BEST ROBIN PARAMETERS FOR OPTIMIZED SCHWARZ METHODS AT CROSS POINTS

MARTIN J. GANDER AND FELIX KWOK

Abstract. Optimized Schwarz methods are domain decomposition methods in which a large-scale PDE problem is solved by subdividing it into smaller subdomain problems, solving the subproblems in parallel, and iterating until one obtains a global solution that is consistent across subdomain boundaries. Fast convergence can be obtained if Robin conditions are used along subdomain boundaries, provided that the Robin parameters p are chosen correctly. In the case of second order elliptic problems such as the Poisson equation, it is well known for two-subdomain problems without overlap that the optimal choice is $p = O(h^{-1/2})$ (where h is the mesh size), with the resulting method having a convergence factor of $\rho = 1 - O(h^{1/2})$. However, when cross points are present, i.e., when several subdomains meet at a single point, this choice leads to a divergent method. In this article, we show for a model problem that convergence can only occur if $p = O(h^{-1})$ at the cross point; thus, a different scaling of the Robin parameter is needed to ensure convergence. In addition, this choice of p allows us to recover the $1 - O(h^{1/2})$ convergence factor in the resulting method.

1. Introduction

When solving large-scale elliptic problems that arise from physical or engineering applications, an attractive way to parallelize their solution is to use optimized Schwarz methods (OSM, [14]). In such methods, one subdivides the physical domain into many subdomains, and then solves the subdomain problems in parallel. One then iterates this process until one finds a global solution that is consistent across subdomain boundaries. For such an algorithm to be well-defined, one must impose boundary conditions along physical as well as *artificial boundaries*, i.e., interfaces between subdomains which were not part of the boundary of the whole domain. A judicious choice of artificial boundary conditions (or *transmission conditions*) can accelerate the convergence of the overall algorithm substantially; see [22] for Laplace's equation, [6] for the Helmholtz equation, [16] for nonlinear elliptic problems, [15] for the time-dependent wave equation, [3, 4] for problems with corners; see also [12] and references therein. In two dimensions or higher, the optimal choice of interface conditions involves nonlocal operators [1, 25, 24], which render the subdomain problems expensive to solve and hard to implement in practice. Various easier-to-implement *local* approximations of the optimal operators have been developed for the Laplace equation [10], the Helmholtz equation [2] and for the convection-diffusion equation, based on Taylor expansions [1], absorbing boundary conditions [26, 17], approximate factorization [24] and equioscillation properties [18, 20, 19]. In the case of non-overlapping decompositions with Robin transmission conditions for $(\eta - \Delta)u = f$, $\eta \geq 0$, [22] shows, using energy estimates, that the continuous optimized Schwarz iteration converges for any Robin parameter $p > 0$; similar arguments appear in [5] for linear second-order elliptic PDEs and in [7] for the harmonic Maxwell equations.

Optimized Schwarz methods can also be formulated algebraically in the discrete setting [29, 27], with properties similar to the continuous case when the mesh is fine enough and no cross points are present, i.e., no grid points are adjacent to three or more subdomains. However, the correct formulation of OSMs in the presence of cross points is a delicate problem, since it is difficult to discretize the PDE and boundary conditions while maintaining continuity of the solution [11]. In [23], Loisel presents a general formulation for OSM and the closely related

2-Lagrange multiplier method (2LM) that handles cross points systematically. In addition, the author shows that for a linear self-adjoint coercive 2nd order elliptic PDE, if the Robin parameter scales like $p = O(h^{-1/2})$, then the condition number of the 2LM-preconditioned system also scales like $O(h^{-1/2})$, which is asymptotically optimal. Even so, the 2LM-preconditioned matrix contains eigenvalues outside the unit disc centered at 1, meaning that the method would diverge unless it is used with a Krylov method. This is in apparent contradiction with the convergence results in [22] and [5], where cross points do not play a role in the convergence of the continuous method. In addition, this divergence prevents OSMs from being used in certain contexts, e.g., as smoothers within a multigrid/multilevel algorithm.

The goal of the current paper is to further understand the spectral properties of OSM/2LM-preconditioned matrices. In particular, we show that when $p = O(h^{-1/2})$ and when cross points are present, the spectrum of the preconditioned system contains large eigenvalues that lead to divergence when the method is used as a stationary iteration; a similar behaviour has been observed in Additive Schwarz with overlap, where the overlap causes stagnation or divergence of the iterative method [9, 13]. However, our analysis shows that if we modify the Robin parameter at the corner to scale like $O(1/h)$, then a convergence factor of $1 - O(\sqrt{h})$ can be restored in the iterative method.

Our paper is organized as follows. In Section 2, we describe the continuous formulation of the optimized Schwarz method, from which we derive the discrete version, which is intimately related to the 2LM method. We also describe symmetry assumptions that will be used for analysis. We then present in Section 3 a fully discrete analysis of the spectral radius of the iteration and show that OSM diverges unless the parameter at the corner scales like $O(1/h)$. In Section 4, we give the asymptotic behavior of the Robin parameters in order for OSM to converge with a factor of $1 - O(\sqrt{h})$. We finally present in Section 5 numerical results confirming the analysis. In addition, we show how choosing a different scaling for the cross point parameters can accelerate the convergence of Krylov methods, especially for three-dimensional problems.

2. Continuous and Discrete Formulations of OSM

2.1. **Continuous Formulation.**

Suppose we want to solve for $d = 2$ or 3

$$\mathcal{L}u = f \quad \text{on } \Omega \subset \mathbb{R}^d, \qquad u = g \quad \text{on } \partial\Omega, \tag{1}$$

with an optimized Schwarz method, where $\mathcal{L} = \eta - \Delta$, $\eta \geq 0$ is the positive definite Helmholtz operator. (Our results also hold, after trivial modifications, for more general coercive second-order elliptic operators.) Assume we have a conformal non-overlapping decomposition $\{\Omega_j\}_{j=1}^N$ i.e., for any i, j, the intersection of the closures of Ω_i and Ω_j, if non-empty, must be either a common vertex or a common edge. When this common edge is non-trivial, we denote its interior (relative to $\partial\Omega_i$) by Γ_{ij}, so that the end points of the edge are excluded from Γ_{ij}. Then the optimized Schwarz method with Robin transmission conditions is defined as follows: for $k = 1, 2, \ldots$, solve for $i = 1, \ldots, N$,

$$\begin{aligned} &\mathcal{L}u_i^k = f \quad \text{on } \Omega_i, \qquad u_i^k = g \quad \text{on } \partial\Omega \cap \partial\Omega_i, \\ &\frac{\partial u_i^k}{\partial n_i} + p_{ij} u_i^k = \frac{\partial u_j^{k-1}}{\partial n_i} + p_{ij} u_j^{k-1} \quad \text{on } \Gamma_{ij} \text{ for all } \Gamma_{ij} \neq \emptyset. \end{aligned} \tag{2}$$

The Robin parameter p_{ij} in (2) is assumed to be strictly positive, but is allowed to vary along Γ_{ij}. We further assume that $p_{ij} = p_{ji}$, i.e., the same Robin parameter is used for communication in both directions. Lions showed in [22], using an energy estimate argument, that for any choice of $p_{ij} > 0$, (2) converges to the unique solution of (1) in $L^2(\Omega_i)$ and weakly in $L^2(\partial\Omega_i)$. However, as noted in [11], the discretization of (2) is delicate when cross points are present, since one must weakly impose two different sets of Robin conditions for each piece of the interface at the cross point. A similar difficulty has been observed in [28] in the optimized Schwarz preconditioner in the context of solving the magnetohydrodynamics equations.

2.2. **Discrete Formulation.** We now consider the discrete formulation of the optimized Schwarz method when cross points are present that is a special case of the 2LM method described in [23]. Suppose we have a finite element mesh for Ω such that each finite element lies within exactly one subdomain Ω_i, i.e., each interface Γ_{ij} must consist of a union of element boundaries. Let R_i be the restriction operator onto the set of degrees of freedom in $\overline{\Omega_i}$, and R_i^T is the corresponding extension operator from Ω_i into the set of all degrees of freedom in Ω. Then a finite-element discretization of (1) leads to a linear system of the form

$$\mathfrak{A}\mathbf{u} = \mathbf{f},$$

where $\mathfrak{A} = \sum_{j=1}^n R_j^T A_j R_j$ is the global stiffness matrix and $A_i = \sum_{e\subset\Omega_i} R_e^T A_e R_e$ is the subdomain stiffness matrix for Ω_i, with A_e being the element stiffness matrix for element e, R_e the restriction operator onto the element e, and the sum running over all elements e that lie within Ω_i. Note that A_i does not yet contain Robin transmission conditions. To incorporate the interface conditions, we need to consider $\tilde{A}_i = A_i + hL_i$, where L_i contains the Robin contributions along the interface (and is thus zero at internal nodes). The discrete optimized Schwarz method is then defined as follows: for $k = 1, 2, \ldots$, solve

$$\tilde{A}_i u_i^k = f_i + \sum_{j\neq i} B_{ij} u_j^{k-1}, \qquad i = 1, \ldots, n, \tag{3}$$

where $f_i = R_i\mathbf{f}$. The B_{ij}, which extract data from neighboring subdomains, can be calculated as follows. At convergence, each subdomain solution u_i must satisfy $u_i = R_i\mathbf{u}$, where $\mathbf{u}$ is the global solution. So $\mathfrak{A}\mathbf{u} = \mathbf{f}$ implies

$$\begin{aligned}
R_i\mathbf{f} &= R_i \sum_{j=1}^n R_j^T A_i R_j \mathbf{u} \\
&= \underbrace{R_i R_i^T}_{=I} A_i R_i \mathbf{u} + \sum_{j\neq i} R_i R_j^T A_j R_i \mathbf{u} \\
&= (A_i + hL_i) R_i\mathbf{u} - hL_i R_i\mathbf{u} + \sum_{j\neq i} R_i R_j^T A_j R_j \mathbf{u} \\
&= \tilde{A}_i u_i + \sum_{j\neq i} (R_i R_j^T A_j - hL_i R_i \tilde{R}_j^T) u_j,
\end{aligned}$$

where $\tilde{R}_j$ is chosen to be identical to R_j except at cross points, where it has a weight of $1/(d-1)$, with d being the number of subdomains meeting at that cross point. This ensures the weights sum to 1 at any interface or cross point, so that

$$\sum_{j\neq i} L_i R_i \tilde{R}_j^T R_j = L_i R_i \qquad \text{for all } i..$$

This gives the definition

$$B_{ij} := -R_i R_j^T A_j + hL_i R_i \tilde{R}_j^T. \tag{4}$$

We further assume that L_i is a diagonal matrix with support along artificial interfaces only; in addition to simplifying our analysis below, this choice arises naturally as part of the discrete integration by parts formula, cf. [11].

2.3. **Equivalence with the 2LM method.** We now show the equivalence of (3) with the 2LM method introduced by Loisel [23]. Suppose each subdomain stiffness matrix A_i is partitioned into internal and interface nodes. Then the subdomain systems (3) can be written as

$$\begin{bmatrix} A_{II,i} & A_{I\Gamma,i} \\ A_{\Gamma I,i} & A_{\Gamma\Gamma,i} + h\hat{L}_i \end{bmatrix} \begin{pmatrix} u_{I,i}^k \\ u_{\Gamma,i}^k \end{pmatrix} = \begin{pmatrix} f_{I,i} \\ f_{\Gamma,i}^e + \lambda_i^{k-1} \end{pmatrix}, \tag{5}$$

where $f^e_{\Gamma,i}$ is the *unassembled* right-hand side corresponding to integration over Ω_i *only*; to obtain the total right-hand side, we need to sum up the contributions from all the neighbouring subdomains. (In other words, the contributions from other subdomains have been absorbed into the unknown λ_i^{k-1}.) When the interior variables are eliminated, we obtain

$$(S_i + h\hat{L}_i)u^k_{\Gamma,i} = g_i + \lambda_i^{k-1}, \tag{6}$$

where $S_i = A_{\Gamma\Gamma,i} - A_{\Gamma I,i}A^{-1}_{II,i}A_{I\Gamma,i}$ is the Schur complement and $g_i = f^e_{\Gamma,i} - A_{\Gamma I,i}A^{-1}_{II,i}f_{I,i}$ is the condensed right-hand side. The 2LM method requires solving a system of the form

$$A_{\text{2LM}}\lambda = c$$

for the unknown Robin traces $\lambda = [\lambda_1^T, \dots, \lambda_n^T]^T$, where

$$A_{\text{2LM}} = h(LM - GL)(S + hL)^{-1} + G, \qquad c = (G - A_{\text{2LM}})g. \tag{7}$$

Here, $g = [g_1^T, \dots, g_n^T]^T$, $L = \text{diag}(\hat{L}_1, \dots, \hat{L}_n)$, $S = \text{diag}(S_1, \dots, S_n)$; M is a matrix with ones on the diagonal and $M_{ij} = -\frac{1}{d_i - 1}$ whenever i is a point adjacent to d_i subdomains and $j \neq i$ corresponds to the same physical point as i, and G has the same non-zero pattern as M, except all entries are $+1$. We claim that the discrete formulation (3) is equivalent to the stationary iteration

$$\lambda^k = (I - A_{\text{2LM}})\lambda^{k-1} + c, \tag{8}$$

which is none other than the Richardson iteration applied to the 2LM system. Indeed, using the definition of A_{2LM} in (7), we can write

$$I - A_{\text{2LM}} = [(I - G)S + hL(I - M)](S + hL)^{-1},$$

from which we can use the definition of S and (6) to infer that

$$(I - A_{\text{2LM}})(g + \lambda^{k-1}) = [(I - G)S + hL(I - M)]u^k_\Gamma.$$

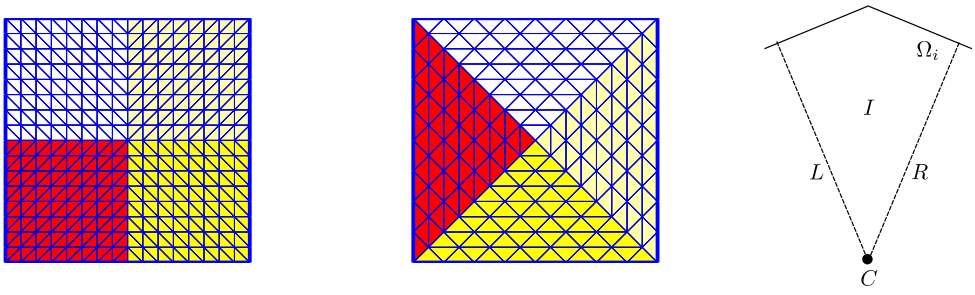
The second equation in (7) then implies

$$[(I - G)S + hL(I - M)]u^k_\Gamma = (I - G)g + c + (I - A_{\text{2LM}})\lambda^{k-1}.$$

But $I - M$ is a matrix with zeros on the diagonal and $1/(d_i - 1)$ at the (i, j) position whenever j is at the same physical point as i. So the term $hL(I - M)u^k_\Gamma$, when restricted to Γ_i, is equal to $h\hat{L}_i R_{\Gamma_i}\sum_{j\neq i}\tilde{R}_j^T u_j^k$, where R_{Γ_i} is the restriction onto the set of interface points in Ω_i. Similarly, $I - G$ has zeros on the diagonal and -1 at the same positions where $I - M$ is non-zero. This implies that the restriction of $(I - G)(Su^k_\Gamma - g)$ onto Γ_i is $-R_{\Gamma_i}\sum_{j\neq i}(R_j^T A_j u_j^k - R^T_{\Gamma_j} f^e_{\Gamma,j})$. Combining the two terms, we get

$$\begin{aligned}
[(I - A_{\text{2LM}})\lambda^{k-1}]_i + c_i &= h\sum_{j\neq i}\hat{L}_i R_{\Gamma_i}\tilde{R}_j^T u_j^k - \sum_{j\neq i} R_{\Gamma_i}(R_j^T A_j u_j^k - R^T_{\Gamma_j} f^e_{\Gamma,j}) \\
&= R_{\Gamma_i}R_i^T\sum_{j\neq i}(hL_i R_i\tilde{R}_j^T - R_i R_j^T A_j)u_j^k + R_{\Gamma_i}\sum_{j\neq i} R^T_{\Gamma_j} f^e_{\Gamma,j} \\
&= R_{\Gamma_i}\Big(R_i^T\sum_{j\neq i} B_{ij}u_j^k + \sum_{j\neq i} R^T_{\Gamma_j} f^e_{\Gamma,j}\Big) \qquad \text{(by (4))} \\
&= R_{\Gamma_i}\Big(R_i^T\tilde{A}_i u_i^{k+1} - R_i^T f_i + \sum_{j\neq i} R^T_{\Gamma_j} f^e_{\Gamma,j}\Big) \qquad \text{(by (3))} \\
&= R_{\Gamma_i}R_i^T\tilde{A}_i u_i^{k+1} - f^e_{\Gamma,i} = \lambda_i^k,
\end{aligned}$$

where we have used the second row of (5) together with the fact that $R_{\Gamma_i}R_i^T f_i = R_{\Gamma_i}\sum_j R^T_{\Gamma_j} f^e_{\Gamma,j}$. Thus, we have shown that the subdomain iteration (3) produces the same iterates as the Richardson iteration (8), (5) as long as the initial guesses are compatible, e.g. when $u_i^0 = 0$ and



(a) (b) (c)

FIGURE 1. (a), (b): two decompositions of the unit square into four subdomains that satisfy the symmetry conditions; (c) a generic wedge with its interior (I), left edge (L), right edge (R) and center (C) nodes identified.

$\lambda_i^0 = R_{\Gamma i} \sum_{j \neq i} R_{\Gamma_j}^T f_{\Gamma,j}^e$. Thus, even though the analysis in the remainder of this paper is on the subdomain iteration (3), the results also apply to the 2LM method.

Remark. No definition made in the above subsection will be reused in the remainder of this paper. In fact, certain letters (e.g., M, G, L, and g) will be redefined to denote other quantities.

2.4. **Model Problem.** For analysis purposes, let us consider the problem of solving (1) in two dimensions on a regular n-gon $\Omega \subset \mathbb{R}^2$ using the optimized Schwarz method (3). We decompose the domain into n identical wedges $\Omega_1, \ldots, \Omega_n$, and each Ω_i is discretized identically using a finite element method on a triangular mesh, see Figure 1(a), (b) for two possibilities. Note that there is exactly one cross point of degree n in the center of the polygon. Since the wedges are identical, we can assume that there exists a local ordering of the nodes such that their stiffness matrices are also identical, i.e.

$$A_1 = \cdots = A_n =: A.$$

In particular, we use the ordering below for the rows and columns of A:

(1) interior nodes,
(2) nodes on the *left* boundary (excluding the center),
(3) nodes on the *right* boundary (excluding the center),
(4) the center node.

This gives the following block representation for A:

$$A = \begin{bmatrix} A_{II} & A_{IL} & A_{IR} & A_{IC} \\ A_{LI} & A_{LL} & A_{LR} & A_{LC} \\ A_{RI} & A_{RL} & A_{RR} & A_{RC} \\ A_{CI} & A_{CL} & A_{CR} & A_{CC} \end{bmatrix}, \tag{9}$$

where the subscripts I, L, R and C denote the interior, left edge, right edge and center nodes respectively, cf. Figure 1(c). Note that A is symmetric positive definite, since $\partial\Omega \cap \partial\Omega_i \neq \emptyset$ for each i. If we eliminate the internal nodes, we obtain the *Schur complement* matrix S_0, which is also symmetric positive definite. In block form, we have

$$S_0 = \begin{bmatrix} S_{LL} & S_{LR} & S_{LC} \\ S_{RL} & S_{RR} & S_{RC} \\ S_{CL} & S_{CR} & S_{CC} \end{bmatrix}, \tag{10}$$

where $S_{LL} = A_{LL} - A_{LI}A_{II}^{-1}A_{IL}$, etc. By symmetry, we assume that the left and right edges are identical in the sense that

$$TS_0 = S_0T, \qquad \text{where } T := \begin{bmatrix} 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{11}$$

Finally, because of rotational symmetry, we know that the operator $R_iR_j^TA - hLR_i\tilde{R}_j^T$ is constant whenever $j - i \equiv \ell \mod n$ for a fixed ℓ. This means we can write the method in terms of the following augmented system:

$$\begin{bmatrix} \tilde{A} & & & \\ & \tilde{A} & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & \tilde{A} \end{bmatrix} \begin{pmatrix} u_1^{k+1} \\ u_2^{k+1} \\ \vdots \\ \vdots \\ u_n^{k+1} \end{pmatrix} = \begin{bmatrix} 0 & B_1 & B_2 & \cdots & B_{n-1} \\ B_{n-1} & 0 & B_1 & \ddots & \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ & \ddots & \ddots & \ddots & B_1 \\ B_1 & B_2 & \cdots & B_{n-1} & 0 \end{bmatrix} \begin{pmatrix} u_1^k \\ u_2^k \\ \vdots \\ \vdots \\ u_n^k \end{pmatrix} + \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ \vdots \\ f_n \end{pmatrix}, \tag{12}$$

where $B_i := B_{1,i}$ in the definition of (4), and $\tilde{A} = A + hL$ with

$$L = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & D & & \\ \hline 0 & & & & D & \\ \hline & & & & & p_C \end{array}\right].$$

Here $D > 0$ is a diagonal matrix and $p_C > 0$, both representing Robin transmission conditions, which are assumed to be the same for all subdomains and for both left and right edges.

3. Spectral Analysis

To analyze the asymptotic convergence rate of the method, we need to calculate the spectral radius of

$$N := \begin{bmatrix} 0 & B_1 & B_2 & \cdots & B_{n-1} \\ B_{n-1} & 0 & B_1 & \ddots & \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ & \ddots & \ddots & \ddots & B_1 \\ B_1 & B_2 & \cdots & B_{n-1} & 0 \end{bmatrix} \begin{bmatrix} \tilde{A} & & & \\ & \tilde{A} & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & \tilde{A} \end{bmatrix}^{-1}, \tag{13}$$

which has the same eigenvalues as the iteration matrix in (12). Since N is a block circulant matrix, it can be block diagonalized as follows. Let ω be any n-th root of unity, i.e., $\omega^n = 1$. Then we get

$$\begin{bmatrix} 0 & B_1\tilde{A}^{-1} & B_2\tilde{A}^{-1} & \cdots & B_{n-1}\tilde{A}^{-1} \\ B_{n-1}\tilde{A}^{-1} & 0 & B_1\tilde{A}^{-1} & \ddots & \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ & \ddots & \ddots & \ddots & B_1\tilde{A}^{-1} \\ B_1\tilde{A}^{-1} & B_2\tilde{A}^{-1} & \cdots & B_{n-1}\tilde{A}^{-1} & 0 \end{bmatrix} \begin{bmatrix} I \\ \omega I \\ \omega^2 I \\ \vdots \\ \omega^{n-1} I \end{bmatrix} = \sum_{i=1}^{n-1} \omega^i B_i \tilde{A}^{-1} \begin{bmatrix} I \\ \omega I \\ \omega^2 I \\ \vdots \\ \omega^{n-1} I \end{bmatrix}.$$

Thus, the spectral radius of N is the largest of the spectral radii of the matrices

$$C_k := \sum_{i=1}^{n-1} \omega^{ik} B_i \tilde{A}^{-1}, \qquad k = 0, 1, \ldots, n-1,$$

where we assume ω is a *primitive* n-th root of unity.

3.1. **Spectral Radius of C_k, $k \neq 0$.** Let us first consider the case where $k \neq 0$, so that C_k is complex. Each of the $B_i \tilde{A}^{-1}$ is of the form

$$\begin{aligned} B_i \tilde{A}^{-1} &= (hLR_1 \tilde{R}_{i+1}^T - R_1 R_{i+1}^T A)\tilde{A}^{-1} \\ &= (hLR_1 \tilde{R}_{i+1}^T + hR_1 R_{i+1}^T L - R_1 R_{i+1}^T \tilde{A})\tilde{A}^{-1} \\ &= h(LR_1 \tilde{R}_{i+1}^T + R_1 R_{i+1}^T L)\tilde{A}^{-1} - R_1 R_{i+1}^T. \end{aligned}$$

This means

$$C_k = h\left[L\Big(\sum_{i=1}^{n-1} \omega^{ik} R_1 \tilde{R}_{i+1}^T\Big) + \Big(\sum_{i=1}^{n-1} \omega^{ik} R_1 R_{i+1}^T\Big)L\right]\tilde{A}^{-1} - \sum_{i=1}^{n-1} \omega^{ik} R_1 R_{i+1}^T.$$

Using the same nodal ordering described in Section 2.4, we can then write $R_1 R_j^T$ in block matrix form for each j:

$$R_1 R_2^T = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & & I & \\ \hline 0 & & & 0 & & \\ \hline & & & & & 1 \end{array}\right], \qquad R_1 R_n^T = (R_1 R_2^T)^T = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & & 0 & \\ \hline 0 & & & I & & \\ \hline & & & & & 1 \end{array}\right]$$

and

$$R_1 R_j^T = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & & 0 & \\ \hline 0 & & & 0 & & \\ \hline & & & & & 1 \end{array}\right] \qquad \text{for } 3 \leq j \leq n-1.$$

Similarly, $R_1 \tilde{R}_j^T$ have exactly the same structure, except the 1 in the bottom right-hand corner is replaced by $1/(n-1)$. Since $\omega^k \neq 1$ for $k \neq 0 \pmod n$, we see that

$$\sum_{i=1}^{n-1} \omega^{ik} R_1 R_{i+1}^T = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & & \omega^k I & \\ \hline 0 & & & \omega^{-k} I & & \\ \hline & & & & & -1 \end{array}\right], \qquad \sum_{i=1}^{n-1} \omega^{ik} R_1 \tilde{R}_{i+1}^T = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & & \omega^k I & \\ \hline 0 & & & \omega^{-k} I & & \\ \hline & & & & & -\frac{1}{n-1} \end{array}\right],$$

where we have used the fact that $\omega^{k(n-1)} = \omega^{-k}$ and $1 + \omega^k + \omega^{2k} + \cdots + \omega^{(n-1)k} = 0$. We can now rewrite C_k as follows:

$$C_k = h\left(\left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & D & & \\ \hline 0 & & D & \\ \hline & & & p_C \end{array}\right]\left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & & \omega^k I & \\ \hline 0 & \omega^{-k} I & & \\ \hline & & & -\frac{1}{n-1} \end{array}\right] + \left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & & \omega^k I & \\ \hline 0 & \omega^{-k} I & & \\ \hline & & & -1 \end{array}\right]\left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & D & & \\ \hline 0 & & D & \\ \hline & & & p_C \end{array}\right]\right)\tilde{A}^{-1} - \left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & & \omega^k I & \\ \hline 0 & \omega^{-k} I & & \\ \hline & & & -1 \end{array}\right].$$

This allows us to conclude that

$$C_k = -\left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & & \omega^k I & \\ \hline 0 & \omega^{-k} I & & \\ \hline & & & -1 \end{array}\right]\left(\left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & I & & \\ \hline 0 & & I & \\ \hline & & & 1 \end{array}\right] - h\left[\begin{array}{c|c|c|c} 0 & & & \\ \ddots & 0 & 0 & \\ 0 & & & \\ \hline 0 & 2D & & \\ \hline 0 & & 2D & \\ \hline & & & \frac{np_C}{n-1} \end{array}\right]\tilde{A}^{-1}\right).$$

If $\tilde{S}$ is the Schur complement of $\tilde{A}$ after eliminating the interior nodes, we can show that

$$\rho(C_k) = \rho\left(\left[\begin{array}{cc|c} 0 & \omega^k I & \\ \omega^{-k} I & 0 & \\ \hline & & -1 \end{array}\right]\left(I - h\begin{bmatrix} 2D & & \\ & 2D & \\ & & \frac{np_C}{n-1} \end{bmatrix}\tilde{S}^{-1}\right)\right). \tag{14}$$

The following lemma gives an upper bound for $\rho(C_k)$ when $k \neq 0$.

Lemma 1. *Suppose $D > 0$, $p_C > 0$ and S_0 is symmetric positive definite. Then for $k = 1, \ldots, n-1$, we have $\rho(C_k) \leq \rho(W) < 1$, where*

$$W = I - h\begin{bmatrix} 2D & & \\ & 2D & \\ & & \frac{np_C}{n-1} \end{bmatrix}\tilde{S}^{-1}. \tag{15}$$

Proof. The matrix on the right-hand side of (14) has the same eigenvalues as

$$\left[\begin{array}{cc|c} 0 & \omega^k I & \\ \omega^{-k} I & 0 & \\ \hline & & -1 \end{array}\right]\left(I - h\begin{bmatrix} 2D & & \\ & 2D & \\ & & \frac{np_C}{n-1} \end{bmatrix}^{1/2}\tilde{S}^{-1}\begin{bmatrix} 2D & & \\ & 2D & \\ & & \frac{np_C}{n-1} \end{bmatrix}^{1/2}\right).$$

Now the spectral radius is bounded above by the 2-norm. The first matrix is unitary and hence has norm 1, whereas the second matrix is real and symmetric, so its 2-norm is equal to its spectral radius. Thus, we have

$$\rho(C_k) \leq \rho\left(I - h\begin{bmatrix} 2D & & \\ & 2D & \\ & & \frac{np_C}{n-1} \end{bmatrix}\tilde{S}^{-1}\right) = \rho(W).$$

Let S_0 denote the Schur complement of A (without the Robin boundary contributions, as defined in (10)) after the interior nodes have been eliminated. Then we can write

$$\tilde{S} = S_0 + h\,\mathrm{diag}(D, D, p_C).$$

It is now possible to rewrite W as

$$\begin{aligned}
W =& I - h\,\mathrm{diag}(2D, 2D, np_C/(n-1))(S_0 + h\,\mathrm{diag}(D, D, p_C))^{-1} \\
=& (S_0 - h\,\mathrm{diag}(D, D, p_c/(n-1)))(S_0 + h\,\mathrm{diag}(D, D, p_C))^{-1} \\
=& \Big[\underbrace{(S_0 + h\,\mathrm{diag}(0, 0, \frac{np_C}{2(n-1)}))}_{S} - h\,\mathrm{diag}(D, D, \frac{(n-2)p_C}{2(n-1)})\Big] \\
&\times \Big[\underbrace{(S_0 + h\,\mathrm{diag}(0, 0, \frac{np_C}{2(n-1)}))}_{S} + h\,\mathrm{diag}(D, D, \frac{(n-2)p_C}{2(n-1)})\Big]^{-1} \\
=& (I - Z)(I + Z)^{-1},
\end{aligned}$$

where

(16)
$$Z = h \begin{bmatrix} D & & \\ & D & \\ & & \frac{(n-2)p_C}{2(n-1)} \end{bmatrix} S^{-1} = h \begin{bmatrix} D & & \\ & D & \\ & & \frac{(n-2)p_C}{2(n-1)} \end{bmatrix} \left(S_0 + \begin{bmatrix} 0 & & \\ & 0 & \\ & & \frac{nhp_C}{2(n-1)} \end{bmatrix} \right)^{-1}.$$

Since $D > 0$, $p_C > 0$ and S_0 is symmetric positive definite, Z is a similarity transformation away from a symmetric positive definite matrix, so it only has positive eigenvalues. Thus, if $\lambda_1, \ldots, \lambda_r$ are the eigenvalues of Z, then the eigenvalues of W have the form

$$\mu_j = \frac{1 - \lambda_j}{1 + \lambda_j},$$

which have absolute value less than 1, since $\lambda_j > 0$. Hence $\rho(W) < 1$, as required. □

3.2. **Spectral radius of C_0.** The situation is completely different when $k = 0$. Here we have

$$\sum_{i=1}^{n-1} R_1 R_{i+1}^T = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & & I & \\ \hline 0 & & & I & & \\ \hline & & & & & n-1 \end{array}\right], \qquad \sum_{i=1}^{n-1} R_1 \tilde{R}_{i+1}^T = \left[\begin{array}{ccc|c|c|c} 0 & & & & & \\ & \ddots & & 0 & 0 & \\ & & 0 & & & \\ \hline 0 & & & & I & \\ \hline 0 & & & I & & \\ \hline & & & & & 1 \end{array}\right],$$

which implies

$$\rho(C_0) = \rho\left(\begin{bmatrix} 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left(\begin{bmatrix} I & & \\ & I & \\ & & n-1 \end{bmatrix} - h \begin{bmatrix} 2D & & \\ & 2D & \\ & & np_C \end{bmatrix} \tilde{S}^{-1} \right)\right) = \rho(TM),$$

with T from (11) and

$$M = \begin{bmatrix} I & & \\ & I & \\ & & n-1 \end{bmatrix} - h \begin{bmatrix} 2D & & \\ & 2D & \\ & & np_C \end{bmatrix} \tilde{S}^{-1}. \tag{17}$$

Lemma 2. *Let $D > 0$ be a diagonal matrix, $p_C > 0$ and S_0 symmetric positive definite. Then $\rho(C_0) = \rho(M)$.*

Proof. First, we observe that T and M commute since $T\tilde{S} = \tilde{S}T$, which follows from the fact that $TS_0 = S_0T$. Thus, it is possible to diagonalize T and M simultaneously with the same eigenbasis, i.e.

$$T = X\Lambda_T X^{-1}, \qquad M = X\Lambda_M X^{-1},$$

so that

$$TM = X(\Lambda_T\Lambda_M)X^{-1},$$

i.e., the eigenvalues of the product are the products of the eigenvalues. But the eigenvalues of T are ± 1; thus, we have

$$|\lambda(C_0)| = |\lambda(TM)| = |\lambda(M)|,$$

which implies $\rho(C_0) = \rho(M)$. □

We now mimic the case $k \neq 0$ and express M as $(I-G)(I+G)^{-1}$ for some G:

$$\begin{aligned}
M &= \operatorname{diag}(I, I, n-1) - h\operatorname{diag}(2D, 2D, np_C)(S_0 + h\operatorname{diag}(D, D, p_C))^{-1} \\
&= (\operatorname{diag}(I, I, n-1)(S_0 + h\operatorname{diag}(D, D, p_C)) - h\operatorname{diag}(2D, 2D, np_C))(S_0 + h\operatorname{diag}(D, D, p_C))^{-1} \\
&= (\operatorname{diag}(I, I, n-1)S_0 - h\operatorname{diag}(D, D, p_C))(S_0 + h\operatorname{diag}(D, D, p_C))^{-1} \\
&= \left[\operatorname{diag}(I, I, \frac{n}{2})S_0 - (h\operatorname{diag}(D, D, p_C) - \operatorname{diag}(0, 0, \frac{n-2}{2})S_0)\right] \\
&\qquad \times \left[\operatorname{diag}(I, I, \frac{n}{2})S_0 + (h\operatorname{diag}(D, D, p_C) - \operatorname{diag}(0, 0, \frac{n-2}{2})S_0)\right]^{-1} \\
&= (I-G)(I+G)^{-1}
\end{aligned}$$

where

$$G = h\begin{bmatrix} D & & \\ & D & \\ & & p_C \end{bmatrix} S_0^{-1} \begin{bmatrix} I & & \\ & I & \\ & & \frac{2}{n} \end{bmatrix} - \begin{bmatrix} 0 & & \\ & 0 & \\ & & 1-\frac{2}{n} \end{bmatrix},$$

which has the same eigenvalues as

$$\tilde{G} = h\begin{bmatrix} D & & \\ & D & \\ & & 2p_C/n \end{bmatrix} S_0^{-1} - \begin{bmatrix} 0 & & \\ & 0 & \\ & & 1-\frac{2}{n} \end{bmatrix}. \tag{18}$$

Again we want to find out what conditions D and p_C should satisfy in order for $\tilde{G}$ to have only positive eigenvalues. Since S_0^{-1} is difficult to compute, we will instead look at $\tilde{G}^{-1}$, which we compute using the Sherman–Morrison–Woodbury formula:

$$(X + uv^T)^{-1} = (I - \frac{1}{1 + v^T X^{-1}u} X^{-1}uv^T)X^{-1},$$

where we let

$$X = h\operatorname{diag}(D, D, 2p_C/n)S_0^{-1}, \qquad u = -\frac{n-2}{n}e, \qquad v = e,$$

with $e^T = (0, \ldots, 0, 1)$. Since $X^{-1} = h^{-1}S_0\operatorname{diag}(D^{-1}, D^{-1}, n/2p_C)$, we have

$$v^T X^{-1}u = -\frac{n-2}{nh}e^T S_0\operatorname{diag}(D^{-1}, D^{-1}, n/2p_C)e = -\frac{(n-2)e^T S_0 e}{2hp_C}.$$

So

$$\begin{aligned}
I - \frac{1}{1 + v^T X^{-1}u} X^{-1}uv^T &= I - \frac{2hp_C}{2hp_C - (n-2)e^T S_0 e} h^{-1}S_0\operatorname{diag}(D^{-1}, D^{-1}, n/2p_C)(-\frac{n-2}{n}e)e^T \\
&= I + \frac{n-2}{2hp_C - (n-2)e^T S_0 e} S_0 ee^T.
\end{aligned}$$

This means

$$\tilde{G}^{-1} = (X + uv^T)^{-1} = (I - \frac{1}{1 + v^T X^{-1} u} X^{-1} uv^T) X^{-1}$$

$$= \Big(S_0 + \frac{n-2}{2hp_C - (n-2)e^T S_0 e} S_0 e e^T S_0\Big) \begin{bmatrix} (hD)^{-1} & & \\ & (hD)^{-1} & \\ & & n/(2hp_C) \end{bmatrix}. \tag{19}$$

We are now ready to show our first main result.

Theorem 3. *Let S_0 be symmetric positive definite, $D > 0$ and $p_C > 0$. Then the optimized Schwarz method* (12) *converges if and only if the corner parameter p_C satisfies*

$$p_C > \frac{(n-2)e^T S_0 e}{2h}. \tag{20}$$

Proof. Since $D > 0$ and $p_C > 0$ by assumption, Lemma 1 shows that $\rho(C_k) < 1$ for $k = 1, \dots n-1$, so the method converges if and only if $\rho(C_0) < 1$. We start by showing that $\rho(C_0) = \rho(M) < 1$ if and only if $\tilde{G}$ (or equivalently $\tilde{G}^{-1}$) has only positive eigenvalues. Indeed, since the eigenvalues λ_j of $\tilde{G}$ are related to those of M (denoted by μ_j) by

$$\mu_j = \frac{1 - \lambda_j}{1 + \lambda_j},$$

we see that $|\mu_j| < 1 \iff \lambda_j > 0$.

We now show that $\tilde{G}$ has positive eigenvalues if and only if (20) holds. We first observe that $\tilde{G}^{-1}$ has the same eigenvalues as the symmetric matrix

$$\begin{bmatrix} (hD)^{-\frac{1}{2}} & & \\ & (hD)^{-\frac{1}{2}} & \\ & & \Big(\frac{n}{2hp_C}\Big)^{\frac{1}{2}} \end{bmatrix} \Big(S_0 + \frac{n-2}{2hp_C - (n-2)e^T S_0 e} S_0 e e^T S_0\Big) \begin{bmatrix} (hD)^{-\frac{1}{2}} & & \\ & (hD)^{-\frac{1}{2}} & \\ & & \Big(\frac{n}{2hp_C}\Big)^{\frac{1}{2}} \end{bmatrix},$$

so it has positive eigenvalues if and only if the middle matrix,

$$K = S_0 + \frac{n-2}{2hp_C - (n-2)e^T S_0 e} S_0 e e^T S_0,$$

is positive definite. If (20) holds, then the denominator in the scalar factor multiplying $S_0 e e^T S_0$ is positive, which implies K is positive definite. On the other hand, if K is positive definite, then $e^T K e > 0$, which implies

$$e^T S_0 e + \frac{n-2}{2hp_C - (n-2)e^T S_0 e} (e^T S_0 e)^2 > 0$$

or

$$\frac{2hp_C e^T S_0 e}{2hp_C - (n-2)e^T S_0 e} > 0,$$

from which we deduce $2hp_C - (n-2)e^T S_0 e > 0$, using the fact that $e^T S_0 e > 0$ (since S_0 is positive definite). □

In other words, the optimal scaling for p_C is necessarily different from the straight interface case, where $p = O(h^{-1/2})$. This result has the following intuitive interpretation. By eliminating the interior unknowns from (12), we obtain an equivalent iteration that is essentially a block-Jacobi method applied to the substructured problem; for such methods, one only expects convergence when the augmented system is diagonally dominant. At the cross point, the implicit part $\tilde{S}$ has weight $\frac{e^T S_0 e}{n} + hp_C$,whereas the explicit parts (B_{ij} with most of the zero rows

removed) have a combined weight of $\frac{(n-1)e^T S_0 e}{n} - hp_C$; thus, the augmented matrix would not be diagonally dominant unless p_C is large enough. For diagonal dominance, we need

$$\frac{e^T S_0 e}{n} + hp_C > \frac{(n-1)e^T S_0 e}{n} - hp_C \implies p_C > \frac{(n-2)e^T S_0 e}{2h}.$$

Remark. If we consider a three-dimensional problem $\Omega \subset \mathbb{R}^3$ decomposed into many subdomains, then there will be two types of cross points, corresponding to points on edges and corners respectively. For such problems, the preceding analysis is no longer directly applicable, since the subdomains are not all in the same plane. However, we can still expect the generalization of $C_0 = \sum_i B_i \tilde{A}^{-1}$ to play a crucial role in determining the spectral radius of the iteration matrix. In particular, we expect that one should choose different parameters for edge and corner points, and both parameters should be large enough to make the augmented system diagonally dominant. As we will see from the numerical experiments in Section 5.3, this choice will be sufficient to make the stationary iteration converge, and will also accelerate the convergence of the associated Krylov method.

4. Asymptotic Behavior of Optimal Robin Parameters

4.1. **Optimality conditions.** While Theorem 3 gives us necessary and sufficient conditions for convergence, it does not tell us how to choose D and p_C optimally and what convergence rate to expect. For optimal asymptotic convergence, we must minimize the spectral radius of N as defined in (13), or equivalently

$$\rho(N) = \max_{0 \le k \le N-1} \rho(C_k).$$

Since C_0 is the one that could potentially cause the method to diverge, we will start by choosing D and p_C to minimize $\rho(C_0)$, which is equivalent to solving the min-max problem

$$\min_{D>0, p_C>0} \max_i \left| \frac{1-\lambda_i}{1+\lambda_i} \right|, \tag{21}$$

where $\lambda_i > 0$ are the eigenvalues of $\tilde{G}$ defined in (18). This is in general a difficult problem, since there are as many parameters to choose as there are points along one edge of the interface. As a simplification, we will make the (possibly suboptimal) choice to use the same parameter $p = f(h)$ for all points on the regular interface, i.e., we will assume that $D = f(h) \cdot I$. This is not an unreasonable choice, since for a straight edge with no cross points, it can be shown that the optimal parameter is constant along the interface [21]. Note however that we reserve the right to choose a different parameter $p_C \neq f(h)$ at the cross point. We will show that there exists a choice of $f(h)$ and $p_C = p_C(h)$ that gives $\rho(C_0) = 1 - ch^{1/2} + O(h)$, a contraction factor that matches the two-subdomain case. We will then show that for this choice of $f(h)$ and p_C, we also have $\rho(C_k) = 1 - \frac{c}{2}h^{1/2} + O(h)$ for $k \neq 0$. Thus, not only does the method converge at the same asymptotic rate as the two-subdomain case, but we do not lose more than a constant by minimizing $\rho(C_0)$ instead of over all the $\rho(C_k)$. Thus, our results show that it is possible to obtain a method that converges with $\rho = 1 - O(h^{1/2})$ simply by using a different Robin parameter at the cross point, leaving the parameters on the other interface points unchanged.

For the remainder of this section, we will use the following partition of S_0 and S_0^{-1} into conforming blocks:

$$S_0 = \begin{bmatrix} S_{EE} & S_{EC} \\ S_{CE} & S_{CC} \end{bmatrix}, \qquad S_0^{-1} = \begin{bmatrix} Y_{EE} & Y_{EC} \\ Y_{CE} & Y_{CC} \end{bmatrix}. \tag{22}$$

The labels E and C correspond to degrees of freedom for the edges and the cross point respectively, so that S_{EE} and Y_{EE} is symmetric positive definite, $S_{CE} = S_{EC}^T$, $Y_{CE} = Y_{EC}^T$, and $S_{CC}, Y_{CC} > 0$ are scalar values.

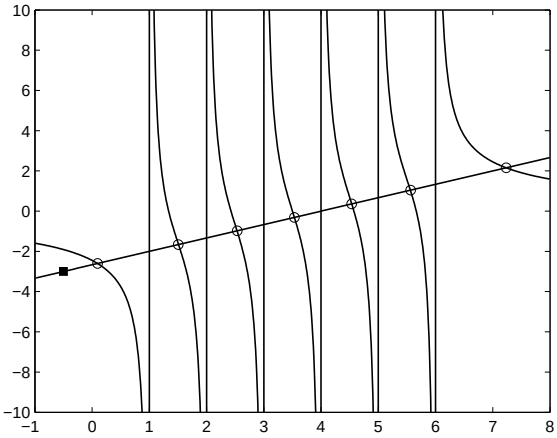


FIGURE 2. A plot of $L(\lambda)$ (the straight line) and $R(\lambda)$ (the function containing vertical asymptotes at $\theta_{1..6} = 1, 2, \ldots, 6$), for $\gamma = 1.5$, $hY_{CC} = 3$. The point $(\frac{2}{n} - 1, -hY_{CC})$ is indicated by the solid square, and the solutions to (29) occur at the intersections marked by circles.

Lemma 4. *Let $D = f \cdot I$, $f = f(h) > 0$ in the definition of $\tilde{G}$ in* (18)*. Let $\lambda_{\max}$ and $\lambda_{\min}$, which are functions of f and p_C, denote the largest and smallest eigenvalues of $\tilde{G}$. If f^* and p_C^* are solutions to the min-max problem* (21) *and $\lambda_{\max}^* = \lambda_{\max}(f^*, p_C^*)$, $\lambda_{\min}^* = \lambda_{\min}(f^*, p_C^*)$ are the optimal values, then the following two properties must hold:*

(i) $\lambda_{\min}^* \lambda_{\max}^* = 1$ *(equioscillation),*

(ii) (f^*, p_C^*) *is a minimizer of* $\kappa(f, p_C) = \dfrac{\lambda_{\max}(f, p_C)}{\lambda_{\min}(f, p_C)}$ *(condition number minimization).*

If both conditions are satisfied, then $\rho(C_0) = 1 - O(\kappa^{-1/2})$.

Proof. To show the equioscillation property, we will follow the classical argument in Wilkinson [31, p.94] to show that for any fixed p_C, the λ_i are continuous, strictly increasing functions of f, and that there exist $f_{\min}, f_{\max} > 0$ such that

$$\lambda_{\min}(f_{\min}, p_C) \cdot \lambda_{\max}(f_{\min}, p_C) < 1, \tag{23}$$

$$\lambda_{\max}(f_{\max}, p_C) \cdot \lambda_{\max}(f_{\max}, p_C) > 1. \tag{24}$$

Then the equioscillation result follows from the generic argument below: suppose for some f we have

$$\frac{1 - \lambda_{\min}}{1 + \lambda_{\min}} > -\frac{1 - \lambda_{\max}}{1 + \lambda_{\max}}, \tag{25}$$

so that $\rho = \dfrac{1 - \lambda_{\min}}{1 + \lambda_{\min}}$. Since $\lambda_{\min}$ and $\lambda_{\max}$ are strictly increasing functions of f, by choosing an $\tilde{f}$ slightly larger than f, we can obtain $\tilde{\lambda}_{\min} > \lambda_{\min}$, $\tilde{\lambda}_{\max} > \lambda_{\max}$ so that

$$-\frac{1 - \tilde{\lambda}_{\max}}{1 + \tilde{\lambda}_{\max}} < \frac{1 - \tilde{\lambda}_{\min}}{1 + \tilde{\lambda}_{\min}} < \frac{1 - \lambda_{\min}}{1 + \lambda_{\min}}.$$

Hence $\tilde{\rho} < \rho$, which means f cannot be optimal. A similar argument shows non-optimality when the inequality sign is reversed in (25), so a necessary condition for optimality is

$$\frac{1-\lambda_{\min}}{1+\lambda_{\min}} = -\frac{1-\lambda_{\max}}{1+\lambda_{\max}},$$

or, after some manipulation,

$$\lambda_{\min}\lambda_{\max} = 1, \tag{26}$$

which, in light of (23), (24) and a continuity argument, must be satisfied for some $f_{\min} < f < f_{\max}$. We now show that the λ_i are strictly increasing functions of f. Suppose S_0^{-1} has the block form (22). Then $\tilde{G}$ has the same eigenvalues as the symmetric matrix

$$\hat{G} = h\begin{bmatrix} f\cdot Y_{EE} & \sqrt{f\gamma}\cdot Y_{EC} \\ \sqrt{f\gamma}\cdot Y_{CE} & \gamma Y_{CC}\end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1-\frac{2}{n}\end{bmatrix}, \tag{27}$$

where $\gamma = 2p_C/n$. Now let Y_{EE} have the spectral decomposition $Y_{EE} = U\Theta U^T$, where $\Theta = \operatorname{diag}(\theta_1,\ldots,\theta_m)$ and $U^TU = I$. Applying the orthogonal transformation $\hat{U} = \operatorname{diag}(U,1)$ to $\hat{G}$ gives

$$\hat{U}^T\hat{G}\hat{U} = h\begin{bmatrix} f\Theta & \sqrt{f\gamma}\cdot U^TY_{EC} \\ \sqrt{f\gamma}\cdot Y_{CE}U & \gamma Y_{CC}\end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1-\frac{2}{n}\end{bmatrix}.$$

To analyze the eigenvalues of $\hat{G}$, we form the characteristic equation of $\hat{U}^T\hat{G}\hat{U}$:

$$0 = (\lambda - h\gamma Y_{CC} + (1-\frac{2}{n}))\prod_{i=1}^m(\lambda - hf\theta_i) - \sum_{i=1}^m h^2f\gamma b_i^2\prod_{j\neq i}(\lambda - hf\theta_j), \tag{28}$$

where the b_i are the components of U^TY_{EC}. To locate the ith eigenvalue λ_i, we need to consider two cases. If $b_i = 0$, then the corresponding eigenvalue is

$$\lambda_i = hf\theta_i.$$

On the other hand, if $b_i \neq 0$, then we can divide (28) by $\gamma\prod_{i=1}^m(\lambda - hf\theta_i)$ and rearrange to obtain the *secular equation*

$$\frac{\lambda + 1 - 2/n}{\gamma} - hY_{CC} = \sum_{i=1}^m \frac{h^2fb_i^2}{\lambda - hf\theta_i}, \tag{29}$$

or

$$L(\lambda) = R(\lambda).$$

Note that $L(\lambda)$ is a straight line through the point $(\frac{2}{n}-1, -hY_{CC})$ with slope $1/\gamma$, and $R(\lambda)$ has poles at $\theta_1,\ldots,\theta_m$. Furthermore, $L(\lambda)$ is independent of f and $R(\lambda)$ of γ. In Figure 2, we show the roots λ_i of (29) as the intersections between $L(\lambda)$ and $R(\lambda)$.

Regardless of whether b_i is zero, each λ_i (and in particular λ_1 and λ_n) is a strictly increasing function of f. It remains to argue that (23) and (24) hold for some $f_{\min}$ and $f_{\max}$. For $f_{\min}$, it suffices to choose f small enough so that either $\lambda_{\min}(f,\gamma) = 0$ (when $b_1 \neq 0$), or $\lambda_{\max} < 1$ (when $b_i = 0$, since $\lambda_1 = f\theta_1 > 0$ in this case). For $f_{\max}$, we can choose f large enough so that $\lambda_{\min} \geq 1/2$ and $\lambda_{\max} \geq f\theta_m > 2$ (see Figure 2).

Now for the condition number criterion, we see that when (26) holds, we have

$$\rho(C_0) = \frac{\lambda_{\max}-1}{\lambda_{\max}+1} = \frac{\lambda_{\max}^{1/2}\lambda_{\max}^{1/2} - \lambda_{\max}^{1/2}\lambda_{\min}^{1/2}}{\lambda_{\max}^{1/2}\lambda_{\max}^{1/2} + \lambda_{\max}^{1/2}\lambda_{\min}^{1/2}} = \frac{\kappa^{1/2}-1}{\kappa^{1/2}+1} = 1 - O(\kappa^{-1/2}),$$

so $\rho(C_0)$ is minimized when $\kappa = \kappa(f,p_C)$ is minimized. □

4.2. **Spectral estimates for $\tilde{G}$.** Lemma 4 gives us two properties that must be satisfied by $\lambda_{\max}(f, p_C)$ and $\lambda_{\min}(f, p_C)$ when the parameters f and p_C are chosen optimally; we now use these two criteria to determine the asymptotic behavior of f and p_C as $h \to 0$. This requires us to estimate $\lambda_{\min}$ and $\lambda_{\max}$, the extremal eigenvalues of $\tilde{G}$, which are identical to those of its symmetrized counterpart $\hat{G}$, as defined in (27). Since $\hat{G}$ is symmetric positive definite for $p_C > \frac{(n-2)e^T S_0 e}{2h}$, we have $\lambda_{\max} = \|\hat{G}\|_2$; similarly, we have $1/\lambda_{\min} = \|\hat{G}^{-1}\|_2$. In other words, we need to estimate the 2-norms of $\hat{G}$ and its inverse. Inspecting (18), we see that the various subblocks of S_0^{-1} are multiplied by different parameters (f and $2p_C/n$), which means we will need estimates for individual subblocks. Thus, our approach is to first estimate the norms of the different subblocks of S_0 and S_0^{-1}, and then estimate the norms of $\hat{G}$ and $\hat{G}^{-1}$ based on the subblock estimates. The latter task will require the following lemma.

Lemma 5. *Let* $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$ *be a symmetric positive definite matrix. Then*

$$\max\{\|M_{11}\|_2, \|M_{22}\|_2\} \le \|M\|_2 \le 2(\|M_{11}\|_2 + \|M_{22}\|_2).$$

Proof. For the lower bound, suppose x is a vector of unit length such that $\|M_{11}x\|_2 = \|M_{11}\|_2$. Then by setting $X = (x^T, 0)^T$, we have

$$\|MX\|_2 = \sqrt{\|M_{11}x\|_2^2 + \|M_{21}x\|_2^2} \ge \|M_{11}x\|_2 = \|M_{11}\|_2,$$

so that $\|M\|_2 \ge \|M_{11}\|_2$. Using the same argument on M_{22} completes the proof for the lower bound. For the upper bound, let $(x^T, y^T)^T$ be the unit eigenvector of M such that

$$M \begin{pmatrix} x \\ y \end{pmatrix} = \|M\|_2 \begin{pmatrix} x \\ y \end{pmatrix}.$$

In this case, we have

$$\begin{aligned} \|M\|_2 &= \begin{pmatrix} x^T & y^T \end{pmatrix} M \begin{pmatrix} x \\ y \end{pmatrix} \\ &\le \begin{pmatrix} x^T & y^T \end{pmatrix} M \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x^T & -y^T \end{pmatrix} M \begin{pmatrix} x \\ -y \end{pmatrix} \\ &= 2(x^T M_{11} x + y^T M_{22} y) \le 2(\|M_{11}\|_2 + \|M_{22}\|_2), \end{aligned}$$

since x and y each have norm less than or equal to 1. □

Thus, to estimate $\|\hat{G}\|_2$ and $\|\hat{G}^{-1}\|_2$ in terms of f and γ, we only need to estimate the norms of the principal subblocks of $\hat{G}$ and $\hat{G}^{-1}$ and apply Lemma 5. From Theorem 3, we know we must have $p_C > \frac{(n-2)Y_{CC}}{2h}$ for convergence, so let us write

$$p_C = \frac{(n-2)S_{CC}}{2h}(1 + g(h))$$

for some $g(h) > 0$. Then from (22), (27) and (19), we can calculate

(30)

$$\hat{G} = \begin{bmatrix} hfY_{EE} & * \\ * & \dfrac{n-2}{n}\big(S_{CC}Y_{CC}(1+g) - 1\big) \end{bmatrix}, \qquad \hat{G}^{-1} = \begin{bmatrix} \dfrac{S_{EE}}{hf} + \dfrac{S_{EC}S_{EC}^T}{hfgS_{CC}} & * \\ * & \dfrac{n}{g(n-2)} \end{bmatrix}.$$

It remains to estimate the norms of S_{EE}, S_{EC}, S_{CC}, Y_{EE} and Y_{CC}, as defined in (22). To do so, we must resort to well-known finite element Sobolev estimates and trace inequalities, which can be found in [30] and are valid for the problem $\mathcal{L} = \eta - \Delta$. In particular, we will use the following results:

Lemma 6. [30, Lemma B.5] *Let ϕ be a nodal basis function associated with a node of an element K in the triangulation of $\Omega \subset \mathbb{R}^n$. Then there exist constants independent of h_K, the diameter of K, such that*

$$c_1 h_K^n \le \|\phi\|_{L^2(K)}^2 \le C_1 h_K^n, \qquad c_2 h_K^{n-1} \le |\phi|_{H^{1/2}(K)}^2 \le C_2 h_K^{n-1}.$$

Lemma 7. [30, Lemma 4.10] *Let u_Γ be a finite element trace along the interface Γ of a subdomain Ω_i, and let S be the Schur complement of the subdomain stiffness matrix with respect to the interface. Then there exists constants c and C such that*

$$c|u_\Gamma|_{H^{1/2}(\Gamma)}^2 \le u^T S u \le C|u_\Gamma|_{H^{1/2}(\Gamma)}^2,$$

where u is the vector of degrees of freedom corresponding to the finite element trace u_Γ.

We are now ready to provide estimates of the individual subblocks of the Schur complement S_0 and those of its inverse.

Lemma 8. *For a piecewise linear finite element discretization on a shape-regular, quasi-uniform triangulation, the following estimates hold for the block decomposition (22) of S_0 as $h \to 0$:*

$$\|S_{EE}\|_2 = \Theta(1), \quad S_{CC} = \Theta(1), \quad \|S_{EC}\|_2 = O(1),$$

where we write $\varphi(h) = \Theta(\psi(h))$ whenever there exist constants c_1 and c_2 independent of h such that $c_1\psi(h) \le |\varphi(h)| \le c_2\psi(h)$. In addition, we have

$$\|Y_{EE}\|_2 = O(h^{-1}), \quad Y_{CC} = O(|\log h|).$$

Proof. For the first three estimates, let u be a vector corresponding to u_Γ, a finite element trace along the interface. Then by Lemmas 6 and 7, there exist constants C_1 and C_2 independent of h such that

$$u^T S_0 u \le C_1 |u_\Gamma|_{H^{1/2}(\Gamma)}^2 \le \frac{C_1 C_2}{h} \|u_\Gamma\|_{L^2(\Gamma)}^2.$$

However, there also exists C_3 independent of h such that

$$u^T u \ge \frac{C_3}{h} \|u_\Gamma\|_{L^2(\Gamma)}^2,$$

yielding

$$\frac{u^T S_0 u}{u^T u} \le \frac{C_1 C_2}{C_3}.$$

By choosing u to span various subblocks of S_0, we obtain

$$\|S_0\|_2 = O(1), \qquad \|S_{EE}\|_2 = O(1), \qquad S_{CC} = O(1). \tag{31}$$

To obtain lower bounds for $\|S_{EE}\|_2$ and S_{CC}, let ϕ be a nodal basis function associated with a node on Γ, and let φ be the corresponding vector representation. Then by Lemma 7 and by letting $n = 1$ in Lemma 6 (interfaces are one-dimensional), we get

$$\varphi^T S_0 \varphi \ge |\phi|_{H^{1/2}(\Gamma)}^2 \ge C, \tag{32}$$

where C is a constant independent of h. But the 2-norm of a matrix must be larger than any one of its entries, since the (i,j)th entry m_{ij} of M satisfies $|m_{ij}| = |e_i^T M e_j| \le \|M\|_2$, where e_i and e_j are unit basis vectors. Thus, (31) and (32) together show that

$$\|S_0\|_2 = \Theta(1), \qquad \|S_{EE}\|_2 = \Theta(1), \qquad S_{CC} = \Theta(1).$$

Finally, we have

$$\|S_{EC}\|_2 = \max_{u,v\neq 0} \frac{|u^T S_{EC} v|}{\|u\|_2 \|v\|_2} = \max_{u,v\neq 0} \frac{|(0,u^T) S_0 (v^T, 0)^T|}{\|u\|_2 \|v\|_2} \le \|S_0\|_2 = O(1).$$

For estimates on $\|Y_{EE}\|_2$ and Y_{CC}, we know by [30, Lemma 4.11] that the condition number of the Schur complement S_0 is bounded by

$$\kappa_2(S_0) \leq \frac{c}{h},$$

for some constant c which, together with the fact that $\|S_0\|_2 = \Theta(1)$, gives $\|S_0^{-1}\|_2 = O(h^{-1})$. To further obtain estimates for the subblocks, consider the vector u defined by

$$u = S_0^{-1}e \iff S_0 u = e,$$

where $e = (0, \ldots, 0, 1)^T$. In addition, let $U : \Omega_i \to \mathbb{R}$ be the finite element extension obtained by harmonically extending u into Ω_i with respect to $\eta - \Delta$. Then [30, Lemma 4.15] states that

$$\|U - \alpha\|^2_{L^\infty(\Omega_i)} \leq C(1 + \log(H/h))|u|^2_{H^1(\Omega_i)},$$

where α is any convex combination of values of $U(x)$ on Ω_i, H is the diameter of the subdomain Ω_i and C is a constant independent of H and h. Thus, by letting $\alpha \to 0$, we conclude that

$$Y_{CC} = e^T u \leq \|U\|_{L^\infty(\Omega_i)} \leq C(1 + |\log h|)^{1/2}|U|_{H^1(\Omega_i)},$$

where C is a constant independent of h. On the other hand, we have

$$e^T u = u^T S_0 u = \int_{\Omega_i} |\nabla U|^2 + \eta |U|^2 \, dx \geq |U|^2_{H^1(\Omega_i)}.$$

Thus,

$$|U|^2_{H^1(\Omega_i)} \leq C(1 + |\log h|)^{1/2}|U|_{H^1(\Omega_i)},$$

which implies

$$|U|_{H^1(\Omega_i)} \leq C(1 + |\log h|)^{1/2},$$

which in turn gives

$$Y_{CC} \leq C^2(1 + |\log h|) = O(|\log h|).$$

Finally, by Lemma 5, we have

$$\max\{\|Y_{EE}\|_2, Y_{CC}\} \leq \|S_0^{-1}\|_2 \leq 2(\|Y_{EE}\|_2 + Y_{CC}).$$

But since $\|S_0^{-1}\|_2 = O(h^{-1})$ and $Y_{CC} = O(|\log h|)$, we must have $\|Y_{EE}\|_2 = O(h^{-1})$, as claimed. □

Using the estimates in Lemma 8, we can put the different subblocks of $\hat{G}$ and $\hat{G}^{-1}$ in (30) together using Lemma 5 to obtain the following estimates for $\lambda_{\max}(\hat{G})$ and $\lambda_{\min}(\hat{G})$:

Lemma 9. *The following estimates hold for $h \to 0$:*

$$\lambda_{\max} = \|\hat{G}\|_2 = O(f) + O((1+g)|\log h|), \qquad \lambda^{-1}_{\min} = \|\hat{G}^{-1}\|_2 = O((1+g)/fgh) + O(1/g).$$

4.3. **Optimal choice of $f(h)$ and $g(h)$.** Since the expressions for $\lambda_{\max}$ and $\lambda_{\min}$ involve terms that scale differently in f and in g, we see that the relative sizes of f and g will lead to different cases in the analysis. It will be convenient to distinguish the cases based on the size of $f/(1+g)$. We distinguish three cases:

(i) $f/(1+g)$ is asymptotically smaller than $|\log h|$.
(ii) $f/(1+g)$ is between $|\log h|$ and $1/h$.
(iii) $f/(1+g)$ is larger than $1/h$.

In order to formalize the notion of smaller, between and larger, we will use the little-o notation: for two positive functions f_1 and f_2, we write $f_1 = o(f_2)$ if $\lim_{h\to 0} f_1/f_2 = 0$; this formalizes the notion of f_1 being "smaller" than f_2. If we exclude borderline cases where f_1/f_2 oscillates between two asymptotics, i.e., cases where $0 = \liminf f_1/f_2 \neq \limsup f_1/f_2 = C$, then the opposite statement is "f_1/f_2 is bounded away from zero", i.e., $f_1 \geq C f_2$ for some constant C. Thus, the above three cases translate to

(i) $f = o((1+g)|\log h|)$,
(ii) $f \geq c_1(1+g)|\log h|$, $f = o((1+g)/h)$,
(iii) $f \geq c_2(1+g)/h$.

Since we expect f and g to be monotonic functions of h, we will not consider the borderline cases; the three cases above cover all the remaining possibilities.

Case (i): $f = o((1+g)|\log h|)$**.** Then $\lambda_{\max} = O((1+g)|\log h|)$. In addition, since

$$\frac{1+g}{fgh} = \frac{(1+g)|\log h|}{f} \cdot \frac{1}{gh|\log h|} \gg \frac{1}{g},$$

we have $\lambda_{\min}^{-1} = O((1+g)/fgh)$. Now, the first optimality condition says $\lambda_{\max}\lambda_{\min} = 1$; thus, we get

$$(1+g)|\log h| \cong \frac{1+g}{fgh} \implies f = O\Big(\frac{1}{gh|\log h|}\Big).$$

But by assumption, $f = o((1+g)|\log h|)$, which implies

$$\frac{(1+g)|\log h|}{f} \gg 1 \implies g(1+g)h|\log h|^2 \gg 1 \implies g \gg \frac{1}{h^{\frac{1}{2}}|\log h|}.$$

This gives

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} = O((1+g)^2|\log h|^2) \gg O(h^{-1}).$$

Case (ii): $f \geq c_1(1+g)|\log h|$, $f = o((1+g)/h)$. Then $\lambda_{\max} = O(f)$ and $\lambda_{\min}^{-1} = O((1+g)/fgh)$. The first optimality condition gives $f^2 = O((1+g)/gh)$, which implies

$$(1+g)^2|\log h|^2 \leq O((1+g)/gh) \implies g(1+g)h|\log h|^2 \leq O(1) \implies g \leq O\Big(\frac{1}{\sqrt{h}|\log h|}\Big).$$

Thus, if $g = o(1)$, then $\kappa = O((1+g)/gh) \gg 1/h$, whereas if $g \geq const.$, we would get $\kappa = O(1/h)$ provided $g \leq O(h^{-\frac{1}{2}}/|\log h|)$.

Case (iii): $f \geq c_2(1+g)/h$**.** Then $\lambda_{\max} = O(f)$, $\lambda_{\min}^{-1} = O(1/g)$, so that $f = O(1/g)$. This implies

$$\frac{1}{g} \geq O((1+g)/h) \implies g(1+g) = O(h) \implies g = O(h).$$

Hence $\kappa = O(f/g) = O(h^{-2})$, which is worse than the first two cases.

Thus, to minimize $\rho(C_0)$, we must choose $f = O(h^{-1/2})$ and $g \leq O(h^{-1/2}/|\log h|)$, which yields $\rho(C_0) = 1 - O(\sqrt{h})$ by Lemma 4. We now show that for this choice of f and g, we also have $\rho(C_k) = 1 - O(\sqrt{h})$ for $1 \leq k \leq n-1$, so that the method (12) itself converges with a factor of $1 - O(\sqrt{h})$, just like the two-subdomain case.

Lemma 10. *Let* $f = c_1/\sqrt{h}$ *and* $p_C = \dfrac{(n-2)S_{CC}}{2h}(1+g(h))$ *with* $c_1 \leq g(h) \leq c_2/h^{1/2}|\log h|$, *so that* $\rho(C_0) \leq 1 - c_3\sqrt{h} + O(h)$. *Then for* $k = 1, \ldots, n-1$, *we have*

$$\rho(C_k) \leq 1 - \frac{c_3}{2}\sqrt{h} + O(h).$$

Proof. By Lemma 1, we have $\rho(C_k) \leq \rho(W)$, where W is defined in (15). Let μ_j be an eigenvalue of W. Then we have $\mu_j = \dfrac{1-\lambda_j}{1+\lambda_j}$, where λ_j is the corresponding eigenvalue of Z defined in (16),

or equivalently, that of its symmetrized version

$$\hat{Z} = \begin{bmatrix} hD & & \\ & hD & \\ & & \tau \end{bmatrix}^{1/2} \left(S_0 + \begin{bmatrix} 0 & & \\ & 0 & \\ & & \sigma \end{bmatrix} \right)^{-1} \begin{bmatrix} hD & & \\ & hD & \\ & & \tau \end{bmatrix}^{1/2},$$

where

$$\sigma = \frac{nhp_C}{2(n-1)}, \qquad \tau = \frac{(n-2)hp_C}{2(n-1)}.$$

A calculation similar to the one in section 4.2 shows that

$$\hat{Z} = \begin{bmatrix} hf(Y_{EE} - \dfrac{\sigma}{1+\sigma Y_{CC}} Y_{EC} Y_{EC}^T) & * \\ * & \dfrac{\tau Y_{CC}}{1+\sigma Y_{CC}} \end{bmatrix}, \qquad \hat{Z}^{-1} = \begin{bmatrix} \dfrac{1}{hf} S_{EE} & * \\ * & \dfrac{S_{CC}+\sigma}{\tau} \end{bmatrix}.$$

Note that $\dfrac{\sigma}{\tau} = \dfrac{n}{n-2}$ is a constant and that

$$\tau = \frac{(n-2)hp_C}{2(n-1)} = \frac{(n-2)^2 S_{CC}}{4(n-1)}(1+g) \geq const.,$$

since $g \geq 0$ and $S_{CC} = \Theta(1)$. Thus, the (2,2) blocks of both $\hat{Z}$ and $\hat{Z}^{-1}$ read

$$\hat{Z}_{22} = \frac{\tau Y_{CC}}{1+\sigma Y_{CC}} \leq \frac{\tau Y_{CC}}{\sigma Y_{CC}} = O(1), \qquad [\hat{Z}^{-1}]_{22} = \frac{S_{CC}+\sigma}{\tau} = \frac{S_{CC}}{\tau} + \frac{\sigma}{\tau} = O(1).$$

On the other hand, the norms of the (1,1) blocks satisfy

$$\|\hat{Z}_{11}\|_2 = hf \left\| Y_{EE} - \frac{\sigma}{1+\sigma Y_{CC}} Y_{EC} Y_{EC}^T \right\|_2 \leq hf \|Y_{EE}\|_2 = \|\hat{G}_{11}\|_2,$$

$$\|[\hat{Z}^{-1}]_{11}\|_2 = \left\| \frac{S_{EE}}{hf} \right\|_2 \leq \left\| \frac{S_{EE}}{hf} + \frac{S_{EC} S_{EC}^T}{hfgS_{CC}} \right\|_2 = \|[\hat{G}^{-1}]_{11}\|_2.$$

Using Lemma 5 to combine the two blocks, we get

$$\|\hat{Z}\|_2 \leq 2(\|\hat{Z}_{11}\|_2 + \hat{Z}_{22}) \leq 2\|\hat{G}\|_2 + O(1),$$
$$\|\hat{Z}^{-1}\|_2 \leq 2(\|[\hat{Z}^{-1}]_{11}\|_2 + [\hat{Z}^{-1}]_{22}) \leq 2\|\hat{G}^{-1}\|_2 + O(1).$$

Thus,

$$\rho(W) = \max\left\{ \frac{1-\lambda_{\min}(\hat{Z})}{1+\lambda_{\min}(\hat{Z})}, \frac{\lambda_{\max}(\hat{Z})-1}{\lambda_{\max}(\hat{Z})+1} \right\} = \max\left\{ \frac{\|\hat{Z}^{-1}\|_2 - 1}{\|\hat{Z}^{-1}\|_2 + 1}, \frac{\|\hat{Z}\|_2 - 1}{\|\hat{Z}\|_2 + 1} \right\}.$$

Since the function $x \mapsto \dfrac{x-1}{x+1}$ is increasing for $x > 1$, we deduce that

$$\rho(W) \leq \max\left\{ \frac{2\|\hat{G}^{-1}\|_2 - 1}{2\|\hat{G}^{-1}\|_2 + 1} + O(\|\hat{G}^{-1}\|_2^{-2}), \frac{2\|\hat{G}\|_2 - 1}{2\|\hat{G}\|_2 + 1} + O(\|\hat{G}\|_2^{-2}) \right\}$$
$$= 1 - \kappa^{-1/2}(\hat{G}) + O(\kappa^{-1}(\hat{G})) = 1 - \frac{c_3}{2}\sqrt{h} + O(h).$$

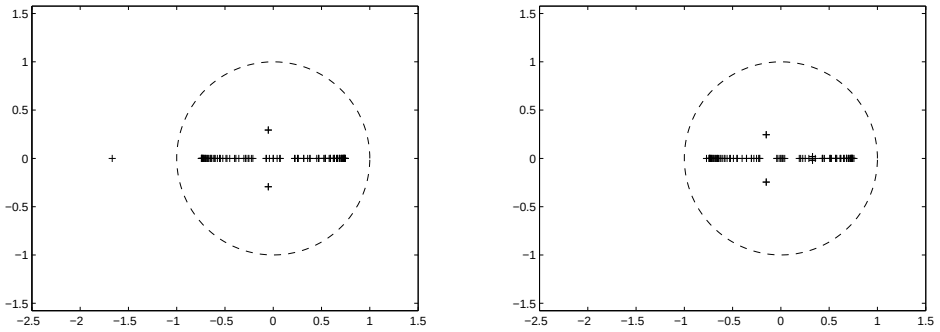
□

Combining the above lemmas, we have finally proved the second main result of this paper.

Theorem 11. *Suppose the optimized Robin parameter is chosen to be constant for the edges, i.e., we have $D = f(h) \cdot I$. Then the optimal parameters f and p_C satisfy*

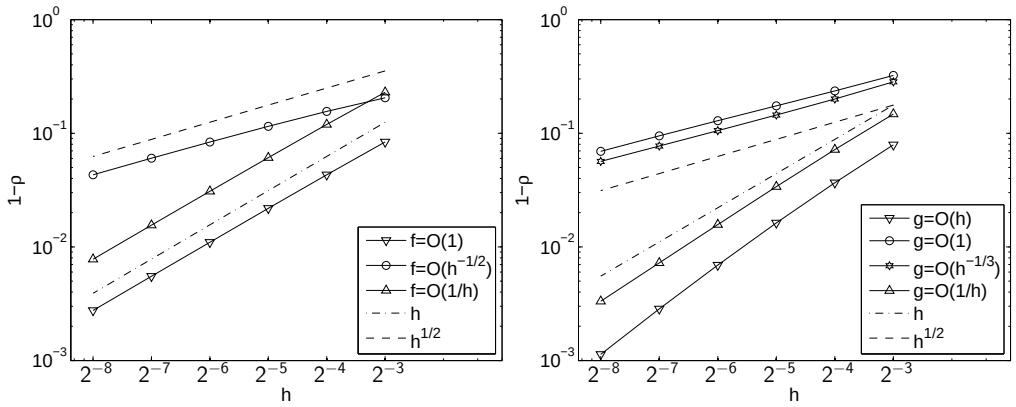
$$f = O(h^{-\frac{1}{2}}), \qquad c_1 h^{-1} \leq p_C \leq \frac{c_2}{h^{3/2}|\log h|}, \qquad c_1 > \frac{(n-2)S_{CC}}{2},$$

and the method (12) *converges at a rate of $\rho = 1 - O(\sqrt{h})$.*



(a) (b)

FIGURE 3. Spectrum of the iteration matrix for the four-subdomain problem with $h = 1/16$, $f(h) = 1.65/\sqrt{h}$ and (a) $g(h) = f(h)$, (b) $g(h) = 0.7/h$.



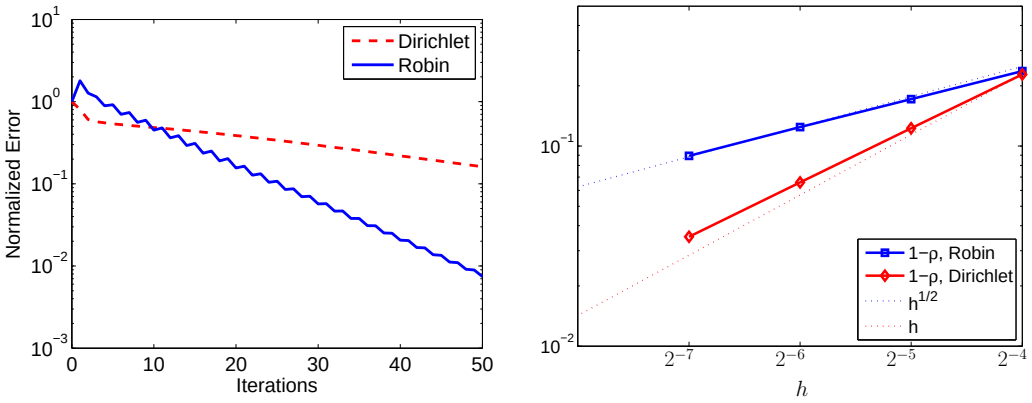
(a) (b)

FIGURE 4. Contraction factor $\rho(C_0)$ for $\mathcal{L} = -\Delta$ for four subdomains, using different scalings for $f(h)$ and $g(h)$. (a) $g(h) = 0.7/h$ fixed, varying $f(h)$, (b) fixed $f(h) = 1.65/\sqrt{h}$, varying $g(h)$.

5. Numerical experiments

5.1. **Four subdomains.** For the first set of experiments, we consider the problem of solving $-\Delta u = f$ on the unit square when it is subdivided into four subdomains, as shown in Figure 1(a). When P^1 finite elements are used, the four subdomains have identical stiffness matrices, and there is a single cross point at the center of the domain. We first verify the analysis of section 4.3 by choosing different scalings for $f(h)$ and $g(h)$. In Figure 4(a), we show the spectral radius of C_0 when we fix $g(h)$ to be equal to $0.7/h$ and vary the scaling of $f(h)$ from $O(1)$ to $O(1/h)$. We see that the best contraction rate is achieved when $f(h) = O(1/\sqrt{h})$, as predicted by the analysis; in this case, $\rho(C_0)$ scales like $1 - O(\sqrt{h})$. In the two other cases, it appears to scale like $1 - O(h)$. This happens when $f(h) = O(1)$, i.e., when we are in case (i) of the analysis; here, the condition number κ behaves as

$$\kappa = O((1+g)^2 |\log h|^2) = O(|\log h|^2/h^2),$$



(a) (b)

FIGURE 5. (a) Convergence of Dirichlet vs. Robin transmission conditions for $h = 1/128$, (b) Contraction factor $(1 - \rho)$ vs. grid parameter (h).

so that $\rho(C_0) \approx 1 - \kappa^{-1/2} = 1 - O(h/|\log h|)$. This is slightly worse than $1 - O(h)$, but hardly discernable from the plot. When $f(h) = O(1/h)$, we are in case (ii), which tells us that $\kappa = O(f^2) = O(1/h^2)$; thus, we have $\rho(C_0) = 1 - O(h)$, which is confirmed by the plot.

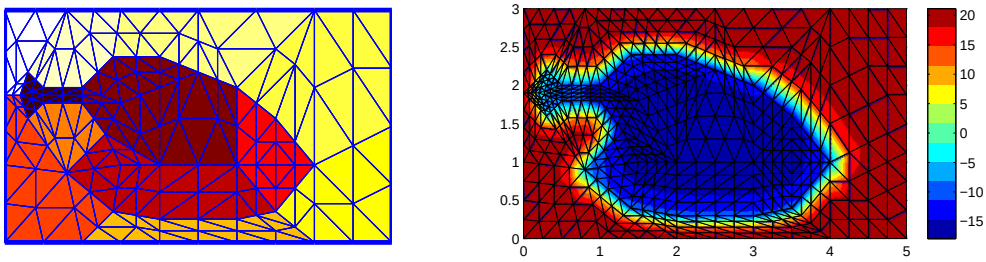
Next, in Figure 4(b), we fix $f = 1.65/\sqrt{h}$ (i.e., the optimal scaling), and vary $g(h)$ from $O(h)$ to $O(1/h)$. When $g = O(1/h)$, this puts us in case (i) of the analysis, where we have

$$\kappa = O((1+g)^2|\log h|^2) \implies \rho(C_0) = 1 - O(h/|\log h|).$$

The other choices correspond to case (ii): when $g = O(h)$ (i.e., when g is too small), we get $\kappa = O((1+g)/gh) = O(1/h^2)$, which matches the $1 - O(h)$ behavior shown in Figure 4(b). On the other hand, when g is between $O(1)$ and $O(h^{-1/2}/|\log h|)$, we get the expected $1 - O(\sqrt{h})$ scaling for $\rho(C_0)$; in fact, the curves for $g = O(1)$ and $g = O(h^{-1/3})$ have the same slope and differ by at most a (rather small) constant. This is why we are unable to deduce the precise optimal scaling for $g(h)$ using asymptotic analysis alone.

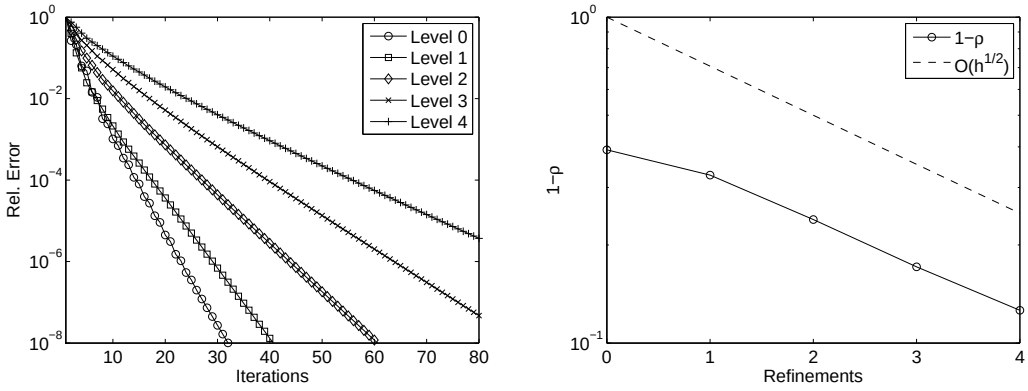
To illustrate Theorem 11, we run the optimized Schwarz method with the optimized parameters $f(h) = 1.65/\sqrt{h}$, $g(h) = 0.7/h$ and compare its convergence rate with the classical Schwarz method (with Dirichlet transmission conditions). Since classical Schwarz does not converge without overlap, we have used one layer of overlap to generate the classical Schwarz curve (even though the optimized Schwarz method still uses non-overlapping subdomains). The results in Figure 5 clearly show that convergence is faster when optimized Robin conditions are used, and the contraction factor behaves as expected under refinement, i.e., $1 - O(h)$ for Dirichlet and $1 - O(\sqrt{h})$ for Robin transmission conditions. We also see from the spectral plots in Figure 3 that if we had used the same parameter for the cross point as for the regular interface, we would get exactly one eigenvalue outside the unit circle (near -1.7 for $h = 1/16$), which means the iteration diverges. This behavior will persist for more general decompositions with cross points, see the next section.

5.2. **Multiple subdomains and cross points.** To show that our choice of parameters also leads to convergent algorithms for more general decompositions, let us consider the problem of defrosting a frozen chicken in room temperature water shown in Figure 6. The rectangular domain is divided into 12 subdomains, with the chicken occupying four subdomains and the water occupying the remaining eight. This leads to a total of 10 cross points of degree 3 or 4. We then refine this grid several times by splitting each triangular element into four smaller ones. The number of elements and degrees of freedom for each refinement level is shown in Table 1.



(a) (b)

FIGURE 6. (a) Initial grid and decomposition into subdomains for the chicken defrosting problem. (b) Temperature for the chicken and the surrounding water after 10 minutes. The computational grid is obtained by refining the initial grid once.



(a) (b)

FIGURE 7. (a) Convergence of the optimized Schwarz method for different refinement levels for the chicken defrosting problem, (b) Contraction factor as a function of the number of refinements.

On each of these grids, we solve the heat equation using backward Euler in time

$$\frac{u^{k+1} - u^k}{\delta t} = D \cdot \Delta u^{k+1},$$

where δt is the time step (always equal to 1 minute for any grid size) and D is the thermal diffusivity, which is $1.2 \times 10^{-6}\,\mathrm{m}^2/\mathrm{s}$ for the chicken and $1.4 \times 10^{-7}\,\mathrm{m}^2/\mathrm{s}$ for the surrounding

TABLE 1. Number of elements and degrees of freedom for each refinement level.

Level	# elements	# dofs
0	252	142
1	1008	535
2	4032	2077
3	16128	8185
4	64512	32497

TABLE 2. Comparison between the relative residuals of the classical and optimized Schwarz methods for the chicken problem. The cross point fix is used for optimized Schwarz to ensure convegence. "Conv." indicates that the relative residual is below 10^{-8}.

	Level 0		Level 1		Level 2		Level 3		Level 4	
Its.	Opt.	Clas.	Opt.	Clas.	Opt.	Clas.	Opt.	Clas.	Opt.	Clas.
5	4.77e-2	2.75e-2	2.45e-2	1.60e-1	9.87e-2	4.48e-1	2.03e-1	6.64e-1	3.04e-1	7.65e-1
10	1.02e-3	4.15e-4	2.13e-3	1.62e-2	1.55e-2	1.57e-1	5.18e-2	4.09e-1	1.10e-1	6.04e-1
15	8.03e-5	4.62e-6	2.62e-4	1.44e-3	3.15e-3	5.44e-2	1.58e-2	2.48e-1	4.49e-2	4.71e-1
20	4.49e-6	1.04e-7	3.64e-5	1.38e-4	7.31e-4	1.93e-2	5.19e-3	1.57e-1	1.94e-2	3.73e-1
25	3.53e-7	conv.	5.01e-6	1.20e-5	1.77e-4	6.58e-3	1.82e-3	9.55e-2	8.71e-3	2.91e-1
30	2.72e-8		6.90e-7	1.13e-6	4.36e-5	2.32e-3	6.58e-4	6.12e-2	4.04e-3	2.35e-1
35	conv.		9.49e-8	9.86e-8	1.10e-5	7.82e-4	2.44e-4	3.69e-2	1.91e-3	1.84e-1
40			1.31e-8	conv.	2.80e-6	2.74e-4	9.16e-5	2.35e-2	9.23e-4	1.49e-1
45			conv.		7.15e-7	9.20e-5	3.51e-5	1.41e-2	4.52e-4	1.17e-1
50					1.82e-7	3.21e-5	1.35e-5	8.96e-3	2.23e-4	9.51e-2
55					4.64e-8	1.08e-5	5.20e-6	5.36e-3	1.11e-4	7.45e-2
60					1.18e-8	3.75e-6	2.03e-6	3.39e-3	5.57e-5	6.04e-2
65					conv.	1.25e-6	7.92e-7	2.02e-3	2.81e-5	4.72e-2
70						4.36e-7	3.09e-7	1.28e-3	1.42e-5	3.82e-2
75						1.46e-7	1.21e-7	7.59e-4	7.25e-6	2.99e-2
80						5.06e-8	4.72e-8	4.79e-4	3.70e-6	2.41e-2

water. At each time step, we need to solve a linear system of the form $(\eta - \Delta_h)u^{k+1} = u^k$, where $\eta = \frac{1}{D\delta t}$. We know from [12] that for the homogeneous Poisson equation, the optimal Robin parameter for the two subdomain case is given by

$$p^* = ((k_{\min}^2 + \eta)(k_{\max}^2 + \eta))^{1/4}, \tag{33}$$

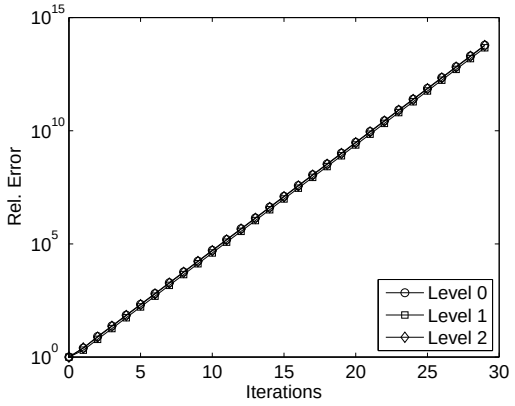
where $k_{\min}$ and $k_{\max}$ are the minimum and maximum frequencies that can be resolved by the spatial grid. For the sake of easy implementation, we have used (33) as a guideline for choosing our parameter p^* away from cross points, even though better choices are available for problems with jumps in the coefficients [8]. We calculate the optimal parameter p^* for different levels of refinement from the coarse mesh using (33); since $k_{\min} = C$ and $k_{\max} = C'/h$ for some constants C and C', we have $p^* = O(1/\sqrt{h})$ for η fixed and h small enough. For cross points, the remark after Theorem 3 tells us that we need to choose p so that the implicit part dominates; furthermore, by Theorem 11, we can choose the Robin parameter to scale like C/h, i.e., the additional weight is $h \cdot p = C$, a constant. Thus, for this experiment, we have chosen p so that the diagonal element of $\tilde{A}_i$ corresponding to the cross point is at least 3/4 of the corresponding element in the global stiffness matrix. The results shown in Figure 7 and Table 2 confirm that this indeed gives a convergent method for any refinement level, and the contraction factor indeed scales like $1 - O(\sqrt{h})$, as expected; thus, optimized Schwarz outperforms classical Schwarz, especially for higher refinement levels. We conclude that even though the analysis required fairly stringent symmetry conditions, we see that the same conclusions hold in much more general settings.

We now examine what would happen if we had used the same Robin parameter everywhere (including cross points). Figure 8(a) and Table 3 show that the method indeed diverges; moreover, the method diverges at the same rate for all refinement levels. In Figure 8(b), we plot the solution after 10 iterations. We see that the solution diverges most quickly at cross points, and that cross points of degree 4 cause faster divergence than cross points of lower degree. All this indicates that the culprit for divergence is indeed the presence of large eigenvalues in the iteration matrix N associated with cross points.

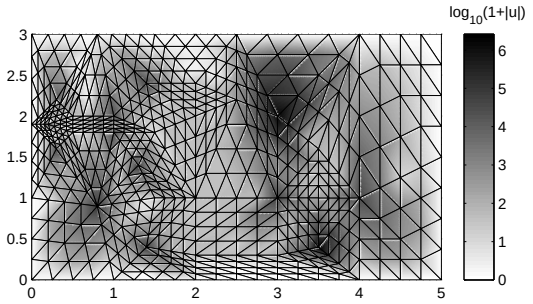
We finally consider what happens when optimized Schwarz is used as a preconditioner within a Krylov subspace method. Figure 9 and Table 4 show the convergence of GMRES both with

TABLE 3. Relative error of optimized Schwarz without the cross point fix.

Its.	Level 0	Level 1	Level 2
5	7.14e+01	5.46e+01	6.82e+01
10	1.73e+04	1.33e+04	1.66e+04
15	4.21e+06	3.22e+06	4.03e+06
20	1.02e+09	7.83e+08	9.78e+08
25	2.49e+11	1.90e+11	2.38e+11
30	6.04e+13	4.62e+13	5.78e+13



(a)



(b)

FIGURE 8. Divergence of the optimized Schwarz method without the cross point fix. The map shows the solution after 10 iterations. The colors are shown in logarithmic scale; darker colors indicate larger values.

TABLE 4. Comparison between optimized Schwarz preconditioned GMRES for the chicken problem, with and without the cross point fix.

	Level 0		Level 1		Level 2		Level 3		Level 4	
Its.	with fix	no fix	with fix	no fix	with fix	no fix	with fix	no fix	with fix	no fix
5	6.05e-3	9.32e-3	3.96e-3	7.51e-3	1.01e-2	9.60e-3	1.59e-2	1.45e-2	1.94e-2	1.91e-2
10	3.28e-5	7.07e-5	5.04e-5	6.62e-5	1.63e-4	1.72e-4	3.18e-4	3.14e-4	6.17e-4	6.80e-4
15	1.22e-7	6.09e-7	7.49e-7	8.96e-7	5.96e-6	6.50e-6	2.18e-5	1.91e-5	6.39e-5	7.05e-5
20	conv.	conv.	6.00e-9	9.95e-9	1.15e-7	1.44e-7	1.67e-6	1.34e-6	6.87e-6	7.31e-6
25			conv.	6.83e-9	6.83e-9	7.71e-9	7.97e-8	6.40e-8	7.21e-7	7.03e-7
30				conv.	conv.	conv.	conv.	conv.	7.93e-8	6.26e-8

and without the cross point fix. Both versions benefit from Krylov acceleration to the same degree, and the similarity of the two plots shows that the cross point fix simply moves the large eigenvalues back into the unit circle, without adversely affecting the rest of the spectrum. One also sees that for 2D problems, the cross point fix is only important for the stationary case; it is not necessary when GMRES is used, since the Krylov method takes care of these outlying eigenvalues automatically. As we will see in the next section, this is not the case for 3D problems, where there will be cross points corresponding to edges, and their corresponding eigenvalues will form non-trivial clusters in the spectrum.

5.3. **3D Example.** We now consider a three-dimensional example in which the cube $[-1,1]^3 \subset \mathbb{R}^3$ is decomposed into 8 smaller cubes meeting at a single cross point at the origin. As mentioned

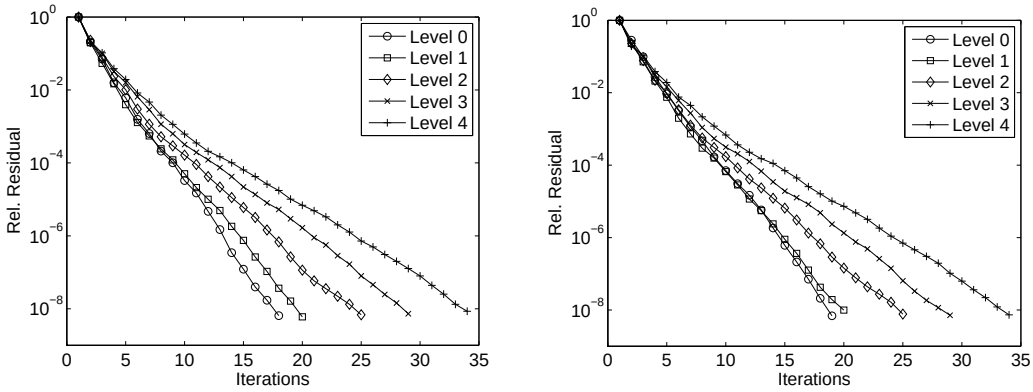


FIGURE 9. (a) Convergence of GMRES with the optimized Schwarz preconditioner for the chicken problem: (a) with the cross point fix, (b) without the cross point fix.

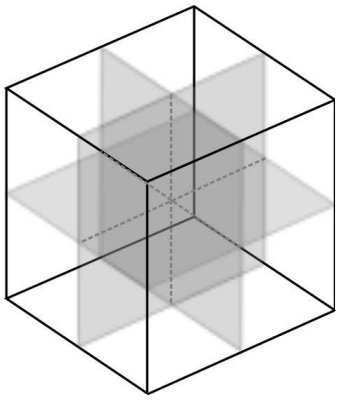


FIGURE 10. A decomposition of a cube into 8 equal subdomains.

in the remark at the end of section 3, we expect to choose different Robin parameters for the faces, edges and corners. For faces, we choose the usual scaling $p = 1/\sqrt{h}$, whereas for edges and faces, we use the same heuristic as the chicken problem, meaning we will choose $p = O(1/h)$ in such a way that the diagonal element of $\tilde{A}_i$ corresponding to the cross point is at least $3/4$ of the corresponding element in the global stiffness matrix. Since the diagonal element in the standard 7-point discretization is 6 for the interior, 3 for the faces, $3/2$ for the edges and $3/4$ for the corner, it suffices to choose $p = 3/h$ for edges and $p = 4/h$ for the corner. We plot the spectrum of the iteration matrix for $h = 1/6$ in Figure 11(b). As a comparison, we also plot the spectrum we would have obtained if we had used $p = 1/\sqrt{h}$ for the edges and the corner. We see that if we had used $p = 1/\sqrt{h}$ everywhere, we would obtain three clusters of eigenvalues corresponding to the faces (inside the unit circle), the edges (the cluster around -2) and the single corner at -4. The presence of the cluster around -2 is especially problematic for GMRES, since it must now spend several iterations removing these components. As we see in Figure 12 and Table 5, GMRES is indeed slowed down by this cluster; it takes about 10 more iterations than the version with the cross point fix to achieve the same relative residual, especially for finer

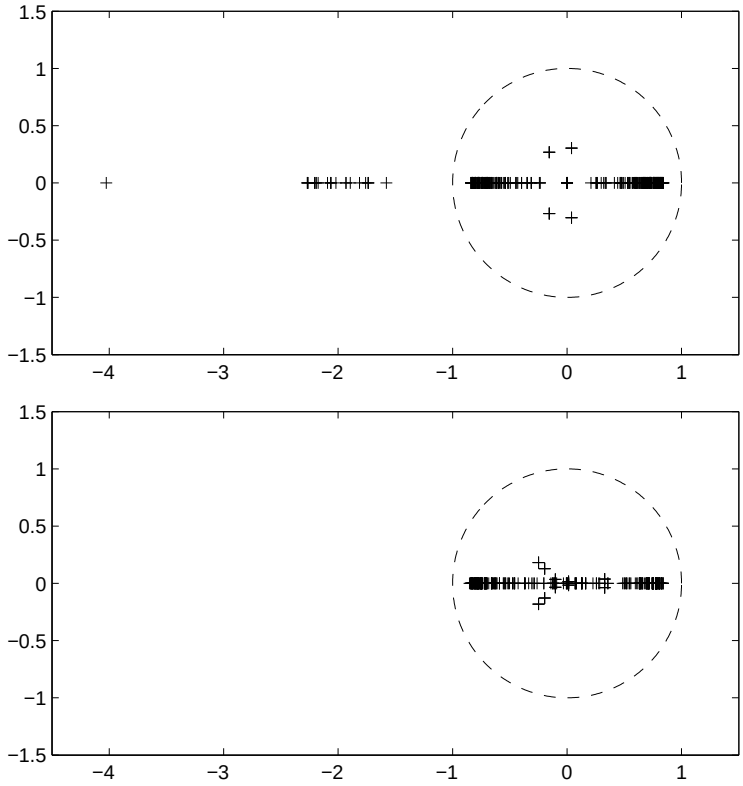


FIGURE 11. Spectrum of the Optimized Schwarz iteration matrix for a $2\times2\times2$ decomposition of the cube. Top: $p=\sqrt{h}$ for all interface points, including cross points. Bottom: $p=\sqrt{h}$ for the regular interface and $p=O(1/h)$ for cross points.

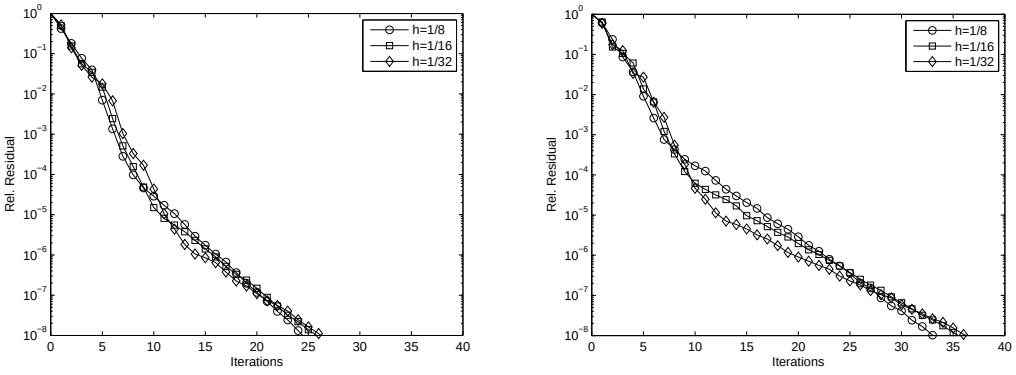
TABLE 5. Comparison between GMRES with the optimized Schwarz preconditioner for the 3D problem: (a) with the cross point fix, (b) without the cross point fix.

	$h=1/8$		$h=1/16$		$h=1/32$	
Its.	with fix	no fix	with fix	no fix	with fix	no fix
5	3.96e-002	3.71e-002	3.46e-002	6.14e-002	2.62e-002	3.35e-002
10	4.63e-005	2.44e-004	4.84e-005	1.21e-004	1.70e-004	1.82e-004
15	2.92e-006	2.98e-005	2.30e-006	1.69e-005	1.06e-006	5.80e-006
20	1.98e-007	4.43e-006	2.36e-007	2.84e-006	1.69e-007	1.18e-006
25	1.32e-008	5.39e-007	2.19e-008	5.44e-007	2.45e-008	3.02e-007
30	conv.	5.48e-008	conv.	9.17e-008	conv.	8.68e-008
35		conv.		1.75e-008		2.13e-008

grids. This example shows the cross point fix is not only of theoretical interest, but can really accelerate convergence of Krylov methods at no extra cost, especially for 3D problem.

REFERENCES

[1] Philippe Charton, Frédéric Nataf, and Francois Rogier. Méthode de décomposition de domaine pour l'équation d'advection-diffusion. *C. R. Acad. Sci.*, 313(9):623–626, 1991.



(a) (b)

FIGURE 12. (a) Convergence of GMRES with the optimized Schwarz preconditioner for the 3D problem: (a) with the cross point fix, (b) without the cross point fix.

[2] Philippe Chevalier and Frédéric Nataf. Symmetrized method with optimized second-order conditions for the Helmholtz equation. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, pages 400–407. Amer. Math. Soc., Providence, RI, 1998.

[3] C. Chniti, F. Nataf, and F. Nier. Improved interface conditions for 2D domain decomposition with corners: a theoretical determination. *Calcolo*, 45:111–147, 2008.

[4] C. Chniti, F. Nataf, and F. Nier. Improved interface conditions for 2D domain decomposition with corners:numerical applications. *Journal of Scientific Computing*, 38:207–228, 2009.

[5] Q. Deng. An analysis for a nonoverlapping domain decomposition iterative procedure. *SIAM J. Sci. Comput.*, 18:1517–1525, 1997.

[6] Bruno Després. *Méthodes de décomposition de domaines pour les problèmes de propagation d'ondes en régime harmonique*. PhD thesis, Univ. Paris IX Dauphine, 1991.

[7] Bruno Després, Patrick Joly, and Jean E. Roberts. A domain decomposition method for the harmonic Maxwell equations. In *Iterative methods in linear algebra (Brussels, 1991)*, pages 475–484. North-Holland, Amsterdam, 1992.

[8] O. Dubois and S. H. Lui. Convergence estimates for an optimized Schwarz method for PDEs with discontinuous coefficients. *Numer Algor*, 51:115131, 2009.

[9] Evridiki Efstathiou and Martin J. Gander. Why restricted additive Schwarz converges faster than additive Schwarz. *BIT*, 43(5):945–959, 2003.

[10] Bjorn Engquist and Hong-Kai Zhao. Absorbing boundary conditions for domain decomposition. *Appl. Numer. Math.*, 27(4):341–365, 1998.

[11] M. J. Gander and F. Kwok. On the applicability of lions' energy estimates in the analysis of discrete optimized Schwarz methods with cross points. In *Domain Decomposition Methods in Science and Engineering XX (submitted)*, 2011.

[12] Martin J. Gander. Optimized Schwarz methods. *SIAM J. Numer. Anal.*, 44(2):699–732, 2006.

[13] Martin J. Gander. Schwarz methods in the course of time. *ETNA*, 31:228–255, 2008.

[14] Martin J. Gander, Laurence Halpern, and Frédéric Nataf. Optimized Schwarz methods. In Tony Chan, Takashi Kako, Hideo Kawarada, and Olivier Pironneau, editors, *Twelfth International Conference on Domain Decomposition Methods, Chiba, Japan*, pages 15–28, Bergen, 2001. Domain Decomposition Press.

[15] Martin J. Gander, Laurence Halpern, and Frédéric Nataf. Optimized Schwarz waveform relaxation for the one dimensional wave equation. *SIAM J. Numer. Anal.*, 41(5):1643–1681, 2003.

[16] Thomas Hagstrom, R. P. Tewarson, and Aron Jazcilevich. Numerical experiments on a domain decomposition algorithm for nonlinear elliptic boundary value problems. *Appl. Math. Lett.*, 1(3), 1988.

[17] C. Japhet and Frédéric Nataf. The best interface conditions for domain decomposition methods: Absorbing boundary conditions. In *Absorbing Boundary and Layers, Domain Decomposition Methods*, pages 348–373. Nova Sci. Publ., 2001.

[18] Caroline Japhet. Optimized Krylov-Ventcell method. Application to convection-diffusion problems. In Petter E. Bjørstad, Magne S. Espedal, and David E. Keyes, editors, *Proceedings of the 9th international conference on domain decomposition methods*, pages 382–389. ddm.org, 1998.

[19] Caroline Japhet, Frédéric Nataf, and Francois Rogier. The optimized order 2 method. application to convection-diffusion problems. *Future Generation Computer Systems FUTURE*, 18, 2001.
[20] Caroline Japhet, Frédéric Nataf, and Francois-Xavier Roux. The Optimized Order 2 Method with a coarse grid preconditioner. application to convection-diffusion problems. In P. Bjorstad, M. Espedal, and D. Keyes, editors, *Ninth International Conference on Domain Decompositon Methods in Science and Engineering*, pages 382–389. John Wiley & Sons, 1998.
[21] Felix Kwok. A note on optimal Robin parameters for two-subdomain problems. *in preparation*, 2011.
[22] Pierre-Louis Lions. On the Schwarz alternating method III: a variant for non-overlapping subdomains. In T.F Chan, R. Glowinski, J. Périaux, and O. Widlund, editors, *Third international symposium on domain decomposition methods for partial differential equations*, pages 47–70, Philadelphia, 1990. SIAM.
[23] Sébastien Loisel. Condition number estimates for the nonoverlapping optimized Schwarz method and the 2-Lagrange multiplier method for general domains and cross points. *submitted*, 2010.
[24] F. Nataf and F. Rogier. Factorization of the convection-diffusion operator and the Schwarz algorithm. *Math. Models Methods Appl. Sci.*, 5(1):67–93, 1995.
[25] F. Nataf, F. Rogier, and E. De Sturler. Optimal interface conditions for domain decomposition methods. Technical report, CMAP, Ecole Polytechnique, Paris, 1994.
[26] Frédéric Nataf. Absorbing boundary conditions in block Gauss-Seidel methods for convection problems. *Math. Models Methods Appl. Sci.*, 6(4):481–502, 1996.
[27] H. San and W. P. Tang. An overdetermined Schwarz alternating method. *SIAM J. Sci. Comput.*, 7:884–905, 1996.
[28] Amik St-Cyr, Duane Rosenberg, and Sang Dong Kim. Optimized Schwarz preconditioning for SEM based magnetohydrodynamics. In Michel Bercovier, Martin J. Gander, Ralf Kornhuber, and Olof Widlund, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, Lecture notes in Computational Science and Engineering 70. Springer-Verlag, 2009.
[29] Wei Pai Tang. Generalized Schwarz splittings. *SIAM J. Sci. Stat. Comp.*, 13(2):573–595, 1992.
[30] Andrea Toselli and Olof B. Widlund. *Domain Decomposition Methods — Algorithms and Theory*, volume 34 of *Springer Series in Computational Mathematics*. Springer, Berlin Heidelberg, 2005.
[31] James Hardy Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, 1965.