Multilevel sequential monte carlo samplers for normalizing constants

DOI:
10.1145/3092841

Document Version
Accepted author manuscript

Citation for published version (APA):
Moral, P. D., Jasra, A., Law, K. J. H., & Zhou, Y. (2017). Multilevel sequential monte carlo samplers for normalizing constants. *ACM Transactions on Modeling and Computer Simulation*, *27*(3), Article 20. https://doi.org/10.1145/3092841

Published in:
ACM Transactions on Modeling and Computer Simulation

Multilevel Sequential Monte Carlo Samplers for Normalizing Constants

Pierre Del Moral* Ajay Jasra† Kody J. H. Law‡
Yan Zhou§

March 13, 2017

Abstract

This article considers the sequential Monte Carlo (SMC) approximation of ratios of normalizing constants associated to posterior distributions which in principle rely on continuum models. Therefore, the Monte Carlo estimation error and the discrete approximation error must be balanced. A multilevel strategy is utilized to substantially reduce the cost to obtain a given error level in the approximation as compared to standard estimators. Two estimators are considered and relative variance bounds are given. The theoretical results are numerically illustrated for two Bayesian inverse problems arising from elliptic partial differential equations (PDEs). The examples involve the inversion of observations of the solution of (i) a 1-dimensional Poisson equation to infer the diffusion coefficient, and (ii) a 2-dimensional Poisson equation to infer the external forcing.

Key words: Multilevel Monte Carlo, Sequential Monte Carlo, Bayesian Inverse Problems.
AMS subject classification: 82C80, 60K35.

1 Introduction

Over the past decades there has been an explosion of interest in accounting for uncertainty in the simulation of systems in science and engineering applications which are governed by continuum limiting systems such as partial differential equations (PDEs) [21, 24].

*Center INRIA Bordeaux Sud-Ouest & Institut de Mathematiques de Bordeaux, Universite de Bordeaux I, 33405, FR
†Department of Statistics & Applied Probability, National University of Singapore, Singapore, 117546, SG
‡Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, 37934 TN, USA
§Department of Statistics & Applied Probability, National University of Singapore, Singapore, 117546, SG

Consider a sequence of probability measures $\{\eta_l\}_{l\geq 0}$ on a common measurable space $(E, \mathcal{E})$; assume that the probabilities have common dominating $\sigma-$finite-measure du. In particular, for some known $\kappa_l : E \to \mathbb{R}^+$, let

$$\eta_l(du) = \frac{\kappa_l(u)}{Z_l} du \tag{1}$$

where the normalizing constant $Z_l = \int_E \kappa_l(u)du$ may be unknown. The context of interest is when the sequence of densities is associated to an 'accuracy' parameter h_l, with $h_l \to 0$ as $l \to \infty$ with $\infty > h_0 > h_1 > \cdots > h_\infty = 0$.

When estimating statistics $\mathbb{E}_{\eta_\infty}[g(U)]$, for $g : E \to \mathbb{R}$, in general one must approximate the limiting measure by η_L and perform statistical estimation with respect to this. For larger L, the approximation of the limit is better, and yet the simulations are more expensive and indeed the measure may also be supported on a subspace of the underlying space E whose dimension is larger.

Monte Carlo methods for statistical estimation are robust and scalable, although they are plagued by a "slow" convergence rate of $\mathcal{O}(N^{-1/2})$ for approximations using N degrees of freedom. Attempts to circumvent this issue, for example using sophisticated deterministic high-dimensional approximation methods typically result in some manifestation of the "curse of dimensionality" , although recent work has indicated potential for the mitigation of such effect for suitably regular problems [5].

The multilevel Monte Carlo (MLMC) framework [12, 13, 14] allows one to leverage in an optimal way the nested problems arising in this context, hence minimizing the necessary cost to obtain a given level of mean square error. In particular, the MLMC method seeks to sample from η_0 as well as a sequence of coupled pairs $(\eta_0, \eta_1), \ldots, (\eta_{L-1}, \eta_L)$ and using a collapsing sum representation of $\mathbb{E}_{\eta_L}[g(U)]$. Then using a suitable trade off of computational effort, one can reduce the amount of work, relative to i.i.d. sampling from η_L and using Monte Carlo integration, for a given amount of error. However, we are concerned with the scenario where such independent sampling is not possible, that is, either η_L or from the sequence of couples. As is well-known, the use of importance sampling to then use the collapsing sum representation, is often not reasonable, in the sense that for importance proposals that can be sampled independently, the associated variance typically explodes exponentially in the dimension of the problem (e.g. [4]). As a result, there has been an extension of MLMC methods in which the approximate target distribution can be sampled from directly, to more sophisticated Monte Carlo techniques for inference; however, this is still in its infancy. Important examples include the preliminary exploration of multilevel Markov chain Monte Carlo (MLMCMC) [15, 20], multilevel sequential Monte Carlo (MLSMC) samplers [3] (see also [18]), multilevel ensemble Kalman filter [16] and multilevel particle filters [19]. It should be noted that MCMC and SMC can perform at a polynomial cost in the dimension; see e.g. [2] and the references therein.

A significant challenge for inference problems is estimation of the normalizing constant Z_L or ratios thereof Z_l/Z_k, $L \geq l > k \geq 0$. Such quantities are central to Bayesian model comparison and choice [17, 25]. In addition, obtaining unbiased estimates (in the sense that the expectation is equal to the value, that is, potentially including discretization bias) are often central in pseudo-marginal algorithms (e.g. [1]). In general the calculation of these quantities are notoriously challenging (see for instance [26]) from a computational perspective.

In this article we extend the framework of [3] to consider the estimation of the ratio of normalizing constants. This is a framework which uses SMC. We consider both the 'standard' unbiased estimator ([9]) used in SMC, adapted to the multilevel setting and an estimator which follows the collapsing sum approach for multilevel methods. For the latter, we introduce a novel decomposition of the normalizing constant of a Feynman-Kac formula, which corresponds to Z_L/Z_0, which facilitates unbiased estimation. We consider new variance bounds for the estimator [9] and our new estimate and show that, in general, both approaches perform in a similar manner. For a given level of error, the cost is less than a Monte Carlo estimate that uses i.i.d. sampling from η_0, to estimate Z_L/Z_0; we assume that the former is possible.

The paper is structured as follows. In Section 2 the setup will be given, along with a description of the multilevel algorithm and the new novel estimator for the normalizing constant. Section 3 contains the theoretical results, including the main theorems of the paper which allow the multilevel theory to carry through. Finally, section 4 presents the results of numerical experiments on an example Bayesian inverse problem. The proofs are housed in the appendix.

2 Estimation

2.1 Notations

Let $(E, \mathcal{E})$ be a measurable space. The notation $\mathcal{B}_b(E)$ denotes the class of bounded and measurable real-valued functions. The supremum norm is written as $\|f\|_\infty = \sup_{u\in E} |f(u)|$ and $\mathcal{P}(E)$ is the set of probability measures on $(E, \mathcal{E})$. We will consider non-negative operators $K : E \times \mathcal{E} \to \mathbb{R}_+$ such that for each $u \in E$ the mapping $A \mapsto K(u, A)$ is a finite non-negative measure on $\mathcal{E}$ and for each $A \in \mathcal{E}$ the function $u \mapsto K(u, A)$ is measurable; the kernel K is Markovian if $K(u, dv)$ is a probability measure for every $u \in E$. For a finite measure μ on $(E, \mathcal{E})$, and a real-valued, measurable $f : E \to \mathbb{R}$, we define the operations:

$$\mu K : A \mapsto \int K(u, A)\, \mu(du) \ ; \quad Kf : u \mapsto \int f(v)\, K(u, dv).$$

We also write $\mu(f) = \int f(u)\mu(du)$. Throughout capital letters are used for random variables and small letters for the realisation of the sample.

2.2 Algorithm

The presentation to follow in this section is a review of the previous work [3], which is necessary such that the material in the present work be self-contained. As described in the Introduction, the context of interest is when a sequence of densities $\{\eta_l\}_{l\geq 0}$, as in (1), are associated to an 'accuracy' parameter h_l, with $h_l \to 0$ as $l \to \infty$, such that $\infty > h_0 > h_1 \cdots > h_\infty = 0$. In practice one cannot treat $h_\infty = 0$ and so must consider these distributions with $h_l > 0$. The laws with large h_l are easy to sample from with low computational cost, but are very different from η_∞, whereas, those distributions with small h_l are hard to sample with relatively high computational cost, but are closer to η_∞. Thus, we choose a maximum level $L \geq 1$ and we will estimate

$$\mathbb{E}_{\eta_L}[g(U)] := \int_E g(u)\eta_L(du) \ .$$

By the standard telescoping identity used in MLMC, one has

$$\mathbb{E}_{\eta_L}[g(U)] = \mathbb{E}_{\eta_0}[g(U)] + \sum_{l=1}^{L} \Big\{\mathbb{E}_{\eta_l}[g(U)] - \mathbb{E}_{\eta_{l-1}}[g(U)]\Big\} \tag{2}$$

$$= \mathbb{E}_{\eta_0}[g(U)] + \sum_{l=1}^{L} \mathbb{E}_{\eta_{l-1}}\Big[\Big(\frac{\kappa_l(U)Z_{l-1}}{\kappa_{l-1}(U)Z_l} - 1\Big)g(U)\Big] \ . \tag{3}$$

Suppose now that one applies an SMC sampler [10] to obtain a collection of samples (particles) that sequentially approximate $\eta_0, \eta_1, \ldots, \eta_L$. We consider the case when one initializes the population of particles by sampling i.i.d. from η_0, then at every step one targets η_l by importance sampling and selects particles according to the weights. An MCMC kernel which keeps η_l invariant is then applied in between to mutate the particles. To be more explicit, denote by $(U_0^{1:N_0}, \ldots, U_{L-1}^{1:N_{L-1}})$, with $+\infty > N_0 \geq N_1 \geq \cdots N_{L-1} \geq 1$, the samples after mutation, where the notation is introduced $U_l^{1:N_l} := \{U_l^i\}_{i=1}^{N_l}$. One resamples $\check{U}_l^{1:N_{l+1}}$ according to the weights $w_l^{1:N_l}$, where

$$w_l^i = \tilde{w}_l^i / (\sum_{j=1}^{N_l} \tilde{w}_l^j), \quad \text{and} \quad \tilde{w}_l^i := G_l(u_l^i) := (\kappa_{l+1}/\kappa_l)(u_l^i), \tag{4}$$

for indices $l \in \{0, \ldots, L-1\}$, yielding corrected samples from η_{l+1}. We will denote by $\{M_l\}_{1\leq l\leq L-1}$ the sequence of MCMC kernels used at stages $1, \ldots, L-1$, such that $\eta_l M_l = \eta_l$; these are discussed below. In particular, after selection by resampling according to $w_l^{1:N_l}$, one then mutates the particles by the MCMC kernel M_{l+1}, so that

$$U_{l+1}^i \sim M_{l+1}(\check{u}_l^i, \cdot) \ .$$

For $\varphi : E \to \mathbb{R}$, $l \in \{1, \ldots, L\}$, we have the following estimator of $\mathbb{E}_{\eta_{l-1}}[\varphi(U)]$:

$$\eta_{l-1}^{N_{l-1}}(\varphi) := \frac{1}{N_{l-1}} \sum_{i=1}^{N_{l-1}} \varphi(u_{l-1}^i) \ .$$

The algorithm is summarized in Table 1. If one considers one more step in the above procedure, it would deliver samples $\{U_L^i\}_{i=1}^{N_L}$. A standard SMC sampler estimate of the quantity of interest in (2) is $\eta_L^{N_L}(g)$; the earlier samples would then be discarded. There are two crucial differences between the algorithm we implement and the standard SMC sampler. First, all the samples along the path will be kept and used. Second, in the standard SMC sampler context, one would typically use $N_l = N$ for all l, for some fixed sample size N. In the present context, we will choose the N_l very carefully such that $N_l > N_{l+1}$ so that the cost is minimized (see equation (16)). A consistent SMC estimate of (3) is

$$\widehat{Y} = \eta_0^{N_0}(g) + \sum_{l=1}^{L} \left\{ \frac{\eta_{l-1}^{N_{l-1}}(gG_{l-1})}{\eta_{l-1}^{N_{l-1}}(G_{l-1})} - \eta_{l-1}^{N_{l-1}}(g) \right\} . \tag{5}$$

The derivation of the estimator from (3) follows from recalling the definition of $G_{l-1} = \kappa_l/\kappa_{l-1}$ from (4), and observing that therefore $\eta_{l-1}(G_{l-1}) = Z_l/Z_{l-1}$. The amount of work required to obtain a given level of error with this estimator is reduced relative to i.i.d. sampling from η_L [3]. Thus the idea of using the approach is clear. It is well known in the literature (e.g. [10]) that SMC samplers can also estimate ratios of normalizing constants as a by-product of the algorithm, using the full trajectory of samples. We now consider this in the multilevel context, and analyze the amount of work to obtain a given level of error.

2.2.1 On the MCMC Kernels

We now briefly discuss the MCMC kernels used in SMC samplers. Technically, M_l need only be η_l-invariant, but of course it is well-known in the literature (e.g. [9, 10]) that if they have fast mixing properties, such as uniform ergodicity, then this leads to improvements in SMC methods. For instance, a reduction in the asymptotic variance in the CLT.

Constructing η_l-invariant kernels is in general not difficult in the context of this article. We have used random walk Metropolis-Hastings (M-H) kernels, with (when required) reparameterizations onto the real line and Gaussian noises for the proposal. If the target is defined in infinite dimensions, then it is often better to define a M-H kernel on this space and discretize; see for instance [8].

2.3 Normalizing Constant

Define, for $l \geq 0$

$$\gamma_l(du_l) = \int_{E^l} \Big(\prod_{p=0}^{l-1} G_p(u_p)\Big)\eta_0(du_0)\prod_{p=1}^{l} M_p(u_{p-1}, du_p).$$

In our context, it is well known that (e.g. [9]):

$$\gamma_l(1) = \frac{Z_l}{Z_0} = \prod_{p=0}^{l-1} \eta_p(G_p).$$

0. Sample $U_0^1, \dots U_0^{N_0}$ i.i.d. from η_0 and compute $G_0(u_0^i)$ for each sample $i \in \{1, \dots, N_0\}$: Set $l = 0$.

1. Sample $\check{U}_l^1, \dots, \check{U}_l^{N_{l+1}}$ with replacement from $u_l^{1:N_l}$ with selection probabilities $\{G_l(u_l^1)/\sum_{j=1}^{N_l} G_l(u_l^j), \dots, (G_l(u_l^{N_l})/\sum_{j=1}^{N_l} G_l(u_l^j)\}$.

2. Sample $U_{l+1}^i|\check{u}_l^i$ from $M_{l+1}(\check{u}_l^i, \cdot)$ and compute $G_{l+1}(u_{l+1}^i)$ for each sample $i \in \{1, \dots, N_{l+1}\}$.

3. Set $l = l + 1$. If $l = L$ stop, otherwise return to the start of Step 1.

Table 1: The SMC algorithm.

This suggests the estimator:

$$\gamma_l^{N_{0:l-1}}(1) := \prod_{p=0}^{l-1} \eta_p^{N_p}(G_p) \tag{6}$$

which is known to be unbiased ([9], see the Martingale representation [9, Chapter 9]). We consider the relative variance of this estimator in Section 3. On appearance (6) may not seem to take advantage of the nature of the ML method: there is no telescopic identity used, and even if the numbers of samples are chosen as prescribed by [3] (or the result in (16) later on), it is not at all obvious that the level of error can be preserved while the N_l decay. However, below we will show that in fact this estimators does provide improved performance, with $N_{0:L-1} := \{N_l\}_{l=0}^{L-1}$ appropriately chosen. In addition, we show that one can use the estimator in subsection 2.5 to remove the discretization bias if that is of interest.

In what follows, we derive a new alternative estimator, which makes clear use of the multilevel framework. Note that one has for any $\varphi \in \mathcal{B}_b(E)$, $l \geq 0$

$$\eta_l(\varphi) = \frac{\gamma_l(\varphi)}{\gamma_l(1)}.$$

In particular, $\gamma_0 = \eta_0$. Now, for any $p \geq 2$

$$\begin{aligned}
\gamma_{p-2}(G_{p-2}G_{p-1}) &= \gamma_{p-2}(1)\eta_{p-2}(G_{p-2}G_{p-1}) \\
&= \frac{Z_{p-2}}{Z_0}\frac{1}{Z_{p-2}}\int_E \kappa_{p-2}(u)\frac{\kappa_{p-1}(u)}{\kappa_{p-2}(u)}\frac{\kappa_p(u)}{\kappa_{p-1}(u)}du \\
&= \frac{Z_p}{Z_0} \\
&= \gamma_p(1).
\end{aligned}$$

Note that this is due to the specific form of G_p, as defined in equation (4), but is not generally true for a given Feynman-Kac formula with arbitrary G_p. Recall that $\gamma_1(1) = \eta_0(G_0)$ and $\gamma_p(1) = \gamma_{p-1}(G_{p-1})$. So for any $l \geq 2$

$$\begin{aligned}\gamma_l(1) &= \eta_0(G_0) + \sum_{p=2}^{l} \Big(\gamma_p(1) - \gamma_{p-1}(1)\Big) \\ &= \eta_0(G_0) + \sum_{p=2}^{l} \Big(\gamma_{p-2}\big(G_{p-2}(G_{p-1} - 1)\big)\Big).\end{aligned}$$

It is the second line that we will approximate with our MLSMC sampler. The proposed approximation is

$$\tilde{\gamma}_l^{N_{0:l-2}}(1) = \eta_0^{N_0}(G_0) + \sum_{p=2}^{l} \Big(\gamma_{p-2}^{N_{0:p-2}}\big(G_{p-2}(G_{p-1} - 1)\big)\Big) \tag{7}$$

where for any $g \in \mathcal{B}_b(E)$, $p \geq 2$

$$\gamma_{p-2}^{N_{0:p-2}}(g) = \Big(\prod_{k=0}^{p-3} \eta_k^{N_k}(G_k)\Big)\eta_{p-2}^{N_{p-2}}(g).$$

Note that for $l \geq 2$, one has, almost surely,

$$\tilde{\gamma}_l^{N_{0:l-2}}(1) \neq \gamma_l^{N_{0:l-1}}(1).$$

For instance,

$$\tilde{\gamma}_2^{N_0}(1) = \eta_0^{N_0}(G_0 G_1)$$

is almost surely not equal to

$$\gamma_2^{N_{0:1}}(1) = \eta_0^{N_0}(G_0)\eta_1^{N_1}(G_1).$$

Using [9] it clearly follows that

$$\gamma_l(1) = \mathbb{E}[\tilde{\gamma}_l^{N_{0:l-2}}(1)]$$

where $\mathbb{E}$ is the expectation w.r.t. the law of the SMC algorithm; the estimator is unbiased. To our knowledge, this is the first time this new unbiased estimator has been defined, at least in the context considered here. It is noted that this estimator can take negative values.

2.4 Biased Estimator

It is important to emphasize why it is not sensible to simply "plug-in" to the naive estimator (5), to derive the following estimator for $\gamma_l(1)$:

$$\prod_{p=0}^{l-1} \left(\eta_0^{N_0}(G_p) + \sum_{l=1}^{p} \Big\{\frac{\eta_{l-1}^{N_{l-1}}(G_p G_{l-1})}{\eta_{l-1}^{N_{l-1}}(G_{l-1})} - \eta_{l-1}^{N_{l-1}}(G_p)\Big\}\right).$$

One can easily prove that this estimate is consistent, but biased, in the sense that

$$\mathbb{E}\Big[\prod_{p=0}^{l-1}\Big(\eta_0^{N_0}(G_p)+\sum_{l=1}^{p}\Big\{\frac{\eta_{l-1}^{N_{l-1}}(G_pG_{l-1})}{\eta_{l-1}^{N_{l-1}}(G_{l-1})}-\eta_{l-1}^{N_{l-1}}(G_p)\Big\}\Big)\Big]\neq\gamma_l(1).$$

However, the main reason why one may not want to consider its use is due to the cost of computing this estimate. If $\sum_{p=0}^{l-1}N_p\mathcal{C}_p$ is the ordinary cost of computing (6), where $\mathcal{C}_p$ is the cost per sample, then the cost of this estimator is $\sum_{p=0}^{l-1}N_p\sum_{q=p}^{l-1}\mathcal{C}_q$. Such a procedure is undesirable in general and this is not investigated further.

2.5 Estimator with no Discretization Bias

We remark that one may consider an estimator which eliminates also discretization bias altogether. However, the utility is potentially limited. Let $M\in\{1,2,\dots\}$ be a random variable that is independent of the MLSMC algorithm with $\mathbb{P}_M(M\geq m)>0\ \forall m>0$. Suppose further that one can prove for $N_0,N_1,\dots$ fixed that

$$\lim_{p\to\infty}\mathbb{E}\Big[\Big(\gamma_{p-2}^{N_{0:p-2}}(G_{p-2}G_{p-1})-\gamma_\infty(1)\Big)^2\Big]^{1/2} = 0 \quad (8)$$

$$\lim_{p\to\infty}\mathbb{E}\Big[\Big(\gamma_{p-2}^{N_{0:p-2}}(G_{p-2})-\gamma_\infty(1)\Big)^2\Big]^{1/2} = 0 \quad (9)$$

$$\sum_{p=2}^{\infty}\frac{1}{\mathbb{P}_M(M\geq p)}\mathbb{E}\Big[\Big(\gamma_{p-2}^{N_{0:p-2}}(G_{p-2}(G_{p-1})-\gamma_\infty(1)\Big)^2\Big] < +\infty \quad (10)$$

then one can use the estimator from [22] to obtain an unbiased estimator for $\gamma_\infty(1)$:

$$\frac{1}{\mathbb{P}_M(M\geq 1)}\eta_0^{N_0}(G_0)+\sum_{p=2}^{M}\frac{1}{\mathbb{P}_M(M\geq p)}\Big(\gamma_{p-2}^{N_{0:p-2}}(G_{p-2}(G_{p-1}-1))\Big).$$

Note that, even if one can prove (8)-(10), one must be prepared to spend an arbitrary amount of computational cost, which is not reasonable in the current context. Hence we do not consider this further here. We further remark that when estimating $\mathbb{E}_{\eta_\infty}[g(U)]$ (as in (3)) there is no unbiased property of the estimators, in the sense of expectations. Hence this approach is unlikely to work in that context.

3 Theory

3.1 Relative Variance Bounds

Throughout E is compact. We make the following assumptions:

(**A1**) There exist $0 < \underline{C} < \overline{C} < +\infty$ such that

$$\begin{aligned} \sup_{0\leq l<\infty} \sup_{u\in E} \kappa_l(u) &\leq \overline{C} ; \\ \inf_{0\leq l<\infty} \inf_{u\in E} \kappa_l(u) &\geq \underline{C} . \end{aligned}$$

(**A2**) There exists a $\rho \in (0,1)$ such that for any $l \geq 1$, $(u,v) \in E^2$, $A \in \mathcal{E}$:

$$\int_A M_l(u, du') \geq \rho \int_A M_l(v, dv') .$$

These assumptions are almost identical to those in [3]. (A1) is different but equivalent to (A1) in [3]. The proofs of the following Theorems are in Appendices B and C respectively. It is remarked that there are other results in the spirit of Theorem 1 below, (see [6, 23]) but the bounds are not sharp enough for the purposes of this work. The assumptions are quite strong, and indeed similar results are expected for weaker assumptions, however the considerable technical difficulty is beyond the scope of the present work; see [11] for a weakening of the conditions for the estimate (5).

Theorem 1. *Assume (A1-2). Then there exists a $c, C < +\infty$ such that for any $L \geq 2$, $N_0 \geq N_1 \geq \cdots \geq N_{L-1} > cL$,*

$$\mathbb{E}\Big[\Big(\frac{\gamma_L^{N_{0:L-1}}(1)}{\gamma_L(1)} - 1\Big)^2\Big] \leq$$

$$C \sum_{p=0}^{L-1} \frac{1}{N_p}\Big(\Big(\sum_{q=p}^{L-1}\Big\|\frac{G_q}{\eta_q(G_q)} - 1\Big\|_\infty\Big)^2 + \Big\|\frac{G_p}{\eta_p(G_p)} - 1\Big\|_\infty \frac{(p+1)}{N_p}\Big).$$

Theorem 2. *Assume (A1-2). Then there exists a $c, C < +\infty$ such that for any $L \geq 2$, $N_0 \geq N_1 \geq \cdots \geq N_{L-2} > c(L-1)$,*

$$\mathbb{E}\Big[\Big(\frac{\tilde{\gamma}_L^{N_{0:L-2}}(1)}{\gamma_L(1)} - 1\Big)^2\Big] \leq$$

$$C\Big(\frac{1}{N_0} + \sum_{p=2}^{L}\frac{(p-1)}{N_{p-2}}\|G_{p-1} - 1\|_\infty^2 + \sum_{p=2}^{L}\sum_{q=2}^{p-1}\frac{(q-1)}{N_{q-2}}\|G_{p-1} - 1\|_\infty \|G_{q-1} - 1\|_\infty\Big).$$

3.2 Cost Analysis

In order to investigate the cost for a given level of error, we introduce the following assumption.

(**A3**) (i) There exist $\alpha, \zeta > 0$, and a $C > 0$ such that for all $p > 0$

$$\begin{cases} |\frac{\gamma_p(1)}{\gamma_\infty(1)} - 1| & \leq C h_p^\alpha; \\ \mathrm{C}(G_{p-1}) & \leq C h_p^{-\zeta}, \end{cases} \tag{11}$$

where $\mathrm{C}(G_{p-1})$ denotes the cost to evaluate G_{p-1}.

(ii) There exist a $\beta > 0$ and a $C > 0$ such that for all $p > 0$

$$\Big\|\frac{G_{p-1}}{\eta_{p-1}(G_{p-1})} - 1\Big\|_\infty^2 \leq Ch_p^\beta.$$

(iii) There exist a $\beta > 0$ and a $C > 0$ such that for all $p > 0$

$$\|G_{p-1} - 1\|_\infty^2 \leq Ch_p^\beta.$$

Corollary 3.1. *Assume (A1,2,3(i)(ii)) and* $2\alpha \geq \max\{\beta, \zeta\}$. *Then for any* $\varepsilon > 0$, *there exist* $L, \{N_l\}_{l=0}^L$ *and* $C > 0$ *such that*

$$\frac{1}{\gamma_\infty(1)^2}\mathbb{E}\Big[\Big(\gamma_L^{N_{0:L-1}}(1) - \gamma_\infty(1)\Big)^2\Big] \leq C\varepsilon^2, \tag{12}$$

for the following cost

$$\mathrm{COST} \leq C\begin{cases} \varepsilon^{-2}|\log(\varepsilon)|, & \text{if} \quad \beta > \zeta, \\ \varepsilon^{-2}|\log(\varepsilon)|^3, & \text{if} \quad \beta = \zeta, \\ \varepsilon^{-\left(2+\frac{\zeta-\beta}{\alpha}\right)}|\log(\varepsilon)|, & \text{if} \quad \beta < \zeta. \end{cases} \tag{13}$$

Corollary 3.2. *Assume (A1,2,3(i)(iii)) and* $2\alpha \geq \max\{\beta, \zeta\}$. *Then for any* $\varepsilon > 0$, *there exist* $L, \{N_l\}_{l=0}^L$ *and* $C > 0$ *such that*

$$\frac{1}{\gamma_\infty(1)^2}\mathbb{E}\Big[\Big(\tilde{\gamma}_L^{N_{0:L-2}}(1) - \gamma_\infty(1)\Big)^2\Big] \leq C\varepsilon^2, \tag{14}$$

for the following cost

$$\mathrm{COST} \leq C\begin{cases} \varepsilon^{-2}|\log(\varepsilon)|, & \text{if} \quad \beta > \zeta, \\ \varepsilon^{-2}|\log(\varepsilon)|^3, & \text{if} \quad \beta = \zeta, \\ \varepsilon^{-\left(2+\frac{\zeta-\beta}{\alpha}\right)}|\log(\varepsilon)|, & \text{if} \quad \beta < \zeta. \end{cases} \tag{15}$$

We give the proof for Corollary 3.2 only. The proof of Corollary 3.1 is almost identical. The only difference is treating the additional term

$$\sum_{p=0}^{L-1}\Big\|\frac{G_p}{\eta_p(G_p)} - 1\Big\|_\infty \frac{(p+1)}{N_p^2},$$

which is bounded by $C\varepsilon^2$, under our assumptions.

Proof of Corollary 3.2. The mean-square error (MSE) can be bounded by

$$\frac{1}{\gamma_\infty(1)^2}\mathbb{E}\Big[\Big(\tilde{\gamma}_L^{N_{0:L-2}}(1) - \gamma_\infty(1)\Big)^2\Big] \leq$$

$$\left(\frac{\gamma_L(1)}{\gamma_\infty(1)}\right)^2 \mathbb{E}\Big[\Big(\frac{\tilde{\gamma}_L^{N_{0:L-2}}(1)}{\gamma_L(1)} - 1\Big)^2\Big] + \left|\left(\frac{\gamma_L(1)}{\gamma_\infty(1)} - 1\right)\right|^2 .$$

Following from (A3(i)), the second term requires that $h_L^\alpha \eqsim \varepsilon$, and assuming $h_L = M^{-L}$ for some $M \geq 2$, this translates to $L \eqsim \log \varepsilon$. Notice that it also follows that $\left(\frac{\gamma_L(1)}{\gamma_\infty(1)}\right)^2 = \mathcal{O}(1)$. Now, defining $V_p = \|G_{p-1} - 1\|_\infty^2$, Theorem 2 provides the following bound for the first term

$$\mathbb{E}\Big[\Big(\frac{\tilde{\gamma}_L^{N_{0:L-2}}(1)}{\gamma_L(1)} - 1\Big)^2\Big] \leq V := C\left(\frac{1}{N_0} + L\sum_{p=1}^{L-1}\frac{V_p}{N_{p-1}}\right).$$

To see this observe that

$$\sum_{p=1}^{L-1}\sum_{q=1}^{p}\frac{q}{N_{q-1}}V_p^{1/2}V_q^{1/2} = \sum_{p=1}^{L-1}\frac{p}{N_{p-1}}V_p^{1/2}\sum_{q=p}^{L-1}V_q^{1/2} \leq CL\sum_{p=1}^{L-1}\frac{V_p}{N_{p-1}}.$$

Optimizing the cost, given that the variance is $\mathcal{O}(\varepsilon^2)$, dictates that $N_l \propto \sqrt{LV_l/\mathrm{C}(G_l)} \eqsim L^{1/2}h_l^{(\beta+\zeta)/2}$. The constraint that $N_l \geq L$ then requires that

$$N_l \propto L\varepsilon^{-2}K_L h_l^{(\beta+\zeta)/2}, \tag{16}$$

where $K_L = \sum_{l=1}^{L-1} h_l^{(\beta-\zeta)/2}$. By assumption $\max\{\beta,\zeta\} \leq 2\alpha$, so $(\beta+\zeta)/2\alpha \leq 2$ and the requirement for all the N_l in Theorem 2 is guaranteed (as long as the proportionality constant is greater than 1). Therefore, the MSE is controlled by $\mathcal{O}(\varepsilon^2)$ with a cost given by

$$\sum_{l=0}^{L-1} N_l \mathrm{C}(G_l) \eqsim L\varepsilon^{-2}K_L^2\,,$$

and the result follows.

□

Remark 3.1. *Suppose one were able to perform i.i.d. sampling from η_0 (denote the samples $u^1, \ldots, u^N$). Then Z_L/Z_0 could be estimated by*

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_L(u^i)}{\gamma_0(u^i)},$$

for a computational effort proportional to $Nh_L^{-\zeta}$, where N is the number of simulated samples. The variance of this MC estimate is $\mathcal{O}(N^{-1})$, independently of L. So to make the overall error (bias squared plus variance) $\mathcal{O}(\varepsilon^2)$, one must take $N \eqsim \varepsilon^{-2}$. The bias requires $h_L = \varepsilon^{1/\alpha}$. This results in a total computational cost of $\mathcal{O}(\epsilon^{-2-\zeta/\alpha})$, which is far worse than the MLSMC sampler estimators presented in the present work.

4 Numerical Example

In this section the performance of the proposed estimators will be demonstrated by two Bayesian inverse problem examples, arising from elliptic PDE. In subsection 4.1 an example in which the PDE is defined on an interval in 1 spatial dimension is considered. In subsection 4.2 an example in which the PDE is defined over a box in 2 spatial dimensions is considered.

4.1 Poisson equation in 1-dimension

The example given in this section is the same one from [3], which introduced the MLSMC algorithm. We reproduce some of the content of that article, to ensure that this one is self-contained. The estimation of normalizing constants is particularly useful in this context when one wants to perform model selection or estimate hyper-parameters. In this case the evaluation of the likelihood for a given value of the model or hyper-parameter requires estimating the normalizing constant. The latter is only considered, as this is the main emphasis of the article.

We define the nested Hilbert spaces $V := H^1(D) \subset L^2(D) \subset H^{-1}(D) =: V^*$, where the domain D will be given later on. Let $D \subset \mathbb{R}^d$ with $\partial D \in C^1$ convex. For $f \in V^*$, consider the following PDE on D:

$$-\nabla \cdot (\hat{u}\nabla p) = f, \qquad \text{on } D, \tag{17}$$

$$p = 0, \qquad \text{on } \partial D, \tag{18}$$

where

$$\hat{u}(x) = \bar{u}(x) + \sum_{k=1}^{K} u_k \sigma_k \phi_k(x). \tag{19}$$

Define $u = \{u_k\}_{k=1}^K$, with $u_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}[-1,1]$ (the uniform distribution on [-1,1]). Assume that $\bar{u}, \phi_k \in C^\infty$ for all k and $\|\phi_k\| = 1$ (the norm on L^2). In particular $\{\sigma_k\}_{k=1}^K$ decays with k. In addition, the following property holds:

$$\inf_x \hat{u}(x) \geq \inf_x \bar{u}(x) - \sum_{k=1}^{K} \sigma_k \geq u_* > 0 \tag{20}$$

so that the operator on the left-hand side of Equation (17) is uniformly elliptic. Let $p(\cdot; u)$ denote the weak solution of Equation (17) for parameter u. Define the following vector-valued function

$$\mathcal{G}(p) = [g_1(p), \ldots, g_M(p)]^\intercal,$$

where g_m are elements of the dual space V^* for $m = 1, \ldots, M$. It is assumed that the data take the form

$$y = \mathcal{G}(p) + \xi, \qquad \xi \sim \mathcal{N}(0, \Xi), \tag{21}$$

where $\mathcal{N}(0,\Xi)$ denotes the Normal distribution with zero mean and covariance Ξ. The likelihood at a given l is therefore given by

$$\kappa_l(u) = \exp\left(-\frac{1}{2}|\Xi^{-\frac{1}{2}}(y - \mathcal{G}_l(p(\cdot;u))|^2\right).$$

The selection function is therefore given by

$$G_l = \frac{\kappa_{l+1}}{\kappa_l} = \exp\left(\frac{1}{2}|\Xi^{-\frac{1}{2}}(y - \mathcal{G}_l(p(\cdot;u))|^2 - \frac{1}{2}|\Xi^{-\frac{1}{2}}(y - \mathcal{G}_{l+1}(p(\cdot;u))|^2\right).$$

The specific setting of the simulations are as the following. Let $D = [0,1]$ and $f(x) = 100x$. Set $K = 50$, $\bar{u}(x) = 0.15 = \text{const.}$, $\sigma_k = (2/5)4^{-k}$ $\phi_k(x) = \sin(k\pi x)$ if k is odd and $\phi_k(x) = \cos(k\pi x)$ if k is even. The forward problem at resolution level l is solved using a finite element method with piecewise linear shape functions on a uniform mesh of with $h_l = 2^{-(l+k)}$, for some starting $k \geq 1$ (so that there are at least two grid-blocks in the coarsest, $l = 0$, case). Thus, on level l the finite element basis functions are $\{\psi_i^l\}_{i=1}^{2^{l+k}-1}$ defined as (for $x_i = i \cdot 2^{-(l+k)}$):

$$\psi_i^l(x) = \begin{cases} (1/h_l)[x - (x_i - h_l)] & \text{if } x \in [x_i - h_l, x_i], \\ (1/h_l)[(x_i + h_l) - x] & \text{if } x \in [x_i, x_i + h_l]. \end{cases}$$

The function of interest g is taken as the solution of the forward problem at the midpoint of the domain, that is $g(u) = p(0.5;u)$. The observation operator is $\mathcal{G}(u) = [p(0.25;u), p(0.75;u)]^\intercal$, and the observational noise covariance is taken to be $\Xi = 0.25^2 I$.

Detailed error rates analysis of this example can be found in [3]. In particular, when the purpose of the study was to estimate $\eta_L(g)$, the variance rate was $\beta = 4$ empirically. Later we will show that for estimating the normalizing constant, the variance rate is very similar.

4.1.1 Verification of Assumptions

Assumptions (A1) and (A3(i)(iii)) (for $|\frac{\gamma_p(1)}{\gamma_\infty(1)} - 1|$), with $\beta = 2\alpha = 2$, follow from Proposition 4.1 of [3]. For (A3(ii)) this follows directly from proving (A3(iii)). The cost at level p is a power of the number degrees of freedom, which is in turn related to h_p^{-1}, verifying (A3(i)) (for $\mathrm{C}(G_{p-1})$). This is because the stiffness matrix of the finite element method is tridiagonal and thus the system can be solved with cost $\mathcal{O}(2^{l+k})$, corresponding to a computational cost rate of $\zeta = 1$. Assumption (A2) is verified for Gibbs sampler in section 4.2 of [3]. As mentioned previously, a Metropolis-Hastings kernel is used which includes an accept/reject step, not present in a Gibbs sampler. This means that (A2) does not hold (see e.g. [11] and the discussion therein) in our numerical examples. However, these numerical examples provide a test of the mathematical theory, in that we believe our results can be extended to the case of M-H kernels and there is only a technical barrier to proving so.

4.1.2 Experiments

We begin by using the theoretical rates $\beta = 2\alpha = 2$ to estimate the MSE and hence the cost ratio. Three cases are considered:

- A standard SMC algorithm, with the estimator $\gamma_L^{N_{0:L-1}}(1)$ and $N_l = N$ for all l.
- MLSMC sampler for $\gamma_L^{N_{0:L-1}}(1)$, as given in equation (6) with $N_{0:L-1}$ chosen as in (16).
- MLSMC sampler for $\tilde{\gamma}_L^{N_{0:L-2}}(1)$, as given in equation (7), with $N_{0:L-1}$ chosen as in (16).

Below, the MSE includse the bias.L is kept fixed with $L \in \{0, , 9\}$ and N_l is chosen accordingly. The bias is of the same order of the variance. The cost vs. MSE is plotted in Figure 1; the theoretical cost of the model is plotted instead of the actual wall clock time. The cost rates are -2.542, -1.934, and -2.076 for the SMC, MLSMC with the standard estimator, and MLSMC with the new estimator, respectively. It is clear that the MLSMC algorithm with both estimators provides superior performance when compared to the standard SMC algorithm. It is interesting that for the given MLSMC ensemble, the performance of the new estimator is comparable to that of the standard estimator, as proven in Corollaries 3.1 and 3.2. It should be noted that in practice, given the same samples $(U_0^{1:N_0}, \ldots, U_{L-1}^{1:N_{L-1}})$, the new estimator is capable of estimating $\gamma_{L+1}(1)$ while the standard one can only estimate the $\gamma_L(1)$, which has a higher bias. The $N_{0:L-1}$ are chosen according to the prescription in (16).

The variance rate β can also be estimated empirically by considering the variance of $\eta_l^{N_l}(G_l)$. The experiment is therefore repeated 100 times to compute the variance of the estimator. This quantity, multiplied by the sample size, is a proxy of V_l and is plotted in Figure 2. The estimated empirical rate is $\hat{\beta} = 4.148$. This is consistent with the rate estimates in [3], and it is explained there that actually under suitable assumptions an L^2 bound can be used in the verification of (A3) to obtain the stronger rate. Since $\alpha = \beta/2$ and $\zeta = 1$, this leads to an MSE cost growth rate of $-2 - \zeta/\alpha = -2.5$ for SMC, and a cost growth rate of -2 (up to log factor) for MLSMC, as verified in Figure 1. Indeed it is found that in practice the cost growth rate is roughly the same when using either value of $\beta = 2$ or 4, in (16). For MLSMC this is expected as the theoretical rate does not change. For SMC we believe this has to do with artificially good performance on the low levels, while asymptotically for $\beta = 2$ the cost rate would approach -3, due to wasted effort from drawing too many samples after already realizing a variance at the same level as the squared bias.

4.2 Poisson equation in 2-dimension

The performance of the proposed algorithms will be demonstrated further with a 2-dimensional Poisson equation example, in which the parameters of the forcing

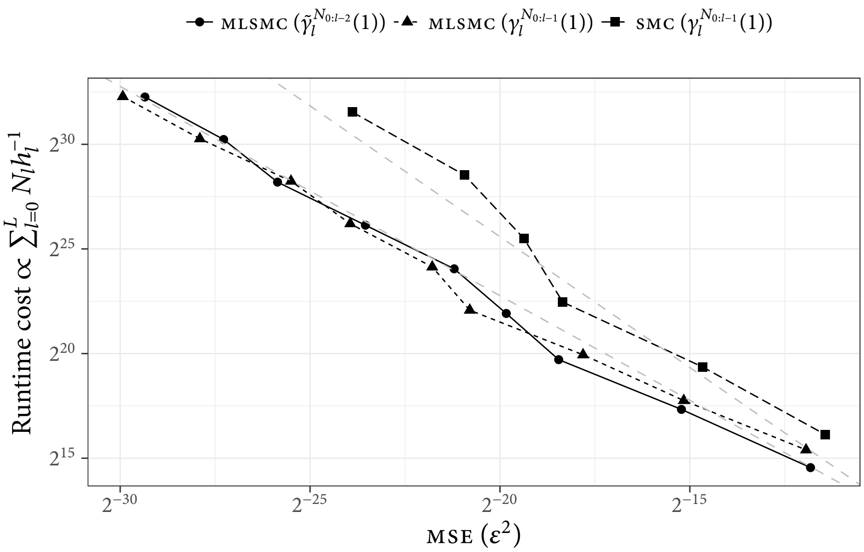

Figure 1: Computational cost against mean squared error of the 1-dimensional Poisson equation inverse problem. For the solver, $k = 2$, i.e. the coarsest mesh ($L = 0$) only partitions the interval into four.

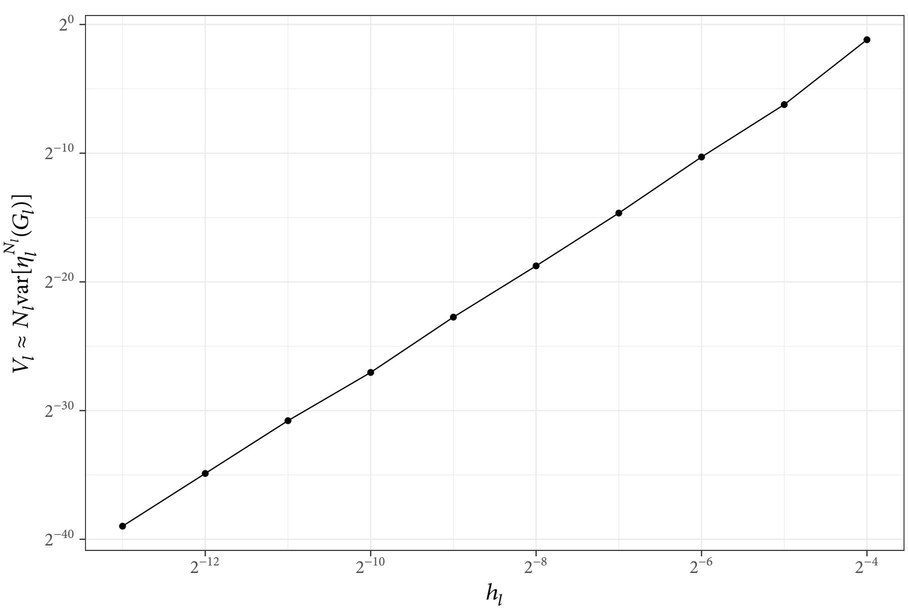

Figure 2: Variance rate estimate of the 1-dimensional Poisson equation inverse problem.

on the right-hand side are inferred. Let $D = [0,1] \times [0,1]$, and consider the following PDE

$$-\nabla^2 p(x,y) + p(x,y) = f(x,y), \qquad \text{on } D \tag{22}$$
$$p(x,y) = 0 \qquad \text{on } \partial D \tag{23}$$

where $f(x,y) = ((u_1^2 + u_2^2)\pi^2 + 1)\sin(u_1\pi x)\cos(u_2\pi y)$. The analytical solution is $p(x,y) = \sin(u_1\pi x)\cos(u_2\pi y) + 1$. The parameters are $u = (u_1, u_2)$, with uniform prior. The data is measured at four points, $\{0.25, 0.75\} \times \{0.25, 0.75\}$ with additive Normal error with standard deviation 0.1. The PDE is numerically solved with finite difference method. The interval at level l along the x-axis, and y-axis, is $h_l = 2^{-(l+k)}$, $k \geq 1$. In this example, we choose $k = 3$. The system can be solved with a cost of $\mathcal{O}(4^{l+k})$, i.e. $\zeta = 2$.

4.2.1 Verification of Assumptions

On can show that the Lax-Milgram lemma [7] holds uniformly for (22), providing the uniform bound of the likelihood. The remaining assumptions are verified in the same manner as section 4.1.1 and are hence omitted. We remark that the variance rate β is the same, but now the cost rate is such that $\zeta = 2$, due to the higher dimension.

4.2.2 Experiments

The variance and bias rates are estimated empirically. The former is shown in Figure 3. The estimated rate is $\hat{\beta} = 3.982$. This rate is used for the calculation of the sample sizes N_l according to (16). In Figure 4 we show the theoretical cost vs. MSE plot. The estimated cost rates are -3.01, -2.06 and -1.986 for the SMC, MLSMC with the standard estimator, and MLSMC with the new estimator, respectively. It is clear that the MLSMC algorithm with both estimators provides superior performance when compared to the standard SMC algorithm. The difference is more profound than the simpler 1-dimensional example, as expected with theoretical cost growth rate of -3 for SMC, while still -2 for each of the MLSMC estimators (up to log factor). This is verified in Figure 4.

Acknowledgements

KJHL was supported in part by DARPA FORMULATE and in part by ORNL LDRD Strategic Hire. AJ & YZ were supported by Ministry of Education AcRF tier 2 grant, R-155-000-161-112. We thank the referees and associate editor for their comments, which have greatly improved the article.

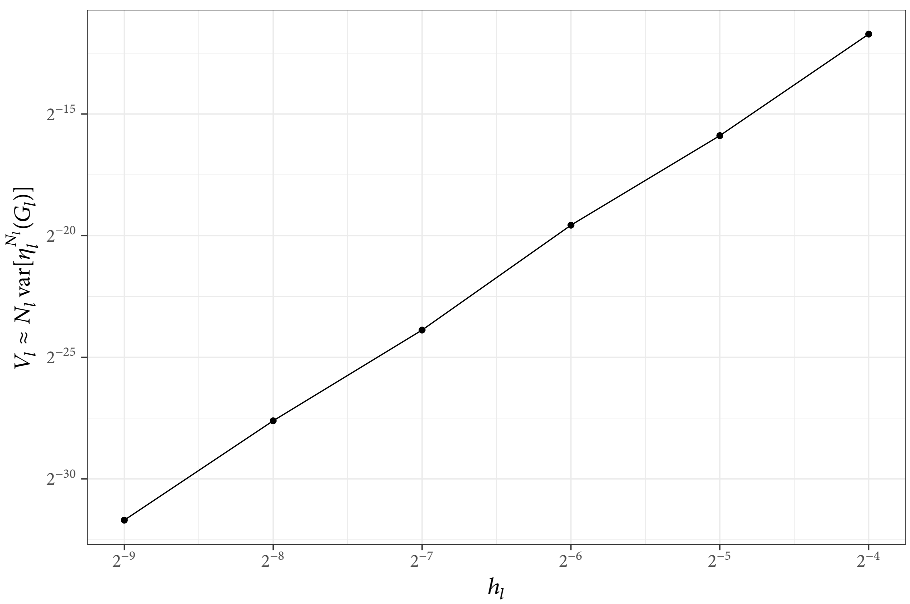

Figure 3: Variance rate estimate of the 2-dimensional Poisson equation inverse problem.

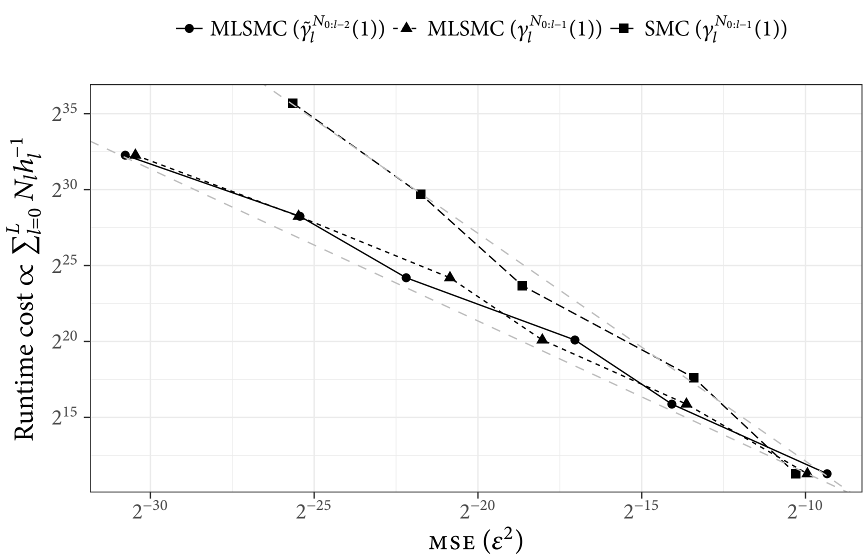

Figure 4: Computational cost against mean squared error of the 2-dimensional Poisson equation inverse problem.

A Notations

We give a collection of defintions which are used in the appendices. Let $n \geq 0$, $F \in \mathcal{B}_b(E \times E)$ and define

$$(\gamma_n^{N_{0:n}})^{\otimes 2}(F) = \Big(\prod_{p=0}^{n-1} \eta_p^{N_p}(G_p)\Big)^2 (\eta_n^{N_n})^{\otimes 2}(F)$$

where for a finite (possibly signed) measure on E, μ, $\mu^{\otimes 2}(d(u_1, u_2)) = \mu(du_1)\mu(du_2)$. We recall the semi-group for $p \leq n$ (for $p = n$ it is the identity operator):

$$Q_{p,n}(x_p, dx_n) = \int_{E^{n-p-1}} Q_{p+1}(x_p, dx_{p+1}) \dots Q_n(x_{n-1}, dx_n)$$

where for $n \geq 1$, $Q_n(x, dy) = G_{n-1}(x)M_n(x, dy)$. We also define the coalescent operator for $F \in \mathcal{B}_b(E \times E)$, $(x, y) \in E \times E$:

$$C(F)(x, y) = F(x, x).$$

Then for $0 \leq s \leq (n + 1)$, $0 \leq i_1 < \cdots < i_n \leq n$, $F \in \mathcal{B}_b(E \times E)$

$$\Gamma_n^{i_1:i_s}(F) = \gamma_{i_1}^{\otimes 2} C Q_{i_1,i_2}^{\otimes 2} C Q_{i_2,i_3}^{\otimes 2} \dots C Q_{i_s,n}^{\otimes 2}(F)$$

and

$$\overline{\Gamma}_n^{i_1:i_s}(F) = \frac{1}{\gamma_n(1)^2} \Gamma_n^{i_1:i_s}(F).$$

The conventions, for $s = 0$, $\Gamma_n^{\emptyset}(F) = \gamma_n^{\otimes 2}(F)$ and $\overline{\Gamma}_n^{\emptyset}(F) = \eta_n^{\otimes 2}(F)$ are adopted. Recall the selection-mutation operator for any $\mu \in \mathcal{P}(E)$, $n \geq 1$

$$\Phi_n(\mu)(dx) = \frac{\mu(G_{n-1}M_n(\cdot, dx))}{\mu(G_{n-1})}.$$

$\mathscr{F}_n^{N_{0:n}}$ denotes the natural filtration generated by the particle system up-to time n. For $f_1, f_2 \in \mathcal{B}_b(E)$ we write the tensor product of functions for every $(x, y) \in E \times E$:

$$f_1 \otimes f_2(x, y) = f_1(x)f_2(y).$$

Unless otherwise stated, the results proved in this appendix are new.

B Proofs for Theorem 1

In order to prove Theorem 1, we first give a technical Lemma, followed by Proposition of individual interest. The proof of Theorem 1 is then given.

Lemma 1. *Assume (A1-2). Then there exist a $C < +\infty$ such that for any $0 \leq p \leq n$, $x \in E$:*

$$\Big|\frac{Q_{p,n}(1)(x)}{\prod_{q=p}^{n-1} \eta_q(G_q)} - 1\Big| \leq C \sum_{q=p}^{n-1} \Big\|\frac{G_q}{\eta_q(G_q)} - 1\Big\|_\infty$$

Proof. We fix n, p and note that the case $p = n$ is trivial, so we suppose $p < n$. We prove the result by induction. We consider $p = n - 1$ and thus

$$\frac{Q_{p,n}(1)(x)}{\prod_{q=p}^{n-1} \eta_q(G_q)} - 1 = \frac{G_{n-1}(x)}{\eta_{n-1}(G_{n-1})} - 1$$

so the initialization is proved. Suppose the result holds at rank p and consider the case $p - 1$. Adding and subtracting the factor:

$$M_p\Big(\frac{Q_{p,n}(1)(\cdot)}{\prod_{q=p}^{n-1} \eta_q(G_q)}\Big)(x) = \int_E \Big(\frac{Q_{p,n}(1)(u)}{\prod_{q=p}^{n-1} \eta_q(G_q)}\Big) M_p(x, du)$$

to the term $Q_{p-1,n}(1)(x)/\prod_{q=p-1}^{n-1} \eta_q(G_q) - 1$ we have

$$\frac{Q_{p-1,n}(1)(x)}{\prod_{q=p-1}^{n-1} \eta_q(G_q)} - 1 =$$

$$\Big(\frac{G_{p-1}(x)}{\eta_{p-1}(G_{p-1})} - 1\Big) M_p\Big(\frac{Q_{p,n}(1)(\cdot)}{\prod_{q=p}^{n-1} \eta_q(G_q)}\Big)(x) + M_p\Big(\frac{Q_{p,n}(1)(\cdot)}{\prod_{q=p}^{n-1} \eta_q(G_q)} - 1\Big)(x).$$

By [6, Lemma 4.1]

$$\frac{Q_{p,n}(1)(x)}{\prod_{q=p}^{n-1} \eta_q(G_q)} \leq C, \tag{24}$$

where C does not depend upon p, n. Thus, by applying the induction hypothesis and the above result it follows that:

$$\Big|\frac{Q_{p-1,n}(1)(x)}{\prod_{q=p-1}^{n-1} \eta_q(G_q)} - 1\Big| \leq C \sum_{q=p-1}^{n-1} \Big\|\frac{G_q}{\eta_q(G_q)} - 1\Big\|_\infty$$

and hence the proof is completed. □

The result below follows one in [6].

Proposition 1. *Assume (A1-2). Then there exists a* $C < +\infty$ *such that for any* $n \geq 0$, $F \in \mathcal{B}_b(E \times E)$ *and* $N_0 \geq \cdots \geq N_n > c(n + 1)$

$$\Big|\mathbb{E}\Big[\frac{(\gamma_n^{N_{0:n}})^{\otimes 2}(F)}{\gamma_n(1)^2}\Big] - \eta_n^{\otimes 2}(F)\Big| \leq 8c\|F\|_\infty \sum_{p=0}^{n} \frac{1}{N_p}.$$

Proof. The case with F constant essentially follows from the proofs of [6]. The only difference is the fact that we have a decreasing number of samples; this does not change the calculations of that paper, so the case of F constant is in [6]. If F is a non-constant function, one has, from the equation above Proposition 3.4 (page 638) of [6]:

$$\Big|\mathbb{E}\Big[\frac{(\gamma_n^{N_{0:n}})^{\otimes 2}(F)}{\gamma_n(1)^2}\Big] - \eta_n^{\otimes 2}(F)\Big| =$$

$$\Big|\sum_{s=1}^{n+1}\sum_{0\le i_1<\dots<i_s\le n}\Big(\prod_{k=1}^{s}\frac{1}{N_{i_k}}\Big)\Big(\prod_{k\notin\{i_1,\dots,i_s\}}(1-\frac{1}{N_k})\Big)\overline{\Gamma}_n^{i_1:i_s}(F-\eta_n^{\otimes 2}(F))\Big|.$$

Following the proof of Theorem 5.1 of [6] and noting that one can allow the function in that paper to be negative, it follows that

$$|\overline{\Gamma}_n^{i_1:i_s}(F-\eta_n^{\otimes 2}(F))| \le \|F-\eta_n^{\otimes 2}(F)\|_\infty\Big(\rho\frac{\overline{C}}{\underline{C}}\Big)^s \le 2\|F\|_\infty\Big(\rho\frac{\overline{C}}{\underline{C}}\Big)^s.$$

Thus one has

$$\Big|\mathbb{E}\Big[\frac{(\gamma_n^{N_{0:n}})^{\otimes 2}(F)}{\gamma_n(1)^2}\Big]-\eta_n^{\otimes 2}(F)\Big| \le 2\|F\|_\infty\sum_{s=1}^{n+1}\sum_{0\le i_1<\dots<i_s\le n}\Big(\prod_{k=1}^{s}\frac{1}{N_{i_k}}\Big)\Big(\rho\frac{\overline{C}}{\underline{C}}\Big)^s.$$

Note that

$$\sum_{s=1}^{n+1}\sum_{0\le i_1<\dots<i_s\le n}\Big(\prod_{k=1}^{s}\frac{1}{N_{i_k}}\Big)\Big(\rho\frac{\overline{C}}{\underline{C}}\Big)^s=\prod_{s=0}^{n}\Big(1+\rho\frac{\overline{C}}{\underline{C}}\frac{1}{N_s}\Big)-1\,,$$

and for $N_0>C(n+1),\dots,N_n>C(n+1)$

$$\prod_{s=0}^{n}\Big(1+\rho\frac{\overline{C}}{\underline{C}}\frac{1}{N_s}\Big)-1\le 2\rho\frac{\overline{C}}{\underline{C}}\sum_{p=0}^{n}\frac{1}{N_p}\,,$$

(see for instance the proofs of Theorem 5.1 and Corollary 5.2 of [6]). It follows that

$$\Big|\mathbb{E}\Big[\frac{(\gamma_n^{N_{0:n}})^{\otimes 2}(F)}{\gamma_n(1)^2}\Big]-\eta_n^{\otimes 2}(F)\Big| \le 8C\|F\|_\infty\sum_{p=0}^{n}\frac{1}{N_p}\,,$$

with $C=\rho\frac{\overline{C}}{\underline{C}}$; the proof is concluded. □

Proof of Theorem 1. Throughout the proof $C<+\infty$ is a constant whose value may change from line-to-line. It will not depend on the level index. By [23, Proposition 2.3]

$$\mathbb{E}\Big[\Big(\frac{\gamma_L^{N_{0:L-1}}(1)}{\gamma_L(1)}-1\Big)^2\Big]=\sum_{p=0}^{L-1}\frac{1}{N_p}\mathbb{E}[T_{p,L}^{N_{0:p}}] \tag{25}$$

where

$$T_{p,L}^{N_{0:p}}=\Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2\Big(\eta_p^{N_p}(h_{p,L}^2)-\eta_p^{N_p}(h_{p,L})^2+\eta_p^{N_p}(h_{p,L})\eta_p^{N_p}\big(\frac{G_p}{\eta_p(G_p)}-1\big)\Big)$$

and we use the short-hand for $0\le p\le n$, $x\in E$:

$$h_{p,n}(x)=\frac{Q_{p,n}(1)(x)}{\prod_{q=p}^{n-1}\eta_q(G_q)}.$$

Now, one has almost surely that

$$T_{p,L}^{N_{0:p}} = \Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2 \times$$

$$\Big(\eta_p^{N_p}([h_{p,L}-1]^2) - \eta_p^{N_p}(h_{p,L}-1)^2 + \eta_p^{N_p}(h_{p,L})\eta_p^{N_p}(\frac{G_p}{\eta_p(G_p)}-1)\Big).$$

As

$$\begin{aligned}\mathbb{E}[T_{p,L}^{N_{0:p}}] &= \mathbb{E}\Bigg[\Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2\Big(\eta_p^{N_p}([h_{p,L}-1]^2) - \eta_p^{N_p}(h_{p,L}-1)^2\Big)\Bigg] + \\ & \quad \mathbb{E}\Bigg[\Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2\eta_p^{N_p}(h_{p,L})\eta_p^{N_p}(\frac{G_p}{\eta_p(G_p)}-1)\Bigg] \qquad (26)\end{aligned}$$

we will consider controlling the two terms on the R.H.S. of (26) separately.
First term on the R.H.S. of (26).
We have, almost surely that

$$\begin{aligned}\eta_p^{N_p}([h_{p,L}-1]^2) - \eta_p^{N_p}(h_{p,L}-1)^2 &\leq C\|h_{p,L}-1\|_\infty^2 \\ &\leq C\Big(\sum_{q=p}^{L-1}\Big\|\frac{G_q}{\eta_p(G_q)}-1\Big\|_\infty\Big)^2\end{aligned}$$

where we have applied Lemma 1 to go to the second line. Then by Proposition 1 as $N_0 > cL, \dots, N_{L-1} > cL$

$$\mathbb{E}\Big[\Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2\Big] \leq C.$$

So we have shown that

$$\mathbb{E}\Bigg[\Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2\Big(\eta_p^{N_p}([h_{p,L}-1]^2) - \eta_p^{N_p}(h_{p,L}-1)^2\Big)\Bigg] \leq$$

$$C\Big(\sum_{q=p}^{L-1}\Big\|\frac{G_q}{\eta_p(G_q)}-1\Big\|_\infty\Big)^2. \qquad (27)$$

Second term on the R.H.S. of (26).
We have almost surely that

$$\Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2\eta_p^{N_p}(h_{p,L})\eta_p^{N_p}(\frac{G_p}{\eta_p(G_p)}-1) =$$

$$\frac{1}{\gamma_p(1)^2}(\gamma_p^{N_{0:p}})^{\otimes 2}\Big(h_{p,L}\otimes\Big(\frac{G_p}{\eta_p(G_p)}-1\Big)\Big)$$

and note that,

$$\eta_p^{\otimes 2}\Big(h_{p,L}\otimes\Big(\frac{G_p}{\eta_p(G_p)}-1\Big)\Big)=0.$$

So by Proposition 1 as $N_0 > cL, \dots, N_{L-1} > cL$ and (24)

$$\Bigg|\mathbb{E}\Bigg[\Big(\frac{\gamma_p^{N_{0:p-1}}(1)}{\gamma_p(1)}\Big)^2\eta_p^{N_p}(h_{p,L})\eta_p^{N_p}\big(\frac{G_p}{\eta_p(G_p)}-1\big)\Bigg]\Bigg| \leq C\Big\|\frac{G_p}{\eta_p(G_p)}-1\Big\|_\infty\frac{(p+1)}{N_p}. \tag{28}$$

Combining (25) with (26) and after applying the triangular inequality, the bounds (27) and (28) complete the proof.

□

C Proofs for Theorem 2

In order to prove Theorem 2, we prove two technical Lemmas. The proof of Theorem 2 is then given. Some of the proofs in this Section will use Proposition 1 in Appendix B.

Lemma 2. *Let $n \geq 1$ and $f_1, f_2 \in \mathcal{B}_b(E)$ then*

$$\mathbb{E}\Big[[\gamma_n^{N_{0:n}}-\gamma_n](f_1)[\gamma_n^{N_{0:n}}-\gamma_n](f_2)\Big] = \mathbb{E}[(\gamma_n^{N_{0:n}})^{\otimes 2}(f_1\otimes f_2)]-\gamma_n(1)^2\eta_n^{\otimes 2}(f_1\otimes f_2).$$

Proof. We have

$$\mathbb{E}\Big[[\gamma_n^{N_{0:n}}-\gamma_n](f_1)[\gamma_n^{N_{0:n}}-\gamma_n](f_2)\Big] =$$

$$\mathbb{E}[\gamma_n^{N_{0:n}}(f_1)\gamma_n^{N_{0:n}}(f_2)]-\gamma_n(f_2)\mathbb{E}[\gamma_n^{N_{0:n}}(f_1)]-\gamma_n(f_1)\mathbb{E}[\gamma_n^{N_{0:n}}(f_2)]+\gamma_n(f_1)\gamma_n(f_2) =$$

$$\mathbb{E}[\gamma_n^{N_{0:n}}(f_1)\gamma_n^{N_{0:n}}(f_2)] - \gamma_n(f_2)\gamma_n(f_1) - \gamma_n(f_1)\gamma_n(f_2) + \gamma_n(f_1)\gamma_n(f_2)$$

where the unbiased property of the normalizing constant has been used to go to the last line. Then it follows that

$$\mathbb{E}[\gamma_n^{N_{0:n}}(f_1)\gamma_n^{N_{0:n}}(f_2)] - \gamma_n(f_2)\gamma_n(f_1) - \gamma_n(f_1)\gamma_n(f_2) + \gamma_n(f_1)\gamma_n(f_2) =$$

$$\mathbb{E}[(\gamma_n^{N_{0:n}})^{\otimes 2}(f_1\otimes f_2)] - \gamma_n(1)^2\eta_n^{\otimes 2}(f_1\otimes f_2)$$

which concludes the proof. □

Lemma 3. *Assume (A1-2). Then there exists a $C < +\infty$ such that for any $2 \leq q < p$, $N_0 \geq N_1 \geq \cdots \geq N_{q-2} > C(q-1)$:*

$$|\mathbb{E}[[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](G_{p-2}(G_{p-1}-1))[\gamma_{q-2}^{N_{0:q-2}} - \gamma_{q-2}](G_{q-2}(G_{q-1}-1))]| \leq$$

$$\frac{c(q-1)\gamma_{q-2}(1)^2}{N_{q-2}}\Big\|G_{q-2}(G_{q-1}-1)Q_{q-2,p-2}(G_{p-2}(G_{p-1}-1))\Big\|_\infty.$$

Proof. From [9, Proposition 7.4.1] we have

$$\mathbb{E}[[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](G_{p-2}(G_{p-1}-1))[\gamma_{q-2}^{N_{0:q-2}} - \gamma_{q-2}](G_{q-2}(G_{q-1}-1))] =$$

$$\sum_{s_1=0}^{p-2}\sum_{s_2=0}^{q-2}\mathbb{E}\Big[\gamma_{s_1}^{N_{0:s_1-1}}(1)[\eta_{s_1}^{N_{s_1}} - \Phi_{s_1}(\eta_{s_1-1}^{N_{s_1-1}})](Q_{s_1,p-2}(\overline{G}_p))\times$$

$$\gamma_{s_2}^{N_{0:s_2-1}}(1)[\eta_{s_2}^{N_{s_2}} - \Phi_{s_2}(\eta_{s_2-1}^{N_{s_2-1}})](Q_{s_2,q-2}(\overline{G}_q))\Big]$$

where we have used the shorthand $\overline{G}_s = G_{s-2}(G_{s-1}-1)$ for any $s \geq 2$. For any $s \geq 0$, $f \in \mathcal{B}_b(E)$

$$\mathbb{E}[\gamma_s^{N_{0:s-1}}(1)[\eta_s^{N_s} - \Phi_s(\eta_{s-1}^{N_{s-1}})](f)|\mathscr{F}_{s-1}^{N_{0:s-1}}] = 0$$

thus, it follows that

$$\mathbb{E}[[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](G_{p-2}(G_{p-1}-1))[\gamma_{q-2}^{N_{0:q-2}} - \gamma_{q-2}](G_{q-2}(G_{q-1}-1))] =$$

$$\sum_{s=0}^{q-2}\mathbb{E}[\gamma_s^{N_{0:s-1}}(1)^2[\eta_s^{N_s} - \Phi_s(\eta_{s-1}^{N_{s-1}})]^{\otimes 2}(Q_{s,p-2}(\overline{G}_p)\otimes Q_{s,q-2}(\overline{G}_q))].$$

Now for any $n \geq 1$, $f_1, f_2 \in \mathcal{B}_b(E)$, one can show, using almost the same calculations as above, that the following holds

$$\sum_{s=0}^{n}\mathbb{E}[\gamma_s^{N_{0:s-1}}(1)^2[\eta_s^{N_s} - \Phi_s(\eta_{s-1}^{N_{s-1}})]^{\otimes 2}(Q_{s,n}(f_1)\otimes Q_{s,n}(f_2))] =$$

$$\mathbb{E}\Big[[\gamma_n^{N_{0:n}} - \gamma_n](f_1)[\gamma_n^{N_{0:n}} - \gamma_n](f_2)\Big].$$

Using this equality with $n = q-2$, and the fact that $Q_{s,p-2} = Q_{s,q-2}Q_{q-2,p-2}$, finally

$$\mathbb{E}[[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](G_{p-2}(G_{p-1}-1))[\gamma_{q-2}^{N_{0:q-2}} - \gamma_{q-2}](G_{q-2}(G_{q-1}-1))] =$$

$$\mathbb{E}\Big[[\gamma_{q-2}^{N_{0:n}} - \gamma_{q-2}](Q_{q-2,p-2}(\overline{G}_p))[\gamma_{q-2}^{N_{0:n}} - \gamma_{q-2}](\overline{G}_q)\Big].$$

Then, by Lemma 2:

$$\mathbb{E}[[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](G_{p-2}(G_{p-1}-1))[\gamma_{q-2}^{N_{0:q-2}} - \gamma_{q-2}](G_{q-2}(G_{q-1}-1))] =$$

$$\mathbb{E}[(\gamma_{q-2}^{N_{0:q-2}})^{\otimes 2}(Q_{q-2,p-2}(\overline{G}_p)\otimes \overline{G}_q)] - \gamma_{q-2}(1)^2\eta_{q-2}^{\otimes 2}(Q_{q-2,p-2}(\overline{G}_p)\otimes \overline{G}_q).$$

Then, one can apply Proposition 1 to obtain that

$$|\mathbb{E}[[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](G_{p-2}(G_{p-1}-1))[\gamma_{q-2}^{N_{0:q-2}} - \gamma_{q-2}](G_{q-2}(G_{q-1}-1))]| \leq$$

$$\frac{C(q-1)\gamma_{q-2}(1)^2}{N_{q-2}}\Big\|G_{q-2}(G_{q-1}-1)Q_{q-2,p-2}(G_{p-2}(G_{p-1}-1))\Big\|_\infty.$$

□

Proof of Theorem 2. Throughout the proof $C < +\infty$ is a constant whose value may change from line-to-line. It will not depend on the level index. We have

$$\mathbb{E}\Big[\Big(\frac{\tilde{\gamma}_L^{N_{0:L-2}}(1)}{\gamma_L(1)} - 1\Big)^2\Big] \leq$$

$$\frac{1}{\gamma_L(1)^2}\mathbb{E}[[\eta_0^{N_0} - \eta_0](G_0)^2] + \frac{1}{\gamma_L(1)^2}\mathbb{E}[(\sum_{p=2}^{L}[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](\overline{G}_p))^2].$$

As $\gamma_L(1) = Z_L/Z_0 \geq \underline{C}/\overline{C}$ it follows by standard results for i.i.d. random variables that one has

$$\frac{1}{\gamma_L(1)^2}\mathbb{E}[[\eta_0^{N_0} - \eta_0](G_0)^2] \leq \frac{C}{N_0}.$$

Now

$$\mathbb{E}[(\sum_{p=2}^{L}[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](\overline{G}_p))^2] = \sum_{p=2}^{L}\gamma_{p-2}(1)^2\mathbb{E}\Big[\frac{\gamma_{p-2}^{N_{0:p-2}}(\overline{G}_p)^2}{\gamma_{p-2}(1)^2} - \eta_{p-2}(\overline{G}_p)^2\Big]$$

$$+2\sum_{p=2}^{L}\sum_{q=2}^{p-1}\mathbb{E}[[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](\overline{G}_p)[\gamma_{q-2}^{N_{0:q-2}} - \gamma_{q-2}](\overline{G}_q)].$$

Applying Propositon 1 to the terms in the single sum and Lemma 3 to the terms in the double sum, we have that

$$\mathbb{E}[(\sum_{p=2}^{L}[\gamma_{p-2}^{N_{0:p-2}} - \gamma_{p-2}](\overline{G}_p))^2] \leq C\Big(\sum_{p=2}^{L}\gamma_{p-2}(1)^2\frac{(p-1)}{N_p}\|\overline{G}_p\|_\infty^2$$

$$+\sum_{p=2}^{L}\sum_{q=2}^{p-1}\frac{(q-1)\gamma_{q-2}(1)^2}{N_{q-2}}\Big\|G_{q-2}(G_{q-1}-1)Q_{q-2,p-2}(G_{p-2}(G_{p-1}-1))\Big\|_\infty\Big).$$

As $\gamma_{p-2}(1) \leq \overline{C}/\underline{C}$, $\gamma_L(1) \geq \underline{C}/\overline{C}$ one has

$$\frac{1}{\gamma_L(1)^2}\sum_{p=2}^{L}\gamma_{p-2}(1)^2\frac{(p-1)}{N_p}\|\overline{G}_p\|_\infty^2 \leq C\sum_{p=2}^{L}\frac{(p-1)}{N_{p-2}}\|G_{p-1}-1\|_\infty^2.$$

We have

$$\frac{1}{\gamma_L(1)^2}\sum_{p=2}^{L}\sum_{q=2}^{p-1}\frac{(q-1)\gamma_{q-2}(1)^2}{N_{q-2}}\Big\|G_{q-2}(G_{q-1}-1)Q_{q-2,p-2}(G_{p-2}(G_{p-1}-1))\Big\|_\infty =$$

$$\sum_{p=2}^{L}\sum_{q=2}^{p-1}\frac{c(q-1)\gamma_{q-2}(1)}{\gamma_L(1)N_{q-2}}\frac{1}{\eta_{q-2}(Q_{q-2,p-2}(1))}\frac{Z_{p-1}}{Z_L}\times$$

$$\Big\| G_{q-2}(G_{q-1}-1)Q_{q-2,p-2}(G_{p-2}(G_{p-1}-1))\Big\|_\infty.$$

Then as $\gamma_{q-2}(1) \leq \overline{C}/\underline{C}$, $\gamma_L(1) \geq \underline{C}/\overline{C}$, $Z_{p-1} \leq C$, $Z_L \geq C$ and by [6, Lemma 4.1]

$$\frac{Q_{q-2,p-2}(G_{p-2}(G_{p-1}-1))}{\eta_{q-2}(Q_{q-2,p-2}(1))} \leq C\|G_{p-1}-1\|_\infty$$

we have

$$\frac{1}{\gamma_L(1)^2}\sum_{p=2}^{L}\sum_{q=2}^{p-1}\frac{(q-1)\gamma_{q-2}(1)^2}{N_{q-2}}\Big\| G_{q-2}(G_{q-1}-1)Q_{q-2,p-2}(G_{p-2}(G_{p-1}-1))\Big\|_\infty \leq$$

$$C\sum_{p=2}^{L}\sum_{q=2}^{p-1}\frac{(q-1)}{N_{q-2}}\|G_{p-1}-1\|_\infty\|G_{q-1}-1\|_\infty.$$

From here one can easily conclude. □

References

[1] Andrieu, C., & Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, **37**, 697-725.

[2] Beskos, A., Crisan, D. & Jasra, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.*, **24**, 1396–1445.

[3] Beskos, A., Jasra, A., Law, K. J. H, Tempone, R. & Zhou, Y. (2017). Multilevel sequential Monte Carlo samplers. *Stoch. Proc. Appl.* (to appear).

[4] Bickel, P., Li, B. & Bengtsson, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions. In *Pushing the Limits of Contemporary Statistics*, B. Clarke & S. Ghosal, Eds, 318–329, IMS.

[5] Bungartz, H-J., & Griebel, M. (2004). Sparse grids. *Acta numerica* **13.1**, 147-269.

[6] Cerou, F., Del Moral, P., & Guyader, A. (2011). A non-asymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincaire*, **47**, 629–649.

[7] Ciarlet, P. G. (1978). *The finite element method for elliptic problems.* Studies in Applied Mathematics, Vol. 4. North-Holland Publishing Company.

[8] Cotter,S. L., Roberts, G.O., Stuart, A.M. & White, D.. (2013). MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, **28**, 424-446.

[9] Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications.* Springer: New York.

[10] Del Moral, P., Doucet, A. & Jasra, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, **68**, 411–436.

[11] Del Moral, P., Jasra, A.& Law, K. J. H. (2017). Multilevel Sequential Monte Carlo: Mean Square Error Bounds under Verifiable Conditions. *Stoch. Anal. Appl.*, **35**, 478-498.

[12] Giles, M. B. (2008). Multilevel Monte Carlo path simulation. *Op. Res.* **56**, 607-617.

[13] Giles, M. B (2015). Multilevel Monte Carlo methods. *Acta Numerica*, **24**, 259-328.

[14] Heinrich, S. (2001). Multilevel Monte Carlo methods. *Large-scale scientific computing.* Springer Berlin Heidelberg, 2001. 58-67.

[15] Hoang, V., Schwab, C. & Stuart, A. (2013). Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Prob.*, **29**, 085010.

[16] Hoel, H., Law, K. J., & Tempone, R. (2016). Multilevel ensemble Kalman filtering. *SIAM J. Numer. Anal.*, **54**, 1813–1839.

[17] Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statist. Sci.*, **14**, 382-401.

[18] Jasra A., Law K. J. H. & Zhou, Y. (2016). Forward and inverse uncertainty quantification using multilevel Monte Carlo algorithms for an elliptic nonlocal equation. *Intl., J. Uncert. Quant.*, **6**, 501-514.

[19] Jasra, A., Kamatani, K., Law, K. J., & Zhou, Y. (2015). Multilevel particle filter. *arXiv preprint arXiv:1510.04977.*

[20] Ketelsen, C., Scheichl, R. & Teckentrup, A. L. (2015). A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA J. Uncer. Quant.*, **3**, 1075-1108

[21] Le Matre, O. P., & Knio, O. M. (2010). *Introduction: Uncertainty Quantification and Propagation.* Springer Netherlands.

[22] Rhee, C. H., & Glynn, P. W. (2015). Unbiased estimation with square root convergence for SDE models. *Op. Res.*, **63**, 1026–1043.

[23] Schweizer, N. (2012). Non-asymptotic error bounds for sequential MCMC and stability of Feynman-Kac operators. *arXiv preprint arXiv:1204.2382v1.*

[24] Walsh, J. B. (1986) *An Introduction to Stochastic Partial Differential Equations.* Springer Berlin Heidelberg.

[25] Wasserman, L. (2000). Bayesian model selection and model averaging. *J. Math. Psych.*, **44**, 92-107.

[26] Zhou, Y., Johansen, A. M. & Aston, J. A. D. (2016). Towards automatic model comparison: An adaptive sequential Monte Carlo approach. *J. Comp. Graph. Statist.*, **25**, 701-726.