SIAM J. NUMER. ANAL.
Vol. 49, No. 3, pp. 1289–1316

OPTIMIZED ADDITIVE SCHWARZ WITH HARMONIC EXTENSION AS A DISCRETIZATION OF THE CONTINUOUS PARALLEL SCHWARZ METHOD*

FELIX KWOK†

Abstract. The additive Schwarz method with harmonic extension (ASH) was introduced by Cai and Sarkis (1999) as an efficient variant of the additive Schwarz method that converges faster and requires less communication. We show that ASH can also be used with optimized transmission conditions to obtain faster convergence. We show that when the decomposition into subdomains contains no cross points, optimized ASH can be reformulated as an iteration that is closely related to the optimized Schwarz method at the continuous level. In fact, the iterates of ASH are identical to the iterates of the discretized parallel Schwarz method outside the overlap, whereas inside the overlap they are linear combinations of previous Schwarz iterates. Thus, one method converges if and only if the other one does, and they do so at the same asymptotic rate, unlike additive Schwarz, which fails to converge inside the overlap. However, when cross points are present, then ASH and the Schwarz methods are incomparable, i.e., there are cases where one method converges and the other diverges, and vice versa.

Key words. domain decomposition, optimized Schwarz methods, additive Schwarz method

AMS subject classifications. 65N55, 65F10, 65N22

DOI. 10.1137/090781632

1. Introduction.

The convergence properties of optimized Schwarz methods, which are Schwarz methods with modified transmission conditions between subdomains, have been studied for a variety of problems; for a nonexhaustive list, see [4, 15, 11, 12, 1] and the references therein. In these works, the convergence rate and optimal parameters are usually derived in the continuous setting, where it is possible to use tools such as Fourier analysis and energy methods. However, the analysis is less straightforward for discrete methods such as additive Schwarz, because there may no longer be a direct correspondence between the discrete iterates and the subdomain solutions in the continuous setting. In fact, it has been shown [9, 13] that additive Schwarz does not converge as an iterative method when an overlapping decomposition is used, even when the corresponding parallel Schwarz method does, and the same difficulties arise when optimized transmission conditions are used. However, if the discrete method can be interpreted as the discretization of the underlying continuous Schwarz method, it would then be possible to estimate the convergence rate using the continuous results, at least when the mesh is fine enough. For the restricted additive Schwarz method, RAS, (defined below), such an interpretation is given in [13]. The goal of this paper is to offer a similar interpretation for a related method, called the additive Schwarz with harmonic extension (ASH), when optimized transmission conditions are used. Once we establish the equivalence between optimized ASH and the continuous optimized Schwarz method, its behavior as a preconditioner can immediately be inferred from the spectral properties of the continuous method. To the best of our knowledge, the optimized ASH method has not been defined nor analyzed in the literature.

*Received by the editors January 4, 2010; accepted for publication (in revised form) March 31, 2011; published electronically June 23, 2011.
http://www.siam.org/journals/sinum/49-3/78163.html
†Department of Mathematics, University of Geneva, 1211 Geneva 4, Switzerland (felix.kwok@unige.ch).

1.1. Lions' method. Let $\Omega \subset\subset \mathbb{R}$ be an open set. Suppose we want to solve the elliptic PDE

$$\mathcal{L}u = f \qquad \text{on } \Omega, \qquad u = g \qquad \text{on } \partial\Omega. \tag{1}$$

Based on the theoretical work of Schwarz [20], Lions introduced in [17] the first domain decomposition methods for solving (1). In the two-subdomain case, let $\Omega_1, \Omega_2 \subset \Omega$ such that $\Omega_1 \cup \Omega_2 = \Omega$ and $\Omega_1 \cap \Omega_2 \neq \emptyset$. Also define

$$\Gamma_1 = \partial\Omega \cap \bar{\Omega}_1, \quad \Gamma_1 = \partial\Omega \cap \bar{\Omega}_2; \qquad \Gamma_{12} = \partial\Omega_1 \cap \bar{\Omega}_2, \quad \Gamma_{21} = \partial\Omega_2 \cap \bar{\Omega}_1.$$

Then Lions' *parallel Schwarz method* defines the subdomain iterates

$$u_1^k : \Omega_1 \to \mathbb{R}, \qquad u_2^k : \Omega_2 \to \mathbb{R}, \qquad k \geq 0,$$

and the method can be written as

$$\begin{aligned} \mathcal{L}u_1^{k+1} &= f && \text{on } \Omega_1, & \mathcal{L}u_2^{k+1} &= f && \text{on } \Omega_2, \\ u_1^{k+1} &= g && \text{on } \Gamma_1, & u_2^{k+1} &= g && \text{on } \Gamma_2, \\ u_1^{k+1} &= u_2^k && \text{on } \Gamma_{12}, & u_2^{k+1} &= u_1^k && \text{on } \Gamma_{21}. \end{aligned} \tag{2}$$

To describe the discretized version of the above method, we introduce some notation. Let R_i be the operator that restricts the set $V = \{1, \ldots, N\}$ of all nodes onto the subset V_i of nodes that lie in Ω_i. Then the discretized parallel Schwarz method becomes

$$\begin{aligned} A_1\mathbf{u}_1^{k+1} &= f_1 - A_{12}\mathbf{u}_2^k, \\ A_2\mathbf{u}_2^{k+1} &= f_2 - A_{21}\mathbf{u}_1^k, \end{aligned}$$

where for $i = 1, 2$,

$$A_i = R_i A R_i^T, \qquad A_{ij} = (R_i A - A_i R_i) R_j^T. \tag{3}$$

The above method trivially generalizes to the case of many subdomains if there are no cross points, i.e., $\Omega_i \cap \Omega_j \cap \Omega_l = \emptyset$ for distinct i, j, and l. The discretized algorithm then becomes

$$A_i\mathbf{u}_i^{k+1} = f_i - \sum_{j \neq i} A_{ij}\mathbf{u}_j^k, \qquad \text{for all } i, \tag{4}$$

with the same definition of A_i and A_{ij} as in (3) extended for all i.

1.2. Optimized parallel Schwarz method. Convergence of Lions' parallel Schwarz method can be improved by introducing *optimized transmission conditions* [12]. In the continuous setting, one can obtain optimized Schwarz methods from (2) by simply replacing the Dirichlet boundary conditions along internal boundaries with more general boundary operators $\mathcal{B}_{ij}$:

$$\begin{aligned} \mathcal{L}u_1^{k+1} &= f && \text{on } \Omega_1, & \mathcal{L}u_2^{k+1} &= f && \text{on } \Omega_2, \\ u_1^{k+1} &= g && \text{on } \Gamma_1, & u_2^{k+1} &= g && \text{on } \Gamma_2, \\ \mathcal{B}_{12}u_1^{k+1} &= \mathcal{B}_{12}u_2^k && \text{on } \Gamma_{12}, & \mathcal{B}_{21}u_2^{k+1} &= \mathcal{B}_{21}u_1^k && \text{on } \Gamma_{21}. \end{aligned} \tag{5}$$

A change in transmission conditions corresponds to changing the operator A_i into $\tilde{A}_i$, i.e., $\tilde{A}_i := A_i + L_i$, where L_i represents a boundary operator that maps traces along $\partial\Omega_i \setminus \partial\Omega$ into itself. If there are no cross points, the discretized version of (4) becomes

$$\tilde{A}_i \mathbf{u}_i^{k+1} = f_i + \sum_{j \neq i} (L_i R_i R_j^T - A_{ij}) \mathbf{u}_j^k \qquad \text{for all } i. \tag{6}$$

Note that both methods (2) and (6) work exclusively on subdomain solutions $\mathbf{u}_j^k$; there is no built-in notion of a global approximation that is valid over the entire domain Ω. In particular, if the subdomains overlap, there is no unique way of defining the global approximations $\mathbf{U}^k$ before the methods converge. Thus, one cannot directly consider parallel Schwarz as a preconditioner for the global system and use it in combination with Krylov subspace methods.

1.3. The methods of additive Schwarz, RAS, and ASH. In order to turn parallel Schwarz into a preconditioner, Dryja and Widlund [8] introduced the *additive Schwarz* method. Starting from an initial guess of the *global* solution $\mathbf{U}^0$, the method calculates successive iterates using

$$\mathbf{U}^{k+1} = \mathbf{U}^k + \sum_j R_j^T A_j^{-1} R_j (f - A\mathbf{U}^k). \tag{7}$$

When the subsets V_j are disjoint, additive Schwarz (AS) is equivalent to a block Jacobi iteration. However, when the subdomains overlap, the method no longer converges inside the overlap [9, 13]. This is because the overlap receives updates from several subdomain solves, leading to a redundancy that prevents convergence of the method. One way of eliminating this redundancy is to use the methods of *restricted additive Schwarz* (RAS) and *additive Schwarz with harmonic extension* (ASH), which have been introduced by Cai and Sarkis [3] as efficient variants of AS. Let $\tilde{\Omega}_j$ be a partition of Ω such that $\tilde{\Omega}_j \subset \Omega_j$. Let $\tilde{V}_j$ be the nodes that lie in $\tilde{\Omega}_j$, and let $\tilde{R}_l$ be a matrix of the same size as R_l such that

$$[\tilde{R}_l]_{ij} = \begin{cases} \delta_{ij} & \text{if } j \in \tilde{V}_l, \\ 0 & \text{otherwise.} \end{cases}$$

Then RAS is defined by

$$\mathbf{U}^{k+1} = \mathbf{U}^k + \sum_j \tilde{R}_j^T A_j^{-1} R_j (f - A\mathbf{U}^k), \tag{8}$$

whereas ASH is defined by

$$\mathbf{U}^{k+1} = \mathbf{U}^k + \sum_j R_j^T A_j^{-1} \tilde{R}_j (f - A\mathbf{U}^k). \tag{9}$$

By restricting either the residual or the update onto $\tilde{V}_j$, RAS and ASH avoid the redundant updates that occur within the overlap when AS is used. There exist other methods capable of eliminating the nonconverging modes in AS, such as the method of restricted additive Schwarz with harmonic overlap (RASHO), which was proposed by [2]. The idea behind RASHO is to construct a symmetric preconditioner (which would then be amenable to conjugate gradients or MINRES, unlike RAS and ASH) by finding a projector that would eliminate the nonconvergent modes from the initial error. A full discussion of this method is beyond the scope of this paper, as we will concentrate on analyzing ASH and its optimized counterpart.

Optimized versions of AS and RAS are obtained by replacing each A_j with $\tilde{A}_j$ in (7) and (8); such methods have been analyzed in [13]. It is clear that the RAS and ASH preconditioners are transposes of each other when A is symmetric; one thus expects the two methods to converge at a similar rate. In the case where A is an M-matrix, the authors of [10] proved that RAS and ASH both converge as an iterative method. For the RAS method, it has been proved [13] that the iterates produced are equivalent to those of the discretized parallel Schwarz method, regardless of the number of subdomains and whether cross points are present. In the case of ASH with classical (Dirichlet) transmission conditions, such an interpretation has been shown in [16].

We note that RAS, ASH, and related methods are often used in combination with a Krylov subspace method for nonsymmetric matrices, such as GMRES [19]; such combinations are among the most efficient parallel iterative methods available for general discretizations. In this paper, our analysis deals mainly with ASH as a stationary iterative method. However, our results are also relevant for understanding the behavior of ASH-GMRES. Since GMRES finds the solution that minimizes the residual over Krylov subspaces, k steps of left-preconditioned ASH-GMRES will always converge faster than k steps of stationary ASH, which, in turn, converges at the same asymptotic rate as (unaccelerated) parallel Schwarz. In addition, knowledge of the spectral radius (and hence eigenvalues) of $I - M^{-1}A$ can be used to derive convergence estimates using complex Chebyshev polynomials [18, Chap. 6] or potential theory [7].

The remainder of this paper is organized as follows. In section 2, we illustrate the type of arguments used to obtain equivalence using a concrete example. In section 3, we state the algebraic conditions that ensure there are no cross points, and then state the main equivalence results. Section 4 is devoted to the proof of the main result. In section 5, we use this equivalence to show that when one of the two methods (optimized ASH and optimized parallel Schwarz) converges, so does the other one, and their asymptotic convergence rates are identical. We finally give numerical examples showing the equivalence of both methods, including one where a system of PDEs is solved. This shows the usefulness and generality of the algebraic conditions, since they carry over trivially to the systems case.

2. An example.

2.1. Two subdomains.

To illustrate the ideas used in the proof, consider the two-subdomain decomposition shown in Figure 1. Assume the initial guess is $\mathbf{U}^0 = 0$. Then, at the first iteration ($k = 1$), ASH solves the following system:

$$\left.\begin{aligned} A_{11}u_1^{1,a} + A_{12}u_2^{1,a} \qquad\qquad\qquad &= f_1 \\ A_{21}u_1^{1,a} + A_{22}u_2^{1,a} + A_{23}u_3^{1,a} \qquad &= f_2 \\ A_{32}u_2^{1,a} + A_{33}u_3^{1,a} \qquad &= 0 \end{aligned}\right\} \quad \text{in } \Omega_a, \tag{10a}$$

$$\left.\begin{aligned} A_{22}u_2^{1,b} + A_{23}u_3^{1,b} \qquad\qquad &= 0 \\ A_{32}u_2^{1,b} + A_{33}u_3^{1,b} + A_{34}u_4^{1,b} &= f_3 \\ A_{43}u_3^{1,b} + A_{44}u_4^{1,b} &= f_4 \end{aligned}\right\} \quad \text{in } \Omega_b. \tag{10b}$$

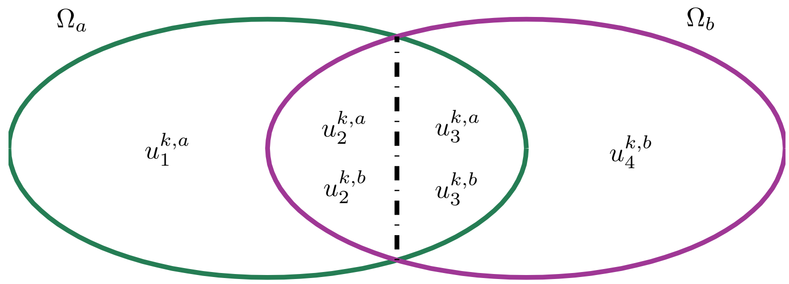


FIG. 1. *A decomposition into two subdomains Ω_a and Ω_b. The subdomain solution $\mathbf{u}_a^k$ contains the subvectors $u_1^{k,a}$, $u_2^{k,a}$, and $u_3^{k,a}$, whereas $\mathbf{u}_b^k$ contains $u_2^{k,b}$, $u_3^{k,b}$, and $u_4^{k,b}$ as components.*

Using the definition of the global solution $\mathbf{U}^1$, which is

$$U_1^1 = u_1^{1,a}, \quad U_2^1 = u_2^{1,a} + u_2^{1,b}, \quad U_3^1 = u_3^{1,a} + u_3^{1,b}, \quad U_4^1 = u_4^{1,b},$$

we calculate the residual $\mathbf{R}^1 = f - A\mathbf{U}^1$:

$$\mathbf{R}^1 = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} - \begin{bmatrix} A_{11}u_1^{1,a} + A_{12}(u_2^{1,a} + u_2^{1,b}) \\ A_{21}u_1^{1,a} + A_{22}(u_2^{1,a} + u_2^{1,b}) + A_{23}(u_3^{1,a} + u_3^{1,b}) \\ A_{32}(u_2^{1,a} + u_2^{1,b}) + A_{33}(u_3^{1,a} + u_3^{1,b}) + A_{34}u_4^{1,b} \\ A_{43}(u_3^{1,a} + u_3^{1,b}) + A_{44}u_4^{1,b} \end{bmatrix} = \begin{bmatrix} -A_{12}u_2^{1,b} \\ 0 \\ 0 \\ -A_{43}u_3^{1,a} \end{bmatrix}.$$

Note that the residual vanishes inside the overlap. At the second iteration ($k = 2$), the global solution is given by

$$\mathbf{U}^2 = \mathbf{U}^1 + \delta\mathbf{u}^{2,a} + \delta\mathbf{u}^{2,b},$$

where

$$\begin{aligned} A_{11}\delta u_1^{2,a} + A_{12}\delta u_2^{2,a} \qquad\qquad &= R_1^1 = -A_{12}u_2^{1,b}, \\ A_{21}\delta u_1^{2,a} + A_{22}\delta u_2^{2,a} + A_{23}\delta u_3^{2,a} &= R_2^2 = 0, \\ A_{32}\delta u_2^{2,a} + A_{33}\delta u_3^{2,a} &= 0. \end{aligned}$$

Adding (10a) to the above equations gives

$$A_{11}u_1^{2,a} + A_{12}u_2^{2,a} = f_1 - A_{12}u_2^{1,b}, \tag{11a}$$

$$A_{21}u_1^{2,a} + A_{22}u_2^{2,a} + A_{23}u_3^{2,a} = f_2, \tag{11b}$$

$$A_{32}u_2^{2,a} + A_{33}u_3^{2,a} = 0. \tag{11c}$$

$\mathbf{u}^{2,a}$ by itself cannot be interpreted as the solution of a continuous subdomain solve because of the term $A_{12}u_2^{1,b}$ in the first equation. To obtain such an interpretation, we will first move the term $A_{12}u_2^{1,b}$ in (11a) to the left-hand side; then we can add the first two equations in (10b) to (11b) and (11c) to get

$$\begin{aligned} A_{11}u_1^{2,a} + A_{12}(u_2^{2,a} + u_2^{1,b}) \qquad\qquad\qquad\qquad &= f_1, \\ A_{21}u_1^{2,a} + A_{22}(u_2^{2,a} + u_2^{1,b}) + A_{23}(u_3^{2,a} + u_3^{1,b}) &= f_2, \\ A_{32}(u_2^{2,a} + u_2^{1,b}) + A_{33}(u_3^{2,a} + u_3^{1,b}) &= f_3 - A_{34}u_4^{1,b}. \end{aligned}$$

Thus, if we define

$$\mathbf{v}^{k,a} = \begin{bmatrix} v_1^{k,a} \\ v_2^{k,a} \\ v_3^{k,a} \end{bmatrix} = \begin{bmatrix} u_1^{k,a} \\ u_2^{k,a} + u_2^{k-1,b} \\ u_3^{k,a} + u_3^{k-1,b} \end{bmatrix}, \qquad \mathbf{v}^{k,b} = \begin{bmatrix} v_2^{k,b} \\ v_3^{k,b} \\ v_4^{k,b} \end{bmatrix} = \begin{bmatrix} u_2^{k,b} + u_2^{k-1,a} \\ u_3^{k,b} + u_3^{k-1,a} \\ u_4^{k,b} \end{bmatrix},$$

we obtain

$$\left.\begin{aligned} A_{11}v_1^{2,a} + A_{12}v_2^{2,a} \qquad\qquad\qquad &= f_1 \\ A_{21}v_1^{2,a} + A_{22}v_2^{2,a} + A_{23}v_3^{2,a} &= f_2 \\ A_{32}v_2^{2,a} + A_{33}v_3^{2,a} &= f_3 - A_{34}v_4^{1,b} \end{aligned}\right\} \text{ in } \Omega_a, \tag{12a}$$

$$\left.\begin{aligned} A_{22}v_2^{2,b} + A_{23}v_3^{2,b} \qquad\qquad &= f_2 - A_{21}v_1^{1,a} \\ A_{32}v_2^{2,b} + A_{33}v_3^{2,b} + A_{34}v_4^{2,b} &= f_3 \\ A_{43}v_3^{2,b} + A_{44}v_4^{2,b} &= f_4 \end{aligned}\right\} \text{ in } \Omega_b, \tag{12b}$$

which is precisely the discretization of the parallel Schwarz method (2) with the u_j^k replaced by the v_j^k. Thus, the ASH iterates are identical to parallel Schwarz iterates outside the overlap, whereas inside the overlap they are linear combinations of the current and previous iterates.

2.2. Cross points. When three or more subdomains have a common overlap, the points within this overlap (regions 4, 7, and 8 in the example in Figure 2(a)) are known as cross points. When cross points are present, one can, in fact, show that ASH and RAS/parallel Schwarz are no longer equivalent. Let $\mathfrak{B}$, B_{RAS}, and B_{ASH} be the iteration matrices of parallel Schwarz, RAS, and ASH, respectively:

$$\mathfrak{B}_{ij} = \begin{cases} 0, & i = j, \\ -A_i^{-1}(R_iA - A_iR_i)\tilde{R}_j^T, & i \neq j; \end{cases}$$

$$B_{RAS} = I - \sum_{j=1}^{N} \tilde{R}_j^T A_j^{-1} R_j A, \qquad B_{ASH} = I - \sum_{j=1}^{N} R_j^T A_j^{-1} \tilde{R}_j A.$$

LEMMA 1. *The nonzero eigenvalues of $\mathfrak{B}$ and B_{RAS} are identical.*

Proof. We first recall the well-known fact that if P and Q are two rectangular matrices such that PQ and QP are both square, then PQ and QP have the same nonzero eigenvalues. Indeed, suppose λ is a nonzero eigenvalue of PQ, i.e., $PQv = \lambda v$ for $v \neq 0$. Then $Qv \neq 0$ (otherwise $PQv = 0 \implies \lambda = 0$), so we have $QP(Qv) = \lambda(Qv)$, so that λ is also an eigenvalue of QP. The other direction is similar. We now show that there exist matrices P and Q such that

$$\mathfrak{B} = PQ, \qquad B_{RAS} = QP.$$

Let

$$P = \begin{bmatrix} -A_1^{-1}(R_1A - A_1R_1) \\ \vdots \\ -A_n^{-1}(R_nA - A_nR_n) \end{bmatrix}, \qquad Q = \left[\tilde{R}_1^T, \dots, \tilde{R}_n^T\right].$$

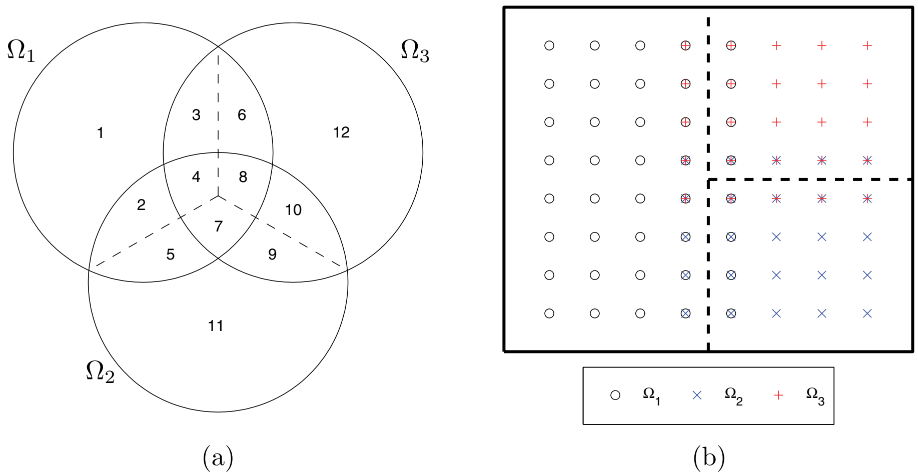


FIG. 2. *Decompositions into three subdomains with cross points. The dotted lines separate the* $\tilde{\Omega}_i$*, the nonoverlapping decomposition.* (a) *A sketch with each subregion numbered, analogous to Figure* 1. (b) *A decomposition for an* 8×8 *grid, used in the spectral study.*

Then it is straightforward to see that $[PQ]_{ij} = \mathfrak{B}_{ij}$ for $i \neq j$. For $i = j$, we calculate

$$\begin{aligned}[PQ]_{ii} &= -A_i^{-1}(R_iA - A_iR_i)\tilde{R}_i^T \\ &= -A_i^{-1}(R_iA\tilde{R}_i^T - \underbrace{R_iAR_i^T}_{A_i} R_i\tilde{R}_i^T) = -A_i^{-1}(R_iA\tilde{R}_i^T - R_iA\tilde{R}_i^T) = 0,\end{aligned}$$

since $R_i^T R_i \tilde{R}_i^T = \tilde{R}_i^T$. Thus, we have $PQ = \mathfrak{B}$. As for QP, we have

$$\begin{aligned}QP &= \sum_j \tilde{R}_j^T A_j^{-1}(A_jR_j - R_jA) \\ &= \sum_j \tilde{R}_j^T A_j^{-1}A_jR_j - \sum_j \tilde{R}_j^T A_j^{-1}R_jA = I - \sum_j \tilde{R}_j^T A_j^{-1}R_jA,\end{aligned}$$

since $\sum_j \tilde{R}_j^T R_j = I$. Thus we get $QP = B_{RAS}$, which means B_{RAS} and $\mathfrak{B}$ have the same nonzero eigenvalues. $\square$

The above lemma immediately implies that parallel Scwharz converges if and only if RAS does; if they do converge, they do so at the same rate. As for ASH, we see that if $A = A^T$, then $B_{ASH} = A^{-1}B_{RAS}^T A$, so that all three matrices have the same nonzero eigenvalues. However, when $A \neq A^T$, such an equivalence is no longer valid; in fact, we will now construct an example for which the spectrum of B_{ASH} is different from that of $\mathfrak{B}$ and B_{RAS}. Given the domain decomposition shown in Figure 2(b), let A be a matrix such that

- $|A_{ij}| = |D_{ij}|$, where D has the same sparsity pattern as the discrete five-point Laplacian matrix, but with 1.9 on the diagonal (instead of 4) and -1 on the off-diagonal;
- $A_{ij} = D_{ij}$ for $j \geq i$;
- $A_{ij} = \pm D_{ij}$ for $j < i$, with the sign chosen randomly with equal probability.

Figure 3 shows the spectra of $\mathfrak{B}$, B_{RAS}, and B_{ASH} for one such A. We see that while the eigenvalues of $\mathfrak{B}$ and B_{RAS} always coincide (as predicted by Lemma 1), the spectrum of B_{ASH} is different from the other two; in this example, we have

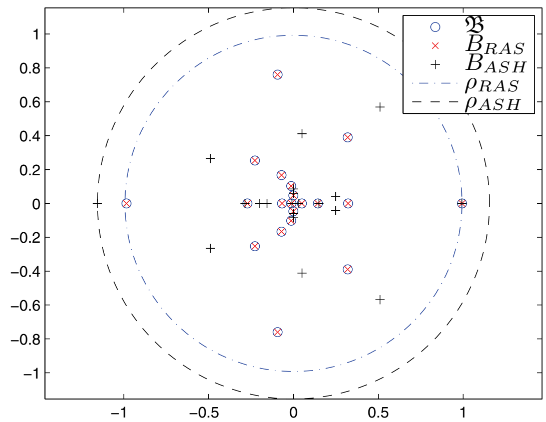


FIG. 3. *Eigenvalues and spectral radii of $\mathfrak{B}$, B_{RAS}, and B_{ASH}.*

$\rho(\mathfrak{B}) = \rho(B_{RAS}) = 0.9929$, whereas $\rho(B_{ASH}) = 1.1552$, so parallel Schwarz and RAS converge, whereas ASH diverges. Of course, if we had considered A^T instead of A as our coefficient matrix, then the exact opposite would happen: RAS and parallel Schwarz would diverge, whereas ASH would converge.

The above example shows that ASH is not equivalent to the multiple-subdomain version of parallel Schwarz proposed by [13]; in fact, the two methods are incomparable. If there were an equivalence between ASH and another Schwarz-like iteration, it would have to be a very different generalization from the two-subdomain case which, to the best of our knowledge, has yet to be proposed. In the absence of such a generalization, we will concentrate on proving the equivalence of ASH and parallel Schwarz without cross points. In the next section, we will state the assumptions necessary to prove this correspondence in the case of multiple subdomains (with no cross points) and with optimized transmission conditions.

Remark. The fact that RAS or ASH may diverge as stationary methods does not mean that they cannot be successful preconditioners when used with Krylov methods; it simply means that their behavior cannot be inferred from properties of the continuous Schwarz method, since their spectra are not equivalent.

3. Assumptions and the main result.

The main theorem stated in this section relates the iterates of optimized ASH with those of the corresponding discretized optimized Schwarz method. This result extends the one stated in [16] to handle optimized transmission conditions. Before stating the result, we make some assumptions that are algebraic manifestations of the fact that there are no cross points. Assumptions 1 and 2 are identical to those that appear in [16], whereas Assumption 3 deals specifically with optimized transmission conditions.

The first assumption ensures that no degree of freedom lies in the intersection of three distinct subdomains and is self-evident based on the definition of the restriction operators R_k.

Assumption 1 (no cross points). For distinct i, j, and l, we have

$$R_i R_j^T R_j R_l^T = 0. \tag{13}$$

The next pair of assumptions ensures that $\partial\Omega_j \setminus \partial\Omega$ are partitioned into r con-

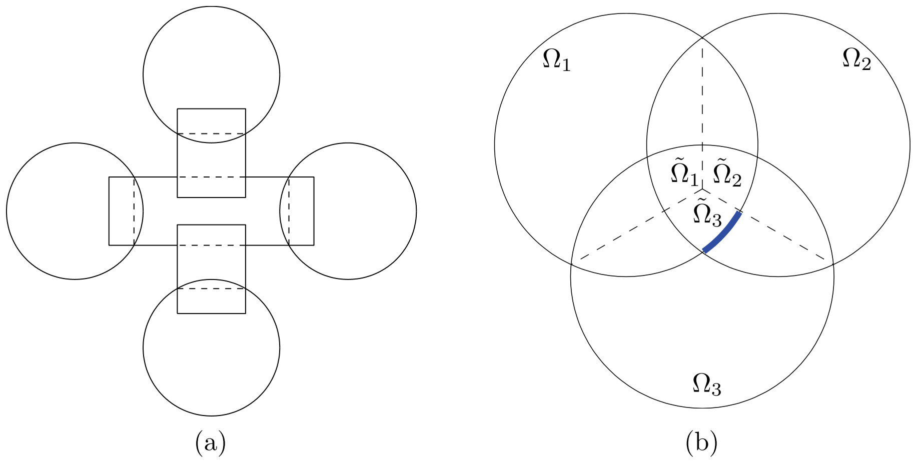


FIG. 4. *Some examples of decompositions into subdomains, with solid lines delimiting* Ω_i *and dashed lines delimiting the* $\tilde{\Omega}_i$. *In* (a), *Assumptions* 1–3 *are satisfied, whereas in* (b) *they are not.*

nected components, each of which must be a subset of $\tilde{\Omega}_i$ for some i (see Figure 4).

Assumption 2 (partition of internal boundaries). For all $i \neq j$, we must have

$$(R_i - \tilde{R}_i)(AR_j^T - R_j^T A_j) = 0, \tag{14a}$$

$$(R_i A - A_i R_i)(R_j^T - \tilde{R}_j^T) = 0. \tag{14b}$$

These two conditions are simply transposes of each other; hence, they will be satisfied simultaneously if A has a symmetric nonzero pattern. Also note that when $i = j$, the two relations are trivially satisfied: since $\tilde{R}_i = \tilde{R}_i R_i^T R_i$, we have

$$\begin{aligned}(R_i - \tilde{R}_i)(AR_i^T - R_i^T A_i) &= \underbrace{R_i A R_i^T}_{A_i} - \underbrace{R_i R_i^T}_{I} A_i - \tilde{R}_i A R_i^T + \tilde{R}_i R_i^T A_i \\ &= 0 - \tilde{R}_i R_i^T \underbrace{R_i A R_i^T}_{A_i} + \tilde{R}_i R_i^T A_i = 0.\end{aligned} \tag{15}$$

The interpretation of (14a) is as follows. For any vector w over Ω_j, the vectors $AR_j^T w$ and $R_j^T A_j w$ must agree inside Ω_j, but $AR_j^T w$ may have nonzero entries outside Ω_j (which $R_j^T A_j w$ cannot have). For a PDE, these entries are generally located along the boundary $\partial\Omega_j$. The assumption then says that these nonzero entries must fall outside the overlap region $\Omega_i \setminus \tilde{\Omega}_i$, i.e., they must be either contained in $\tilde{\Omega}_i$ or completely outside Ω_i, as in Figure 4(a). In Figure 4(b), the thick blue portion of $\partial\Omega_1$ is inside $\Omega_2 \setminus \tilde{\Omega}_2$, violating (14a) and (14b).

The next set of assumptions characterizes the optimized transmission conditions. They are analogous to (14a), (14b), and essentially require that L_i operate along the internal subdomain boundaries.

Assumption 3 (optimized transmission operators). For each i, L_i must satisfy

$$\tilde{R}_i^T L_i = 0, \qquad L_i \tilde{R}_i = 0. \tag{16}$$

Moreover, for $i \neq j$, we must have

$$(R_i - \tilde{R}_i) R_j^T L_j = 0, \qquad L_j R_j (R_i - \tilde{R}_i)^T = 0, \tag{17}$$

and for distinct i, j, and l, we must have

$$R_i R_j^T L_j R_j R_l^T = 0. \tag{18}$$

The above conditions are motivated by the observation in [13] that a change in transmission condition corresponds to a change in the diagonal blocks corresponding to degrees of freedom on (or near) the boundary. Thus, conditions (16) require that L_i have support along the internal boundaries only, which must lie completely outside $\tilde{\Omega}_i$. (In other words, we have implicitly assumed that the overlaps are large enough, so that the boundary nodes do not lie within the nonoverlapping portion of the subdomain.) Conditions (17) then require that the boundary be completely contained in $\tilde{\Omega}_j$ for some j, just as in (14a) and (14b); this is usually the case when no cross points are present. Again Figure 4(b) violates this condition because of the thick blue portion of $\partial\Omega_1$. Finally, condition (18) prohibits direct coupling between disconnected parts of the boundary via the optimized transmission conditions L_i. This assumption is reasonable, since the transmission conditions are supposed to be local operators and should not introduce far away coupling. Note that for classical ASH (Dirichlet transmission conditions), we have $L_i = 0$; in this case, the assumptions are trivially satisfied, so our results also apply to classical ASH.

We are now ready to state our main result.

THEOREM 2. *Suppose* $\mathbf{U}^0 = 0$ *and Assumptions* 1–3 *are satisfied. Then the iterates* $\mathbf{U}^k$ *of the ASH method are related to the iterates* $\mathbf{v}_i^k$ *of the discretized parallel Schwarz method*

$$\begin{aligned} \mathbf{v}_i^0 &= 0, \\ \tilde{A}_i \mathbf{v}_i^1 &= \tilde{R}_i f, \\ \tilde{A}_i \mathbf{v}_i^k &= R_i f + \sum_{j \neq i} (L_i R_i R_j^T - A_{ij}) \mathbf{v}_j^{k-1} \qquad \textit{for } k \geq 2 \end{aligned}$$

via the relation

$$\sum_{j=1}^{N} R_j^T \mathbf{v}_j^k = \mathbf{U}^k + \left(\sum_j R_j^T R_j - I \right) \mathbf{U}^{k-1}. \tag{19}$$

In particular, $\mathbf{U}^k$ *agree with* $\mathbf{v}_j^k$ *outside overlapping regions for all* j*.*

Since $\sum_j R_j R_j^T - I$ is zero outside overlapping regions, we see that the iterates of ASH and parallel Schwarz are identical outside the overlap, whereas inside they are linear combinations of the current and previous iterates. The following corollary gives the exact relationship between $\mathbf{U}^k$ and $\mathbf{v}_i^k$ inside the overlap.

COROLLARY 3. *Let Assumptions* 1–3 *be satisfied, and let* $\mathbf{U}^k$ *and* $\mathbf{v}_i^k$ *be defined as in Theorem* 2*. Then the following relations hold:*

$$\mathbf{U}^k = \sum_{j=1}^{N} R_j^T \mathbf{v}_j^k + \left(\sum_{j=1}^{N} R_j^T R_j - I \right) \sum_{l=1}^{k} \sum_{j=1}^{N} (-1)^l R_j^T \mathbf{v}_j^{k-l}, \tag{20}$$

$$\mathbf{v}_i^k = R_i \mathbf{U}^{k-1} + \tilde{A}_i^{-1} \tilde{R}_i (f - A\mathbf{U}^{k-1}) \tag{21}$$

$$= R_i \mathbf{U}^k - \sum_{j \neq i} R_i R_j^T \tilde{A}_j^{-1} \tilde{R}_j (f - A\mathbf{U}^{k-1}). \tag{22}$$

Thus, within the overlap, $\mathbf{U}^k$ is a linear combination of the previous Schwarz iterates $\mathbf{v}_j^l$, $l = 0, \ldots, k$. In contrast, the Schwarz iterates $\mathbf{v}_i^k$ cannot be obtained by simply restricting $\mathbf{U}^k$ to V_i; nonetheless, they are linear combinations of $\mathbf{U}^k$ and the subdomain solutions of the residual problem $\tilde{A}_j^{-1}\tilde{R}_j\mathbf{r}^{k-1}$.

4. Proof of Theorem 2. We assume throughout this section that Assumptions 1–3 hold. We begin by introducing some notation:

$$\mathbf{r}^k := f - A\mathbf{U}^k, \tag{23}$$

$$\delta\mathbf{u}_j^k := \tilde{A}_j^{-1}\tilde{R}_j(f - A\mathbf{U}^k), \tag{24}$$

$$\mathbf{u}_j^k := \sum_{l=0}^{k-1} \delta\mathbf{u}_j^l. \tag{25}$$

From the definition of ASH (see (9)), we have $\mathbf{U}^k = \sum_j R_j^T \mathbf{u}_j^k$. We also define the vectors $\mathbf{v}_j^k$ such that $\mathbf{v}_j^0 = \mathbf{u}_j^0 = 0$, $\mathbf{v}_j^1 = \mathbf{u}_j^1$, and for $k \geq 2$,

$$\mathbf{v}_j^k = \mathbf{u}_j^k + \sum_{i \neq j} R_j R_i^T \mathbf{u}_i^{k-1}. \tag{26}$$

The following properties are elementary and will be used often:

(a) $R_i R_i^T = I$ for all i,

(b) $\tilde{R}_i^T R_i = R_i^T \tilde{R}_i = \tilde{R}_i^T \tilde{R}_i$ for all i,

(c) $\tilde{R}_i R_i^T = R_i \tilde{R}_i^T = \tilde{R}_i \tilde{R}_i^T$ for all i,

(d) $\sum_j \tilde{R}_j^T \tilde{R}_j = I$.

We first characterize the residual $\mathbf{r}^k$.

LEMMA 4. *For all $k \geq 1$ and for all i, we have*

$$(R_i - \tilde{R}_i)\mathbf{r}^k = L_i \delta\mathbf{u}_i^{k-1}. \tag{27}$$

Proof. Fix i and let $k \geq 0$. We have

$$(R_i - \tilde{R}_i)\mathbf{r}^k = (R_i - \tilde{R}_i)\sum_j R_j^T \tilde{R}_j \mathbf{r}^k,$$

since $\sum_j R_j^T \tilde{R}_j = I$. We can now use (24) to rewrite $\tilde{R}_j\mathbf{r}^k$ as $\tilde{A}_j\delta\mathbf{u}_j^k$, giving

$$\begin{aligned}(R_i - \tilde{R}_i)\mathbf{r}^k &= (R_i - \tilde{R}_i)\sum_j R_j^T \tilde{A}_j \delta\mathbf{u}_j^k \\ &= (R_i - \tilde{R}_i)\sum_j R_j^T A_j \delta\mathbf{u}_j^k + (R_i - \tilde{R}_i)\sum_j R_j^T L_j \delta\mathbf{u}_j^k.\end{aligned}$$

For the first term on the right-hand side, we use (14a) and (15) to obtain

$$\begin{aligned}(R_i - \tilde{R}_i)\sum_j A R_j^T \delta\mathbf{u}_j^k &= (R_i - \tilde{R}_i)A\sum_j R_j^T \delta\mathbf{u}_j^k \\ &= (R_i - \tilde{R}_i)A\sum_j R_j^T (\mathbf{u}_j^{k+1} - \mathbf{u}_j^k) \\ &= (R_i - \tilde{R}_i)A(\mathbf{U}^{k+1} - \mathbf{U}^k) \qquad \left(\text{since } \textstyle\sum_j R_j^T u_j^k = \mathbf{U}^k\right) \\ &= (R_i - \tilde{R}_i)(\mathbf{r}^k - \mathbf{r}^{k+1}) \qquad (\text{since } \mathbf{r}^k = \mathbf{f}^k - A\mathbf{U}^k).\end{aligned}$$

For the second term, note that the first identity in (17) implies

$$\begin{aligned}(R_i - \tilde{R}_i)\sum_j R_j^T L_j \delta \mathbf{u}_j^k &= (R_i - \tilde{R}_i) R_i^T L_i \delta \mathbf{u}_i^k \\ &= \underbrace{R_i R_i^T}_{=I} L_i \delta \mathbf{u}_i^k - R_i \underbrace{\tilde{R}_i^T L_i}_{=0} \delta \mathbf{u}_i^k = L_i \delta \mathbf{u}_i^k.\end{aligned}$$

Thus,

$$(R_i - \tilde{R}_i)\mathbf{r}^k = (R_i - \tilde{R}_i)(\mathbf{r}^k - \mathbf{r}^{k+1}) + L_i \delta \mathbf{u}_i^k.$$

Canceling $(R_i - \tilde{R}_i)\mathbf{r}^k$ from both sides gives

$$(R_i - \tilde{R}_i)\mathbf{r}^{k+1} = L_i \delta \mathbf{u}_i^k \qquad \text{for } k \geq 0,$$

and the result follows. ∎

The next two lemmas will be needed for Lemma 7.

LEMMA 5. *For all i,*

$$\sum_{j \neq i} L_i R_i R_j^T R_j R_i^T = L_i.$$

Proof. The second identity in (17) shows that for $i \neq j$,

$$L_i R_i (R_j - \tilde{R}_j)^T = 0 \implies L_i R_i R_j^T = L_i R_i \tilde{R}_j^T.$$

So

$$\begin{aligned}L_i = L_i R_i R_i^T &= \sum_j L_i R_i \tilde{R}_j^T \tilde{R}_j R_i^T \\ &= L_i R_i \tilde{R}_i^T \tilde{R}_i R_i^T + \sum_{j \neq i} L_i R_i \tilde{R}_j^T R_j R_i^T \\ &= \underbrace{L_i \tilde{R}_i}_{=0} \tilde{R}_i^T \tilde{R}_i R_i^T + \sum_{j \neq i} L_i R_i R_j^T R_j R_i^T,\end{aligned}$$

since $L_i \tilde{R}_i = 0$ by (16). ∎

LEMMA 6. *For all i and $k \geq 1$,*

$$L_i R_i \sum_{j \neq i} R_j^T \mathbf{v}_j^k = L_i R_i \mathbf{U}^k - L_i \delta \mathbf{u}_i^{k-1}.$$

Proof. We have

$$L_i R_i \sum_{j \neq i} R_j^T \mathbf{v}_j^k = L_i R_i \sum_{j \neq i} R_j^T \mathbf{u}_j^k + L_i R_i \sum_{j \neq i} \sum_{l \neq j} R_j^T R_j R_l^T \mathbf{u}_l^{k-1}.$$

The double sum can be simplified by noting that if i, j, and l are all distinct, then the term $L_i R_i R_j^T R_j R_l^T \mathbf{u}_l^{k-1} = 0$ by (13) (the no-cross-points assumption). Thus, for a fixed i, all the terms within the double sum vanish, except for $l = i$, $j \neq i$. So we have

$$\begin{aligned}
L_i R_i \sum_{j\neq i} R_j^T \mathbf{v}_j^k &= L_i R_i \sum_{j\neq i} R_j^T \mathbf{u}_j^k + L_i R_i \sum_{j\neq i} R_j^T R_j R_i^T \mathbf{u}_i^{k-1} \\
&= L_i R_i \sum_{j\neq i} R_j^T \mathbf{u}_j^k + L_i \mathbf{u}_i^{k-1} \qquad\qquad \text{(by Lemma 5)} \\
&= L_i R_i \sum_{j} R_j^T \mathbf{u}_j^k - L_i \mathbf{u}_i^k + L_i \mathbf{u}_i^{k-1} \\
&= L_i R_i \mathbf{U}^k - L_i \mathbf{u}_i^k + L_i \mathbf{u}_i^{k-1} \\
&= L_i R_i \mathbf{U}^k - L_i \delta \mathbf{u}_i^{k-1}. \qquad \square
\end{aligned}$$

We can now prove the following lemma.

LEMMA 7. *The vectors* $\mathbf{v}_i^k$ *satisfy the following equations:*

$$\tilde{A}_i \mathbf{v}_i^1 = \tilde{R}_i f, \tag{28}$$

$$\tilde{A}_i \mathbf{v}_i^k = R_i f + \sum_{j\neq i} L_i R_i R_j^T \mathbf{v}_j^{k-1} - A_{i\Gamma} \mathbf{U}^{k-1} \qquad (k \geq 2), \tag{29}$$

where $A_{i\Gamma} := R_i A - A_i R_i$ *is the boundary operator.*

Proof. For $k \geq 0$, we have

$$\begin{aligned}
\tilde{A}_i \mathbf{v}_i^{k+1} &= \tilde{A}_i \mathbf{u}_i^{k+1} + \tilde{A}_i \sum_{j\neq i} R_i R_j^T \mathbf{u}_j^k \\
&= \tilde{A}_i \delta \mathbf{u}_i^k + \tilde{A}_i \mathbf{u}_i^k + \tilde{A}_i \sum_{j\neq i} R_i R_j^T \mathbf{u}_j^k \\
&= \tilde{R}_i (f - A\mathbf{U}^k) + \tilde{A}_i \sum_{j} R_i R_j^T \mathbf{u}_j^k \\
&= \tilde{R}_i f - \tilde{R}_i A \mathbf{U}^k + (A_i R_i + L_i R_i) \mathbf{U}^k \quad \left(\tilde{A}_i = A_i + L_i \text{ and } \sum R_j^T \mathbf{u}_j^k = \mathbf{U}^k\right).
\end{aligned}$$

If $k = 0$, then all terms other than $\tilde{R}_i f$ vanish because $\mathbf{U}^0 = 0$; we thus obtain (28). We continue by assuming $k \geq 1$:

$$\begin{aligned}
\tilde{A}_i \mathbf{v}_i^{k+1} &= \tilde{R}_i f - \tilde{R}_i A \mathbf{U}^k + (R_i A - A_{i\Gamma} + L_i R_i) \mathbf{U}^k \\
&= \tilde{R}_i f + (R_i - \tilde{R}_i) A \mathbf{U}^k + (L_i R_i - A_{i\Gamma}) \mathbf{U}^k \\
&= \tilde{R}_i f + (R_i - \tilde{R}_i)(f - \mathbf{r}^k) + (L_i R_i - A_{i\Gamma}) \mathbf{U}^k \\
&= R_i f - (R_i - \tilde{R}_i) \mathbf{r}^k + (L_i R_i - A_{i\Gamma}) \mathbf{U}^k \\
&= R_i f - L_i \delta \mathbf{u}_i^{k-1} + L_i R_i \mathbf{U}^k - A_{i\Gamma} \mathbf{U}^k \qquad\qquad \text{(by Lemma 4)} \\
&= R_i f + L_i R_i \sum_{j\neq i} R_j^T \mathbf{v}_j^k - A_{i\Gamma} \mathbf{U}^k \qquad\qquad \text{(by Lemma 6)},
\end{aligned}$$

from which (29) follows. $\square$

Note that (28) is precisely the first iteration specified by Theorem 2; the proof for $k \geq 2$ follows.

Proof of Theorem 2. We first prove the relation (19), which follows from the definition of $\mathbf{v}_j^k$. We have

$$\begin{aligned}\sum_j R_j^T \mathbf{v}_j^k &= \sum_j R_j^T \mathbf{u}_j^k + \sum_j R_j^T \sum_{l\neq j} R_j R_l^T \mathbf{u}_l^{k-1} \\ &= \mathbf{U}^k + \sum_j R_j^T R_j \left(\sum_l R_l^T \mathbf{u}_l^{k-1} - R_j^T \mathbf{u}_j^{k-1}\right) \\ &= \mathbf{U}^k + \sum_j R_j^T R_j \left(\mathbf{U}^{k-1} - R_j^T \mathbf{u}_j^{k-1}\right) \\ &= \mathbf{U}^k + \left(\sum_j R_j^T R_j\right)\mathbf{U}^{k-1} - \sum_j R_j^T \mathbf{u}_j^{k-1} \\ &= \mathbf{U}^k + \left(\sum_j R_j^T R_j\right)\mathbf{U}^{k-1} - \mathbf{U}^{k-1} \\ &= \mathbf{U}^k + \left(\sum_j R_j^T R_j - I\right)\mathbf{U}^{k-1},\end{aligned}$$

as required. It remains for us to show that

$$\tilde{A}_i \mathbf{v}_i^k = R_i f + \sum_{j\neq i} (L_i R_i R_j^T - A_{ij})\mathbf{v}_j^{k-1}$$

for $k \geq 2$. Multiplying both sides of (19) by $A_{i\Gamma}$ on the left gives

$$\begin{aligned}A_{i\Gamma} \sum_j R_j^T \mathbf{v}_j^k &= A_{i\Gamma}\mathbf{U}^k + A_{i\Gamma}\left(\sum_j R_j^T R_j - I\right)\mathbf{U}^{k-1} \\ &= A_{i\Gamma}\mathbf{U}^k + A_{i\Gamma}\left(\sum_j (R_j^T R_j - \tilde{R}_j^T R_j)\right)\mathbf{U}^{k-1} \\ &= A_{i\Gamma}\mathbf{U}^k + A_{i\Gamma}\left(\sum_j (R_j - \tilde{R}_j)^T R_j\right)\mathbf{U}^{k-1} = A_{i\Gamma}\mathbf{U}^k,\end{aligned}$$

since $A_{i\Gamma}(R_j - \tilde{R}_j)^T = 0$ for all i and j by (14b) and (15). When $i \neq j$, $A_{i\Gamma} R_j^T = (R_i A - A_i R_i) R_j^T = A_{ij}$ by definition (see (3)), and when $i = j$, we have

$$A_{i\Gamma} R_i^T = (R_i A - A_i R_i) R_i^T = R_i A R_i^T - A_i R_i R_i^T = A_i - A_i \cdot I = 0.$$

So, in fact, we have

$$A_{i\Gamma}\mathbf{U}^k = A_{i\Gamma} \sum_j R_j^T \mathbf{v}_j^k = \sum_{j\neq i} A_{ij} \mathbf{v}_j^k.$$

Substituting into (29) gives the required result. ∎

We can now prove Corollary 3.

Proof of Corollary 3. To prove (20), we simply unroll (19): by defining $P = \sum_j R_j^T R_j - I$, we get

$$\begin{aligned}\mathbf{U}^k &= \sum_j R_j^T \mathbf{v}_j^k - P\mathbf{U}^{k-1} \\ &= \sum_j R_j^T \mathbf{v}_j^k - P \sum_j R_j^T \mathbf{v}_j^{k-1} + P^2 \mathbf{U}^{k-2} \\ &= \cdots = \sum_j R_j^T \mathbf{v}_j^k + \sum_{l=1}^{k} \sum_j (-1)^l P^l R_j^T \mathbf{v}_j^{k-l},\end{aligned}$$

using the fact that $\mathbf{U}^0 = 0$. But P acts as the identity on the overlap and zero outside the overlap; thus, $P^2 = P$, which allows us to simplify the last equation to

$$\mathbf{U}^k = \sum_j R_j^T \mathbf{v}_j^k + P \cdot \sum_{l=1}^{k} \sum_j (-1)^l R_j^T \mathbf{v}_j^{k-l}.$$

We now prove (21). By Lemma 5, we have for $k \geq 2$,

$$\tilde{A}_i \mathbf{v}_i^k = R_i f + \sum_{j \neq i} L_i R_i R_j^T \mathbf{v}_j^{k-1} - A_{i\Gamma} \mathbf{U}^{k-1},$$

where $A_{i\Gamma} = R_i A - A_i R_i$. But Lemma 4 says for all $k \geq 1$,

$$L_i R_i \sum_{j \neq i} R_j^T \mathbf{v}_j^k = L_i R_i \mathbf{U}^k - L_i \delta \mathbf{u}_i^{k-1},$$

and Lemma 4 says for all $k \geq 1$,

$$(R_i - \tilde{R}_i)(f - A\mathbf{U}^k) = L_i \delta \mathbf{u}^{k-1}.$$

Substituting these two results yields

$$\begin{aligned}\tilde{A}_i \mathbf{v}_i^k &= R_i f + L_i R_i \mathbf{U}^{k-1} - (R_i - \tilde{R}_i)(f - A\mathbf{U}^{k-1}) - A_{i\Gamma}\mathbf{U}^{k-1} \\ &= R_i(f - A\mathbf{U}^{k-1}) + R_i A \mathbf{U}^{k-1} + L_i R_i \mathbf{U}^{k-1} \\ &\quad - (R_i - \tilde{R}_i)(f - A\mathbf{U}^{k-1}) - A_{i\Gamma}\mathbf{U}^{k-1} \\ &= \tilde{R}_i(f - A\mathbf{U}^{k-1}) + R_i A\mathbf{U}^{k-1} - A_{i\Gamma}\mathbf{U}^{k-1} + L_i R_i \mathbf{U}^{k-1} \\ &= \tilde{R}_i(f - A\mathbf{U}^{k-1}) + A_i R_i \mathbf{U}^{k-1} + L_i R_i \mathbf{U}^{k-1} \\ &= \tilde{R}_i(f - A\mathbf{U}^{k-1}) + \tilde{A}_i R_i \mathbf{U}^{k-1}.\end{aligned}$$

Multiplying both sides by $\tilde{A}_i^{-1}$ gives the desired result. From (21), we can easily obtain (22) using the fact (cf. (9)) that

$$\mathbf{U}^k - \mathbf{U}^{k-1} = \sum_j R_j^T \tilde{A}_j^{-1}(f - A\mathbf{U}^{k-1}). \qquad \square$$

5. Convergence rate. Given the close relationship between ASH and parallel Schwarz, one would expect that the two methods converge at the same rate. The goal of this section is to show that this is indeed the case. Recall that we can interpret the optimized parallel Schwarz method as a stationary iteration on the *augmented system*

$$\begin{bmatrix} \tilde{A}_1 & \tilde{A}_{12} & \cdots & \tilde{A}_{1N} \\ \tilde{A}_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \tilde{A}_{N1} & \cdots & \cdots & \tilde{A}_N \end{bmatrix} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_N \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} \qquad \text{or} \qquad \mathfrak{A}\mathbf{v} = \mathbf{f}, \tag{30}$$

where $\tilde{A}_{ij} = A_{ij} - L_i R_i R_j^T = (R_i A - \tilde{A}_i R_i) R_j^T$ and $f_i = R_i f$. In fact, optimized parallel Schwarz is simply a block Jacobi splitting of the system $\mathfrak{A}\mathbf{v} = \mathbf{f}$, i.e.,

$$\mathfrak{M}\mathbf{v}^{k+1} = (\mathfrak{M} - \mathfrak{A})\mathbf{v}^k + \mathbf{f} \qquad \text{for } k \geq 1,$$

where $\mathfrak{M} = \operatorname{diag}(\tilde{A}_1, \dots, \tilde{A}_N)$. It is possible for $\mathfrak{A}$ to be singular even when A is nonsingular; in this case, the method can produce spurious solutions in which the $\mathbf{v}_j$ do not agree in the overlap, even when the method converges.

To facilitate later discussions, we now define the operator R_o, which restricts the set of all nodes onto the set of nodes in the overlap (more precisely, the union of all overlapping regions). In other words, R_o has full row rank and satisfies

$$\sum_{j=1}^N R_j^T R_j = I + R_o^T R_o.$$

Note that since $\sum_j R_i^T \tilde{R}_i = I$, the above definition implies that

$$R_o^T R_o = \sum_{j=1}^N R_j^T (R_j - \tilde{R}_j) = \sum_{j=1}^N (R_j - \tilde{R}_j)^T R_j. \tag{31}$$

We now state the main theorem of this section.

THEOREM 8. *Suppose Assumptions* 1–3 *are satisfied, and that A and $\mathfrak{A}$ are both nonsingular. Then the optimized ASH method*

$$\mathbf{U}^{k+1} = \mathbf{U}^k + \sum_{j=1}^N R_j \tilde{A}_j^{-1} \tilde{R}_j (f - A\mathbf{U}^k)$$

converges to the exact solution $\mathbf{U}^*$ of $A\mathbf{U} = f$ for any f if and only if the parallel Schwarz method with $\mathbf{v}_i^0 = 0$ and

$$\begin{aligned} \tilde{A}_i \mathbf{v}_i^1 &= \tilde{R}_i f, \\ \tilde{A}_i \mathbf{v}_i^{k+1} &= R_i f - \sum_{j \neq i} \tilde{A}_{ij} \mathbf{v}_j^k \qquad (k \geq 2) \end{aligned} \tag{32}$$

converges for all right-hand side f. In addition, when both methods converge, they do so with the same asymptotic contraction rate $r = \rho(I - \sum_j R_j A_j^{-1} \tilde{R}_j A) < 1$.

To prove Theorem 8, we need the following two technical lemmas, which are shown in the appendix.

LEMMA 9. *Suppose Assumptions* 1–3 *hold. Let B_{ASH} be the ASH iteration matrix*

$$B_{ASH} = I - M^{-1}A, \qquad M^{-1} = \sum_{j=1}^N R_j^T \tilde{A}_j^{-1} \tilde{R}_j.$$

If there exists $\mathbf{w}_o \neq 0$ *such that* $B_{ASH}R_o^T\mathbf{w}_o = -R_o^T\mathbf{w}_o$, *then one of the following must be true:*

(i) $R_o\tilde{A}R_o^T$ *is singular, where* $\tilde{A} = A + \sum_{k=1}^N R_k^T L_k R_k$, *or*

(ii) *there exists some* f *for which the iteration* (32) *fails to converge.*

LEMMA 10. *Suppose Assumptions* 1–3 *hold. Then whenever* $R_o\tilde{A}R_o^T$ *is singular, so is* $\mathfrak{A}$, *the augmented matrix in* (30).

Proof of Theorem 8. Let $\mathbf{U}^*$ be the unique solution of the global problem, i.e., we have $A\mathbf{U}^* = f$. First, suppose that ASH converges with an asymptotic rate of $\rho = \rho(B_{ASH}) < 1$, i.e., there exists a constant C such that

$$\|\mathbf{U}^k - \mathbf{U}^*\|_2 \leq C\rho^k$$

for large enough k. We will show that $\mathbf{v}_i^k \to R_i\mathbf{U}^*$ with rate $r \leq \rho(B_{ASH}) < 1$. By (21), we have

$$\mathbf{v}_i^k = R_i\mathbf{U}^{k-1} + \tilde{A}_i^{-1}\tilde{R}_i(f - A\mathbf{U}^{k-1}).$$

Subtracting $R_i\mathbf{U}^*$ from both sides gives

$$\begin{aligned}\mathbf{v}_i^k - R_i\mathbf{U}^* &= R_i(\mathbf{U}^{k-1} - \mathbf{U}^*) + \tilde{A}_i^{-1}\tilde{R}_iA(\mathbf{U}^* - \mathbf{U}^{k-1}) \\ &= (R_i - \tilde{A}_i^{-1}\tilde{R}_iA)(\mathbf{U}^{k-1} - \mathbf{U}^*).\end{aligned}$$

Taking two-norms on both sides yields

$$\|\mathbf{v}_i^k - R_i\mathbf{U}^*\|_2 \leq \|(R_i - \tilde{A}_i^{-1}\tilde{R}_iA)\|_2\|\mathbf{U}^{k-1} - \mathbf{U}^*\|_2 \leq \bar{C}\rho^{k-1}.$$

Thus, optimized parallel Schwarz converges with rate $r \leq \rho(B_{ASH})$.

Now suppose optimized parallel Schwarz (32) converges for any right-hand side f. At convergence, the subdomain solutions $\{\mathbf{v}_i\}$ must satisfy the augmented system (30); since $\mathbf{v}_i^k = R_i\mathbf{U}^*$ also satisfies (30) and $\mathfrak{A}$ is invertible, we must have $\mathbf{v}_i^k \to R_i\mathbf{U}^*$; thus, there exists a constant C independent of k and $0 \leq r < 1$ such that

$$\|\mathbf{v}_i^k - R_i\mathbf{U}^*\|_2 \leq Cr^k\|\mathbf{v}_i^0 - R_i\mathbf{U}^*\|_2.$$

We now show that the spectral radius of the ASH iteration matrix $B_{ASH} = I - M^{-1}A$ satisfies $\rho(B_{ASH}) \leq r$. If $\rho(B_{ASH}) = 0$, there is nothing to show. Otherwise, suppose we choose an f so that the initial error $\mathbf{E}^1$ satisfies

$$\mathbf{E}^1 = \mathbf{U}^1 - \mathbf{U}^* = -A^{-1}\mathbf{f} = \lambda\mathbf{W},$$

where $\mathbf{W}$ is the eigenvector of B_{ASH} corresponding to λ, the largest eigenvalue of B_{ASH} in magnitude. Then for all k, we have

$$\mathbf{E}^k = \mathbf{U}^k - \mathbf{U}^* = \lambda^{k-1}\mathbf{E}^1 = \lambda^k\mathbf{W}.$$

Taking (19) and subtracting $\sum_j R_j^TR_j\mathbf{U}^*$ from both sides gives

$$\sum_{j=1}^N R_j^T(\mathbf{v}_j^k - R_j\mathbf{U}^*) = \mathbf{E}^k + \left(\sum_j R_j^TR_j - I\right)\mathbf{E}^{k-1} = \lambda^k\mathbf{W} + \lambda^{k-1}R_o^TR_o\mathbf{W}.$$

If we take two-norms on both sides, we see that for large enough k, we have

$$|\lambda|^{k-1}\left\|\lambda\mathbf{W} + R_o^TR_o\mathbf{W}\right\|_2 \leq \bar{C}r^k \tag{33}$$

for some constant $\bar{C}$ independent of k. Note that $R_o^T R_o$ simply projects the vector $\mathbf{W}$ onto its components belonging to the overlapping regions; thus, if we reorder the vector $\mathbf{W}$ and write the nonoverlapping and overlapping components separately as $\mathbf{W} = [\mathbf{W}_{no}^T, \mathbf{W}_o^T]^T$, then we get

$$\lambda\mathbf{W} + R_o^T R_o \mathbf{W} = \begin{bmatrix} \lambda\mathbf{W}_{no} \\ (\lambda+1)\mathbf{W}_o \end{bmatrix}.$$

Then squaring (33) and expressing it in terms of $\mathbf{W}_{no}$ and $\mathbf{W}_o$ gives

$$|\lambda|^{2k}\|\mathbf{W}_{no}\|_2^2 + |\lambda|^{2k-2}|\lambda+1|^2\|\mathbf{W}_o\|_2^2 \le \bar{C}r^{2k}.$$

Thus, we obtain the inequalities

$$|\lambda|^{2k}\|\mathbf{W}_{no}\|_2^2 \le \bar{C}r^{2k}, \qquad |\lambda|^{2k-2}|\lambda+1|^2\|\mathbf{W}_o\|_2^2 \le \bar{C}r^{2k}. \tag{34}$$

If $\|\mathbf{W}_{no}\|_2 \neq 0$, then the first inequality leads to

$$|\lambda|^{2k} \le \left(\frac{\bar{C}}{\|\mathbf{W}_{no}\|_2^2}\right) r^{2k}.$$

Taking the $(2k)$th root and letting $k \to \infty$ shows that $|\lambda| \le r$.

If $\|\mathbf{W}_{no}\|_2 = 0$, then $\mathbf{W}_o = R_o^T \mathbf{w}_o \neq 0$. Since $R_o\tilde{A}R_o^T$ is nonsingular by Lemma 10 and (32) converges for all f, we know by Lemma 9 that $\lambda \neq -1$, so there must exist $\varepsilon > 0$ such that $|\lambda+1| > \varepsilon$. Then the second inequality in (34) implies

$$|\lambda|^{2k-2}\varepsilon^2\|\mathbf{W}_o\|_2^2 \le \bar{C}r^{2k}.$$

Taking $(2k-2)$th roots on both sides and letting $k \to \infty$ show once again that $|\lambda| \le r$. Thus in both cases we have $\rho(B_{ASH}) \le r$, as required. Hence, we have shown that when one method converges, so does the other, and we have $\rho(B_{ASH}) \le r$ and $r \le \rho(B_{ASH})$, which implies $r = \rho(B_{ASH})$, i.e., both methods converge at the same asymptotic rate. ∎

Remark. The main difficulty in proving Theorem 8 stems from the need to ensure that the optimized ASH method does not contain oscillatory modes, i.e., we must ascertain that its iteration matrix does not have -1 as an eigenvalue. If -1 does belong to the spectrum of B, Lemmas 9 and 10 assert that this is because either the augmented system is singular, or because optimized parallel Schwarz itself has an oscillatory mode, which optimized ASH has inherited.

6. Numerical examples.

In this section, we verify Theorem 8 by presenting three examples in which optimized Schwarz and ASH have identical convergence rates.

6.1. A two-subdomain example.

For the first test problem, we solve Poisson's equation on the unit square with homogeneous Dirichlet boundary condition. We use a 20×20 grid, which is divided into two subdomains with a two-row overlap (Figure 5(a)). In the first case, we simply use Dirichlet transmission conditions as in [16], which correspond to the original ASH method as defined in [3]. In the second case, we use Robin conditions with the optimal parameter $p^* = (\pi^2/2h)^{1/3}$, as given in [21]. The convergence results are shown in Figure 5(b). As expected, the optimized method converges much faster than the classical method with Dirichlet transmission conditions. We also see that regardless of the type of transmission conditions used, the convergence curves for both methods are very close to each other, and the slopes are asymptotically equal. Thus, as iterative methods, ASH and parallel Schwarz converge at the same rate, just as we expected.

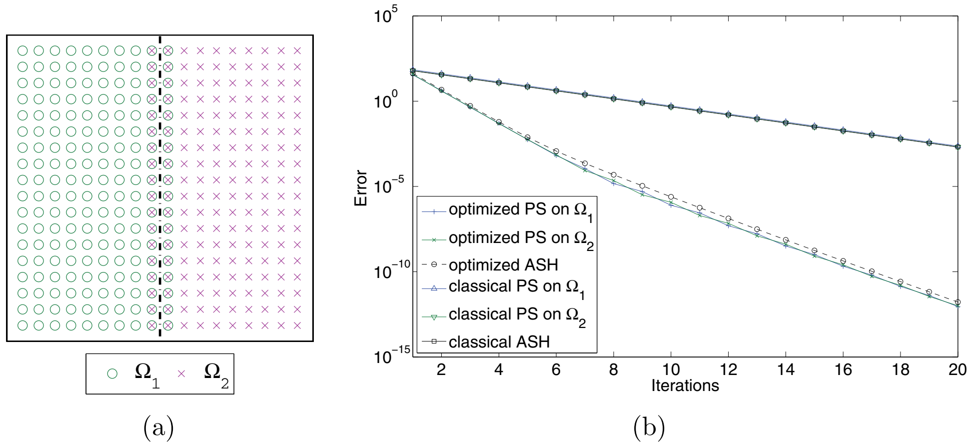


FIG. 5. (a) *A two-subdomain decomposition;* (b) *convergence of parallel Schwarz versus ASH with classical (Dirichlet) transmission conditions and optimized Robin conditions.*

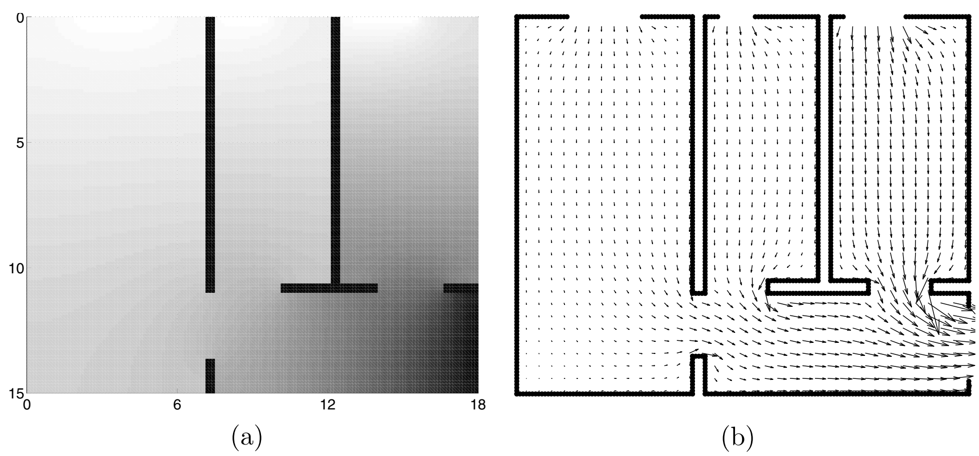


FIG. 6. *Simulation of air flow in an apartment:* (a) *steady state pressure field (light = high pressure exterior, dark = low pressure interior);* (b) *flow directions and speeds calculated based on pressure gradients.*

6.2. An example with multiple subdomains. We now present a more involved example illustrating the convergence of ASH and parallel Schwarz. Here, we would like to calculate the air flow in the apartment shown in Figure 6 when there is a pressure difference between the exterior ($P = 1$ at the open windows) and the interior of the building ($P = 0$ at the entrance of the apartment). We impose Neumann boundary conditions everywhere (including along the walls separating the rooms) except at the windows and the entrance, where we use Dirichlet conditions as indicated above. At steady state, the pressure field within the apartment satisfies the Laplace equation and is shown in Figure 6(a), with the induced air currents shown in 6(b). We decompose the domain into four subdomains (one per room), with an overlapping structure similar to the one in Figure 5(a) to ensure that Assumptions 1–3 are satisfied. The optimal Robin parameter for this problem is $p^* = (2a^2h)^{-1/3}$, where a is the distance between the interface and the closest Dirichlet boundary (see [14]).

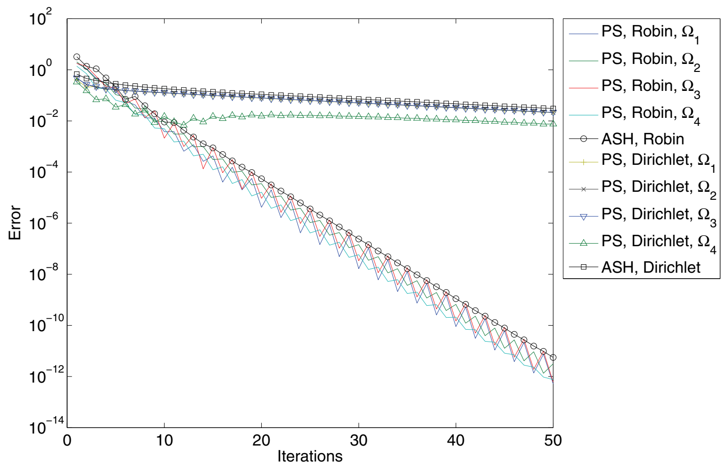


FIG. 7. *Convergence of classical and optimized ASH for the apartment problem.*

The convergence of both methods is shown in Figure 7. We see once again that both optimized ASH and optimized parallel Schwarz converge much faster than their classical counterparts, and that regardless of the boundary conditions used, the ASH curve closely follows the parallel Schwarz curves and the two methods have the same asymptotic convergence rates.

6.3. An example with systems of PDEs. We end this section with an example involving a system of PDEs. We solve the Cauchy–Riemann equations

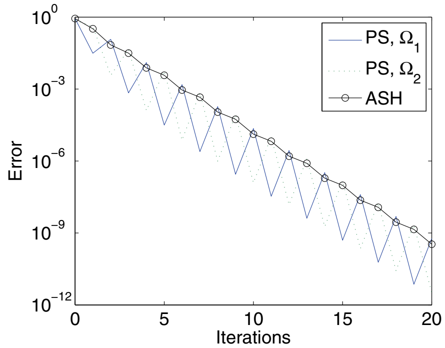
$$\mathcal{L}\mathbf{u} = \sqrt{\eta}\mathbf{u} + A\partial_x \mathbf{u} + B\partial_y \mathbf{u} = \mathbf{f}, \quad A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

on the square $\Omega = [0,1] \times [0,1]$, together with boundary conditions

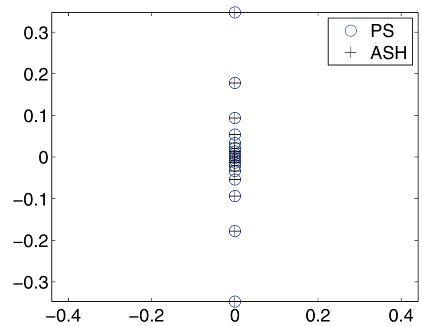
$$u_1(1,y) = f_1, \quad u_2(0,y) = f_2, \quad u_1(x,0) + u_2(x,0) = u_1(x,1) - u_2(x,1) = 0.$$

The convergence of the parallel Schwarz method on this system has been analyzed in [5, 6]. We use the same discretization as in [6] and the same decomposition into two subdomains as in subsection 6.1 (cf. Figure 5(a)). We show only results in which characteristic (Dirichlet) data are used as transmission conditions, although results using optimized conditions are similar. We see from Figure 8(a) that once again, ASH converges at the same asymptotic rate as parallel Schwarz. Whereas the convergence rate of parallel Schwarz within each subdomain exhibits two-cyclic behavior (which is typical for two-subdomain problems), the ASH error curve takes into account errors over the whole domain; hence, its error at each iteration is the maximum of the two errors. We also see from Figure 8(b) that the nonzero eigenvalues of both methods coincide perfectly, just like for RAS. This shows that our results are just as valid for systems as for scalar PDEs. Note that the identical spectra can only occur thanks to the no-cross-point assumption, since we have already shown that the spectra can be quite different when cross points are present (cf. Figure 3).

7. Conclusion. We have extended ASH to take advantage of optimized transmission conditions. In the absence of cross points, this new method is closely related



(a)



(b)

FIG. 8. *Example on Cauchy–Riemann equations:* (a) *convergence of parallel Schwarz and ASH;* (b) *eigenvalues of the iteration matrices for parallel Schwarz and ASH.*

to optimized versions of Lions' method: when the domain decomposition contains no cross points, the iterates of the optimized ASH method can be obtained by taking linear combinations of the corresponding parallel Schwarz method. In fact, the iterates of the two methods are identical outside the overlap. Such insight can be used to determine the convergence rate of the optimized ASH methods by analyzing the underlying optimized Schwarz methods, for which more convergence results are available (e.g., for systems of PDEs). Thus, this work complements the known convergence results for RAS and ASH, such as those in [10]. It would be interesting to see whether similar ideas can be applied to RASHO to relate it to the parallel Schwarz method. When cross points are present, the equivalence between ASH and parallel Schwarz no longer holds; in fact, the two stationary iterations are incomparable from a spectral point of view. Thus, another interesting prospect for future work would be to clarify whether the divergence is caused by a single outlying eigenvalue or whether there could be whole clusters lying outside the unit disc. This would give more insight into whether the two methods have similar behavior when used as preconditioners under GMRES.

Appendix. Proof of Lemma 9. The proof proceeds in several steps.

Step 1. Suppose there exists $\mathbf{w}_o \neq 0$ such that $(I - \sum_j R_j^T \tilde{A}_j^{-1} \tilde{R}_j A) R_o^T \mathbf{w}_o = -R_o^T \mathbf{w}_o$. Then

$$2R_o^T \mathbf{w}_o = \sum_j R_j^T \tilde{A}_j^{-1} \tilde{R}_j A R_o^T \mathbf{w}_o. \tag{35}$$

We now define $\mathbf{w}_j = \tilde{A}_j^{-1} \tilde{R}_j A R_o^T \mathbf{w}_o$, so that $2R_o^T \mathbf{w}_o = \sum_j R_j^T \mathbf{w}_j$. Moreover,

$$\sum_j R_j^T \tilde{A}_j \mathbf{w}_j = \sum_j R_j^T \tilde{R}_j A R_o^T \mathbf{w}_o = A R_o^T \mathbf{w}_o. \tag{36}$$

Thus,

$$\sum_j (A R_j^T - R_j^T \tilde{A}_j) \mathbf{w}_j = A \sum_j R_j \mathbf{w}_j - \sum_j R_j^T \tilde{A}_j \mathbf{w}_j = 2A R_o^T \mathbf{w}_o - A R_o^T \mathbf{w}_o = A R_o^T \mathbf{w}_o.$$

Now if we multiply from the left-hand side by $R_i - \tilde{R}_i$, then we see that for any i,

$$\begin{aligned}(R_i - \tilde{R}_i)AR_o^T\mathbf{w}_o &= \sum_j (R_i - \tilde{R}_i)(AR_j^T - R_j^T\tilde{A}_j)\mathbf{w}_j \\ &= \sum_j \underbrace{(R_i - \tilde{R}_i)(AR_j^T - R_j^T A_j)}_{=0 \text{ by Ass. 2}}\mathbf{w}_j - \sum_j \underbrace{(R_i - \tilde{R}_i)R_j^T L_j}_{=0 \text{ for } j\neq i \text{ by (17)}}\mathbf{w}_j \\ &= -L_i\mathbf{w}_i.\end{aligned}$$

Step 2. Another way of writing (35) is to note that

$$\sum_j R_j^T R_j R_o^T\mathbf{w}_o = (I + R_o^T R_o)R_o^T\mathbf{w}_o = 2R_o^T\mathbf{w}_o,$$

so (35) becomes

$$\sum_j (R_j^T R_j - R_j^T\tilde{A}_j^{-1}\tilde{R}_j A)R_o^T\mathbf{w}_o = \sum_j R_j^T\tilde{A}_j^{-1}(\tilde{A}_j R_j - \tilde{R}_j A)R_o^T\mathbf{w}_o = 0.$$

Define $\mathbf{z}_j = \tilde{A}_j^{-1}(\tilde{A}_j R_j - \tilde{R}_j A)R_o^T\mathbf{w}_o$, so that $\sum_j R_j^T\mathbf{z}_j = 0$. It is straightforward to see that $\mathbf{w}_j + \mathbf{z}_j = R_j R_o^T\mathbf{w}_o$ for all j. The result of the first step then implies

$$(R_i - \tilde{R}_i)AR_o^T\mathbf{w_o} = L_i(\mathbf{z}_i - R_i R_o^T\mathbf{w}_o),$$

which then gives

$$\begin{aligned}L_i\mathbf{z}_i &= (R_i - \tilde{R}_i)AR_o^T\mathbf{w_o} + L_i R_i R_o^T\mathbf{w}_o \\ &= (R_i - \tilde{R}_i)(A + R_i^T L_i R_i)R_o^T\mathbf{w}_o,\end{aligned} \tag{37}$$

since $\tilde{R}_i^T L_i = 0$ by (16).

We now give another representation of $\mathbf{z}_i$ which will be useful later. We calculate

$$\begin{aligned}(\tilde{A}_j R_j - \tilde{R}_j A)R_o^T &= L_j R_j R_o^T + (A_j R_j - \tilde{R}_j A)R_o^T \\ &= L_j R_j R_o^T + (A_j R_j - R_j A)R_o^T + (R_j A - \tilde{R}_j A)R_o^T \\ &= L_j R_j R_o^T + (A_j R_j - R_j A)R_o^T R_o R_o^T + (R_j A - \tilde{R}_j A)R_o^T \\ &= L_j R_j R_o^T + \sum_k \underbrace{(A_j R_j - R_j A)(R_k - \tilde{R}_k)^T}_{=0 \text{ by Ass. 2}} R_k R_o^T \\ &\quad + (R_j - \tilde{R}_j)AR_o^T,\end{aligned}$$

so that

$$(\tilde{A}_j R_j - \tilde{R}_j A)R_o^T = L_j R_j R_o^T + (R_j - \tilde{R}_j)AR_o^T = (R_j - \tilde{R}_j)(A + R_j^T L_j R_j)R_o^T.$$

Thus, combining the definition of $\mathbf{z}_j$ and (37) gives

$$\tilde{A}_j\mathbf{z}_j = (\tilde{A}_j R_j - \tilde{R}_j A)R_o^T\mathbf{w}_o = (R_j - \tilde{R}_j)(A + R_j^T L_j R_j)R_o^T\mathbf{w}_o = L_j\mathbf{z}_j. \tag{38}$$

Incidentally, (38) implies $A_j\mathbf{z}_j = 0$, but this does not imply $\mathbf{z}_j = 0$ because A_j does not need to be invertible. Indeed, the method is well defined whenever $\tilde{A}_j$ is nonsingular, so the nonsingularity of A_j is not a natural requirement unless classical (Dirichlet) transmission conditions are used.

Step 3. We now need to consider two cases: either $\mathbf{z}_j = 0$ for all j, or $\mathbf{z}_i \neq 0$ for at least one i. In the first case, $\mathbf{z}_i = 0$ implies $\mathbf{w}_i = R_i R_o^T \mathbf{w}_o$ for all i; the last line of Step 1, which reads $(R_i - \tilde{R}_i) A R_o^T \mathbf{w}_o = -L_i \mathbf{w}_i$, can be rewritten as

$$\begin{aligned} 0 &= (R_i - \tilde{R}_i) A R_o^T \mathbf{w}_o + L_i \mathbf{w}_i \\ &= (R_i - \tilde{R}_i)(A + R_i^T L_i R_i) R_o^T \mathbf{w}_o \\ &= (R_i - \tilde{R}_i)\left(A + \sum_k R_k^T L_k R_k\right) R_o^T \mathbf{w}_o, \end{aligned} \tag{39}$$

since $(R_i - \tilde{R}_i) R_k^T L_k = 0$ for $k \neq i$ by (17). Since (39) is true for all i, we can sum through the i and get

$$0 = \sum_i R_i^T (R_i - \tilde{R}_i)\left(A + \sum_k R_k^T L_k R_k\right) R_o^T \mathbf{w}_o = R_o^T R_o \tilde{A} R_o^T \mathbf{w}_o.$$

Multiplying the above by R_o shows that $R_o \tilde{A} R_o^T$ is singular, with a nullspace containing the nonzero vector $R_o^T \mathbf{w}_o$.

Step 4. Now suppose at least one of the $\mathbf{z}_j$ is nonzero. Then let us run the iteration (32) with the right-hand side $f = A R_o^T \mathbf{w}_o$. We claim that for all $k \geq 1$, we have

$$\mathbf{v}_i^k = R_i R_o^T \mathbf{w}_o + (-1)^k \mathbf{z}_i. \tag{40}$$

Since $\mathbf{z}_i \neq 0$ for some i, (40) would imply that parallel optimized Schwarz does not converge, as stated in Lemma 9. For the first iterate, we have by definition

$$\mathbf{v}_i^1 = \tilde{A}_i^{-1} \tilde{R}_i A R_o^T \mathbf{w}_o = \mathbf{w}_i = R_i R_o^T \mathbf{w}_o - \mathbf{z}_i.$$

For $k = 2$, we subtract the equation for $k = 1$ from the one for $k = 2$:

$$\begin{aligned} \tilde{A}_i(\mathbf{v}_i^2 - \mathbf{v}_i^1) &= (R_i - \tilde{R}_i) f + \sum_{j \neq i} (L_i R_i R_j^T - A_{ij}) \mathbf{v}_j^1 \\ &= (R_i - \tilde{R}_i) A R_o^T \mathbf{w}_o - \sum_{j \neq i} (R_i A - \tilde{A}_i R_i) R_j^T \mathbf{w}_j \\ &= (R_i - \tilde{R}_i) A R_o^T \mathbf{w}_o - \sum_j (R_i A - \tilde{A}_i R_i) R_j^T \mathbf{w}_j + (R_i A - \tilde{A}_i R_i) R_i^T \mathbf{w}_i. \end{aligned}$$

The second term in the right-hand side above can be simplified using $\sum_j R_j \mathbf{w}_j = 2 R_o^T \mathbf{w}_o$:

$$\begin{aligned} \tilde{A}_i(\mathbf{v}_i^2 - \mathbf{v}_i^1) &= (R_i - \tilde{R}_i) A R_o^T \mathbf{w}_o - 2(R_i A - \tilde{A}_i R_i) R_o^T \mathbf{w}_o \\ &\quad + \underbrace{(R_i A - A_i R_i) R_i^T}_{=0} \mathbf{w}_i - L_i \mathbf{w}_i \\ &= (R_i - \tilde{R}_i) A R_o^T \mathbf{w}_o + 2 L_i R_i R_o^T \mathbf{w_o} - L_i \mathbf{w}_i \\ &= (R_i - \tilde{R}_i)(A + R_i^T L_i R_i) R_o^T \mathbf{w}_o + L_i (R_i R_o^T \mathbf{w_o} - \mathbf{w}_i) = 2 \tilde{A}_i \mathbf{z}_i \end{aligned}$$

by (38). Since $\tilde{A}_i$ is nonsingular, we conclude that

$$\mathbf{v}_i^2 = \mathbf{v}_i^1 + 2\mathbf{z}_i = R_i R_o^T \mathbf{w}_o + \mathbf{z}_i.$$

Finally, for $k > 2$, we assume inductively that $\mathbf{v}_i^k = R_i R_o^T \mathbf{w}_o + (-1)^k \mathbf{z}_i$. Then to show the $(k+1)$st step, we subtract the equation for k from the one for $k+1$:

$$\begin{aligned}
\tilde{A}_i(\mathbf{v}_i^{k+1} - \mathbf{v}_i^k) &= \sum_{j \neq i} (L_i R_i R_j^T - A_{ij})(\mathbf{v}_j^k - \mathbf{v}_j^{k-1}) \\
&= 2(-1)^k \sum_{j \neq i} (\tilde{A}_i R_i - R_i A) R_j^T \mathbf{z}_j \\
&= 2(-1)^k \Big[-(\tilde{A}_i R_i - R_i A) R_i^T \mathbf{z}_i + \underbrace{\sum_j (\tilde{A}_i R_i - R_i A) R_j^T \mathbf{z}_j}_{=0 \text{ since } \sum_j R_j^T \mathbf{z}_j = 0} \Big] \\
&= -2(-1)^k L_i \mathbf{z}_i = -2(-1)^k \tilde{A}_i \mathbf{z}_i.
\end{aligned}$$

Hence, we get

$$\mathbf{v}_i^{k+1} = \mathbf{v}_i^k - 2(-1)^k \mathbf{z}_i = R_i R_o^T \mathbf{w}_o + (-1)^{k+1} \mathbf{z}_i.$$

Proof of Lemma 10. To prove Lemma 10, we will first need the following lemma, which is the analogue of the no-cross-point assumption for the stencil of A.

LEMMA 11. *For distinct i, j, and l, we have*

$$R_i R_j^T R_j A R_l^T = R_i A R_j^T R_j R_l^T = 0, \tag{41}$$

$$R_i R_j^T R_j \tilde{A} R_l^T = R_i \tilde{A} R_j^T R_j R_l^T = 0. \tag{42}$$

Proof. We only show the first inequalities in (41) and (42); the proof for the second inequalities is similar. We start by showing (41). Assumption 2 (partition of internal boundaries) implies

$$R_j A R_l^T = R_j R_l^T A_l + \tilde{R}_j (A R_l^T - R_l^T A_l),$$

which means

$$R_i R_j^T R_j A R_l^T = R_i R_j^T R_j R_l^T A_l + R_i R_j^T \tilde{R}_j (A R_l^T - R_l^T A_l).$$

Since $R_i R_j^T R_j R_l^T = R_i R_j^T \tilde{R}_j R_l^T = 0$ for distinct i, j, and l (no cross points), two out of the three terms on the right-hand side vanish; multiplying the remaining terms from the left by R_i^T gives

$$R_i^T R_i R_j^T R_j A R_l^T = R_i^T R_i R_j^T \tilde{R}_j A R_l^T.$$

Thus, $R_i^T R_i R_j^T R_j A R_l^T$ has support inside $V_i \cap \tilde{V}_j$. By interchanging the roles of i and j, we see that $R_j^T R_j R_i^T R_i A R_l^T$ has support inside $V_j \cap \tilde{V}_i$. But

$$(R_i^T R_i)(R_j^T R_j) A R_l^T = (R_j^T R_j)(R_i^T R_i) A R_l^T,$$

since diagonal matrices commute; so $R_i^T R_i R_j^T R_j A R_l^T$, in fact, has support only inside $\tilde{V}_i \cap \tilde{V}_j = \emptyset$. Hence, we have shown that

$$R_i R_j^T R_j A R_l^T = R_i (R_i^T R_i R_j^T R_j A R_l^T) = 0.$$

We now show that $R_i R_j^T R_j R_k^T L_k R_k R_l^T = 0$ for all k, which together with (41) would imply that

$$R_i R_j^T R_j \tilde{A} R_l^T = R_i R_j^T R_j \left(\sum_k R_k^T L_k R_k \right) R_l^T = 0.$$

By (18) in Assumption 3, we know that $R_j R_k^T L_k R_k R_l^T = 0$ for distinct j, k, and l; thus, we have

$$R_i R_j^T R_j R_k^T L_k R_k R_l^T = \begin{cases} 0, & k \notin \{j, l\}, \\ R_i R_j^T L_j R_j R_l^T = 0, & k = j, \\ \underline{R_i R_j^T R_j R_l} L_l = 0, & k = l, \end{cases}$$

where the last case gives zero because of the no-cross-point assumption. ◻

Let $P_{ij} = R_i^T R_i R_j^T R_j = R_j^T R_j R_i^T R_i$, i.e., P_{ij} replaces any component outside $V_i \cap V_j$ by zero. Then Lemma 11 says that unless $l \in \{i, j\}$, we must have $P_{ij} A R_l^T = 0$, i.e., only stencils within V_i or V_j can extend into the overlap $V_i \cap V_j$.

COROLLARY 12. *For $i \neq j$, we have*

$$P_{ij} A R_o^T = P_{ij} A P_{ij} R_o^T, \qquad R_o A P_{ij} = R_o P_{ij} A P_{ij}, \tag{43}$$

$$P_{ij} \tilde{A} R_o^T = P_{ij} \tilde{A} P_{ij} R_o^T, \qquad R_o \tilde{A} P_{ij} = R_o P_{ij} \tilde{A} P_{ij}. \tag{44}$$

Proof. We prove only the first inequality; the other is similar. This can be done either purely algebraically using the properties of R_i, or by the following geometric argument. For any $\mathbf{w}_o \neq 0$, the nonzero elements of $R_o^T \mathbf{w_o}$ must lie within two of the subdomains V_l. However, for any $l \notin \{i, j\}$, we have $P_{ij} \tilde{A} R_l^T = 0$. Thus, only nonzero elements that lie within $V_i \cap V_j$ can contribute to the result, i.e.,

$$P_{ij} \tilde{A} R_o^T = P_{ij} \tilde{A} P_{ij} R_o^T. \qquad ◻$$

Proof of Lemma 10. Suppose $R_o \tilde{A} R_o^T$ is singular, i.e., there exists $\mathbf{y}^T \neq 0$ such that $\mathbf{y}^T R_o \tilde{A} R_o^T = 0$. We show that there must be distinct i and j such that $\mathbf{y}^T R_o R_i^T R_i R_j^T R_j \neq 0$. Assume the contrary, i.e., $\mathbf{y}^T R_o R_i^T R_i R_j^T R_j = 0$ for all $i \neq j$. Then we must have

$$\begin{aligned} 0 = \sum_i \sum_{j \neq i} \mathbf{y}^T R_o R_i^T R_i R_j^T R_j &= \mathbf{y}^T R_o \sum_i R_i^T R_i \left(\sum_j R_j^T R_j - R_i^T R_i \right) \\ &= \mathbf{y}^T R_o \left(\sum_i R_i^T R_i \sum_j R_j^T R_j - \sum_i R_i^T R_i \right) \\ &= \mathbf{y}^T R_o \Big[(I + R_o^T R_o)(I + R_o^T R_o) - (I + R_o^T R_o) \Big] \\ &= 2 \mathbf{y}^T R_o, \end{aligned}$$

which contradicts the fact that $\mathbf{y}^T \neq 0$, since R_o has full row rank. Thus, by renaming the subdomains if necessary, we can assume without loss of generality that $\mathbf{y}^T R_o R_1^T R_1 R_2^T R_2 \neq 0$. Now consider the nonzero vector

$$\mathbf{z}^T = \left[\mathbf{y}^T R_o R_2^T R_2 R_1^T, \quad -\mathbf{y}^T R_o R_1^T R_1 R_2^T, \quad 0, \ldots, 0 \right].$$

We argue that $\mathbf{z}^T\mathfrak{A} = 0$, which would imply that $\mathfrak{A}$ is singular. We verify this by calculating the first, second, and jth component ($j \neq 1,2$) of $\mathbf{z}^T\mathfrak{A}$. For the first component, we have

$$\begin{aligned}
(\mathbf{z}^T\mathfrak{A})_1 &= \mathbf{y}^T R_o(R_2^T R_2 R_1^T \tilde{A}_1 - R_1^T R_1 R_2^T \tilde{A}_{21}) \\
&= \mathbf{y}^T R_o\Big[R_2^T R_2 R_1^T R_1 A R_1^T + R_2^T R_2 R_1^T L_1 \\
&\qquad\qquad - R_1^T R_1 R_2^T (R_2 A - A_2 R_2 - L_2 R_2) R_1^T\Big] \\
&= \mathbf{y}^T R_o\Big[R_2^T R_2 R_1^T L_1 + R_1^T R_1 R_2^T L_2 R_2 R_1^T + R_1^T R_1 R_2^T A_2 R_2 R_1^T\Big] \\
&= \mathbf{y}^T R_o\Big[R_2^T R_2 R_1^T L_1 + R_1^T R_1 R_2^T L_2 R_2 R_1^T + R_1^T R_1 R_2^T R_2 A R_2^T R_2 R_1^T\Big] \\
&\overset{\text{(Lem. 5)}}{=} \mathbf{y}^T R_o \Bigg[R_2^T R_2 R_1^T \sum_{k\neq 1} L_1 R_1 R_k^T R_k R_1^T \\
&\qquad\qquad + R_1^T R_1 R_2^T L_2 R_2 R_1^T + R_1^T R_1 R_2^T R_2 A R_2^T R_2 R_1^T\Bigg] \\
&\overset{(18)}{=} \mathbf{y}^T R_o\Big[R_2^T R_2 R_1^T L_1 R_1 R_2^T R_2 R_1^T + R_1^T R_1 R_2^T L_2 R_2 R_1^T \\
&\qquad\qquad + R_1^T R_1 R_2^T R_2 A R_2^T R_2 R_1^T\Big] \\
&= \mathbf{y}^T R_o R_1^T R_1 R_2^T R_2 \tilde{A} R_2^T R_2 R_1^T \\
&\overset{(44)}{=} \mathbf{y}^T R_o \tilde{A} R_2^T R_2 R_1^T.
\end{aligned}$$

Using the fact that $R_2^T R_2 R_1^T R_1 = R_o^T R_o R_2^T R_2 R_1^T R_1$ (since $R_o^T R_o$ projects onto the union of all overlapping regions, which is a superset of $V_1 \cap V_2$), we get

$$(\mathbf{z}^T\mathfrak{A})_1 = \underbrace{y^T R_o \tilde{A}(R_o^T}_{=0} R_o) R_2^T R_2 R_1^T R_1 R_1^T = 0.$$

A similar calculation shows that the second component of $\mathbf{z}^T\mathfrak{A}$ also vanishes. For the jth component with $j \neq 1,2$, we see that

$$\begin{aligned}
(\mathbf{z}^T\mathfrak{A})_j &= \mathbf{y}^T R_o(R_2^T R_2 R_1^T \tilde{A}_{1j} - R_1^T R_1 R_2^T \tilde{A}_{2j}) \\
&= \mathbf{y}^T R_o\Big[R_2^T R_2 R_1^T (R_1 A - A_1 R_1 - L_1 R_1) \\
&\qquad\qquad - R_1^T R_1 R_2^T (R_2 A - A_2 R_2 - L_2 R_2)\Big] R_j^T.
\end{aligned}$$

Since $R_2^T R_2 R_1^T L_1 R_1 R_j^T = R_1^T R_1 R_2^T L_2 R_2 R_j^T = 0$ for $j \neq 1,2$, we, in fact, have

$$\begin{aligned}
(\mathbf{z}^T\mathfrak{A})_j &= \mathbf{y}^T R_o\Big[R_2^T R_2 R_1^T (R_1 A - A_1 R_1) - R_1^T R_1 R_2^T (R_2 A - A_2 R_2)\Big] R_j^T \\
&= \mathbf{y}^T R_o (R_2^T R_2 R_1^T A_1 R_1 - R_1^T R_1 R_2^T A_2 R_2) R_j^T \\
&= \mathbf{y}^T R_o R_1^T R_1 R_2^T R_2 (A R_1^T R_1 - A R_2^T R_2) R_j^T \\
&= \mathbf{y}^T R_o P_{12} A R_1^T R_1 R_j^T - \mathbf{y}^T R_o P_{12} A R_2^T R_2 R_j^T.
\end{aligned}$$

We now show that $\mathbf{y}^T R_o P_{12} A R_i^T R_i R_j^T = 0$ for $i = 1,2$. Using the fact that $R_i^T R_i R_j^T =$

$(R_o^T R_o)R_i^T R_i R_j^T$ (the same argument as above, since $i \neq j$), we get

$$\begin{aligned}
\mathbf{y}^T R_o P_{12} A R_i^T R_i R_j^T &= \mathbf{y}^T R_o P_{12} A (R_o^T R_o) R_i^T R_i R_j^T \\
&= \mathbf{y}^T R_o P_{12} A P_{12} R_o^T R_o R_i^T R_i R_j^T && \text{(by (43))} \\
&= \mathbf{y}^T R_o A P_{12} R_o^T R_o R_i^T R_i R_j^T && \text{(by (43))} \\
&= \mathbf{y}^T R_o A R_o^T R_o P_{12} R_i^T R_i R_j^T = 0.
\end{aligned}$$

Thus, we conclude that $\mathbf{z}^T \mathfrak{A} = 0$, i.e., $\mathfrak{A}$ is singular. $\square$

Acknowledgments. We thank Martin J. Gander for suggesting that we work on a continuous interpretation of ASH and for his invaluable comments. We would also like to thank the anonymous referee who made many good suggestions that have strengthened this paper considerably.

REFERENCES

[1] D. Bennequin, M. J. Gander, and L. Halpern, *A homographic best approximation problem with application to optimized Schwarz waveform relaxation*, Math. Comp., 78 (2009), pp. 185–223.

[2] X.-C. Cai, M. Dryja, and M. Sarkis, *Restricted additive Schwarz preconditioners with harmonic overlap for symmetric positive definite linear systems*, SIAM J. Numer. Anal., 41 (2003), pp. 1209–1231.

[3] X.-C. Cai and M. Sarkis, *A restricted additive Schwarz preconditioner for general sparse linear systems*, SIAM J. Sci. Comput., 21 (1999), pp. 792–797.

[4] B. Després, *Décomposition de domaine et problème de Helmholtz*, C.R. Acad. Sci. Paris Sér. I Math., 311 (1990), pp. 313–316.

[5] V. Dolean and M. J. Gander, *Why classical Schwarz methods applied to hyperbolic systems can converge even without overlap*, in Proceedings of the 17th International Conference on Domain Decomposition Methods, Springer, Berlin, 2008, pp. 467–475.

[6] V. Dolean and M. J. Gander, *Can the discretization modify the performance of Schwarz methods?*, in Domain Decomposition Methods in Science and Engineering XIX, Lect. Notes Comput. Sci. Eng. 78, Springer-Verlag, Berlin, 2010, pp. 117–124.

[7] T. A. Driscoll, K.-C. Toh, and L. N. Trefethen, *From potential theory to matrix iterations in six steps*, SIAM Rev., 40 (1998), pp. 547–578.

[8] M. Dryja and O. B. Widlund, *Some domain decomposition algorithms for elliptic problems*, in Iterative Methods for Large Linear Systems, L. Hayes and D. Kincaid, eds., Academic Press, Boston, 1989, pp. 273–291.

[9] E. Efstathiou and M. J. Gander, *Why restricted additive Schwarz converges faster than additive Schwarz*, BIT, 43 (2003), pp. 945–959.

[10] A. Frommer and D. B. Szyld, *An algebraic convergence theory for restricted additive Schwarz methods using weighted max norms*, SIAM J. Numer. Anal., 39 (2001), pp. 463–479.

[11] M. J. Gander, F. Magoulès, and F. Nataf, *Optimized Schwarz methods without overlap for the Helmholtz equation*, SIAM J. Sci. Comput., 24 (2002), pp. 38–60.

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

[13] M. J. Gander, *Schwarz methods in the course of time*, Electron. Trans. Numer. Anal., 31 (2008), pp. 228–255.

[14] M. J. Gander, *On the influence of geometry on Schwarz methods*, Bol. Soc. Esp. Mat. Apl., 53 (2011), pp. 71–78.

[15] C. Japhet, *Méthode de décomposition de domaine et conditions aux limites artificielles en mécanique des fluides: méthode optimisée d'Ordre* 2, Ph.D. thesis, Université Paris XIII, Paris, France, 1998.

[16] F. Kwok, *Is additive Schwarz with harmonic extension just Lions' method in disguise?*, in Domain Decomposition Methods in Science and Engineering XIX, Huang, Kornhuber, Widlund, and Xu, eds., Lect. Notes Comput. Sci. Eng. 78, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 101–108.

[17] P.-L. Lions, *On the Schwarz alternating method*. I, in the First International Symposium on Domain Decomposition Methods for Partial Differential Equations, SIAM, Philadelphia, 1988.

[18] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.
[19] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comp., 7 (1986), pp. 856–869.
[20] H. A. Schwarz, *Über einen Grenzübergang durch alternierendes Verfahren*, Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, 15 (1870), pp. 272–286.
[21] A. St-Cyr, M. J. Gander, and S. J. Thomas, *Multiplicative, additive, and restricted additive Schwarz preconditioning*, SIAM J. Sci. Comput., 29 (2007), pp. 2402–2425.