Special Quasirandom Structures: a selection approach for stochastic homogenization

C. Le Bris, F. Legoll and W. Minvielle

École des Ponts and INRIA,

6 et 8 avenue Blaise Pascal, 77455 Marne-La-Vallée Cedex 2, France

`{lebris,william.minvielle}@cermics.enpc.fr`

`legoll@lami.enpc.fr`

August 2, 2018

Abstract

We adapt and study a variance reduction approach for the homogenization of elliptic equations in divergence form. The approach, borrowed from atomistic simulations and solid-state science [23, 24, 25], consists in selecting random realizations that best satisfy some statistical properties (such as the volume fraction of each phase in a composite material) usually only obtained asymptotically.

We study the approach theoretically in some simplified settings (one-dimensional setting, perturbative setting in higher dimensions), and numerically demonstrate its efficiency in more general cases.

1   Introduction

1.1   Overview

In this article, we adapt, theoretically study and numerically test a specific variance reduction approach for the numerical homogenization of an elliptic equation with heterogeneous random coefficients.

The equation we consider is the following scalar elliptic equation in divergence form

$$- \operatorname{div}\left(A\left(\frac{\cdot}{\varepsilon},\omega\right) \nabla u^{\varepsilon}(\cdot,\omega)\right) = f \ \text{ in } \mathcal{D}, \qquad u^{\varepsilon}(\cdot,\omega) = 0 \ \text{ on } \partial\mathcal{D}, \ (1)$$

1

set on a bounded regular domain $\mathcal{D}$ in $\mathbb{R}^d$ (for some $d \geq 1$), with a deterministic function $f \in H^{-1}(\mathcal{D})$ in the right-hand side. The field A is a fixed matrix-valued random field. It is assumed to be uniformly elliptic, uniformly bounded and stationary in a discrete sense. All this is made precise in Section 1.2. Since the parameter ε in (1) is assumed small, the coefficient $A\left(\frac{\cdot}{\varepsilon}, \omega\right)$ is oscillatory and (1) is challenging to solve numerically. On the other hand, the problem is theoretically well understood, as is recalled below.

In the numerical practice, the traditional approach to approximate the solution $u^\varepsilon(\cdot, \omega)$ to (1) is to consider (for any $p \in \mathbb{R}^d$), and solve, the so-called corrector problem

$$\begin{cases} -\operatorname{div}\left[A(p + \nabla w_p)\right] = 0 \quad \text{in } \mathbb{R}^d \text{ almost surely,} \\ \int_Q \mathbb{E}(\nabla w_p) = 0, \quad \nabla w_p \text{ is stationary in the sense of (5) below,} \end{cases} \tag{2}$$

associated to (1). The solution to (2) gives the deterministic and constant coefficient $A^\star$ of the homogenized equation that in turn serves for the approximation of (1). We refer to Section 1.2 below for details.

Since (2) is a problem set on the entire space $\mathbb{R}^d$, it is necessary to truncate it on a bounded domain, and to complement it with appropriate boundary conditions. In practice, it is standard to consider the problem

$$-\operatorname{div}\left(A(\cdot, \omega)\left(p + \nabla w_p^N(\cdot, \omega)\right)\right) = 0, \quad w_p^N(\cdot, \omega) \text{ is } Q_N\text{-periodic,} \tag{3}$$

where, say, $Q_N = (0, N)^d$. The deterministic homogenized matrix $A^\star$ is then approximated by the random variable $A_N^\star(\omega)$ defined by

$$\forall p \in \mathbb{R}^d, \quad A_N^\star(\omega)\, p = \frac{1}{|Q_N|} \int_{Q_N} A(\cdot, \omega)(p + \nabla w_p^N(\cdot, \omega)). \tag{4}$$

This approximate homogenized coefficient $A_N^\star(\omega)$ is then evaluated using the Monte-Carlo method. Random realizations of the environment, namely the matrix coefficient $A(y, \omega)$, are considered within the truncated domain Q_N. For each of these environments, (3) is solved and the matrix $A_N^\star(\omega)$ is computed using (4). The homogenized coefficient $A^\star$ is eventually approximated as an empirical mean over several realizations of $A_N^\star(\omega)$. More details are given below in Section 1.3.

The purpose of this article is to reduce the variance of the approximation of $A^\star$.

2

For this purpose, we borrow a variance reduction approach originally introduced in a completely different context, namely that of atomistic simulations for microscopic solid state science. In the series of articles [23, 24, 25], an approach is indeed described that selects some particular random realizations of the environment, based on some selection criteria derived from asymptotic properties. Intuitively, the approach aims at considering only realizations that, for N fixed, *already* satisfy properties that are usually only obtained in the asymptotic limit $N \to \infty$. The approach carries the name SQS, abbreviation of *Special Quasirandom Structures*. Its principles share some similarity with those underlying another classical variance reduction technique, namely *stratified sampling*.

We aim at adapting this approach to our context, at studying it theoretically in some simple situations, and testing it numerically in more general situations.

For the sake of completeness, we mention that we have already studied the theoretical properties and the practical performance of several variance reduction methods for numerical random homogenization in some previous works of ours. The classical approach of *antithetic variables*, an approach that is quite generic and does not require nor exploit knowledge of the specific structure of the random problem at hand, has been considered in [4, 5, 9, 18]. The significantly more elaborate (and thus more efficient) approach of *control variates* is the subject of [17]. That approach requires a better knowledge of the problem considered, and is not always amenable to fully generic situations.

Our article is articulated as follows.

In the remainder of this introductory section, we present the basics of the theoretical setting (in Section 1.2) and of the numerical approximation method (in Section 1.3) for the homogenization of the random equation (1).

In Section 2, we introduce the variance reduction approach we consider. For pedagogic purposes, we first briefly expose the approach in the context of solid state physics it has originally been introduced in. This is the purpose of Section 2.1. In Section 2.2, we formally derive the specifics of our variance reduction approach using a perturbative setting. This formal derivation provides the motivation for the general so-called SQS conditions that we use in the sequel of the work. Section 2.3 presents how we compute these conditions in practice. Section 2.4 contains the pseudo-code of our approach, along with some

3

comments.

The theoretical analysis of the approach is the purpose of Section 3. We begin by proving, in a fairly general situation (in any ambient dimension), that the approximation provided by our approach (at least the simplest variant of our approach) converges to the homogenized coefficient $A^\star$ when the truncated domain converges to the whole space (see Theorem 8 in Section 3.1). Next, in Section 3.2, we investigate more thoroughly the one-dimensional setting, where we can indeed completely analyze our approach and actually prove its efficiency.

Our final Section 4 contains numerical tests. First, since it is often necessary to enforce the desired conditions up to some tolerance (see Remark 3 below), we investigate in Section 4.1 how this tolerance affects the quality of the approximation and the efficiency of the approach. We observe there that the approach is robust in this respect.

In Section 4.2, we illustrate on a prototypical situation the efficiency of our approach and scrutinize its sensitivity and the various sources of error involved. The conclusions are the following. The systematic error is kept approximately constant by the approach (it might even be reduced), while the variance is reduced by several orders of magnitude. The more conditions we impose on the microstructures, the smaller the variance. The total error is always reduced. Such an efficiency is achieved at almost no additional cost with respect to the classical Monte Carlo algorithm.

In order to demonstrate the versatility of the approach, we apply it in Section 4.3 to a case with a way more general geometry of microstructures. There again, the approach provides a significant reduction of the variance.

We conclude this overview by emphasizing that, although the approach introduced in this article is applied to the simple linear elliptic equation (1), there is no reason to believe that it cannot be applied for a large class of partial differential equations with random coefficients. Indeed, the principles of the approach do not depend upon the specific form of the equation.

1.2 Theoretical setting

To begin with, we introduce the basic setting of stochastic homogenization. We refer to the seminal works [15, 21], to [11] for a general, numerically oriented presentation and to [2, 7, 14] for classical textbooks. We also refer to [16] and the review article [1] (and the extensive

bibliography therein) for a presentation of our particular setting.

Throughout this article, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and we denote by $\mathbb{E}(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega)$ the expectation of any random variable $X \in L^1(\Omega, d\mathbb{P})$. We next fix $d \in \mathbb{N}^\star$ (the ambient physical dimension), and assume that the group $(\mathbb{Z}^d, +)$ acts on Ω. We denote by $(\tau_k)_{k \in \mathbb{Z}^d}$ this action, and assume that it preserves the measure $\mathbb{P}$, that is, for all $k \in \mathbb{Z}^d$ and all $E \in \mathcal{F}$, $\mathbb{P}(\tau_k E) = \mathbb{P}(E)$. We assume that the action τ is *ergodic*, that is, if $E \in \mathcal{F}$ is such that $\tau_k E = E$ for any $k \in \mathbb{Z}^d$, then $\mathbb{P}(E) = 0$ or 1. In addition, we define the following notion of stationarity (see [16]): a function $F \in L^1_{\text{loc}}\left(\mathbb{R}^d, L^1(\Omega)\right)$ is *stationary* if

$$\forall k \in \mathbb{Z}^d, \quad F(x + k, \omega) = F(x, \tau_k \omega) \quad \text{a.e. in } x \text{ and a.s.} \qquad (5)$$

In this setting, the ergodic theorem [22] can be stated as follows:

Let $F \in L^\infty\left(\mathbb{R}^d, L^1(\Omega)\right)$ be a stationary random variable in the above sense. For $k = (k_1, k_2, \ldots, k_d) \in \mathbb{Z}^d$, we set $|k|_\infty = \max\limits_{1 \le i \le d} |k_i|$. Then

$$\frac{1}{(2N+1)^d} \sum_{|k|_\infty \le N} F(x, \tau_k \omega) \underset{N \to \infty}{\longrightarrow} \mathbb{E}\left(F(x, \cdot)\right) \quad \text{in } L^\infty(\mathbb{R}^d), \text{ almost surely.}$$

This implies (denoting by $Q = (0,1)^d$ the unit cube in $\mathbb{R}^d$) that

$$F\left(\frac{x}{\varepsilon}, \omega\right) \xrightarrow[\varepsilon \to 0]{\star} \mathbb{E}\left(\int_Q F(x, \cdot) dx\right) \quad \text{in } L^\infty(\mathbb{R}^d), \text{ almost surely.}$$

Besides technicalities, the purpose of the above setting is simply to formalize that, even though realizations may vary, the function F at point $x \in \mathbb{R}^d$ and the function F at point $x + k$, $k \in \mathbb{Z}^d$, share the same law. In the homogenization context, this means that the local, microscopic environment (encoded in the matrix field A in (1)) is everywhere the same *on average*. From this, homogenized, macroscopic properties follow.

We consider problem (1), which we recall here for convenience:

$$-\operatorname{div}\left(A\left(\frac{\cdot}{\varepsilon}, \omega\right) \nabla u^\varepsilon(\cdot, \omega)\right) = f \text{ in } \mathcal{D}, \qquad u^\varepsilon(\cdot, \omega) = 0 \text{ on } \partial\mathcal{D}.$$

The random matrix A is assumed stationary in the sense of (5). We also assume that A is bounded and coercive, that is, there exist two scalars $0 < c \le C < \infty$ such that, almost surely,

$$\|A(\cdot, \omega)\|_{L^\infty(\mathbb{R}^d)} \le C \quad \text{and} \quad \forall \xi \in \mathbb{R}^d, \quad \xi^T A(x, \omega) \xi \ge c\, \xi^T \xi \quad \text{a.e.}$$

5

In this specific setting, the solution $u^\varepsilon(\cdot, \omega)$ to (1) almost surely converges (when ε goes to 0) to the solution $u^\star$ to the homogenized problem

$$-\operatorname{div}(A^\star \nabla u^\star) = f \quad \text{in } \mathcal{D}, \qquad u^\star = 0 \quad \text{on } \partial\mathcal{D}. \tag{6}$$

The convergence of $u^\varepsilon(\cdot, \omega)$ to $u^\star$ holds weakly in $H^1(\mathcal{D})$ and strongly in $L^2(\mathcal{D})$.

The homogenized matrix $A^\star$ in (6) is deterministic, and given by an expectation of an integral involving the so-called corrector function, that solves a random auxiliary problem set on the *entire* space. It is given by

$$\forall p \in \mathbb{R}^d, \quad A^\star p = \mathbb{E}\left[\int_Q A(x, \cdot)\,(p + \nabla w_p(x, \cdot))\,dx\right], \tag{7}$$

where we recall that $Q = (0,1)^d$ and where, for any vector $p \in \mathbb{R}^d$, the *corrector* w_p is the unique solution (up to the addition of a random constant) in $L^2(\Omega; L^2_{\text{loc}}(\mathbb{R}^d))$ with gradient in $L^2(\Omega; L^2_{\text{unif}}(\mathbb{R}^d))^d$ of the corrector problem (2). We have used the notation $L^2_{\text{unif}}(\mathbb{R}^d)$ for the *uniform* L^2 space, that is the space of functions for which, say, the L^2 norm on a ball of unit size is bounded from above independently of the center of the ball.

1.3 Numerical approximation of the homogenized matrix

As briefly mentioned above, the corrector problem (2) is set on the *entire* space $\mathbb{R}^d$, and is therefore challenging to solve. Approximations are in order. In practice, the deterministic matrix $A^\star$ is approximated by the random matrix $A_N^\star(\omega)$ defined by (4), which is obtained by solving the corrector problem (3) on a *truncated* domain, say the cube $Q_N = (0, N)^d$. Although $A^\star$ itself is a deterministic object, its practical approximation $A_N^\star$ is random. It is only in the limit of infinitely large domains Q_N that the deterministic value is attained. As shown in [6], we indeed have

$$\lim_{N \to \infty} A_N^\star(\omega) = A^\star \quad \text{almost surely.} \tag{8}$$

As usual, the error $A^\star - A_N^\star(\omega)$ may be expanded as

$$A^\star - A_N^\star(\omega) = \left(A^\star - \mathbb{E}\left[A_N^\star\right]\right) + \left(\mathbb{E}\left[A_N^\star\right] - A_N^\star(\omega)\right), \tag{9}$$

that is the sum of a *systematic* error and of a *statistical* error (the first and second terms in the above right-hand side, respectively).

A standard technique to compute an approximation of $\mathbb{E}\left[A_N^\star\right]$ is to consider M independent and identically distributed realizations of the field A, solve for each of them the corrector problem (3) (thereby obtaining i.i.d. realizations $A_N^{\star,m}(\omega)$, for $1 \leq m \leq M$) and compute the Monte Carlo approximation

$$\mathbb{E}\left[(A_N^\star)_{ij}\right] \approx I_M^{\mathrm{MC}}(\omega) := \frac{1}{M} \sum_{m=1}^{M} \left(A_N^{\star,m}(\omega)\right)_{ij} \qquad (10)$$

for any $1 \leq i,j \leq d$. In view of the Central Limit Theorem, we know that $\mathbb{E}\left[(A_N^\star)_{ij}\right]$ asymptotically lies within the confidence interval

$$\left[I_M^{\mathrm{MC}} - 1.96 \frac{\sqrt{\mathbb{V}\mathrm{ar}\left[(A_N^\star)_{ij}\right]}}{\sqrt{M}}, I_M^{\mathrm{MC}} + 1.96 \frac{\sqrt{\mathbb{V}\mathrm{ar}\left[(A_N^\star)_{ij}\right]}}{\sqrt{M}} \right]$$

with a probability equal to 95 %.

For simplicity, and because this is overwhelmingly the case in the numerical practice, we have considered in (3) *periodic* boundary conditions. These are the conditions we adopt throughout our study. It is to be remarked, however, that other boundary conditions may be employed. Likewise, other slightly modified forms of equation (3) may be considered. The specific choice of approximation technique is motivated by considerations about the decrease of the systematic error in (9). Several recent mathematical studies have clarified this issue. In addition, in the particular case of periodic boundary conditions (3), it has been recently established in [12, Theorem 2] that the statistical error in (9) decays like $N^{-d/2}$ while the systematic error in (9) scales as $N^{-d}(\log N)^d$. Both estimates have been established for the *discrete variant* of the problem. A similar decay of the statistical error has also been established for the continuous case we consider in the present article (see [13, Theorem 1] and [20, Theorem 1.3 and Proposition 1.4]).

2 Variance reduction approach

2.1 Original formulation of the SQS approach

The variance reduction approach we elaborate upon in this article has been originally introduced for a slightly different purpose in atomistic solid-state science [23, 24, 25].

In order to convey to the reader the intuition of the original approach, we consider here a simple one-dimensional setting, which nevertheless illustrates the difficulties of a generic problem. We consider a linear chain of atomistic sites of two species A and B which interact by the interaction potentials V_{AA}, V_{AB} and V_{BB} with obvious notation. For simplicity we consider only nearest neighbour interaction. The atomic sites are occupied by a single species randomly chosen between A and B. A typical random configuration of the "material" therefore reads as an infinite sequence of the type $\cdots ABBAAABBAAAA\cdots$

In order to compute the energy per unit particle of that atomistic system, one has to consider all possible such infinite sequences, and for each of them its normalized energy

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} V_{X_{i+1}X_i}, \tag{11}$$

where X_i denotes the species present at the i-th site for that particular configuration ($X_i \equiv A$ or B). The "energy" of the system is then defined as the *expectation* of (11) over all possible configurations. Other quantities than (11) may be considered, or may be simultaneously considered.

In practice, one considers a presumably extremely large, finite N, truncates the infinite sequence over the finite length $2N+1$, and compute

$$\frac{1}{2N+1} \sum_{i=-N}^{N} V_{X_{i+1}X_i}$$

for many (say M, where M is also presumably large) configurations.

The approach introduced in [23, 24, 25] consists in *selecting* specific configurations $(X_i)_{-N\leq i\leq N}$ of atomic sites that satisfy statistical properties usually obtained only in the limit of infinitely large N.

The first such statistical property is the volume fraction, namely the proportion of species (A, B) present on average. If the sites are all occupied randomly with probability $1/2$ of A and $1/2$ of B (and

8

assuming that all these random variables are independent), then obviously the volume fraction of A is $1/2$ and so is that of B. Then, one only consider truncated sequences $(X_i)_{-N \leq i \leq N}$ that *exactly* reproduce that volume fraction.

Similarly, again for such an evenly distributed proportion of A and B, the energy of the entire infinite system evidently reads as

$$\mathcal{E} = \frac{1}{4}\left[V_{AA} + 2V_{AB} + V_{BB}\right]$$

(recall that we only consider nearest-neighbour interactions). Thus, one only considers truncated sequences $(X_i)_{-N \leq i \leq N}$ which, in addition to exhibiting the exact volume fraction, have an average energy $\dfrac{1}{2N+1}\displaystyle\sum_{i=-N}^{N} V_{X_{i+1}X_i}$ which is *equal* to $\mathcal{E}$. And so on and so forth for other quantities of interest.

Mathematically, this *selection* of suitable configurations among all the possible configurations classically considered in a Monte-Carlo sample amounts to replacing the computation of an expectation by that of a *conditional expectation*.

The simplistic model we have just considered for pedagogic purposes can of course be replaced by more elaborate models, with more sophisticated quantities to compute, and more demanding statistical quantities to condition the computations with. The bottom line of the approach remains the same, and we adapt it to design a variance reduction approach for numerical random homogenization.

In the next section, we derive the appropriate conditions, which we call the SQS conditions, for our specific context.

2.2 Formal derivation of the SQS conditions using a perturbative setting

The purpose of this Section is to formally derive the SQS conditions that we use in the sequel. Such conditions can be easily intuitively understood. We however believe it is interesting to (formally) *derive* them in a particular case. The case we proceed with is a perturbative setting (although, we emphasize it, the conditions will be employed in the full general, not necessarily perturbative, setting).

We assume throughout this section that the matrix valued coeffi-

cient A in (1) reads as

$$A_\eta(x, \omega) = C_0(x, \omega) + \eta \ \chi(x, \omega) C_1(x, \omega) \tag{12}$$

for some presumably small scalar coefficient η, where

- C_0 and C_1 are two stationary, uniformly bounded matrix fields,
- $C_0(\cdot, \omega) - C_1(\cdot, \omega)$ and $C_0(\cdot, \omega) + C_1(\cdot, \omega)$ are almost surely coercive,
- χ is a stationary scalar field with values in $[-1, 1]$.

Under these assumptions, for any $\eta \in (-1, 1)$, the matrix A_η is stationary, bounded and coercive. Intuitively, when η is small, A_η is a perturbation of the matrix-valued field $C_0(x, \omega)$.

Remark 1. *The expression* (12) *models e.g. a two-phase composite material, where the phases are modelled by the coefficients C_0 and C_1, while χ is the indicator function of the first phase.*

Let $p \in \mathbb{R}^d$. The corrector problem (2) reads, in this particular setting, as

$$\begin{cases} -\operatorname{div}\left[(C_0 + \eta \chi C_1)(p + \nabla w_\eta)\right] = 0 & \text{in } \mathbb{R}^d, \\ \mathbb{E} \int_Q \nabla w_\eta = 0, \quad \nabla w_\eta \text{ is stationary in the sense of (5),} \end{cases} \tag{13}$$

and the homogenized matrix (7) is given by

$$\forall p \in \mathbb{R}^d, \quad A_\eta^\star p = \mathbb{E} \int_Q A_\eta(p + \nabla w_\eta). \tag{14}$$

Note that, for the sake of clarity, we omit to write the dependency of w_η with respect to p.

The truncated version of (13) on the domain Q_N is

$$\begin{cases} -\operatorname{div}\left[(C_0 + \eta \chi C_1)(p + \nabla w_\eta^N)\right] = 0 & \text{in } Q_N, \\ w_\eta^N(\cdot, \omega) \text{ is } Q_N\text{-periodic,} \end{cases} \tag{15}$$

and we approach the homogenized matrix (14) by

$$\forall p \in \mathbb{R}^d, \quad A_\eta^{\star,N}(\omega)\, p = \frac{1}{|Q_N|} \int_{Q_N} A_\eta(\cdot, \omega)\left(p + \nabla w_\eta^N(\cdot, \omega)\right). \tag{16}$$

2.2.1 Expansion in powers of η

As η goes to 0, we may now expand $A_\eta^{\star,N}(\omega)$ and $A_\eta^\star$ in powers of η. This expansion is classical (see for instance [5, 8]). We only provide it here for the sake of consistency. The corrector expands as

$$\nabla w_\eta = \nabla w_0 + \eta \nabla u_1 + \eta^2 \nabla u_2 + o(\eta^2). \tag{17}$$

This expansion holds in $L^2(\Omega; L^2_{\mathrm{unif}}(\mathbb{R}^d))$. The functions w_0, u_1 and u_2 appearing in the expansion are respectively defined by the following systems of equations:

$$\begin{cases} -\operatorname{div}\left[C_0(p + \nabla w_0)\right] = 0 \quad \text{in } \mathbb{R}^d, \\ \mathbb{E}\int_Q \nabla w_0 = 0, \quad \nabla w_0 \text{ is stationary,} \end{cases} \tag{18}$$

$$\begin{cases} -\operatorname{div}\left[C_0\nabla u_1\right] = \operatorname{div}\left[\chi C_1(p + \nabla w_0)\right] \quad \text{in } \mathbb{R}^d, \\ \mathbb{E}\int_Q \nabla u_1 = 0, \quad \nabla u_1 \text{ is stationary,} \end{cases} \tag{19}$$

and

$$\begin{cases} -\operatorname{div}\left[C_0\nabla u_2\right] = \operatorname{div}\left[\chi C_1\nabla u_1\right] \quad \text{in } \mathbb{R}^d, \\ \mathbb{E}\int_Q \nabla u_2 = 0, \quad \nabla u_2 \text{ is stationary.} \end{cases}$$

Inserting the expansion (12) of A_η and (17) of w_η in (14), we obtain

$$A_\eta^\star = A_0^\star + \eta A_1^\star + \eta^2 A_2^\star + o(\eta^2), \tag{20}$$

with, for any $p \in \mathbb{R}^d$,

$$\begin{aligned} A_0^\star p &= \mathbb{E}\left[\int_Q C_0(p + \nabla w_0)\right], \\ A_1^\star p &= \mathbb{E}\left[\int_Q \chi C_1(p + \nabla w_0)\right] + \mathbb{E}\left[\int_Q C_0\nabla u_1\right], \\ A_2^\star p &= \mathbb{E}\left[\int_Q \chi C_1\nabla u_1\right] + \mathbb{E}\left[\int_Q C_0\nabla u_2\right]. \end{aligned} \tag{21}$$

Likewise, we expand w_η^N as

$$\nabla w_\eta^N = \nabla w_0^N + \eta \nabla u_1^N + \eta^2 \nabla u_2^N + o(\eta^2),$$

with

$$\begin{cases} -\operatorname{div}\left[C_0(p + \nabla w_0^N)\right] = 0 \quad \text{in } Q_N, \\ w_0^N(\cdot, \omega) \text{ is } Q_N\text{-periodic,} \end{cases} \tag{22}$$

11

$$\begin{cases} -\operatorname{div}\left[C_0 \nabla u_1^N\right] = \operatorname{div}\left[\chi C_1(p + \nabla w_0^N)\right] & \text{in } Q_N, \\ u_1^N(\cdot, \omega) \text{ is } Q_N\text{-periodic}, \end{cases} \quad (23)$$

and

$$\begin{cases} -\operatorname{div}\left[C_0 \nabla u_2^N\right] = \operatorname{div}\left[\chi C_1 \nabla u_1^N\right] & \text{in } Q_N, \\ u_2^N(\cdot, \omega) \text{ is } Q_N\text{-periodic}. \end{cases}$$

The homogenized matrix $A_\eta^{\star,N}(\omega)$ therefore satisfies

$$\left| A_\eta^{\star,N}(\omega) - \left[A_0^{\star,N}(\omega) + \eta A_1^{\star,N}(\omega) + \eta^2 A_2^{\star,N}(\omega) \right] \right| \le C\eta^3, \quad (24)$$

where C is independent of η, N and ω, and where the matrices $A_0^{\star,N}(\omega)$, $A_1^{\star,N}(\omega)$ and $A_2^{\star,N}(\omega)$ are defined by

$$A_0^{\star,N}(\omega)\, p \;=\; \frac{1}{|Q_N|} \int_{Q_N} C_0(p + \nabla w_0^N),$$

$$A_1^{\star,N}(\omega)\, p \;=\; \frac{1}{|Q_N|} \int_{Q_N} \chi C_1(p + \nabla w_0^N) + \frac{1}{|Q_N|} \int_{Q_N} C_0 \nabla u_1^N, (25)$$

$$A_2^{\star,N}(\omega)\, p \;=\; \frac{1}{|Q_N|} \int_{Q_N} \chi C_1 \nabla u_1^N + \frac{1}{|Q_N|} \int_{Q_N} C_0 \nabla u_2^N.$$

2.2.2 SQS conditions

In line with the motivation we have mentioned above in Section 1.3, we are now in position to introduce the conditions that we use to *select* particular configurations of the environment within Q_N for which we compute the solution to (15), and, in turn, compute the approximation (16) of $A_\eta^\star$. Our conditions are based upon the comparison of (21) and (25).

Definition 2. *For finite fixed N, we say that an environment $\omega \in \Omega$ satisfies the SQS condition of*

- *order 0 if $A_0^{\star,N}(\omega) = A_0^\star$, that is to say, for any $p \in \mathbb{R}^d$,*

$$\frac{1}{|Q_N|} \int_{Q_N} C_0(\cdot, \omega)(p + \nabla w_0^N(\cdot, \omega)) = \mathbb{E}\left[\int_Q C_0(p + \nabla w_0) \right], \quad (26)$$

- *order 1 if $A_1^{\star,N}(\omega) = A_1^\star$, that is to say, for any $p \in \mathbb{R}^d$,*

$$\frac{1}{|Q_N|} \int_{Q_N} \left[\chi(\cdot, \omega) C_1(\cdot, \omega)(p + \nabla w_0^N(\cdot, \omega)) + C_0(\cdot, \omega) \nabla u_1^N(\cdot, \omega) \right]$$

$$= \mathbb{E}\left[\int_Q \chi C_1(p + \nabla w_0) + C_0 \nabla u_1 \right], \quad (27)$$

12

- *order 2 if $A_2^{\star,N}(\omega) = A_2^\star$, that is to say, for any $p \in \mathbb{R}^d$,*

$$\frac{1}{|Q_N|} \int_{Q_N} \left[\chi(\cdot,\omega)C_1(\cdot,\omega)\nabla u_1^N(\cdot,\omega) + C_0(\cdot,\omega)\nabla u_2^N(\cdot,\omega) \right]$$
$$= \mathbb{E}\left[\int_Q \chi C_1 \nabla u_1 + C_0 \nabla u_2 \right]. \quad (28)$$

Remark 3. *In full generality, we* do not *claim that there exist environments that satisfy these conditions. This might be the case that no such environment exists. One may for instance simply remark that a random variable that takes value -1 and $+1$ both with probability $1/2$ never has value zero, which is its expectation! In some situations, we therefore have to* relax *the above conditions (see Section 2.4 below), but we temporarily leave these technicalities aside and* assume *that suitable environments exist.*

Consider now the two expansions (20) and (24). It is immediate to see, by subtraction, that

$$A_\eta^{\star,N}(\omega) - A_\eta^\star = (A_0^{\star,N}(\omega) - A_0^\star) + \eta(A_1^{\star,N}(\omega) - A_1^\star) + \eta^2(A_2^{\star,N}(\omega) - A_2^\star) + o(\eta^2).$$

Therefore it is readily seen that, if the configuration ω satisfies the SQS conditions of Definition 2 up to the order k included ($k = 0$, 1, 2 in our definition, but clearly one could consider higher order conditions derived likewise), then

$$A_\eta^{\star,N}(\omega) - A_\eta^\star = o(\eta^k), \quad (29)$$

where the constant in the right-hand side is independent of η, N and ω. Taking the expectation over such configurations therefore formally provides a more accurate approximation of $A_\eta^\star$.

Now that we have derived the conditions (26)–(27)–(28) (which we henceforth call the *SQS conditions*) in the perturbative setting, we will actually use them in the non perturbative setting, namely for a similar two-phase composite material, but with η *not* small. Of course, a property such as (29) cannot be expected any longer since the homogenized matrix $A^\star$ is no longer a series in a small coefficient that encodes a perturbation. Nevertheless, it can be expected that selecting the configurations using these conditions may improve the approximation, in particular by reducing the variance. We show in Sections 3 and 4 that it is indeed the case, theoretically and experimentally.

For the time being, we need to make a *practical* observation. The right-hand side of conditions (26)–(27)–(28) need to be evaluated in order to practically encode the SQS conditions. In principle, those right-hand sides are exact expectations, that can only be determined using an asymptotic limit, and are therefore as challenging to compute in practice as $A^\star$ itself.

We therefore need to restrict the generality of our setting (12) and consider cases where those right-hand sides are indeed amenable to a simple, inexpensive computation. This is the purpose of the next section.

2.3 Practical evaluation of the SQS conditions

In order to make our approach practical, we need, as mentioned above, to consider settings where the expectations present in the right-hand sides of (26)–(27)–(28) may be computed effectively.

2.3.1 Condition of order 0

We first consider (26) and its right-hand side

$$\mathbb{E}\left[\int_Q C_0(x,\cdot)(p + \nabla w_0(x,\cdot))dx\right]. \qquad (30)$$

A natural assumption, which already covers a large portion of practically relevant situations, is

$$C_0(x,\omega) = C_0(x) \quad \text{is a deterministic, } \mathbb{Z}^d\text{-periodic matrix.} \qquad (31)$$

The computation of (30) is then inexpensive since the solution w_0 to (18) is in fact the deterministic solution to

$$-\operatorname{div}\left[C_0(p + \nabla w_0)\right] = 0 \ \text{ in } \mathbb{R}^d, \quad w_0 \text{ is } \mathbb{Z}^d\text{-periodic,}$$

which is unique up to the addition of a constant.

In addition, when N is an integer (and when the approximation chosen for (2) is the *periodic* approximation (3), as is indeed the case throughout this work), the solution to (22) is $w_0^N \equiv w_0$ (up to an additive constant), and hence the condition (26) is systematically satisfied.

We henceforth assume that (31) holds, that N is an integer, and that we proceed with the periodic approximation (3).

2.3.2 Condition of order 1

We next consider the SQS condition (27). One possible assumption to make that condition practical is

$$C_0(x,\omega) = C_0 \quad \text{is a deterministic, } constant \text{ matrix.} \qquad (32)$$

Since $\nabla w_0 = 0$, the right-hand side of (27) reads

$$\mathbb{E}\left[\int_Q \chi C_1(p + \nabla w_0) + C_0 \nabla u_1\right] = \int_Q \mathbb{E}\left[\chi C_1\right] p + C_0 \mathbb{E}\int_Q \nabla u_1,$$

where the rightmost term vanishes in view of (19) and where the first term of the right-hand side may be computed using only characteristic properties of the environment considered. The condition (27) thus reads

$$\frac{1}{|Q_N|}\int_{Q_N} \chi(\cdot,\omega)C_1(\cdot,\omega) = \mathbb{E}\left[\int_Q \chi C_1\right]. \qquad (33)$$

For instance, in a two-phase composite material mixing two *constant* and *deterministic* matrices C_0 and C_1, we have

$$\mathbb{E}\left[\int_Q \chi C_1\right] = \mathbb{E}\left[\int_Q \chi\right] C_1.$$

This quantity obviously only depends on the *volume fraction* of the two phases (recall (12)). Proceeding likewise with the left-hand side of the condition (27), we see that this condition reads

$$\frac{1}{|Q_N|}\int_{Q_N} \chi(x,\omega)dx = \mathbb{E}\left[\int_Q \chi\right].$$

Interestingly (and not unexpectedly), we notice here that this condition on the volume fraction agrees with the condition we used to consider in the simple atomistic system of Section 2.1.

2.3.3 Condition of order 2

We next proceed with condition (28). In addition to (32), we assume that

$$C_1(x,\omega) = C_1(x) \quad \text{is a deterministic, } \mathbb{Z}^d\text{-periodic matrix,} \qquad (34)$$

and that

$$\chi(y,\omega) = \sum_{k\in\mathbb{Z}^d} X_k(\omega)\mathbb{1}_{Q+k}(y), \qquad (35)$$

15

where X_k are identically distributed scalar random variables taking their values in $[-1, 1]$. We also assume that

$$\mathcal{C} = \sum_{k \in \mathbb{Z}^d} |\mathbb{C}\text{ov}(X_0, X_k)| < \infty, \tag{36}$$

which is obviously satisfied if X_k are independent one from each other.

We then have the following result, which will be useful to make condition (28) practical. Its proof is postponed until Appendix A.

Lemma 4. *Under the assumptions* (32), (34), (35) *and* (36), *the solution* u_1 *to* (19) *satisfies*

$$\nabla u_1(y, \omega) = \mathbb{E}[X_0] \nabla \overline{u_1}(y) + \sum_{k \in \mathbb{Z}^d} \left(X_k(\omega) - \mathbb{E}[X_k] \right) \nabla \phi_1(y - k), \tag{37}$$

where ϕ_1 *is the (unique up to the addition of a constant) solution in* $\{v \in L^2_{\text{loc}}(\mathbb{R}^d), \quad \nabla v \in (L^2(\mathbb{R}^d))^d\}$ *to*

$$- \text{div}\left[C_0 \nabla \phi_1 \right] = \text{div}\left[\mathbb{1}_Q C_1 p \right] \quad \text{in } \mathbb{R}^d \tag{38}$$

and $\overline{u_1}$ *is the (unique up to the addition of a constant) solution to*

$$- \text{div}\left[C_0 \nabla \overline{u_1} \right] = \text{div}\left[C_1 p \right] \quad \text{in } \mathbb{R}^d, \quad \overline{u_1} \text{ is } \mathbb{Z}^d\text{-periodic.} \tag{39}$$

The sum in (37) *is a convergent series in* $L^2(Q \times \Omega)$.

Using simpler arguments, we see that the solution u_1^N to (23) satisfies

$$\nabla u_1^N(y, \omega) = \mathbb{E}[X_0] \nabla \overline{u_1}(y) + \sum_{k \in \mathbb{Z}^d \cap Q_N} \left(X_k(\omega) - \mathbb{E}[X_k] \right) \nabla \phi_1^N(y - k),$$
$$\tag{40}$$

where $\overline{u_1}$ is defined by (39) and ϕ_1^N is the (unique up to the addition of a constant) solution to

$$- \text{div}\left[C_0 \nabla \phi_1^N \right] = \text{div}\left[\mathbb{1}_Q C_1 p \right] \quad \text{in } Q_N, \quad \phi_1^N \text{ is } Q_N\text{-periodic.} \tag{41}$$

In practice, we can easily obtain an accurate approximation of ϕ_1 since the right-hand side of (38) has compact support. Truncating (38) over a sufficiently large bounded domain (with homogeneous Dirichlet boundary conditions) provides such an accurate approximation. Given (32), the right-hand side of Condition (28) rewrites

16

$$\mathbb{E}\left[\int_Q \chi C_1 \nabla u_1\right] \text{ since } \mathbb{E}\left[\int_Q \nabla u_2\right] = 0. \text{ In view of (37), this quantity}$$
is in turn expanded as

$$\mathbb{E}\left[\int_Q \chi C_1 \nabla u_1\right]$$

$$= (\mathbb{E}[X_0])^2 \int_Q C_1 \nabla \overline{u_1} + \sum_{k \in \mathbb{Z}^d} \mathbb{E}\left[\int_Q X_0\Big(X_k - \mathbb{E}[X_k]\Big)C_1 \nabla \phi_1(\cdot - k)\right]$$

$$= (\mathbb{E}[X_0])^2 \int_Q C_1 \nabla \overline{u_1}$$

$$+ \sum_{k \in \mathbb{Z}^d} \mathbb{E}\left[\int_Q \Big(X_0 - \mathbb{E}[X_0]\Big)\Big(X_k - \mathbb{E}[X_k]\Big)C_1 \nabla \phi_1(\cdot - k)\right], \quad (42)$$

where, as mentioned above, $\nabla \phi_1$ can be easily and accurately computed, while the series in $k \in \mathbb{Z}^d$ may be truncated in an efficient manner because of the rapid decay at infinity of $\nabla \phi_1$ (see [5, Lemma 3.1]).

We correspondingly expand the left-hand side of (28). The second term vanishes, while the first term reads, in view of (40),

$$\frac{1}{|Q_N|}\int_{Q_N} \chi(y,\omega)C_1(y)\nabla u_1^N(y,\omega)\,dy$$

$$= \sum_{j \in \mathbb{Z}^d \cap Q_N} \frac{1}{|Q_N|}\int_{Q_N} X_j(\omega)\mathbb{1}_{Q+j}C_1\mathbb{E}[X_0]\nabla\overline{u_1}$$

$$+ \sum_{k,j \in \mathbb{Z}^d \cap Q_N} \frac{1}{|Q_N|}\int_{Q_N} X_j(\omega)\mathbb{1}_{Q+j}C_1\Big(X_k(\omega) - \mathbb{E}[X_k]\Big)\nabla\phi_1^N(\cdot - k)$$

$$= (\mathbb{E}[X_0])^2 \int_Q C_1 \nabla\overline{u_1}$$

$$+ \mathbb{E}[X_0]\left(\frac{1}{|Q_N|}\sum_{j \in \mathbb{Z}^d \cap Q_N}\Big(X_j(\omega) - \mathbb{E}[X_j]\Big)\right)\int_Q C_1\nabla\overline{u_1}$$

$$+ \mathbb{E}[X_0]\sum_{k \in \mathbb{Z}^d \cap Q_N}\frac{1}{|Q_N|}\int_{Q_N}C_1\Big(X_k(\omega) - \mathbb{E}[X_k]\Big)\nabla\phi_1^N(\cdot - k)$$

$$+ \sum_{k,j \in \mathbb{Z}^d \cap Q_N}\frac{1}{|Q_N|}\int_{Q+j}\Big(X_j(\omega) - \mathbb{E}[X_j]\Big)C_1\Big(X_k(\omega) - \mathbb{E}[X_k]\Big)\nabla\phi_1^N(\cdot - k).$$

$$(43)$$

In this particular (however still very generic) setting, we infer from (42) and (43) that Condition (28) reads as

$$\frac{1}{|Q_N|} \sum_{k,j \in Q_N \cap \mathbb{Z}^d} \left(X_k(\omega) - \mathbb{E}[X_k] \right) \left(X_j(\omega) - \mathbb{E}[X_j] \right) I_{k,j}^N$$

$$+ \frac{1}{|Q_N|} \mathbb{E}[X_0] \sum_{k \in Q_N \cap \mathbb{Z}^d} \left(X_k(\omega) - \mathbb{E}[X_k] \right) I_k^N = \sum_{k \in \mathbb{Z}^d} \mathbb{Cov}(X_0, X_k) I_k^\infty,$$

$$(44)$$

where

$$I_k^\infty = \int_{Q+k} C_1(y) \nabla \phi_1(y), \tag{45}$$

$$I_{k,j}^N = \int_{Q+j} C_1(y) \nabla \phi_1^N(y-k) dy, \tag{46}$$

$$I_k^N = \int_{Q_N} C_1(y) \nabla \phi_1^N(y-k) dy + \int_Q C_1(y) \nabla \overline{u_1}(y) dy. \tag{47}$$

2.3.4   Summary

In the prototypical case where

$$A(x,\omega) = C_0 + \chi(x,\omega) C_1(x),$$

where C_0 is constant, C_1 is $\mathbb{Z}^d$ periodic and χ takes the form (35) (and where we consider the periodic approximation (3) of (2)), we have that:

- The condition (26) (SQS condition of order 0) is systematically fulfilled.

- In view of (33) and (35), the condition (27) (SQS condition of order 1) rewrites as

$$\frac{1}{|Q_N|} \sum_{k \in \mathbb{Z}^d \cap Q_N} X_k(\omega) = \mathbb{E}[X_0]. \tag{48}$$

- In view of (44), the condition (28) (SQS condition of order 2) writes as

$$\frac{1}{|Q_N|} \sum_{k,j \in Q_N \cap \mathbb{Z}^d} \overline{X}_k(\omega) \overline{X}_j(\omega) I_{k,j}^N$$

$$+ \frac{1}{|Q_N|} \mathbb{E}[X_0] \sum_{k \in Q_N \cap \mathbb{Z}^d} \overline{X}_k(\omega) I_k^N = \sum_{k \in \mathbb{Z}^d} \mathbb{Cov}(X_0, X_k) I_k^\infty, \quad (49)$$

18

where $\overline{X}_k(\omega) = X_k(\omega) - \mathbb{E}[X_k]$.

The conditions (48) and (49) are henceforth called the SQS 1 and SQS 2 conditions, respectively.

Remark 5. *If (48) is satisfied, then the coefficient I_k^N in (49) can be replaced by*

$$\overline{I}_k^N = \int_{Q_N} C_1(y) \nabla \phi_1^N (y - k) dy$$

and there is no need to compute $\overline{u_1}$.

2.4   Selection Monte Carlo sampling

We are now in position to describe the selection Monte Carlo sampling we employ. We recall that the classical Monte Carlo sampling reads as follows:

Algorithm 1 (Classical Monte Carlo).
For $m = 1, \ldots, M$,

1. *Generate a random environment ω_m.*

2. *Solve the truncated corrector problem (3).*

3. *Compute $A_N^\star(\omega_m)$.*

Compute the approximation $\mathcal{I}_{MC}^M = \dfrac{1}{M} \displaystyle\sum_{m=1}^M A_N^\star(\omega_m)$ of $A^\star$.

In contrast, our selection Monte Carlo sampling algorithm, in the particular case described in Section 2.3.4, reads as follows:

Algorithm 2.
The algorithm requires a tolerance $\mathrm{tol} > 0$, fixed by the user.

1. **Offline stage**

   (a) *Solve the equation (38).*

   (b) *Compute $(I_k^\infty)_{k \in \mathbb{Z}^d}$ defined by (45).*

   (c) *Compute the right-hand side of the SQS conditions (48) and (49).*

   (d) *Solve the equations (39) and (41).*

   (e) *Compute $(I_{k,j}^N)_{k,j \in \mathbb{Z}^d \cap Q_N}$ and $(I_k^N)_{k \in \mathbb{Z}^d \cap Q_N}$ defined by (46) and (47).*

2. **Online stage**
   For $m = 1, \ldots, M$,

(a) *Generate a random environment ω_m.*

(b) *Using $I_{k,j}^N$ and I_k^N, compute the left-hand sides of (48) and (49).*

(c) *If the left-hand sides differ from the right-hand sides by more than* tol, *return to Step 2a.*

(d) *Solve the truncated corrector problem (3).*

(e) *Compute $A_N^\star(\omega_m)$.*

Compute the approximation $\mathcal{I}_{SQS}^M = \dfrac{1}{M} \displaystyle\sum_{m=1}^{M} A_N^\star(\omega_m)$ of $A^\star$.

Remark 6. *As pointed out above, the series in $k \in \mathbb{Z}^d$ in the right-hand side of (49) may be truncated in an efficient manner because of the rapid decay at infinity of $\nabla \phi_1$. Therefore only a few factors I_k^∞ have to be computed at Step 1b.*

Remark 7. *When several SQS conditions (in practice SQS 1 and SQS 2) have to be simultaneously satisfied, we simply add them up using some weighting parameter. We have not observed any particular sensitivity of our numerical results (collected in Section 4 below) with respect to the adjustment of this parameter, provided it remains not too close to 0 and 1.*

We have already mentioned that, in many situations, there might not be *any* random environments that satisfy some, or all, of the SQS conditions (26)–(27)–(28) we wish to enforce. Therefore, some adaptation is in order, and we have used in Algorithm 2 a tolerance parameter tol > 0 for the SQS conditions to be satisfied.

However, if these conditions are enforced within some given tolerance as in Algorithm 2, the following issue arises. Since the motivation for precisely considering the SQS conditions is that they are fulfilled *asymptotically*, the larger the truncated computational domain we consider (that is, the larger N), the less restrictive the conditions are, and therefore the less effective the variance reduction is likely to be. To circumvent this difficulty, a first possibility is to consider a tolerance that decreases when the size of Q_N increases. We consider this variant in our theoretical study of Section 3.2.1 below (see formula (69)). More precisely, we require in Proposition 14 that

$$\text{the SQS condition is satisfied with the tolerance } \frac{\lambda}{\sqrt{|Q_N|}}$$

for some λ. In practice, implementing such a threshold is not an easy matter, as the rate and the constants need to be adequately adjusted. In order to avoid such technicalities, we prefer to take a slightly different perspective, the purpose of which is to always select a *fixed proportion* of the original sample of the $\mathcal{M}$ environments drawn. Practically, we pick the M configurations that best satisfy the SQS conditions among the $\mathcal{M}$ configurations that have been drawn.

The practical algorithm we employ is therefore as follows:

Algorithm 3 (Selection Monte Carlo sampling).
The algorithm requires a number of trials $\mathcal{M}$, fixed by the user.

1. **Offline stage 1***: same as the offline stage of Algorithm 2.*

2. **Offline stage 2: selection step**
   For $m = 1, \ldots, \mathcal{M}$,

   (a) *Generate a random environment ω_m.*

   (b) *Using $I_{k,j}^N$ and I_k^N, compute the left-hand sides of* (48) *and* (49).

   (c) *Compute the error error_m between the left-hand sides and the right-hand sides of* (48) *and* (49).

   Sort the random environments $(\omega_m)_{1 \leq m \leq \mathcal{M}}$ according to error_m. Keep the M best realizations, and reject the others.

3. **Online stage: resolution**
   For $m = 1, \ldots, M$,

   (a) *Solve the truncated corrector problem* (3).

   (b) *Compute $A_N^\star(\omega_m)$.*

Compute the approximation $\mathcal{I}_{SQS}^M = \dfrac{1}{M} \displaystyle\sum_{m=1}^M A_N^\star(\omega_m)$ of $A^\star$.

We wish to make a couple of comments about this selection Monte Carlo approach.

In full generality, the cost of Monte Carlo approaches is usually dominated by the cost of draws, and therefore selection algorithms are targeted to reject as few draws as possible.

In the present context, where boundary value problems such as (3) are to be solved repeatedly, the cost of draws for the environment is negligible in front of the cost of the solution procedure for such boundary value problems. Likewise, evaluating the quantities present in e.g. (49) is not expensive. Therefore, the purpose of the selection mechanism is to limit the number of boundary value problems to be

solved, even though this comes at the (tiny) price of rejecting many environments. This also explains why we employ a simplistic rejection procedure for the selection, while in other situations of Monte Carlo samplings, one would invest in a more clever selection procedure.

A second observation is that, as potentially for any selection procedure, our selection introduces a bias (i.e. a modification of the systematic error in (9)). The point is to ensure that the gain in variance superseeds the bias introduced by the variance reduction approach.

Our next section addresses some theoretical aspects of our approach.

3 Elements of theoretical analysis

This section contains some elements of analysis that we are able to provide. We begin with a (somewhat) general result of convergence, and next, in some simplified cases, study our approach more thoroughly.

3.1 Proof of convergence of the approach

Formally, our approach consists in replacing an empirical average provided by the classical Monte Carlo approach to compute $\mathbb{E}[A_N^\star]$ by an empirical average *restricted* to some environments within Q_N satisfying some additional condition(s) (see Section 2.4). We work at a fixed size N of the truncation domain Q_N and recall that $A_N^\star(\omega)$ is defined by (4). Mathematically, our approach amounts to considering conditional expectations of the type $\mathbb{E}[A_N^\star \mid \text{SQS}]$, where SQS encodes that one, or several, of the conditions summarized in (48)–(49) are satisfied.

The least we can expect from our approach is that it converges to the correct limit when $N \to \infty$, namely $A^\star$, as in (8).

The theorem we now state establishes this fact. In order to prove it, we need to make some assumptions on our setting (see the details below), and also to make specific the SQS conditions we use. In Theorem 8 below, we specifically use the SQS 1 condition, in the form (48).

In order to state a result as general as possible, we therefore consider a condition that reads $\dfrac{1}{|Q_N|} \displaystyle\sum_{k \in \mathbb{Z}^d \cap Q_N} f(X_k) = \mathbb{E}[f(X_0)]$ for some function f. In practice, our specific SQS 1 condition (48) corresponds to the choice $f(x) = x$.

Theorem 8. *Let $(X_k)_{k \in \mathbb{Z}^d}$ be a sequence of independent and identically distributed scalar random variables following a common law μ. We assume that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, and that, for any $k \in \mathbb{Z}^d$, $X_k(\omega) \in [-1, 1]$ almost surely. We consider the stationary random field*

$$A(y, \omega) = C_0 + \sum_{k \in \mathbb{Z}^d} X_k(\omega) \, \mathbb{1}_{Q+k}(y) \, C_1(y),$$

where C_0 is constant and C_1 is $\mathbb{Z}^d$-periodic and bounded. We also assume that $C_0 + C_1(y)$ and $C_0 - C_1(y)$ are uniformly coercive, and that C_0 and C_1 are symmetric.

Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a measurable function with compact level sets. We assume that f is not constant. Then we have

$$\mathbb{E}\left[A_N^\star \; \Big| \; \frac{1}{|Q_N|} \sum_{k \in Q_N \cap \mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right] \xrightarrow[N \to \infty]{} A^\star, \qquad (50)$$

where $A_N^\star(\omega)$ is defined by (4) and $A^\star$ is defined by (7).

Some remarks are in order.

Remark 9. *As is the case throughout this article, we have considered the* periodic *approximation (3) of (2). The proof of Theorem 8 actually carries over to the case of Neumann or Dirichlet boundary conditions, or any alternate truncation problem that provides some $A^{\star,N}(\omega)$ such that $A_{\mathrm{Neu}}^{\star,N}(\omega) \leq A^{\star,N}(\omega) \leq A_{\mathrm{Dir}}^{\star,N}(\omega)$ (see additional details in [19, Appendix]).*

Remark 10. *The assumptions regarding independence of the X_k, absolute continuity of their common law with respect to the Lebesgue measure and compactness of the level sets of f are necessary for technical reasons, since we need to apply a general result from [3]. See below for details.*

The proof of Theorem 8 is based on the following result, which is a particular case of a more general result due to C. Bernardin and S. Olla (see [3, Theorem B.2.2]):

Theorem 11 (C. Bernardin and S. Olla, [3]). *Consider n scalar random variables X_1, ..., X_n, that are independent and that all share the same probability distribution $\mu(x) \, dx$ on $\mathbb{R}$. Consider a measurable*

function $f : \mathbb{R} \mapsto \mathbb{R}$, which is assumed to be not constant and to have compact level sets. Let $f_0 = \mathbb{E}[f(X_1)] = \int_{\mathbb{R}} f(x)\,\mu(x)\,dx$. Consider also a bounded and continuous function $F : \mathbb{R}^k \mapsto \mathbb{R}$. Then

$$\lim_{n \to \infty} \mathbb{E}\left[F(X_1, \ldots, X_k) \,\Big|\, \frac{1}{n} \sum_{i=1}^{n} f(X_i) = f_0 \right] = \mathbb{E}\left[F(X_1, \ldots, X_k) \right].$$

$$(51)$$

Note that, when $n \to \infty$, the quantity $\dfrac{1}{n} \sum_{i=1}^{n} f(X_i(\omega))$ almost surely converges to f_0. Theorem 11 shows that conditioning on the manifold $\dfrac{1}{n} \sum_{i=1}^{n} f(X_i(\omega)) = f_0$ does not change the value (when $n \to \infty$) of the expectation of a function F of a *finite* number k of random variables. Note that the condition that k is independent of n can be somewhat relaxed. It is indeed shown in [10] that one can take $k = o(n)$ in some cases. It is also shown there that one cannot take $k = n$.

In our context, the variable X_i is the value of the field A on the cell $Q + i$. The conditioning in the left-hand side of (51) is identical to the conditioning in the left-hand side of (50).

The difference between Theorem 11 and our result lies in the quantity of which we compute the expectation. In our case, this quantity is $A_N^\star(\omega)$, which is (asymptotically when $N \to \infty$) a function of *all* the variables X_i and not only of a *finite* number of them. We hence cannot directly use Theorem 11. The proof of our result essentially amounts to introducing an upper bound and a lower bound on $A_N^\star(\omega)$ that both read as a sum of functions that depend on a *finite* number of random variables (see e.g. (54) below). We will then be in position to apply Theorem 11 on these functions.

Proof of Theorem 8. We fix some $p \in \mathbb{R}^d$. For the sake of clarity, the approximate homogenized matrix $A_N^\star(\omega)$ defined by (4) is here denoted $A_{\mathrm{per}}^{\star,N}(\omega)$, to emphasize that we have considered periodic boundary conditions. Since the matrix A is symmetric, we have

$$p^T A_{\mathrm{per}}^{\star,N}(\omega)p = \inf\left\{ \mathcal{J}_{Q_N}(v,\omega), \;\; v \in H_{\mathrm{per}}^1(Q_N) \right\},$$

where

$$\mathcal{J}_{Q_N}(v,\omega) = \frac{1}{|Q_N|} \int_{Q_N} (p + \nabla v)^T A(\cdot,\omega)(p + \nabla v).$$

24

We have considered in (3) periodic boundary conditions. As is well-known, other boundary conditions can be used, and these alternate approximations will be useful for the proof.

Step 1: Upper bound. We first introduce an approximation of $A^\star$ using a truncated corrector problem complemented with homogeneous Dirichlet boundary conditions. We consider the problem

$$\begin{cases} -\operatorname{div}\left(A(\cdot,\omega)\left(p + \nabla w^N_{p,\mathrm{Dir}}(\cdot,\omega)\right)\right) = 0 & \text{in } Q_N, \\ w^N_{p,\mathrm{Dir}}(\cdot,\omega) = 0 & \text{on } \partial Q_N, \end{cases}$$

which yields an approximation of $A^\star$ that we denote $A^{\star,N}_{\mathrm{Dir}}(\omega)$ and which is defined by

$$\forall p \in \mathbb{R}^d, \quad A^{\star,N}_{\mathrm{Dir}}(\omega)p = \frac{1}{|Q_N|}\int_{Q_N} A(\cdot,\omega)(p + \nabla w^N_{p,\mathrm{Dir}}(\cdot,\omega)).$$

As shown in [6], we know that

$$\lim_{N \to \infty} A^{\star,N}_{\mathrm{Dir}}(\omega) = A^\star \quad \text{a.s.} \tag{52}$$

Since A is symmetric, we have

$$p^T A^{\star,N}_{\mathrm{Dir}}(\omega)p = \inf\left\{\mathcal{J}_{Q_N}(v,\omega), \quad v \in H^1_0(Q_N)\right\}.$$

The matrix $A^{\star,N}_{\mathrm{Dir}}(\omega)$ is always larger (in the sense of symmetric matrices) than $A^{\star,N}_{\mathrm{per}}(\omega)$. Indeed, let $v \in H^1_0(Q_N)$, and consider its Q_N-periodic extension $\widetilde{v}$. Then this function belongs to $H^1_{\mathrm{per}}(Q_N)$. We hence have that

$$p^T A^{\star,N}_{\mathrm{per}}(\omega)p \leq \mathcal{J}_{Q_N}(\widetilde{v},\omega) = \mathcal{J}_{Q_N}(v,\omega).$$

Minimizing over $v \in H^1_0(Q_N)$, we get that

$$p^T A^{\star,N}_{\mathrm{per}}(\omega)p \leq p^T A^{\star,N}_{\mathrm{Dir}}(\omega)p \quad \text{a.s.} \tag{53}$$

Just as $A^{\star,N}_{\mathrm{per}}(\omega)$, the matrix $A^{\star,N}_{\mathrm{Dir}}(\omega)$ depends on all the random variables $X_i(\omega)$, $i \in Q_N \cap \mathbb{Z}^d$. But, thanks to the use of homogeneous Dirichlet boundary conditions, it can be bounded from above by a sum of matrices that depend only on a finite number of random variables. To show this, we proceed as follows.

Figure 1: The domain Q_N (here represented for $N = 11$) is split into domains of size R^d (here $R = 2$; one of them is shown in red on the figure), up to some boundary layer $B_{N,R}$ (shown in light gray).

For any positive integers N and R, we introduce the integer part M of N/R. Then Q_N can decomposed into a set of cubes of size R^d, up to some boundary layer $B_{N,R}$:

$$Q_N = \left(\cup_{j \in \mathbb{Z}^d, \, |j| \leq M} \, R\,j + Q_R \right) \cup B_{N,R}.$$

For any $j \in \mathbb{Z}^d$, $|j| \leq M$, consider a function $v_j \in H_0^1(R\,j + Q_R)$. We now define the function v on Q_N as:

- for any $x \in R\,j + Q_R$, we set $v(x) = v_j(x)$;
- if $x \in B_{N,R}$, we set $v(x) = 0$.

The function v belongs to $H_0^1(Q_N)$. We hence write that

$$p^T A_{\mathrm{Dir}}^{\star,N}(\omega)p \leq \mathcal{J}_{Q_N}(v,\omega) = \frac{|Q_R|}{|Q_N|} \sum_{j \in \mathbb{Z}^d, \, |j| \leq M} \mathcal{J}_{R\,j+Q_R}(v_j,\omega).$$

Minimizing over the functions $v_j \in H_0^1(R\,j + Q_R)$, we hence get that

$$p^T A_{\mathrm{Dir}}^{\star,N}(\omega)p \leq \frac{|Q_R|}{|Q_N|} \sum_{j \in \mathbb{Z}^d, \, |j| \leq M} Y_j(\omega) \quad \text{a.s.} \tag{54}$$

where

$$Y_j(\omega) = \inf \left\{ \mathcal{J}_{R\,j+Q_R}(v,\omega), \quad v \in H_0^1(R\,j + Q_R) \right\}.$$

Since A is stationary, we note that all the random variables $Y_j(\omega)$ share the same law. Moreover, we observe that $Y_0(\omega) = p^T A_{\mathrm{Dir}}^{\star,R}(\omega)p$, which is the approximation of the homogenized matrix using Dirichlet boundary conditions on Q_R.

We now take the conditional expectation of (54), and use the fact that the variables Y_j all share the same law:

$$\mathbb{E} \left[p^T A_{\mathrm{Dir}}^{\star,N}(\omega)p \ \Big| \ \frac{1}{|Q_N|} \sum_{k \in Q_N \cap \mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right]$$

$$\leq \frac{R^d\,M^d}{N^d} \, \mathbb{E} \left[p^T A_{\mathrm{Dir}}^{\star,R}(\omega)p \ \Big| \ \frac{1}{|Q_N|} \sum_{k \in Q_N \cap \mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right].$$

We next observe that $p^T A_{\mathrm{Dir}}^{\star,R}(\omega)p$ only depends on a *finite* number of random variables, namely only on $X_k(\omega)$ with $k \in Q_R \cap \mathbb{Z}^d$. We are

thus in position to use Theorem 11, which yields the limit of the above right-hand side when $N \to \infty$. Hence, for any fixed R, we have

$$\limsup_{N\to\infty} \mathbb{E}\left[p^T A_{\mathrm{Dir}}^{\star,N}(\omega)p \,\Big|\, \frac{1}{|Q_N|} \sum_{k\in Q_N\cap\mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right]$$
$$\leq \mathbb{E}\left[p^T A_{\mathrm{Dir}}^{\star,R}(\omega)p \right].$$

Letting R go to ∞ in the above bound and using (52), we obtain that

$$\limsup_{N\to\infty} \mathbb{E}\left[p^T A_{\mathrm{Dir}}^{\star,N}(\omega)p \,\Big|\, \frac{1}{|Q_N|} \sum_{k\in Q_N\cap\mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right] \leq p^T A^\star p.$$

Using (53), we deduce that

$$\forall p \in \mathbb{R}^d, \quad \limsup_{N\to\infty} p^T U_N p \leq p^T A^\star p, \tag{55}$$

where

$$U_N = \mathbb{E}\left[A_{\mathrm{per}}^{\star,N}(\omega) \,\Big|\, \frac{1}{|Q_N|} \sum_{k\in Q_N\cap\mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right]. \tag{56}$$

Step 2: Lower bound. We now introduce an approximation of $A^\star$ using a truncated problem complemented with Neumann boundary conditions. We consider the problem

$$\begin{cases} -\operatorname{div}\left(A(\cdot,\omega)\left(p + \nabla w_{p,\mathrm{Neu}}^N(\cdot,\omega) \right) \right) = 0 & \text{in } Q_N, \\ n^T A(\cdot,\omega)(p + \nabla w_{p,\mathrm{Neu}}^N(\cdot,\omega)) = n^T p & \text{on } \partial Q_N, \end{cases} \tag{57}$$

which yields an approximation of $A^\star$ that we denote $A_{\mathrm{Neu}}^{\star,N}(\omega)$ and which is defined by

$$A_{\mathrm{Neu}}^{\star,N}(\omega) = \left(S_{\mathrm{Neu}}^{\star,N}(\omega) \right)^{-1}, \tag{58}$$

where $S_{\mathrm{Neu}}^{\star,N}(\omega)$ is defined by

$$\forall p \in \mathbb{R}^d, \quad S_{\mathrm{Neu}}^{\star,N}(\omega)p = \frac{1}{|Q_N|} \int_{Q_N} p + \nabla w_{p,\mathrm{Neu}}^N(\cdot,\omega). \tag{59}$$

28

We refer to Remark 12 below for some heuristic justification of (58)–(59).

As recalled in [19, Appendix], we have that

$$\lim_{N \to \infty} A_{\text{Neu}}^{\star,N}(\omega) = A^\star \quad \text{a.s.} \tag{60}$$

and

$$p^T A_{\text{Neu}}^{\star,N}(\omega) p \leq p^T A_{\text{per}}^{\star,N}(\omega) p \quad \text{a.s.} \tag{61}$$

In addition, we have the following variational characterization:

$$p^T S_{\text{Neu}}^{\star,N}(\omega) p = \inf \left\{ \mathcal{E}_{Q_N}(\sigma, \omega), \quad \sigma \in V(Q_N) \right\}, \tag{62}$$

where

$$\mathcal{E}_{Q_N}(\sigma, \omega) = \frac{1}{|Q_N|} \int_{Q_N} (p + \sigma)^T A^{-1}(\cdot, \omega)(p + \sigma)$$

and

$$V(Q_N) = \left\{ \sigma \in (L^2(Q_N))^d, \quad \text{div}\, \sigma = 0 \text{ in } Q_N, \quad n^T \sigma = 0 \text{ on } \partial Q_N \right\}.$$

The matrix $S_{\text{Neu}}^{\star,N}(\omega)$ (and hence the matrix $A_{\text{Neu}}^{\star,N}(\omega)$) depends on all the variables $X_i(\omega)$, $i \in Q_N \cap \mathbb{Z}^d$. However, thanks to the characterization (62), it can be bounded from above by a sum of matrices that depend only on a finite number of random variables.

To show this, we proceed as in Step 1 of the proof. For any positive integers N and R, we introduce the integer part M of N/R, and decompose Q_N into a set of cubes of size R^d, up to some boundary layer $B_{N,R}$ (see Figure 1):

$$Q_N = \left(\cup_{j \in \mathbb{Z}^d, \, |j| \leq M} R j + Q_R \right) \cup B_{N,R}.$$

For any $j \in \mathbb{Z}^d$, $|j| \leq M$, consider a function $\sigma_j \in V(Rj + Q_R)$. We now define the function σ on Q_N as:

- for any $x \in Rj + Q_R$, we set $\sigma(x) = \sigma_j(x)$;

- if $x \in B_{N,R}$, we set $\sigma(x) = 0$.

We claim that $\sigma \in V(Q_N)$. We indeed first have that $\sigma \in (L^2(Q_N))^d$.

We next consider $\varphi \in C_0^\infty(Q_N)$ and compute that

$$
\begin{aligned}
\langle \operatorname{div} \sigma, \varphi \rangle &= -\langle \sigma, \nabla \varphi \rangle \\
&= -\sum_{j \in \mathbb{Z}^d, \, |j| \leq M} \int_{Rj+Q_R} \sigma_j \cdot \nabla \varphi \\
&= -\sum_{j \in \mathbb{Z}^d, \, |j| \leq M} \int_{\partial(Rj+Q_R)} n_j^T \sigma_j \, \varphi \\
&= 0,
\end{aligned}
$$

where n_j is the outward normal to the domain $Rj + Q_R$. We hence have checked that $\sigma \in V(Q_N)$.

We next write that

$$
p^T S_{\mathrm{Neu}}^{\star,N}(\omega)p \leq \mathcal{E}_{Q_N}(\sigma, \omega) = \frac{|Q_R|}{|Q_N|} \sum_{j \in \mathbb{Z}^d, \, |j| \leq M} \mathcal{E}_{Rj+Q_R}(\sigma_j, \omega).
$$

Minimizing over the functions $\sigma_j \in V(Rj + Q_R)$, we hence get that

$$
p^T S_{\mathrm{Neu}}^{\star,N}(\omega)p \leq \frac{|Q_R|}{|Q_N|} \sum_{j \in \mathbb{Z}^d, \, |j| \leq M} Z_j(\omega) \quad \text{a.s.} \tag{63}
$$

where

$$
Z_j(\omega) = \inf \left\{ \mathcal{E}_{Rj+Q_R}(\sigma, \omega), \quad \sigma \in V(Rj + Q_R) \right\}.
$$

Since A is stationary, we note that all the random variables $Z_j(\omega)$ share the same law. Moreover, we observe that $Z_0(\omega) = p^T S_{\mathrm{Neu}}^{\star,R}(\omega)p$.

We now take the conditional expectation of (63), and use the fact that the variables Z_j all share the same law:

$$
\mathbb{E}\left[p^T S_{\mathrm{Neu}}^{\star,N}(\omega)p \;\Big|\; \frac{1}{|Q_N|} \sum_{k \in Q_N \cap \mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right]
$$

$$
\leq \frac{R^d \, M^d}{N^d} \mathbb{E}\left[p^T S_{\mathrm{Neu}}^{\star,R}(\omega)p \;\Big|\; \frac{1}{|Q_N|} \sum_{k \in Q_N \cap \mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)] \right].
$$

We observe that $p^T S_{\mathrm{Neu}}^{\star,R}(\omega)p$ only depends on a *finite* number of random variables, namely only on X_k with $k \in Q_R \cap \mathbb{Z}^d$. We are thus in position to use Theorem 11, which yields the limit of the above

right-hand side when $N \to \infty$. Hence, for any fixed R, we have

$$\limsup_{N\to\infty} \mathbb{E}\left[p^T S_{\text{Neu}}^{\star,N}(\omega)p \;\Big|\; \frac{1}{|Q_N|}\sum_{k\in Q_N\cap\mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)]\right]$$
$$\leq \mathbb{E}\left[p^T S_{\text{Neu}}^{\star,R}(\omega)p\right].$$

Letting R go to ∞ in the above bound and using (58) and (60), we obtain that

$$\limsup_{N\to\infty} \mathbb{E}\left[p^T S_{\text{Neu}}^{\star,N}(\omega)p \;\Big|\; \frac{1}{|Q_N|}\sum_{k\in Q_N\cap\mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)]\right]$$
$$\leq p^T(A^\star)^{-1}p.$$

Using (58) and (61), we deduce that

$$\limsup_{N\to\infty} \mathbb{E}\left[p^T\left(A_{\text{per}}^{\star,N}(\omega)\right)^{-1}p \;\Big|\; \frac{1}{|Q_N|}\sum_{k\in Q_N\cap\mathbb{Z}^d} f(X_k) = \mathbb{E}[f(X_0)]\right]$$
$$\leq p^T(A^\star)^{-1}p.$$

Using Jensen inequality, we infer from the above bound that

$$\forall p \in \mathbb{R}^d, \quad \limsup_{N\to\infty} p^T(U_N)^{-1}p \leq p^T(A^\star)^{-1}p, \tag{64}$$

where the matrix U_N is defined by (56).

Step 3: Conclusion. We eventually show that (55) and (64) imply that U_N converges to $A^\star$ when $N \to \infty$.

From the assumptions on A, we know that there exists $0 < a_- \leq a_+ < \infty$ such that, for any N and almost surely, $a_- \leq A_{\text{per}}^{\star,N}(\omega) \leq a_+$. Hence, for any N, the symmetric matrix U_N satisfies $a_- \leq U_N \leq a_+$. We can thus extract a subsequence $U_{\varphi(N)}$ that converges to some symmetrix matrix B. Let us show that $B = A^\star$.

Let $p \in \mathbb{R}^d$. We first observe that, by definition,

$$\limsup_{k\to\infty} p^T U_k p \geq \lim_{k\to\infty} p^T U_{\varphi(k)} p = p^T B p.$$

We thus infer from (55) that

$$\forall p \in \mathbb{R}^d, \quad p^T B p \leq p^T A^\star p. \tag{65}$$

31

We now proceed likewise with U_k^{-1}. We observe that,

$$\limsup_{k \to \infty} \, p^T U_k^{-1} p \geq \lim_{k \to \infty} \, p^T U_{\varphi(k)}^{-1} p = p^T B^{-1} p.$$

We thus infer from (64) that

$$\forall p \in \mathbb{R}^d, \quad p^T B^{-1} p \leq p^T (A^\star)^{-1} p. \tag{66}$$

Collecting (65) and (66), we deduce that $B = A^\star$.

The sequence U_N is bounded, and we have shown that any converging subsequence converges to $A^\star$. This implies that U_N converges to $A^\star$ when $N \to \infty$, which is exactly the result (50). This concludes the proof of Theorem 8. $\qquad\square$

Remark 12. *In view of* (57), *we can check that*

$$\frac{1}{|Q_N|} \int_{Q_N} A(\cdot, \omega) \left(p + \nabla w_{p,\mathrm{Neu}}^N(\cdot, \omega) \right) = p.$$

The definition (58)–(59) *can hence be understood as*

$$\left\langle A(\cdot, \omega) \left(p + \nabla w_{p,\mathrm{Neu}}^N(\cdot, \omega) \right) \right\rangle = A_{\mathrm{Neu}}^{\star, N}(\omega) \left\langle p + \nabla w_{p,\mathrm{Neu}}^N(\cdot, \omega) \right\rangle,$$

where $\langle \cdot \rangle = |Q_N|^{-1} \int_{Q_N} \cdot$ *is the average on* Q_N.

3.2 Complete analysis in some simple cases

In this section, we aim at improving the convergence result (50) of the previous section by quantifying both the statistical and systematic errors, in order to assess the efficiency of our approach. We are only able to proceed in simple situations where all the quantities are indeed accessible using analytic calculations. These two situations are examined in Sections 3.2.1 and 3.2.2 respectively. For the sake of brevity, and because the proofs are not very enlightening and are not likely to carry over to more general cases, we do not provide the proofs of our claims here. We refer to [19] where they are presented in details.

We establish below that our approach preserves the *rate* of decay of the standard Monte Carlo sampling both for the systematic and the statistical error (and thus, in particular, the systematic error remains, in rate, smaller than the statistical error). Furthermore, the *prefactor* in the statistical error is significantly reduced by our approach.

3.2.1 "Zero-dimensional" homogenization

As simplest possible situation, we consider a function $g : \mathbb{R} \mapsto \mathbb{R}$ and the random variables $(X_i)_{1 \leq i \leq n}$. We assume that these random variables are independent and that they are all centered Gaussian random variables with unit variance. We also assume that $g \in C^1(\mathbb{R})$ and that $\mathbb{E}\left[|g(X_1)| + |g'(X_1)|\right] < \infty$. Note that it is not surprising to make some smoothness assumptions on g as we are here after *rates* of convergence, and not only a convergence result as in Section 3.1.

We set

$$\xi : x \mapsto \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Assume we want to compute $\mathbb{E}[g(X_1)]$. A classical Monte Carlo approach would approximate this by the limit of the empirical mean $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(X_i(\omega))$. In this particular instance, the simplest version of our variance reduction approach instead considers $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(X_i(\omega))$ for realizations $X(\omega)$ that satisfy $\xi(X(\omega)) = 0$.

In this simple case, the bias of the classical approach is actually identically zero: of course, $\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} g(X_i)\right]$ does not depend on n. The statistical error is controlled by the Central Limit Theorem and is asymptotically of order $\sqrt{\frac{\mathbb{V}\mathrm{ar}[g(X_1)]}{n}}$.

Proposition 13. *Under the assumptions of this section, the bias of the selection method is of order $1/n$. More specifically,*

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} g(X_i) \ \Big| \ \xi(X) = 0\right] - \mathbb{E}[g(X_1)] = -\frac{1}{2n}\mathbb{E}[g'(X_1)] + O\left(\frac{1}{n^2}\right).$$

$$(67)$$

The variance of the selection method is reduced by a factor asymptotically independent of n. More specifically,

$$\frac{\mathbb{V}\mathrm{ar}\left[\frac{1}{n} \sum_{i=1}^{n} g(X_i) \ \Big| \ \xi(X) = 0\right]}{\mathbb{V}\mathrm{ar}\left[\frac{1}{n} \sum_{i=1}^{n} g(X_i)\right]} = 1 - \frac{(\mathbb{E}[g'(X_1)])^2}{\mathbb{V}\mathrm{ar}[g(X_1)]} + O\left(\frac{1}{n}\right). \quad (68)$$

In view of (67)–(68), we observe that, at the price of introducing a bias of order $O\left(1/n\right)$, our approach reduces the statistical error from

$\dfrac{\lambda_{\mathrm{MC}}}{\sqrt{n}}$ to $\dfrac{\lambda_{\mathrm{SQS}}}{\sqrt{n}}$ (with $\lambda_{\mathrm{SQS}} < \lambda_{\mathrm{MC}}$), and therefore, for sufficiently large n, reduces the total error.

The following result covers the case where we insert a non-zero tolerance in Algorithm 2.

Proposition 14. *Under the assumptions of this section, consider the selection method where we condition on the realizations such that $\dfrac{z_0}{\sqrt{n}} \leq \xi(X(\omega)) \leq \dfrac{z_1}{\sqrt{n}}$, for some z_0 and $z_1 > z_0$ in $\mathbb{R}$. Then, for any choice of z_0 and $z_1 > z_0$, the variance of the selection method is reduced by a factor asymptotically independent of n:*

$$\frac{\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n} g(X_i) \ \Big| \ \frac{z_0}{\sqrt{n}} \leq \xi(X) \leq \frac{z_1}{\sqrt{n}}\right]}{\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right]}$$

$$= 1 - (1 - C)\frac{(\mathbb{E}[g'(X_1)])^2}{\mathbb{V}\mathrm{ar}[g(X_1)]} + O\left(\frac{1}{n}\right), \quad (69)$$

where $C = \mathbb{V}\mathrm{ar}\left[X_1 \ \Big| \ z_0 \leq X_1 \leq z_1\right]$.

The conditioning $z_0/\sqrt{n} \leq \xi(X(\omega)) \leq z_1/\sqrt{n}$ is deliberately chosen in order to match the rate of the Central Limit Theorem. It corresponds to the selection of a fixed *proportion* of samples (as in Algorithm 3 when $\mathcal{M}$ is proportional to M). Note that $C > 0$, hence the variance is less reduced than when conditioning at $\xi(X) = 0$ (which is the case considered in Proposition 13). Note also that the variance is reduced (with respect to the standard Monte Carlo sampling) if, and only if, $1 - C \geq 0$. We are yet unable to conclude that this is the case in general. We simply note that, when $z_1 = -z_0 > 0$, then $C = 1$, yielding no gain.

3.2.2 One-dimensional homogenization

In the one-dimensional case, the homogenization of a random field $a : (y, \omega) \mapsto \sum_{i \in \mathbb{Z}} \overline{g}(X_i(\omega)) \mathbb{1}_{(i,i+1)}(y)$ (where $\overline{g}$ is valued, say, in $[a_-, a_+]$ with $a_- > 0$) is a simple harmonic average. It is readily seen that

$$a_N^{\star}(\omega) = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\overline{g}(X_k)}\right)^{-1} = \varphi\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\overline{g}(X_k)}\right) \quad \text{with } \varphi(x) = 1/x.$$

Formally, the problem is thus analogous to that of the previous section, for a certain $\varphi : \mathbb{R} \mapsto \mathbb{R}$ instead of $\varphi = \mathrm{Id}$. Therefore, it is sufficient to prove consistency and variance reduction for quantities of the form

$$\varphi \left(\frac{1}{N} \sum_{i=1}^{N} g(X_i) \right).$$

Proposition 15. *Consider a smooth function $\varphi : \mathbb{R} \mapsto \mathbb{R}$. Under the assumptions of this section, the bias of the standard method and that of the selection method respectively are*

$$\mathbb{E} \left[\varphi \left(\frac{1}{N} \sum_{i=1}^{N} g(X_i) \right) \right] - \varphi(g_0) = \frac{\varphi''(g_0)}{2N} \mathbb{V}\mathrm{ar}[g(X_1)] + O\left(\frac{1}{N^2}\right) \quad (70)$$

and

$$\mathbb{E} \left[\varphi \left(\frac{1}{N} \sum_{i=1}^{N} g(X_i) \right) \ \Big| \ \xi(X) = 0 \right] - \varphi(g_0)$$

$$= \frac{\varphi''(g_0)}{2N} \left(\mathbb{V}\mathrm{ar}[g(X_1)] - (\mathbb{E}[g'(X_1)])^2 \right) - \frac{\varphi'(g_0)}{2N} \mathbb{E}[X_1 g'(X_1)] + o\left(\frac{1}{N}\right),$$

$$(71)$$

with $g_0 = \mathbb{E}[g(X_1)]$.

The variance of the selection method is reduced by a factor asymptotically independent of N:

$$\frac{\mathbb{V}\mathrm{ar} \left[\varphi \left(\frac{1}{N} \sum_{i=1}^{N} g(X_i) \right) \ \Big| \ \xi(X) = 0 \right]}{\mathbb{V}\mathrm{ar} \left[\varphi \left(\frac{1}{N} \sum_{i=1}^{N} g(X_i) \right) \right]} = 1 - \frac{(\mathbb{E}[g'(X_1)])^2}{\mathbb{V}\mathrm{ar}[g(X_1)]} + o(1). \quad (72)$$

To keep things simple, we do not investigate whether a more general result, accounting for some tolerance in the manner our condition is fulfilled (in the spirit of Proposition 14), holds here.

Proposition 15 shows that the bias is unchanged in rate, while the prefactor for the variance is reduced. Since the variance only decays at the rate $1/\sqrt{N}$ while the bias decays at the rate $1/N$, we see that our approach indeed reduces the total error for sufficiently large N.

In the numerical practice (mimicking in this one-dimensional setting what is actually performed for higher dimensional settings – although it is in some sense unnecessary here), we generate several, independent realizations of the N-tuples $(X_i)_{1 \leq i \leq N}$ corresponding to as many draws of environments within the "cube" Q_N. In the classical

Monte Carlo approach, we keep all such N-tuples. In our approach, we only consider those that satisfy an additional criterion.

An empirical mean (aimed at approximating $A^\star$) is then computed. The systematic error and the statistical error of the latter approximation are precisely related to the errors estimated in (70)–(71)–(72) respectively. Thus a theoretical assessment of our practical approach.

4 Numerical experiments

We first present in this section some numerical experiments that show the robustness of our variance reduction approach with respect to the tolerance with which we enforce the SQS conditions (see Section 4.1). We next turn to studying the performance of our approach in Section 4.2.

We consider the test-case when A reads as in (12), that is

$$A_\eta(x,\omega) = C_0(x,\omega) + \eta\ \chi(x,\omega)C_1(x,\omega),$$

with $\eta = 1/2$, $C_0 = C_1 = \mathrm{Id}$, and χ is of the form (35), that is

$$\chi(x,\omega) = \sum_{k\in\mathbb{Z}^d} X_k(\omega)\mathbb{1}_{Q+k}(x).$$

The random variables X_k are i.i.d. and follow a Bernoulli law of parameter $1/2$ valued in $\{-1,+1\}$. The contrast (i.e. the ratio of the largest value of A divided by its minimum value) is equal to 3. The influence of the contrast on the efficiency of our approach is investigated at the end of Section 4.2 (see Table 1). We consider there much larger values of the contrast (however all smaller than 20).

In what follows, we only consider Algorithm 3, where we take $M = 100$ and $\mathcal{M} = 2000$ (thus an acceptance ratio of 5%).

In this setting, the SQS 1 condition as stated in (48) is satisfied if and only if the numbers of cells within which $X_k(\omega) = 1$ is equal to the number of cells within which $X_k(\omega) = -1$. It is thus possible to enforce (48) by randomly selecting $|Q_N|/2$ cells within the $|Q_N|$ cells that are in Q_N, and setting $X_k = 1$ on these cells and $X_k = -1$ on the others.

In all our tests, we have kept the computational time fixed, or almost fixed, since the additional time needed by the selection step (namely Steps 1 and 2 of Algorithm 3) is roughly 5% of the total original computational time.

We conclude this section with some numerical tests on a problem involving a more general geometry of microstructures (see Section 4.3). On such a problem, we again obtain a significant reduction of the variance, at no additional computational cost.

4.1 Robustness of the approach

As pointed out above, the SQS 2 condition as stated in (49) is only enforced in Algorithm 3 up to some tolerance. In this section, we experimentally investigate how this tolerance affects the quality of the approximation and the efficiency of the approach. To mimick the difficulty associated with the SQS 2 condition, we have also performed some tests where we only enforce the SQS 1 condition up to some tolerance, and not exactly. The results of our numerical tests are displayed in Figures 2 through 5.

Figures 2 and 3 show the sensitivity of the variance reduction ratio upon the first order condition (48). Using Algorithm 3, we generate $\mathcal{M} = 2000$ realizations. To investigate the robustness of our approach, we sort these realizations with respect to the error in (48), and successively consider 20 groups of 100 realizations that less and less accurately satisfy (48). On Figure 2, the left-most circle displays the ratio $V_{\text{SQS 1}}/V_{\text{MC}}$ between the empirical variance $V_{\text{SQS 1}}$ among the best $M = 100$ realizations and the reference Monte Carlo variance $V_{\text{MC}} = \mathbb{V}\text{ar}\left[\left(A_N^\star\right)_{11}\right]$. The second circle shows the ratio between the empirical variance among the next best $M = 100$ realizations and the reference Monte Carlo variance V_{MC}. We proceed similarly with all the subsequent groups of $M = 100$ realizations.

On Figure 3, we display the same ratio of variances in function, for each group of $M = 100$ realizations, of the maximum error with which the first order condition (48) is satisfied. Hence, the first group (left-most circle) corresponds to exactly satisfying the condition, the second group corresponds to an error between 0 and tol, the third group corresponds to an error between tol and $2\,tol$, and so on and so forth.

Figures 4 and 5 show the sensitivity upon the second order condition (49). Here, we only consider realizations that satisfy (48). Using Algorithm 3, we again generate $\mathcal{M} = 2000$ realizations and sort them according to the error in (49). We again successively consider 20 groups of 100 realizations that all satisfy (48) but that less and less accurately satisfy (49). We present the results on Figures 4 and 5

following the same procedure as for Figures 2 and 3. For instance, for the left-most circle, we plot the ratio $\dfrac{V_{\text{SQS 2}}}{V_{\text{exact SQS 1}}}$ between the variance $V_{\text{SQS 2}}$ among the $M = 100$ realizations that exactly satisfy the SQS 1 condition and best satisfy the SQS 2 condition on the one hand, and, on the other hand, the variance $V_{\text{exact SQS 1}}$ of the realizations that exactly satisfy the SQS 1 condition.

We observe that, even if the SQS conditions (48)–(49) are not *exactly* satisfied, but only with some small tolerance, we obtain a significant variance reduction. We conclude that our approach is robust in this respect.

Figure 2: Variance ratio $V_{\text{SQS 1}}/V_{\text{MC}}$ for the 20 groups of realizations (sorted according to their SQS 1 error).

We next investigate whether enforcing the SQS 1 condition (48) affects the probability distribution function of the left-hand side of the SQS 2 condition (49). Results are shown on Figure 6, where we plot two histograms:

38

Figure 3: Variance ratio $V_{\text{SQS }1}/V_{\text{MC}}$ as a function of the error in (48). Results for only the best 7 groups (out of the 20 groups) are shown.

Figure 4: Variance ratio $\dfrac{V_{\text{SQS 2}}}{V_{\text{exact SQS 1}}}$ for the distinct groups of realizations (sorted according to their SQS 2 error; the SQS 1 condition is exactly satisfied). Only the 10 best groups are shown.

Figure 5: Variance ratio $\dfrac{V_{\text{SQS 2}}}{V_{\text{exact SQS 1}}}$ as a function of the error in (49) (the condition (48) is exactly satisfied). Only the 10 best groups are shown.

- the distribution of the criterion SQS 2 (namely, the left-hand side of (49)) among all realizations.

- the conditional distribution of the criterion SQS 2 among realizations that exactly satisfy the SQS 1 condition (48).

The two histograms are sufficiently close to each other to state that enforcing the SQS 1 condition does not change the distribution of the SQS 2 criterion.

PSfrag replacements

Figure 6: Empirical probability distribution function of the SQS 2 criterion (black histogram: no conditioning; red histogram: the samples exactly satisfy the SQS 1 criterion). Both histograms have been computed using 100 realizations.

4.2 Efficiency of the approach

In this section, we investigate how the efficiency of our approach depends (i) on the size of the truncated domain Q_N and (ii) on the contrast in A.

4.2.1 Experimental error analysis

Figure 7 shows the set of approximations of the first entry $[A^\star]_{11}$ of the homogenized matrix and their respective confidence intervals. We show three curves (along with their respective confidence intervals):

- The standard Monte Carlo approximation, which is defined by (10). The variance is large.

- The approximation obtained by selecting realizations that exactly satisfy the SQS 1 condition. The variance is much smaller, leading in turn to a narrower confidence interval.

- The approximation obtained with realizations satisfying exactly the SQS 1 condition and selected according to the SQS 2 condition (see Algorithm 3). The variance is much smaller than when using the SQS 1 approach, even when the size of the domain Q_N is small.

PSfrag replacements

Figure 7: Approximations of $[A^\star]_{11}$ (along with confidence intervals) as a function of N. Black curve: Monte Carlo method. Red curve: SQS 1 method. Blue curve: SQS 2 method (see text).

Figure 8 shows a representation of the total error as a function of the size of Q_N. As often in a Monte Carlo approach, computing the total error is challenging, precisely because the reference value is,

by definition, in general unknown. In the specific case considered, namely the random checkerboard, the value of $A^\star$ is actually known and equal to $A^\star = \sqrt{(1+\eta)(1-\eta)}\,\mathrm{Id}$. But for the large number of realizations and the large size of the truncated domains that we aim at considering, the total error is so small that we cannot neglect the contribution of the specific error due the finiteness of the meshsize used to solve (3). Therefore, we are bound to obtain and use as reference an approximate value $A_h^\star$ of $A^\star$ corresponding to an hypothetical finite element approximation on the whole space. As a surrogate for this $A_h^\star$, which is unknown in practice, we choose the empirical expectation of $A_{N_{\mathrm{ref}}}^\star(\omega)$ over $\mathcal{M}_{\mathrm{ref}} = 2000$ random realizations exactly satisfying the SQS 1 condition (with a view to use a value with the lowest possible statistical error), and for the largest domain $Q_{N_{\mathrm{ref}}}$ we can consider given the computing facilities we have access to, that is $N_{\mathrm{ref}} = 50$.

On Figure 8, we display three curves:

- The total error of the standard Monte Carlo method, defined as

$$\text{total error} = \left| \frac{1}{M} \sum_{m=1}^{M} A_N^\star(\omega_m) - A_{\mathrm{ref}}^\star \right|.$$

- The other two curves show the same quantity, where the M environments considered now satisfy either the SQS 1 condition or that condition together with the SQS 2 condition.

The results obtained using the SQS 2 approach are in general comparable to, and often better than, those obtained with the SQS 1 approach. More accurate estimates of the reference value $A_h^\star$ would probably help in clarifying the superiority of SQS 2 over SQS 1 in terms of total accuracy. As will now be seen, the superiority of SQS 2 in terms of variance (which, in some sense, is the key point for practice) is definite.

Figure 9 shows the empirical variance of the different approximations of $[A^\star]_{11}$ as a function of the size of Q_N. We again display three curves:

- The standard Monte Carlo approximation defined by (10).

- The approximation obtained by selecting realizations that exactly satisfy the SQS 1 condition.

- The approximation obtained with realizations exactly satisfying the SQS 1 condition and selected according to the SQS 2 condition (see Algorithm 3).

44

Figure 8: log-log plot of the total error as a function of N (natural logarithm). Black curve: Monte Carlo method. Red curve: SQS 1 method. Blue curve: SQS 2 method (see text).

We observe that, each time we consider an additional SQS condition, the empirical variance of the approximation is significantly reduced (even if this SQS condition is not exactly enforced; recall that we consider here only the 5 % best samples in terms of the SQS 2 condition, but that we are unable to enforce it exactly). On our test-case, enforcing the SQS 1 condition leads to a variance 20 times smaller than that of the standard Monte Carlo approach, while additionally enforcing the SQS 2 condition leads to an additional variance reduction of a factor of 10.

We also observe on Figure 9 that all variances decay as $\lambda/|Q_N|$, where

$$\lambda_{\text{SQS 2}} < \lambda_{\text{exact SQS 1}} < \lambda_{\text{MC}}.$$

This corroborates in higher dimension the behaviour predicted in Section 3.2. In particular, the gain in variance does not decrease when the size of Q_N becomes larger.

4.2.2 Sensitivity to the contrast

We eventually investigate how the contrast in the field A affects the gain in variance. Results are shown in Table 1. We observe that the gain decreases when the contrast increases. Note that this is also the case with the antithetic variable and the control variate techniques that we have previously studied (see [4, 18, 17]).

However, our SQS 2 approach still yields a significant gain of a factor of 10 when the contrast is equal to 20.

4.3 A more general geometry

In order to show that the approach carries over to other settings involving more general geometries than the setting considered above, we briefly consider in the present final section a linear elasticity problem, for a two-phase composite material with random inclusions. The radii $r_j(\omega)$ of the inclusions are i.i.d. random variables satisfying

$$r_j(\omega) = \sqrt{(M - m)\sqrt{U_j(\omega)} + m},$$

where $U_j(\omega)$ are i.i.d. variables uniformly distributed in $[0, 1]$. The parameters M and m are such that the minimum (resp. maximum) inclusion radius is 0.125 (resp. 0.45). The inclusions centers are distributed according to a Poisson point process, and we additionally

46

Figure 9: log-log plot of the variance as a function of N (natural logarithm). Black curve: Monte Carlo method. Red curve: SQS 1 method. Blue curve: SQS 2 method.

47

Contrast	V_{MC}	$V_{\text{exact SQS 1}}$	$V_{\text{SQS 2}}$	$\dfrac{V_{\text{MC}}}{V_{\text{exact SQS 1}}}$	$\dfrac{V_{\text{MC}}}{V_{\text{SQS 2}}}$
1.22	0.0000273	5.801e-08	6.858e-10	470	39821
1.50	0.0001097	0.0000009	1.585e-08	118	6921
1.86	0.0002488	0.0000047	0.0000001	52.6	1996
2.33	0.0004478	0.0000151	0.0000006	29.5	720
3.00	0.0007118	0.0000379	0.0000024	18.8	296
4.00	0.0010496	0.0000814	0.0000080	12.8	131
5.67	0.0014769	0.0001600	0.0000244	9.23	60.5
9.00	0.0020289	0.0003021	0.0000739	6.71	27.4
19.0	0.0028330	0.0006061	0.0002554	4.67	11.1

Table 1: For various values of the contrast, we show the Monte Carlo variance (column #2), the variance of the SQS 1 method (column #3) and the variance of the SQS 2 method (column #4). We next show the variance ratio $V_{\text{MC}}/V_{\text{exact SQS 1}}$ for the SQS 1 approach (column #5) and the variance ratio $V_{\text{MC}}/V_{\text{SQS 2}}$ for the SQS 2 approach (column #6). The size of Q_N is fixed at $N = 20$.

impose that inclusions do not overlap (see Figure 10). We consider the truncated domain $Q_N = [0,5]^2$. Each microstructure contains 25 inclusions. If some part of an inclusion falls outside of Q_N, it is reproduced on the other side of Q_N by periodicity.

The inclusions (resp. the background) are modeled by a isotropic linear elasticity tensor, with a uniform Poisson ratio $\nu = 0.3$ and a Young modulus $E = 7$ (resp. $E = 1$).

For that problem, in the spirit of the SQS 1 approach described above, we select the microstructures such that the volumic fraction $\theta_N(\omega)$ of inclusions within the domain Q_N, for each random realization considered, agrees as accurately as possible with its asymtotic value $\theta^\star = \lim_{N \to \infty} \theta_N(\omega)$.

The results are shown in Table 2. We examine two entries of the homogenized elasticity tensor, and two methods:

- The classical Monte Carlo approach, for which we generate $M = 100$ i.i.d. microstructures.

- The SQS approach, in which we generate $\mathcal{M} = 2000$ i.i.d. microstructures, and next consider the $M = 100$ microstructures

Figure 10: Two realizations of the microstructure geometry.

for which $|\theta_N(\omega) - \theta^\star|$ is the smallest.

In both cases, we solve M correctors problems and we compute the empirical expectation and variance of the homogenized elasticity tensor. We observe that our approach provides a variance reduction of a factor close to 20, while the bias is essentially constant.

	$[A_N^\star(\omega)]_{1111}$	$[A_N^\star(\omega)]_{1111}$	$[A_N^\star(\omega)]_{1122}$	$[A_N^\star(\omega)]_{1122}$
	Exp.	Var.	Exp.	Var.
MC approach	2.522	0.0136	1.016	0.00184
SQS approach	2.519	0.000532	1.014	0.000101

Table 2: Empirical expectation (Exp.) and variance (Var.) for two entries of the homogenized elasticity tensor, computed with the Monte Carlo approach (MC) or our approach (SQS).

Acknowledgements

A   Proof of Lemma 4

We follow the arguments of the proof of [5, Lemma 3.2].

The existence and uniqueness (up to the addition of a constant) of ϕ_1 solution to (38) is established in [5, Lemma 3.1]. We next point out that (39) admits a unique (up to the addition of a constant) solution in $H^1_{\text{per}}(Q)$. It is a simple consequence of the Lax-Milgram lemma.

We now prove that the sum in (37) is a convergent series in $L^2(Q \times \Omega)$. For this purpose, we compute the norm of the remainder of the series, using the notation $\overline{X}_k(\omega) = X_k(\omega) - \mathbb{E}[X_k]$:

$$\left\| \sum_{|k| \geq N+1} \overline{X}_k \nabla \phi_1(\cdot - k) \right\|^2_{L^2(Q \times \Omega)}$$

$$= \sum_{|k| \geq N+1} \sum_{|\ell| \geq N+1} \mathbb{E}\left[\overline{X}_k \overline{X}_\ell \right] \int_Q \nabla \phi_1(y - k) \cdot \nabla \phi(y - \ell) \, dy$$

$$\leq \sum_{|k| \geq N+1} \sum_{|\ell| \geq N+1} |\mathbb{C}\text{ov}(X_k, X_\ell)| \; \|\nabla \phi_1\|_{L^2(Q-k)} \, \|\nabla \phi_1\|_{L^2(Q-\ell)}$$

$$\leq \sum_{|k| \geq N+1} \sum_{|\ell| \geq N+1} |\mathbb{C}\text{ov}(X_k, X_\ell)| \; \|\nabla \phi_1\|^2_{L^2(Q-k)},$$

where we have used at the last line the discrete Cauchy-Schwarz inequality between the sequences $|\mathbb{C}\text{ov}(X_k, X_\ell)|^{1/2} \; \|\nabla \phi_1\|_{L^2(Q-k)}$ and $|\mathbb{C}\text{ov}(X_k, X_\ell)|^{1/2} \; \|\nabla \phi_1\|_{L^2(Q-\ell)}$. We next write, using the stationarity of X_k and (36), that

$$\left\| \sum_{|k| \geq N+1} \overline{X}_k \nabla \phi_1(\cdot - k) \right\|^2_{L^2(Q \times \Omega)}$$

$$\leq \sum_{|k| \geq N+1} \|\nabla \phi_1\|^2_{L^2(Q-k)} \sum_{|\ell| \geq N+1} |\mathbb{C}\text{ov}(X_k, X_\ell)|$$

$$\leq \mathcal{C} \sum_{|k| \geq N+1} \|\nabla \phi_1\|^2_{L^2(Q-k)}.$$

50

The above right-hand side converges to 0 as $N \to \infty$ since $\nabla \phi_1 \in \left(L^2(\mathbb{R}^d)\right)^d$.

Hence, the right-hand side of (37) defines a function $T \in \left(L^2(Q \times \Omega)\right)^d$. As $\partial_i T_j = \partial_j T_i$, there exists a function $\widetilde{u}_1$ such that

$$\nabla \widetilde{u}_1 = T = \mathbb{E}[X_0]\nabla \overline{u_1} + \sum_{k \in \mathbb{Z}^d} \Big(X_k(\omega) - \mathbb{E}[X_k] \Big) \nabla \phi_1(\cdot - k).$$

As $\overline{u_1}$ is $\mathbb{Z}^d$-periodic, we infer from the above equality that

$$\nabla \widetilde{u}_1 \text{ is stationary and } \int_Q \mathbb{E}(\nabla \widetilde{u}_1) = 0. \tag{73}$$

Next, we compute

$$C_0 \nabla \widetilde{u}_1 = \mathbb{E}[X_0] C_0 \nabla \overline{u_1} + \sum_{k \in \mathbb{Z}^d} \Big(X_k(\omega) - \mathbb{E}[X_k] \Big) C_0 \nabla \phi_1(\cdot - k).$$

Taking the divergence of this equation and using (32) and (34), we thus find that, in the distribution sense,

$$\begin{aligned} &- \operatorname{div}\left[C_0 \nabla \widetilde{u}_1\right] \\ = \; &\sum_{k \in \mathbb{Z}^d} -\Big(X_k(\omega) - \mathbb{E}[X_k] \Big) \operatorname{div}\left[C_0 \nabla \phi_1(\cdot - k)\right] - \mathbb{E}[X_0] \operatorname{div}\left[C_0 \nabla \overline{u_1}\right] \\ = \; &\sum_{k \in \mathbb{Z}^d} \Big(X_k(\omega) - \mathbb{E}[X_k] \Big) \operatorname{div}\left[\mathbb{1}_{Q+k} C_1 p\right] + \mathbb{E}[X_0] \operatorname{div}\left[C_1 p\right] \\ = \; &\operatorname{div}\left[\chi(\cdot, \omega) C_1 p\right]. \end{aligned} \tag{74}$$

Collecting (73) and (74), we see that $\widetilde{u}_1$ solves (19). As the solution to this equation is unique up to the addition of a (possibly random) constant $C(\omega)$, we obtain that $\widetilde{u}_1 = u_1 + C(\omega)$, hence proving (37).

References

[1] A. Anantharaman, R. Costaouec, C. Le Bris, F. Legoll, and F. Thomines. Introduction to numerical stochastic homogenization and the related computational challenges: some recent developments. In W. Bao and Q. Du, editors, *Multiscale modeling and analysis for materials simulation*, volume 22 of *Lect. Notes Series, Institute for Mathematical Sciences, National University of Singapore*, pages 197–272. World Sci. Publ., Hackensack, NJ, 2011.

[2] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic methods in periodic structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-New York, 1978.

[3] C. Bernardin and S. Olla. *Thermodynamics and non-equilibrium macroscopic dynamics of chains of anharmonic oscillators*. 2014. Lecture notes available at `https://www.ceremade.dauphine.fr/~olla`.

[4] X. Blanc, R. Costaouec, C. Le Bris, and F. Legoll. Variance reduction in stochastic homogenization: the technique of antithetic variables. In B. Engquist, O. Runborg, and R. Tsai, editors, *Numerical Analysis of Multiscale Computations*, volume 82 of *Lect. Notes Comput. Sci. Eng.*, pages 47–70. Springer, 2012.

[5] X. Blanc, R. Costaouec, C. Le Bris, and F. Legoll. Variance reduction in stochastic homogenization using antithetic variables. *Markov Processes and Related Fields*, 18(1):31–66, 2012. (preliminary version available at `http://cermics.enpc.fr/~legoll/hdr/FL24.pdf`).

[6] A. Bourgeat and A. Piatnitski. Approximations of effective coefficients in stochastic homogenization. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 40(2):153–165, 2004.

[7] D. Cioranescu and P. Donato. *An introduction to homogenization*, volume 17 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1999.

[8] R. Costaouec. Asymptotic expansion of the homogenized matrix in two weakly stochastic homogenization settings. *Applied Mathematics Research eXpress*, 2012(1):76–104, 2012.

[9] R. Costaouec, C. Le Bris, and F. Legoll. Variance reduction in stochastic homogenization: proof of concept, using antithetic variables. *Boletin Soc. Esp. Mat. Apl.*, 50:9–27, 2010.

[10] A. Dembo and O. Zeitouni. Refinements of the Gibbs conditioning principle. *Probability Theory and Related Fields*, 104(1):1–14, 1996.

[11] B. Engquist and P. E. Souganidis. Asymptotic and numerical homogenization. *Acta Numerica*, 17:147–190, 2008.

[12] 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.

[13] A. Gloria and F. Otto. Quantitative estimates on the periodic approximation of the corrector in stochastic homogenization. *ESAIM: Proc.*, 48:80–97, 2015.

[14] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, Berlin, 1994.

[15] S. M. Kozlov. Averaging of random structures. *USSR Doklady*, 241(5):1016–1019, 1978.

[16] C. Le Bris. Some numerical approaches for "weakly" random homogenization. In G. Kreiss, P. Lötstedt, A. Malqvist, and M. Neytcheva, editors, *Numerical Mathematics and Advanced Applications*, Proceedings of ENUMATH 2009, Lect. Notes Comput. Sci. Eng., pages 29–45. Springer, 2010.

[17] F. Legoll and W. Minvielle. A control variate approach based on a defect-type theory for variance reduction in stochastic homogenization. *Multiscale Modeling & Simulation*, 13(2):519–550, 2015.

[18] F. Legoll and W. Minvielle. Variance reduction using antithetic variables for a nonlinear convex stochastic homogenization problem. *Discrete and Continuous Dynamical Systems - Series S*, 8(1):1–27, 2015.

[19] W. Minvielle. *Quelques problèmes liés à l'erreur statistique en homogénéisation stochastique*. PhD thesis, Université Paris-Est, 2015. (available at http://cermics.enpc.fr/~minvielw/Thesis_manuscript.pdf).

[20] J. Nolen. Normal approximation for a random elliptic equation. *Probability Theory and Related Fields*, 159(3-4):661–700, 2014.

[21] G. C. Papanicolaou and S. R. S. Varadhan. Boundary value problems with rapidly oscillating random coefficients. In J. Fritz, J. L. Lebaritz, and D. Szasz, editors, *Proc. Colloq. on Random fields: Rigorous results in statistical mechanics and quantum field theory*, volume 10 of *Colloq. Math. Soc. János Bolyai*, pages 835–873. North-Holland, Amsterdam-New York, 1981.

[22] A. N. Shiryayev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1984.

[23] J. von Pezold, A. Dick, M. Friák, and J. Neugebauer. Generation and performance of special quasirandom structures for studying the elastic properties of random alloys: Application to Al-Ti. *Physical Review B*, 81(9):094203, 2010.

[24] S.-H. Wei, L. G. Ferreira, J. E. Bernard, and A. Zunger. Electronic properties of random alloys: Special quasirandom structures. *Physical Review B*, 42(15):9622, 1990.

[25] A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard. Special quasirandom structures. *Physical Review Letters*, 65(3):353, 1990.