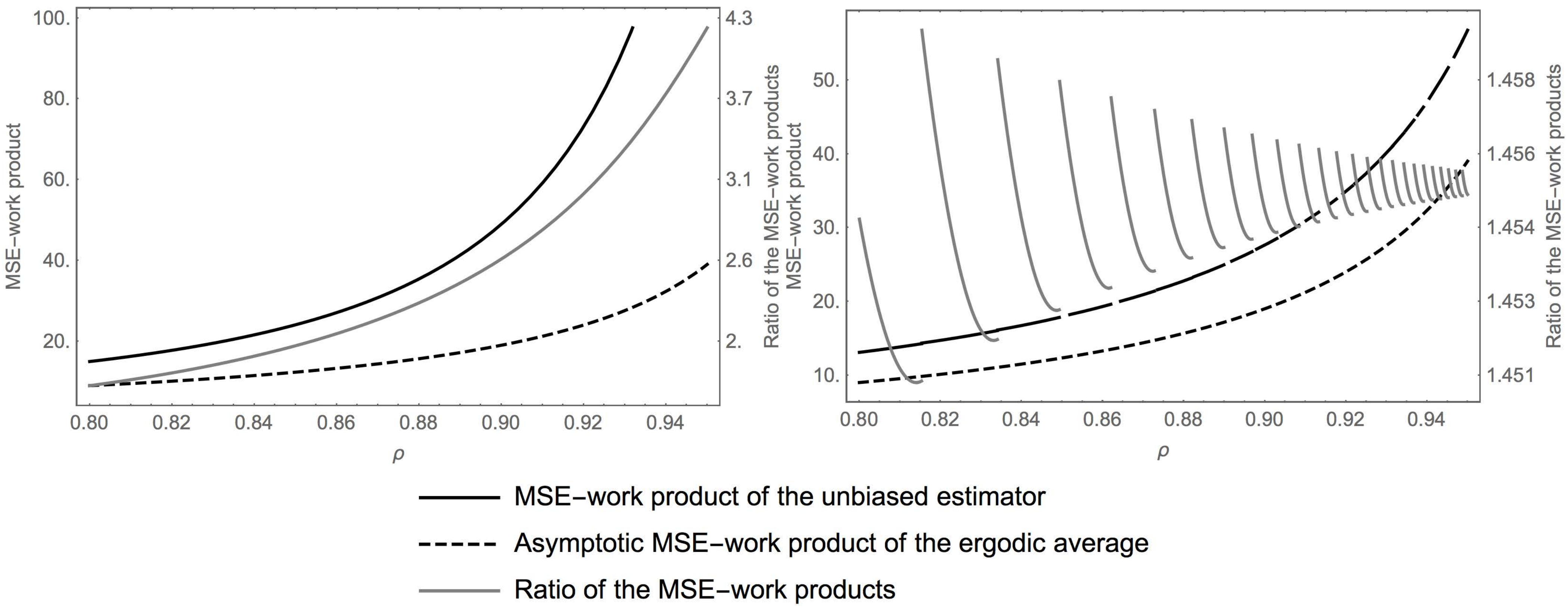


Figure: 1) Left: optimized $\mathbb{P}(N \geq i)$ for fixed $a_i = 4(i+1)$ (as in GR13), 2) Right: optimized $\mathbb{P}(N \geq i)$ and a_i over subclass $a_i = m(i+1)$.

Conclusions - further work

- UE is often feasible.
- Optimization wrt parameters is crucial especially in function space setting.
- UE easily parallelizable: a) use independent copies of Z, b) Δ_i's independent.
- UE seems competitive. Looking forward to comparisons in problems of higher complexity (e.g. BIP-UQ example).

http://www.sergiosagapiou.com/

- S. Agapiou, G. O. Roberts and S. J. Vollmer, *Unbiased Monte Carlo: posterior estimation for intractable/infinite dimensional models*, arXiv:1411.7713
- C. H. Rhee, *Unbiased estimation with biased samples*, PhD thesis, Stanford University, 2013, (supervisor P. W. Glynn).
- J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms, 1996.
- M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, arXiv:1302.6989.
- M. Hairer, A. M. Stuart and S. J. Vollmer *Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions*, The Annals of Applied Probability, 2014.