Received: 26 April 2019 | Revised: 28 January 2020 | Accepted: 13 February 2020

DOI: 10.1002/nla.2296

RESEARCH ARTICLE

WILEY

Numerical study in stochastic homogenization for elliptic partial differential equations: Convergence rate in the size of representative volume elements

Venera Khoromskaia[1,2] | **Boris N. Khoromskij**[1] | **Felix Otto**[1]

[1]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

[2]Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

Correspondence
Boris N. Khoromskij,
Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22-26, D-04103 Leipzig, Germany.
Email: bokh@mis.mpg.de

Summary

We describe the numerical scheme for the discretization and solution of 2D elliptic equations with strongly varying piecewise constant coefficients arising in the stochastic homogenization of multiscale composite materials. An efficient stiffness matrix generation scheme based on assembling the local Kronecker product matrices is introduced. The resulting large linear systems of equations are solved by the preconditioned conjugate gradient iteration with a convergence rate that is independent of the grid size and the variation in jumping coefficients (contrast). Using this solver, we numerically investigate the convergence of the representative volume element (RVE) method in stochastic homogenization that extracts the effective behavior of the random coefficient field. Our numerical experiments confirm the asymptotic convergence rate of systematic error and standard deviation in the size of RVE rigorously established in Gloria et al. The asymptotic behavior of covariances of the homogenized matrix in the form of a quartic tensor is also studied numerically. Our approach allows laptop computation of sufficiently large number of stochastic realizations even for large sizes of the RVE.

KEYWORDS

covariance of homogenization matrix, elliptic problem solver, empirical variance, homogenized matrix, Kronecker product, PCG iteration, representative volume element, stochastic homogenization

MOS SUBJECT CLASSIFICATION

65F30; 65F50; 65N35; 65F10

1 | INTRODUCTION

Homogenization methods allow to derive the effective mechanical and physical properties of highly heterogeneous materials from the knowledge of the spatial distribution of their components.[1–3] In particular, stochastic homogenization via the representative volume element (RVE) methods provide means for calculating the effective large-scale characteristics related to structural and geometric properties of random composites, by utilizing a possibly large number of probabilistic

realizations.[4–9] The numerical investigation of the effective characteristics of random structures is a challenging problem since the underlying elliptic equation (the corrector problem) with randomly generated coefficients should be solved for many thousands realizations and for domains with substantial structural complexity to obtain sufficient statistics. Note that for every realization of the random medium one should construct a new stiffness matrix and right-hand side to solve the discretized corrector problem. Therefore, construction of fast solvers (which allow to confirm numerically the quantitative results in the stochastic homogenization) using the conventional computing facilities is a challenge.

This article presents the numerical study of the stochastic homogenization of an elliptic system with randomly generated coefficients. Our approach is based on the finite element method (FEM)-Galerkin approximation of the 2D elliptic equations in a periodic setting by using fast assembling of the FEM stiffness matrix in a sparse matrix format, which is performed by agglomerating the Kronecker tensor products of simple 1D FEM discrete operators.[10] We use the product piecewise linear finite elements on the rectangular grid assuming that strongly varying piecewise constant equation coefficients are resolved on that grid. This scheme provides efficient approximation of equations with complicated jumping coefficients. The numerical analysis of the error in the Galerkin FEM approximation indicates the convergence rate $O(h^\beta)$ in the L^2-norm with $3/2 \le \beta \le 2$.

The resulting large linear system of equations is solved by the preconditioned conjugate gradient (PCG) iteration, where the convergence rate is proven to be independent on the grid size and the relative variation in jumping coefficients, that is, on the contrast. The preconditioned iterative solvers for the discrete elliptic systems of equations with variable coefficients have been long studied in the literature on numerical methods for multidimensional and stochastic partial differential equations (PDEs), see References 11–14 and the literature therein. The review article[15] on sparse tensor approximation of high-dimensional stochastic PDEs discusses many different numerical schemes on the topic.

In this article, we consider an ensemble of two-valued random coefficient fields, which is based on independently and uniformly placed (and thus overlapping) axis-parallel square inclusions of fixed side length. We investigate the RVE method that (approximately) extracts the effective (i.e., large-scale) behavior of the medium in form of the deterministic and homogeneous matrix $\mathbb{A}_{\rm hom}$ from a given (stationary and ergodic) ensemble. This method produces an approximation to $\mathbb{A}_{\rm hom}$ by solving two-dimensional elliptic equations on a square of (lateral) size L with periodic boundary conditions and a specific right-hand side (the corrector equation), by taking the spatial average of the flux of these solutions, and by taking the empirical mean over N independent realizations of this coefficient field under the naturally periodized version of the ensemble. This is an approximation in so far as the outcome is still random (as quantified by the standard deviation of the outcome of a single realization) and that the periodic boundary conditions affect the statistics (which we call the systematic error, because the periodization introduces artificial long-range correlations, and which can be considered as a bias). In Reference 8, Gloria, Neukamm, and the last author rigorously derived upper bounds how the standard deviation and the systematic error decrease with increasing RVE size L. Our numerical experiments confirm the scaling of these bounds. Since numerically, there is no access to exact values of the variance (or standard deviation) or the expectation, we replace these quantities by their empirical counterparts for a large number of realizations N. We thus first provide numerical evidence that these quantities have saturated in N (i.e., reach convergence of a Monte Carlo estimate of variance), and second that their limiting values display the predicted scaling in L.

In work[16] by Duerinckx, Gloria, and the last author, it was worked out that the properly rescaled variance of the output of the RVE converges as $L\uparrow\infty$ to a quartic tensor Q that governs the leading-order fluctuations of any solution. In this article, we show how the symmetry properties of the ensemble yields symmetry properties of Q (and its approximation). Also, a convergence rate was rigorously established in that work, and is being numerically investigated here. In the range of investigated parameters N and L the numerical findings confirm the theoretic results.[16]

In numerical tests on the stochastic properties of the 2D RVE method we study the asymptotic of empirical variance versus the size of RVE $L \le 128$, and of the systematic error versus the number of realizations N up to $N = 10^5$. Furthermore, we estimate the convergence of the quartic tensor by implementing a large number of stochastic realizations. The proposed techniques allow to compute a sufficiently large number of realizations of random coefficient fields with a large number of overlapping inclusions up to $L^2 = 128^2$ corresponding to the stiffness matrix size $513^2 \times 513^2$ using MATLAB on a moderate computer cluster.

The numerical investigation of the stochastic homogenization problem attracts interest and becomes an active field of research, see the survey[1] and references therein. Recently, the numerical solution of the corrector-type problem, in the context of homogenization of the diffusion equation with spherical inclusions by using boundary element methods and the fast multipole techniques has been considered in Reference 17.

The rest of the article is organized as follows. In Section 2, we address the problem setting and define the elliptic equations of stochastic homogenization. Section 3 describes the Galerkin-FEM discretization scheme based on the

fast matrix generation by using sums of Kronecker products of single-dimensional matrices. We also outline the preconditioned conjugate gradient (PCG) iteration applied in the computer simulations and provide numerics on the FEM discretization error (see Appendix). Section 4 introduces the computational scheme for the stochastic average coefficient matrix. Furthermore, in Section 4.3 we describe the construction and properties of the covariances of the homogenized matrix in the form of a quartics tensor. Section 5 presents results of numerical experiments on the empirical average and systematic error at the limit of a large number of stochastic realizations. The asymptotic of the quartic tensor versus the leading order variances is analyzed numerically in Section 5.3. Conclusions outline the main results of the article.

2 | ELLIPTIC EQUATIONS IN STOCHASTIC HOMOGENIZATION

In this section, we describe the problem setting in the stochastic homogenization theory. For given $f \in \mathcal{L}^2(\Omega)$ such that $\int_\Omega f(x)dx = 0$, we consider the class of model elliptic boundary value problems on the d-dimensional hypercube $\Omega := [0, L)^d$ of side-length $L \in \mathbb{N}$ with periodic boundary conditions: find $\phi \in H^1(\Omega)$, s.t.

$$\mathcal{A}\phi := -\nabla \cdot \mathbb{A}(x)\nabla\phi = f(x), \quad x = (x_1, \dots, x_d) \in \Omega, \tag{1}$$

with the diagonal $d \times d$ uniformly elliptic coefficient matrix $\mathbb{A}(x)$, $\infty > \beta_0 I \geq \mathbb{A}(x) \geq \alpha_0 I > 0$. In this article, we consider the case $d = 2$ and focus on the special class of elliptic problems (1) arising in stochastic homogenization theory for the corrector problem, where the highly varying coefficient matrix and the right-hand side (RHS) are defined by a sequence of stochastic realizations as described in References 5–9, see details in Sections 3 and 4.

In what follows, we present the numerical analysis for 2D stochastic homogenization problems (1) with periodic boundary conditions on $\Gamma = \partial\Omega$, in the form

$$\begin{aligned} &\phi(0, x_2) = \phi(L, x_2), \quad \frac{\partial}{\partial x_1}\phi(0, x_2) = \frac{\partial}{\partial x_1}\phi(L, x_2), \quad x_2 \in [0, L), \\ &\phi(x_1, 0) = \phi(x_1, L), \quad \frac{\partial}{\partial x_2}\phi(x_1, 0) = \frac{\partial}{\partial x_2}\phi(x_1, L), \quad x_1 \in [0, L). \end{aligned}$$

The diagonal 2×2 coefficient matrix $\mathbb{A}(x) = \mathbb{A}(x, \omega)$ is defined by

$$\mathbb{A}(x, \omega) = \begin{pmatrix} a(x, \omega) & 0 \\ 0 & a(x, \omega) \end{pmatrix}, \quad x \in \Omega,$$

where the scalar function $a(x, \omega) > 0$ is piecewise constant in the domain Ω and the randomness is encoded in the coefficient $a(x, \omega)$ via stochastic realizations as described in Sections 3 and 4. The efficient numerical simulation presupposes the fast numerical solution of Equation (1) in the case of many different realizations of the coefficients $\mathbb{A}(x)$ and RHSs, generated by certain stochastic procedure, and in the calculation of various functionals on the sequence of solutions ϕ. In this problem setting, the bottleneck task is fast and accurate generation of the FEM stiffness matrix in the sparse matrix form, which should be recalculated many hundred if not thousand times in the course of stochastic realizations.

In asymptotic analysis of stochastic homogenization problems the coefficient and the RHS are chosen in a specific way, see References 5 and 7 for the particular problem setting. In this article, we describe the conventional 2D FEM discretization scheme in the domain $\Omega = [0, L)^2$. Given the size of RVEs $L = 2, 3, \dots$, we randomly and independently pick L^2 points on the discretization grid according to the uniform distribution, which for $L \to \infty$ approximate the Poisson point process with uniform density. Then, we consider the union of L^2 equal unit cells G_s, $s = 1, \dots, L^2$, each of size $2\alpha \times 2\alpha$, centered at these points (which of course often overlap, see Figure 1). For numerical convenience, we further rescale the computational domain to the unit square, $[0, L)^2 \mapsto [0, 1)^2$. In our numerical experiments the occupation factor α is chosen in the range $0 < \alpha \leq 1/2$, so that the size of unit cell rescales to $\frac{2\alpha}{L} \times \frac{2\alpha}{L}$. Stochastic characteristics of the system can be estimated at the limit of a large number L.

We consider a sequence of random coefficient realizations $\{G_s\}_n$, numbered by $n = 1, \dots, N$, where the particular set $\{G_s\} = \{G_s\}_n$ for fixed n will be called a realization. For any fixed realization define the covered domain

$$\hat{G} = \hat{G}_n := \bigcup_{s=1}^{L^2} G_s, \tag{2}$$

and the respective coefficient

$$\hat{a}(x) = \hat{a}^{(n)}(x) = \begin{cases} 1 & \text{if } x \in \hat{G}_n, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The stochastic model is specified by the choice of the overlap constant $\alpha \in (0, 1/2]$ and the scaling factor $\lambda \in (0, 1]$. In the following, the constant λ will be fixed in the interval $0.1 \le \lambda \le 0.8$. Given the model parameters α and λ, we denote the "stochastic" elliptic operator for the particular realization by $\mathcal{A}^{(n)}$ or just $\mathcal{A}$ (if n is fixed) so that

$$\mathcal{A}^{(n)} = -\nabla \cdot \mathbb{A}^{(n)}(x)\nabla,$$

where the corresponding 2×2 coefficient matrix $\mathbb{A}^{(n)}(x) = \mathbb{A}^{(n)}(x, \lambda, \alpha, \{G_n\})$ is defined by

$$\mathbb{A}^{(n)}(x) = \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} \hat{a}^{(n)}(x) & 0 \\ 0 & \hat{a}^{(n)}(x) \end{pmatrix} := \begin{pmatrix} a^{(n)}(x) & 0 \\ 0 & a^{(n)}(x) \end{pmatrix}, \quad x \in \Omega, \tag{4}$$

and the diagonal matrix coefficient takes the form

$$a^{(n)}(x) = \lambda + (1-\lambda)\hat{a}^{(n)}(x). \tag{5}$$

We use the notation $\hat{\mathbb{A}}^{(n)}(x)$ for the "stochastic part" of a matrix associated with the diagonal coefficient $\hat{a}^{(n)}(x)$, that is,

$$\hat{\mathbb{A}}^{(n)}(x) = \begin{pmatrix} \hat{a}^{(n)}(x) & 0 \\ 0 & \hat{a}^{(n)}(x) \end{pmatrix}.$$

Now the elliptic equations in stochastic homogenization are formulated as follows. Fixed the realization of coefficient $\hat{\mathbb{A}}^{(n)}(x)$, for $i = 1, 2$ solve the periodic elliptic problems in Ω,

$$-\lambda\Delta\phi_i - (1-\lambda)\nabla \cdot \hat{\mathbb{A}}^{(n)}(\mathbf{e}_i + \nabla\phi_i) = 0, \tag{6}$$

where the directional unit vectors $\mathbf{e}_i$, $i = 1, 2$, are given by $\mathbf{e}_1 = (1, 0)^T$ and $\mathbf{e}_2 = (0, 1)^T$, see Section 4 for more details. For given realization of the coefficient $\hat{\mathbb{A}}^{(n)}(x)$, the variational form of the deterministic elliptic equation (6) reads as follows

$$\int_\Omega (\lambda\nabla\phi_i \cdot \nabla\psi + (1-\lambda)\hat{a}^{(n)}(x)[\mathbf{e}_i + \nabla\phi_i] \cdot \nabla\psi)dx = 0 \quad \forall\, \psi \in H^1(\Omega). \tag{7}$$

Equation (6) can be also written in the classical form (1)

$$\mathcal{A}^{(n)}\phi_i = f_i, \quad \text{with} \quad f_i = (1-\lambda)\nabla \cdot \hat{a}^{(n)}\, \mathbf{e}_i. \tag{8}$$

Figure 1 illustrates an example of the particular realization of stochastic coefficient $a^{(n)}(x)$ in the case $L = 8$, $\lambda = 0.2$, and $\alpha = 1/4$ visualized on $m_1 \times m_1$ grid with $m_1 = 97$. In this example the stochastic part of the coefficient varies in the range $[0.2; 1]$.

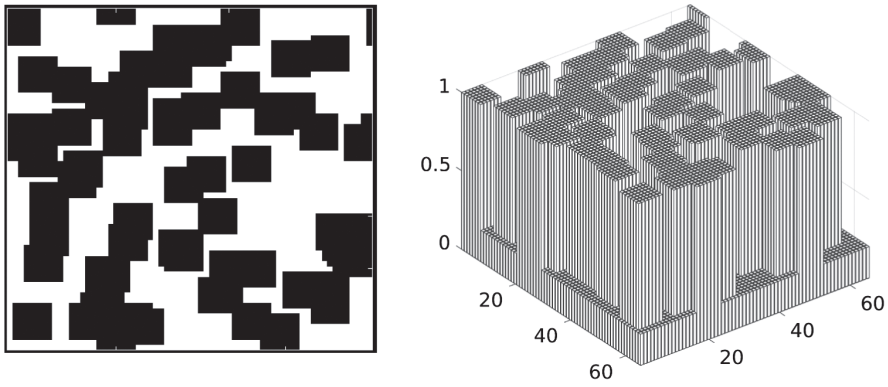

FIGURE 1 A realization of a stochastic coefficient with L^2 overlapping cells for $L = 8$, $m_0 = 8$, $\alpha = 3/8$, $\lambda = 0.2$

The problem setting remains verbatim in the d-dimensional case, $d > 2$. In this case, Equation (8) takes the same form, where a $d \times d$ coefficient matrix is given by

$$\mathbb{A}^{(n)}(x) = \mathrm{diag}\{a^{(n)}(x), \dots, a^{(n)}(x)\}, \quad x \in \Omega$$

and $\mathbf{e}_i \in \mathbb{R}^d$, $i = 1, \dots, d$, represents the set of directional unit vectors in $\mathbb{R}^d$.

3 | MATRIX GENERATION AND ITERATIVE SOLUTION

In this section, we describe the FEM discretization scheme and the fast matrix generation approach based on the use of tensor Kronecker products of "univariate" matrices.

3.1 | Galerkin FEM discretization

First, we introduce the uniform $m_s \times m_s$ rectangular grid Ω_{h_s} in Ω with the grid size $h_s = \frac{1}{m_s - 1}$, such that $m_s = m_0 L + 1$, $m_0 = 2^{p_0}$, that is, $h_s = \frac{1}{m_0 L}$. We assume that the unit cell G_s, $s = 1, \dots L^2$, of size $\frac{2\alpha}{L} \times \frac{2\alpha}{L}$ is adjusted to the square grid Ω_{h_s}, such that the center c_s of G_s belongs to the set of grid points in Ω_s, while the overlap factor α may take values $\alpha \in \left\{ \frac{1}{m_0}, \frac{2}{m_0}, \dots \frac{2^{p_0 - 1}}{m_0} \right\}$. In this construction, the univariate size of the unit cell varies as

$$\frac{2\alpha}{L} = \frac{2\alpha m_0}{m_0 L} = k h_s, \quad \text{with} \quad k = 2,\ 4, \dots m_0.$$

In the following numerical examples we normally use the occupation constant $\alpha = 1/4$. In the case of $\alpha = 1/2$, the unit cell of the size $\frac{1}{L} \times \frac{1}{L}$ contains $m_0 + 1$ grid points in each spatial direction leading to $m_s \times m_s$ rectangular grid with $m_s = m_0 L + 1$.

The FEM discretization of the elliptic PDE in (8) can be constructed, in general, on the finer grid Ω_h compared with Ω_s, which serves for the resolution of jumping coefficients. To that end, we introduce the $m_1 \times m_1$ rectangular grid Ω_h with the mesh size $h = \frac{1}{m_1 - 1}$, $m_1 \geq m_s$, that is obtained by a dyadic refinement of the grid Ω_s, such that the relation

$$m_1 - 1 = (m_s - 1)2^p, \quad \text{with} \quad p = 0, 1, 2, \dots \tag{9}$$

holds, implying $h_s = 2^p h$. Now the grid-size of the unit cell G_s on the finer grid Ω_h is given by $(m_0 2^p + 1) \times (m_0 2^p + 1)$.

Given a finite dimensional space $X \subset H^1(\Omega)$ of tensor product piecewise linear finite elements

$$X = \mathrm{span}\{\psi_\mu(x)\},$$

associated with the grid Ω_h, with $\mu = 1, \dots, M_d$, $M_d = m_1^d$, for $d = 2$ incorporating periodic boundary conditions, we are looking for the traditional FEM Galerkin approximation of the exact solution in the form

$$\phi(x) \approx \phi_X(x) = \sum_{\mu=1}^{M_d} u_\mu \psi_\mu(x) \in X,$$

where $\mathbf{u} = (u_1, \dots, u_{M_d})^T \in \mathbb{R}^{M_d}$ is the unknown coefficients vector. Fixed realization of the coefficient $a^{(n)}(x)$, for $i = 1, 2$ we define the Galerkin-FEM discretization (with respect to above defined finite dimensional space X) of the variational equation (7) by

$$A\mathbf{u}_i = \mathbf{f}_i, \quad A = [a_{\mu\nu}] \in \mathbb{R}^{M_d \times M_d}, \quad \mathbf{f}_i = \mathbf{f} = [f_\mu] \in \mathbb{R}^{M_d}, \tag{10}$$

where the Galerkin-FEM matrix A generated by the equation coefficient $\mathbb{A}^{(n)}(x)$ is calculated by using the associated bilinear form

$$a_{\mu\nu} = \langle \mathcal{A}\psi_\mu, \psi_\nu \rangle = \int_\Omega (\lambda \nabla\psi_\mu \cdot \nabla\psi_\nu + (1 - \lambda)\hat{a}^{(n)}(x) \nabla\psi_\mu \cdot \nabla\psi_\nu) dx, \tag{11}$$

and

$$f_\mu = \langle f, \psi_\mu \rangle = \int_\Omega (1-\lambda)\nabla \cdot \hat{a}^{(n)}(x)\ \mathbf{e}_i \psi_\mu\ dx = -(1-\lambda)\int_\Omega \hat{a}^{(n)}(x)\frac{\partial \psi_\mu}{\partial x_i}dx. \tag{12}$$

Corresponding to (8) and (11), we represent the stiffness matrix A in the additive form

$$A = \lambda A_\Delta + (1-\lambda)\hat{A}_s, \tag{13}$$

where A_Δ represents the $M_d \times M_d$ FEM Laplacian matrix in periodic setting that has the standard two-term Kronecker product form. Here matrix $\hat{A}_s$ provides the FEM approximation to the "stochastic part" in the elliptic operator corresponding to the coefficient $\hat{a}^{(n)}(x)$, see (4). The latter is determined by the sequence of random coefficients distribution in the course of stochastic realizations, numbered by $n = 1, \dots, N$.

In the case of complicated jumping coefficients the stiffness matrix generation in the elliptic FEM usually constitutes the dominating part of the overall solution cost. In the course of stochastic realizations, Equation (10) is to be solved many hundred or even thousand times, so that every time one has to update the stiffness matrix A and the RHS f.

Our discretization scheme computes all matrix entries at the low cost by assembling of the local Kronecker product matrices obtained by representation of $\hat{a}^{(n)}(x)$ as a sum of separable functions. This allows to store the resultant stiffness matrix in the sparse matrix format. Such a construction only includes the precomputing of tridiagonal matrices representing 1D elliptic operators with jumping coefficients in periodic setting. In the next sections, we shall describe the efficient construction of the "stochastic" term A_s.

3.2 | Matrix generation by using Kronecker product sums

To enhance the time-consuming matrix assembling process we apply the FEM Galerkin discretization (11) of Equation (8) by means of the tensor-product piecewise linear finite elements

$$\{\psi_{\boldsymbol{\mu}}(x) := \psi_{\mu_1}(x_1)\dots\psi_{\mu_d}(x_d)\}, \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_d), \quad \mu_\ell \in \mathcal{I}_\ell = \{1, \dots, m_\ell\}, \quad \ell = 1, \dots, d,$$

where $\psi_{\mu_\ell}(x_\ell)$ are the univariate piecewise linear hat functions*. The $M_d \times M_d$ stiffness matrix is constructed by the standard mapping of the multi-index $\boldsymbol{\mu}$ into the long univariate index $1 \le \mu \le M_d$ for the active degrees of freedom in periodic setting. For instance, we use the so-called big-endian convention for $d = 3$ and $d = 2$

$$\boldsymbol{\mu} \mapsto \mu := \mu_3 + (\mu_2 - 1)m_3 + (\mu_1 - 1)m_2 m_3, \quad \boldsymbol{\mu} \mapsto \mu := \mu_2 + (\mu_1 - 1)m_2,$$

respectively. In what follows, we consider the case $d = 2$ in more detail.

In our discretization scheme, we calculate the stiffness matrix by assembling of the local Kronecker product terms by using representation of the "stochastic part" in the coefficient $\hat{a}^{(n)}(x)$ as an R-term sum of separable functions. To that end, let us assume for the moment that the scalar diffusion coefficient $a(x_1, x_2)$ can be represented in the separate form (rank-1 representation)

$$a(x_1, x_2) = a^{(1)}(x_1)a^{(2)}(x_2).$$

Then the entries of the Galerkin stiffness matrix $A = [a_{\mu\nu}] \in \mathbb{R}^{M_d \times M_d}$ can be represented by

$$\begin{aligned} a_{\mu\nu} = \langle \mathcal{A}\psi_\mu, \psi_\nu \rangle &= \int_\Omega a^{(1)}(x_1)a^{(2)}(x_2)\nabla\psi_\mu \cdot \nabla\psi_\nu dx \\ &= \int_{(0,1)} a^{(1)}(x_1)\frac{\partial \psi_{\mu_1}(x_1)}{\partial x_1}\frac{\partial \psi_{\nu_1}(x_1)}{\partial x_1}dx_1 \int_{(0,1)} a^{(2)}(x_2)\psi_{\mu_2}(x_2)\psi_{\nu_2}(x_2)dx_2 \\ &\quad + \int_{(0,1)} a^{(1)}(x_1)\psi_{\mu_1}(x_1)\psi_{\nu_1}(x_1)dx_1 \int_{(0,1)} a^{(2)}(x_2)\frac{\partial \psi_{\mu_2}(x_2)}{\partial x_2}\frac{\partial \psi_{\nu_2}(x_2)}{\partial x_2}dx_2, \end{aligned}$$

*Notice that the univariate grid size m_ℓ is of the order of $m_\ell = O(1/\epsilon)$, where the small homogenization parameter is given by $\epsilon \approx 1/(m_0 L)$, designating the total problem size $M_d = m_1 m_2 \dots m_d = O(1/\epsilon^d)$.

which leads to the rank-2 Kronecker product representation

$$A = A_1 \otimes S_2 + S_1 \otimes A_2,$$

where $\otimes$ denotes the conventional Kronecker product of matrices, see Definition 1 below. Here $A_1 = [a_{\mu_1\nu_1}] \in \mathbb{R}^{m_1\times m_1}$ and $A_2 = [a_{\mu_2\nu_2}] \in \mathbb{R}^{m_2\times m_2}$ denote the univariate stiffness matrices and $S_1 = [s_{\mu_1\nu_1}] \in \mathbb{R}^{m_1\times m_1}$ and $S_2 = [s_{\mu_2\nu_2}] \in \mathbb{R}^{m_2\times m_2}$ define the weighted mass matrices, for example

$$a_{\mu_1\nu_1} = \int_{(0,1)} a^{(1)}(x_1)\frac{\partial\psi_{\mu_1}(x_1)}{\partial x_1}\frac{\partial\psi_{\nu_1}(x_1)}{\partial x_1}dx_1, \quad s_{\mu_1\nu_1} = \int_{(0,1)} a^{(1)}(x_1)\psi_{\mu_1}(x_1)\psi_{\nu_1}(x_1)dx_1.$$

Definition 1. Recall that given $p_1 \times q_1$ matrix A and $p_2 \times q_2$ matrix B, their Kronecker product is defined as a $p_1p_2 \times q_1q_2$ matrix C via the block representation

$$C = A \otimes B = [a_{ij}B], \quad i = 1, \dots, p_1, \quad j = 1, \dots, q_1.$$

Let us discuss in more detail the calculation of the 1D stiffness matrices A_1 and A_2 in the case of variable 1D coefficients. We choose the Galerkin FEM with $m = m_1$ piecewise-linear hat functions $\{\psi_{\mu_1}\}$ in periodic setting in $\Omega = [0, 1)$, constructed on a uniform grid with a step size $h = 1/m$, and nodes $x_{\mu_1} = h\ \mu_1$, $\mu_1 = 1, \dots, m$. If we denote the diffusion coefficient by $a(x_1)$, then the entries of the exact stiffness matrix A_1 read as

$$(a)_{\mu_1,\mu_1'} = \langle a(x)\nabla\psi_{\mu_1}(x), \quad \nabla\psi_{\mu_1'}(x)\rangle_{L_2(D)}, \quad \mu_1, \mu_1' = 1, \dots, m.$$

We assume that the coefficient remains constant at each spatial interval $[x_{\mu_1-1}, x_{\mu_1}]$, which corresponds to the evaluation of the scalar product above via the midpoint quadrature rule yielding the approximation order $O(h^2)$.

Introducing the coefficient vector $\mathbf{a} = [a_{\mu_1}] \in \mathbb{R}^m$, $a_{\mu_1} = a(x_{\mu_1-1/2})$, $\mu_1 = 1, \dots, m$, where $x_{i_1-1/2}$ is the middle point of the integration interval, the symmetric tridiagonal matrix of interest can be represented by

$$A_1 = -\frac{1}{h}\begin{bmatrix} a_1+a_2 & -a_2 & & & & -a_1 \\ -a_2 & a_2+a_3 & -a_3 & & & \\ & \ddots & \ddots & \ddots & & \\ & & -a_{m-1} & a_{m-1}+a_m & -a_m & \\ -a_1 & & & -a_m & a_m+a_1 & \end{bmatrix}. \tag{14}$$

By simple algebraic transformations (e.g., by lamping of the mass matrices) the matrix A can be simplified to the form (without loss of approximation order)

$$A \mapsto A = A_1 \otimes D_2 + D_1 \otimes A_2, \tag{15}$$

where D_1, D_2 are the diagonal matrices with positive entries. This representation applies in particular to the periodic Laplacian. We notice that the product finite element space X of the univariate piecewise linear hat functions applied in this article guarantees the low Kronecker rank representation for stiffness matrix of the elliptic operator with separable coefficients. In the case of Laplacian we use the standard finite element scheme.

In the general case, the piecewise constant stochastic coefficient can be represented as an R-term sum of separable coefficients. This leads to the linear system of equations

$$A\mathbf{u} = \mathbf{f}, \tag{16}$$

constructed for the general R-term separable coefficient $a(x_1, x_2)$ with $R > 1$.

For preconditioning needs, we use the simplified version of (15) in the form of anisotropic Laplacian type matrix

$$A \mapsto B = \alpha_2 A_1 \otimes I_2 + \alpha_1 I_1 \otimes A_2,$$

where α_1 and α_2 define the average values of the diagonal entries of matrices D_1 and D_2, respectively. The matrix B will be used as a prototype preconditioner in the PCG iteration for solving the target linear system (16).

Taking into account the rectangular structure of the grid, we use the simple finite-difference (FD) scheme for the matrix representation of the negative Laplacian operator $-\Delta$. In this case the discrete Laplacian incorporating periodic boundary conditions takes the form

$$A_\Delta = \Delta_1 \otimes I_{m_2} + I_{m_1} \otimes \Delta_2, \tag{17}$$

where

$$\Delta_1 = \frac{1}{h^2}\left(\text{tridiag}\{1,-2,1\} + P^{(1)}\right) \in \mathbb{R}^{m_1\times m_1},$$

such that the entries of the "periodization" matrix $P^{(1)} \in \mathbb{R}^{m_1\times m_1}$ are all zeros except

$$P^{(1)}_{1,m_1} = P^{(1)}_{m_1,1} = 1.$$

Here $I_{m_1} \in \mathbb{R}^{m_1\times m_1}$ is the identity matrix, $\Delta_1 = \Delta_2$ is the 1D FD Laplacian (endorsed with the periodic boundary conditions), and $\otimes$ denotes the Kronecker product of matrices, see Definition 1. We say that the Kronecker rank of both A in (15) and A_Δ in (17) equals to 2.

Notice that the $m_1 \times m_1$ Laplacian matrices for the Neumann and periodic boundary conditions in 1D read as

$$\Delta_N = \frac{1}{h^2}\begin{bmatrix} -1 & 1 & \dots & 0 & 0 \\ 1 & -2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -2 & 1 \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix} \quad \text{and} \quad \Delta_P = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & \dots & 0 & 1 \\ 1 & -2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -2 & 1 \\ 1 & 0 & \dots & 1 & -2 \end{bmatrix}, \tag{18}$$

respectively.

In the d-dimensional case we have the similar Kronecker rank-d representations. For example, in the case $d = 3$ the "periodic" Laplacian $M_d \times M_d$ matrix A_Δ takes a form

$$A_\Delta = A_{1,P} \otimes I_2 \otimes I_3 + I_1 \otimes A_{2,P} \otimes I_3 + I_1 \otimes I_2 \otimes A_{3,P},$$

such that its Kronecker rank equals to 3, and similar for the arbitrary $d \geq 3$.

3.3 | Fast matrix assembling for the stochastic part

The Kronecker form representation of the "stochastic" term in (13) further denoted by $\hat{A}_s$ is more involved. For given stochastically chosen distribution of overlapping cells G_s, $s = 1, \dots, L^2$, we construct the minimal nonoverlapping decomposition of the full covered grid domain $\hat{G} = \cup_{s=1}^{L^2} G_s$ colored by gray in Figure 1 (we have $a(x) = 1$ for $x \in \hat{G}$ and $a(x) = \lambda$ for $x \in \Omega \setminus \hat{G}$) in a form of a union of elementary square cells S_k, $k = 1, \dots, K$, $K \geq L^2$, each of the grid-size $\overline{m}_0 \times \overline{m}_0$,

$$\hat{G} = \cup_{k=1}^{K} S_k. \tag{19}$$

Here $\overline{m}_0 = 2^p + 1$, and $p = 0, 1, 2, \dots$, is fixed as above by relation $m_1 - 1 = (m_s - 1)2^p$, see (9). Recall that the grid with m_s grid points specifies the the construction of stochastic realization (coarse grid), while the possibly finer grid with m_1 points defines the FEM discretization space. Hence, in the case $p = 0$ we have $m_1 = m_s$, while in general there holds $m_1 \geq m_s$.

In this construction, the nonoverlapping elementary cells S_k for different k are allowed to have the only common edges of size $\overline{m}_0$. Notice that in the case of nonoverlapping decomposition (2) the set of cells $\{S_k\}$ coincides with the initial set $\{G_s\}$ which allows to maximize the size $\overline{m}_0 \times \overline{m}_0$ of each S_k, $k = 1, \dots, L^2$, to the largest possible, that is, to $\overline{m}_0 = m_0 2^p + 1$.

To finalize the matrix generation procedure for $\hat{A}_s$, we define the local $\overline{m}_0 \times \overline{m}_0$ matrices representing the discrete Laplacian with Neumann boundary conditions,

$$\hat{Q}_{\overline{m}_0} := \text{tridiag}\{1,-2,1\} + \text{diag}\{1,0,\dots,0,1\} \in \mathbb{R}^{\overline{m}_0\times\overline{m}_0},$$

and the diagonal lamped matrix

$$\hat{I}_{\overline{m}_0} := \text{diag}\{1/2, 1, \dots, 1, 1/2\} \in \mathbb{R}^{\overline{m}_0\times\overline{m}_0},$$

see the visualization in (18). Here, we may select $\overline{m}_0 = 2, 3, 5, \dots$ that corresponds to the choice $p = 0, 1, 2, \dots$ in (9). In the case of $\overline{m}_0 \times \overline{m}_0$ matrix with minimal size $\overline{m}_0 = 2$, both discrete Laplacians in (18) simplify to

$$\Delta_N = \frac{1}{h^2}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad \Delta_P = \frac{1}{h^2}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{20}$$

Let the subdomain S_k be supported by the index set $I_k^{(1)} \times I_k^{(2)}$ of size $\overline{m}_0 \times \overline{m}_0$ for $k = 1, \dots, K$. Introduce the block-diagonal matrices $\overline{Q}_k \in \mathbb{R}^{m_1 \times m_1}$ and $\overline{I}_k \in \mathbb{R}^{m_1 \times m_1}$ by inserting matrices $\hat{Q}_{\overline{m}_0}$ and $\hat{I}_{\overline{m}_0}$ as diagonal blocks into $m_1 \times m_1$ zero matrix in the positions $I_k^{(1)} \times I_k^{(1)}$ and $I_k^{(2)} \times I_k^{(2)}$, respectively.

Now the stiffness matrix $\hat{A}_s$ is represented in the form of a Kronecker product sum as follows,

$$\hat{A}_s = \frac{1}{h^2}\left(\sum_{k=1}^{K}(\overline{Q}_k \otimes \bar{I}_k + \bar{I}_k \otimes \overline{Q}_k) + P^{(2)}\right), \tag{21}$$

where

$$P^{(2)} = P^{(1)} \otimes I_{m_1} + I_{m_1} \otimes P^{(1)} \in \mathbb{R}^{M_d \times M_d}$$

is the "periodization" matrix in 2D, where the entries of the "periodization" matrix $P^{(1)} \in \mathbb{R}^{m_1 \times m_1}$ for the case of discretization with Neumann boundary conditions on elementary cells S_k, $k = 1, \dots, K$, are all zeros except (cf. (18))

$$P^{(1)}_{1,m_1} = P^{(1)}_{m_1,1} = 1 \quad \text{and} \quad P^{(1)}_{1,1} = P^{(1)}_{m_1,m_1} = -1.$$

In a d-dimensional case the representation (21) generalizes to a sum of d-factor Kronecker products

$$\hat{A}_s = \frac{1}{h^2}\left(\sum_{k=1}^{K}(\overline{Q}_k \otimes \bar{I}_k \otimes \cdots \otimes \bar{I}_k + \cdots + \bar{I}_k \otimes \cdots \otimes \bar{I}_k \otimes \overline{Q}_k) + P^{(d)}\right), \tag{22}$$

where $P^{(d)}$ is the "periodization" matrix in d dimensions, constructed as the d-term Kronecker sum similar to the case $d = 2$.

The Kronecker product form of (17) and (21) leads to the corresponding Kronecker sum representation for the total stiffness matrix A. This allows an efficient implementation of the matrix assembly and low storage for the stiffness matrix preserving the Kronecker sparsity. Hence, it proves the following storage complexity for the matrix A.

Lemma 1. *Let K be the number of elementary cells in the nonoverlapping decomposition of the domain $\hat{G}$, see (19), then the storage size for the Kronecker factors composing the stiffness matrix A is bounded by*

$$Stor(A) \approx Stor(\hat{A}_s) = O(d\overline{m}_0 K + dm_1).$$

Here, in general, the number K of elementary cells[†] is larger than L^2, and it coincides with L^2 only in the case of nonoverlapping decomposition $\hat{G} = \cup_{s=1}^{L^2} G_s$, where different patches G_s are allowed to have joint pieces of boundary but no overlapping area.

In the general case $d \geq 2$ and $K \geq L^d$, the Kronecker rank of the matrix A is bounded by

$$\text{rank}_{\text{Kron}}(A) \leq d\ K.$$

The Kronecker rank of the stiffness matrix reduces dramatically in two cases:

(a) For the case of nonoverlapping cells G_s, $s = 1, \dots, L^2$, we have

$$\text{rank}_{\text{Kron}}(A) \leq L^d.$$

[†]For example, for cells of minimal size, $\overline{m}_0 \times \overline{m}_0$ with $\overline{m}_0 = 2$, as in (20), we have $K = O(m_1^2)$.

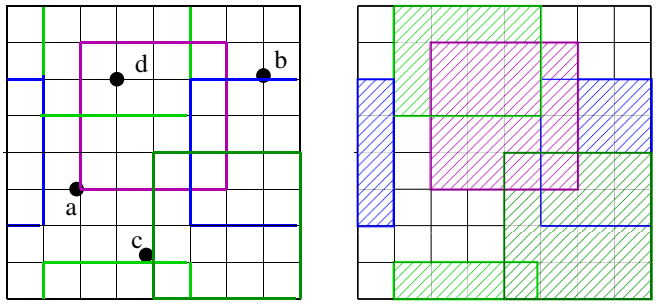

FIGURE 2 Example of the covering domain $\hat{G}$ (right) and the typical locations of sampling points for the grid representation of $\mathbb{A}_h(x_h)$ (left)

(b) In the case of cell-centered locations of subdomains G_s (special case of geometric homogenization) there holds

$$\mathrm{rank}_{\mathrm{Kron}}(A) \le L^{d-1}.$$

The corresponding vector representation $\mathbf{f}_i \in \mathbb{R}^{M_d}$ of the right-hand side $f_i(x)$ is computed by multiplication of the discrete upwind gradient matrix ∇_h with a vector $\mathbf{y}_i \in \mathbb{R}^{M_d}$. Here the vector $\mathbf{y}_i$ represents the multiple of the vector $\mathbf{e}_i$, $i = 1, 2$, and the equation coefficient $\mathbb{A}^{(n)} = \mathbb{A}(x) = \mathrm{diag}\{a^{(n)}(x), a^{(n)}(x)\}$, discretized on the grid Ω_h, that is, each block-entry of the "discretized" matrix coefficient $\mathbb{A}(x) \mapsto A_h(x_h)$ is given by an M_d-vector array with $M_d = m_1^2$,

$$A_h(x_h) = \mathrm{diag}\{a^{(n)}(x_h), a^{(n)}(x_h)\}, \quad a^{(n)}(x_h)_{|x_h \in \Omega_h} \in \mathbb{R}^{M_d}.$$

Hence, we finally arrive at

$$\mathbf{f}_i = (1-\lambda)\nabla_h \cdot \mathbf{y}_i, \quad \mathbf{y}_i = [\mathbf{y}_i(x_h)] \in \mathbb{R}^{M_d} \ \text{ with } \ \mathbf{y}_i(x_h) = A_h(x_h)\mathbf{e}_i, \quad x_h \in \Omega_h,$$

for $i = 1, 2$. Specifically, given the grid-point $x_h \in \Omega_h$, the corresponding diagonal value of $A_h(x_h)$ is defined by $a^{(n)}(x_h)$, see (5). Here the variable part $\hat{a}^{(n)}(x_h)$, describing the jumping coefficient, is assigned by 1 for interior points in $\hat{G}$, (d), by $1/2$ for interface points, (b), (the angle equals to $\pi/2$), by $3/4$ for the "interior" L-shaped corners, (c), (the angle equals to $3\pi/4$) and by $1/4$ for the "exterior" corner of $\hat{G}$, (a), (the angle equals to $\pi/4$), see points (d), (b), (c), and (a) in Figure 2, respectively. This figure corresponds to $L = 2$, the discretization parameter $n_0 = 4$ and periodic completion of the geometry. One observes the complicated shape of the strongly jumping coefficients.

3.4 | Preconditioned CG iteration

Let the RHS in (10) satisfy $\langle \mathbf{f}, 1\rangle = 0$, then for a fixed m, the equation

$$A^{(n)}\mathbf{u} = (\lambda A_\Delta + (1-\lambda)A_s^{(n)})\mathbf{u} = \mathbf{f}, \tag{23}$$

has the unique solution under the same constraints $\langle \mathbf{u}, 1\rangle = 0$. We solve this equation by the PCG iteration (routine *pcg* in MATLAB library) with the preconditioner

$$B = \frac{1+\lambda}{2}A_\Delta + \delta I = \frac{1+\lambda}{2}\Delta_h + \delta I,$$

where $\delta > 0$ is a small regularization parameter introduced only for stability reasons (can be ignored in the theory) and I is the $M_d \times M_d$ identity matrix.

It can be proven that the condition number of preconditioned matrix is uniformly bounded in m_1, L and in the number of stochastic realizations $n = 1, \dots, N$. The particular estimates on the condition number in terms of a parameter λ can be derived by introducing the average coefficient

$$a_0(x) = \frac{1}{2}(a^+(x) + a^-(x)),$$

where $a^+(x)$ and $a^-(x)$ are chosen as *majorants and minorants* of $a^{(n)}(x)$ in (4), respectively. The following simple result holds.

Lemma 2. *Given the preconditioner B with $\delta = 0$, then the condition number of the preconditioned matrix $B^{-1}A^{(n)}$ is bounded by*

$$cond\{B^{-1}A^{(n)}\} \le C\lambda^{-1}.$$

Proof. Lemma 4.1 in Reference 18 shows that the preconditioner A_0 generated by the coefficient $a_0(x) = \frac{1}{2}(a^+(x) + a^-(x))$ allows the condition number estimate

$$\mathrm{cond}\{A_0^{-1}A^{(n)}\} \le C\max\frac{1+q}{1-q}, \quad \text{with} \quad q := \max(a^+(x) - a_0(x))/a_0(x) < 1.$$

The preconditioner B corresponds to the choice $a^+(x) = 1$ and $a^-(x) = \lambda$, hence, we obtain $a_0(x) = \frac{1+\lambda}{2}$ and the result follows. ∎

The PCG solver for the system of equations (16) with the shifted discrete Laplacian as the preconditioner demonstrates robust convergence with the rate $q \ll 1$. In the practically interesting case $\alpha \approx 0.5$ we found that q does not depend on λ. This can be explained by the fact that in this case the total overlap in all subdomains covers the large portion of the computational box Ω. In all numerical examples considered so far the number of PCG iterations was smaller than 10 for the residual stopping criteria $\delta = 10^{-8}$. We use the univariate grid size $m_1 = m_s$, corresponding to the choice $p = 0$ in (9) which is fine enough to resolve geometry for larger L.

To complete this section we notice that the numerical complexity of the presented algorithm scales at least quadratically in L, that is, by $O(L^2)$. On the other hand the larger parameter $\alpha > 0$, that controls the density of filled in area in random media (covered domain $\hat{G}$), enforces the larger number K of elementary square cells S_k each corresponding to the one Kronecker product term in the matrix assembling process. Hence this leads to the larger numerical cost of the order $O(dK)$ for the matrix generation, with the lower bound $K \ge L^d$, see the discussion in Lemma 1.

We demonstrate the numerical performance of our scheme in Section 5.

4 | ASYMPTOTIC CONVERGENCE TO THE STOCHASTIC AVERAGE

In this section, we describe the computational scheme for calculation of the homogenized coefficient matrix for each stochastic realization.

4.1 | Computational scheme for the stochastic average

For fixed stochastic realizations specifying the variable part in the 2×2 coefficient matrix $\hat{\mathbb{A}}^{(n)}(x)$, $n = 1, \dots, N$, we consider the problems

$$-\lambda\Delta\phi_i - (1-\lambda)\nabla \cdot \hat{\mathbb{A}}^{(n)}(\cdot)(\mathbf{e}_i + \nabla\phi_i) = 0, \tag{24}$$

for $i = 1, 2$. The RHS in Equation (24), rewritten in the canonical form (8), reads as

$$f_i(x) = (1-\lambda)\nabla \cdot \hat{\mathbb{A}}^{(n)}(x)\mathbf{e}_i.$$

Taking into account (4), where the diagonal of $\hat{\mathbb{A}}^{(n)}(x)$ is defined in terms of the scalar function $\hat{a}^{(n)}(x)$, we arrive at

$$f_1(x) = (1-\lambda)\frac{\partial\hat{a}^{(n)}(x)}{\partial x_1}, \quad f_2(x) = (1-\lambda)\frac{\partial\hat{a}^{(n)}(x)}{\partial x_2}.$$

Figure 3 illustrates an example of the calculated (reshaped) RHS vector in $\mathbb{R}^{m_1\times m_1}$ and the respective solution ϕ_1 for $L = 12$, $m_1 = 193$, and $m_0 = 16$.

Fixed L, for the particular realization $\mathbb{A}^{(n)}$, by definition, the averaged coefficient matrix $\overline{\mathbb{A}}_L^{(n)} = \overline{\mathbb{A}}^{(n)} = [\bar{a}_{ij}^{(n)}] \in \mathbb{R}^{2\times 2}$, $i, j = 1, 2$, with the constant entries is given by

$$\overline{\mathbb{A}}^{(n)}\mathbf{e}_i = \int_\Omega \mathbb{A}^{(n)}(x)(\mathbf{e}_i + \nabla\phi_i)dx, \tag{25}$$

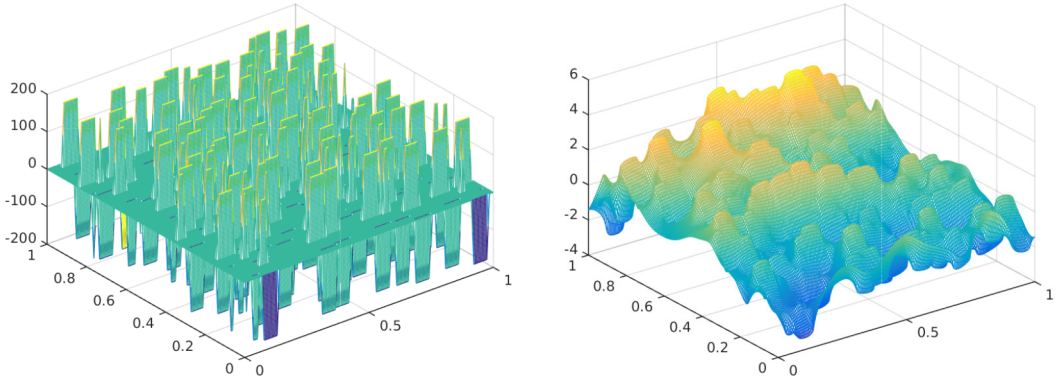

FIGURE 3 Right-hand side (left) and the solution ϕ_1 (right) for $L = 12$, $m_1 = 193$, $m_0 = 16$

which implies the representation for matrix elements

$$\bar{a}^{(n)}_{L,ij} \equiv \bar{a}^{(n)}_{ij} = \int_\Omega [(\lambda I_{2\times 2} + (1-\lambda)\hat{\mathbb{A}}^{(n)}(x))(\mathbf{e}_i + \nabla\phi_i)]_j dx, \quad i,j = 1,2.$$

The latter leads to the entry-wise representation of the matrix $\overline{\mathbb{A}}^{(n)} = [\bar{a}^{(n)}_{ij}]$, $i,j = 1,2$,

$$\begin{aligned}
\bar{a}^{(n)}_{11} &= \int_\Omega a^{(n)}(x)\left(\frac{\partial\phi_1}{\partial x_1} + 1\right)dx,\\
\bar{a}^{(n)}_{12} &= \int_\Omega a^{(n)}(x)\frac{\partial\phi_1}{\partial x_2}dx,\\
\bar{a}^{(n)}_{21} &= \int_\Omega a^{(n)}(x)\frac{\partial\phi_2}{\partial x_1}dx,\\
\bar{a}^{(n)}_{22} &= \int_\Omega a^{(n)}(x)\left(\frac{\partial\phi_2}{\partial x_2} + 1\right)dx.
\end{aligned} \tag{26}$$

The representation (26) ensures the symmetry of the homogenized matrix $\overline{\mathbb{A}}^{(n)}$, that is, $\bar{a}^{(n)}_{ij} = \bar{a}^{(n)}_{ji}$. Indeed, we calculate the difference between the scalar product of the first equation in Equation (24) with ϕ_2,

$$\langle\lambda\nabla\phi_1 + (1-\lambda)\hat{\mathbb{A}}^{(n)}\nabla\phi_1, \nabla\phi_2\rangle - (1-\lambda)\langle\nabla\cdot\hat{\mathbb{A}}^{(n)}(\cdot)\mathbf{e}_1, \phi_2\rangle = 0,$$

and the second equation in Equation (24) with ϕ_1,

$$\langle\lambda\nabla\phi_2 + (1-\lambda)\hat{\mathbb{A}}^{(n)}\nabla\phi_2, \nabla\phi_1\rangle - (1-\lambda)\langle\nabla\cdot\hat{\mathbb{A}}^{(n)}(\cdot)\mathbf{e}_2, \phi_1\rangle = 0,$$

and get the relation

$$\left\langle\frac{\partial\hat{a}^{(n)}}{\partial x_1}, \phi_2\right\rangle - \left\langle\frac{\partial\hat{a}^{(n)}}{\partial x_2}, \phi_1\right\rangle = 0,$$

which then implies the desired property via integration by parts, and taking into account the relation (5),

$$\left\langle a^{(n)}, \frac{\partial\phi_2}{\partial x_1}\right\rangle = \left\langle a^{(n)}, \frac{\partial\phi_1}{\partial x_2}\right\rangle.$$

In numerical implementation, we apply the Galerkin scheme for FEM discretization of Equation (24) its RHS. We use the same quadrature rule for computation of integrals in (26) thus preserving the symmetry in the matrix $\mathbb{A}^{(n)}$ inherited from the exact variational formulation (see argument above and Section 4.3 for the more detailed discussion).

Integrals over Ω in (25), (26) for the matrix entries $(\overline{\mathbb{A}}^{(n)})_{i,j}$, $i,j = 1,2$, are calculated (approximately) by the scalar product of the N-vector of all-ones with the discrete representation of integrand on the grid Ω_h, see Figure 2.

TABLE 1 Symmetry in the matrix $\overline{\mathbb{A}}_L^{(n)}$, with fixed n, versus residual stopping criteria δ

Tol. δ	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}
$\|\mathbb{A}-\mathbb{A}^T\|$	10^{-6}	3×10^{-7}	10^{-7}	3×10^{-9}	10^{-10}	3.6×10^{-11}	10^{-11}	10^{-12}	5.7×10^{-15}

To complete this section, we check numerically that the FEM discretization scheme preserves the symmetry in the matrix $\overline{\mathbb{A}}_L^{(n)}$ for fixed L if the discrete system of equations (10) is solved accurately enough. Table 1 demonstrates that the symmetry in the matrix $\overline{\mathbb{A}}_L^{(n)}$ with fixed n is recovered on the level of residual stopping criteria $\delta > 0$ in the preconditioned iteration for solving the discrete system of equations. For this calculation we set $L = 4$, $m_0 = 8$, $\alpha = 0.5$, and $\lambda = 0.2$.

4.2 | Asymptotic of systematic error and standard deviation

The set of numerical approximations $\{\overline{\mathbb{A}}_L^{(n)}\}$ to the homogenized matrix $\mathbb{A}_{\text{hom}}$ is calculated by (25) for the sequence $\{\mathbb{A}_L^{(n)}(x)\}$ of $n = 1, \dots, N$ realizations, where N is large enough, and the artificial period L defines the size of RVEs. For a fixed L, the approximation $\overline{\mathbb{A}}_L^N$ is computed as the *empirical average* of the sequence $\{\overline{\mathbb{A}}_L^{(n)}\}_{n=1}^N$,

$$\overline{\mathbb{A}}_L^N = \frac{1}{N}\sum_{n=1}^{N} \overline{\mathbb{A}}_L^{(n)}. \tag{27}$$

By the law of large numbers we have that the empirical average converges almost surely to the *ensemble average* (expectation)

$$\langle \overline{\mathbb{A}}_L \rangle_L = \lim_{N\to\infty} \overline{\mathbb{A}}_L^N. \tag{28}$$

Furthermore, by qualitative homogenization theory, as the artificial period $L \to \infty$, this converges to the homogenized matrix

$$\mathbb{A}_{\text{hom}} := \lim_{L\to\infty} \langle \overline{\mathbb{A}}_L \rangle_L. \tag{29}$$

In what follows, we use the entrywise notation for $d \times d$ matrices $\mathbb{A} = [a_{ij}]$, $i, j = 1, \dots, d$, for example, $\langle \overline{\mathbb{A}}_L \rangle = [\bar{a}_{L,ij}]$ and $\overline{\mathbb{A}}_L^{(n)} = [\bar{a}_{L,ij}^{(n)}]$, and so forth.

In terms of square expectations, the convergence rate for the computable quantities can be estimated by, see Reference 8,

$$\left\langle |\overline{\mathbb{A}}_L^N - \mathbb{A}_{\text{hom}}|^2 \right\rangle_L^{1/2} \le \frac{C_1}{\sqrt{N}} L^{-d/2} + C_2 L^{-d} \log^d L. \tag{30}$$

We numerically study the asymptotic of both terms on the RHS of (30) separately by considering the *random part* of the error,

$$\text{var}_L^{1/2}(\overline{\mathbb{A}}_L) = \left\langle \left| \mathbb{A}_{\text{hom}} - \langle \overline{\mathbb{A}}_L \rangle_L \right|^2 \right\rangle_L^{1/2} \le C_1 L^{-d/2}, \tag{31}$$

and the *systematic error*

$$\left| \mathbb{A}_{\text{hom}} - \langle \overline{\mathbb{A}}_L \rangle_L \right| \le C_2 L^{-d} \log^d L, \tag{32}$$

where $\langle \overline{\mathbb{A}}_L \rangle_L$ is approximated by $\langle \overline{\mathbb{A}}_L^{(N)} \rangle_L$ for large enough N.

4.3 | Covariances of the homogenized matrix in the form of quartic tensor

Let $\langle \cdot \rangle_L$ be an ensemble of uniformly elliptic symmetric coefficient fields on the d-dimensional hypercube $[0, L)^d$ with periodicity constraints. Assume that it is invariant under translation (stationary) and under the group $\mathcal{G}$ of all orthogonal

transformations R of $\mathbb{R}^d$ that leave the (hyper-)cube $[0,L)^d$ invariant (this is generated by rotations in one of the Cartesian two-dimensional planes and reflections along any Cartesian hyperplane) in the sense of (A1) below. In case of isotropic (i.e., scalar) coefficient fields $\mathbb{A}(x)$, (A1) turns into

$$\mathbb{A}(R\cdot) \text{ and } \mathbb{A} \text{ have the same distribution under } \langle\cdot\rangle_L,$$

which is certainly the case for the ensembles we consider numerically.

Let X be a finite-dimensional space of functions on the periodic cell $[0,L)^d$ of side-length L with square-integrable gradients, for example, coming from continuous, piecewise affine Finite Elements. For a given realization $\mathbb{A}(x) = \mathbb{A}^{(n)}(x)$ (see (4) and (24)) of the coefficient field and any direction $i = 1,\dots,d$, we consider $\phi_i \in X$ defined through

$$\forall\, \psi \in X \quad \int_{[0,L)^d} \nabla\psi \cdot \mathbb{A}(\mathbf{e}_i + \nabla\phi_i) = 0, \tag{33}$$

where $\mathbf{e}_i$ denotes the unit vector in direction i. If X contains the constant functions (as would be the case for the Finite Element space), ϕ_i has to be normalized to be unique, for example, by imposing $\int_{[0,L)^d}\phi_i = 0$, but this should be irrelevant since we are only interested in $\nabla\phi_i$. If X is indeed a Finite Element space, and if $\{\psi_\alpha\}_{\text{nodes}\alpha}$ denotes the standard basis of piecewise linear functions, then the (stiffness) matrix $A = [a_{\alpha\beta}]$ and the right-hand side $\mathbf{f} = [f_\alpha]$ are given by

$$a_{\alpha\beta} = \int_{[0,L)^d} \nabla\psi_\alpha \cdot \mathbb{A}\nabla\psi_\beta \quad \text{and} \quad f_\alpha = -\int_{[0,L)^d} \nabla\psi_\alpha \cdot \mathbb{A}\mathbf{e}_i. \tag{34}$$

Here it is important to treat periodicity correctly: In practice, one identifies functions on $[0,L)^d$ with functions on $\mathbb{R}^d$ that are periodic in each (Cartesian) argument of period L, hence if the node α is such that one of the adjacent triangles crosses the boundary of the periodic cell $[0,L)^d \subset \mathbb{R}^d$, then there is a piece of ϕ_α that appears on the other side. If a quadrature rule is used for computing the stiffness matrix, it is important that the same one is used for approximation of the RHS.

Let us consider the $d\times d$ matrix $\overline{\mathbb{A}}_L = [\bar{a}_{L,ij}] = \overline{\mathbb{A}}_L(\mathbb{A})$ defined through (see also (25))

$$\bar{a}_{L,ij} := \mathbf{e}_j \cdot \int_{[0,L)^d} \mathbb{A}(\mathbf{e}_i + \nabla\phi_i), \tag{35}$$

(where again, the same quadrature rule should be used). Then we have for every realization

$$\overline{\mathbb{A}}_L \text{ is symmetric, that is, } \bar{a}_{L,ij} = \bar{a}_{L,ji}. \tag{36}$$

Let us consider the ensemble average $\langle\overline{\mathbb{A}}_L\rangle_L$, which by the law of large numbers is given by (see also (28))

$$\langle\overline{\mathbb{A}}_L\rangle_L = \lim_{N\uparrow\infty} \frac{1}{N}\sum_{n=1}^{N} \overline{\mathbb{A}}_L^{(n)}, \tag{37}$$

almost surely, where $\overline{\mathbb{A}}_L^{(n)}$ come via (35) from independent realizations $\mathbb{A} = \mathbb{A}^{(n)}$ according to the distribution $\langle\cdot\rangle_L$. Suppose that the finite-dimensional space X is invariant under reflections in the coordinate directions in the sense of (A2) below. This imposes a more serious restriction on the Finite Element space, namely that it is based on a subdivision of the torus $[0,L)^d$ into axi-parallel cubes (instead of triangles) and that the function space on each cube is spanned by functions that are multilinear in the Cartesian coordinates (as opposed to affine). If this condition is satisfied, then we have

$$\langle\overline{\mathbb{A}}_L\rangle_L \text{ is isotropic, that is } \langle\bar{a}_{L,ij}\rangle_L = \lambda_L\delta_{ij}, \tag{38}$$

for some $\lambda_L \in (0,L)$.

We are interested in the covariances of the entries of $\overline{\mathbb{A}}_L$, and note that by the law of large numbers

$$\mathrm{cov}_{\langle\cdot\rangle_L}[\bar{a}_{L,ij},\bar{a}_{L,i'j'}] := \Big\langle (\bar{a}_{L,ij}-\langle\bar{a}_{L,ij}\rangle_L)(\bar{a}_{L,i'j'}-\langle\bar{a}_{L,i'j'}\rangle_L)\Big\rangle_L$$
$$= \lim_{N\uparrow\infty}\frac{1}{N-1}\sum_{n=1}^{N}\left(\bar{a}^{(n)}_{L,ij}-\frac{1}{N}\sum_{m=1}^{N}\bar{a}^{(m)}_{L,ij}\right)\left(\bar{a}^{(n)}_{L,i'j'}-\frac{1}{N}\sum_{m'=1}^{N}\bar{a}^{(m')}_{L,i'j'}\right).$$

More precisely, we are interested in its rescaled version

$$\overline{Q}_{L,iji'j'} := L^d\mathrm{cov}_{\langle\cdot\rangle_L}[\bar{a}_{L,ij},\bar{a}_{L,i'j'}]$$

which is easier to understand as the four-linear form

$$\overline{Q}_L(\eta,\xi,\eta',\xi') := L^d\mathrm{cov}_{\langle\cdot\rangle_L}[\eta\cdot\overline{\mathbb{A}}_L\xi,\eta'\cdot\overline{\mathbb{A}}_L\xi'].$$

We claim that it has the invariance property

$$\overline{Q}_L(R\eta,R\xi,R\eta',R\xi') = \overline{Q}_L(\eta,\xi,\eta',\xi'). \tag{39}$$

In the case of $d=2$, this implies that $\overline{Q}_L$ is just characterized by three different numbers:

$$\overline{Q}_L(e_1,e_1,e_1,e_2) = \overline{Q}_L(e_1,e_1,e_2,e_1) = \overline{Q}_L(e_1,e_2,e_1,e_1) = \overline{Q}_L(e_2,e_1,e_1,e_1) = 0, \tag{40}$$

$$\overline{Q}_L(e_1,e_2,e_2,e_2) = \overline{Q}_L(e_2,e_1,e_2,e_2) = \overline{Q}_L(e_2,e_2,e_1,e_2) = \overline{Q}_L(e_2,e_2,e_2,e_1) = 0, \tag{41}$$

$$\overline{Q}_L(e_1,e_2,e_1,e_2) = \overline{Q}_L(e_1,e_2,e_2,e_1) = \overline{Q}_L(e_2,e_1,e_1,e_2) = \overline{Q}_L(e_2,e_1,e_2,e_1), \tag{42}$$

$$\overline{Q}_L(e_1,e_1,e_2,e_2) = \overline{Q}_L(e_2,e_2,e_1,e_1), \tag{43}$$

$$\overline{Q}_L(e_1,e_1,e_1,e_1) = \overline{Q}_L(e_2,e_2,e_2,e_2). \tag{44}$$

Proofs of the properties of quartics tensor are presented in Appendix A2.

5 | NUMERICAL STUDY OF STOCHASTIC HOMOGENIZATION

In this section, we estimate numerically the mean constant coefficient in the system (8) depending on L and other model parameters at the limit of $N\to\infty$, see References 5 and 7 for the respective problem setting.

5.1 | Tests on performance of the numerical method

In this paragraph we demonstrate the numerical performance of our numerical scheme.

Recall that the homogenization problem is solved in the unit square $\Omega=[0,1]^2$ with the grid size $m_s\times m_s$, where $m_s = m_0 L+1$. Due to tensor-based construction of the stiffness matrix and sparse representation of matrix entities, in our numerical experiments using MATLAB, the largest number of generated homogenization cells in the domain Ω reaches the value up to $L^2=128^2$. It corresponds to the problem (vector) size $M_d = 263169$ ($m_s = 513$ with $m_0=4$).

Figure 4 illustrates examples of distributions of L^2 randomly located (overlapping) cells specifying the equation coefficient in the cases of moderate and large size of the *RVEs* for $L=4,8,\dots,128$, used in the study of asymptotic of empirical variance/average versus the size of the RVE, L.

Table 2 presents the central processing unit (CPU) times for generating the stiffness matrix, the RHS, and for the solution of the discretized system for the case of overlapping inclusions, for tolerance $\varepsilon=10^{-8}$. Number of inclusions (L^2) varies from 16 to 16,384. The latter is computed on a mesh of size 513×513. We observe that matrix generation takes the

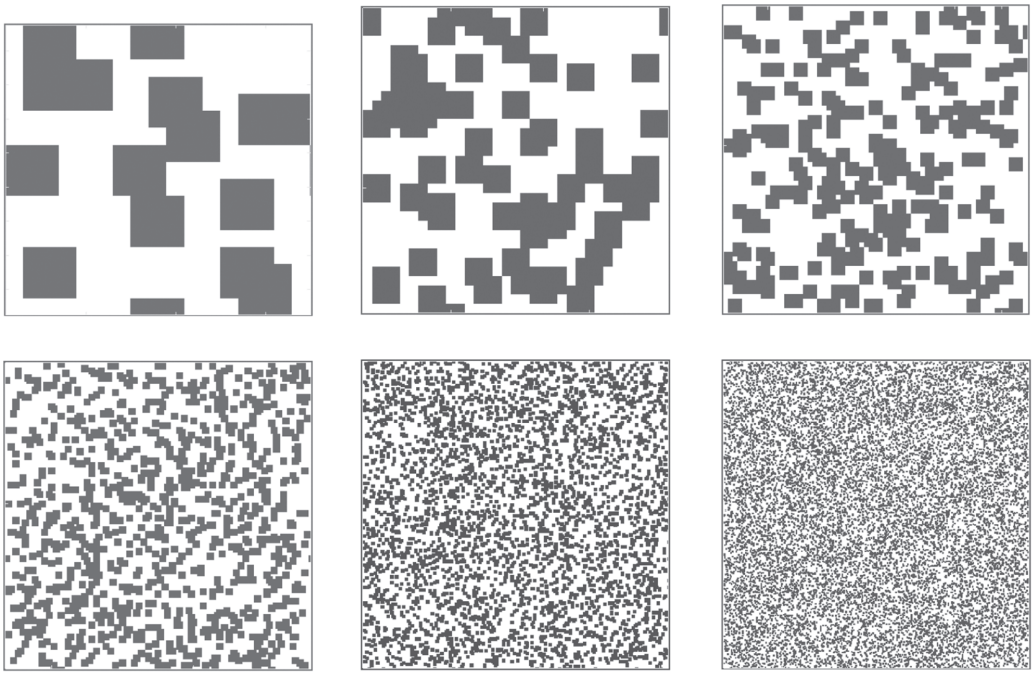

FIGURE 4 Realization examples of a stochastic process with L^2 overlapping cells for $L = 4, 8, 16$ (top) and $L = 32, 64, 128$ (bottom), $m_0 = 4$, $\alpha = 1/4$

L^2	m/m^2	Matrix	RHS	PCG time
4^2	17/289	0.012	0.01	0.006
8^2	33/1089	0.06	0.045	0.137
16^2	65/4225	0.34	0.19	0.11
32^2	129/16,641	3.0	0.8	0.5
64^2	257/66,049	36	3.7	2.6
128^2	513/263,169	561	22	13.8

TABLE 2 CPU times (s) versus the number of inclusions (i.e., L^2) for generating the stiffness matrix, the RHS, and for the solution of the discretized system for the case of overlapping inclusions

Note: PCG stopping criteria is $\varepsilon = 10^{-8}$.
Abbreviations: CPU, central processing unit; PCG, preconditioned conjugate gradient; RHS, right-hand side.

dominating time. In case of large grids, the time growth factor for increasing values of L approaches 16 which corresponds to a product of $O(L^2)$ summation terms in (21) and $O(m_1^2)$ operations for the computation of each Kronecker product.

5.2 | Systematic error and empirical variance versus L

In what follows, we numerically check the theoretical convergence rate (30), in form of checking (31) and (32) separately.

Figure 5 serves to illustrate the asymptotic convergence of the systematic error see (32), at the limit of large L. Since we do not have access to the ensemble averages $\langle \overline{\mathbb{A}}_L \rangle_L$, we take empirical averages $\overline{\mathbb{A}}_L^N$ for large enough N, (cf. (29)) as a proxy. Furthermore, due to the fact that $\mathbb{A}_{\text{hom}}$ is not computable we compare the differences in $\langle \bar{a}_{L,11} \rangle_L$ computed on a sequence of increasing values of L.

Figure 5 shows the differences in matrix entries $\langle \bar{a}_{L,11} \rangle_L - \langle \bar{a}_{2L,11} \rangle_{2L}$ for increasing sizes of the RVE, that is, for $L = 2^p$, $p = 1, 2, \ldots, 5$, computed with $N = 10^5$, $N = 10^4$, and $N = 10^3$ stochastic realizations. It illustrates the asymptotic convergence of the *systematic error*, see (32),

$$\left| \langle \overline{\mathbb{A}}_L \rangle_L - \mathbb{A}_{\text{hom}} \right| \lesssim L^{-d} \log^d L,$$

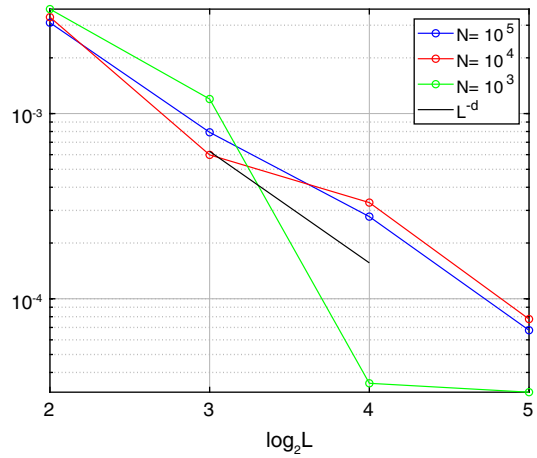

FIGURE 5 Systematic error $\langle \bar{a}_{L,11}\rangle_L - \langle \bar{a}_{2L,11}\rangle_{2L}$ versus L, for increasing $L = 2^p$, $p = 1, 2, \ldots, 5$ computed for the largest number of realizations $N = 10^5$.

TABLE 3 Systematic error $\langle \bar{a}_{L,11}\rangle_L - \langle \bar{a}_{2L,11}\rangle_{2L}$ versus L, for increasing $L = 2^p$, $p = 1, 2, \ldots, 5$, computed for $N = 10^5, 10^4$, and 10^3 realizations.

L/N	10^5	10^4	10^3
4	0.003095	0.003316	0.003665
8	0.000792	0.000598	0.001198
16	0.000277	0.000330	−0.000034
32	0.000067	0.000077	0.000031

at the limit of large L. Calculations are performed with $m_0 = 4$, $\alpha = \frac{1}{4}$, and $\lambda = 0.4$ and tolerance $\varepsilon = 10^{-8}$. The black line corresponds to the curve L^{-d}, with $d = 2$.

The largest size of RVE with $p = \log_2 L = 5$, presented in statistics in Figure 5, corresponds to the most left picture in the bottom row in Figure 4. In this example the jumping coefficient contains 32^2 (overlapping) inclusions, and the discrete problem of size $m_s^2 = 129^2$ (i.e., vector size is 16,641) has been solved $N = 10^5$ times for providing the representative statistics. For readers convenience, Table 3 presents the same data visualized in Figure 5.

We now turn to the random error, that is, the variance of $\langle \overline{\mathbb{A}}_L\rangle_L = [\bar{a}_{ij}]$. Since by symmetry considerations, $\langle \bar{a}_{11}\rangle_L = \langle \bar{a}_{22}\rangle_L$ and $\langle \bar{a}_{12}\rangle_L = 0$, we monitor

$$\langle (\bar{a}_{L,11} - \bar{a}_{L,22})^2\rangle_L^{1/2} \quad \text{and} \quad \langle (\bar{a}_{L,12})^2\rangle_L^{1/2},$$

which should decay as $1/L$. Again, since we do not have access to the ensemble averages defining the standard deviation, we replace them by their empirical approximation for large enough N,

$$\left(\frac{1}{N}\sum_{n=1}^{N} (\bar{a}_{L,12}^{(n)})^2\right)^{1/2} \approx \langle (\bar{a}_{L,12})^2\rangle_L^{1/2} \le C_1 L^{-d/2}, \tag{45}$$

$$\left(\frac{1}{N}\sum_{n=1}^{N} (\bar{a}_{L,11}^{(n)} - \bar{a}_{L,22}^{(n)})^2\right)^{1/2} \approx \langle (\bar{a}_{L,11} - \bar{a}_{L,22})^2\rangle_L^{1/2} \le C_1 L^{-d/2}. \tag{46}$$

Figure 6 presents the empirical average (standard deviation) for $\bar{a}_{L,12}$ and $\bar{a}_{L,11} - \bar{a}_{L,22}$ versus $L = 2, 4, \ldots, 64$, corresponding to $N = 500, 10^3$, and 10^4 realizations ($\alpha = \frac{1}{4}$, $\lambda = 0.4$), confirming the estimates (45) and (46). Notice that starting from $N = 500$ the values of empirical average for different number of realizations practically coincide. The results for the homogenized matrix for the largest $p = 6$ presented in Figure 6 correspond to ensembles with 4096 overlapping cells, and the size of the discrete problem (i.e., vector/matrix size) is 66,049. These systems of equations have been solved 10^4 times. An example of realization for $p = 6$ is shown in Figure 4 (middle bottom panel).

Related to Figure 6, Table 4 presents standard deviation of $\bar{a}_{L,12}$ (left) and $\bar{a}_{L,11} - \bar{a}_{L,22}$ (right) versus L, with $L = 2^p$, $p = 1, 2, \ldots, 6$, for $N = 500$ and $N = 10^4$. We choose the following discretization and model parameters $m_0 = 4$, $\varepsilon = 10^{-8}$, $\alpha = \frac{1}{4}$, and $\lambda = 0.4$.

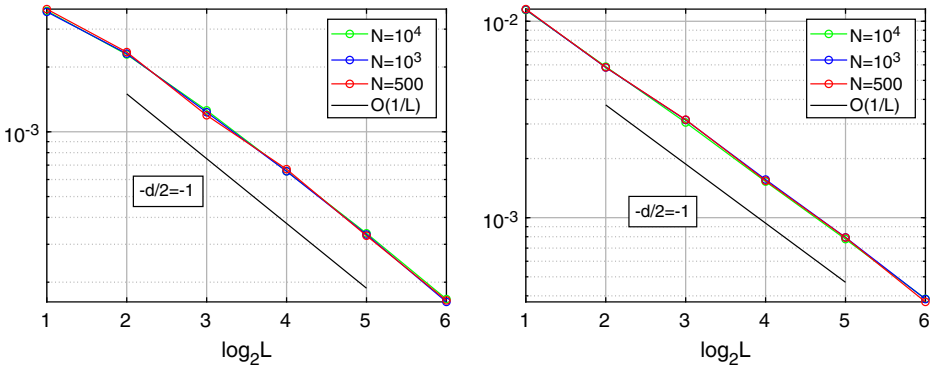

FIGURE 6 Standard deviation of $\bar{a}_{L,12}$ (left) and of $(\bar{a}_{L,11} - \bar{a}_{L,22})$ (right) versus L, with $L = 2^p$, $p = 1, 2, \ldots, 6$, for $N = 500, 10^3$, and 10^4

	$\bar{a}_{L,12}$		$\bar{a}_{L,11} - \bar{a}_{L,22}$	
L/N	10^4	500	10^4	500
2	0.003643	0.003716	0.011402	0.011531
4	0.002287	0.002346	0.005875	0.005828
8	0.001258	0.001193	0.003052	0.003156
16	0.000656	0.000670	0.001527	0.001543
32	0.000337	0.000329	0.000778	0.000792
64	0.000167	0.000165	0.000386	0.000372

TABLE 4 Standard deviation of $\bar{a}_{L,12}$ (left) and $\bar{a}_{L,11} - \bar{a}_{L,22}$ (right) versus L, with $L = 2^p$, $p = 1, 2, \ldots, 6$, for $N = 500$ and $N = 10^4$

We summarize that numerical results presented in Figures 5 and 6 (see also Tables 3 and 6) confirm the asymptotic convergence rates of the systematic error (32) and the empirical average (45), (46) in the size L of RVE.

5.3 | The asymptotic of quartic tensor versus leading order variances

In this section, we consider the convergence of the *quartic tensor* $\overline{Q}_L$, representing *covariances* of the matrix $\overline{\mathbb{A}}_L$, to its leading order variances $Q_{\rm hom}$, see Section 4.3. For the large number of realizations N, the computable approximation, $\overline{Q}_L^N \in \mathbb{R}^{2\times2\times2\times2}$, to the scaled quartic tensor is defined by

$$\overline{Q}_L^N = \frac{L^d}{N-1} \sum_{n=1}^{N} \left(\overline{\mathbb{A}}_L^{(n)} - \frac{1}{N} \sum_{n'=1}^{N} \overline{\mathbb{A}}_L^{(n')} \right)^{\otimes 2}, \tag{47}$$

so that by the central limit theorem

$$\overline{Q}_L := \lim_{N\to\infty} \overline{Q}_L^N.$$

The equivalent matrix representation of $\overline{Q}_L^N$ is obtained by setting the operation $\otimes 2$ in (47) as the Kronecker product of matrices (see Definition 1), further denoted by

$$\overline{Q}_L^N = [\overline{q}_{L,ij}] \in \mathbb{R}^{4\times4}, \quad i,j = 1, \ldots, 4.$$

In our numerical tests we shall check the asymptotic behavior

$$\left\langle \left| \frac{L^d}{N-1} \sum_{n=1}^{N} \left(\overline{\mathbb{A}}_L^{(n)} - \frac{1}{N} \sum_{n'=1}^{N} \overline{\mathbb{A}}_L^{(n')} \right)^{\otimes 2} - Q_{\rm hom} \right|^2 \right\rangle_L \lesssim L^{-d} \ln^d L,$$

which can be expected at the limit of large size L of the RVE, see References 8 and 16.

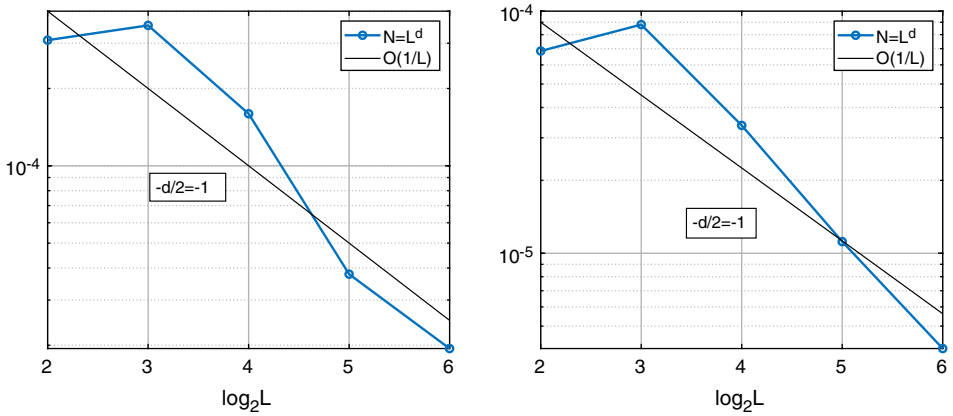

FIGURE 7 $\overline{q}_{L,11} - \overline{q}_{2L,11}$ (left) and $\overline{q}_{L,14} - \overline{q}_{2L,14}$ (right) versus $L = 2^p$, $p = 1, \dots, 6$, $N = L^d$

It is worth to note that the quartic tensor $\overline{Q}_L$ can be calculated at no further cost than the effective homogenized matrix $\overline{\mathbb{A}}_L$.

Figure 7 shows the diagonal elements, $\overline{q}_{L,11} - \overline{q}_{2L,11}$ and $\overline{q}_{L,14} - \overline{q}_{2L,14}$, in quartic tensor $\overline{Q}_L^N$ versus increasing size of RVE $L = 2^p$, $p = 1, \dots, 6$, (see (47)), for $N = L^d$. Figure 7 confirms convergence rate of $\overline{q}_{L,11} - \overline{q}_{2L,11}$ in RVE L as $O(L^{d/2})$.

6 | CONCLUSIONS

We present the numerical scheme for discretization and solution of 2D elliptic equations with strongly varying piecewise constant coefficients arising in stochastic homogenization of multiscale random materials. The resulting large linear system of equations is solved by the PCG iteration with the convergence rate that is independent of the grid size and of the variation in jumping coefficients. For a fixed size of the RVE, our approach allows to avoid the generation of the new FEM space in each stochastic realization. For every realization, fast assembling of the FEM stiffness matrix is performed by agglomerating the Kronecker tensor products of 1D FEM discretization matrices. The resultant stiffness matrix is maintained in a sparse matrix format.

Our numerical scheme allows to investigate the asymptotic convergence rate of significant quantities of stochastic homogenization process in the course of a large number of realizations (of the order of $N = 10^5$) and for large sizes of the RVEs up to $L = 128$, corresponding to the number of inclusions 16,384 and matrix size $513^2 \times 513^2$. Note that for every realization a new matrix generation and solution of the respective linear system is performed.

Our numerical experiments study the asymptotic convergence rate of systematic error and standard deviation in the size of RVE, rigorously established in Reference 8. In particular, we confirm in various numerical tests the theoretical asymptotic estimates, see Section 4.2, concerning the convergence rate $O(1/L)$ for the empirical variance at the limit of large L, but with a moderate number of stochastic realizations N, and the asymptotic $CL^{-2}\ln^2 L$ in the case of large N. Our numerical scheme applies to a stationary ergodic problems where the asymptotic convergence is subject to the central limit theorem. The model elliptic problem is posed in periodic setting. The randomness is encoded by geometry which is described by the set of square inclusions specified on the Cartesian grid, which allows the fast matrix generation by using a sum of the Kronecker product terms. The respective FEM discretizations are constructed on the refined tensor grid.

The asymptotic behavior of covariances of the homogenized matrix in the form of quartic tensor are studied numerically. In particular, we consider the asymptotic of the quartic tensor versus the leading order variances, computed for the large number of stochastic realizations up to $N = 10^4$. In this way, the asymptotic $O(L^{-d}\ln^d L)$, for $d = 2$, is confirmed on a sequence of increasing sizes of the RVE, up to $L = 64$.

The stochastic characteristics of the system are analyzed for a range of intrinsic model parameters like the number of realizations, the size of periodic RVE, the jump-ratio in the stochastic equation coefficients (contrast) and various grid discretization parameters. The presented numerical scheme allows to perform large scale simulations using MATLAB

on a moderate computer cluster. The uniform convergence of the PCG iteration can be expected for a class of nondegenerate elliptic equations which excludes situations with the very small parameter $\lambda \to 0$. The tensor-based numerical techniques for matrix generation presented in this article can be extended to 3D and higher dimensional problems. The

application of our approach is limited by the class of random geometries described on the tensor product Cartesian grids. Here we apply these techniques to the scalar-valued linear elliptic problems in divergent form.

ACKNOWLEDGEMENTS

The authors would like to thank Julian Fischer (IST Austria, Wien) and Ronald Kriemann (MPI MiS, Leipzig) for useful discussions concerning the problem setting.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Boris N. Khoromskij https://orcid.org/0000-0002-5853-5521

REFERENCES

1. Anantharaman A, Costaouec R, Le Bris C, Legoll F, Thomines F. Introduction to numerical stochastic homogenization and the related computational challenges: some recent developments. In: Bao W, Du Q, editors. Lecture notes series, institute for mathematical sciences. Volume 22. Singapore, Asia: National University of Singapore, 2011; p. 197–272.
2. Kanit T, Forest S, Galliet I, Mounoury V, Jeulin D. Determination of the size of the representative volume element for random composites: Statistical and numerical approach. Int J Solids Struct. 2003;40:3647–3679.
3. Le Bris C, Legoll F. Examples of computational approaches for elliptic, possibly multiscale PDEs with random inputs. J Comput Phys. 2017;328:455–473.
4. Gloria A, Otto F. An optimal variance estimate in stochastic homogenization of discrete elliptic equations. Ann Probab. 2011;39(3):779–856.
5. Gloria A, Otto F. The corrector in stochastic homogenization: near-optimal rates with optimal stochastic integrability; 2016. http://arxiv.org/abs/1510.08290.
6. Gloria A, Otto F. Quantitative estimates on the periodic approximation of the corrector in stochastic homogenization. ESAIM: Proc Surv. 2015;48:80–97.
7. Gloria A, Otto F. An optimal error estimate in stochastic homogenization of discrete elliptic equations. Ann Appl Probab. 2012;22(1):1–28. http://de.arxiv.org/abs/1203.0908.
8. Gloria A, Neukamm S, Otto F. Quantification of ergodicity in stochastic homogenization: optimal bounds via spectral gap on Glauber dynamics. Invent Math. 2015;199(2):455–515. https://doi.org/10.1007/s00222-014-0518-z.
9. Fischer J. The choice of representative volumes in the approximation of the effective properties of random materials; 2018. arXiv:1807.00834.
10. Khoromskaia V, Khoromskij BN Otto F. A numerical primer in 2D stochastic homogenization: CLT scaling in the representative volume element. Preprint 47/2017, Max-Planck Institute for Mathematics in the Sciences, Leipzig, Germany; 2017.
11. Khoromskij BN, Wittum G. Numerical solution of elliptic differential equations by reduction to the interface. Research monograph, LNCSE, No. 36. New York, NY: Springer-Verlag, 2004.
12. Khoromskij BN. Tensor numerical methods in scientific computing. Research monograph. Berlin, Germany: De Gruyter Verlag, 2018.
13. Khoromskij BN, Schwab C. Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs. SIAM J Sci Comput. 2011;33(1):364–385.
14. Dolgov S, Khoromskij BN, Litvinenko A, Matthies HG. Polynomial chaos expansion of random coefficients and the solution of stochastic partial differential equations in the tensor train format. SIAM/ASA J Uncertain Quantif. 2015;3(1):1109–1135.
15. Schwab C, Gittelson CJ. Sparse tensor discretization of high-dimensional parametric and stochastic PDEs. Acta Numer. 2011;20:291–467.
16. Duerinckx M, Gloria A, Otto F. The structure of fluctations in stochastic homogenization; 2017. arXiv: 1602.01717v3.
17. Cances E, Ehrlacher V, Legoll F, Stamm B, Xiang S. An embedded corrector problem for homogenization. Part II: Algorithms and discretization; 2018. http://arxiv.org/abs/1810.09885v1.
18. Khoromskij BN, Repin S. A fast iteration method for solving elliptic problems with quasi-periodic coefficients. Russ J Numer Anal Math Modell. 2015;30(6):329–344. arXiv:1510.00284, 2015.

How to cite this article: Khoromskaia V, Khoromskij BN, Otto F. Numerical study in stochastic homogenization for elliptic partial differential equations: Convergence rate in the size of representative volume elements. *Numer Linear Algebra Appl*. 2020;27:e2296. https://doi.org/10.1002/nla.2296

APPENDIX A1. NUMERICAL ANALYSIS OF THE FEM APPROXIMATION ERROR

We tested convergence of the solutions on a sequence of dyadic refined grids, for the fixed configuration of coefficients and the right-hand side given by $f = \sin(2\pi x)\cos(6\pi y)$. Test examples are performed for $L = 2, 4, 8$, corresponding to 4, 16, and 64 bumps in the coefficients, respectively. For each fixed L, we compare the solution vectors $\mathbf{u}_p$ calculated on a sequence of five dyadic refined grids with the grid size $M_d = m_p^2 = M_{d,p}$, with $m_p = 2^{4+p} - 1$, $p = 1, \dots, 5$, equal to $M_{d,p} = 31^2$, 63^2, 127^2, 255^2, 511^2, and 1023^2, respectively. The matrix size is given by $M_d \times M_d$. A FEM interpolation error in the $H^r(\Omega)$-norm is expected of the order of $O(m^{-\beta+r})$ for $\beta \in (0, 2]$ and $r \in [0, 1]$, where $h = O(1/m)$ and β measures the regularity of the solution $u \in H^\beta(\Omega)$.

Table A1 shows the decay of the solution error in L_2-norm estimated on a sequence of dyadic refined grids, and for different values of $L = 2, 4, 8$. The solution is supposed to be represented on a sequence of grid in the form $\mathbf{u}_p = \mathbf{c}_0 + \mathbf{c}_1 h_p^\beta$ up to higher order terms. We expect the asymptotic error behavior $O(h^\beta)$ with $1 \le \beta \le 2$, where in our case β is close to $3/2$ that corresponds to decay factor $2\sqrt{2} \approx 2.8$. The latter can be expected in the case of reduced regularity in the solution caused by cusps by the multiple interior corners in the configuration of coefficient jumps. The respective convergence rate in the H^1-norm is of the order of $O(h^{\beta-1})$.

Figure A1 illustrates examples of the solution $\mathbf{u}$ discretized over $m \times m$ grid with the univariate grid size $m = 255$ and for $L = 2$ and $L = 4$.

Figure A2 represents differences in solutions on pair of $m \times m$ grids with $m = 127, 255$ (left) and $m = 5, 111, 023$ (right) for $f = \sin(2\pi x)\cos(6\pi y)$ and fixed $L = 4$. One can observe the expected increase in the approximation error towards the interior corners in the geometry specifying the jumping coefficient function. The error decays by a factor about 10 which agrees with the expected decay by 2.8^2. For ease of comparison both solutions are interpolated onto the common grid with $m = 63$.

TABLE A1 Differences in the relative norms of solutions $\|\mathbf{u}_p - \mathbf{u}_{p-1}\|_2 / \|\mathbf{u}_{p-1}\|_2$, $p = 1, \dots, 5$, on dyadic refined grids computed in L_2-norm for $L = 2, 4, 8$ and $\alpha = 0.5$, $\lambda = 0.1$.

Grid size m_p	63	127	255	511	1023
Vector size $M_{d,p}$	3969	16,129	65,025	261,121	1,046,529
$L = 2$	—	0.0051	0.0015	5.03e−04	1.806e−04
$L = 4$	—	0.0057	0.0020	7.13e−04	2.638e−04
$L = 8$	—	—	0.0035	1.40e−03	5.257e−04

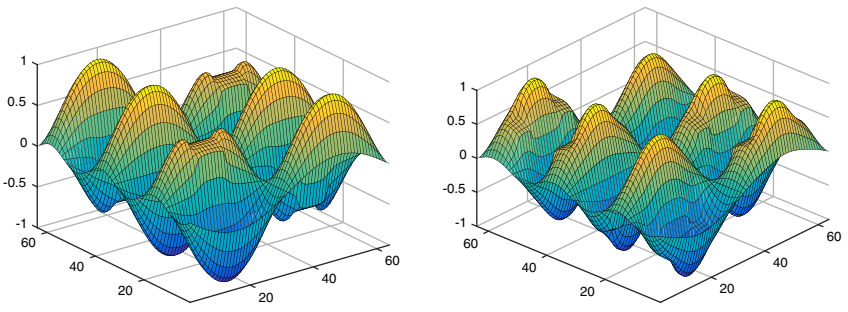

FIGURE A1 Examples of solutions u for $m = 255$, where $L = 2$ (left) and $L = 4$ (right)

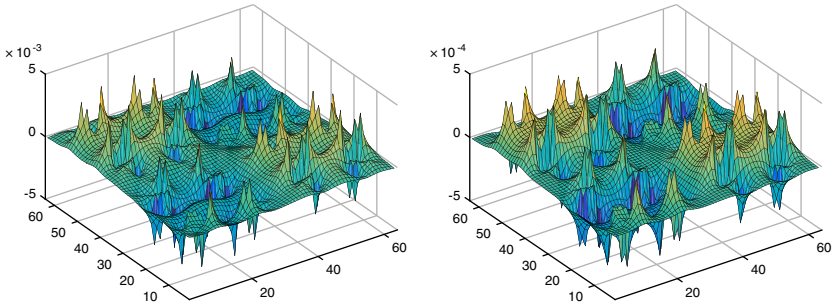

FIGURE A2 Differences in solutions on the $m \times m$ grids with $m = 127, 255$ (left) and $m = 511, 1023$ (right) for $L = 4$

APPENDIX A2. PROOFS FOR THE PROPERTIES OF THE QUARTIC TENSOR

ARGUMENT FOR (36). According to (33), definition (35) may be reformulated as

$$\bar{a}_{L,ij} = \int_{[0,L)^d} (\mathbf{e}_j + \nabla\phi_j) \cdot \mathbb{A}(\mathbf{e}_i + \nabla\phi_i),$$

so that the symmetry of $\mathbb{A}$ yields the symmetry of $\overline{\mathbb{A}}_L$.

ARGUMENT FOR (38). Identifying the points on the periodic cell with $[0,L)^d \subset \mathbb{R}^d$, let $\mathcal{G}$ denote the subgroup of the orthogonal group that leaves $[0,L)^d$ invariant. According to our assumption, for any $R \in \mathcal{G}$,

$$R^t\mathbb{A}(R\cdot)R \text{ and } \mathbb{A} \text{ have the same distribution under } \langle\cdot\rangle_L, \tag{A1}$$

where $R^t\mathbb{A}(R\cdot)R$ denotes the matrix field $[0,L)^d \ni x \mapsto R^t\mathbb{A}(Rx)R$. According to our assumption on X we have

$$\psi \in X \;\Rightarrow\; \psi(R\cdot) \in X, \tag{A2}$$

where $\psi(R\cdot)$ denotes the function $[0,L)^d \ni x \mapsto \psi(Rx)$.

For a fixed vector $\xi \in \mathbb{R}^d$, we consider $\phi_\xi := \xi_i\phi_i$ (Einstein's summation convention) and note that in view of (33), for given realization $\mathbb{A} = \mathbb{A}^{(n)}$, the function $\phi_\xi = \phi_\xi(\mathbb{A})$ (at least up to additive constants) is characterized by

$$\forall\, \psi \in X \quad \int_{[0,L)^d} \nabla\psi \cdot \mathbb{A}(\xi + \nabla\phi_\xi) = 0, \tag{A3}$$

We now argue that ϕ transforms under $R \in \mathcal{G}$ as follows

$$\phi_{R\xi}(\mathbb{A}; Rx) = \phi_\xi(R^t\mathbb{A}(R\cdot)R; x). \tag{A4}$$

Indeed, this relies on the straightforward orthogonal transformation rule

$$\begin{aligned}&\int_{[0,L)^d} \nabla_y[\psi(R^Ty)] \cdot \mathbb{A}(y)(R\xi + \nabla\phi_{R\xi}(y))dy\\ &\overset{y=Rx}{=} \int_{[0,L)^d} \nabla\psi(x) \cdot R^t\mathbb{A}(Rx)R(\xi + \nabla_x[\phi_{R\xi}(Rx)])dx.\end{aligned}$$

According to (A2) and (A3) (with ξ replaced by $R\xi$) the left-hand side vanishes for all $\psi \in X$; hence by the characterization (A3) applied to the RHS, we obtain (A4).

We now argue note that from (A4) we obtain for the gradient $\nabla\phi_{R\xi}(\mathbb{A}; Rx) = \nabla\phi_\xi(R^t\mathbb{A}(R\cdot)R; x)$ and thus for the flux $q_\xi(\mathbb{A}; x) := \mathbb{A}(\xi + \nabla\phi_\xi(\mathbb{A}; x))$ the transformation rule

$$q_{R\xi}(\mathbb{A}; Rx) = Rq_\xi(R^t\mathbb{A}(R\cdot)R; x),$$

from which we obtain by definition (35) that

$$\overline{\mathbb{A}}_L(\mathbb{A})R\xi = R\overline{\mathbb{A}}_L(R^t\mathbb{A}(R\cdot)R)\xi. \tag{A5}$$

According to (A1) this yields the following invariance property for the symmetric matrix $\langle\overline{\mathbb{A}}_L\rangle_L$

$$\langle\overline{\mathbb{A}}_L\rangle_L R\xi = R\langle\overline{\mathbb{A}}_L\rangle_L\xi.$$

Since this holds for all $\xi \in \mathbb{R}^d$ and all $R \in \mathcal{G}$, by an argument of elementary algebra, we obtain the isotropy of $\langle\overline{\mathbb{A}}_L\rangle_L$, cf (38).

ARGUMENT FOR (39). This follows from (A5) in form of

$$(R\eta) \cdot \overline{\mathbb{A}}_L(\mathbb{A})R\xi = \eta \cdot \overline{\mathbb{A}}_L(R^t\mathbb{A}(R\cdot)R)\xi$$

and from (A1).

ARGUMENT FOR (44)–(40). The four identities in (42) on the variances just follow from the symmetry of the underlying random variable $\overline{\mathbb{A}}_L$, compare (36), in form of

$$\mathbf{e}_1 \cdot \overline{\mathbb{A}}_L\mathbf{e}_2 = \mathbf{e}_2 \cdot \overline{\mathbb{A}}_L\mathbf{e}_1.$$

The identity (43) follows from the symmetry of the covariance in its two arguments. The vanishing of the eight entries stated in (40) and (41) follows from (39) applied to the reflection $R \in \mathcal{G}$ given by $R\mathbf{e}_1 = -\mathbf{e}_1$ and $R\mathbf{e}_2 = \mathbf{e}_2$. The identity (44) follows from (39) applied to the reflection $R \in \mathcal{G}$ given by $R\mathbf{e}_1 = \mathbf{e}_2$ and $R\mathbf{e}_2 = \mathbf{e}_1$.