

EFFICIENT METHODS FOR THE ESTIMATION OF HOMOGENIZED COEFFICIENTS

J.-C. MOURRAT

ABSTRACT. The main goal of this paper is to define and study new methods for the computation of effective coefficients in the homogenization of divergence-form operators with random coefficients. The methods introduced here are proved to have optimal computational complexity, and are shown numerically to display small constant prefactors. In the spirit of multiscale methods, the main idea is to rely on a progressive coarsening of the problem, which we implement via a generalization of the Green-Kubo formula. The technique can be applied more generally to compute the effective diffusivity of any additive functional of a Markov process. In this broader context, we also discuss the alternative possibility of using Monte-Carlo sampling, and show how a simple one-step extrapolation can considerably improve the performance of this alternative method.

1. INTRODUCTION

Let $-\nabla \cdot \mathbf{a}(x)\nabla$ be a divergence-form operator with random coefficients. If the law of the matrix field $(\mathbf{a}(x))$ is stationary and ergodic, then the operator homogenizes over large scales to an operator with constant, deterministic coefficients $-\nabla \cdot \bar{\mathbf{a}}\nabla$. The main goal of this paper is to explore numerical methods to estimate the homogenized matrix $\bar{\mathbf{a}}$. We can formalize this task as follows.

Problem. *Find an algorithm that, given $\delta > 0$ and a realization of the random coefficient field $(\mathbf{a}(x))$, outputs a matrix $\hat{\mathbf{a}}_\delta$ such that*

$$(1.1) \quad \mathbb{E} [|\bar{\mathbf{a}} - \hat{\mathbf{a}}_\delta|^2]^{\frac{1}{2}} \leq \delta.$$

In (1.1), we chose to measure the quality of the approximation $\hat{\mathbf{a}}_\delta$ in an L^2 sense for definiteness, but other choices (e.g. convergence in probability, or higher moments) are equally valid, and in fact would only marginally alter the results presented below. The point is that $\hat{\mathbf{a}}_\delta$ should typically be within δ or less from $\bar{\mathbf{a}}$.

Naturally, the real question is to find an *efficient* algorithm. We focus on a discrete-space setting, and assume throughout that the coefficients are uniformly elliptic, independent and identically distributed (i.i.d.). In order to measure computational effort, we use a slightly loosely defined notion of “elementary” operation: any memory access, boolean operation, floating point addition or

Date: April 6, 2018.

2010 Mathematics Subject Classification. 82B80; 35B27; 82B28.

Key words and phrases. homogenization, multiscale methods.

multiplication counts as one operation. We first give a lower bound on the number of operations any algorithm must use.

Proposition 1.1 (Complexity lower bound). *No algorithm can output an approximation $\widehat{\mathbf{a}}_\delta$ of $\bar{\mathbf{a}}$ in the sense of (1.1) within $o(\delta^{-2})$ operations.*

The main method we propose to compute an approximation of $\bar{\mathbf{a}}$ is as follows. We fix $\xi \in \mathbb{R}^d$ of unit norm, set

$$v_{-1}(x) := (\nabla \cdot \mathbf{a}\xi)(x) \quad (x \in \mathbb{Z}^d),$$

and for each $k \in \mathbb{N}$, we let v_k be the solution to

$$(1.2) \quad (2^{-k} - \nabla \cdot \mathbf{a}\nabla)v_k = 2^{-k}v_{k-1} \quad \text{on } \mathbb{Z}^d.$$

(We refer to Section 8 for precise notation and assumptions.) Note that the functions $(v_k)_{k \geq -1}$ depend on the coefficient field \mathbf{a} , although we keep this dependence implicit. For each $r \geq 0$, we write

$$B_r := \{-\lfloor r \rfloor, \dots, \lfloor r \rfloor\}^d.$$

Theorem 1.2 (Efficient approximation of $\bar{\mathbf{a}}$). *For each $\varepsilon \in (0, \frac{d-1}{2d})$, there exists a constant $C < \infty$ such that the following holds. Fix $n \in \mathbb{N}$, and denote*

$$(1.3) \quad r_k := 2^{n - (\frac{1}{2} - \varepsilon)k} \quad (k \in \{0, \dots, n\}),$$

$$(1.4) \quad \widehat{\sigma}_n^2 := \sum_{k=0}^n \frac{1}{|B_{r_k}|} \sum_{x \in B_{r_k}} 2^k (v_{k-1}(x)v_k(x) + v_k^2(x)).$$

We have

$$(1.5) \quad \mathbb{E} \left[\left(\xi \cdot \bar{\mathbf{a}}\xi - \mathbb{E}[\xi \cdot \mathbf{a}\xi] + \widehat{\sigma}_n^2 \right)^2 \right]^{\frac{1}{2}} \leq C 2^{-\frac{dn}{2}}.$$

Moreover, the quantity $\widehat{\sigma}_n^2$ can be computed in at most $Cn2^{dn}$ operations.

By the central limit theorem, the quantity $\mathbb{E}[\xi \cdot \mathbf{a}\xi]$ can be computed at precision $\delta > 0$ in $C\delta^{-2}$ operations. Hence, Theorem 1.2 gives us a method that requires $\delta^{-2} \log(\delta^{-1})$ operations to compute an approximation of $\bar{\mathbf{a}}$ at precision $\delta > 0$ in the sense of (1.1), essentially matching the lower bound given by Proposition 1.1.

For $\varepsilon = 0$, the statement of Theorem 1.2 remains valid provided that we replace the right side of (1.5) by $Cn2^{-\frac{dn}{2}}$. I expect that the sharp upper bound in this case is $Cn^{\frac{1}{2}}2^{-\frac{dn}{2}}$. Whatever may be, this additional multiplicative factor of n or $n^{\frac{1}{2}}$ does not degrade the bound much. For simplicity, it is the version with $\varepsilon = 0$ that was implemented numerically.

The proof of Theorem 1.2 relies in particular on the assumption that the law of the coefficient field \mathbb{P} is invariant under the action of \mathbb{Z}^d . If this assumption is weakened and the law is only assumed to be invariant under the action of a sublattice of \mathbb{Z}^d which may be unknown to us a priori, then one can use masks in the spatial averages over B_{r_k} appearing in (1.4) in order to avoid boundary layer effects. When weakening the independence assumption on the coefficients, I expect Theorem 1.2 to remain unchanged provided that the field $x \mapsto \mathbf{a}(x)$ is

sufficiently uncorrelated for partial sums over boxes to satisfy a central limit theorem.

The power of the method presented in Theorem 1.2 comes from the fact that the domain of interest shrinks rapidly as k increases. Moreover, when k is of order one, the condition number of the elliptic problems we need to solve and the size of the associated boundary layers are also of order one. As k increases, condition numbers and boundary layers become larger, but these adverse effects are more than compensated by the reduction of size of the domain of interest. In fact, in order to prove the complexity upper bound in Theorem 1.2, we do not need to use any preconditioner for the resolution of the elliptic problems: a direct application of the conjugate gradient method suffices. Moreover, the overhead caused by the presence of boundary layers only participates to a negligible fraction of the total computational time: this fraction of time is of the order of $2^{-\frac{n}{2}}$ in dimension $d = 2$, and of the order of 2^{-n} in dimension $d \geq 3$.

It is straightforward to adapt the method of Theorem 1.2 to accommodate for parallel computing. Indeed, we will see that the correlation length of the random field

$$(1.6) \quad x \mapsto v_{k-1}(x)v_k(x) + v_k^2(x)$$

is of the order of $2^{\frac{k}{2}}$. Computing spatial averages of this random field is the best strategy if all computations are done sequentially. On the other hand, if parallel computing is available, then we may proceed differently and compute separately $v_k^{(1)}, v_k^{(2)}, \dots$, based on independent copies of the coefficient field but otherwise defined in the same way as v_k , and then average the results. The general rule is that we may replace any spatial average

$$\frac{1}{|B_r|} \sum_{x \in B_r} (v_{k-1}(x)v_k(x) + v_k^2(x))$$

by

$$\frac{1}{\lceil 2^{-\frac{kd}{2}} r^d \rceil} \sum_{i=1}^{\lceil 2^{-\frac{kd}{2}} r^d \rceil} \left(v_{k-1}^{(i)}(0)v_k^{(i)}(0) + \left(v_k^{(i)}(0) \right)^2 \right),$$

or with any other intermediate combination of spatial averaging and independent sampling.

Disregarding logarithmic factors, the previously known best procedure for computing $\bar{\mathbf{a}}$ at precision $\delta > 0$ consists in averaging δ^{-1} samples of the energy of an approximate corrector calculated on a box of side length $\delta^{-\frac{1}{d}}$ —see [24] and [23, Section 3.3.4] for details. We will refer to this technique as the representative volume element (RVE) method. If no preconditioner is used, then the computational cost of this method is of the order of $\delta^{-(2+\frac{1}{d})}$, up to logarithmic factors. While efficient preconditioners can bring this estimate back to being of the order of δ^{-2} , up to logarithmic factors, the prefactors involved may be very large in practice. I did some modest tests with an incomplete Cholesky factorization as a preconditioner. While this did reduce the number of iterations involved in the conjugate gradient method, it was not sufficient to compensate for the induced overhead and resulted in degraded performance.

Other preconditioners will certainly show better performance (using for instance the homogeneous Laplacian, which can be solved in almost linear time by fast Fourier transform), and I did not investigate this question further. In any case, the computational overhead caused by boundary layers for the RVE method in fact takes up an asymptotically full proportion of the computational time, since boundary layers occupy a volume of the order of $\delta^{-1} \log^d(\delta^{-1})$ on each of the δ^{-1} domains. Naturally, this observation calls for an adjustment of the strategy: one should solve fewer problems, each of larger volume. But this runs counter to the fact that it is computationally best to solve smaller elliptic problems, and only suboptimal compromises can be found in this way.

An additional advantage of the method described in Theorem 1.2 is that it is cumulative: all computations involved in the evaluation of $\widehat{\sigma}_n^2$ remain useful if the approximation needs to be refined. In contrast, in the older method, computations done for a coarse precision must be thrown away if greater precision is then desired.

The theoretical arguments presented here are supplemented by numerical tests. The numerical results confirm the theoretical predictions, and display small constant prefactors.

As was said above, we assume that the coefficient field ($\mathbf{a}(x)$) is uniformly elliptic and i.i.d., and we describe algorithms that perform well under these assumptions. In practice, it seems more appropriate to implement more “agnostic” algorithms, that would evaluate on the fly the behavior of the mean and variance of the spatial averages appearing in (1.4) as r and k vary, and adapt the numerical scheme dynamically. We will not discuss this possibility any further here.

It would be interesting to investigate how the new method presented here adapts to the continuous-space setting. In this case, I expect that the error analysis remains unchanged. The complexity analysis would however also have to take into account the cost of the small-scale resolution of the equations, below the typical length scale of the correlations of the coefficient field. This problem is separate from the homogenization phenomenon, and is likely to come as the same multiplicative factor in the complexity analysis of any method.

Another interesting line of further research concerns the behavior of the effective conductivity of percolation clusters. We refer to [36, 44] for very nice surveys on this problem. The method presented here should allow for a sharper numerical analysis of the effective conductivity as the percolation probability approaches criticality. Despite immense progress on the understanding of critical and near-critical percolation in two dimensions, it is still not known whether the effective conductivity behaves as a power law near criticality (let alone compute the exponent) in small dimensions. See however [38, 37, 19], as well as [13, 43, 14, 15] for very fine results in high dimensions.

It turns out that the underlying idea of the method can be recast in the more general framework of additive functionals of Markov processes. The paper is thus split into two parts. In the first part, we describe the general strategy in this broader context. We also discuss alternative methods based on Monte-Carlo sampling, and show that a simple one-step extrapolation can in

general improve the efficiency of this approach dramatically. The second part of the paper is concerned with the computation of $\bar{\mathbf{a}}$ per se. In particular, we prove Theorem 1.2 and report on numerical results there. Readers who are only interested in the problem of computing $\bar{\mathbf{a}}$ are encouraged to go directly to the second part of this paper, where we explain how to extract relevant information from the first part without having to refer to Markov processes.

We conclude this introduction with a brief review of related work. The core idea of the method presented in Theorem 1.2 is to decompose the problem into a series of scales. The approach is inspired by the renormalization scheme introduced in [10], which progressively coarsens the operator and relies on a “quadratic response” behavior around the homogenization limit—see also the discussion below (8.7). Other relevant recent references for quantitative homogenization include [53, 30, 31, 28, 50, 12, 29, 11, 9, 32, 8].

More generally, the approach presented here is similar in spirit to multiscale methods [18, 39, 1]. In this context, it was quickly realized that the resolution of cell problems or the computation of adapted finite-element base functions can display so-called “resonant errors” due to inappropriate boundary condition [34, 35, 60, 22]. A powerful approach was introduced in [17, 31, 24, 25, 53, 26], based on the introduction of a massive term that smoothly makes the faraway boundary condition irrelevant. The method presented here is faithful to this idea; see also [49, 33, 6].

A different line of research focused on developing more efficient methods for computing homogenized coefficients of small random perturbations of a simpler, e.g. homogeneous, medium, in the spirit of earlier fundamental insights [51, 59]. A typical example is that of spherical inclusions randomly placed at a small volume fraction $p > 0$ in an otherwise constant medium. One looks for asymptotic formulas of the homogenized coefficients involving easily computable coefficients, in the regime $p \ll 1$. We refer to [42, 56, 16, 4, 5, 2, 3, 54, 21] for a mathematical analysis of the problem. Note that the seemingly simpler task of determining the value of the volume fraction p at precision δ from a sample of the coefficient field already requires of the order of δ^{-2} operations (the proof of the lower bound is a minor variant of the proof of Proposition 1.1).

Several techniques have been explored to reduce the size of the fluctuations of quantities of interest, typically by a constant multiplicative factor. We refer to [45] for a recent survey.

The knowledge of the homogenized matrix allows to produce approximations of elliptic equations with rapidly oscillating coefficients that become sharp in the limit of infinite separation of scales. One may however want to actually compute the solution of the elliptic problem at high precision, in a regime of large but finite scale separation. For this purpose, a multigrid-type numerical scheme was discovered in [7] that uses the homogenized operator as coarsened operator.

Part I. General methods

2. MAIN RESULTS

In this first part of the paper, we explore the problem of computing the effective diffusivity of general additive functionals of a Markov process. Let $(X(t))_{t \geq 0}$ be a Feller process taking values in a metric space \mathcal{X} , and denote the associated family of probability laws by $(\mathbf{P}_x)_{x \in \mathcal{X}}$, with corresponding expectations $(\mathbf{E}_x)_{x \in \mathcal{X}}$. We assume that there exists a probability measure \mathbb{P} on \mathcal{X} (with corresponding expectation \mathbb{E}) such that the process (X_t) is reversible and ergodic with respect to \mathbb{P} ; we denote the law of $(X(t))$ started from the measure \mathbb{P} by $\mathbf{P}_{\mathbb{P}}$ (with corresponding expectation $\mathbf{E}_{\mathbb{P}}$). The assumption of reversibility will be dropped shortly, and is only meant to facilitate the exposition here. Given $f \in L^2(\mathbb{P})$ of mean zero, one can check using reversibility that the limit

$$(2.1) \quad \sigma^2(f) := \lim_{t \rightarrow +\infty} \mathbf{E}_{\mathbb{P}} \left[\left(\frac{1}{\sqrt{t}} \int_0^t f(X(s)) \, ds \right)^2 \right] \in [0, +\infty]$$

always exists. Denote by \mathcal{L} the (non-negative) infinitesimal generator of $(X(t))$, and $\langle f, g \rangle := \mathbb{E}[fg]$. An equivalent characterization of $\sigma^2(f)$ is given by

$$(2.2) \quad \sigma^2(f) = 2 \langle f, \mathcal{L}^{-1} f \rangle := 2 \lim_{\lambda \rightarrow 0} \langle f, (\lambda + \mathcal{L})^{-1} f \rangle.$$

Kipnis and Varadhan [40] showed that the condition $\sigma^2(f) < \infty$ implies that the additive functional

$$(2.3) \quad \frac{1}{\sqrt{t}} \int_0^t f(X(s)) \, ds$$

converges in law under $\mathbf{P}_{\mathbb{P}}$ to a centered Gaussian random variable of variance $\sigma^2(f)$ as t tends to infinity.

Our goal is to discuss practical ways to compute $\sigma^2(f)$. There are two standard methods to address this question (see for instance [23]). The first one uses the formula (2.1) as the basis for a Markov chain Monte-Carlo (MCMC) algorithm. The second method dispenses with the simulation of the Markov chain, and uses the formula (2.2) instead. The problem then becomes that of computing an approximation of $\mathcal{L}^{-1}f$, and computing the average.

We study MCMC methods in Section 4. The main contribution of this section is to show that a one-step extrapolation of the naive MCMC algorithm yields significant improvement of the approximation in most cases.

We then explore methods based on the formula (2.2). Similarly to multiscale methods, we aim to split the computational effort involved in the computation of $\mathcal{L}^{-1}f$ into a hierarchy of scales. A good starting point for witnessing this

multiscale decomposition is the Green-Kubo formula, which in our present context is a direct consequence of (2.2) and reads

$$(2.4) \quad \sigma^2(f) = 2 \int_0^{+\infty} \langle f, e^{-t\mathcal{L}} f \rangle dt = 2 \int_0^{+\infty} \mathbf{E}_{\mathbb{P}} [f(X(0))f(X(t))] dt.$$

We think of the part of the integral with $t \lesssim 1$ as carrying the small-scale information of the process, while the part with $t \gg 1$ carries the larger-scale information. The first, very elementary but crucial, modification to (2.4) we then bring forward is that, by reversibility,

$$(2.5) \quad \langle f, e^{-t\mathcal{L}} f \rangle = \left\langle e^{-\frac{t}{2}\mathcal{L}} f, e^{-\frac{t}{2}\mathcal{L}} f \right\rangle$$

so that, after a change of variables,

$$(2.6) \quad \sigma^2(f) = 4 \int_0^{+\infty} \langle e^{-t\mathcal{L}} f, e^{-t\mathcal{L}} f \rangle dt.$$

The point of this simple modification is that in order to compute $\mathbb{E}[(e^{-t\mathcal{L}} f)^2]$, we need to average over samples of $(e^{-t\mathcal{L}} f)^2$, and for large $t > 0$, each of these samples will be much smaller than realizations of $f e^{-t\mathcal{L}} f$ (by a square factor). For large $t > 0$, we thus need much fewer samples to compute the average on the right side of (2.5) than we do to compute the average on the left side of (2.5). This is relevant for practical computations since it is usually more difficult to compute samples of $e^{-t\mathcal{L}} f$ when t is large. In other words, we decomposed the problem of computing $\sigma^2(f)$ into a small-scale part ($t \lesssim 1$) which has a contribution of order one and is an average over many samples that are relatively easy to compute, and a large-scale part ($t \gg 1$) which refines the estimate, is computationally more demanding, but requires much fewer samples due to the own smallness of each of these samples. This is what will ultimately allow for the rapid shrinking of the domain of interest in (1.4) as k increases.

The second modification to the Green-Kubo formula we consider is meant to replace the integral over time by a sum over a moderate number of terms. The starting point is the resolvent formula, which states that for every $\lambda, \mu > 0$,

$$(2.7) \quad R_\lambda = R_\mu + (\mu - \lambda)R_\lambda R_\mu, \quad \text{where } R_\lambda := (\lambda + \mathcal{L})^{-1}.$$

For any given sequence $(\mu_k)_{k \in \mathbb{N}}$ of real numbers in $(0, 1]$, we thus apply (2.7) recursively and obtain, at least formally,

$$(2.8) \quad R_0 = \sum_{k=0}^{+\infty} \mu_0 \cdots \mu_{k-1} R_{\mu_0} \cdots R_{\mu_k}.$$

As will be explained in more details below (see Theorem 5.1), defining recursively

$$f_{-1} := f \quad \text{and} \quad \forall k \in \mathbb{N}, \quad f_k := \mu_k R_{\mu_k} f_{k-1},$$

we can use the formula (2.8) and reversibility to get

$$(2.9) \quad \langle f, \mathcal{L}^{-1} f \rangle = \sum_{k=0}^{+\infty} \mu_k^{-1} (\langle f_{k-1}, f_k \rangle + \langle f_k, f_k \rangle).$$

This is the variant of the Green-Kubo formula that we will use. The choice of the parameters μ_k can be adjusted according to the problem at hand. Here we

mostly have in mind dynamics that relax to equilibrium at a polynomial rate. In this case, it seems appropriate to choose μ_k to scale to 0 geometrically, e.g. by fixing $\mu_k \equiv 2^{-k}$. While the series is infinite, we will show that the partial sums typically converge to their limit exponentially fast.

In Section 3, we drop the reversibility assumption and define our notation precisely. We discuss Monte-Carlo methods in Section 4. We then prove (2.9) and, under suitable assumptions, give estimates on the rate of convergence of the partial sums to their limit in Section 5. We show how to adapt these considerations to discrete-time Markov chains in Section 6. Finally, in Section 7, we discuss briefly and at a heuristic level the comparative advantages of the different methods for interacting particle systems and Langevin dynamics in random potential.

3. NOTATION AND ASSUMPTIONS

In this section, we drop the reversibility assumption and clarify some of the notation used in the previous section.

From now on, we only assume that the measure \mathbb{P} is invariant for the Markov process X . We still denote by \mathcal{L} the infinitesimal generator of the process, and choose the sign convention so that $(e^{-t\mathcal{L}})_{t \geq 0}$ is the associated semigroup. That is, for every bounded and continuous function $f : \mathcal{X} \rightarrow \mathbb{R}$ and every $t \geq 0$, we have

$$(e^{-t\mathcal{L}}f)(x) = \mathbf{E}_x[f(X_t)].$$

By Jensen's inequality, for each $t \geq 0$ and $p \in [1, \infty]$, we have that $e^{-t\mathcal{L}}$ is a contraction from $L^p(\mathbb{P})$ to itself. It thus follows that for each $\lambda > 0$ and $p \in [1, \infty]$, the resolvent

$$(3.1) \quad R_\lambda := \int_0^{+\infty} e^{-\lambda t} e^{-t\mathcal{L}} dt$$

is a well-defined operator from $L^p(\mathbb{P})$ to itself, and moreover, λR_λ is a contraction from $L^p(\mathbb{P})$ to itself. One can then verify that $R_\lambda = (\lambda + \mathcal{L})^{-1}$ and that the resolvent formula (2.7) holds (see e.g. [47, Theorem 3.16 and (3.7)]). In particular, for every $\lambda > 0$ and $\mu > 0$, the resolvents R_λ and R_μ commute. We denote by \mathcal{L}^* the adjoint of \mathcal{L} in $L^2(\mathbb{P})$, and by $(e^{-t\mathcal{L}^*})_{t \geq 0}$ and $(R_\lambda^*)_{\lambda > 0}$ the associated semigroup and resolvent.

We fix a mean-zero function $f \in L^2(\mathbb{P})$. In the non-reversible setting, it is more difficult to assert whether the additive functional

$$(3.2) \quad \frac{1}{\sqrt{t}} \int_0^t f(X(s)) ds$$

satisfies a central limit theorem. We refer to [41, Chapter 2] for several results in this direction. We only record here that under relatively weak conditions, the limit

$$(3.3) \quad \sigma^2(f) := 2 \lim_{\lambda \rightarrow 0} \langle f, (\lambda + \mathcal{L})^{-1} f \rangle$$

is well-defined, and as t tends to infinity, the random variable in (3.2) converges in law under $\mathbf{P}_\mathbb{P}$ to a centered Gaussian of variance $\sigma^2(f)$ (see in particular [41, Theorem 2.7 and (2.14)]).

Part of the difficulty with non-reversible dynamics is that quantities such as $\langle f, (\lambda + \mathcal{L})^{-1} f \rangle$ or $\langle f, e^{-t\mathcal{L}} f \rangle$ are no longer necessarily nonnegative. Since the emphasis of the present paper is on numerical methods for the computation of $\sigma^2(f)$ rather than on subtleties that may arise in borderline cases, we make the simplifying assumption that

$$(3.4) \quad \int_0^{+\infty} |\langle f, e^{-t\mathcal{L}} f \rangle| dt < \infty.$$

Under this assumption, the existence and finiteness of the limit in the definition of $\sigma^2(f)$, see (3.3), are easily shown using (3.1). The validity of the classical Green-Kubo formula (2.4) is obtained similarly.

Whenever we want to go beyond the derivation of identities and want to discuss rates of convergence, we will assume that there exists $\alpha > 1$ such that for every $t \geq 0$,

$$(3.5) \quad |\langle f, e^{-t\mathcal{L}} f \rangle| \leq (1+t)^{-\alpha}.$$

Note that a possible multiplicative constant in this inequality could be absorbed after a redefinition of f . We will indicate explicitly whenever (3.4) or (3.5) is assumed.

4. MONTE-CARLO METHODS

In this section, we show that a single-step extrapolation of the naive Monte-Carlo estimator for $\sigma^2(f)$ allows for a significant improvement in computational complexity. In addition to the quantitative assumption (3.5), we also assume that

$$(4.1) \quad \sup_{t \geq 1} \mathbf{E}_{\mathbb{P}} \left[\left(\frac{1}{\sqrt{t}} \int_0^t f(X(s)) ds \right)^4 \right] \leq 1.$$

The assumption of (3.4) implies that the supremum on the left side of (4.1), with the fourth power replaced by the second power, is finite. The assumption of finiteness of the fourth moment in (4.1) (which can then be normalized to be of unit size, up to a redefinition of f) will be used to control the fluctuating part of the Monte-Carlo estimator.

In order to describe MCMC methods, we assume that we are given a family $X^{(0)}, X^{(1)}, \dots$ of processes that are independent and are each distributed according to the law $\mathbf{P}_{\mathbb{P}}$ (that is, the law of X started from the measure \mathbb{P}). We denote their joint law by $\mathbf{P}_{\mathbb{P}}^{\otimes}$ (expectation $\mathbf{E}_{\mathbb{P}}^{\otimes}$). The next proposition quantifies the accuracy of the naive Monte-Carlo estimator.

Proposition 4.1 (naive MCMC method). *For each $\alpha > 1$, there exists a constant $C < \infty$ such that the following holds. Let $f \in L^2(\mathbb{P})$ satisfy (3.5) and (4.1), and for each $t \geq 2$ and positive integer N , let*

$$(4.2) \quad \hat{\sigma}_0^2(N, t, f) := \frac{1}{N} \sum_{k=1}^N \left(\frac{1}{\sqrt{t}} \int_0^t f(X^{(k)}(s)) ds \right)^2.$$

We have

$$(4.3) \quad \mathbf{E}_{\mathbb{P}}^{\otimes} \left[\left(\widehat{\sigma}_0^2(N, t, f) - \sigma^2(f) \right)^2 \right]^{\frac{1}{2}} \leq C \begin{cases} N^{-\frac{1}{2}} + t^{-(\alpha-1)} & \text{if } \alpha < 2, \\ N^{-\frac{1}{2}} + t^{-1} \log(t) & \text{if } \alpha = 2, \\ N^{-\frac{1}{2}} + t^{-1} & \text{if } \alpha > 2. \end{cases}$$

Proof. We decompose the left side of (4.3) into a variance and a bias terms:

$$(4.4) \quad \mathbf{E}_{\mathbb{P}}^{\otimes} \left[\left(\widehat{\sigma}_0^2(N, t, f) - \sigma^2(f) \right)^2 \right] \\ = \mathbf{E}_{\mathbb{P}}^{\otimes} \left[\left(\widehat{\sigma}_0^2(N, t, f) - \mathbf{E}_{\mathbb{P}}^{\otimes} \left[\widehat{\sigma}_0^2(N, t, f) \right] \right)^2 \right] + \left| \mathbf{E}_{\mathbb{P}}^{\otimes} \left[\widehat{\sigma}_0^2(N, t, f) \right] - \sigma^2(f) \right|^2.$$

By (4.1), it is clear that the first term on the right side above is bounded by N^{-1} . By stationarity, the expectation appearing in the second term can be rewritten as

$$(4.5) \quad \mathbf{E}_{\mathbb{P}} \left[\left(\frac{1}{\sqrt{t}} \int_0^t f(X(s)) \, ds \right)^2 \right] = \frac{2}{t} \int_{0 \leq u < v \leq t} \mathbf{E}_{\mathbb{P}} [f(X(u))f(X(v))] \, du \, dv \\ = \frac{2}{t} \int_{0 \leq u < v \leq t} \mathbf{E}_{\mathbb{P}} [f(X(0))f(X(v-u))] \, du \, dv \\ = \frac{2}{t} \int_0^t (t-s) \langle f, e^{-s\mathcal{L}} f \rangle \, ds.$$

Recalling the formula (2.4) for $\sigma^2(f)$ and using (3.5), we can bound the difference between the integral above and $\sigma^2(f)$ by

$$\frac{2}{t} \int_0^t s(1+s)^{-\alpha} \, ds + 2 \int_t^{+\infty} (1+s)^{-\alpha} \, ds \leq C \begin{cases} t^{-(\alpha-1)} & \text{if } \alpha < 2, \\ t^{-1} \log(t) & \text{if } \alpha = 2, \\ t^{-1} & \text{if } \alpha > 2. \end{cases}$$

This completes the proof. \square

The next proposition, which is the main result of this section, shows that a simple one-step extrapolation of the naive MCMC estimator yields an improved convergence rate in most cases.

Proposition 4.2 (Improved MCMC method). *For each $\alpha > 1$, there exists a constant $C < \infty$ such that the following holds. Let $f \in L^2(\mathbb{P})$ satisfy (3.5) and (4.1), and for each $t \geq 2$ and positive integer N , let*

$$\widehat{\sigma}_1^2(N, t, f) := 2\widehat{\sigma}_0^2(N, 2t, f) - \widehat{\sigma}_0^2(N, t, f),$$

where $\widehat{\sigma}_0^2(N, t, f)$ is defined in (4.2). We have

$$\mathbf{E}_{\mathbb{P}}^{\otimes} \left[\left(\widehat{\sigma}_1^2(N, t, f) - \sigma^2(f) \right)^2 \right]^{\frac{1}{2}} \leq C \left(N^{-\frac{1}{2}} + t^{-(\alpha-1)} \right).$$

Proof. We use the variance-bias decomposition (4.4) (with $\widehat{\sigma}_0^2$ replaced by $\widehat{\sigma}_1^2$) and estimate the variance as in the proof of Proposition 4.1. There remains to estimate the bias term, which takes the form

$$\left| 2\mathbf{E}_{\mathbb{P}} \left[\left(\frac{1}{\sqrt{2t}} \int_0^{2t} f(X(s)) \, ds \right)^2 \right] - \mathbf{E}_{\mathbb{P}} \left[\left(\frac{1}{\sqrt{t}} \int_0^t f(X(s)) \, ds \right)^2 \right] - \sigma^2(f) \right|.$$

In view of (2.4) and (4.5), we can rewrite this quantity as

$$2 \left| \int_0^{+\infty} \left[2 \left(1 - \frac{s}{2t} \right)_+ - \left(1 - \frac{s}{t} \right)_+ - 1 \right] \langle f, e^{-s\mathcal{L}} f \rangle ds \right|,$$

where we denote $x_+ = x \vee 0 = \max(x, 0)$ for the positive part of x . We also write $x \wedge y = \min(x, y)$. Simplifying and using (3.5), we get that the expression above is

$$(4.6) \quad 2 \left| \int_t^{+\infty} \left[\left(\frac{s}{t} - 1 \right) \wedge 1 \right] \langle f, e^{-s\mathcal{L}} f \rangle ds \right| \leq 2 \int_t^{+\infty} s^{-\alpha} ds \leq Ct^{-(\alpha-1)},$$

as announced. \square

The one-step extrapolated scheme described in Proposition 4.2 is therefore more efficient than the naive scheme described in Proposition 4.1 whenever $\alpha \geq 2$. To make this point more precise, we may measure the computational complexity of each method in terms of its “total simulated time”: if an algorithm samples n trajectories, each respectively up to time t_1, \dots, t_n , then the total simulated time of the algorithm is $\sum_{i=1}^n t_i$. For any fixed $\alpha > 2$, we compute the total simulated time required for each method to produce an output of the order of $\delta \in (0, 1]$ away from $\sigma^2(f)$. For the naive method, we need to choose $N \simeq \delta^{-2}$ and $t \simeq \delta^{-1}$, so that the total simulated time is of the order of δ^{-3} . For the one-step extrapolated method, we choose instead $t \simeq \delta^{-\frac{1}{\alpha-1}}$, still with $N \simeq \delta^{-2}$, so that the total simulated time is of the order of $\delta^{-(2+\frac{1}{\alpha-1})} \ll \delta^{-3}$. If we replace the polynomial convergence to equilibrium in (3.5) by an exponential one, then one can show that the one-step extrapolated method requires only of the order of $\delta^{-2} \log(\delta^{-1})$ of simulated time, while the naive method still requires a total simulated time of the order of δ^{-3} . It is striking that a single-step extrapolation suffices to get rid of “saturation” phenomena as seen in Proposition 4.1 at all orders at once.

5. VARIANTS OF THE GREEN-KUBO FORMULA

In this section, we prove the non-reversible version of the variant of the Green-Kubo formula presented in (2.9). We then discuss the convergence rate of the series under the additional assumption of (3.5).

Theorem 5.1. *Let $f \in L^2(\mathbb{P})$ be such that (3.4) holds, let $(\mu_k)_{k \in \mathbb{N}}$ be a sequence of real numbers in $(0, 1]$, and define recursively*

$$(5.1) \quad \begin{cases} f_{-1} = f_{-1}^* := f, \\ \forall k \in \mathbb{N}, \quad f_k := \mu_k(\mu_k + \mathcal{L})^{-1} f_{k-1} \quad \text{and} \quad f_k^* := \mu_k(\mu_k + \mathcal{L}^*)^{-1} f_{k-1}^*. \end{cases}$$

For every $n \in \mathbb{N}$, the quantity

$$(5.2) \quad \langle f_n^*, \mathcal{L}^{-1} f_n \rangle := \lim_{\lambda \rightarrow 0} \langle f_n^*, (\lambda + \mathcal{L})^{-1} f_n \rangle$$

is well-defined and finite, and

$$(5.3) \quad \lim_{n \rightarrow \infty} \langle f_n^*, \mathcal{L}^{-1} f_n \rangle = 0.$$

Moreover, for every $n \in \mathbb{N}$, we have

$$(5.4) \quad \langle f, \mathcal{L}^{-1} f \rangle = \sum_{k=0}^n \mu_k^{-1} (\langle f_{k-1}^*, f_k \rangle + \langle f_k^*, f_k \rangle) + \langle f_n^*, \mathcal{L}^{-1} f_n \rangle.$$

Proof. Let $\lambda > 0$. From the formula (3.1), we see that if T_λ is an exponential random variable of parameter λ , independent of any other quantity in the problem, then we can represent the resolvent operator R_λ as

$$\lambda R_\lambda g = E [e^{-T_\lambda \mathcal{L}} g],$$

where E denotes the expectation over the random variable T_λ . By induction, we see that if S_n has the law of a sum of $(n+1)$ independent exponential random variables of parameter μ_0, \dots, μ_n respectively, then

$$f_n = E [e^{-S_n \mathcal{L}} f].$$

Here again the random variable S_n is independent of any other quantity in the problem, and E computes the average of this random variable. Denoting by S'_n an independent copy of S_n , we have

$$\begin{aligned} \langle f_n^*, (\lambda + \mathcal{L})^{-1} f_n \rangle &= \int_0^{+\infty} e^{-\lambda t} E \left[\langle e^{-S'_n \mathcal{L}^*} f, e^{-(S_n+t) \mathcal{L}} f \rangle \right] dt \\ &= \int_0^{+\infty} e^{-\lambda t} E \left[\langle f, e^{-(S_n+S'_n+t) \mathcal{L}} f \rangle \right] dt. \end{aligned}$$

The assumption of (3.4) ensures that for each fixed $s \geq 0$,

$$\int_0^{+\infty} |\langle f, e^{-(s+t) \mathcal{L}} f \rangle| dt \leq \int_0^{+\infty} |\langle f, e^{-t \mathcal{L}} f \rangle| dt < +\infty.$$

Hence, the quantity

$$\int_0^{+\infty} e^{-\lambda t} \langle f, e^{-(s+t) \mathcal{L}} f \rangle dt$$

is bounded uniformly over $s \geq 0$ and $\lambda > 0$, and converges to

$$\int_0^{+\infty} \langle f, e^{-(s+t) \mathcal{L}} f \rangle dt$$

as λ tends to 0. By the dominated convergence theorem, this implies that

$$\langle f_n^*, (\lambda + \mathcal{L})^{-1} f_n \rangle$$

converges as λ tends to infinity to

$$\int_0^{+\infty} E \left[\langle f, e^{-(S_n+S'_n+t) \mathcal{L}} f \rangle \right] dt \in \mathbb{R}.$$

We now verify that the integral above tends to 0 as n tends to infinity. It is clear that we can realize the sequences S_n, S'_n such that $S_{n+1} \geq S_n$ and $S'_{n+1} \geq S'_n$, and moreover, since the sequence (μ_k) is bounded by 1, we have that S_n and S'_n tend to infinity almost surely. For each $M \geq 1$, we can bound the absolute value of the integral above by

$$(5.5) \quad P[S_n + S'_n \leq M] \int_0^{+\infty} |\langle f, e^{-t \mathcal{L}} f \rangle| dt + \int_M^{+\infty} |\langle f, e^{-t \mathcal{L}} f \rangle| dt.$$

By the assumption of (3.4), we may fix M sufficiently large to make the second integral in (5.5) as small as desired. Then taking n sufficiently large ensures that the probability on the left of (5.5) is arbitrarily small. This completes the proof of (5.3).

There remains to show (5.4). For each $\lambda > 0$, we introduce the shorthand notation

$$\mu_k^\lambda := \mu_k - \lambda.$$

Applying the resolvent formula (2.7) recursively, we see that

$$(5.6) \quad R_\lambda = \left(\sum_{k=0}^n \mu_0^\lambda \cdots \mu_{k-1}^\lambda R_{\mu_0} \cdots R_{\mu_k} \right) + \mu_0^\lambda \cdots \mu_n^\lambda R_{\mu_0} \cdots R_{\mu_n} R_\lambda.$$

We define recursively

$$f_{-1}^\lambda = f_{-1}^{\lambda,*} := f,$$

$$\forall k \in \mathbb{N}, \quad f_k^\lambda := \mu_k^\lambda R_{\mu_k} f_{k-1}^\lambda \quad \text{and} \quad f_k^{\lambda,*} := \mu_k^\lambda R_{\mu_k}^* f_{k-1}^{\lambda,*}.$$

Using the formula (5.6) with the sequence $(\mu_k)_{k \in \mathbb{N}}$ replaced by $(\mu_0, \mu_0, \mu_1, \mu_1, \dots)$, we get that

$$(5.7) \quad \langle f, R_\lambda f \rangle = \sum_{k=0}^n \frac{1}{\mu_k^\lambda} \left(\langle f_{k-1}^{\lambda,*}, f_k^\lambda \rangle + \langle f_k^{\lambda,*}, f_k^\lambda \rangle \right) + \langle f_n^{\lambda,*}, R_\lambda f_n^\lambda \rangle,$$

where for simplicity, we may assume that $\lambda > 0$ is sufficiently small that $\mu_k^\lambda \neq 0$ for each $k \leq n$. Note that f_n^λ differs from f_n by a scalar factor only, and this factor tends to 1 as λ tends to 0; and similarly for $f_n^{\lambda,*}$. It is therefore clear that

$$\lim_{\lambda \rightarrow 0} \langle f_n^{\lambda,*}, R_\lambda f_n^\lambda \rangle = \langle f_n^*, \mathcal{L}^{-1} f_n \rangle.$$

Passing to the limit $\lambda \rightarrow 0$ in the sum on the right side of (5.7) is similar (only simpler), and we obtain (5.4). \square

We now give some simple criteria for evaluating the size of the remainder term in the decomposition 5.4. This will depend on the sequence (μ_k) we choose. The simplest case is when we choose this sequence to be constant.

Proposition 5.2 (Remainder estimate, constant step size). *For each $\alpha > 1$, there exists a constant $C < \infty$ such that the following holds. Let $f \in L^2(\mathbb{P})$ satisfy (3.5), and let $(f_k)_{k \in \mathbb{N}}, (f_k^*)_{k \in \mathbb{N}}$ be defined according to (5.1) with the choice $\mu_k \equiv 1$. For every $n \in \mathbb{N}$, we have*

$$(5.8) \quad |\langle f_n^*, \mathcal{L}^{-1} f_n \rangle| \leq C(1+n)^{-(\alpha-1)}.$$

Proof. Denote by S_n and S'_n two independent random variables, each having the law of a sum of n independent exponential random variables of parameter 1. We assume that these random variables are also independent of any other quantity discussed so far, and denote by E the expectation with respect to these random variables. We have

$$f_n = E \left[e^{-S_n \mathcal{L}} f \right] \quad \text{and} \quad f_n^* = E \left[e^{-S'_n \mathcal{L}^*} f \right].$$

Moreover,

$$\begin{aligned} \langle f_n^*, \mathcal{L}^{-1} f_n \rangle &= \int_0^{+\infty} \langle f_n^*, e^{-t\mathcal{L}} f_n \rangle dt \\ &= \int_0^{+\infty} E \left[\left\langle e^{-S'_n \mathcal{L}^*} f, e^{-(t+S_n)\mathcal{L}} f \right\rangle \right] dt \\ &= \int_0^{+\infty} E \left[\left\langle f, e^{-(t+S_n+S'_n)\mathcal{L}} f \right\rangle \right] dt. \end{aligned}$$

By an elementary large deviation estimate, there exists a constant $C < \infty$ such that

$$P[|S_n + S'_n - 2n| \leq n] \leq C \exp(-n/C).$$

Using the assumption (3.5), we thus get that

$$\begin{aligned} |\langle f_n^*, \mathcal{L}^{-1} f_n \rangle| &\leq \int_0^{+\infty} E[(1+t+S_n+S'_n)^{-\alpha}] dt \\ &\leq C \exp(-n/C) + \int_0^{+\infty} (1+t+n)^{-\alpha} dt \\ &\leq C(1+n)^{-(\alpha-1)}, \end{aligned}$$

as announced. \square

Under the assumption (3.5), it is also possible to use geometrically increasing step sizes with no degradation of the convergence rate, as the next proposition demonstrates. This can be interesting, since it reduces considerably the number of Poisson equations one needs to solve: for a desired precision of $\delta > 0$, using geometrically increasing step sizes requires only the resolution of a constant times $\log \delta^{-1}$ such equations, as opposed to $\delta^{-1/(\alpha-1)}$ with the constant-step-size scheme of Proposition 5.2. (On the other hand, if the assumption (3.5) was replaced by exponentially fast convergence, then the constant-step-size scheme should be preferred.)

Proposition 5.3 (Remainder estimate, geometric step size). *For each $\alpha > 1$, there exists a constant $C < \infty$ such that the following holds. Let $f \in L^2(\mathbb{P})$ satisfy (3.5), and let $(f_k)_{k \in \mathbb{N}}$, $(f_k^*)_{k \in \mathbb{N}}$ be defined according to (5.1) with the choice $\mu_k \equiv 2^{-k}$. For every $n \in \mathbb{N}$, we have*

$$(5.9) \quad |\langle f_n^*, \mathcal{L}^{-1} f_n \rangle| \leq C 2^{-n(\alpha-1)}.$$

Remark 5.4. In both (5.8) and (5.9), the right-hand side is a multiple of $(\mu_0^{-1} + \dots + \mu_n^{-1})^{-(\alpha-1)}$.

Proof. Let $(T_k)_{k \in \mathbb{N}}$ be independent exponential random variables of respective parameter $(2^{-k})_{k \in \mathbb{N}}$, and for each $n \in \mathbb{N}$, let

$$(5.10) \quad S_n = \sum_{k=0}^n T_k.$$

We assume that the random variables (T_k) are independent of the other quantities of our problem, and denote by E the expectation with respect to these

random variables. We also give ourselves S'_n an independent copy of S_n , and recall that

$$f_n = E [e^{-S_n \mathcal{L}} f] \quad \text{and} \quad f_n^* = E [e^{-S_n \mathcal{L}^*} f].$$

We decompose the rest of the proof into three steps.

Step 1. In this first step, we show that for each $K < \infty$, there exists a constant $C < \infty$ such that for every $k \leq n$,

$$(5.11) \quad P [S_n \leq 2^k] \leq C \exp(-K(n-k)).$$

By Chebyshev's inequality, we have

$$\begin{aligned} P [S_n \leq 2^k] &\leq \mathbb{E} [\exp(1 - 2^{-k}(T_1 + \dots + T_n))] \\ &\leq \exp\left(1 - \sum_{j=0}^n \log(1 + 2^{j-k})\right). \end{aligned}$$

We select $k_0(K) < \infty$ such that

$$\log(1 + 2^{k_0}) \geq K,$$

and deduce that

$$P [S_n \leq 2^k] \leq \exp(1 - K(n - k - k_0)_+).$$

This shows (5.11).

Step 2. We show that for every $\beta \in (0, \infty)$, there exists a constant $C < \infty$ such that for every n ,

$$(5.12) \quad E [(1 + S_n)^{-\beta}] \leq C 2^{-n\beta}.$$

We rewrite the expectation on the left side above as

$$E [(1 + S_n)^{-\beta}] = \beta \int_1^{+\infty} t^{-(\beta+1)} P[1 + S_n \leq t] dt,$$

and decompose the integral into several parts. The easiest part is

$$\beta \int_{2^n}^{\infty} t^{-(\beta+1)} P[1 + S_n \leq t] dt \leq \beta \int_{2^n}^{\infty} t^{-(\beta+1)} dt \leq 2^{-n\beta}.$$

We split the remaining part along a dyadic sequence:

$$\begin{aligned} \int_1^{2^n} s^{-(\beta+1)} P[1 + S_n \leq s] ds &\leq \sum_{k=0}^{n-1} \int_{2^k}^{2^{k+1}} s^{-(\beta+1)} P[1 + S_n \leq s] ds \\ &\leq \sum_{k=0}^{n-1} 2^{-k\beta} P[S_n \leq 2^{k+1}]. \end{aligned}$$

Using the result of the first step with the choice of $K = (\beta + 1) \log 2$, we obtain that the sum above is bounded by

$$C \sum_{k=0}^{n-1} 2^{-k\beta} 2^{-(n-k)(\beta+1)} \leq C 2^{-n\beta}.$$

This completes the proof of (5.12).

Step 3. We complete the proof. As in the proofs of Theorem 5.1 and Proposition 5.2, our starting point is to observe that

$$\langle f_n^*, \mathcal{L}^{-1} f_n \rangle = \int_0^{+\infty} E \left[\langle f, e^{-(t+S_n+S'_n)\mathcal{L}} f \rangle \right] dt.$$

Using (3.5), Fubini's theorem and a change of variables, we infer that

$$\begin{aligned} |\langle f_n^*, \mathcal{L}^{-1} f_n \rangle| &\leq \int_0^{+\infty} E [(1+t+S_n+S'_n)^{-\alpha}] dt \\ &\leq \int_0^{+\infty} E [(1+t+S_n)^{-\alpha}] dt \\ &\leq (\alpha-1)^{-1} E [(1+S_n)^{-(\alpha-1)}]. \end{aligned}$$

The conclusion then follows from (5.12). \square

Propositions 5.2 and 5.3 provide us with guidelines for the choice of the terminal index n in the series expansion of $\langle f, \mathcal{L}^{-1} f \rangle$ given in (5.4). One central feature of this series expansion is the fact that as the index k of the series increases and f_k and f_k^* become more difficult to compute, these quantities also become smaller, and we therefore need much fewer samples to evaluate the average $\langle f_k^*, f_k \rangle$ accurately. The smallness of f_k is most clearly seen if we specialize the proposition below to the reversible case.

Proposition 5.5. *There exists a constant $C < \infty$ such that the following holds. Let $f \in L^2(\mathbb{P})$ satisfy (3.5), and let $(f_k)_{k \in \mathbb{N}}$, $(f_k^*)_{k \in \mathbb{N}}$ be defined according to (5.1) with the choice $\mu_k \equiv 1$. For every $k \in \mathbb{N}$, we have*

$$|\langle f_{k-1}^*, f_k \rangle| + |\langle f_k^*, f_k \rangle| \leq C(1+k)^{-\alpha}.$$

If instead, the sequences $(f_k)_{k \in \mathbb{N}}$ and $(f_k^)_{k \in \mathbb{N}}$ are defined according to (5.1) with the choice $\mu_k \equiv 2^{-k}$, then for every $k \in \mathbb{N}$,*

$$|\langle f_{k-1}^*, f_k \rangle| + |\langle f_k^*, f_k \rangle| \leq C2^{-k\alpha}.$$

The proof of this proposition is almost identical to the proofs of Propositions 5.2 and 5.3, so we omit it. The point we wish to stress here is that at least under the assumption of reversibility (that is, $\mathcal{L}^* = \mathcal{L}$ and thus $f_k^* = f_k$), the typical size of f_k (as measured in $L^2(\mathbb{P})$) is small: it is of the order of $k^{-\alpha/2}$ for constant step sizes, and of the order of $2^{-k\alpha/2}$ for geometrically increasing step sizes.

In the non-reversible setting, the situation is more subtle. Indeed, the assumption of (3.5) no longer informs us on the size of $e^{-t\mathcal{L}} f$ in $L^2(\mathbb{P})$; similarly, we can no longer sharply estimate the size of f_k (whatever the choice of μ_k) from this assumption. In Proposition 6.1 below, we consider dynamical systems (which in some sense is the “worst-case scenario” for these bounds) and show, under an assumption comparable to (3.5), that f_k is of the order of $k^{-\frac{1}{4}}$ in the case of $\mu_k \equiv 1$, and that f_k is of the order of $2^{-\frac{k}{2}}$ in the case of geometrically increasing step sizes. This gives further motivation for using geometrically increasing step sizes.

6. DISCRETE-TIME MARKOV CHAINS

In this section, we explain how to adapt the arguments of the previous section to the setting of discrete-time Markov chains. We then discuss the rate of decay to zero of the remainder in the series expansion (5.4) in the “extreme” case of dynamical systems.

Let $(X_n)_{n \in \mathbb{N}}$ be a discrete-time Markov chain on \mathcal{X} , and let P be its transition operator. That is, with hopefully transparent notation,

$$Pf(x) := \mathbf{E}_x[f(X_1)].$$

We define

$$(6.1) \quad \mathcal{L} := \text{Id} - P.$$

We assume that the measure \mathbb{P} is invariant for (X_n) , and continue denoting by $\langle \cdot, \cdot \rangle$ the scalar product in $L^2(\mathbb{P})$. We fix $f \in L^2(\mathbb{P})$ of mean zero. As in the continuous case, there is no simple and general theorem asserting whether the additive functional

$$(6.2) \quad \frac{1}{\sqrt{n}} \sum_{k=0}^n f(X_k)$$

converges in law to a Gaussian as n tends to infinity. Note that at this level of generality, the question covers dynamical systems as a particular case; we refer to [48, 20, 52] and references therein for results in this context. In analogy with the continuous case, we assume throughout that

$$(6.3) \quad \sum_{k=0}^{+\infty} |\langle f, P^k f \rangle| < \infty.$$

Under some additional assumptions (e.g. reversibility is sufficient [40]), the random variable in (6.2) converges in law to a centered Gaussian of variance

$$(6.4) \quad \begin{aligned} \sigma^2(f) &:= \langle f, f \rangle + 2 \lim_{\lambda \rightarrow 0} \langle f, (\lambda + \mathcal{L})^{-1} f \rangle \\ &= -\langle f, f \rangle + 2 \sum_{k=0}^{+\infty} \langle f, P^k f \rangle. \end{aligned}$$

The analogue of the formula (2.6), adapted to the non-reversible and discrete-time setting, reads

$$(6.5) \quad \sigma^2(f) = -\langle f, f \rangle + 2 \sum_{k=0}^{+\infty} (\langle (P^*)^k f, P^k f \rangle + \langle (P^*)^k f, P^{k+1} f \rangle),$$

where P^* is the adjoint of P in $L^2(\mathbb{P})$. This formula can be used directly to evaluate $\sigma^2(f)$ (this is essentially the formula used in Section 12 below). We also have that the exact statement of Theorem 5.1, with \mathcal{L} now defined by (6.1), holds—and the proof given there applies with essentially no change besides a minor adaptation to the discrete-time setting. The minor adaptation boils

down to replacing the formula for the resolvent in (3.1) with

$$(6.6) \quad R_\lambda = \sum_{k=0}^{+\infty} (\lambda + 1)^{-(k+1)} P^k.$$

Strengthening (6.3), and in analogy with (3.5), we may assume that there exists an exponent $\alpha > 1$ such that for every $k \in \mathbb{N}$,

$$(6.7) \quad |\langle f, P^k f \rangle| \leq (1 + k)^{-\alpha}.$$

This assumption implies the validity of Propositions 5.2, 5.3 and 5.5. The proofs are almost identical except for the use of (6.6) in place of (3.1).

With reversible dynamics, the assumption of (6.7) directly translates into information on the size of $P^k f$ itself. This is not so in general. We can illustrate this phenomenon most clearly in the particular case of dynamical systems. Indeed, assume that there exists a measure-preserving mapping $\mathbb{T} : \mathcal{X} \rightarrow \mathcal{X}$ such that for every n , $X_n = \mathbb{T}^n(X_0)$, or equivalently,

$$(6.8) \quad Pf = f \circ \mathbb{T}.$$

Then the L^2 norm of $P^k f$ does not depend on k , since \mathbb{T} is measure-preserving. In this case, we thus need many samples of $\mathbb{T}^k f$ and $(\mathbb{T}^*)^k f$ if we wish to estimate the average $\langle (\mathbb{T}^*)^k f, \mathbb{T}^k f \rangle$ using independent draws. On the other hand, we may expect the quantity f_k appearing in Theorem 5.1 to display some cancellations that make it smaller. The next proposition quantifies this more precisely.

Proposition 6.1. *There exists a constant $C < \infty$ such that the following holds. Let P be the transition operator of a dynamical system, see (6.8), let $f \in L^2(\mathbb{P})$ satisfy (6.7), and let $(f_k)_{k \in \mathbb{N}}$, $(f_k^*)_{k \in \mathbb{N}}$ be defined according to (5.1) with the choice $\mu_k \equiv 1$. For every $k \in \mathbb{N}$, we have*

$$(6.9) \quad |\langle f_k^*, f_k^* \rangle| + |\langle f_k, f_k \rangle| \leq C(1 + k)^{-\frac{1}{2}}.$$

If instead, the sequences $(f_k)_{k \in \mathbb{N}}$ and $(f_k^)_{k \in \mathbb{N}}$ are defined according to (5.1) with the choice $\mu_k \equiv 2^{-k}$, then for every $k \in \mathbb{N}$,*

$$(6.10) \quad |\langle f_k^*, f_k^* \rangle| + |\langle f_k, f_k \rangle| \leq C2^{-k}.$$

We stress again that the exact statement of Proposition 5.5 is true under the assumption of (6.7). Proposition 6.1 is a statement about the respective sizes of f_k^* and f_k , *not* about the size of the correlation between these quantities.

Proof of Proposition 6.1. Let T_1, T_2, \dots be independent random variables following the same law of geometric type:

$$(6.11) \quad \forall k \in \mathbb{N}, \quad P[T_i = k] = \mu_i(\mu_i + 1)^{-(k+1)}.$$

We assume that these random variables are independent of the other randomness of the problem, and denote by E the expectation with respect to these variables. We also set, for each $k \in \mathbb{N}$,

$$S_k := T_1 + \dots + T_k,$$

and give ourselves an independent copy S'_k of S_k . In view of the formula (6.6), we have the representation

$$f_k = E [P^{S_k} f],$$

and thus

$$\langle f_k, f_k \rangle = E \left[\left\langle P^{S_k} f, P^{S'_k} f \right\rangle \right].$$

Since the mapping \mathbb{T} is measure-preserving, we deduce that

$$\langle f_k, f_k \rangle = E \left[\left\langle f, P^{|S_k - S'_k|} f \right\rangle \right],$$

and by the assumption of (6.7) and the triangle inequality, we get

$$\begin{aligned} |\langle f_k, f_k \rangle| &\leq E [(1 + |S_k - S'_k|)^{-\alpha}] \\ &\leq CE [(1 + |S_k - E[S_k]|)^{-\alpha}]. \end{aligned}$$

We now use the decomposition

$$(6.12) \quad E [(1 + |S_k - E[S_k]|)^{-\alpha}] = \int_1^\infty \frac{\alpha}{s^{\alpha+1}} P [1 + |S_k - E[S_k]| \leq s] ds.$$

We first analyze the case of constant step size, that is $\mu_k \equiv 1$. In this case, the random variable S_k is a sum of independent and identically distributed random variables. Denote by τ^2 the variance of T_1 . By the Berry-Esseen theorem, see [57, Theorem 5.5], there exists a constant $C < \infty$ such that for every $k \in \mathbb{N}$ and $s \in \mathbb{R}$,

$$\left| P [|S_k - E[S_k]| \leq s\sqrt{k}] - \frac{2}{\tau\sqrt{2\pi}} \int_0^s \exp\left(-\frac{x^2}{2\tau^2}\right) dx \right| \leq \frac{C}{\sqrt{k}}.$$

We thus obtain that

$$\begin{aligned} E [(1 + |S_k - E[S_k]|)^{-\alpha}] &\leq \frac{C}{\sqrt{k}} + C \int_1^\infty \frac{1}{s^{\alpha+1}} \int_0^{\frac{s}{\sqrt{k}}} \exp\left(-\frac{x^2}{2\tau^2}\right) dx ds \\ &\leq \frac{C}{\sqrt{k}} + C \int_1^\infty \frac{1}{s^{\alpha+1}} \frac{s}{\sqrt{k}} ds \\ &\leq \frac{C}{\sqrt{k}}, \end{aligned}$$

as announced.

We now turn to the case of geometrically increasing step sizes, that is, $\mu_k \equiv 2^{-k}$. We first show that there exists a constant $C < \infty$ such that for every $k \in \mathbb{N}$ and $s \geq 0$,

$$(6.13) \quad P [|S_k - E[S_k]| \leq s] \leq C2^{-k} (s + 1).$$

In order to do so, it suffices to verify that there exists $C < \infty$ such that for every $k \in \mathbb{N}$ and $x \in \mathbb{N}$,

$$P [S_k = x] \leq C2^{-k}.$$

It follows from the definition of T_k that this property is satisfied if S_k is replaced by T_k . By independence between T_k and S_{k-1} , we deduce

$$\begin{aligned} P[S_k = x] &= E[P[T_k = x - S_{k-1} \mid S_{k-1}]] \\ &\leq E\left[\sup_{z \in \mathbb{Z}} P[T_k = z \mid S_{k-1}]\right] \\ &\leq C2^{-k}. \end{aligned}$$

This implies (6.13). Combining this with (6.12) yields

$$E[(1 + |S_k - E[S_k]|)^{-\alpha}] \leq C2^{-k} \int_1^\infty \frac{1}{s^\alpha} ds \leq C2^{-k},$$

completing the proof. \square

Remark 6.2. One can verify that the estimates obtained in Proposition 6.1 are sharp, in the sense that assuming the converse inequality to (6.7), one can show the converse inequalities to (6.9) and (6.10), up to multiplicative constants.

7. EXAMPLES

In the second part of the paper, we explore thoroughly the example of a reversible random walk in a random environment, or equivalently the homogenization of discrete divergence-form operators with random coefficients. Under some assumptions, the relaxation property (3.5) holds with $\alpha = \frac{d}{2} + 1$. We will explain how to implement numerical schemes derived from Theorem 5.1 by leveraging on the mixing properties of the functions f_k under spatial translations. We will show that such numerical schemes achieve essentially optimal complexity, requiring $C\delta^{-2} \log(\delta^{-1})$ computations to compute the effective diffusivity at precision $\delta > 0$.

An alternative route to solve the same problem would be to use Monte-Carlo methods. Computing the effective diffusivity at precision $\delta > 0$ using the method described in Proposition 4.2 would call for the simulation of δ^{-2} random walk trajectories that run up to time $\delta^{-\frac{2}{d}}$, bringing the total cost of the method to a suboptimal $\delta^{-2-\frac{2}{d}}$. (See [27, Theorem 4.1] for the verification that a large deviation estimate much stronger than (4.1) holds.)

In this section, we discuss briefly other examples of interest where the methods described here can be applied.

7.1. Tagged particle in exclusion process. One example where the approach based on Theorem 5.1 seems very costly to implement is the problem of computing the effective diffusivity of a tagged particle in a system of particles evolving according to, say, exclusion dynamics on \mathbb{Z}^d . We refer to [46] for the definition of this and other models of interacting particle systems. Indeed, the Poisson equations appearing in (5.1) are posed in $\{0, 1\}^{\mathbb{Z}^d}$. Even if we restrict the dynamics to a finite box $B \subseteq \mathbb{Z}^d$, the size of each ergodic component of the dynamics is exponentially large as a function of the volume of B , which seems to preclude any computationally efficient approach to the resolution of these equations. In contrast, the simulation of the dynamics on B is feasible, and by exchangeability, the trajectory of each of the simulated particles can be used as

a sample of the trajectory of a tagged particle. One would then need to assert precisely the independence properties of these trajectories, but it is clear that the diffusivity can be computed in polynomial time using an MCMC scheme, and possibly in about $\delta^{-2-\frac{2}{d}}$ operations.

7.2. Langevin dynamics. Hypoelliptic dynamics can also be considered. We explore briefly here the case of Langevin dynamics in a random potential. Let ϕ be a smooth stationary and ergodic function over \mathbb{R}^d . We consider the Markov process $(X_t, V_t)_{t \geq 0}$, taking values in $\mathbb{R}^d \times \mathbb{R}^d$, whose infinitesimal generator is given by

$$(7.1) \quad Lf := \Delta_v f - v \cdot \nabla_v f - \nabla_x \phi \cdot \nabla_v f + v \cdot \nabla_x f.$$

Note that in this case, the process (X_t) itself is an additive functional, since

$$X_t = \int_0^t V_s ds,$$

while for random walks or tagged particles the connection between the dynamics and additive functionals is less immediate (but see [40]). The fact that $(\varepsilon X_{\varepsilon^{-2}t})_{t \geq 0}$ converges in law to Brownian motion as ε tends to zero was proved in [55]. Moreover, the invariant measure for the process of the environment as seen from X is explicit. Under suitable assumptions on the mixing properties of ϕ , we may expect, in analogy with the simpler setting explore in the second part of the paper, that the polynomial relaxation (3.5) holds with $\alpha = \frac{d}{2} + 1$.

It can be relatively costly (but not as prohibitive as for interacting particle systems) to compute approximate solutions to the Poisson equations appearing in (5.1). Compared with the situation encountered for random walks, the additional difficulty stems in part from the fact that the operator is hypoelliptic, but also simply from the fact that there are more variables to solve for (due to the addition of velocity coordinates). The superiority of this approach compared to that based on Monte-Carlo sampling is thus unclear, and probably depends on the situation. In one extreme case, if the dimension is small and ϕ is a deterministic periodic function (whose periods need not be known), then an approach similar to that exposed in the second part of the paper is most probably best (and it should be preferable to choose $\mu_k \equiv 1$ there, since the convergence to equilibrium is exponentially rapid in the periodic setting). At the other extreme, in the random case and in high dimension, the Monte-Carlo approach probably becomes superior.

Part II. Stochastic homogenization

8. NOTATION AND BASIC PROPERTIES

In this second part of the paper, we study the problem of approximating the homogenized matrix of divergence-form operators with random coefficients, as stated in the introduction. In the present section, we set up notation and assumptions and recall basic facts of homogenization. We then proceed to prove the error estimate (1.5) in Section 9. In Section 10, we complete the proof of

Theorem 1.2 by estimating the computational complexity of the method, and prove the complexity lower bound given by Proposition 1.1. We report on the results of numerical tests in Section 11. Finally, we explore a variant of the method based on discrete-time parabolic equations in Section 12.

We focus on the discrete space \mathbb{Z}^d , $d \geq 2$. We say that $x, y \in \mathbb{Z}^d$ are neighbors, and write $x \sim y$, if $|x - y| = 1$. This turns \mathbb{Z}^d into a graph, and we denote by \mathbb{B} the corresponding set of undirected edges. For a fixed ellipticity parameter $\Lambda \in (1, \infty)$, we let $\Omega = [1, \Lambda]^{\mathbb{B}}$ denote the set of coefficient fields, equipped with the product σ -algebra. We denote by $\mathbf{a} = (\mathbf{a}_e)_{e \in \mathbb{B}}$ the canonical random variable on Ω . We let \mathbb{P} be a probability measure on Ω under which the random variables $(\mathbf{a}_e)_{e \in \mathbb{B}}$ are independent and identically distributed (i.i.d.), and denote by \mathbb{E} the associated expectation. Besides uniform ellipticity, this independence assumption is our main quantitative hypothesis. The group \mathbb{Z}^d acts on Ω by translations via

$$(\theta_x \mathbf{a})_e = \mathbf{a}_{x+e},$$

where $x+e$ denotes the edge e translated by the vector x . Let $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ be the canonical basis of \mathbb{Z}^d . We may identify each vector \mathbf{e}_i with the corresponding edge $\{0, \mathbf{e}_i\}$ when convenient. The discrete gradient ∇ acts on functions $f : \mathbb{Z}^d \rightarrow \mathbb{R}$ according to

$$\nabla f(x) := (f(x + \mathbf{e}_1) - f(x), \dots, f(x + \mathbf{e}_d) - f(x)),$$

while the discrete divergence acts on function $F = (F_1, \dots, F_d) : \Omega \rightarrow \mathbb{R}^d$ via

$$\nabla \cdot F(x) := \sum_{i=1}^d (F_i(x) - F_i(x - \mathbf{e}_i)).$$

One can check that up to a minus sign, the gradient and divergence operators are formal adjoints of one another with respect to the counting measure on \mathbb{Z}^d . The operator of interest to us is defined by

$$L := -\nabla \cdot \mathbf{a} \nabla,$$

where here we interpret $\mathbf{a}(x)$ as the diagonal matrix $\text{diag}(\mathbf{a}_{x+\mathbf{e}_1}, \dots, \mathbf{a}_{x+\mathbf{e}_d})$. More explicitly, for every $f : \mathbb{Z}^d \rightarrow \mathbb{R}$,

$$Lf(x) = \sum_{y \sim x} \mathbf{a}_{xy} (f(x) - f(y)).$$

It is useful to lift this operator to stationary functions on Ω . We first define the corresponding discrete gradient D , which acts on functions $f : \Omega \rightarrow \mathbb{R}$ via

$$Df(\mathbf{a}) := (f(\theta_{\mathbf{e}_1} \mathbf{a}) - f(\mathbf{a}), \dots, f(\theta_{\mathbf{e}_d} \mathbf{a}) - f(\mathbf{a})),$$

and the discrete divergence, which acts on functions $F = (F_1, \dots, F_d) : \Omega \rightarrow \mathbb{R}^d$ via

$$D \cdot F(\mathbf{a}) := \sum_{i=1}^d (F_i(\mathbf{a}) - F_i(\theta_{-\mathbf{e}_i} \mathbf{a})).$$

One can verify that up to a minus sign, the gradient and divergence operators are formal adjoints of one another under the measure \mathbb{P} . The “lifted” operator is

$$(8.1) \quad \mathcal{L} := -D \cdot \mathbf{a} D,$$

where we understand here that $\mathbf{a} = \text{diag}(\mathbf{a}_{\mathbf{e}_1}, \dots, \mathbf{a}_{\mathbf{e}_d})$. In a more explicit form, for every $f : \Omega \rightarrow \mathbb{R}$,

$$\mathcal{L}f(\mathbf{a}) = \sum_{z \sim 0} \mathbf{a}_{0z} (f(\mathbf{a}) - f(\theta_z \mathbf{a})).$$

The operator \mathcal{L} is self-adjoint on $L^2(\Omega)$. With each function $f : \Omega \rightarrow \mathbb{R}$, one can associate a stationary extension $\tilde{f} : \Omega \times \mathbb{Z}^d \rightarrow \mathbb{R}$ defined by

$$\tilde{f}(\mathbf{a}, x) := f(\theta_x \mathbf{a}).$$

Any function on $\Omega \times \mathbb{Z}^d$ that can be written in this way is said to be stationary.

We let $L^2_{\text{pot}}(\Omega)$ denote the closure in $L^2(\Omega, \mathbb{R}^d)$ of the set of gradient fields $\{Df : f \in L^2(\Omega, \mathbb{R}^d)\}$. Each element of $L^2_{\text{pot}}(\Omega)$ can be realized as the gradient of a possibly non-stationary function $\Omega \times \mathbb{Z}^d \rightarrow \mathbb{R}$. Naturally, the gradient determines the function up to an additive constant. We denote by Df any element of $L^2_{\text{pot}}(\Omega)$, keeping in mind that f may be a non-stationary function. We fix once and for all a unit vector $\xi \in \mathbb{R}^d$, and define the gradient of the corrector in the direction of ξ as the unique $D\phi \in L^2_{\text{pot}}(\Omega)$ solution to

$$(8.2) \quad -D \cdot \mathbf{a}(\xi + D\phi) = 0.$$

The equation (8.2) is interpreted in the weak sense of

$$(8.3) \quad \forall Df \in L^2_{\text{pot}}(\Omega), \quad \mathbb{E}[Df \cdot \mathbf{a}(\xi + D\phi)] = 0.$$

The existence of $D\phi$ can be showed by considering first the approximate correctors $(\phi_\mu)_{\mu > 0}$. For each $\mu > 0$, the approximate corrector $\phi_\mu \in L^2(\Omega)$ solves an equation with an additional massive term:

$$(8.4) \quad \mu\phi_\mu - D \cdot \mathbf{a}(\xi + D\phi_\mu) = 0,$$

and its gradient converges to $D\phi$:

$$(8.5) \quad D\phi_\mu \xrightarrow[\mu \rightarrow 0]{L^2(\Omega)} D\phi.$$

Indeed, energy estimates show that $D\phi_\mu$ is bounded in $L^2(\Omega)$, and therefore converges weakly to some element of $L^2_{\text{pot}}(\Omega)$ up to extraction. Using also that $\mu^{\frac{1}{2}}\phi_\mu$ is bounded in $L^2(\Omega)$, we infer that the convergence in (8.5) holds in the weak sense. In order to justify strong convergence in $L^2(\Omega)$, it then suffices to verify that

$$\limsup_{\mu \rightarrow 0} \mathbb{E}[D\phi_\mu \cdot \mathbf{a}D\phi_\mu] \leq \mathbb{E}[D\phi \cdot \mathbf{a}D\phi].$$

This follows from

$$\mathbb{E}[\mu\phi_\mu^2 + D\phi_\mu \cdot \mathbf{a}D\phi_\mu] = -\mathbb{E}[D\phi_\mu \cdot \mathbf{a}\xi]$$

and the fact that the right side converges to

$$-\mathbb{E}[D\phi \cdot \mathbf{a}\xi] = \mathbb{E}[D\phi \cdot \mathbf{a}D\phi].$$

As is well-known, the discrete operator $-\nabla \cdot \mathbf{a}\nabla$ homogenizes to the continuous differential operator $-\nabla \cdot \bar{\mathbf{a}}\nabla$, where $\bar{\mathbf{a}}$ is the constant-in-space, deterministic matrix characterized by the relation

$$(8.6) \quad \xi \cdot \bar{\mathbf{a}}\xi = \mathbb{E}[(\xi + D\phi) \cdot \mathbf{a}(\xi + D\phi)].$$

Using the weak formulation (8.3), we can rewrite this formula as

$$(8.7) \quad \xi \cdot \bar{\mathbf{a}}\xi = \mathbb{E}[\xi \cdot \mathbf{a}\xi] - \mathbb{E}[D\phi \cdot \mathbf{a}D\phi].$$

As was stated above, it is easy to compute the first term on the right side of (8.7). This term is the linear part of the mapping going from coefficient fields to homogenized matrices. One way to understand why formula (8.7) may be better suited than (8.6) for the development of numerical methods is by considering the case of small ellipticity contrast. If we assume, say, that $1 \leq \mathbf{a} \leq 1 + \eta$ for some small constant $\eta > 0$, then the gradient $|D\phi|$ will be of the order of η , and thus enters as a perturbation of the order of η in the computation of $(\xi + D\phi) \cdot \mathbf{a}(\xi + D\phi)$. It is only after we take expectations that we can discover that the difference $|\bar{\mathbf{a}} - \mathbb{E}[\mathbf{a}]|$ is in fact of the order of η^2 . In contrast, this property is directly apparent in (8.7). The method presented in Theorem 1.2 can be understood as a way to leverage on this phenomenon in general, through a coarsening of the problem which progressively inherits the properties of the small-ellipticity-contrast regime.

Denoting by

$$(8.8) \quad f(\mathbf{a}) := D \cdot \mathbf{a}\xi,$$

and by $\langle \cdot, \cdot \rangle$ the scalar product in $L^2(\mathbb{P})$, we can rewrite the second term on the right side of (8.7) as

$$(8.9) \quad \mathbb{E}[D\phi \cdot \mathbf{a}D\phi] = \lim_{\mu \rightarrow 0} \langle f, (\mu + \mathcal{L})^{-1}f \rangle = \langle f, \mathcal{L}^{-1}f \rangle.$$

Indeed, we have that $(\mu + \mathcal{L})^{-1}f = \phi_\mu$ and

$$\langle f, \phi_\mu \rangle = -\mathbb{E}[D\phi_\mu \cdot \mathbf{a}\xi] \xrightarrow{\mu \rightarrow 0} -\mathbb{E}[D\phi \cdot \mathbf{a}\xi] = \mathbb{E}[D\phi \cdot \mathbf{a}D\phi].$$

The problem of computing $\langle f, \mathcal{L}^{-1}f \rangle$ is precisely of the type discussed in the first part of the paper. Since the homogenization of the operator $-\nabla \cdot \mathbf{a}\nabla$ is essentially equivalent to the large-scale diffusive behavior of the corresponding random walk (which itself can be reduced to the study of an additive functional [40]), this is not surprising. We point out that no reference to random walks is necessary to prove the results of Section 5 there, which are the only results of the first part we will need here. Indeed, all the arguments in Section 5 only rely on the classical resolvent formula (2.7) and on the property (3.4). That the latter is true follows from (8.9) and the fact that, by reversibility, $\langle f, e^{-t\mathcal{L}}f \rangle \geq 0$. Hence, the reader who may have skipped the first part of the paper can read Section 5 independently of the rest of the first part, with the understanding that \mathcal{L} and f are defined by (8.1) and (8.8) respectively.

In order to compute $\langle f, \mathcal{L}^{-1}f \rangle$, we employ the strategy described in Theorem 5.1 with $\mu_k \equiv 2^{-k}$. That is, we set $f_{-1} = f$ and, for each $k \in \mathbb{N}$, denote by $f_k \in L^2(\Omega)$ the unique solution to

$$(8.10) \quad (2^{-k} + \mathcal{L})f_k = 2^{-k}f_{k-1}.$$

We thus infer from Theorem 5.1 that, for each $n \in \mathbb{N}$,

$$(8.11) \quad \langle f, \mathcal{L}^{-1}f \rangle = \sum_{k=0}^n 2^k (\langle f_{k-1}, f_k \rangle + \langle f_k, f_k \rangle) + \langle f_n, \mathcal{L}^{-1}f_n \rangle.$$

For each $k \geq -1$, we denote by $(v_k(x), x \in \mathbb{Z}^d)$ the stationary extension of f_k , that is,

$$v_k(x) := f_k(\theta_x \mathbf{a}).$$

The dependence of v_k on the coefficient field \mathbf{a} is left implicit in the notation. Using (8.10), one can check that v_k solves (1.2).

9. ERROR ANALYSIS

The goal of this section is to prove the first part of Theorem 1.2, which gives an estimate of the error in the approximation provided by $\widehat{\sigma}_n^2$. Recall that f is defined in (8.8) and the identities (8.7) and (8.9).

Theorem 9.1 (Error estimate). *For each $\varepsilon \in (0, \frac{1}{2})$, there exists a constant $C < \infty$ such that the following holds. Fix $n \in \mathbb{N}$, and for each $k \leq n$, denote*

$$(9.1) \quad r_k := 2^{n - (\frac{1}{2} - \varepsilon)k}.$$

We have

$$(9.2) \quad \mathbb{E} \left[\left(\langle f, \mathcal{L}^{-1} f \rangle - \sum_{k=0}^n \frac{1}{|B_{r_k}|} \sum_{x \in B_{r_k}} 2^k (v_{k-1}(x)v_k(x) + v_k^2(x)) \right)^2 \right]^{\frac{1}{2}} \leq C 2^{-\frac{nd}{2}}.$$

Remark 9.2. The proof given below in fact gives the existence of a constant $C < \infty$ such that for any choice of $(r_k)_{0 \leq k \leq n}$, we have the inequality (9.2) with the right side replaced by

$$C 2^{-\frac{dn}{2}} + C \sum_{k=0}^n 2^{-\frac{dk}{4}} r_k^{-\frac{d}{2}}.$$

Fixing r_k according to (9.1) for some $\varepsilon > 0$ indeed allows to bound the sum above by a constant times $2^{-\frac{dn}{2}}$, but other choices are possible, such as, say,

$$r_k = 2^{n - \frac{k}{2}} \left(1 + \frac{k^{\frac{3}{d}}}{100} \right).$$

For $\varepsilon = 0$, we obtain an upper bound of the form $Cn 2^{-\frac{nd}{2}}$. I conjecture that the sharp upper bound in (9.2) is

$$(9.3) \quad C 2^{-\frac{dn}{2}} + C 2^{-\frac{nd}{2}} \left(\sum_{k=0}^n 2^{d(\frac{k}{2} - n)} r_k^d \right)^{\frac{1}{2}},$$

which for the choice of r_k as in (9.1) with $\varepsilon = 0$ gives the upper bound $Cn^{\frac{1}{2}} 2^{-\frac{nd}{2}}$. As will be seen in the next section, for r_k chosen as in (9.1), the asymptotic complexity of the method described in Theorem 1.2 does not depend on the choice of $\varepsilon \in [0, \frac{d-1}{2d})$.

In order to prove Theorem 1.2, our first task is to estimate the size of the remainder in the series expansion displayed in (8.11).

Proposition 9.3 (Remainder estimate). *There exists a constant $C < \infty$ such that for every $n \in \mathbb{N}$,*

$$0 \leq \langle f_n, f_n \rangle \leq C2^{-n(\frac{d}{2}+1)}$$

and

$$0 \leq \langle f_n, \mathcal{L}^{-1} f_n \rangle \leq C2^{-\frac{nd}{2}}.$$

Proposition 9.3 is an immediate consequence of Propositions 5.3 and 5.5, together with the second moment bound proved in [28, Theorem 1] and recalled in the next proposition.

Proposition 9.4 ([28]). *Recall that $f(\mathbf{a}) = D \cdot \mathbf{a}\xi$. For each $p \in [1, \infty)$, there exists a constant $C < \infty$ such that for every $t \geq 0$,*

$$\mathbb{E} \left[|e^{-t\mathcal{L}} f|^p \right]^{\frac{1}{p}} \leq C(1+t)^{-\left(\frac{d}{4} + \frac{1}{2}\right)}.$$

We rely on the spatial mixing property of the random field in (1.6) to compute the expected values in the series expansion in (8.11). This mixing property is quantified by the next proposition. For every random variable X , we denote by $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$ the variance of X .

Proposition 9.5 (Spatial mixing of v_k). *There exists $C(d, \Lambda) < \infty$ such that for every $k \in \mathbb{N}$ and $r \geq 0$,*

$$(9.4) \quad \text{Var} \left[\frac{1}{|B_r|} \sum_{x \in B_r} (v_{k-1}(x)v_k(x) + v_k^2(x)) \right] \leq C2^{-2k(\frac{d}{2}+1)} \left(1 + 2^{-\frac{k}{2}} r\right)^{-d}.$$

As will be seen shortly, we have

$$(9.5) \quad \mathbb{E}[v_k^4(x)] = \mathbb{E}[f_k^4] \leq C2^{-2k(\frac{d}{2}+1)},$$

which is the same scaling as the square of the second moment, see Proposition 9.3. Roughly speaking, Proposition 9.5 shows that the correlation length of the random field $x \mapsto v_{k-1}(x)v_k(x) + v_k^2(x)$ is of the order of $2^{\frac{k}{2}}$. Indeed, the fact that the left side of (9.4) is bounded by $C2^{-2k(\frac{d}{2}+1)}$ is a direct consequence of the bound (9.5). The additional term on the right side of (9.4) witnesses CLT-type cancellations at length scale $2^{\frac{k}{2}}$.

In order to prove Proposition 9.5, we define, for each $t \geq 0$ and $x \in \mathbb{Z}^d$,

$$u(t, x) := (e^{-t\mathcal{L}} f)(\theta_x \mathbf{a}).$$

The function u solves the parabolic equation

$$(9.6) \quad \begin{cases} \partial_t u = \nabla \cdot \mathbf{a} \nabla u & \text{on } \mathbb{R}_+ \times \mathbb{Z}^d, \\ u(0, \cdot) = \nabla \cdot \mathbf{a} \xi & \text{on } \mathbb{Z}^d. \end{cases}$$

We also denote by $(G(t, x, y), t \geq 0, x, y \in \mathbb{Z}^d)$ the parabolic Green function associated with the operator $-\nabla \cdot \mathbf{a} \nabla$. We will make use of [28, Theorem 3], which we now recall. For each $e \in \mathbb{B}$, $x \in \mathbb{Z}^d$ and $t \geq 0$, we introduce the weight function

$$w_e(t, x) := 1 + \frac{|x - e|}{(1+t)^{\frac{1}{2}}},$$

where we use the notation $|x - e|$ to denote the distance between x and the endpoint of the edge e closest to x .

Proposition 9.6 ([28]). *There exists an exponent $p_0(d, \Lambda) > 1$ and, for every $\alpha < \infty$, a constant $C(d, \Lambda, \alpha) < \infty$ such that for every $1 \leq p \leq p_0$, $t \geq 0$ and $e \in \mathbb{B}$,*

$$\left(\sum_{x \in \mathbb{Z}^d} (w_e^\alpha(t, x) |\nabla G(t, x, e)|^2)^p \right)^{\frac{1}{2p}} \leq C(1+t)^{-\left(\frac{d}{2} + \frac{1}{2}\right) + \frac{d}{2} \frac{1}{2p}}.$$

We will also rely on the following observation.

Lemma 9.7. *There exists $C(d, \Lambda) < \infty$ such that for every $t \geq 0$ and $e \in \mathbb{B}$,*

$$\mathbb{E} [|\nabla u(t, e)|^2] \leq C(1+t)^{-\left(\frac{d}{2} + 2\right)}.$$

Proof. We first observe that the mapping $t \mapsto \mathbb{E}[(\nabla u \cdot \mathbf{a} \nabla u)(t, x)]$ is nonincreasing. Indeed, by stationarity, and dropping the variables (t, x) from the notation,

$$\begin{aligned} \partial_t \mathbb{E}[\nabla u \cdot \mathbf{a} \nabla u] &= 2\mathbb{E}[\nabla(\nabla \cdot \mathbf{a} \nabla u) \cdot \mathbf{a} \nabla u] \\ &= -2\mathbb{E}[(\nabla \cdot \mathbf{a} \nabla u)^2] \leq 0. \end{aligned}$$

Moreover,

$$\partial_t \mathbb{E}[u^2] = -2\mathbb{E}[\nabla u \cdot \mathbf{a} \nabla u].$$

Using also Proposition 9.4 with $p = 2$, we obtain

$$\begin{aligned} \mathbb{E}[(\nabla u \cdot \mathbf{a} \nabla u)(t, x)] &\leq \frac{2}{t} \int_{\frac{t}{2}}^t \mathbb{E}[(\nabla u \cdot \mathbf{a} \nabla u)(s, x)] ds \\ &= t^{-1} (\mathbb{E}[u^2(\frac{t}{2}, x)] - \mathbb{E}[u^2(t, x)]) \\ &\leq C(1+t)^{-\left(\frac{d}{2} + 2\right)}. \end{aligned}$$

The announced result then follows by the assumption of uniform ellipticity of the coefficient field. \square

Proof of Proposition 9.5. For notational convenience, we focus on showing that

$$(9.7) \quad \mathbb{V}\text{ar} \left[\frac{1}{|B_r|} \sum_{x \in B_r} v_k^2(x) \right] \leq C 2^{-2k\left(\frac{d}{2} + 1\right)} \left(1 + 2^{-\frac{k}{2}} r\right)^{-d}.$$

The proof that

$$\mathbb{V}\text{ar} \left[\frac{1}{|B_r|} \sum_{x \in B_r} v_{k-1}(x) v_k(x) \right] \leq C 2^{-2k\left(\frac{d}{2} + 1\right)} \left(1 + 2^{-\frac{k}{2}} r\right)^{-d}$$

is essentially identical, and we therefore omit it. We decompose the proof into four steps.

Step 1. We show that there exists a constant $C < \infty$ such that for every $k \in \mathbb{N}$ and $x \in \mathbb{Z}^d$,

$$(9.8) \quad \mathbb{E}[v_k^4(x)] \leq C 2^{-k(d+2)}.$$

Denote by S_k the random variable in (5.10). We assume that this random variable is independent of the coefficient field \mathbf{a} , and denote the expectation with respect to this random variable by E . We recall that we have the representation

$$(9.9) \quad v_k(x) = E[u(S_k, x)].$$

By Jensen's inequality and Proposition 9.4 (with $p = 4$), we have

$$\mathbb{E}[v_k^4(x)] \leq E E[u^4(S_k, x)] \leq E[(1 + S_k)^{-(d+2)}].$$

The conclusion (9.8) then follows from (5.12).

Step 2. By Jensen's inequality and (9.8), we obtain the inequality (9.7) provided that $r \leq 2^{\frac{k}{2}}$. From now on, we therefore assume that $r \geq 2^{\frac{k}{2}}$. We start by recalling the Efron-Stein (or spectral gap) inequality. Let $(\mathbf{a}'_e)_{e \in \mathbb{B}}$ be an independent copy of the environment $(\mathbf{a}_e)_{e \in \mathbb{B}}$, defined on the same probability space. For each edge $e \in \mathbb{B}$, we let \mathbf{a}^e be the environment defined by

$$(9.10) \quad \mathbf{a}_b^e := \begin{cases} \mathbf{a}_b & \text{if } b \neq e, \\ \mathbf{a}'_e & \text{if } b = e. \end{cases}$$

In other words, the environment \mathbf{a}^e is obtained from the environment \mathbf{a} by resampling the conductance at the edge e . By the independence assumption on the conductances (\mathbf{a}_e) , every random variable X satisfies the Efron-Stein inequality

$$(9.11) \quad \text{Var}[X] \leq \frac{1}{2} \sum_{e \in \mathbb{B}} \mathbb{E}[(\partial_e X)^2],$$

where $\partial_e X$ denotes the Glauber derivative:

$$\partial_e X(a) := X(\mathbf{a}^e) - X(\mathbf{a}).$$

Applying this inequality to our problem yields that

$$\text{Var} \left[\sum_{x \in B_r} v_k^2(x) \right] \leq \frac{1}{2} \sum_{e \in \mathbb{B}} \mathbb{E} \left[\left(\sum_{x \in B_r} \partial_e (v_k^2(x)) \right)^2 \right].$$

For each edge e , we write

$$v_k^{(e)}(x) := |v_k(\mathbf{a}^e, x)| + |v_k(\mathbf{a}, x)|.$$

where we stressed the dependency on the environment \mathbf{a} or \mathbf{a}^e in the notation. Note that the environment \mathbf{a}^e has the same law as \mathbf{a} . Moreover,

$$|\partial_e (v_k^2(x))| \leq v_k^{(e)}(x) |\partial_e v_k(x)|.$$

We fix an exponent $\alpha > d$. By Hölder's inequality,

$$\begin{aligned} \sum_{x \in B_r} \partial_e (v_k^2(x)) &\leq \sum_{x \in B_r} v_k^{(e)}(x) |\partial_e v_k(x)| \\ &\leq \left(\sum_{x \in B_r} w_e^{-\alpha}(2^k, x) |v_k^{(e)}(x)|^2 \right)^{\frac{1}{2}} \left(\sum_{x \in B_r} w_e^{\alpha}(2^k, x) |\partial_e v_k(x)|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Therefore,

$$(9.12) \quad \mathbb{V}\text{ar} \left[\sum_{x \in B_r} v_k^2(x) \right] \\ \leq \sum_{e \in \mathbb{B}} \mathbb{E} \left[\left(\sum_{x \in B_r} w_e^{-\alpha}(2^k, x) |v_k^{(e)}(x)|^2 \right)^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\left(\sum_{x \in B_r} w_e^{\alpha}(2^k, x) |\partial_e v_k(x)|^2 \right)^2 \right]^{\frac{1}{2}}.$$

By Hölder's inequality,

$$\left(\sum_{x \in B_r} w_e^{-\alpha}(2^k, x) |v_k^{(e)}(x)|^2 \right)^2 \leq \left(\sum_{x \in B_r} w_e^{-\alpha}(2^k, x) \right) \left(\sum_{x \in B_r} w_e^{-\alpha}(2^k, x) |v_k^{(e)}(x)|^4 \right).$$

Hence, by the definition of $v_k^{(e)}$ and (9.8),

$$(9.13) \quad \mathbb{E} \left[\left(\sum_{x \in B_r} w_e^{-\alpha}(2^k, x) |v_k^{(e)}(x)|^2 \right)^2 \right]^{\frac{1}{2}} \leq C 2^{-k} \left(1 + 2^{-\frac{k}{2}} \text{dist}(e, B_r) \right)^{-\alpha}.$$

If we temporarily admit that there exists a constant $C(d, \Lambda, \alpha) < \infty$ such that for every $e \in \mathbb{B}$,

$$(9.14) \quad \mathbb{E} \left[\left(\sum_{x \in B_r} w_e^{\alpha}(2^k, x) |\partial_e v_k(x)|^2 \right)^2 \right]^{\frac{1}{2}} \leq C 2^{-k(\frac{d}{2}+1)},$$

then we can combine the two bounds (9.13) and (9.14) into (9.12) and thus complete the proof of the inequality (9.7). It therefore only remains to show that (9.14) holds. We decompose the proof of this latter statement into two steps.

Step 3. We show that there exists a constant $C < \infty$ such that for every $r, t \geq 0$ and $e \in \mathbb{B}$,

$$(9.15) \quad \mathbb{E} \left[\left(\sum_{x \in B_r} w_e^{\alpha}(t, x) |\partial_e u(t, x)|^2 \right)^2 \right]^{\frac{1}{2}} \leq C(1+t)^{-\left(\frac{d}{2}+1\right)}.$$

We recall from [28, Lemma 3] that there exists $C(d, \Lambda) < \infty$ such that

$$(9.16) \quad |\partial_e u(t, x)| \leq C |\nabla G(t, x, e)| + \int_0^t |\nabla G(t-s, x, e) g(s, e)| ds,$$

where $g(s, e)$ is a random variable that satisfies, for each $p < \infty$,

$$(9.17) \quad \mathbb{E} [|g(s, e)|^p]^{\frac{1}{p}} \leq C \mathbb{E} [|\nabla u(s, e)|^p]^{\frac{1}{p}}.$$

By Proposition 9.6 with $p = 1$, the first term on the right side of (9.16) is handled without difficulty. It therefore suffices to show that

$$(9.18) \quad \mathbb{E} \left[\left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) \left| \int_0^t |\nabla G(t-s, x, e) g(s, e)| \, ds \right|^2 \right)^2 \right]^{\frac{1}{2}} \leq C(1+t)^{-\left(\frac{d}{2}+1\right)}.$$

We first note for future reference that by Hölder's inequality, (9.17), Lemma 9.7 and Proposition 9.4, there exists a constant $C(d, \Lambda) < \infty$ such that

$$(9.19) \quad \mathbb{E} [|g(s, e)|^4]^{\frac{1}{4}} \leq \mathbb{E} [|g(s, e)|^2]^{\frac{1}{2}} \mathbb{E} [|g(s, e)|^6]^{\frac{1}{6}} \leq Ct^{-\left(\frac{d}{4}+1\right)}.$$

We now proceed to estimate the sum on the left side of (9.18). By the integral triangle inequality,

$$\begin{aligned} & \left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) \left| \int_0^t |\nabla G(t-s, x, e) g(s, e)| \, ds \right|^2 \right)^{\frac{1}{2}} \\ & \leq \int_0^t \left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) |\nabla G(t-s, x, e) g(s, e)|^2 \right)^{\frac{1}{2}} \, ds. \end{aligned}$$

Applying the triangle inequality again, we arrive at

$$(9.20) \quad \mathbb{E} \left[\left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) \left| \int_0^t |\nabla G(t-s, x, e) g(s, e)| \, ds \right|^2 \right)^2 \right]^{\frac{1}{4}} \leq \int_0^t \mathbb{E} \left[\left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) |\nabla G(t-s, x, e) g(s, e)|^2 \right)^2 \right]^{\frac{1}{4}} \, ds.$$

We fix $p = p_0(d, \Lambda) > 1$ as given by Proposition 9.6, and denote by $q = \frac{p}{p-1}$ its conjugate exponent. We also give ourselves an exponent $\beta > 0$ to be specified later. By Hölder's inequality,

$$\begin{aligned} & \sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) |\nabla G(t-s, x, e) g(s, e)|^2 \\ & \leq \left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) w_e^{\beta p}(t-s, x) |\nabla G(t-s, x, e)|^{2p} \right)^{\frac{1}{p}} \\ & \quad \times \left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) w_e^{-\beta q}(t-s, x) |g(s, e)|^{2q} \right)^{\frac{1}{q}} \end{aligned}$$

By Proposition 9.6, the first term is deterministically bounded by a constant times

$$(1 + |t-s|)^{-(d+1) + \frac{d}{2} \frac{1}{p}}.$$

Moreover, using the triangle inequality once more, we get

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{x \in \mathbb{Z}^d} w_e^\alpha(t, x) w_e^{-\beta q}(t-s, x) |g(s, e)|^{2q} \right)^{\frac{2}{q}} \right]^{\frac{q}{2}} \\ & \leq \sum_{x \in \mathbb{Z}^d} \mathbb{E} \left[\left(w_e^\alpha(t, x) w_e^{-\beta q}(t-s, x) |g(s, e)|^{2q} \right)^{\frac{2}{q}} \right]^{\frac{q}{2}} \\ & \leq C (1 + |t-s|)^{\frac{d}{2}} (1+s)^{-2q(\frac{d}{4}+1)}, \end{aligned}$$

where we chose $\beta < \infty$ sufficiently large and used (9.19) and Proposition 9.4 in the last step. The quantity on the left of (9.20) is therefore bounded by a constant times

$$\begin{aligned} & \int_0^t (1 + |t-s|)^{-\left(\frac{d}{2}+\frac{1}{2}\right)+\frac{d}{2p}} (1 + |t-s|)^{\frac{d}{2} \frac{1}{2q}} (1+s)^{-\left(\frac{d}{4}+1\right)} ds \\ & = \int_0^t (1 + |t-s|)^{-\left(\frac{d}{4}+\frac{1}{2}\right)} (1+s)^{-\left(\frac{d}{4}+1\right)} ds \\ & \leq C(1+t)^{-\left(\frac{d}{4}+\frac{1}{2}\right)}. \end{aligned}$$

This completes the proof of the inequality (9.18).

Step 4. In order to complete the proof, there remains to show that (9.15) implies (9.14). We use the representation (9.9) and Jensen's and the triangle inequalities to bound

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{x \in B_r} w_e^\alpha(2^k, x) |\partial_e v_k(x)|^2 \right)^2 \right]^{\frac{1}{2}} \\ & \leq \mathbb{E} \left[\left(E \left[\sum_{x \in B_r} w_e^\alpha(2^k, x) |\partial_e u(S_k, x)|^2 \right] \right)^2 \right]^{\frac{1}{2}} \\ & \leq E \left[\mathbb{E} \left[\left(\sum_{x \in B_r} w_e^\alpha(2^k, x) |\partial_e u(S_k, x)|^2 \right)^2 \right]^{\frac{1}{2}} \right]. \end{aligned}$$

By the result of the previous step and (5.12), this last term is bounded by a constant times

$$E \left[(1 + S_k)^{-\left(\frac{d}{2}+1\right)} \right] \leq C 2^{-k\left(\frac{d}{2}+1\right)}.$$

This completes the proof. \square

Proof of Theorem 9.1. We decompose the left side of (9.2) into bias and variance. Using Theorem 5.1 and Proposition 9.3, we can bound the bias by

$$\begin{aligned} & \left| \langle f, \mathcal{L}^{-1} f \rangle - \sum_{k=0}^n \frac{1}{|B_{r_k}|} \sum_{x \in B_{r_k}} 2^k \mathbb{E} [v_{k-1}(x) v_k(x) + v_k^2(x)] \right| = |\langle f_n, \mathcal{L}^{-1} f_n \rangle| \\ & \leq C 2^{-\frac{dn}{2}}. \end{aligned}$$

By Proposition 9.5, we have

$$\begin{aligned} \mathbb{V}\text{ar} \left[\frac{1}{|B_{r_k}|} \sum_{x \in B_{r_k}} 2^k (v_{k-1}(x)v_k(x) + v_k^2(x)) \right] &\leq C 2^{-\frac{dk}{2}} r_k^{-d} \\ &\leq C 2^{-dn - \varepsilon dk}. \end{aligned}$$

By the triangle inequality, we deduce that

$$\begin{aligned} \mathbb{V}\text{ar} \left[\sum_{k=0}^n \frac{1}{|B_{r_k}|} \sum_{x \in B_{r_k}} 2^k (v_{k-1}(x)v_k(x) + v_k^2(x)) \right] &\leq C \left(\sum_{k=0}^n 2^{-\frac{1}{2}(dn + \varepsilon dk)} \right)^2 \\ &\leq C 2^{-dn}. \end{aligned}$$

This completes the proof. \square

10. COMPLEXITY ANALYSIS

In this section, we first complete the proof of Theorem 1.2 by evaluating the computational complexity of the method described there. We then prove the complexity lower bound announced in Proposition 1.1.

We recall that the method described in Theorem 1.2 is a family of approximations of the quantity of interest given, for each $n \in \mathbb{N}$, by

$$(10.1) \quad \sum_{k=0}^n \frac{1}{|B_{r_k}|} \sum_{x \in B_{r_k}} 2^k (v_{k-1}(x)v_k(x) + v_k^2(x)),$$

where v_k are functions on \mathbb{Z}^d defined by (1.2), and r_k is defined in (1.3) and depends implicitly on n and on an exponent $\varepsilon \in [0, \frac{1}{2})$. (In view of Remark 9.2, we will include the case $\varepsilon = 0$ in the complexity analysis.) We investigate the complexity of this method under the additional assumption

$$(10.2) \quad \varepsilon < \frac{d-1}{2d}.$$

The question concerns the cost of computing the functions $(v_k, k \in \{0, \dots, n\})$ on B_{r_k} . These functions are solutions of elliptic equations. As we now argue, direct conjugate gradient methods, without any preconditioning, already yield the optimal scaling for the complexity of the method. Indeed, by standard Green function upper bounds (see e.g. [54, Proposition 3.6]), if we solve the equation for v_k with Dirichlet boundary condition outside of $B_{r_k + Cn2^{\frac{k}{2}}}$ for a sufficiently large constant $C < \infty$ (instead of the full space), then we can ensure that the solution thus obtained is 2^{-100dn} away from the true solution v_k . The problem of computing (10.1) therefore boils down to the resolution of a series of elliptic problems, indexed by $k \in \{0, \dots, n\}$, where for each k , we need to invert the operator $(2^{-k} + \nabla \cdot \mathbf{a} \nabla)$ on a domain of side length $r_k + Cn2^{\frac{k}{2}}$. The condition number for this problem is of the order of 2^k . Recall that using the conjugate gradient method, the computational cost of computing the solution to an elliptic problem with N unknowns and condition number κ at precision δ is of the order of $N\sqrt{\kappa} \log(\delta^{-1})$ (see e.g. [58, Theorem 4.12]). Hence, for the problem of computing (10.1), the total complexity using conjugate gradient

methods without preconditioning, and with a precision fixed at, say, 2^{-100dn} , is bounded by a constant times

$$\begin{aligned} n \sum_{k=0}^n \left(r_k + Cn2^{\frac{k}{2}} \right)^d 2^{\frac{k}{2}} &= n \sum_{k=0}^n \left(2^{n-\frac{k}{2}+\varepsilon k} + Cn2^{\frac{k}{2}} \right)^d 2^{\frac{k}{2}} \\ &\leq Cn \sum_{k=0}^n 2^{nd - \left(\frac{d-1}{2} - \varepsilon d\right)k} \\ &\leq Cn2^{nd}, \end{aligned}$$

where we used the assumption of (10.2) in the last step. This completes the proof of Theorem 1.2.

One may argue that, under the specific i.i.d. assumption we make on the coefficient field, periodization would allow to simply get rid of any boundary layer. However, the idea of periodizing the medium is in general not suitable for solving the problem as stated at the onset of the introduction to this paper. Indeed, as soon as some short range of correlation is present in the medium, a periodization of the medium will create boundary defects that cause significant errors. This is discussed in more details in [23], see in particular Figure 9 there.

In the method described in Theorem 1.2, the computational overhead caused by the necessity of boundary layers is an asymptotically small fraction of the total computational cost. More precisely, if we compare, say for $\varepsilon = 0$, the actual cost of the method with the fictitious cost one would face if boundary layers were not computed (i.e. where only values that enter into the formula (10.1) are computed), we find that the difference is bounded by a constant times

$$n^2 \sum_{k=0}^n 2^{\frac{k}{2} + (d-1)\left(n - \frac{k}{2}\right) + \frac{k}{2}} \leq Cn^2 2^{(d-1)n} \times \begin{cases} 2^{\frac{n}{2}} & \text{if } d = 2, \\ n & \text{if } d = 3, \\ 1 & \text{if } d \geq 4. \end{cases}$$

Hence, up to logarithmic corrections, the computational overhead of boundary layers is only of the order of $2^{(d-1)n}$ in dimension $d \geq 3$. As we will argue shortly, no method can compute the homogenized coefficients at precision $2^{-\frac{nd}{2}}$ without evaluating the coefficients in a region of space of volume 2^{nd} , and in this case a boundary layer of unit size already gives a computational overhead of $2^{(d-1)n}$, so this estimate is optimal.

We now turn to proving the complexity lower bound announced in Proposition 1.1. Since this proposition was stated relatively loosely, we first make it more precise. The number of operations of a given algorithm is a random variable; at precision $\delta > 0$, we call this random variable \mathcal{N}_δ . The assumption we wish to contradict is that $\delta^2 \mathcal{N}_\delta$ tends to 0 in probability, as δ tends to 0. We will contradict it for a specific choice of the law of the random environment. However, it will be clear that the proof is very generic, and that the specific choice we make is only a matter of convenient notation and computation. Indeed, the only key point is to observe that “generically”, the Kullback-Leibler divergence (a.k.a. the relative entropy) responds quadratically to perturbations.

Proof of Proposition 1.1. For each $p \in [0, 1]$, let \mathbb{P}_p denote the product measure

$$\mathbb{P}_p = \bigotimes_{e \in \mathbb{B}} (p\delta_1 + (1-p)\delta_2),$$

where δ_x denotes the Dirac probability measure at $x \in \mathbb{R}$. We denote by \mathbb{E}_p the associated expectation, and take $(\mathbf{a}_e)_{e \in \mathbb{B}}$ to be the canonical random variable on $\mathbb{R}^{\mathbb{B}}$. Let $\bar{\mathbf{a}}(p)$ be the homogenized matrix associated with the law \mathbb{P}_p . By [54], the mapping $p \mapsto \bar{\mathbf{a}}(p)$ is C^1 on $[0, 1]$, and it is not constant since $\bar{\mathbf{a}}(0) \neq \bar{\mathbf{a}}(1)$. We will argue that in order for $\bar{\mathbf{a}}(p)$ to be determined within an error of δ , the algorithm must query the values of at least $\simeq \delta^{-2}$ conductances, and therefore perform at least $\simeq \delta^{-2}$ operations. Since the law is of product form, we may assume without loss of generality that the algorithm queries $\mathbf{a}_{e_1}, \mathbf{a}_{e_2}, \dots$ in order, where e_1, e_2, \dots is an arbitrary sequence of distinct elements of \mathbb{B} .

For every $p, q \in (0, 1)$, every $n \in \mathbb{N}$ and every event $A \in \sigma(\mathbf{a}_{e_1}, \dots, \mathbf{a}_{e_n}, e \in \mathcal{E})$,

$$\begin{aligned} \mathbb{P}_q[A] &= \mathbb{E}_p \left[\mathbf{1}_A \prod_{i=1}^n \left(\frac{q}{p} \mathbf{1}_{\mathbf{a}_{e_i}=1} + \frac{1-q}{1-p} \mathbf{1}_{\mathbf{a}_{e_i}=2} \right) \right] \\ (10.3) \quad &= \mathbb{E}_p \left[\mathbf{1}_A \exp \left(\widehat{\text{kl}}(q, p, n) \right) \right], \end{aligned}$$

where we introduced the ‘‘empirical’’ Kullback-Leibler divergence

$$\widehat{\text{kl}}(q, p, n) := \sum_{i=1}^n \log \left(\frac{q}{p} \mathbf{1}_{\mathbf{a}_{e_i}=1} + \frac{1-q}{1-p} \mathbf{1}_{\mathbf{a}_{e_i}=2} \right).$$

The ‘‘true’’ Kullback-Leibler divergence between Bernoulli random variables with parameters q and p is the expectation of each of these summands, that is,

$$\text{kl}(q, p) := p \log \frac{q}{p} + (1-p) \log \frac{1-q}{1-p} \geq 0.$$

Differentiating with respect to q , we verify that for each $\eta \in (0, 1/2)$, there exists $C(\eta) < \infty$ such that

$$\forall p, q \in (\eta, 1-\eta), \quad \text{kl}(q, p) \leq C(q-p)^2.$$

We also verify that

$$\forall p, q \in (\eta, 1-\eta), \quad \log \left(\frac{q}{p} \mathbf{1}_{\mathbf{a}_{e_i}=1} + \frac{1-q}{1-p} \mathbf{1}_{\mathbf{a}_{e_i}=2} \right) \leq C|q-p|.$$

In particular, uniformly over $p, q \in (\eta, 1-\eta)$ and $n \geq 1$,

$$(10.4) \quad \mathbb{E}_p \left[\left(\widehat{\text{kl}}(q, p, n) \right)^2 \right] \leq Cn|q-p|^2 + Cn^2|q-p|^4.$$

If an algorithm can compute $\widehat{\mathbf{a}}_\delta$ in $o(\delta^{-2})$ operations, we can define a deterministic $N_\delta = o(\delta^{-2})$ such that the algorithm queries only the edges e_1, \dots, e_{N_δ} with probability at least $1/2$. Denoting by \mathcal{E}_δ the event that only edges e_1, \dots, e_{N_δ} are queried, we have

$$\mathbb{E}_q \left[(\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(q))^2 \right] \geq \mathbb{E}_p \left[(\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(q))^2 \mathbf{1}_{\mathcal{E}_\delta} \exp \left(\widehat{\text{kl}}(q, p, N_\delta) \right) \right].$$

We choose q_δ in such a way that

$$(10.5) \quad \delta \ll q_\delta - p \ll N_\delta^{-1/2},$$

in the sense that the ratio of the right to the left side of both inequalities diverge to infinity as $\delta \rightarrow 0$. Let \mathcal{E}'_δ be the conjunction of \mathcal{E}_δ with the event that $\widehat{\text{kl}}(q_\delta, p, N_\delta) \geq -1$. Combining the three previous displays, we obtain, for $\delta > 0$ sufficiently small,

$$\mathbb{E}_{q_\delta} [|\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(q_\delta)|^2] \geq \mathbb{E}_p [|\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(q_\delta)|^2 \mathbf{1}_{\mathcal{E}'_\delta}] \exp(-1), \quad \mathbb{P}_p [\mathcal{E}'_\delta] \geq 1/4.$$

Since $\mathbb{E}_p [|\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(p)|^2] \leq \delta^2$, we deduce that

$$\begin{aligned} |\bar{\mathbf{a}}(p) - \bar{\mathbf{a}}(q_\delta)| &\leq 4\mathbb{E}_p [|\bar{\mathbf{a}}(p) - \bar{\mathbf{a}}(q_\delta)| \mathbf{1}_{\mathcal{E}'_\delta}] \\ &\leq 4\mathbb{E}_p [|\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(q_\delta)|^2 \mathbf{1}_{\mathcal{E}'_\delta}]^{1/2} + 4\mathbb{E}_p [|\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(p)|^2 \mathbf{1}_{\mathcal{E}'_\delta}]^{1/2} \\ &\leq 8\mathbb{E}_{q_\delta} [|\widehat{\mathbf{a}}_\delta - \bar{\mathbf{a}}(q_\delta)|^2]^{1/2} + 4\delta \\ &\leq 12\delta. \end{aligned}$$

Since our choice of $p \in (0, 1)$ was arbitrary, and $p \mapsto \bar{\mathbf{a}}(p)$ is C^1 and non-constant over the interval $[0, 1]$, this contradicts (10.5). \square

11. NUMERICAL TESTS

In this section, we report on numerical results for the method presented in Theorem 1.2. For comparison, we also discuss the performance of the RVE method. The code was written in the Julia language, and can be downloaded as part of the source files of the arXiv posting of this paper.

We focus on a two-dimensional example, with conductances taking the values $c_- = 1$ and $c_+ = 9$ with probability $1/2$ each. This is a rare example for which the homogenized matrix is known analytically: it is $\bar{\mathbf{a}} = \sqrt{c_- c_+} \text{Id} = 3 \text{Id}$ (see e.g. [24, Appendix A] or [8, Exercise 2.3]). In this case, we have $\mathbb{E}[\mathbf{a}] = 5 \text{Id}$, and we thus expect that $\widehat{\sigma}_n^2$ converges to 2. We fix $\xi = (1, 0)$.

The best numerical method should probably optimize the parameter r_k in a dynamic way rather than rely on a length scale obtained a priori from theoretical arguments. Here, I strived instead to avoid any fine-tuning and simply chose $\varepsilon = 0$ in the method described in Theorem 9.1. Whenever the operator $(\mu - \nabla \cdot \mathbf{a} \nabla)$ needs to be inverted, I chose the size of the boundary layer to be

$$(11.1) \quad 5 \left(1 \vee \mu^{-\frac{1}{2}}\right) \left(1 \vee \frac{\log\left(\mu^{-\frac{1}{2}}\right)}{\log(2)}\right),$$

where we recall that $x \vee y := \max(x, y)$.

The base-2 logarithm of the left side of (1.5) (with $\varepsilon = 0$), as a function of n , is represented on Figure 11.1. The slope of the regression line there is -1.0 , in full agreement with the theoretical prediction of Theorem 1.2. Moreover, the intercept of the regression line is 1.0 , which indicates that the prefactor constant in (1.5) can be chosen as $C \simeq 2.0$.

The code was run on a laptop computer with 16 Go of memory and a processor clocking at 2.40 GHz. The base-2 logarithm of the time it takes to compute $\widehat{\sigma}_n^2$ on this machine, as a function of n , is reported on Figure 11.2. The slope of the regression line there is 1.9 , very closely agreeing with the

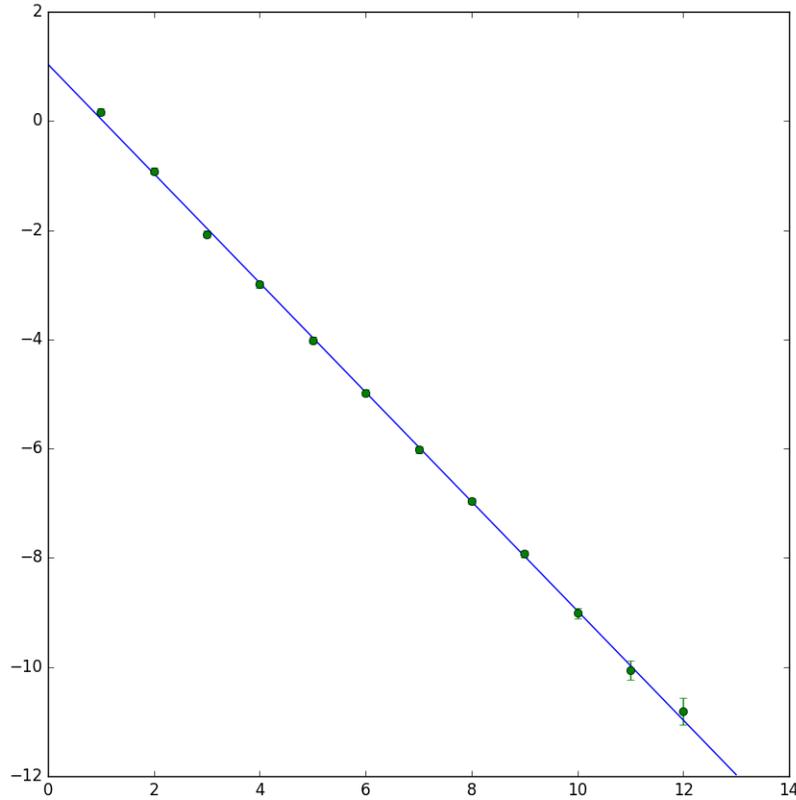


FIGURE 11.1. The base-2 logarithm of the left side of (1.5) (with $\varepsilon = 0$) as a function of n . The error bars are 95% confidence intervals. The slope of the regression line is -1.0 and the intercept is 1.0 .

prediction of Theorem 1.2. The intercept is -15.2 . It takes less than 3 minutes to compute $\widehat{\sigma}_{12}^2$.

We compare these results with the RVE approach, which we now describe more precisely. Recall the definition of the approximate corrector ϕ_μ in (8.4). For each $n \in \mathbb{N}$, we denote $\mu_n := 2^{-n}$ and compute 2^n independent copies $\phi_{\mu_n}^{(1)}, \dots, \phi_{\mu_n}^{(2^n)}$ of ϕ_{μ_n} on a (two-dimensional) box of side length $2^{\frac{n}{2}}$. (This involves the management of boundary layers of side length $Cn2^{\frac{n}{2}}$.) The RVE algorithm then outputs

$$\widetilde{\sigma}_n^2 := \frac{1}{k_n} \sum_{i=1}^{k_n} \frac{1}{|B_{2^{\frac{n}{2}}}|} \sum_{x \in B_{2^{\frac{n}{2}}}} (\xi + \nabla \phi_{\mu_n}^{(i)}) \cdot \mathbf{a}(\xi + \nabla \phi_{\mu_n}^{(i)})(x).$$

It was shown in [30, 31] that there exists a constant $C < \infty$ such that, for every $n \in \mathbb{N}$,

$$(11.2) \quad \mathbb{E} \left[(\xi \cdot \bar{\mathbf{a}}\xi - \widetilde{\sigma}_n^2)^2 \right]^{\frac{1}{2}} \leq C2^{-n}.$$

(Recall that we assume $d = 2$ here.) We stick with the choice of (11.1) for the size of the boundary layer.

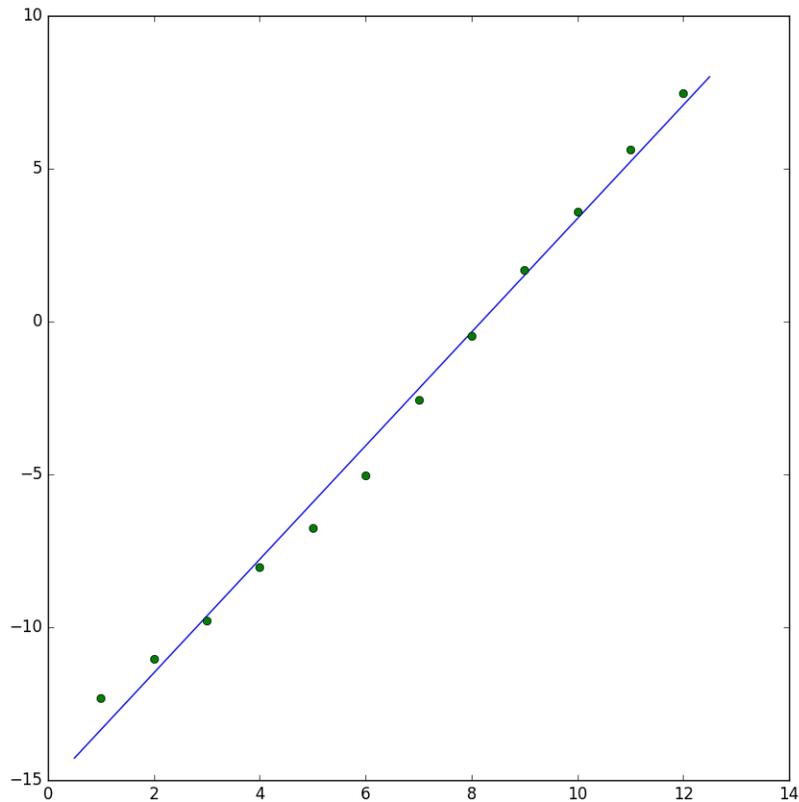


FIGURE 11.2. The base-2 logarithm of the CPU time used to compute $\hat{\sigma}_n^2$, in seconds, as a function of n . The slope of the regression line is 1.9 and the intercept is -15.2 .

Figure 11.3 displays the base-2 logarithm of the left side of (11.2), as a function of n . The slope of the regression line is -0.9 , in very good agreement with the theoretical prediction. Figure 11.4 reports on the cost of computing $\tilde{\sigma}_n^2$, as a function of n . Two implementations were tried. The first uses no preconditioning, and in this case the slope of the regression line in Figure 11.4 is about 2.7, in close agreement with the theoretical prediction of $\frac{5}{2}$. I also implemented a version which uses an incomplete Cholesky factorization as a preconditioner. While this does reduce the number of iterations of the conjugate gradient method, this gain was not sufficient to offset the overhead caused by the preconditioning. The slope of the regression line on Figure 11.4 in this case is 2.6, but the intercept is -12.1 , while the regression line for the non-preconditioned method has an intercept of -15.0 . It takes more than 2 minutes to compute $\tilde{\sigma}_8^2$ using the non-preconditioned method, while computing the quantity $\hat{\sigma}_8^2$ introduced in the present paper takes less than 1 second and yields a better approximation.

The computational complexity reported here for the RVE method seems to be consistent with the numerical findings presented in [24]. Indeed, [24, Figure 14] displays the computational cost of the method as a function of the error, and

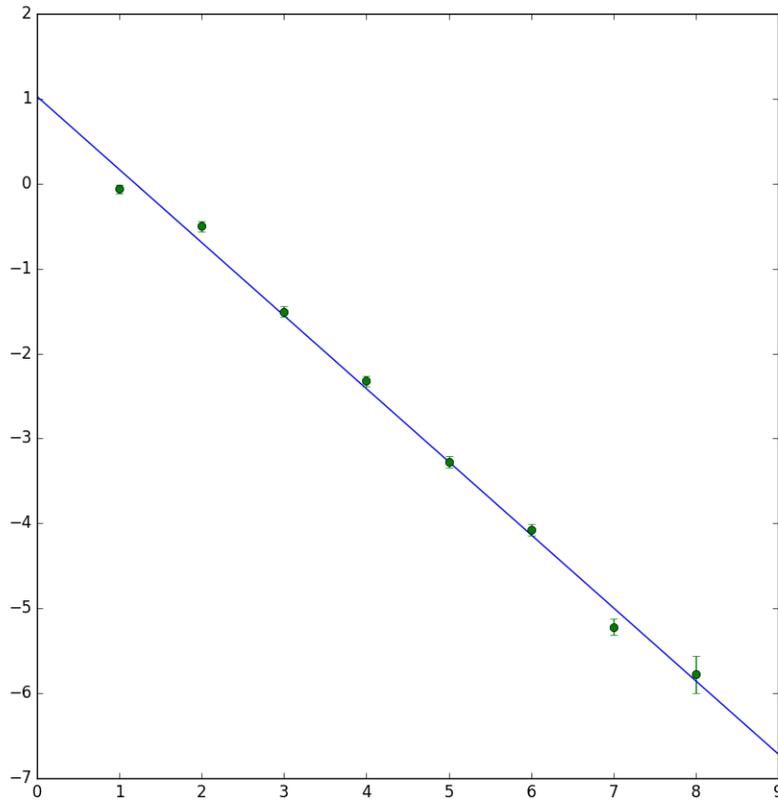


FIGURE 11.3. The base-2 logarithm of the left side of (11.2) as a function of n . The error bars are 95% confidence intervals. The slope of the regression line is -0.9 and the intercept is 1.0 .

suggests that the computational time to reach precision δ scales like $\delta^{-2.6}$, in close agreement with the results reported on Figures 11.3 and 11.4¹.

12. DISCRETE-TIME APPROACH

In this final section, we explore an alternative method based on a discrete-time dynamics associated with the operator \mathcal{L} . Contrary to the approach presented in Theorem 1.2, the method explored here does not seem to generalize to the continuous-space setting. For discrete-space problems, it provides with an approach that does not involve the resolution of linear equations, and which I conjecture to have a computational cost for a precision of $\delta > 0$ of the order of $C\delta^{-2}$ (with no logarithmic correction) in dimension $d \geq 3$. This would match

¹In order to estimate the slope of the regression line of the set of blue dots in [24, Figure 14], I first drew the straight line going through the extremal blue dots, and observed that all the other dots are very close to this line. I then measured the coordinates of the extremal dots to be $(3.59, -0.62)$ and $(9.89, -3.06)$ respectively, which yields that the dx/dy slope of the line is about -2.58 . Moreover, the seven dots closest to the leftmost blue point are below this line, while the three dots closest to the rightmost blue point are above this line, and thus essentially any other reasonable choice of pair of points to draw a line and measure the slope from would yield a larger absolute slope.

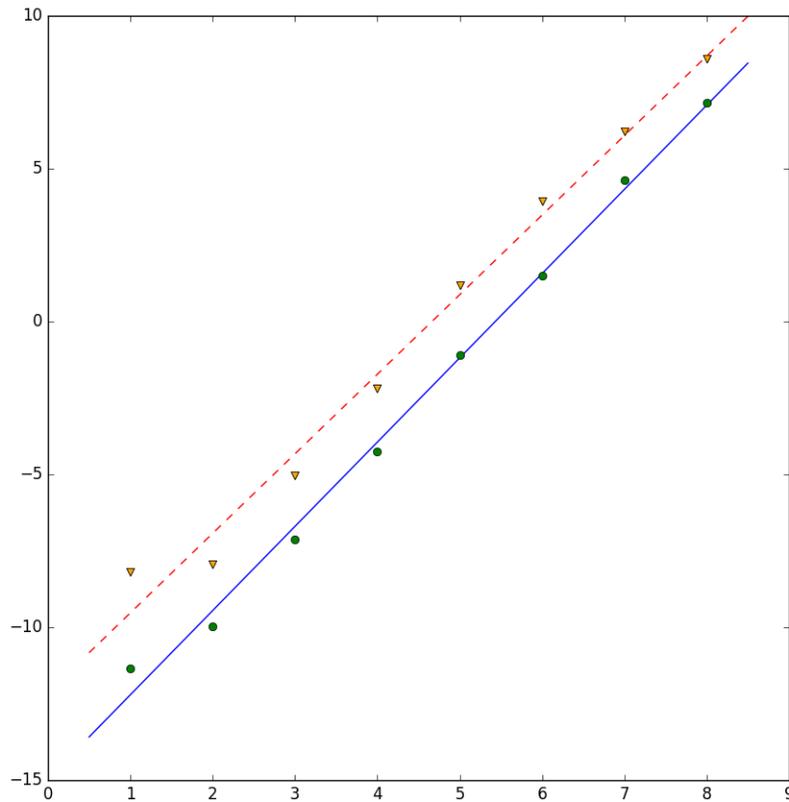


FIGURE 11.4. The base-2 logarithm of the CPU time used to compute $\tilde{\sigma}_n^2$, in seconds, as a function of n , for the non-preconditioned (green dots and solid regression line) and the preconditioned (orange triangles and dashed regression line) methods. The slope and intercept of the regression line are 2.7 and -15.0 respectively for the non-preconditioned method, and 2.6 and -12.1 for the preconditioned method.

exactly the complexity lower bound of Proposition 1.1. In dimension 2, I expect the number of operations to scale like $\delta^{-2} \log(\delta^{-1})$ to yield a precision of $\delta > 0$.

We set

$$\pi(\mathbf{a}) := \sum_{x \sim 0} \mathbf{a}_{0x},$$

and consider the operator

$$\mathcal{L}_1 := -\pi^{-1} D \cdot \mathbf{a} D.$$

This operator is the generator of a discrete-time Markov chain on Ω , and is self-adjoint on $L^2(\Omega, \pi(\mathbf{a}) d\mathbb{P}(\mathbf{a}))$. We denote by $\langle \cdot, \cdot \rangle_\pi$ the scalar product in this L^2 space and set

$$(12.1) \quad g(\mathbf{a}) := \pi^{-1} D \cdot \mathbf{a} \xi.$$

Proposition 12.1. *We have*

$$(12.2) \quad \xi \cdot \bar{\mathbf{a}} \xi = \mathbb{E}[\xi \cdot \mathbf{a} \xi] - \langle g, \mathcal{L}_1^{-1} g \rangle_\pi.$$

Proof. We recall that the rightmost term in (12.2) is interpreted as

$$\lim_{\mu \rightarrow 0} \langle g, (\mu + \mathcal{L}_1)^{-1} g \rangle_\pi,$$

and that

$$\mathbb{E}[\xi \cdot \mathbf{a}\xi] - \xi \cdot \bar{\mathbf{a}}\xi = \mathbb{E}[D\phi \cdot \mathbf{a}D\phi],$$

where $D\phi$ is the L^2 limit of $D\phi_\mu$, see (8.5). Let ψ_μ solve

$$(\mu + \mathcal{L}_1) \psi_\mu = g,$$

that is,

$$(\mu - \pi^{-1}D \cdot \mathbf{a}D) \psi_\mu = \pi^{-1}D \cdot \mathbf{a}\xi.$$

By the same arguments allowing to justify (8.5), we can verify that

$$D\psi_\mu \xrightarrow[\mu \rightarrow 0]{L^2(\Omega)} D\phi,$$

and that $\mu^{\frac{1}{2}}\psi_\mu$ converges to 0 in $L^2(\mathbb{P})$. (This uses that π is uniformly bounded from above and below.) In particular,

$$\mathbb{E}[D\phi \cdot \mathbf{a}D\phi] = \lim_{\mu \rightarrow 0} \mathbb{E}[D\psi_\mu \cdot \mathbf{a}D\psi_\mu].$$

Moreover, by the weak formulation of the equation for ψ_μ , we have

$$\begin{aligned} \langle g, (\mu + \mathcal{L}_1)^{-1} g \rangle_\pi &= \langle g, \psi_\mu \rangle_\pi \\ &= -\mathbb{E}[\mathbf{a}\xi \cdot D\psi_\mu] \\ &= \mathbb{E}[D\psi_\mu \cdot \mathbf{a}D\psi_\mu] + \mu \mathbb{E}[\pi \psi_\mu^2] \\ &\xrightarrow[\mu \rightarrow 0]{} \mathbb{E}[D\phi \cdot \mathbf{a}D\phi]. \end{aligned} \quad \square$$

Note that the operator $\text{Id} - \mathcal{L}_1$ is a contraction in $L^\infty(\Omega)$. It is also self-adjoint, and its spectrum is therefore contained in $[-1, 1]$. The spectrum of \mathcal{L}_1 is thus contained in $[0, 2]$. In order to avoid problems related to the periodicity of the underlying Markov chain, we modify slightly the formula (6.4) for computing $\langle g, \mathcal{L}_1^{-1} g \rangle_\pi$. We use instead that, for every $\lambda \in [0, 2]$,

$$(12.3) \quad \frac{1}{\lambda} = \frac{1}{2} \sum_{k=0}^{+\infty} \left(1 - \frac{\lambda}{2}\right)^k,$$

so that, setting

$$\mathcal{P} := \text{Id} - \frac{1}{2}\mathcal{L}_1,$$

we have by spectral calculus that

$$\begin{aligned} \langle g, \mathcal{L}_1^{-1} g \rangle_\pi &= \frac{1}{2} \sum_{k=0}^{+\infty} \langle g, \mathcal{P}^k g \rangle_\pi \\ (12.4) \quad &= \frac{1}{2} \sum_{k=0}^{+\infty} (\langle \mathcal{P}^k g, \mathcal{P}^k g \rangle_\pi + \langle \mathcal{P}^k g, \mathcal{P}^{k+1} g \rangle_\pi). \end{aligned}$$

In view of the rigorous results in the continuous-time case, we expect that

$$(12.5) \quad |\langle g, \mathcal{P}^k g \rangle_\pi| \leq C(1+k)^{-1-\frac{d}{2}}.$$

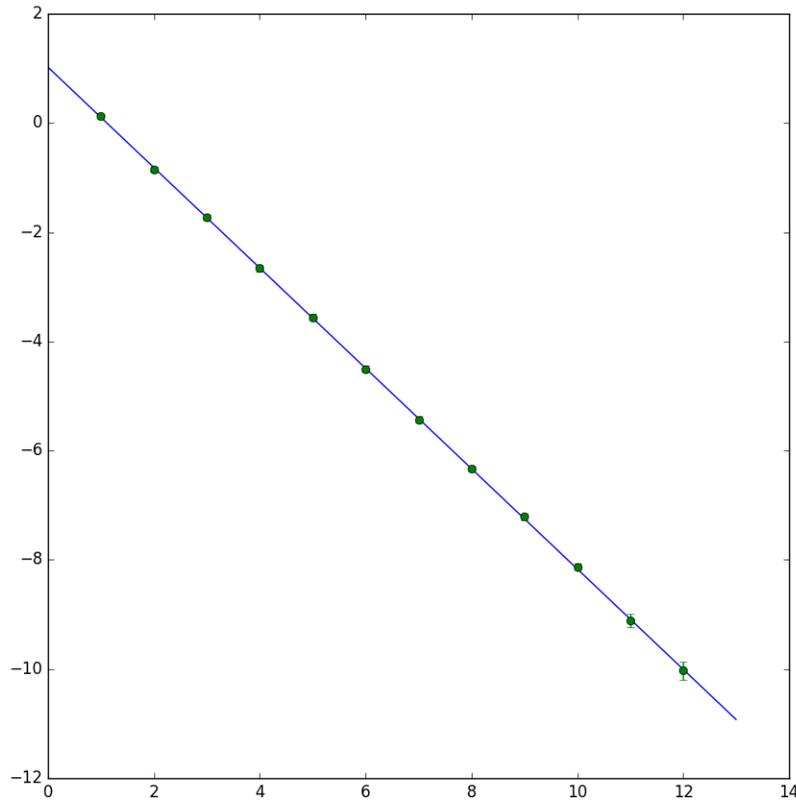


FIGURE 12.1. The base-2 logarithm of the quantity on the left side of (12.8), as a function of L . The error bars are 95% confidence intervals. The slope of the regression line is -0.9 and the intercept is 1.0 .

If we write $w(k, x) := (\mathcal{P}^k g)(\theta_x \mathbf{a})$, then the function w solves the discrete-time parabolic equation

$$(12.6) \quad \begin{cases} w(k+1, x) = w(k, x) + \frac{1}{2\pi(x)} \nabla \cdot \mathbf{a} \nabla w(k, x) & \text{for } (k, x) \in \mathbb{N} \times \mathbb{Z}^d, \\ w(0, x) = \frac{1}{\pi(x)} (\nabla \cdot \mathbf{a} \xi)(x), & \text{for } x \in \mathbb{Z}^d, \end{cases}$$

where we write $\pi(x) := \pi(\theta_x \mathbf{a})$. In analogy with Theorem 1.2 and Remark 9.2, we define, for each $L \in \mathbb{N}$,

$$(12.7) \quad D_L := \sum_{\ell=0}^{L-1} \sum_{k=2^\ell-1}^{2^{\ell+1}-2} \frac{1}{|B_{r(L,\ell)}|} \sum_{x \in B_{r(L,\ell)}} \pi(x) (w^2(k, x) + w(k, x)w(k+1, x)),$$

where

$$r(L, \ell) := 2^{L-\frac{\ell}{2}},$$

and we expect that

$$(12.8) \quad \mathbb{E} \left[(\langle g, \mathcal{L}^{-1} g \rangle_\pi - D_L)^2 \right]^{\frac{1}{2}} \leq CL^{\frac{1}{2}} 2^{-\frac{dL}{2}}.$$

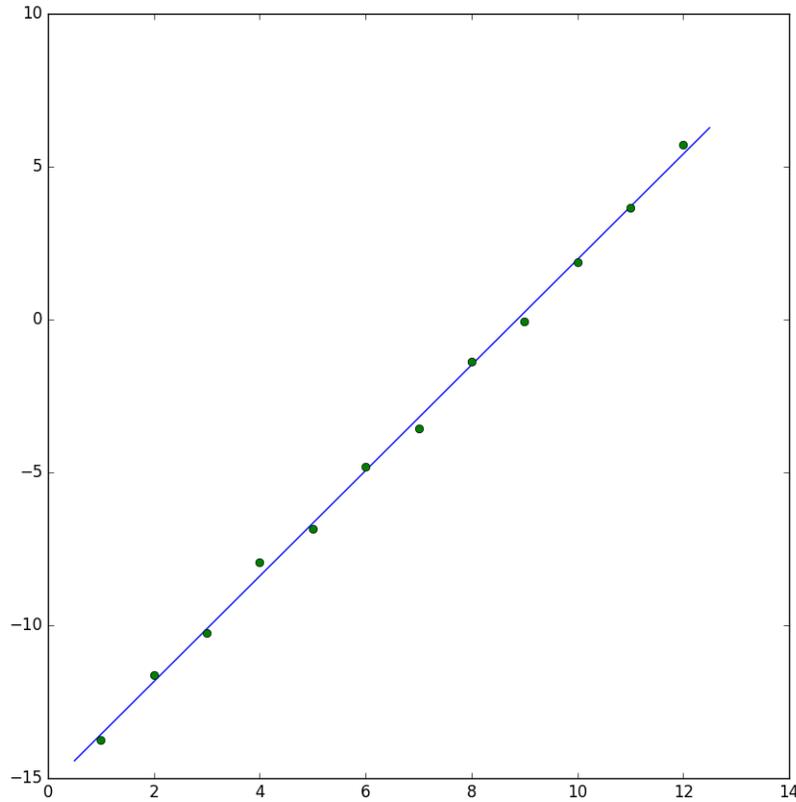


FIGURE 12.2. The base-2 logarithm of the CPU time used to compute D_L , in seconds, as a function of L . The slope of the regression line is 1.7 and the intercept is -15.3 .

Moreover, the extraneous factor of $L^{\frac{1}{2}}$ is expected to be absent if we redefine $r(L, \ell)$ to be $2^{L-\frac{\ell}{2}+\varepsilon\ell}$, for $\varepsilon > 0$ arbitrarily small. Finally, the computational cost of this scheme is given by

$$\sum_{l=0}^{L-1} \sum_{k=2^\ell-1}^{2^{\ell+1}-2} |B_{r(L,\ell)}| \leq C \sum_{l=0}^{L-1} \sum_{k=2^\ell-1}^{2^{\ell+1}-2} 2^{d(L-\frac{\ell}{2})} \leq C 2^{dL} \begin{cases} L & \text{if } d = 2, \\ 1 & \text{if } d \geq 3. \end{cases}$$

If we replace $r(L, \ell)$ by $2^{L-\frac{\ell}{2}+\varepsilon\ell}$, then the complexity of this method degrades in $d = 2$, no matter how small $\varepsilon > 0$ is chosen. This is in contrast with the method explored in the previous sections. On the other hand, in dimension $d \geq 3$ and for $\varepsilon > 0$ sufficiently small, the complexity remains bounded by $C 2^{dL}$ and we that the error bound in (12.8) holds with right-hand side replaced by $2^{-\frac{dL}{2}}$. We thus expect this scheme to match exactly the complexity lower bound given by Proposition 1.1 in dimensions $d \geq 3$.

We now report on numerical tests related to the quantity D_L defined in (12.7). The code can be downloaded as part of the source files of the arXiv posting of this paper. We explore the same two-dimensional example as in Section 12, and use the same computer to run the code. Figure 12.1 displays the base-2 logarithm of the left side of (12.8), as a function of L , while Figure 12.2 displays

the base-2 logarithm of the required CPU time to compute D_L . The numerical findings are very consistent with the heuristic predictions above. It takes less than one minute to compute D_{12} . The ratio of the slopes in Figures 12.2 and 12.1 is about 1.9, in close agreement with the predicted theoretical complexity. The fact that we observe an exponent slightly below 2 may be explained by the diminishing importance taken up by boundary layers as we move to larger values of L .

Acknowledgments. I would like to thank Josselin Garnier for an inspiring talk which motivated me to revisit this problem, Tony Lelièvre for his helpful feedback, and Harmen Stoppels for his precious help with the Julia language. This work has been partially supported by the ANR Grant LSD (ANR-15-CE40-0020-03).

REFERENCES

- [1] A. Abdulle, W. E. B. Engquist, and E. Vanden-Eijnden. The heterogeneous multiscale method. *Acta Numer.*, 21:1–87, 2012.
- [2] Y. Almog. Averaging of dilute random media: a rigorous proof of the Clausius-Mossotti formula. *Arch. Ration. Mech. Anal.*, 207(3):785–812, 2013.
- [3] Y. Almog. The Clausius-Mossotti formula in a dilute random medium with fixed volume fraction. *Multiscale Model. Simul.*, 12(4):1777–1799, 2014.
- [4] A. Anantharaman and C. Le Bris. A numerical approach related to defect-type theories for some weakly random problems in homogenization. *Multiscale Model. Simul.*, 9(2):513–544, 2011.
- [5] A. Anantharaman and C. Le Bris. Elements of mathematical foundations for numerical approaches for weakly random homogenization problems. *Commun. Comput. Phys.*, 11(4):1103–1143, 2012.
- [6] D. Arjmand and O. Runborg. A time dependent approach for removing the cell boundary error in elliptic homogenization problems. *J. Comput. Phys.*, 314:206–227, 2016.
- [7] S. Armstrong, A. Hannukainen, T. Kuusi, and J. C. Mourrat. An iterative method for elliptic problems with rapidly oscillating coefficients, preprint, arXiv:1803.03551.
- [8] S. Armstrong, T. Kuusi, and J.-C. Mourrat. *Quantitative stochastic homogenization and large-scale regularity*. Preliminary version available at www.math.ens.fr/~mourrat/lecturenotes.pdf (2018).
- [9] S. Armstrong, T. Kuusi, and J.-C. Mourrat. Mesoscopic higher regularity and subadditivity in elliptic homogenization. *Comm. Math. Phys.*, 347(2):315–361, 2016.
- [10] S. Armstrong, T. Kuusi, and J.-C. Mourrat. The additive structure of elliptic homogenization. *Invent. Math.*, 208(3):999–1154, 2017.
- [11] S. N. Armstrong and J.-C. Mourrat. Lipschitz regularity for elliptic equations with random coefficients. *Arch. Ration. Mech. Anal.*, 219(1):255–348, 2016.
- [12] S. N. Armstrong and C. K. Smart. Quantitative stochastic homogenization of convex integral functionals. *Ann. Sci. Éc. Norm. Supér. (4)*, 49(2):423–481, 2016.
- [13] M. T. Barlow, A. A. Járai, T. Kumagai, and G. Slade. Random walk on the incipient infinite cluster for oriented percolation in high dimensions. *Comm. Math. Phys.*, 278(2):385–431, 2008.
- [14] G. Ben Arous, M. Cabezas, and A. Fribergh. Scaling limit for the ant in high-dimensional labyrinths, preprint, arXiv:1609.03977.
- [15] G. Ben Arous, M. Cabezas, and A. Fribergh. Scaling limit for the ant in a simple labyrinth, preprint, arXiv:1609.03980.
- [16] L. Berlyand and V. Mityushev. Generalized Clausius-Mossotti formula for random composite with circular fibers. *J. Statist. Phys.*, 102(1-2):115–145, 2001.
- [17] X. Blanc and C. Le Bris. Improving on computation of homogenized coefficients in the periodic and quasi-periodic settings. *Netw. Heterog. Media*, 5(1):1–29, 2010.

- [18] A. Brandt. Multiscale scientific computation: review 2001. In *Multiscale and multiresolution methods*, volume 20 of *Lect. Notes Comput. Sci. Eng.*, pages 3–95. Springer, Berlin, 2002.
- [19] M. Damron, J. Hanson, and P. Sosoe. Subdiffusivity of random walk on the 2D invasion percolation cluster. *Stochastic Process. Appl.*, 123(9):3588–3621, 2013.
- [20] D. Dolgopyat. Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.*, 356(4):1637–1689, 2004.
- [21] M. Duerinckx and A. Gloria. Analyticity of homogenized coefficients under Bernoulli perturbations and the Clausius-Mossotti formulas. *Arch. Ration. Mech. Anal.*, 220(1):297–361, 2016.
- [22] Y. Efendiev and T. Y. Hou. *Multiscale finite element methods*, volume 4 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2009.
- [23] A.-C. Eglhoff, A. Gloria, J.-C. Mourrat, and T. N. Nguyen. Random walk in random environment, corrector equation and homogenized coefficients: from theory to numerics, back and forth. *IMA J. Numer. Anal.*, 35(2):499–545, 2015.
- [24] A. Gloria. Numerical approximation of effective coefficients in stochastic homogenization of discrete elliptic equations. *ESAIM Math. Model. Numer. Anal.*, 46(1):1–38, 2012.
- [25] A. Gloria and Z. Habibi. Reduction in the resonance error in numerical homogenization II: Correctors and extrapolation. *Found. Comput. Math.*, 16(1):217–296, 2016.
- [26] A. Gloria and J.-C. Mourrat. Spectral measure and approximation of homogenized coefficients. *Probab. Theory Related Fields*, 154(1-2):287–326, 2012.
- [27] A. Gloria and J.-C. Mourrat. Quantitative version of the Kipnis-Varadhan theorem and Monte Carlo approximation of homogenized coefficients. *Ann. Appl. Probab.*, 23(4):1544–1583, 2013.
- [28] A. Gloria, S. Neukamm, and F. Otto. Quantification of ergodicity in stochastic homogenization: optimal bounds via spectral gap on Glauber dynamics. *Invent. Math.*, 199(2):455–515, 2015.
- [29] A. Gloria, S. Neukamm, and F. Otto. A regularity theory for random elliptic operators, preprint, arXiv:1409.2678.
- [30] A. Gloria and F. Otto. An optimal variance estimate in stochastic homogenization of discrete elliptic equations. *Ann. Probab.*, 39(3):779–856, 2011.
- [31] A. Gloria and F. Otto. An optimal error estimate in stochastic homogenization of discrete elliptic equations. *Ann. Appl. Probab.*, 22(1):1–28, 2012.
- [32] A. Gloria and F. Otto. The corrector in stochastic homogenization: optimal rates, stochastic integrability, and fluctuations, preprint, arXiv:1510.08290.
- [33] P. Henning and D. Peterseim. Oversampling for the multiscale finite element method. *Multiscale Model. Simul.*, 11(4):1149–1175, 2013.
- [34] T. Y. Hou and X.-H. Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.*, 134(1):169–189, 1997.
- [35] T. Y. Hou, X.-H. Wu, and Z. Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comp.*, 68(227):913–943, 1999.
- [36] B. D. Hughes. Conduction and diffusion in percolating systems. In *Encyclopedia of complexity and systems science*, pages 1395–1424. Springer, 2009.
- [37] A. A. Járai and A. Nachmias. Electrical resistance of the low dimensional critical branching random walk. *Comm. Math. Phys.*, 331(1):67–109, 2014.
- [38] H. Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.*, 22(4):425–487, 1986.
- [39] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and C. Theodoropoulos. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.*, 1(4):715–762, 2003.
- [40] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.

- [41] T. Komorowski, C. Landim, and S. Olla. *Fluctuations in Markov processes*, volume 345 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Heidelberg, 2012.
- [42] S. M. Kozlov. Geometric aspects of averaging. *Uspekhi Mat. Nauk*, 44(2(266)):79–120, 1989.
- [43] G. Kozma and A. Nachmias. The Alexander-Orbach conjecture holds in high dimensions. *Invent. Math.*, 178(3):635–654, 2009.
- [44] T. Kumagai. *Random walks on disordered media and their scaling limits*, volume 2101 of *Lecture Notes in Mathematics*. Springer, Cham, 2014.
- [45] C. Le Bris and F. Legoll. Examples of computational approaches for elliptic, possibly multiscale PDEs with random inputs. *J. Comput. Phys.*, 328:455–473, 2017.
- [46] T. M. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes*, volume 324 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1999.
- [47] T. M. Liggett. *Continuous time Markov processes*, volume 113 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.
- [48] C. Liverani. Central limit theorem for deterministic systems. In *International Conference on Dynamical Systems (Montevideo, 1995)*, volume 362 of *Pitman Res. Notes Math. Ser.*, pages 56–75. Longman, Harlow, 1996.
- [49] A. Målqvist and D. Peterseim. Localization of elliptic multiscale problems. *Math. Comp.*, 83(290):2583–2603, 2014.
- [50] D. Marahrens and F. Otto. Annealed estimates on the Green function. *Probab. Theory Related Fields*, 163(3-4):527–573, 2015.
- [51] J. C. Maxwell. Medium in which small spheres are uniformly disseminated. *A treatise on electricity and magnetism*, part II, chapter IX, article 314. Clarendon Press, 3d ed., 1891.
- [52] I. Melbourne and M. Nicol. Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys.*, 260(1):131–146, 2005.
- [53] J.-C. Mourrat. Variance decay for functionals of the environment viewed by the particle. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(1):294–327, 2011.
- [54] J.-C. Mourrat. First-order expansion of homogenized coefficients under Bernoulli perturbations. *J. Math. Pures Appl. (9)*, 103(1):68–101, 2015.
- [55] G. Papanicolaou and S. R. S. Varadhan. Ornstein-Uhlenbeck process in a random potential. *Comm. Pure Appl. Math.*, 38(6):819–834, 1985.
- [56] G. C. Papanicolaou. Diffusion in random media. In *Surveys in applied mathematics, Vol. 1*, pages 205–253. Plenum, New York, 1995.
- [57] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1995.
- [58] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical mathematics*, volume 37 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2000.
- [59] J. W. Strutt, 3d Baron Rayleigh. On the influence of obstacles arranged in rectangular order upon the properties of a medium. *Philos. mag.*, 34(211):481–502, 1892.
- [60] X. Yue and W. E. The local microscale problem in the multiscale modeling of strongly heterogeneous media: effects of boundary conditions and cell size. *J. Comput. Phys.*, 222(2):556–572, 2007.

(J.-C. Mourrat) DMA, ECOLE NORMALE SUPÉRIEURE, CNRS, PSL RESEARCH UNIVERSITY, PARIS, FRANCE

E-mail address: mourrat@dma.ens.fr