An a posteriori-based adaptive preconditioner for controlling a local algebraic error norm

Ani Anciaux-Sedrakian, Laura Grigori, Zakariae Jorti, Soleiman Yousef

manuscript No.
(will be inserted by the editor)

An a posteriori-based adaptive preconditioner for controlling a local algebraic error norm

A. Anciaux-Sedrakian · L. Grigori · Z. Jorti(✉) · S. Yousef

Received: date / Accepted: date

Abstract This paper introduces an adaptive preconditioner for iterative solution of sparse linear systems arising from partial differential equations with self-adjoint operators. This preconditioner allows to control the growth rate of a dominant part of the algebraic error within a fixed point iteration scheme. Several numerical results that illustrate the efficiency of this adaptive preconditioner with a PCG solver are presented and the preconditioner is also compared with a previous variant in the literature.

Keywords Adaptive preconditioner · Fixed-point iteration · Error's growth rate · Block partitioning

Mathematics Subject Classification (2000) 65F08 · 65F10 · 65N22

1 Introduction

In this paper, we focus on solving linear systems of equations arising from computing the numerical solution of a partial differential equation. The accuracy of this numerical solution is estimated by evaluating the error between the obtained approximate solution and the exact solution. There are several different sources that contribute to this error, the discretization error, the linearization error, and the algebraic error. We

✉ Z. Jorti
IFP Energies nouvelles, 1-4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France
INRIA Paris, Alpines, and Sorbonne Université, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, Paris, France
Tel.: +33-642-755614
E-mail: zjorti@hotmail.fr

A. Anciaux-Sedrakian · S. Yousef
IFP Energies nouvelles, 1-4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

L. Grigori
INRIA Paris, Alpines, and Sorbonne Université, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, Paris, France

are in particular interested in the algebraic error which arises from computing iteratively the solution of the resulting linear system of equations. A means for measuring the accuracy of this result is to evaluate the error, which is the difference between the approximate and the exact solutions. Also in the realm of linear algebra, an *algebraic error* originates from the iterative solution of the linear system of equations resulting from the discretization of the problem. Several works were carried out with the primary objective of identifying the *algebraic error* during the iterative solution of a linear system, raised from a numerical approximation of partial differential equations [8,6,5,7,21–23,18]. Subsequent works used the theory of a posteriori error estimates in order to derive rigorous upper bounds of the global error, which includes the algebraic error [3,2,16,12,10,20]. More recent works focused on deriving appropriate guaranteed upper bound directly on the algebraic error using equilibrated flux reconstructions [20,17,19]. In the context of using a posteriori error estimates, an adaptive preconditioner, which is used in combination with a specific initial guess and based on the estimated local distribution of the algebraic error, is derived in [1]. To the best of our knowledge, that is the first work that focused on preconditioners that take into account the distribution of the error.

In this work, we are interested in solving a linear system $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ by an iterative method. The system arises from the discretization of a partial differential equation for which an error analysis has been undertaken and a posteriori algebraic error estimates are available. The magnitudes of these estimates indicate that the algebraic error is concentrated on some specific areas of the domain, which leads to a certain error-driven domain decomposition. If the indices of the vertices where the algebraic error is high are gathered in a subset L, whereas the remaining vertex indices are grouped into another subset R, a first step to efficiently reduce the energy norm of the error is obtained, according to [1], by expressing it as a sum of two terms: a first term that is specific to the unknowns of L, called L-term and dominant –according to the information stemming from a posteriori error estimates– and a second term that does not depend on the unknowns of L, called R-term (see [1, Formula (20)]). A second step consists in making the projection of the residual on the unknowns of L nil. This implies that the L-term is cancelled (equal to zero). This lead to the introduction of an adaptive solve procedure and its equivalent preconditioner and initial guess for PCG solver (see [1, Theorem 1]).

In this paper, we introduce an approach that allows to control the evolution of the dominant part of the algebraic error, from one iteration to another of an iterative scheme. We define a seminorm of the error, that we call $\underline{\mathbf{A}}_L$-seminorm, that bounds the dominant part of the algebraic error (localized on L). Then we derive a preconditioner, which we refer to as L-adaptive, that bounds the growth rate of this quantity in a fixed-point iteration scheme. Even though the theoretical results are derived for a fixed-point iteration scheme, our numerical results use preconditioned CG solver for faster convergence. We consider several test cases, in particular those for which the adaptive solve procedure and its equivalent preconditioner, which we refer to as R-adaptive, studied in [1] is not efficient. We observe that the configurations of the L-adaptive preconditioner for which the $\underline{\mathbf{A}}_L$-seminorm of the error is strictly decreasing in a fixed-point iteration scheme, perform well when a PCG solver is used as well. We notice that even though the number of iterations of PCG is reduced when

going from an initial Block Jacobi preconditioner to the R-adaptive preconditioner in [1, Section 5], we still get an improvement by using the L-adaptive one. In fact, the decrease in the number of iterations with the L-adaptive preconditioner is more important in the test cases where the R-adaptive preconditioner is not sufficient to significantly reduce the number of iterations.
This article is organized as follows. Section 2 presents the model problem and recalls the starting assumption on the distribution of the algebraic error, then we analyze in Section 3 the behavior of a partial algebraic error within a fixed-point iteration scheme. In Section 4, we derive specific preconditioners that would ensure that from an iteration i to the next iteration $i+1$, the evolution of that algebraic error localized on the targeted subdomains is controlled. More precisely, the growth rate of the local algebraic error between those two iterations can be bounded by a fixed coefficient. Section 5 makes the connection with the preconditioner derived in [1]. The numerical behavior of both preconditioners is studied on several different test cases in Section 6.

2 Preliminaries

For $1 \leq d \leq 3$, let $\Omega \subset \mathbb{R}^d$ be a non-empty, open, bounded set. We assume Ω is connected. $\overline{\Omega}$, $\mathring{\Omega}$ and $\partial\Omega$ denote respectively the closure, interior and boundary of Ω. Our model problem is a second-order elliptic equation which seeks an unknown function $\underline{u}: \Omega \to \mathbb{R}$ such that:

$$\begin{cases} -\nabla \cdot (\underline{\mathbf{K}} \nabla \underline{u}) = \underline{f} & \text{in } \Omega, \\ \underline{u} = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.1}$$

where $\underline{\mathbf{K}}$ is a positive definite and uniformly bounded diffusion tensor, and $\underline{f}: \Omega \to \mathbb{R}$ is a source term in $\mathrm{L}^2(\Omega)$, which is the space of square integrable functions over Ω. Let $\mathscr{T}_h$ be a matching simplicial mesh of Ω and V_h be the usual finite element space of continuous piecewise p-th order polynomial functions ($p \geq 1$). For simplicity, it is assumed that $\underline{\mathbf{K}}$ and $\underline{f}$ are piecewise constant with respect to the mesh $\mathscr{T}_h$. The linear algebraic system arising from the discretization by finite element method of (2.1) on $\mathscr{T}_h$ using the basis functions of V_h is expressed as:

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}, \tag{2.2}$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a symmetric positive definite (SPD) matrix, $\mathbf{b} \in \mathbb{R}^n$ is the right hand side vector, and $\mathbf{x} \in \mathbb{R}^n$ is the solution vector. The Galerkin solution can then be written as $\underline{u}_h = \sum_{l=1}^{n} x_l \varphi_l \in V_h$, where $(\varphi_l)_{1 \leq l \leq n}$ is a basis of V_h.
We consider solving the system (2.2) by an iterative solver. We denote by $\mathbf{x}^{(i)}$ the approximate solution after i iterations and by $\underline{u}_h^{(i)} = \sum_{l=1}^{n} x_l^{(i)} \varphi_l$ the associated function from V_h. Our main assumption is that an estimation of the local distribution of the error $\|\underline{\mathbf{K}}^{\frac{1}{2}} \nabla(\underline{u}_h - \underline{u}_h^{(i)})\|_{\mathrm{L}^2(K)}$ is available on all mesh elements $K \in \mathscr{T}_h$ ($\overline{\Omega} = \cup \overline{K}$), and

consequently the main domain Ω is decomposed into two disjoint open subdomains (they are aggregates of mesh elements) Ω_1 and Ω_2 such that:

$$\begin{cases} \overline{\Omega}_1 \cup \overline{\Omega}_2 & = \overline{\Omega}, \\ \mathring{\Omega}_1 \cap \mathring{\Omega}_2 & = \emptyset, \end{cases} \tag{2.3}$$

where Ω_1 is composed of the elements with a high algebraic error:

$$\boxed{||\underline{\mathbf{K}}^{1/2}\nabla(\underline{u}_h - \underline{u}_h^{(i)})||^2_{\mathrm{L}^2(\Omega_1)} \gg ||\underline{\mathbf{K}}^{1/2}\nabla(\underline{u}_h - \underline{u}_h^{(i)})||^2_{\mathrm{L}^2(\Omega_2)}.} \tag{2.4}$$

Following the domain decomposition of (2.3), let $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ be the local stiffness matrices for the subdomains Ω_1 and Ω_2, respectively. They are defined by:

$$\mathbf{A}^{(1)}_{jk} = (\underline{\mathbf{K}}^{\frac{1}{2}}\nabla\varphi_k, \underline{\mathbf{K}}^{\frac{1}{2}}\nabla\varphi_j)_{\Omega_1},\ 1 \le j,k \le n,\ \mathrm{supp}\,\varphi_k \cap \Omega_1 \neq \emptyset,\ \mathrm{supp}\,\varphi_j \cap \Omega_1 \neq \emptyset,$$

$$\mathbf{A}^{(2)}_{jk} = (\underline{\mathbf{K}}^{\frac{1}{2}}\nabla\varphi_k, \underline{\mathbf{K}}^{\frac{1}{2}}\nabla\varphi_j)_{\Omega_2},\ 1 \le j,k \le n,\ \mathrm{supp}\,\varphi_k \cap \Omega_2 \neq \emptyset,\ \mathrm{supp}\,\varphi_j \cap \Omega_2 \neq \emptyset.$$

For ease of presentation, we consider the following ordering. The unknowns corresponding to the vertices of $\mathring{\Omega}_1$ are numbered first, those of the interface between the two subdomains second, and those of $\mathring{\Omega}_2$ last. Let $n_L \in \mathbb{N}$ and $n_2 \in \mathbb{N}$ be the number of vertices in $\overline{\Omega}_1$ and $\overline{\Omega}_2$ resp., we define the restriction matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ from the global set of degrees of freedom to the set of degrees of freedom related to $\overline{\Omega}_1$ and to $\overline{\Omega}_2$:

$$\forall(\mathbf{x}_L, \mathbf{x}_R, \mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^{n_L} \times \mathbb{R}^{n-n_L} \times \mathbb{R}^{n-n_2} \times \mathbb{R}^{n_2} :\ \mathbf{R}_1 \cdot \begin{pmatrix} \mathbf{x}_L \\ \mathbf{x}_R \end{pmatrix} = \mathbf{x}_L,\ \mathbf{R}_2 \cdot \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \mathbf{x}_2, \tag{2.5}$$

such that:

- $\mathbf{x}_L$ is the vector containing the degrees of freedom related to $\overline{\Omega}_1$;
- $\mathbf{x}_R$ is the vector containing the degrees of freedom related to $\mathring{\Omega}_2$;
- $\mathbf{x}_1$ is the vector containing the degrees of freedom related to $\mathring{\Omega}_1$;
- $\mathbf{x}_2$ is the vector containing the degrees of freedom related to $\overline{\Omega}_2$.

Therefore, for the same vector $\begin{pmatrix} \mathbf{x}_L \\ \mathbf{x}_R \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$, the entries of $\mathbf{x}_1$ constitute a part of $\mathbf{x}_L$. The same can be said about $\mathbf{x}_R$ and $\mathbf{x}_2$, respectively.
Then we can split the operator $\mathbf{A}$ as follows,

$$\mathbf{A} = \mathbf{A}_p^{(1)} + \mathbf{A}_p^{(2)}, \quad \mathbf{A}_p^{(1)} = \mathbf{R}_1^{\mathrm{T}}\mathbf{A}^{(1)}\mathbf{R}_1, \quad \mathbf{A}_p^{(2)} = \mathbf{R}_2^{\mathrm{T}}\mathbf{A}^{(2)}\mathbf{R}_2. \tag{2.6}$$

$\mathbf{A}_p^{(1)}$ and $\mathbf{A}_p^{(2)}$ are the extensions of the local stiffness matrices (also called Neumann matrices [9]) $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ to the whole domain. They are symmetric positive semidefinite (SPSD).
Furthermore, we obtain the equivalent formulation to (2.4) in the realm of matrices:

$$\boxed{(\mathbf{x} - \mathbf{x}^{(i)})^{\mathrm{T}} \cdot \mathbf{A}_p^{(1)} \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \gg (\mathbf{x} - \mathbf{x}^{(i)})^{\mathrm{T}} \cdot \mathbf{A}_p^{(2)} \cdot (\mathbf{x} - \mathbf{x}^{(i)})}, \tag{2.7}$$

where $\mathbf{x}^{(i)}$ is the approximate solution at iteration i. Note that summing both sides of inequality (2.7) gives the energy norm of the error, $(\mathbf{x}-\mathbf{x}^{(i)})^{\mathrm{T}}\cdot\mathbf{A}\cdot(\mathbf{x}-\mathbf{x}^{(i)})$. This inequality expresses that $\mathbf{A}_p^{(1)}$-seminorm of the error, also referred to as L-norm (see [1, Section 5.1]), is the dominant part of the energy norm of the error. Thus, it is this quantity that should be decreased.

3 Controlling the local algebraic error in fixed-point iteration scheme

In this section we study an approach that allows to maintain the evolution of the left hand side of (2.7) limited, that is to keep it under a given threshold from one iteration to another. This is equivalent to ensuring the following property,

$$\exists\tau>0,\quad \forall\mathbf{i}\in\mathbb{N},\quad (\mathbf{x}-\mathbf{x}^{(i+1)})^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot(\mathbf{x}-\mathbf{x}^{(i+1)})\leq\tau(\mathbf{x}-\mathbf{x}^{(i)})^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot(\mathbf{x}-\mathbf{x}^{(i)}).$$

Thus, considering a fixed-point iteration scheme,

$$\mathbf{x}^{(i+1)}:=\mathbf{x}^{(i)}+\mathbf{M}^{-1}\cdot(\mathbf{b}-\mathbf{A}\cdot\mathbf{x}^{(i)}),\forall i\in\mathbb{N},\tag{3.1}$$

with an arbitrary initial guess $\mathbf{x}^{(0)}$, we seek a preconditioner $\mathbf{M}^{-1}$ that satisfies the property,

$$\exists\tau>0,\ \forall\mathbf{u}\in\mathbb{R}^n:\ \ ((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u})^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u})\leq\tau\mathbf{u}^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot\mathbf{u},\tag{3.2}$$

because we have $\mathbf{x}-\mathbf{x}^{(i+1)}=(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot(\mathbf{x}-\mathbf{x}^{(i)})$ from (3.1).
In what follows, we state several lemmas that are useful for establishing the necessary and sufficient conditions for property (3.2).

Lemma 3.1 *Let $\mathbf{P}\in\mathbb{R}^{n\times n}$ be a symmetric positive semi-definite matrix, $\mathbf{Q}\in\mathbb{R}^{n\times n}$ be an invertible matrix and V be an arbitrary subspace of $\mathbb{R}^n$. The following two assertions are equivalent:*

- $\exists\tau_1>0,\quad\forall\mathbf{u}\in V:\quad((\mathbf{I}-\mathbf{Q})\cdot\mathbf{u})^{\mathrm{T}}\cdot\mathbf{P}\cdot((\mathbf{I}-\mathbf{Q})\cdot\mathbf{u})\leq\tau_1\mathbf{u}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{u}$,
- $\exists\tau_2>0,\quad\forall\mathbf{u}\in V:\quad(\mathbf{Q}\cdot\mathbf{u})^{\mathrm{T}}\cdot\mathbf{P}\cdot(\mathbf{Q}\cdot\mathbf{u})\leq\tau_2\mathbf{u}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{u}$.

Proof We denote by $||\ ||_{\mathbf{P}}:\mathbf{x}\mapsto\sqrt{\mathbf{x}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{x}}$ the seminorm defined by $\mathbf{P}$ on $\mathbb{R}^n$. To prove the equivalence of the assertions, it suffices to notice that for any $\mathbf{u}\in\mathbb{R}^n$ we have:

$$||\mathbf{u}-\mathbf{Q}\cdot\mathbf{u}||_{\mathbf{P}}^2\leq 2||\mathbf{u}||_{\mathbf{P}}^2+2||\mathbf{Q}\cdot\mathbf{u}||_{\mathbf{P}}^2$$

and

$$||\mathbf{Q}\cdot\mathbf{u}||_{\mathbf{P}}^2=||\mathbf{u}-(\mathbf{u}-\mathbf{Q}\cdot\mathbf{u})||_{\mathbf{P}}^2\leq 2||\mathbf{u}||_{\mathbf{P}}^2+2||\mathbf{u}-\mathbf{Q}\cdot\mathbf{u}||_{\mathbf{P}}^2.$$

□

Lemma 3.2 *Let $\mathbf{P}\in\mathbb{R}^{n\times n}$ be a symmetric positive semi-definite matrix and $\mathbf{Q}\in\mathbb{R}^{n\times n}$ be an invertible matrix. If*

$$\exists\tau>0,\quad\forall\mathbf{u}\in\mathbb{R}^n:\quad((\mathbf{I}-\mathbf{Q})\cdot\mathbf{u})^{\mathrm{T}}\cdot\mathbf{P}\cdot((\mathbf{I}-\mathbf{Q})\cdot\mathbf{u})\leq\tau\mathbf{u}^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{u},$$

then $\mathrm{Ker}(\mathbf{P})$ *is invariant of* $\mathbf{Q}$*, i.e.* $\quad\forall\mathbf{v}\in\mathrm{Ker}(\mathbf{P}):\mathbf{Q}\cdot\mathbf{v}\in\mathrm{Ker}(\mathbf{P})$.

Proof We give a proof by contrapositive of this lemma. If $\mathrm{Ker}(\mathbf{P})$ is not invariant of $\mathbf{Q}$, then there exists a vector $\mathbf{u}_0 \in \mathrm{Ker}(\mathbf{P})$ such that $\mathbf{Q}\cdot\mathbf{u}_0 \notin \mathrm{Ker}(\mathbf{P})$. Thus,

$$\forall \tau > 0, \exists \mathbf{u}_0 \in \mathbb{R}^n : ((\mathbf{I}-\mathbf{Q})\cdot\mathbf{u}_0)^{\mathrm{T}}\cdot\mathbf{P}\cdot((\mathbf{I}-\mathbf{Q})\cdot\mathbf{u}_0) = ||\mathbf{Q}\cdot\mathbf{u}_0||^2_{\mathbf{P}} > 0 = \tau(\mathbf{u}_0^{\mathrm{T}}\cdot\mathbf{P}\cdot\mathbf{u}_0).$$

□

The next corollary follows from the above lemma by taking $\mathbf{P} := \mathbf{A}_p^{(1)}$ and $\mathbf{Q} := \mathbf{M}^{-1}\mathbf{A}$.

Corollary 3.1 *Let* $\mathbf{M}^{-1}$ *be a preconditioner of the matrix* $\mathbf{A}$ *that satisfies,*

$$\exists \tau > 0, \quad \forall \mathbf{u} \in \mathbb{R}^n : \quad ((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u})^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u}) \leq \tau\mathbf{u}^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot\mathbf{u},$$

then $\mathrm{Ker}(\mathbf{A}_p^{(1)})$ *is invariant of* $\mathbf{M}^{-1}\mathbf{A}$.

This corollary provides a necessary condition for the property (3.2) to be satisfied. On the other hand, we aim at proving sufficient condition for that property with the following lemma.

Lemma 3.3 *Let* $\mathbf{M}^{-1}$ *be a preconditioner of the matrix* $\mathbf{A}$. *If* $\mathrm{Range}(\mathbf{A}_p^{(1)})$ *is invariant of* $(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})$ *then:*

$$\exists \tau > 0, \forall \mathbf{u} \in \mathrm{Range}(\mathbf{A}_p^{(1)}) : ((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u})^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u}) \leq \tau\mathbf{u}^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot\mathbf{u}.$$

Proof $\mathbf{A}_p^{(1)}$ and $(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})$ are symmetric positive semidefinite. In addition, if $\mathrm{Range}(\mathbf{A}_p^{(1)})$ is invariant of $(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})$, then $(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})$ and $\mathbf{A}_p^{(1)}$, seen as linear operators from $\mathrm{Range}(\mathbf{A}_p^{(1)})$ to $\mathrm{Range}(\mathbf{A}_p^{(1)})$, are symmetric positive semidefinite and symmetric positive definite respectively. In this case, we can introduce the following generalized eigenvalue problem:

$$\text{Find } (\mathbf{y}_k, \mu_k) \in \mathrm{Range}(\mathbf{A}_p^{(1)}) \times \mathbb{R} \text{ such that } (\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{y}_k = \mu_k\mathbf{A}_p^{(1)}\cdot\mathbf{y}_k. \tag{3.3}$$

Since $\mathbf{A}_p^{(1)}$ is an SPD operator on $\mathrm{Range}(\mathbf{A}_p^{(1)})$, the eigenvalues of (3.3) can be chosen so that they form a basis of $\mathrm{Range}(\mathbf{A}_p^{(1)})$ that is both $\mathbf{A}_p^{(1)}$-orthonormal and $(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})$-orthogonal. Let $m = \dim(\mathrm{Range}(\mathbf{A}_p^{(1)})) = \mathrm{rank}(\mathbf{A}_p^{(1)})$ and let $\mathbf{u} \in \mathrm{Range}(\mathbf{A}_p^{(1)})$ then:

$$\begin{aligned}
\mathbf{u} &= \sum_{k=1}^{m} \langle \mathbf{A}_p^{(1)}\cdot\mathbf{u}, \mathbf{y}_k\rangle \mathbf{y}_k \\
\implies (\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u} &= \sum_{k=1}^{m} \langle \mathbf{A}_p^{(1)}\cdot\mathbf{u}, \mathbf{y}_k\rangle (\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{y}_k \\
&= \sum_{k=1}^{m} \mu_k \langle \mathbf{A}_p^{(1)}\cdot\mathbf{u}, \mathbf{y}_k\rangle \mathbf{A}_p^{(1)}\cdot\mathbf{y}_k \\
\implies \mathbf{u}^{\mathrm{T}}\cdot(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\mathbf{A}_p^{(1)}(\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u} &= \sum_{k=1}^{m} \mu_k \langle \mathbf{A}_p^{(1)}\cdot\mathbf{u}, \mathbf{y}_k\rangle^2 \\
&\leq (\max_{1\leq i\leq m} \mu_i)^2 \sum_{k=1}^{m} \langle \mathbf{A}_p^{(1)}\cdot\mathbf{u}, \mathbf{y}_k\rangle^2 = (\max_{1\leq i\leq m} \mu_i)^2 \mathbf{u}^{\mathrm{T}}\cdot\mathbf{A}_p^{(1)}\cdot\mathbf{u}.
\end{aligned}$$

And applying the result of Lemma 3.1 on the last inequality ends the proof. □

It must be emphasised that Lemma 3.3 gives a sufficient condition for Property (3.2) to be true not on the whole space $\mathbb{R}^n$, but only on a subspace of it. However, the error $\mathbf{e}^{(i)} := \mathbf{x} - \mathbf{x}^{(i)}$ is not guaranteed to belong to that subspace as the initial guess $\mathbf{x}^{(0)}$ is arbitrarily chosen. Therefore, this sufficient condition is too restrictive, since its effect is valid only for the iterations when the error lies in the range of $\mathbf{A}_p^{(1)}$. It is also worth mentioning that the necessary condition of Corollary 3.1 and the sufficient condition of Lemma 3.3 do not match. In the sequel, we will derive a second property that is similar to (3.2), that involves a definite matrix, and for which we can derive a sufficient and necessary condition.

4 Deriving a block partitioning and controlling the corresponding algebraic error norm

By proceeding in the same way as in [1, Section 3.3], we replace the sum-splitting of the operator, as in (2.6), by a block-partitioning of the matrix. By considering the ordering of the unknowns described in Section 2, $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ can be expressed as,

$$\mathbf{A}^{(1)} = \begin{pmatrix} \mathbf{F} & \mathbf{E}^{(1)} \\ \mathbf{E}^{(1)^{\mathrm{T}}} & \mathbf{A}_{\mathrm{int}}^{(1)} \end{pmatrix}, \qquad \mathbf{A}^{(2)} = \begin{pmatrix} \mathbf{A}_{\mathrm{int}}^{(2)} & \mathbf{E}^{(2)} \\ \mathbf{E}^{(2)^{\mathrm{T}}} & \mathbf{A}_R \end{pmatrix},$$

where $\mathbf{A}_{\mathrm{int}}^{(1)}$ and $\mathbf{A}_{\mathrm{int}}^{(2)}$ are the blocks related to the vertices in the interface, i.e., the vertices in $\overline{\Omega}_1 \cap \overline{\Omega}_2$. If we take,

$$\mathbf{A}_{LR} = \mathbf{A}_{RL}^{\mathrm{T}} = \begin{pmatrix} \mathbf{0} \\ \mathbf{E}^{(2)} \end{pmatrix}, \qquad \mathbf{A}_L = \mathbf{A}^{(1)} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{\mathrm{int}}^{(2)} \end{pmatrix},$$

then we get the following block-partitioning

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_L & \mathbf{A}_{LR} \\ \mathbf{A}_{RL} & \mathbf{A}_R \end{pmatrix}, \tag{4.1}$$

where the L-part stands for the set of vertices that belong to $\overline{\Omega}_1$ and the R-part for the vertices of $\mathring{\Omega}_2$. If we denote by n_R the number of vertices of $\mathring{\Omega}_2$, then $n_R = n - n_L$. n_L and n_R are equal to the sizes of the diagonal SPD blocks $\mathbf{A}_L$ and $\mathbf{A}_R$, respectively. The right hand side vector $\mathbf{b}$ can be partitioned accordingly, i.e., $\mathbf{b} = \begin{bmatrix} \mathbf{b}_L \\ \mathbf{b}_R \end{bmatrix}$.

We recall that in this case, if we denote,

$$\underline{\mathbf{A}}_L := \mathbf{R}_1^{\mathrm{T}} \mathbf{A}_L \mathbf{R}_1 = \begin{pmatrix} \mathbf{A}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{4.2}$$

then by combining the formulation of (2.7) with the superiority property of [1, Lemma 3], which states that the $\mathbf{A}_L$-norm is higher than the $\mathbf{A}^{(1)}$-seminorm, we have:

$$(\mathbf{x} - \mathbf{x}^{(i)})^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \geq (\mathbf{x} - \mathbf{x}^{(i)})^{\mathrm{T}} \cdot \mathbf{A}_p^{(1)} \cdot (\mathbf{x} - \mathbf{x}^{(i)}) \gg (\mathbf{x} - \mathbf{x}^{(i)})^{\mathrm{T}} \cdot \mathbf{A}_p^{(2)} \cdot (\mathbf{x} - \mathbf{x}^{(i)}). \tag{4.3}$$

Therefore, instead of controlling the dominant $\mathbf{A}_p^{(1)}$-seminorm of the error, we will focus on the $\underline{\mathbf{A}}_L$-seminorm of the error which is larger. To ensure that the evolution of this latter in a fixed-point iteration scheme from one iteration to another is limited and kept below a fixed tolerance, we have to find a preconditioner $\mathbf{M}^{-1}$ such that,

$$\exists \tau > 0,\ \forall \mathbf{u} \in \mathbb{R}^n :\ ((\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) \cdot \mathbf{u})^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot ((\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) \cdot \mathbf{u}) \leq \tau \mathbf{u}^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot \mathbf{u}. \quad (4.4)$$

Lemma 3.2 applied to the matrices $\mathbf{P} := \underline{\mathbf{A}}_L$ and $\mathbf{Q} := \mathbf{M}^{-1}\mathbf{A}$ yields the following corollary.

Corollary 4.1 *Let $\mathbf{M}^{-1}$ be a preconditioner of the matrix $\mathbf{A}$ that satisfies,*

$$\exists \tau > 0, \quad \forall \mathbf{u} \in \mathbb{R}^n : \quad ((\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) \cdot \mathbf{u})^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot ((\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) \cdot \mathbf{u}) \leq \tau \mathbf{u}^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot \mathbf{u},$$

then $\mathrm{Ker}(\underline{\mathbf{A}}_L)$ *is invariant of* $\mathbf{M}^{-1}\mathbf{A}$.

Lemma 4.1 $\mathrm{Ker}(\underline{\mathbf{A}}_L)$ *is invariant of* $\mathbf{M}^{-1}\mathbf{A}$ *if and only if*

$$\exists \mathbf{M}_1 \in \mathbb{R}^{n_L \times n_L}, \exists \mathbf{M}_3 \in \mathbb{R}^{n_R \times n_L}, \exists \mathbf{M}_4 \in \mathbb{R}^{n_R \times n_R} : \mathbf{M}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{0} \\ \mathbf{M}_3 & \mathbf{M}_4 \end{pmatrix}.$$

Proof As $\underline{\mathbf{A}}_L = \begin{pmatrix} \mathbf{A}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ and $\mathbf{A}_L$ is SPD, we have $\mathrm{Ker}(\underline{\mathbf{A}}_L) = \left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{y}_R \end{pmatrix} | \mathbf{y}_R \in \mathbb{R}^{n_R} \right\}$. Let $\mathbf{M}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_3 & \mathbf{M}_4 \end{pmatrix}$, then $\underline{\mathbf{A}}_L\mathbf{M}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{A}_L\mathbf{M}_1 & \mathbf{A}_L\mathbf{M}_2 \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. Let $\mathbf{y}_R \in \mathbb{R}^{n_R}$, we have

$$\underline{\mathbf{A}}_L\mathbf{M}^{-1}\mathbf{A} \cdot \begin{pmatrix} \mathbf{0} \\ \mathbf{y}_R \end{pmatrix} = \begin{pmatrix} \mathbf{A}_L\mathbf{M}_2\mathbf{y}_R \\ \mathbf{0} \end{pmatrix}.$$

Recall that $\mathrm{Ker}(\underline{\mathbf{A}}_L)$ is invariant of $\mathbf{M}^{-1}\mathbf{A}$ if and only if

$$\forall \mathbf{y} \in \mathrm{Ker}(\underline{\mathbf{A}}_L), \quad \underline{\mathbf{A}}_L\mathbf{M}^{-1}\mathbf{A} \cdot \mathbf{y} = \mathbf{0}.$$

Therefore, $\mathrm{Ker}(\underline{\mathbf{A}}_L)$ is invariant of $\mathbf{M}^{-1}\mathbf{A}$ if and only if

$$\forall \mathbf{y}_R \in \mathbb{R}^{n_R}, \quad \mathbf{A}_L\mathbf{M}_2 \cdot \mathbf{y}_R = \mathbf{0}.$$

Since $\mathbf{A}_L$ is SPD, this latter property is equivalent to

$$\forall \mathbf{y}_R \in \mathbb{R}^{n_R}, \quad \mathbf{M}_2 \cdot \mathbf{y}_R = \mathbf{0},$$

which means that $\mathbf{M}_2 = \mathbf{0}$. □

Lemma 4.2 *Let $\mathbf{M}^{-1}$ be a preconditioner of the matrix $\mathbf{A}$. If $\mathbf{M}^{-1}\mathbf{A}$ can be written as*

$$\mathbf{M}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{M}_1 & \mathbf{0} \\ \mathbf{M}_3 & \mathbf{M}_4 \end{pmatrix}, \quad (4.5)$$

where $\mathbf{M}_1 \in \mathbb{R}^{n_L \times n_L}, \mathbf{M}_3 \in \mathbb{R}^{n_R \times n_L}, \mathbf{M}_4 \in \mathbb{R}^{n_R \times n_R}$, then

$$\exists \tau > 0, \quad \forall \mathbf{u} \in \mathbb{R}^n : \quad (\mathbf{M}^{-1}\mathbf{A} \cdot \mathbf{u})^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot (\mathbf{M}^{-1}\mathbf{A} \cdot \mathbf{u}) \leq \tau \mathbf{u}^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot \mathbf{u}.$$

Proof With (4.5), we have

$$(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\underline{\mathbf{A}}_L(\mathbf{M}^{-1}\mathbf{A}) = \begin{pmatrix}\mathbf{M}_1^{\mathrm{T}} & \mathbf{M}_3^{\mathrm{T}}\\ \mathbf{0} & \mathbf{M}_4^{\mathrm{T}}\end{pmatrix}\begin{pmatrix}\mathbf{A}_L & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{M}_1 & \mathbf{0}\\ \mathbf{M}_3 & \mathbf{M}_4\end{pmatrix} = \begin{pmatrix}\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}.$$

Let $\mathbf{u} = \begin{pmatrix}\mathbf{u}_L\\ \mathbf{u}_R\end{pmatrix} \in \mathbb{R}^n$, we have:

$$\mathbf{u}^{\mathrm{T}}\cdot(\mathbf{M}^{-1}\mathbf{A})^{\mathrm{T}}\underline{\mathbf{A}}_L(\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u} = \begin{pmatrix}\mathbf{u}_L\\ \mathbf{u}_R\end{pmatrix}^{\mathrm{T}}\cdot\begin{pmatrix}\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{u}_L\\ \mathbf{u}_R\end{pmatrix} = \mathbf{u}_L^{\mathrm{T}}\cdot\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1\cdot\mathbf{u}_L$$

and

$$\mathbf{u}^{\mathrm{T}}\cdot\underline{\mathbf{A}}_L\cdot\mathbf{u} = \begin{pmatrix}\mathbf{u}_L\\ \mathbf{u}_R\end{pmatrix}^{\mathrm{T}}\cdot\begin{pmatrix}\mathbf{A}_L & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{u}_L\\ \mathbf{u}_R\end{pmatrix} = \mathbf{u}_L^{\mathrm{T}}\cdot\mathbf{A}_L\cdot\mathbf{u}_L.$$

We consider the following generalized eigenvalue problem:

$$\text{Find}\quad (\lambda_k,\mathbf{y}_k)\in\mathbb{R}\times\mathbb{R}^{n_L}\quad\text{such that}\quad \mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1\cdot\mathbf{y}_k = \lambda_k\mathbf{A}_L\cdot\mathbf{y}_k. \tag{4.6}$$

$\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1$ and $\mathbf{A}_L$ are both symmetric positive definite matrices, therefore the eigenvectors $\mathbf{y}_k$ can be chosen so that they form a basis of $\mathbb{R}^{n_L}$ that is both $\mathbf{A}_L$-orthonormal and $\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1$-orthogonal. As a consequence, we can write:

$$\begin{aligned}&\mathbf{u}_L = \sum_{k=1}^{n_L}\langle\mathbf{A}_L\cdot\mathbf{u}_L,\mathbf{y}_k\rangle\mathbf{y}_k\\ \iff\ &\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1\cdot\mathbf{u}_L = \sum_{k=1}^{n_L}\langle\mathbf{A}_L\cdot\mathbf{u}_L,\mathbf{y}_k\rangle\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1\cdot\mathbf{y}_k = \sum_{k=1}^{n_L}\langle\mathbf{A}_L\cdot\mathbf{u}_L,\mathbf{y}_k\rangle\lambda_k\mathbf{A}_L\cdot\mathbf{y}_k.\end{aligned}$$

Therefore,

$$\mathbf{u}_L^{\mathrm{T}}\cdot\mathbf{M}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{M}_1\cdot\mathbf{u}_L = \sum_{k=1}^{n_L}\lambda_k\langle\mathbf{A}_L\cdot\mathbf{u}_L,\mathbf{y}_k\rangle^2 \le \max_{1\le i\le n_L}\lambda_i\sum_{k=1}^{n_L}\langle\mathbf{A}_L\cdot\mathbf{u}_L,\mathbf{y}_k\rangle^2 = \left(\max_{1\le i\le n_L}\lambda_i\right)\mathbf{u}_L^{\mathrm{T}}\cdot\mathbf{A}_L\cdot\mathbf{u}_L. \tag{4.7}$$

Hence $$(\mathbf{M}^{-1}\mathbf{A}\cdot\mathbf{u})^{\mathrm{T}}\cdot\underline{\mathbf{A}}_L\cdot(\mathbf{M}^{-1}\mathbf{A}\cdot\mathbf{u}) \le \left(\max_{1\le i\le n_L}\lambda_i\right)\mathbf{u}^{\mathrm{T}}\cdot\underline{\mathbf{A}}_L\cdot\mathbf{u}.$$ □

Theorem 4.1 *Let* $\mathbf{A} = \begin{pmatrix}\mathbf{A}_L & \mathbf{A}_{LR}\\ \mathbf{A}_{RL} & \mathbf{A}_R\end{pmatrix}$ *be an SPD matrix,* $\mathbf{M}^{-1}$ *be a preconditioner of* $\mathbf{A}$ *and* $\underline{\mathbf{A}}_L$ *be defined as in* (4.2)*. The property,*

$$\exists\tau>0,\ \forall\mathbf{u}\in\mathbb{R}^n:\quad((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u})^{\mathrm{T}}\cdot\underline{\mathbf{A}}_L\cdot((\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\cdot\mathbf{u}) \le \tau\mathbf{u}^{\mathrm{T}}\cdot\underline{\mathbf{A}}_L\cdot\mathbf{u} \tag{4.8}$$

is satisfied if and only if

$$\exists\mathbf{W}_1\in\mathbb{R}^{n_L\times n_L},\mathbf{W}_3\in\mathbb{R}^{n_R\times n_L},\mathbf{W}_4\in\mathbb{R}^{n_R\times n_R}:\mathbf{M}^{-1} = \begin{pmatrix}\mathbf{W}_1 & -\mathbf{W}_1\mathbf{A}_{LR}\mathbf{A}_R^{-1}\\ \mathbf{W}_3 & \mathbf{W}_4\end{pmatrix}. \tag{4.9}$$

In that case, the minimum coefficient τ *that satisfies* (4.8) *is the maximum eigenvalue* λ_k *of* (4.6)*.*

Proof According to Corollary 4.1, Lemmas 4.1 and 4.2, we have:

$$\exists \tau > 0, \forall \mathbf{u} \in \mathbb{R}^n : ((\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) \cdot \mathbf{u})^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot ((\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) \cdot \mathbf{u}) \leq \tau \mathbf{u}^{\mathrm{T}} \cdot \underline{\mathbf{A}}_L \cdot \mathbf{u}$$

if and only if $\mathbf{R}_1\mathbf{M}^{-1}\mathbf{A}\mathbf{R}_3{}^{\mathrm{T}} = \mathbf{0}$, where $\mathbf{R}_1 \in \mathbb{R}^{n_L \times n}, \mathbf{R}_3 \in \mathbb{R}^{n_R \times n}$ are restriction matrices defined such that $\mathbf{R}_1 : \begin{pmatrix} \mathbf{x}_L \\ \mathbf{x}_R \end{pmatrix} \in \mathbb{R}^n \mapsto \mathbf{x}_L \in \mathbb{R}^{n_L}$, and $\mathbf{R}_3 : \begin{pmatrix} \mathbf{x}_L \\ \mathbf{x}_R \end{pmatrix} \in \mathbb{R}^n \mapsto \mathbf{x}_R \in \mathbb{R}^{n_R}$. Let $\mathbf{W}_1 \in \mathbb{R}^{n_L \times n_L}, \mathbf{W}_2 \in \mathbb{R}^{n_L \times n_R}, \mathbf{W}_3 \in \mathbb{R}^{n_R \times n_L}, \mathbf{W}_4 \in \mathbb{R}^{n_R \times n_R}$ such that $\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_2 \\ \mathbf{W}_3 & \mathbf{W}_4 \end{pmatrix}$. Then $\mathbf{R}_1\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_2 \end{pmatrix}$ and $\mathbf{A}\mathbf{R}_3{}^{\mathrm{T}} = \begin{pmatrix} \mathbf{A}_{LR} \\ \mathbf{A}_R \end{pmatrix}$. Hence:

$$\mathbf{R}_1\mathbf{M}^{-1}\mathbf{A}\mathbf{R}_3{}^{\mathrm{T}} = \mathbf{0} \iff \mathbf{W}_1\mathbf{A}_{LR} + \mathbf{W}_2\mathbf{A}_R = \mathbf{0} \iff \mathbf{W}_2 = -\mathbf{W}_1\mathbf{A}_{LR}\mathbf{A}_R^{-1}.$$

The minimum coefficient τ is deduced from the proof of Lemma 4.2 as inequality (4.7) is sharp. □

It is worth mentioning that the submatrix $\mathbf{A}_R^{-1}$ involved in the preconditioner $\mathbf{M}^{-1}$ in (4.9) cannot easily be computed in practice and requires itself an iterative scheme. Furthermore, it is noteworthy that Theorem 4.1 does not impose any special requirements on matrices $\mathbf{W}_1$, $\mathbf{W}_3$ and $\mathbf{W}_4$. For instance, the choice of a matrix $\mathbf{W}_3 := -\mathbf{A}_R^{-1}\mathbf{A}_{RL}\mathbf{W}_1^{\mathrm{T}}$ and symmetric diagonal blocks $\mathbf{W}_1$ and $\mathbf{W}_4$ may be considered for the purposes of symmetry. In this case, it is possible to use $\mathbf{M}^{-1}$ as a preconditioner outside the framework of fixed-point iteration schemes, for a PCG solver for example. As far as the block $\mathbf{W}_4$ is concerned, we can choose it to be equal to $\mathbf{A}_R^{-1}$ as this inverse is already needed for an off-diagonal block. Furthermore, if we denote $\mathbf{S}_L := \mathbf{A}_L - \mathbf{A}_{LR}\mathbf{A}_R^{-1}\mathbf{A}_{RL}$, we can deduce the expression of the upper left block $\mathbf{M}_1$ of the matrix $\mathbf{M}^{-1}\mathbf{A}$ from (4.9) in Theorem 4.1. Indeed, we have

$$\mathbf{M}_1 = \mathbf{R}_1\mathbf{M}^{-1}\mathbf{A}\mathbf{R}_1^{\mathrm{T}} = \begin{pmatrix} \mathbf{W}_1 & -\mathbf{W}_1\mathbf{A}_{LR}\mathbf{A}_R^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{A}_L \\ \mathbf{A}_{RL} \end{pmatrix} = \mathbf{W}_1\mathbf{S}_L;$$

and the generalized eigenvalue problem (4.6) becomes,

$$\text{Find} \quad (\lambda_k, \mathbf{y}_k) \in \mathbb{R} \times \mathbb{R}^{n_L} \quad \text{such that} \quad \mathbf{S}_L\mathbf{W}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L \cdot \mathbf{y}_k = \lambda_k \mathbf{A}_L \cdot \mathbf{y}_k. \quad (4.10)$$

We are interested in the maximum λ_k as it gives the minimum value of τ in (4.8), i.e. the minimum upperbound of the error's growth rate. It is straightforward that the eigenvalues λ_k form the spectrum of the matrix $\mathbf{S}_L\mathbf{W}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L\mathbf{A}_L^{-1}$. To render this spectrum bounded from above, three options can be considered:

- $\mathbf{W}_1 = \mathbf{S}_L^{-1}$: This choice reduces the spectrum of $\mathbf{S}_L\mathbf{W}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L\mathbf{A}_L^{-1}$ to 1 since

$$\mathbf{S}_L\mathbf{W}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L = \mathbf{A}_L.$$

- $\mathbf{W}_1 = \mathbf{A}_L^{-1}$: This choice makes the spectrum of $\mathbf{S}_L\mathbf{W}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L\mathbf{A}_L^{-1}$ bounded from above by 1. In fact,

$$\mathbf{S}_L\mathbf{W}_1^{\mathrm{T}}\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L\mathbf{A}_L^{-1} = (\mathbf{S}_L\mathbf{A}_L^{-1})^2,$$

and $\mathrm{Sp}(\mathbf{S}_L\mathbf{A}_L^{-1}) \subset]0, 1]$ (see [11, Theorem 3.1]).

– Taking $\mathbf{W}_1$ as a SPD preconditioner such that the eigenvalues of $\mathbf{W}_1\mathbf{A}_L$ are below a fixed scalar $\nu > 0$ implies that the maximum λ_k (and hence the minimal τ) is less than ν^2 (see Lemma 4.3 below).

It should be outlined that the first two choices are more theoretical than practical as it is often costly to compute the exact inverse of a large block $\mathbf{A}_L$ or a dense Schur complement matrix $\mathbf{S}_L$. The third choice is more affordable in practice, as there are many preconditioning strategies that allow to bound the maximum eigenvalue of the preconditioned operator, e.g. domain decomposition-based preconditioners (like one-level Additive Schwarz, BDD or two-level Schwarz) [9] or LORASC [11].

Lemma 4.3 *Let $\mathbf{W}_1$ be an SPD preconditioner of $\mathbf{A}_L$ such that the eigenvalues of $\mathbf{W}_1\mathbf{A}_L$ are below a fixed scalar $\nu > 0$,*

$$\lambda_{\max}(\mathbf{W}_1\mathbf{A}_L) \leq \nu. \tag{4.11}$$

Then it holds that:

$$\lambda_{\max}(\mathbf{S}_L\mathbf{W}_1\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L\mathbf{A}_L^{-1}) \leq \nu^2. \tag{4.12}$$

Proof We know that

$$\mathbf{S}_L\mathbf{W}_1 = \mathbf{A}_L\mathbf{W}_1 - \mathbf{A}_{LR}\mathbf{A}_R^{-1}\mathbf{A}_{RL}\mathbf{W}_1.$$

Therefore since $\mathbf{W}_1$ and $\mathbf{A}_{LR}\mathbf{A}_R^{-1}\mathbf{A}_{RL}$ are SPD matrices, the eigenvalues of $\mathbf{A}_{LR}\mathbf{A}_R^{-1}\mathbf{A}_{RL}\mathbf{W}_1$ are nonnegative therefore

$$\lambda_{\max}(\mathbf{S}_L\mathbf{W}_1) \leq \lambda_{\max}(\mathbf{A}_L\mathbf{W}_1).$$

And also from [11, Theorem 3.1],

$$\lambda_{\max}(\mathbf{S}_L\mathbf{A}_L^{-1}) \leq 1.$$

Thus,

$$\begin{aligned}\lambda_{\max}(\mathbf{S}_L\mathbf{W}_1\mathbf{A}_L\mathbf{W}_1\mathbf{S}_L\mathbf{A}_L^{-1}) &\leq \lambda_{\max}(\mathbf{S}_L\mathbf{W}_1)\lambda_{\max}(\mathbf{A}_L\mathbf{W}_1)\lambda_{\max}(\mathbf{S}_L\mathbf{A}_L^{-1}) \\ &\leq \lambda^2_{\max}(\mathbf{A}_L\mathbf{W}_1) \\ &\leq \nu^2.\end{aligned}$$

□

5 Link with the adaptive preconditioner for PCG based on local error indicators

The main goal of Section 4 was to identify a relevant seminorm (the $\underline{\mathbf{A}}_L$-seminorm), which is superior to the dominant $A_p^{(1)}$-seminorm, and to derive a preconditioner that enables to control that seminorm in a fixed-point iteration scheme. In this section, we connect the results obtained in the previous sections with the preconditioner introduced in [1, Section 4] by:

❶ verifying if its main feature, which is cancelling the residual on the L-part, is still true in a fixed-point iteration scheme (Lemmas 5.1 and 5.2),
❷ identifying the seminorm of the error whose growth rate is controlled by this preconditioner within a fixed-point iteration scheme (Corollary 5.1).

For that, we exploit the lemmas proven in Section 4 to derive the properties satisfied by the adaptive preconditioner introduced in [1].

First, we start by the preconditioner $\mathbf{M} = \begin{pmatrix} \mathbf{A}_L & \mathbf{A}_{LR} \\ \mathbf{A}_{RL} & \mathbf{M}_S + \mathbf{A}_{RL}\mathbf{A}_L^{-1}\mathbf{A}_{LR} \end{pmatrix}$ suggested in [1, Theorem 1], where $\mathbf{M}_S$ is a preconditioner for $\mathbf{S}_R := \mathbf{A}_R - \mathbf{A}_{RL}\mathbf{A}_L^{-1}\mathbf{A}_{LR}$. That theorem states that when preconditioned by this preconditioner, the PCG solver on the whole system becomes equivalent to a PCG solver on a reduced Schur complement system resulting in a nil residual on the unknowns of Ω_1 (i.e. the L-part) at each iteration:

$$\forall i > 0, \quad \mathbf{R}_1 \cdot \mathbf{r}^{(i)} = \mathbf{R}_1\mathbf{A} \cdot (\mathbf{x} - \mathbf{x}^{(i)}) = \mathbf{0}, \tag{5.1}$$

where $\mathbf{R}_1$ is the restriction matrix of (2.5). In the following lemma, we prove that this property is still satisfied with a fixed-point iteration scheme (3.1).

Lemma 5.1 *Let* $\mathbf{M} = \begin{pmatrix} \mathbf{A}_L & \mathbf{A}_{LR} \\ \mathbf{A}_{RL} & \mathbf{M}_S + \mathbf{A}_{RL}\mathbf{A}_L^{-1}\mathbf{A}_{LR} \end{pmatrix}$ *where* $\mathbf{M}_S$ *is invertible. When* $\mathbf{M}^{-1}$ *is used as the preconditioner of the fixed-point iteration scheme defined in* (3.1)*, then property* (5.1) *holds regardless of the choice of the initial guess* $\mathbf{x}^{(0)}$.

Proof The inverse of $\mathbf{M}$ is expressed as

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{A}_L^{-1} + \mathbf{A}_L^{-1}\mathbf{A}_{LR}\mathbf{M}_S^{-1}\mathbf{A}_{RL}\mathbf{A}_L^{-1} & -\mathbf{A}_L^{-1}\mathbf{A}_{LR}\mathbf{M}_S^{-1} \\ -\mathbf{M}_S^{-1}\mathbf{A}_{RL}\mathbf{A}_L^{-1} & \mathbf{M}_S^{-1} \end{pmatrix}.$$

Therefore,

$$\mathbf{M}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{A}_L^{-1}\mathbf{A}_{LR}(\mathbf{I} - \mathbf{M}_S^{-1}\mathbf{S}_R) \\ \mathbf{0} & \mathbf{M}_S^{-1}\mathbf{S}_R \end{pmatrix},$$

and

$$\mathbf{A}(\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_R - \mathbf{S}_R\mathbf{M}_S^{-1}\mathbf{S}_R \end{pmatrix}.$$

Thus, $\mathbf{R}_1\mathbf{A}(\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) = \mathbf{0}$.
Consequently, for any iteration $i > 0$ we have

$$\mathbf{R}_1\mathbf{r}^{(i)} = \mathbf{R}_1\mathbf{A} \cdot (\mathbf{x} - \mathbf{x}^{(i)}) = \mathbf{R}_1\mathbf{A}(\mathbf{I} - \mathbf{M}^{-1}\mathbf{A}) \cdot (\mathbf{x} - \mathbf{x}^{(i-1)}) = \mathbf{0}.$$

□

For the more general shape of the adaptive preconditioner proposed in [1, Theorem 2], we prove in what follows that Property (5.1) still holds, but this time for a specific initial guess $\mathbf{x}^{(0)}$.

Lemma 5.2 *Let $\mathbf{x}_R^{(0)}$ be an arbitrary vector of $\mathbb{R}^{n_R}$ and $\mathbf{W}_1 \in \mathbb{R}^{n_L \times n_L}$, $\mathbf{W}_2 \in \mathbb{R}^{n_R \times n_R}$ two invertible matrices. Let the linear system $\mathbf{A}\cdot\mathbf{x} = \mathbf{b}$ be solved by a fixed-point iteration scheme* (3.1) *with a preconditioner $\mathbf{M} = \mathbf{W}\mathbf{W}^{\mathrm{T}}$ and an initial guess $\mathbf{x}^{(0)}$ such that*

$$\mathbf{x}^{(0)} = \begin{bmatrix} \mathbf{A}_L^{-1}\cdot(\mathbf{b}_L - \mathbf{A}_{LR}\cdot\mathbf{x}_R^{(0)}) \\ \mathbf{x}_R^{(0)} \end{bmatrix}, \quad \mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} \\ \mathbf{A}_{RL}\mathbf{A}_L^{-1}\mathbf{W}_1 & \mathbf{W}_2 \end{pmatrix}; \tag{5.2}$$

then $\mathbf{R}_1\cdot(\mathbf{b}-\mathbf{A}\cdot\mathbf{x}^{(i)}) = \mathbf{0}$ at each iteration $i > 0$.

Proof The definition of $\mathbf{x}^{(0)}$ yields

$$\mathbf{x}-\mathbf{x}^{(0)} = \begin{bmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR}\cdot(\mathbf{x}_R-\mathbf{x}_R^{(0)}) \\ \mathbf{x}_R-\mathbf{x}_R^{(0)} \end{bmatrix} = \begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR} \\ \mathbf{I} \end{pmatrix}\cdot(\mathbf{x}_R-\mathbf{x}_R^{(0)}).$$

We know that due to (3.1), we have for any i,

$$\mathbf{R}_1\cdot(\mathbf{b}-\mathbf{A}\cdot\mathbf{x}^{(i)}) = \mathbf{R}_1\mathbf{A}\cdot(\mathbf{x}-\mathbf{x}^{(i)}) = \mathbf{R}_1\mathbf{A}(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})^i\cdot(\mathbf{x}-\mathbf{x}^{(0)}).$$

Therefore, proving the lemma amounts to prove that

$$\forall i\in\mathbb{N}, \quad \mathbf{R}_1\mathbf{A}(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})^i\begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR} \\ \mathbf{I} \end{pmatrix} = \mathbf{0}. \tag{5.3}$$

Let us demonstrate that by induction.
For $i = 0$,

$$\mathbf{R}_1\mathbf{A}\begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR} \\ \mathbf{I} \end{pmatrix} = \begin{pmatrix}\mathbf{A}_L & \mathbf{A}_{LR}\end{pmatrix}\begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR} \\ \mathbf{I} \end{pmatrix} = \mathbf{0}.$$

We denote $\mathbf{M}_1 := \mathbf{W}_1\mathbf{W}_1^{\mathrm{T}}$, and $\mathbf{M}_S := \mathbf{W}_2\mathbf{W}_2^{\mathrm{T}}$. A quick computation of the inverse of $\mathbf{W}$ gives

$$\mathbf{M}^{-1} = \mathbf{W}^{-\mathrm{T}}\mathbf{W}^{-1} = \begin{pmatrix} \mathbf{M}_1^{-1}+\mathbf{A}_L^{-1}\mathbf{A}_{LR}\mathbf{M}_S^{-1}\mathbf{A}_{RL}\mathbf{A}_L^{-1} & -\mathbf{A}_L^{-1}\mathbf{A}_{LR}\mathbf{M}_S^{-1} \\ -\mathbf{M}_S^{-1}\mathbf{A}_{RL}\mathbf{A}_L^{-1} & \mathbf{M}_S^{-1} \end{pmatrix}.$$

Thus we obtain,

$$\mathbf{I}-\mathbf{M}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{I}-\mathbf{M}_1^{-1}\mathbf{A}_L & -\mathbf{M}_1^{-1}\mathbf{A}_{LR}+\mathbf{A}_L^{-1}\mathbf{A}_{LR}\mathbf{M}_S^{-1}\mathbf{S}_R \\ \mathbf{0} & \mathbf{I}-\mathbf{M}_S^{-1}\mathbf{S}_R \end{pmatrix} \tag{5.4}$$

Let $i\in\mathbb{N}$, we assume the induction hypothesis (5.3) for i, then due to (5.4) we have

$$\begin{aligned}
\mathbf{R}_1\mathbf{A}(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})^{i+1}\begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR} \\ \mathbf{I} \end{pmatrix} &= \begin{pmatrix}\mathbf{A}_L & \mathbf{A}_{LR}\end{pmatrix}(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})^i(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})\begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR} \\ \mathbf{I} \end{pmatrix} \\
&= \begin{pmatrix}\mathbf{A}_L & \mathbf{A}_{LR}\end{pmatrix}(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})^i\begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR}(\mathbf{I}-\mathbf{M}_S^{-1}\mathbf{S}_R) \\ (\mathbf{I}-\mathbf{M}_S^{-1}\mathbf{S}_R) \end{pmatrix} \\
&= \begin{pmatrix}\mathbf{A}_L & \mathbf{A}_{LR}\end{pmatrix}(\mathbf{I}-\mathbf{M}^{-1}\mathbf{A})^i\begin{pmatrix} -\mathbf{A}_L^{-1}\mathbf{A}_{LR} \\ \mathbf{I} \end{pmatrix}(\mathbf{I}-\mathbf{M}_S^{-1}\mathbf{S}_R) \\
&= 0.
\end{aligned}$$

Thus, we can apply the induction hypothesis for i to show that (5.3) is true for $i+1$.
□

Moreover, as far as the algebraic error norm is concerned, if we denote by

$$\underline{\mathbf{A}}_R := \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_R \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{5.5}$$

we can derive the subsequent corollary for these adaptive preconditioners.

Corollary 5.1 *Let $\mathbf{M}^{-1}$ be an adaptive preconditioner of the matrix $\mathbf{A}$ as defined in Lemma 5.2. When $\mathbf{M}^{-1}$ is used as the preconditioner of the fixed-point iteration scheme defined in* (3.1)*, then it holds that:*

$$\exists \tau > 0, \quad \forall \mathbf{u} \in \mathbb{R}^n : \quad (\mathbf{M}^{-1}\mathbf{A}\cdot\mathbf{u})^{\mathrm{T}} \cdot \underline{\mathbf{A}}_R \cdot (\mathbf{M}^{-1}\mathbf{A}\cdot\mathbf{u}) \leq \tau \mathbf{u}^{\mathrm{T}} \cdot \underline{\mathbf{A}}_R \cdot \mathbf{u}.$$

The minimum value of τ is the maximum eigenvalue of the generalized eigenvalue problem:

$$\textit{Find} \quad (\lambda_k, \mathbf{y}_k) \in \mathbb{R}\times\mathbb{R}^{n_R} \quad \textit{such that} \quad \mathbf{M}_4^{\mathrm{T}}\mathbf{A}_R\mathbf{M}_4 \cdot \mathbf{y}_k = \lambda_k \mathbf{A}_R \cdot \mathbf{y}_k, \tag{5.6}$$

where $\mathbf{M}_4$ is the bottom right block of size $n_R \times n_R$ of the matrix $\mathbf{M}^{-1}\mathbf{A}$.

Proof First, if we look at the shape of the adaptive preconditioners either in its general shape (as in Lemma 5.2) or in its particular shape (as in Lemma 5.1) we notice that the preconditioned operator takes the following block form:

$$\exists (\mathbf{M}_1', \mathbf{M}_3, \mathbf{M}_4) \in \mathbb{R}^{n_L\times n_L} \times \mathbb{R}^{n_L\times n_R} \times \mathbb{R}^{n_R\times n_R} : \mathbf{M}^{-1}\mathbf{A} = \begin{pmatrix} \mathbf{M}_1' & \mathbf{M}_3 \\ \mathbf{0} & \mathbf{M}_4 \end{pmatrix}.$$

Therefore, we can apply Lemma 4.2 by switching or commuting the roles of the subsets L and R. In this case, the result follows immediately. □

Summary and comments

We provide here a summary of the results of the lemmas and corollary of this section. We prove in Lemmas 5.1 and 5.2 that the main feature of the adaptive preconditioner introduced in [1] for PCG solver, that is the projection of the residual on the unknowns of subdomain Ω_1 (L-part) is nil at each iteration, is still valid with a fixed-point iteration scheme. We also prove in Corollary 5.1 that the adaptive preconditioner satisfies with the same scheme a criterion that expresses that the $\underline{\mathbf{A}}_R$-seminorm of the error does not increase more than τ-times from iteration i to iteration $i+1$, where τ is the maximum eigenvalue of the generalized eigenvalue problem (5.6). The discussion of the choices of the diagonal blocks of the preconditioner in the end of Section 4 holds for this part as well, i.e., the quality of the preconditioner $\mathbf{M}_S$ affects the value taken by τ. The smaller the eigenvalues of $\mathbf{M}_S^{-1}\mathbf{S}_R$ are, the smaller τ is. Besides, the $\underline{\mathbf{A}}_R$-seminorm of the error represents a share of the global energy norm of the error. The evolution of this share from an iteration to another is controlled by the adaptive preconditioner and the growth rate is kept under a fixed threshold. That being said, the starting hypothesis (2.7) does not enable us to determine whether this share is dominant. Indeed, (2.7) expresses that the local algebraic error on subdomain Ω_1,

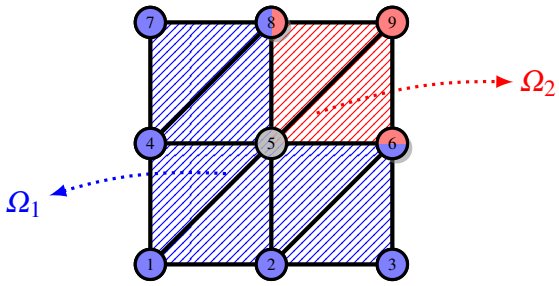


Fig. 5.1 Simple example of the decomposition with a 2×2 mesh grid.

which is the $\mathbf{A}_p^{(1)}$-seminorm of the error, is dominant over the local algebraic error on subdomain Ω_2, which is the $\mathbf{A}_p^{(2)}$-seminorm of the error. Yet, in general we cannot prove any partial order or superiority relationship, in the sense of matrix positiveness, between the matrices $\underline{\mathbf{A}}_R$ and $\mathbf{A}_p^{(2)}$ as we did between the matrices $\underline{\mathbf{A}}_L$ and $\mathbf{A}_p^{(1)}$. Indeed, we demonstrate that with the following counterexample. We consider Poisson's equation with Dirichlet boundary conditions on the square $[0,1] \times [0,1]$. We discretize this PDE by FEM on the uniform grid shown in Figure 5.1. Note that with FEM, the boundary conditions are taken into account in the evaluation of the right hand side vector of the linear system. The global matrix obtained for the natural ordering of the degrees of freedom (from 1 to 9) is:

$$\mathbf{A} = \begin{pmatrix} 1 & -1/2 & 0 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ -1/2 & 2 & -1/2 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & 1 & 0 & 0 & -1/2 & 0 & 0 & 0 \\ -1/2 & 0 & 0 & 2 & -1 & 0 & -1/2 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1/2 & 0 & -1 & 2 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & -1/2 & 0 & 0 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1/2 & 2 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & -1/2 & 0 & -1/2 & 1 \end{pmatrix}.$$

If we consider the domain decomposition illustrated in Figure 5.1, then the set of interior vertices of subdomain Ω_2 comprises only the vertex 9 whereas the vertices 5, 6 and 8 are in the interface. Therefore, if we keep the natural ordering then the local

stiffness matrix for Ω_1 and Ω_2, respectively are:

$$\mathbf{A}^{(1)} = \begin{pmatrix} 1 & -1/2 & 0 & -1/2 & 0 & 0 & 0 & 0 \\ -1/2 & 2 & -1/2 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1/2 & 1 & 0 & 0 & -1/2 & 0 & 0 \\ -1/2 & 0 & 0 & 2 & -1 & 0 & -1/2 & 0 \\ 0 & -1 & 0 & -1 & 3 & -1/2 & 0 & -1/2 \\ 0 & 0 & -1/2 & 0 & -1/2 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/2 & 0 & 0 & 1 & -1/2 \\ 0 & 0 & 0 & 0 & -1/2 & 0 & -1/2 & 1 \end{pmatrix},$$

$$\mathbf{A}^{(2)} = \begin{pmatrix} 1 & -1/2 & -1/2 & 0 \\ -1/2 & 1 & 0 & -1/2 \\ -1/2 & 0 & 1 & -1/2 \\ 0 & -1/2 & -1/2 & 1 \end{pmatrix}.$$

and the extended matrices $\mathbf{A}_p^{(1)}$, $\mathbf{A}_p^{(2)}$ and $\underline{\mathbf{A}}_R$ are equal to:

$$\mathbf{A}_p^{(1)} = \begin{pmatrix} 1 & -1/2 & 0 & -1/2 & 0 & 0 & 0 & 0 & 0 \\ -1/2 & 2 & -1/2 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1/2 & 1 & 0 & 0 & -1/2 & 0 & 0 & 0 \\ -1/2 & 0 & 0 & 2 & -1 & 0 & -1/2 & 0 & 0 \\ 0 & -1 & 0 & -1 & 3 & -1/2 & 0 & -1/2 & 0 \\ 0 & 0 & -1/2 & 0 & -1/2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/2 & 0 & 0 & 1 & -1/2 & 0 \\ 0 & 0 & 0 & 0 & -1/2 & 0 & -1/2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{A}_p^{(2)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1/2 & 0 & -1/2 & 0 \\ 0 & 0 & 0 & 0 & -1/2 & 1 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1/2 & 0 & 0 & 1 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & -1/2 & 0 & -1/2 & 1 \end{pmatrix}, \quad \underline{\mathbf{A}}_R = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore, the difference is:

$$\mathbf{A}_p^{(2)} - \underline{\mathbf{A}}_R = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1/2 & 0 & -1/2 & 0 \\ 0 & 0 & 0 & 0 & -1/2 & 1 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1/2 & 0 & 0 & 1 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & -1/2 & 0 & -1/2 & 0 \end{pmatrix},$$

and its eigenvalues can be computed: $\lambda_1 \approx -0.45161$; $\lambda_2 = 0$; $\lambda_3 \approx 0.59697$; $\lambda_4 = 1$; $\lambda_5 \approx 1.8546$. We can notice that they are not all nonnegative. Therefore, neither of the assertions ($\underline{\mathbf{A}}_R \leq \mathbf{A}_p^{(2)}$ or $\mathbf{A}_p^{(2)} \leq \underline{\mathbf{A}}_R$) is true in the sense of matrix positiveness.

Furthermore, it should be stressed that the properties of the adaptive preconditioner introduced in [1, Theorem 1], and the preconditioner introduced in Lemma 4.1 are complementary, in the sense that each allows to control a part of the global energy norm of the error. The first one requires inverting the L-block corresponding to a subdomain Ω_1 with a high algebraic error, reduces the $\underline{\mathbf{A}}_R$-seminorm of the error and cancels the residual on the unknowns of L, whereas the second one requires inverting the R-block corresponding to a subdomain with low algebraic error ($\mathring{\Omega}_2$) and ensures the growth rate of the $\underline{\mathbf{A}}_L$-seminorm of the error from iteration i to iteration $i+1$ stays below a given threshold. Therefore, the choice between those two depends on the size of submatrix $\mathbf{A}_R$ with respect to $\mathbf{A}_L$. For the cases where the subdomain Ω_1 with high errors is confined to a small area, the first adaptive preconditioner could be an acceptable choice as the submatrix to invert $\mathbf{A}_L$ is relatively small. In the cases where the high errors are scattered in wide regions of the domain, the second adaptive preconditioner seems to be a good alternative as the size of the submatrix to invert $\mathbf{A}_R$ is increasingly reduced as the submatrix $\mathbf{A}_L$ gets larger and larger. Note also that the exact inverse of the matrix $\mathbf{A}$ satisfies the shapes of both preconditioners.

6 Numerical results

In this section, we consider the following experimental framework. The tests are carried in Matlab. We generate an uniform mesh and use PDE toolbox to solve the considered PDE on the domain Ω. Once the linear system is built, we run a few iterations of the linear solver (20 iterations of PCG preconditioned by a Block-Jacobi preconditioner) to get an initial estimation of the algebraic error on all the elements of the mesh. From the values of these a posteriori error estimates, we determine the subsets of indices L and R. In the sequel, on the plots of the initial distribution of a posteriori algebraic error estimates, a thick horizontal straight labeled with θ on the color bar indicates the extent of the errors' range covered with the Dörfler rate considered (see [1, Section 5.1]), i.e. all elements represented in color shades above the corresponding threshold θ form Ω_1.

We check first with a fixed-point iteration scheme that the result of Theorem 4.1 holds in practice and that the growth rate of the $\underline{\mathbf{A}}_L$-seminorm of the error is controlled by the choice of the block $\mathbf{W}_1$ in the preconditioner defined in (4.9). We denote this preconditioner as the L-adaptive preconditioner. The L-adaptive preconditioner implemented in our Matlab prototype is nested (see [14]) in the sense that it does not build the inverse of the block $\mathbf{A}_R$, but the application of the preconditioner is carried out by a call to an inner solver (PCG in our configuration) with a reduced relative tolerance of 10^{-2} and a reduced maximum number of iterations (set to 100). As a preconditioner for the inner solve, we reuse the initial Block-Jacobi preconditioner restricted to the R-indices. For checking purposes, we consider two different preconditioners for the block $\mathbf{W}_1$ for which an upper bound for the eigenvalues of the preconditioned

operator is known: a LORASC preconditioner and a Block-Jacobi preconditioner for $\mathbf{A}_L$. For the latter one, we vary the number of diagonal blocks (2, 4, or 8). For the blocks $\mathbf{W}_3$ and $\mathbf{W}_4$, the choices proposed in Section 4, i.e. $\mathbf{W}_3 := -\mathbf{A}_R^{-1}\mathbf{A}_{RL}\mathbf{W}_1^{\mathrm{T}}$ and $\mathbf{W}_4 := \mathbf{A}_R^{-1}$, are the ones considered for the numerical experiments.
Second, we evaluate the L-adaptive preconditioner with PCG solver and compare its results to the preconditioner introduced in [1], that we denote as the R-adaptive preconditioner. For the numerical experiments with PCG, we consider a fixed block-size (≈ 5000) for the diagonal blocks used to build Block-Jacobi preconditioners, and a stopping threshold value of 10^{-6} for the euclidean norm of the residual.
The three test cases considered here are the following. First, we deal with Poisson equations of the form

$$-\Delta \underline{u} = -\frac{\partial^2 \underline{u}}{\partial x^2} - \frac{\partial^2 \underline{u}}{\partial y^2} = \underline{f}(x,y) \quad \text{in} \quad \Omega =]-1,1[\times]-1,1[\tag{6.1}$$

with homogeneous Dirichlet boundary condition

$$\underline{u} = 0 \quad \text{on} \quad \partial\Omega. \tag{6.2}$$

In the first two cases, we consider two examples with given smooth solutions $\underline{u}$:

$$\underline{u}^{(1)} = (x+1)\times(x-1)\times(y+1)\times(y-1)\times\exp(-\alpha\times(x^2+y^2)), \tag{6.3}$$

$$\underline{u}^{(2)} = (x+1)\times(x-1)\times(y+1)\times(y-1)\times(\exp(-\alpha\times((x+0.5)^2 + (y+0.5)^2)) - \exp(-\beta\times((x-0.5)^2+(y-0.5)^2))), \tag{6.4}$$

with $\alpha = 4000$ and $\beta = 3000$. Then, we tackle a diffusion problem with inhomogeneous diffusion tensor

$$-\nabla\cdot(\underline{\mathbf{K}}\nabla\underline{u}) = 1 \quad \text{in} \quad \Omega =]0,1[\times]0,1[\tag{6.5}$$

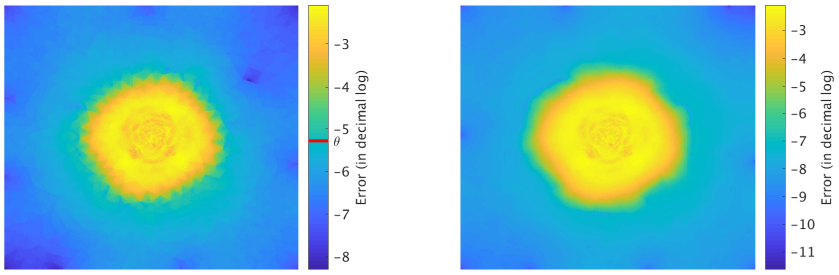
with Dirichlet boundary condition

$$\underline{u}(x,y) = \sqrt{x} \quad \text{on} \quad \partial\Omega. \tag{6.6}$$

The diffusion tensor is defined as, $\underline{\mathbf{K}} = c\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and c is the diffusivity that varies through the domain Ω. In the third test case, the diffusivity is defined as:

$$c^{(3)}(x,y) = \begin{cases} 10^5(\lfloor 9x\rfloor + 1) & \text{if} \quad \lfloor(9x)\rfloor \equiv 0 \pmod 2 \text{ and } \lfloor(9y)\rfloor \equiv 0 \pmod 2, \\ 1 & \text{otherwise.} \end{cases}$$

As far as the mesh configuration is concerned, we take an uniform mesh with maximum edge size $H_{\max} = 0.1$ (for the first two test cases) and $H_{\max} = 0.01$ (for the third test case), and the total number of elements is equal to 87552 (for the first two test cases) and 32544 (for the third test case). After discretization, the size of the matrix $\mathbf{A}$ is 43457×43457 (for the first two test cases) and 16057×16057 (for the third test case). For the first test case, the initial distribution and a posteriori estimation of the algebraic error (after $j_0 = 20$ iterations) over the domain Ω are shown in Figure 6.1



(a) Algebraic a posteriori error estimates after 20 iterations

(b) Algebraic errors after 20 iterations

Fig. 6.1 Initial distribution and a posteriori estimation of the algebraic error for test case n°1.

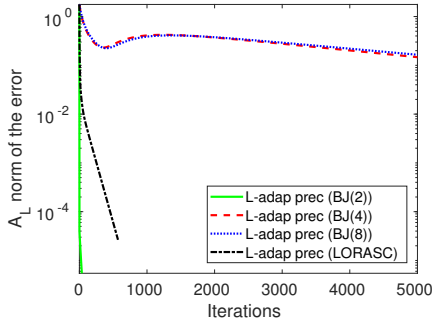
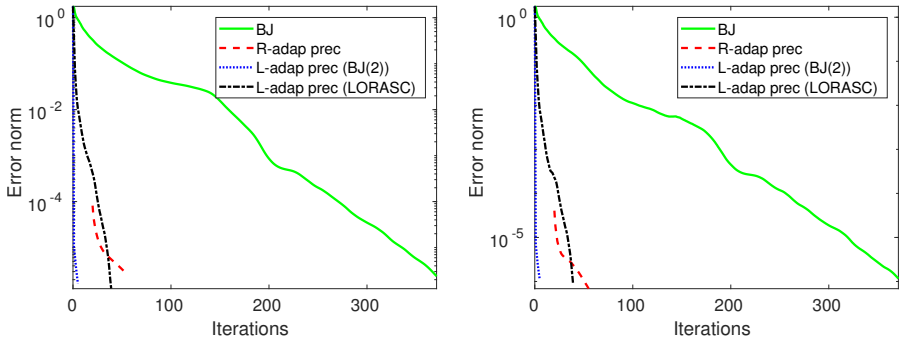


Fig. 6.2 Evolution of the $\underline{\mathbf{A}}_L$-seminorm of the error with a fixed-point iteration scheme for test case n°1.

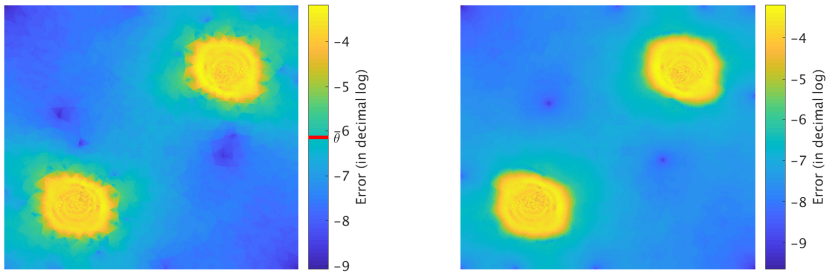
where the color shades selected in the subdomain Ω_1 are indicated by the red marker on the left subfigure.
Figure 6.2 displays the $\underline{\mathbf{A}}_L$-seminorm of the error when using a fixed-point iteration scheme preconditioned with four different configurations of the L-adaptive preconditioner, whereas the evolution of the global energy norm and the L-norm of the error when using a PCG solver preconditioned by a Block-Jacobi preconditioner, the R-adaptive preconditioner and the L-adaptive preconditioner is displayed in Figure 6.3.

Likewise, we present the results obtained for the second test case. Figure 6.4 displays the initial error distribution over the mesh. Figure 6.5 shows the $\underline{\mathbf{A}}_L$-seminorm of the error when using a fixed-point iteration scheme with the L-adaptive preconditioner. The global energy norm and the L-norm of the error are plotted in Figure 6.6 for the PCG solves with the L-adaptive and the R-adaptive preconditioners.
The results of the third test case are as follows. Figure 6.7 displays the initial algebraic error distribution over the mesh (the subdomain Ω_1 is formed by mesh elements whose color shades are above the purple marker θ on the color bar), and the evolution of the $\underline{\mathbf{A}}_L$-seminorm of the error with a fixed-point iteration scheme preconditioned by L-adaptive preconditioners. Then the evolution of the global energy norm and



(a) Energy norm of the algebraic error (b) L-norm of the algebraic error

Fig. 6.3 Error evolution with PCG for test case n°1.



(a) Algebraic a posteriori error estimates after 20 iterations (b) Algebraic errors after 20 iterations

Fig. 6.4 Initial distribution and a posteriori estimation of the algebraic error for test case n°2.

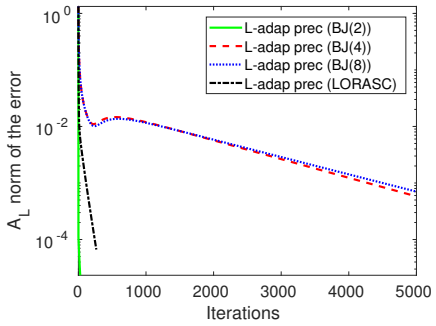
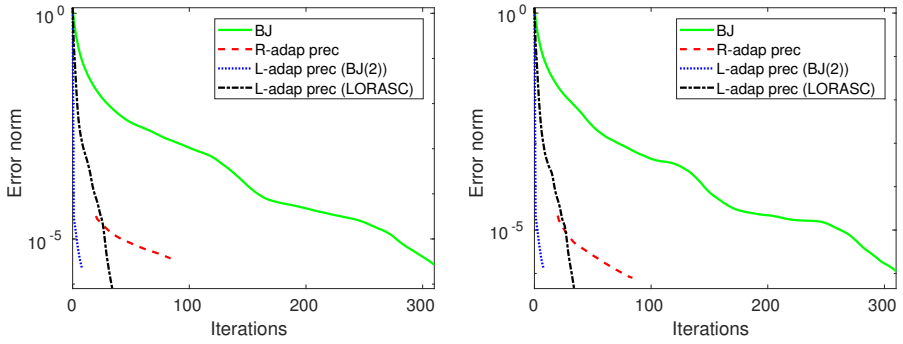
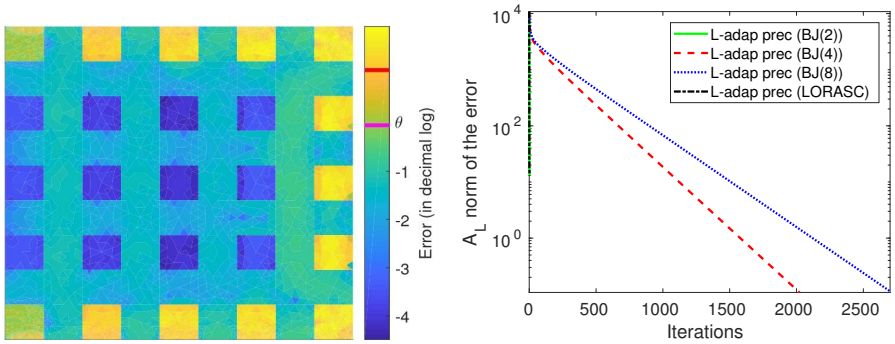


Fig. 6.5 Evolution of the $\underline{\mathbf{A}}_L$-seminorm of the error with a fixed-point iteration scheme for test case n°2.



(a) Energy norm of the algebraic error (b) L-norm of the algebraic error

Fig. 6.6 Error evolution with PCG for test case n°2.

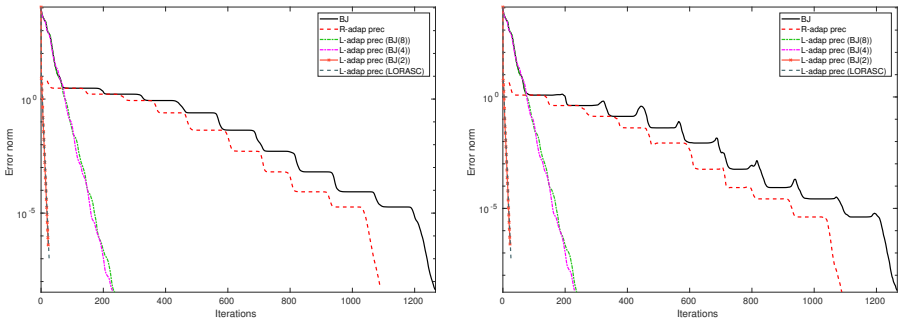


(a) Algebraic a posteriori error estimates after 20 iterations (b) Evolution of the $\underline{\mathbf{A}}_L$-seminorm of the error

Fig. 6.7 Initial distribution of algebraic error and evolution of the $\underline{\mathbf{A}}_L$-seminorm of the error with a fixed-point iteration scheme for test case n°3.

the L-norm of the error when using a PCG solver preconditioned by a Block-Jacobi preconditioner, the R-adaptive preconditioner and the L-adaptive preconditioner is displayed in Figure 6.8.

Figures 6.2 and 6.5 show that the $\underline{\mathbf{A}}_L$-seminorm of the error is monotonically decreasing when a LORASC preconditioner or a two-block diagonal preconditioner is used to approximate $\mathbf{A}_L$. But this is not the case when we use a block diagonal preconditioner with more than two blocks. This fact relates to ν, the upper bound of the largest eigenvalue of $\mathbf{W}_1\mathbf{A}_L$. Indeed, it follows from Lemma 4.3 that the growth rate of the $\underline{\mathbf{A}}_L$-seminorm of the error (i.e. the minimal τ in (4.4)) is less than or equal to ν^2. Lemma 7.1 (see Appendix) and [11, Theorem 3.1] yield the value of the bound ν obtained when $\mathbf{W}_1$ is taken as an m-block diagonal preconditioner and a LORASC preconditioner of $\mathbf{A}_L$ respectively. We note that ν is equal to $1 + \max\limits_{1\leq i\leq m} \sum\limits_{\substack{j=1 \\ j\neq i}}^{m} \gamma_{ij}$ for the former and to 1 for the latter, where the Cauchy-Bunyakovsky-Schwarz (C.B.S.) con-



(a) Energy norm of the algebraic error (b) L-norm of the algebraic error

Fig. 6.8 Error evolution with PCG for test case n°3.

Table 6.1 The total number of iterations (IT) needed for the convergence of the preconditioned solve of the linear systems stemming from test cases 1, 2 and 3.

	Test case n°1	**Test case n°2**	**Test case n°3**
initial BJ prec.	371	310	1 267
R-adap prec.[1]	36 (+20)	65 (+20)	1 069 (+20)
L-adap prec. (LORASC)	40	35	26
L-adap prec (BJ(2))	6	9	23

stants γ_{ij} are defined by (7.2) adapted to $\mathbf{A}_L$. Since $\mathbf{A}_L$ is positive definite, then the $(\gamma_{ij})_{ij}$ are less than 1. In practice, and for the matrices tested in this section, a single C.B.S. constant is very small, thus for a two-block diagonal preconditioner, the expression of ν contains only one C.B.S. constant. As a consequence, $\nu^2 = (1+\gamma_{12})^2$ is, to a certain extent, close to 1, resulting in monotonic $\underline{\mathbf{A}}_L$-seminorm of the error. The same finding applies when a LORASC preconditioner is selected for $\mathbf{W}_1$ as ν equals 1. However, for larger number of blocks (here for example $m \in \{4,8\}$), ν reflects the sum of $m-1$ C.B.S. constants. Therefore, ν^2 greatly exceeds 1 and the $\underline{\mathbf{A}}_L$-seminorm of the error is not decreasing monotonically this time. In summary, the magnitude of ν^2 determines the rate of decrease of the $\underline{\mathbf{A}}_L$-seminorm of the error. In other words, the growth rate τ is controlled by the quality of the preconditioner for the L-block. All this is reflected in Figures 6.2, 6.5 and 6.7.

Furthermore, one observes that those preconditioners, which we refer to as L-adap prec (BJ(2)) and L-adap prec (LORASC) in the legends of Figures 6.2, 6.5 and 6.7 and for which the $\underline{\mathbf{A}}_L$-seminorm of the error is strictly decreasing in a fixed-point iteration scheme, perform well when a PCG solver is used as well (Figures 6.3, 6.6 and 6.8). Table 6.1 gives the number of iterations for the PCG solve with the L-adaptive preconditioners, with the R-adaptive preconditioner and with an initial Block-Jacobi preconditioner. We notice that even though the number of iterations of PCG is reduced when going from the initial BJ preconditioner to the R-adaptive

[1] The 20 iterations count here is due to the fact that the intermediate solution is used to compute the new initial guess for the R-adaptive preconditioner.

preconditioner, we still get an improvement by using the L-adaptive one, in both variants L-adap prec (LORASC) and L-adap prec (BJ(2)). In fact, the drop in the number of iterations is more important in the third test case, where the R-adaptive preconditioner is not sufficient to significantly reduce the number of iterations, whereas the L-adaptive one manages to converge within around twenty iterations only. This can be due to the fact that the algebraic error is more scattered in this third test case, and that the size of $\mathbf{A}_L$ is almost twice that of $\mathbf{A}_R$ for this test case.

7 Conclusions

In this article, we have presented an adaptive preconditioner that is designed to control the growth rate of the $\underline{\mathbf{A}}_L$-seminorm of the error when used within a fixed-point iteration scheme. Indeed, we have proven the relationship between that growth rate and the largest eigenvalue of the L-block of the preconditioned operator $\mathbf{M}^{-1}\mathbf{A}$. The proposed preconditioner has some similarities with the one already proposed in [1]. We have discussed the link between the two preconditioners and proven the properties satisfied for this type of preconditioners in a fixed-point iteration scheme. We have tested and compared the two approaches when used as preconditioners for a CG solver. A significant speedup is observed with the adaptive preconditioner proposed here with fewer iterations needed for convergence.

References

1. Anciaux-Sedrakian, A., Grigori, L., Jorti, Z., Papež, J., Yousef, S.: Adaptive solution of linear systems of equations based on a posteriori error estimators. Numerical Algorithms (2019). DOI 10.1007/s11075-019-00757-z. URL https://doi.org/10.1007/s11075-019-00757-z
2. Arioli, M., Georgoulis, E.H., Loghin, D.: Stopping criteria for adaptive finite element solvers. SIAM J. Sci. Comput. **35**(3), A1537–A1559 (2013). DOI 10.1137/120867421. URL http://dx.doi.org/10.1137/120867421
3. Arioli, M., Loghin, D., Wathen, A.J.: Stopping criteria for iterations in finite element methods. Numer. Math. **99**(3), 381–410 (2005). DOI 10.1007/s00211-004-0568-z. URL http://dx.doi.org/10.1007/s00211-004-0568-z
4. Axelsson, O.: Iterative Solution Methods. Cambridge University Press, New York (1994). URL https://books.google.com/books?isbn=0521555698
5. Bai, D., Brandt, A.: Local mesh refinement multilevel techniques. SIAM J. Sci. Statist. Comput. **8**(2), 109–134 (1987). DOI 10.1137/0908025. URL http://dx.doi.org/10.1137/0908025
6. Bank, R.E., Sherman, A.H.: An adaptive, multilevel method for elliptic boundary value problems. Computing **26**(2), 91–105 (1981). DOI 10.1007/BF02241777. URL http://dx.doi.org/10.1007/BF02241777
7. Bank, R.E., Smith, R.K.: A posteriori error estimates based on hierarchical bases. SIAM J. Numer. Anal. **30**(4), 921–935 (1993). DOI 10.1137/0730048. URL http://dx.doi.org/10.1137/0730048
8. Brandt, A.: Multi-level adaptive solutions to boundary-value problems. Math. Comp. **31**(138), 333–390 (1977)
9. Dolean, V., Jolivet, P., Nataf, F.: An Introduction to Domain Decomposition Methods: Algorithms, Theory, and Parallel Implementation. SIAM (2015)
10. Ern, A., Vohralík, M.: Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. SIAM J. Sci. Comput. **35**(4), A1761–A1791 (2013). DOI 10.1137/120896918
11. Grigori, L., Nataf, F., Yousef, S.: Robust algebraic Schur complement preconditioners based on low rank corrections. Research Report RR-8557, INRIA (2014). URL https://hal.inria.fr/hal-01017448

12. Jiránek, P., Strakoš, Z., Vohralík, M.: A posteriori error estimates including algebraic error and stopping criteria for iterative solvers. SIAM J. Sci. Comput. **32**(3), 1567–1590 (2010). DOI 10.1137/08073706X. URL http://dx.doi.org/10.1137/08073706X
13. Kolotilina, L.Y.: Bounds for eigenvalues of symmetric block jacobi scaled matrices. Journal of Mathematical Sciences **79**(3), 1043–1047 (1996). DOI 10.1007/BF02366126
14. Liu, J., Marsden, A.L.: A robust and efficient iterative method for hyper-elastodynamics with nested block preconditioning. Journal of Computational Physics **383**, 72 – 93 (2019). DOI https://doi.org/10.1016/j.jcp.2019.01.019
15. Mandel, J.: On block diagonal and schur complement preconditioning. Numerische Mathematik **58**(1), 79–93 (1990)
16. Meidner, D., Rannacher, R., Vihharev, J.: Goal-oriented error control of the iterative solution of finite element equations. J. Numer. Math. **17**(2), 143–172 (2009). DOI 10.1515/JNUM.2009.009. URL http://dx.doi.org/10.1515/JNUM.2009.009
17. Miraçi, A., Papež, J., Vohralík, M.: A multilevel algebraic error estimator and the corresponding iterative solver with p-robust behavior (2019). URL https://hal.archives-ouvertes.fr/hal-02070981. Working paper or preprint
18. Oswald, P.: Multilevel Finite Element Approximation: Theory & Applications. Teubner Skripten zur Numerik. Teubner, Stuttgart (1994)
19. Papež, J., Rüde, U., Vohralík, M., Wohlmuth, B.: Sharp algebraic and total a posteriori error bounds for h and p finite elements via a multilevel approach (2017). URL https://hal.inria.fr/hal-01662944. HAL-preprint
20. Papež, J., Strakoš, Z., Vohralík, M.: Estimating and localizing the algebraic and total numerical errors using flux reconstructions. Numerische Mathematik **138**(3), 681–721 (2018). DOI 10.1007/s00211-017-0915-5
21. Rüde, U.: Fully adaptive multigrid methods. SIAM J. Numer. Anal. **30**(1), 230–248 (1993)
22. Rüde, U.: Mathematical and computational techniques for multilevel adaptive methods, *Frontiers in Applied Mathematics*, vol. 13. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1993). DOI 10.1137/1.9781611970968. URL http://dx.doi.org/10.1137/1.9781611970968
23. Rüde, U.: Error estimates based on stable splittings. In: Domain decomposition methods in scientific and engineering computing (University Park, PA, 1993), *Contemp. Math.*, vol. 180, pp. 111–118. Amer. Math. Soc., Providence, RI (1994)

Appendix

Block-Jacobi preconditioning

In this section, we recall a property of Block-Jacobi preconditioners. We follow the theory developed in [4] and generalize the results to the case with m blocks.

Let the symmetric positive definite matrix $\mathbf{A}$ be partitioned into m blocks as follows: $\mathbf{A} = [\mathbf{A}_{ij}]$ for $i, j = 1, \ldots, m$. We denote by n_i the size of the diagonal block $\mathbf{A}_{ii}$, and by $(V_i)_{1 \leq i \leq n}$ the finite dimensional spaces consistent with the above block partition of $\mathbf{A}$. Let $\mathbf{M}$ be the Block-Jacobi preconditioner of $\mathbf{A}$: $\mathbf{M} = [\mathbf{M}_{ij}]$ for $i, j = 1, \ldots, m$ where:

$$\mathbf{M}_{ij} = \begin{cases} \mathbf{A}_{ii} \text{ if } i = j \\ \mathbf{0}_{ij} \text{ otherwise} \end{cases} \tag{7.1}$$

Let γ_{ij} be the Cauchy-Schwarz-Bunyakowski (C.B.S.) constant [4] defined for the 2×2 block matrix composed by $\mathbf{A}_{ii}$ and $\mathbf{A}_{jj}$ as diagonal blocks, $\mathbf{A}_{ij}$ and $\mathbf{A}_{ji}$ as off-

diagonal blocks:

$$\gamma_{ij} := \sup_{\mathbf{v}_i \in V_i, \mathbf{v}_j \in V_j} \frac{\mathbf{v}_i^{\mathrm{T}} \mathbf{A}_{ij} \mathbf{v}_j}{\left(\mathbf{v}_i^{\mathrm{T}} \mathbf{A}_{ii} \mathbf{v}_i \mathbf{v}_j^{\mathrm{T}} \mathbf{A}_{jj} \mathbf{v}_j\right)^{\frac{1}{2}}} \tag{7.2}$$

Lemma 7.1 *Let λ be an eigenvalue of $\mathbf{M}^{-1}\mathbf{A}$. We have*

$$1 - \max_{1 \le i \le m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \gamma_{ij} \le \lambda \le 1 + \max_{1 \le i \le m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \gamma_{ij} \tag{7.3}$$

Proof The extreme eigenvalues of $\mathbf{A} \cdot \mathbf{x} = \lambda \mathbf{M} \cdot \mathbf{x}$ are the extreme values of

$$\frac{\mathbf{x}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{x}}{\mathbf{x}^{\mathrm{T}} \cdot \mathbf{M} \cdot \mathbf{x}} = \frac{\sum_{j=1}^{m} \mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j + \sum_{j=1}^{m} \sum_{\substack{i=1 \\ i \neq j}}^{m} \mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{ji} \cdot \mathbf{x}_i}{\sum_{j=1}^{m} \mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j}; \qquad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}$$

For each pair of distinct indices (i, j), we have:

$$|\mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{ji} \cdot \mathbf{x}_i| \le \gamma_{ij} \left(\mathbf{x}_i^{\mathrm{T}} \mathbf{A}_{ii} \mathbf{x}_i \mathbf{x}_j^{\mathrm{T}} \mathbf{A}_{jj} \mathbf{x}_j\right)^{\frac{1}{2}} \le \frac{\gamma_{ij}}{2} (\mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j + \mathbf{x}_i^{\mathrm{T}} \cdot \mathbf{A}_{ii} \cdot \mathbf{x}_i)$$

Therefore

$$\mathbf{x}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{x} - \sum_{j=1}^{m} \mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j \le \sum_{j=1}^{m} \sum_{\substack{i=1 \\ i \neq j}}^{m} \frac{\gamma_{ji}}{2} (\mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j + \mathbf{x}_i^{\mathrm{T}} \cdot \mathbf{A}_{ii} \cdot \mathbf{x}_i)$$

However, due to the symmetry we notice that:

$$\sum_{j=1}^{m} \sum_{\substack{i=1 \\ i \neq j}}^{m} \frac{\gamma_{ji}}{2} (\mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j + \mathbf{x}_i^{\mathrm{T}} \cdot \mathbf{A}_{ii} \cdot \mathbf{x}_i) = \sum_{j=1}^{m} \mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j \sum_{\substack{i=1 \\ i \neq j}}^{m} \gamma_{ji}$$

Consequently

$$\mathbf{x}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{x} \le \sum_{j=1}^{m} \mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j (1 + \sum_{\substack{i=1 \\ i \neq j}}^{m} \gamma_{ji})$$

$$\mathbf{x}^{\mathrm{T}} \cdot \mathbf{A} \cdot \mathbf{x} \le (1 + \max_{1 \le i \le m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \gamma_{ij}) \sum_{j=1}^{m} \mathbf{x}_j^{\mathrm{T}} \cdot \mathbf{A}_{jj} \cdot \mathbf{x}_j$$

This latter inequality completes the proof. □

Here, we retrieve a known feature of Block-Jacobi preconditioners: they bound the maximum eigenvalue of the preconditioned matrix (see for example [13,4,15] and references therein). Indeed, according to Lemma 7.1, a Block-Jacobi preconditioner allows to keep the maximum eigenvalue of the preconditioned operator $\mathbf{M}^{-1}\mathbf{A}$ bounded by a constant that depends on the blocks of matrix. In fact, as the C.B.S. constants are less than or equal to 1 (because $\mathbf{A}$ is SPD [4]), we can deduce that the maximum eigenvalue is bounded by m the number of blocks.