EFFICIENT NUMERICAL TREATMENT OF HIGH-CONTRAST DIFFUSION PROBLEMS

A Dissertation Presented to
the Faculty of the Department of Mathematics
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Daria Kurzanova
May 2017

EFFICIENT NUMERICAL TREATMENT OF HIGH-CONTRAST DIFFUSION PROBLEMS

Daria Kurzanova

APPROVED:

Dr. Yuliya Gorb (Committee Chair)
Department of Mathematics, University of Houston

Dr. Yuri A. Kuznetsov
Department of Mathematics, University of Houston

Dr. Giles Auchmuty
Department of Mathematics, University of Houston

Dr. Dmitri Kuzmin
Institute of Applied Mathematics,
Dortmund University of Technology

Dean, College of Natural Sciences and Mathematics
University of Houston

Acknowledgements

I would like to thank my advisor, Dr. Yuliya Gorb for her expertise, direction, and assistance throughout my Ph.D. program. Without her constant support and encouragement, this work would not have been possible.

I would also like to thank the other members of my committee, Dr. Yuri Kuznetsov, Dr. Giles Auchmuty and Dr. Dmitri Kuzmin, for their service. Last two years Dr. Yuri Kuznetsov served as an excellent mentor, always willing to provide guidance and advice. His commitment to my success was invaluable.

I would like to thank my family for all their love, patience and support throughout my years at the University of Houston.

Last but definitely not least, I dedicate this dissertation to my husband Sergii and our son Ethan. I am indebted to them for being my source of inspiration and strength.

This research was supported by the NSF grant DMS-1350248.

EFFICIENT NUMERICAL TREATMENT OF HIGH-CONTRAST DIFFUSION PROBLEMS

An Abstract of a Dissertation
Presented to
the Faculty of the Department of Mathematics
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Daria Kurzanova
May 2017

Abstract

This dissertation concerns efficient numerical treatment of the elliptic partial differential equations with *high-contrast* coefficients. High-contrast means that the ratio between highest and lowest values of the coefficients is very high, or even infinite. A finite-element discretization of such equations yields a linear system with an ill-conditioned matrix which leads to significant issues in numerical methods.

The research in Chapter 2 introduces a procedure by which the discrete system obtained from a linear finite-element discretization of the given continuum problem is converted into an equivalent linear system of a saddle point type. Then a robust preconditioner for the Lancsoz method of minimized iterations for solving the derived saddle point problem is proposed. Numerical experiments demonstrate effectiveness and robustness of the proposed preconditioner and show that the number of iterations is independent of the contrast and the discretization size.

The research in Chapter 3 concerns the case of infinite-contrast problems with almost touching injections. The Dirichlet-Neumann domain decomposition algorithm yields a Schur complement linear system. The issue is that the block corresponding to the highly-dense part of the domain is impossible to obtain in practice. An approximation of this block is proposed by using a discrete Dirichlet-to-Neumann map, introduced in [11]. The process of construction of a discrete map together with all its properties is described and numerical illustrations with comparison to the solution obtained by the direct method are provided.

Contents

Chapter 1

Introduction

Composites, that are materials made from two or more constituents with different physical characteristics, are very common in both nature and engineering. These materials are more in demand because of their new properties, they could be stronger, lighter or cheaper as compared with traditional materials. The first obvious reason for studying composite materials is because of their usefulness, they are widely used in engineering. The second important reason is that what we learn from the theory of composites could be extended to other fields. While we work on challenging problems from composite field, we can develop new mathematical tools.

We distinguish composites whose two phases are described by different sets of equations and those described by equations of the same type. In the research proposed hereafter, we concern with the later. In this case, the phases are differentiated by coefficients of the corresponding partial differential equations (PDE). Moreover, we focus on mathematical models of so-called *high-contrast* composites. High-contrast means that the ratio between highest and lowest values of the coefficients is very high, even infinite. The example of such composites is a conductive medium with insulating inclusions.

Mathematical modeling of these type of composites poses significant challenges likewise, a numerical approximation for these composite materials results in a very large system of algebraic equations with an ill-conditioned matrix [17]. For example, the ill-conditioning of the discrete problem, describing composites with closely-spaced inclusions, is a consequence of the small thickness of the length between the inhomogeneities. Solving this system of equations is computationally expensive. Since there is a need in solving problems associated with the high-contrast composites with complex geometry new methods and tools have to be developed.

In this dissertation, we develop an efficient numerical treatment of the linear system arising from the discretization of the Poisson problem

$$-\nabla \cdot [\sigma(x)\nabla u] = f, \quad x \in \Omega \tag{1.1}$$

with appropriate boundary conditions on $\partial\Omega$. We assume that Ω is a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, that contains $m \geq 1$ polygonal or polyhedral subdomains $\mathcal{D}^i$, see Fig. 2.1. The main focus of this work is on the case where the coefficient function

$\sigma(x) \in L^\infty(\Omega)$ varies largely within the domain Ω, that is,

$$\kappa = \frac{\sup_{x\in\Omega} \sigma(x)}{\inf_{x\in\Omega} \sigma(x)} \gg 1.$$

The finite element method (FEM) discretization of this problem results in a linear system

$$\boldsymbol{\mathcal{K}}\overline{u} = \overline{F}, \tag{1.2}$$

with a large and sparse matrix $\boldsymbol{\mathcal{K}}$. A major issue in numerical treatments of (1.1), with the coefficient σ discussed above, is that the high-contrast leads to an ill-conditioned matrix $\boldsymbol{\mathcal{K}}$ in (1.2). If h is the discretization scale, then the condition number of the resulting stiffness matrix $\boldsymbol{\mathcal{K}}$ grows proportionally to h^{-2} with the coefficient of proportionality depending on κ. Because of that result, the high-contrast problems have been a subject of recent active research recently, see, e.g., [1, 2].

In Chapter 2 we assume that inclusions are separated by distances comparable to their sizes, while the key aspect in Chapter 3 is that injections are located very close, almost touching each other.

We remark that the work in Chapter 2 has been submitted for publication and the work in Chapter 3 is being prepared for submission.

Chapter 2

Robust preconditioner for high-contrast problems with moderate density of inclusions

2.1 Introduction

If $\boldsymbol{\mathcal{K}}$ of (1.2) is symmetric and positive definite, then (1.2) is typically solved with the Conjugate Gradient (CG) method, see e.g. [3], if $\boldsymbol{\mathcal{K}}$ is nonsymmetric the most common solver for (1.2) is the Generalized Minimal Residual Algorithm (GMRES), see e.g. [30]. In this dissertation, the introduction of an additional variable allows us to replace (1.2) with an equivalent formulation of the form

$$\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{\mathcal{F}} \tag{2.1}$$

with a *saddle point matrix* $\boldsymbol{\mathcal{A}}$ written in the block form:

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\boldsymbol{\Sigma} \end{bmatrix}, \tag{2.2}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric positive definite, $\mathbf{B} \in \mathbb{R}^{k\times n}$ is rank deficient, and $\boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}$ is symmetric and positive semidefinite, so that the corresponding linear system is singular but consistent. Unfortunately, the Krylov space iterative methods tend to converge very slowly when applied to systems with saddle point matrices and preconditioners are needed to achieve faster convergence.

The CG method, which was mainly developed for the iterative solution of linear systems with symmetric definite matrices is not in general robust for systems with indefinite matrices, [35]. The *Lanczos algorithm* of minimized iterations does not have such a restriction and has been utilized in this dissertation. Below in this chapter, we introduce a construction of a robust preconditioner for solving (2.1) by the Lanczos iterative scheme, whose convergence rate is independent of the contrast parameter $\kappa \gg 1$ and the discretization size $h > 0$.

Also, the special case of (1.2) with (2.2) tackled in this chapter is when $\boldsymbol{\Sigma} \equiv \mathbf{0}$. The problem of this type has received considerable attention over the years. But the most studied case is when $\boldsymbol{\mathcal{A}}$ is *nonsingular*, in which case $\mathbf{B}$ must be of full rank, see, e.g., [22, 27] and references therein. The main focus of this research is on singular $\boldsymbol{\mathcal{A}}$ with the rank deficient block $\mathbf{B}$. Below we construct a block-diagonal preconditioner for the Lanczos method employed to solve the problem (2.1), and this preconditioner is also singular. We also provide numerical experiments demonstrating the robustness of the proposed approach with respect to the contrast κ and mesh size $h > 0$.

The rest of this chapter is organized as follows. In Section 2.2 the mathematical problem formulation is presented and main results are stated. Section 2.3 discusses proofs of main results, and numerical results of the proposed procedure are given in Section 2.4. Conclusions are presented in Section 2.5.

2.2 Problem Formulation and Main Results

Consider an open, bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$ with piece-wise smooth boundary $\Gamma = \partial\Omega$, that contains $m \geq 1$ subdomains $\mathcal{D}^i$, which are located at distances comparable to their sizes from one another, see Fig. 2.1. For simplicity, we assume that Ω and $\mathcal{D}^i$ are polygons if $d = 2$ or polyhedra if $d = 3$. The union of $\mathcal{D}^i$ is denoted by $\mathcal{D}$.

In the domain Ω we consider the following elliptic problem

$$\begin{cases} -\nabla \cdot [\sigma(x)\nabla u] &= f, \quad x \in \Omega \\ u &= 0, \quad x \in \Gamma \end{cases} \tag{2.3}$$

with the coefficient σ that largely varies inside the domain Ω. For simplicity of the presentation, the focus of this case is where σ is a piecewise constant function given by

$$\sigma(x) = \begin{cases} 1, & x \in \Omega \setminus \overline{\mathcal{D}} \\ 1 + \dfrac{1}{\varepsilon_i}, & x \in \mathcal{D}_i, \ i \in \{1, \ldots, m\} \end{cases} \tag{2.4}$$

with $\max\limits_i \varepsilon_i \ll 1$. We also assume that the source term in (2.3) is $f \in L^2(\Omega)$.

Figure 2.1: The domain Ω with highly conducting inclusions $\mathcal{D}^i$, $i \in \{1, \dots, m\}$

When performing a FEM discretization of (2.3) with (2.4) with polynomials of first order, we choose a FEM space $V_h \subset H_0^1(\Omega)$ to be the space of linear finite-element functions defined on a conforming quasi-uniform triangulation Ω_h of Ω of the size $h \ll 1$. The mesh is adapted to inclusions, that is, nodes are required to fit $\partial\mathcal{D}^i$ for all inclusions. For simplicity, we assume that $\partial\Omega_h = \Gamma$. With that, the classical FEM discretization results in the system of the type (1.2). We proceed differently and derive another discretized system of the saddle point type as shown below.

2.2.1 Derivation of a Singular Saddle Point Problem

If $\mathcal{D}_h^i = \Omega_h|_{\mathcal{D}^i}$ then we denote $V_h^i := V_h|_{\mathcal{D}_h^i}$ and $\mathcal{D}_h := \cup_{i=1}^m \mathcal{D}_h^i$. The FEM formulation of (2.3)-(2.4) is

$$\text{Find} \quad u_h \in V_h \quad \text{and} \quad \lambda_h = (\lambda_h^1, \dots, \lambda_h^m) \quad \text{with} \quad \lambda_h^i \in V_h^i \quad \text{such that}$$

$$\int\limits_{\Omega_h} \nabla u_h \cdot \nabla v_h \, dx + \int\limits_{\mathcal{D}_h} \nabla \lambda_h \cdot \nabla v_h \, dx = \int\limits_{\Omega_h} f v_h \, dx, \quad \forall v_h \in V_h, \tag{2.5}$$

provided

$$u_h = \varepsilon_i \lambda_h^i + c_i \quad \text{in} \quad \mathcal{D}_h^i, \quad i \in \{1, \dots, m\}, \tag{2.6}$$

where c_i is an arbitrary constant. First, we turn out attention to the FEM discretization of (2.5) that yields a system of linear equations

$$\mathbf{A}\overline{u} + \mathbf{B}^T\overline{\lambda} = \overline{\mathrm{F}}, \tag{2.7}$$

and then discuss implications of (2.6).

To provide the comprehensive description of all elements of the system (2.7), we introduce the following notations for the number of degrees of freedom in different parts of Ω_h. Let N be the total number of nodes in Ω_h, and n be the number of nodes in $\overline{\mathcal{D}}_h$ so that

$$n = \sum_{i=1}^{m} n_i,$$

where n_i denotes the number of degrees of freedom in $\overline{\mathcal{D}}_h^i$, and, finally, n_0 is the number of nodes in $\Omega_h \setminus \overline{\mathcal{D}_h}$, so that we have

$$N = n_0 + n = n_0 + \sum_{i=1}^{m} n_i.$$

Then in (2.7), the vector $\overline{u} \in \mathbb{R}^N$ has entries $u_i = u_h(x_i)$ with $x_i \in \overline{\Omega}_h$. We count the entries of $\overline{u}$ in such a way that its first n elements correspond to the nodes of $\overline{\mathcal{D}}_h$, and the remaining n_0 entries correspond to the nodes of $\overline{\Omega}_h \setminus \overline{\mathcal{D}}_h$. Similarly, the vector $\overline{\lambda} \in \mathbb{R}^n$ has entries $\lambda_i = \lambda_h(x_i)$ where $x_i \in \overline{\mathcal{D}}_h$.

The symmetric positive definite matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ of (2.7) is the stiffness matrix that arises from the discretization of the Laplace operator with the homogeneous Dirichlet boundary conditions on Γ. Entries of $\mathbf{A}$ are defined by

$$(\mathbf{A}\overline{u}, \overline{v}) = \int\limits_{\Omega_h} \nabla u_h \cdot \nabla v_h \; dx, \quad \text{where} \quad \overline{u}, \overline{v} \in \mathbb{R}^N, \quad u_h, v_h \in V_h, \tag{2.8}$$

where $(\cdot,\cdot)$ is the standard dot-product of vectors. This matrix can also be partitioned into

$$\mathbf{A} = \begin{bmatrix} \mathrm{A}_{\mathcal{D}\mathcal{D}} & \mathrm{A}_{\mathcal{D}0} \\ \mathrm{A}_{0\mathcal{D}} & \mathrm{A}_{00} \end{bmatrix}, \tag{2.9}$$

where the block $\mathrm{A}_{\mathcal{D}\mathcal{D}} \in \mathbb{R}^{n\times n}$ is the stiffness matrix corresponding to the highly conducting inclusions $\overline{\mathcal{D}}_h^i$, $i \in \{1,\dots,m\}$, the block $\mathrm{A}_{00} \in \mathbb{R}^{n_0\times n_0}$ corresponds to the region outside of $\overline{\mathcal{D}}_h$, and the entries of $\mathrm{A}_{\mathcal{D}0} \in \mathbb{R}^{n\times n_0}$ and $\mathrm{A}_{0\mathcal{D}} = \mathrm{A}_{\mathcal{D}0}^T$ are assembled from contributions both from finite elements in $\overline{\mathcal{D}}_h$ and $\overline{\Omega}_h \setminus \overline{\mathcal{D}}_h$.

The matrix $\mathbf{B} \in \mathbb{R}^{n\times N}$ of (2.7) is also written in the block form as

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\mathcal{B}}_{\mathcal{D}} & \mathbf{0} \end{bmatrix} \tag{2.10}$$

with zero-matrix $\mathbf{0} \in \mathbb{R}^{n\times n_0}$ and $\boldsymbol{\mathcal{B}}_{\mathcal{D}} \in \mathbb{R}^{n\times n}$ that corresponds to the highly conducting inclusions. The matrix $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ is the stiffness matrix corresponding to the discretization of the Laplace operator in the domain $\overline{\mathcal{D}}_h$ with the Neumann boundary conditions on $\partial\mathcal{D}_h$. In its turn, $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ is written in the block form by

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}} = \begin{bmatrix} \mathrm{B}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathrm{B}_m \end{bmatrix} = \mathrm{diag}\,(\mathrm{B}_1,\dots,\mathrm{B}_m)$$

with matrices $\mathrm{B}_i \in \mathbb{R}^{n_i\times n_i}$, whose entries are similarly defined by

$$(\mathrm{B}_i\overline{u},\overline{v}) = \int\limits_{\mathcal{D}_h^i} \nabla u_h \cdot \nabla v_h \; dx, \quad \text{where} \quad \overline{u},\overline{v} \in \mathbb{R}^{n_i}, \quad u_h, v_h \in V_h^i. \tag{2.11}$$

We remark that each B_i is positive semidefinite with

$$\ker \mathrm{B}_i = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\}. \tag{2.12}$$

Finally, the vector $\overline{\mathrm{F}} \in \mathbb{R}^N$ of (2.7) is defined in a similar way by

$$(\overline{\mathrm{F}}, \overline{v}) = \int\limits_{\Omega_h} f v_h \, dx, \quad \text{where} \quad \overline{v} \in \mathbb{R}^N, \quad v_h \in V_h.$$

To complete the derivation of the linear system corresponding to (2.5)-(2.6), we rewrite (2.6) in the weak form that is as follows:

$$\int\limits_{\mathcal{D}_h^i} \nabla u_h \cdot \nabla v_h^i \, dx - \varepsilon_i \int\limits_{\mathcal{D}_h^i} \nabla \lambda_h^i \cdot \nabla v_h^i \, dx = 0, \quad i \in \{1, \dots, m\} \quad \forall \, v_h^i \in V_h^i, \tag{2.13}$$

and add the discrete analog of (2.6) to the system (2.7). For that, denote

$$\mathbf{\Sigma}_\varepsilon = \begin{bmatrix} \varepsilon_1 \mathrm{B}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \varepsilon_m \mathrm{B}_m \end{bmatrix} = \mathrm{diag}\,(\varepsilon_1 \mathrm{B}_1, \dots, \varepsilon_m \mathrm{B}_m),$$

then (2.13) implies

$$\mathbf{\Sigma}_\varepsilon \overline{\lambda} = \mathbf{B} \overline{u}. \tag{2.14}$$

This together with (2.7) yields

$$\begin{cases} \mathbf{A}\overline{u} + \mathbf{B}^T \overline{\lambda} &= \overline{\mathrm{F}}, \\ \mathbf{B}\overline{u} - \mathbf{\Sigma}_\varepsilon \overline{\lambda} &= \overline{0}, \end{cases} \qquad \overline{u} \in \mathbb{R}^N, \quad \mathbb{R}^n \ni \overline{\lambda} \perp \ker \boldsymbol{\mathcal{B}}_{\mathcal{D}}, \tag{2.15}$$

or

$$\boldsymbol{\mathcal{A}}_\varepsilon \mathbf{x}_\varepsilon = \overline{\mathcal{F}}, \tag{2.16}$$

where

$$\boldsymbol{\mathcal{A}}_\varepsilon = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\boldsymbol{\Sigma}_\varepsilon \end{bmatrix} = \begin{bmatrix} \mathrm{A}_{\mathcal{D}\mathcal{D}} & \mathrm{A}_{\mathcal{D}0} & \boldsymbol{\mathcal{B}}_{\mathcal{D}} \\ \mathrm{A}_{0\mathcal{D}} & \mathrm{A}_{00} & \mathbf{0} \\ \boldsymbol{\mathcal{B}}_{\mathcal{D}} & \mathbf{0} & -\boldsymbol{\Sigma}_\varepsilon \end{bmatrix}, \quad \mathbf{x}_\varepsilon = \begin{bmatrix} \overline{u} \\ \overline{\lambda} \end{bmatrix}, \quad \overline{\mathcal{F}} = \begin{bmatrix} \overline{\mathrm{F}} \\ \overline{0} \end{bmatrix}. \tag{2.17}$$

This saddle point formulation (2.16)-(2.17) for the PDE (2.3)-(2.4) was first proposed in [25]. Since $\mathbf{A}$ is positive definite matrix, there exists a unique solution $\overline{u} \in \mathbb{R}^N$ of (2.16)-(2.17).

2.2.2 Discussions on the system (2.15)

Denote the solution of (2.16)-(2.17) by

$$\mathbf{x}_\varepsilon = \begin{bmatrix} \overline{u}_\varepsilon \\ \overline{\lambda}_\varepsilon \end{bmatrix},$$

and consider an auxiliary linear system

$$\boldsymbol{\mathcal{A}}_0 \mathbf{x}_0 = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \overline{u}_0 \\ \overline{\lambda}_0 \end{bmatrix} = \begin{bmatrix} \overline{\mathrm{F}} \\ \overline{0} \end{bmatrix}, \tag{2.18}$$

or

$$\begin{cases} \mathbf{A}\overline{u}_0 + \mathbf{B}^T\overline{\lambda}_0 & = \overline{\mathrm{F}}, \\ \mathbf{B}\overline{u}_0 & = \overline{0}. \end{cases} \tag{2.19}$$

where matrices $\mathbf{A}$, $\mathbf{B}$ and the vector $\overline{\mathrm{F}}$ are the same as above. The linear system (2.18) or, equivalently, (2.19) emerges in a FEM discretization of the diffusion problem posed in the domain Ω whose inclusions are *infinitely conducting*, where $\varepsilon = 0$ in (2.4). The corresponding PDE formulation for problem (2.19) might be as follows (see e.g. [13])

$$\begin{cases} \triangle u &= f, \qquad x \in \Omega \setminus \overline{\mathcal{D}} \\ u &= \text{const}, \quad x \in \partial\mathcal{D}^i, \ i \in \{1, \dots, m\} \\ \int\limits_{\partial\mathcal{D}^i} \nabla u \cdot \mathbf{n}_i \, ds &= 0, \qquad i \in \{1, \dots, m\} \\ u &= 0, \qquad x \in \Gamma \end{cases} \tag{2.20}$$

where $\mathbf{n}_i$ is the outer unit normal to the surface $\partial\mathcal{D}^i$. If $u \in H_0^1(\Omega \setminus \overline{\mathcal{D}})$ is an electric potential then it attains constant values on the inclusions $\mathcal{D}^i$ and these constants are not known a priori, and are unknowns of the problem (2.20), together with u.

Formulation (2.18) or (2.19) also arises in constrained quadratic optimization problem and solving the Stokes equations for an incompressible fluid [18], and solving elliptic problems using methods combining a fictitious domain and a distributed Lagrange multiplier techniques to force boundary conditions [19].

Then the following relation between solutions of systems (2.15) and (2.19) holds true.

Lemma 1. *Let* $\boldsymbol{x}_0 = \begin{bmatrix} \overline{u}_0 \\ \overline{\lambda}_0 \end{bmatrix} \in \mathbb{R}^{N+n}$ *be the solution of the linear system* (2.19)*, and*

$\boldsymbol{x}_\varepsilon = \begin{bmatrix} \overline{u}_\varepsilon \\ \overline{\lambda}_\varepsilon \end{bmatrix} \in \mathbb{R}^{N+n}$ *the solution of* (2.15). *Then*

$$\overline{u}_\varepsilon \to \overline{u}_0 \quad \textit{as} \quad \varepsilon \to 0.$$

This lemma asserts that the discrete approximation for the problem (2.3)-(2.4) converges to the discrete approximation of the solution of (2.20) as $\varepsilon \to 0$. Note, the continuum version of this fact was shown in [13].

Proof. Without loss of generality, assume that all $\varepsilon_i = \varepsilon$, $i \in \{1, \ldots, m\}$. Hereafter, denote by C a positive constant that is independent of ε.

Subtract first equations of (2.15) and (2.19) and multiply by $\overline{u}_\varepsilon - \overline{u}_0$ to obtain

$$\left(\mathbf{A}(\overline{u}_\varepsilon - \overline{u}_0, \overline{u}_\varepsilon - \overline{u}_0)\right) + \Big(\mathbf{B}^T(\overline{\lambda}_\varepsilon - \overline{\lambda}_0), \overline{u}_\varepsilon - \overline{u}_0\Big) = \overline{0}.$$

Recall, the matrix $\mathbf{A}$ is SPD then

$$(\mathbf{A}\xi, \xi) \geq \mu_1(\mathbf{A})\|\xi\|^2, \quad \forall \xi \in \mathbb{R}^N, \tag{2.21}$$

where $\mu_1(\mathbf{A}) > 0$ is the *minimal eigenvalue* of $\mathbf{A}$, and $\|\cdot\| = (\cdot, \cdot)$.

Making use of the second equation of (2.15) we have

$$\mu_1(\mathbf{A})\|\overline{u}_\varepsilon - \overline{u}_0\|^2 \leq -\left(\varepsilon \boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\lambda}_\varepsilon, \overline{\lambda}_\varepsilon\right) + \left(\varepsilon \boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\lambda}_\varepsilon, \overline{\lambda}_0\right) \leq \left(\varepsilon \boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\lambda}_\varepsilon, \overline{\lambda}_0\right),$$

where we used the fact that $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ is positive semidefinite. Then

$$\|\overline{u}_\varepsilon - \overline{u}_0\|^2 \leq \varepsilon \|\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\lambda}_\varepsilon\|. \tag{2.22}$$

Now the goal is to bound $\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\lambda}_{\varepsilon}$ by a constant independent of ε. To show it we multiply the first equation of (2.15) by $\mathbf{AB}^T\overline{\lambda}_{\varepsilon}$:

$$\left(\mathbf{A}\overline{u}_{\varepsilon}, \mathbf{AB}^T\overline{\lambda}_{\varepsilon}\right) + \left(\mathbf{B}^T\overline{\lambda}_{\varepsilon}, \mathbf{AB}^T\overline{\lambda}_{\varepsilon}\right) = \left(\overline{\mathrm{F}}, \mathbf{AB}^T\overline{\lambda}_{\varepsilon}\right),$$

which yields

$$\mu_1(\mathbf{A})\|\mathbf{B}^T\overline{\lambda}_{\varepsilon}\|^2 \leq C\|\overline{\mathrm{F}} - \mathbf{A}\overline{u}_{\varepsilon}\|\|\mathbf{B}^T\overline{\lambda}_{\varepsilon}\|.$$

Note that $\|\mathbf{B}^T\overline{\lambda}_{\varepsilon}\| = \|\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\lambda}_{\varepsilon}\|$, hence,

$$\|\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\lambda}_{\varepsilon}\| \leq C\|\overline{\mathrm{F}} - \mathbf{A}\overline{u}_{\varepsilon}\|, \tag{2.23}$$

so collecting estimates (2.22) and (2.23), it remains to show $\|\overline{u}_{\varepsilon}\|$ is bounded. For that we multiply the first equation of (2.15) by $\overline{u}_{\varepsilon}$ and obtain

$$(\mathbf{A}\overline{u}_{\varepsilon}, \overline{u}_{\varepsilon}) + \left(\mathbf{B}^T\overline{\lambda}_{\varepsilon}, \overline{u}_{\varepsilon}\right) = \left(\overline{\mathrm{F}}, \overline{u}_{\varepsilon}\right),$$

which yields

$$\mu_1(\mathbf{A})\|\overline{u}_{\varepsilon}\|^2 + \left(\mathbf{B}^T\overline{\lambda}_{\varepsilon}, \overline{u}_{\varepsilon}\right) \leq \|\overline{F}\|\|\overline{u}_{\varepsilon}\|,$$

where we used (2.21) and Cauchy-Schwarz inequality.

Making use of the second equation of (2.15) and the fact that $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ is positive semidefinite we have

$$\mu_1(\mathbf{A})\|\overline{u}_{\varepsilon}\|^2 \leq \|\overline{F}\|\|\overline{u}_{\varepsilon}\|.$$

Hence,

$$\|\overline{u}_{\varepsilon}\| \leq \frac{\|\overline{F}\|}{\mu_1(\mathbf{A})}. \tag{2.24}$$

This shows that boundness of $\|\overline{u}_\varepsilon\|$.

Collecting estimates (2.22), (2.23) and (2.24), we conclude

$$\|\overline{u}_\varepsilon - \overline{u}_0\|^2 \leq C\varepsilon,$$

hence $\overline{u}_\varepsilon \to \overline{u}_0$ as $\varepsilon \to 0$. □

2.2.3 Spectral Properties of the Matrix $\boldsymbol{\mathcal{A}}_0$ of the Auxiliary Problem (2.19)

As previously observed, see, e.g., [24], the following matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T \end{bmatrix}, \tag{2.25}$$

is the best choice for a preconditioner of $\boldsymbol{\mathcal{A}}_0$. This is because there are exactly three eigenvalues of $\boldsymbol{\mathcal{A}}_0$ associated with the following generalized eigenvalue problem

$$\boldsymbol{\mathcal{A}}_0 \begin{bmatrix} \overline{u} \\ \overline{\lambda} \end{bmatrix} = \mu \mathbf{P} \begin{bmatrix} \overline{u} \\ \overline{\lambda} \end{bmatrix}, \tag{2.26}$$

and which are: $\mu_1 < 0$, $\mu_2 = 1$ and $\mu_3 > 1$; hence, a Krylov subspace iteration method applied for a preconditioned system for solving (2.26) with (2.25) *converges to the exact solution in three iterations.*

The preconditioner (2.25) is also the best choice for our original problem (2.16)-(2.17) with $\varepsilon > 0$ as the eigenvalue of the generalized eigenvalue problem

$$\boldsymbol{\mathcal{A}}_\varepsilon \mathbf{x} = \mu \mathbf{P} \mathbf{x}$$

belonging to the union of $[c_1, c_2] \cup [c_3, c_4]$ with $c_1 \leq c_2 < 0$ and $0 < c_3 \leq c_4$, with numbers c_i being dependent on eigenvalues of (2.26) but not h, see [25].

Since expensive evaluation of $\mathbf{A}^{-1}$ in (2.25) makes $\mathbf{P}$ of limited practical use, $\mathbf{P}$ is a subject of primarily theoretical interest. To construct a preconditioner that one can actually use in practice, we seek a matrix

$$\boldsymbol{\mathcal{P}} = \begin{bmatrix} \mathcal{P}_{\mathrm{A}} & 0 \\ 0 & \mathcal{P}_{\mathrm{B}} \end{bmatrix}, \tag{2.27}$$

such that there are constants α, β independent of the mesh size h such that

$$\alpha(\mathbf{P}\mathbf{x}, \mathbf{x}) \leq (\boldsymbol{\mathcal{P}}\mathbf{x}, \mathbf{x}) \leq \beta(\mathbf{P}\mathbf{x}, \mathbf{x}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^N. \tag{2.28}$$

This property (2.28) is hereafter referred to as *spectral equivalence* of $\boldsymbol{\mathcal{P}}$ to $\mathbf{P}$ of (2.25). Below, we construct $\boldsymbol{\mathcal{P}}$ of the form (2.27) in such a way that the block $\mathcal{P}_{\mathrm{A}}$ is spectrally equivalent to $\mathbf{A}$, whereas $\mathcal{P}_{\mathrm{B}}$ is spectrally equivalent to $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$. For the former one, use any existing priconditioner developed for symmetric and positive definite matrices. Our primary aim is to construct a preconditioner $\mathcal{P}_{\mathrm{B}}$ that could be effectively used in solving (2.15).

2.2.4 Main Result: Block-Diagonal Preconditioner

The main theoretical result of this chapter establishes a robust preconditioner for solving (2.18) or, equivalently (2.19), and is given in the following theorem.

Theorem 1. *Let the triangulation Ω_h for* (2.20) *be conforming and quasi-uniform. Then the matrix $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ is spectrally equivalent to the matrix $\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^T$, that is, there* *exist constants $\mu_\star, \mu^\star > 0$ independent of h and such that*

$$\mu_\star \leq \frac{\left(\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\psi}, \overline{\psi}\right)}{\left(\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{T}\overline{\psi}, \overline{\psi}\right)} \leq \mu^\star, \quad \textit{for all} \quad 0 \neq \overline{\psi} \in \mathbb{R}^n,\ \overline{\psi} \perp \ker \boldsymbol{\mathcal{B}}_{\mathcal{D}}. \tag{2.29}$$

This theorem asserts that the nonzero eigenvalues of the generalized eigenproblem

$$\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{T}\overline{\psi} = \mu \boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\psi}, \quad \overline{\psi} \in \mathbb{R}^n, \tag{2.30}$$

are bounded. Hence, its proof is based on the construction of the **upper** and **lower** bounds for μ in (2.30) and is comprised of the following facts many of which are proven in the next section.

Lemma 2. *The following equality of matrices holds*

$$\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}^{T} = \boldsymbol{\mathcal{B}}_{\mathcal{D}}\mathrm{S}_{00}^{-1}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{T}, \tag{2.31}$$

where

$$\mathrm{S}_{00} = \mathrm{A}_{\mathcal{D}\mathcal{D}} - \mathrm{A}_{\mathcal{D}0}\mathrm{A}_{00}^{-1}\mathrm{A}_{0\mathcal{D}},$$

is the Schur complement to the block A_{00} of the matrix $\boldsymbol{A}$ of (2.18)*.*

This fact is straightforward and comes from the block structure of matrices $\mathbf{A}$ of (2.9) and $\mathbf{B}$ of (2.10). Indeed, using this, the generalized eigenproblem (2.30) can be rewritten as

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}\mathrm{S}_{00}^{-1}\boldsymbol{\mathcal{B}}_{\mathcal{D}}\,\overline{\psi} = \mu \boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{\psi}, \qquad \overline{\psi} \in \mathbb{R}^n. \tag{2.32}$$

Introduce a matrix $\mathrm{B}_{\mathcal{D}}^{1/2}$ via $\boldsymbol{\mathcal{B}}_{\mathcal{D}} = \mathrm{B}_{\mathcal{D}}^{1/2}\mathrm{B}_{\mathcal{D}}^{1/2}$ and note that $\ker \boldsymbol{\mathcal{B}}_{\mathcal{D}} = \ker \mathrm{B}_{\mathcal{D}}^{1/2}$.

Lemma 3. *The generalized eigenvalue problem* (2.32) *is equivalent to*

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2} \mathrm{S}_{00}^{-1} \boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\, \overline{\varphi} = \mu\, \overline{\varphi}, \tag{2.33}$$

in the sense that they both have the same eigenvalues μ's, and the corresponding eigenvectors are related via $\overline{\varphi} = \boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2} \overline{\psi} \in \mathbb{R}^n$.

Lemma 4. *The generalized eigenvalue problem* (2.33) *is equivalent to*

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}\, \overline{u}_{\mathcal{D}} = \mu \mathrm{S}_{00}\, \overline{u}_{\mathcal{D}}, \tag{2.34}$$

in the sense that both problems have the same eigenvalues μ's, and the corresponding eigenvectors are related via $\overline{u}_{\mathcal{D}} = \mathrm{S}_{00}^{-1} \boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2} \overline{\varphi} \in \mathbb{R}^n$.

This result is also straightforward and can be obtained multiplying (2.33) by $\mathrm{S}_{00}^{-1} \boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}$.

To that end, establishing the upper and lower bounds for the eigenvalues of (2.34) and due to equivalence of (2.34) with (2.33), and hence (2.32), we obtain that the eigenvalues of (2.30) are bounded. Our interest is in the nonzero eigenvalues of (2.34), for which the following result holds.

Lemma 5. *Let the triangulation Ω_h for* (2.20) *be conforming and quasi-uniform. Then there exists $\hat{\mu}_\star > 0$ independent of the mesh size $h > 0$ such that*

$$\hat{\mu}_\star \leq \frac{(\boldsymbol{\mathcal{B}}_{\mathcal{D}}\, \overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}})}{(\mathrm{S}_{00} \overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}})} \leq 1, \quad \textit{for all} \quad 0 \neq \overline{u}_{\mathcal{D}} \in \mathbb{R}^n,\ \overline{u}_{\mathcal{D}} \perp \ker \boldsymbol{\mathcal{B}}_{\mathcal{D}}. \tag{2.35}$$

2.3 Proofs of statements in Chapter 2.2.4

2.3.1 Harmonic extensions

Hereafter, we will use the index $\mathcal{D}$ to indicate vectors or functions associated with the domain $\mathcal{D}$ that is the union of all inclusions, and index 0 to indicate quantities that are associated with the domain outside the inclusions $\Omega \setminus \overline{\mathcal{D}}$.

Now we recall some classical results from the theory of elliptic PDEs. Suppose a function $u^{\mathcal{D}} \in H^1(\mathcal{D})$, then consider its harmonic extension $u^0 \in H^1(\Omega \setminus \overline{\mathcal{D}})$ that satisfies

$$\begin{cases} - \triangle u^0 &= 0, \quad \text{in } \Omega \setminus \overline{\mathcal{D}}, \\ u^0 &= u^{\mathcal{D}}, \quad \text{on } \partial\mathcal{D}, \\ u^0 &= 0, \quad \text{on } \Gamma. \end{cases} \tag{2.36}$$

For such functions the following holds true:

$$\int\limits_{\Omega} |\nabla u|^2 \, \mathrm{d}x = \min_{v \in H_0^1(\Omega)} \int\limits_{\Omega} |\nabla v|^2 \, \mathrm{d}x, \tag{2.37}$$

where

$$u = \begin{cases} u^{\mathcal{D}}, & \text{in } \mathcal{D} \\ u^0, & \text{in } \Omega \setminus \overline{\mathcal{D}} \end{cases} \qquad \text{and} \qquad v = \begin{cases} u^{\mathcal{D}}, & \text{in } \mathcal{D} \\ v^0, & \text{in } \Omega \setminus \overline{\mathcal{D}} \end{cases} \tag{2.38}$$

where the function $v^0 \in H^1(\Omega \setminus \overline{\mathcal{D}})$ such that $v^0|_\Gamma = 0$, and

$$\|u\|_{H_0^1(\Omega)} \leq C \|u^{\mathcal{D}}\|_{H^1(\mathcal{D})} \quad \text{with the constant } C \text{ independent of } \ u^{\mathcal{D}}, \tag{2.39}$$

where $\|\cdot\|_{H^1(\Omega)}$ denotes the standard norm of $H^1(\Omega)$:

$$\|v\|^2_{H^1(\Omega)} = \int\limits_{\Omega} |\nabla v|^2 dx + \int\limits_{\Omega} v^2 dx, \tag{2.40}$$

and $\|v\|^2_{H^1_0(\Omega)} = \int\limits_{\Omega} |\nabla v|^2 dx.$

In other words, the function u^0 of (2.38) is the *best extension* of $u^{\mathcal{D}} \in H^1(\mathcal{D})$ among all $H^1(\Omega \setminus \overline{\mathcal{D}})$ functions that vanish on Γ, because it minimizes the energy functional (2.37). The algebraic linear system that corresponds to (2.37) satisfies the similar property. Namely, if the vector $\overline{u}_0 \in \mathbb{R}^{n_0}$ is a FEM discretization of the function $u^0 \in H^1_0(\Omega \setminus \overline{\mathcal{D}})$ of (2.36), then for a given $\overline{u}_{\mathcal{D}} \in \mathbb{R}^n$, the best extension $\overline{u}_0 \in \mathbb{R}^{n_0}$ would satisfy

$$\mathrm{A}_{0\mathcal{D}}\, \overline{u}_{\mathcal{D}} + \mathrm{A}_{00}\, \overline{u}_0 = 0, \tag{2.41}$$

and

$$\left(\mathbf{A} \begin{bmatrix} \overline{u}_{\mathcal{D}} \\ \overline{u}_0 \end{bmatrix}, \begin{bmatrix} \overline{u}_{\mathcal{D}} \\ \overline{u}_0 \end{bmatrix} \right) = \min_{\overline{v}_0 \in \mathbb{R}^{n_0}} \left(\mathbf{A} \begin{bmatrix} \overline{u}_{\mathcal{D}} \\ \overline{v}_0 \end{bmatrix}, \begin{bmatrix} \overline{u}_{\mathcal{D}} \\ \overline{v}_0 \end{bmatrix} \right). \tag{2.42}$$

2.3.2 Proof of Lemma 3

Consider generalized eigenvalue problem (2.32) and replace $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ with $\boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}} \boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}}$ there, then

$$\boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}} \boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}} \mathrm{S}^{-1}_{00} \boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}} \boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}}\, \overline{\psi} = \mu \boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}} \boldsymbol{\mathcal{B}}^{1/2}_{\mathcal{D}}\, \overline{\psi}.$$

Now multiply both sides by the Moore-Penrose pseudo inverse[1] $\left[\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\right]^{\dagger}$, see e.g. [3]:

$$\left[\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\right]^{\dagger}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\mathrm{S}_{00}^{-1}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\,\overline{\psi} = \mu\left[\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\right]^{\dagger}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\,\overline{\psi}.$$

This pseudo inverse has the property that

$$\left[\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\right]^{\dagger}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2} = \mathrm{P}_{\mathrm{im}},$$

where P_{im} is an orthogonal projector onto the image $\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}$, hence, $\mathrm{P}_{\mathrm{im}}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2} = \boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}$ and therefore,

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\mathrm{S}_{00}^{-1}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\overline{\varphi} = \mu\overline{\varphi}, \quad \text{where} \quad \overline{\varphi} = \boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\overline{\psi}.$$

Conversely, consider the eigenvalue problem (2.33), and multiply its both sides by $\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}$. Then

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\mathrm{S}_{00}^{-1}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\,\overline{\varphi} = \mu\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\,\overline{\varphi},$$

where we replace $\overline{\varphi}$ by $\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\,\overline{\psi}$

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\mathrm{S}_{00}^{-1}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\,\overline{\psi} = \mu\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\,\overline{\psi}$$

to obtain (2.32). □

2.3.3 Proof of Lemma 5

I. Upper Bound for the Generalized Eigenvalues of (2.30)

[1] $\mathrm{M}^{\dagger}$ is the Moore-Penrose pseudo inverse of M if and only if it satisfies the following Moore-Penrose equations:

$$\text{(i) } \mathrm{M}^{\dagger}\mathrm{M}\mathrm{M}^{\dagger} = \mathrm{M}^{\dagger}, \quad \text{(ii) } \mathrm{M}\mathrm{M}^{\dagger}\mathrm{M} = \mathrm{M}, \quad \text{(iii) } \mathrm{M}\mathrm{M}^{\dagger} \text{ and } \mathrm{M}^{\dagger}\mathrm{M} \text{ are symmetric.}$$

Consider $\overline{u} = \begin{bmatrix} \overline{u}_{\mathcal{D}} \\ \\ \overline{u}_0 \end{bmatrix} \in \mathbb{R}^N$ with $\overline{u}_{\mathcal{D}} \in \mathbb{R}^n$, $\overline{u}_{\mathcal{D}} \perp \ker \boldsymbol{\mathcal{B}}_{\mathcal{D}}$, and $\overline{u}_0 \in \mathbb{R}^{n_0}$ satisfying (2.41), then

$$(\mathrm{S}_{00}\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}}) = (\mathbf{A}\overline{u}, \overline{u})\,. \tag{2.43}$$

Using (2.8) and (2.11) we obtain from (2.43):

$$\mu = \frac{(\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}})}{(\mathrm{S}_{00}\,\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}})} = \frac{(\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}})}{(\mathbf{A}\overline{u}, \overline{u})} = \frac{\displaystyle\int_{\mathcal{D}_h} |\nabla u_h^{\mathcal{D}}|^2 \,\mathrm{d}x}{\displaystyle\int_{\Omega_h} |\nabla u_h|^2 \,\mathrm{d}x} \leq 1, \tag{2.44}$$

with

$$u_h = \begin{cases} u_h^{\mathcal{D}}, & \text{in } \mathcal{D}_h \\ \\ u_h^0, & \text{in } \Omega \setminus \overline{\mathcal{D}}_h \end{cases} \tag{2.45}$$

where u_h^0 is the harmonic extension of $u_h^{\mathcal{D}}$ into $\Omega_h \setminus \overline{\mathcal{D}}_h$ in the sense (2.36). □

II. Lower Bound for the Generalized Eigenvalues of (2.30)

Before providing the proofs, we introduce one more construction to simplify our consideration below. Because all inclusions are located at distances that are comparable to their sizes, we construct new domains $\hat{\mathcal{D}}^i$, $i \in \{1, \ldots, m\}$, see Fig. 2.2, centered at the centers of the original inclusions $\mathcal{D}^i$, $i \in \{1, \ldots, m\}$, but of sizes much larger of those of $\mathcal{D}^i$ and such that

$$\hat{\mathcal{D}}^i \cap \hat{\mathcal{D}}^j = \emptyset, \quad \text{for} \quad i \neq j.$$

With that, one can see that the problem (2.20) might be partitioned into m independent subproblems, hence, without loss of generality, has only one inclusion, that is, $m = 1$.

Figure 2.2: New domains $\hat{\mathcal{D}}^i$ for our construction of the lower bound of μ

We also recall a few important results from classical PDE theory analogs of which will be used below. Namely, for a given $v \in H^1(\mathcal{D})$ there exists an extension v_0 of v to $\Omega \setminus \overline{\mathcal{D}}$ so that

$$\|v_0\|_{H^1(\Omega\setminus\mathcal{D})} \leq C\|v\|_{H^1(\mathcal{D})}, \quad \text{with} \quad C = C(d, \mathcal{D}, \Omega). \tag{2.46}$$

One can also introduce a number of norms equivalent to (2.40), and, in particular, below we will use

$$\|v\|^2_{\mathcal{D}} := \int\limits_{\mathcal{D}} |\nabla v|^2 dx + \frac{1}{R^2} \int\limits_{\mathcal{D}} v^2 dx, \tag{2.47}$$

where R is the radius of the particle $\mathcal{D} = \mathcal{D}_1$. The scaling factor $1/R^2$ is needed for transforming the classical results from a reference (i.e. unit) disk to the disk of radius $R \neq 1$.

We note that the FEM analog of the extension result of (2.46) for a regular grid was shown in [34], from which it also follows that the constant C of (2.46) is independent of the mesh size $h > 0$. We utilize this observation in our construction below.

Consider $u_h \in V_h$ given by (2.45). Introduce a space $\hat{V}_h = \Big\{ v_h \in V_h : \; v_h = 0 \text{ in } \Omega_h \setminus \overline{\hat{\mathcal{D}}}_h \Big\}$. Similarly to (2.45), define

$$\hat{V}_h \ni \hat{u}_h = \begin{cases} u_h^{\mathcal{D}}, & \text{in } \mathcal{D}_h \\ \\ \hat{u}_h^0, & \text{in } \Omega_h \setminus \overline{\mathcal{D}}_h \end{cases}, \tag{2.48}$$

where $\hat{u}_h^0$ is the harmonic extension of $u_h^{\mathcal{D}}$ into $\hat{\mathcal{D}}_h \setminus \overline{\mathcal{D}}_h$ in the sense (2.36) and $\hat{u}_h^0 = 0$ on $\partial\hat{\mathcal{D}}_h$. Also, by (2.37) we have

$$\int\limits_{\Omega_h \setminus \mathcal{D}_h} |\nabla u_h^0|^2 dx \leq \int\limits_{\Omega_h \setminus \mathcal{D}_h} |\nabla \hat{u}_h^0|^2 dx.$$

Define the matrix

$$\hat{\mathbf{A}} := \begin{bmatrix} \mathrm{A}_{\mathcal{D}\mathcal{D}} & \hat{\mathrm{A}}_{\mathcal{D}0} \\ \hat{\mathrm{A}}_{0\mathcal{D}} & \hat{\mathrm{A}}_{00} \end{bmatrix}$$

by

$$\left(\hat{\mathbf{A}}\overline{v}, \overline{w}\right) = \int\limits_{\Omega_h} \nabla v_h \cdot \nabla w_h dx, \quad \text{where } \overline{v}, \overline{w} \in \mathbb{R}^N, \quad v_h, w_h \in \hat{V}_h.$$

As before, introduce the Schur complement to the block $\hat{\mathrm{A}}_{00}$ of $\hat{\mathbf{A}}$:

$$\hat{\mathrm{S}}_{00} = \mathrm{A}_{\mathcal{D}\mathcal{D}} - \hat{\mathrm{A}}_{\mathcal{D}0}\hat{\mathrm{A}}_{00}^{-1}\hat{\mathrm{A}}_{0\mathcal{D}}, \tag{2.49}$$

and consider a new generalized eigenvalue problem

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}\, \overline{u}_{\mathcal{D}} = \hat{\mu}\hat{\mathrm{S}}_{00}\, \overline{u}_{\mathcal{D}} \quad \text{with} \quad \mathbb{R}^n \ni \overline{u}_{\mathcal{D}} \perp \ker \boldsymbol{\mathcal{B}}_{\mathcal{D}}. \tag{2.50}$$

By (2.42) and (2.43) we have

$$(\mathrm{S}_{00}\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}}) \leq \left(\hat{\mathrm{S}}_{00}\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}}\right) \quad \text{for all} \quad \overline{u}_{\mathcal{D}} \in \mathbb{R}^n. \tag{2.51}$$

Now, we consider a new generalized eigenvalue problem similar to one in (2.33), namely,

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\hat{\mathrm{S}}_{00}^{-1}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\, \overline{\varphi} = \hat{\mu}\, \overline{\varphi}, \qquad \overline{\varphi} \in \mathbb{R}^n. \tag{2.52}$$

We plan to replace $\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}$ in (2.52) with a new symmetric positive-definite matrix $\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}^{1/2}$, given below in (2.55), so that

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\overline{\xi} = \boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}^{1/2}\overline{\xi} = \hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}^{1/2}\boldsymbol{\mathcal{B}}_{\mathcal{D}}^{1/2}\overline{\xi} \quad \text{for all} \quad \mathbb{R}^n \ni \overline{\xi} \perp \ker \boldsymbol{\mathcal{B}}_{\mathcal{D}}, \tag{2.53}$$

with what (2.52) has the same nonzero eigenvalues as the problem

$$\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}^{1/2}\hat{\mathrm{S}}_{00}^{-1}\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}^{1/2}\,\overline{\varphi} = \hat{\mu}\,\overline{\varphi}, \quad \overline{\varphi} \in \mathbb{R}^n. \tag{2.54}$$

For this purpose, we consider the decomposition:

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}} = \mathrm{W}\Lambda\mathrm{W}^T,$$

where $\mathrm{W} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix composed of eigenvectors $\overline{w}_i$, $i \in \{0, 1, \dots, n-1\}$, of

$$\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{w} = \nu\overline{w}, \quad \overline{w} \in \mathbb{R}^n,$$

and

$$\Lambda = \operatorname{diag}\left[\nu_0,\, \nu_1, \dots, \nu_{n-1}\right].$$

Then $\overline{w}_0$ is an eigenvector of $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ corresponding to $\nu_0 = 0$ and

$$\overline{w}_0 = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

To that end, we choose

$$\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}} = \boldsymbol{\mathcal{B}}_{\mathcal{D}} + \beta\,\overline{w}_0 \otimes \overline{w}_0 = \boldsymbol{\mathcal{B}}_{\mathcal{D}} + \beta\,\overline{w}_0\overline{w}_0^T, \tag{2.55}$$

where $\beta > 0$ is some constant parameter chosen below. Note that the matrix $\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}$ is symmetric and positive-definite, and satisfies (2.53). It is trivial to show that $\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}$ given by (2.55) is spectrally equivalent to $\boldsymbol{\mathcal{B}}_{\mathcal{D}} + \beta \mathrm{I}$ for any $\beta > 0$. Also, for quasi-uniform grids, the matrix $h^2\mathrm{I}$ (in 3-dim case, $h^3\mathrm{I}$) is spectrally equivalent to the mass matrix $\mathrm{M}_{\mathcal{D}}$ given by

$$(\mathrm{M}_{\mathcal{D}}\overline{u}, \overline{v}) = \int\limits_{\mathcal{D}_h^1} u_h v_h \, dx, \quad \text{where} \quad \overline{u}, \overline{v} \in \mathbb{R}^{n_1}, \quad u_h, v_h \in V_h^1,$$

see e.g. [31]. This implies there exists a constant $C > 0$ independent of h, such that

$$\left(\hat{\boldsymbol{\mathcal{B}}}_{\mathcal{D}}\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}}\right) \geq C\left(\left(\boldsymbol{\mathcal{B}}_{\mathcal{D}} + \frac{1}{R^2}\mathrm{M}_{\mathcal{D}}\right)\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}}\right), \quad \text{with} \quad \beta = \frac{h^2}{R^2}. \tag{2.56}$$

The choice of the matrix $\boldsymbol{\mathcal{B}}_{\mathcal{D}} + \frac{1}{R^2}\mathrm{M}_{\mathcal{D}}$ for the spectral equivalence was motivated by the fact that the right hand side of (2.56) describes $\|\cdot\|_{\mathcal{D}_h}$-norm (2.47) of the FEM function $u_h^{\mathcal{D}} \in V_h^1$ that corresponds to the vector $\overline{u}_{\mathcal{D}} \in \mathbb{R}^n$.

Now consider $\overline{u} = \begin{bmatrix} \overline{u}_{\mathcal{D}} \\ \overline{u}_0 \end{bmatrix} \in \mathbb{R}^N$ with $\overline{u}_{\mathcal{D}} \in \mathbb{R}^n$, $\overline{u}_{\mathcal{D}} \perp \ker \boldsymbol{\mathcal{B}}_{\mathcal{D}}$, and $\overline{u}_0 \in \mathbb{R}^{n_0}$ satisfying (2.41), and similarly choose $\overline{\hat{u}} = \begin{bmatrix} \overline{u}_{\mathcal{D}} \\ \overline{\hat{u}}_0 \end{bmatrix} \in \mathbb{R}^N$ with $\overline{\hat{u}}_0 \in \mathbb{R}^{n_0}$ satisfying $\hat{\mathrm{A}}_{0\mathcal{D}}\,\overline{u}_{\mathcal{D}} + \hat{\mathrm{A}}_{00}\,\overline{\hat{u}}_0 = 0$, which implies

$$\left(\hat{S}_{00}\overline{u}_{\mathcal{D}}, \overline{u}_{\mathcal{D}}\right) = \left(\hat{\mathbf{A}}\overline{\hat{u}}, \overline{\hat{u}}\right). \tag{2.57}$$

Then

$$\left(\hat{\mathbf{A}}\overline{\hat{u}}, \overline{\hat{u}}\right) = \int\limits_{\Omega_h} |\nabla \hat{u}_h|^2 dx = \int\limits_{\hat{\mathcal{D}}_h \setminus \mathcal{D}_h} |\nabla \hat{u}_h^0|^2 dx + \int\limits_{\mathcal{D}_h} |\nabla u_h^{\mathcal{D}}|^2 dx \leq (C^* + 1)\|u_h^{\mathcal{D}}\|_{\mathcal{D}_h}^2, \tag{2.58}$$

where $\hat{u}_h \in \hat{V}_h$ is the same extension of $u_h^{\mathcal{P}}$ from $\overline{\mathcal{D}}_h$ to $\Omega_h \setminus \overline{\mathcal{D}}_h$ as defined in (2.48). For the inequality of (2.58), we applied the FEM analog of the extension result of (2.46) by [34], that yields that the constant C^* in (2.58) is independent of h.

With all the above, we have the following chain of inequalities:

$$\frac{(\boldsymbol{\mathcal{B}}_{\mathcal{D}}\overline{u}_{\mathcal{D}},\overline{u}_{\mathcal{D}})}{(\mathrm{S}_{00}\overline{u}_{\mathcal{D}},\overline{u}_{\mathcal{D}})} \underset{(2.53),(2.55)}{=} \frac{((\boldsymbol{\mathcal{B}}_{\mathcal{D}}+\beta\,\overline{w}_0\otimes\overline{w}_0)\,\overline{u}_{\mathcal{D}},\overline{u}_{\mathcal{D}})}{(\mathrm{S}_{00}\overline{u}_{\mathcal{D}},\overline{u}_{\mathcal{D}})} \underset{(2.51)}{\geq} \frac{((\boldsymbol{\mathcal{B}}_{\mathcal{D}}+\beta\,\overline{w}_0\otimes\overline{w}_0)\,\overline{u}_{\mathcal{D}},\overline{u}_{\mathcal{D}})}{\left(\hat{\mathrm{S}}_{00}\overline{u}_{\mathcal{D}},\overline{u}_{\mathcal{D}}\right)}$$

$$\underset{(2.57),(2.56)}{\geq} C\frac{\left(\left(\boldsymbol{\mathcal{B}}_{\mathcal{D}}+\frac{1}{R^2}M_{\mathcal{D}}\right)\overline{u}_{\mathcal{D}},\overline{u}_{\mathcal{D}}\right)}{\left(\hat{\mathbf{A}}\overline{\hat{u}},\overline{\hat{u}}\right)} \underset{(2.58)}{\geq} \frac{C\|u_h^{\mathcal{P}}\|^2_{\mathcal{D}_h}}{(C^*+1)\|u_h^{\mathcal{P}}\|^2_{\mathcal{D}_h}} = \frac{C}{(C^*+1)} =: \hat{\mu}_\star, \quad \text{with} \quad \beta = \frac{h^2}{R^2}$$

where $\mu_\star$ is independent of $h > 0$.

From the obtained above bounds, we have (2.35).

□

2.3.4 Notes on Lanczos algorithm with the block-diagonal preconditioner $\boldsymbol{\mathcal{P}}$

The preconditioned Lanczos procedure of minimized iterations can be used for solving algebraic systems with symmetric and positive semidefinite matrices. In this section, we propose a preconditioner for solving (2.18).

The theoretical justification of the usage of a preconditioner (2.27) where the blocks $\mathcal{P}_{\mathrm{A}}$ and $\mathcal{P}_{\mathrm{B}}$ are spectrally equivalent to $\mathbf{A}$ and $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$, respectively, was shown in [21]. With theoretical considerations provided above, in our practical implementation of the generalized Lanczos method of minimized iterations, we use the following block-diagonal preconditioner:

$$\boldsymbol{\mathcal{P}} = \begin{bmatrix} \mathcal{P}_{\mathrm{A}} & 0 \\ 0 & \boldsymbol{\mathcal{B}}_{\mathcal{D}} \end{bmatrix}, \tag{2.59}$$

where one can choose any typical preconditioner $\mathcal{P}_{\mathrm{A}}$ for the symmetric and positive definite matrix $\mathbf{A}$. Define

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{P}}^{\dagger} = \begin{bmatrix} \mathcal{P}_{\mathrm{A}}^{-1} & 0 \\ 0 & [\boldsymbol{\mathcal{B}}_{\mathcal{D}}]^{\dagger} \end{bmatrix}, \tag{2.60}$$

and a new scalar product

$$(\overline{x}, \overline{y})_{\mathcal{H}} := (\boldsymbol{\mathcal{H}}\overline{x}, \overline{y}), \quad \text{for all} \quad \overline{x}, \overline{y} \in \mathbb{R}^{N+n},$$

and consider the preconditioned Lancsoz iterations $\overline{z}^k = \begin{bmatrix} \overline{u}^k \\ \overline{\lambda}^k \end{bmatrix} \in \mathbb{R}^{N+n}$, $k \geq 1$:

$$\overline{z}^k = \overline{z}^{k-1} - \beta_k \overline{y}_k,$$

where

$$\beta_k = \frac{(\boldsymbol{\mathcal{A}}_\varepsilon \overline{z}^{k-1} - \overline{\mathcal{F}}, \boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_k)_{\mathcal{H}}}{(\boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_k, \boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_k)_{\mathcal{H}}}.$$

and

$$y_k = \begin{cases} \boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{A}}_\varepsilon \overline{z}^0 - \overline{\mathcal{F}}), & k = 1 \\ \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_1 - \alpha_2 \overline{y}_1, & k = 2 \\ \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-1} - \alpha_k \overline{y}_{k-1} - \gamma_k \overline{y}_{k-2}, & k > 2, \end{cases}$$

with

$$\alpha_k = \frac{(\boldsymbol{\mathcal{A}}_\varepsilon \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-1}, \boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-1})_{\mathcal{H}}}{(\boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-1}, \boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-1})_{\mathcal{H}}}, \qquad \gamma_k = \frac{(\boldsymbol{\mathcal{A}}_\varepsilon \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-1}, \boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-1})_{\mathcal{H}}}{(\boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-2}, \boldsymbol{\mathcal{A}}_\varepsilon \overline{y}_{k-2})_{\mathcal{H}}}.$$

Here, we recall that the matrix $\boldsymbol{\mathcal{B}}_{\mathcal{D}}$ is singular, however, as evident from the algorithm above one actually never needs to use its pseudo inverse at all. Indeed, this is due to the block-diagonal structure of $\boldsymbol{\mathcal{H}}$ (2.60), and block form of the original matrix $\boldsymbol{\mathcal{A}}_{\varepsilon}$ (2.16)-(2.17).

2.4 Numerical Results

In this section, we use four examples to show the numerical advantages of the Lanczos iterative scheme with the preconditioner $\boldsymbol{\mathcal{P}}$ defined in (2.59) over the existing preconditioned conjugate gradient method.

Our numerical experiments are performed by implementing the described above Lancsoz algorithm for the problem (2.3)-(2.4), where the domain Ω is chosen to be a disk of radius 5 with $m = 37$ identical circular inclusions $\mathcal{D}^i$, $i \in \{1, \dots, m\}$. Inclusions are equally spaced. The function f of the right hand side of (2.3) is chosen to be a constant, $f = 50$.

In the first set of experiments the values of ε_i's of (2.4) are going to be identical in all inclusions and vary from 10^{-1} to 10^{-8}. In the second set of experiments we consider four groups of particles with the same values of ε in each group that vary from 10^{-4} to 10^{-7}. In the third set of experiments we consider the case when all inclusions have different values of ε_i's that vary from 10^{-1} to 10^{-9}. Finally, in the fourth set of experiments we decrease the distance between neighboring inclusions.

The initial guess z^0 is a random vector that was fixed for all experiments. The stopping criteria is the relative error

$$\frac{\|\boldsymbol{\mathcal{A}}_\varepsilon \bar{z}^k - \overline{\mathcal{F}}\|_2}{\|\overline{\mathcal{F}}\|_2}$$

being less than a fixed tolerance constant.

We test our results agains standard `pcg` function of MATLAB® with $\mathcal{P}_{\mathrm{A}} = \mathbf{A}$. The same matrix is also used in the implementation of the described above Lancsoz algorithm. In the following tables **PCG** stands for preconditioned conjugate gradient method by MATLAB® and **PL** stands for preconditioned Lancsoz method previously.

Experiment 1. For the first set of experiments we consider particles $\mathcal{D}^i$ of radius $R = 0.45$ in the disk Ω. This choice makes distance d between neighboring inclusions approximately equal to the radius R of inclusions. The triangular mesh Ω_h has $N = 32,567$ nodes. Tolerance is chosen to be equal to 10^{-4}. This experiment concerns the described problem with parameter ε being the same in each inclusion. Table 2.1 shows the number of iterations corresponding to the different values of ε.

	Values of ε							
	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
PCG	10	20	32	40	56	183	302	776
PL	33	37	37	37	37	37	37	37

Table 2.1: Number of iterations in *Experiment 1*, $N = 32,567$

Based on these results, we first observe that our **PL** method requires fewer iterations as ε goes less than 10^{-4}. We also notice that number of iterations in the

Lancsoz algorithm does not depend on ε.

Figure 2.3: The domain Ω with highly conducting inclusions $\mathcal{D}^i$ of fours groups

Experiment 2. In this experiment we leave radii of the inclusions to be the same, namely, $R = 0.45$. Tolerance is chosen to be 10^{-6}. We now distinguish four groups of particles of different ε's. The first group consists of one inclusion – in the center – with the coefficient $\varepsilon = \varepsilon_1$, whereas the second, third, and fourth groups are comprised of the disks in the second, third, and fourth circular layers of inclusions with coefficients ε_2, ε_3, and ε_4 respectively, see Fig. 2.3 (particles of the same group are indicated with the same color). We perform this type of experiments for three different triangular meshes with the total number of nodes $N = 5,249$, $N = 12,189$ and $N = 32,567$. Tables 2.2, 2.3, and 2.4 below show the number of iterations corresponding to three meshes respectively.

Values of ε					
ε_1	ε_2	ε_3	ε_4	**PCG**	**PL**
10^{-5}	10^{-5}	10^{-4}	10^{-4}	217	39
10^{-5}	10^{-5}	10^{-4}	10^{-3}	208	39
10^{-6}	10^{-5}	10^{-4}	10^{-3}	716	39
10^{-7}	10^{-6}	10^{-5}	10^{-4}	571	39

Table 2.2: Number of iterations in *Experiment 2*, $N = 5,249$

These results yield that **PL** requires much less iterations than the corresponding

Values of ε					
ε_1	ε_2	ε_3	ε_4	**PCG**	**PL**
10^{-5}	10^{-5}	10^{-4}	10^{-4}	116	39
10^{-5}	10^{-5}	10^{-4}	10^{-3}	208	39
10^{-6}	10^{-5}	10^{-4}	10^{-3}	457	39
10^{-7}	10^{-6}	10^{-5}	10^{-4}	454	39

Table 2.3: Number of iterations in *Experiment 2*, $N = 12,189$

Values of ε					
ε_1	ε_2	ε_3	ε_4	**PCG**	**PL**
10^{-5}	10^{-5}	10^{-4}	10^{-4}	217	35
10^{-5}	10^{-5}	10^{-4}	10^{-3}	208	35
10^{-6}	10^{-5}	10^{-4}	10^{-3}	716	35
10^{-7}	10^{-6}	10^{-5}	10^{-4}	571	35

Table 2.4: Number of iterations in *Experiment 2*, $N = 32,567$

PCG with the number of iterations still being independent of both the contrast ε and the mesh size h for **PL**.

Experiment 3. The next point of interest is to assign different value of ε for each of 37 inclusions. The geometrical setup is the same as in *Experiment 2*. The value of ε_i, $i \in \{1, \dots, 37\}$, is randomly assigned to each particle and is chosen from the range of ε's reported in Table 2.5 below. The tolerance is 10^{-6} as above. The triangular mesh Ω_h has $12,189$ nodes. We run **ten** tests for each range of contrasts and obtain the **same number of iterations** in every case, and that number is being reported in Table 2.5.

Range of ε	**PL**
10^{-1} to 10^{-8}	53
10^{-1} to 10^{-3}	53
10^{-7} to 10^{-9}	39

Table 2.5: Number of iterations in *Experiment 3*, $N = 12,189$

We also observe that as the contrast between conductivities in the background domain $\Omega \setminus \overline{\mathcal{D}}$ and the one inside particles $\mathcal{D}_i$, $i \in \{1, \dots, 37\}$, becomes larger our preconditioner demonstrates better convergence, as the third row of Table 2.5 reports. This is expected since the preconditioner constructed above was chosen for the case of absolutely conductive particles. These sets of tests are not compared against the **PCG** due to the large number of considered contrasts that prevent this test to converge in a reasonable amount of time.

Experiment 4. In the next set of experiments we intend to test how well our algorithm performs if the distance between particles decreases. Recall that the assumption made for our procedure to work is that the interparticle distance d is of order of the particles' radius R. With that, we take the same setup as in *Experiment 2* and decrease the distance between particles by making radius of each disk larger. We set $R = 0.56$ obtaining that the radius of each inclusion is now twice larger than the distance d, and also consider $R = 0.59$ so that the radius of an inclusion is three times larger than d. The triangular mesh Ω_h has $N = 6,329$ and $N = 6,497$ nodes, respectively. The tolerance is chosen to be 10^{-6}. Tables 2.6 and 2.7 show the number of iterations in each case.

Values of ε					
ε_1	ε_2	ε_3	ε_4	**PCG**	**PL**
10^{-5}	10^{-5}	10^{-4}	10^{-4}	799	61
10^{-7}	10^{-6}	10^{-5}	10^{-4}	859	61

Table 2.6: Number of iterations in *Experiment 4*, $N = 6,329$

Values of ε					
ε_1	ε_2	ε_3	ε_4	**PCG**	**PL**
10^{-5}	10^{-5}	10^{-4}	10^{-4}	311	73
10^{-7}	10^{-6}	10^{-5}	10^{-4}	890	73

Table 2.7: Number of iterations in *Experiment 4*, $N = 6,497$

Here we observe that number of iterations increases for both **PCG** and **PL**, while this number still remains independent of ε for **PL**.

We then continue to decrease the distance d, and set $R = 0.62$ that is approximately four times larger than the distance between two neighboring inclusions d. Choose the same tolerance 10^{-6} as above, and the triangular mesh Ω_h of $N = 6,699$ nodes, and we observed that our **PL** method does not reach the desired tolerance in $1,128$ iterations, that confirms our expectations. Further research is needed to develop novel techniques for the case of closely spaced particles. This fact inspired research presented in Chapter 3.

2.5 Conclusions

This chapter focuses on a construction of the robust preconditioner (2.59) for the Lancsoz iterative scheme that can be used in order to solve high-contrast PDEs of the type (2.3)-(2.4). A typical FEM discretization yields an ill-conditioning matrix when the contrast in σ becomes high (i.e., $\varepsilon \ll 1$). We propose a saddle point formulation of the given problem with the symmetric and indefinite matrix and consequently construct the corresponding preconditioner that yields a robust numerical approximation of (2.3)-(2.4). The main feature of this novel and elegant approach is that we precondition the given linear system with a *symmetric and indefinite matrix*. Our numerical results have shown the effectiveness of the proposed preconditioner for these type of problems, and demonstrated convergence of the constructed **PL** scheme independently on the contrast ε and mesh size h.

Chapter 3

Efficient numerical scheme for high-contrast problems modeling highly dense composites

3.1 Introduction

This chapter concerns the case when injections are almost touching each other. This feature leads to *rapidly oscillatory* coefficients meaning that values alternate on very small length scales. This leads to challenges in numerical methods due to small mesh size needed in the gaps between injections.

We use two types of discretization for continuous problems: *numerical* and *structural* ones. The examples of numerical discretization are finite element, finite difference and other methods where the mesh size is adjustable depending on the desired precision. The structural discretization is based on physical features of the considered domains. An example of such a discretization results in a finite-dimensional discrete network, which is a graph whose edges and nodes match to the physical objects. In this case the discretization scale is determined by the natural size of inhomogeneities and distances between them.

The main goal of this chapter is a numerical treatment of problems associated with high-contrast composite materials with complex geometry. The *novel* idea is to take advantage of properties of structured materials to build new numerically efficient schemes. In particular, a focus is on the *domain decomposition* methods for the problems, which describe media whose parts have high-contrast constituents. The key step here is to split a large domain into subdomains in a natural way to deal separately with homogeneous and high-contrast parts. A coupled problem is obtained where subdomains are bridged though the interface. To avoid solving the problem in a nonhomogeneous part we use the discrete approximation of the Dirichlet-to-Neumann (DtN) map that was developed in [11]. Then, we build an effective iterative method based on the resulted partition.

The rest of this chapter is organized as follows. In Section 3.2 the mathematical problem formulation is presented and numerical algorithm is described. Section 3.3 discusses the results on DtN map, and numerical results of the proposed scheme are given in Section 3.4. Conclusions are presented in Section 3.5.

3.2 Problem formulation and domain decomposition method

Consider an open, a bounded domain $\Omega \subset \mathbb{R}^2$ with piece-wise smooth boundary $\partial\Omega$, that contains $m \geq 1$ subdomains $\mathcal{D}^i$, which are located at distances much smaller than their sizes from one another. For simplicity, the assumption is that Ω and $\mathcal{D}^i$ are polygons. The union of $\mathcal{D}^i$ is denoted by $\mathcal{D}$. In the domain Ω we consider the following problem:

$$\begin{cases} -\triangle u &= f, \qquad x \in \Omega \setminus \overline{\mathcal{D}}, \\ u &= \text{const}, \; x \in \partial\mathcal{D}^i, \; i \in \{1, \dots, m\}, \\ \int\limits_{\partial\mathcal{D}^i} \nabla u \cdot \mathbf{n}_i \; ds &= 0, \qquad i \in \{1, \dots, m\}, \\ u &= 0, \qquad x \in \partial\Omega, \end{cases} \tag{3.1}$$

where $\mathbf{n}_i$ is the outer unit normal to the surface $\partial\mathcal{D}^i$. If $u \in H_0^1(\Omega \setminus \overline{\mathcal{D}})$ is an electric potential that attains constant values on the inclusions $\mathcal{D}^i$ and these constants are not known a priori so that they are unknowns of the problem (3.1) together with u.

Problem (3.1) describes the case of infinitely conducting injections. With slight abuse of terminology we refer to this problem as high-contrast one as it is commonly used in literature.

The assumption is to split the domain $\Omega := \Omega_1 \cup \Omega_2 \cup \Gamma$ in a way that high-contrast part is separated in Ω_1, while subdomain Ω_2 has constant conductivity equal to 1, see Fig. 3.1. Here Γ denotes the interface between the two subdomains. The goal is to take advantage of this partition to build an effective domain decomposition algorithm.

Figure 3.1: The domain Ω with highly conducting inclusions $\mathcal{D}^i$ concentrated in one region of the domain

3.2.1 Discretization of continuous problem

A triangulation Ω_h of domain Ω is considered, the nodes of triangulation are required to match the interface Γ. Classical FEM discretization of (3.1) with piecewise linear functions results in a linear system

$$A\overline{u} = \overline{f}, \tag{3.2}$$

with a symmetric, positive definite matrix A.

The degrees of freedom are split into the degrees belonging to Ω_1, and to Ω_2, and those belonging to the interface Γ. With that partition system (3.2) can be written in a block form

$$\begin{pmatrix} A_{11} & 0 & A_{1\Gamma} \\ 0 & A_{22} & A_{2\Gamma} \\ A_{1\Gamma}^T & A_{2\Gamma}^T & A_{\Gamma\Gamma} \end{pmatrix} \begin{pmatrix} \overline{u}_1 \\ \overline{u}_2 \\ \overline{u}_\Gamma \end{pmatrix} = \begin{pmatrix} \overline{f}_1 \\ \overline{f}_2 \\ \overline{f}_\Gamma \end{pmatrix}. \tag{3.3}$$

The stiffness matrix and load vector can be obtained by assembling the corresponding components contributed by the subdomains, denoting

$$A^{(i)} = \begin{pmatrix} A_{ii} & A_{i\Gamma} \\ A_{i\Gamma}^T & A_{\Gamma\Gamma}^{(i)} \end{pmatrix}, \qquad \overline{f}^{(i)} = \begin{pmatrix} \overline{f}_i \\ \overline{f}_\Gamma^{(i)} \end{pmatrix}, \qquad i = 1, 2,$$

then

$$A_{\Gamma\Gamma} = A_{\Gamma\Gamma}^{(1)} + A_{\Gamma\Gamma}^{(2)}, \qquad \overline{f}_\Gamma = \overline{f}_\Gamma^{(1)} + \overline{f}_\Gamma^{(2)}.$$

3.2.2 Schur complement system

The usual first step of many iterative domain decomposition methods is the elimination of interior unknowns $\overline{u}_1$ and $\overline{u}_2$, which reduces the system (3.2) with (3.3) to the *Schur complement system* for $\overline{u}_\Gamma$

$$S\overline{u}_\Gamma = \overline{g}_\Gamma, \tag{3.4}$$

where

$$S = S^{(1)} + S^{(2)}, \qquad \overline{g}_\Gamma = \overline{g}_\Gamma^{(1)} + \overline{g}_\Gamma^{(2)},$$

with

$$S^{(i)} = A_{\Gamma\Gamma}^{(i)} - A_{i\Gamma}^T A_{ii}^{-1} A_{i\Gamma}, \qquad \overline{g}_\Gamma^{(i)} = \overline{f}_\Gamma^{(i)} - A_{i\Gamma}^T A_{ii}^{-1} \overline{f}_i, \qquad i = 1, 2.$$

Matrix S is usually referred to as Schur complement to the unknowns on Γ. Once system (3.4) is solved, the internal components could be found from

$$\overline{u}_i = A_{ii}^{-1}\overline{f}_i - A_{ii}^{-1} A_{i\Gamma}\overline{u}_\Gamma, \qquad i = 1, 2. \tag{3.5}$$

Next, we derive one more auxiliary approximation that we use later in the description of the method. Consider local Neumann problem in Ω_1

$$\begin{pmatrix} A_{11} & A_{1\Gamma} \\ A_{1\Gamma}^T & A_{\Gamma\Gamma}^{(1)} \end{pmatrix} \begin{pmatrix} \overline{u}_1 \\ \overline{u}_\Gamma \end{pmatrix} = \begin{pmatrix} \overline{f}_1 \\ \overline{f}_\Gamma^{(1)} + \overline{\lambda}_\Gamma^{(1)} \end{pmatrix},$$

with $\overline{\lambda}_\Gamma^{(1)}$ being an approximation for weak normal derivative on the interface Γ. With the definition of $S^{(1)}$ as previously introduced the formula becomes

$$\overline{\lambda}_\Gamma^{(1)} = S^{(1)}\overline{u}_\Gamma - \overline{g}_\Gamma^{(1)}. \tag{3.6}$$

3.2.3 The Dirichlet-Neumann algorithm

The classical Dirichlet-Neumann domain decomposition method was described in [31]. In this research the algorithm is adjusted to the case when Ω_1 is embedded in Ω_2.

The iteration step consists of two fractional steps: Dirichlet problem in subdomain Ω_1 and mixed Neumann-Dirichlet problem in subdomain Ω_2 with a Neumann condition on the interface as determined by the solution Ω_1 obtained in the previous step and with Dirichlet data on $\partial\Omega_2 \setminus \Gamma$. The next iterate is chosen as a linear combination of trace of the solution in Ω_2 and data on the interface obtained on previous iteration with a suitably chosen relaxation parameter $\theta \in (0, \theta_{max})$ to ensure convergence of the method. In terms of differential operators the above algorithm looks as follows:

$$(D): \quad \begin{cases} -\triangle u_1^{n+1/2} = f & \text{in } \Omega_1, \\ u_1^{n+1/2} = u_\Gamma^n & \text{on } \Gamma, \end{cases}$$

$$(D+N): \quad \begin{cases} -\triangle u_2^{n+1} = f & \text{in } \Omega_2, \\ u_2^{n+1} = 0 & \text{on } \partial\Omega_2 \setminus \Gamma, \\ \frac{\partial u_2^{n+1}}{\partial n_2} = -\frac{\partial u_1^{n+1/2}}{\partial n_1} & \text{on } \Gamma, \end{cases}$$

$$u_\Gamma^{n+1} = \theta u_2^{n+1} + (1-\theta) u_\Gamma^n \text{ on } \Gamma.$$

With use of the approximations given above, the corresponding iteration for the discrete problem is as follows

$$(D): \quad A_{11}\overline{u}_1^{n+1/2} + A_{1\Gamma}\overline{u}_\Gamma^n = \overline{f}_1,$$

$$(D+N): \quad \begin{pmatrix} A_{\Gamma\Gamma}^{(2)} & A_{\Gamma 2}^T \\ A_{\Gamma 2} & A_{22} \end{pmatrix} \begin{pmatrix} \overline{u}_\Gamma^{n+1} \\ \overline{u}_2^{n+1} \end{pmatrix} = \begin{pmatrix} \overline{f}_\Gamma^{(2)} - \overline{\lambda}_\Gamma^{(1)n+1/2} \\ \overline{f}_2 \end{pmatrix},$$

$$\overline{u}_\Gamma^{n+1} = \theta \overline{u}_2^{n+1} + (1-\theta)\overline{u}_\Gamma^n \text{ on } \Gamma.$$

Next, elimination of $\overline{u}_1^{n+1/2}$ and $\overline{u}_2^{n+1}$ yields the following equation:

$$S^{(2)}\left(\overline{u}_\Gamma^{n+1} - \overline{u}_\Gamma^{n}\right) = \theta\left(\overline{g}_\Gamma - S\overline{u}_\Gamma^{n}\right). \tag{3.7}$$

This shows that the Dirichlet-Neumann algorithm is a preconditioned Richardson iteration (3.7) for the Schur complement system (3.4), with the preconditioner $S^{(2)}$.

We remark that $S^{(1)}$ is spectrally equivalent to $S^{(2)}$ due to existence of discrete harmonic extensions from the interface into the subdomains Ω_1 and Ω_2. Therefore, the condition number of $S^{(2)-1}S$ is ensured to stay uniformly bounded.

3.2.4 Challenges of the problem with a densely packed subdomain Ω_1

The key feature of the problem, which is a densely packed subdomain Ω_1, requires a very fine mesh in the gaps between the inclusions and causes a large size of the matrix A_{11}. The condition number of that matrix worsens as $1/h^2$, where h is a size of the mesh. As a result it is impossible to contract matrix $S^{(1)}$, as this requires inversion of the block A_{11}.

In this research we propose an approximation of $S^{(1)}$ that could be used to make the described algorithm applicable in practice. The replacement we suggest uses the approximation of the DtN map by a discrete one, as introduced in the next section.

3.3 Introduction to the discrete DtN map

3.3.1 Asymptotic approximation of the DtN map

In this subsection we review the results obtained in [11].

Continuous DtN map of an elliptic PDE maps the boundary trace of the solution to its normal derivative at the boundary Γ, i.e. $\Lambda : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$. This paper discusses the DtN map of the following equation

$$-\nabla \cdot [\sigma(\mathbf{x})\nabla u(\mathbf{x})] = 0, \quad \mathbf{x} \in \Omega_1,$$

was studied, where Ω_1 is a bounded, simply connected domain in $\mathbb{R}^d$. Coefficient $\sigma(\mathbf{x})$ has high contrast and varies rapidly within the domain. The map Λ, defined by

$$\Lambda\psi(\mathbf{x}) = \sigma(\mathbf{x})\nabla u(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}), \quad \mathbf{x} \in \Gamma,$$

where $\mathbf{n}(\mathbf{x})$ is the outer normal vector to Γ, is self-adjoint. Consequently the map is determined by its quadratic form

$$\langle \psi, \Lambda\psi \rangle = \int\limits_{\Gamma} \psi(\mathbf{x})\Lambda\psi(\mathbf{x})\ \mathrm{d}s(\mathbf{x}), \quad \forall\psi \in H^{1/2}(\Gamma).$$

Through integration by parts the map can be related to the energy

$$\langle \psi, \Lambda\psi \rangle = \int\limits_{\Omega_1} \sigma(\mathbf{x})|\nabla u(\mathbf{x})|^2\ \mathrm{d}\mathbf{x}. \tag{3.8}$$

It was shown that Λ can be approximated by the matrix valued DtN map.

3.3.1.1 Problem formulation and the main result

The following setup was considered: domain Ω_1 is a disk of radius L packed with m perfectly conductive inclusions $\mathcal{D}_i$. Each inclusion is a disk of radius $R \ll L$. There is one more important length scale in this problem: a typical distance between inclusions $\delta \ll R$. To define δ first we need to specify what it means for two inclusions to be neighbors.

Let $\mathcal{V}_i$ be the Voronoi cell constructed for inclusion $\mathcal{D}_i$, $i = 1, \ldots, m$

$$\mathcal{V}_i = \{\mathbf{x} \in \Omega_1 \text{ such that } |\mathbf{x} - \mathbf{x}_i| \le |\mathbf{x} - \mathbf{x}_j| \; \forall j = 1, \ldots, m, j \neq i\}.$$

Note that each cell is a convex polygon. The inclusions $\mathcal{D}_i$ and $\mathcal{D}_j$ are said to be neighbors if their cells share an edge. For each inclusion $\mathcal{D}_i$ denote a set of indices of the neighboring inclusions

$$\mathcal{M}_i = \{j \in \{1, \ldots, m\}, \; \mathcal{D}_j \text{ is a neighbor to } \mathcal{D}_i\}.$$

Let the typical distance between two neighboring inclusion be defined as

$$\delta_{ij} = \operatorname{dist}\{\mathcal{D}_i, \mathcal{D}_j\}, \quad \delta_{ij} \ll R,$$

where $i = 1, \ldots, m$ and $j \in \mathcal{M}_i$.

Similarly, the inclusion $\mathcal{D}_i$ neighbors the boundary if $\mathcal{V}_i \cap \Gamma \neq \emptyset$ and define the typical distance between inclusion and the boundary as

$$\delta_i = \operatorname{dist}\{\mathcal{D}_i, \Gamma\}, \quad \delta_i \ll R.$$

The inclusions are numbered starting with those who neighbors the boundary and going counter clockwise. Hence, inclusion $\mathcal{D}_i$ neighbors the boundary if $i = 1, \ldots, m^\Gamma$ and $\mathcal{D}_i$ is an interior inclusion if $i = m^\Gamma + 1, \ldots, m$.

Then the conductivity function in the necks between inclusions is defined by

$$\sigma_{ij} = \pi\sqrt{\frac{R}{\delta_{ij}}},\ i = 1, \dots, m,\ j \in \mathcal{M}_i,$$

$$\sigma_i = \pi\sqrt{\frac{2R}{\delta_i}},\ i = 1, \dots, m^\Gamma.$$

Since Γ is a circle we parametrize it by angle $\theta \in [0, 2\pi]$ and present a boundary potential ψ as a truncated Fourier series

$$\psi(\theta) = \sum_{k=0}^{K} a_k \cos k\theta + b_k \sin k\theta =: \sum_{k=0}^{K} \psi_k(\theta). \tag{3.9}$$

The main result is given in the theorem below.

Theorem 2. *For a potential ψ of the form (3.9) we have that*

$$\langle \psi, \Lambda\psi \rangle = 2\left[E^{net}(\Psi(\psi)) + \frac{1}{2}\langle \psi, \Lambda_0\psi \rangle + \mathcal{R}(\psi)\right][1 + o(1)]. \tag{3.10}$$

The first term is the discrete energy $E^{net}(\Psi(\psi))$ of the resistor network given by

$$E^{net}(\Psi) = \min_{\mathcal{U}\in\mathbb{R}^m}\left\{\sum_{i=1}^{m^\Gamma} \frac{\sigma_i}{2}\left[\mathcal{U}_i - \Psi_i\right]^2 + \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{M}_j} \frac{\sigma_{ij}}{2}(\mathcal{U}_i - \mathcal{U}_j)^2\right\}, \tag{3.11}$$

with vector $\Psi = (\Psi_1, \dots, \Psi_{m^\Gamma})^T$ of boundary potentials defined by

$$\Psi_i(\psi) = \sum_{k=0}^{K} \psi_k(\theta_i) e^{-\frac{k\sqrt{2R\delta_i}}{L}}, \quad i = 1, \dots, m^\Gamma,$$

where θ_i, $i = 1, \dots, m^\Gamma$ is the closest points on Γ to the inclusion $\mathcal{D}_i$.

The second term is the quadratic form of the DtN map Λ_0 of the reference medium, with uniform conductivity $\sigma = 1$. For boundary potential given by a single Fourier mode $\psi = \cos kx$ this term is defined by

$$\langle \psi, \Lambda_0\psi \rangle = k\pi.$$

The last term $\mathcal{R}$ is given by

$$\mathcal{R} = \sum_{i=1}^{m^\Gamma} \sum_{k,m=0}^{K} e^{-|k-m|\frac{\sqrt{2R\delta_i}}{L}} \mathcal{R}_{i,k\wedge m} \left\{ (a_k a_m + b_k b_m) \cos\left[(k-m)\theta_i\right] \right.$$
$$\left. + (b_k a_m - a_k b_m) \sin\left[(k-m)\theta_i\right], \right.$$

where $k \wedge m = \min\{k, m\}$, and $\mathcal{R}_{i,k}$ is defined by

$$\mathcal{R}_{i,k} = \frac{\sigma_i}{4} \left[\sqrt{\frac{2k\delta_i}{\pi L}} \boldsymbol{Li}_{\mathbf{1/2}} \left(e^{-\frac{2k\delta_i}{L}} \right) - e^{-\frac{2k\sqrt{2R\delta_i}}{L}} \right], \tag{3.12}$$

in terms of the polylogarithm function $\boldsymbol{Li}_{\mathbf{1/2}}$.

Proof of the theorem is given in [11].

3.3.2 Construction of discrete DtN map

In this section we use the result of Theorem 2 to construct a matrix-valued DtN map Λ. We discretize the boundary Γ with M points $\theta_i = \frac{(i-1)2\pi}{M}, i = 1, \dots, M$. Then the discrete DtN map $\Lambda \in \mathbb{R}^{M \times M}$ is a symmetric positive definite matrix. To find the entries of this matrix we use the approximation to quadratic form (3.10) and auxiliary fact

$$\langle \overline{\hat{\varphi}}, \Lambda \overline{\tilde{\varphi}} \rangle = \frac{1}{4} \left[\langle \overline{\hat{\varphi}} + \overline{\tilde{\varphi}}, \Lambda(\overline{\hat{\varphi}} + \overline{\tilde{\varphi}} \rangle - \langle \overline{\hat{\varphi}} - \overline{\tilde{\varphi}}, \Lambda(\overline{\hat{\varphi}} - \overline{\tilde{\varphi}} \rangle \right], \tag{3.13}$$

to construct a system of equations

$$\Phi^T \Lambda \Phi = \Upsilon, \tag{3.14}$$

where $\Phi = [\overline{\varphi}_1, \dots, \overline{\varphi}_M]$ form a basis in $\mathbb{R}^M$. Hereafter the 'bar' indicates vectors in $\mathbb{R}^M$. A set of M linearly independent functions $\{\varphi_1, \dots, \varphi_M\}$ is selected to be

$$\left\{ \frac{1}{\sqrt{\pi}}; \frac{\cos\theta}{\sqrt{\pi}}; \dots; \frac{\cos\left(\frac{M}{2} - 1\right)\theta}{\sqrt{\pi}}; \frac{\sin\theta}{\sqrt{\pi}}; \dots; \frac{\sin\frac{M}{2}\theta}{\sqrt{\pi}} \right\},$$

Thus Φ is given by

$$\Phi = \begin{pmatrix} \frac{1}{\sqrt{\pi}} & \frac{\cos\theta_1}{\sqrt{\pi}} & \dots & \frac{\cos\left(\frac{M}{2}-1\right)\theta_1}{\sqrt{\pi}} & \frac{\sin\theta_1}{\sqrt{\pi}} & \dots & \frac{\sin\frac{M}{2}\theta_1}{\sqrt{\pi}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{\pi}} & \frac{\cos\theta_M}{\sqrt{\pi}} & \dots & \frac{\cos\left(\frac{M}{2}-1\right)\theta_M}{\sqrt{\pi}} & \frac{\sin\theta_M}{\sqrt{\pi}} & \dots & \frac{\sin\frac{M}{2}\theta_M}{\sqrt{\pi}} \end{pmatrix}.$$

Entries of the right hand side of (3.14) are given by

$$\begin{aligned} \upsilon_{ij} =& \frac{1}{2}\left\{E^{net}(\Psi(\varphi_i+\varphi_j)) + \frac{1}{2}\langle\overline{\varphi}_i+\overline{\varphi}_j, \Lambda_0(\overline{\varphi}_i+\overline{\varphi}_j)\rangle + \mathcal{R}(\varphi_i+\varphi_j)\right\} \\ &-\frac{1}{2}\left\{E^{net}(\Psi(\varphi_i-\varphi_j) + \frac{1}{2}\langle\overline{\varphi}_i-\overline{\varphi}_j, \Lambda_0(\overline{\varphi}-\overline{\varphi}_j)\rangle + \mathcal{R}(\overline{\varphi}_i-\overline{\varphi}_j)\right\}. \end{aligned}$$

We split υ_{ij} into a sum of three terms to treat them separately

$$\begin{aligned} \upsilon_{ij}^{(1)} &= \frac{1}{2}\left\{E^{net}(\Psi(\varphi_i+\varphi_j)) - E^{net}(\Psi(\varphi_i-\varphi_j)\right\}, \\ \upsilon_{ij}^{(2)} &= \frac{1}{4}\left\{\langle\overline{\varphi}_i+\overline{\varphi}_j, \Lambda_0(\overline{\varphi}_i+\overline{\varphi}_j)\rangle - \langle\overline{\varphi}_i-\overline{\varphi}_j, \Lambda_0(\overline{\varphi}_i-\overline{\varphi}_j)\rangle\right\}, \\ \upsilon_{ij}^{(3)} &= \frac{1}{2}\left\{\mathcal{R}(\varphi_i+\varphi_j) - \mathcal{R}(\varphi_i-\varphi_j)\right\}. \end{aligned}$$

Later we show how much influence each term has depending on k for the boundary function given by a single mode $\varphi = \cos k\theta$.

Once matrix Υ is constructed we obtain DtN map Λ by $\Lambda = (\Phi^T)^{-1}\Upsilon\Phi^{-1}$. A detailed description in the following subsections shows how to compute entries $\upsilon_{ij}^{(1)}, \upsilon_{ij}^{(2)}$ and $\upsilon_{ij}^{(3)}$.

3.3.2.1 Compute entries $\upsilon_{ij}^{(1)}$

The discrete energy from a boundary function given by a single Fourier mode is given by (3.11). To use this formula for the boundary functions given by a sum or a difference of two Fourier modes we need two lemmas.

Lemma 6. *A vector $\overline{\mathcal{U}}$ is the minimizer of* (3.11) *iff $\overline{\mathcal{U}}$ if and only if $\overline{\mathcal{U}}$ is the solution of equation $M\overline{\mathcal{U}} = \overline{\mathcal{P}}$, where entries of M are defined by*

$$M_{ij} = \begin{cases} \sigma_i + \sum\limits_{l\in\mathcal{M}_i} \sigma_{il}, & \textit{if } i \in \{1,\dots,m^\Gamma\} \textit{ and } i = j, \\ \sum\limits_{l\in\mathcal{M}_i} \sigma_{il}, & \textit{if } i \in \{m^\Gamma+1,\dots,m\} \textit{ and } i = j, \\ -\sigma_{ij}, & \textit{if } i \in \{1,\dots,m\} \textit{ and } j \in \mathcal{M}_i, \\ 0, & \textit{if } i \in \{1,\dots,m\} \textit{ and } j \notin \mathcal{M}_i, \end{cases} \tag{3.15}$$

and entries of $\overline{\mathcal{P}}$ are defined by

$$\mathcal{P}_i = \begin{cases} \sigma_i\Psi_i, & \textit{if } i \in \{1,\dots,m^\Gamma\}, \\ 0, & \textit{if } i \in \{m^\Gamma+1,\dots,m\}. \end{cases} \tag{3.16}$$

Proof. ($\Rightarrow$)To minimize (3.11) take partial derivatives $\forall i$, $i = 1,\dots,m$.

$$\frac{\partial E^{\text{net}}}{\partial \mathcal{U}_i} = \begin{cases} \sigma_i[\mathcal{U}_i - \Psi_i] + \sum\limits_{j\in\mathcal{M}_i} \sigma_{ij}[\mathcal{U}_i - \mathcal{U}_j], & \text{if } i \in \{1,\dots,m^\Gamma\}, \\ \sum\limits_{j\in\mathcal{N}_i} \sigma_{ij}[\mathcal{U}_i - \mathcal{U}_j], & \text{if } i \in \{m^\Gamma+1,\dots,m\}. \end{cases}$$

Relation $\frac{\partial E^{\text{net}}}{\partial \mathcal{U}_i} = 0$ gives rise to the following equations

$$\begin{cases} \left(\sigma_i + \sum\limits_{j\in\mathcal{M}_i} \sigma_{ij}\right)\mathcal{U}_i - \sum\limits_{j\in\mathcal{M}_i} \sigma_{ij}\mathcal{U}_j = \sigma_i\Psi_i, & \text{if } i \in \{1,\dots,m^\Gamma\}, \\ \sum\limits_{j\in\mathcal{M}_i} \sigma_{ij}\mathcal{U}_i - \sum\limits_{j\in\mathcal{M}_i} \sigma_{ij}\mathcal{U}_j = 0, & \text{if } i \in \{m^\Gamma+1,\dots,m\}. \end{cases}$$

The last equations yield (3.15) and (3.16).

($\Leftarrow$) To show that this is sufficient condition, matrix R of second derivatives $\frac{\partial^2 E^{\text{net}}}{\partial \mathcal{U}_i \partial \mathcal{U}_j}$ must be positive definite. All inclusions are assumed adjacent to the boundary and each other. If not, some σ_i and σ_{ij} would be zero, which does not affect positive definiteness of the matrix. Entries of the matrix R are the following

$$R_{ij} = \begin{cases} \sigma_i + \sum\limits_{l=1}^{m} \sigma_{il}, & \text{if } i = j, \\ -\sigma_{ij}, & \text{if } i \neq j. \end{cases}$$

We show that this matrix is positive definite by induction over the number of inclusions. Consider case of two inclusions.

$$\begin{pmatrix} 1 & \alpha \end{pmatrix} \begin{pmatrix} \sigma_1 + \sigma_{11} + \sigma_{12} & -\sigma_{12} \\ -\sigma_{12} & \sigma_2 + \sigma_{12} + \sigma_{22} \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \end{pmatrix}$$

$$= \sigma_1 + \sigma_{11} + \sigma_{12} - 2\alpha\sigma_{12} + \alpha^2(\sigma_2 + \sigma_{12} + \sigma_{22}).$$

The minimizer of this quadratic form is

$$\alpha_{\min} = \frac{\sigma_{12}}{\sigma_2 + \sigma_{12} + \sigma_{22}},$$

and the minimum of quadratic form is

$$\frac{\sigma_1\sigma_2 + \sigma_1\sigma_{12} + \sigma_1\sigma_{22} + \sigma_{11}\sigma_2 + \sigma_{11}\sigma_{12} + \sigma_{11}\sigma_{22} + \sigma_{12}\sigma_2 + \sigma_{12}\sigma_{22}}{\sigma_2 + \sigma_{12} + \sigma_{22}} > 0.$$

So quadratic form is positive definite for any $\alpha \in \mathbb{R}$ in case of two inclusions. Assume that matrix is positive definite for l inclusions, thus $\alpha^T R \alpha > 0$, $\forall \alpha \in \mathbb{R}^l$ or

$$\sum_{j=1}^{l} \alpha_j \sum_{i=1}^{l} \alpha_i R_{ij} > 0.$$

Consider the case with $l+1$ inclusions

$$\begin{pmatrix} \alpha^T & 1 \end{pmatrix} \begin{pmatrix} R^* & -\beta \\ -\beta^T & \sigma_{l+1} + \sum\limits_{i=1}^{l+1} \sigma_{i\,l+1} \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \end{pmatrix}, \tag{3.17}$$

where $\beta = (\sigma_{1\,l+1}, \dots, \sigma_{l\,l+1}) \in \mathbb{R}^l$ and

$$R^*_{ij} = \begin{cases} R_{ij}, & \text{if } i \neq j, \\ R_{ij} + \sigma_{i\,l+1}, & \text{if } i = j. \end{cases}$$

Expanding (3.17) we get

$$\begin{aligned}
&\sum_{j=1}^{l} \alpha_j \sum_{i=1}^{l} \alpha_i R^*_{ij} - 2\sum_{i=1}^{l} \alpha_i \beta_i + \sigma_{l+1} + \sum_{i=1}^{l+1} \sigma_{i\,l+1} \\
&= \sum_{j=1}^{l} \alpha_j \left(\sum_{i=1}^{l} \alpha_i R_{ij} + \alpha_j \sigma_{j\,l+1} \right) - 2\sum_{i=1}^{l} \alpha_i \sigma_{i\,l+1} + \sigma_{l+1} + \sum_{i=1}^{l+1} \sigma_{i\,l+1} \\
&= \sum_{j=1}^{l} \alpha_j \sum_{i=1}^{l} \alpha_i R_{ij} + \sum_{i=1}^{l} \alpha_i^2 \sigma_{i\,l+1} - 2\sum_{i=1}^{l} \alpha_i \sigma_{i\,l+1} + \sigma_{l+1} + \sigma_{l+1\,l+1} + \sum_{i=1}^{l} \sigma_{i\,l+1} \\
&= \sum_{j=1}^{l} \alpha_j \sum_{i=1}^{l} \alpha_i R_{ij} + \sigma_{l+1}\sigma_{l+1\,l+1} + \sum_{i=1}^{l} \sigma_{i\,l+1}(\alpha_i^2 - 2\alpha_i + 1).
\end{aligned}$$

Since $\alpha_i^2 - 2\alpha_i + 1 \geq 0$ and $\sum_{j=1}^{l} \alpha_j \sum_{i=1}^{l} \alpha_i R_{ij} > 0$ by the l case, we have (3.17) is positive $\forall \alpha \in \mathbb{R}^l$. Thus matrix R is positive definite for any number of inclusions and $\overline{\mathcal{U}}$ is indeed the minimizer of (3.11). □

Lemma 7. *If the boundary potential* Ψ *in* (3.11) *is a sum of two terms* $\Psi = \Psi^{(1)} + \Psi^{(2)}$ *then the minimizer of* (3.11) *is a sum of two terms* $\overline{\mathcal{U}} = \overline{\mathcal{U}}^{(1)} + \overline{\mathcal{U}}^{(2)}$ *where* $\overline{\mathcal{U}}^{(i)}$ *is a minimizer of* (3.11) *with boundary potential* $\Psi^{(i)}$.

Proof. By Lemma 6 minimizers of (3.11) with boundary potentials $\Psi^{(1)}$ and $\Psi^{(2)}$ satisfy $M\overline{\mathcal{U}}^{(1)} = \overline{\mathcal{P}}^{(1)}$ and $M\overline{\mathcal{U}}^{(2)} = \overline{\mathcal{P}}^{(2)}$ respectively. So $M\overline{\mathcal{U}}^{(1)} + M\overline{\mathcal{U}}^{(2)} = \overline{\mathcal{P}}^{(1)} + \overline{\mathcal{P}}^{(2)}$. And hence $\overline{\mathcal{U}}^{(1)} + \overline{\mathcal{U}}^{(2)}$ is a minimizer of (3.11) with boundary potential $\Psi^{(1)} + \Psi^{(2)}$. □

With the use of two lemmas above and formula (3.11) we conclude that

$$\upsilon_{ij}^{(1)} = \frac{1}{2}\Bigg\{ \sum_{l=1}^{m^\Gamma} \frac{\sigma_l}{2} \left[(\mathcal{U}_l^i + \mathcal{U}_l^j) - (\Psi_l^i + \Psi_l^j)\right]^2 + \frac{1}{2}\sum_{l=1}^{m} \sum_{g\in\mathcal{M}_l} \frac{\sigma_{lg}}{2} \left[(\mathcal{U}_l^i + \mathcal{U}_l^j) - (\mathcal{U}_g^i + \mathcal{U}_g^j)\right]^2$$
$$- \sum_{l=1}^{m^\Gamma} \frac{\sigma_l}{2} \left[(\mathcal{U}_l^i - \mathcal{U}_l^j) - (\Psi_l^i - \Psi_l^j)\right]^2 - \frac{1}{2}\sum_{l=1}^{m} \sum_{g\in\mathcal{M}_l} \frac{\sigma_{lg}}{2} \left[(\mathcal{U}_l^i - \mathcal{U}_l^j) - (\mathcal{U}_g^i - \mathcal{U}_g^j)\right]^2 \Bigg\},$$

where $\mathcal{U}_l^i$ is the lth entry of vector $\overline{\mathcal{U}}^i$, with $\overline{\mathcal{U}}^i$ being a minimizer of (3.11) when Ψ^i is given by the ith function from (3.15).

3.3.2.2 Compute entries $\upsilon_{ij}^{(2)}$

Consider $\langle \overline{\varphi}_i \pm \overline{\varphi}_j, \Lambda_0(\overline{\varphi}_i \pm \overline{\varphi}_j)\rangle$ and recall that Λ_0 is the DtN map of the reference medium with uniform conductivity $\sigma = 1$. Similarly to (3.8), quadratic form is defined by

$$\langle \overline{\varphi}_i \pm \overline{\varphi}_j, \Lambda_0(\overline{\varphi}_i \pm \overline{\varphi}_j)\rangle = \int\limits_{\Omega_1} |\nabla u(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x},$$

where u is a solution to

$$-\nabla \cdot [\nabla u(\mathbf{x})] = 0, \quad \mathbf{x} \in \Omega_1,$$

with Dirichlet boundary condition $u = \varphi_i \pm \varphi_j$ on $\partial\Omega_1$. It is known that the solution u to this problem in a disk could be found explicitly. Integration of the gradient of u yields

$$\upsilon_{ij}^{(2)} = \begin{cases} k\pi, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

3.3.2.3 Compute entries $\upsilon_{ij}^{(3)}$

Last, we treat values $\upsilon_{ij}^{(3)}$. In our case all gaps between inclusions $\mathcal{D}_i$ and the boundary Γ are identical, i.e. $\delta_i = \delta_1$. Therefore, $\mathcal{R}_{i,k} = \mathcal{R}_{1,k}$ for $i = 1, \dots, m^\Gamma$, and is defined by (3.12). With that υ_{ij}^{3} simplifies to

$$\upsilon_{ij}^{(3)} = m^\Gamma \sum_{k=0}^{K} R_{1,k} \left(a_k^i a_k^j + b_k^i b_k^j\right) + 2m^\Gamma \sum_{k=0}^{K} R_{1,k} \sum_{q \in \mathbb{Z}^+} e^{-qm^\Gamma \frac{\sqrt{2R\delta_1}}{L}} 1_{[0,K]}(k + qm^\Gamma)$$
$$\left[a_k^i a_{k+qm^\Gamma}^j + a_k^j a_{k+qm^\Gamma}^i + b_k^i b_{k+qm^\Gamma}^j + b_k^j b_{k+qm^\Gamma}^i\right],$$

where a_k^i, b_k^i stands for kth coefficients in a truncated Fourier series for function φ_i. Note that because of the choice of basis functions first term contributes nonzero values only to the diagonal elements $\upsilon_{ii}^{(3)}$.

3.4 Modification of the Schur complement system and numerical illustrations

3.4.1 Modification of the Schur complement system (3.4)

With matrix Λ described in a previous section, $\Lambda \overline{u}_\Gamma$ is an approximation for normal derivative on the interface Γ. Then $P\Lambda \overline{u}_\Gamma$ is an approximation for the weak normal derivative on Γ, where P is the matrix corresponding to the integral over the interface. With that

$$P\Lambda \overline{u}_\Gamma = \overline{\lambda}_\Gamma^{(1)},$$

where $\overline{\lambda}_\Gamma^{(1)}$ is an approximation of weak normal derivative on Γ introduced in section 3.2.2. We use relation (3.6) to modify Schur complement system (3.4) to obtain the following linear system

$$(P\Lambda + S^{(2)})\overline{u}_\Gamma = \overline{g}_\Gamma^{(2)}, \tag{3.18}$$

with the preconditioner $S^{(2)}$.

3.4.2 Matrix Λ and its properties

As a test problem to build matrix Λ, the domain Ω_1 is chosen to be a disk of radius $L = 1$. Insert $m = 19$ inclusions inside the domain Ω_1. All inclusions are identical disks of radii $R = 0.198$. Inclusions are evenly distributed in the domain, see Fig. 3.2. The smallest distance between neighboring inclusions is $\delta_{ij} = 0.004$, while the distance between inclusions and the boundary $\partial\Omega_1 = \Gamma$ is $\delta_i = 0.002$.

Figure 3.2: The domain Ω_1 with highly conducting inclusions $\mathcal{D}^i$ of fours groups

Discretize the boundary Γ with M points selected in a prescribed way. First, the closest points on Γ to the boundary neighboring inclusions are required to be in a discretization set of the points. Second, select points equidistantly over Γ. Following the procedure described in section 3.3.2 we obtain matrix $\Lambda \in \mathbb{R}^{M\times M}$.

The resulted matrix Λ is symmetric and positive definite. Also, the diagonal elements of Λ are periodic with the same frequency as the points of discretization match the closest points on Γ to the boundary neighboring inclusions.

Finally, three regimes of parameters R, δ and k are distinguished in [11]. When the boundary function ψ has low oscillations, the resistor network gets excited and determines the leading order of the energy. If the boundary potential ψ is very oscillatory, the network plays no role, because it is not excited. In this case, the energy is approximately equal to that in the reference medium. The resonant term $\mathcal{R}_k$ plays an important role in the approximation of energy when k gets intermediate values. Table 3.1 shows numerical illustration on how much influence each term has when boundary potential is given by a single Fourier mode $\psi = \cos k\theta$.

k	E^{net}	$\langle \psi, \Lambda_0 \psi \rangle$	$\mathcal{R}_k$
1	17.72	6.28	0.23
10	9.51	31.42	32
50	0.6	157.08	57.39
100	0.02	314.16	44.14
200	$1.88 \cdot 10^{-5}$	628.32	25.42
500	$1.8 \cdot 10^{-14}$	1570.8	5.28

Table 3.1: Energy terms for $\psi = \cos kx$

3.4.3 Numerical results

Figure 3.3: The domain Ω partitioned into subdomains

To perform domain decomposition method, the domain Ω_1 is embedded in a disk of radius $\widetilde{L} = 3$, see Fig. 3.3. Conductivity σ outside of Ω_1 is equal to 1. Function

f of the right hand side of (3.1) satisfies the compatibility condition. Number of discretization points on the interface $M = 912$.

The standard `pcg` function of MATLAB® with the preconditioner $S^{(2)}$ is used to solve (3.18). The initial guess $\overline{u}_\Gamma^0$ is a zero vector. The stopping criteria is the relative error

$$\frac{\|(P\Lambda + S^{(2)})\overline{u}_\Gamma^k - \overline{g}_\Gamma^{(2)}\|_2}{\|\overline{g}_\Gamma^{(2)}\|_2}$$

being less than 10^{-6}. PCG algorithm converges to a solution with the desired tolerance in 24 iterations.

Because an analytical solution of (3.1) is not available, we compare the solution $\overline{u}_\Gamma^{\mathrm{DD}}$ obtained by the domain decomposition method to the solution $\overline{u}_\Gamma^{\mathrm{PL}}$, produced by the technique from Chapter 2. We run experiments for different functions f, see Fig. 3.4, Fig. 3.5 and Fig. 3.6. Table 3.2 shows the CPU time for both domain decomposition (DD) and preconditioned Lancsoz (PL) methods. Reported time corresponds to iterative procedure itself, and does not include processioning steps.

f	DD	PL
y^3	0.36	931.28
x	0.34	915.45
$x^3 + y^5$	0.29	793.78

Table 3.2: CPU time in seconds for domain decomposition (DD) and preconditioned Lancsoz (PL) methods

Once solutions on the interface are found we retrieve internal components of the solution. Solution $\overline{u}_2^{\mathrm{DD}}$ is defined by (3.5) with $i = 2$. Solution inside domain Ω_1 is found as a linear interpolation of constant potentials on the inclusions $\mathcal{U}_i$ for $i = 1, \dots, m$ given by (3.11). Since $\overline{u}_\Gamma^{\mathrm{DD}}$ and $\overline{u}_\Gamma^{\mathrm{PL}}$ are close and the matrix of system (3.2) is positive definite, internal solutions $\overline{u}_i^{\mathrm{DD}}$ and $\overline{u}_i^{\mathrm{PL}}$ in corresponding subdomains are close.

Figure 3.4: Solutions $\overline{u}_\Gamma^{\mathrm{DD}}$ and $\overline{u}_\Gamma^{\mathrm{PL}}$ for $f = y^3$

Figure 3.5: Solutions $\overline{u}_\Gamma^{\mathrm{DD}}$ and $\overline{u}_\Gamma^{\mathrm{PL}}$ for $f = x$

Figure 3.6: Solutions $\overline{u}_\Gamma^{\mathrm{DD}}$ and $\overline{u}_\Gamma^{\mathrm{PL}}$ for $f = x^3 + y^5$

3.5 Conclusions

This chapter focuses on a construction of the efficient numerical scheme that can be used to solve high-contrast PDEs of the type (3.1). A typical FEM discretization yields an ill-conditioned matrix when the mesh size h becomes very small in the gaps between inclusions. We propose an approximation for *Schur complement* matrix in Ω_1, which is built using the discrete DtN map. The numerical illustration shows that the proposed algorithm gives qualitatively accurate solution while, being computationally efficient.

Bibliography

[1] J. Aarnes, and T. Y. Hou, “Multiscale domain decomposition methods for elliptic problems with high aspect ratios”, *Acta Mathematicae Applicatae Sinica. English Series*, **18:1**, 2002, pp. 63–76

[2] B. Aksoylu, I. G. Graham, H. Klie, and R. Scheichl, “Towards a rigorously justified algebraic preconditioner for high-contrast diffusion problems”, *Computing and Visualization in Science*, **11:4-6**, 2008, pp. 319–331

[3] O. Axelsson, “Iterative Solution Methods”, Cambridge University Press, 1994

[4] Z.-Z. Bai, M. K. Ng, and Z.-Q. Wang, “Constraint preconditioners for symmetric indefinite matrices”, *SIAM Journal on Matrix Analysis and Applications*, **31:2**, 2009, pp. 410–433

[5] M. Benzi, G. H. Golub, and J. Liesen, “Numerical solution of saddle point problems”, *Acta Numerica*, **14**, 2005, pp. 1–137

[6] M. Benzi, and A. J. Wathen, “Some preconditioning techniques for saddle point problems”, in *Model order reduction: theory, research aspects and applications*, Springer, Berlin, **13**, 2008, pp. 195–211

[7] L. Berlyand, Y. Gorb, and A. Novikov, “Discrete network approximation for highly-packed composites with irregular geometry in three dimensions”, *Multiscale Methods in Science and Engineering. Lecture Notes in Computational Science and Engineering*, Springer, Berlin, Heidelberg, **44**, 2005, pp. 21–57

[8] L. Berlyand, and A. Kolpakov, “Network approximation in the limit of small interparticle distance of the effective properties of a high-contrast random dispersed composite”, *Archive for Rational Mechanics and Analysis*, **159:3**, 2001, pp. 179–227

[9] L. Berlaynd, A. G. Kolpakov, and A. Novikov, "Introduction to the network approximation method for materials modeling", Cambridge University Press, 2012

[10] L. Berlyand, and A. Novikov, "Error of the network approximation for densely packed composites with irregular geometry", *SIAM Journal on Mathematical Analysis*, **34:2**, 2002, pp. 385–408

[11] L. Borcea, Y. Gorb, and Y. Wang, "Asympthotic approximation of the Dirichlet to Neumann map of high contrast conductive media", *SIAM Multiscale modeling & simulation*, **12:4**, 2014, pp.1494 –1532

[12] L. Borcea, and G. Papanicolaou, "Network approximation for transport properties of high contrast materials", *SIAM Journal on Applied Mathematics*, **58:2**, 1998, pp.501–539

[13]

[14] Y. Efendiev, and J. Galvis, "Domain decomposition preconditioners for multiscale flows in high-contast media", *SIAM Multoscale Modeling and Simulation*, **8:4**, 2010, pp.1461–1483

[15] Y. Efendiev, and J. Galvis, "Domain decomposition preconditioners for multiscale flows in high-contrast media: reduced dimension coarse spaces", *SIAM Multiscale Modeling and Simulation*, **8:5**, 2010, pp.1621–1644

[16] Y. Efendiev, and J. Galvis, "Domain decomposition preconditioner for multiscale high-contrast problems", *Huang Y., Kornhuber R., Widlund O., Xu J. (eds) Domain Decomposition Methods in Science and Engineering XIX. Lecture Notes in Computational Science and Engineering*, **78**, 2011, pp.189–196

[17] Y. Efendiev, J. Galvis, and X. Wu, "Multiscale finite element methods: theory and applications", Surveys and Tutorials in the Applied Mathematical Sciences, Springer New York, 2009

[18] H. C. Elman, D. J. Silvester, and A. J. Wathen, Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics, in *Numerical Mathematics and Scientific Computation*, Oxford University Press, New York, 2005

[19] R. Glowinski, and Yu. Kuznetsov, "On the solution of the Dirichlet problem for linear elliptic operators by a distributed Lagrange multiplier method", *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, **327:7**, 1998, pp. 693–698

[20] I. G. Graham, P. O. Lechner, and R. Scheichl, "Domain decomposition for multiscale PDEs", *Numerische Mathematik*, **106**, 2007, pp. 589–626

[21] Yu. Iliash, T. Rossi, and J. Toivanen, "Two iterative methods to solve the Stokes problem", *Technical Report No. 2. Lab. Sci. Comp.*, Dept. Mathematics, University of Jyväskylä. Jyväskylä, Finland, 1993.

[22] C. Keller, N. I. M. Gould, and A. J. Wathen, "Constraint preconditioning for indefinite linear systems", *SIAM Journal on Matrix Analysis and Applications*, **21:4**, 2000, pp. 1300–1317

[23] J. B. Keller, "Conductivity of a medium containing a dense array of perfectly conducting spheres or cylinders or nonconducting cylinders", *Journal of Applied Physics*, **34**, 1963, pp. 991–993

[24] Yu. Kuznetsov, "Efficient iterative solvers for elliptic finite element problems on nonmatching grids", *Russian Journal of Numerical Analysis and Mathematical Modelling*, **10:3**, 1995, pp. 187–211

[25] Yu. Kuznetsov, "Preconditioned iterative methods for algebraic saddle-point problems", *Journal of Numerical Mathematics*, **17:1**, 2009, pp. 67-75

[26] Yu. Kuznetsov, and G. Marchuk, "Iterative methods and quadratic functionals". In *Méthodes de l'Informatique-4*, eds. J.-L. Lions and G. Marchuk, pp. 3–132, Paris, 1974 (In French)

[27] L. Lukšan, and J. Vlček, "Indefinitely preconditioned inexact Newton method for large sparse equality constrained non-linear programming problems", *Numerical Linear Algebra with Applications*, **5:3**, 1998, pp. 219–247

[28] G. W. Milton, "The Theory of Composites", Cambridge University Press, 2002

[29] C. C. Paige, "Computational variants of the Lanczos method for the eigenproblem", *Journal of the Institute of Mathematics and its Applications*, **10**, 1972, pp. 373–381

[30] Y. Saad, and M.H. Schultz, "GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems", *Society for Industrial and Applied Mathematics. Journal on Scientific and Statistical Computing*, **7:3**, 1986, pp. 856–869

[31] A. Toselli, and O. Widlund, "Domain decomposition methods – algorithms and theory", Ŝpringer Series in Computational Mathematics, **34**, Springer-Verlag, Berlin, 2005

[32] E. L. Wachspress, “Iterative solution of elliptic systems, and applications to the neutron diffusion equations of reactor physics”, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966

[33] A. J. Wathen, “Preconditioning”, *Acta Numerica*, **24**, 2015, pp. 329–376

[34] O. B. Widlund, “An Extension Theorem for Finite Element Spaces with Three Applications”, Chapter *Numerical Techniques in Continuum Mechanics* in *Notes on Numerical Fluid Mechanics*, **16**, 1987, pp. 110–122

[35] X. Wu, B. P. B. Silva, and J. Y. Yuan, “Conjugate gradient method for rank deficient saddle point problems”, *Numerical Algorithms*, **35:2–4**, 2004, pp. 139–154