HAL
open science

Tensor-based multiscale method for diffusion problems in quasi-periodic heterogeneous media

Quentin Ayoul-Guilmard, Anthony Nouy, Christophe Binetruy

To cite this version:

Quentin Ayoul-Guilmard, Anthony Nouy, Christophe Binetruy. Tensor-based multiscale method for diffusion problems in quasi-periodic heterogeneous media . ESAIM: Mathematical Modelling and Numerical Analysis, 2018, 52 (3), pp.869-891. 10.1051/m2an/2018022 . hal-02922752

HAL Id: hal-02922752

https://hal.science/hal-02922752v1

Submitted on 26 Aug 2020

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.

ESAIM: M2AN 52 (2018) 869–891
https://doi.org/10.1051/m2an/2018022

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

TENSOR-BASED MULTISCALE METHOD FOR DIFFUSION PROBLEMS IN QUASI-PERIODIC HETEROGENEOUS MEDIA☆

Quentin Ayoul-Guilmard[1,*], Anthony Nouy[2] and Christophe Binetruy[1]

Abstract. This paper proposes to address the issue of complexity reduction for the numerical simulation of multiscale media in a quasi-periodic setting. We consider a stationary elliptic diffusion equation defined on a domain D such that $\overline{D}$ is the union of cells $\{\overline{D_i}\}_{i\in I}$ and we introduce a two-scale representation by identifying any function $v(x)$ defined on D with a bi-variate function $v(i, y)$, where $i \in I$ relates to the index of the cell containing the point x and $y \in Y$ relates to a local coordinate in a reference cell Y. We introduce a weak formulation of the problem in a broken Sobolev space $V(D)$ using a discontinuous Galerkin framework. The problem is then interpreted as a tensor-structured equation by identifying $V(D)$ with a tensor product space $\mathbb{R}^I \otimes V(Y)$ of functions defined over the product set $I \times Y$. Tensor numerical methods are then used in order to exploit approximability properties of quasi-periodic solutions by low-rank tensors.

Mathematics Subject Classification. 15A69, 35B15, 65N30

Received October 23, 2017. Accepted March 27, 2018.

1. Introduction

Heterogeneous periodic media are increasingly common in the industry, particularly owing to the use of architectured microstructure (*e.g.* composite materials). Their complex behaviour calls for thorough and expensive experimental investigations. As an alternative, numerical simulations involve fine-scale models which often require heavy computations. Periodicity assumption on the medium means that all its information is contained within a single cell, which can be exploited in practical resolutions (*e.g.* homogenisation). Nonetheless, the need to withdraw this assumption arises with situations such as defect impact studies; this raises a computational challenge.

To the best of the authors' knowledge, there exist currently two families of approaches available to tackle such problems more efficiently than brute fine-scale computation—such as typical finite element method. First is the set of multiscale methods such as multiscale finite element method (MsFEM) [3, 5, 12, 22], heterogeneous multiscale method (HMM) [2, 5, 11] or patch methods [9, 17, 18, 34, 35]. Although these are designed to address the issue of multiscale complexity, they are intended for broader purposes than our particular case of interest;

☆The authors gratefully acknowledge the financial support from the Fondation CETIM.

Keywords and phrases: Quasi-periodicity, tensor approximation, discontinuous Galerkin, multiscale, heterogeneous diffusion.

[1] École Centrale de Nantes, GeM UMR CNRS 6183, Nantes, France.
[2] École Centrale de Nantes, LMJL UMR CNRS 6629, Nantes, France.

* Corresponding author: quentin.ayoul-guilmard@centraliens-nantes.net

Article published by EDP Sciences
© EDP Sciences, SMAI 2018

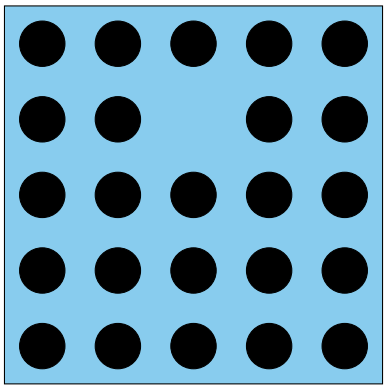


FIGURE 1. Periodic medium with one defect.

as such, they fail to achieve the complexity reduction one could expect from a quasi-periodicity assumption. Secondly, progress has been made over the past few years towards exploitation of quasi-periodicity in stochastic homogenisation methods. These works focus on computational cost reduction of classical stochastic homogenisation through suitable assumption on the stochastic model [4, 6, 24], as well as specific variance reduction schemes [7, 28, 29] and an adaptation to special quasirandom structures used in atomistic simulations [25]. The aforementioned methods exploit quasi-periodicity in order to reduce the number of supercell problems to solve, comparatively to classical stochastic homogenisation. Consequently, they are cost-efficient to compute good approximations of homogenised quantities of a material ideally periodic yet perturbed by random imperfections. They do not, however, reduce complexity of a given deterministic, quasi-periodic supercell problem such as those they involve. To address this computational bottleneck, various adaptations of aforementioned general multiscale methods have been developed (*e.g.* [26]). Several noteworthy approaches based on reduced basis methods (whose principle is explained in [31]) have been developed to exploit quasi-periodic patterns, such as [1, 8, 27]. We propose here a multiscale method designed specifically to address such quasi-periodic problems.

Section 2 will introduce the reference problem, a two-scale representation and the related discontinuous Galerkin formulation. In Section 3, we identify the problem as an operator equation in a Hilbert tensor space and we use a greedy algorithm for the construction of a sequence of low-rank approximations of the solution. Finally, Section 4 illustrates the efficiency of the proposed method through a number of representative numerical experiments.

2. Reference problem and discontinuous Galerkin formulation

Let $D \subset \mathbb{R}^d$ be an open rectangular cuboid. We consider a stationary diffusion equation

$$-\nabla \cdot (K \nabla u) = f \quad \text{in } D, \tag{2.1}$$

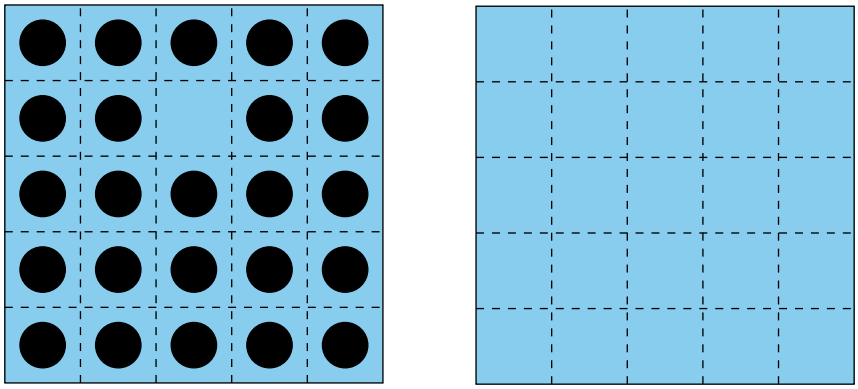
with periodic boundary conditions, where K is the diffusion (or "conductivity") coefficient and f is a source term. An example of quasi-periodic heterogeneous two-phase material is given in Figure 1.

We assume that $K \in \mathrm{L}^\infty(D)$ and $f \in \mathrm{L}^2(D)$. A weak solution

$$u \in \mathrm{H}^1_{per}(D) = \left\{ v \in \mathrm{H}^1(D) : v \ D\text{-periodic} \right\}$$

of (2.1) is such that

$$\int_D K \nabla u \cdot \nabla v = \int_D f v, \quad \forall v \in \mathrm{H}^1_{per}(D).$$



(A) (B)

FIGURE 2. Mesoscopic mesh $\mathcal{T}(D)$ of domain D. (A) $\overline{D} = \bigcup_{i\in I} \overline{D}_i$. (B) $\mathcal{T}(D)$.

2.1. Mesoscopic discretisation

We introduce a partition of $\overline{D}$ into closed domains $\{\overline{D_i}\}_{i\in I}$, where I is a totally ordered set. The subsets $(D_i)_{i\in I}$ are open and identical up to a translation. They will be called "cells" and are chosen so as to fit the quasi-periodically repeated patterns (see Fig. 2A).

The set of cells defines a mesoscopic mesh $\mathcal{T}(D) = \{D_i : i \in I\}$ over D (see Fig. 2B). We denote by $\mathcal{F}(D)$ the set of *faces*, by $\mathcal{F}_e(D) = \{F \in \mathcal{F}(D) \ : \ F \subset \partial D\}$ the set of *external faces*, and by $\mathcal{F}_i(D) = \mathcal{F}(D) \setminus \mathcal{F}_e(D)$ the set of *internal faces*. We define in the same way $\mathcal{F}(D_i)$, the set of faces of cell D_i for any $i \in I$.

For a face $F \in \mathcal{F}_i(D)$, we let $(i, j) \in I^2$ be the unique ordered pair of indices such that $F = \partial D_i \cap \partial D_j$. We denote by n_F the unit normal vector of face F, outwards from D_i. For a function v defined over $D_i \cup D_j$, we denote—if it exists—the trace $(v_{|D_i})_{|F}$ on F of its restriction $v_{|D_i}$. We then define the average operator over face F $\{\cdot\}_F$ by $\{v\}_F = \frac{1}{2}((v_{|D_i})_{|F} + (v_{|D_j})_{|F})$, and the jump operator over face F $[\cdot]_F$ by $[v]_F = ((v_{|D_i})_{|F} - (v_{|D_i})_{|F})$.

For the sake of simplicity, the subscript F will be omitted whenever the face related to is obvious. Periodic boundary conditions allow to extend these definitions to external faces by identifying a face $F = \partial D_i \cap \partial D \in \mathcal{F}_e(D)$ with the opposite face $F' = \partial D_j \cap \partial D \in \mathcal{F}_e(D)$, which we will use below (see Rem. 2.2). For the definition of the normal n_F and the jump operator, we use the convention $i < j$.

We then introduce the broken Sobolev space

$$\mathrm{H}^1\left(\bigcup_{i\in I} D_i\right) = \left\{v \in \mathrm{L}^2(\mathbb{R}) : \forall i \in I, v_{|D_i} \in \mathrm{H}^1(D_i)\right\}.$$

It should be noted that, since the cells D_i are open, $\bigcup_{i\in I} D_i \neq D$.

2.2. Symmetric weighted interior penalty (SWIP) formulation

We make the following assumption on the regularity of the solution.

Assumption 2.1. We assume that the solution u of (2.1) is in $\mathrm{H}^1_{per}(D) \cap \mathrm{H}^2(D)$ so that, for all $F \in \mathcal{F}(D)$,

$$\int_F [u] = 0 \quad \text{and} \quad \int_F [K\nabla u] \cdot n = 0.$$

From Theorem 1, (Sect. 6.3.1, p. 309 of [15]), if $-\nabla \cdot (K\nabla u) = f$ with $K \in \mathrm{C}^1(D)$ and $f \in \mathrm{L}^2(D)$, then $u \in \mathrm{H}^2_{loc}(D)$ whatever the boundary conditions. If u is D-periodic, then $u \in \mathrm{H}^2(D)$ since

$$\left\{v \in \mathrm{H}^2_{loc}(D) : v \text{ is } D\text{-periodic}\right\} = \left\{v \in \mathrm{H}^2(D) : v \text{ is } D\text{-periodic}\right\}.$$

Therefore, Assumption 2.1 is verified in our case if $K \in \mathrm{C}^1(D)$. For Dirichlet and Neumann boundary conditions, we refer the reader to Theorem 3.12, page 119 of [14].

For the discontinuous Galerkin formulation to come, we introduce a subset of $\mathrm{H}^1(\bigcup_{i\in I} D_i)$ defined as

$$V(D) := \left\{ v \in \mathrm{H}^1\left(\bigcup_{i\in I} D_i\right) : \forall i \in I, (\nabla v_{|D_i})_{|\partial D_i} \in \mathrm{L}^2(\partial D_i)^d \right\}.$$

Then the solution u of (2.1) satisfies

$$\forall v \in V(D), \quad a(u,v) - c(u,v) = b(v),$$

where a and c are bilinear forms over $V(D)$ respectively defined by

$$a(u,v) = \sum_{i\in I} \int_{D_i} K\nabla u \cdot \nabla v, \quad c(u,v) = \sum_{F\in\mathcal{F}(D)} \int_F n \cdot \{K\nabla u\}[v],$$

and b is a linear form defined by

$$b(v) = \int_D fv.$$

From Assumption 2.1 and from the D-periodicity of u, we have that u also satisfies

$$\forall v \in V(D), \quad a(u,v) - c(u,v) - c(v,u) + \sum_{F\in\mathcal{F}(D)} \frac{\sigma}{|F|} \int_F [u][v] = b(v), \tag{2.2}$$

with $|F|$ the measure of face F, and with σ a positive penalty parameter. Equation (2.2) corresponds to the symmetric interior penalty (SIP) formulation of (2.1) (see [10] for a detailed explanation), which involves a coercive bilinear form for a sufficiently high value of the penalisation parameter σ.

In the present context, K may show strong heterogeneities. The symmetric weighted interior penalty (SWIP) method [10], a variant of SIP, is designed to account for this by introducing weights in the definition of averages on faces and in the penalty term. For a cell D_i, we let k_i^+ and k_i^- be the constants defined by

$$k_i^- = \inf_{x\in D_i} \lambda_{min}(K(x)) \quad \text{and} \quad k_i^+ = \sup_{x\in D_i} \lambda_{max}(K(x)), \tag{2.3}$$

where $\lambda_{min}(A)$ and $\lambda_{max}(A)$ respectively denote the minimum and maximum eigenvalues of a symmetric matrix A. For a face $F = \partial D_i \cap \partial D_j \in \mathcal{F}(D)$, we define a stabilisation weight ω_F and average weights β_F^- and β_F^+ as

$$\omega_F = \frac{2k_i^+ k_j^+}{k_i^+ + k_j^+}, \quad \beta_F^- = \frac{k_i^+}{k_i^+ + k_j^+}, \quad \beta_F^+ = \frac{k_j^+}{k_i^+ + k_j^+}. \tag{2.4}$$

Then, we redefine the average operator $\{\cdot\}_F$ over F by

$$\{v\}_F = \beta_F^-(v_{|D_i})_{|F} + \beta_F^+(v_{|D_j})_{|F}, \tag{2.5}$$

and we introduce a stabilisation bilinear form

$$s(v,w) = \sum_{F\in\mathcal{F}(D)} \sigma \frac{\omega_F}{|F|} \int_F [w]_F [v]_F. \tag{2.6}$$

The problem with periodic boundary conditions admits infinitely many solutions that differ by a constant. We decide to fix this constant by choosing a particular solution in the kernel of the linear form $\phi(v) = \int_D v$. This is achieved by introducing a symmetric bilinear form

$$m(u,v) = \phi(u)\phi(v)$$

whose left kernel is the kernel of ϕ.

Finally, we achieve a consistent SWIP formulation, *i.e.* the solution $u \in \mathrm{H}^1_{per}(D) \cap \mathrm{H}^2(D)$ of (2.1) verifies

$$\forall v \in V(D), \quad a^{swip}(u,v) = b(v), \tag{2.7}$$

with $a^{swip}(u,v) = a(u,v) - c(u,v) - c(v,u) + s(u,v) + m(u,v)$.

Remark 2.2 (Periodic boundary conditions' enforcement)**.** As explained in Section 2.1, periodic boundary conditions give meaning to an extension of face jump and face average operators to external faces. As far as D-periodic functions are concerned, these external faces can be considered as internal faces. Thus, in formulation (2.7), periodic boundary conditions are weakly enforced through the terms $\frac{\sigma\omega_F}{|F|}\int_F [u][v]$ associated with faces $F \in \mathcal{F}_e(D)$ in the bilinear form $a^{swip}(u,v)$.

2.3. Coercivity

We choose a finite dimensional subspace $V_h(D) \subset V(D)$ and consider the problem whose solution $u_h \in V_h(D)$ satisfies

$$\forall v_h \in V_h(D), \quad a^{swip}(u_h, v_h) = b(v_h). \tag{2.8}$$

As a closed subspace of a Hilbert space, $V_h(D)$ is a Hilbert space itself; therefore, problem (2.8) is well posed if a^{swip} is coercive on $V_h(D)$. Then u_h would be a Galerkin approximation of u.

The bilinear form a^{swip} can be proven to be coercive on $V_h(D)$ for a sufficiently high value of parameter σ in the stabilisation form (2.6) [10, 30]. For meshes of simplices and when using polynomial spaces $V_h(D_i)$, a lower bound for σ can be found in [13]. In this section we provide a lower bound on σ to have the coercivity of the bilinear form a^{swip} on $V_h(D)$, with an explicit expression of the coercivity constant allowing its evaluation for any finite dimensional approximation subspace of $V(D)$.

We equip the broken Sobolev space $\mathrm{H}^1(\bigcup_{i\in I} D_i)$ with the norm $\|\cdot\|_E$ defined by

$$\|v\|_E^2 = a(v,v) + s(v,v) + m(v,v).$$

The application $v \mapsto (a(v,v) + s(v,v))^{1/2}$ defines a semi-norm on $\mathrm{H}^1(\bigcup_{i\in I} D_i)$, and the addition of m ensures that $\|\cdot\|_E$ is a norm. It is, *a fortiori*, a norm on $V_h(D) \subset \mathrm{H}^1(\bigcup_{i\in I} D_i)$.

Proposition 2.3 (Discrete trace inequality)**.** *Let $D_i \in \mathcal{T}(D)$ and $F \in \mathcal{F}(D_i)$. Then*

$$\exists C(V_h(D_i), F) > 0, \forall v \in V_h(D_i), \quad \|\nabla v_{|F}\|_{\mathrm{L}^2(F)^d} \leqslant C(V_h(D_i), F)\|\nabla v\|_{\mathrm{L}^2(D_i)^d},$$

where $C(V_h(D_i), F)$ depends on $V_h(D_i)$ and F.

Proof of Proposition 2.3. Let $i \in I$, $v \in V_h(D_i)$ and $F \in \mathcal{F}(D_i)$. From the definition of $V_h(D_i)$ above, we have

$$\|\nabla v_{|F}\|_{\mathrm{L}^2(F)^d} \leqslant \|\nabla v_{|F}\|_{\mathrm{L}^2(F)^d} + \|\nabla v\|_{\mathrm{L}^2(D_i)^d}.$$

The application $\|\cdot\|_{\mathrm{L}^2(F)^d} + \|\cdot\|_{\mathrm{L}^2(D_i)^d}$ is a norm on the subspace $W = \{\nabla v : v \in V_h(D_i)\}$ of $\mathrm{L}^2(D_i)$. Since W is of finite dimension, this norm is equivalent to $\|\cdot\|_{\mathrm{L}^2(D_i)^d}$ on W, which means that there exists $C(V_h(D_i), F) > 0$, independent of v, such that

$$\|\nabla v_{|F}\|_{\mathrm{L}^2(F)^d} + \|\nabla v\|_{\mathrm{L}^2(D_i)^d} \leqslant C(V_h(D_i), F)\|\nabla v\|_{\mathrm{L}^2(D_i)^d}.$$

□

Before stating the next result, we introduce some notations. We denote the maximum number of faces of elements in $\mathcal{T}(D)$ by $N_{\mathcal{F}} = \max\{\#\mathcal{F}(D_i) : i \in I\}$, the upper bound of face measures by $|\mathcal{F}|^+ = \max\{|F| : F \in \mathcal{F}(D)\}$, the upper and lower bounds for the eigenvalues of the diffusion operator by $k^+_{max} = \max\{k^+_i : i \in I\}$ and $k^-_{min} = \min\{k^-_i : i \in I\}$, the upper bound of the average weights by $\beta_{max} = \max\left\{\max\{\beta^+_F, \beta^-_F\} : F \in \mathcal{F}(D)\right\}$, the lower bound of the weights in the stabilisation form by $\omega_{min} = \min\{\omega_F : F \in \mathcal{F}(D)\}$, and the upper bound of the constant in the discrete trace inequality by $C(V_h(D)) = \max\{C(V_h(D_i), F) : i \in I, F \in \mathcal{F}(D_i)\}$.

Proposition 2.4 (SWIP coercivity)**.** *If*

$$\sigma > \sigma_- := C(V_h(D))^2 \beta^2_{max} N_{\mathcal{F}} |\mathcal{F}|^+ \frac{k^+_{max}}{\omega_{min}} \frac{k^+_{max}}{k^-_{min}}, \tag{2.9}$$

then

$$\forall v \in V_h(D), \quad a^{swip}(v, v) \geqslant \left(1 - \sqrt{\frac{\sigma_-}{\sigma}}\right) \|v\|^2_E, \tag{2.10}$$

i.e. *a^{swip} is coercive with coercivity constant $C_{swip} = 1 - \sqrt{\frac{\sigma_-}{\sigma}}$.*

Proof of Proposition 2.4. First, let us assume

$$\exists C < \frac{1}{2}, \forall v \in V_h(D), \quad c(v, v) \leqslant C(a(v, v) + s(v, v)). \tag{2.11}$$

Consequently, for all $v \in V_h(D)$,

$$\begin{aligned} a^{swip}(v, v) &= a(v, v) + s(v, v) - 2c(v, v) + m(v, v) \\ &\geqslant (1 - 2C)(a(v, v) + s(v, v)) + m(v, v) \\ &\geqslant (1 - 2C)\|v\|^2_E. \end{aligned}$$

Therefore, it is enough that (2.11) holds with $2C = \sqrt{\frac{\sigma_-}{\sigma}}$ to prove (2.10). Since (2.9) would then ensue from the necessary condition $C < \frac{1}{2}$, it would complete the proof.

Let us consider a face $F = \partial D_i \cap \partial D_j \in \mathcal{F}(D)$. We let $\alpha > 0$ and, applying successively Cauchy-Schwarz's and Young's inequalities, we have that

$$\begin{aligned}\int_F n \cdot \{K\nabla v\}[v] &\leqslant \|n \cdot \{K\nabla v\}\|_{\mathrm{L}^2(F)} \|[v]\|_{\mathrm{L}^2(F)} \\ &= \frac{\alpha}{\alpha} \|n \cdot \{K\nabla v\}\|_{\mathrm{L}^2(F)} \|[v]\|_{\mathrm{L}^2(F)} \\ &\leqslant \frac{1}{2\alpha^2} \|n \cdot \{K\nabla v\}\|^2_{\mathrm{L}^2(F)} + \frac{\alpha^2}{2} \|[v]\|^2_{\mathrm{L}^2(F)}.\end{aligned}$$

From Proposition 2.3 and the definitions of the weighted average operator and of k_i^+ (in Eqs. (2.5) and (2.3)), we get

$$\begin{aligned}\|n \cdot \{K\nabla v\}\|^2_{\mathrm{L}^2(F)} &= \|n \cdot \left(\beta_F^+ (K\nabla v)_{|F^+} + \beta_F^- (K\nabla v)_{|F^-}\right)\|^2_{\mathrm{L}^2(F)} \\ &\leqslant \|n \cdot \beta_F^+ (K\nabla v)_{|F^+}\|^2_{\mathrm{L}^2(F)} + \|n \cdot \beta_F^- (K\nabla v)_{|F^-}\|^2_{\mathrm{L}^2(F)} \\ &\leqslant \beta_{max}^2 {k_{max}^+}^2 \left(\|(\nabla v)_{|F^+}\|^2_{\mathrm{L}^2(F)^d} + \|(\nabla v)_{|F^-}\|^2_{\mathrm{L}^2(F)^d}\right) \\ &\leqslant C(V_h(D))^2 \beta_{max}^2 {k_{max}^+}^2 \left(\|(\nabla v)_{|D_i}\|^2_{\mathrm{L}^2(D_i)^d} + \|(\nabla v)_{|D_j}\|^2_{\mathrm{L}^2(D_j)^d}\right).\end{aligned}$$

Now we let $\epsilon > 0$ and choose $\alpha = C(V_h(D))\beta_{max} k_{max}^+ \sqrt{N_{\mathcal{F}}(\epsilon k_{min}^-)^{-1}}$, so that

$$\begin{aligned}\int_F n \cdot \{K\nabla v\}[v] &\leqslant \frac{1}{2\alpha^2} C(V_h(D))^2 \beta_{max}^2 {k_{max}^+}^2 \left(\|(\nabla v)_{|D_i}\|^2_{\mathrm{L}^2(D_i)^d} + \|(\nabla v)_{|D_j}\|^2_{\mathrm{L}^2(D_j)^d}\right) \\ &\quad + \frac{\alpha}{2} \|[v]\|^2_{\mathrm{L}^2(F)} \\ &= \frac{\epsilon k_{min}^-}{2N_{\mathcal{F}}} \|(\nabla v)_{|D_i}\|^2_{\mathrm{L}^2(D_i)^d} + \frac{N_{\mathcal{F}}}{2\epsilon k_{min}^-} C(V_h(D))^2 \beta_{max}^2 {k_{max}^+}^2 \|[v]\|^2_{\mathrm{L}^2(F)}. \qquad (2.12)\end{aligned}$$

Noting that

$$k_{min}^- \|\nabla v\|^2_{\mathrm{L}^2(D_i)^d} \leqslant k_i^- \|\nabla v\|^2_{\mathrm{L}^2(D_i)^d} \leqslant \int_{D_i} \nabla v \cdot K \nabla v$$

and that[1] $\sum_{F = D_i \cap D_j \in \mathcal{F}(D)} (p_i + p_j) = N_{\mathcal{F}} \sum_{i \in I} p_i$, we find

$$\frac{k_{min}^-}{N_{\mathcal{F}}} \sum_{F = D_i \cap D_j \in \mathcal{F}(D)} \left(\|\nabla v\|^2_{\mathrm{L}^2(D_i)^d} + \|\nabla v\|^2_{\mathrm{L}^2(D_j)^d}\right) = k_{min}^- \sum_{i \in I} \|\nabla v\|^2_{\mathrm{L}^2(D_i)^d} \leqslant a(v, v). \qquad (2.13)$$

From the definition of ω_{min} and $|\mathcal{F}|^+$, we also have

$$\frac{\sigma \omega_{min}}{|\mathcal{F}|^+} \sum_{F \in \mathcal{F}(D)} \|[v]\|^2_{\mathrm{L}^2(F)} \leqslant \sum_{F \in \mathcal{F}(D)} \frac{\sigma \omega_F}{|F|} \int_F [v]^2 = s(v, v). \qquad (2.14)$$

[1]Only with periodic boundary conditions (see Rem. 2.2), although inequality (2.13) is still verified without them.

We put together (2.13) and (2.14) in (2.12) and obtain

$$\begin{aligned} c(v,v) = \sum_{F\in\mathcal{F}(D)} \int_F n_F \cdot \{K\nabla v\}[v] &\leqslant \frac{C(V_h(D))^2 \beta_{max}^2 {k_{max}^+}^2 N_{\mathcal{F}} |\mathcal{F}|^+}{2\epsilon k_{min}^- \sigma \omega_{min}} s(v,v) + \frac{\epsilon}{2} a(v,v) \\ &= \frac{\sigma_-}{2\epsilon\sigma} s(v,v) + \frac{\epsilon}{2} a(v,v) \\ &\leqslant C_\epsilon (s(v,v) + a(v,v)), \end{aligned}$$

with $C_\epsilon := \max\left\{\frac{\sigma_-}{2\epsilon\sigma}, \frac{\epsilon}{2}\right\}$. Therefore, (2.11) holds with $C := \inf_{\epsilon>0} C_\epsilon = \frac{1}{2}\sqrt{\frac{\sigma_-}{\sigma}}$, which concludes the proof. □

The result of Proposition 2.4 is of major interest in choosing a suitable value for stabilisation parameter σ: too high a value degrades the performance of the algorithm that will be presented in Section 3.2, due to poor conditioning of discrete operators associated with a^{swip}; on the other hand, σ must be high enough for a^{swip} to be coercive. Consequently, knowledge of lower bound σ_- enables us to set not too low a value for σ. However, one should keep in mind that "$\sigma > \sigma_-$" is only a sufficient condition, since σ_- is not necessarily the lowest value above which σ ensures coercivity. A choice of σ lower than σ_- may improve the performance of the aforementioned algorithm. Alternatively, to improve conditioning while retaining coercivity, one could replace the stabilisation form $s(v,w)$ by $\sum_{F\in\mathcal{F}(D)} \int_F \sigma_F [w][v]$, where $(\sigma_F)_{F\in\mathcal{F}(D)}$ is a set of penalisation parameters defined face-wise. Incidentally, the weights functions ω and β added from SIP to SWIP formulations are a way of tuning the stabilisation face-wise according to conductivity.

It should be noted that, unlike typical discontinuous Galerkin settings, there are two level of discretisation here: first the mesoscopic level, at which the domain is partitioned in "cells" and where discontinuities occur; then the microscopic level, *i.e.* the mesh within each cell, which relates to $V_h(D)$. The characteristic length of the former appears in formula (2.9) as $|\mathcal{F}|^+$, while the latter is accounted for in $C(V_h(D))$, whose computation is discussed below. Section 4 features examples of approximation spaces with their associated trace constant's value.

Remark 2.5 (Trace constant computation)**.** The evaluation of the lower bound σ_- according to formula (2.9) requires the evaluation of $C(V_h(D))$ which, in turn, calls for the value of $C(V_h(D_i), F)$ for all $i \in I$ and $F \in \mathcal{F}(D_i)$. The evaluation of $C(V_h(D_i), F)$, defined by

$$C(V_h(D_i), F)^2 = \max\left\{ \frac{\|\nabla v_{|F}\|^2_{\mathrm{L}^2(F)^d}}{\|\nabla v\|^2_{\mathrm{L}^2(D_i)^d}} : v \in V_h(D_i), \|\nabla v\|_{\mathrm{L}^2(D_i)^d} > 0 \right\},$$

requires computing the maximum eigenvalue of a generalised eigenvalue problem. Let us assume that there exists a diffeomorphism ξ_i which maps D_i onto a reference domain Y, *i.e.* $\xi_i(D_i) = Y$, and that $V_h(D_i) = \{v \circ \xi_i : x \in D_i \mapsto v(\xi_i(x)) : v \in V_h(Y)\}$, with $V_h(Y) \subset \mathrm{H}^1(Y)$. If the domains D_i are obtained by translations of a particular domain $D_{i^\star} = Y$, $i^\star \in I$, then $C(V_h(D_i), F) = C(V_h(Y), F_Y)$, with $F_Y = \xi_i(F)$, is independent of i. If $Y =]0,1[^d$ and ξ_i is an affine transformation, *i.e.* $\xi_i(x) := A_i x + b_i$ for a certain invertible matrix $A_i \in \mathbb{R}^{d\times d}$ with positive determinant and a certain vector $b_i \in \mathbb{R}^d$, then

$$C(V_h(D_i), F)^2 = \max\left\{ \frac{|F|}{|D_i|} \frac{\|A_i^T \nabla v_{|F_Y}\|^2_{\mathrm{L}^2(F_Y)^d}}{\|A_i^T \nabla v\|^2_{\mathrm{L}^2(Y)^d}} : v \in V_h(Y), \|\nabla v\|_{\mathrm{L}^2(Y)^d} > 0 \right\},$$

so that

$$C(V_h(D_i), F) \leqslant \frac{\varsigma_{max}(A_i)}{\varsigma_{min}(A_i)} \frac{|F|^{1/2}}{|D_i|^{1/2}} C(V_h(Y), F_Y)^2,$$

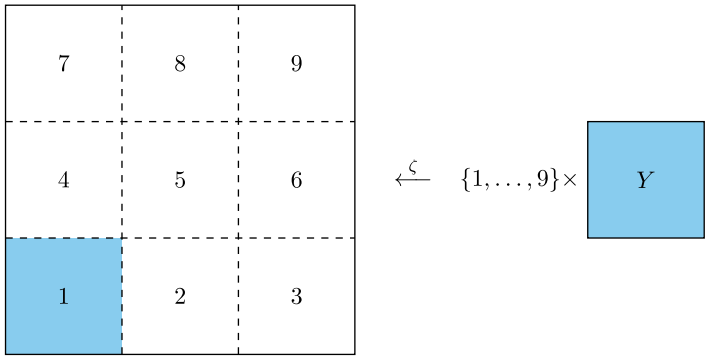


FIGURE 3. Bijection ζ between $I \times Y$ and $\bigcup_{i\in I} D_i$.

where $\varsigma_{max}(A_i)$ and $\varsigma_{min}(A_i)$ are respectively the maximum and minimum singular values of A_i.

Remark 2.6 (K's eigenvalues computation)**.** Evaluation of the bounds $\{k_i^-, k_i^+\}_{i\in I}$ of eigenvalues of K, defined in (2.3), is required for a^{swip}: these bounds yield the face-wise weights ω, β^- and β^+ as expressed in (2.4), used in bilinear forms s and c, and are involved in formula (2.9) of lower bound σ_-.

If such bounds are not explicitly given, they are evaluated numerically. Assuming that $K(x) = \sum_{n\in\mathcal{N}_h(D_i)} K(x_n)\phi_n(x)$ where $\{x_n : n \in \mathcal{N}_h(D_i)\}$ is the set of nodes of the mesh $\mathcal{T}_h(D_i)$, and where $\{\phi_n(x) : n \in \mathcal{N}_h(D_i)\}$ forms a partition of unity, *i.e.*, are non-negative functions such that $\sum_{n\in\mathcal{N}_h(D_i)} \phi_n(x) = 1$ for all x, then $k_i^- = \min\{\lambda_{min}(K(x_n)) : n \in \mathcal{N}_h(D_i\}$ and $k_i^+ = \max\{\lambda_{max}(K(x_n)) : n \in \mathcal{N}_h(D_i)\}$. For a general $K(x)$, it can be approximated under the above form with a sufficiently fine mesh, and k_i^- and k_i^+ are estimated from its approximation.

Although, for the sake of simplicity, we consider numerical examples with a scalar-valued diffusion operator $K \in \mathrm{L}^\infty(D)$, there is no objection to its being matrix-valued, *i.e.* $K \in \mathrm{L}^\infty(D)^{d\times d}$. If K is diagonal, the evaluation cost of $\{k_i^-, k_i^+\}_{i\in I}$ is insignificant—*a fortiori* if it is scalar. If K is not diagonal, the extreme eigenvalues of a d-by-d matrix must be computed at every node. Our simulations found this latter cost, albeit not negligible, to remain small compared to the overall resolution cost. A parallelisation strategy would considerably reduce the cost of these evaluations.

3. Tensor-structured method

3.1. Formulation over a tensor product space

We here assume that the domains D_i, $i \in I$, are obtained by translations of a reference domain $Y = \xi_i(D_i)$, with $\xi_i(x) = x + b_i$ for a certain vector $b_i \in \mathbb{R}^d$. Then there exists a bijection ζ between $I \times Y$ and $\bigcup_{i\in I} D_i$ (see Fig. 3) given by

$$\zeta(i, y) = \xi_i^{-1}(y) = y - b_i, \quad (i, y) \in I \times Y.$$

We define

$$V(Y) := \left\{v \in \mathrm{H}^1(Y) : (\nabla v)_{|\partial Y} \in \mathrm{L}^2(\partial Y)^d\right\}.$$

Then we denote by $X = \mathbb{R}^I \otimes V(Y)$ the tensor space of functions defined on $I \times Y$ which is the linear span of elementary tensors $v^I \otimes v^Y$, with $v^I \in \mathbb{R}^I$ and $v^Y \in V(Y)$. This tensor product space is equipped with an inner

product $\langle \cdot, \cdot \rangle$ and associated norm $\|\cdot\|$ such that

$$\|v^I \otimes v^Y\| = \|v^I\|_{\mathbb{R}^I} \|v^Y\|_{V(Y)}.$$

We denote by Υ the map which associates to a function $v : \bigcup_{i\in I} D_i \to \mathbb{R}$ the function $\Upsilon(v) = v \circ \zeta : I \times Y \to \mathbb{R}$. This allows us to identify a function $v \in V(D)$ with a tensor $\Upsilon(v) \in X$ such that $\Upsilon(v) = \sum_{i\in I} e_i \otimes v_i^Y$, where $\{e_i\}_{i\in I}$ is the canonical orthonormal basis of $\mathbb{R}^I$, and $v_i^Y = v_{|D_i} \circ \xi_i^{-1}$. Noting that

$$\|\Upsilon(v)\|^2 = \sum_{i,j\in I} \langle e_i \otimes v_i^Y, e_j \otimes v_j^Y \rangle = \sum_{i\in I} \|v_i^Y\|^2_{\mathrm{H}^1(Y)} = \sum_{i\in I} \|v_{|D_i}\|^2_{\mathrm{H}^1(D_i)},$$

we have that Υ defines a linear isometry between $V(D)$ and X, with $V(D)$ equipped with the norm $\|\cdot\|_{V(D)}$ defined by $\|v\|^2_{V(D)} = \sum_{i\in I} \|v_{|D_i}\|^2_{\mathrm{H}^1(D_i)}$, which is equivalent to the energy norm $\|\cdot\|_E$. Then a^{swip} can be identified with a bilinear form on $X \times X$ of the form

$$a^{swip} = \sum_{k=1}^{r_{swip}} a_k^I \otimes a_k^Y \tag{3.1}$$

for some bilinear forms $a_k^I : \mathbb{R}^I \times \mathbb{R}^I \to \mathbb{R}$ and $a_k^Y : V(Y) \times V(Y) \to \mathbb{R}$ to be determined. The bilinear forms a_k^I are here identified with matrices in $\mathbb{R}^{I\times I}$. Similarly, b can be identified with a linear form on X of the form

$$b = \sum_{k=1}^{r_b} b_k^I \otimes b_k^Y \tag{3.2}$$

for some linear forms $b_k^I : \mathbb{R}^I \to \mathbb{R}$ and $b_k^Y : V(Y) \to \mathbb{R}$ to be determined. The linear forms b_k^I are identified with vectors in $\mathbb{R}^I$. Subsequently, as with (2.7) the tensor representation $u \in X$ of the solution to problem (2.1) verifies $a^{swip}(u, v) = b(v)$ for all $v \in X$.

We choose a finite dimensional subspace $V_h(Y) \subset V(Y)$, as we did with $V_h(D)$ in Section 2.3. This defines another finite dimensional subspace $X_h := \mathbb{R}^I \otimes V_h(Y) \subset X$. Approximation subspaces $V_h(D)$ and $X_h(D)$ are linearly isometric, and problem (2.8) is then equivalent to finding a tensor $u_h \in X_h$ such that

$$\forall v_h \in X_h, \quad a^{swip}(u_h, v_h) = b(v_h). \tag{3.3}$$

For a comprehensive introduction to tensor numerical calculus and problems formulated over tensor spaces, we refer the reader to the monograph [20].

Representation of the linear form b on X. To obtain a representation of the linear form b in the form (3.2), it is sufficient to consider the restriction of b to elementary tensors. Let us assume that the source term f is such that

$$\Upsilon(f) = \sum_{k=1}^{r_f} f_k^I \otimes f_k^Y \ \in \mathbb{R}^I \otimes \mathrm{L}^2(Y); \tag{3.4}$$

see Remark 3.1 on this representation. For $v \in V(D)$ such that $\Upsilon(v) = v^I \otimes v^Y$, with $v^I \in \mathbb{R}^I$ and $v^Y \in V(Y)$, we then have

$$b(v) = \sum_{i\in I} \int_{D_i} f v = \sum_{k=1}^{r_f} \sum_{i\in I} v^I(i) f_k^I(i) \int_Y f_k^Y v^Y,$$

which yields a representation of the form (3.2) with $r_b = r_f$ and linear forms $b_k^I(v^I) = \sum_{i\in I} v^I(i) f_k^I(i)$ and $b_k^Y(v^Y) = \int_Y f_k^Y v^Y$. Note that b_k^I can be identified with the vector f_k^I.

Representation of a^{swip} on $X \times X$. To obtain a representation of the bilinear form a^{swip} in the form (3.1), it is sufficient to consider the restriction of a^{swip} to elementary tensors. We first consider the representation of the diffusion form a. Let us assume that the conductivity field K is such that

$$\Upsilon(K) = \sum_{n=1}^{r_K} K_n^I \otimes K_n^Y \ \in \mathbb{R}^I \otimes \mathrm{L}^\infty(Y) \tag{3.5}$$

(see Rem. 3.1). Then, for any v, w in $V(D)$ such that $\Upsilon(v) = v^I \otimes v^Y$ and $\Upsilon(w) = w^I \otimes w^Y$ are elementary tensors in X, we have

$$a(v,w) = \sum_{i\in I} \int_{D_i} K\nabla v \cdot \nabla w = \sum_{n=1}^{r_K} \sum_{i\in I} K_n^I(i) v^I(i) w^I(i) \int_Y K_n^Y \nabla v^Y \cdot \nabla w^Y,$$

which yields

$$a = \sum_{n=1}^{r_K} \mathrm{diag}(K_n^I) \otimes N[K_n^Y],$$

with $\mathrm{diag}(K_n^I)$ the diagonal matrix in $\mathbb{R}^{I\times I}$ with diagonal K_n^I, and $N[\psi]$ the bilinear form defined for $\psi \in \mathrm{L}^\infty(Y)$ by

$$N[\psi](v^Y, w^Y) = \int_Y \psi \nabla v^Y \cdot \nabla w^Y.$$

In a similar way, we obtain

$$\begin{aligned} c = \sum_{n=1}^{r_K} \sum_{q=1}^{d} & l\left(\chi^q[K_n^I]^T\right) \otimes N_0^q[K_n^Y] + l\left(\chi^q[K_n^I]\right) \otimes N_0^{-q}[K_n^Y] \\ & -\chi^q[K_n^I]^T \otimes N_1^q[K_n^Y] - \chi^q[K_n^I] \otimes N_1^{-q}[K_n^Y], \end{aligned}$$

and

$$\begin{aligned} s = \sigma \sum_{q=1}^{d} & l\left(\chi^q\left[\frac{\omega_K}{|\partial Y_q|}\right]^T\right) \otimes M_0^q + l\left(\chi^q\left[\frac{\omega_K}{|\partial Y_q|}\right]\right) \otimes M_0^{-q} \\ & -\chi^q\left[\frac{\omega_K}{|\partial Y_q|}\right]^T \otimes M_1^q - \chi^q\left[\frac{\omega_K}{|\partial Y_q|}\right] \otimes (M_1^q)^T, \end{aligned}$$

where the bilinear forms M_0^q, M_1^q, $N_0^q[\psi]$ and $N_1^q[\psi]$ are respectively defined, for $q \in \{-d, \dots, d\} \setminus \{0\}$, by

$$M_0^q(v^Y, w^Y) = \int_{\partial Y_q} v^Y w^Y, \qquad M_1^q(v^Y, w^Y) = \int_{\partial Y_q} v^Y (w^Y \circ \tau_q),$$

$$N_0^q[\psi](v^Y, w^Y) = \int_{\partial Y_q} \psi \frac{e_q}{2} \cdot (\nabla v^Y) w^Y, \qquad N_1^q[\psi](v^Y, w^Y) = \int_{\partial Y_q} \psi \frac{e_q}{2} \cdot \nabla v^Y (w^Y \circ \tau_q),$$

with $(e_q)_{q\in\{1,\dots,d\}}$ the canonical basis of $\mathbb{R}^d$ and $e_{-q} := -e_q$, ∂Y_q the face of Y whose outward normal is e_q and τ_q the translation that maps ∂Y_q onto ∂Y_{-q}, where the matrix $\chi^q[\psi]$ is defined by

$$(\chi^q[\psi])_{ij} = \begin{cases} \psi(i,j) & \text{if } \xi_i(\partial D_i \cap \partial D_j) = \partial Y_q \\ 0 & \text{else} \end{cases},$$

and where for a matrix $A \in \mathbb{R}^{I\times I}$, $l(A)$ is the diagonal matrix such that $l\,(A)_{ij} = \delta_{ij}\sum_{k\in I}(A)_{ik}$. Finally, we have

$$m \equiv \mathbb{1}^I \otimes M,$$

with $\mathbb{1}^I$ the identity matrix in $\mathbb{R}^{I\times I}$ and

$$M(v^Y, w^Y) = \int_Y v^Y w^Y.$$

Remark 3.1 (Tensor representations of K and f)**.** The formulation of problem (3.3) over tensor product space X_h requires knowledge of tensor representations $\Upsilon(K) \in \mathbb{R}^I \otimes \mathrm{L}^\infty(Y)$ and $\Upsilon(f) \in \mathbb{R}^I \otimes \mathrm{L}^2(Y)$, yet they are generally known as elements of $\mathrm{L}^\infty(D)$ and $\mathrm{L}^2(D)$, respectively. We showed that there is a straightforward identification of K with $\sum_{k=1}^{\#I} e_k \otimes (K_{|D_k} \circ \xi_k^{-1})$, and likewise for f. This representation of K involved the sum of $\#I$ elementary tensor products, which would lead to representations of a^{swip} and b with an even greater number of terms, hence high storage and computational complexities. This would degrade the performance of the algorithm that is to be introduced in Section 3.2. Therefore, it is desirable to look for tensor representations in the form (3.5) and (3.4) with a small number of terms, *i.e.*, low rank(K) and rank(f), respectively.

Apart from rare simple cases (such as the examples in Sect. 4), K and f have full rank, so that low-rank approximations have to be introduced. Such approximations can be sought by using truncated singular value decomposition or empirical interpolation method [32]. Thus the ranks of a^{swip} and b are curbed, which improves computational efficiency. From the quasi-periodicity assumption, K is expected to have a low rank or, at least, to admit an accurate low-rank approximation.

3.2. Low-rank approximation

Tensor-based approaches have already been successfully used to reduce multiscale complexity, *e.g.* by exploiting sparsity in [21]. The novelty of the method presented here lies in the tensor representation designed specifically to exploit quasi-periodicity *via* low-rank approximation techniques.

In order to get some insight into the relation between quasi-periodicity and low-rankness, we first note that a periodic function $v : D \to \mathbb{R}$ is such that for all $i \in I$, $v_{|D_i} = v^Y \circ \xi_i$, with $v^Y : Y \to \mathbb{R}$. Such a function is identified with the rank-1 tensor $\Upsilon(v) = 1_I \otimes v^Y$, where $1_I(i) = 1$ for all $i \in I$. Let us now consider a function v which coincides with a periodic function except on a subset of cells indexed by $\Lambda \subset I$. The function v is such that $v_{|D_i} = v_0^Y \circ \xi_i$ for all $i \in I \setminus \Lambda$, and $v_{|D_i} = v_i^Y \circ \xi_i$ for $i \in \Lambda$, where v_0^Y and v_i^Y, $i \in \Lambda$, are scalar functions defined over Y. Then, v can be identified with a tensor

$$\Upsilon(v) = 1_{I\setminus\Lambda} \otimes v_0^Y + \sum_{i\in\Lambda} 1_{\{i\}} \otimes v_i^Y,$$

where, for $A \subset I$, 1_A is such that $1_A(i) = 1$ if $i \in A$ and 0 if $i \notin A$; the rank of this tensor is bounded by $1 + \#\Lambda$. A function which coincides with a periodic function except on a small number of cells will therefore admit a representation as a tensor with low-rank. Figure 4 illustrates this case for $\#\Lambda = 1$. Also, note that even if $\#\Lambda$ is large but many of the functions v_i^Y are the same, then the rank may be low. More precisely,

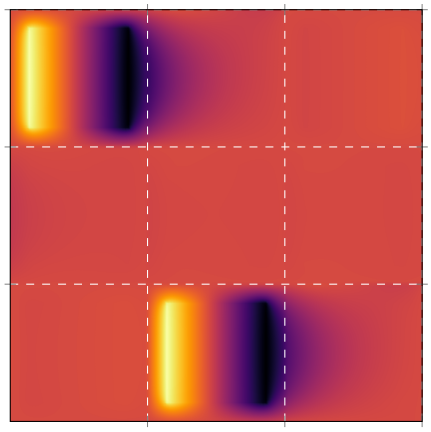


FIGURE 4. Example of rank-2 function.

$\mathrm{rank}(v) \leqslant 1 + \dim(\mathrm{span}\{v_i\}_{i\in\Lambda})$. We expect that the solution of (3.3), for a quasi-periodic medium and for some right-hand sides, will admit an accurate approximation with such a function.

In order to build a low-rank approximation of the solution of (3.3), various algorithms are available in the literature; the reader may consult surveys [19, 23] for a presentation of existing methods. We here rely on an adaptive algorithm detailed in [33], which we will outline below. Let J be the convex functional given by

$$J(v) = \frac{1}{2}a^{swip}(v,v) - b(v,v),$$

whose unique minimiser over X_h is the solution u of (3.3). This algorithm constructs a sequence of approximations $(u_n)_{n\geqslant 1}$ with increasing rank, starting with $u_0 = 0$. At each step $n \geqslant 1$, it proceeds as follows. A rank-one correction $u_n^I \otimes u_n^Y$ of u_{n-1} is first computed by solving the optimisation problem

$$\min_{v^I\in\mathbb{R}^I, v^Y\in V_h(Y)} J(u_{n-1} + v^I \otimes v^Y).$$

In practice, we perform a few iterations of an alternating minimisation algorithm which consists in minimising alternatively over u^I and u^Y. This first step yields an approximation u_n of the form $u_n = \sum_{k=1}^n u_k^I \otimes u_k^Y$. Then we compute the Galerkin projection of the solution in $\mathbb{R}^I \otimes U_n$, with $U_n = \mathrm{span}\{u_1^Y, \dots, u_n^Y\}$, which is solution of

$$\min_{v\in\mathbb{R}^I\otimes U_n} J(v). \tag{3.6}$$

This is equivalent to updating the functions u_k^I in the representation of u_n by minimising $J(u_n)$ over the functions u_k^I with fixed functions u_k^Y. Finally, we compute the Galerkin projection of the solution in $S_n \otimes V_h(Y)$, with $S_n = \mathrm{span}\{u_1^I, \dots u_n^I\}$, which is solution of

$$\min_{v\in S_n\otimes V_h(Y)} J(v).$$

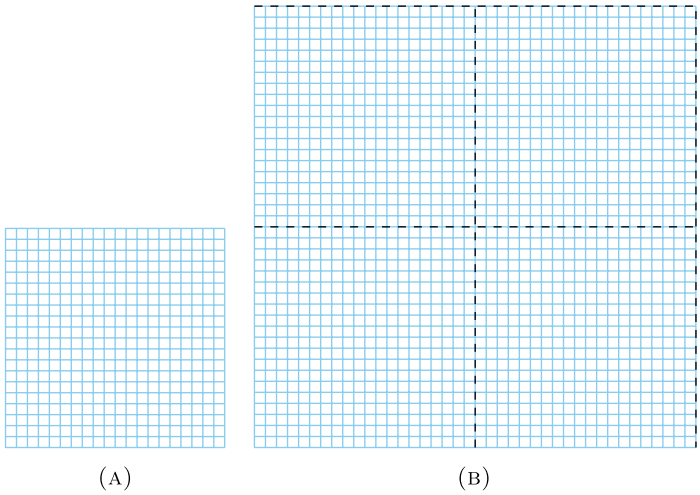


FIGURE 5. Meshes comparison between MsLRM and FEM. (A) Mesh for $V_h(Y)$. (B) Mesh for $V^c_{h,per}(D)$ (4 cells).

This is equivalent to updating the functions u_k^Y in the representation of u_n by minimising $J(u_n)$ over the functions u_k^Y for fixed functions u_k^I. Finally, we stop the algorithm when the residual error criterion

$$\frac{\|a^{swip}(u_n,\cdot)-b\|_{X'_h}}{\|b\|_{X'_h}} \leqslant \text{tolerance} \tag{3.7}$$

is verified. This method is a particular case of one of the class of algorithms whose convergence analysis can be found in [16].

We may give some insight into the complexity reduction through problems sizes. A direct resolution of (2.7) requires the resolution of a linear system of size $\#I \times \dim(V_h(Y))$. One step of the proposed algorithm requires the alternate resolution of problems of size $\#I$ and $\dim(V_h(Y))$ in the rank-one correction step, the resolution of one problem of size $n \times \#I$ and finally, the resolution of one problem of size $n \times \dim(V_h(Y))$. The cost of one iteration therefore increases with n but, for moderate ranks, it remains small compared to a direct resolution method. Note also that compared to a direct resolution method, the tensor-structured approach may allow a significant reduction in the storage of the operator.

4. Numerical results

The proposed multiscale low-rank approximation method, here denoted MsLRM, has been tested on two-dimensional problems with quasi-periodic diffusion operator K of the form

$$\Upsilon(K) = B \otimes K_1^Y + (1-B) \otimes K_2^Y, \tag{4.1}$$

where the $\{B(i) : i \in I\}$ are independent and identically distributed Bernoulli random variables with values in $\{0,1\}$. This means that K is a random function whose restriction to any cell D_i is K_1^Y if $B(i)=1$, and K_2^Y if $B(i)=0$. The conductivity field K can be interpreted as a random perturbation of an ideal periodic medium, where a cell D_i displays the material property of the reference periodic medium if $B(i)=1$, and a "perturbed" property if $B(i)=0$. This arbitrary interpretation means that K_1^Y represents the conductivity of a *sound* cell and K_2^Y the conductivity of a *faulty* one. Since the $\{B(i) : i \in I\}$ are identically distributed, the defect probability is the same for every cell and we note it $p := \mathbb{P}(B(i)=0)$.

TABLE 1. Default parameters.

$\dim(V_h(Y))$	Tolerance	p	$\sup(K)/\inf(K)$
441	10^{-3}	0.1	100

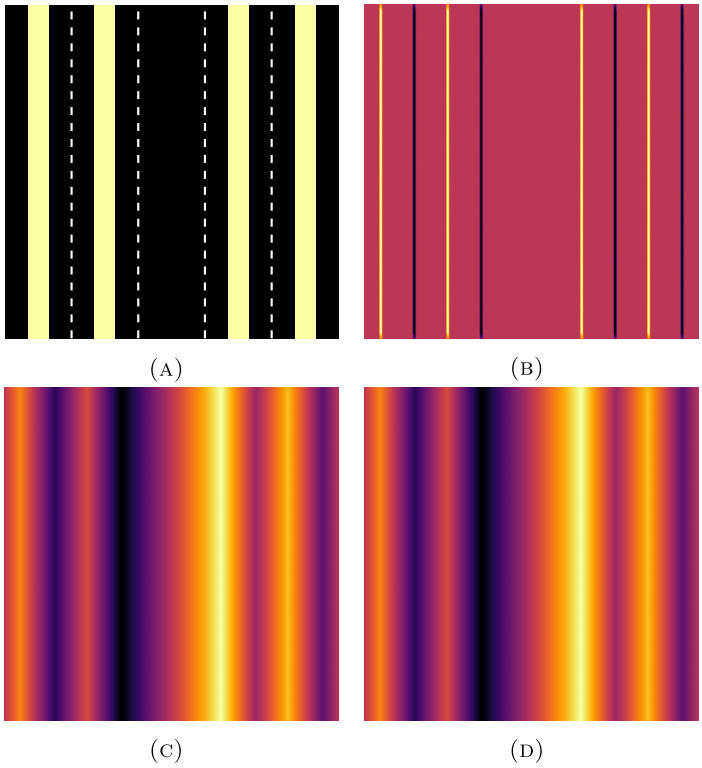


FIGURE 6. Missing fibres test case: example with five cells. (A) Conductivity. (B) Source term. (C) Rank-3 approximation. (D) Reference solution.

The source term chosen is the same for all experiments and was inspired by corrector problems in stochastic homogenisation [4]. We define it over D as

$$f = \nabla \cdot (Ke_1), \tag{4.2}$$

where the choice of direction e_1 is arbitrary. The boundary conditions remain periodic.

We choose an approximation space $V_h(Y)$ of continuous, piecewise affine functions[2] based on a mesh of isoparametric quadrangle elements. This mesh is a regular grid of 20×20 elements; Figure 5A shows an isotropic example for $Y :=]0,1[^d$. The associated trace constant $C(V_h(Y)) \approx 7$ (unless specified otherwise), computed accordingly to Remark 2.5. For comparison, we use as a reference method a standard continuous Galerkin finite element method with an approximation space $V^c_{h,per}(D) = V_h(D) \cap \mathrm{C}^0(D) \cap \mathrm{H}^1_{per}(D)$ (continuous and periodic functions in $V_h(D)$).

Unless specified otherwise, the default parameter values given in Table 1 apply.

[2]Those are piecewise Lagrange polynomials of degree at most 1.

TABLE 2. Missing fibres test case.

$\#I$	FEM	MsLRM	
	Time (s)	Time (s)	Rank
25	1.3	0.57	3
100	9	0.60	3
225	35	0.56	3

Remark 4.1 (Approximation spaces' dimensions compared)**.** Where elements of $V^c_{h,per}(D)$ are concerned, each cell is meshed as Y is (see an example in Fig. 5B) and therefore $\dim(V^c_{h,per}(D))$ is of the same order as $\dim(V_h(D))$. More precisely, $\dim(V^c_{h,per}(D)) < \dim(V_h(D))$ because of the continuity constraints at cell interfaces—including half of external faces, due to periodic boundary conditions. Those cell interfaces are outlined on the example of Figure 5B; each node located along those lines would have one more degree of freedom in $V_h(D)$ than in $V^c_{h,per}(D)$. For example, a square domain of 1024 cells with $\dim(V_h(Y)) = 441$ yields $\dim(V^c_{h,per}(D)) = 410\,881$, whereas $\dim(V_h(D)) = 451\,584$.

All computations were run on the same workstation, *viz.* a Dell™ Optiplex™ 7010 with:

- 8 GiB[3] (2×4) RAM DDR3 1600 MHz;
- Intel® Core™ i7-3770 CPU: 4 cores at 3.40 GHz with 2 threads each.

4.1. Various conductivity patterns

Here we compare the computational time between FEM and MsLRM on three test cases. These differ by their diffusion operator K, reference cell Y and connectivity between cells. For each case, the comparison spans three values of $\#I$ (*viz.* 25, 100 and 225) to give a small insight into computational cost sensitivity to an increase in number of cells.

Missing fibres

This test case was directly inspired by composite materials with unidirectional fibre reinforcements. The mesoscopic mesh is also unidirectional since every cell spans the entire width of the domain, with $Y :=]0,1[\times]0,5[$ and $D := [0,\#I] \times [0,5]$. Fibre and matrix both have uniform conductivities and the faulty cells have no fibre, thus K is expressed as (4.1) with $K_1^Y = 1 + 99\chi$ and $K_2^Y = 1$, where $\chi \in \mathrm{C}^0(Y,[0,1])$ is the continuous indicator function of the fibre, *i.e.* $[0.25, 0.75]\times]0,1[$. An example of such conductivity with five cells, of which the middle one is faulty, is displayed in Figure 6A.

Y is meshed with the same number of elements as in Figure 5A, resulting in an anisotropic mesh. We thus keep $\dim(V_h(Y))$ at the value in Table 1 but, due to these particular cell size and mesh, the trace constant here is $C(V_h(Y)) \approx 4.47$.

The results are shown in Table 2. Only a rank 3 is required to reach the desired precision and therefore we hardly see any effect of the increase in domain size on computational time. These results are mainly due to the unidimensionality of the mesoscopic mesh. A faulty cell essentially affects the solution in the two neighbouring cells, as is visible in Figure 6C and D which display the reference FEM solution and its MsLRM approximation for the conductivity from Figure 6A.

For the FEM resolution, we observe a more significant increase in computational time as $\#I$ increases.

Undulating fibres

This test was inspired by woven composite materials. Unlike the previous test, there are fibres in two orthogonal directions and the mesoscopic mesh is bidimensional with reference cell $Y :=]0,1[^d$. Faulty cells show an

[3] 5.5 to 6.5 of which were usually available for the simulations.

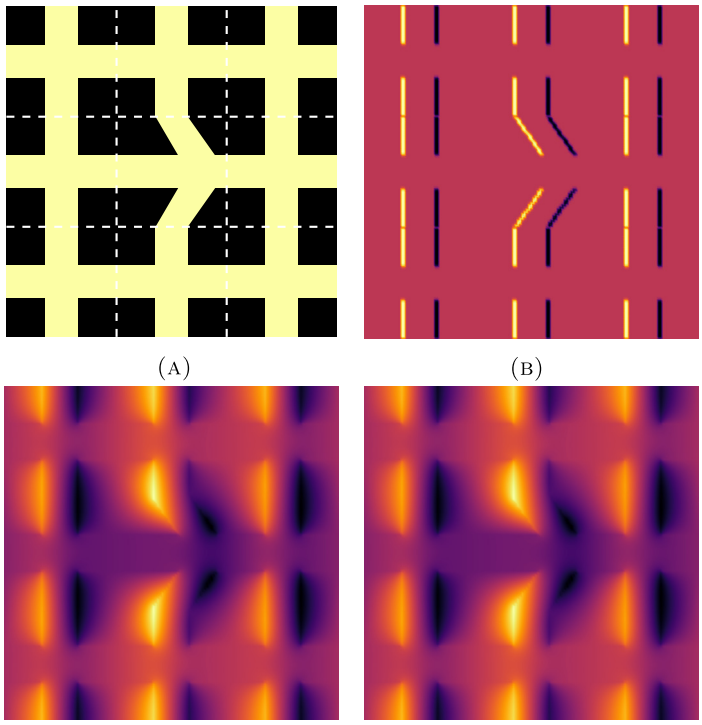


FIGURE 7. Undulating fibres test case: example with nine cells. (A) Conductivity. (B) Source term. (C) MsLRM approximation. (D) Reference solution.

TABLE 3. Undulating fibres test case.

$\#I$	FEM	MsLRM	
	Time (s)	Time (s)	Rank
25	1.3	2.16	14
100	9	5.1	20
225	35	6.4	21

undulation in a fibre, as illustrated in Figure 7A. Consequently, K is expressed as in (4.1) with $K_1^Y = 1 + 99\chi_1$ and $K_2^Y = 1 + 99\chi_2$, where $\chi_1, \chi_2 \in \mathrm{C}^0(Y, [0,1])$ are indicator functions of the straight cross and cross with bent fibre, respectively; the crosses' arms have a width of $1 - 2^{-1/2}$ so that they occupy half the surface.

The results in Table 3 show that, compared to the first test case, a higher rank of approximation is necessary to achieve the same precision. This is mainly due to the bidimensionality of the mesoscopic mesh: each cell has eight neighbours, whereas it had only two in the first test case. The impact of a defect requires more functions in $V_h(Y)$ to be represented. Reference solution and its approximation are displayed in Figure 7C and D.

Consequently, the computational time is more affected by an increase in the number of cells. This increase in computational time remains, however, considerably smaller than that of the reference resolution method.

Missing inclusions

This test, sketched in Figure 8A, echoes the example shown in Figures 1 and 2B: a square inclusion is present in sound cells and absent from faulty ones. Therefore, $Y :=]0,1[^d$ and K is expressed as (4.1) with

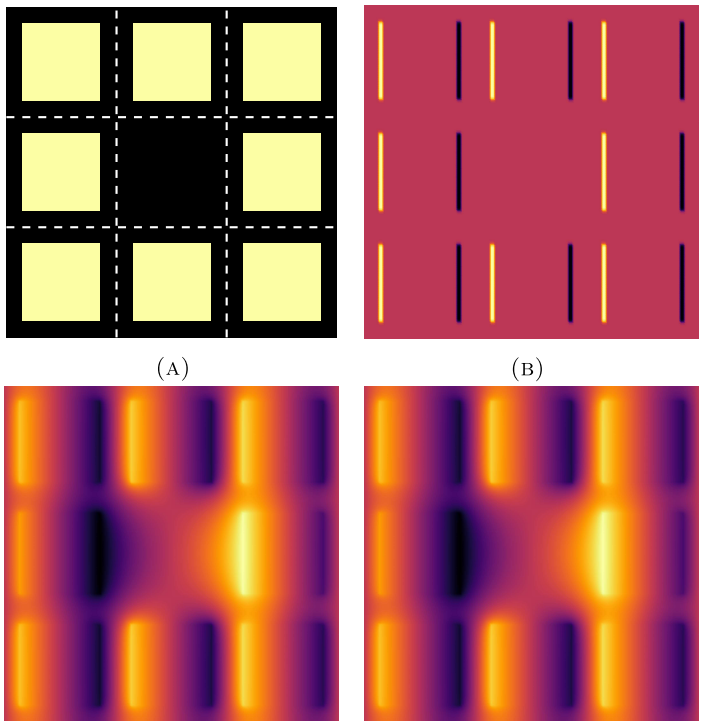


FIGURE 8. Missing inclusions test case: example with nine cells. (A) Conductivity. (B) Source. (C) MsLRM approximation. (D) Reference solution.

TABLE 4. Missing inclusions test case.

$\#I$	FEM	MsLRM	
	Time (s)	Time (s)	Rank
25	1.43	1.25	10
100	9.46	3.6	17
225	36.2	3.8	17

$K_1^Y = 1 + 99\chi$ and $K_2^Y = 1$, where $\chi \in \mathrm{C}^0(Y, [0,1])$ is the continuous indicator function of the square inclusion, *i.e.* $[(2-\sqrt{2})/4, (2+\sqrt{2})/4]^2$; the square's dimensions were chosen so as to have the same occupied surface in sound cells as the undulating fibres test case.

The results are shown in Table 4. The slight difference between this case and the previous one can only be ascribed to the change in conductivity pattern, since both are bidimensional at the mesoscopic scale. Although this case has a higher conductivity contrast between sound and faulty cells than the previous one, it shows significant complexity reduction compared with the reference method.

For the sake of consistency, we retain only this conductivity pattern for the following experiments in Sections 4.2–4.4.

4.2. Influence of domain size and source term

The three initial tests of Section 4.1 gave hints on the complexity reduction of the low-rank approximation method compared to a direct resolution method. To get a better insight into this, we observed the computational

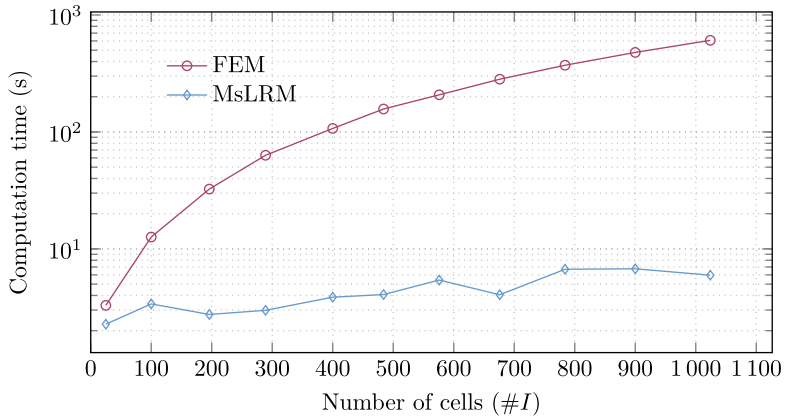


FIGURE 9. Domain size influence on MsLRM compared to FEM.

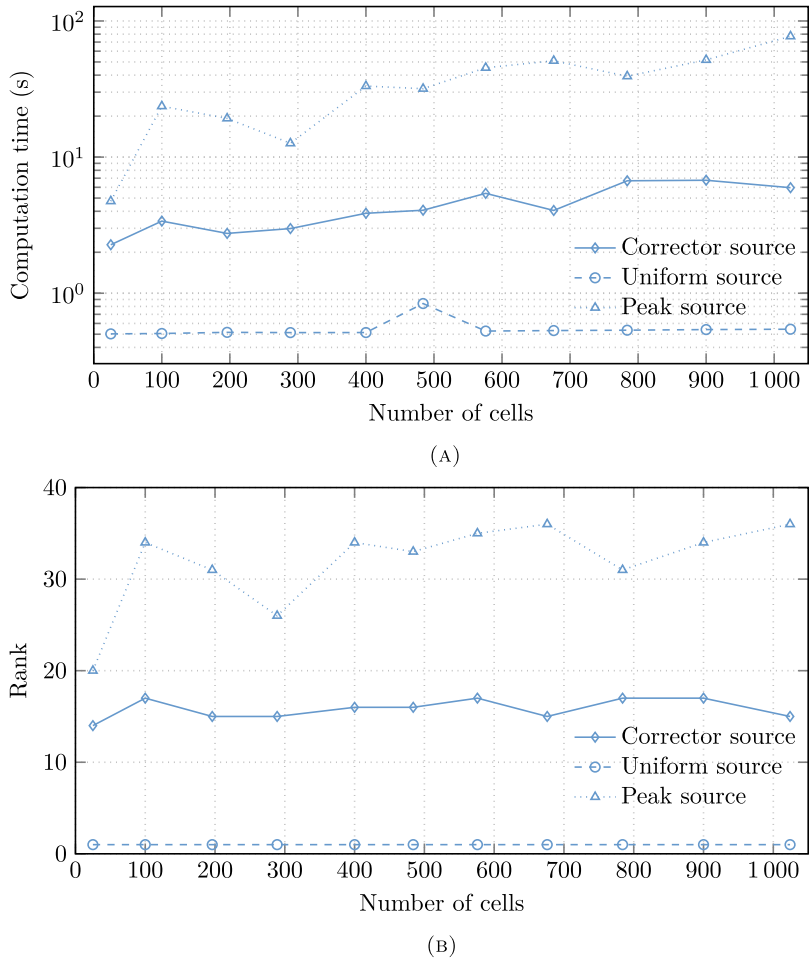
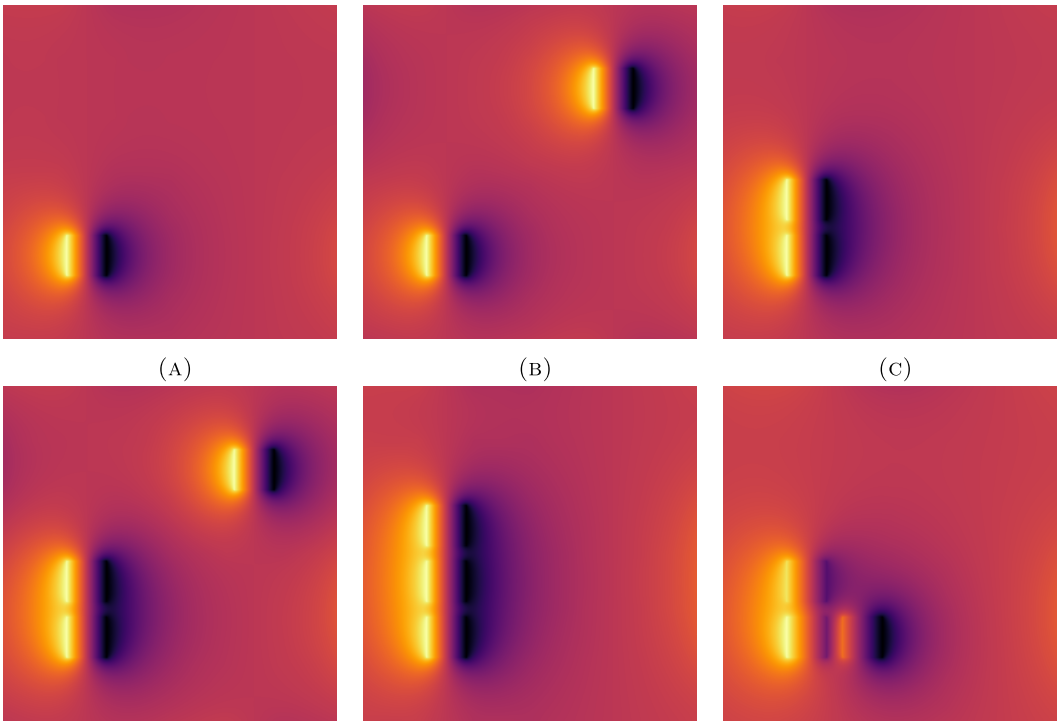


FIGURE 10. Source term influence on MsLRM. (A) Impact on computational cost. (B) Impact on approximation rank.



(A) (B) (C)

(D) (E) (F)

FIGURE 11. Approximation rank for various configurations of square inclusion defects. (A) Rank 10. (B) Rank 10. (C) Rank 13. (D) Rank 14. (E) Rank 13. (F) Rank 14.

time of missing inclusions problems for a larger range of values of $\#I$, which resulted in Figure 9. The difference in complexity is made obvious.

These results were obtained with a quasi-periodic source term given by equation (4.2). We investigated the influence of the source term by running identical computations with two other source terms. The first one is a uniform source term which smooths defect influence. The second one is a centred peak given for $x \in D$ by

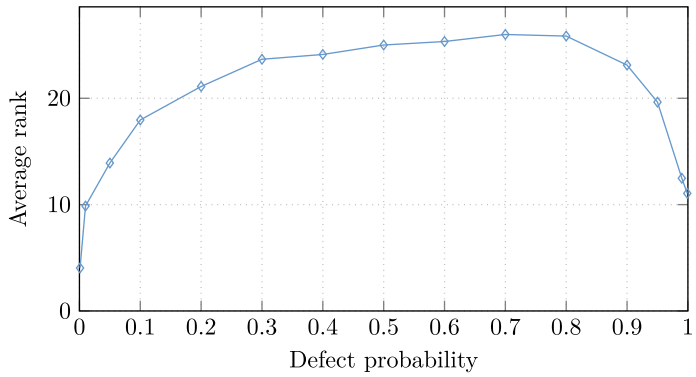
$$f(x) = e^{-10\|x-\theta\|},$$

where θ is the centre of D. It is chosen so as to break periodicity.

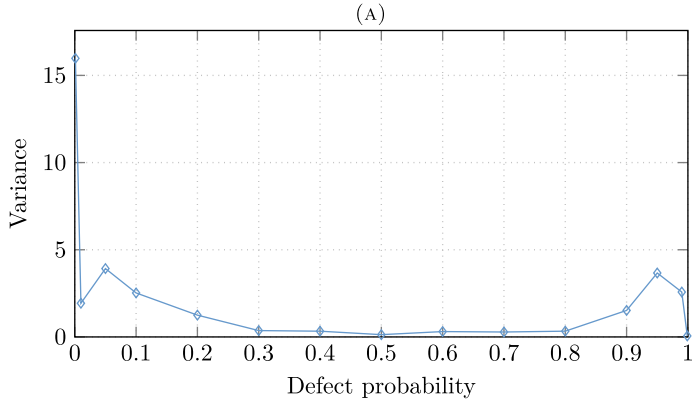
As expected, Figure 10A shows that with the uniform source term, the solution is quasi-periodic and the proposed method yields a high complexity reduction. The peak source term problems have a higher complexity, as far as the MsLRM is concerned. However, we see from Figure 10B that the approximation rank is bounded even in this latter case: there is still an underlying structure to the solution that allows an accurate approximation with low rank regardless of the domain size.

4.3. Influence of the probability of defects

On the previous tests, the approximation rank as a function of the number of cells $\#I$ seemed to rapidly reach a plateau. This was most obvious in Figure 10. One interpretation, illustrated in Figure 11A, is that new patterns in the solution are caused by new configurations of defects, which increase the approximation rank for a given tolerance. This plateau is a consequence of the medium's ergodicity: the larger the domain, the higher the probability to observe every possible configuration. The number of cells before reaching the plateau depends



(A)



(B)

FIGURE 12. Effect of square inclusion probability on approximation rank ($\#I = 400$, 100 samples). (A) Average approximation rank. (B) Variance of average approximation rank.

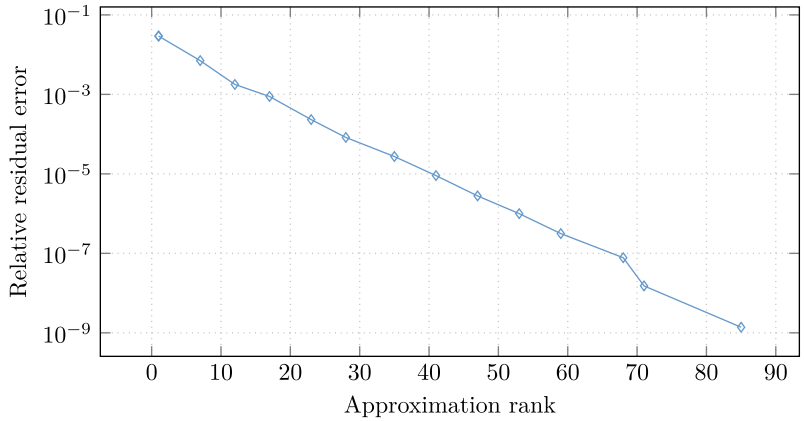


FIGURE 13. Approximation rank with respect to precision ($\#I = 400$).

on a number of parameters: the rank of the conductivity field (related to the number of cell types), the area of influence of a defect (*cf*. Fig. 11A–D), and the probability of a defect.

Furthermore, we observed the influence of the probability of defect p on the approximation rank. For each value of p, we observed the average rank over 100 computations. Here, the defect is a square inclusion, as in Figure 11A. The results are plotted in Figure 12A and display the expected low values when p goes to 0 or 1, where we tend to a periodic medium. The graph is slightly asymmetric: the highest approximation ranks were encountered when cells with inclusions were more likely. A missing inclusion in a medium with periodic inclusions has less effect than an inclusion in a uniform medium.

To investigate the variability of ranks, we plotted their variance for each value of p in Figure 12B. As for the average rank, this graph is slightly skewed. The highest values are when p goes to 0 or 1, *i.e*. when the probability of getting a periodic medium and the probability of having at least one defect are of similar order. This can be mostly explained by considering Figure 11A: in this case the solution associated with a perfectly periodic medium would be of rank 1; one defect yields an approximation of rank[4] 10 in Figure 11A; Figure 11B–F show that additional defects cause a much smaller increase in rank—none if no new pattern appears (*cf*. Fig. 11C and E). Therefore, the approximation rank reaches its highest variance for values of p that make a periodic medium as likely as a medium with at least one defect.

4.4. Rank and precision

All previous results were obtained for a tolerance of 10^{-3}. We have seen the influence of conductivity patterns, problem size and source terms on the rank of the approximation. Now, we analyse the convergence of the approximation with respect to the rank. We consider a problem of missing inclusions as in Section 4.1, over a square domain of 400 cells, and observe the evolution of the relative residual error as defined in equation (3.7) with respect to the approximation rank.

Figure 13 presents the results. We observe an exponential convergence of the error with respect to the rank. Tolerance remains a major factor in computational cost of the proposed low-rank method: for small domain size and high precision, a direct resolution method would be more efficient.

5. Conclusion

We have presented an approximation method to reduce the complexity of the resolution of stationary diffusion problems in quasi-periodic media. The method relies on a two-scale representation of the resolution, which is identified with a tensor. The method then exploits the fact that the solution admits accurate low-rank approximations. A greedy algorithm is employed to build a non-optimal yet cost-efficient low-rank approximation with a desired precision. The proposed method can be easily adapted to a larger class of linear elliptic PDEs.

Cost-efficiency has been illustrated comparatively to a direct resolution method in numerical experiments with several conductivity patterns which are typical in composite materials. Complexity reduction compared to the direct resolution method has been observed on the different experiments. Finally, the validity of the low-rank assumption has been tested with respect to precision and perturbation of periodicity. A plateau in approximation rank with respect to domain size increase, attributed to the medium's ergodicity, has been observed and suggests good performance for computations on large domains, even in case of low periodicity.

References

[1] A. Abdulle and Y. Bai, Reduced basis finite element heterogeneous multiscale method for high-order discretizations of elliptic homogenization problems. *J. Comput. Phys.* **231** (2012) 7014–7036.

[2] A. Abdulle, E. Weinan, B. Engquist and E. Vanden-Eijnden, The heterogeneous multiscale method. *Acta Numer.* **21** (2012).

[3] G. Allaire and R. Brizzi, A multiscale finite element method for numerical homogenization. *Multiscale Model. Simul.* **4** (2005) 790–812.

[4] A. Anantharaman, R. Costaouec, C. Le Bris, F. Legoll and F. Thomines, Introduction to Numerical Stochastic Homogenization 311 and the Related Computational Challenges: Some Recent Developments, Vol. 22. World Scientific, Singapore (2011).

[4] All ranks given here are for approximations with the tolerance value in Table 1.

[5] G. Bal, J. Wenjia and W. Jing, Corrector theory for MsFEM and HMM in random media. *Multiscale Model. Simul.* **9** (2011) 1549–1587.

[6] X. Blanc, C. Le Bris and P.-L. Lions, Une variante de la théorie de l'homogénéisation stochastique des opérateurs elliptiques. *C. R. Math.* **343** (2006) 717–724.

[7] X. Blanc, R. Costaouec, C. Le Bris and F. Legoll, Variance reduction in stochastic homogenization using antithetic variables. *Markov Processes Relat. Fields* **66** (2012) 31–66.

[8] S. Boyaval, Reduced-basis approach for homogenization beyond the periodic setting. *Multiscale Model. Simul.* **7** (2008).

[9] M. Chevreuil, A. Nouy and E. Safatly, A multiscale method with patch for the solution of stochastic partial differential equations with localized uncertainties. *Comput. Methods Appl. Mech. Eng.* **255** (2013) 255–274.

[10] D.A. Di Pietro and A. Ern, Mathematical Aspects of Discontinuous Galerkin Methods, Vol. 69. Springer Science & Business Media (2011).

[11] W. E, B. Engquist, X. Li, W. Ren and E. Vanden-Eijnden, Heterogeneous multiscale methods: a review. *Commun. Comput. Phys.* **2** (2007) 367–450.

[12] Y. Efendiev and T.Y. Hou, Multiscale Finite Element Methods. *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, NY (2009).

[13] Y. Epshteyn and B. Rivière, Estimation of penalty parameters for symmetric interior penalty Galerkin methods. *J. Comput. Appl. Math.* **206** (2007) 843–872.

[14] A. Ern and J.-L. Guermond, Theory and Practice of Finite Elements. *Applied Mathematical Sciences*. Springer, New York (2004).

[15] L.C. Evans, Partial Differential Equations. *Graduate Studies in Mathematics*. American Mathematical Society (1998).

[16] A. Falcó and A. Nouy, Proper generalized decomposition for nonlinear convex problems in tensor Banach spaces. *Numer. Math.* **121** (2012) 503–530.

[17] L. Gendre, O. Allix and P. Gosselet, A two-scale approximation of the Schur complement and its use for non-intrusive coupling. *Int. J. Numer. Methods Eng.* **87** (2011) 889–905.

[18] R. Glowinski, J. He, J. Rappaz and J. Wagner, Approximation of multi-scale elliptic problems using patches of finite elements. *C. R. Math.* **337** (2003) 679–684.

[19] L. Grasedyck, D. Kressner and C. Tobler, A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen* **36** (2013) 53–78.

[20] W. Hackbusch, Tensor Spaces and Numerical Tensor Calculus. Vol. 42 of *Springer Series in Computational Mathematics*. Springer, Heidelberg (2012).

[21] V. Ha Hoang and C. Schwab, High-dimensional finite elements for elliptic problems with multiple scales. *Multiscale Model. Simul.* **3** (2005).

[22] T.Y. Hou and X.-H. Wu, A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.* **134** (1997) 169–189.

[23] B.N. Khoromskij, Tensors-structured numerical methods in scientific computing: survey on recent advances. *Chemom. Intell. Lab. Syst.* **110** (2012) 1–19.

[24] C. Le Bris, Some numerical approaches for weakly random homogenization, in Numerical Mathematics and Advanced Applications, edited by G. Kreiss, P. Lötstedt, A. Målqvist and M. Neytcheva. Springer, Berlin, Heidelberg (2009) 29–45.

[25] C. Le Bris, F. Legoll and W. Minvielle, Special quasirandom structures: a selection approach for stochastic homogenization. *Monte Carlo Methods Appl.* **22** (2016) 25–54.

[26] C. Le Bris, F. Legoll and F. Thomines, Multiscale finite element approach for "weakly" random problems and related issues. *ESAIM: M2AN* **48** (2014) 815–858.

[27] C. Le Bris and F. Thomines, A reduced basis approach for some weakly stochastic multiscale problems. *Chin. Ann. Math. Ser. B* **33** (2012) 657–672.

[28] F. Legoll and W. Minvielle, A control variate approach based on a defect-type theory for variance reduction in stochastic homogenization. *Multiscale Model. Simul.* **13** (2015).

[29] F. Legoll and W. Minvielle, Variance reduction using antithetic variables for a nonlinear convex stochastic homogenization problem. *Discrete Contin. Dyn. Syst. Ser. S* **8** (2015) 1–27.

[30] L. Lin and B. Stamm, A posteriori error estimates for discontinuous Galerkin methods using non-polynomial basis functions. Part I: Second order linear PDE. *ESAIM: M2AN* **50** (2016) 1193–1222.

[31] Y. Maday, Reduced basis method for the rapid and reliable solution of partial differential equations, in *International Congress of Mathematicians, Madrid*. European Mathematical Society (2006). 1255–1270.

[32] Y. Maday, N.C. Nguyen, A.T. Patera and G.S.H. Pau, A general multipurpose interpolation procedure: the magic points. *Commun. Pure Appl. Anal.* **8** (2009) 383–404.

[33] A. Nouy, Low-rank methods for high-dimensional approximation and model order reduction, in Chapter 4 of Model Reduction and Approximation. SIAM (2017) 171–226.

[34] O. Pironneau and J.-L. Lions, Domain decomposition methods for CAD. *C. R. Acad. Sci. Ser. I – Math.* **328** (1999) 73–80.

[35] V. Rezzonico, A. Lozinski, M. Picasso, J. Rappaz and J. Wagner, Multiscale algorithm with patches of finite elements. *Math. Comput. Simul.* **76** (2007) 181–187.