Convergence of iterative methods based on Neumann series for composite materials: theory and practice

Hervé Moulinec, Pierre Suquet, Graeme Milton

Convergence of iterative methods based on Neumann series for composite materials: theory and practice.

Hervé Moulinec[a], Pierre Suquet[a,*], Graeme W. Milton[b]

[a]*Aix-Marseille Univ, CNRS, Centrale Marseille, LMA,*
4 Impasse Nikola Tesla, CS 40006, 13453 Marseille Cedex 13, France.
[b] *Univ Utah, Dept Math, Salt Lake City, UT 84112, USA.*

Abstract

Iterative Fast Fourier Transform methods are useful for calculating the fields in composite materials and their macroscopic response. By iterating back and forth until convergence, the differential constraints are satisfied in Fourier space, and the constitutive law in real space. The methods correspond to series expansions of appropriate operators and to series expansions for the effective tensor as a function of the component moduli. It is shown that the singularity structure of this function can shed much light on the convergence properties of the iterative Fast Fourier Transform methods. We look at a model example of a square array of conducting square inclusions for which there is an exact formula for the effective conductivity (Obnosov). Theoretically some of the methods converge when the inclusions have zero or even negative conductivity. However, the numerics do not always confirm this extended range of convergence and show that accuracy is lost after relatively few iterations. There is little point in iterating beyond this. Accuracy improves when the grid size is reduced, showing that the discrepancy is linked to the discretization. Finally, it is shown that none of the three iterative schemes investigated over-performs the others for all possible microstructures and all contrasts.

Keywords: Fourier Transforms; homogenization; heterogeneous media

1. Introduction

It is a well-established result in the theory of linear composites that the local fields in such materials satisfy an integral equation which can be written symbolically as (Kröner [1], Willis [2], Milton [3])

$$\left(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right) \boldsymbol{\varepsilon} = \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the local field under investigation, $\boldsymbol{E}$ is its average, $\boldsymbol{\Gamma}^0$ is the Green's operator associated with a homogeneous reference medium $\boldsymbol{L}^0$ and $\boldsymbol{\delta L} = \boldsymbol{L} - \boldsymbol{L}^0$ is the deviation of the actual material properties of the composite from the homogeneous reference medium. As recognized by Kröner [1], this integral equation has the same form as the Lippmann-Schwinger equation of scattering theory.

*Corresponding author: suquet@lma.cnrs-mrs.fr

The resolution of (1) requires the inverse of $\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}$ which can be expanded in Neumann-Liouville series. This expansion was used by Brown [4] with one of the phases as reference material (see also Kröner [1]).

The Neumann-Liouville series is again the basis of the fixed-point scheme proposed by Moulinec and Suquet [5, 6] which differs from previous works by two important aspects. First, the reference medium is not chosen to be one of the phases, but optimized to enhance the convergence of the series. Second, the Discrete Fourier Transform is used to compute efficiently the Green's operator $\boldsymbol{\Gamma}^0$ and solve iteratively the Lippmann-Schwinger equation (1). This fixed-point algorithm (sometimes called the "basic" algorithm) alternates between real space (where the unit-cell is discretized along a regular grid) and Fourier space (where the spatial frequency vector takes discrete values dual to the spatial discretization in real space). The Fourier transform of the Green's operator $\boldsymbol{\Gamma}^0$ is known for $\boldsymbol{L}^0$ with arbitrary anisotropy and can be given an explicit form for several classes of symmetry (Khatchaturyan [7], Mura [8], Nemat-Nasser *et al* [9]). The choice of the reference tensor $\boldsymbol{L}^0$ and the rates of convergence that were observed with this fixed-point algorithm for various microstructures were consistent with the conditions for "unconditional" convergence derived by Michel *et al* [10].

However, several issues were raised subsequently by the authors themselves or by other authors concerning the rate of convergence of the basic scheme, which can be poor when the contrast between the phases becomes large. In addition, spurious oscillations of the local fields were sometimes observed. The criticisms, and some of the progress made over the years, can be schematically listed under three main categories.

1. Rate of convergence. To improve on the rate of convergence of the basic scheme, accelerated methods have been proposed. The earliest one is due to Eyre and Milton [11] (generalized in Milton [3]) and consists in a change of the field on which the algorithm operates, and a shift of $\boldsymbol{\Gamma}^0$ in an effort to reduce the norm of the operator product entering the interations. This accelerated scheme can also be seen as the summation of a Neumann-Liouville series. The shifting and associated series expansion were first introduced in [12], sect.5, in the context of the conductivity problem, and its convergence properties, without assuming the conductivity tensor was symmetric, were studied in [13], sect.3. The series is related to a series that forms the basis for establishing microstructure independent relations satisfied by effective tensors (see, e.g., sections 14.10 and 17.3 in [3] and [14]). Let us mention that several other accelerated schemes (polarization scheme [15, 16], conjugate gradient [17], Galerkin approaches [18]) have also been proposed but are not discussed here. Some comparisons were recently made in [19], suggesting the conjugate gradient method may often have the fastest convergence.
2. Convergence criterion. As in every iterative method, choosing a sensible test to decide when the iterations should be stopped is of crucial importance. The initial criterion of Moulinec and Suquet [6] was based on the L^2 norm of one of the equations to be satisfied (equilibrium equation). Other criteria, such as the difference between two iterates, have also been proposed (Milton [3]).
3. Spectral derivatives and Green's operators. The Green's operator used in the basic scheme is the continuous Green's operator. Spurious oscillations in the solutions have been ob-

served in the computed fields. Gibbs phenomena are sometimes invoked to explain these oscillations, but in our opinion their origin is different, as they appear gradually along the iterations. A possible explanation is that computing the Fourier transform of the derivative of a nonsmooth function by means of the Fourier transform requires some care. This has motivated Müller [20] and subsequently several authors (Brown *et al* [21], Willot [22], Schneider *et al* [23] among others) to use discrete Green's operators.

The present article discusses several aspects of the convergence of three Neumann-Liouville series for three different schemes present in the literature and presented here in a unified form in section 2.2. Section 2 presents theoretical results on the convergence of the series with no restriction on the microstructure. The integral operator entering the power series is split into an integral operator with norm 1 and a local operator depending only on the contrast between the constituents. Section 3 examines the theoretical convergence of the series when more information about the microstructure is available. The integral operator is split differently into an integral operator depending on the microstructure and a local operator depending only on the contrast between the phases. Theoretical estimates for the radius of convergence of the iterative methods are derived. It is also shown that information on the singularities of the effective moduli in the complex plane can be used to improve the range of phase contrast for which these iterative methods converge and the theoretical prediction of their rate of convergence. These theoretical estimates are compared in section 4 with numerical simulations of the effective conductivity of a square array of square inclusions for which an explicit solution is available (Obnosov [24]). The range of contrast for which convergence is observed numerically appears to be smaller than what it should be theoretically. It is found that the origin of this discrepancy is the reduction of the continuous problem to a finite-dimensional one. The discretized solution being different from the exact continuous one, the higher- order terms in the series have to be different from their exact value. Finally the rate of convergence of the three iterative schemes is examined for more general microstructures when the location of the singularities and the contrast between the phases are varied. It is found that none of the three iterative schemes over-performs the other two for all possible microstructures and all phase contrasts.

2. Estimates for the convergence radius of Neumann-Liouville series with no information on the microstructure

2.1. The Lippman-Schwinger equation

Consider a unit-cell V of a periodic composite material. The composite is made up of N homogeneous phases $V^{(r)}, r = 1, .., N$, whose distribution is defined by characteristic functions $\chi^{(r)}$. Let $\langle . \rangle$ and $\langle . \rangle^{(r)}$ denote spatial averaging over V and $V^{(r)}$ respectively. The material property under consideration is characterized by a tensor field $\boldsymbol{L}(\boldsymbol{x})$ relating two local fields $\boldsymbol{\sigma}(\boldsymbol{x})$ and $\boldsymbol{\varepsilon}(\boldsymbol{x})$ satisfying partial differential equations expressing, in elasticity, balance equations and

compatibility conditions :

$$\left.\begin{array}{l} \boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{L}(\boldsymbol{x}) : \boldsymbol{\varepsilon}(\boldsymbol{x}), \quad \text{div } \boldsymbol{\sigma}(\boldsymbol{x}) = 0 \quad \text{in } V, \\ \boldsymbol{\varepsilon}(\boldsymbol{x}) = \boldsymbol{E} + \boldsymbol{\varepsilon}^*(\boldsymbol{x}), \quad \boldsymbol{\varepsilon}^*(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{\nabla} \boldsymbol{u}^*(\boldsymbol{x}) + \boldsymbol{\nabla} \boldsymbol{u}^{*T}(\boldsymbol{x})), \\ \boldsymbol{u}^* \text{ periodic on } \partial V, \quad \boldsymbol{\sigma}.\boldsymbol{n} \text{ anti-periodic on } \partial V. \end{array}\right\} \tag{2}$$

In elasticity $\boldsymbol{L}$ is the fourth-order stiffness tensor, $\boldsymbol{\sigma}$ is the stress field, $\boldsymbol{\varepsilon}$ is the strain field, while in conductivity, $\boldsymbol{L}$ is the second-order conductivity tensor, $\boldsymbol{\sigma}$ is the current field, $\boldsymbol{\varepsilon}$ is the electric field which is the gradient of the electrical potential u. $\boldsymbol{L}$ is assumed to be a symmetric tensor.

Introducing a reference tensor $\boldsymbol{L}^0$ (symmetric and definite positive), the constitutive relation between $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ is rewritten with a polarization field $\boldsymbol{\tau}$ as

$$\boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{L}^0 : \boldsymbol{\varepsilon}(\boldsymbol{x}) + \boldsymbol{\tau}(\boldsymbol{x}), \quad \boldsymbol{\tau}(\boldsymbol{x}) = \boldsymbol{\delta L}(\boldsymbol{x}) : \boldsymbol{\varepsilon}(\boldsymbol{x}), \quad \boldsymbol{\delta L}(\boldsymbol{x}) = \boldsymbol{L}(\boldsymbol{x}) - \boldsymbol{L}^0. \tag{3}$$

The solution of (2) can be expressed with the Green's operator associated with $\boldsymbol{L}^0$ as

$$\boldsymbol{\varepsilon} = \boldsymbol{E} - \boldsymbol{\Gamma}^0 \boldsymbol{\tau},$$

where $\boldsymbol{\Gamma}^0$ is an integral operator (see appendix A) and $\boldsymbol{\Gamma}^0 \boldsymbol{\tau}$ is to be understood as the action of this integral operator on the field $\boldsymbol{\tau}$ given by (3). Replacing $\boldsymbol{\tau}$ by its expression (3) leads to the Lippmann-Schwinger integral equation for the field $\boldsymbol{\varepsilon}$:

$$\left(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right) \boldsymbol{\varepsilon} = \boldsymbol{E}, \tag{4}$$

the solution of which can be written as

$$\boldsymbol{\varepsilon} = \left(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right)^{-1} \boldsymbol{E}. \tag{5}$$

The effective moduli $\widetilde{\boldsymbol{L}}$ relating the average stress $\langle \boldsymbol{\sigma} \rangle$ and the average strain $\boldsymbol{E} = \langle \boldsymbol{\varepsilon} \rangle$ read as

$$\widetilde{\boldsymbol{L}} = \langle \boldsymbol{L} \left(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right)^{-1} \rangle = \boldsymbol{L}^0 + \langle \boldsymbol{\delta L} \left(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right)^{-1} \rangle. \tag{6}$$

2.2. Neumann-Liouville series

The operator $(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L})^{-1}$ can be *formally* expanded in power series of $\boldsymbol{\Gamma}^0 \boldsymbol{\delta L}$ and the corresponding expansion for the strain field $\boldsymbol{\varepsilon}$ and the effective moduli $\widetilde{\boldsymbol{L}}$ are:

$$\boldsymbol{\varepsilon} = \sum_{j=0}^{\infty} \left(-\boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right)^j \boldsymbol{E}, \quad \widetilde{\boldsymbol{L}} = \boldsymbol{L}^0 + \sum_{j=0}^{\infty} \langle \boldsymbol{\delta L} \left(-\boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right)^j \rangle. \tag{7}$$

The successive terms in these series correspond to successive iterates in the resolution of the Lippmann-Schwinger equation by Picard iterations [1]

$$\boldsymbol{\varepsilon}^{(k+1)} = -\boldsymbol{\Gamma}^0 \boldsymbol{\delta L} \boldsymbol{\varepsilon}^{(k)} + \boldsymbol{E}, \text{ where } \boldsymbol{\varepsilon}^{(k)} = \sum_{j=0}^{k} \left(-\boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right)^j \boldsymbol{E}. \tag{8}$$

[1] Here (as in Milton [3] chap. 14) $\boldsymbol{\Gamma}^0$ and $\boldsymbol{\delta L}$ should be interpreted as *operators* and $\left(-\boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right)^j$ as the operator $-\boldsymbol{\Gamma}^0 \boldsymbol{\delta L}$ applied j times (and not as the field $-\boldsymbol{\Gamma}^0 \boldsymbol{\delta L}$ raised to the j-th power).

Similarly the k-th approximation of the effective tensor $\widetilde{\boldsymbol{L}}$ is given as

$$\widetilde{\boldsymbol{L}}^{(k)} = \boldsymbol{L}^0 + \sum_{j=0}^{k} \langle \boldsymbol{\delta L} \left(-\boldsymbol{\Gamma}^0 \boldsymbol{\delta L} \right)^j \rangle. \tag{9}$$

Convergence of the Liouville-Neumann series (7), or equivalently of the iterative procedure (8), requires specific choices for the reference tensor $\boldsymbol{L}^0$ and/or limitations on the contrast $\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L}$ between the phases. To investigate these conditions it is useful to re-write $\boldsymbol{\Gamma}^0\boldsymbol{\delta L}$ as the composition of two operators

$$\boldsymbol{\Gamma}^0\boldsymbol{\delta L} = \boldsymbol{\Gamma}^0\boldsymbol{L}^0 \; \boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} = \boldsymbol{\Gamma}^1\boldsymbol{Z}, \quad \boldsymbol{\Gamma}^1 = \boldsymbol{\Gamma}^0\boldsymbol{L}^0, \quad \boldsymbol{Z} = \boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L}. \tag{10}$$

$\boldsymbol{\Gamma}^1$ is a projection (for an energetic scalar product, see appendix A) and its norm is therefore equal to 1. The power series $(7)_1$ and $(7)_2$ take the form

$$\boldsymbol{\varepsilon} = \sum_{j=0}^{\infty} \left(-\boldsymbol{\Gamma}^1\boldsymbol{Z} \right)^j \boldsymbol{E}, \quad \boldsymbol{L}^{0^{-1}}\widetilde{\boldsymbol{L}} = \boldsymbol{I} + \sum_{j=0}^{\infty} \langle \boldsymbol{Z} \left(-\boldsymbol{\Gamma}^1\boldsymbol{Z} \right)^j \rangle, \tag{11}$$

where it is seen that, in this approach, the two pivotal operators are $\boldsymbol{\Gamma}^1$ and $\boldsymbol{Z}$.

Eyre and Milton [11] noticed that the power series (7) makes use of powers of the operator $\boldsymbol{\Gamma}^1 = \boldsymbol{\Gamma}^0\boldsymbol{L}^0$. Given that this operator is positive with norm equal to 1 (in energetic norm), they suggested to iterate with a shifted operator and introduced (with different notations)

$$\boldsymbol{H}^1 = 2\boldsymbol{\Gamma}^0\boldsymbol{L}^0 - \boldsymbol{I}. \tag{12}$$

$\boldsymbol{H}^1$ is a self-adjoint operator (for an appropriate scalar product), with all its eigenvalues between -1 and 1 and can therefore replace $\boldsymbol{\Gamma}^1$ in the Lippmann-Schwinger equation. Following Milton [3] an equivalent form of the Lippmann-Schwinger equation and of its associated Neumann-Liouville series can be derived after some algebra. Noting that $\boldsymbol{\Gamma}^0\boldsymbol{L}^0 = \frac{1}{2}(\boldsymbol{H}^1 + \boldsymbol{I})$, one obtains that

$$\boldsymbol{I} + \boldsymbol{\Gamma}^0\boldsymbol{\delta L} = \boldsymbol{I} + \frac{1}{2}(\boldsymbol{H}^1 + \boldsymbol{I})\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} = \boldsymbol{I} + \frac{1}{2}\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} + \frac{1}{2}\boldsymbol{H}^1\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L}. \tag{13}$$

Note that

$$\begin{aligned} \boldsymbol{I} + \tfrac{1}{2}\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} &= \boldsymbol{L}^{0^{-1}}\boldsymbol{L}^0 + \tfrac{1}{2}\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} \\ &= \tfrac{1}{2}\boldsymbol{L}^{0^{-1}}(\boldsymbol{L} + \boldsymbol{L}^0), \end{aligned} \tag{14}$$

and that

$$\begin{aligned} \boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} \left(\boldsymbol{L} + \boldsymbol{L}^0 \right)^{-1} \boldsymbol{L}^0 &= \boldsymbol{L}^{0^{-1}} \left(\boldsymbol{L} + \boldsymbol{L}^0 - 2\boldsymbol{L}^0 \right) \left(\boldsymbol{L} + \boldsymbol{L}^0 \right)^{-1} \boldsymbol{L}^0 = \boldsymbol{I} - 2 \left(\boldsymbol{L} + \boldsymbol{L}^0 \right)^{-1} \boldsymbol{L}^0 \\ &= \left(\boldsymbol{L} + \boldsymbol{L}^0 \right)^{-1} \left(\boldsymbol{L} + \boldsymbol{L}^0 \right) - 2 \left(\boldsymbol{L} + \boldsymbol{L}^0 \right)^{-1} \boldsymbol{L}^0 = \left(\boldsymbol{L} + \boldsymbol{L}^0 \right)^{-1} \boldsymbol{\delta L}. \end{aligned}$$

and therefore

$$\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} = \boldsymbol{W}\boldsymbol{L}^{0^{-1}} \left(\boldsymbol{L} + \boldsymbol{L}^0 \right), \quad \frac{1}{2}\boldsymbol{H}^1\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L} = \frac{1}{2}\boldsymbol{H}^1\boldsymbol{W}\frac{1}{2}\boldsymbol{L}^{0^{-1}}(\boldsymbol{L}+\boldsymbol{L}^0) \quad \text{with} \quad \boldsymbol{W} = \left(\boldsymbol{L} + \boldsymbol{L}^0 \right)^{-1} \boldsymbol{\delta L}. \tag{15}$$

Substituting (14) and (15) into (13) yields

$$\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L} = (\boldsymbol{I} + \boldsymbol{H}^1 \boldsymbol{W}) \frac{1}{2} \boldsymbol{L}^{0^{-1}} (\boldsymbol{L} + \boldsymbol{L}^0). \tag{16}$$

The Lippmann-Schwinger equation (4) and its associated Neumann series (7) become

$$\left(\boldsymbol{I} + \boldsymbol{H}^1 \boldsymbol{W}\right) \frac{1}{2} \boldsymbol{L}^{0^{-1}} (\boldsymbol{L} + \boldsymbol{L}^0) \boldsymbol{\varepsilon} = \boldsymbol{E}, \quad \boldsymbol{\varepsilon} = 2(\boldsymbol{L} + \boldsymbol{L}^0)^{-1} \boldsymbol{L}^0 \sum_{j=0}^{\infty} (-\boldsymbol{H}^1 \boldsymbol{W})^j \boldsymbol{E}. \tag{17}$$

The corresponding series expansion for the effective moduli is obtained by writing that

$$\boldsymbol{\sigma} = \boldsymbol{L}^0 : \boldsymbol{\varepsilon} + \boldsymbol{\delta L} : \boldsymbol{\varepsilon}, \quad \text{i.e.} \quad \langle \boldsymbol{\sigma} \rangle = \boldsymbol{L}^0 : \boldsymbol{E} + \langle \boldsymbol{\delta L} : \boldsymbol{\varepsilon} \rangle$$

and using again (14) and (15), one gets

$$\boldsymbol{L}^{0^{-1}} \widetilde{\boldsymbol{L}} = \boldsymbol{I} + 2 \sum_{j=0}^{\infty} \langle \boldsymbol{W} \left(-\boldsymbol{H}^1 \boldsymbol{W}\right)^j \rangle, \tag{18}$$

which evidences the similarity with (11), $\boldsymbol{Z}$ being substituted with $\boldsymbol{W}$, and $\boldsymbol{\Gamma}^1$ with $\boldsymbol{H}^1$.
Remark 1: Truncated sums corresponding to $\boldsymbol{\varepsilon}^{(k)}$ in (8) can also be considered within the Eyre-Milton scheme. Defining [2]

$$\boldsymbol{e}^{(k)} = 2(\boldsymbol{L} + \boldsymbol{L}^0)^{-1} \boldsymbol{L}^0 \sum_{j=0}^{k} (-\boldsymbol{H}^1 \boldsymbol{W})^j \boldsymbol{E}, \tag{19}$$

and using (16), the following recurrence relation is obtained

$$\left(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}\right) \boldsymbol{e}^{(k)} = \left(\boldsymbol{I} + \boldsymbol{H}^1 \boldsymbol{W}\right) \sum_{j=0}^{k} (-\boldsymbol{H}^1 \boldsymbol{W})^j = \boldsymbol{E} + \frac{1}{2} \boldsymbol{L}^{0^{-1}} (\boldsymbol{L} + \boldsymbol{L}^0) \left(\boldsymbol{e}^{(k)} - \boldsymbol{e}^{(k+1)}\right),$$

or equivalently,

$$\boldsymbol{e}^{(k+1)} = \boldsymbol{e}^{(k)} - 2(\boldsymbol{L} + \boldsymbol{L}^0)^{-1} \boldsymbol{L}^0 \left[(\boldsymbol{I} + \boldsymbol{\Gamma}^0 \boldsymbol{\delta L}) \boldsymbol{e}^{(k)} - \boldsymbol{E}\right], \tag{20}$$

which is equivalent to the writing of the Eyre-Milton algorithm in Michel *et al* [10].
Remark 2: Noting that $\boldsymbol{\Gamma}^0 \boldsymbol{\delta L} \boldsymbol{\varepsilon} = \boldsymbol{E} - \boldsymbol{\varepsilon}$, the last equation in (17) can be rewritten as

$$\boldsymbol{\varepsilon} = \boldsymbol{E} - 2\boldsymbol{\Gamma}^0 \boldsymbol{W} \boldsymbol{L}^0 \sum_{j=0}^{\infty} (-\boldsymbol{H}^1 \boldsymbol{W})^j \boldsymbol{E}. \tag{21}$$

[2]The notation $\boldsymbol{e}^{(k)}$ is used here instead of $\boldsymbol{\varepsilon}^{(k)}$ to highlight the fact that the truncated sums are not guaranteed compatible strain fields (except at convergence), unlike the truncated series (8) and (22).

The partial sums

$$\boldsymbol{\varepsilon}^{(k)} = \boldsymbol{E} - 2\boldsymbol{\Gamma}^0 \boldsymbol{W} \boldsymbol{L}^0 \sum_{j=0}^{k} (-\boldsymbol{H}^1 \boldsymbol{W})^j \boldsymbol{E} = \boldsymbol{E} - \boldsymbol{\Gamma}^0 \boldsymbol{\delta} \boldsymbol{L} \boldsymbol{e}^{(k)}, \tag{22}$$

then define a sequence of conforming approximations to $\boldsymbol{\varepsilon}$, in the sense that each field $\boldsymbol{\varepsilon}^{(k)}$ is a compatible strain field, i.e., the gradient of some symmetrized displacement gradient. Note that in the process of computing the iterates in (20) one calculates $\boldsymbol{\Gamma}^0 \boldsymbol{\delta} \boldsymbol{L} \boldsymbol{e}^{(k)}$ thus immediately giving $\boldsymbol{\varepsilon}^{(k)}$ from (22) with no extra work. In conclusion, it is inaccurate to say that the Eyre-Milton algorithm is a non-conforming approximation scheme —one only needs to keep track of the fields $\boldsymbol{\varepsilon}^{(k)}$.

Remark 3: An alternative Neumann series for the Eyre-Milton scheme can be obtained by writing the iterative algorithm (20) in terms of the polarization field

$$\boldsymbol{\tau} = (\boldsymbol{L} + \boldsymbol{L}^0)\boldsymbol{e}. \tag{23}$$

Relation (20), re-written as

$$(\boldsymbol{L} + \boldsymbol{L}^0)\boldsymbol{e}^{k+1} = (\boldsymbol{L} + \boldsymbol{L}^0)\boldsymbol{e}^k - 2\boldsymbol{L}^0\boldsymbol{e}^k - 2\boldsymbol{L}^0\boldsymbol{\Gamma}^0 \left(\boldsymbol{L} - \boldsymbol{L}^0\right) \boldsymbol{e}^k + 2\boldsymbol{L}^0\boldsymbol{E}\,, \tag{24}$$

can be expressed with the polarization field as

$$\boldsymbol{\tau}^{k+1} = -\mathcal{H}^1 \mathcal{W} \boldsymbol{\tau}^k + 2\boldsymbol{L}^0\boldsymbol{E}\,,\ \mathcal{H}^1 = 2\boldsymbol{L}^0\boldsymbol{\Gamma}^0 - \boldsymbol{I},\ \mathcal{W} = (\boldsymbol{L} - \boldsymbol{L}^0)(\boldsymbol{L} + \boldsymbol{L}^0)^{-1}. \tag{25}$$

Note that the operators $\mathcal{H}^1$ and $\mathcal{W}$ are slightly different from the operators $\boldsymbol{H}^1$ and $\boldsymbol{W}$: see (12) and (16).

The series expansion corresponding to (25) now reads

$$\boldsymbol{\tau}^k = \sum_{k=0}^{k} 2\left(-\mathcal{H}^1\mathcal{W}\right)^k\ \boldsymbol{L}^0\boldsymbol{E}\,. \tag{26}$$

At convergence $\boldsymbol{e} = \boldsymbol{\varepsilon}$ and the following two expressions for $\langle\boldsymbol{\tau}\rangle$ resulting from (23) and (26),

$$\boldsymbol{T} = <\boldsymbol{\tau}> = \sum_{k=0}^{\infty} 2 < \left(-\mathcal{H}^1\mathcal{W}\right)^k > \ \boldsymbol{L}^0\boldsymbol{E}$$

and

$$\boldsymbol{T} = <\boldsymbol{\tau}> = \boldsymbol{L}^0 <\boldsymbol{\varepsilon}> + <\boldsymbol{\sigma}> = \boldsymbol{L}^0\boldsymbol{E} + \tilde{\boldsymbol{L}}\boldsymbol{E}, \tag{27}$$

lead to

$$\tilde{\boldsymbol{L}}\boldsymbol{L}^{0^{-1}} = \boldsymbol{I} + \sum_{k=1}^{\infty} 2\langle\left(-\mathcal{H}^1\mathcal{W}\right)^k\rangle. \tag{28}$$

Noting that

$$\begin{aligned} \langle\left(\mathcal{H}^1\mathcal{W}\right)^k\rangle &= \langle(2\boldsymbol{L}^0\boldsymbol{\Gamma}^0 - \boldsymbol{I})\mathcal{W}\left(\mathcal{H}^1\mathcal{W}\right)^{k-1}\rangle \\ &= 2\boldsymbol{L}^0\langle\boldsymbol{\Gamma}^0\mathcal{W}\left(\mathcal{H}^1\mathcal{W}\right)^{k-1}\rangle - \langle\mathcal{W}\left(\mathcal{H}^1\mathcal{W}\right)^{k-1}\rangle = -\langle\mathcal{W}\left(\mathcal{H}^1\mathcal{W}\right)^{k-1}\rangle, \end{aligned} \tag{29}$$

one gets from (28)

$$\tilde{\boldsymbol{L}}\boldsymbol{L}^{0^{-1}} = \boldsymbol{I} + 2\sum_{j=0}^{\infty}\langle \boldsymbol{\mathcal{W}}\big(-\mathfrak{H}^1\boldsymbol{\mathcal{W}}\big)^j\rangle, \tag{30}$$

which is similar to (18).

2.3. Unconditional conditions for convergence of the power series

"Unconditional" refers to sufficient conditions which are independent of the microstructure under consideration. Conservative estimates on the range of contrast for which the power series $(7)_1$ converges, can be obtained by estimating the norm (defined in an appropriate way) of $\boldsymbol{\Gamma}^0\boldsymbol{\delta L}$. This can be done in a first attempt by using the decomposition (10).

It has been already recalled that the operator $\boldsymbol{\Gamma}^1$ has norm equal to 1 independently of the microstructure (appendix A). Therefore

$$\big|\big|\boldsymbol{\Gamma}^0\boldsymbol{\delta L}\big|\big| \leq \Big|\Big|\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L}\Big|\Big|. \tag{31}$$

A sufficient condition for convergence of the power series is therefore that

$$\Big|\Big|\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L}\Big|\Big| < 1. \tag{32}$$

When the reference medium is chosen to be one of the phases, as proposed by Brown [4], the inequality (32) imposes a restriction on the moduli of the other phases which, roughly speaking, have to be lower than twice the moduli of the phase taken as reference. The corresponding scheme (7) will be called B-scheme (B for Brown) in the sequel.

Moulinec and Suquet [5] suggested to consider $\boldsymbol{L}^0$ as a free parameter and to choose it in a such a way that (32) is satisfied. There is a wide range of possible reference media. For instance taking $\boldsymbol{L}^0$ greater (in the sense of quadratic forms) than all the individual moduli $\boldsymbol{L}^{(r)}, r = 1, .., N$ of the phases ensures that (32) is satisfied with no additional restriction on the contrast between the phases. However, even when the convergence of the power series (7) is ensured, its convergence rate, which is actually strongly dependent on the reference medium, can be low. This convergence rate is directly related to the spectral radius of $\boldsymbol{\Gamma}^0\boldsymbol{\delta L}$ which is bounded from above by $\Big|\Big|\boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L}\Big|\Big|$ independently of the microstructure. The "unconditionally" optimal reference medium (with no information on the microstructure) is obtained by solving the optimization problem

$$\underset{\boldsymbol{L}^0}{\text{Min}}\ \Big|\Big|\boldsymbol{L}^{0^{-1}} : \boldsymbol{\delta L}\Big|\Big|, \tag{33}$$

or equivalently for a N-phase composite, the discrete optimization problem

$$\underset{\boldsymbol{L}^0}{\text{Min}}\ \underset{r\,=\,1,...,N}{\text{Max}}\ \Big|\Big|\boldsymbol{L}^{0^{-1}} : (\boldsymbol{L}^{(r)} - \boldsymbol{L}^0)\Big|\Big|. \tag{34}$$

This optimization problem leads for instance in a two phase composite with isotropic phases to (Moulinec and Suquet [6], Michel *et al* [10], Milton [3])

$$\boldsymbol{L}^{(0)} = \frac{1}{2}(\boldsymbol{L}^{(1)} + \boldsymbol{L}^{(2)}). \tag{35}$$

The corresponding scheme (7) with the choice (35) will be referred to as the MS-scheme (for Moulinec-Suquet) in the sequel.

Similar sufficient conditions can be obtained for the Eyre-Milton version (17) of the Lippmann-Schwinger equation. The operator $\boldsymbol{H}^1$ has norm 1 (with an appropriate choice of the scalar product), and the spectral radius of $\boldsymbol{H}^1\boldsymbol{W}$ satisfies

$$\left|\left|\boldsymbol{H}^1\boldsymbol{W}\right|\right| \leq ||\boldsymbol{W}||\,. \tag{36}$$

Therefore convergence of the power series $(17)_2$ is assured whenever

$$||\boldsymbol{W}|| = \left|\left|(\boldsymbol{L} + \boldsymbol{L}^0)^{-1}(\boldsymbol{L} - \boldsymbol{L}^0)\right|\right| < 1. \tag{37}$$

This condition is always satisfied when $\boldsymbol{L}$ and $\boldsymbol{L}^0$ are definite positive, bounded, and bounded from below by some positive constant of coercivity. Again a large choice of reference media is possible. The "unconditionally" optimal choice for $\boldsymbol{L}^0$ for the requirement (37) is obtained by minimizing the left-hand side (37)

$$\underset{\boldsymbol{L}^0}{\text{Min}}\ \left|\left|(\boldsymbol{L} + \boldsymbol{L}^0)^{-1}(\boldsymbol{L} - \boldsymbol{L}^0)\right|\right|. \tag{38}$$

An explicit solution to this minimization problem can be derived when the tensors $\boldsymbol{L}$ and $\boldsymbol{L}^0$ have a more specific form, as will be illustrated in section 2.4. The scheme (17) with this optimized reference medium will be referred to as the EM-scheme (for Eyre-Milton) in the sequel.

The microstructure does not enter the optimization problems (33) and (38). Solving these minimization problems ensures that the spectral radius of the two operators $\boldsymbol{\Gamma}^0\boldsymbol{\delta L} = \boldsymbol{\Gamma}^1\boldsymbol{Z}$ and $\boldsymbol{H}^1\boldsymbol{W}$ is always smaller than 1, for any microstructure and any strictly positive and finite moduli in the phases.

However in the extreme case of one phase being void (with zero moduli), the upper bounds (31) and (36) on the spectral radii of the iterative methods are exactly 1. Convergence is not ensured. The situation is even worse when the modulus of one of the phases is nonpositive while the moduli of the other phases are positive. In the other extreme case, one of the phases having infinite moduli, the "optimal" reference medium itself has infinite moduli (and it does not make sense to consider its Green's operator $\boldsymbol{\Gamma}^0$).

A sharper estimate of this spectral radius can be obtained by taking into account the microstructure of the composite. This is the aim of section 3 where the class of composites under consideration will be restricted in order to avoid complicated notations.

2.4. A class of composite materials with a scalar measure of the contrast

In general, the material properties of the phases depend on several moduli, either because the phases are anisotropic or because the tensors describing the material properties under consideration, even when they are isotropic, depend on several moduli (isotropic elastic tensors depend on two moduli, a bulk modulus and a shear modulus). Therefore the contrast between the different phases depend on several contrast variables.

The results of the forthcoming sections are better understood when the contrast between the different phases is measured with a single scalar parameter, called z in the sequel. This is in particular the case when all tensorial moduli $\boldsymbol{L}^{(r)}$ are proportional to the same tensor $\boldsymbol{\Lambda}$ and when the moduli of the reference medium are also chosen proportional to the same tensor $\boldsymbol{\Lambda}$. For simplicity again, it will be sufficient for our purpose here to restrict attention to two-phase composites

$$\boldsymbol{L}^{(1)} = z\boldsymbol{\Lambda}, \quad \boldsymbol{L}^{(2)} = \boldsymbol{\Lambda}, \quad \boldsymbol{L}^{0} = z_0\boldsymbol{\Lambda}, \tag{39}$$

where after a proper rescaling, the prefactor of $\boldsymbol{\Lambda}$ in the phases can be set equal to 1 in phase 2, to z in phase 1 and to z_0 in the reference medium. Two typical cases are

1. Effective conductivity of two-phase composites with isotropic phases. Then

$$\boldsymbol{\Lambda} = \boldsymbol{I}, \tag{40}$$

   and z is the conductivity of phase 1 (the conductivity of phase 2 being 1).

2. Effective elastic moduli of two-phase composites with isotropic phases sharing the same Poisson ratio ν. Then

$$\boldsymbol{\Lambda} = \frac{1}{1-2\nu}\boldsymbol{J} + \frac{1}{1+\nu}\boldsymbol{K}, \tag{41}$$

   where $\boldsymbol{J}$ and $\boldsymbol{K}$ are the usual projectors on second-order spherical and deviatoric symmetric tensors respectively, z is the Young modulus of phase 1 (phase 2 having a normalized Young modulus equal to 1). The case of incompressible elastic phases considered in [25] is a special case of (41) with, as usual, some care to be taken to handle properly incompressible phases ($\nu = 1/2$).

Under these assumptions the operators $\boldsymbol{\Gamma}^1$, $\boldsymbol{H}^1$ are independent of the modulus z_0 of the reference medium, and they take a simple form,

$$\boldsymbol{\Gamma}^1 = \boldsymbol{\Gamma}^{(\boldsymbol{\Lambda})}\boldsymbol{\Lambda}, \quad \boldsymbol{H}^1 = 2\boldsymbol{\Gamma}^{(\boldsymbol{\Lambda})}\boldsymbol{\Lambda} - \boldsymbol{I}, \tag{42}$$

where $\boldsymbol{\Gamma}^{(\boldsymbol{\Lambda})}$ is the Green's operator $\boldsymbol{\Gamma}^0$ for $\boldsymbol{\Lambda}$. The local tensor fields $\boldsymbol{Z}$ and $\boldsymbol{W}$ depend on the contrast z as

$$\left.\begin{aligned} \boldsymbol{Z}(\boldsymbol{x}) &= \left[\frac{z-z_0}{z_0}\chi^{(1)}(\boldsymbol{x}) + \frac{1-z_0}{z_0}\chi^{(2)}(\boldsymbol{x})\right]\boldsymbol{I}, \\ \boldsymbol{W}(\boldsymbol{x}) &= \left[\frac{z-z_0}{z+z_0}\chi^{(1)}(\boldsymbol{x}) + \frac{1-z_0}{1+z_0}\chi^{(2)}(\boldsymbol{x})\right]\boldsymbol{I}. \end{aligned}\right\} \tag{43}$$

The operators $\boldsymbol{Z}$ and $\boldsymbol{W}$ are local so their operator norm is a function of their pointwise norm in each phase

$$||\boldsymbol{Z}|| = \text{Max}\left(\left|\frac{z-z_0}{z_0}\right|, \left|\frac{1-z_0}{z_0}\right|\right)|\boldsymbol{I}|\,, \quad ||\boldsymbol{W}|| = \text{Max}\left(\left|\frac{z-z_0}{z+z_0}\right|, \left|\frac{1-z_0}{1+z_0}\right|\right)|\boldsymbol{I}|\,, \tag{44}$$

where $|\boldsymbol{I}|$ is the Euclidian norm of the identity which can be set equal to 1 by a proper rescaling. The convergence of three different series expansions will be investigated in the remainder of the paper. The first two series correspond to the expansion (7) with the two variants mentioned above, B-scheme when the reference medium is one of the phases, MS-scheme when the reference medium minimizes $||\boldsymbol{Z}||$ (optimization problem (33)). The third series is the EM-scheme obtained with the expansion (17) and the reference medium minimizing $||\boldsymbol{W}||$ (optimization problem (38)).

1. *B-scheme:* When the reference medium is phase 2 ($z_0 = 1$), then $||\boldsymbol{Z}|| = |z-1|$ and the sufficient condition (32) (independent of the microstructure) is

$$|z-1| < 1. \tag{45}$$

2. *MS-scheme:* The optimal reference medium minimizing $||\boldsymbol{Z}||$ is found by solving the minimization problem

$$\underset{z_0}{\text{Min}}\ \text{Max}\left(\left|\frac{z-z_0}{z_0}\right|, \left|\frac{1-z_0}{z_0}\right|\right). \tag{46}$$

   The optimal choice is found by studying the variations with z_0 of the two functions of $\left|\frac{z-z_0}{z_0}\right|$ and $\left|\frac{1-z_0}{z_0}\right|$ and the result is

$$z_0 = \frac{z+1}{2}, \quad ||\boldsymbol{Z}|| = \left|\frac{z-1}{z+1}\right|, \quad ||\boldsymbol{Z}|| < 1 \ \Leftrightarrow\ z \in]0, +\infty). \tag{47}$$

   The spectral radius of $\boldsymbol{\Gamma^0 \delta L}$ being bounded by $||\boldsymbol{Z}||$, the series (7)- (9) are convergent for all $0 < z < +\infty$. However the criterion (32) does not ensure convergence of the series when $z = 0$, or $z < 0$, or $z = +\infty$.

3. *EM-scheme:* Finally the optimal reference medium minimizing $||\boldsymbol{W}||$ is found by solving the minimization problem

$$\underset{z_0}{\text{Min}}\ \text{Max}\left(\left|\frac{z-z_0}{z+z_0}\right|, \left|\frac{1-z_0}{1+z_0}\right|\right). \tag{48}$$

   The optimal choice is found by studying the variations with z_0 of the two functions of $\left|\frac{z-z_0}{z+z_0}\right|$ and $\left|\frac{1-z_0}{1+z_0}\right|$ and the result is

$$z_0 = \sqrt{z}, \quad ||\boldsymbol{W}|| = \left|\frac{\sqrt{z}-1}{\sqrt{z}+1}\right|, \quad ||\boldsymbol{W}|| < 1 \ \Leftrightarrow\ z \in]0, +\infty). \tag{49}$$

   The spectral radius of $\boldsymbol{H^1 W}$ being bounded by $||\boldsymbol{W}||$, the series (17)- (18) are convergent for all $0 < z < +\infty$. Again the criterion (37) does not ensure convergence of the series when $z = 0$, or $z < 0$, or $z = +\infty$.

The aim of the following section is to show that, with additional information on the microstructure, it is possible to improve on the conservative estimates (45), (47) and (49) for the radius of convergence of the power series.

3. Extending the estimation of the domain of convergence of iterative methods with information on the microstructure

3.1. What analytic properties of effective properties tell about the radius of convergence of iterative methods

Analytic properties of the effective moduli as function of the contrast between the phases can be used to extend the domain of convergence of the power series. In the interest of simplicity we will consider here only composites following the requirements of section 2.4, where the contrast can be measured by a scalar parameter. Two features of analytic functions are helpful in estimating the domain of convergence of iterative methods:

1. If f is an analytic function on a domain $\mathcal{D}$ of the complex plane, it coincides with its Taylor series at any point of $\mathcal{D}$, in any disk centered at that point and lying within $\mathcal{D}$.
2. If f is analytic on $\mathcal{D}$ and g is analytic on Ω containing the range of f then $g(f(z))$ is analytic on $\mathcal{D}$.

Bergman [26] proposed that in conductivity problems $\widetilde{\boldsymbol{L}}$ is an analytic function of z in the complex plane minus the negative real axis. A correct argument validating this was given in Milton [27] and a more rigorous proof was given by Golden and Papanicolaou [28]. Therefore $\widetilde{\boldsymbol{L}}(z)$ can be expanded into a power series in the neighborhood of any point in the complex plane minus the negative real axis. This result is valid for *any microstructure*. The radius of convergence of this series is that of the largest disk contained in the complex plane minus the real negative axis. For instance, the expansion of $\widetilde{\boldsymbol{L}}(z)$ in the neighborhood of $z = 1$ has a radius of convergence at least equal to 1.

Interestingly, the domain of analyticity of $\widetilde{\boldsymbol{L}}(z)$ can be extended when more information about the singularities (poles or branch-cuts) of the function $\widetilde{\boldsymbol{L}}(z)$ is available. This additional information often comes from an explicit expression of $\widetilde{\boldsymbol{L}}$ as a function of z. For instance, the two Hashin-Strikman bounds on the effective conductivity of isotropic two-phase composites, which are attained by a wide class of isotropic microstructures, have a single pole on the negative axis. Their typical form is

$$\widetilde{\boldsymbol{L}}(z) = z^* \boldsymbol{I}, \quad z^* = 1 + \alpha \frac{z-1}{z+\beta}, \tag{50}$$

where α and β are positive scalars depending on the volume fraction of the phases, on the dimension of space and on the type of bound (upper or lower) (see Milton [3]). Let us also mention that if the Hashin-Shtrikman bound (50) has a single isolated pole, there are other microstructures for which the whole real negative axis is singular. The checkerboard, with effective conductivity $z^* = \sqrt{z}$, is an example of such a situation. In between, there are microstructures where the poles or branch-cuts are localized. Even in the absence of complete

information about the microstructure one can sometimes restrict the interval on the real negative axis where singularities occur [29]. If inclusions have sharp corners then this too gives information about the singularities (Hetherington and Thorpe [30] and Milton [3], sect.18.3).

Therefore, when the locations of the singularities of $\widetilde{\boldsymbol{L}}(z)$ are known or partially known for a given microstructure, its known domain of analyticity may be significantly larger than the complex plane minus the negative real axis. Consequently the radius of convergence of power series can be automatically extended to all disks contained in this domain. The counterpart for Neumann series is that their radius of convergence might be significantly larger than the conservative estimates of section 2.3. An example of such a situation will be considered in section 4.

3.2. Power-series expansions accounting explicitly for the microstructure

Apart from the analyticity of the effective moduli with respect to the contrast, it is possible to account explicitly for the microstructure by incorporating it in the operators entering the decomposition of $\boldsymbol{\Gamma}^0\boldsymbol{\delta L}$. Instead of the decomposition (10) one can write

$$\boldsymbol{\Gamma}^0\boldsymbol{\delta L} = \sum_{r=1}^{N} \boldsymbol{\Gamma}^1\chi^{(r)}\boldsymbol{Z}^{(r)}, \quad \boldsymbol{Z}^{(r)} = \boldsymbol{L}^{0^{-1}}\boldsymbol{\delta L}^{(r)}, \tag{51}$$

where $\chi^{(r)}$ is the characteristic function of phase r. The operator $\boldsymbol{\Gamma}^0\boldsymbol{\delta L}$ entering the Liouville-Neumann series (7)-(9) is therefore the composition of two operators:

1. The operators $\boldsymbol{Z}^{(r)}$ depend only the *material properties* of the phases (and on the reference medium). They serve as measures of the contrast of the composite.
2. The operators $\boldsymbol{\Gamma}^1\chi^{(r)}$ depend on the *microstructure* of the composite through $\chi^{(r)}$ (but not exclusively since $\boldsymbol{\Gamma}^1$ may depend on the reference medium). These operators express the way in which the operator $\boldsymbol{\Gamma}^1$ "sees" the microstructure.

The decomposition (51) highlights the contribution of the material properties on one hand and of the microstructure on the other hand. In a quite similar way the operator entering the Eyre-Milton scheme can be written as

$$\boldsymbol{H}^1\boldsymbol{W} = \sum_{r=1}^{N} \boldsymbol{H}^1\chi^{(r)}\boldsymbol{W}^{(r)}, \quad \boldsymbol{W}^{(r)} = (\boldsymbol{L}^{(r)} + \boldsymbol{L}^0)^{-1}\boldsymbol{\delta L}^{(r)}. \tag{52}$$

The general case being rather technical, attention is limited here to two-phase composite whose contrast can be measured by a single scalar z, as described in section 2.4. The expression of $\boldsymbol{Z}$ and $\boldsymbol{W}$ are given in (43) and the decompositions (51) and (52) read as

$$\left.\begin{aligned} \boldsymbol{\Gamma}^0\boldsymbol{\delta L} = \boldsymbol{\Gamma}^1\boldsymbol{Z} &= \frac{z-z_0}{z_0}\boldsymbol{\Gamma}_{\chi^{(1)}} + \frac{1-z_0}{z_0}\boldsymbol{\Gamma}_{\chi^{(2)}}, \\ \boldsymbol{H}^1\boldsymbol{W} &= \frac{z-z_0}{z+z_0}\boldsymbol{H}_{\chi^{(1)}} + \frac{1-z_0}{1+z_0}\boldsymbol{H}_{\chi^{(2)}}, \end{aligned}\right\} \tag{53}$$

where

$$\left.\begin{array}{c} \boldsymbol{\Gamma}_{\chi^{(r)}} = \boldsymbol{\Gamma}^1 \chi^{(r)} \boldsymbol{I}, \quad \boldsymbol{\Gamma}_{\chi^{(1)}} + \boldsymbol{\Gamma}_{\chi^{(2)}} = \boldsymbol{\Gamma}, \quad \boldsymbol{\Gamma} = \boldsymbol{\Gamma}^1 \boldsymbol{I}, \\ \boldsymbol{H}_{\chi^{(1)}} = \boldsymbol{H}^1 \chi^{(1)} \boldsymbol{I}, \quad \boldsymbol{H}_{\chi^{(1)}} + \boldsymbol{H}_{\chi^{(2)}} = \boldsymbol{H}, \quad \boldsymbol{H} = \boldsymbol{H}^1 \boldsymbol{I}. \end{array}\right\} \tag{54}$$

It follows from (53) that the key operators in the iterations for the three cases of interest identified in section 2.4 are:

Reference medium= Matrix (B-scheme):

$$\left.\begin{array}{l} z_0 = 1, \quad \boldsymbol{\Gamma}^0 \boldsymbol{\delta L} = t \boldsymbol{\Gamma}_{\chi^{(1)}}, \quad t = z - 1 \\ \boldsymbol{L}^{0^{-1}} \widetilde{\boldsymbol{L}} = \boldsymbol{I} + \sum_{j=0}^{\infty} (-1)^j \langle \chi^{(1)} (\boldsymbol{\Gamma}_{\chi^{(1)}})^j \rangle t^{j+1}. \end{array}\right\} \tag{55}$$

Reference medium= Arithmetic mean (MS-scheme):

$$\left.\begin{array}{l} z_0 = \dfrac{z+1}{2}, \quad \boldsymbol{\Gamma}^0 \boldsymbol{\delta L} = t \left(\boldsymbol{\Gamma}_{\chi^{(1)}} - \boldsymbol{\Gamma}_{\chi^{(2)}} \right), \quad t = \dfrac{z-1}{z+1}, \\ \boldsymbol{L}^{0^{-1}} \widetilde{\boldsymbol{L}} = \boldsymbol{I} + \sum_{j=0}^{\infty} (-1)^j \langle (\chi^{(1)} - \chi^{(2)}) (\boldsymbol{\Gamma}_{\chi^{(1)}} - \boldsymbol{\Gamma}_{\chi^{(2)}})^j \rangle t^{j+1}. \end{array}\right\} \tag{56}$$

Reference medium= Geometric mean (EM-scheme):

$$\left.\begin{array}{l} z_0 = \sqrt{z}, \boldsymbol{H}^1 \boldsymbol{W} = t \left(\boldsymbol{H}_{\chi^{(1)}} - \boldsymbol{H}_{\chi^{(2)}} \right), \quad t = \dfrac{\sqrt{z}-1}{\sqrt{z}+1}, \\ \boldsymbol{L}^{0^{-1}} \widetilde{\boldsymbol{L}} = \boldsymbol{I} + 2 \sum_{j=0}^{\infty} (-1)^j \langle (\chi^{(1)} - \chi^{(2)}) (\boldsymbol{H}_{\chi^{(1)}} - \boldsymbol{H}_{\chi^{(2)}})^j \rangle t^{j+1}. \end{array}\right\} \tag{57}$$

The expressions of the effective moduli in (56) and (57) evidence the symmetric role played by the two phases (they also play a symmetric role in the reference medium), whereas the first expression is asymmetric (because the reference medium is one of the two phases). The characteristic function of one phase can be eliminated from the last two relations in (56) by noting that

$$\chi^{(1)} - \chi^{(2)} = 2\chi^{(1)} - 1 \stackrel{def}{=} \chi', \quad \boldsymbol{\Gamma}_{\chi^{(1)}} - \boldsymbol{\Gamma}_{\chi^{(2)}} = \boldsymbol{\Gamma}_{\chi'}, \quad \boldsymbol{H}_{\chi^{(1)}} - \boldsymbol{H}_{\chi^{(2)}} = \boldsymbol{H}_{\chi'}. \tag{58}$$

It is remarkable that, in the series expansion (55), (56) and (57), the variable t measuring the contrast between the phases and the microstructure are decoupled. In other words, one could compute the micro-structural coefficients $\langle \chi^{(1)} (\boldsymbol{\Gamma}_{\chi^{(1)}})^j \rangle$, $\langle \chi' (\boldsymbol{\Gamma}_{\chi'})^j \rangle$, $\langle \chi' (\boldsymbol{H}_{\chi'})^j \rangle$ once for all. And for any arbitrary contrast t the effective property would be obtained by summing the series (55), (56) or (57), without having to solve the Lippmann-Schwinger equation. A similar observation has previously been made by Hoang and Bonnet [25] for incompressible phases and recently by To *et al*[16] for elastic phases with the same Poisson ratio. It is also closely related to the role of correlation functions of increasing order into the series expansions of the effective moduli (Milton [3] chap. 15).

4. A study case: the Obnosov formula for the conductivity of a square array of square inclusions

Obnosov [24] derived a closed form expression for the conductivity of a square array of square inclusions with volume fraction 0.25 (cf Figure 1). This microstructure has square sym-

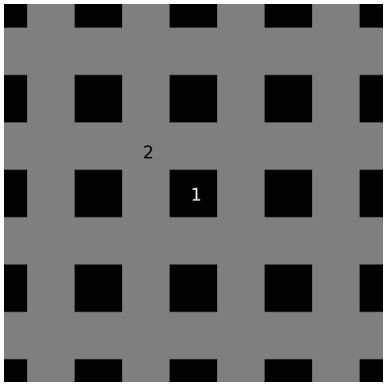


Figure 1: Obnosov microstructure: square array of square inclusions with volume fraction 0.25.

metry and its effective conductivity tensor is isotropic

$$\widetilde{\boldsymbol{L}}(z) = \widetilde{z}\boldsymbol{I}, \quad \widetilde{z} = \sqrt{\frac{1+3z}{3+z}}, \tag{59}$$

where, after proper rescaling the conductivity of phase 1 (inclusion) is z and that of phase 2 (matrix) is 1.

4.1. Convergence of the exact power series

Exact power series expansions for $\widetilde{z}$, or alternatively for $\widetilde{z}/z_0$ in the form

$$\widetilde{z} = \sum_{k=0}^{\infty} d_k t^k, \quad \frac{\widetilde{z}}{z_0} = \sum_{k=0}^{\infty} b_k t^k, \tag{60}$$

can be derived from (59) in the three cases of interest described in section 2.4, where the contrast variable t depends on the choice of the reference medium. Typically, $t = z-1$ or $t = Z = \dfrac{z-1}{z+1}$ or $t = w = \dfrac{\sqrt{z}-1}{\sqrt{z}+1}$. These power series are given in appendix Appendix B for the three cases of interest.

The conservative estimates of section 2.3 ensure convergence of these power series for $z \in \mathbb{C} - \mathbb{R}^-$, *i.e.* when z satisfies the conditions (45) or (47) (which coincides with (49)). These conservative estimates can be improved by means of the analytic properties of the effective conductivity (59).

Consider first the expansion with respect to $z-1$ (corresponding to the matrix as reference medium). $\widetilde{z}(z)$ has a branch-cut $[-3,-\frac{1}{3}]$ on the negative real axis but is analytic in the whole complex plane outside this branch-cut. Therefore the power series of $\widetilde{z}$ in the neighborhood of $z=1$ will converge in all disks centered at $z=1$ and not intersecting the branch-cut, *i.e.* for $|z-1|<4/3$. Considering only real values for z, the power series converges when $-1/3<z<7/3$ and in particular when $z=0$. This extended domain of convergence is confirmed in Figure 2(a) where the deviation from the exact result (59) as a function of the number n of terms in the series is measured by

$$\epsilon_n = \left| \widetilde{z}^{(n)} - \sqrt{\frac{1+3z}{3+z}} \right|, \quad \widetilde{z}^{(n)} = \sum_{k=0}^{n} d_k t^k. \tag{61}$$

Note in particular that the series converges (although slowly) even when $z=-0.3$, in accordance with the theoretical predictions.

The situation is similar for the power series in $Z=\frac{z-1}{z+1}$. The effective conductivity (59) reads as

$$\widetilde{z} = \sqrt{\frac{1+Z/2}{1-Z/2}}, \tag{62}$$

which has two branch-cuts $(-\infty,-2]\cup[2,+\infty)$. The disk of convergence of its power series expansion in the neighborhood of $Z=0$ is therefore $|Z|<2$.

Limiting attention to z on the real axis, the power series in Z converges when $z\in(-\infty,-3[\cup]-1/3,+\infty)$. Therefore the range of z for which this series converges is larger than in the previous expansion with respect to $z-1$. This theoretical prediction is confirmed by the numerical experiments reported in figure 2(b) where the error between the partial series and the exact effective conductivity is still measured by (61), but with $t=\frac{z-1}{z+1}$. Note in particular that the series converges for $z=-10$.

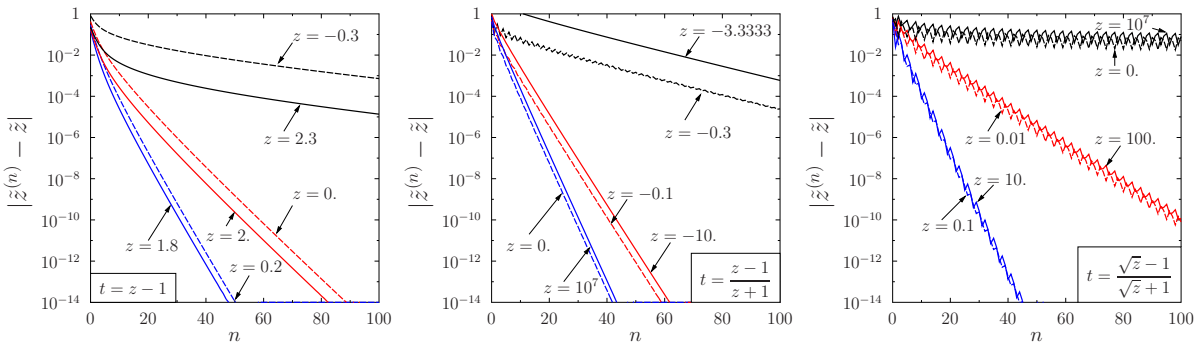
The expansion with respect to $t=(\sqrt{z}-1)/(\sqrt{z}+1)$ is slightly different since, in order for $\sqrt{z}$ to be properly defined, z has to be in the cut complex plane (outside the negative real axis). The zig-zags are due to the absolute value in the error (the successive truncated sums are located above and below the exact effective conductivity). The convergence of the corresponding series is quite fast at moderate contrast, but is not very fast at high contrast ($z=0$ and $z=10^7$). This is expected as the square root in $\sqrt{z}$ introduces a branch cut of singularities extending from zero to infinity.

4.2. Convergence of numerical series

As already mentioned in section 3.2 the power series expansion of the effective tensor $\widetilde{\boldsymbol{L}}$ can be obtained by successive applications of the operators $\boldsymbol{\Gamma}^1$ or $\boldsymbol{H}^1$. Taking into account the specific form of the effective tensor $\widetilde{\boldsymbol{L}}=\widetilde{z}\boldsymbol{I}$ and $\boldsymbol{L}=z_0\boldsymbol{I}$ into (55), (56) and (57) the coefficients d_k of the power series (60) can be related to the characteristic function of the phases and to the operators $\boldsymbol{\Gamma}^1$ or $\boldsymbol{H}^1$ through the following relations:

Reference medium= Matrix (B-scheme): $t=z-1$

$$d_0=1,\ d_k=b_k=(-1)^{k-1}\langle\chi^{(1)}(\boldsymbol{\Gamma}_{\chi^{(1)}})^{k-1}\rangle.\boldsymbol{e}_1 \quad k\geq 1. \tag{63}$$



(a) (b) (c)

Figure 2: Convergence of the theoretical power series (60). Error (61) as a function of the order of truncation. Expansion with respect to (a) $t = z - 1$, (b) $t = \dfrac{z-1}{z+1}$, and (c) $t = \dfrac{\sqrt{z}-1}{\sqrt{z}+1}$. Solid lines: $t > 0$, dashed lines $t < 0$.

Reference medium= Arithmetic mean (MS-scheme): $t = \dfrac{z-1}{z+1}$

$$d_0 = 1,\ d_k - d_{k-1} = b_k,\ b_k = (-1)^{k-1}\langle \chi'(\boldsymbol{\Gamma}_{\chi'})^{k-1}\rangle.\boldsymbol{e}_1, \quad k \geq 1. \tag{64}$$

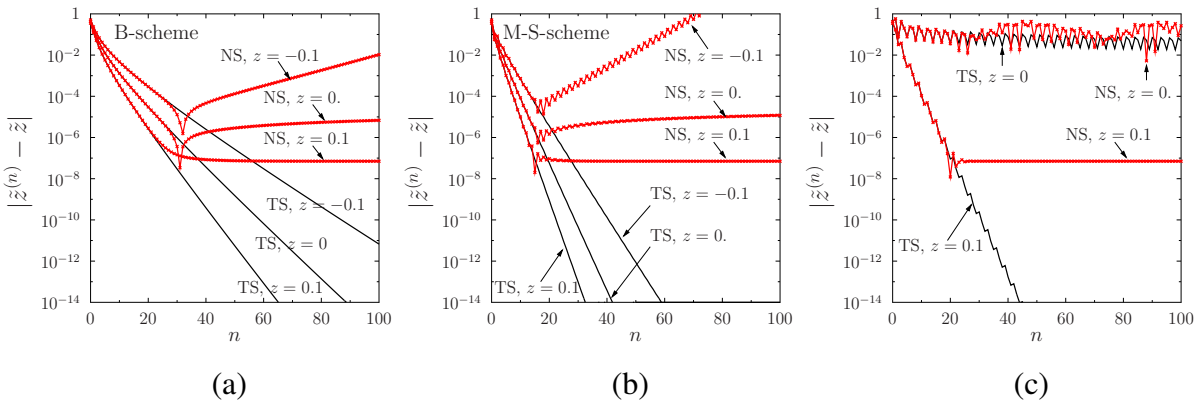
Reference medium= Geometric mean (EM-scheme): $t = \dfrac{\sqrt{z}-1}{\sqrt{z}+1}$

$$d_0 = 1,\ d_1 = 2(\langle \chi' \rangle + 1),\ d_k - d_{k-1} = b_k + b_{k-1},\ b_k = 2(-1)^{k-1}\langle \chi'(\boldsymbol{H}_{\chi'})^{k-1}\rangle.\boldsymbol{e}_1, \quad k \geq 2, \tag{65}$$

where χ', $\boldsymbol{\Gamma}_{\chi'}$ and $\boldsymbol{H}_{\chi'}$ are defined in (58). The $\boldsymbol{\Gamma}_{\chi^{(1)}}$, $\boldsymbol{\Gamma}_{\chi'}$ and $\boldsymbol{H}_{\chi'}$ and the operator $\boldsymbol{\Gamma}^0$ from which they are formed, are more easily handled in Fourier space where they are explicitly known. In conductivity and elasticity the operator $\boldsymbol{\Gamma}^0$ can be expressed in Fourier space as

$$\hat{\boldsymbol{\Gamma}}^0(\boldsymbol{\xi}) = \boldsymbol{\xi} \otimes \left(\boldsymbol{\xi}.\boldsymbol{L}^0.\boldsymbol{\xi}\right)^{-1} \otimes \boldsymbol{\xi}. \tag{66}$$

In the above expression of the Green's operator, $\boldsymbol{\xi}$ is the actual frequency in Fourier space. Therefore, in practice, the coefficients of the series expansion (60) can be computed by successive applications of the same operator $\boldsymbol{\Gamma}_{\chi^{(1)}}$, $\boldsymbol{\Gamma}_{\chi'}$ or $\boldsymbol{H}_{\chi'}$ depending on the reference medium. Each application of this operator involves a multiplication by a characteristic function followed by a Fourier transform. The latter operation is performed on a discretized image of the microstructure by Fast Fourier Transform. Therefore computing d_k will require, in all three cases, $k-1$ applications of the FFT to a nonsmooth field (the nonsmoothness stemming from the characteristic function). The "theoretical power series" (TS) (60) (where the exact coefficients b_k and d_k are given for each scheme in appendix Appendix B) and the "numerical power series" (NS) (63, 64, 65) are compared in figure 3 for the three schemes of section 2.4. The following observations can be made:



(a) (b) (c)

Figure 3: Obnosov problem. Comparison between the convergence of the theoretical series (TS) and the numerical series (NS) for different contrasts. (a) Reference medium: matrix (B-scheme). (b) Reference medium: arithmetic mean (MS-scheme). (c) Reference medium: geometric mean (EM-scheme). Discretization: 512×512 pixels.

1. For all three contrast variables, the sums of the first 10 to 20 terms of the theoretical and numerical series coincide.
2. However, after a certain number of iterations, the deviation of the computed effective conductivity from the exact result decreases only slowly (for positive contrast) or saturates, or even increases (for negative contrast) when more and more terms are added.
3. Figure 4 shows that the point of deviation between the two series depends on the discretization of the image. Three discretization for the image, 128^2, 512^2 and 2048^2 pixels respectively, have been considered. The finer the discretization, the later the deviation and the better the prediction of effective conductivity. In the limit of an infinitely fine discretization, it is expected that both series will coincide at any order.

4.3. Discussion

Even with the discretized equations, the truncated series expansions are rational functions of the moduli and therefore in the domain of convergence they converge to an analytic function ([31], theorem 10.28). The analytic continuation of this function will have singularities at the radius of convergence. Such singularities might include, for example, a pole with very small residue that is not necessarily located on the negative real axis. These spurious singularities in the discrete approximation dictate the convergence properties of the discrete approximation.

The only difference between the theoretical and the numerical series $\widetilde{z}^{(n)}$ resides in the coefficients d_k. The numerical coefficients, which will be denoted d_k^{NS} to distinguish them from the exact ones d_k, are only an approximation of the exact coefficients, the accuracy of the approximation depending on the refinement of the discretization and on the choice of the $\boldsymbol{\Gamma}^0$ operator (continuous or discrete).

The role of the coefficients d_k in the discrepancy between the theoretical and numerical predictions of the effective conductivity is confirmed by figure 6 and table C.1 where the first

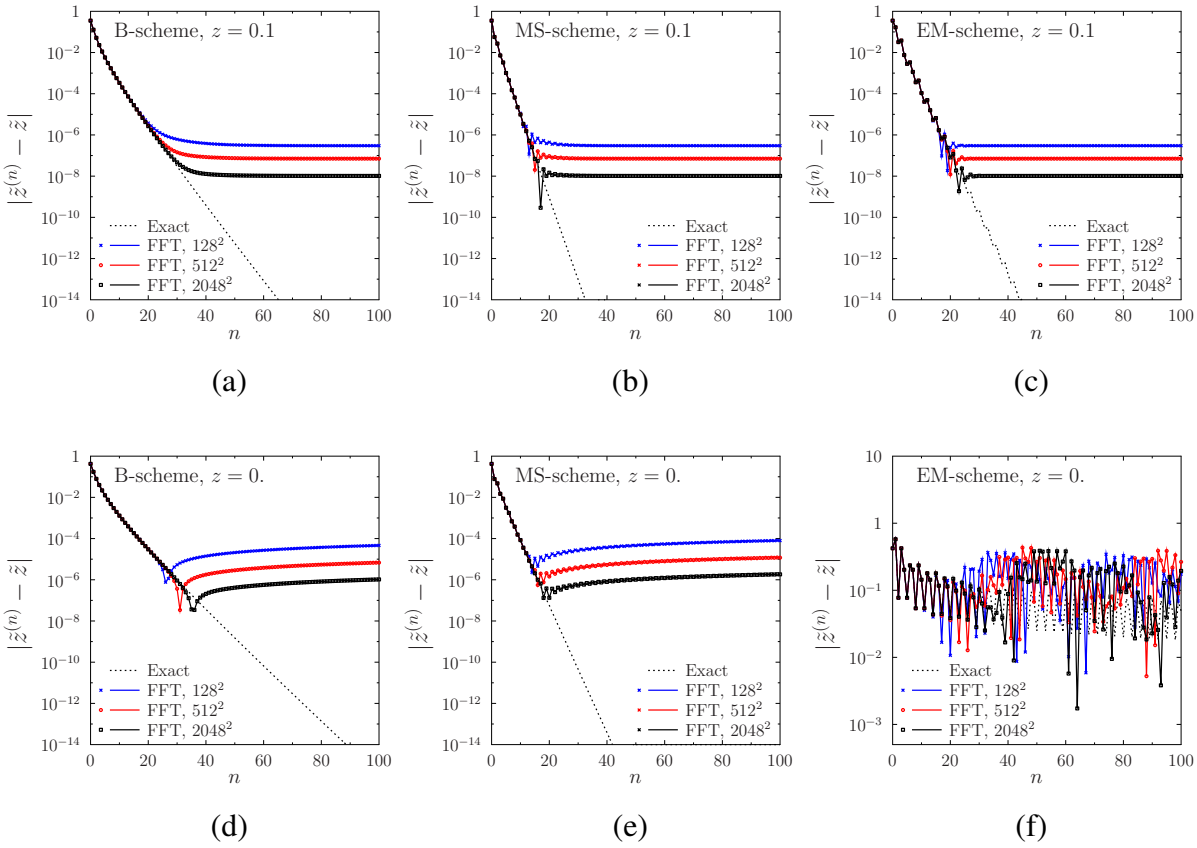


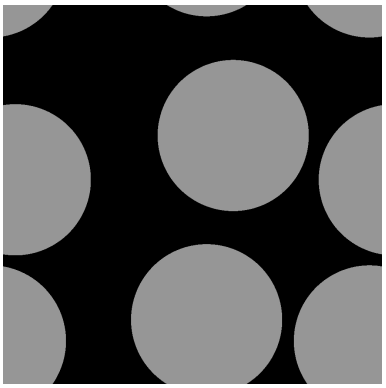
Figure 4: Influence of the discretization on the convergence of the numerical scheme. (a)(b)(c): contrast $z = 0.1$, (d)(e)(f): $z = 0$. (a)(d) Reference medium: matrix (B-scheme). (b)(e) Reference medium: arithmetic mean (MS-scheme). (c)(f) Reference medium: geometric mean (EM-scheme).

d_k coefficients in the theoretical and numerical series are compared. It is seen by comparing figure 3 and figure 6 that the deviation in the effective properties observed in figure 3 coincides with a similar deviation in the coefficients. Table C.1 shows the variation of the individual coefficients with the order in the series (or with the number of iterations). It clearly appears that beyond a certain order in the expansion (or beyond a certain number of iterations), hereafter denoted by $\mathcal{K}$, there is a strong discrepancy between the theoretical and numerical d_k's. The first terms (up to $\mathcal{K}$) in the numerical series can be considered as exact, whereas the terms of higher order differ from the exact values. Therefore the series expansion for $\widetilde{z}$ may be written as

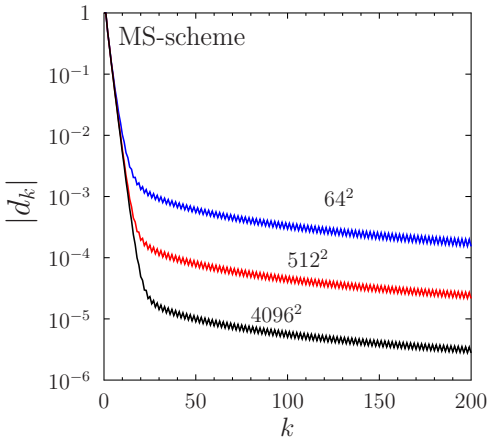
$$\widetilde{z}_n^{NS} \simeq \sum_{k=0}^{\mathcal{K}} d_k t^k + \sum_{k=\mathcal{K}+1}^{n} d_k^{NS} t^k \ . \tag{67}$$

The coefficients d_k^{NS} terms ($k > \mathcal{K}$) decrease slowly (at best, or even remain constant) when k increases but remain much larger than the exact d_k terms. This discrepancy is indeed not surprising, since if all coefficients were accurately computed at any order, the sum of the series would be the *exact* effective property. However, the use of any computational method means that the solution of the continuous problem, which belongs in general to a vector space of infinite dimension, is approximated in a space of finite dimension. Therefore an intrinsic error due to the spatial discretization exists, resulting in a gap between the exact solution and its finite dimensional approximation, regardless of the computational method. And since the effective property is computed here by means of its variational characterization (through the associated energy), this gap is always non zero regardless of the computational method (except when the solution belongs to the finite dimensional space of approximation, which is rarely the case). Therefore the sum of the approximate series differs from the exact solution and the coefficients of the numerical series have to deviate, sooner or later, from their theoretical value. Clearly enough, when the spatial discretization of the microstructure is refined, the approximate solution is a better approximation of the exact one, the effective energy is lower and the gap reduces. Therefore the deviation from the theoretical values is delayed by refining the discretization (but occurs eventually). This is confirmed in figure 6.

Our interpretation of this deviation from exactness goes as follows. In the course of the iterations, the successive approximations (partial sums of the series) converge towards the "best" approximation of the solution in the finite dimensional space associated with a given discretization. When the iterates are close to this best approximation, the effective energy cannot be further lowered and the effective properties reach a plateau as can be seen in figure 4 (a)-(c) for instance. Iterating further neither improves the plateau, which is an energetic barrier, nor the local solution in this finite dimensional space. In conclusion, all iterations beyond $\mathcal{K}$ introduced in (67) do not improve convergence and can even deteriorate the local field (as observed sometimes). It is therefore proposed in section 4.4 to stop iterating at $\mathcal{K}$, provided $\mathcal{K}$ can be properly defined. The effect of the discretization on the loss of accuracy along the iterations is not restricted to the specific microstructure investigated here, it is indeed a generic observation. To illustrate this point, the MS scheme has been applied to the 2D microstructure of figure 5(a) and the d_k terms of the series (56) have been computed numerically for different resolutions, from 64×64 pixels to 4096×4096 pixels. As for the Obnosov microstructure, the d_k terms begin to



(a)



(b)

Figure 5: (a) Microstructure: 4 circular inclusions, volume fraction 50%. (b) Coefficients d_k of the series (56) at different resolutions. MS scheme.

deviate from each other after a certain number of iterations which depends on the discretization. The point of deviation between a coarse and a fine discretization is probably the point where the iterations for the coarser discretization should be stopped, since all iterations beyond that point are affected by a significant discretization error.

The deviation from exactness of the coefficients of the series expansions is essential to understand why iterative methods fail to converge when the contrast between the constituents is very large or negative, although theoretically they should converge for certain microstructures (the Obnosov microstructure is one of them) even when one of the phases has a nonpositive modulus in a certain range. In such situations of extreme contrast, the absolute value of the parameter t measuring the contrast is close to 1 or even larger than 1. Therefore the convergence of the series $d_k t^k$ relies crucially on the accuracy of the d_k's. But since these coefficients must deviate from their exact value beyond a given order $\mathcal{K}$ it is not surprising that the series does not converge towards its exact sum, or does not converge at all. Conversely, when the contrast parameter t is strictly less than 1 (in absolute value), the deviation from exactness of the coefficients at high order does not really influence the convergence of the series $d_k t^k$ which is governed by t^k. This is the situation for moderate contrast between the constituents.

The threshold $\mathcal{K}$ depends on the computational method used to perform the iterations, or equivalently to compute the coefficients of the series expansions with the help of the relations of section 3.2. The Fast Fourier Transform, in discrete form, is used in the present study. Its role in the deviation of the coefficients remains to quantified. However it is clear, qualitatively, that the deviation will occur in general with *any* other computational method operating within a finite dimensional space.

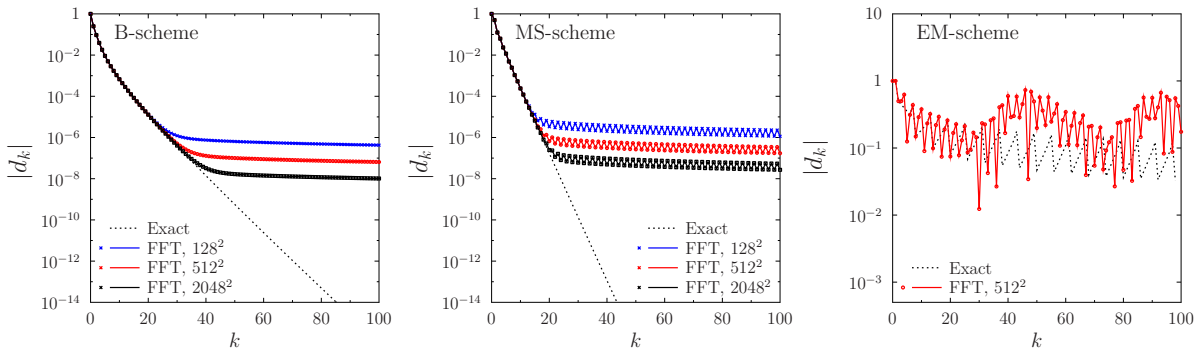


Figure 6: Comparison between the coefficients d_k of the power series. a) Reference medium: matrix (B scheme). (b) Reference medium: arithmetic mean (MS-scheme). (c) Reference medium: geometric mean (EM-scheme).

4.4. Stopping criteria

The iterative procedure $(8)_1$ requires the choice of a criterion to stop the iterations, or alternatively to truncate the series $(8)_2$. Two error indicators which are popular in the literature are

- equilibrium residual [6]:

$$\delta_1^{(k)} = \frac{\left|\left|\mathrm{div}(\boldsymbol{\sigma}^{(k)})\right|\right|_{L^2}}{||\boldsymbol{\Sigma}||} = \frac{\langle\left|\mathrm{div}(\boldsymbol{\sigma}^{(k)}(\boldsymbol{x}))\right|^2\rangle^{1/2}}{||\boldsymbol{\Sigma}||}, \tag{68}$$

  where $\boldsymbol{\Sigma}$ is a stress which is natural in the problem at hand and used in (68) to arrive at a non-dimensional criterion. The value commonly used for $\boldsymbol{\Sigma}$ is the average of the stress field: $\boldsymbol{\Sigma} =< \boldsymbol{\sigma} >$, but in the present study, for simplicity, it has been chosen with unit norm, i.e. $||\boldsymbol{\Sigma}|| = 1$.

- Difference between two successive iterates [32]:

$$\delta_2^{(k)} = \left|\left|\boldsymbol{\varepsilon}^{(k+1)} - \boldsymbol{\varepsilon}^{(k)}\right|\right|_{L^2} = \langle\left|\boldsymbol{\varepsilon}^{(k+1)}(\boldsymbol{x}) - \boldsymbol{\varepsilon}^{(k)}(\boldsymbol{x})\right|^2\rangle^{1/2}. \tag{69}$$

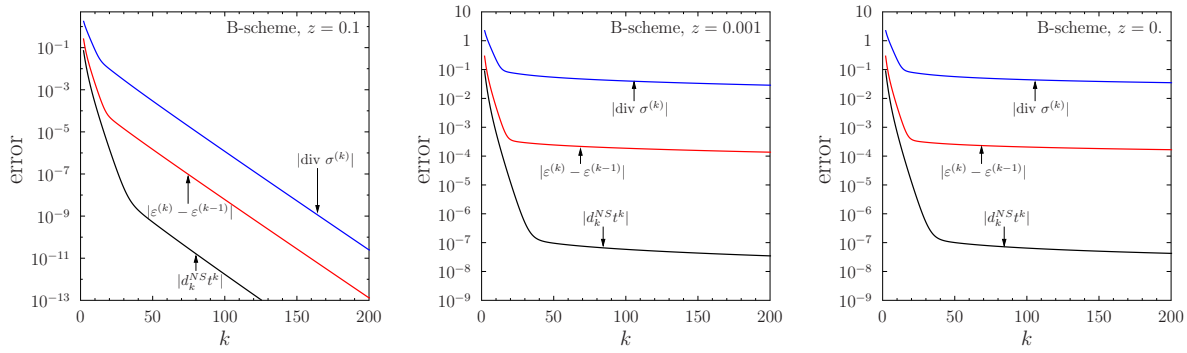
Note that

$$\boldsymbol{\varepsilon}^{(k+1)} - \boldsymbol{\varepsilon}^{(k)} = -\boldsymbol{\Gamma}^0\boldsymbol{\sigma}^{(k)},$$

from which it follows that the second criterion is equivalent to a criterion on the norm of $\boldsymbol{\Gamma}^0\boldsymbol{\sigma}^{(k)}$.

The iterative algorithm is stopped (or equivalently the Neumann series is truncated) as soon as the chosen error indicator is less than a given tolerance δ:

$$\delta_i^{(k)} \leq \delta, \ i = 1 \text{ or } i = 2.$$



(a) (b) (c)

Figure 7: Comparison between the error indicators (68), (69) and $|d_k^{NS} t^k|$. B-scheme (matrix as reference). Three different contrasts: (a) $z = 0.1$, (b) $z = 0.001$ and (c) $z = 0$.

The variations with k of these two error indicators are compared in figure 7 with the k-th term of the numerical power series for the fixed-point scheme $(8)_1$ which can be used as a third error indicator

$$\left|d_k t^k\right| ||\boldsymbol{E}|| \,. \tag{70}$$

The following comments can be made.

1. The most demanding criterion is the one based on equilibrium, *i.e.* the L^2 norm of $\text{div}(\boldsymbol{\sigma})$. For a given error, it is the criterion requiring the larger number of iterations. Then comes the criterion based on the difference of two iterates (which is equivalently a check on equilibrium via the Green's operator $\boldsymbol{\Gamma}^0$). The third criterion (70), based on the difference between two successive partial sums in the Neumann expansion of the effective moduli is a global criterion, by contrast with the two others. It is therefore the less demanding criterion.
2. The three curves for the three criteria are almost parallel. As observed in section 4.3 the knee in the third curve (70) is probably the point where the numerical error on the coefficients d_k^{NS} due to discretization becomes significant and where the iterations should be stopped for this particular microstructure. It is interesting to note that the two other convergence criteria show the same knee. It is therefore expected that, whatever the criterion, all iterations beyond the knee do not improve the accuracy of the fields or of the effective properties. It is also expected that the threshold $\mathcal{K}$ would be of the same order for all three criteria.
3. A possible way to detect the threshold $\mathcal{K}$, for any of these criteria, would be to conduct the simulations with two different discretizations and to stop iterating when the criterion for the two different discretizations deviate from each other.

4.5. Discrete Green's operators

Spurious oscillations are sometimes observed in the local fields when the continuous Green's operator (66) is used. Although certain authors attribute these oscillations to a Gibb's phenomenon, they are more likely due to the fact that the derivative of a *nonsmooth function* (not C^1) is poorly approximated by the inverse Fourier transform of the Fourier transform of the function multiplied by the frequency. This has motivated several authors to introduce discrete Green's operators having a better behaviour with respect to derivation. Instead of using directly the expression of the continuous Green's operator (66)), the partial derivatives are approximated by finite differences and incorporated into the Green's operator. Several variants have been proposed depending on the operator used to approximate the derivation ([20], [21], [33], [34], [22], [35], [23]).

The variations with k of the coefficients d_k are compared in figure 8 for three different choices of the Green's operator: the continuous Green's operator used in Moulinec and Suquet [5, 6], the modified Green's operators proposed by Müller [20] and by Willot and Pellegrini [36].

The expressions used for the modified Green's operators are obtained by simply replacing in relation (66) the angular frequency $\boldsymbol{\xi}$ by $\boldsymbol{\xi}^M$ (as proposed by Müller), or $\boldsymbol{\xi}^W$ (as proposed by Willot and Pellegrini), where

$$\xi_j^M = \frac{N_j}{L_j}\sin\left(\frac{L_j}{N_j}\xi_j\right) \quad (j = 1, 2 \; or \; 3), \quad \xi_j^W = \imath\frac{N_j}{L_j}\exp\left(-\imath\frac{L_j}{N_j}\xi_j\right) \quad (j = 1, 2 \; or \; 3),$$

(where ξ_j is the j-th component of $\boldsymbol{\xi}$, where $\imath = \sqrt{-1}$, and where L_j and N_j are, respectively, the size and the number of pixels of the unit-cell in the j-th direction).

The MS scheme is used for this comparison but the same conclusions hold for the other schemes. The main observations are the following:

1. In the first few iterations the predictions with the three Green's operator are in good agreement with the theoretical values, with slightly better predictions by the continuous Green's operator up to higher order (see the close-up in figure 8(b)). The continuous Green's operator seems to give the best estimation when k is lower than 17, while the two discrete schemes deviate from the theoretical predictions at about $k = 12$.
2. The numerical d_k^{NS} deviate from the theoretical d_k between iterations $k = 15$ and $k = 20$ for all three operators, the deviation occurring earlier for the discrete operators as already noticed. However, beyond that point, the predictions of the continuous Green's operator deviate more significantly from the theoretical coefficients and decrease only slowly, whereas the two modified Green's operators exhibit a more rapid decrease, although remaining much larger than it should be theoretically. As an illustration, in figure 8 at iteration $k = 30$, the coefficient obtained with Willot-Pellegrini modified operator is approximately 18 times as large as the theoretical coefficient, the coefficient of Müller's operator is 32 times as large and the coefficient obtained with the continuous operator is 2772 times as large as the theoretical coefficient. In this sense the coefficients d_k^{NS} for large k are better predicted by the modified operators.

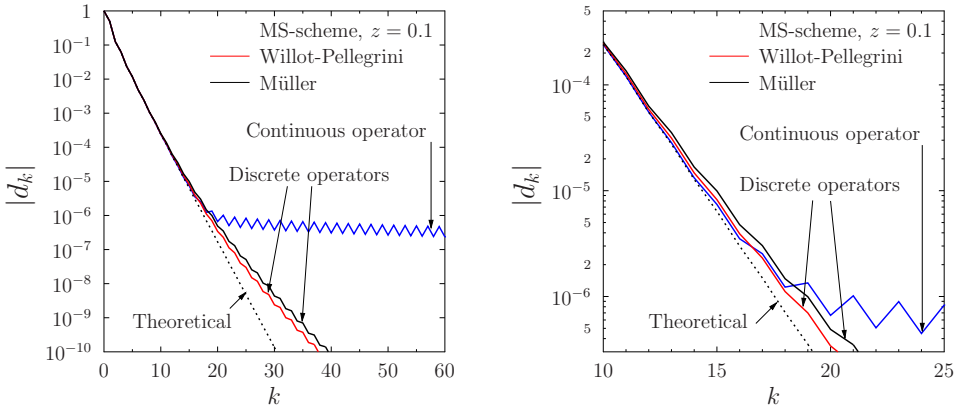
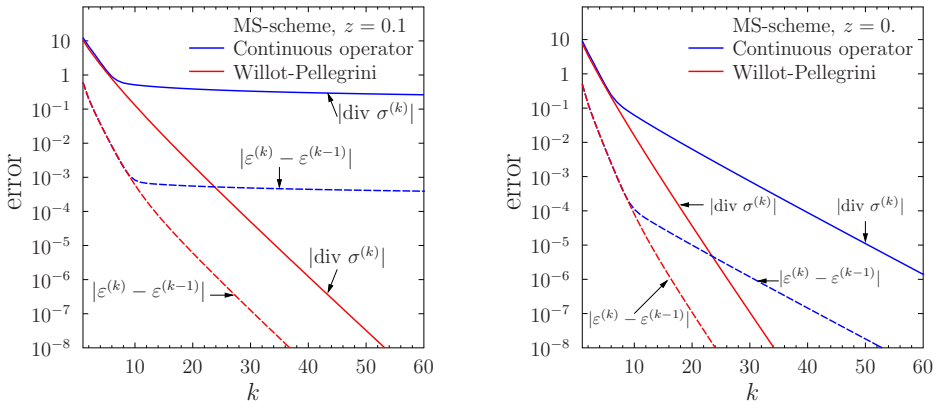


Figure 8: Obnosov's microstructure, discretization 512×512 pixels. MS scheme. Contrast $z = 0.1$. Coefficients d_k of the series (60) with different Green's operators: continuous operator (blue), Müller's operator [20] (black), Willot-Pellegrini operator [22] (red). (a): iterations 1 to 60. (b): close-up on iterations 10 to 25.

3. This more accurate prediction of the d_k beyond the threshold $\mathcal{K}$ goes together with a faster decrease of the two classical error indicators (68, 69) for the discrete operators as shown in figure 9. It is therefore very likely that the stopping criterion based on the discrete Green's operator (all iterates being computed with this discrete operator) is less sensitive to the discretization error than the continuous operator.
4. These observations seem to favour the use of discrete Green's operators. However, the comparison of the effective properties predicted by the different Green's operators in figure 10 goes in the opposite direction. The effective properties computed with the continuous Green's operator are in better agreement with the theoretical value, except in the extreme case $z = 0$. This can be understood by going back to the expansion (67). The first coefficients of the series are better predicted by the continuous operator up to a threshold $\mathcal{K}$ which is larger than for the discrete operators. And although the coefficients for larger k are better predicted by the discrete operators, these coefficients enter the expansion (67) with a factor t^k which is very small for large k when $|t| < 1$. So the coefficients which are crucial for an accurate prediction of the effective property are the coefficients of lower order, which are better predicted with the continuous operator. The gain in accuracy provided by the discrete operators for the coefficients of higher order is lost by the multiplication by t^k when $|t| < 1$. However when t approaches 1, or even become larger than 1, these coefficients become again crucial and the predictions of the discrete operators become more accurate. As a tentative conclusion, one could say that the continuous Green's operator performs better for moderate contrast whereas discrete operators should be used for large contrast.



(a) (b)

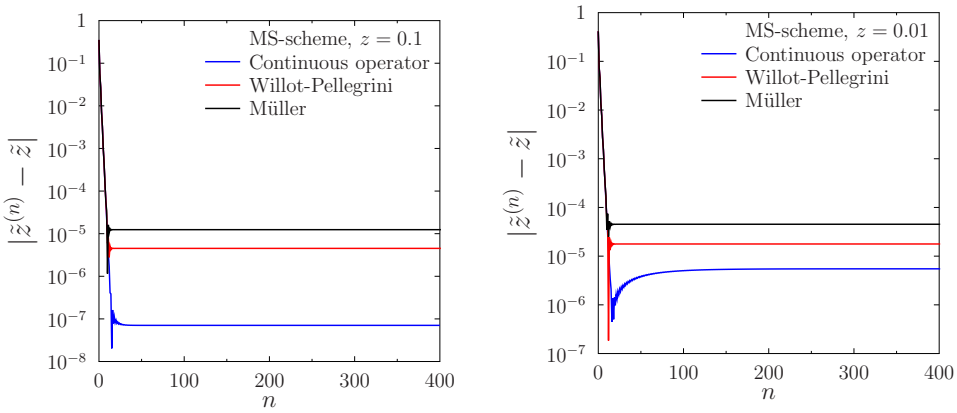
Figure 9: Obnosov's microstructure, discretization 512×512 pixels. MS scheme. Error indicators (68) (solid line), and (69) (dashed line). Comparison between the continuous Green's operator (blue) and Willot-Pellegrini discrete operator [22] (red). (a): Contrast $z = 0.1$. (b) Contrast $z = 0$.

4.6. Theoretical rate of convergence of the three iterative schemes

We end this section by a discussion of the (theoretical) convergence rate of the three iterative schemes (Brown, Moulinec-Suquet, Eyre-Milton) for microstructures which are more general that Obnosov's microstructure. These rates of convergence depend on the microstructure and on the contrast between the phases. The present section shows that none of these iterative methods will converge faster than the others for all possible microstructures and phase contrasts. The choice of the fastest method depends on the material data. Therefore any comparison between these methods (and others), as often reported in the literature, should be taken with care, as the conclusions may depend on the microstructure under consideration and on the contrast between the phases.

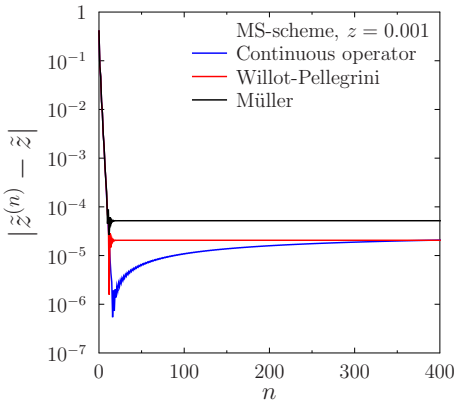
This dependence can be understood on a simple example, that of the effective conductivity of a two-phase composite with two isotropic phases. The singularities of the complex conductivity $\widetilde{z}$ as a function of z are located on the real negative axis. It is further assumed that these singularities lie in an interval $[-\beta, -\frac{1}{\beta}]$ with $\beta \geq 1$. Specifically, we assume that there is a singularity at $z = -1/\beta$ and no singularities for $z < -\beta$ (implying no pole at $z = \infty$). Obnosov's microstructure corresponds to $\beta = 3$. The checkerboard microstructure corresponds to $\beta = +\infty$.

It is to be cautioned that the analysis here is the theoretical rate of convergence in the asymptotic regime where the number of iterates is very large. In particular, suppose the singularity at $z = -1/\beta$ is say a pole with extremely small residue, and there is a pole at $z = -1/\gamma$ with a significant residue, and no singularities in $(-1/\gamma, -1/\beta)$, nor for $z < -\gamma$ (γ is supposed to be larger than 1). Then the initial rate of convergence of the iterates should be dictated by the pole at $z = -1/\gamma$, with the pole at $z = -1/\beta$ only slowing down the convergence rate after

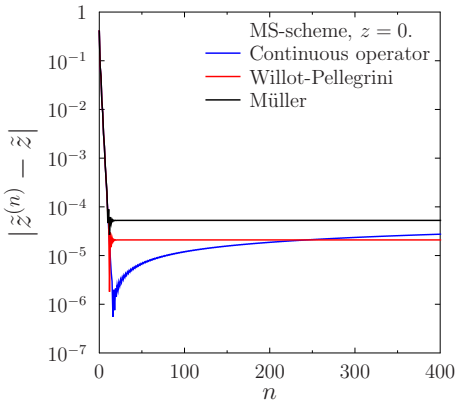


(a)

(b)



(c)



(d)

Figure 10: Obnosov's microstructure, discretization 512×512 pixels. MS scheme. Error on the effective property with different Green's operators: continuous operator (blue), Willot-Pellegrini operator [22] (red), Müller's operator [20] (black). (a): $z = 0.1$. (b): $z = 0.01$. (c): $z = 0.001$. (d): $z = 0$.

many iterates (when the error is of the same magnitude as the effect due to the extremely small residue of the pole at $z = -1/\beta$). One then expects a knee in the convergence rates, similar to our numerical simulations. As in our numerical simulations, such knees can also be due to spurious singularities caused by the discretization, and these could swamp the effect of an extremely small residue at $z = -1/\beta$. We ignore such considerations in the following analysis, thus assuming the singularity at $z = -1/\beta$ to be significant enough to control the convergence rate during the crucial initial stage of iterations. The three schemes (Brown, Moulinec-Suquet,

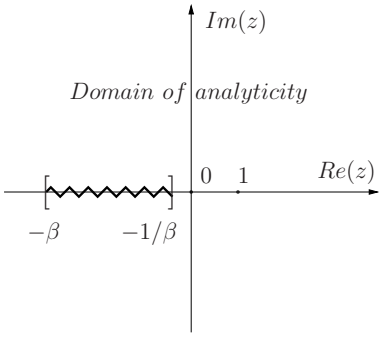


Figure 11: Domain of analyticity of $\widetilde{z}$.

Eyre-Milton) and the corresponding iterative Neumann series are associated with a contrast variable t which reads in the three respective cases

$$\text{Brown: } t = z - 1, \quad \text{Moulinec-Suquet: } t = \frac{z-1}{z+1}, \quad \text{Eyre-Milton: } t = \frac{\sqrt{z}-1}{\sqrt{z}+1}. \tag{71}$$

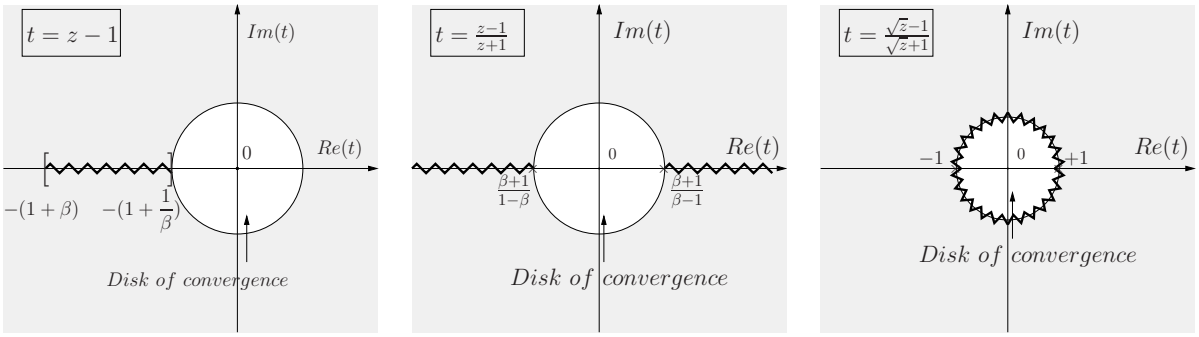
The domains of z for which the iterative schemes converge is the radius of convergence of the power series in the t-plane in the neighborhood of $t = 0$. It is the largest disk contained in the domain of analyticity of $\widetilde{z}$ as a function of t. These domains, schematically represented in figure 12, are respectively

$$\mathcal{D}_B = \{|t| < 1 + \frac{1}{\beta}\}, \quad \mathcal{D}_{MS} = \{|t| < \frac{\beta+1}{\beta-1}\}, \quad \mathcal{D}_{EM} = \{|t| < 1\}. \tag{72}$$

The radius of convergence of the corresponding series in power of t are given

$$\rho_B = 1 + \frac{1}{\beta}, \quad \rho_{MS} = \frac{\beta+1}{\beta-1}, \quad \rho_{EM} = 1. \tag{73}$$

The three power series $\sum_{k=0}^{+\infty} d_k t^k$ (and the associated iterative schemes) converge when $|t| < \rho$, where $t = z - 1$, $\frac{z-1}{z+1}$, and $\frac{\sqrt{z}-1}{\sqrt{z}+1}$, and $\rho = \rho_B$, ρ_{MS}, ρ_{EM} respectively. It is therefore natural to evaluate, for a given contrast (characterized by z) and a given microstructure (associated with β), the rate of convergence of a series expansion through the ratio $r = \rho/|t|$. The larger r, the



(a) (b) (c)

Figure 12: Domain of convergence of the power series in the neighborhood of $t = 0$. (a) $t = z - 1$ (B-scheme. (b): MS-scheme. (c): EM-scheme.

faster the convergence of the series expansion (and of the associated iterative schemes). This ratio reads, for each choice of the variable t:

$$r_B = \left|\frac{\beta+1}{\beta(z-1)}\right|, \quad r_{MS} = \left|\frac{(\beta+1)(z+1)}{(\beta-1)(z-1)}\right|, \quad r_{RM} = \left|\frac{\sqrt{z}+1}{\sqrt{z}-1}\right|. \tag{74}$$

Snapshots of the three ratios are shown in Figure 13, for small, moderate and large values of z and β. The snapshots corresponding to a given scheme are placed horizontally. The three schemes can be compared by looking vertically at the columns, the color table at the bottom being the same for a given column. The larger the ratio, the better the rate of convergence.

The ratios can be compared analytically two-by-two in the domains of z where they are defined. Attention will be limited to real values of z.

Brown vs Moulinec-Suquet. The joint domain of definition of the two power expansions in the (β, z)-plane is

$$z > -\frac{1}{\beta}, \quad \beta \geq 1. \tag{75}$$

Straightforward algebra shows that

$$\frac{r_{MS}}{r_B} = \left|\frac{\beta z + 1}{\beta - 1} + 1\right| > 1, \tag{76}$$

where the last inequality comes from (75). In other words, the MS-scheme always over-performs the B-scheme.

Brown vs Eyre-Milton. The joint domain of definition of the two power expansions in the (β, z)-plane is

$$z \geq 0, \quad \beta \geq 1. \tag{77}$$

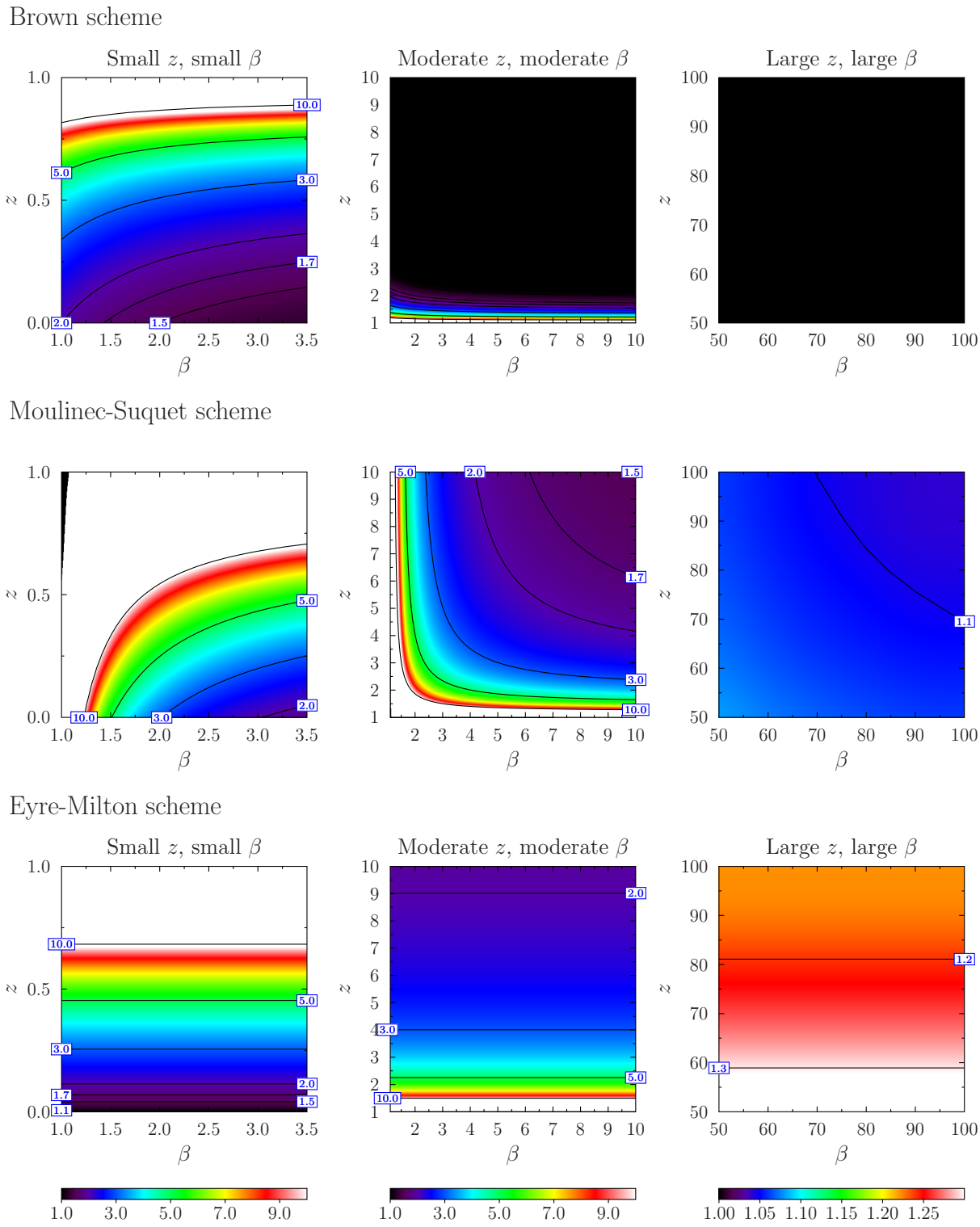
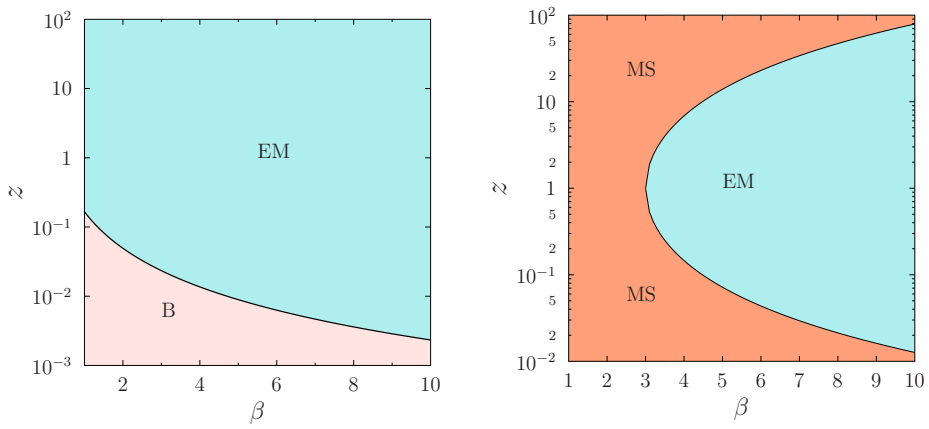


Figure 13: Snapshots of the rate of convergence for the 3 schemes in the plane (β, z). The brighter the color, the faster the rate of convergence.



(a) (b)

Figure 14: Best rate of convergence. (a) EM-scheme vs B-scheme. (b): All 3 schemes.

After straightforward algebra, one obtains that

$$\frac{r_{EM}}{r_B} = (\sqrt{z}+1)^2 \frac{\beta}{\beta+1}, \tag{78}$$

and

$$r_B \geq r_{EM} \quad \text{when } 0 \leq z \leq z_0 = \left(\sqrt{\frac{\beta+1}{\beta}} - 1\right)^2, \quad r_{EM} \geq r_B \quad \text{when } z \geq z_0. \tag{79}$$

In other words, for small enough z (less than z_0) the Brown scheme over-performs the Eyre-Milton scheme, but the Eyre-Milton scheme is faster when $z \geq z_0$. For the Obnosov's microstructure $\beta = 3$, $z_0 \simeq 0.23932$.

Eyre-Milton vs Moulinec-Suquet. The joint domain of definition of the two power expansions in the (β, z)-plane is

$$z \geq 0, \quad \beta \geq 1. \tag{80}$$

Straightforward algebra shows that

$$\frac{r_{EM}}{r_{MS}} = \frac{1 + \frac{2\sqrt{z}}{z+1}}{1 + \frac{2}{\beta-1}}, \tag{81}$$

Two different cases are found

1. When $1 \leq \beta \leq 3$, then

$$\frac{2}{\beta-1} \geq 1 \geq \frac{2\sqrt{z}}{z+1} \quad \Rightarrow \quad \frac{r_{EM}}{r_{MS}} \leq 1,$$

and the MS-scheme is faster than the EM-scheme, regardless of z. The only case where $r_{EM}/r_{MS} = 1$ is when $\beta = 3$ and $z = 1$ (homogeneous material).

2. When $\beta > 3$ then

$$r_{EM} \geq r_{MS} \quad \text{when } z_1 \leq z \leq z_2, \quad r_{MS} \geq r_{EM} \quad \text{when } 0 \leq z \leq z_1 \text{ or } z \geq z_2,$$

$$\text{with } z_1 = \frac{1}{4}\left[\beta - 1 - \sqrt{(\beta-1)^2 - 4}\right]^2, \quad z_2 = \frac{1}{4}\left[\beta - 1 + \sqrt{(\beta-1)^2 - 4}\right]^2. \quad (82)$$

The scheme offering the best ratio is therefore dependent on the problem at hand. It depends on the microstructure through β and on the contrast between the phases through z. A map of the comparison between the 3 schemes is given in figure 14. When β is large, then $z_1 \sim 1/(\beta-1)^2$ tends to 0, and $z_2 \sim (\beta-1)^2$ tends to $+\infty$. In particular when $\beta = +\infty$ (checkerboard for instance), the EM-scheme has a better ratio of convergence than the MS-scheme. This confirms the column on the right of figure 13.

Remark 4: As we have just seen, the relative performance of the different methods depends on the conductivity ratio z and on the interval on the negative real z-axis outside of which there are no singularities in the conductivity function (for the Obnosov microstructure this is the interval $[-3, -1/3]$). If this interval is known the new accelerated FFT algorithm described in Milton [37] chap. 8, should be superior to all three methods. However it is to be noted that calculating, or estimating, this interval for an arbitrary geometry is a highly nontrivial problem. Some progress was made in Bruno [29].

5. Conclusion

The starting point of this study is the classical observation that the convergence of certain iterative methods used to invert the Lippmann-Schwinger operator in heterogeneous media, is directly related to the convergence of series expansions of the effective properties in powers of a contrast variable. Three different such Neumann series are investigated in details, first from a theoretical perspective and then numerically. A specific example (Obnosov's microstructure) where effective properties are available in closed form serves for the comparison.

A first result of the present study is that the range of convergence of the series given in the literature with no information on the microstructure (recalled here in section 2) can be extended when additional information about the microstructure is available (section 3). In particular, when the location of the singularities of the effective moduli in the complex plane of phase contrast is known, the theoretical prediction of the range of contrast for which the iterative methods converge and the prediction of the rate of convergence of these methods can be improved significantly. Unfortunately these theoretical improvements are not always confirmed by the numerics.

Second, in an attempt to understand this discrepancy between the theory and the numerics, it is noted that the approximation of the field solutions in a finite dimensional space introduces a systematic error on the effective properties (because of their variational character). A direct consequence of this error is that beyond a certain order $\mathcal{K}$ the numerical coefficients of the series differ from their exact value. This is confirmed by the observation that the deviation between the theory and the numerics occurs later when the discretization is refined. The theoretical results for convergence are for a continuous problem, whereas the numerical results are for a discrete,

finite-dimensional approximation of it. The error on the coefficients of high order in the series do not influence much the convergence of the series at moderate contrast, but has a dramatic influence at high contrast. This explains why Neumann series fail to converge numerically for certain values of the contrast for which they should converge theoretically.

Thirdly, The rate of convergence of the three iterative schemes are compared when the contrast z between the phases and a parameter β related to the location of the singularities of the effective moduli are varied. It is found that none of the three schemes over-performs the others for all values of z and β. The fastest scheme depends on the microstructure and on the phase contrast.

Finally, when discrete Green's operators are used, the detection of the moment where the iterations should be stopped seems to be more accurate. This does not always result in an improvement of the effective properties but it is likely that the local fields are better captured by avoiding superfluous iterations which may only add noise to the fields.

[1] Kröner E.. *Statistical Continuum Mechanics.* CISM Lecture Notes, vol. **92**: . Wien: Springer-Verlag; 1972.

[2] Willis J.R.. Variational and related methods for the overall properties of composites. In: Yih C.S., ed. *Advances in Applied Mechanics 21*, :1-78Academic Press; 1981; New-York.

[3] Milton G.W.. *The Theory of Composites*. Cambridge: Cambridge University Press; 2002.

[4] Brown W.F.. Solid Mixture Permittivities. *J. Chem. Phys.*. 1955;**23**:1514-1517. doi: 10.1063/1.1742339.

[5] Moulinec H., Suquet P.. A fast numerical method for computing the linear and nonlinear properties of composites. *C. R. Acad. Sc. Paris II.* 1994;**318**:1417-1423.

[6] Moulinec H., Suquet P.. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Comp. Meth. Appl. Mech. Engng.*. 1998;**157**:69-94.

[7] Khatchaturyan A.G.. *Theory of Structural Transformations in Solids*. New York: John Wiley and Sons; 1983.

[8] Mura T.. *Micromechanics of Defects in Solids*. Dordrecht: Martinus Nijhoff Publishers; 1987.

[9] Nemat-Nasser S., Iwakuma I., Hejazi M.. On composites with periodic microstructure. *Mech. Materials.* 1982;:239-267.

[10] Michel JC., Moulinec H., Suquet P.. A computational method for linear and nonlinear composites with arbitrary phase contrast. *Int. J. Numer. Meth. Engng.* 2001;**52**:139-160.

[11] Eyre D.J., Milton G.W.. A fast numerical scheme for computing the response of composites using grid refinement. *European Physical Journal. Applied Physics.* 1999;**6**:41-47.

[12] Milton G. W., Golden K. M.. Representations for the conductivity functions of multicomponent composites. *Commun. Pure Appl. Math.*. 1990;**43**:647–671.

[13] Clark K. E., Milton G. W.. Modeling the effective conductivity function of an arbitrary two-dimensional polycrystal using sequential laminates. *Proceedings of the Royal Society of Edinburgh.* 1994;124A(4):757–783.

[14] Grabovsky Y.: *Composite Materials: Mathematical Theory and Exact Relations*. Bristol, UK: IOP Publishing; 2017.

[15] Monchiet V., Bonnet G.: A polarization-based FFT iterative scheme for computing the effective properties of elastic composites with arbitrary contrast. *Int. J. Numer. Meth. Engng.* 2012;89:1419-1436. doi:10.1002/nme.3295.

[16] To Q.D., Nguyen M.T., Bonnet G., Monchiet V., TO V.T.: Overall elastic properties of composites from optimal strong contrast expansion. *Int. J. Solids Structures* 2017; 120:245-256. doi:10.1016/j.ijsolstr.2017.05.006.

[17] Zeman J., Vondřejc J., Novák J., Marek I.: Accelerating a FFT-based solver for numerical homogenization of periodic media by conjugate gradients. *Journal of Computational Physics.* 2010;229(21):8065–8071.

[18] Brisard S., Dormieux L.. Combining Galerkin approximation techniques with the principle of Hashin and Shtrikman to derive a new FFT-based numerical method for the homogenization of composites. *Comput. Methods Appl. Mech. Engrg.*. 2012;**217-220**:197 - 212. doi:10.1016/j.cma.2012.01.003.

[19] Mishra N., Vondřejc J., Zeman J.. A comparative study on low-memory iterative solvers for FFT-based homogenization of periodic media. *Journal of Computational Physics.* 2016;321(Supplement C):151–168.

[20] Müller W.H.. Mathematical versus experimental stress analysis of inhomogeneities in solids. *J. Physique IV.* 1996;**6**:C1.139-C1-148.

[21] Brown C.M, Dreyer W., Müller W.H.. Discrete Fourier transforms and their application to stress-strain problems in composite mechanics: A convergence study. *Proc. R. Soc. London A.* 2002;**458**:1-21.

[22] Willot F., Abdallah B., Pellegrini Y.P.. Fourier-based schemes with modified Green operator for computing the electrical response of heterogeneous media with accurate local fields. *Comp. Meth. Appl. Mech. Engng.*. 2014;98(7):518-533.

[23] Schneider M., Ospald F., Kabel M.. Computational homogenization of elasticity on a staggered grid. *Int. J. Numer. Meth. Engng.* 2016;105(9):693–720.

[24] Obnosov Y.. Periodic heterogeneous structures: new explicit solutions and effective characteristics of refraction of an imposed field. *SIAM J. Appl. Math.*. 1999;59(4):1267–1287.

[25] Hoang-Duc H., Bonnet G.. A series solution for the effective properties of incompressible viscoelastic media. *Int. J. Solids Structures.* 2014;**51**:381-391.

[26] Bergman D.. The dielectric constant of a composite material - A problem in classical Pphysics. *Physics Reports.* 1978;**43**:377-407.

[27] Milton G. W.. Bounds on the complex permittivity of a two-component composite material. *Journal of Applied Physics.* 1981;**52**(8):5286–5293.

[28] Golden K., Papanicolaou G.. Bounds for Effective Parameters of Heterogeneous Media by Analytic Continuation. *Commun. Math. Phys.*. 1983;**90**:473-491.

[29] Bruno O.P.. The effective conductivity of strongly heterogeneous composites. *Proc. R. Soc. London A.* 1991;**433**:353–381.

[30] Hetherington J. H., Thorpe M. F.. The conductivity of a sheet containing inclusions with sharp corners. *Proc Royal Soc. London A.* 1992;**438**:591–604.

[31] Rudin W.. *Real and complex analysis*. New York: McGraw-Hill Inc.; 1987.

[32] Kabel M., Böhlke T., Schneider M.. Efficient fixed point and Newton-Krylov solvers for FFT-based homogenization of elasticity at large deformations. *Comput. Mech.*. 2014;54:1497-1514. doi:10.1006/s00466-014-1071-8.

[33] Neumann S., Herrmann K.P., Müller W.H.. Stress/strain computation in heterogeneous bodies with discrete Fourier transforms - Different approaches. *Computational Materials Science.* 2002;**25**:151-158.

[34] Brisard S., Dormieux L.. FFT-based methods for the mechanics of composites: A general variational framework. *Computational Materials Science.* 2010;**49**(3):663 - 671.

[35] Willot F.. Fourier-based schemes for computing the mechanical response of composites with accurate local fields. *Comptes Rendus Mécanique.* 2015;343(3):232–245.

[36] Willot F., Pellegrini Y.-P.. Fast Fourier Transform computations and build-up of plastic deformation in 2D, elastic-perfectly plastic, pixelwise disordered porous media. In: D. Jeulin S. Forest, ed. *Continuum Models and Discrete Systems CMDS 11*, :443–449Mines ParisTech Les Presses; 2008; Paris.

[37] Milton (editor) G. W.. *Extending the Theory of Composites to Other Areas of Science*. P.O. Box 581077, Salt Lake City, UT 85148, USA: Milton–Patton Publishers; 2016.

Appendix A. The Green's operator $\boldsymbol{\Gamma^0}$ and useful properties of the key operator $\boldsymbol{\Gamma^0 L^0}$

The space of periodic L^2 tensor fields on V is denoted by $\mathscr{H} = \mathcal{L}^2_\sharp(V)$. It is a Hilbert space when equipped with the scalar product

$$\langle \boldsymbol{e}, \tilde{\boldsymbol{e}} \rangle = \frac{1}{|V|} \int_V \boldsymbol{e}(\boldsymbol{x}).\tilde{\boldsymbol{e}}(\boldsymbol{x}) d\boldsymbol{x},$$

with

$$\boldsymbol{e}.\tilde{\boldsymbol{e}} = \sum_{i=1}^{N} e_i \tilde{e}_i \text{ for 1st order tensors, or } \quad \boldsymbol{e}.\tilde{\boldsymbol{e}} = \sum_{i,j=1}^{N} e_{ij} \tilde{e}_{ij} \text{ for 2nd order tensors.}$$

Let $\boldsymbol{L}^0$ denote a uniform (no spatial dependence) fourth-order tensor, positive definite with major and minor symmetries. For a given field $\boldsymbol{\tau}$ in $\mathscr{H}$, consider the following Eshelby problem:

$$\textit{Find } \boldsymbol{\sigma} \in \mathscr{S} \textit{ and } \boldsymbol{\varepsilon}^* \in \mathscr{E}_0 \textit{ such that: } \boldsymbol{\sigma} = \boldsymbol{L}^0 : \boldsymbol{\varepsilon}^* - \boldsymbol{\tau}, \tag{A.1}$$

where

$$\left.\begin{aligned} \mathscr{E}_0 &= \Big\{ \boldsymbol{\varepsilon}^* \in \mathscr{H} \text{ such that: } \exists\, \boldsymbol{u}^* \in \boldsymbol{H}^1_\sharp(V)\,,\ \boldsymbol{\varepsilon}^* = \tfrac{1}{2}(\boldsymbol{\nabla u}^* + \boldsymbol{\nabla u}^{*\top}) \Big\}, \\ \mathscr{S} &= \Big\{ \boldsymbol{\sigma} \in \mathscr{H} \text{ such that: } \operatorname{div} \boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{0} \text{ in } V,\ \boldsymbol{\sigma}.\boldsymbol{n} \text{ anti-periodic on } \partial V \Big\}. \end{aligned}\right\} \tag{A.2}$$

The problem (A.1) has a unique solution $\boldsymbol{\sigma} \in \mathscr{S}$ and $\boldsymbol{\varepsilon} \in \mathscr{E}_0$ and the periodic strain Green's operator $\boldsymbol{\Gamma}^0$ of the reference medium with stiffness $\boldsymbol{L}^0$ is defined as

$$\boldsymbol{\Gamma}^0 :\ \boldsymbol{\tau} \in \mathscr{H} \longmapsto \boldsymbol{\Gamma}^0 \boldsymbol{\tau} = \boldsymbol{\varepsilon}^* \text{ solution of (A.1).}$$

$\mathscr{H}$ can be alternatively equipped with an energetic scalar product

$$\langle\langle \boldsymbol{e}, \tilde{\boldsymbol{e}} \rangle\rangle = \langle \boldsymbol{e} : \boldsymbol{L}^0 : \tilde{\boldsymbol{e}} \rangle, \tag{A.3}$$

The operator $\boldsymbol{\Gamma}^0 \boldsymbol{L}^0$ has the following properties on $\mathscr{H}$ endowed with the energetic scalar product (A.3).

1. $\boldsymbol{\Gamma}^0 \boldsymbol{L}^0$ is a self-adjoint operator from $\mathscr{H}$ endowed with the scalar product (A.3) into itself:

$$\langle\langle \boldsymbol{e}_1, \boldsymbol{\Gamma}^0 \boldsymbol{L}^0 \boldsymbol{e}_2 \rangle\rangle = \langle\langle \boldsymbol{\Gamma}^0 \boldsymbol{L}^0 \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle\rangle. \tag{A.4}$$

2. $\boldsymbol{\Gamma}^0 \boldsymbol{L}^0$ is a positive operator from $\mathscr{H}$ endowed with the scalar product (A.3) into itself:

$$\langle\langle \boldsymbol{e}, \boldsymbol{\Gamma}^0 \boldsymbol{L}^0 \boldsymbol{e} \rangle\rangle \geq 0, \quad \forall\, \boldsymbol{e} \in \mathscr{H}. \tag{A.5}$$

3. $\boldsymbol{\Gamma}^0 \boldsymbol{L}^0$ is the orthogonal projector from $\mathscr{H}$ onto $\mathscr{E}_0$ for the energetic scalar product (A.3)

$$\forall \boldsymbol{e} \in \mathscr{H} :\ \boldsymbol{\Gamma}^0(\boldsymbol{L}^0 \boldsymbol{e}) \in\ \mathscr{E}_0,\ \boldsymbol{\Gamma}^0 \boldsymbol{L}^0 \boldsymbol{\varepsilon}^* = \boldsymbol{\varepsilon}^* \quad \forall \boldsymbol{\varepsilon}^* \in \mathscr{E}_0,\ \langle\langle \boldsymbol{e} - \boldsymbol{\Gamma}^0 \boldsymbol{L}^0 \boldsymbol{e}, \boldsymbol{\Gamma}^0 \boldsymbol{L}^0 \boldsymbol{e} \rangle\rangle = 0. \tag{A.6}$$

As such it is a contraction on $\mathscr{H}$ and its operator norm is exactly 1 when $\mathscr{H}$ is endowed with the energetic scalar product (A.3).

Appendix B. Series expansions of Obnosov's analytical expression

The Obnosov relation (59)

$$\widetilde{\boldsymbol{L}}(z) = \widetilde{z}\boldsymbol{I}, \quad \widetilde{z} = \sqrt{\frac{1+3z}{3+z}},$$

can be expanded in power series with respect to the three variables of interest $t = z-1$, $t = Z = \dfrac{z-1}{z+1}$ or $t = w = \dfrac{\sqrt{z}-1}{\sqrt{z}+1}$.

Two types of expansions are of interest:

$$\frac{\widetilde{z}}{z_0} = \sum_{k=0}^{\infty} b_k t^k, \quad \widetilde{z} = \sum_{k=0}^{\infty} d_k t^k. \tag{B.1}$$

The first quantity will serve for comparison with the general expansion of $\boldsymbol{L}^{0^{-1}}\widetilde{\boldsymbol{L}}$, whereas the second expansion serves to assess the accuracy of the numerical simulations.

A useful trick. First, because of the square root in (59), it is noticed that an expansion of $(\widetilde{z}/z_0)^2$ or $\widetilde{z}^2$ will be easier to derive than that of $\widetilde{z}/z_0$ or $\widetilde{z}$. Writing

$$\left(\frac{\widetilde{z}}{z_0}\right)^2 = \sum_{n=0}^{\infty} a_n t^n, \quad \frac{\widetilde{z}}{z_0} = \sum_{k=0}^{\infty} b_k t^k, \quad \widetilde{z}^2 = \sum_{n=0}^{\infty} c_n t^n, \quad \widetilde{z} = \sum_{k=0}^{\infty} d_k t^k, \tag{B.2}$$

the coefficients $(a_n)|_{n\in\mathbb{N}}$ and $(b_k)|_{k\in\mathbb{N}}$ are related by:

$$a_n = \sum_{k=0}^{n} b_k b_{n-k}, \text{ i.e. } b_0^2 = a_0,\ 2b_0b_1 = a_1,\ 2b_0b_n + \sum_{k=1}^{n-1} b_k b_{n-k} = a_n \quad \forall n \geq 2. \tag{B.3}$$

which gives

$$b_0 = \sqrt{a_0},\ b_1 = \frac{1}{2}\frac{a_1}{\sqrt{a_0}},\ b_k = \frac{1}{2b_0}(a_k - \sum_{i=1}^{k-1} b_i b_{k-i}) \quad \forall k \geq 2, \tag{B.4}$$

with similar relations between c_n and d_k.

Reference=matrix. $z_0 = 1,\ t = z-1$. With $t = z-1$, the relation (59) can be rewritten as:

$$\widetilde{z}^2 = \frac{1+3t/4}{1+t/4} = 1 + \sum_{k=1}^{\infty} 2(-1)^{k-1}\left(\frac{1}{4}\right)^k t^k, \tag{B.5}$$

Therefore (with the notations of (B.2)),

$$a_0 = 1, \quad a_k = 2(-1)^{k-1}4^{-k} \quad \forall k \geq 1 \tag{B.6}$$

and following (B.4),

$$b_0 = 1,\ \ b_1 = \frac{1}{4},\ b_k = (-1)^{k-1}\left(\frac{1}{4}\right)^k - \frac{1}{2}\sum_{i=1}^{k-1} b_i b_{k-i} \quad \forall n \geq 2 \tag{B.7}$$

The reference between the matrix ($z_0 = 1$), the coefficients c_n and d_k coincide with a_n and b_k respectively.

Reference=arithmetic mean. $z_0 = \dfrac{z+1}{2},\ t = Z = \dfrac{z-1}{z+1}$. Noting that

$$t = \frac{z-1}{z+1}, \quad z = \frac{1+t}{1-t}, \quad z_0 = \frac{z+1}{2} = \frac{1}{1-t} \tag{B.8}$$

it follows that:

$$\left(\frac{\widetilde{z}}{z_0}\right)^2 = \frac{1+t/2}{1-t/2}(1-t)^2 = 1 - t - \frac{1}{2}t^2 + \sum_{k=3}^{\infty} 2^{1-k}t^k. \tag{B.9}$$

Therefore

$$a_0 = 1,\ a_1 = -1,\ a_2 = -\frac{1}{2},\ a_k = 2^{1-k} \quad \forall k \geq 3. \tag{B.10}$$

The b_k's can then be obtained from (B.10) by (B.4). Similarly

$$\widetilde{z}^2 = \frac{1+t/2}{1-t/2} = 1 + 2\sum_{k=1}^{\infty}\left(\frac{t}{2}\right)^k. \tag{B.11}$$

Therefore

$$c_0 = 1, \quad c_k = 2^{1-k} \quad \forall k \geq 1. \tag{B.12}$$

The d_k's can then be obtained from (B.12) by (B.4).

Reference= geometric mean. $z_0 = \sqrt{z},\ t = w = \dfrac{\sqrt{z}-1}{\sqrt{z}+1}$. With

$$t = \frac{\sqrt{z}-1}{\sqrt{z}+1}, \quad z = \left(\frac{1+t}{1-t}\right)^2 = z_0^2, \tag{B.13}$$

one can write

$$\left(\frac{\widetilde{z}}{z_0}\right)^2 = \frac{1+t+t^2}{1-t+t^2}\left(\frac{1-t}{1+t}\right)^2 = \frac{1-t^3}{1+t^3}\frac{1-t}{1+t}. \tag{B.14}$$

Then, noting that

$$\frac{1-t}{1+t} = 1 + 2\sum_{k=1}^{\infty}(-t)^k, \quad \frac{1-t^3}{1+t^3} = 1 + 2\sum_{k=1}^{\infty}(-t)^{3k},$$

one arrives at

$$\left(\frac{\widetilde{z}}{z_0}\right)^2 = \sum_{k=0}^{\infty} a_k t^k, \quad a_0 = a'_0 a''_0, \quad a_k = \sum_{i=0}^{k} a'_i a''_{k-i}, \tag{B.15}$$

with

$$\left.\begin{aligned} &a'_0 = 1, \quad a'_i = 2(-1)^i \quad \forall i \geq 1, \\ &a''_0 = 1,\ a''_1 = 0,\ a''_2 = 0,\ a''_{3j} = 2(-1)^{3j},\ a''_{3j+1} = a''_{3j+2} = 0 \quad \forall j \geq 1 \end{aligned}\right\} \tag{B.16}$$

It is found that

$$\left.\begin{aligned} &a_0 = 1,\ a_1 = -2,\ a_2 = 2, \\ &a_{3j} = (-1)^{3j}(4j),\ a_{3j+1} = (-1)^{3j+1}(4j+2),\ a'_{3j+2} = (-1)^{3j+2}(4j+2) \quad \forall j \geq 1. \end{aligned}\right\} \tag{B.17}$$

Similarly

$$\widetilde{z}^2 = \frac{1+t+t^2}{1-t+t^2} = \frac{1-t^3}{1+t^3}\frac{1+t}{1-t}. \tag{B.18}$$

$$\widetilde{z}^2 = \sum_{k=0}^{\infty} c_k t^k, \quad c_0 = c'_0 c''_0, \quad c_k = \sum_{i=0}^{k} c'_i c''_{k-i}, \tag{B.19}$$

with

$$\left.\begin{aligned} &c'_0 = 1, \quad c'_i = 2 \quad \forall i \geq 1, \\ &c''_0 = 1,\ c''_1 = 0,\ c''_2 = 0,\ c''_{3j} = 2(-1)^{3j},\ c''_{3j+1} = c''_{3j+2} = 0 \quad \forall j \geq 1 \end{aligned}\right\} \tag{B.20}$$

It is found that

$$\left.\begin{aligned} &c_0 = 1,\ c_1 = 2,\ c_2 = 2, \\ &c_{3j} = 0,\ c_{3j+1} = c_{3j+2} = 2(-1)^{3j} \quad \forall j \geq 1. \end{aligned}\right\} \tag{B.21}$$

Appendix C. Comparison between the exact and numerical coefficients of the power series expansion for the Obnosov microstructure

The theoretical and numerical coefficients d_k for the expansion of $\widetilde{z}$ in powers of $t = \dfrac{z-1}{z+1}$ are given in the table below.

k	d_k (theoretical)	d_k^{NS} (numerical)	$\dfrac{\lvert d_k - d_k^{NS}\rvert}{d_k}$ (%)
0	1.000000000	1.000000000	0.00000000
1	0.500000000	0.500000000	0.00000000
2	0.125000000	0.125000000	0.00000000
3	0.625000000E-01	0.625000000E-01	0.00000000
4	0.234375000E-01	0.234375000E-01	0.00000000
5	0.117187500E-01	0.117141463E-01	0.03928491
6	0.488281250E-02	0.488051063E-02	0.04714230
7	0.244140625E-02	0.243735662E-02	0.16587284
8	0.106811523E-02	0.106609042E-02	0.18956850
9	0.534057617E-03	0.535119341E-03	0.19880327
10	0.240325928E-03	0.240856801E-03	0.22089710
11	0.120162964E-03	0.125219714E-03	4.20824340
12	0.550746918E-04	0.576030820E-04	4.59083858
13	0.275373459E-04	0.342665252E-04	24.4365573
14	0.127851963E-04	0.161497831E-04	26.3162702
15	0.639259815E-05	0.133939187E-04	109.522300
16	0.299653038E-05	0.649716274E-05	116.822856
17	0.149826519E-05	0.822353999E-05	448.870790
18	0.707514118E-06	0.407011143E-05	475.269288
19	0.353757059E-06	0.666877547E-05	1785.12859
20	0.168034603E-06	0.332550840E-05	1879.06166
21	0.840173016E-07	0.600176714E-05	7043.48953
22	0.400991667E-07	0.299895396E-05	7378.84359
23	0.200495833E-07	0.559017674E-05	27781.7602
24	0.960709201E-08	0.279466942E-05	28989.6498
25	0.480354601E-08	0.527750045E-05	109766.762

Table C.1: Obnosov problem. Discretization 128×128 pixels. MS-scheme. Comparison between the theoretical and numerical coefficients d_k of the power series with $t = \dfrac{z-1}{z+1}$.