Received: 5 July 2018 | Revised: 22 January 2019 | Accepted: 5 March 2019

DOI: 10.1002/nla.2243

Check for updates

RESEARCH ARTICLE

WILEY

Preconditioned iterative methods for diffusion problems with high-contrast inclusions

Yuliya Gorb[1] | Vasiliy Kramarenko[2] | Yuri Kuznetsov[1]

[1]Department of Mathematics, University of Houston, Houston, Texas

[2]Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia

Correspondence
Yuliya Gorb, Department of Mathematics, University of Houston, Houston, TX 77204.
Email: gorb@math.uh.edu

Funding information
National Science Foundation, Division of Mathematical Sciences, Grant/Award Number: 1350248

Summary

This paper is concerned with robust numerical treatment of an elliptic PDE with high-contrast coefficients, for which classical finite-element discretizations yield ill-conditioned linear systems. This paper 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 the saddle-point type. Three preconditioned iterative procedures—preconditioned Uzawa, preconditioned Lanczos, and preconditioned conjugate gradient for the square of the matrix—are discussed for a special type of the application, namely, highly conducting particles distributed in the domain. Robust preconditioners for solving the derived saddle-point problem are proposed and investigated. Robustness with respect to the contrast parameter and the discretization scale is also justified. Numerical examples support theoretical results and demonstrate independence of the number of iterations of the proposed iterative schemes on the contrast in parameters of the problem and the mesh size.

KEYWORDS

high contrast, Lanczos method, robust preconditioning, saddle-point problem, Schur complement, Uzawa method

1 | INTRODUCTION

We consider iterative solutions of the linear system arising from the discretization of a diffusion problem

$$-\nabla \cdot [\sigma(x)\nabla u] = f, \quad x \in \Omega \tag{1}$$

with appropriate boundary conditions on $\Gamma = \partial\Omega$. In what follows, both in our theoretical consideration and numerical tests, we will assume the homogeneous Dirichlet boundary conditions on Γ. The main focus of this work is on the case when the coefficient function $\sigma(x) \in L^\infty(\Omega)$ varies largely within the domain Ω, that is,

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

We assume that Ω is a bounded domain $\Omega \subset \mathbb{R}^2$ that contains $m \geq 1$ disjoint polygonal subdomains D^s, $s \in \{1, \ldots, m\}$ (see Figure 1) in which σ is "large," for example, of order $O(\kappa)$, but remains of $O(1)$ in the domain outside of $D := \cup_{s=1}^m D^s$.

A P1-FEM discretization of this problem results in a linear system

$$\mathbf{A}_\sigma \overline{u} = \overline{f}, \tag{2}$$

 © 2019 John Wiley & Sons, Ltd.

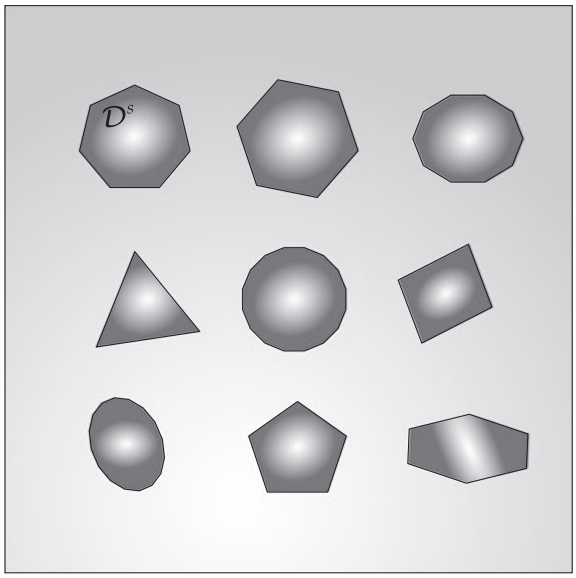


FIGURE 1 An example of D^s

with a large, sparse, symmetric, and positive definite (SPD) matrix $\mathbf{A}_\sigma$. A major issue in numerical treatments of (1) with the coefficient σ discussed above is that the high contrast leads to an ill-conditioned matrix $\mathbf{A}_\sigma$. Indeed, if h is the discretization scale, then the condition number of the resulting stiffness matrix $\mathbf{A}_\sigma$ grows proportionally to h^{-2} with the coefficient of proportionality linearly depending on κ. Accordingly, the high-contrast problems have been a subject of an active research recently.[1–4] There are a few methods in preconditioning linear systems describing problems with a high-contrast heterogeneous coefficient proposed in the exiting literature. The most common are the domain decomposition (e.g., the works of Farhat et al.[5] and Galvis et al.[6]) for the standard Galerkin FEM, and multilevel methods for the hybridized mixed system (e.g., the work of Kraus[7] and the references therein).

Our objective here is robust numerical treatment of the described problem. Accordingly, we introduce an additional variable that allows us to replace (2) with an equivalent formulation in terms of a linear system

$$\boldsymbol{\mathcal{A}}\overline{x} = \overline{F}, \quad \text{with} \quad \overline{F} = \begin{bmatrix} \bar{\mathrm{f}} \\ \hline 0 \end{bmatrix}, \tag{3}$$

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

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\mathbf{C} \end{bmatrix}, \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{N\times N}$ is SPD, $\mathbf{B} \in \mathbb{R}^{n\times N}$ is rank deficient, and $\mathbf{C} \in \mathbb{R}^{n\times n}$ is an SPD matrix. Below, we discuss three iterative procedures: preconditioned Uzawa (PU) method for the system with an SPD Schur complement matrix, preconditioned Lanczos (PL) method for solving (3), and preconditioned conjugate gradient (PCG) method for an equivalent system with an SPD matrix. Then, we propose a robust block-diagonal preconditioner

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathcal{H}_{\mathrm{A}} & 0 \\ 0 & \mathcal{H}_{\mathrm{S}} \end{bmatrix}$$

for solving (3)–(4) with these three iterative methods. The main feature of the proposed preconditioners is that convergence rates of the discussed iterative schemes are independent of the contrast parameter $\kappa \gg 1$ and the discretization size $h > 0$. A rigorous justification of the latter statement is based on the evaluation of the eigenvalues of the matrix $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}$, which are shown to belong to the union of two intervals $[\mu_-^1, \mu_-^2] \cup [\mu_+^1, \mu_+^2]$, where $\mu_-^1 < \mu_-^2 < 0 < \mu_+^1 < \mu_+^2$. Assuming that the mesh on Ω is regularly shaped and quasi-uniform, we demonstrate that constants $\mu_\pm^i$ $(i = 1, 2)$ are independent of the discretization scale h and the number of inclusions. If, in addition, we assume that particles are located at distances comparable to their sizes, then $\mu_\pm^i$ $(i = 1, 2)$ are independent of the diameters of $\mathcal{D}^s$, $s \in \{1, \ldots, m\}$, their locations, and distances between them. The numerical experiments on simple test cases support theoretical findings and demonstrate independence of convergence rates of the proposed iterative schemes on parameters indicated above. These numerical tests are performed for a two-dimensional problem, whereas theoretical results remain true for three dimensions as well.

An *optimal* preconditioner is a matrix operator that allows to speed up an iterative scheme so that the convergence rate is independent of the mesh size. The development of optimal preconditioners for saddle-point problems has been an active area of research since the early 1990s; see for example other works[8–11] and the references therein. The saddle-point algebraic systems in the cited literature arise from various discretization scenarios, for example, constrained minimization,[10] mixed FEM approximation for an elliptic problem,[9] or mixed FEM for Stokes problem.[11] In this paper, the saddle-point linear system is obtained from the FEM discretization of (1) augmented with an additional equation that yields (3)–(4) that is equivalent to (2) produced by the classical P1-FEM discretization.

The main feature of the problem considered in this paper is that we deal with a special type of saddle-point matrices that, in particular, contains a rank-deficient block **B**. In addition, this paper proposes a very special form for the block $\mathcal{H}_S$ of $\boldsymbol{\mathcal{H}}$ (see (29) in Section 3), utilized in three methods that yield theoretical results mentioned above. Moreover, one of the iterative procedures that we employed in this paper is the *Lanczos method*,[12,13] which, as would be evident from our numerical experiments below, has demonstrated significant advantages over the other methods with respect to the arithmetic cost.

Finally, we point out that robust numerical treatment of the described problem is crucial in developing the mutiscale strategies for models of composite materials with highly conducting particles. The latter finds their application in particulate flows, subsurface flows in natural porous formations, electrical conduction in composite materials, and medical and geophysical imaging.

This paper is organized as follows. In Section 2, the mathematical problem formulation is presented including the derivation of the saddle-point problem of the type (3)–(4). Section 3 discusses three iterative methods (PU, PL, and PCG, mentioned above) for solving system (3)–(4) and proposes efficient preconditioners for all of them. The main theoretical results, which are the estimates for the eigenvalues of the matrix $\boldsymbol{\mathcal{H}\mathcal{A}}$, are stated and proven in Section 4. Numerical experiments based on simple test scenarios are presented in Section 5. Conclusions are discussed in Section 6.

2 | PROBLEM FORMULATION

2.1 | Equivalent variational formulations

Consider an open, bounded domain $\Omega \subset \mathbb{R}^2$ with a piece-wise smooth boundary $\Gamma := \partial\Omega$ that contains $m \geq 1$ subdomains D^s with piece-wise smooth boundaries $\Gamma_s := \partial\Omega_s, s \in \{1, \dots, m\}$ (see Figure 1). Assume that $\Gamma_s \cap \Gamma_t = \emptyset$ when $s \neq t$, and $\Gamma \cap \Gamma_s = \emptyset, s \in \{1, \dots, m\}$. For simplicity, we assume that Ω and D^s are polygons. The union of D^s is denoted by D. In the domain Ω, we consider the following elliptic problem:

$$\begin{cases} -\nabla \cdot [\sigma(x)\nabla u] = f, & x \in \Omega \\ u = 0, & x \in \Gamma \end{cases} \tag{5}$$

with the source term $f \in L^2(\Omega)$, and the coefficient σ that varies largely inside the domain Ω. In this paper, we are focused on the case when σ is a piecewise constant function given by

$$\sigma(x) = \begin{cases} 1, & x \in \Omega \backslash \overline{D} \\ 1 + \frac{1}{\varepsilon_s}, & x \in D^s, \ \ s \in \{1, \dots, m\} \end{cases} \tag{6}$$

with $0 < \varepsilon_s \equiv \text{const} \leqslant 1, s \in \{1, \dots, m\}$. The standard variational formulation of (5) is

$$\text{find } u \in V := H_0^1(\Omega) \text{ such that } \int_\Omega \nabla u \cdot \nabla v \, dx + \sum_{s=1}^{m} \frac{1}{\varepsilon_s} \int_{D^s} \nabla u \cdot \nabla v \, dx = \int_\Omega f v \ dx, \ \forall v \in V. \tag{7}$$

We introduce new variables $p_s \in H^1(D^s)$ via

$$p_s = \frac{1}{\varepsilon_s} u_s + c_s \ \text{ in } \ D^s, \quad s \in \{1, \dots, m\}, \tag{8}$$

where $u_s = u|_{D^s}$ and the c_s's are arbitrary constants, $s \in \{1, \dots, m\}$. We therefore replace the following variational formulation (7) with

$$\text{find } u \in V \text{ and } p_s \in V_s := H^1(D^s) = V\big|_{D^s}, \ s \in \{1, \dots, m\} \text{ such that}$$

$$\int_\Omega \nabla u \cdot \nabla v \, dx + \sum_{t=1}^{m} \int_{D^t} \nabla p_t \cdot \nabla v \, dx = \int_\Omega f v \, dx, \ \forall v \in V, \tag{9}$$

$$\int_{D^s} \nabla u \cdot \nabla w \, dx - \varepsilon_s \int_{D^s} \nabla p_s \cdot \nabla w \, dx = 0, \forall w \in V_s, s \in \{1, \dots, m\}. \tag{10}$$

Two formulations (7) and (9)–(10) are equivalent in the sense that their solutions $u \in H^1(\Omega)$ coincide, and any solution $p_s \in V_s$ of (9)–(10) is equal to the function $\frac{1}{\varepsilon_s} u_s + c_s$ with an appropriate constant $c_s, s \in \{1, \dots, m\}$. For the uniqueness of p_s, we can either require

$$\int_{D^s} p_s \, dx = 0, \quad s \in \{1, \dots, m\} \tag{11}$$

or modify the formulation (10) as follows:

$$\begin{aligned} &\text{find } u \in V \text{ and } p_s \in V_s \text{ such that } \int_\Omega \nabla u \cdot \nabla v \, dx + \sum_{t=1}^{m} \int_{D^t} \nabla p_t \cdot \nabla v \, dx = \int_\Omega f v \, dx, \forall v \in V, \\ &\int_{D^s} \nabla u \cdot \nabla w \, dx - \varepsilon_s \int_{D^s} \nabla p_s \cdot \nabla w \, dx - \frac{1}{|D^s|} \left[\int_{D^s} p_s \, dx \right] \left[\int_{D^s} w \, dx \right] = 0, \forall w \in V_s, \ s \in \{1, \dots, m\}, \end{aligned} \tag{12}$$

where $|D^s|$ is the area of the particle D^s. It is obvious that solutions $p_s, s \in \{1, \dots, m\}$, of (12) satisfy condition (11), and the above constants c_s are defined by

$$c_s = -\frac{1}{\varepsilon_s} \int_{D^s} u \, dx, \quad s \in \{1, \dots, m\}.$$

2.2 | Discretization of (12) and description of the saddle-point problem

Let Ω_h be a triangular mesh on Ω. We assume that Ω_h conforms with boundaries Γ and $\Gamma_s, s \in \{1, \dots, m\}$. By that, we mean that Γ and Γ_s are the unions of the triangular sides. We define $D_h^s = \Omega_h|_{D^s}, s \in \{1, \dots, m\}$, and $D_h := \cup_{s=1}^m D_h^s$.

We now choose a FEM space $V_h \subset H_0^1(\Omega)$ to be the space of linear finite-element functions defined on Ω_h, and $V_h^s := V_h|_{D_h^s}, s \in \{1, \dots, m\}$. The FEM discretization[14] of (12) may be described as follows:

$$\begin{aligned} &\text{find } u_h \in V_h \text{ and } p_h = \left(p_h^1, \dots, p_h^m\right) \text{ with } p_h^s \in V_h^s \text{ such that} \\ &\int_{\Omega_h} \nabla u_h \cdot \nabla v_h \, dx + \int_{D_h} \nabla p_h \cdot \nabla v_h \, dx = \int_{\Omega_h} f v_h \, dx, \quad \forall v_h \in V_h, \\ &\int_{D_h^s} \nabla u_h \cdot \nabla w_h^s \, dx - \varepsilon_s \int_{D_h^s} \nabla p_h^s \cdot \nabla w_h^s \, dx - \frac{1}{|D_h^s|} \left[\int_{D_h^s} p_h^s \, dx \right] \left[\int_{D^s} w_h^s \, dx \right] = 0, \ \forall w_h^s \in V_h^s, \end{aligned} \tag{13}$$

for $s \in \{1, \dots, m\}$. Because Ω_h conforms with boundaries Γ and Γ_s and $D_h^s = D^s$, we drop the subscript h in the notation of domains. Now, we observe that the FEM problem (13) is fully equivalent to the classical FEM problem:

$$\text{find } u_h \in V_h \text{ such that } \int_\Omega \nabla u_h \cdot \nabla v_h \, dx + \sum_{s=1}^{m} \frac{1}{\varepsilon_s} \int_{D^s} \nabla u_h \cdot \nabla v_h \, dx = \int_\Omega f v_h \, dx, \ \forall v_h \in V_h,$$

as it yields the same solution u_h from the same FEM space V_h as u_h of (13). To that end, all the robust error estimates for the classical P1-FEM approximation of the continuum problem (5)–(6) stay true for the case of FEM approximation (13).

FEM formulation (13) produces the following linear system of equations:

$$\begin{cases} \mathbf{A}\overline{u} + \mathbf{B}^T \overline{p} = \overline{\mathrm{f}}, \\ \mathbf{B}\overline{u} - [\mathbf{\Sigma}_\varepsilon \boldsymbol{\mathcal{B}}_D + \mathbf{Q}]\overline{p} = \overline{0}, \end{cases} \qquad \overline{u} \in \mathbb{R}^N, \ \overline{p} \in \mathbb{R}^n, \tag{14}$$

or equivalently,

$$\boldsymbol{\mathcal{A}}_\varepsilon \mathbf{z}_\varepsilon = \overline{\mathrm{F}}, \tag{15}$$

with the *saddle-point* matrix

$$\boldsymbol{\mathcal{A}}_\varepsilon = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\mathbf{\Sigma}_\varepsilon \boldsymbol{\mathcal{B}}_D - \mathbf{Q} \end{bmatrix} \in \mathbb{R}^{(N+n)\times(N+n)}$$

and vectors

$$\mathbf{z}_\varepsilon = \begin{bmatrix} \overline{u} \\ \overline{p} \end{bmatrix} \in \mathbb{R}^{N+n}, \quad \overline{\mathrm{F}} = \begin{bmatrix} \overline{\mathrm{f}} \\ \overline{0} \end{bmatrix} \in \mathbb{R}^{N+n}.$$

To provide the comprehensive description of the linear system (14) or (15), we introduce the following notation 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 let n be the number of nodes in $\overline{D}_h$ so that

$$n = \sum_{s=1}^{m} n_s,$$

where n_s denotes the number of nodes in $\overline{D}_h^s$, and finally, n_0 is the number of nodes in $\Omega_h \backslash \overline{D_h}$. We have

$$N = n_0 + n.$$

The vector $\overline{u} \in \mathbb{R}^N$ in (14) has entries $u_i = u_h(x_i)$ for $x_i \in \Omega_h$. We enumerate the entries of $\overline{u}$ in such a way that its first n entries correspond to the nodes of $\overline{D}_h$, and the remaining n_0 entries correspond to the nodes of $\Omega_h \backslash \overline{D}_h$. Entries of the first group can be further partitioned into m subgroups such that there are n_s entries in the sth group corresponding to $\overline{D}_h^s$, $s \in \{1, \ldots, m\}$. Similarly, the vector $\overline{p} \in \mathbb{R}^n$ has entries $p_i = p_h(x_i)$ where $x_i \in \overline{D}_h$. We can write

$$\overline{p} = \begin{bmatrix} \overline{p}_1 \\ \vdots \\ \overline{p}_n \end{bmatrix} \in \mathbb{R}^n, \quad \text{where } \overline{p}_s \in \mathbb{R}^{n_s}, \quad s \in \{1, \ldots, m\}.$$

The symmetric positive-definite matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ of (14) is the stiffness matrix that arises from the discretization of the Laplace operator with the homogeneous Dirichlet boundary conditions on Γ, that is,

$$(\mathbf{A}\overline{u}, \overline{v}) = \int_{\Omega_h} \nabla u_h \cdot \nabla v_h \, dx, \quad \text{where } \overline{u}, \overline{v} \in \mathbb{R}^N, \quad u_h, v_h \in V_h, \tag{16}$$

where $(\cdot, \cdot)$ is the standard dot product of vectors. With the above orderings, the matrix $\mathbf{A}$ of (16) can be presented as a 2×2 block-matrix as follows:

$$\mathbf{A} = \begin{bmatrix} \mathrm{A}_{DD} & \mathrm{A}_{D0} \\ \mathrm{A}_{0D} & \mathrm{A}_{00} \end{bmatrix}, \tag{17}$$

where the block $\mathrm{A}_{DD} \in \mathbb{R}^{n \times n}$ corresponds to the inclusions $\overline{D}_h^s$, $s \in \{1, \ldots, m\}$, the block $\mathrm{A}_{00} \in \mathbb{R}^{n_0 \times n_0}$ corresponds to the region outside of $\overline{D}_h$, and the entries of $\mathrm{A}_{D0} \in \mathbb{R}^{n \times n_0}$ and $\mathrm{A}_{0D} = \mathrm{A}_{D0}^T$ are assembled from entries associated with both $\overline{D}_h$ and $\Omega_h \backslash \overline{D}_h$.

The matrix $\boldsymbol{\mathcal{B}}_D \in \mathbb{R}^{n \times n}$ in (14), which corresponds to the highly conducting inclusions is the $m \times m$ block-diagonal matrix

$$\boldsymbol{\mathcal{B}}_D = \mathrm{diag}\,(\mathrm{B}_1, \ldots, \mathrm{B}_\mathrm{m}), \tag{18}$$

whose blocks $\mathrm{B_s} \in \mathbb{R}^{n_s \times n_s}$ are defined by

$$(\mathrm{B_s}\overline{u}, \overline{v}) = \int_{D^s} \nabla u_h \cdot \nabla v_h \, dx, \quad \text{where } \overline{u}, \overline{v} \in \mathbb{R}^{n_s}, \quad u_h, v_h \in V_h^s. \tag{19}$$

Note that the matrix $\mathrm{B_s}$ is the stiffness matrix in the discretization of the Laplace operator in the domain D^s with the Neumann boundary conditions on Γ_s, $s \in \{1, \ldots, m\}$. In addition, we remark that each matrix $\mathrm{B_s}$ is *positive semidefinite* with

$$\ker\, \mathrm{B_s} = \mathrm{span}\{\overline{e}_s\}, \quad \text{where} \quad \overline{e}_s = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^{n_s}. \tag{20}$$

To this end, we have

$$\dim \ker\, \boldsymbol{\mathcal{B}}_D = m.$$

Consequently, the matrix $\mathbf{B} \in \mathbb{R}^{n \times N}$ of (14) is written in the block form as

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\mathcal{B}}_D & \mathbf{0} \end{bmatrix}, \tag{21}$$

with zero-matrix $\mathbf{0} \in \mathbb{R}^{n \times n_0}$ and $\boldsymbol{\mathcal{B}}_D \in \mathbb{R}^{n \times n}$. The vector $\overline{\mathrm{f}} \in \mathbb{R}^N$ of (14) is defined in a similar way by

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

The foregoing implies that the first equation of (13) results in the first equation of (14). We now denote

$$\boldsymbol{\Sigma}_\varepsilon = \mathrm{diag}\,(\varepsilon_1 \mathrm{I}_1, \ldots, \varepsilon_m \mathrm{I}_\mathrm{m}),$$

where $\mathrm{I_s} \in \mathbb{R}^{n_s \times n_s}$ is the identity matrix. Finally, we construct the matrix $\mathbf{Q}$ in (13) using

$$\mathbf{Q} = \mathrm{diag}\,(\mathrm{Q_1}, \ldots, \mathrm{Q_m}), \tag{22}$$

whose blocks $\mathrm{Q_s} \in \mathbb{R}^{n_s \times n_s}$, $s \in \{1, \ldots, m\}$, are defined by

$$(\mathrm{Q_s}\,\overline{p}, \overline{q}) = \frac{1}{|D_h^s|}\left[\int_{D_h^s} p_h\,dx\right]\left[\int_{D_h^s} q_h\,dx\right], \quad \text{where } \overline{p}, \overline{q} \in \mathbb{R}^{n_s},\ p_h,\ q_h \in V_h^s. \tag{23}$$

It should be evident from the considerations below that an alternative representation of the matrix $\mathrm{Q_s}$ is obtained via

$$\mathrm{Q_s} = \frac{1}{d_s^2}\left[\mathrm{M_s}\overline{w}_s^1 \otimes \mathrm{M_s}\overline{w}_s^1\right], \quad \text{where} \quad d_s = \left|D_h^s\right|^{1/2} \quad \text{and} \quad \overline{w}_s^1 := \frac{1}{d_s}\overline{e}_s \in \mathbb{R}^{n_s}. \tag{24}$$

The matrix $\mathrm{M_s} \in \mathbb{R}^{n_s \times n_s}$, *mass matrix* associated with the inclusion $\mathcal{D}^s$, is given by

$$(\mathrm{M_s}\,\overline{p}_s, \overline{q}_s) = \int_{D_h^s} p_h^s\, q_h^s\,dx, \quad \text{for all } \overline{p}_s, \overline{q}_s, \in \mathbb{R}^{n_s}, \quad p_h^s,\, q_h^s \in V_h^s,\ \ s \in \{1, \ldots, m\}. \tag{25}$$

In (24), $\overline{p} \otimes \overline{q} = \overline{p}\,\overline{q}^T$ denotes the *outer product* of vectors $\overline{p}$ and $\overline{q}$. The matrix $\mathrm{Q_s}$ is a symmetric and positive-semidefinite rank-one matrix generated by the $\mathrm{M_s}$-normal vector $\overline{w}_s^1$, that is, $(\mathrm{M_s}\overline{w}_s^1, \overline{w}_s^1) = 1$, $s \in \{1, \ldots, m\}$.

By virtue of (19)–(25), the second equation of (12) yields the second equation in the system (14). Note that, with (17), the symmetric and indefinite matrix $\boldsymbol{\mathcal{A}}_\epsilon$ defined in (15) is then

$$\boldsymbol{\mathcal{A}}_\epsilon = \begin{bmatrix} \mathrm{A}_{DD} & \mathrm{A}_{D0} & \boldsymbol{\mathcal{B}}_D \\ \mathrm{A}_{0D} & \mathrm{A}_{00} & \mathbf{0}^T \\ \boldsymbol{\mathcal{B}}_D & \mathbf{0} & -\boldsymbol{\Sigma}_\epsilon \boldsymbol{\mathcal{B}}_D - \mathbf{Q} \end{bmatrix}. \tag{26}$$

This concludes the derivation of the saddle-point formulation (14). Clearly, there exists a unique solution $\overline{u} \in \mathbb{R}^N, \overline{p} \in \mathbb{R}^n$, or equivalently, $\mathbf{z}_\epsilon \in \mathbb{R}^{N+n}$.

System (14) was proposed in other works[15–17] for the case when $\mathbf{Q} = \mathbf{0}$, where it was also demonstrated that (14) can be derived in a purely algebraic way.

3 | PRECONDITIONED ITERATIVE METHODS

In what follows, we consider and investigate three iterative methods for solving system (15). The first one is the PCG method or PU for the Schur complement system

$$\mathbf{S}_\epsilon\,\overline{p} = \overline{\mathrm{g}} =: \mathbf{B}\mathbf{A}^{-1}\,\overline{\mathrm{f}}, \tag{27}$$

where

$$\mathbf{S}_\epsilon := \boldsymbol{\Sigma}_\epsilon \boldsymbol{\mathcal{B}}_D + \mathbf{Q} + \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T, \tag{28}$$

with the preconditioner

$$\mathcal{H}_\mathrm{S} = [\boldsymbol{\mathcal{B}}_D + \mathbf{Q}]^{-1} \in \mathbb{R}^{n \times n}. \tag{29}$$

The second method is the preconditioned Lanzcos (PL) method with the preconditioner

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathcal{H}_\mathrm{A} & 0 \\ 0 & \mathcal{H}_\mathrm{S} \end{bmatrix}, \tag{30}$$

where $\mathcal{H}_\mathrm{A} \in \mathbb{R}^{N \times N}$ is a given symmetric positive-definite matrix introduced below, and $\mathcal{H}_\mathrm{S}$ is specified in (29). We note that the main focus of this paper is on the choice of $\mathcal{H}_\mathrm{S}$ given by (29), whereas for $\mathcal{H}_\mathrm{A}$, one can choose any preconditioner for the discrete Laplacian $\mathbf{A}$.

The third method is the PCG method with the preconditioner $\boldsymbol{\mathcal{H}}$ defined in (30) for a modified system obtained from (15) as follows:

$$\mathbf{K}_\epsilon \mathbf{z}_\epsilon = \boldsymbol{\mathcal{G}}_\epsilon, \tag{31}$$

where

$$\mathbf{K}_\epsilon = \boldsymbol{\mathcal{A}}_\epsilon \boldsymbol{\mathcal{H}} \boldsymbol{\mathcal{A}}_\epsilon, \quad \boldsymbol{\mathcal{G}}_\epsilon = \boldsymbol{\mathcal{A}}_\epsilon \boldsymbol{\mathcal{H}}\,\overline{\mathrm{F}}. \tag{32}$$

3.1 | PU method

The PU algorithm combined with the PCG method is well known.[18,19] It is defined by

$$\overline{p}^k = \overline{p}^{k-1} - \beta_k \overline{\xi}_k, \quad k = 1, 2, \ldots, \tag{33}$$

where

$$\overline{\xi}_k = \begin{cases} \mathcal{H}_S\left(\mathbf{S}_\varepsilon \overline{p}^0 - \overline{g}\right), & k = 1 \\ \mathcal{H}_S\left(\mathbf{S}_\varepsilon \overline{p}^{k-1} - \overline{g}\right) - \alpha_k \overline{\xi}_{k-1}, & k \geq 2, \end{cases} \tag{34}$$

and

$$\beta_k = \frac{\left(\mathbf{S}_\varepsilon \overline{p}^{k-1} - \overline{g}, \overline{\xi}_k\right)}{\left(\mathbf{S}_\varepsilon \overline{\xi}_k, \overline{\xi}_k\right)}, \quad \alpha_k = \frac{\left(\mathcal{H}_S\left(\mathbf{S}_\varepsilon \overline{p}^{k-1} - \overline{g}\right), \mathbf{S}_\varepsilon \overline{\xi}_{k-1}\right)}{\left(\mathbf{S}_\varepsilon \overline{\xi}_{k-1}, \overline{\xi}_{k-1}\right)}, \quad k = 1, 2, \ldots. \tag{35}$$

Here, $\overline{p}^0$ is an initial guess and $\mathcal{H}_S$ is given by (29).

We denote by $\overline{p}^*$ the solution of (27); then, the convergence estimate for (31)–(33) is given by[20,21]

$$\|\overline{p}^k - \overline{p}^*\|_{\mathbf{S}_\varepsilon} \leqslant \frac{1}{C_k\left(\frac{b+a}{b-a}\right)} \|\overline{p}^0 - \overline{p}^*\|_{\mathbf{S}_\varepsilon}, \quad k = 0, 1, 2, \ldots,$$

where $\|\cdot\|_{\mathbf{S}_\varepsilon}$ is the elliptic norm generated by the matrix $\mathbf{S}_\varepsilon$, $C_k(t)$ is the Chebyshev polynomial of degree k, and b and a are the upper and lower estimate of the eigenvalues of the matrix $\mathcal{H}_S\mathbf{S}_\varepsilon$, respectively.

Turning our attention to the eigenvalue problem

$$\mathcal{H}_S\mathbf{S}_\varepsilon\overline{\psi} = \mu\overline{\psi},$$

we observe that

$$\mathcal{H}_S\boldsymbol{\mathcal{B}}_D = \mathbf{I} - \tilde{\mathbf{Q}}, \tag{36}$$

$$\mathcal{H}_S\mathbf{Q} = \tilde{\mathbf{Q}}, \tag{37}$$

where $\tilde{\mathbf{Q}}$ is an $m \times m$ block diagonal matrix

$$\tilde{\mathbf{Q}} = \mathrm{diag}\,(\tilde{\mathrm{Q}}_1, \ldots, \tilde{\mathrm{Q}}_m),$$

with M_s-orthogonal projectors

$$\tilde{\mathrm{Q}}_s = \overline{w}_s^1 \otimes \left(\mathrm{M}_s\overline{w}_s^1\right) \in \mathbb{R}^{n_s \times n_s}, \quad s \in \{1, \ldots, m\},$$

where $\overline{w}_s^1$ and M_s were introduced in (24) and (25), respectively.

Remark 1. It follows from (36), (37) that implementation of the matrix–vector products $\mathcal{H}_S\boldsymbol{\mathcal{B}}_D\overline{y}$ and $\mathcal{H}_S\mathbf{Q}\overline{y}$ requires only $2n$ arithmetical operations for any vector $\overline{y} \in \mathbb{R}^n$, and we do not need to solve a system with the matrix $\boldsymbol{\mathcal{B}}_D + \mathbf{Q}$.

Simple algebraic analysis[15,17] shows that

$$a \geqslant \min\{a_0 + \varepsilon_{\min}; 1\} \tag{38}$$

and

$$b \geqslant \max\{b_0 + \varepsilon_{\max}; 1\}, \tag{39}$$

where $a_0 > 0$ and b_0 are lower and upper estimates, respectively, for the eigenvalues of the matrix

$$\mathcal{H}_S\mathbf{S}_0 \equiv \mathcal{H}_S\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T,$$

where $\mathbf{S}_0 = \mathbf{S}_\varepsilon$ when $\varepsilon_1 = \ldots \varepsilon_m = 0$, and $\varepsilon_{\min} = \min_{1\leqslant t\leqslant m}\varepsilon_t$, $\varepsilon_{\max} = \max_{1\leqslant t\leqslant m}\varepsilon_t$. The values of a_0 and b_0 will be derived in Section 4.

3.2 | PL method

The PL method for systems with symmetric indefinite matrices was proposed in the late 1960s; see the work of Marchuk et al.[21] and references therein. Herein, we consider the PL method for the saddle-point system (15) preconditioned by a symmetric positive-definite matrix $\boldsymbol{\mathcal{H}}$ of (30) with the given symmetric positive-definite matrix $\mathcal{H}_A$ introduced

below, and $\mathcal{H}_S$ defined by (29). The PL method is described as follows[21]:

$$\overline{z}^k = \overline{z}^{k-1} - \beta_k \overline{\xi}_k, \quad k = 1, 2, \ldots, \tag{40}$$

where

$$\overline{\xi}_k = \begin{cases} \boldsymbol{\mathcal{H}}\left(\boldsymbol{\mathcal{A}}_\varepsilon \overline{z}^0 - \overline{\mathrm{F}}\right), & k = 1 \\ \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_1 - \alpha_2 \overline{\xi}_1, & k = 2 \\ \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-1} - \alpha_k \overline{\xi}_{k-1} - \gamma_k \overline{\xi}_{k-2}, & k \geq 3, \end{cases} \tag{41}$$

and

$$\alpha_k = \frac{\left(\boldsymbol{\mathcal{A}}_\varepsilon \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-1}, \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-1}\right)}{\left(\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-1}, \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-1}\right)}, \qquad \gamma_k = \frac{\left(\boldsymbol{\mathcal{A}}_\varepsilon \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-1}, \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-2}\right)}{\left(\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-2}, \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_{k-2}\right)}, \quad k = 1, 2, \ldots, \tag{42}$$

$$\beta_k = \frac{\left(\boldsymbol{\mathcal{A}}_\varepsilon \overline{z}^{k-1} - \overline{\mathrm{F}}, \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_k\right)}{\left(\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_k, \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon \overline{\xi}_k\right)}, \quad k = 1, 2, \ldots. \tag{43}$$

Let $\overline{z}^0$ be an initial guess and let $\overline{z}^*$ be the solution of (15); then, the following convergence estimate holds[21]:

$$\|\overline{z}^k - \overline{z}^*\|_{\mathbf{K}_\varepsilon} \leqslant \frac{1}{C_{k/2}\left(\frac{b^2+a^2}{b^2-a^2}\right)} \|\overline{z}^0 - \overline{z}^*\|_{\mathbf{K}_\varepsilon}, \quad k = 2, 4, \ldots, \tag{44}$$

where $\mathbf{K}_\varepsilon$ is given by (32) and $\|\cdot\|_{\mathbf{K}_\varepsilon}$ is the elliptic norm generated by the matrix $\mathbf{K}_\varepsilon = \mathbf{K}_\varepsilon^T > 0$. Here, $C_{k/2}$ is the Chebyshev polynomial of degree $k/2$, and b^2 and $a^2 > 0$ are estimates from above and from below for eigenvalues of the matrix $(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon)^2$, respectively.

3.3 | PCG method

We apply the PCG method with the preconditioner $\boldsymbol{\mathcal{H}}$ defined by (30) to system (31):

$$\overline{z}^k = \overline{z}^{k-1} - \beta_k \overline{\xi}_k, \quad k = 1, 2, \ldots, \tag{45}$$

where

$$\overline{\xi}_k = \begin{cases} \boldsymbol{\mathcal{H}}\left(\mathbf{K}_\varepsilon \overline{z}^0 - \mathcal{G}_\varepsilon\right), & k = 1 \\ \boldsymbol{\mathcal{H}}\left(\mathbf{K}_\varepsilon \overline{z}^{k-1} - \mathcal{G}_\varepsilon\right) - \alpha_k \overline{\xi}^{k-1}, & k \geq 2, \end{cases} \tag{46}$$

and

$$\alpha_k = \frac{\left(\boldsymbol{\mathcal{H}}\left[\mathbf{K}_\varepsilon \overline{z}^{k-1} - \mathcal{G}_\varepsilon\right], \mathbf{K}_\varepsilon \overline{\xi}_{k-1}\right)}{\left(\mathbf{K}_\varepsilon \overline{\xi}_{k-1}, \overline{\xi}_{k-1}\right)}, \qquad \beta_k = \frac{\left(\mathbf{K}_\varepsilon \overline{z}^{k-1} - \mathcal{G}_\varepsilon, \overline{\xi}_k\right)}{\left(\mathbf{K}_\varepsilon \overline{\xi}_k, \overline{\xi}_k\right)}, \quad k = 1, 2, \ldots. \tag{47}$$

The convergence estimate for the method is as follows[20,21]:

$$\|\overline{z}^k - \overline{z}^*\|_{\mathbf{K}_\varepsilon} \leqslant \frac{1}{C_k\left(\frac{b^2+a^2}{b^2-a^2}\right)} \|\overline{z}^0 - \overline{z}^*\|_{\mathbf{K}_\varepsilon}, \ k = 1, 2, \ldots, \tag{48}$$

where the same matrix $\mathbf{K}_\varepsilon$ is defined in (32) and values a^2 and b^2 are those of estimate (44).

4 | EIGENVALUE ESTIMATES

4.1 | Eigenvalue estimates for the matrix $\mathcal{H}_S\mathbf{S}_0$

We consider the eigenvalue problem

$$\mathbf{S}_0 \overline{\psi}_D = \mu \mathcal{H}_S^{-1} \overline{\psi}_D, \tag{49}$$

where

$$\mathbf{S}_0 = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T = \boldsymbol{\mathcal{B}}_D \mathrm{S}_{00}^{-1} \boldsymbol{\mathcal{B}}_D, \tag{50}$$

$$\mathcal{H}_{\mathrm{S}}^{-1} = \boldsymbol{\mathcal{B}}_D + \mathbf{Q}, \tag{51}$$

and

$$\mathrm{S}_{00} = \mathrm{A}_{DD} - \mathrm{A}_{D0}\mathrm{A}_{00}^{-1}\mathrm{A}_{0D}$$

is the Schur complement of A_{00}. It is obvious that $\mu = 1$ if $\overline{\psi}_D \in \ker \boldsymbol{\mathcal{B}}_D$, and $\mathbf{Q}\overline{\psi}_D = \mathbf{0}$ for any $\mu \neq 1$; then, $\overline{\psi}_D$ is $\mathbf{M}$-orthogonal to $\ker \boldsymbol{\mathcal{B}}_D$, with

$$\mathbf{M} = \mathrm{diag}\left\{\mathrm{M}_1, \ldots, \mathrm{M}_{\mathrm{m}}\right\}, \quad \text{where } \mathrm{M}_{\mathrm{s}} \text{ is given by (25)}, \quad s \in \{1, \ldots, m\}.$$

Thus, to derive a_0 and b_0 of (38)–(39), instead of (49), we can consider the following eigenvalue problem:

$$\mathrm{S}_0\overline{\psi}_D = \mu\, \boldsymbol{\mathcal{B}}_D\overline{\psi}_D, \tag{52}$$

under the condition $\left(\mathbf{M}\overline{\psi}_D, \overline{w}\right) = 0$ for all $\overline{w} \in \ker \boldsymbol{\mathcal{B}}_D$.

If μ is an eigenvalue of (52) and $\overline{\psi}_D$ a corresponding eigenvector, then

$$\mu = \frac{(\mathrm{S}_0\overline{\psi}_D, \overline{\psi}_D)}{(\boldsymbol{\mathcal{B}}_D\overline{\psi}_D, \overline{\psi}_D)} = \frac{(\mathbf{A}\overline{\psi}, \overline{\psi})}{(\boldsymbol{\mathcal{B}}_D\overline{\psi}_D, \overline{\psi}_D)} = \frac{\int\limits_{\Omega_h} |\nabla\psi_h|^2\, dx}{\int\limits_{\mathcal{D}_h} |\nabla\psi_h|^2\, dx}, \tag{53}$$

where

$$\overline{\psi} = \begin{bmatrix} \overline{\psi}_D \\ \overline{\psi}_0 \end{bmatrix}$$

such that

$$\mathrm{A}_{0D}\overline{\psi}_D + \mathrm{A}_{00}\overline{\psi}_0 = 0,$$

and $\psi_h \in V_h$. The vector $\overline{\psi}_0 \in \mathbb{R}^{n_0}$ corresponds to an FEM function $\psi_{0,h} \in V_h|_{\Omega_h\setminus\mathcal{D}_h}$ called the *continuous h-harmonic extension* of $\psi_{D,h} \in V_h|_{\mathcal{D}_h}$ from $\mathcal{D}_h$ into $\Omega_h\setminus\mathcal{D}_h$, and the FEM function $\psi_{D,h}$ corresponds to the vector $\overline{\psi}_D \in \mathbb{R}^n$. Note that $\psi_{0,h}$ is the solution of the following variational finite element problem:

$$\text{find } u_h \in V_h|_{\Omega_h\setminus\mathcal{D}_h} \text{ satisfying } u_h = \psi_{D,h} \text{ on } \partial\mathcal{D}_h \text{ such that}$$
$$\int\limits_{\Omega_h\setminus\mathcal{D}_h} |\nabla u_h|^2\, dx = \min_{\substack{v_h \in V_h|_{\Omega_h\setminus\mathcal{D}_h} \\ v_h|_{\partial\mathcal{D}_h} = \psi_{D,h}}} \int\limits_{\Omega_h\setminus\mathcal{D}_h} |\nabla v_h|^2\, dx.$$

Subsequently, we will write $\mathcal{D}$ instead of $\mathcal{D}_h$, and $\mathcal{D}^s$ instead of $\mathcal{D}_h^s$, $s \in \{1, \ldots, m\}$ because they are identical by virtue of the foregoing assumptions.

To estimate the value of a_0 in (38) from below, we consider the eigenvalue problem (52) using the spectral decomposition of $\mathrm{B}_{\mathrm{s}} \in \mathbb{R}^{n_s}$, $s \in \{1, \ldots, m\}$ that comes from

$$\mathrm{B}_{\mathrm{s}}\overline{w} = \lambda \mathrm{M}_{\mathrm{s}}\overline{w}, \tag{54}$$

for

$$\mathrm{B}_{\mathrm{s}} = \mathrm{M}_{\mathrm{s}}\mathrm{W}_{\mathrm{s}}\Lambda_s\mathrm{W}_{\mathrm{s}}^T\mathrm{M}_{\mathrm{s}}, \tag{55}$$

with

$$\mathrm{W}_{\mathrm{s}} = \left[\overline{w}_s^1, \ldots, \overline{w}_s^{n_s}\right], \quad \text{and} \quad \Lambda_s = \mathrm{diag}\left\{\lambda_s^1, \ldots, \lambda_s^{n_s}\right\},$$

where $0 = \lambda_s^1 < \lambda_s^2 \leqslant \ldots \leqslant \lambda_s^{n_s}$ are the eigenvalues in (54) and $\overline{w}_s^1, \ldots, \overline{w}_s^{n_s}$ are the corresponding M_{s}-orthonormal eigenvectors, $s \in \{1, \ldots, m\}$. We define the matrices

$$\hat{\mathrm{B}}_{\mathrm{s}} = \mathrm{M}_{\mathrm{s}}{}^{\frac{1}{2}}\mathrm{W}_{\mathrm{s}}\Lambda_s\mathrm{W}_{\mathrm{s}}{}^T\mathrm{M}_{\mathrm{s}}{}^{\frac{1}{2}} \tag{56}$$

and

$$\hat{\mathrm{B}}_s^{\frac{1}{2}} = \mathrm{M}_{\mathrm{s}}{}^{\frac{1}{2}}\mathrm{W}_{\mathrm{s}}\Lambda_s^{\frac{1}{2}}\mathrm{W}_{\mathrm{s}}{}^T\mathrm{M}_{\mathrm{s}}{}^{\frac{1}{2}}. \tag{57}$$

It is obvious that the $\hat{\mathrm{B}}_s^{\frac{1}{2}}$'s are symmetric positive-semidefinite matrices and $\hat{\mathrm{B}}_s^{\frac{1}{2}}\hat{\mathrm{B}}_s^{\frac{1}{2}} = \hat{\mathrm{B}}_s$, $s \in \{1, \ldots, m\}$. We also observe that $\overline{w}_s^1 \in \ker \mathrm{B}_s$ and is precisely the one that is given by (24). In addition, we define the matrices

$$\hat{\mathrm{B}}_{d,s}^{\frac{1}{2}} = \hat{\mathrm{B}}_s^{\frac{1}{2}} + \frac{1}{d_s}\mathrm{M}_s^{\frac{1}{2}}\overline{w}_s^1 \otimes \mathrm{M}_s^{\frac{1}{2}}\overline{w}_s^1,$$

where d_s, $s \in \{1, \dots, m\}$, was introduced in (24). Straightforward multiplications show that

$$\hat{\mathrm{B}}_s^{\frac{1}{2}} \hat{\mathrm{B}}_{d,s}^{\frac{1}{2}} = \hat{\mathrm{B}}_{d,s}^{\frac{1}{2}} \hat{\mathrm{B}}_s^{\frac{1}{2}} = \hat{\mathrm{B}}_s, \quad s \in \{1, \dots, m\}.$$

The latter observation guarantees that eigenvalue problem (52) is equivalent to the eigenvalue problem

$$\mathbf{M}^{\frac{1}{2}} \hat{\boldsymbol{\mathcal{B}}}^{\frac{1}{2}} \hat{\boldsymbol{\mathcal{B}}}_d^{\frac{1}{2}} \mathbf{M}^{\frac{1}{2}} \mathrm{S}_{00}^{-1} \mathbf{M}^{\frac{1}{2}} \hat{\boldsymbol{\mathcal{B}}}_d^{\frac{1}{2}} \hat{\boldsymbol{\mathcal{B}}}^{\frac{1}{2}} \mathbf{M}^{\frac{1}{2}} \overline{w} = \mu \boldsymbol{\mathcal{B}}_D \overline{w}, \tag{58}$$

and

$$\hat{\boldsymbol{\mathcal{B}}}^{\frac{1}{2}} = \mathrm{diag}\left(\hat{\mathrm{B}}_1^{\frac{1}{2}}, \dots, \hat{\mathrm{B}}_m^{\frac{1}{2}} \right), \quad \hat{\boldsymbol{\mathcal{B}}}_d^{\frac{1}{2}} = \mathrm{diag}\left(\hat{\mathrm{B}}_{d,1}^{\frac{1}{2}}, \dots, \hat{\mathrm{B}}_{d,m}^{\frac{1}{2}} \right),$$

are $m \times m$ block diagonal matrices. Clearly, the minimal eigenvalue in (58) is bounded from below by the minimal eigenvalue of the matrix

$$\hat{\boldsymbol{\mathcal{B}}}_d^{\frac{1}{2}} \mathbf{M}^{\frac{1}{2}} \mathrm{S}_{00}^{-1} \mathbf{M}^{\frac{1}{2}} \hat{\boldsymbol{\mathcal{B}}}_d^{\frac{1}{2}},$$

which is equal to the minimal eigenvalue of the similar matrix $\mathrm{S}_{00}^{-1} \boldsymbol{\mathcal{B}}_d$ with

$$\boldsymbol{\mathcal{B}}_d = \mathbf{M}^{\frac{1}{2}} \hat{\boldsymbol{\mathcal{B}}}_d \mathbf{M}^{\frac{1}{2}} = \mathrm{diag}\left(\mathrm{B}_1 + \mathrm{Q}_1, \dots, \mathrm{B}_s + \mathrm{Q}_\mathrm{m} \right),$$

where Q_s, $s \in \{1, \dots, m\}$ is defined in (24). If $(\mu, \overline{w})$ is an eigenpair of the matrix $\mathrm{S}_{00}^{-1} \boldsymbol{\mathcal{B}}_d$, then arguments similar to those used to obtain (53) yield

$$\mu = \max_{\substack{v_h \in V_h|_{\Omega_h \setminus D} \\ v_h|_{\partial D} = w_h}} \frac{\int\limits_D |\nabla w_h|^2 dx + \sum\limits_{s=1}^m \frac{1}{d_s^2} \left[\int\limits_{D^s} w_h dx \right]^2}{\int\limits_D |\nabla w_h|^2 dx + \int\limits_{\Omega_h \setminus D} |\nabla v_h|^2 dx} \geqslant \frac{\|w_h\|_d^2}{\|w_h\|_d^2 + \int\limits_{\Omega_h \setminus D} |\nabla v_h|^2 dx}, \tag{59}$$

for any $v_h \in V_h|_{\Omega_h \setminus D}$, such that $v_h = w_h$ on ∂D, where

$$\|w_h\|_d^2 = \int\limits_D |\nabla w_h|^2 dx + \sum_{s=1}^m \frac{1}{d_s^2} \left[\int\limits_{D^s} w_h \, dx \right]^2. \tag{60}$$

Following the works of Kuznetsov,[17,22] we embed subdomains D^s into subdomains $\tilde{D}^s$ with the conforming boundary $\tilde{\Gamma} = \partial \tilde{D}^s$ (see Figure 2) so that

$$\min_{x \in \overline{D}^s, \ y \in \tilde{D}^s \cup \Gamma} |x - y| \geqslant c d_s, \tag{61}$$

with a positive constant c independent of d_s, $s \in \{1, \dots, m\}$. We assume that $\tilde{D}^s \cap \tilde{D}^t = \varnothing$ for any $s \neq t, s, t \in \{1, \dots, m\}$. We define $\tilde{D} = \bigcup_{s=1}^m \tilde{D}^s$ and assume that v_h in (59) vanishes in $\Omega_h \setminus \tilde{D}$. Consequently, we obtain the following estimate:

$$\mu \geqslant \min_{s \in \{1, \dots, m\}} \frac{\|w_{h,s}\|_{d,s}^2}{\|w_{h,s}\|_{d,s}^2 + \int\limits_{\tilde{D}^s \setminus D^s} |\nabla v_h|^2 dx},$$

for any $v_h \in V_h|_{\tilde{D}^s \setminus D^s}$ such that $v_h|_{\Gamma_s} = w_{h,s}$, $v_h|_{\tilde{\Gamma}_s} = 0$, where $w_{h,s} := w_h|_{D^s}$, and

$$\|w_{h,s}\|_{d,s}^2 = \int\limits_{D^s} |\nabla w_{h,s}|^2 dx + \frac{1}{d_s^2} \left[\int\limits_{D^s} w_{h,s} dx \right]^2.$$

If we assume that, for any $w_{h,s} \in V_s$, its finite element extension $\tilde{w}_h \in V_h|_{\tilde{D}^s \setminus D^s}$ with $\tilde{w}_h|_{\Gamma_s} = w_{h,s}$, $\tilde{w}_h|_{\tilde{\Gamma}_s} = 0$ exists such that

$$\int\limits_{\tilde{D}^s \setminus D^s} |\nabla \tilde{w}_h|^2 dx \leqslant C^2 \|w_{h,s}\|_{d,s}^2, \tag{62}$$

with a positive constant C independent of Ω_h and values of d_s, $s \in \{1, \dots, m\}$, then we arrive at the estimate

$$\mu \geqslant \frac{1}{1 + C^2}. \tag{63}$$

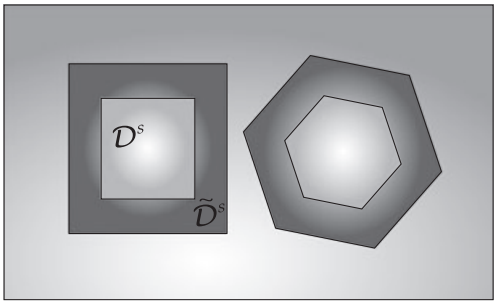


FIGURE 2 An example of D^s and $\tilde{D}^s$

The existence of norm-preserving finite element extensions on quasi-uniform regular-shaped triangular meshes was established in the work of Toselli et al.[23] To utilize the latter result to (62), we have to assume that the mesh Ω_h is quasi-uniform and regular shaped in subdomains $\tilde{D}^s \backslash \overline{D}^s$ and need to apply the transformation $x' = \frac{1}{d_s} x$ for each of the subdomains $\tilde{D}^s$, as it was proposed in the work of Kuznetsov,[17] $s \in \{1, \ldots, m\}$.

Thus, under the assumptions made, the estimate

$$a_0 \geqslant \frac{1}{1 + C^2}$$

holds, where C is a positive constant independent on Ω_h and values of d_s, $s \in \{1, \ldots, m\}$.

Hence, we have proven the following result.

Theorem 1. *If the mesh Ω_h be regularly shaped and quasi-uniform, and distances between $\mathcal{D}^s$ and $\mathcal{D}^t$ satisfy (61) with a constant c independent of Ω_h as well as the shape and location of inclusions, then the lower $a_0 > 0$ and upper b_0 estimates, respectively, for the eigenvalues of the matrix $\mathcal{H}_S \mathbf{S}_0 \equiv \mathcal{H}_S \mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T$ satisfy*

$$a_0 \geqslant \frac{1}{1 + C^2}, \quad b_0 \leq 1,$$

with a positive constant C independent of Ω_h and values of d_s, $s \in \{1, \ldots, m\}$ from the norm-preserving extension theorem (62).

Remark 2. There is an alternative proof of the lower estimate for μ given by (63) (see the work of Kuznetsov[22]) that does not use the algebraic technique (54)–(58) proposed in this paper.

4.2 | Eigenvalue estimates for the matrix $\mathcal{H}\mathcal{A}_\varepsilon$

In this section, we assume that the assumptions made in the end of the Section 4.1 are still valid, namely, the mesh Ω_h in $\tilde{D}^s$ is regularly shaped and quasi-uniform, $s \in \{1, \ldots, m\}$, and the distances between $\mathcal{D}^s$ and $\mathcal{D}^t$ satisfy (61) with a constant c independent of Ω_h, as well as the shape and location of inclusions. In other words, we assume that

$$\mathbf{S}_0 \leqslant \boldsymbol{\mathcal{B}}_D \leqslant (1 + C^2)\mathbf{S}_0, \tag{64}$$

where C^2 is a positive constant independent of Ω_h, and the shape and locations of $\mathcal{D}^s$, $s \in \{1, \ldots, m\}$. We now consider the eigenvalue problem

$$\boldsymbol{\mathcal{A}}_\varepsilon \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix} = \mu \boldsymbol{\mathcal{H}}_0^{-1} \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix} \tag{65}$$

and two additional eigenvalue problems

$$\hat{\boldsymbol{\mathcal{A}}} \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix} = \hat{\mu} \boldsymbol{\mathcal{H}}_0^{-1} \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix}, \tag{66}$$

$$\check{\boldsymbol{\mathcal{A}}} \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix} = \check{\mu} \boldsymbol{\mathcal{H}}_0^{-1} \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix}, \tag{67}$$

with the matrices

$$\hat{\boldsymbol{\mathcal{A}}} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\mathbf{Q} \end{bmatrix}, \quad \text{and} \quad \check{\boldsymbol{\mathcal{A}}} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -r_{\max}\mathbf{S}_0 - \mathbf{Q} \end{bmatrix},$$

respectively, where

$$r_{\max} = (1 + C^2)\varepsilon_{\max},$$

and

$$\boldsymbol{\mathcal{H}}_0^{-1} = \begin{pmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T + \mathbf{Q} \end{pmatrix}.$$

It is obvious that

$$\check{\boldsymbol{\mathcal{A}}} \leqslant \boldsymbol{\mathcal{A}}_\varepsilon \leqslant \hat{\boldsymbol{\mathcal{A}}},$$

and three eigenproblems (65), (66), and (67) have equal numbers of negative and positive eigenvalues. It is also obvious that all three eigenproblems have the same multiplicity of the eigenvalue $\mu = \check{\mu} = \hat{\mu} = 1$ and the underlying eigenvectors $\begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix}$ satisfy the conditions

$$\overline{v} \in \ker\ \mathbf{B}, \quad \overline{w} \in \ker\ \mathbf{B}^T = \ker\ \boldsymbol{\mathcal{B}}_D. \tag{68}$$

The latter condition in (68) implies that, for the eigenvalues $\mu, \check{\mu}, \hat{\mu}$ not equal to one in (65)–(67), we can impose additional conditions on the vector $\overline{w} \in \mathbb{R}^n$:

$$(\mathbf{M}\overline{w}, \overline{\xi}) = 0, \quad \forall \overline{\xi} \in \ker \boldsymbol{\mathcal{B}}_D. \tag{69}$$

Assume that $\hat{\mu} \neq 1$ in (66) and $\check{\mu} \neq 1$ in (67); then, eliminating the vector $\overline{v} \in \mathbb{R}^N$ in (66) and (67) (see also the work of Kuznetsov[24]) yields the equations

$$-\frac{1}{1-\hat{\mu}} = \hat{\mu}, \quad \text{and} \quad -\frac{1}{1-\check{\mu}} - r_{\max} = \check{\mu},$$

respectively. It follows that each of eigenproblems (66) and (67) under the condition (69) has only two different eigenvalues

$$\hat{\mu}_{1,2} = \frac{1 \mp \sqrt{5}}{2}, \quad \text{and} \quad \check{\mu}_{1,2} = \frac{1 - r_{\max} \mp \sqrt{(1 - r_{\max})^2 + 4\,(1 + r_{\max})}}{2}, \tag{70}$$

respectively.

Remark 3. It is obvious that $\check{\mu}_1$ tends to $\hat{\mu}_1 = \frac{1}{2}(1 - \sqrt{5})$ and $\check{\mu}_2$ tends to $\hat{\mu}_2 = \frac{1}{2}(1 + \sqrt{5})$ as $\varepsilon_{\max}$ tends to zero.

Straightforward analysis of (70) shows that

$$\check{\mu}_1 < \hat{\mu}_1 < 0 < \check{\mu}_2 < \hat{\mu}_2.$$

Using inequalities (64) and the results of the work of Bellman,[25] we conclude that all eigenvalues of (66), which are not equal one, belong to the union of two disjoint segments

$$[\check{\mu}_1, \hat{\mu}_1] \cup [\check{\mu}_2, \hat{\mu}_2]\,,$$

with the endpoints independent of Ω_h, and the shape and location of the inclusions (see the assumption in the beginning of this section). Simple analysis shows that $\check{\mu}_2 > 1$ for any $\varepsilon_{\max} \geqslant 0$. To this end, we conclude that all the eigenvalues of (65) belong to the set

$$[\check{\mu}_1, \hat{\mu}_1] \cup [1, \hat{\mu}_2]\,.$$

Now, we consider the eigenvalue problem

$$\boldsymbol{\mathcal{A}}_\varepsilon \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix} = \mu \boldsymbol{\mathcal{H}}^{-1} \begin{bmatrix} \overline{v} \\ \overline{w} \end{bmatrix}, \tag{71}$$

where

$$\boldsymbol{\mathcal{H}}^{-1} = \begin{bmatrix} \mathcal{H}_{\mathrm{A}}^{-1} & 0 \\ 0 & \boldsymbol{\mathcal{B}}_D + \mathbf{Q} \end{bmatrix}, \tag{72}$$

and $\mathcal{H}_{\mathrm{A}}$ is a symmetric positive-definite matrix satisfying the condition

$$\beta_1 \mathbf{A} \leqslant \mathcal{H}_{\mathrm{A}}^{-1} \leqslant \beta_2 \mathbf{A}, \tag{73}$$

with positive constants β_1 and β_2. We assume that β_1 and β_2 are independent of Ω_h. For example, $\mathcal{H}_{\mathrm{A}}^{-1}$ could be a Bramble–Pasciak–Xu (BPX) or an algebraic multigrid (AMG) preconditioner.[14,18,26,27] Using (64) and (72), we obtain

$$\alpha_{\min} \boldsymbol{\mathcal{H}}_0^{-1} \leqslant \boldsymbol{\mathcal{H}}^{-1} \leqslant \alpha_{\max} \boldsymbol{\mathcal{H}}_0^{-1},$$

where

$$\alpha_{\min} = \min\left\{\beta_1; \frac{1}{1 + C^2}\right\}, \quad \text{and} \quad \alpha_{\max} = \max\{\beta_2; 1\}.$$

Straightforward analysis shows that the eigenvalues of the matrix $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon$ belong to the set

$$[C_1, C_2] \cup [C_3, C_4],$$

where

$$C_1 = \frac{\check{\mu}_1}{\alpha_{\min}} \leqslant C_2 = \frac{\hat{\mu}_1}{\alpha_{\max}} < 0, \quad \text{and} \quad C_4 = \frac{\hat{\mu}_2}{\alpha_{\min}} > C_3 = \frac{1}{\alpha_{\max}} > 0. \tag{74}$$

Thus, we have established the following result.

Theorem 2. *If the mesh Ω_h be regularly shaped and quasi-uniform, and distances between $\mathcal{D}^s$ and $\mathcal{D}^t$ satisfy (61) with a constant c independent of Ω_h as well as the shape and location of inclusions, then the eigenvalues of the matrix $(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\varepsilon)^2$ belong to the segment $[a^2, b^2]$, where*

$$a = \min\{|C_2|; C_3\}, \quad b = \max\{|C_1|; C_4\},$$

where C_i, $i \in \{1, \dots, 4\}$, are given by (74).

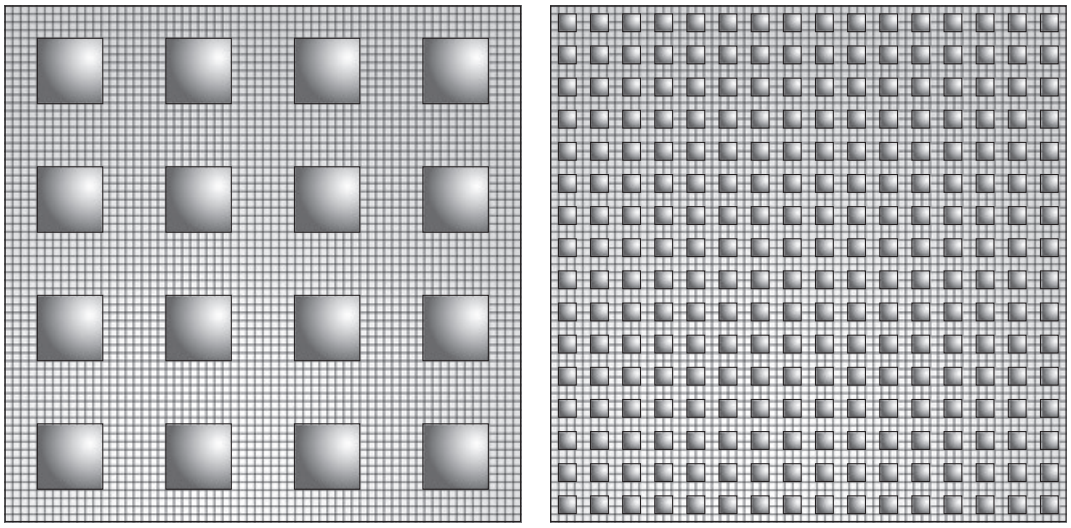
We note that, under the assumption made, the values of a and b are independent of Ω_h, as well as the shape and location of inclusions $\mathcal{D}^s$ in Ω, $s \in \{1, \dots, m\}$.

Remark 4. The results of this section can be easily extended to the case of the 3D diffusion problem, as well as to the problems with a nonzero reaction coefficient and to different types of boundary conditions (Neumann, Robin, and mixed).

5 | NUMERICAL RESULTS

To evaluate and verify methods proposed in Sections 2 and 3, and the theoretical results justified in Section 4, we consider the following simple model problem. Let Ω be a unit square, and let Ω_h be a triangulated square mesh with mesh step size $h = \frac{1}{\sqrt{N}-1}$. We consider two types of particles' distribution in Ω, $s \in \{1, \dots, m\}$. The first one, called *"periodic,"* is shown in Figures 3A and 3B. The second one, called *"random,"* is obtained by removing $\hat{m} < \bar{m}$ inclusions randomly chosen from the periodic array of $\bar{m}$ particles so that $m = \bar{m} - \hat{m}$ (see Figure 4). The values of ε_s in $\mathcal{D}^s$, $s \in \{1, \dots, m\}$, are chosen either randomly from the segment $[\varepsilon_{\min}, 10^{-2}]$, where $\varepsilon_{\min} < 1$, or uniformly $\varepsilon_s = \varepsilon$, $s \in \{1, \dots, m\}$.

In our numerical tests, the inclusions are represented by $d \times d$ squares separated by the distance $d \equiv d_s, s \in \{1, \dots, m\}$, between neighboring inclusions so that the minimal distance between the inclusions and the boundary $\partial\Omega$ equals $d/2$ as shown in Figure 3B.



(a) (b)

FIGURE 3 Periodic distributions of particles. (a) $m = 16$, $h = \frac{1}{64}$, $d = 8h$. (b) $m = 256$, $h = \frac{1}{64}$, $d = 2h$

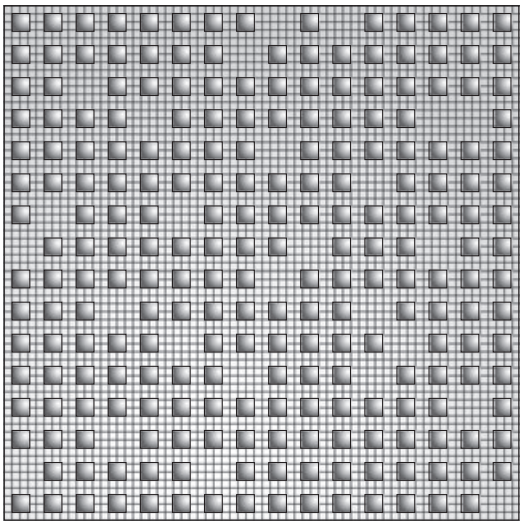


FIGURE 4 Random distribution of particles ($m = 230$, $h = \frac{1}{64}$, $d = 2h$)

The matrix $\mathcal{H}_{\mathrm{A}}$ is the W-cycle algebraic multigrid preconditioner, proposed and investigated in the works of Kuznetsov.[27,28] It was shown in the work of Kuznetsov[28] that the eigenvalues of the matrix $\mathcal{H}_{\mathrm{A}}\mathbf{A}$ lie in the segment $\left[\frac{1}{2}(3-\sqrt{3}), \frac{3}{2}(1+\sqrt{3})\right]$, that is, we obtain a union of the form (73) by setting

$$\beta_1 = \frac{1}{2}\left(3-\sqrt{3}\right), \quad \beta_2 = \frac{3}{2}\left(1+\sqrt{3}\right).$$

Therefore, the number of arithmetical operations (flops) for calculation of the matrix–vector product $\mathcal{H}_{\mathrm{A}}\overline{\xi}$ with $\overline{\xi} \in \mathbb{R}^N$ is bounded above by $5 \times N$; hence, the arithmetical costs of multiplication of a vector by $\mathcal{H}_{\mathrm{A}}$ and $\mathbf{A}$ are almost equal.

The main goal of our numerical experiments is to evaluate the minimal number of iterations sufficient for the minimization of initial errors in δ^{-1} times, $\delta < 1$. To this end, in our numerical tests, we consider the homogeneous systems with a randomly chosen initial guess.

For the PU method (34)–(35), the stopping criteria were

$$\|\overline{p}^k\|_{\mathbf{S}_\varepsilon} \leqslant \delta \|\overline{p}^0\|_{\mathbf{S}_\varepsilon}, \tag{75}$$

whereas for the PL method (41)–(43) and the PCG method (46)–(47), the stopping criteria were

$$\|\overline{z}^k\|_{\mathbf{K}_\varepsilon} \leqslant \delta \|\overline{z}^0\|_{\mathbf{K}_\varepsilon}. \tag{76}$$

In Table 1, we display the number of PCG iterations with the preconditioner $\mathcal{H}_{\mathrm{A}}$ mentioned at the beginning of this section for the homogeneous system

$$\mathbf{A}\overline{x} = \overline{0},$$

and randomly chosen initial guesses $\overline{x}^0$. Here, the stopping criteria were

$$\|\overline{x}^k\|_{\mathbf{A}} \leqslant \delta \|\overline{x}^0\|_{\mathbf{A}}.$$

We observe that 12 iterations are sufficient to minimize the $\mathbf{A}$-norm of the error in 10^7 times.

In Table 2, we display the number of iterations of the PU method with $\delta = 10^{-6}$, which is independent of a random choice of $\varepsilon \in [\varepsilon_{\min}, 10^{-2}]$ in the algebraic system, and the distribution of the inclusions. To perform the product $\mathcal{H}_{\mathrm{A}}\overline{\xi}$, $\overline{\xi} \in \mathbb{R}^N$, we used 12 iterations of the PCG method for systems with the matrix $\mathbf{A}$.

In Tables 3 and 4, we display the number of iterations for the PL and PCG methods described in Sections 3.2 and 3.3, respectively. The tests are done for various numbers of particles m, and the two types of particles' distribution: periodic and random ones. As it is clearly seen, the number of iterations does not depend on $\varepsilon_{\min}$ nor on distribution of the particles, their number, or the mesh size h in Ω_h.

Using results of the tests presented in Tables 2, 3 , and 4, we compare all three respective methods (PU, PL, and PCG) in terms of their arithmetical costs in Table 5. We observe that, due to Remark 1, the major computational effort is associated with multiplications by the matrices $\mathcal{H}_{\mathrm{A}}$ and $\mathbf{A}$. Hence, this table presents the number of multiplications by $\mathcal{H}_{\mathrm{A}}$ and $\mathbf{A}$

TABLE 1 The number of preconditioned conjugate gradient iterations

N / δ	65,025	261,121	1,046,529	4,190,209
10^{-2}	4	4	4	4
10^{-4}	7	7	7	7
10^{-6}	10	10	10	10
10^{-7}	12	12	12	12
10^{-8}	14	14	14	14

TABLE 2 The number of preconditioned Uzawa iterations

m	65,536		16,384		4,096	
$\epsilon_{\min}$	Period	Rand	Period	Rand	Period	Rand
10^{-2}	11	11	11	10	10	10
10^{-4}	11	11	11	11	10	10
10^{-6}	11	11	11	11	10	10

TABLE 3 The number of preconditioned Lanczos iterations

m	65,536		16,384		4,096	
$\epsilon_{\min}$	Period	Rand	Period	Rand	Period	Rand
10^{-2}	40	40	43	43	46	44
10^{-4}	40	40	44	44	46	46
10^{-6}	40	40	44	44	46	46

TABLE 4 The number of preconditioned conjugate gradient iterations

m	65,536		16,384		4,096	
$\epsilon_{\min}$	Period	Rand	Period	Rand	Period	Rand
10^{-2}	90	90	88	88	89	89
10^{-4}	93	93	92	92	92	92
10^{-6}	93	93	92	92	92	92

TABLE 5 Arithmetical cost

Method / $\epsilon_{\min}$	PL	PCG	PU
10^{-2}	**44**	176	120
10^{-4}	**46**	184	132
10^{-6}	**46**	184	132

Note. PL = preconditioned Lanczos; PCG = preconditioned conjugate gradient; PU = preconditioned Uzawa.

needed to solve the underlying systems with accuracy δ due to criteria (75) and (76). Based on these results, we may conclude that, for the above test problems, the PL method is almost three times faster than the PU method, and almost four times faster than the PCG method. Obviously, the results and conclusions may be different for other test problems and different choices of a preconditioner $\mathcal{H}_{\mathrm{A}}$ for the matrix $\mathbf{A}$.

Finally we note that Table 2 presents results for PU method that uses $\mathcal{H}_{\mathrm{A}} = \mathbf{A}$, which is precisely the discrete Laplacian. Because of that, this method converges in fewer iterations than the other two methods. The difference between the results of Table 3 and Table 4 are also consistent: For the latter one, we had to execute PCG for the linear system with the matrix $(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_{\varepsilon})^2$ that requires roughly twice as much operations than the corresponding PL method for the linear system with $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_{\varepsilon}$.

6 | CONCLUSIONS

We have proposed three preconditioned iterative methods for solving a linear system of the saddle-point type arising in the discretization of the diffusion problem (5) that involves large variation of its coefficient (6). The latter feature is typically called *high contrast*. The main theoretical outcome presented in Theorem 2 yields that, with the proposed preconditioner $\boldsymbol{\mathcal{H}}$, the condition numbers of the preconditioned matrix $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_\epsilon$ are of $O(1)$. This implies robustness of the proposed preconditioners. The assumption about regularly shaped and quasi-uniform mesh Ω_h is needed to apply the norm-preserving extension theorem of Widlund[29] that yields independence of convergence rates of the mesh size h. In order to claim independence of convergence rates of the diameter of $\mathcal{D}^s$, $s \in \{1, \dots, m\}$, and their locations, we need assumption (61). Our numerical experiments based on simple test scenarios presented in Section 5 confirm theoretical findings of this paper and demonstrate convergence rates of the proposed iterative schemes to be independent of the contrast, discretization size, and the number of inclusions and their sizes. The very important feature of the discussed procedures is that they are computationally inexpensive with the arithmetical cost being proportional to the size of the linear system. This makes the proposed methodology attractive for the type of applications that deals with high-contrast particles.

ACKNOWLEDGEMENT

Y. Gorb has been supported by NSF Grant DMS-1350248. There are no conflicts of interest to this work.

ORCID

Yuliya Gorb https://orcid.org/0000-0002-9968-4494

REFERENCES

1. Aarnes J, Hou TY. Multiscale domain decomposition methods for elliptic problems with high aspect ratios. Acta Math Appl Sin. 2002;18(1):63–76.
2. Aksoylu B, Graham IG, Klie H, Scheichl R. Towards a rigorously justified algebraic preconditioner for high-contrast diffusion problems. Comput Vis Sci. 2008;11(4-6):319–331.
3. Borcea L, Gorb Y, Wang Y. Asymptotic approximation of the Dirichlet to Neumann map of high contrast conductive media. Multiscale Model Simul. 2014;12(4):1494–1532.
4. Gorb Y, Kurzanova D. Heterogeneous domain decomposition method for high contrast dense composites. J Comput Appl Math. 2018;337:135–149.
5. Farhat C, Lesoinne M, LeTallec P, Pierson K, Rixen D. FETI-DP: a dual-primal unified FETI method—Part I. A faster alternative to the two-level FETI method. Int J Numer Methods Eng. 2001;50(7):1523–1544.
6. Galvis J, Efendiev Y. Domain decomposition preconditioners for multiscale flows in high-contrast media. SIAM Multiscale Model Simul. 2010;8(4):1461–1483.
7. Kraus J. Additive Schur complement approximation and application to multilevel preconditioning. SIAM J Sci Comput. 2012;34(6):A2872–A2895.
8. Ewing RE, Lazarov RD, Lu P, Vassilevski PS. Preconditioning indefinite systems arising from mixed finite element discretization of second-order elliptic problems. In: Preconditioned conjugate gradient methods. Berlin, Germany: Springer, 1990; p. 28–43.
9. Powell CE, Silvester D. Optimal preconditioning for Raviart-Thomas mixed formulation of second-order elliptic problems. SIAM J Matrix Anal Appl. 2004;25:718–738.
10. Rusten T, Winther R. A preconditioned iterative method for saddlepoint problems. SIAM J Matrix Anal Appl. 1991;13(3):887–904.
11. Wathen A, Silvester D. Fast iterative solution of stabilised stokes systems. Part I: Using simple diagonal preconditioners. SIAM J Numer Anal. 1991;30(3):630–649.
12. Lanczos C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. J Res Natl Bureau Stand. 1950;45:255–282.
13. Paige CC. Computational variants of the Lanczos method for the eigenproblem. J Inst Math Appl. 1972;10:373–381.
14. Brenner SC, Scott LR. The mathematical theory of finite element methods. 3rd ed. Vol. 15. New York, NY: Springer; 2008. Texts in applied mathematics.
15. Gorb Y, Kurzanova D, Kuznetsov Y. A robust preconditioner for high-contrast problems. arXiv:1801.01578; 2018.
16. Kuznetsov YA. New iterative methods for singular perturbed positive definite matrices. Russian J Numer Anal Math Modelling. 2000;15(1):65–71.
17. Kuznetsov Y. Preconditioned iterative methods for algebraic saddle-point problems. J Numer Math. 2009;17(1):67–75.
18. Bramble JH, Pasciak JE, Xu J. Parallel multilevel preconditioners. Math Comp. 1990;55(191):1–22.

19. Glowinski R. Numerical methods for nonlinear variational problems. New York, NY: Springer-Verlag; 1984.
20. Axelsson O. Iterative solution methods. Cambridge, UK: Cambridge University Press; 1994.
21. Marchuk G, Kuznetsov Y. Iterative methods and quadratic functionals. In: Lions JL, Marchuk G, editors. Méthodes de l'informatique–4. 1974. p. 3–132. In French.
22. Kuznetsov Y. New homogenization method for diffusion equations. Russ J Numer Anal Math Modelling. 2018;33(2):85–93.
23. Toselli A, Widlund O. Domain decomposition methods – algorithms and theory. Berlin, Germany: Springer-Verlag; 2005. Springer series in computational mathematics, No. 34.
24. Kuznetsov YA. Efficient iterative solvers for elliptic finite element problems on nonmatching grids. Russian J Numer Anal Math Modelling. 1995;10(3):187–212.
25. Bellman R. Introduction to matrix analysis. 2nd ed. Philadelphia, PA: SIAM; 1997.
26. Bramble JH, Pasciak JE, Vassilev AT. Analysis of the inexact Uzawa algorithm for saddle point problems. SIAM J Numer Anal. 1997;34(3):1072–1092.
27. Kuznetsov Y. Algebraic multigrid domain decomposition methods. Sov Jour Num Meth Math Modelling. 1989;4:351–380.
28. Kuznetsov Y. Multigrid domain decomposition method. In: Third international symposium on domain decomposition methods for partial differential equations. Philadelphia, PA: SIAM, 1990; p. 290–313.
29. Widlund OB. An extension theorem for finite element spaces with three applications. In: Numerical techniques in continuum mechanics: Proceedings of the Second GAMM-Seminar, Kiel, January 17 to 19, 1986. Wiesbaden, Germany: Vieweg+Teubner Verlag, 1987; p. 110–122. Notes on numerical fluid mechanics, No. 16.

How to cite this article: Gorb Y, Kramarenko V, Kuznetsov Y. Preconditioned iterative methods for diffusion problems with high-contrast inclusions. *Numer Linear Algebra Appl*. 2019;e2243. https://doi.org/10.1002/nla.2243