Analysis of the SORAS domain decomposition preconditioner for non-self-adjoint or indefinite problems

Marcella Bonazzoli, Xavier Claeys, Frédéric Nataf, Pierre-Henri Tournier

Noname manuscript No.
(will be inserted by the editor)

Analysis of the SORAS domain decomposition preconditioner for non-self-adjoint or indefinite problems

Marcella Bonazzoli · Xavier Claeys · Frédéric Nataf · Pierre-Henri Tournier

Received: date / Accepted: date

Abstract We analyze the convergence of the one-level overlapping domain decomposition preconditioner SORAS (Symmetrized Optimized Restricted Additive Schwarz) applied to a generic linear system whose matrix is not necessarily symmetric/self-adjoint nor positive definite. By generalizing the theory for the Helmholtz equation developed in [I.G. Graham, E.A. Spence, and J. Zou, *SIAM J. Numer. Anal.*, 2020], we identify a list of assumptions and estimates that are sufficient to obtain an upper bound on the norm of the preconditioned matrix, and a lower bound on the distance of its field of values from the origin. We stress that our theory is general in the sense that it is not specific to one particular boundary value problem. Moreover, it does not rely on a coarse mesh whose elements are sufficiently small. As an illustration of this framework, we prove new estimates for overlapping domain decomposition methods with Robin-type transmission conditions for the heterogeneous reaction-convection-diffusion equation (to prove the stability assumption for this equation we consider the case of a coercive bilinear form, which is non-symmetric, though).

Keywords Non-self-adjoint problems · indefinite problems · domain decomposition · preconditioners · field of values · reaction-convection-diffusion equation

Mathematics Subject Classification (2010) 65N55 · 65F08 · 65F10 · 76R99

1 Introduction

The discretization of several partial differential equations relevant in applications, such as the Helmholtz equation, the time-harmonic Maxwell equations or the

M. Bonazzoli
Inria, Centre de Mathématiques Appliquées, CNRS, École Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France.
E-mail: marcella.bonazzoli@inria.fr, corresponding author

X. Claeys · F. Nataf · P.-H. Tournier
Sorbonne Université, CNRS, Université de Paris, Inria, Laboratoire Jacques-Louis Lions, F-75005 Paris, France. E-mail: xavier.claeys@sorbonne-universite.fr, frederic.nataf@sorbonne-universite.fr, tournier@ljll.math.upmc.fr

reaction-convection-diffusion equation, yields linear systems whose matrices are not symmetric/self-adjoint or indefinite. The rigorous analysis of the convergence of preconditioned iterative methods for such problems is harder than for symmetric positive definite (SPD) problems. Indeed, in the SPD case, Hilbert space theorems such as the Fictitious Space lemma (see e.g. [27,19]) yield a powerful general framework of spectral analysis for domain decomposition preconditioners. In addition, in the non-SPD case the conjugate gradient method cannot be used, and the analysis of the spectrum of the preconditioned matrix is not sufficient for iterative methods such as GMRES suited for non-self-adjoint matrices. In fact, as stated in [18], "any nonincreasing convergence curve can be obtained with GMRES applied to a matrix having any desired eigenvalues". In the literature, GMRES convergence estimates are based for instance on the field of values [13,12,4] or on the pseudo-spectrum (see [31] and references therein) of the preconditioned operator. For example, field of values bounds were derived for overlapping domain decomposition preconditioners for non-symmetric parabolic problems which are small perturbations of SPD operators [9], and later for the high-frequency Helmholtz [16, 17,15] and time-harmonic Maxwell [5] equations.

Here, by generalizing the work of [17], we analyze for generic problems the convergence of the preconditioned GMRES method in its weighted version [14]. We identify a list of assumptions and estimates that are sufficient to obtain an upper bound on the norm of the preconditioned matrix, and a lower bound on the distance of its field of values from the origin. This analysis applies to a class of one-level overlapping Schwarz domain decomposition preconditioners, with Robin-type or more general absorbing transmission conditions on the interfaces between subdomains. This type of preconditioners with the basic Robin-type transmission conditions was first introduced in ([24], 2007) for the Helmholtz equation and called OBDD-H (*Overlapping Balancing Domain Decomposition for Helmholtz*). It was later studied in ([21], 2015) for generic symmetric positive definite problems and viewed as a symmetric variant of the ORAS preconditioner ([30], 2007), hence called SORAS (*Symmetrized Optimized Restricted Additive Schwarz*). Note that in [24] several one-level and two-level versions, with a coarse space based on plane waves, were tested numerically, and more than ten years later the one-level OBDD-H version was rigorously analyzed in [17], for the Helmholtz equation. In [21] a two-level version, with a spectral coarse space, was rigorously analyzed for generic SPD problems. The present article gives, to the best of our knowledge, the first rigorous analysis of such one-level preconditioners for *generic* non-SPD problems.

Furthermore, we apply our general framework to the case of convection-diffusion equations to obtain, for the first time, convergence bounds for one-level overlapping Schwarz domain decomposition preconditioners with Robin-type transmission conditions. For these equations, the two-level overlapping case, but with standard Dirichlet transmission conditions, was analyzed in [7,8], where a coarse space is built from a coarse mesh whose elements are sufficiently small. As for the one-level non-overlapping case, it was studied with Robin or more general transmission conditions in e.g. [25,26], see also [22] for some numerical results. Apart from Schwarz methods, the Neumann–Neumann algorithm [6], which belongs to the substructuring family of domain decomposition methods, was generalized to convection-diffusion equations in [1], and a coarse space not based on a coarse mesh was proposed in [2] although without convergence analysis.

The paper is structured as follows. In section 2 we first describe in detail the considered class of domain decomposition preconditioners and introduce notation for the global and local inner products and norms. In section 3 we state and prove the main theorem, which provides a general and practical tool for the rigorous convergence analysis of the preconditioner. This framework is applied in section 4 to the case of the heterogeneous reaction-convection-diffusion equation. After specifying the global and local bilinear forms, inner products and norms and the discretization, we prove estimates for the assumptions of the theorem for this equation, without making any a priori assumption on the regime of the physical coefficients nor of the numerical parameters; to prove the stability assumption for this equation we consider the case of a coercive bilinear form (which is non-symmetric, though). Finally, we discuss for a particular regime the resulting lower bound on the field of values, and we test numerically the performance of the preconditioner.

2 Setting

Let A denote the $n \times n$ (potentially complex-valued) matrix arising from the discretization of the problem to be solved, posed in an open domain $\Omega \subset \mathbb{R}^d$. The matrix A is not necessarily positive definite nor self-adjoint. This means that here we do *not* necessarily require $A^* = A$, where $A^* := \overline{A^T}$; note that "self-adjoint matrix" is a synonym for "Hermitian matrix". In particular, if A is real-valued this means that here it does not need to be symmetric.

The definition of the preconditioner is based on a set of overlapping open subdomains $\Omega_j, j = 1, \ldots, N$, such that $\Omega = \cup_{j=1}^N \Omega_j$ and each $\overline{\Omega_j}$ is a union of elements of the mesh $\mathcal{T}^h$ of Ω. Then we consider the set $\mathcal{N}$ of the unknowns on the whole domain, so $\#\mathcal{N} = n$, and its decomposition $\mathcal{N} = \bigcup_{j=1}^N \mathcal{N}_j$ into the non-disjoint subsets corresponding to the different overlapping subdomains Ω_j, with $\#\mathcal{N}_j = n_j$. Then one builds the following matrices (see e.g. [11, §1.3]):

- the *restriction* matrices R_j from Ω to the subdomain Ω_j: they are $n_j \times n$ Boolean matrices whose (i, i') entry equals 1 if the i-th unknown in $\mathcal{N}_j$ is the i'-th one in $\mathcal{N}$ and vanishes otherwise;
- the *extension* by zero matrices from the subdomain Ω_j to Ω, which are $n \times n_j$ Boolean matrices given by R_j^T;
- the *partition of unity* matrices D_j, which are $n_j \times n_j$ diagonal matrices with real non-negative entries such that $\sum_{j=1}^N R_j^T D_j R_j = I$. They can be seen as matrices that properly weight the unknowns belonging to the overlap between subdomains;
- the *local matrices* B_j, of size $n_j \times n_j$, arising from the discretization of subproblems posed in Ω_j, with for instance Robin-type or absorbing[1] transmission conditions on the interfaces $\partial\Omega_j \setminus \partial\Omega$.

[1] Absorbing boundary conditions are approximations of transparent boundary conditions. Basic absorbing boundary conditions are Robin-type boundary conditions, which consist in a weighted combination of Neumann-type and Dirichlet-type boundary conditions. Their precise definition depends on the specific problem. For instance, for Maxwell equations impedance boundary conditions are Robin-type absorbing boundary conditions.

Finally, the one-level Symmetrized Optimized Restricted Additive Schwarz (SORAS) preconditioner is defined as

$$M^{-1} \coloneqq \sum_{j=1}^{N} R_j^T D_j B_j^{-1} D_j R_j. \tag{2.1}$$

Note that here the preconditioner is not self-adjoint when B_j is not self-adjoint, even if we maintain the SORAS name, where S stands for 'Symmetrized'. In fact, this denomination was introduced in [21] for SPD problems, since in that case the SORAS preconditioner is a symmetric variant of the ORAS preconditioner $\sum_{j=1}^{N} R_j^T D_j B_j^{-1} R_j$. Thus, the adjective 'Symmetrized' stands for the presence of the rightmost partition of unity D_j. We recall that the adjective 'Restricted' indicates the presence of the leftmost partition of unity D_j. The adjective 'Optimized' refers to the choice of transmission conditions other than standard Dirichlet conditions in the local matrices B_j, which can be better suited to the problem at hand and accelerate the convergence of the method.

The weighted GMRES method [14] differs from the standard one in the norm used for the residual minimization, which is not the standard Hermitian norm but a more general weighted norm. For vectors of degrees of freedom $\mathbf{V}, \mathbf{W} \in \mathbb{C}^n$, using the notation $(\mathbf{V}, \mathbf{W}) \coloneqq \mathbf{W}^* \mathbf{V}$ to indicate the Hermitian inner product, given a $n \times n$ self-adjoint positive definite matrix F_Ω, we consider the weighted norm

$$\|\mathbf{V}\|_\Omega \coloneqq (\mathbf{V}, \mathbf{V})_{F_\Omega}^{1/2}, \quad \text{where } (\mathbf{V}, \mathbf{W})_{F_\Omega} \coloneqq (F_\Omega \mathbf{V}, \mathbf{W}) = \mathbf{W}^* F_\Omega \mathbf{V}.$$

Locally, on the subdomain Ω_j, we consider a weighted norm represented by a $n_j \times n_j$ self-adjoint positive definite matrix F_{Ω_j}: for vectors of degrees of freedom $\mathbf{V}^j, \mathbf{W}^j \in \mathbb{C}^{n_j}$ local to Ω_j, we define

$$\|\mathbf{V}^j\|_{\Omega_j} \coloneqq (\mathbf{V}^j, \mathbf{V}^j)_{F_{\Omega_j}}^{1/2}, \quad \text{where } (\mathbf{V}^j, \mathbf{W}^j)_{F_{\Omega_j}} \coloneqq (F_{\Omega_j} \mathbf{V}^j, \mathbf{W}^j) = (\mathbf{W}^j)^* F_{\Omega_j} \mathbf{V}^j.$$

Typically F_{Ω_j} is a Neumann-type matrix on Ω_j, that is, coming from an inner product at the continuous level with no boundary integral.

3 General theory

In order to apply Elman-type estimates for the convergence of weighted GMRES [14], such as [16, Theorem 5.1] or its improvement [5, Theorem 5.3], we need to prove an upper bound on the weighted norm of the preconditioned matrix, and a lower bound on the distance of its weighted field of values from the origin. Recall that the *field of values (or numerical range)* of a matrix C *with respect to* the inner product induced by a matrix F is the set defined as

$$W_F(C) = \{ (\mathbf{V}, C\mathbf{V})_F \mid \mathbf{V} \in \mathbb{C}^n, \|\mathbf{V}\|_F = 1 \}.$$

(Note that the convergence estimate for GMRES based on the field of values can be used only when this latter does not contain 0.)

The following theorem, which generalizes the theory for the Helmholtz equation developed in [17], identifies assumptions that are sufficient to obtain the two bounds. In particular, the proof was inspired by the one of [17, Theorem 3.11] and by the analysis in subsection [17, §3.2].

We will need the notation for the commutator $[P, Q] \coloneqq PQ - QP$.

Theorem 3.1. *For $j = 1, \ldots, N$, assume that for all global vectors of degrees of freedom $\mathbf{V} \in \mathbb{C}^n$ and local vectors of degrees of freedom $\mathbf{W}^j \in \mathbb{C}^{n_j}$ in Ω_j*

$$(D_j R_j A \mathbf{V}, \mathbf{W}^j) = (D_j B_j R_j \mathbf{V}, \mathbf{W}^j). \tag{3.1}$$

Suppose that there exists $\Lambda_0 > 0$ such that for all local vectors of degrees of freedom $\mathbf{W}^j \in \mathbb{C}^{n_j}$ in Ω_j, $j = 1, \ldots, N$, we have

$$\left\| \sum_{j=1}^{N} R_j^T \mathbf{W}^j \right\|_{\Omega}^2 \leq \Lambda_0 \sum_{j=1}^{N} \|\mathbf{W}^j\|_{\Omega_j}^2, \tag{3.2}$$

and $\Lambda_1 > 0$ such that for all global vectors of degrees of freedom $\mathbf{V} \in \mathbb{C}^n$

$$\sum_{j=1}^{N} \|R_j \mathbf{V}\|_{\Omega_j}^2 \leq \Lambda_1 \|\mathbf{V}\|_{\Omega}^2. \tag{3.3}$$

For $j = 1, \ldots, N$, suppose also that there exist $C_{D,j}, C_{DB,j} > 0$ such that for all local vectors of degrees of freedom $\mathbf{W}^j, \mathbf{V}^j \in \mathbb{C}^{n_j}$ in Ω_j

$$\|D_j \mathbf{W}^j\|_{\Omega_j} \leq C_{D,j} \|\mathbf{W}^j\|_{\Omega_j}, \tag{3.4}$$

$$|([D_j, B_j]\mathbf{V}^j, \mathbf{W}^j)| \leq C_{DB,j} \|\mathbf{V}^j\|_{\Omega_j} \|\mathbf{W}^j\|_{\Omega_j}, \tag{3.5}$$

and that B_j satisfies the following inf-sup condition: there exists $C_{\mathrm{stab},j} > 0$ such that for all local vectors of degrees of freedom $\mathbf{U}^j \in \mathbb{C}^{n_j}$

$$\|\mathbf{U}^j\|_{\Omega_j} \leq C_{\mathrm{stab},j} \max_{\mathbf{W}^j \in \mathbb{C}^{n_j} \setminus \{0\}} \left(\frac{|(B_j \mathbf{U}^j, \mathbf{W}^j)|}{\|\mathbf{W}^j\|_{\Omega_j}} \right). \tag{3.6}$$

Then, we obtain the following upper bound on the norm of the preconditioned matrix:

$$\boxed{\max_{\mathbf{V} \in \mathbb{C}^n} \frac{\|M^{-1} A \mathbf{V}\|_{\Omega}}{\|\mathbf{V}\|_{\Omega}} \leq \sqrt{\Lambda_0 \Lambda_1} \max_{j=1,\ldots,N} \{C_{D,j}(C_{\mathrm{stab},j} C_{DB,j} + C_{D,j})\}.} \tag{3.7}$$

If in addition, for $j = 1, \ldots, N$, for all global vectors of degrees of freedom $\mathbf{V} \in \mathbb{C}^n$ and local vectors of degrees of freedom $\mathbf{W}^j \in \mathbb{C}^{n_j}$ in Ω_j

$$(D_j R_j F_{\Omega} \mathbf{V}, \mathbf{W}^j) = (D_j F_{\Omega_j} R_j \mathbf{V}, \mathbf{W}^j), \tag{3.8}$$

and there exists $C_{DF,j} > 0$ such that for all local vectors of degrees of freedom $\mathbf{V}^j, \mathbf{W}^j \in \mathbb{C}^{n_j}$ in Ω_j

$$|([D_j, F_{\Omega_j}]\mathbf{V}^j, \mathbf{W}^j)| \leq C_{DF,j} \|\mathbf{V}^j\|_{\Omega_j} \|\mathbf{W}^j\|_{\Omega_j}, \tag{3.9}$$

then we obtain the following lower bound on the distance of the field of values of the preconditioned matrix from the origin:

$$\boxed{\begin{aligned} \min_{\mathbf{V} \in \mathbb{C}^n} \frac{|(F_{\Omega} \mathbf{V}, M^{-1} A \mathbf{V})|}{\|\mathbf{V}\|_{\Omega}^2} \geq \frac{1}{\Lambda_0} &- \Lambda_1 \max_{j=1,\ldots,N} \{C_{D,j} C_{\mathrm{stab},j} C_{DB,j}\} \\ &- \Lambda_1 \max_{j=1,\ldots,N} \{C_{DF,j}(C_{\mathrm{stab},j} C_{DB,j} + C_{D,j})\}. \end{aligned}} \tag{3.10}$$

Remark 3.2. We will comment on assumptions (3.1), (3.2), (3.3), (3.8) in subsection 3.1. Note that in finite dimension, the constants in assumptions (3.4), (3.5), (3.6), (3.9) are finite, and in the statement of the theorem we actually mean that we are able to estimate these constants.

Proof. To obtain both bounds an important quantity is

$$\|(B_j^{-1}D_jR_jA - D_jR_j)\mathbf{V}\|_{\Omega_j}.$$

For its estimate, for any vector of degrees of freedom $\mathbf{W}^j \in \mathbb{C}^{n_j}$ local to Ω_j, write

$$\begin{aligned}(B_j(B_j^{-1}D_jR_jA - D_jR_j)\mathbf{V}, \mathbf{W}^j) &= (D_jR_jA\mathbf{V}, \mathbf{W}^j) - (B_jD_jR_j\mathbf{V}, \mathbf{W}^j)\\ &\overset{(3.1)}{=} (D_jB_jR_j\mathbf{V}, \mathbf{W}^j) - (B_jD_jR_j\mathbf{V}, \mathbf{W}^j)\\ &= ([D_j, B_j]R_j\mathbf{V}, \mathbf{W}^j),\end{aligned}$$

where assumption (3.1) was used. Thus we have found that $(B_j^{-1}D_jR_jA - D_jR_j)\mathbf{V}$ is the solution to a local problem with a right-hand side involving the commutator between the partition of unity and the local matrix. So by the stability bound (3.6), we have:

$$\|(B_j^{-1}D_jR_jA - D_jR_j)\mathbf{V}\|_{\Omega_j} \leq C_{\mathrm{stab},j} \max_{\mathbf{W}^j \in \mathbb{C}^{n_j}\setminus\{0\}} \left(\frac{|([D_j, B_j]R_j\mathbf{V}, \mathbf{W}^j)|}{\|\mathbf{W}^j\|_{\Omega_j}}\right).$$

Moreover by assumption (3.5)

$$|([D_j, B_j]R_j\mathbf{V}, \mathbf{W}^j)| \leq C_{DB,j}\|R_j\mathbf{V}\|_{\Omega_j}\|\mathbf{W}^j\|_{\Omega_j} \;\forall\, \mathbf{W}^j.$$

Therefore

$$\|(B_j^{-1}D_jR_jA - D_jR_j)\mathbf{V}\|_{\Omega_j} \leq C_{\mathrm{stab},j}C_{DB,j}\|R_j\mathbf{V}\|_{\Omega_j}. \tag{3.11}$$

Together with (3.11), a direct consequence of (3.11) itself and assumption (3.4) will be also used repeatedly:

$$\|B_j^{-1}D_jR_jA\mathbf{V}\|_{\Omega_j} \leq (C_{\mathrm{stab},j}C_{DB,j} + C_{D,j})\|R_j\mathbf{V}\|_{\Omega_j}. \tag{3.12}$$

Now, it is easy to obtain the upper bound (3.7): for $\mathbf{V} \in \mathbb{C}^n$ we have

$$\begin{aligned}\left\|\sum_{j=1}^N R_j^T D_j B_j^{-1} D_j R_j A\mathbf{V}\right\|_\Omega^2 &\overset{(3.2)}{\leq} \Lambda_0 \sum_{j=1}^N \|D_jB_j^{-1}D_jR_jA\mathbf{V}\|_{\Omega_j}^2\\ &\overset{(3.4)}{\leq} \Lambda_0 \sum_{j=1}^N C_{D,j}^2\|B_j^{-1}D_jR_jA\mathbf{V}\|_{\Omega_j}^2\\ &\overset{(3.12)}{\leq} \Lambda_0 \sum_{j=1}^N C_{D,j}^2(C_{\mathrm{stab},j}C_{DB,j} + C_{D,j})^2\|R_j\mathbf{V}\|_{\Omega_j}^2\\ &\overset{(3.3)}{\leq} \Lambda_0\Lambda_1 \max_{j=1,\dots,N}\{C_{D,j}^2(C_{\mathrm{stab},j}C_{DB,j} + C_{D,j})^2\}\|\mathbf{V}\|_\Omega^2,\end{aligned}$$

where we have indicated above each inequality sign which equation was used.

The derivation of the lower bound (3.10) is more involved. First of all write

$$(F_\Omega \mathbf{V}, \sum_{j=1}^{N} R_j^T D_j B_j^{-1} D_j R_j A\mathbf{V}) = \sum_{j=1}^{N} (F_\Omega \mathbf{V}, R_j^T D_j B_j^{-1} D_j R_j A\mathbf{V})$$
$$= \sum_{j=1}^{N} (D_j R_j F_\Omega \mathbf{V}, B_j^{-1} D_j R_j A\mathbf{V}) \overset{(3.8)}{=} \sum_{j=1}^{N} (D_j F_{\Omega_j} R_j \mathbf{V}, B_j^{-1} D_j R_j A\mathbf{V}),$$

where, beside applying assumption (3.8), we have used the fact that the partition of unity matrices D_j are real-valued and diagonal, hence symmetric, and the restriction matrices R_j satisfy $(\mathbf{V}, R_j^T \mathbf{W}^j) = (R_j \mathbf{V}, \mathbf{W}^j)$. Now, we make appear the commutator between the partition of unity and the local inner product matrix, and also the quantity $(B_j^{-1} D_j R_j A - D_j R_j)\mathbf{V}$:

$$\begin{aligned}
&(D_j F_{\Omega_j} R_j \mathbf{V}, B_j^{-1} D_j R_j A\mathbf{V}) \\
&= (F_{\Omega_j} D_j R_j \mathbf{V}, B_j^{-1} D_j R_j A\mathbf{V}) + ([D_j, F_{\Omega_j}] R_j \mathbf{V}, B_j^{-1} D_j R_j A\mathbf{V}) \\
&= (F_{\Omega_j} D_j R_j \mathbf{V}, D_j R_j \mathbf{V}) + (F_{\Omega_j} D_j R_j \mathbf{V}, (B_j^{-1} D_j R_j A - D_j R_j)\mathbf{V}) \\
&\quad + ([D_j, F_{\Omega_j}] R_j \mathbf{V}, B_j^{-1} D_j R_j A\mathbf{V}).
\end{aligned}$$

Therefore

$$\begin{aligned}
&|(F_\Omega \mathbf{V}, M^{-1} A\mathbf{V})| \geq \\
&\sum_{j=1}^{N} \|D_j R_j \mathbf{V}\|_{\Omega_j}^2 - \sum_{j=1}^{N} |(F_{\Omega_j} D_j R_j \mathbf{V}, (B_j^{-1} D_j R_j A - D_j R_j)\mathbf{V})| \\
&- \sum_{j=1}^{N} |([D_j, F_{\Omega_j}] R_j \mathbf{V}, B_j^{-1} D_j R_j A\mathbf{V})|.
\end{aligned} \tag{3.13}$$

For the first term in (3.13) we use the partition of unity property $\sum_{j=1}^{N} R_j^T D_j R_j = I$ and assumption (3.2) with $\mathbf{W}^j = D_j R_j \mathbf{V}$:

$$\|\mathbf{V}\|_\Omega^2 = \left\| \sum_{j=1}^{N} R_j^T (D_j R_j \mathbf{V}) \right\|_\Omega^2 \overset{(3.2)}{\leq} \Lambda_0 \sum_{j=1}^{N} \|D_j R_j \mathbf{V}\|_{\Omega_j}^2,$$

so

$$\sum_{j=1}^{N} \|D_j R_j \mathbf{V}\|_{\Omega_j}^2 \geq \frac{1}{\Lambda_0} \|\mathbf{V}\|_\Omega^2.$$

For the second term in (3.13), we use first the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\sum_{j=1}^{N} |(F_{\Omega_j} D_j R_j \mathbf{V}, (B_j^{-1} D_j R_j A - D_j R_j)\mathbf{V})| \\
&\quad \leq \sum_{j=1}^{N} \|D_j R_j \mathbf{V}\|_{\Omega_j} \|(B_j^{-1} D_j R_j A - D_j R_j)\mathbf{V}\|_{\Omega_j} \\
&\quad \overset{(3.4),(3.11)}{\leq} \sum_{j=1}^{N} C_{D,j} C_{\mathrm{stab},j} C_{DB,j} \|R_j \mathbf{V}\|_{\Omega_j}^2 \\
&\quad \overset{(3.3)}{\leq} \Lambda_1 \max_{j=1,\ldots,N} \{C_{D,j} C_{\mathrm{stab},j} C_{DB,j}\} \|\mathbf{V}\|_\Omega^2.
\end{aligned}$$

Finally for the third term in (3.13) we write

$$\begin{aligned}
&\sum_{j=1}^{N} |([D_j, F_{\Omega_j}]R_j\mathbf{V}, B_j^{-1}D_jR_jA\mathbf{V})| \\
&\overset{(3.9)}{\leq} \sum_{j=1}^{N} C_{DF,j}\|R_j\mathbf{V}\|_{\Omega_j}\|B_j^{-1}D_jR_jA\mathbf{V}\|_{\Omega_j} \\
&\overset{(3.12)}{\leq} \sum_{j=1}^{N} C_{DF,j}(C_{\mathrm{stab},j}C_{DB,j}+C_{D,j})\|R_j\mathbf{V}\|_{\Omega_j}^2 \\
&\overset{(3.3)}{\leq} \Lambda_1 \max_{j=1,\dots,N}\{C_{DF,j}(C_{\mathrm{stab},j}C_{DB,j}+C_{D,j})\}\|\mathbf{V}\|_{\Omega}^2.
\end{aligned}$$

In conclusion, inserting these estimations in (3.13) we obtain the lower bound (3.10).

□

3.1 Comments on the assumptions of Theorem 3.1

Assumptions (3.1) and (3.8) relate the global matrices with the local ones through the partition of unity and restriction matrices. They may appear unconventional at first glance, but they are satisfied for quite natural choices of the local sesquilinear form and continuous norm on the subdomains. More precisely, if the i-th entry of the diagonal of D_j is not zero, assumption (3.1) requires that the i-th rows of R_jA and B_jR_j are equal; likewise assumption (3.8) requires that the i-th rows of R_jF_Ω and $F_{\Omega_j}R_j$ are equal. First of all, note that typically the entries corresponding to $\partial\Omega_j \setminus \partial\Omega$ of the partition of unity D_j are zero. Moreover, B_j arises from the discretization of a local sesquilinear form that usually is like the global sesquilinear form yielding A but with the integrals on Ω_j instead of Ω and with an additional boundary integral on $\partial\Omega_j \setminus \partial\Omega$. In this case assumption (3.1) is satisfied. Likewise, assumption (3.8) is satisfied if the local continuous norm yielding F_{Ω_j} is obtained from the global continuous norm yielding F_Ω just by replacing Ω with Ω_j in the integration domain. As an illustration, see the bilinear forms a, a_j and the continuous norms $\|\cdot\|_{1,c}$, $\|\cdot\|_{1,c,\Omega_j}$ defined in §4 for the reaction-convection-diffusion equation and the proof of Lemma 4.7: in this case the essential properties on the continuous level are those expressed in Remak 4.1.

Assumptions (3.2) and (3.3) are classical inequalities in the domain decomposition framework. Inequality (3.2) is dubbed in [17] 'a kind of converse to the stable splitting result', and it can be viewed as a continuity property of the reconstruction operator $\{\mathbf{W}^j\}_{j=1}^N \mapsto \sum_{j=1}^N R_j^T\mathbf{W}^j$. In [17, Lemma 3.6] the inequality is proved at the continuous level for the Helmholtz energy norm (see [17, eq. (1.15)]) with

$$\Lambda_0 = \max_{j=1,\dots,N} \#\Lambda(j), \quad \text{where } \Lambda(j) \coloneqq \{\, i \mid \Omega_j \cap \Omega_i \neq \emptyset \,\},$$

in other words, Λ_0 is the maximum number of neighboring subdomains. Note that the proof in [17, Lemma 3.6] (essentially consisting in the one in [16, eq. (4.8)]) is more generally valid, for instance whenever the local continuous norm can be obtained from the global continuous norm just by replacing Ω with Ω_j in the

integration domain, as before. The equivalent of assumption (3.2) at the continuous level can be found in Lemma 4.8.

When the local and the global continuous norms are related as above again, it is immediate to prove inequality (3.3) with

$$\Lambda_1 = \max\{\, m \mid \exists\, j_1 \neq \ldots \neq j_m \text{ such that } \operatorname{meas}(\Omega_{j_1} \cap \cdots \cap \Omega_{j_m}) \neq 0 \,\},$$

that is Λ_1 is the maximal multiplicity of the subdomain intersection (this constant is like the one defined in [11, Lemma 7.13] and is slightly more precise than Λ_0 that was used in [17, eq. (2.10)]). The equivalent of assumption (3.3) at the continuous level can be found in Lemma 4.9. Note that Λ_0 and Λ_1 are geometric constants, related to the decomposition into overlapping subdomains.

The remaining assumptions can be also expressed in the finite element language, which is introduced in the next section: see equation (4.21) for the stability assumption (3.6); equation (4.23) for assumption (3.4) on the partition of unity; equations (4.26), (4.29) for assumptions (3.9),(3.5) on the commutators between the partition of unity and the local (inner product and problem) matrices.

4 The reaction-convection-diffusion equation

As an illustration of the general theory, we apply Theorem 3.1 to the case of the heterogeneous reaction-convection-diffusion equation; recall that the convergence theory for the homogeneous, respectively heterogeneous, Helmholtz equation was developed in [17], respectively [15]. Let $\Omega \subset \mathbb{R}^d$ be an open bounded polyhedral domain. We study the heterogeneous reaction-convection-diffusion problem in conservative form, with Robin-type and Dirichlet boundary conditions:

$$\begin{cases} c_0 u + \operatorname{div}(\mathbf{a}u) - \operatorname{div}(\nu \nabla u) = f & \text{in } \Omega, \\ \nu \frac{\partial u}{\partial n} - \frac{1}{2}\mathbf{a}\cdot\mathbf{n}\, u + \alpha u = g & \text{on } \Gamma_R, \\ u = 0 & \text{on } \Gamma_D, \end{cases} \tag{4.1}$$

where $\partial\Omega = \Gamma = \Gamma_R \cup \Gamma_D$, $\mathbf{n}$ is the outward-pointing unit normal vector to Γ, $c_0 \in \mathrm{L}^\infty(\Omega)$, $\mathbf{a} \in \mathrm{L}^\infty(\Omega)^d$, $\operatorname{div}\mathbf{a} \in \mathrm{L}^\infty(\Omega)$, $\nu \in \mathrm{L}^\infty(\Omega)$, $f \in \mathrm{L}^2(\Omega)$, $g \in \mathrm{L}^2(\Gamma_R)$, $\alpha \in \mathrm{L}^\infty(\Omega)$ (in this case all quantities are real-valued). With the notation

$$\tilde{c} \coloneqq c_0 + \frac{1}{2}\operatorname{div}\mathbf{a},$$

suppose that there exist $\tilde{c}_- > 0$, $\tilde{c}_+ > 0$ such that

$$\tilde{c}_- \leq \tilde{c}(\mathbf{x}) \leq \tilde{c}_+ \text{ a.e. in } \Omega, \tag{4.2}$$

(the positiveness of $\tilde{c}(\mathbf{x})$ is a classical assumption in reaction-convection-diffusion equation literature), and there exist $\nu_- > 0$, $\nu_+ > 0$ such that

$$\nu_- \leq \nu(\mathbf{x}) \leq \nu_+ \text{ a.e. in } \Omega,$$

and $\alpha(\mathbf{x}) \geq 0$ a.e. in Ω. Note that the appropriate Robin-type boundary condition (on Γ_R) here is not simply $\nu\frac{\partial u}{\partial n} + \alpha u = g$; we will comment below about a possible choice of α, see (4.3). Now, set $\mathrm{H}^1_{0,D}(\Omega) \coloneqq \{\, v \in \mathrm{H}^1(\Omega) \mid v = 0 \text{ on } \Gamma_D \,\}$. In order

to find the variational formulation, multiply the equation by a test function $v \in \mathrm{H}^1_{0,D}(\Omega)$ and integrate over Ω:

$$\int_\Omega \Big(c_0 uv + \frac{1}{2}\operatorname{div}(\mathbf{a}u)v + \frac{1}{2}\operatorname{div}(\mathbf{a}u)v - \operatorname{div}(\nu\nabla u)\, v \Big) = \int_\Omega fv.$$

For the first divergence term use the identity $\operatorname{div}(\mathbf{a}u) = \operatorname{div}(\mathbf{a})u + \mathbf{a}\cdot\nabla u$, while for the second integrate by parts:

$$\int_\Omega \frac{1}{2}\operatorname{div}(\mathbf{a}u)v = \int_\Omega -\frac{1}{2}u\,\mathbf{a}\cdot\nabla v + \int_{\partial\Omega}\frac{1}{2}\mathbf{a}\cdot\mathbf{n}\,uv,$$

and, also by integration by parts,

$$\int_\Omega -\operatorname{div}(\nu\nabla u)\, v = \int_\Omega \nu\nabla u\cdot\nabla v - \int_{\partial\Omega}\nu\frac{\partial u}{\partial n}v.$$

Therefore, imposing the boundary conditions, the variational formulation is: find $u \in \mathrm{H}^1_{0,D}(\Omega)$ such that

$$a(u,v) = F(v), \quad \text{for all } v \in \mathrm{H}^1_{0,D}(\Omega),$$

where a is a non-symmetric bilinear form defined as

$$a(u,v) = \int_\Omega \Big(\tilde{c}uv + \frac{1}{2}\mathbf{a}\cdot\nabla u\, v - \frac{1}{2}u\,\mathbf{a}\cdot\nabla v + \nu\nabla u\cdot\nabla v \Big) + \int_{\Gamma_R}\alpha uv.$$

and

$$F(v) \coloneqq \int_\Omega fv + \int_{\Gamma_R} gv.$$

Define the weighted scalar product and norm

$$(u,v)_{1,c} \coloneqq \int_\Omega \Big(\tilde{c}uv + \nu\nabla u\cdot\nabla v \Big), \qquad \|u\|_{1,c} \coloneqq (u,u)^{1/2}_{1,c}.$$

On each subdomain Ω_j we consider the local problem with bilinear form

$$a_j(u,v) \coloneqq \int_{\Omega_j} \Big(\tilde{c}uv + \frac{1}{2}\mathbf{a}\cdot\nabla u\, v - \frac{1}{2}u\,\mathbf{a}\cdot\nabla v + \nu\nabla u\cdot\nabla v \Big) + \int_{\partial\Omega_j\setminus\Gamma_D}\alpha uv,$$

where we impose absorbing transmission conditions on the subdomain interface $\partial\Omega_j \setminus \partial\Omega$: for instance, we can choose a zeroth-order Taylor approximation of transparent conditions given by

$$\alpha = \sqrt{(\mathbf{a}\cdot\mathbf{n})^2 + 4c_0\nu}/2 \tag{4.3}$$

(see e.g. [22] and the references therein). We define the local weighted scalar product and norm

$$(u,v)_{1,c,\Omega_j} \coloneqq \int_{\Omega_j} \Big(\tilde{c}uv + \nu\nabla u\cdot\nabla v \Big), \qquad \|u\|_{1,c,\Omega_j} \coloneqq (u,u)^{1/2}_{1,c,\Omega_j},$$

which would correspond to Neumann-type boundary conditions on $\partial\Omega_j$. Set

$$\tilde{c}_{+,j} \coloneqq \|\tilde{c}\|_{\mathrm{L}^\infty(\Omega_j)}, \quad \tilde{c}_{-,j} \coloneqq \|\tilde{c}^{-1}\|^{-1}_{\mathrm{L}^\infty(\Omega_j)}, \quad \text{so } \tilde{c}_{-,j} \le \tilde{c}(\mathbf{x}) \le \tilde{c}_{+,j} \text{ a.e. in } \Omega_j,$$
$$\nu_{+,j} \coloneqq \|\nu\|_{\mathrm{L}^\infty(\Omega_j)}, \quad \nu_{-,j} \coloneqq \|\nu^{-1}\|^{-1}_{\mathrm{L}^\infty(\Omega_j)}, \quad \text{so } \nu_{-,j} \le \nu(\mathbf{x}) \le \nu_{+,j} \text{ a.e. in } \Omega_j.$$

Remark 4.1. For $u, v \in \mathrm{H}^1(\Omega)$, if u or v are supported in $\overline{\Omega_j}$ and thus vanish on $\partial\Omega_j \setminus \partial\Omega$, then

$$a(u,v) = a_j(u,v), \quad \text{and} \quad (u,v)_{1,c} = (u,v)_{1,c,\Omega_j}.$$

For the finite element discretization, let $\mathcal{T}^h$ be a family of conforming simplicial meshes of Ω that are h-uniformly shape regular as the mesh diameter h tends to zero. We consider finite elements of order r

$$\mathcal{V}^h = \{\, v_h \in C^0(\overline{\Omega}), v_h|_\tau \in \mathbb{P}_{r-1}(\tau)\, \forall\, \tau \in \mathcal{T}^h, v_h|_{\Gamma_D} = 0 \,\} \subset \mathrm{H}^1_{0,D}(\Omega).$$

Consider nodal basis functions $\varphi_i, i = 1, \ldots, n$ (for example Lagrange basis functions), in duality with the degrees of freedom associated with nodes $\mathbf{x}_j, j = 1, \ldots, n$, that is $\varphi_i(\mathbf{x}_j) = \delta_{ij}$. Thus we can define the standard nodal Lagrange interpolation operator $\Pi^h v = \sum_{i=1}^n v(\mathbf{x}_i)\varphi_i$. Assume that $\mathcal{V}^h$ satisfies the standard interpolation error estimate (see e.g. [10, §3.1]): for $\tau \in \mathcal{T}^h$, provided $v \in \mathrm{H}^r(\tau)$

$$\|(I - \Pi^h)v\|_{\mathrm{L}^2(\tau)} + h|(I - \Pi^h)v|_{\mathrm{H}^1(\tau)} \leq C_\Pi h^r |v|_{\mathrm{H}^r(\tau)}. \tag{4.4}$$

Assume that the subdomains Ω_j are polyhedra with characteristic length scale H_{sub}, which means

Definition 4.2 (Characteristic length scale)**.** A domain has characteristic length scale L if its diameter $\sim L$, its surface area $\sim L^{d-1}$, and its volume $\sim L^d$, where $\sim$ means uniformly bounded from below and above.

For each $j = 1, \ldots, N$, denote by $\mathcal{V}_j^h$ the space of functions in $\mathcal{V}^h$ restricted to $\overline{\Omega}_j$. So, A, F_Ω, B_j, F_{Ω_j} are defined as the matrices arising, respectively, from the finite element discretization of a, $(\cdot,\cdot)_{1,c}$ on $\mathcal{V}^h$, and a_j, $(\cdot,\cdot)_{1,c,\Omega_j}$ on $\mathcal{V}_j^h$: for $v_h, w_h \in \mathcal{V}^h$ with vectors of degrees of freedom $\mathbf{V}, \mathbf{W} \in \mathbb{R}^n$, and for $v_h^j, w_h^j \in \mathcal{V}_j^h$ with vectors of degrees of freedom $\mathbf{V}^j, \mathbf{W}^j \in \mathbb{R}^{n_j}$

$$a(v_h, w_h) = (A\mathbf{V}, \mathbf{W}), \qquad a_j(v_h^j, w_h^j) = (B_j\mathbf{V}^j, \mathbf{W}^j), \tag{4.5}$$

$$(v_h, w_h)_{1,c} = (F_\Omega\mathbf{V}, \mathbf{W}), \qquad (v_h^j, w_h^j)_{1,c,\Omega_j} = (F_{\Omega_j}\mathbf{V}^j, \mathbf{W}^j). \tag{4.6}$$

Consider partition of unity functions χ_j, $j = 1, \ldots, N$, such that $\sum_{j=1}^N \chi_j = 1$ in $\overline{\Omega}$, and $\mathrm{supp}(\chi_j) \subset \Omega_j$, so in particular they are zero on $\partial\Omega_j \setminus \partial\Omega$. Assume that

$$\|\partial_{\mathbf{x}}^\beta \chi_j\|_{\infty,\tau} \leq C_{\mathrm{dPU}} \frac{1}{\delta^{|\beta|}} \quad \text{for all } \tau \in \mathcal{T}_h \text{ and multi-index } \beta \text{ with } |\beta| \leq r, \tag{4.7}$$

where δ is the size of the overlap between subdomains, and C_{dPU} is required to be independent of the simplex τ and of the derivative multi-index β. The diagonal matrices D_j are constructed by interpolation of the functions χ_j, so the vector of degrees of freedom of $\Pi^h(\chi_j v_h)$ is $D_j R_j \mathbf{V}$.

Next we need to introduce a technical ingredient, namely so-called multiplicative trace inequalities. Such estimates can be found e.g. in [20].

Lemma 4.3 (Multiplicative trace inequality, [20, last eq. on page 41])**.** *For any bounded Lipschitz open subset $\omega \subset \mathbb{R}^d$ there exists $C_{\mathrm{tr}}(\omega) > 0$ such that, for all $u \in \mathrm{H}^1(\omega)$, we have $\|u\|^2_{\mathrm{L}^2(\partial\omega)} \leq C_{\mathrm{tr}}(\omega)(\|u\|_{\mathrm{L}^2(\omega)}\|\nabla u\|_{\mathrm{L}^2(\omega)} + \|u\|^2_{\mathrm{L}^2(\omega)}/\mathrm{diam}(\omega))$.*

Although the constant $C_{\mathrm{tr}}(\omega)$ above does a priori depend on the shape of ω, it does not depend on its diameter (it is invariant under homothety). In the sequel we shall assume that there exists a *fixed* constant $C_{\mathrm{tr}} > 0$ such that we have $C_{\mathrm{tr}}(\Omega_j) < C_{\mathrm{tr}}$. This holds for example if the subdomains are assumed to be uniformly star-shaped i.e. there exists a fixed constant $\mu > 0$ such that, for each j there exists $\boldsymbol{x}_{\Omega_j} \in \Omega_j$ satisfying

$$\begin{aligned} &\forall \boldsymbol{x} \in \partial\Omega_j,\ [\boldsymbol{x}, \boldsymbol{x}_{\Omega_j}] \subset \overline{\Omega}_j \quad \text{and} \\ &\boldsymbol{n}_j(\boldsymbol{x}) \cdot (\boldsymbol{x} - \boldsymbol{x}_{\Omega_j}) \geq \mu |\boldsymbol{x} - \boldsymbol{x}_{\Omega_j}| \end{aligned} \tag{4.8}$$

Assumption 4.4. The multiplicative trace estimates of Lemma 4.3 hold uniformly for all subdomains.

This assumption allows to derive uniform upper bounds for the continuity modulus of the bilinear forms $a(\ ,\ )$ and $a_j(\ ,\ )$.

Lemma 4.5 (Continuity of the bilinear forms a and a_j)**.** *Assume that Ω has characteristic length scale L in the sense of Definition 4.2. Then for all $u, v \in \mathrm{H}^1(\Omega)$*

$$a(u,v) \leq C_{\mathrm{cont}} \|u\|_{1,c} \|v\|_{1,c},$$

where

$$C_{\mathrm{cont}} = \frac{\tilde{c}_+}{\tilde{c}_-}\frac{\nu_+}{\nu_-} + \frac{1}{2}\frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega)}}{\sqrt{\nu_- \tilde{c}_-}} + \frac{\|\alpha\|_{\mathrm{L}^\infty(\Omega)} C_{\mathrm{tr}}}{\sqrt{\tilde{c}_-}} \left(\frac{1}{L\sqrt{\tilde{c}_-}} + \frac{1}{2\sqrt{\nu_-}} \right).$$

Similarly for all $u, v \in \mathrm{H}^1(\Omega_j)$

$$a_j(u,v) \leq C_{\mathrm{cont},j} \|u\|_{1,c,\Omega_j} \|v\|_{1,c,\Omega_j}, \tag{4.9}$$

where

$$C_{\mathrm{cont},j} = \frac{\tilde{c}_{+,j}}{\tilde{c}_{-,j}}\frac{\nu_{+,j}}{\nu_{-,j}} + \frac{1}{2}\frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\sqrt{\nu_{-,j} \tilde{c}_{-,j}}} + \frac{\|\alpha\|_{\mathrm{L}^\infty(\Omega_j)} C_{\mathrm{tr}}}{\sqrt{\tilde{c}_{-,j}}} \left(\frac{1}{H_{\mathrm{sub}}\sqrt{\tilde{c}_{-,j}}} + \frac{1}{2\sqrt{\nu_{-,j}}} \right). \tag{4.10}$$

Proof. By Cauchy-Schwarz inequality

$$\begin{aligned} a(u,v) \leq\ & \tilde{c}_+ \|u\|_{\mathrm{L}^2(\Omega)} \|v\|_{\mathrm{L}^2(\Omega)} + \nu_+ \|\nabla u\|_{\mathrm{L}^2(\Omega)} \|\nabla v\|_{\mathrm{L}^2(\Omega)} \\ &+ \frac{1}{2}\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega)} \left(\|\nabla u\|_{\mathrm{L}^2(\Omega)} \|v\|_{\mathrm{L}^2(\Omega)} + \|u\|_{\mathrm{L}^2(\Omega)} \|\nabla v\|_{\mathrm{L}^2(\Omega)} \right) \\ &+ \|\alpha\|_{\mathrm{L}^\infty(\Omega)} \|u\|_{\mathrm{L}^2(\Gamma_R)} \|v\|_{\mathrm{L}^2(\Gamma_R)}. \end{aligned}$$

First, using the Cauchy-Schwarz inequality with respect to the Euclidean inner product in $\mathbb{R}^2$ and $1 \leq (\tilde{c}_+/\tilde{c}_-)$, $1 \leq (\nu_+/\nu_-)$, we get

$$\begin{aligned} &\tilde{c}_+ \|u\|_{\mathrm{L}^2(\Omega)} \|v\|_{\mathrm{L}^2(\Omega)} + \nu_+ \|\nabla u\|_{\mathrm{L}^2(\Omega)} \|\nabla v\|_{\mathrm{L}^2(\Omega)} \\ &= \begin{pmatrix} \frac{\tilde{c}_+}{\tilde{c}_-}\sqrt{\tilde{c}_-}\|u\|_{\mathrm{L}^2(\Omega)} & \frac{\nu_+}{\nu_-}\sqrt{\nu_-}\|\nabla u\|_{\mathrm{L}^2(\Omega)} \end{pmatrix} \begin{pmatrix} \sqrt{\tilde{c}_-}\|v\|_{\mathrm{L}^2(\Omega)} \\ \sqrt{\nu_-}\|\nabla v\|_{\mathrm{L}^2(\Omega)} \end{pmatrix} \\ &\leq \frac{\tilde{c}_+}{\tilde{c}_-}\frac{\nu_+}{\nu_-} \left(\tilde{c}_- \|u\|^2_{\mathrm{L}^2(\Omega)} + \nu_- \|\nabla u\|^2_{\mathrm{L}^2(\Omega)} \right)^{1/2} \left(\tilde{c}_- \|v\|^2_{\mathrm{L}^2(\Omega)} + \nu_- \|\nabla v\|^2_{\mathrm{L}^2(\Omega)} \right)^{1/2} \\ &\leq \frac{\tilde{c}_+}{\tilde{c}_-}\frac{\nu_+}{\nu_-} \|u\|_{1,c} \|v\|_{1,c}. \end{aligned}$$

Second

$$\begin{aligned}
&\|\nabla u\|_{\mathrm{L}^2(\Omega)}\|v\|_{\mathrm{L}^2(\Omega)} + \|u\|_{\mathrm{L}^2(\Omega)}\|\nabla v\|_{\mathrm{L}^2(\Omega)} \\
&= \frac{1}{\sqrt{\nu_-\tilde{c}_-}} \left(\sqrt{\nu_-}\|\nabla u\|_{\mathrm{L}^2(\Omega)}\ \sqrt{\tilde{c}_-}\|u\|_{\mathrm{L}^2(\Omega)}\right) \begin{pmatrix} \sqrt{\tilde{c}_-}\|v\|_{\mathrm{L}^2(\Omega)} \\ \sqrt{\nu_-}\|\nabla v\|_{\mathrm{L}^2(\Omega)} \end{pmatrix} \\
&\leq \frac{1}{\sqrt{\nu_-\tilde{c}_-}} \left(\tilde{c}_-\|u\|^2_{\mathrm{L}^2(\Omega)} + \nu_-\|\nabla u\|^2_{\mathrm{L}^2(\Omega)}\right)^{1/2} \left(\tilde{c}_-\|v\|^2_{\mathrm{L}^2(\Omega)} + \nu_-\|\nabla v\|^2_{\mathrm{L}^2(\Omega)}\right)^{1/2} \\
&\leq \frac{1}{\sqrt{\nu_-\tilde{c}_-}}\|u\|_{1,c}\|v\|_{1,c}.
\end{aligned}$$

Third, for the boundary term, using the multiplicative trace inequality recalled in Lemma 4.3 and using also the inequality $ab \leq (a^2+b^2)/2$ valid for all $a, b > 0$, we have

$$\begin{aligned}
&\|u\|_{\mathrm{L}^2(\Gamma_R)} \\
&\leq \sqrt{C_{\mathrm{tr}}}\frac{1}{\sqrt[4]{\tilde{c}_-}} \left(\frac{1}{L\sqrt{\tilde{c}_-}}\tilde{c}_-\|u\|^2_{\mathrm{L}^2(\Omega)} + \frac{1}{\sqrt{\nu_-}}\sqrt{\nu_-}\|\nabla u\|_{\mathrm{L}^2(\Omega)}\sqrt{\tilde{c}_-}\|u\|_{\mathrm{L}^2(\Omega)}\right)^{1/2} \\
&\leq \sqrt{C_{\mathrm{tr}}}\frac{1}{\sqrt[4]{\tilde{c}_-}} \left(\frac{1}{L\sqrt{\tilde{c}_-}}\|u\|^2_{1,c} + \frac{1}{2\sqrt{\nu_-}}\|u\|^2_{1,c}\right)^{1/2} \\
&= \sqrt{C_{\mathrm{tr}}}\frac{1}{\sqrt[4]{\tilde{c}_-}} \left(\frac{1}{L\sqrt{\tilde{c}_-}} + \frac{1}{2\sqrt{\nu_-}}\right)^{1/2} \|u\|_{1,c}
\end{aligned}$$

and

$$\|\alpha\|_{\mathrm{L}^\infty(\Omega)}\|u\|_{\mathrm{L}^2(\Gamma_R)}\|v\|_{\mathrm{L}^2(\Gamma_R)} \leq \|\alpha\|_{\mathrm{L}^\infty(\Omega)}C_{\mathrm{tr}}\frac{1}{\sqrt{\tilde{c}_-}}\left(\frac{1}{L\sqrt{\tilde{c}_-}}+\frac{1}{2\sqrt{\nu_-}}\right)\|u\|_{1,c}\|v\|_{1,c}.$$

In conclusion

$$a(u,v) \leq \left(\frac{\tilde{c}_+}{\tilde{c}_-}\frac{\nu_+}{\nu_-} + \frac{1}{2}\frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega)}}{\sqrt{\nu_-\tilde{c}_-}} + \frac{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}C_{\mathrm{tr}}}{\sqrt{\tilde{c}_-}}\left(\frac{1}{L\sqrt{\tilde{c}_-}} + \frac{1}{2\sqrt{\nu_-}}\right)\right)\|u\|_{1,c}\|v\|_{1,c}.$$

Finally, note that the local bilinear form a_j has the same form as the bilinear form a, so the analogous inequality holds (with $L = H_{\mathrm{sub}}$). □

Lemma 4.6 (Coercivity of the bilinear forms a and a_j)**.** *We have*

$$a(v,v) \geq \|v\|^2_{1,c} \quad \textit{for all } v \in \mathrm{H}^1(\Omega), \tag{4.11}$$

$$a_j(v,v) \geq \|v\|^2_{1,c,\Omega_j} \quad \textit{for all } v \in \mathrm{H}^1(\Omega_j). \tag{4.12}$$

Proof. Note that

$$a(v,v) = \int_\Omega \left(\tilde{c}v^2 + \nu|\nabla v|^2\right) + \int_{\Gamma_R} \alpha v^2,$$

and

$$a_j(v,v) = \int_{\Omega_j} \left(\tilde{c}v^2 + \nu|\nabla v|^2\right) + \int_{\partial\Omega_j\setminus\Gamma_D} \alpha v^2,$$

because the anti-symmetric terms cancel out. Thus properties (4.11)-(4.12) follow. □

Note that the good constant in the coercivity estimates is a result of careful choices made in the derivation of the bilinear forms (see the beginning of section 4), such as the handling of the $\mathrm{div}(\mathbf{a}u)v$ term (split into two parts with different treatments) and the definition of suitable Robin-type boundary conditions.

4.1 Estimates for the assumptions of Theorem 3.1

Now we prove, for the heterogeneous reaction-convection-diffusion problem (4.1), the equalities and inequalities that have been identified in Theorem 3.1 as the assumptions for the convergence analysis. In the proofs we do not make any assumption on the regime of the physical coefficients of the equation nor of the numerical parameters. Note that to prove the stability assumption (3.6) for this problem we have considered the case of a coercive bilinear form (which is non-symmetric, though), see Lemma 4.11. However, in general the problem does not need to be positive definite for Theorem 3.1 to be valid.

In what follows, we prove equalities and estimates in the continuous setting, which can be translated into results in the discrete setting recalling relations (4.5) between the continuous and discrete bilinear forms, relations (4.6) between the continuous and discrete inner products (hence between the norms), and the fact that the vector of degrees of freedom of $\Pi^h(\chi_j v_h)$ is $D_j R_j \mathbf{V}$.

First of all, note that the partition of unity, the global and local bilinear forms and norms fit the typical framework identified in §3.1: the entries corresponding to $\partial\Omega_j \setminus \partial\Omega$ of the partition of unity matrix D_j are zero; the local bilinear form is like the global bilinear form but with the integrals on Ω_j instead of Ω and with an additional boundary integral on $\partial\Omega_j \setminus \partial\Omega$; the local norm can be obtained from the global norm just by replacing Ω with Ω_j in the integration domain. Therefore it is not surprising that assumptions (3.1), (3.8), (3.2), (3.3) are verified. As a more precise illustration of the general remarks in §3.1, we first provide the detailed proof of assumptions (3.1) and (3.8), which is essentially based on Remark 4.1:

Lemma 4.7. *For all global vectors of degrees of freedom $\mathbf{U} \in \mathbb{R}^n$ and local vectors of degrees of freedom $\mathbf{V}^j \in \mathbb{R}^{n_j}$ in Ω_j, $j = 1, \dots, N$, we have*

$$(D_j R_j A\mathbf{U}, \mathbf{V}^j) = (D_j B_j R_j \mathbf{U}, \mathbf{V}^j),$$
$$(D_j R_j F_\Omega \mathbf{U}, \mathbf{V}^j) = (D_j F_{\Omega_j} R_j \mathbf{U}, \mathbf{V}^j).$$

Proof. Since the partition of unity matrices D_j are diagonal, hence symmetric, and the restriction matrices R_j satisfy $(\mathbf{V}, R_j^T \mathbf{W}^j) = (R_j \mathbf{V}, \mathbf{W}^j)$ and $R_j R_j^T \mathbf{V}^j = \mathbf{V}^j$, we can write

$$(D_j R_j A\mathbf{U}, \mathbf{V}^j) = (A\mathbf{U}, R_j^T D_j \mathbf{V}^j) = (A\mathbf{U}, R_j^T D_j R_j R_j^T \mathbf{V}^j).$$

Now, call $\widetilde{\mathbf{V}}^j := R_j^T \mathbf{V}^j$ and $\widetilde{v}_j \in \mathcal{V}^h$ the function with degrees of freedom given by $\widetilde{\mathbf{V}}^j$, so $D_j R_j R_j^T \mathbf{V}^j$ is the local vector of degrees of freedom of $\Pi^h(\chi_j \widetilde{v}_j)$, and $R_j^T D_j R_j R_j^T \mathbf{V}^j$ is the global vector of degrees of freedom of $\Pi^h(\chi_j \widetilde{v}_j)$. Call $u \in \mathcal{V}^h$ the function with degrees of freedom given by $\mathbf{U}$. Therefore

$$(A\mathbf{U}, R_j^T D_j R_j R_j^T \mathbf{V}^j) = a(u, \Pi^h(\chi_j \widetilde{v}_j)).$$

Moreover, observe that $\chi_j \widetilde{v}_j$ is supported in Ω_j and vanishes on $\partial\Omega_j \setminus \partial\Omega$, thus the same is true for its interpolant $\Pi^h(\chi_j \widetilde{v}_j)$, and by applying Remark 4.1 we obtain

$$a(u, \Pi^h(\chi_j \widetilde{v}_j)) = a_j(u, \Pi^h(\chi_j \widetilde{v}_j)).$$

Finally

$$a_j(u, \Pi^h(\chi_j \widetilde{v}_j)) = (B_j R_j \mathbf{U}, D_j R_j R_j^T \mathbf{V}^j) = (B_j R_j \mathbf{U}, D_j \mathbf{V}^j) = (D_j B_j R_j \mathbf{U}, \mathbf{V}^j).$$

The proof of $(D_j R_j F_\Omega \mathbf{U}, \mathbf{V}^j) = (D_j F_{\Omega_j} R_j \mathbf{U}, \mathbf{V}^j)$ proceeds in the same way. □

Now we prove that assumptions (3.2) and (3.3) are indeed verified with the geometric constants Λ_0 and Λ_1 defined in §3.1:

Lemma 4.8 (Continuous version of assumption (3.2))**.** *For all $w^j \in \mathrm{H}^1(\Omega_j)$, $j = 1, \ldots, N$, denoting by $\widetilde{w}^j$ their extensions by zero to Ω, we have*

$$\left\| \sum_{j=1}^{N} \widetilde{w}^j \right\|_{1,c}^2 \leq \Lambda_0 \sum_{j=1}^{N} \|w^j\|_{1,c,\Omega_j}^2,$$

where Λ_0 is the maximum number of neighboring subdomains:

$$\Lambda_0 = \max_{j=1,\ldots,N} \#\Lambda(j), \quad \text{where } \Lambda(j) \coloneqq \{\, i \mid \Omega_j \cap \Omega_i \neq \emptyset \,\}.$$

Proof. By applying several times the Cauchy-Schwarz inequality (first for the scalar product $(\ ,\ )_{1,c}$ and then twice for the dot product of the Euclidean space), and by Remark 4.1, we get

$$\begin{aligned}
\left\| \sum_{j=1}^{N} \widetilde{w}^j \right\|_{1,c}^2 &= \left(\sum_{j=1}^{N} \widetilde{w}^j, \sum_{j'=1}^{N} \widetilde{w}^{j'} \right)_{1,c} = \sum_{j=1}^{N} \sum_{j' \in \Lambda(j)} (\widetilde{w}^j, \widetilde{w}^{j'})_{1,c} \\
&\leq \sum_{j=1}^{N} \left(\|w^j\|_{1,c,\Omega_j} \sum_{j' \in \Lambda(j)} \|w^{j'}\|_{1,c,\Omega_{j'}} \right) \\
&\leq \left(\sum_{j=1}^{N} \|w^j\|_{1,c,\Omega_j}^2 \right)^{1/2} \left(\sum_{j=1}^{N} \Big(\sum_{j' \in \Lambda(j)} \|w^{j'}\|_{1,c,\Omega_{j'}} \Big)^2 \right)^{1/2} \\
&\leq \left(\sum_{j=1}^{N} \|w^j\|_{1,c,\Omega_j}^2 \right)^{1/2} \left(\sum_{j=1}^{N} \Big(\sum_{j' \in \Lambda(j)} 1^2 \sum_{j' \in \Lambda(j)} \|w^{j'}\|_{1,c,\Omega_{j'}}^2 \Big) \right)^{1/2}.
\end{aligned}$$

Now we have

$$\begin{aligned}
&\sum_{j=1}^{N} \Big(\sum_{j' \in \Lambda(j)} 1^2 \sum_{j' \in \Lambda(j)} \|w^{j'}\|_{1,c,\Omega_{j'}}^2 \Big) = \sum_{j=1}^{N} \Big(\#\Lambda(j) \sum_{j' \in \Lambda(j)} \|w^{j'}\|_{1,c,\Omega_{j'}}^2 \Big) \\
&\leq \Lambda_0 \sum_{j=1}^{N} \sum_{j' \in \Lambda(j)} \|w^{j'}\|_{1,c,\Omega_{j'}}^2 = \Lambda_0 \sum_{j'=1}^{N} \#\Lambda(j') \|w^{j'}\|_{1,c,\Omega_{j'}}^2 \leq \Lambda_0^2 \sum_{j'=1}^{N} \|w^{j'}\|_{1,c,\Omega_{j'}}^2.
\end{aligned}$$

Therefore in summary

$$\left\| \sum_{j=1}^{N} \widetilde{w}^j \right\|_{1,c}^2 \leq \left(\sum_{j=1}^{N} \|w^j\|_{1,c,\Omega_j}^2 \right)^{1/2} \left(\Lambda_0^2 \sum_{j'=1}^{N} \|w^{j'}\|_{1,c,\Omega_{j'}}^2 \right)^{1/2} = \Lambda_0 \sum_{j=1}^{N} \|w^j\|_{1,c,\Omega_j}^2.$$

□

Lemma 4.9 (Continuous version of assumption (3.3))**.** *For all $v \in \mathrm{H}^1(\Omega)$*

$$\sum_{j=1}^{N} \|v|_{\Omega_j}\|_{1,c,\Omega_j}^2 \leq \Lambda_1 \|v\|_{1,c}^2,$$

where Λ_1 is the maximal multiplicity of the subdomain intersection:

$$\Lambda_1 = \max \{\, m \mid \exists\, j_1 \neq \ldots \neq j_m \text{ such that } \mathrm{meas}(\Omega_{j_1} \cap \cdots \cap \Omega_{j_m}) \neq 0 \,\}.$$

Proof. The result is immediate by the definition of the norms and of Λ_1:

$$\sum_{j=1}^{N} \|v|_{\Omega_j}\|_{1,c,\Omega_j}^2 = \sum_{j=1}^{N} \int_{\Omega_j} \left(\tilde{c} (v|_{\Omega_j})^2 + \nu \left| \nabla (v|_{\Omega_j}) \right|^2 \right) \leq \Lambda_1 \int_{\Omega} \left(\tilde{c} v^2 + \nu |\nabla v|^2 \right).$$

□

For the remaining assumptions, for the translation from the continuous to the discrete setting we also need to consider the error in interpolation of $\chi_j v_h$, studied in the following lemma.

Lemma 4.10. *For any $j = 1, \dots, N$, let $v_h \in \mathcal{V}_j^h$. Then*

$$\|(\mathrm{I} - \Pi^h)(\chi_j v_h)\|_{1,c,\Omega_j} \leq C_{\mathrm{err},j} \|v_h\|_{1,c,\Omega_j}, \tag{4.13}$$

where

$$C_{\mathrm{err},j} = C_\Pi \, c(r,d) \, C_{\mathrm{dPU}} \sqrt{C_{\mathrm{inv}}} \left(\sqrt{\frac{\nu_{+,j}}{\nu_{-,j}}} + \sqrt{\frac{\tilde{c}_{+,j}}{\nu_{-,j}}} h \right) \frac{h}{\delta}, \tag{4.14}$$

and C_Π appears in (4.4)*, C_{dPU} in* (4.7)*, C_{inv} is a standard inverse inequality constant (see the proof for more details), and $c(r,d) = \max_{|\gamma|=r} \sum_{\beta \,|\, 0<\beta\leq\gamma} \binom{\gamma}{\beta}$.*

Proof. For each simplex $\tau \in \mathcal{T}^h$, $\tau \subset \Omega_j$, from (4.4) we have

$$\|(I - \Pi^h)(\chi_j v_h)\|_{\mathrm{L}^2(\tau)} + h |(I - \Pi^h)(\chi_j v_h)|_{\mathrm{H}^1(\tau)} \leq C_\Pi h^r |\chi_j v_h|_{\mathrm{H}^r(\tau)}. \tag{4.15}$$

In order to estimate $|\chi_j v_h|_{\mathrm{H}^r(\tau)}$, let $\gamma \in \mathbb{N}^d$ be a multi-index of order r, i.e. $|\gamma| = r$. By the multivariate Leibniz rule and observing that $\partial_{\mathbf{x}}^\gamma v_h = 0$ since $v_h|_\tau$ is a polynomial of degree $r-1$, we have

$$\partial_{\mathbf{x}}^\gamma (\chi_j v_h) = \sum_{\beta \,|\, 0\leq\beta\leq\gamma} \binom{\gamma}{\beta} (\partial_{\mathbf{x}}^\beta \chi_j)(\partial_{\mathbf{x}}^{\gamma-\beta} v_h) = \sum_{\beta \,|\, 0<\beta\leq\gamma} \binom{\gamma}{\beta} (\partial_{\mathbf{x}}^\beta \chi_j)(\partial_{\mathbf{x}}^{\gamma-\beta} v_h),$$

(note that in the last equality the multi-index $0 = (0, \dots, 0) \in \mathbb{N}^d$ is excluded). Then, setting $c(r,d) = \max_{|\gamma|=r} \sum_{\beta \,|\, 0<\beta\leq\gamma} \binom{\gamma}{\beta}$, and using (4.7), we get

$$\|\partial_{\mathbf{x}}^\gamma (\chi_j v_h)\|_{\mathrm{L}^2(\tau)} \leq c(r,d) C_{\mathrm{dPU}} \max_{\beta \,|\, 0<\beta\leq\gamma} \delta^{-|\beta|} |v_h|_{\mathrm{H}^{r-|\beta|}(\tau)}. \tag{4.16}$$

Now we want to estimate $|v_h|_{\mathrm{H}^{r-|\beta|}(\tau)}$ using an inverse inequality, but in terms of the weighted norm $\| \, \|_{1,c,\tau}$ instead of the standard $\| \, \|_{H^1(\tau)}$ norm, and without making regime assumptions on the coefficients of the equation. First of all, note that, performing the change of variables $\mathbf{y} = \sqrt{\frac{\tilde{c}_{-,j}}{\nu_{-,j}}} \mathbf{x}$ and setting

$$\tau_c \coloneqq \left\{ \sqrt{\frac{\tilde{c}_{-,j}}{\nu_{-,j}}} \mathbf{x} \,\middle|\, \mathbf{x} \in \tau \right\}, \quad \phi_c(v_h)(\mathbf{y}) \coloneqq v_h(\mathbf{x}) = v_h \left(\mathbf{y} \sqrt{\frac{\nu_{-,j}}{\tilde{c}_{-,j}}} \right),$$

we can rewrite

$$\begin{aligned}\|v_h\|_{1,c,\tau}^2 &\geq \int_\tau \left(\tilde{c}_{-,j} v_h^2 + \nu_{-,j}|\nabla_{\mathbf{x}} v_h|^2\right) d\mathbf{x} \\ &= \int_{\tau_c} \left(\tilde{c}_{-,j}(\phi_c(v_h))^2 + \nu_{-,j}\frac{\tilde{c}_{-,j}}{\nu_{-,j}}|\nabla_{\mathbf{y}}\phi_c(v_h)|^2\right)\left(\sqrt{\frac{\tilde{c}_{-,j}}{\nu_{-,j}}}\right)^{-d} d\mathbf{y} \\ &= \nu_{-,j}\left(\frac{\tilde{c}_{-,j}}{\nu_{-,j}}\right)^{1-d/2} \|\phi_c(v_h)\|_{\mathrm{H}^1(\tau_c)}^2.\end{aligned} \tag{4.17}$$

Performing the same change of variables, we examine $|v_h|_{\mathrm{H}^{r-|\beta|}(\tau)}$:

$$\begin{aligned}|v_h|_{\mathrm{H}^{r-|\beta|}(\tau)}^2 &= \sum_{\xi \,|\, |\xi|=r-|\beta|} \int_\tau |\partial_{\mathbf{x}}^\xi v_h|^2 d\mathbf{x} \\ &= \sum_{\xi \,|\, |\xi|=r-|\beta|} \int_{\tau_c} \left(\frac{\tilde{c}_{-,j}}{\nu_{-,j}}\right)^{r-|\beta|} |\partial_{\mathbf{y}}^\xi \phi_c(v_h)|^2 \left(\sqrt{\frac{\tilde{c}_{-,j}}{\nu_{-,j}}}\right)^{-d} d\mathbf{y} \\ &= \left(\frac{\tilde{c}_{-,j}}{\nu_{-,j}}\right)^{r-|\beta|-d/2} |\phi_c(v_h)|_{\mathrm{H}^{r-|\beta|}(\tau_c)}^2,\end{aligned}$$

so, using a standard inverse inequality (see e.g. [10, Theorem 3.2.6]), applied with $\sqrt{\frac{\tilde{c}_{-,j}}{\nu_{-,j}}}h$ (diameter of τ_c), we get

$$\begin{aligned}|v_h|_{\mathrm{H}^{r-|\beta|}(\tau)}^2 &\leq C_{\mathrm{inv}} \left(\frac{\tilde{c}_{-,j}}{\nu_{-,j}}\right)^{r-|\beta|-d/2} \left(\sqrt{\frac{\tilde{c}_{-,j}}{\nu_{-,j}}}h\right)^{-2(r-|\beta|-1)} \|\phi_c(v_h)\|_{\mathrm{H}^1(\tau_c)}^2 \\ &= C_{\mathrm{inv}} \left(\frac{\tilde{c}_{-,j}}{\nu_{-,j}}\right)^{1-d/2} h^{-2(r-|\beta|-1)} \|\phi_c(v_h)\|_{\mathrm{H}^1(\tau_c)}^2 \\ &\leq C_{\mathrm{inv}} h^{-2(r-|\beta|-1)} \frac{1}{\nu_{-,j}} \|v_h\|_{1,c,\tau}^2,\end{aligned}$$

where the last inequality comes from (4.17) (reversed). Therefore (4.16) becomes:

$$\begin{aligned}\|\partial_{\mathbf{x}}^\gamma(\chi_j v_h)\|_{\mathrm{L}^2(\tau)} &\leq c(r,d) C_{\mathrm{dPU}} \sqrt{C_{\mathrm{inv}}} \max_{m=1,\dots,r} \delta^{-m} h^{-(r-m-1)} \frac{1}{\sqrt{\nu_{-,j}}} \|v_h\|_{1,c,\tau} \\ &= c(r,d) C_{\mathrm{dPU}} \sqrt{C_{\mathrm{inv}}} \delta^{-1} h^{-r+2} \frac{1}{\sqrt{\nu_{-,j}}} \|v_h\|_{1,c,\tau},\end{aligned} \tag{4.18}$$

where we have used the fact that $(h/\delta) \leq 1$, so that the maximum is attained for $m = 1$.

Finally, combining (4.15) and (4.18), and summing over all simplices $\tau \subset \Omega_j$, we obtain

$$\|(\mathrm{I} - \Pi^h)(\chi_j v_h)\|_{\mathrm{L}^2(\Omega_j)} \leq C_\Pi c(r,d) C_{\mathrm{dPU}} \sqrt{C_{\mathrm{inv}}} \frac{h^2}{\delta} \frac{1}{\sqrt{\nu_{-,j}}} \|v_h\|_{1,c,\Omega_j}, \tag{4.19}$$

$$|(\mathrm{I} - \Pi^h)(\chi_j v_h)|_{\mathrm{H}^1(\Omega_j)} \leq C_\Pi c(r,d) C_{\mathrm{dPU}} \sqrt{C_{\mathrm{inv}}} \frac{h}{\delta} \frac{1}{\sqrt{\nu_{-,j}}} \|v_h\|_{1,c,\Omega_j}. \tag{4.20}$$

Now, applying $\sqrt{a^2+b^2} \leq a+b$ with a the left-hand side of (4.19) multiplied by $\sqrt{\tilde{c}_{+,j}}$ and b the left-hand side of (4.20) multiplied by $\sqrt{\nu_{+,j}}$ in order to recover the weighted norm, we obtain

$$\|(\mathrm{I}-\Pi^h)(\chi_j v_h)\|_{1,c,\Omega_j} \leq C_\Pi c(r,d) C_{\mathrm{dPU}} \sqrt{C_{\mathrm{inv}}}\big(\sqrt{\tilde{c}_{+,j}}h+\sqrt{\nu_{+,j}}\big)\frac{h}{\delta}\frac{1}{\sqrt{\nu_{-,j}}}\|v_h\|_{1,c,\Omega_j}.$$

□

We prove now the stability bound (3.6).

Lemma 4.11. *(Stability bound for the local problems) For all $u_h^j \in \mathcal{V}_j^h$, we have*

$$\|u_h^j\|_{1,c,\Omega_j} \leq \sup_{v_h^j \in \mathcal{V}_j^h \setminus \{0\}} \left(\frac{|a_j(u_h^j, v_h^j)|}{\|v_h^j\|_{1,c,\Omega_j}} \right). \tag{4.21}$$

Therefore, recalling the relation in (4.5) *between the local continuous and discrete bilinear forms, assumption* (3.6) *is satisfied with*

$$C_{\mathrm{stab},j} = 1.$$

Proof. This is a consequence of Lemmas 4.5–4.6 and Lax-Milgram theorem (see e.g. [29, Theorem 5.14]): note that the constant in the stability bound is the reciprocal of the constant in the coercivity bound (4.12), which is 1. □

The good constant obtained in the stability estimate is a result of careful choices made in the derivation of the bilinear form, as already pointed out for the coercivity estimate (4.12).

Next, we prove estimates for assumption (3.4).

Lemma 4.12 ($C_{D,j}$ in (3.4))**.** *For all $v \in \mathrm{H}^1(\Omega_j)$*

$$\|\chi_j v\|_{1,c,\Omega_j} \leq \sqrt{2}\left(1 + C_{\mathrm{dPU}}\sqrt{\frac{\nu_{+,j}}{\tilde{c}_{-,j}}}\frac{1}{\delta}\right)\|v\|_{1,c,\Omega_j}, \tag{4.22}$$

where C_{dPU} appears in (4.7)*. Moreover, for all $v_h \in \mathcal{V}_j^h$,*

$$\|\Pi^h(\chi_j v_h)\|_{1,c,\Omega_j} \leq C_{D,j}\|v_h\|_{1,c,\Omega_j}, \tag{4.23}$$

which is the finite element expression of assumption (3.4)*, where*

$$C_{D,j} = \sqrt{2}\left(1 + C_{\mathrm{dPU}}\sqrt{\frac{\nu_{+,j}}{\tilde{c}_{-,j}}}\frac{1}{\delta}\right) + C_{\mathrm{err},j}, \tag{4.24}$$

with $C_{\mathrm{err},j}$ given by (4.14)*.*

Proof. We have

$$\|\chi_j v\|_{1,c,\Omega_j}^2 \leq \int_{\Omega_j} \tilde{c}|\chi_j v|^2 + 2\int_{\Omega_j} \nu|(\nabla\chi_j)v|^2 + 2\int_{\Omega_j} \nu|\chi_j \nabla v|^2$$

and using $|\chi_j| \leq 1$ and (4.7) we get

$$\begin{aligned}\|\chi_j v\|^2_{1,c,\Omega_j} &\leq \int_{\Omega_j} \tilde{c}|v|^2 + 2\int_{\Omega_j} \nu\, C^2_{\mathrm{dPU}} \frac{1}{\delta^2}|v|^2 + 2\int_{\Omega_j} \nu|\nabla v|^2 \\ &\leq 2\Big(1 + C^2_{\mathrm{dPU}} \frac{\nu_{+,j}}{\tilde{c}_{-,j}} \frac{1}{\delta^2}\Big)\|v\|^2_{1,c,\Omega_j}.\end{aligned}$$

Now, for the second estimate, using the triangle inequality, the newly found inequality (4.22) and (4.13), we get

$$\begin{aligned}\|\Pi^h(\chi_j v_h)\|_{1,c,\Omega_j} &\leq \|\chi_j v_h\|_{1,c,\Omega_j} + \|(\mathrm{I} - \Pi^h)(\chi_j v_h)\|_{1,c,\Omega_j} \\ &\leq \left[\sqrt{2}\left(1 + C_{\mathrm{dPU}}\sqrt{\frac{\nu_{+,j}}{\tilde{c}_{-,j}}}\frac{1}{\delta}\right) + C_{\mathrm{err},j}\right]\|v_h\|_{1,c,\Omega_j}.\end{aligned}$$

□

Next, we prove estimates for assumption (3.9), which involves a commutator between the partition of unity and the local inner product matrix.

Lemma 4.13 ($C_{DF,j}$ in (3.9))**.** *For all $v, w \in \mathrm{H}^1(\Omega_j)$*

$$|(v, \chi_j w)_{1,c,\Omega_j} - (\chi_j v, w)_{1,c,\Omega_j}| \leq C_{\mathrm{dPU}} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \frac{1}{\delta}\|v\|_{1,c,\Omega_j}\|w\|_{1,c,\Omega_j}, \quad (4.25)$$

where C_{dPU} appears in (4.7)*. Moreover, for all $v_h, w_h \in \mathcal{V}_j^h$*

$$|(v_h, \Pi^h(\chi_j w_h))_{1,c,\Omega_j} - (\Pi^h(\chi_j v_h), w_h)_{1,c,\Omega_j}| \leq C_{DF,j}\|v_h\|_{1,c,\Omega_j}\|w_h\|_{1,c,\Omega_j} \quad (4.26)$$

which is the finite element expression of assumption (3.9)*, where*

$$C_{DF,j} = C_{\mathrm{dPU}} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \frac{1}{\delta} + 2C_{\mathrm{err},j}, \quad (4.27)$$

with $C_{\mathrm{err},j}$ given by (4.14)*.*

Proof. Note that

$$\begin{aligned}&(v, \chi_j w)_{1,c,\Omega_j} - (\chi_j v, w)_{1,c,\Omega_j} \\ &= \int_{\Omega_j} \nu\nabla v \cdot (w\nabla\chi_j + \chi_j\nabla w) - \int_{\Omega_j} \nu(v\nabla\chi_j + \chi_j\nabla v)\cdot\nabla w \\ &= \int_{\Omega_j} \nu\nabla\chi_j \cdot (w\nabla v - v\nabla w).\end{aligned}$$

Then, by the Cauchy-Schwarz inequality and (4.7)

$$
\begin{aligned}
&|(v, \chi_j w)_{1,c,\Omega_j} - (\chi_j v, w)_{1,c,\Omega_j}| \\
&\leq \nu_{+,j}\, C_{\mathrm{dPU}} \frac{1}{\delta} \big(\|w\|_{\mathrm{L}^2(\Omega_j)} \|\nabla v\|_{\mathrm{L}^2(\Omega_j)} + \|v\|_{\mathrm{L}^2(\Omega_j)} \|\nabla w\|_{\mathrm{L}^2(\Omega_j)} \big) \\
&= \frac{C_{\mathrm{dPU}}}{\delta} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \Big(\sqrt{\tilde{c}_{-,j}} \|w\|_{\mathrm{L}^2(\Omega_j)} \sqrt{\nu_{-,j}} \|\nabla v\|_{\mathrm{L}^2(\Omega_j)} \\
&\qquad\qquad + \sqrt{\tilde{c}_{-,j}} \|v\|_{\mathrm{L}^2(\Omega_j)} \sqrt{\nu_{-,j}} \|\nabla w\|_{\mathrm{L}^2(\Omega_j)} \Big) \\
&= \frac{C_{\mathrm{dPU}}}{\delta} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \begin{pmatrix} \sqrt{\tilde{c}_{-,j}} \|w\|_{\mathrm{L}^2(\Omega_j)} & \sqrt{\nu_{-,j}} \|\nabla w\|_{\mathrm{L}^2(\Omega_j)} \end{pmatrix} \begin{pmatrix} \sqrt{\nu_{-,j}} \|\nabla v\|_{\mathrm{L}^2(\Omega_j)} \\ \sqrt{\tilde{c}_{-,j}} \|v\|_{\mathrm{L}^2(\Omega_j)} \end{pmatrix} \\
&\leq \frac{C_{\mathrm{dPU}}}{\delta} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \Big(\tilde{c}_{-,j} \|w\|^2_{\mathrm{L}^2(\Omega_j)} + \nu_{-,j} \|\nabla w\|^2_{\mathrm{L}^2(\Omega_j)} \Big)^{1/2} \\
&\qquad\qquad \cdot \Big(\tilde{c}_{-,j} \|v\|^2_{\mathrm{L}^2(\Omega_j)} + \nu_{-,j} \|\nabla v\|^2_{\mathrm{L}^2(\Omega_j)} \Big)^{1/2} \\
&\leq \frac{C_{\mathrm{dPU}}}{\delta} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \|v\|_{1,c,\Omega_j} \|w\|_{1,c,\Omega_j},
\end{aligned}
$$

where at the end we have used the Cauchy-Schwarz inequality with respect to the Euclidean inner product in $\mathbb{R}^2$.

For $C_{DF,j}$ we find the continuous analogue of the left-hand side in (3.9): for $\mathbf{V}^j, \mathbf{W}^j \in \mathbb{R}^{n_j}$ vectors of degrees of freedom for local functions $v_h, w_h \in \mathcal{V}_j^h$

$$
\begin{aligned}
|([D_j, F_{\Omega_j}]\mathbf{V}^j, \mathbf{W}^j)| &= |(F_{\Omega_j}\mathbf{V}^j, D_j\mathbf{W}^j) - (F_{\Omega_j} D_j \mathbf{V}^j, \mathbf{W}^j)| \\
&= |(v_h, \Pi^h(\chi_j w_h))_{1,c,\Omega_j} - (\Pi^h(\chi_j v_h), w_h)_{1,c,\Omega_j}| \\
&= |((I - \Pi^h)(\chi_j v_h), w_h)_{1,c,\Omega_j} - (v_h, (I - \Pi^h)(\chi_j w_h))_{1,c,\Omega_j} \\
&\qquad + (v_h, \chi_j w_h)_{1,c,\Omega_j} - (\chi_j v_h, w_h)_{1,c,\Omega_j}|.
\end{aligned}
$$

Now, by the Cauchy-Schwarz inequality and (4.13)

$$
|((I - \Pi^h)(\chi_j v_h), w_h)_{1,c,\Omega_j}| \leq C_{\mathrm{err},j} \|v_h\|_{1,c,\Omega_j} \|w_h\|_{1,c,\Omega_j}
$$

and similarly for $|(v_h, (I - \Pi^h)(\chi_j w_h))_{1,c,\Omega_j}|$, so, combining with (4.25), we get

$$
\begin{aligned}
|([D_j, F_{\Omega_j}]\mathbf{V}^j, \mathbf{W}^j)| &\leq \left(C_{\mathrm{dPU}} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \frac{1}{\delta} + 2C_{\mathrm{err},j} \right) \|v_h\|_{1,c,\Omega_j} \|w_h\|_{1,c,\Omega_j} \\
&= \left(C_{\mathrm{dPU}} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \frac{1}{\delta} + 2C_{\mathrm{err},j} \right) \|\mathbf{V}^j\|_{\Omega_j} \|\mathbf{W}^j\|_{\Omega_j}.
\end{aligned}
$$

□

Finally, for assumption (3.5) let us study the commutator between the partition of unity matrix and the local problem matrix.

Lemma 4.14 ($C_{DB,j}$ in (3.5))**.** *For all* $v, w \in \mathrm{H}^1(\Omega_j)$

$$|a_j(v,\chi_j w) - a_j(\chi_j v, w)| \leq C_{\mathrm{dPU}} \left(\frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} + \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}_{-,j}} \right) \frac{1}{\delta} \|v\|_{1,c,\Omega_j} \|w\|_{1,c,\Omega_j} \tag{4.28}$$

where C_{dPU} *appears in* (4.7)*. Moreover, for all* $v_h, w_h \in \mathcal{V}_j^h$

$$|a_j(v_h, \Pi^h(\chi_j w_h)) - a_j(\Pi^h(\chi_j v_h), w_h)| \leq C_{DB,j} \|v_h\|_{1,c,\Omega_j} \|w_h\|_{1,c,\Omega_j} \tag{4.29}$$

which is the finite element expression of assumption (3.5)*, where*

$$C_{DB,j} = C_{\mathrm{dPU}} \left(\frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} + \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}_{-,j}} \right) \frac{1}{\delta} + 2C_{\mathrm{cont},j} C_{\mathrm{err},j}, \tag{4.30}$$

with $C_{\mathrm{cont},j}$*,* $C_{\mathrm{err},j}$ *given by* (4.10)*,* (4.14)*.*

Proof. Note that

$$\begin{aligned}
a_j(v,\chi_j w) - a_j(\chi_j v, w) &= \frac{1}{2}\int_{\Omega_j} \chi_j w \mathbf{a}\cdot\nabla v - v\mathbf{a}\cdot(w\nabla\chi_j + \chi_j\nabla w) + \\
&- \frac{1}{2}\int_{\Omega_j} w\mathbf{a}(v\nabla\chi_j + \chi_j\nabla v) - \chi_j v\mathbf{a}\cdot\nabla w \\
&+ \int_{\Omega_j} \nu\nabla v\cdot(w\nabla\chi_j + \chi_j\nabla w) - \int_{\Omega_j} \nu(v\nabla\chi_j + \chi_j\nabla v)\cdot\nabla w \\
&= -\int_{\Omega_j} vw\,\mathbf{a}\cdot\nabla\chi_j + \int_{\Omega_j} \nu\nabla\chi_j\cdot(w\nabla v - v\nabla w).
\end{aligned}$$

By the Cauchy-Schwarz inequality and (4.7)

$$\begin{aligned}
\left| \int_{\Omega_j} vw\,\mathbf{a}\cdot\nabla\chi_j \right| &\leq C_{\mathrm{dPU}} \frac{1}{\delta} \|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)} \|v\|_{\mathrm{L}^2(\Omega_j)} \|w\|_{\mathrm{L}^2(\Omega_j)} \\
&\leq C_{\mathrm{dPU}} \frac{1}{\delta} \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}_{-,j}} \|v\|_{1,c,\Omega_j} \|w\|_{1,c,\Omega_j}.
\end{aligned}$$

Therefore, proceeding for the other term as in Lemma 4.13,

$$|a_j(v,\chi_j w) - a_j(\chi_j v, w)| \leq C_{\mathrm{dPU}} \left(\frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} + \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}_{-,j}} \right) \frac{1}{\delta} \|v\|_{1,c,\Omega_j} \|w\|_{1,c,\Omega_j}.$$

For $C_{DB,j}$ we find the continuous analogue of the left-hand side in (3.5): for $\mathbf{V}^j, \mathbf{W}^j \in \mathbb{R}^{n_j}$ vectors of degrees of freedom for local functions $v_h, w_h \in \mathcal{V}_j^h$

$$\begin{aligned}
&|([D_j, B_j]\mathbf{V}^j, \mathbf{W}^j)| = |(B_j\mathbf{V}^j, D_j\mathbf{W}^j) - (B_j D_j \mathbf{V}^j, \mathbf{W}^j)| \\
&= |a_j(v_h, \Pi^h(\chi_j w_h)) - a_j(\Pi^h(\chi_j v_h), w_h)| \\
&= |a_j((I - \Pi^h)(\chi_j v_h), w_h) - a_j(v_h, (I - \Pi^h)(\chi_j w_h)) \\
&\qquad + a_j(v_h, \chi_j w_h) - a_j(\chi_j v_h, w_h)|.
\end{aligned}$$

Now, by the continuity property (4.9) of a_j and (4.13)

$$|a_j((I-\Pi^h)(\chi_j v_h), w_h)| \leq C_{\mathrm{cont},j} C_{\mathrm{err},j} \|v_h\|_{1,c,\Omega_j} \|w_h\|_{1,c,\Omega_j}$$

and similarly for $|a_j(v_h, (I-\Pi^h)(\chi_j w_h))|$, so, combining with (4.28), we get

$$\begin{aligned}
&|([D_j, B_j]\mathbf{V}^j, \mathbf{W}^j)| \\
&\leq \left[C_{\mathrm{dPU}} \left(\frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} + \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}_{-,j}} \right) \frac{1}{\delta} + 2 C_{\mathrm{cont},j} C_{\mathrm{err},j} \right] \|v_h\|_{1,c,\Omega_j} \|w_h\|_{1,c,\Omega_j} \\
&= \left[C_{\mathrm{dPU}} \left(\frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} + \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}_{-,j}} \right) \frac{1}{\delta} + 2 C_{\mathrm{cont},j} C_{\mathrm{err},j} \right] \|\mathbf{V}^j\|_{\Omega_j} \|\mathbf{W}^j\|_{\Omega_j}.
\end{aligned}$$

□

4.2 Summary of the constants

For the heterogeneous reaction-convection-diffusion problem (4.1) we have proved that the upper and lower bounds of Theorem 3.1

$$\max_{\mathbf{V}\in\mathbb{R}^n} \frac{\|M^{-1}A\mathbf{V}\|_\Omega}{\|\mathbf{V}\|_\Omega} \leq \sqrt{\Lambda_0 \Lambda_1} \max_{j=1,\dots,N} \{ C_{D,j} (C_{\mathrm{stab},j} C_{DB,j} + C_{D,j}) \}$$

$$\begin{aligned}
\min_{\mathbf{V}\in\mathbb{R}^n} \frac{|(F_\Omega \mathbf{V}, M^{-1}A\mathbf{V})|}{\|\mathbf{V}\|_\Omega^2} \geq \frac{1}{\Lambda_0} &- \Lambda_1 \max_{j=1,\dots,N} \{ C_{D,j} C_{\mathrm{stab},j} C_{DB,j} \} \\
&- \Lambda_1 \max_{j=1,\dots,N} \{ C_{DF,j} (C_{\mathrm{stab},j} C_{DB,j} + C_{D,j}) \}
\end{aligned}$$

hold with the constants Λ_0, Λ_1 from Lemma 4.8, Lemma 4.9:

$$\Lambda_0 = \max_{j=1,\dots,N} \#\Lambda(j), \quad \text{where } \Lambda(j) = \{ j' \mid \Omega_j \cap \Omega_{j'} \neq \emptyset \}$$

$$\Lambda_1 = \max \{ m \mid \exists\, j_1 \neq \dots \neq j_m \text{ such that } \mathrm{meas}(\Omega_{j_1} \cap \dots \cap \Omega_{j_m}) \neq 0 \}$$

$C_{\mathrm{stab},j}$ from Lemma 4.11:

$$C_{\mathrm{stab},j} = 1$$

$C_{D,j}$ from (4.24):

$$C_{D,j} = \sqrt{2} \left(1 + C_{\mathrm{dPU}} \sqrt{\frac{\nu_{+,j}}{\tilde{c}_{-,j}}} \frac{1}{\delta} \right) + C_{\mathrm{err},j}$$

$C_{DF,j}$ from (4.27):

$$C_{DF,j} = C_{\mathrm{dPU}} \frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} \frac{1}{\delta} + 2 C_{\mathrm{err},j}$$

$C_{DB,j}$ from (4.30):

$$C_{DB,j} = C_{\mathrm{dPU}} \left(\frac{\nu_{+,j}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} + \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}_{-,j}} \right) \frac{1}{\delta} + 2 C_{\mathrm{cont},j} C_{\mathrm{err},j}$$

where from (4.10)

$$C_{\mathrm{cont},j} = \frac{\tilde{c}_{+,j}}{\tilde{c}_{-,j}}\frac{\nu_{+,j}}{\nu_{-,j}} + \frac{1}{2}\frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\sqrt{\tilde{c}_{-,j}\nu_{-,j}}} + \frac{\|\alpha\|_{\mathrm{L}^\infty(\Omega)}C_{\mathrm{tr}}}{\sqrt{\tilde{c}_{-,j}}}\left(\frac{1}{H_{\mathrm{sub}}\sqrt{\tilde{c}_{-,j}}} + \frac{1}{2\sqrt{\nu_{-,j}}}\right)$$

and from (4.14)

$$C_{\mathrm{err},j} = C_{\Pi}\, c(r,d)\, C_{\mathrm{dPU}}\, \sqrt{C_{\mathrm{inv}}}\left(\sqrt{\frac{\nu_{+,j}}{\nu_{-,j}}} + \sqrt{\frac{\tilde{c}_{+,j}}{\nu_{-,j}}}h\right)\frac{h}{\delta},$$

and C_{tr} appears in Lemma 4.3, C_{Π} in (4.4), C_{dPU} in (4.7), and C_{inv} is a standard inverse inequality constant (see the proof of Lemma 4.10 for more details), and $c(r,d) = \max_{|\gamma|=r}\sum_{\beta\,|\,0<\beta\leq\gamma}\binom{\gamma}{\beta}$.

These estimates can be then specialized for particular regimes of the physical coefficients of the equation or of the numerical parameters. Note that the lower bound is interesting only if the positive term dominates the negative ones in the considered regime. In particular, if the overlap δ is sufficiently generous, both negative terms can be made arbitrarily small. So we have proved for the SORAS algorithm that a larger overlap helps the convergence of the domain decomposition preconditioner, as expected.

For instance, if the equation in (4.1) derives from a backward Euler scheme for solving the time-dependent convection-diffusion problem, we would have $\tilde{c} = 1/\Delta t$, where Δt is the time step of the scheme. Now, note that the constants $C_{D,j}, C_{DB,j}, C_{DF,j}$ appearing in the negative terms contain the adimensional quantities

$$\sqrt{\frac{\nu}{\tilde{c}}}\frac{1}{\delta}, \qquad \frac{\|\mathbf{a}\|_{\mathrm{L}^\infty(\Omega_j)}}{\tilde{c}}\frac{1}{\delta},$$

(where we have considered the homogeneous case for simplicity). Hence for these quantities to be small, the overlap δ should be asymptotically bigger than the square root of the diffusion area covered in a time step, and than the convection distance covered in a time step. Therefore, on the one hand when the diffusion coefficient or the convection velocity grow, the overlap size should be increased; on the other hand if the time discretization step shrinks, one could take a smaller overlap. Furthermore, the interpolation constant $C_{\mathrm{err},j}$, also appearing in $C_{D,j}, C_{DB,j}, C_{DF,j}$, leads to restrictions involving the mesh size h and the overlap δ.

The lower bound on the field of values could be improved by designing a suitable coarse space to add a second level to the standard SORAS preconditioner. Note that for generic symmetric positive definite problems, robust lower bounds *on the spectrum* can be indeed obtained in this manner [21], but for generic non-self-adjoint or indefinite problems this currently constitutes a major challenge.

4.3 Numerical experiments

To conclude, we test numerically the performance of the preconditioner on the reaction-convection-diffusion problem (4.1) with Ω a rectangle $[0, N\cdot 0.2]\times[0,0.2]$ (where N is the number of subdomains), $\Gamma_D = \Gamma$, $\Gamma_R = \emptyset$; for the local problems of the preconditioner, Robin transmission conditions with parameter α as in (4.3)

	#SORAS(ORAS)			
$\mathbf{a} = 2\pi[-(y-0.1),(x-0.5)]^T$	$\delta = 2h$	$\delta = 4h$	$\delta = 6h$	$\delta = 8h$
$c_0 = 1,\ \nu = 1$	21(18)	20(14)	20(12)	19(11)
$c_0 = 1,\ \nu = 0.001$	14(9)	13(6)	12(5)	12(5)
$c_0 = 0.001,\ \nu = 1$	21(20)	20(15)	20(13)	19(11)
$c_0 = 0.001,\ \nu = 0.001$	15(10)	14(7)	13(5)	13(5)

Table 1 Iteration numbers for SORAS(ORAS) preconditioners in the case of a convection field $\mathbf{a} = 2\pi[-(y-0.1),(x-0.5)]^T$, for different values of the overlap δ, the reaction coefficient c_0 and the viscosity ν. The domain is decomposed into $N = 5$ overlapping vertical strips and the global problem has 18361 degrees of freedom.

are imposed on the subdomain interfaces. In Tables 1,2,3 we take $N = 5$ and $f = 100\exp\{-10((x-0.5)^2+(y-0.1)^2)\}$, which is centered at the barycenter of Ω. In Tables 4, 5 we vary N to test weak scaling, where the size of the problem increases about proportionally to N (note that the width of Ω above is proportional to N), and we take $f = 100\exp\{-10((x-0.1)^2+(y-0.1)^2)\}$, which is centered on the left of Ω. The problem is discretized by piece-wise linear Lagrange finite elements on a uniform triangular mesh with 60 points on the vertical side of the rectangle and $N \cdot 60$ points on the horizontal one, resulting in 18361 degrees of freedom for $N = 5$, and 7381, 14701, 29341, 58621, 117181, 234301 degrees of freedom for $N = 2, 4, 8, 16, 32, 64$ respectively. In Tables 1–4 the domain is partitioned into N vertical strips, while in Table 5 we consider arbitrary partitions into N irregular subdomains obtained using the automatic mesh partitioner METIS [23]; then each subdomain is augmented with mesh elements layers of size $\delta/2$ to obtain the overlapping decomposition (the total width of the overlap between two subdomains is then δ).

GMRES with right preconditioning is stopped when the relative residual is reduced by 10^{-6}. In Tables 1,2,3 we take a zero initial guess, while in Tables 4, 5 we take a random initial guess. We test SORAS preconditioner (2.1) and also ORAS preconditioner:

$$M_{ORAS}^{-1} \coloneqq \sum_{j=1}^{N} R_j^T D_j B_j^{-1} R_j.$$

In the tables we use # to denote the number of iterations for convergence. To apply the preconditioner, the local problems in each subdomain are solved with the direct solver MUMPS [3]. All the computations are done in the ffddm framework of FreeFEM, an open source domain specific language (DSL) specialised for solving boundary value problems with variational methods.

We examine several configurations for the coefficients in (4.1). First, in Table 1 we consider a rotating convection field $\mathbf{a} = 2\pi[-(y-0.1),(x-0.5)]^T$ and small/large values for the reaction coefficient c_0 and the viscosity ν. We can see that a larger overlap helps the convergence of the preconditioners, as expected. The number of iterations appears not very sensitive to c_0, while it increases when ν is larger. ORAS preconditioner performs better than SORAS preconditioner, but currently the convergence of ORAS preconditioner can not be rigorously analyzed.

Then, in Table 2 we take $\mathbf{a} = [-x, -y]^T$, which has negative divergence $\operatorname{div}\mathbf{a} = -2$, to test the robustness of the method when condition (4.2) on the positiveness

	#SORAS(ORAS)			
$\mathbf{a} = [-x, -y]^T$	$\delta = 2h$	$\delta = 4h$	$\delta = 6h$	$\delta = 8h$
$c_0 = 1,\ \nu = 1$	21(19)	21(14)	20(13)	20(11)
$c_0 = 1,\ \nu = 0.001$	16(7)	16(7)	16(6)	16(6)
$c_0 = 0.001,\ \nu = 1$	22(24)	22(18)	22(15)	21(13)
$c_0 = 0.001,\ \nu = 0.001$	17(8)	16(7)	16(7)	16(6)

Table 2 Repeat of Table 1 but with $\mathbf{a} = [-x, -y]^T$. In this case $\operatorname{div} \mathbf{a} = -2$ is negative and $\tilde{c} = c_0 - 1$ does not verify condition (4.2).

	#SORAS(ORAS)			
$\mathbf{a} = [1, 0]^T$	$\delta = 2h$	$\delta = 4h$	$\delta = 6h$	$\delta = 8h$
$c_0 = 1,\ \nu = 1$	20(18)	20(15)	20(13)	20(12)
$c_0 = 1,\ \nu = 0.001$	11(6)	11(5)	11(5)	11(5)
$c_0 = 0.001,\ \nu = 1$	20(20)	20(16)	20(14)	20(13)
$c_0 = 0.001,\ \nu = 0.001$	12(6)	12(5)	12(5)	12(5)

Table 3 Repeat of Table 1 but with $\mathbf{a} = [1, 0]^T$ and with Streamline Upwind Petrov-Galerkin stabilization for the Galerkin approximation.

of $\tilde{c}$ is violated: in this case, $\tilde{c} = c_0 - 1$, so $\tilde{c} = 0$, $\tilde{c} = -0.999$ for $c_0 = 1$, $c_0 = 0.001$ respectively. We can observe that both preconditioners still perform well.

Finally, in Table 3 we consider a horizontal convecting field $\mathbf{a} = [1, 0]^T$, which is normal to the interfaces between subdomains. In this case non-physical numerical instabilities appear in the solution. Note that this is a discretization issue, not related to the preconditioner: the robust direct solver MUMPS yields the same instabilities in the numerical solution as the preconditioned GMRES solver. To stabilize the discrete variational formulation, we use the Streamline Upwind Petrov-Galerkin (SUPG) method, which adds to the Galerkin approximation the following term (see for instance [28, §11.8.6]):

$$\mathcal{L}_h(u_h, f; v_h) = \theta \sum_{\tau \in \mathcal{T}^h} \int_\tau (\mathcal{L} u_h - f) \frac{h_\tau}{|\mathbf{a}|} \mathcal{L}_{SS} v_h,$$

where θ is a stabilization parameter (here we choose $\theta = 0.15$), h_τ is the diameter of the mesh element τ, and

$$\mathcal{L} u_h = c_0 u_h + \operatorname{div}(\mathbf{a} u_h) - \operatorname{div}(\nu \nabla u_h), \quad \mathcal{L}_{SS} v_h = \frac{1}{2} \operatorname{div}(\mathbf{a} v_h) + \frac{1}{2} \mathbf{a} \cdot \nabla v_h.$$

We can see that ORAS preconditioner performs better than SORAS preconditioner, but depends more on the overlap size.

Now, again in this third configuration with $\mathbf{a} = [1, 0]^T$ and SUPG stabilization, we perform a weak scaling test by taking $\Omega = [0, N \cdot 0.2] \times [0, 0.2]$ for increasing number of subdomains N. First we consider a regular decomposition into vertical strips (Table 4) and then an arbitrary decomposition made by METIS (Table 5). We fix the overlap $\delta = 4h$. Comparing Table 4 with Table 5, we can see that the number of iterations is higher when taking arbitrary-shaped subdomains. Moreover, in the cases with $\nu = 0.001$, convergence deteriorates with N, which shows the need of designing robust two-level preconditioners.

	#SORAS(ORAS)					
$\mathbf{a} = [1, 0]^T$	$N = 2$	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$
$c_0 = 1,\ \nu = 1$	18(15)	23(18)	28(19)	35(19)	36(19)	36(19)
$c_0 = 1,\ \nu = 0.001$	8(3)	10(5)	16(8)	23(16)	37(32)	63(62)
$c_0 = 0.001,\ \nu = 1$	18(15)	23(19)	29(21)	35(21)	36(21)	36(21)
$c_0 = 0.001,\ \nu = 0.001$	8(3)	10(5)	16(8)	24(16)	40(32)	71(64)

Table 4 Weak scaling test, with a regular decomposition into N vertical strips ($\delta = 4h$). The global problem has 7381, 14701, 29341, 58621, 117181, 234301 degrees of freedom for $N = 2, 4, 8, 16, 32, 64$ subdomains respectively.

	#SORAS(ORAS)					
$\mathbf{a} = [1, 0]^T$	$N = 2$	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$
$c_0 = 1,\ \nu = 1$	21(17)	30(22)	40(23)	48(23)	53(23)	55(23)
$c_0 = 1,\ \nu = 0.001$	10(4)	12(5)	17(9)	25(17)	38(32)	63(63)
$c_0 = 0.001,\ \nu = 1$	21(18)	30(25)	40(28)	48(27)	54(28)	57(29)
$c_0 = 0.001,\ \nu = 0.001$	10(4)	12(5)	18(9)	26(17)	42(33)	73(65)

Table 5 Weak scaling test as in Table 4, but with METIS decomposition into N arbitrary-shaped subdomains.

References

1. Achdou, Y., Le Tallec, P., Nataf, F., Vidrascu, M.: A domain decomposition preconditioner for an advection-diffusion problem. Comput. Methods Appl. Mech. Engrg. **184**, 145–170 (2000)
2. Alart, P., Barboteu, M., Le Tallec, P., Vidrascu, M.: Méthode de Schwarz additive avec solveur grossier pour problèmes non symétriques. Comptes Rendus de l'Académie des Sciences-Series I-Mathematics **331**(5), 399–404 (2000)
3. Amestoy, P., Duff, I., L'Excellent, J., Koster, J.: A fully asynchronous multifrontal solver using distributed dynamic scheduling. SIAM Journal on Matrix Analysis and Applications **23**(1), 15–41 (2001)
4. Beckermann, B., Goreinov, S.A., Tyrtyshnikov, E.E.: Some remarks on the Elman estimate for GMRES. SIAM journal on Matrix Analysis and Applications **27**(3), 772–778 (2005)
5. Bonazzoli, M., Dolean, V., Graham, I.G., Spence, E.A., Tournier, P.H.: Domain decomposition preconditioning for the high-frequency time-harmonic Maxwell equations with absorption. Math. Comp. **88**(320), 2559–2604 (2019). DOI 10.1090/mcom/3447
6. Bourgat, J.F., Glowinski, R., Le Tallec, P., Vidrascu, M.: Variational formulation and algorithm for trace operator in domain decomposition calculations. In: Domain Decomposition Methods, pp. 3–16. SIAM, Philadelphia, PA (1989)
7. Cai, X.C.: Additive Schwarz algorithms for parabolic convection-diffusion equations. Numerische Mathematik **60**(1), 41–61 (1991)
8. Cai, X.C., Widlund, O.B.: Domain decomposition algorithms for indefinite elliptic problems. SIAM J. Sci. Statist. Comput. **13**(1), 243–258 (1992)
9. Chan, T.F., Zou, J.: A convergence theory of multilevel additive Schwarz methods on unstructured meshes. Numerical Algorithms **13**(2), 365–398 (1996). DOI 10.1007/BF02207701
10. Ciarlet, P.G.: The finite element method for elliptic problems. North-Holland Publishing Co. (1978)
11. Dolean, V., Jolivet, P., Nataf, F.: An introduction to domain decomposition methods: algorithms, theory and parallel implementation. SIAM, Philadelphia, PA (2015). DOI 10.1137/1.9781611974065.ch1
12. Eisenstat, S.C., Elman, H.C., Schultz, M.H.: Variational iterative methods for nonsymmetric systems of linear equations. SIAM Journal on Numerical Analysis **20**(2), 345–357 (1983)
13. Elman, H.C.: Iterative methods for large, sparse, nonsymmetric systems of linear equations. Ph.D. thesis, Yale University New Haven, Conn (1982)

14. Essai, A.: Weighted FOM and GMRES for solving nonsymmetric linear systems. Numer. Algorithms **18**(3-4), 277–292 (1998). DOI 10.1023/A:1019177600806
15. Gong, S., Graham, I.G., Spence, E.A.: Domain decomposition preconditioners for high-order discretizations of the heterogeneous Helmholtz equation. IMA Journal of Numerical Analysis (2020). DOI 10.1093/imanum/draa080
16. Graham, I.G., Spence, E.A., Vainikko, E.: Domain decomposition preconditioning for high-frequency Helmholtz problems with absorption. Math. Comp. **86**(307), 2089–2127 (2017). DOI 10.1090/mcom/3190
17. Graham, I.G., Spence, E.A., Zou, J.: Domain Decomposition with Local Impedance Conditions for the Helmholtz Equation with Absorption. SIAM J. Numer. Anal. **58**(5), 2515–2543 (2020). DOI 10.1137/19M1272512
18. Greenbaum, A., Ptak, V., Strakoš, Z.: Any nonincreasing convergence curve is possible for GMRES. SIAM Journal on Matrix Analysis and Applications **17**(3), 465–469 (1996)
19. Griebel, M., Oswald, P.: On the abstract theory of additive and multiplicative Schwarz algorithms. Numer. Math. **70**(2), 163–180 (1995). DOI 10.1007/s002110050115
20. Grisvard, P.: Elliptic problems in nonsmooth domains, *Monographs and Studies in Mathematics*, vol. 24. Pitman, Boston, MA (1985)
21. Haferssas, R., Jolivet, P., Nataf, F.: A robust coarse space for optimized Schwarz methods: SORAS-GenEO-2. C. R. Math. Acad. Sci. Paris **353**(10), 959–963 (2015). DOI 10.1016/j.crma.2015.07.014
22. Japhet, C., Nataf, F., Rogier, F.: The optimized order 2 method: Application to convection–diffusion problems. Future Generation Computer Systems **18**(1), 17 – 30 (2001). DOI https://doi.org/10.1016/S0167-739X(00)00072-8
23. Karypis, G., Kumar, V.: A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM Journal on Scientific Computing **20**(1), 359–392 (1998)
24. Kimn, J.H., Sarkis, M.: Restricted overlapping balancing domain decomposition methods and restricted coarse problems for the Helmholtz problem. Comput. Methods Appl. Mech. Engrg. **196**(8), 1507–1514 (2007). DOI 10.1016/j.cma.2006.03.016
25. Lube, G., Mueller, L., Otto, F.C.: A non-overlapping domain decomposition method for the advection-diffusion problem. Computing **64**, 49–68 (2000)
26. Nataf, F., Rogier, F.: Factorization of the convection-diffusion operator and the Schwarz algorithm. M^3AS **5**(1), 67–93 (1995)
27. Nepomnyaschikh, S.V.: Mesh theorems of traces, normalizations of function traces and their inversions. Sov. J. Numer. Anal. Math. Modeling **6**, 1–25 (1991)
28. Quarteroni, A.M.: Numerical models for differential problems, vol. 2. Springer (2009)
29. Spence, E.A.: "When all else fails, integrate by parts" - an overview of new and old variational formulations for linear elliptic PDEs, pp. 93–159. SIAM (2015). DOI 10.1137/1.9781611973822.ch6
30. St-Cyr, A., Gander, M.J., Thomas, S.J.: Optimized multiplicative, additive, and restricted additive Schwarz preconditioning. SIAM J. Sci. Comput. **29**(6), 2402–2425 (2007). DOI 10.1137/060652610
31. Trefethen, L.N., Embree, M.: Spectra and pseudospectra: the behavior of nonnormal matrices and operators. Princeton University Press (2005)