



HAL
open science

Numerical approximation of effective coefficients in stochastic homogenization of discrete elliptic equations

Antoine Gloria

► **To cite this version:**

Antoine Gloria. Numerical approximation of effective coefficients in stochastic homogenization of discrete elliptic equations. [Research Report] 2010, pp.37. inria-00510514v1

HAL Id: inria-00510514

<https://inria.hal.science/inria-00510514v1>

Submitted on 19 Aug 2010 (v1), last revised 18 Dec 2012 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NUMERICAL APPROXIMATION OF EFFECTIVE COEFFICIENTS IN STOCHASTIC HOMOGENIZATION OF DISCRETE ELLIPTIC EQUATIONS.

ANTOINE GLORIA

Abstract. We introduce and analyze a numerical strategy to approximate effective coefficients in stochastic homogenization of discrete elliptic equations. In particular, we consider the simplest case possible: An elliptic equation on the d -dimensional lattice \mathbb{Z}^d with independent and identically distributed conductivities on the associated edges. Recent results by Otto and the author quantify the error made by approximating the homogenized coefficient by the averaged energy of a regularized corrector (with parameter T) on some box of finite size L . In this article, we replace the regularized corrector (which is the solution of a problem posed on \mathbb{Z}^d) by some practically computable proxy on some box of size $R \geq L$, and quantify the associated additional error. In order to improve the convergence, one may also consider N independent realizations of the computable proxy, and take the arithmetic mean of the associated approximate homogenized coefficients. A natural optimization problem consists in properly choosing T, R, L and N in order to reduce the error at given computational complexity. Our analysis is sharp and allows us to give a clear answer to this question. In particular, we propose and analyze a numerical algorithm to approximate the homogenized coefficients, taking advantage of the (nearly) optimal scalings of the errors we derive. The efficiency of the approach is illustrated by a numerical study in dimension 2.

Keywords: stochastic homogenization, effective coefficients, difference operator, numerical method.

2010 Mathematics Subject Classification: 35B27, 39A70, 60H25, 65N99 .

1. INTRODUCTION

In this article, we continue the analysis begun with Otto in [9] & [10] on stochastic homogenization of discrete elliptic equations. More precisely, we consider real functions u of the sites x in a d -dimensional Cartesian lattice \mathbb{Z}^d . Every edge e of the lattice is endowed with a “conductivity” $a(e) > 0$. This defines a discrete elliptic differential operator $-\nabla^* \cdot A \nabla$ via

$$-\nabla^* \cdot (A \nabla u)(x) := \sum_y a(e)(u(x) - u(y)),$$

where the sum is over the $2d$ sites y which are connected by an edge $e = [x, y] = [y, x]$ to the site x . We assume the conductivities a to be uniformly elliptic in the sense of

$$\alpha \leq a(e) \leq \beta \quad \text{for all edges } e$$

for some fixed constants $0 < \alpha \leq \beta < \infty$.

Date: August 10, 2010.

We are interested in random coefficients. To fix ideas, we consider the simplest situation possible:

$$\{a(e)\}_e \quad \text{are independently and identically distributed (i. i. d.).}$$

Hence the statistics are described by a distribution on the finite interval $[\alpha, \beta]$.

Classical results in stochastic homogenization of linear elliptic equations (see [14] and [20] for the continuous case, and [16] and [15] for the discrete case) state that there exist *homogeneous and deterministic* coefficients A_{hom} such that the solution operator of the continuum differential operator $-\nabla \cdot A_{\text{hom}} \nabla$ describes the large scale behavior of the solution operator of the discrete differential operator $-\nabla^* \cdot A \nabla$. As a by product of this homogenization result, one obtains a characterization of the homogenized coefficients A_{hom} : It is shown that for every direction $\xi \in \mathbb{R}^d$, there exists a unique scalar field ϕ such that $\nabla \phi$ is stationary (stationarity means that the fields $\nabla \phi(\cdot)$ and $\nabla \phi(\cdot + z)$ have the same statistics for all shifts $z \in \mathbb{Z}^d$) and $\langle \nabla \phi \rangle = 0$, solving the equation

$$-\nabla^* \cdot (A(\xi + \nabla \phi)) = 0 \quad \text{in } \mathbb{Z}^d, \quad (1.1)$$

and normalized by $\phi(0) = 0$. As in periodic homogenization, the function $\mathbb{Z}^d \ni x \mapsto \xi \cdot x + \phi(x)$ can be seen as the A -harmonic function which macroscopically behaves as the affine function $\mathbb{Z}^d \ni x \mapsto \xi \cdot x$. With this ‘‘corrector’’ ϕ , the homogenized coefficients A_{hom} (which in general form a symmetric matrix and for our simple statistics in fact a multiple of the identity: $A_{\text{hom}} = a_{\text{hom}} \text{Id}$) can be characterized as follows:

$$\xi \cdot A_{\text{hom}} \xi = \langle (\xi + \nabla \phi) \cdot A(\xi + \nabla \phi) \rangle. \quad (1.2)$$

Since the scalar field $(\xi + \nabla \phi) \cdot A(\xi + \nabla \phi)$ is stationary, it does not matter (in terms of the distribution) at which site x it is evaluated in the formula (1.2), so that we suppress the argument x in our notation.

When one is interested in explicit values for A_{hom} , one has to solve (1.1) and compute (1.2). Since this is not possible in practice, one has to make approximations. For a discussion of the literature on error estimates, in particular the pertinent work by Yurinskii [22] and Naddaf & Spencer [18], we refer to [9, Section 1.2]. As recalled in [10], a standard approach used in practice consists in solving (1.1) in a box $Q_L = [-L, L]^d$ with periodic boundary conditions

$$-\nabla^* \cdot (A(\xi + \nabla \phi_{L,\#})) = 0 \quad \text{in } Q_L, \quad (1.3)$$

and replacing (1.2) by a space average

$$\xi \cdot A_{L,\#} \xi = \int_{Q_L} (\xi + \nabla \phi_{L,\#}) \cdot A(\xi + \nabla \phi_{L,\#}). \quad (1.4)$$

Such an approach is consistent in the sense that

$$\lim_{L \rightarrow \infty} A_{L,\#} = A_{\text{hom}}$$

almost surely, as proved in [19] for both the continuous and discrete cases (see also [3] and [4]). Numerical experiments tend to show that the use of periodic boundary conditions gives better results than other choices such as homogeneous Dirichlet boundary conditions, see [21]. As argued in [10], the error analysis for $\langle |A_{L,\#} - A_{\text{hom}}|^2 \rangle^{1/2}$ is however not obvious a priori since $\nabla \phi$ and $\nabla \phi_{L,\#}$ are not jointly stationary. In [10], we have followed

a somewhat different route by considering the standard regularization of (1.1) to prove existence of correctors. In particular, we have introduced a zero-order term in (1.1):

$$T^{-1}\phi_T - \nabla^* \cdot A(\xi + \nabla\phi_T) = 0 \quad \text{in } \mathbb{Z}^d. \quad (1.5)$$

The advantage of (1.5) for the analysis is that $\nabla\phi$ and $\nabla\phi_T$ are jointly stationary and solve an equation of the same type as (1.1) and (1.5):

$$-\nabla^* \cdot A(\nabla\phi - \nabla\phi_T) = T^{-1}\phi_T \quad \text{in } \mathbb{Z}^d.$$

This has been of great help to estimate $|A_T - A_{\text{hom}}|$ via $\langle |\nabla\phi_T - \nabla\phi|^2 \rangle^{1/2}$ in [10, Theorem 1], where

$$\xi \cdot A_T \xi := \langle (\xi + \nabla\phi_T) \cdot A(\xi + \nabla\phi_T) \rangle. \quad (1.6)$$

Yet, the defining equation (1.5) for ϕ_T is still posed on the whole space \mathbb{Z}^d , which is a handicap for the numerical practice.

Turning back to the idea leading to (1.3), one may approximate the regularized corrector ϕ_T by solving (1.5) in a box $Q_R = [-R, R]^d$, $R \geq L$, with periodic boundary conditions

$$T^{-1}\phi_{T,R,\#} - \nabla^* \cdot A(\xi + \nabla\phi_{T,R,\#}) = 0 \quad \text{in } Q_R, \quad (1.7)$$

and replace (1.6) by

$$\xi \cdot A_{T,R,L,\#} \xi := \int_{\mathbb{Z}^d} (\xi + \nabla\phi_{T,R,\#}) \cdot A(\xi + \nabla\phi_{T,R,\#}) \mu_L,$$

where μ_L is a suitable mask with support in Q_L (see Theorem 1). As opposed to the case without the zero-order term, estimating $|\nabla\phi_T - \nabla\phi_{T,R,\#}|$ in a box Q_L is made easy if $R - L \gg \sqrt{T}$ due to the exponential decay of the Green function associated with $T^{-1} - \nabla^* \cdot A \nabla$ (see Lemma 3). Therefore, $\langle |A_{T,R,L,\#} - A_{T,L}|^2 \rangle^{1/2}$ is expected to be of infinite order in terms of $\frac{R-L}{\sqrt{T}}$, where

$$\xi \cdot A_{T,L} \xi := \int_{\mathbb{Z}^d} (\xi + \nabla\phi_T) \cdot A(\xi + \nabla\phi_T) \mu_L.$$

One crucial feature of the zero-order term is to make the dependence of $\nabla\phi_{T,L,\#}$ upon the boundary value be exponentially small in terms of the distance to the boundary measured in units of \sqrt{T} . Hence, although the zero-order term in (1.5) has been introduced for the ‘‘convenience’’ of the analysis, it turns out that such a term is also very pertinent from the numerical point of view, as further illustrated at the end of this article in our discrete stochastic case. Even in the much more studied continuous periodic case (for which the addition of a zero-order term is not needed for the analysis), such a term yields a striking improvement of the order of convergence for the approximation of the homogenized coefficients (see [8]). Let us now give the argument to conclude the numerical analysis. To pass from an estimate on $\langle |A_{T,R,L,\#} - A_{T,L}|^2 \rangle^{1/2}$ to an estimate on $\langle |A_{T,R,L,\#} - A_{\text{hom}}|^2 \rangle^{1/2}$, we use the triangle inequality in the form

$$\begin{aligned} & \langle |A_{T,R,L,\#} - A_{\text{hom}}|^2 \rangle^{1/2} \\ & \leq \langle |A_{T,R,L,\#} - A_{T,L}|^2 \rangle^{1/2} + \langle |A_{T,L} - A_T|^2 \rangle^{1/2} + \langle |A_T - A_{\text{hom}}|^2 \rangle^{1/2}, \end{aligned}$$

and then appeal to [9, Theorem 2.1] to deal with the second term of the r. h. s. (which is the variance of $A_{T,L}$), and to [10, Theorem 1] for the last term (which is the systematic error due to the zero-order perturbation).

Since the zero-order term reduces the dependence of the solution upon the boundary conditions far from the boundary, the precise nature of the boundary conditions is somewhat irrelevant (in contrast to the numerical evidence in [21] without the zero-order term). Hence, one may safely replace the periodic boundary conditions of (1.7) by homogeneous Dirichlet boundary conditions, i. e. without changing the order of convergence of the method. For the numerical practice the use of Dirichlet boundary conditions is an advantage (sparsity of the matrix, efficient preconditioner, and so on). In this article we will therefore focus on the following proxy for ϕ_T in the box Q_R : The unique solution $\phi_{T,R}$ to

$$\begin{cases} T^{-1}\phi_{T,R} - \nabla^* \cdot A(\xi + \nabla\phi_{T,R}) = 0 & \text{in } Q_R, \\ \phi_{T,R} = 0 & \text{in } \mathbb{Z}^d \setminus Q_R, \end{cases}$$

and define an approximation $A_{T,R,L}$ of A_{hom} by

$$\xi \cdot A_{T,R,L}\xi = \int_{\mathbb{Z}^d} (\xi + \nabla\phi_{T,R}) \cdot A(\xi + \nabla\phi_{T,R})\mu_L. \quad (1.8)$$

As we shall prove in Theorem 1, there exists $c > 0$ (depending only on d and the ellipticity constants α, β) such that for all $R \sim L \sim R - L$,

$$\langle |A_{T,R,L} - A_{T,L}|^2 \rangle^{1/2} \lesssim T^{3/4} \exp\left(-c \frac{R-L}{\sqrt{T}}\right).$$

In combination with [9, Theorem 2.1] and [10, Theorem 1] (see the argument hereafter), this yields the following complete error estimate

$$\langle |A_{T,R,L} - A_{\text{hom}}|^2 \rangle^{1/2} \lesssim \begin{cases} d = 2 & : L^{-1} \ln^q L, \\ d = 3 & : L^{-3/2}, \\ d = 4 & : L^{-2} \ln L, \\ d > 4 & : L^{-2}, \end{cases}$$

provided in addition $\sqrt{T} \ll R$. This result is the first sharp estimate of the convergence rate in stochastic homogenization for $d > 1$ (suboptimal estimates for $d > 2$ were obtained in [7] and [3] using Yurinskii's results in [22]).

In the applied mechanics community, the periodization approach is usually combined with an empirical average: N independent realizations $\{A_{L,\#,k}\}_{k \in \{1, \dots, N\}}$ of $A_{L,\#}$ are computed, and A_{hom} is approximated by

$$A_{L,\#}^N = \frac{1}{N} \sum_{k=1}^N A_{L,\#,k}.$$

Some numerical experiments on such a method with partial conclusions are reported on in [13]. Proceeding the same way, we may consider N independent realizations of $\{A_k\}_{k \in \{1, \dots, N\}}$ of $A|_{Q_R}$ and approximate A_{hom} by

$$A_{T,R,L}^N = \frac{1}{N} \sum_{k=1}^N A_{T,R,L,k},$$

where $A_{T,R,L,k}$ is the approximation (1.8) for the realization A_k of A , $k \in \{1, \dots, N\}$.

An important question for practical purposes is to quantify the dependence of the error $\langle |A_{T,R,L}^N - A_{\text{hom}}|^2 \rangle^{1/2}$ in terms of T, R, L and N . Let us first neglect the fact that ϕ_T has to be approximated on a finite box, and detail some more the error we make by

approximating A_{hom} by $A_{T,L}^N$. As shown in [9, Introduction] when $N = 1$ (the argument does depend on N), the error is made of two contributions, a “systematic error” and a “random error”:

$$\begin{aligned} \langle |A_{T,R,L}^N - A_{\text{hom}}|^2 \rangle &= \underbrace{\langle (\nabla \phi_T - \nabla \phi) \cdot A(\nabla \phi_T - \nabla \phi) \rangle^2}_{=:\text{Error}_{\text{sys}}(T)^2} \\ &+ \underbrace{\left\langle \left(\langle (\xi + \nabla \phi_T) \cdot A(\xi + \nabla \phi_T) \rangle - \frac{1}{N} \sum_{k=1}^N \int_{\mathbb{Z}^d} (\xi + \nabla \phi_{T,k}) \cdot A_k(\xi + \nabla \phi_{T,k}) dx \right)^2 \right\rangle}_{=:\text{Error}_{\text{rand}}(T, L, N)^2}. \end{aligned}$$

The systematic error has been estimated in [10, Theorem 1]

$$\text{Error}_{\text{sys}}(T) \lesssim \begin{cases} d = 2 & : T^{-1} \ln^q T, \\ d = 3 & : T^{-3/2}, \\ d = 4 & : T^{-2} \ln T, \\ d > 4 & : T^{-2}, \end{cases} \quad (1.9)$$

for some q depending only on α, β , where “ \lesssim ” stands for “ \leq ” up to a multiplicative constant depending only on α, β , and d . It vanishes when $T \uparrow \infty$. Using the independence of the $\phi_{T,k}$, we may rewrite the random error as

$$\begin{aligned} &\text{Error}_{\text{rand}}(T, L, N)^2 \\ &= \left\langle \left(\frac{1}{N} \sum_{k=1}^N \left(\left\langle \int_{\mathbb{Z}^d} (\xi + \nabla \phi_{T,k}) \cdot A_k(\xi + \nabla \phi_{T,k}) \mu_L dx \right\rangle - \int_{\mathbb{Z}^d} (\xi + \nabla \phi_{T,k}) \cdot A_k(\xi + \nabla \phi_{T,k}) \mu_L dx \right) \right)^2 \right\rangle \\ &= \frac{1}{N} \text{var} \left[\int_{\mathbb{Z}^d} (\xi + \nabla \phi_T) \cdot A(\xi + \nabla \phi_T) \mu_L dx \right] \end{aligned} \quad (1.10)$$

It measures the fluctuations of the energy density. This error vanishes as $L \uparrow \infty$, but also when the number of realizations $N \uparrow \infty$. In [9, Theorem 2.1], we have proved that

$$\text{var} \left[\int_{\mathbb{Z}^d} (\xi + \nabla \phi_T) \cdot A(\xi + \nabla \phi_T) \mu_L dx \right]^{1/2} \lesssim \begin{cases} d = 2 & : L^{-1} \ln^q T, \\ d > 2 & : L^{-d/2}. \end{cases} \quad (1.11)$$

Hence,

$$\text{Error}_{\text{rand}}(T, L, N) \lesssim \begin{cases} d = 2 & : (N^{1/2}L)^{-1} \ln^q T, \\ d > 2 & : (N^{1/d}L)^{-d/2}. \end{cases}$$

where q only depends on the coercivity constants. These two estimates then give a clear suggestion on the choice of N and L (and therefore T via $T = L^2$ at first order). There are two extreme strategies:

- (A) compute $N_A = N$ realizations of a system of size $L_A = L, T_A = L^2$;
- (B) compute $N_B = 1$ realization of a system of size $L_B = N^{1/d}L, T_B = N^{2/d}L^2$.

For both methods, the number of unknowns is $M = NL^d$ and the overall errors are:

$$\begin{aligned} \text{Error}_A(N, M) &:= \text{Error}_{\text{sys}}(L^2) + \text{Error}_{\text{rand}}(L^2, L, N) \\ &\lesssim \begin{cases} d = 2 &: NM^{-1} \ln^q(NM^{-1}) + M^{-1/2} \ln^q M, \\ d = 3 &: NM^{-1} + M^{-1/2}, \\ d = 4 &: NM^{-1} \ln(NM^{-1}) + M^{-1/2}, \\ d > 4 &: (NM^{-1})^{4/d} + M^{-1/2}, \end{cases} \end{aligned}$$

and

$$\begin{aligned} \text{Error}_B(M) &:= \text{Error}_{\text{sys}}(N^{2/d}L^2) + \text{Error}_{\text{rand}}(N^{2/d}L^2, N^{1/d}L, 1) \\ &\lesssim \begin{cases} d = 2 &: M^{-1/2} \ln^q M, \\ 2 < d \leq 8 &: M^{-1/2}, \\ d > 8 &: M^{-4/d}. \end{cases} \end{aligned}$$

For $d = 2, 3, 4$ and $N = \sqrt{M}$, $\text{Error}_A(\sqrt{M}, M)$ and $\text{Error}_B(M)$ are of the same order (up to a logarithm for $d = 4$). Since the cost of solving a linear system is superlinear in terms of the number of unknowns, it is more favourable to solve N systems of L^d unknowns than 1 system of NL^d unknowns. Hence, strategy (A) may be preferable for $d \leq 4$. *In particular, for $d \leq 4$ it seems best to evenly split a given number M of unknowns into the number N of realizations and the number L^d of unknowns per realization, i. e. $N = L^d = \sqrt{M}$.*

Yet, this coarse complexity analysis does not take into account the fact that one needs to approximate ϕ_T on Q_R with $R - L \gg \sqrt{T}$, and not on Q_L . With this respect, the error analysis we develop in this paper (see Theorem 1) is precise enough to solve the nontrivial optimization process for the choice of the repartition of the M unknowns into N subproblems of size $L_N = (N^{-1}M)^{1/d}$, at fixed convergence rate.

The article is organized as follows: In Section 2, we introduce the general framework and state the main result of this paper, i. e. a sharp estimate of $\langle |A_{T,R,L}^N - A_{\text{hom}}|^2 \rangle^{1/2}$, whose proof is the object of Section 3. In Section 4, we take advantage of this error analysis to address the optimization of the number N of subproblems given a fixed total number M of unknowns. This allows us in particular to check the sharpness of our result on a two-dimensional example.

To conclude this introduction, let us mention that we'd like to see the discrete stochastic elliptic operator under investigation here as a good model problem for continuum elliptic operators with random coefficients of correlation length unity. The extension of the results of [9, 10] to the continuous stochastic case with mixing coefficients is in progress, [11]. In particular, the adaptation of the strategy, results, and proofs of this article to the continuous case is straightforward provided the mentioned extensions.

Throughout the paper, we make use of the following notation:

- $d \geq 2$ is the dimension;
- $\int_{\mathbb{Z}^d} dx$ denotes the sum over $x \in \mathbb{Z}^d$, and $\int_D dx$ denotes the sum over $x \in \mathbb{Z}^d$ such that $x \in D$, D of \mathbb{R}^d ;
- $\langle \cdot \rangle$ is the ensemble average, or equivalently the expectation in the underlying probability space;
- $\text{var}[\cdot]$ is the variance associated with the ensemble average;

- \lesssim and \gtrsim stand for \leq and \geq up to a multiplicative constant which only depends on the dimension d and the constants α, β (see Definition 1 below) if not otherwise stated;
- when both \lesssim and \gtrsim hold, we simply write \sim ;
- we use \gg instead of \gtrsim when the multiplicative constant is (much) larger than 1;
- $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ denotes the canonical basis of \mathbb{Z}^d .

2. MAIN RESULT

2.1. General framework.

Definition 1. We say that a is a conductivity function if there exist $0 < \alpha \leq \beta < \infty$ such that for every edge e of the square lattice generated by \mathbb{Z}^d , one has $a(e) \in [\alpha, \beta]$. We denote by $\mathcal{A}_{\alpha\beta}$ the set of such conductivity functions.

Definition 2. The elliptic operator L associated with a conductivity function $a \in \mathcal{A}_{\alpha\beta}$ is defined for all $u : \mathbb{Z}^d \rightarrow \mathbb{R}$ and $x \in \mathbb{Z}^d$ by

$$(Lu)(x) = -\nabla^* \cdot A(x) \nabla u(x) \quad (2.1)$$

where

$$\nabla u(x) := \begin{bmatrix} u(x + \mathbf{e}_1) - u(x) \\ \vdots \\ u(x + \mathbf{e}_d) - u(x) \end{bmatrix}, \quad \nabla^* u(x) := \begin{bmatrix} u(x) - u(x - \mathbf{e}_1) \\ \vdots \\ u(x) - u(x - \mathbf{e}_d) \end{bmatrix},$$

and

$$A(x) := \text{diag}[a(e_1), \dots, a(e_d)],$$

$$e_1 = [x, x + \mathbf{e}_1], \dots, e_d = [x, x + \mathbf{e}_d].$$

We now turn to the definition of the statistics of the conductivity function.

Definition 3. A conductivity function is said to be independent and identically distributed (i. i. d.) if the coefficients $a(e)$ are i. i. d. random variables.

Definition 4. The conductivity matrix A is obviously stationary in the sense that for all $z \in \mathbb{Z}^d$, $A(\cdot + z)$ and $A(\cdot)$ have the same statistics, so that for all $x, z \in \mathbb{Z}^d$,

$$\langle A(x + z) \rangle = \langle A(x) \rangle.$$

Therefore, any translation invariant function of A , such as the modified corrector ϕ_T (see Lemma 2), is jointly stationary with A . In particular, not only are ϕ_T and its gradient $\nabla \phi_T$ stationary, but also any function of A , ϕ_T and $\nabla \phi_T$. A useful such example is the energy density $(\xi + \nabla \phi_T) \cdot A(\xi + \nabla \phi_T)$, which is stationary by joint stationarity of A and $\nabla \phi_T$.

Lemma 1 (corrector). [16, Theorem 3] *Let $a \in \mathcal{A}_{\alpha\beta}$ be an i. i. d. conductivity function, then for all $\xi \in \mathbb{R}^d$, there exists a unique random function $\phi : \mathbb{Z}^d \rightarrow \mathbb{R}$ which satisfies the corrector equation*

$$-\nabla^* \cdot A(x) (\xi + \nabla \phi(x)) = 0 \quad \text{in } \mathbb{Z}^d, \quad (2.2)$$

and such that $\phi(0) = 0$, $\nabla \phi$ is stationary and $\langle \nabla \phi \rangle = 0$. In addition, $\langle |\nabla \phi|^2 \rangle \lesssim |\xi|^2$.

We also define an ‘‘approximation’’ of the corrector as follows:

Lemma 2 (approximate corrector). [16, Proof of Theorem 3] *Let $a \in \mathcal{A}_{\alpha\beta}$ be an i. i. d. conductivity function, then for all $T > 0$ and $\xi \in \mathbb{R}^d$, there exists a unique stationary random function $\phi_T : \mathbb{Z}^d \rightarrow \mathbb{R}$ which satisfies the “approximate” corrector equation*

$$T^{-1}\phi_T(x) - \nabla^* \cdot A(x) (\xi + \nabla\phi_T(x)) = 0 \quad \text{in } \mathbb{Z}^d, \quad (2.3)$$

and such that $\langle \phi_T \rangle = 0$. In addition, $T^{-1} \langle \phi_T^2 \rangle + \langle |\nabla\phi_T|^2 \rangle \lesssim |\xi|^2$.

Definition 5 (homogenized coefficients). Let $a \in \mathcal{A}_{\alpha\beta}$ be an i. i. d. conductivity function and let $\xi \in \mathbb{R}^d$ and ϕ be as in Lemma 1. We define the homogenized $d \times d$ -matrix A_{hom} as

$$\xi \cdot A_{\text{hom}} \xi = \langle (\xi + \nabla\phi) \cdot A(\xi + \nabla\phi) \rangle. \quad (2.4)$$

Note that (2.4) fully characterizes A_{hom} since A_{hom} is a symmetric matrix (it is in particular of the form $a_{\text{hom}}\text{Id}$ for an i. i. d. conductivity function).

2.2. Statement of the main result. We replace ϕ_T by the computable function $\phi_{T,R}$, which is an approximation of ϕ_T on a bounded domain of size $2R$. Let $a \in \mathcal{A}_{\alpha\beta}$, $T > 0$, $R \gg 1$ and $\xi \in \mathbb{R}^d$. We set $Q_R := [-R, R]^d \cap \mathbb{Z}^d$ and we let $\phi_{T,R}$ be the solution in $L^2(\mathbb{Z}^d)$ to

$$\begin{cases} T^{-1}\phi_{T,R} - \nabla^* \cdot A(\xi + \nabla\phi_{T,R}) = 0 & \text{in } Q_R, \\ \phi_{T,R} = 0 & \text{on } \mathbb{Z}^d \setminus Q_R. \end{cases} \quad (2.5)$$

We then quantify the error we make by replacing ϕ_T by the computable $\phi_{T,R}$. It is given by the following:

Proposition 1. *Let $a \in \mathcal{A}_{\alpha\beta}$ be an i. i. d. conductivity function, $T > 0$, $R \gg 1$ and $\xi \in \mathbb{R}^d$, $|\xi| = 1$. Let ϕ_T denote the approximate corrector, and $\phi_{T,R}$ be the solution of (2.5). Then there exists $c > 0$ depending only on α, β , and d , such that for all $R - L \geq \sqrt{T}$, we have*

$$\int_{Q_L} |\nabla\phi_{T,R}(x) - \nabla\phi_T(x)|^2 dx \lesssim L^d \left(\left(\frac{R}{R-L} \right)^{d-1/2} T^{3/4} \exp\left(-c \frac{R-L}{\sqrt{T}}\right) \right)^2. \quad (2.6)$$

From this proposition, we deduce the main result of this paper.

Theorem 1. *Let $a \in \mathcal{A}_{\alpha\beta}$ be an i. i. d. conductivity function, $T > 0$, $R \gg 1$ and $\xi \in \mathbb{R}^d$, $|\xi| = 1$. Let $\{\phi_{T,R,k}\}_{k=1,\dots,N}$ be the solutions of (2.5) with $N \geq 1$ independent realizations A_k of A , and A_{hom} be the homogenized matrix. For all $R - L \geq \sqrt{T}$, we denote by $\mu_L : \mathbb{Z}^d \rightarrow [0, 1]$ a mask such that $\int_{\mathbb{Z}^d} \mu_L(x) dx = 1$, $|\nabla\mu_L(x)| \lesssim L^{-d-1}$ and $\text{supp}(\mu_L) \subset Q_L$. For all $k \in \{1, \dots, N\}$ we define*

$$\xi \cdot A_{T,R,L,k} \xi = \int_{\mathbb{Z}^d} (\xi + \nabla\phi_{T,R,k}(x)) \cdot A_k(x) (\xi + \nabla\phi_{T,R,k}(x)) \mu_L(x) dx,$$

and set

$$\xi \cdot A_{T,R,L}^N \xi := \xi \cdot \left(\frac{1}{N} \sum_{k=1}^N A_{T,R,L,k} \right) \xi.$$

Then, we have the following error estimate

$$\begin{aligned} \left\langle (\xi \cdot A_{T,R,L}^N \xi - \xi \cdot A_{\text{hom}} \xi)^2 \right\rangle^{1/2} &\lesssim \left(\frac{R}{L} \right)^{d/2} \left(\frac{R}{R-L} \right)^{d-1/2} T^{3/4} \exp \left(-c \frac{R-L}{\sqrt{T}} \right) \\ &+ \begin{cases} d=2 & : (T^{-1} + (NL^2)^{-1/2}) (\ln T)^q \\ d=3 & : T^{-3/2} + (NL^3)^{-1/2} \\ d=4 & : T^{-2} \ln T + (NL^4)^{-1/2} \\ d>4 & : T^{-2} + (NL^d)^{-1/2} \end{cases} \end{aligned} \quad (2.7)$$

for some q depending only on α, β , and some c depending further on d .

Let us apply Theorem 1 to the two extreme strategies described in the introduction. We focus on the case $d \leq 4$, and let M denote a reference number of unknowns. To apply Theorem 1 to Strategy (B), we choose the following set of parameters:

$$\begin{cases} N_B = 1, \\ 2L_B = M^{1/d}, \\ T_B = M^{1/d}, \\ 2R_B = 2 \left(1 + \frac{\ln^2(2L_B)}{\sqrt{2L_B}} \right) L_B. \end{cases} \quad (2.8)$$

This choice of parameters amounts to solving one equation with $(2R_B)^d \simeq M$ unknowns. Theorem 1 then provides the (nearly) optimal error estimate in dimensions 2, 3 & 4:

$$\left\langle (\xi \cdot A_{T_B, R_B, L_B} \xi - \xi \cdot A_{\text{hom}} \xi)^2 \right\rangle^{1/2} \lesssim \begin{cases} d=2 & : M^{-1/2} (\ln M)^q, \\ d=3 & : M^{-1/2}, \\ d=4 & : M^{-1/2} \ln M, \end{cases} \quad (2.9)$$

since for all $\gamma > 0$

$$\exp(-c \ln^2 M) = o(M^{-\gamma}).$$

For Strategy (A), the set of parameters is now

$$\begin{cases} N_A = \sqrt{M} \ln^{-2d} M, \\ 2L_A = \sqrt{M}^{1/d} \ln^2 M, \\ T_A = M^{1/d}, \\ 2R_A = 3\sqrt{M}^{1/d} \ln^2 M. \end{cases} \quad (2.10)$$

This choice of parameters amounts to solving N_A equations with $(2R_A)^d$ unknowns each, which gives a total number of $3^d M$ unknowns. Theorem 1 then provides the (nearly) optimal error estimate in dimensions 2, 3 & 4:

$$\left\langle (\xi \cdot A_{T_A, R_A, L_A}^{N_A} \xi - \xi \cdot A_{\text{hom}} \xi)^2 \right\rangle^{1/2} \lesssim \begin{cases} d=2 & : M^{-1/2} (\ln M)^q, \\ d=3 & : M^{-1/2}, \\ d=4 & : M^{-1/2} \ln M. \end{cases} \quad (2.11)$$

Compared to the informal statement of introduction, the effect of the boundary conditions makes Strategy (A) slightly more expansive than expected (3^d times as many unknowns), and the comparison of Strategies (A) and (B) not so clear. Yet, Strategy (A) is intrinsically parallelizable, which is a big advantage. The optimization of N (and R, L) in (2.7) at fixed complexity and fixed rate of convergence is the object of Section 4.

3. PROOFS OF THE RESULTS

We define discrete Green's functions as follows:

Definition 6 (discrete Green's function). Let $d \geq 2$. For all $T > 0$, the Green function $G_T : \mathcal{A}_{\alpha\beta} \times \mathbb{Z}^d \times \mathbb{Z}^d \rightarrow \mathbb{Z}^d$, $(a, x, y) \mapsto G_T(x, y; a)$ associated with the conductivity function a is defined for all $y \in \mathbb{Z}^d$ and $a \in \mathcal{A}_{\alpha\beta}$ as the unique solution in $L_x^2(\mathbb{Z}^d)$ to

$$\int_{\mathbb{Z}^d} T^{-1} G_T(x, y; a) v(x) dx + \int_{\mathbb{Z}^d} \nabla v(x) \cdot A(x) \nabla_x G_T(x, y; a) dx = v(y), \quad \forall v \in L^2(\mathbb{Z}^d), \quad (3.1)$$

where A is as in (2.1).

Throughout this paper, we use the shorthand notation $G_T(x, y)$ for $G_T(x, y; a)$.

3.1. Proof of Proposition 1. This proof is inspired by the analysis by Bourgeat & Piatnitski in [3], that we adapt here to the discrete setting. In order to prove Proposition 1, we need the optimal pointwise decay of the Green function G_T and a uniform bound on the approximate corrector field ϕ_T . These are given by the following two auxiliary lemmas.

Lemma 3 (optimal pointwise decay estimates). *There exists $c > 0$ depending only on α, β , and d , such that for all $a \in \mathcal{A}_{\alpha\beta}$ and $T > 0$, the Green function G_T satisfies the pointwise estimates: There exists $c > 0$ depending only α, β , and d , such that for all $x, y \in \mathbb{Z}^d$,*

$$\text{for } d > 2 : G_T(x, y) \lesssim (1 + |x - y|)^{2-d} \exp\left(-c \frac{|x - y|}{\sqrt{T}}\right), \quad (3.2)$$

$$\text{for } d = 2 : G_T(x, y) \lesssim \ln\left(\frac{\sqrt{T}}{1 + |x - y|}\right) \exp\left(-c \frac{|x - y|}{\sqrt{T}}\right). \quad (3.3)$$

Lemma 4. *Let $a \in \mathcal{A}_{\alpha\beta}$, $T \gg 1$, and $\xi \in \mathbb{R}^d$, $|\xi| = 1$. The approximate corrector ϕ_T satisfies the following uniform bound*

$$\sup_{\mathbb{Z}^d} |\phi_T| \lesssim \sqrt{T}. \quad (3.4)$$

Note that this bound is sharper than the one used for the continuous case in [3, Formula (25)]. We first prove Proposition 1, and then turn to Lemma 4. The proof of Lemma 3 is postponed to Appendix B.

Proof of Proposition 1. We divide the proof in two steps.

Step 1. Proof of the estimate

$$|\phi_T(x) - \phi_{T,R}(x)| \lesssim \left(\frac{R}{\rho}\right)^{d-1/2} T^{3/4} \exp\left(-c \frac{\rho}{\sqrt{T}}\right) \quad (3.5)$$

for all $x \in Q_{R-\rho}$, with $\rho \sim R \sim R - \rho \gtrsim \sqrt{T}$, and some c depending only on d and the ellipticity constants α, β .

The function $\phi_T - \phi_{T,R}$ is solution to

$$\begin{cases} T^{-1}(\phi_T - \phi_{T,R}) - \nabla^* \cdot A(\nabla(\phi_T - \phi_{T,R})) = 0 & \text{in } Q_R, \\ \phi_T - \phi_{T,R} = \phi_T & \text{on } \mathbb{Z}^d \setminus Q_R. \end{cases} \quad (3.6)$$

Let φ_0 denote the trivial lifting of $\phi_T|_{\mathbb{Z}^d \setminus Q_R}$ in Q_R :

$$\begin{cases} \varphi_0(x) = 0 & \text{in } Q_R, \\ \varphi_0(x) = \phi_T(x) & \text{on } \mathbb{Z}^d \setminus Q_R. \end{cases}$$

We may then write $\phi_T - \phi_{T,R} = \varphi_0 + \varphi_1$, where φ_1 is the solution to

$$\begin{cases} T^{-1}\varphi_1 - \nabla^* \cdot A \nabla \varphi_1 = \underbrace{-T^{-1}\varphi_0 + \nabla^* \cdot A \nabla \varphi_0}_{=: \tilde{\varphi}_0} & \text{in } Q_R, \\ \varphi_1 = 0 & \text{on } \mathbb{Z}^d \setminus Q_R. \end{cases} \quad (3.7)$$

Note that

$$\begin{cases} \nabla \varphi_0(x) = 0 & \text{for } x \in Q_R, d(x, \mathbb{Z}^d \setminus Q_R) \geq 2, \\ \sup_{x \in Q_R} |\varphi_0(x)| \lesssim \sup_{(3.4)} |\phi_T| \lesssim \sqrt{T}. \end{cases} \quad (3.8)$$

We next use the Green representation formula. To this aim, we define the Green function $G_{T,R}(\cdot, y) : \mathbb{Z}^d \rightarrow \mathbb{R}$ for all $y \in Q_R$ as the unique solution to

$$\begin{cases} T^{-1}G_{T,R}(x, y) - \nabla^* \cdot A(\nabla G_{T,R}(x, y)) = \delta(y - x) & \text{in } Q_R, \\ G_{T,R}(x, y) = 0 & \text{on } \mathbb{Z}^d \setminus Q_R. \end{cases} \quad (3.9)$$

By the maximum principle, for all $x \in Q_R$ we have

$$0 \leq G_{T,R}(x, y) \leq G_T(x, y). \quad (3.10)$$

For all $x \in Q_R$, we then rewrite $\varphi_1(x)$ as

$$\begin{aligned} \varphi_1(x) &= \int_{Q_R} G_{T,R}(x, y) \tilde{\varphi}_0(y) dy \\ &= - \int_{Q_R} G_{T,R}(x, y) T^{-1} \varphi_0(y) dy + \int_{Q_R} G_{T,R}(x, y) \nabla^* \cdot A(y) \nabla \varphi_0(y) dy. \end{aligned}$$

By integration by parts (recall that $G_{T,R}(x, y) = 0$ for $y \in \mathbb{Z}^d \setminus Q_R$) and using the fact that φ_0 is supported on the boundary of Q_R , this yields

$$\varphi_1(x) = -T^{-1} \int_{Q_R \setminus Q_{R-1}} G_{T,R}(x, y) \varphi_0(y) dy - \int_{Q_R \setminus Q_{R-2}} \nabla_y G_{T,R}(x, y) \cdot A(y) \nabla \varphi_0(y) dy. \quad (3.11)$$

We first deal with the first term of the r. h. s. of (3.11), that we estimate for all $x \in Q_{R-\rho}$ using the uniform bound (3.8) on φ_0 , the maximum principle (3.10), and the optimal pointwise estimates on G_T from Lemma 3:

$$\begin{aligned} & \left| \int_{Q_R \setminus Q_{R-1}} G_{T,R}(x, y) \varphi_0(y) dy \right| \\ & \lesssim R^{d-1} (1 + d(x, \mathbb{Z}^d \setminus Q_R))^{2-d} \exp\left(-c \frac{d(x, \mathbb{Z}^d \setminus Q_R)}{\sqrt{T}}\right) \sqrt{T} \\ & \leq \left(\frac{R}{\rho}\right)^{d-1} \frac{\rho}{\sqrt{T}} \exp\left(-c \frac{\rho}{\sqrt{T}}\right) T. \end{aligned} \quad (3.12)$$

For the second term of the r. h. s. of (3.11), we use Cauchy-Schwarz' inequality, the boundedness of A , and (3.8) in the form of the uniform bound $\sup |\nabla \varphi_0| \lesssim \sup |\varphi_0| \lesssim \sqrt{T}$

to obtain

$$\left| \int_{Q_R \setminus Q_{R-2}} \nabla_y G_{T,R}(x,y) \cdot A(y) \nabla \varphi_0(y) dy \right| \lesssim R^{(d-1)/2} \sqrt{T} \left(\int_{Q_R \setminus Q_{R-2}} |\nabla_y G_{T,R}(x,y)|^2 dy \right)^{1/2}. \quad (3.13)$$

We then appeal to Cacciopoli's inequality. To this aim, we recall that $\rho \sim R \sim R - \rho$, and we let $\eta_\rho : Q_R \rightarrow [0, 1]$ be a cut-off function such that

$$\begin{aligned} \text{for } y \in Q_{R-\rho/2} & : \eta_\rho(y) = 0, \\ \text{for } y \in Q_R \setminus Q_{R-2} & : \eta_\rho(y) = 1, \\ \text{for } y \in Q_R & : |\nabla \eta_\rho(y)| \lesssim \rho^{-1}. \end{aligned} \quad (3.14)$$

Since $G_{T,R} = 0$ on $\mathbb{Z}^d \setminus Q_R$, multiplying the defining equation (3.9) for $G_{T,R}$ by $\eta_\rho^2 G_{T,R}$, and integrating by parts on Q_R yield the following discrete Cacciopoli estimate (see [9, Proof of Lemma 2.8, Step 1] for details)

$$\int_{Q_R} \eta_\rho^2(y) |\nabla G_{T,R}(x,y)|^2 dy \lesssim \int_{Q_R} G_{T,R}^2(x,y) |\nabla \eta_\rho(y)|^2 dy,$$

provided that $x \in Q_{R-\rho}$. By the properties (3.14) of η_ρ , this implies

$$\int_{Q_R \setminus Q_{R-2}} |\nabla G_{T,R}(x,y)|^2 dy \lesssim \rho^{-2} \int_{Q_R \setminus Q_{R-\rho/2}} G_{T,R}^2(x,y) dy. \quad (3.15)$$

We are now in position to estimate (3.13) for all $x \in Q_{R-\rho}$. By the Cacciopoli estimate (3.15), the maximum principle (3.10), and the optimal pointwise estimates on G_T from Lemma 3, (3.13) turns into

$$\begin{aligned} & \left| \int_{Q_R \setminus Q_{R-2}} \nabla_y G_{T,R}(x,y) \cdot A(y) \nabla \varphi_0(y) dy \right| \\ & \lesssim R^{(d-1)/2} \left(\rho^{-2} R^d \rho^{2(2-d)} \exp\left(-2c \frac{\rho/2}{\sqrt{T}}\right) \right)^{1/2} \sqrt{T} \\ & = \left(\frac{R}{\rho} \right)^{d-1/2} T^{3/4} \left(\frac{\rho}{\sqrt{T}} \right)^{1/2} \exp\left(-c \frac{\rho}{2\sqrt{T}}\right) \end{aligned} \quad (3.16)$$

for all $x \in Q_{R-\rho}$.

The combination of (3.11), (3.12), (3.16), and the definition of φ_1 shows (3.5) for some constant $c > 0$ depending only on d , and α, β .

Step 2. Proof of (2.6).

We first bound $|\nabla \phi_T(x)|$ by $\sum_{i=1}^d |\phi_T(x)| + |\phi_T(x + \mathbf{e}_i)|$ and integrate inequality (3.5) over Q_L for $L = R - \rho \sim R \sim R - L$, obtaining

$$\int_{Q_L} |\nabla \phi_{T,R}(x) - \nabla \phi_T(x)|^2 dx \lesssim L^d \left(\left(\frac{R}{R-L} \right)^{d-1/2} T^{3/4} \exp\left(-c \frac{R-L}{\sqrt{T}}\right) \right)^2,$$

as desired. \square

Proof of Lemma 4. We start with the Green representation formula, and perform an integration by parts using that G_T is in $L^1(\mathbb{Z}^d)$ by [9, Corollary 2.2]:

$$\begin{aligned}\phi_T(x) &= \int_{\mathbb{Z}^d} G_T(x, y) \nabla^* \cdot A(z) \xi \, dy \\ &= - \int_{\mathbb{Z}^d} \nabla G_T(x, y) \cdot A(z) \xi \, dy.\end{aligned}$$

This yields

$$|\phi_T(x)| \lesssim \int_{|x-y| \leq \sqrt{T}} |\nabla G_T(x, y)| \, dy + \int_{|x-y| > \sqrt{T}} |\nabla G_T(x, y)| \, dy. \quad (3.17)$$

We begin with the second term of the r. h. s. of (3.17). We divide the integration on $\{y : |x-y| > \sqrt{T}\}$ as the integration on annuli of the form $\{y : 2^i \sqrt{T} < |x-y| \leq 2^{i+1} \sqrt{T}\}$ for $i \in \mathbb{N}$, and appeal to the optimal decay of ∇G_T on such annuli from [9, Lemma 2.9] for $q = 2$ and $k = 4$. This yields by Cauchy-Schwarz' inequality

$$\begin{aligned}& \int_{|x-y| > \sqrt{T}} |\nabla G_T(x, y)| \, dy \\ & \leq \sum_{i \in \mathbb{N}} \left(\int_{2^i \sqrt{T} < |x-y| \leq 2^{i+1} \sqrt{T}} |\nabla G_T(x, y)|^2 \, dy \right)^{1/2} (2^i \sqrt{T})^{d/2} \\ & \stackrel{[9, \text{Lemma 2.9}]}{\lesssim} \sum_{i \in \mathbb{N}} \left((2^i \sqrt{T})^{d+2(1-d)} (2^i)^{-4} \right)^{1/2} (2^i \sqrt{T})^{d/2} \\ & = \sqrt{T} \sum_{i \in \mathbb{N}} (2^i)^{-1} \lesssim \sqrt{T}.\end{aligned} \quad (3.18)$$

For the first term of the r. h. s. of (3.17), we also make use of a dyadic decomposition of space. Let $R = 2^{-I} \sqrt{T} \sim 1$, $I \in \mathbb{N}$, be such that [9, Lemma 2.9] applies on annuli of the form $\{y : 2^{-i-1} \sqrt{T} < |x-y| \leq 2^{-i} \sqrt{T}\}$ for $i \leq I-1$ (R has to be large enough although of order unity). We then split the integration on $\{y : |x-y| \leq \sqrt{T}\}$ as the integration on the ball of radius $R \sim 1$, and the integration over annuli of the form $\{y : 2^{-i-1} \sqrt{T} < |x-y| \leq 2^{-i} \sqrt{T}\}$ for $i \in \{0, \dots, I-1\}$. For the integral on the ball, we appeal to the uniform estimate $|\nabla G_T| \lesssim 1$ from [9, Corollary 2.3], and for the integrals

on the annuli, we appeal once more to the optimal decay of [9, Lemma 2.9]. By Cauchy-Schwarz' inequality, this yields

$$\begin{aligned}
& \int_{|x-y| \leq \sqrt{T}} |\nabla G_T(x, y)| dy \\
& \lesssim \int_{|x-y| \leq R} |\nabla G_T(x, y)| dy \\
& \quad + \sum_{i=0}^{I-1} \left(\int_{2^{-i-1}\sqrt{T} < |x-y| \leq 2^{-i}\sqrt{T}} |\nabla G_T(x, y)|^2 dy \right)^{1/2} (2^{-i}\sqrt{T})^{d/2} \\
& \lesssim 1 + \sum_{i=0}^{I-1} \left((2^{-i}\sqrt{T})^{d+2(1-d)} \right)^{1/2} (2^{-i}\sqrt{T})^{d/2} \\
& = 1 + \sqrt{T} \sum_{i=0}^{I-1} 2^{-i} \lesssim \sqrt{T}. \tag{3.19}
\end{aligned}$$

The claim of the lemma now follows from the combination of (3.17), (3.18), and (3.19). \square

3.2. Proof of Theorem 1. To prove Theorem 1, we combine the variance estimate of [9, Theorem 2.1] and the estimate of the systematic error in [10, Theorem 1] with Proposition 1.

We start with the triangle inequality

$$\begin{aligned}
& \left\langle \left(\frac{1}{N} \sum_{k=1}^N \int_{\mathbb{Z}^d} (\xi + \nabla \phi_{T,R,k}) \cdot A_k (\xi + \nabla \phi_{T,R,k}) \mu_L - \xi \cdot A_{\text{hom}} \xi \right)^2 \right\rangle^{1/2} \\
& \leq \left\langle \left(\frac{1}{N} \sum_{k=1}^N \int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R,k}) \cdot A_k (\xi + \nabla \phi_{T,R,k}) \right. \right. \right. \\
& \quad \left. \left. \left. - (\xi + \nabla \phi_{T,k}) \cdot A_k (\xi + \nabla \phi_{T,k}) \right) \mu_L \right)^2 \right\rangle^{1/2} \\
& \quad + \left\langle \left(\frac{1}{N} \sum_{k=1}^N \int_{\mathbb{Z}^d} (\xi + \nabla \phi_{T,k}) \cdot A_k (\xi + \nabla \phi_{T,k}) \mu_L - \xi \cdot A_{\text{hom}} \xi \right)^2 \right\rangle^{1/2}. \tag{3.20}
\end{aligned}$$

We first deal with the first term of the r. h. s. of (3.20). To this aim, we expand the square, which yields N^2 terms. By Cauchy-Schwarz' inequality in probability, each term

is bounded by the same single term: For all $k, k' \in \{1, \dots, N\}$,

$$\begin{aligned}
 & \left\langle \left(\int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R,k}) \cdot A_k(\xi + \nabla \phi_{T,R,k}) - (\xi + \nabla \phi_{T,k}) \cdot A_k(\xi + \nabla \phi_{T,k}) \right) \mu_L \right) \right. \\
 & \quad \times \left. \left(\int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R,k'}) \cdot A_{k'}(\xi + \nabla \phi_{T,R,k'}) - (\xi + \nabla \phi_{T,k'}) \cdot A_{k'}(\xi + \nabla \phi_{T,k'}) \right) \mu_L \right) \right\rangle \\
 & \leq \left\langle \left(\int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R,k}) \cdot A_k(\xi + \nabla \phi_{T,R,k}) - (\xi + \nabla \phi_{T,k}) \cdot A_k(\xi + \nabla \phi_{T,k}) \right) \mu_L \right)^2 \right\rangle^{1/2} \\
 & \quad \times \left\langle \left(\int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R,k'}) \cdot A_{k'}(\xi + \nabla \phi_{T,R,k'}) - (\xi + \nabla \phi_{T,k'}) \cdot A_{k'}(\xi + \nabla \phi_{T,k'}) \right) \mu_L \right)^2 \right\rangle^{1/2} \\
 & = \left\langle \left(\int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R}) \cdot A(\xi + \nabla \phi_{T,R}) - (\xi + \nabla \phi_T) \cdot A(\xi + \nabla \phi_T) \right) \mu_L \right)^2 \right\rangle,
 \end{aligned}$$

where we have dropped the subscripts k and k' since the A_k have the same law. From this and the symmetry of A , we deduce

$$\begin{aligned}
 & \left\langle \left(\frac{1}{N} \sum_{k=1}^N \int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R,k}) \cdot A_k(\xi + \nabla \phi_{T,R,k}) \right. \right. \right. \\
 & \quad \left. \left. \left. - (\xi + \nabla \phi_{T,k}) \cdot A_k(\xi + \nabla \phi_{T,k}) \right) \mu_L \right)^2 \right\rangle^{1/2} \\
 & \leq \left\langle \left(\int_{\mathbb{Z}^d} \left((\xi + \nabla \phi_{T,R}) \cdot A(\xi + \nabla \phi_{T,R}) - (\xi + \nabla \phi_T) \cdot A(\xi + \nabla \phi_T) \right) \mu_L \right)^2 \right\rangle^{1/2} \\
 & = \left\langle \left(\int_{\mathbb{Z}^d} (\nabla \phi_{T,R} - \nabla \phi_T) \cdot A(2\xi + \nabla \phi_{T,R} + \nabla \phi_T) \mu_L \right)^2 \right\rangle^{1/2}.
 \end{aligned}$$

To bound this term, we make use of the a priori estimate

$$\int_{\mathbb{Z}^d} |\nabla \phi_{T,R}(x)|^2 dx \lesssim R^d, \tag{3.21}$$

that we obtain by integration by parts after testing (2.5) with $\phi_{T,R}$ itself. We then use Cauchy-Schwarz' inequality in \mathbb{Z}^d , Proposition 1, the properties of μ_L , and the a priori

estimates on $\nabla\phi_T$ and $\nabla\phi_{T,R}$ to bound the r. h. s.

$$\begin{aligned}
& \left\langle \left(\int_{\mathbb{Z}^d} (\nabla\phi_{T,R}(x) - \nabla\phi_T(x)) \cdot A(x) (2\xi + \nabla\phi_{T,R}(x) + \nabla\phi_T(x)) \mu_L(x) dx \right)^2 \right\rangle \\
& \lesssim \left\langle \frac{1}{L^d} \int_{Q_L} |\nabla\phi_{T,R}(x') - \nabla\phi_T(x')|^2 dx' \int_{\mathbb{Z}^d} (1 + |\nabla\phi_{T,R}(x)|^2 + |\nabla\phi_T(x)|^2) \mu_L(x) dx \right\rangle \\
& \stackrel{(2.6)}{\lesssim} \left(\left(\frac{R}{R-L} \right)^{d-1/2} T^{3/4} \exp\left(-c \frac{R-L}{\sqrt{T}}\right) \right)^2 \\
& \quad \times \left(1 + \underbrace{\left\langle \int_{\mathbb{Z}^d} |\nabla\phi_{T,R}(x)|^2 \mu_L(x) dx \right\rangle}_{\stackrel{(3.21)}{\lesssim} \left(\frac{R}{L}\right)^d} + \int_{\mathbb{Z}^d} \underbrace{\langle |\nabla\phi_T|^2 \rangle}_{\stackrel{\text{Lemma 2}}{\lesssim} 1} \mu_L(x) dx \right) \\
& \lesssim \left(\frac{R}{L} \right)^d \left(\left(\frac{R}{R-L} \right)^{d-1/2} T^{3/4} \exp\left(-c \frac{R-L}{\sqrt{T}}\right) \right)^2. \tag{3.22}
\end{aligned}$$

For the second term of the r. h. s. of (3.20), we recall that it can be split into a variance part and a systematic error, using the independence of the A_k (see [10, (1.8)&(1.9)] for details):

$$\begin{aligned}
& \left\langle \left(\frac{1}{N} \sum_{k=1}^N \int_{\mathbb{Z}^d} (\xi + \nabla\phi_{T,k}) \cdot A_k (\xi + \nabla\phi_{T,k}) \mu_L - \xi \cdot A_{\text{hom}} \xi \right)^2 \right\rangle \\
& = \frac{1}{N} \text{var} \left[\int_{\mathbb{Z}^d} (\xi + \nabla\phi_T) \cdot A (\xi + \nabla\phi_T) \mu_L \right] \\
& \quad + \left(\left\langle \int_{\mathbb{Z}^d} (\xi + \nabla\phi_T) \cdot A (\xi + \nabla\phi_T) \mu_L \right\rangle - \xi \cdot A_{\text{hom}} \xi \right)^2.
\end{aligned}$$

For the variance term, we appeal to [9, Theorem 2.1]

$$\text{var} \left[\int_{\mathbb{Z}^d} (\xi + \nabla\phi_T) \cdot A (\xi + \nabla\phi_T) \mu_L \right] \stackrel{[9, \text{Theorem 2.1}]}{\lesssim} \begin{cases} d=2 & : L^{-2} \ln^q T \\ d>2 & : L^{-d} \end{cases},$$

and for the systematic error, we appeal to [10, Theorem 1]

$$\left(\left\langle \int_{\mathbb{Z}^d} (\xi + \nabla\phi_T) \cdot A (\xi + \nabla\phi_T) \mu_L \right\rangle - \xi \cdot A_{\text{hom}} \xi \right)^2 \stackrel{[10, \text{Theorem 1}]}{\lesssim} \begin{cases} d=2 & : T^{-2} \ln^q T \\ d=3 & : T^{-3} \\ d=4 & : T^{-4} \ln^2 T \\ d>4 & : T^{-4} \end{cases},$$

so that

$$\left\langle \left(\frac{1}{N} \sum_{k=1}^N \int_{\mathbb{Z}^d} (\xi + \nabla \phi_{T,k}) \cdot A_k (\xi + \nabla \phi_{T,k}) \mu_L - \xi \cdot A_{\text{hom}} \xi \right)^2 \right\rangle \lesssim \begin{cases} d = 2 & : (N^{-1}L^{-2} + T^{-2}) \ln^q T, \\ d = 3 & : N^{-1}L^{-3} + T^{-3}, \\ d > 4 & : N^{-1}L^{-4} + T^{-4} \ln^2 T, \\ d > 4 & : N^{-1}L^{-d} + T^{-4}. \end{cases} \quad (3.23)$$

The combination of (3.20), (3.22), and (3.23) concludes the proof of (2.7).

4. NUMERICAL STRATEGY AND VALIDATION

In this section, we propose a complexity analysis for the computation of $A_{T,L}^N$. In particular, we identify the number of realizations N_{opt} (and the associated parameters T, R, L) which minimizes the computational cost to approximate A_{hom} at a given precision. *Precision* is understood here as the *scaling of the error* in terms of the “effective number of unknowns” M (see below) of the approximation, as in (2.9) and (2.11) (in particular, we disregard prefactors). The answer depends on the dimension and on the linear solver used. We treat the cases $d = 2, 3$, with a preconditioned conjugate gradient method, and a Cholesky method to solve the linear problems. Whereas the preconditioned conjugate gradient method is the most efficient solution method for this problem, we also provide the analysis of the Cholesky method in view of its application to linear elasticity. Although our techniques of proofs crucially rely on the scalar character of the equation, we believe that the results of this paper are “likely to hold” in the case of linear elasticity considered in [13]. Another application of interest to us is the numerical approximation of the discrete model for rubber studied in [2], which is a nonlinear version of the discrete elliptic equation dealt with here. In those cases, the linear system is ill-conditioned, and direct solvers such as the Cholesky method are to be used. This motivates us to consider direct solvers for the complexity analysis.

In order to illustrate our main result and check the accuracy of this complexity analysis, we have conducted a series of numerical tests on the following problem: $d = 2$, the coefficients a are i. i. d. taking values $\alpha = 1$ and $\beta = 9$ with probability $1/2$. As proved in Appendix A, Dykhne’s formula (see the original paper [6], and the monograph [12, Section 1.5]) holds true in this particular *discrete* case, so that the associated homogenized matrix A_{hom} is given by $A_{\text{hom}} = \sqrt{\alpha\beta} \text{Id} = 3 \text{Id}$. We then apply Strategy (A) and Strategy (B) to this problem and estimate their relative cost. In the last subsection, we compare these two methods to standard approaches used in the literature. We focus in particular on the importance of the zero-order term in the equation from a numerical point of view.

4.1. Complexity analysis. We first set some notation for Strategy (B). Let $L \in \mathbb{N}$. If we take $T = 2L$ and $R = L + b_f \sqrt{T} \ln^2 T$ for some $b_f > 0$ (b_f for buffer zone), Theorem 1 ensures that the error on the approximation of A_{hom} by Strategy (B) is of order $M^{-1/2}$ (up to the logarithmic correction for $d = 2$), with $M := (2L)^d$. Since the associated approximation of A_{hom} is given by a weighted sum of the energy density $(\xi + \nabla \phi_{T,R}(x)) \cdot A(x)(\xi + \nabla \phi_{T,R}(x))$ at exactly $M = (2L)^d$ sites, we will say that M is the “effective number of unknowns”. Note that it differs from the total number of unknowns, which is $(2R)^d > (2L)^d = M$.

As discussed in the introduction of this paper, one may distribute the effective number of unknowns on several smaller independent domains, while keeping the error on the approximation of A_{hom} of order $M^{-1/2}$. This gives rise (but not only) to Strategy (A). Let N be a number of domains. On the one hand, in order to keep the effective number of unknowns fixed, we take $L_N := LN^{-1/d}$. On the other hand, in order to keep the precision unchanged, we still need $T \sim L$, and the buffer zone to be of the same order as for Strategy (B). Hence we set $T_N := T = 2L$, and $R_N := L_N + b_f \sqrt{T} \ln^2 T$. The precision is therefore $M^{-1/2}$, the effective number of unknowns is $N(2L_N)^d = M$, whereas the total number of unknowns is now $N(2R_N)^d > (2R)^d > M$.

In order to make a complexity analysis, one needs to make precise the linear systems to be solved depending on N and M . Let \mathcal{L} denote a symmetric positive definite matrix of order $l \in \mathbb{N}$. Assume further that it has a fixed number of diagonals (in our discrete case: 5 for $d = 2$, 7 for $d = 3$) and a bandwidth $b \leq l$. Then, solving the system

$$\mathcal{L}X = B$$

in \mathbb{R}^l by a conjugate gradient method requires *approximately* $\mathcal{C}_{\text{CG}}(\mathcal{L}) \sim bl$ operations. When solved by a Cholesky method, it *exactly* requires $\mathcal{C}_{\text{Chol}}(\mathcal{L}) \sim b^2 l$ operations.

Let us now use these complexity estimates in the homogenization problem under investigation. For our difference operators, l should be replaced by $(2R_N)^d$ and b by $(2R_N)^{d-1}$. Hence, the overall number of operations to solve the N problems is

$$\begin{aligned} \Gamma_{\text{CG}}(N, 2L) &\sim N \left(2LN^{-1/d} + 2b_f \sqrt{2L} \ln^2(2L) \right)^{2d-1} \\ &= N^{-1+1/d} \left(2L + 2b_f N^{1/d} \sqrt{2L} \ln^2(2L) \right)^{2d-1}, \\ \Gamma_{\text{Chol}}(N, 2L) &\sim N \left(2LN^{-1/d} + 2b_f \sqrt{2L} \ln^2(2L) \right)^{2(d-1)+d} \\ &= N^{-2+2/d} \left(2L + 2b_f N^{1/d} \sqrt{2L} \ln^2(2L) \right)^{3d-2}. \end{aligned}$$

We have essentially two extreme strategies for the choice of N , the effective number of unknowns $N^d(2LN^{-1/d})^d = M$ being fixed. Either we choose N_{opt} which minimizes the cost of solving N problems of size $R_N = 2LN^{-1/d} + b_f \sqrt{2L} \ln^2(2L)$, i. e. such that

$$\Gamma(N_{\text{opt}}, 2L) \leq \Gamma(N, 2L)$$

for all $N \leq 2L$, where Γ denotes a cost function (Γ_{CG} or Γ_{Chol}). Or, given an arbitrary number of processors, we choose N in order to minimize the effective time to solve the N problems using parallel computing (recall that the N problems are completely independent). The first option consists in minimizing the time on one single processor, the second option, on an arbitrary number of processors.

In order to make the complexity analysis concrete, we fix $b_f = 0.1$, since this is the value we use in the numerical tests. We first treat the case $d = 2$, and then the case $d = 3$. Note that the complexity analysis for the Cholesky method is *exact* since the number of operations involved is known a priori. For the conjugate gradient, this is not the case and the number of operations depends on the number of iterations (which in turn does not only depend on the tolerance required but also on the preconditioner used). The present

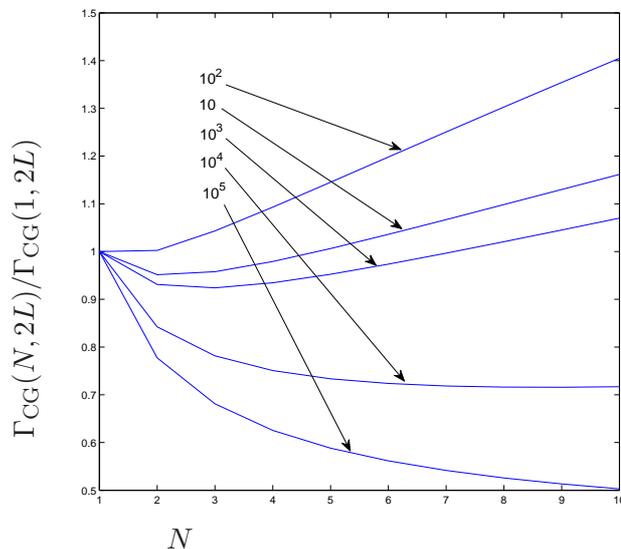


FIGURE 1. Complexity for $d = 2$ with a conjugate gradient method, and $2L = 10^1, \dots, 10^5$.

discussion is therefore only *qualitative* for the conjugate gradient method. Numerical tests will complete the discussion for the conjugate gradient method in dimension 2.

Dimension $d = 2$. The ratio $\Gamma_{\text{CG}}(N, 2L)/\Gamma_{\text{CG}}(1, 2L)$ is plotted on Figure 1 for $N \in \{1, \dots, 10\}$ and $2L \in \{10, 10^2, 10^3, 10^4, 10^5\}$ (that is from 100 to 10^{10} effective unknowns). Except for $2L = 10^2$, this ratio is minimal for some $N_{\text{opt}} \neq 1$. Hence it seems more advantageous to make several computations (say 3 for a typical number unknowns of $M = 10^6$) on smaller domains. For the Cholesky method (see Figure 2 for $N \in \{1, \dots, 100\}$), this is even more clear, and the gain is much larger.

For the second strategy, the objective is to minimize the effective computational time given a fixed number S of processors. Roughly speaking, it is reasonable to take $N \geq S$, whatever the method. There are then two regimes. In the first one (small number of processors), the gain is linear in the number of processors: If the number of processors is doubled, the effective time is divided by two. This is the scalable regime. Eventually, it saturates and the relative gain in effective time decreases. On Figures 3 & 4, the effective time is plotted in function of the number of processors (for which N is optimized) in logarithmic scale, for the two methods and with $2L = 10^3$ (which corresponds to $M = 10^6$ unknowns). The thick line is the effective time, whereas the thin straight line is the perfect scalable regime. It is a lower bound. For the conjugate gradient method, the problem is scalable up to 15 processors, whereas for the Cholesky method, the problem can be considered scalable up to 50 processors.

Dimension $d = 3$. The conclusions are essentially the same as for $d = 2$. The ratios $\Gamma_{\text{CG}}(N, 2L)/\Gamma_{\text{CG}}(1, 2L)$ and $\Gamma_{\text{Chol}}(N, 2L)/\Gamma_{\text{Chol}}(1, 2L)$ are plotted on Figures 5 & 6 respectively, for $N \in \{1, \dots, 100\}$. Note that $N = 1$ is never the optimal choice (especially for the Cholesky method, for which the cost can be reduced by a factor 3 for $M = 10^6$ by taking $N = 31$). The effective time in function of the number of processors is plotted in

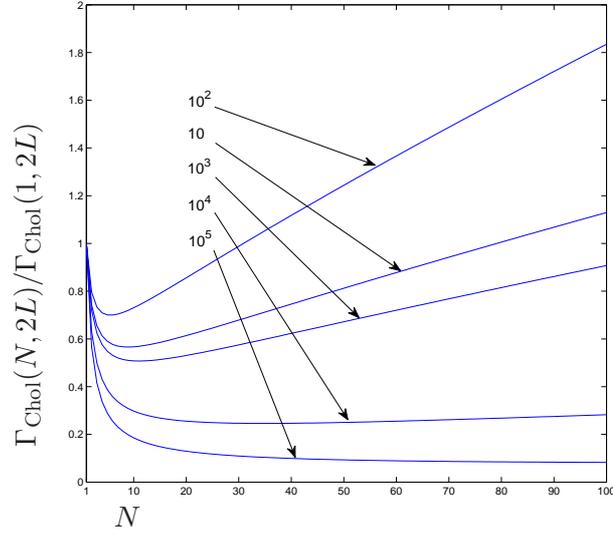


FIGURE 2. Complexity for $d = 2$ with a Cholesky method, and $2L = 10^1, \dots, 10^5$.

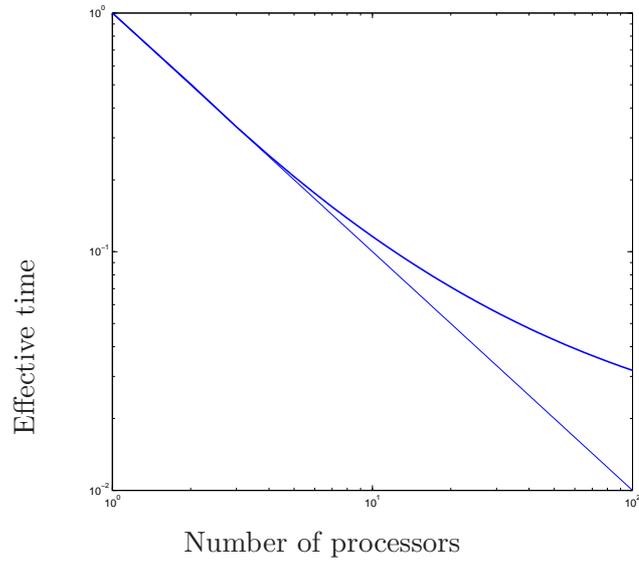


FIGURE 3. Effective time in function of the number of processors for $d = 2$ with a conjugate gradient method (dots), and $2L = 10^3$.

logarithmic scale for $M = 10^6$ unknowns on Figures 7 & 8, for both the conjugate gradient and Cholesky methods. For the conjugate gradient method, although the gain in time remains important, the problem is only scalable up to 8 processors. For the Cholesky method, the method is perfectly scalable in the regime considered.

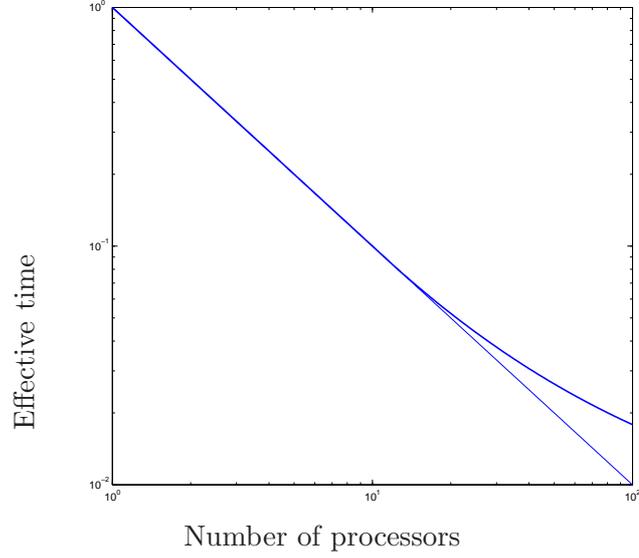


FIGURE 4. Effective time in function of the number of processors for $d = 2$ with a Cholesky method (dots), and $2L = 10^3$.

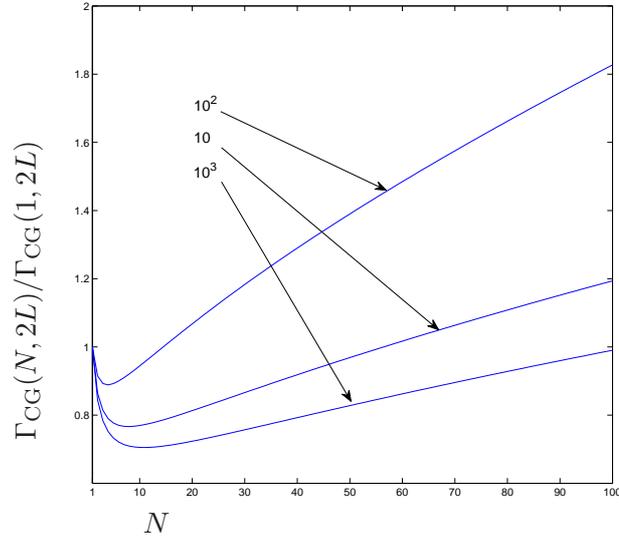


FIGURE 5. Complexity for $d = 3$ with a conjugate gradient method, and $2L = 10^1, 10^2, 10^3$.

4.2. **Numerical tests for Strategy (B).** The first series of tests illustrates (2.9) in Theorem 1 for $N = 1$, with

$$\begin{cases} 2L_B = \sqrt{M} \\ T_B = 2L_B + 3, \\ R_B = \left(1 + 0.1 \frac{\ln^2(2L_B)}{\sqrt{2L_B}}\right) (L_B + 1). \end{cases}$$

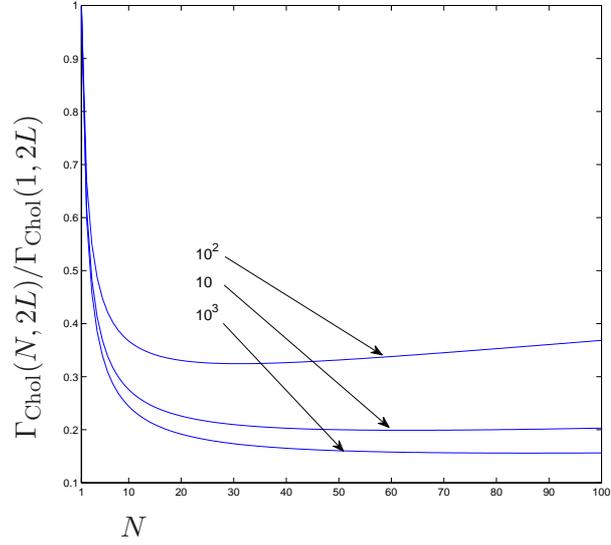


FIGURE 6. Complexity for $d = 3$ with a Cholesky method, and $2L = 10^1, 10^2, 10^3$.

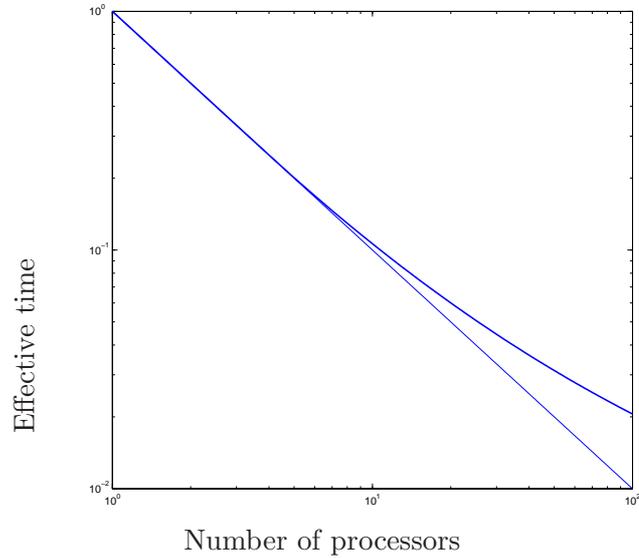


FIGURE 7. Effective time in function of the number of processors for $d = 3$ with a conjugate gradient method (dots), and $2L = 10^2$.

The mask has been chosen as follows:

$$\mu_{L_B}(x) = \tilde{\mu}_{L_B}(x_1)\tilde{\mu}_{L_B}(x_2),$$

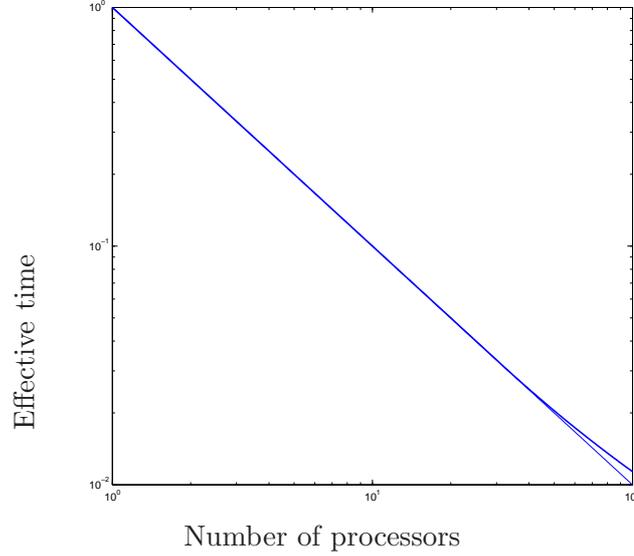


FIGURE 8. Effective time in function of the number of processors for $d = 3$ with a Cholesky method (dots), and $2L = 10^2$.

where $x = (x_1, x_2) \in \mathbb{Z}^2$, and $\tilde{\mu}_{L_B} : \mathbb{Z} \rightarrow \mathbb{R}^+$ is given by

$$\tilde{\mu}_{L_B}(t) = \gamma_{L_B} \begin{cases} L_B \leq |t| & : & 0, \\ \frac{L_B}{3} \leq |t| \leq L_B & : & \frac{3}{2} - \frac{3}{2L_B}|t|, \\ |t| \leq \frac{L_B}{3} & : & 1, \end{cases}$$

and γ_{L_B} is such that $\int_{\mathbb{Z}} \tilde{\mu}_{L_B}(t) dt = 1$. The linear system (2.5) is solved by a preconditioned conjugate gradient method, whose preconditioner is the incomplete Cholesky factorization $IC(2)$ (see [17]). For a uniform sampling of $\log L_B$, $L_B \in [10, 2000]$, we approximate the expectation

$$\left\langle \left(\int_{\mathbb{Z}^d} (\mathbf{e}_1 + \nabla \phi_{T_B, R_B}(x)) \cdot A(x) (\mathbf{e}_1 + \nabla \phi_{T_B, R_B}(x)) \mu_{L_B}(x) dx - 3 \right)^2 \right\rangle$$

by an empirical average over $r(M)$ realizations, and define the error by

Error(M) :=

$$\sqrt{\frac{1}{r(M)} \sum_{j=1}^{r(M)} \left(\int_{\mathbb{Z}^d} (\mathbf{e}_1 + \nabla \phi_{T_B, R_B}^j(x)) \cdot A_j(x) (\mathbf{e}_1 + \nabla \phi_{T_B, R_B}^j(x)) \mu_{L_B}(x) dx - 3 \right)^2}, \quad (4.1)$$

where $\{\phi_{T_B, R_B}^j\}$ are the solutions of (2.5) for the $r(M)$ different realizations A_j of the coefficients A . The number of realizations for each M considered and the associated error (4.1), are reported on in Table 1.

M	1.6E+02	4.0E+02	9.0E+02	2.1E+03	5.6E+03	1.5E+04	4.0E+04
$r(M)$	10000	10000	10000	10000	5000	3000	1600
Error(M)	2.39E-01	1.48E-01	9.35E-02	5.97E-02	3.65E-02	2.28E-02	1.41E-02
M	1.1E+05	3.0E+05	8.1E+05	2.2E+06	6.0E+06	1.6E+07	
$r(M)$	1000	600	360	200	150	100	
Error(M)	9.11E-03	5.95E-03	3.95E-03	2.46E-03	1.55E-03	8.78E-04	

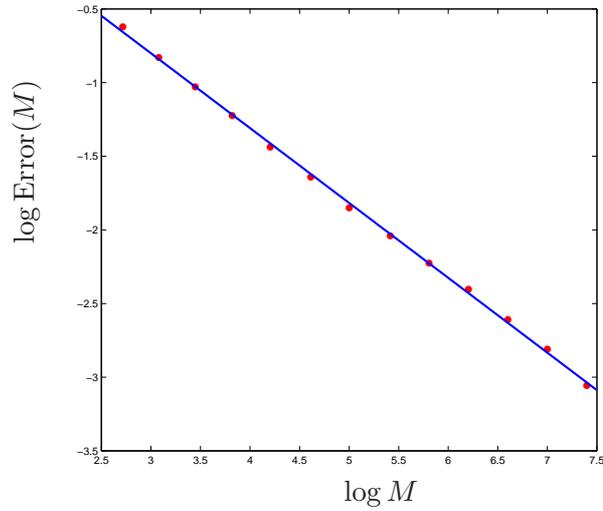
TABLE 1. Error (4.1) for different M , and $r(M)$ realizations.

FIGURE 9. Error (4.1) in log scale

The error is also plotted in function of M in logarithmic scale on Figure 9. The dots (which indicate calculations) are in very good agreement with the straight line of slope $-1/2$ corresponding to the decay provided by Theorem 1 in the form (2.9).

In order to compare a priori Strategies (A) & (B), one needs to know the number of iterations of the conjugate gradient method (for the fixed tolerance 10^{-9}). The number of iterations is plotted in function of the bandwidth of the matrix on Figure 10 (in log-log). The dots represent the numerical experiments, the straight line represents a linear dependence as assumed in Subsection 4.1 (b times a matrix-vector multiplication which costs $O(l)$ operations), whereas the dashed line is a linear fitting of the numerical experiments (equation: $y(x) = 0.64x + 0.36$).

4.3. Numerical tests for Strategy (A). This second series of tests confirms (2.9) in Theorem 1 with $N > 1$. As for Subsection 4.2, we still consider several M , and the parameters $2L_B = \sqrt{M}$, and $T_B = 2L_B + 3$. We distribute the number of unknowns on K^2 subdomains, $K \in \mathbb{N}$. We then set

$$\begin{cases} 2L_A = \sqrt{M}/K, \\ T_A = T_B, \\ R_A = L_A + 1 + 0.1 \ln^2(2L_B) \sqrt{2L_B}, \end{cases}$$

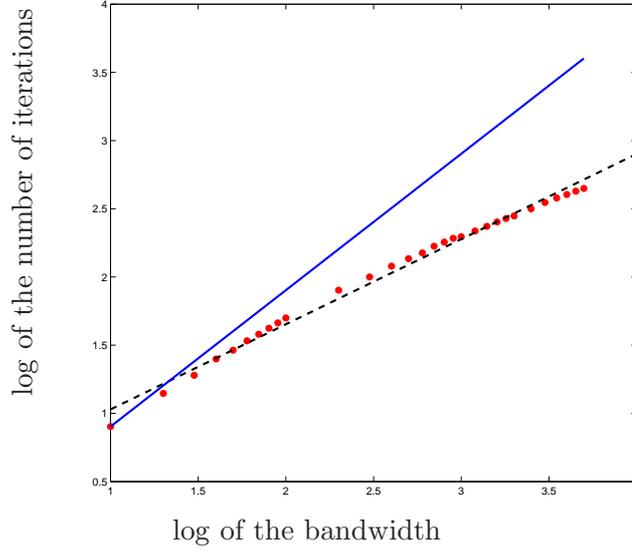


FIGURE 10. Number of iterations of the conjugate gradient method in function of the bandwidth of the matrix

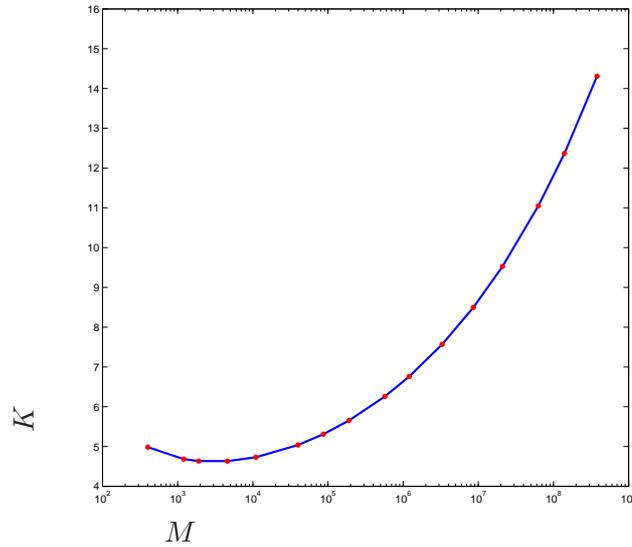


FIGURE 11. Number of subdomains K per dimension in function of the total number of unknowns M according to formula (4.2)

where the number of subdomains is $N_A = K^2$, and K is given by

$$K := \frac{2L_B}{0.1\sqrt{2L_B}\ln^2(2L_B)} = \frac{\sqrt{M}}{0.1M^{1/4}\ln^2\sqrt{M}} = \frac{40M^{1/4}}{\ln^2 M}. \quad (4.2)$$

The values of K are plotted on Figure 11 for M in $[4 \cdot 10^2, 3.8 \cdot 10^8]$. We have also gathered

M	4.0E+02	1.2E+03	1.9E+03	4.6E+03	1.1E+04	4.0E+04	8.7E+04	1.9E+05
$r(M)$	10000	10000	10000	10000	5000	3000	1600	1000
N_A	5×5	5×5	4×4	4×4	4×4	5×5	5×5	5×5
L_A	4	7	11	17	26	40	59	87
R_A	12	21	33	51	78	120	177	261
Error(M)	4.83E-01	2.26E-01	2.02E-01	1.22E-01	8.45E-02	3.41E-02	2.31E-02	1.53E-02
M	5.7E+05	1.2E+06	3.3E+06	8.6E+06	2.1E+07	6.3E+07	1.4E+08	3.8E+08
$r(M)$	600	360	200	150	100	30	10	4
N_A	6×6	6×6	7×7	8×8	9×9	11×11	12×12	14×14
L_A	126	181	258	366	514	720	1001	1358
R_A	378	543	774	1098	1542	2160	3003	4074
Error(M)	8.18E-03	6.13E-03	3.41E-03	2.06E-03	1.28E-03	8.57E-04	5.53E-04	1.69E-04

TABLE 2. Error (4.3) for different M , and $r(M)$ realizations.

in Table 2 the description of the tests in function of the values of M : The sizes of the domain R_A and L_A , the number N_A of such domains, the number of realizations $r(M)$, and the associated error

Error(M) :=

$$\sqrt{\frac{1}{r(M)} \sum_{j=1}^{r(M)} \left(\sum_{k=1}^{N_A} \int_{\mathbb{Z}^d} (\mathbf{e}_1 + \nabla \phi_{T_A, R_A}^{j,k}(x)) \cdot A_{j,k}(x) (\mathbf{e}_1 + \nabla \phi_{T_A, R_A}^{j,k}(x)) \mu_{L_A}(x) dx - 3 \right)^2}, \quad (4.3)$$

where the $\phi_{T_A, R_A}^{j,k}$ are the solutions of (2.5) for the $N_A \times r(M)$ different realizations $A_{j,k}$ of the coefficients A .

As expected, the convergence rate has the scaling of the central limit theorem $-1/2$, as can be seen on Figure 13, where the logarithm of the error is plotted in function of the logarithm of M .

In Table 3, we compare the overall time of computation for Strategy (A) with respect to the reference time of Strategy (B). Prediction 1 is given by our complexity analysis, assuming a number of iterations for the conjugate gradient method of the order of the bandwidth of the matrix. Eventually, as can be seen on Figure 10 this assumption turns out to be pessimistic and the number of iterations does not increase as much as expected using the incomplete Cholesky factorization to precondition the system. Therefore, Prediction 1 in Table 3 underestimates the relative cost of Strategy (A) with respect to Strategy (B). The use of the dashed line in Figure 10 to estimate the number of iterations of the conjugate gradient yields Prediction 2, which is much closer to the numerical experiments.

Note that in the range of effective unknowns considered (M from 10^2 to 10^7), numerical experiments show that $N_{\text{opt}} = 1$ for Strategy (B). The reason is once more the efficiency of the preconditioner. To illustrate this, we have plotted on Figure 12 the ratio $\Gamma_{\text{CG}}(N, 2L)/\Gamma_{\text{CG}}(1, 2L)$ using this time the number of iterations of the conjugate gradient method obtained in the numerical tests for $N \in \{1, \dots, 10\}$. As can be seen, $2L = 10^{3.5}$ (that is $M = 10^7$) is the critical number of unknowns under which $N_{\text{opt}} = 1$ for Strategy (B).

M	4.0E+04	8.7E+04	1.9E+05	5.7E+05	1.2E+06	3.3E+06	8.6E+06	2.1E+07
Numerical test	2.61	2.72	2.12	2.95	2.36	2.59	2.40	3.06
Prediction 1	2.0	1.8	1.7	1.8	1.7	1.6	1.6	1.5
Prediction 2	2.65	2.51	2.36	2.58	2.38	2.46	2.48	2.43

TABLE 3. Ratio of the overall times of computation for Strategies (B) & (A) using a conjugate gradient method: Numerical experiment versus predictions 1 & 2.

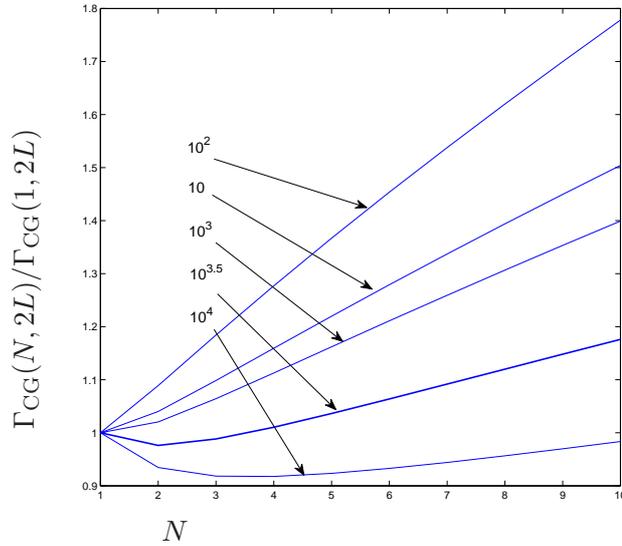


FIGURE 12. Complexity for $d = 2$ with a conjugate gradient method (number of iterations observed numerically), for $2L = 10, 10^2, 10^3, 10^{3.5}$.

Concerning the issue of memory, it is worth noticing that the larger N the less memory needed. Hence, larger numbers of unknowns can be reached using larger N . In the present case, the computation for $3.8 \cdot 10^8$ unknowns could not have been done on one computer without taking $N = 196$.

4.4. Comparison with standard approaches and comments. In order to approximate effective coefficients in stochastic homogenization (see for instance [21] for the continuous case), it is standard to replace the abstract corrector field by the solution to

$$\begin{cases} -\nabla^* \cdot A(\xi + \nabla \phi_R) = 0 & \text{in } Q_R, \\ \phi_R = 0 & \text{on } \mathbb{Z}^d \setminus Q_R, \end{cases} \quad (4.4)$$

that is (2.5) without the zero-order term. Let M be the effective number of unknowns. The formula for the approximation of the homogenized coefficients is then

$$\text{Error}(M) := \sqrt{\frac{1}{r(M)} \sum_{j=1}^{r(M)} \left(\int_{\mathbb{Z}^d} (\mathbf{e}_1 + \nabla \phi_R^j(x)) \cdot A_j(x) (\mathbf{e}_1 + \nabla \phi_R^j(x)) \mu_L(x) dx - 3 \right)^2}, \quad (4.5)$$

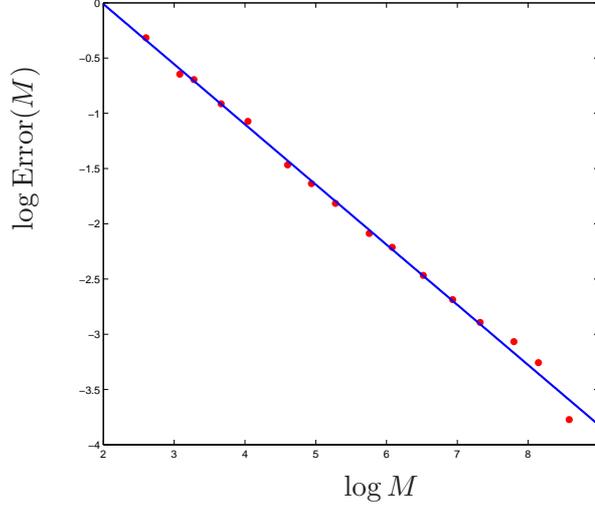


FIGURE 13. Error (4.3) in log scale

where $r(M)$ is the number of independent realizations. Let us make a *formal* error analysis of such an approach.

Assuming that the corrector field ϕ is uniformly bounded (which we do not know a priori since we only control $\langle |\phi|^q \rangle$ for all $q < \infty$ in [9, Proposition 1]), the error we make by replacing ϕ by ϕ_R is due to the use of the Dirichlet boundary conditions, which are not exact. This error typically scales as a surface term in the energy, that is $R^{d-1}/R^d = 1/R$ in any dimension. In dimension 2, the effect due to the boundary conditions and the central limit theorem have the same scaling $R/R^2 = 1/R = M^{-1/2}$. Hence the addition of the zero order term may not be crucial in dimension 2 to obtain the optimal scaling $M^{-1/2}$ (although we are not able to turn this into a rigorous argument). This is confirmed by numerical tests, as illustrated on Figure 14, where the proxy for ϕ is the solution to (4.4).

On the contrary, in dimension 3 (and more), the effect of the boundary conditions now scales as $1/R = M^{-1/3}$, whereas the central limit theorem scaling is still $M^{-1/2} \ll M^{-1/3}$. Hence the use of the zero-order term is crucial to observe the optimal scaling. Another case for which the zero-order term is crucial is Strategy (A), even in dimension 2. If the proxy for ϕ on the domain Q_R is approximated independently on subdomains of size R^γ (with $1 \geq \gamma \geq 1/2$), the error due to the boundary conditions scales as $1/R^\gamma$ without the zero-order term, whereas it will remain of order $1/R$ with the zero-order term.

Compared to [21], another practical advantage of the approach is the type of boundary conditions used in (2.5), that is homogeneous Dirichlet boundary conditions. Compared to periodic boundary conditions (which are “widely recognised” as less perturbative than Dirichlet boundary conditions for the cell problem without the zero-order term), the former has the advantage not to destroy the band structure of the stiffness matrix. On the one hand, periodic boundary conditions change the profile of the matrix, making the triangle matrix in the Cholesky factorization less sparse (only the band structure is preserved by the algorithm) and the factorization more expensive. On the other hand, there is no

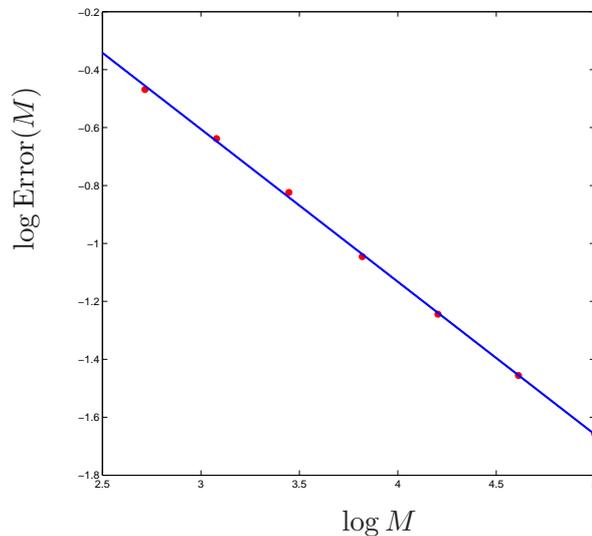


FIGURE 14. Error (4.5) in log scale

optimal preconditioner for periodic boundary conditions. With this respect, the proposed strategy with the zero-order term and Dirichlet boundary conditions allows us to use efficient methods to solve the linear systems, without sacrificing the convergence rate.

As a conclusion, we have proposed and fully analyzed a numerical method to approximate effective coefficients for the stochastic homogenization of discrete elliptic equations. The analysis is sharp, and the numerical method effective. It crucially relies on the introduction of a zero-order term in the corrector equation, which is not only essential for the analysis, but also for the numerical practice.

ACKNOWLEDGEMENTS

The research of the author was partly supported by INRIA, under the grant “Action de Recherche Collaborative” DISCO. It is a pleasure to thank Felix Otto for stimulating discussions on the subject.

APPENDIX A. PROOF OF DYKHNE’S FORMULA IN THE DISCRETE STOCHASTIC CASE

In periodic homogenization of elliptic differential operators in dimension two, Dykhne’s formula is as follows. Let $A : \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 2}$ be a periodic function taking values in a subspace of uniformly bounded and elliptic matrices, and such that $A(x)A(R \cdot x) = \gamma \text{Id}$, where R denotes the rotation by $\pi/2$ in \mathbb{R}^2 and $\gamma > 0$. Then the homogenized matrix associated with A is $A_{\text{hom}} = \sqrt{\gamma} \text{Id}$. This formula is a particular case of general duality relations. Its proof (see for instance [12, Section 1.5]) makes use of the following two facts:

- (i) if $\chi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a gradient field, then $x \mapsto \chi(R \cdot x)$ is divergence-free;
- (ii) if $\chi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is divergence-free, then $x \mapsto \chi(R \cdot x)$ is a gradient field.

In particular, let ϕ be periodic and satisfy the corrector equation $\nabla \cdot A(x) \nabla \phi(x) = 0$. We further set $\xi = \langle \nabla \phi \rangle$ (where $\langle \cdot \rangle$ denotes the average in this periodic case). Since $\nabla \phi$ is a

gradient, (i) implies that

$$\nabla \cdot A(x)A(R \cdot x)\nabla\phi(R \cdot x) = \gamma\nabla \cdot \nabla\phi(R \cdot x) = 0. \quad (\text{A.1})$$

On the other hand $\nabla \cdot A(x)\nabla\phi(x) = 0$ implies by (ii) that $x \mapsto A(R \cdot x)\nabla\phi(R \cdot x)$ is a gradient field, so that, by definition of the homogenized coefficients, (A.1) yields

$$\langle A(x)A(R \cdot x)\nabla\phi(R \cdot x) \rangle = A_{\text{hom}} \langle A(R \cdot x)\nabla\phi(R \cdot x) \rangle.$$

Using now that the average does not change by rotation, this turns into

$$\gamma\xi = \langle A(x)A(R \cdot x)\nabla\phi(R \cdot x) \rangle = A_{\text{hom}} \langle A\nabla\phi \rangle = A_{\text{hom}}^2\xi.$$

from which we deduce the claim by symmetry and coercivity of A_{hom} .

In the two-dimensional discrete case, the following counterparts to (i) and (ii) hold:

- (i') if $\chi : \mathbb{Z}^2 \rightarrow \mathbb{R}^2$ is a gradient field, then $x \mapsto \chi(R \cdot x)$ is divergence free, that is $\nabla \cdot \chi \equiv 0$;
- (ii') if $\chi : \mathbb{Z}^2 \rightarrow \mathbb{R}^2$ is divergence-free, then $x \mapsto \chi(R \cdot x)$ is a gradient field $\nabla\phi$.

However, if one considers ϕ such that $\nabla^* \cdot A(x)\nabla\phi(x) \equiv 0$, one has

$$\nabla^* \cdot A(x)A(R \cdot x)\nabla\phi(R \cdot x) = \gamma\nabla^* \cdot \nabla\phi(R \cdot x) \not\equiv \gamma\nabla \cdot \nabla\phi(R \cdot x) \equiv 0$$

in general. Similarly, $x \mapsto A(R \cdot x)\nabla\phi(R \cdot x)$ is not a gradient field either. Hence, the arguments of the proof do not carry out to difference operators in the periodic case. Actually, counterexamples to Dykhne's formula are easily constructed in this case (see for instance the discrete example in [8, Section 4.1] for which $\gamma = 10 < 26.240099009901\dots = a_{\text{hom}}$).

As opposed to the periodic case, the proof of Dykhne's formula carries out from the stochastic continuous case to the stochastic discrete case, as we quickly show now. Let $d = 2$ and $A \in \mathcal{A}_{\alpha\beta}$ be an i. i. d. conductivity matrix whose entries, denoted by $x \mapsto a_1(x)$ and $x \mapsto a_2(x)$, take values $\alpha > 0$ with probability 1/2 and $\beta > 0$ with probability 1/2. We then introduce the following two auxiliary conductivity matrices: $\bar{A} := \alpha\beta A^{-1}$, and \tilde{A} defined by $\tilde{A} : x \mapsto \alpha\beta \text{diag} [a_2(x + \mathbf{e}_2)^{-1}, a_1(x + \mathbf{e}_1)^{-1}]$. Note that A, \bar{A} and \tilde{A} have the same law, so that they yield the same homogenized matrix A_{hom} . Let then $\xi = \xi_1\mathbf{e}_1 + \xi_2\mathbf{e}_2 \in \mathbb{R}^2$ and $\tilde{\phi}$ be the corrector associated with \tilde{A} and ξ . We introduce the vector field

$$W : x \mapsto \bar{A}(x) \begin{bmatrix} \xi_2 + \nabla_2^* \tilde{\phi}(x) \\ -\xi_1 - \nabla_1^* \tilde{\phi}(x) \end{bmatrix}.$$

The field W satisfies the following three properties.

Property 1.

$$\nabla^* \cdot A(x)W(x) = 0. \quad (\text{A.2})$$

A direct computation actually shows

$$\begin{aligned} \nabla^* \cdot A(x)W(x) &= \nabla^* \cdot A(x)\bar{A}(x) \begin{bmatrix} \xi_2 + \nabla_2^* \tilde{\phi}(x) \\ -\xi_1 - \nabla_1^* \tilde{\phi}(x) \end{bmatrix} \\ &= \alpha\beta \nabla^* \cdot \begin{bmatrix} \nabla_2^* \tilde{\phi}(x) \\ -\nabla_1^* \tilde{\phi}(x) \end{bmatrix} \\ &\equiv 0. \end{aligned}$$

Property 2.

$$\nabla \times W(x) = 0. \quad (\text{A.3})$$

Using the defining equation for the corrector, and the definitions of \bar{A} and \tilde{A} , one has

$$\begin{aligned} \nabla \times W(x) &= \nabla \times \bar{A}(x) \begin{bmatrix} \xi_2 + \nabla_2^* \tilde{\phi}(x) \\ -\xi_1 - \nabla_1^* \tilde{\phi}(x) \end{bmatrix} \\ &= \nabla_2(\bar{a}_1(x)(\xi_2 + \nabla_2^* \tilde{\phi}(x))) + \nabla_1(\bar{a}_2(x)(\xi_1 + \nabla_1^* \tilde{\phi}(x))) \\ &= \nabla_2(\tilde{a}_2(x - \mathbf{e}_2)(\xi_2 + \nabla_2^* \tilde{\phi}(x))) + \nabla_1(\tilde{a}_1(x - \mathbf{e}_1)(\xi_1 + \nabla_1^* \tilde{\phi}(x))) \\ &= \nabla_2^*(\tilde{a}_2(x)(\xi_2 + \nabla_2 \tilde{\phi}(x))) + \nabla_1^*(\tilde{a}_1(x)(\xi_1 + \nabla_1 \tilde{\phi}(x))) \\ &= \nabla^* \cdot \tilde{A}(x)(\xi + \tilde{\phi}) \\ &\equiv 0. \end{aligned}$$

From (A.3) we deduce that $x \mapsto W(x)$ is a gradient field.

Property 3.

$$\langle W(x) \rangle = A_{\text{hom}}(\xi_2 \mathbf{e}_1 - \xi_1 \mathbf{e}_2). \quad (\text{A.4})$$

To prove this property, we use the definitions of \bar{A} and \tilde{A} , and the joint stationarity of \tilde{A} and $\nabla \tilde{\phi}$.

$$\begin{aligned} \langle W(x) \rangle &= \left\langle \bar{A}(x) \begin{bmatrix} \xi_2 + \nabla_2^* \tilde{\phi}(x) \\ -\xi_1 - \nabla_1^* \tilde{\phi}(x) \end{bmatrix} \right\rangle \\ &= \left\langle \tilde{a}_2(x - \mathbf{e}_2)(\xi_2 + \nabla_2^* \tilde{\phi}(x)) \mathbf{e}_1 - \tilde{a}_1(x - \mathbf{e}_1)(\xi_1 + \nabla_1^* \tilde{\phi}(x)) \mathbf{e}_2 \right\rangle \\ &= \left\langle \tilde{a}_2(x)(\xi_2 + \nabla_2 \tilde{\phi}(x)) \mathbf{e}_1 - \tilde{a}_1(x)(\xi_1 + \nabla_1 \tilde{\phi}(x)) \mathbf{e}_2 \right\rangle \\ &= A_{\text{hom}}(\xi_2 \mathbf{e}_1 - \xi_1 \mathbf{e}_2). \end{aligned}$$

In the last equality, we have used that \tilde{a}_1 and \tilde{a}_2 have the same law.

We are now in position to conclude. Since $x \mapsto W(x)$ is a gradient field and satisfies (A.2), one has by definition of the homogenized matrix

$$A_{\text{hom}} \langle W \rangle = \langle AW \rangle. \quad (\text{A.5})$$

We use (A.4) to rewrite the l. h. s. of (A.5) as

$$A_{\text{hom}} \langle W \rangle = A_{\text{hom}}^2(\xi_2 \mathbf{e}_1 - \xi_1 \mathbf{e}_2), \quad (\text{A.6})$$

and the definition of \bar{A} to rewrite the r. h. s. of (A.5) as

$$\langle AW \rangle = \alpha \beta \left\langle \begin{bmatrix} \xi_2 + \nabla_2^* \tilde{\phi} \\ -\xi_1 - \nabla_1^* \tilde{\phi} \end{bmatrix} \right\rangle = \alpha \beta (\xi_2 \mathbf{e}_1 - \xi_1 \mathbf{e}_2), \quad (\text{A.7})$$

by stationarity of $\tilde{\phi}$. The combination of (A.5), (A.6) & (A.7) proves that

$$A_{\text{hom}}^2 = \alpha \beta \text{Id},$$

from which we deduce Dykhne's formula

$$A_{\text{hom}} = \sqrt{\alpha \beta} \text{Id}$$

since A_{hom} is symmetric positive definite.

APPENDIX B. PROOF OF LEMMA 3

The proof of Lemma 3 is a discrete version of the corresponding proof in the continuous case in [8, Appendix]. The discreteness compels us to slightly modify the definitions of $g_{T,j,k}$ and $\chi_{T,j}$ (see below). Likewise, we have to use a discrete Leibniz' rule, which complexifies notation. As for the continuous case, we combine the decay estimates of [9, Lemma 8] with Harnack's inequality and Agmon's positivity method (see [1]). We recall here the Harnack inequality on graphs due to Delmotte (see [5, Proposition 5.3]).

Lemma 5 (Harnack's inequality). *Let $a \in \mathcal{A}_{\alpha\beta}$ and $R \gg 1$. If $g : \mathbb{Z}^d \rightarrow \mathbb{R}^+$ satisfies*

$$-\nabla^* \cdot A \nabla g(x) \leq 0 \quad (\text{B.1})$$

in the annulus $\{R/2 < |x| \leq 4R\}$ (that is g_T is a nonnegative subsolution), then

$$\sup_{R < |x| \leq 2R} g(x) \lesssim \left(R^{-d} \int_{R/2 < |x| \leq 4R} g(x)^2 dx \right)^{1/2}, \quad (\text{B.2})$$

where the multiplicative constant does not depend on R , nor on A .

For $|x - y| \leq \sqrt{T}$, (3.2) & (3.3) coincide with [10, Lemma 5, (2.10) & (2.11)], and we only treat the case $|x - y| \geq \sqrt{T}$.

Step 1. Operator positivity method. We first prove the exponential decay using the classical Leibniz rule, for which the algebra is simpler. We shall deal with the modifications due to the discrete Leibniz rule in the last step of this proof.

For $b > 0$, $j, k \in \mathbb{N}$, let $\tilde{\chi}_{T,j} : \mathbb{Z} \rightarrow [0, 1]$ and $\tilde{g}_{T,j,k} : \mathbb{Z} \rightarrow \mathbb{R}^+$ be given by

$$\tilde{\chi}_{T,j}(x) = \begin{cases} \text{for } |x| \leq 2^j \sqrt{T} & : 0 \\ \text{for } 2^j \sqrt{T} \leq |x| \leq 2^{j+1} \sqrt{T} & : (2^j \sqrt{T})^{-1} (|x| - 2^j \sqrt{T}) \\ \text{for } |x| \geq 2^{j+1} \sqrt{T} & : 1 \end{cases},$$

and

$$\tilde{g}_{T,j,k}(x) = \begin{cases} \text{for } |x| \leq 2^{j+1} \sqrt{T} & : 0 \\ \text{for } 2^{j+1} \sqrt{T} \leq |x| \leq 2^{j+1} \sqrt{T} + 2^k & : b(|x| - 2^{j+1} \sqrt{T}) \\ \text{for } 2^{j+1} \sqrt{T} + 2^k \leq |x| \leq 2^{j+1} \sqrt{T} + 2^{k+1} & : 2^{k+1} b + b(2^{j+1} \sqrt{T} - |x|) \\ \text{for } 2^{j+1} \sqrt{T} + 2^{k+1} \leq |x| & : 0 \end{cases}.$$

These functions are plotted for convenience on Figure 15. We then set

$$g_{T,j,k} : \mathbb{Z}^d \rightarrow \mathbb{R}^+ \\ x \mapsto \sum_{i=1}^d \tilde{g}_{T,j,k}(x_i),$$

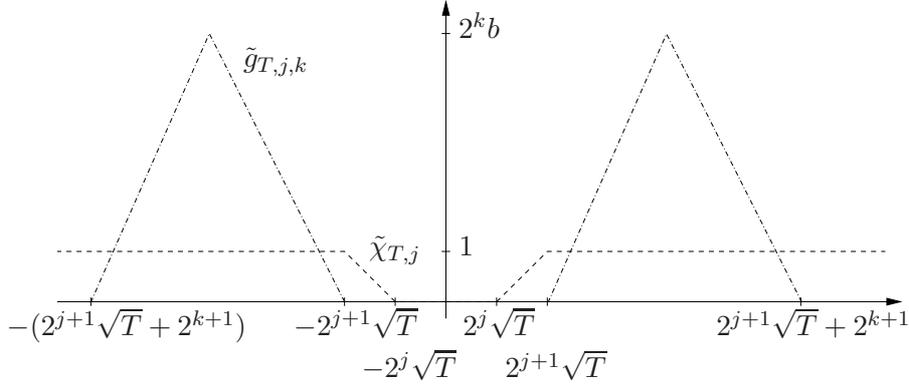
and

$$\chi_{T,j} : \mathbb{Z}^d \rightarrow [0, 1] \\ x \mapsto 1 - \prod_{i=1}^d (1 - \tilde{\chi}_{T,j}(x_i)).$$

Note that $|\nabla g_{T,j,k}(x)| \leq \sqrt{db}$ for all $x \in \mathbb{Z}^d$, $j \in \{1, \dots, d\}$, $k \in \mathbb{N}$. With the notation $|x|_\infty := \max\{x_1, \dots, x_d\}$, $\chi_{T,j}$ satisfies: $\chi_{T,j}|_{|x|_\infty \leq 2^j \sqrt{T}} \equiv 0$ and $\chi_{T,j}|_{|x|_\infty \geq 2^{j+1} \sqrt{T}} \equiv 1$.

Let $s \in \mathbb{N} \setminus \{0\}$. We multiply the defining equation for G_T by the test function

$$x \mapsto \chi_{T,j}(x)^2 \exp(2g_{T,j,k}(x)) G_T(x).$$


 FIGURE 15. Functions $\tilde{g}_{T,j,k}$ and $\tilde{\chi}_{T,j}$.

Since this function is in $L^2(\mathbb{Z}^d)$ due to [9, Lemma 9, (2.23)], we may integrate by parts the equation on \mathbb{Z}^d , obtaining

$$\begin{aligned} T^{-1} \int_{\mathbb{Z}^d} \left(\chi_{T,j}(x) \exp(g_{T,j,k}(x)) G_T(x) \right)^2 dx \\ + \int_{\mathbb{Z}^d} \nabla \left(\chi_{T,j}(x)^2 \exp(2g_{T,j,k}(x)) G_T(x) \right) \cdot A(x) \nabla G_T(x) dx = 0. \end{aligned} \quad (\text{B.3})$$

We focus on the second term of the equation and use the (classical) Leibniz rule. For the sake of clarity, we drop the subscripts and variables in the following calculation.

$$\begin{aligned} & \nabla \left(\chi^2 \exp(2g) G \right) \cdot A \nabla G \\ &= \nabla \left(\chi \exp(g) G \right) \cdot A \chi \exp(g) \nabla G + \chi \exp(g) G \nabla \left(\chi \exp(g) \right) \cdot A \nabla G \\ &= \nabla \left(\chi \exp(g) G \right) \cdot A \nabla \left(\chi \exp(g) G \right) \\ & \quad - \underbrace{\nabla \left(\chi \exp(g) G \right) \cdot A \nabla \left(\chi \exp(g) \right) G + \chi \exp(g) G \nabla \left(\chi \exp(g) \right) \cdot A \nabla G}_{\clubsuit}. \end{aligned}$$

We rewrite the second term of the r. h. s. as follows:

$$\begin{aligned} \clubsuit &= -\chi \exp(g) \nabla G \cdot A \nabla \left(\chi \exp(g) \right) G - G^2 \nabla \left(\chi \exp(g) \right) \cdot A \nabla \left(\chi \exp(g) \right) \\ &= -G^2 \nabla \left(\chi \exp(g) \right) \cdot A \nabla \left(\chi \exp(g) \right) - \chi \exp(g) G \nabla \left(\chi \exp(g) \right) \cdot A \nabla G, \end{aligned}$$

by symmetry of A . The combination of these two equalities yields

$$\begin{aligned} & \nabla \left(\chi^2 \exp(2g) G \right) \cdot A \nabla G \\ &= \nabla \left(\chi \exp(g) G \right) \cdot A \nabla \left(\chi \exp(g) G \right) - G^2 \nabla \left(\chi \exp(g) \right) \cdot A \nabla \left(\chi \exp(g) \right) \\ &\geq -\beta G^2 \left| \nabla \left(\chi \exp(g) \right) \right|^2 \end{aligned} \quad (\text{B.4})$$

by the uniform bounds on A .

Setting $\psi_{T,j,k} : x \mapsto \exp(g_{T,j,k}(x))G_T(x)$, we insert (B.4) into (B.3) to get

$$\int_{\mathbb{Z}^d} \left(T^{-1} \chi_{T,j}(x)^2 \psi_{T,j,k}(x)^2 - \beta G_T(x)^2 \left| \nabla \left(\chi_{T,j}(x) \exp(g_{T,j,k}(x)) \right) \right|^2 \right) dx \leq 0.$$

Using the properties of $\chi_{T,j}$ and $g_{T,j,k}$, this yields

$$\begin{aligned} & \int_{|x|_\infty \geq 2^j \sqrt{T}} \left(T^{-1} \psi_{T,j,k}(x)^2 - \beta G_T(x)^2 \left| \nabla \exp(g_{T,j,k}(x)) \right|^2 \right) \chi_{T,j}(x)^2 dx \\ & \leq \int_{2^j \sqrt{T} \leq |x|_\infty < 2^{j+1} \sqrt{T}} \beta \left| \nabla \chi_{T,j}(x) \right|^2 G_T(x)^2 dx, \end{aligned}$$

and finally

$$\begin{aligned} & \int_{|x|_\infty \geq 2^j \sqrt{T}} \left(T^{-1} - \beta \left| \nabla g_{T,j,k}(x) \right|^2 \right) \psi_{T,j,k}(x)^2 \chi_{T,j}(x)^2 dx \\ & \leq \beta T^{-1} 2^{-2j} \int_{2^j \sqrt{T} \leq |x|_\infty < 2^{j+1} \sqrt{T}} G_T(x)^2 dx. \end{aligned}$$

Choosing $b = (2d\beta T)^{-1/2}$ and using the definition of $\chi_{T,j}$, this turns into

$$\begin{aligned} \int_{|x|_\infty \geq 2^{j+1} \sqrt{T}} \psi_{T,j,k}(x)^2 dx & \leq \int_{|x|_\infty \geq 2^j \sqrt{T}} \psi_{T,j,k}(x)^2 \chi_{T,j}(x)^2 dx \\ & \lesssim 2^{-2j} \int_{2^j \sqrt{T} \leq |x|_\infty < 2^{j+1} \sqrt{T}} G_T(x)^2 dx. \end{aligned}$$

We then pass to the limit $k \rightarrow \infty$ by the monotone convergence theorem and use the definition of $\psi_{T,j,k}$ for $|x|_\infty \geq 2^{j+2} \sqrt{T}$ to prove the main result of this step:

$$\int_{|x|_\infty \geq 2^{j+2} \sqrt{T}} G_T(x)^2 dx \lesssim 2^{-2j} \exp(-b2^{j+1} \sqrt{T}) \int_{2^j \sqrt{T} \leq |x|_\infty < 2^{j+1} \sqrt{T}} G_T(x)^2 dx \quad (\text{B.5})$$

for all $s \in \mathbb{N} \setminus \{0\}$.

Step 2. Decay estimate and Harnack inequality. We now use the decay estimates of [9, Lemma 8, (2.23)] with $q = 2$: For all $j \in \mathbb{N}$ and $d \geq 2$, we have (recall that $|\cdot|$ and $|\cdot|_\infty$ are equivalent norms)

$$\int_{2^j \sqrt{T} \leq |x|_\infty < 2^{j+1} \sqrt{T}} G_T(x)^2 dx \lesssim (2^j \sqrt{T})^{d+(2-d)2}. \quad (\text{B.6})$$

Since

$$-\nabla^* \cdot A \nabla G_T(x) = -T G_T(x) \leq 0$$

for $|x| \gg 1$, Lemma 5 implies

$$\begin{aligned}
 & \sup_{2^{j+3}\sqrt{T} \leq |x|_\infty \leq 2^{j+4}\sqrt{T}} G_T(x) \\
 & \lesssim \left((2^j \sqrt{T})^{-d} \int_{2^{j+2}\sqrt{T} \leq |x|_\infty \leq 2^{j+5}\sqrt{T}} G_T(x)^2 dx \right)^{1/2} \\
 & \stackrel{(B.5)}{\lesssim} \left((2^j \sqrt{T})^{-d} 2^{-2j} \exp(-b2^{j+1}\sqrt{T}) \int_{2^j \sqrt{T} \leq |x|_\infty < 2^{j+1}\sqrt{T}} G_T(x)^2 dx \right)^{1/2} \\
 & \stackrel{(B.6)}{\lesssim} \left((2^j \sqrt{T})^{-d} 2^{-2j} \exp(-b2^{j+1}\sqrt{T}) (2^j \sqrt{T})^{d+(2-d)2} \right)^{1/2} \\
 & \leq (2^j \sqrt{T})^{2-d} \exp(-b2^j \sqrt{T}),
 \end{aligned}$$

which is the claim (3.2) for $d > 2$ and (3.3) for $d = 2$.

Step 3. Modifications due to the discreteness. In Step 1, we have used the standard Leibniz rule. In this step, we complete the proof by turning to the discrete Leibniz rule, and prove the following inequality corresponding to (B.4):

$$\begin{aligned}
 & \nabla(\chi_{T,j}^2 \exp(2g_{T,j,k}) G_T(x)) \cdot A \nabla G_T(x) \\
 & \geq -4\beta(G_T(x + \mathbf{e}_i)^2 + G_T(x)^2) \left| \nabla \left(\chi_{T,j}(x)^2 \exp(2g_{T,j,k}(x)) \right) \right|^2. \quad (B.7)
 \end{aligned}$$

This inequality is a consequence of the following version of [9, Proof of Lemma 8, Step 5, (4.27)]

$$\begin{aligned}
 & \nabla(\eta^2 G_T^{q-1}) \cdot A \nabla G_T(x) \\
 & \geq (1 - 2C^{-1}) \sum_{i=1}^d a_i(x) \frac{\eta^2(x + \mathbf{e}_i) + \eta^2(x)}{2} \underbrace{(G_T^{q-1}(x + \mathbf{e}_i) - G_T^{q-1}(x)) \nabla_i G_T(x)}_{\geq 0} \\
 & \quad - C \sum_{i=1}^d a_i(x) (G_T(x + \mathbf{e}_i)^q + G_T(x)^q) |\nabla_i \eta(x)|^2,
 \end{aligned}$$

with the values $q = 2$, $C = 4$, $\eta(x) = \chi_{T,j}(x) \exp(g_{T,j,k}(x))$, and using that $a_i \leq \beta$. To prove this version of [9, (4.26)], one just needs to keep track of a_i and not use the lower and upper bounds on a_i in [9, Proof of Lemma 9, Step 5]. Setting $\psi_{T,j,k} : x \mapsto \exp(g_{T,j,k}(x)) G_T(x)^{2s}$, we insert (B.7) into (B.3) to get

$$\begin{aligned}
 & \int_{\mathbb{Z}^d} \left(T^{-1} \chi_{T,j}(x)^2 \psi_{T,j,k}(x)^2 - 4\beta \sum_{i=1}^d (G_T(x + \mathbf{e}_i)^2 + G_T(x)^2) \right. \\
 & \quad \left. (\nabla_i (\chi_{T,j}(x) \exp(g_{T,j,k}(x))))^2 \right) dx \leq 0.
 \end{aligned}$$

By the properties of $\chi_{T,j}$ and $g_{T,j,k}$, this turns into

$$\begin{aligned} & \int_{|x|_\infty \geq 2^j \sqrt{T}} \left(T^{-1} \psi_{T,j,k}(x)^2 - 4\beta \sum_{i=1}^d (G_T(x + \mathbf{e}_i)^2 + G_T(x)^2) \right. \\ & \qquad \qquad \qquad \left. (\nabla_i \exp(g_{T,j,k}(x)))^2 \right) \chi_{T,j}(x)^2 dx \\ & \leq \int_{2^j \sqrt{T} \leq |x|_\infty \leq 2^{j+1} \sqrt{T}} 4\beta (1 + \exp(2b)) \sum_{i=1}^d (\nabla_i \chi_{T,j}(x))^2 (G_T(x + \mathbf{e}_i)^2 + G_T(x)^2) dx. \end{aligned} \quad (\text{B.8})$$

We then use the specific structures of $g_{T,j,k}$: For all $x_i \in [-2^{j+1}\sqrt{T} - 2^{k+1}, -2^{j+1}\sqrt{T} - 2^k] \cup [2^{j+1}\sqrt{T}, 2^{j+1}\sqrt{T} + 2^k]$, one has $\tilde{g}_{T,j,k}(x_i + 1) = \tilde{g}_{T,j,k}(x_i) + b$, and

$$\begin{aligned} \nabla_i \exp(g_{T,j,k}(x)) &= \exp \left(\sum_{l \neq i} \tilde{g}_{T,j,k}(x_l) \right) \left(\exp(\tilde{g}_{T,j,k}(x_i + 1)) - \exp(\tilde{g}_{T,j,k}(x_i)) \right) \\ &= \exp \left(\sum_{l \neq i} \tilde{g}_{T,j,k}(x_l) \right) \exp(\tilde{g}_{T,j,k}(x_i)) (\exp(b) - 1) \\ &= \exp \left(\sum_{l \neq i} \tilde{g}_{T,j,k}(x_l) \right) \exp(\tilde{g}_{T,j,k}(x_i + 1)) (1 - \exp(-b)), \end{aligned}$$

so that

$$\left(\nabla_i \exp(g_{T,j,k}(x)) \right)^2 = \exp(g_{T,j,k}(x))^2 (\exp(b) - 1)^2 \quad (\text{B.9})$$

$$= \exp(g_{T,j,k}(x + \mathbf{e}_i))^2 (1 - \exp(-b))^2 \quad (\text{B.10})$$

for $x_i \in [-2^{j+1}\sqrt{T} - 2^{k+1}, -2^{j+1}\sqrt{T} - 2^k] \cup [2^{j+1}\sqrt{T}, 2^{j+1}\sqrt{T} + 2^k]$.

For $x_i \in [-2^{j+1}\sqrt{T} - 2^k, -2^{j+1}\sqrt{T}] \cup [2^{j+1}\sqrt{T} + 2^k, 2^{j+1}\sqrt{T} + 2^{k+1}]$, the same chain of arguments also yields (B.9) & (B.10).

Finally, for

$$\begin{aligned} x_i &\geq 2^{j+1}\sqrt{T} + 2^{k+1}, \\ x_i &< -(2^{j+1}\sqrt{T} + 2^{k+1}), \\ -2^{j+1}\sqrt{T} &\leq x_i < 2^{j+1}\sqrt{T}, \end{aligned}$$

one has $\left(\nabla_i \exp(g_{T,j,k}(x)) \right)^2 = 0$.

We then deduce the following bound for all $|x_i| \geq 2^j \sqrt{T}$:

$$\begin{aligned} & \left(\nabla_i \exp(g_{T,j,k}(x)) \right)^2 (G_T(x + \mathbf{e}_i)^2 + G_T(x)^2) \\ & \stackrel{(\text{B.9}) \& (\text{B.10})}{=} \psi_{T,j,k}(x + \mathbf{e}_i)^2 (1 - \exp(-b))^2 + \psi_{T,j,k}(x)^2 (\exp(b) - 1)^2 \\ & \lesssim \left(\psi_{T,j,k}(x + \mathbf{e}_i)^2 + \psi_{T,j,k}(x)^2 \right) b^2, \end{aligned} \quad (\text{B.11})$$

for all $0 < b \leq 1$. Inserting (B.11) into (B.8) yields

$$\int_{|x|_\infty \geq 2^{j+1}\sqrt{T}} (T^{-1} - 4\beta db^2) \psi_{T,j,k}(x)^2 dx \lesssim \int_{2^j\sqrt{T} \leq |x|_\infty \leq 2^{j+1}\sqrt{T}+1} G_T(x)^2 dx,$$

and we may conclude as before, taking this time $b = (8\beta dT)^{-1/2}$.

REFERENCES

[1] S. Agmon. *Lectures on exponential decay of solutions of second-order elliptic equations: bounds on eigenfunctions of N-body Schrödinger operators*, volume 29 of *Mathematical Notes*. Princeton University Press, Princeton, NJ, 1982.

[2] R. Alicandro, M. Cicalese, and A. Gloria. Integral representation results for energies defined on stochastic lattices and application to nonlinear elasticity. 2009. Preprint available at <http://hal.archives-ouvertes.fr/inria-00437765/en/>.

[3] A. Bourgeat and A. Piatnitski. Approximations of effective coefficients in stochastic homogenization. *Ann. I. H. Poincaré*, 40:153–165, 2005.

[4] P. Caputo and D. Ioffe. Finite volume approximation of the effective diffusion matrix: the case of independent bond disorder. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(3):505–525, 2003.

[5] T. Delmotte. Inégalité de Harnack elliptique sur les graphes. *Colloq. Math.*, 72(1):19–37, 1997.

[6] A. Dykhne. Conductivity of a two-dimensional two-phase system. *Sov. Phys. JETP.*, 32:63–65, 1971. Russian version: *Zh Eksp Teor Fiz* 1970;59:110–5.

[7] W. E, P.B. Ming, and P.W. Zhang. Analysis of the heterogeneous multiscale method for elliptic homogenization problems. *J. Amer. Math. Soc.*, 18:121–156, 2005.

[8] A. Gloria. Reduction of the resonance error - Part 1: Approximation of homogenized coefficients. Preprint. Available at <http://hal.archives-ouvertes.fr/inria-00457159/en/>.

[9] A. Gloria and F. Otto. An optimal variance estimate in stochastic homogenization of discrete elliptic equations. *Ann. Probab.*, to appear.

[10] A. Gloria and F. Otto. An optimal error estimate in stochastic homogenization of discrete elliptic equations. Preprint available at <http://hal.archives-ouvertes.fr/inria-00457020/en/>.

[11] A. Gloria and F. Otto. In preparation.

[12] V.V. Jikov, S.M. Kozlov, and O.A. Oleinik. *Homogenization of Differential Operators and Integral Functionals*. Springer-Verlag, Berlin, 1994.

[13] T. Kanit, S. Forest, I. Galliet, V. Mounoury, and D. Jeulin. Determination of the size of the representative volume element for random composites: statistical and numerical approach. *Int. J. Sol. Struct.*, 40:3647–3679, 2003.

[14] S.M. Kozlov. The averaging of random operators. *Mat. Sb. (N.S.)*, 109(151)(2):188–202, 327, 1979.

[15] S.M. Kozlov. Averaging of difference schemes. *Math. USSR Sbornik*, 57(2):351–369, 1987.

[16] R. Künnemann. The diffusion limit for reversible jump processes on \mathbb{Z}^d with ergodic random bond conductivities. *Commun. Math. Phys.*, 90:27–68, 1983.

[17] J. A. Meijerink and H. A. van der Vorst. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M -matrix. *Math. Comp.*, 31(137):148–162, 1977.

[18] A. Naddaf and T. Spencer. Estimates on the variance of some homogenization problems. Preprint, 1998.

[19] H. Owhadi. Approximation of the effective conductivity of ergodic media by periodization. *Probab. Theory Relat. Fields*, 125:225–258, 2003.

[20] G.C. Papanicolaou and S.R.S. Varadhan. Boundary value problems with rapidly oscillating random coefficients. In *Random fields, Vol. I, II (Esztergom, 1979)*, volume 27 of *Colloq. Math. Soc. János Bolyai*, pages 835–873. North-Holland, Amsterdam, 1981.

[21] X. Yue and W. E. The local microscale problem in the multiscale modeling of strongly heterogeneous media: effects of boundary conditions and cell size. *J. Comput. Phys.*, 222(2):556–572, 2007.

[22] V.V. Yurinskii. Averaging of symmetric diffusion in random medium. *Sibirskii Matematicheskii Zhurnal*, 27(4):167–180, 1986.

(Antoine Gloria) PROJET SIMPAF, INRIA LILLE-NORD EUROPE, FRANCE
E-mail address: antoine.gloria@inria.fr