SCHWARZ METHODS: TO SYMMETRIZE OR NOT TO SYMMETRIZE[1]

Michael Holst
Applied Mathematics 217-50
Caltech, Pasadena, CA 91125

Stefan Vandewalle
Applied Mathematics 217-50
Caltech, Pasadena, CA 91125

SUMMARY

A preconditioning theory for Schwarz methods is presented. The theory establishes sufficient conditions for multiplicative and additive Schwarz algorithms to yield self-adjoint positive definite preconditioners. It allows for the analysis and use of non-variational and non-convergent linear methods as preconditioners for conjugate gradient methods, and it is applied to domain decomposition and multigrid. This paper illustrates why symmetrizing may be a bad idea for linear methods. Numerical examples are presented for a test problem.

INTRODUCTION

In this paper, we consider additive and multiplicative Schwarz methods and their acceleration with Krylov methods, for the numerical solution of self-adjoint positive definite (SPD) operator equations arising from the discretization of elliptic partial differential equations. The standard theory of conjugate gradient acceleration of linear methods requires that a certain operator associated with the linear method—the preconditioner—be symmetric and positive definite. Often, however, as in the case of Schwarz-based preconditioners, the preconditioner is known only implicitly, and symmetry and positive definiteness are not easily verified. Here, we try to construct natural sets of sufficient conditions that are easily verified and do not require the explicit formulation of the preconditioner. More precisely, we derive conditions for the constituent components of MG and DD algorithms (smoother, subdomain solver, transfer operators, etc.), that guarantee symmetry and positive definiteness of the preconditioning operator which is (explicitly or implicitly) defined by the resulting Schwarz method. We examine the implications of these conditions for various formulations of the standard DD and MG algorithms.

The outline of the paper is as follows. We begin in the next section by reviewing basic linear methods for SPD linear operator equations and by examining Krylov acceleration strategies. A simple lemma will illustrate why symmetrizing may be a bad idea for linear methods. In the third and fourth sections, we analyze multiplicative and additive Schwarz preconditioners. We develop a theory that establishes sufficient conditions for the multiplicative and additive algorithms to yield SPD preconditioners. This theory is used to establish sufficient conditions for multiplicative and additive DD and MG methods, and it allows for analysis of non-variational and even non-convergent linear methods as preconditioners. In the final section, we report results of numerical experiments with finite-element-based DD and MG methods applied to a difficult test problem with discontinuous coefficients to illustrate the theory and conjectures.

[1]This work was supported in part by the NSF under Cooperative Agreement No. CCR-9120008.

Notation. Let $\mathcal{H}$ be a real finite-dimensional Hilbert space equipped with the inner-product $(\cdot,\cdot)$ inducing the norm $\|\cdot\| = (\cdot,\cdot)^{1/2}$. The *adjoint* of a linear operator $A \in \mathbf{L}(\mathcal{H},\mathcal{H})$ with respect to $(\cdot,\cdot)$ is the unique operator A^T satisfying $(Au,v) = (u, A^T v)$, $\forall u,v \in \mathcal{H}$. An operator A is called *self-adjoint* or *symmetric* if $A = A^T$; a self-adjoint operator A is called *positive definite* or simply *positive* if $(Au,u) > 0$, $\forall u \in \mathcal{H},\ u \neq 0$. If A is self-adjoint positive definite (SPD), then the bilinear form (Au,v) defines another inner-product, which we denote as $(\cdot,\cdot)_A$. It induces the norm $\|\cdot\|_A = (\cdot,\cdot)_A^{1/2}$.

The adjoint of an operator M with respect to $(\cdot,\cdot)_A$, the A-*adjoint*, is the unique operator M^* satisfying $(Mu,v)_A = (u, M^*v)_A$, $\forall u,v \in \mathcal{H}$. From this definition it follows that

$$M^* = A^{-1} M^T A \ . \tag{1}$$

M is called A-*self-adjoint* if $M = M^*$ and A-*positive* if $(Mu,u)_A > 0$, $\forall u \in \mathcal{H},\ u \neq 0$.

If $N \in \mathbf{L}(\mathcal{H}_1,\mathcal{H}_2)$, then $N^T \in \mathbf{L}(\mathcal{H}_2,\mathcal{H}_1)$ is defined as the unique operator relating the inner-products in $\mathcal{H}_1$ and $\mathcal{H}_2$ as follows:

$$(Nu,v)_{\mathcal{H}_2} = (u, N^T v)_{\mathcal{H}_1} \ , \quad \forall u \in \mathcal{H}_1 \ , \quad \forall v \in \mathcal{H}_2 \ . \tag{2}$$

Since it is usually clear from the arguments which inner-product is involved, we shall often drop the subscripts on inner-products (and norms) throughout the paper.

We denote the spectrum of an operator M as $\sigma(M)$. The spectral theory for self-adjoint linear operators states that the eigenvalues of the self-adjoint operator M are real and lie in the closed interval $[\lambda_{\min}(M), \lambda_{\max}(M)]$ defined by the Raleigh quotients:

$$\lambda_{\min}(M) = \min_{u\neq 0} \frac{(Mu,u)}{(u,u)}, \qquad \lambda_{\max}(M) = \max_{u \neq 0} \frac{(Mu,u)}{(u,u)}.$$

Similarly, if an operator M is A-self-adjoint, then its eigenvalues are real and lie in the interval defined by the Raleigh quotients generated by the A-inner-product. A well-known property is that if M is self-adjoint, then the spectral radius of M, denoted as $\rho(M)$, satisfies $\rho(M) = \|M\|$. This property can also be shown to hold in the A-norm for A-self-adjoint operators.

Lemma 1. *If A is SPD and M is A-self-adjoint, then* $\rho(M) = \|M\|_A$.

Linear methods. Given the equation $Au = f$, where $A \in \mathbf{L}(\mathcal{H},\mathcal{H})$ is SPD, consider the *preconditioned* equation $BAu = Bf$, with $B \in \mathbf{L}(\mathcal{H},\mathcal{H})$. The operator B, the *preconditioner*, is usually chosen so that the linear iteration

$$u^{n+1} = u^n - BAu^n + Bf = (I - BA)u^n + Bf, \tag{3}$$

has some desired convergence properties. The convergence of (3) is determined by the properties of the so-called *error propagation operator*, $E = I - BA$.

We now state a series of simple lemmas that we shall use repeatedly in the following sections. Their short proofs and further references can be found in [5].

Lemma 2. *If A is SPD, then BA is A-self-adjoint if and only if B is self-adjoint.*

Lemma 3. *If A is SPD, then E is A-self-adjoint if and only if B is self-adjoint.*

Lemma 4. *If A and B are SPD, then BA is A-SPD.*

Lemma 5. *If A is SPD and B is self-adjoint, then $\|E\|_A = \rho(E)$.*

Lemma 6. *If E^* is the A-adjoint of E, then $\|E\|_A^2 = \|EE^*\|_A$.*

Lemma 7. *If A and B are SPD and E is A-non-negative, then $\|E\|_A < 1$.*

Lemma 8. *If A is SPD and B is self-adjoint, and E is such that*

$$-C_1(u,u)_A \leq (Eu,u)_A \leq C_2(u,u)_A, \qquad \forall u \in \mathcal{H},$$

for $C_1 \geq 0$ and $C_2 \geq 0$, then $\rho(E) = \|E\|_A \leq \max\{C_1, C_2\}$.

Lemma 9. *If A and B are SPD, then Lemma 8 holds for some $C_2 < 1$.*

The following lemma illustrates why symmetrizing is a bad idea for linear methods. It exposes the convergence rate penalty incurred by symmetrization of a linear method.

Lemma 10. *For any $E \in \mathbf{L}(\mathcal{H}, \mathcal{H})$, it holds that:*

$$\rho(EE) \leq \|EE\|_A \leq \|E\|_A^2 = \|EE^*\|_A = \rho(EE^*).$$

Proof. The first and second inequalities hold for any norm. The first equality follows from Lemma 6, and the second follows from Lemma 1. □

Note that this is an inequality not only for the spectral radii but also for the A-norms of the nonsymmetric and symmetrized error propagators. The lemma illustrates that one may actually see the differing convergence rates early in the iteration as well.

Krylov acceleration of SPD linear methods. The conjugate gradient method was developed by Hestenes and Stiefel [4] as a method for solving linear systems $Au = f$, with SPD operators A. In order to improve convergence, it is common to *precondition* the linear system by an SPD *preconditioning operator* $B \approx A^{-1}$, in which case the generalized or preconditioned conjugate gradient method results. Our goal in this section is to briefly review some relationships between the contraction number of a basic linear preconditioner and that of the resulting preconditioned conjugate gradient algorithm.

We start with the well-known conjugate gradient contraction bound [3]

$$\|e^{i+1}\|_A \;\leq\; 2\left(1 - \frac{2}{1+\sqrt{\kappa_A(BA)}}\right)^{i+1} \|e^0\|_A \;=\; 2\,\delta_{\rm cg}^{i+1}\,\|e^0\|_A,$$

where $\kappa_A(BA)$, the A-condition number of BA, is the ratio of extreme eigenvalues of BA.

The following result gives a bound on the condition number of the operator BA in terms of the extreme eigenvalues of the error propagator $E = I - BA$; such bounds are often used in the analysis of linear preconditioners (cf. Proposition 5.1 in [9]).

Lemma 11. *If A and B are SPD and E is such that*

$$-C_1(u,u)_A \leq (Eu,u)_A \leq C_2(u,u)_A, \qquad \forall u \in \mathcal{H},$$

for $C_1 \geq 0$ and $C_2 \geq 0$, then the above must hold with $C_2 < 1$, and it follows that

$$\kappa_A(BA) \leq \frac{1+C_1}{1-C_2}.$$

Remark 1. Even if a linear method is not convergent, it may still be a good preconditioner. If $C_2 << 1$ and if $C_1 > 1$ does not become too large, then $\kappa_A(BA)$ will be small and the conjugate gradient method will converge rapidly, even though the linear method diverges.

The next result connects the contraction number of the preconditioner to the contraction number of the preconditioned conjugate gradient method (see [10] for a proof).

Lemma 12. *If A and B are SPD and $\|I - BA\|_A \leq \delta < 1$, then $\delta_{\text{cg}} < \delta$.*

Krylov acceleration of nonsymmetric linear methods. The convergence theory of the conjugate gradient iteration requires that the preconditioned operator BA be A-self-adjoint (see [1] for more general conditions), which from Lemma 2 requires that B be self-adjoint. If a Schwarz method is employed which produces a nonsymmetric operator B, then although A is SPD, the theory of the previous section does not apply and a nonsymmetric solver such as conjugate gradients on the normal equations [1], GMRES [6], CGS [7], or Bi-CGstab [8] must be used. Further on, we shall use the preconditioned Bi-CGstab algorithm to accelerate nonsymmetric Schwarz methods. In a sequence of numerical experiments, we shall compare the effectiveness of this approach with unaccelerated symmetric and nonsymmetric Schwarz methods, and with symmetric Schwarz methods accelerated with conjugate gradients.

MULTIPLICATIVE SCHWARZ METHODS

Consider a product operator of the form:

$$E = I - BA = (I - \bar{B}_1 A)(I - B_0 A)(I - B_1 A), \tag{4}$$

where $\bar{B}_1, B_0$, and B_1 are linear operators on $\mathcal{H}$, and where A is, as before, an SPD operator on $\mathcal{H}$. We are interested in conditions for $\bar{B}_1, B_0$, and B_1, which guarantee that the implicitly defined operator B is self-adjoint and positive definite and, hence, can be accelerated by using the conjugate gradient method.

Lemma 13. *Sufficient conditions for symmetry and positivity of operator B, defined by (4), are:*

1. $\bar{B}_1 = B_1^T$;
2. $B_0 = B_0^T$;
3. $\|I - B_1 A\|_A < 1$;

4. B_0 non-negative on $\mathcal{H}$.

Proof. By Lemma 3, in order to prove symmetry of B, it is sufficient to prove that E is A-self-adjoint. By using (1), we get

$$E^* = A^{-1}E^TA = (I - B_1^TA)(I - B_0^TA)(I - \bar{B}_1^TA),$$

which equals E following from conditions 1 and 2.

Next, we prove that $(Bu, u) > 0$, $\forall u \in \mathcal{H}$, $u \neq 0$. Since A is non-singular, this is equivalent to proving that $(BAu, Au) > 0$. Using condition 1, we have that

$$\begin{aligned}
(BAu, Au) &= ((I - E)u, Au) \\
&= (u, Au) - ((I - B_1^TA)(I - B_0A)(I - B_1A)u, Au) \\
&= (u, Au) - ((I - B_0A)(I - B_1A)u, A(I - B_1A)u) \\
&= (u, Au) - ((I - B_1A)u, A(I - B_1A)u) + (B_0w, w),
\end{aligned}$$

where $w = A(I - B_1A)u$. By condition 4, we have that $(B_0w, w) \geq 0$. Condition 3 implies that $((I - B_1A)u, A(I - B_1A)u) < (u, Au)$ for $u \neq 0$. Thus, the first two terms in the sum above are together positive, while the third is non-negative, so that B is positive. □

Multiplicative domain decomposition. Given the finite-dimensional Hilbert space $\mathcal{H}$, consider J spaces $\mathcal{H}_k$, $k = 1, \ldots, J$, together with linear operators $I_k \in \mathbf{L}(\mathcal{H}_k, \mathcal{H})$, $\text{null}(I_k) = \{0\}$, such that $I_k\mathcal{H}_k \subseteq \mathcal{H} = \sum_{k=1}^J I_k\mathcal{H}_k$. We also assume the existence of another space $\mathcal{H}_0$, an associated operator I_0 such that $I_0\mathcal{H}_0 \subseteq \mathcal{H}$, and some linear operators $I^k \in \mathbf{L}(\mathcal{H}, \mathcal{H}_k), k = 0, \ldots, J$. For notational convenience, we shall denote the inner-products on $\mathcal{H}_k$ by $(\cdot, \cdot)$ (without explicit reference to the particular space). Note that the inner products on different spaces need not be related.

In a domain decomposition context, the spaces $\mathcal{H}_k$, $k = 1, \ldots, J$ are typically associated with *local subdomains* of the original domain on which the partial differential equation is defined. The space $\mathcal{H}_0$ is then a space associated with some global coarse mesh. The operators $I_k, k = 1, \ldots, J$ are usually inclusion operators, while I_0 is an interpolation or prolongation operator (as in a two-level MG method). The operators $I^k, k = 1, \ldots, J$ are usually orthogonal projection operators, while I^0 is a restriction operator (again, as in a two-level MG method).

The error propagator of a multiplicative DD method on the space $\mathcal{H}$ employing the subspaces $I_k\mathcal{H}_k$ has the general form [2]

$$E = I - BA = (I - I_J\bar{R}_JI^JA) \cdots (I - I_0R_0I^0A) \cdots (I - I_JR_JI^JA) \,, \tag{5}$$

where $\bar{R}_k$ and R_k, $k = 1, \ldots, J$, are linear operators on $\mathcal{H}_k$ and R_0 is a linear operator on $\mathcal{H}_0$. Usually the operators $\bar{R}_k$ and R_k are constructed so that $\bar{R}_k \approx A_k^{-1}$ and $R_k \approx A_k^{-1}$, where A_k is the operator defining the subdomain problem in $\mathcal{H}_k$. Similarly, R_0 is constructed so that $R_0 \approx A_0^{-1}$. Actually, quite often R_0 is a "direct solve", i.e., $R_0 = A_0^{-1}$. The subdomain problem operator A_k is related to the restriction of A to $\mathcal{H}_k$. We say that A_k satisfies the *Galerkin conditions* or, in a finite element setting, that it is *variationally* defined when

$$A_k = I^kAI_k, \qquad I^k = I_k^T. \tag{6}$$

Recall that the superscript "T" is to be interpreted as the adjoint in the sense of (2), i.e., with respect to the inner-products in $\mathcal{H}$ and $\mathcal{H}_k$.

Propagator (5) can be thought of as the product operator (4) by choosing

$$I - \bar{B}_1 A = \prod_{k=J}^{1} (I - I_k \bar{R}_k I^k A) \ , \quad B_0 = I_0 R_0 I^0 \ , \quad I - B_1 A = \prod_{k=1}^{J} (I - I_k R_k I^k A) \ ,$$

where $\bar{B}_1$ and B_1 are known only implicitly. This identification allows for the use of Lemma 13 to establish sufficient conditions on the subdomain operators $\bar{R}_k$, R_k, and R_0 to guarantee that multiplicative domain decomposition yields an SPD operator B.

Theorem 1. *Sufficient conditions for symmetry and positivity of the multiplicative domain decomposition operator B, defined by (5), are:*

1. $I^k = c_k I_k^T \ , \quad c_k > 0 \ , \quad k = 0, \cdots, J$;
2. $\bar{R}_k = R_k^T \ , \quad k = 1, \cdots, J$;
3. $R_0 = R_0^T$;
4. $\left\| \prod_{k=1}^{J} (I - I_k R_k I^k A) \right\|_A < 1$;
5. R_0 *non-negative on* $\mathcal{H}_0$.

Proof. We show that the conditions of Lemma 13 are satisfied. First, we prove that $\bar{B}_1 = B_1^T$, which, by Lemma 3, is equivalent to proving that $(I - B_1 A)^* = (I - \bar{B}_1 A)$. By using (1), we have

$$\left(\prod_{k=1}^{J} (I - I_k R_k I^k A) \right)^* = A^{-1} \left(\prod_{k=1}^{J} (I - I_k R_k I^k A) \right)^T A = \prod_{k=J}^{1} (I - (I^k)^T R_k^T (I_k)^T A) \ ,$$

which equals $(I - \bar{B}_1 A)$ under conditions 1 and 2 of the theorem. The symmetry of B_0 follows immediately from conditions 1 and 3; indeed,

$$B_0^T = (I_0 R_0 I^0)^T = (I^0)^T R_0^T (I_0)^T = (c_0 I_0) R_0 (c_0^{-1} I^0) = I_0 R_0 I^0 = B_0 \ .$$

By condition 4 of the theorem, condition 3 of Lemma 13 holds trivially. The theorem follows if one realizes that condition 4 of Lemma 13 is also satisfied, since,

$$(B_0 u, u) = (I_0 R_0 I^0 u, u) = (R_0 I^0 u, I_0^T u) = c_0^{-1} (R_0 I^0 u, I^0 u) \geq 0 \ , \quad \forall u \in \mathcal{H} \ .$$

□

Remark 2. Note that one sweep through the subdomains, followed by a coarse problem solve, followed by another sweep through the subdomains in reverse order, gives rise to an error propagator of the form (5). Also, note that no conditions are imposed on the nature of the operators A_k associated with each subdomain. In particular, the theorem *does not* require that the variational conditions be satisfied. The theorem also does not require that the overall multiplicative DD method be convergent.

Remark 3. The results of the theorem apply for operators on general finite-dimensional Hilbert spaces with arbitrary inner-products. They hold in particular for matrix operators on $\mathbb{R}^N$, equipped with the Euclidean inner-product or the discrete L_2 inner-product. In the former case the superscript "T" corresponds to the standard matrix transpose. In the latter case, the matrix representation of the adjoint is a scalar multiple of the matrix transpose; the scalar may be different from unity when the adjoint involves two different spaces, as in the case of prolongation and restriction. This possible constant in the case of the discrete L_2 inner-product is absorbed in the factor c_k in condition 1. This allows for an easy verification of the conditions of the theorem in an actual implementation, where the operators are represented as matrices and where the inner-products do not explicitly appear in the algorithm.

Remark 4. Condition 1 of the theorem (with $c_k = 1$) for $k = 1, \ldots, J$ is usually satisfied trivially for domain decomposition methods. For $k = 0$, it may have to be imposed explicitly. Condition 2 of the theorem allows for several alternatives which give rise to an SPD preconditioner, namely: (1) use of exact subdomain solvers (if A_k is a symmetric operator); (2) use of identical symmetric subdomain solvers in the forward and backward sweeps; and (3) use of the adjoint of the subdomain solver on the second sweep. Condition 3 is satisfied when the coarse problem is symmetric and the solve is an exact one, which is usually the case. If not, the coarse problem solve has to be symmetric. Condition 4 in Theorem 1 is clearly a non-trivial one; it is essentially the assumption that the multiplicative DD method without a coarse space is convergent. Condition 5 is satisfied, for example, when the coarse problem is SPD and the solve is exact.

Multiplicative multigrid. Consider the Hilbert space $\mathcal{H}$ and J spaces $\mathcal{H}_k$ together with operators $I_k \in \mathbf{L}(\mathcal{H}_k, \mathcal{H})$, $\text{null}(I_k) = 0$, such that the spaces $I_k\mathcal{H}_k$ are nested and satisfy $I_1\mathcal{H}_1 \subseteq I_2\mathcal{H}_2 \subseteq \cdots \subseteq I_{J-1}\mathcal{H}_{J-1} \subseteq \mathcal{H}_J \equiv \mathcal{H}$. As before, we denote the $\mathcal{H}_k$-inner-products by $(\cdot, \cdot)$, since it will be clear from the arguments which inner-product is intended. Again, the inner-products are not necessarily related in any way. We assume also the existence of operators $I^k \in \mathbf{L}(\mathcal{H}, \mathcal{H}_k)$.

In a multigrid context, the spaces $\mathcal{H}_k$ are typically associated with a nested hierarchy of successively refined meshes, with $\mathcal{H}_1$ being the coarsest mesh and $\mathcal{H}_J$ being the fine mesh on which the PDE solution is desired. The linear operators I_k are prolongation operators, constructed from given interpolation or prolongation operators that operate between subspaces, i.e., $I_{k-1}^k \in \mathbf{L}(\mathcal{H}_{k-1}, \mathcal{H}_k)$. The operator I_k is then constructed (only as a theoretical tool) as a composite operator

$$I_k = I_{J-1}^J I_{J-2}^{J-1} \cdots I_{k+1}^{k+2} I_k^{k+1}, \quad k = 1, \ldots, J-1. \tag{7}$$

The composite restriction operators I^k, $k = 1, \ldots, J-1$, are constructed similarly from some given restriction operators $I_k^{k-1} \in \mathbf{L}(\mathcal{H}_k, \mathcal{H}_{k-1})$. The coarse problem operators A_k are related to the restriction of A to $\mathcal{H}_k$. As in the case of DD methods, we say that A_k is *variationally* defined or satisfies the *Galerkin conditions* when conditions (6) hold. It is not difficult to see that conditions (6) are equivalent to the following recursively defined variational conditions:

$$A_k = I_{k+1}^k A_{k+1} I_k^{k+1}, \qquad I_{k+1}^k = (I_k^{k+1})^T. \tag{8}$$

when the composite operators I_k appearing in (6) are defined as in (7).

In a finite element setting, conditions (8) can be shown to hold in ideal situations, for both the stiffness matrices and the abstract weak form operators, for a nested sequence of successively refined finite element meshes. In the finite difference or finite volume method setting, conditions (8) must often be imposed algebraically, in a recursive fashion.

The error propagator of a multiplicative V-cycle MG method is defined implicitly as

$$E = I - BA = I - D_J A_J, \tag{9}$$

where $A_J = A$ and where operators $D_k,\ k = 2, \ldots, J$, are defined recursively:

$$\begin{aligned} I - D_k A_k &= (I - \bar{R}_k A_k)(I - I_{k-1}^{k} D_{k-1} I_k^{k-1} A_k)(I - R_k A_k),\ k = 2, \ldots, J, & (10) \\ D_1 &= R_1 \ . & (11) \end{aligned}$$

Operators $\bar{R}_k$ and R_k are linear operators on $\mathcal{H}_k$, usually called *smoothers*. The linear operators $A_k \in L(\mathcal{H}_k, \mathcal{H}_k)$ define the coarse problems. They often satisfy the variational condition (8).

The error propagator (9) can be thought of as an operator of the form (4) with

$$\bar{B}_1 = \bar{R}_J \ , \quad B_0 = I_{J-1}^{J} D_{J-1} I_J^{J-1} \ , \quad B_1 = R_J \ .$$

Such an identification with the product method allows for the use of Lemma 13. The following theorem establishes sufficient conditions for the subspace operators R_k, $\bar{R}_k$, and A_k in order to generate an (implicitly defined) SPD operator B that can be accelerated with conjugate gradients.

Theorem 2. *Sufficient conditions for symmetry and positivity of the multiplicative multigrid operator B, implicitly defined by (9), (10), and (11), are*

1. A_k *is SPD on* $\mathcal{H}_k$, $\quad k = 2, \ldots, J$;
2. $I_k^{k-1} = c_k (I_{k-1}^{k})^T, \quad c_k > 0, \quad k = 2, \ldots, J$;
3. $\bar{R}_k = R_k^T, \quad k = 2, \ldots, J$;
4. $R_1 = R_1^T$;
5. $\|I - R_J A\|_A < 1$;
6. $\|I - R_k A_k\|_{A_k} \leq 1, \quad k = 2, \ldots, J-1$;
7. R_1 *non-negative on* $\mathcal{H}_1$.

Proof. Since $\bar{R}_J = R_J^T$, we have that $\bar{B}_1 = B_1^T$, which gives condition 1 of Lemma 13. Now, B_0 is symmetric if and only if

$$B_0 = I_{J-1}^{J} D_{J-1} I_J^{J-1} = (c_J^{-1} I_J^{J-1})^T D_{J-1}^T (c_J I_{J-1}^{J})^T = B_0^T,$$

which holds under condition 2 and a symmetry requirement for D_{J-1}. We will prove that $D_{J-1} = D_{J-1}^T$ by induction. First, $D_1 = D_1^T$ since $R_1 = R_1^T$. By Lemma 3 and condition 1, D_k is symmetric if and only if $E_k = I - D_k A_k$ is A_k-self-adjoint. By using (1), we have that

$$\begin{aligned} E_k^* &= A_k^{-1} \left((I - \bar{R}_k A_k)(I - I_{k-1}^{k} D_{k-1} I_k^{k-1} A_k)(I - R_k A_k) \right)^T A_k \\ &= (I - \bar{R}_k A_k)(I - (c_k I_{k-1}^{k}) D_{k-1}^T (c_k^{-1} I_k^{k-1}) A_k)(I - R_k A_k), \end{aligned}$$

where we have used conditions 1, 2, and 3. Therefore, $E_k^* = E_k$, if $D_{k-1} = D_{k-1}^T$. Hence, the result follows by induction on k.

Condition 3 of Lemma 13 follows trivially by condition 5 of the theorem.

It remains to verify condition 4 of Lemma 13, namely that B_0 is non-negative. This is equivalent to showing that D_{J-1} is non-negative on $\mathcal{H}_{J-1}$. This will follow again from an induction argument. First, note that $D_1 = R_1$ is non-negative on $\mathcal{H}_1$. Next, we prove that $(D_k v_k, v_k) \geq 0$, $\forall v_k \in \mathcal{H}_k$, or, equivalently, since A_k is non-singular, that $(D_k A_k v_k, A_k v_k) \geq 0$. So, for all $v_k \in \mathcal{H}_k$,

$$
\begin{aligned}
(D_k A_k v_k, A_k v_k) &= (A_k v_k, v_k) - (A_k E_k v_k, v_k) \\
&= (A_k v_k, v_k) - (A_k(I - I_{k-1}^k D_{k-1} I_k^{k-1} A_k)(I - R_k A_k)v_k, (I - R_k A_k)v_k) \\
&= (A_k v_k, v_k) - (A_k(I - R_k A_k)v_k, (I - R_k A_k)v_k) \\
&\quad + (A_k I_{k-1}^k D_{k-1} I_k^{k-1} A_k (I - R_k A_k)v_k, (I - R_k A_k)v_k) \\
&= (v_k, v_k)_{A_k} - (S_k v_k, S_k v_k)_{A_k} + c_k^{-1}(D_{k-1} v_{k-1}, v_{k-1})
\end{aligned}
$$

where $S_k = I - R_k A_k$ and $v_{k-1} = I_k^{k-1} A_k (I - R_k A_k) v_k \in \mathcal{H}_{k-1}$. By condition 6, the first two terms add up to a non-negative value. Hence, D_k is non-negative if D_{k-1} is non-negative. □

Remark 5. As noted earlier in Remark 3, the conditions and conclusions of the theorem can be interpreted completely in terms of the usual matrix representations of the multigrid operators.

Remark 6. Condition 1 of the theorem requires all but the coarsest grid operator to be SPD. This is easily satisfied when they are constructed either by discretization or by explicitly enforcing the Galerkin condition. Condition 2 requires restriction and prolongation to be adjoints, possibly multiplied by an arbitrary constant. Condition 3 of the theorem is satisfied when the number of pre-smoothing steps equals the number of post-smoothing steps and, in addition, one of the following is imposed: (1) use of the same symmetric smoother for both pre- and post-smoothing; or (2) use of the adjoint of the pre-smoothing operator as the post-smoother. Condition 4 requires a symmetric coarsest mesh solver. When the coarsest mesh problem is SPD, the symmetry of R_1 is satisfied when it corresponds to an exact solve (as is typical for MG methods). Condition 5 is a convergence requirement on the fine space smoother. Condition 6 requires the coarse grid smoothers to be non-divergent. The non-negativity requirement for R_1 is a non-trivial one; however, if A_1 is SPD, it is immediately satisfied when the operator corresponds to an exact solve.

ADDITIVE SCHWARZ METHODS

Consider a sum operator of the following form:

$$E = I - BA = I - \omega(B_0 + B_1)A, \qquad \omega > 0, \tag{12}$$

where, as before, A is an SPD operator and B_0 and B_1 are linear operators on $\mathcal{H}$.

Lemma 14. *Sufficient conditions for symmetry and positivity of B, defined in (12), are*

1. B_1 *is SPD in* $\mathcal{H}$*;*

2. B_0 *is symmetric and non-negative on* $\mathcal{H}$*.*

Proof. We have that $B = \omega(B_0 + B_1)$, which is symmetric by the symmetry of B_0 and B_1. Positivity follows since $(B_0 u, u) \geq 0$ and $(B_1 u, u) > 0$, $\forall u \in \mathcal{H}$, $u \neq 0$. □

Additive domain decomposition. We consider the space $\mathcal{H}$ and the J subspaces $I_k\mathcal{H}_k$ such that $I_k\mathcal{H}_k \subseteq \mathcal{H} = \sum_{k=1}^{J} I_k\mathcal{H}_k$. Again, we allow for a "coarse" subspace $I_0\mathcal{H}_0 \subseteq \mathcal{H}$.

The error propagator of an additive DD method on the space $\mathcal{H}$ employing the subspaces $I_k\mathcal{H}_k$ has the general form (see [10])

$$E = I - BA = I - \omega(I_0 R_0 I^0 + I_1 R_1 I^1 + \cdots + I_J R_J I^J)A. \tag{13}$$

The operators R_k are constructed in such a way that $R_k \approx A_k^{-1}$, where the A_k are the subdomain problem operators. Propagator (13) can be thought of as the sum method (12) by taking $B_0 = I_0 R_0 I^0$ and $B_1 = \sum_{k=1}^{J} I_k R_k I^k$. This identification allows for the use of Lemma 14 in order to establish conditions to guarantee that additive domain decomposition yields an SPD preconditioner. Before we state the main theorem, we need the following lemma, which characterizes the splitting of $\mathcal{H}$ into subspaces $I_k\mathcal{H}_k$ in terms of a positive *splitting constant* S_0.

Lemma 15. *Given any* $v \in \mathcal{H}$*, there exists a splitting* $v = \sum_{k=1}^{J} I_k v_k$*,* $v_k \in \mathcal{H}_k$*, and a constant* $S_0 > 0$ *such that*

$$\sum_{k=1}^{J} \|I_k v_k\|_A^2 \leq S_0 \|v\|_A^2. \tag{14}$$

Proof. Since $\sum_{k=1}^{J} I_k\mathcal{H}_k = \mathcal{H}$, we can construct subspaces $\mathcal{V}_k \subseteq \mathcal{H}_k$ such that $I_k\mathcal{V}_k \cap I_l\mathcal{V}_l = \{0\}$, for $k \neq l$ and $\mathcal{H} = \sum_{k=1}^{J} I_k\mathcal{V}_k$. Any $v \in \mathcal{H}$, can be decomposed uniquely as $v = \sum_{k=1}^{J} I_k v_k$, $v_k \in \mathcal{V}_k$. Define the projectors $Q_k \in L(\mathcal{H}, I_k\mathcal{V}_k)$ such that $Q_k v = I_k v_k$. Then,

$$\sum_{k=1}^{J} \|I_k v_k\|_A^2 = \sum_{k=1}^{J} \|Q_k v\|_A^2 \leq \sum_{k=1}^{J} \ \|Q_k\|_A^2 \ \|v\|_A^2.$$

Hence, the result follows with $S_0 = \sum_{k=1}^{J} \|Q_k\|_A^2$. □

Theorem 3. *Sufficient conditions for symmetry and positivity of the additive domain decomposition operator* B*, defined in (13), are*

1. $I^k = c_k I_k^T, \quad c_k > 0, \quad k = 0, \ldots, J$*;*

2. R_k *is SPD on* $\mathcal{H}_k$*,* $\quad k = 1, \ldots, J$*;*

3. R_0 *is symmetric and non-negative on* $\mathcal{H}_0$*.*

Proof. Symmetry of B_0 and B_1 follow trivially from the symmetry of R_k and R_0 and from $I^k = c_k I_k^T$. That B_0 is non-negative on $\mathcal{H}$ follows immediately from the non-negativity of R_0 on $\mathcal{H}_0$.

Finally, we prove positivity of B_1. Define $A_k = I^k A I_k$, $k = 1, \ldots, J$. By condition 1 and the full rank nature of I_k, we have that A_k is SPD. Now, since R_k is also SPD, the product $R_k A_k$ is A_k-SPD. Hence, there exists an $\omega_0 > 0$ such that $0 < \omega_0 < \lambda_i(R_k A_k),\ k = 1, \ldots, J$. This is used together with (14) to bound the sum

$$\sum_{k=1}^{J} c_k^{-1}(R_k^{-1}v_k, v_k) = \sum_{k=1}^{J} c_k^{-1}(A_k A_k^{-1} R_k^{-1} v_k, v_k) \leq \sum_{k=1}^{J} c_k^{-1}(A_k v_k, v_k) \max_{v_k \neq 0} \frac{(A_k A_k^{-1} R_k^{-1} v_k, v_k)}{(A_k v_k, v_k)}$$

$$\leq \sum_{k=1}^{J} c_k^{-1}\omega_0^{-1}(A_k v_k, v_k) = \sum_{k=1}^{J} \omega_0^{-1}(A I_k v_k, I_k v_k) = \sum_{k=1}^{J} \omega_0^{-1}\|I_k v_k\|_A^2 \leq \left(\frac{S_0}{\omega_0}\right)\|v\|_A^2,$$

with $v = \sum_{k=1}^{J} I_k v_k$. We can now employ this result to establish positivity of B_1:

$$\|v\|_A^2 = (Av, v) = \sum_{k=1}^{J}(Av, I_k v_k) = \sum_{k=1}^{J}(I_k^T A v, v_k) = \sum_{k=1}^{J}(R_k c_k^{1/2} I_k^T A v, R_k^{-1} c_k^{-1/2} v_k).$$

By using the Cauchy-Schwarz inequality first in the R_k-inner-product and then in $\mathbb{R}^J$, we have that

$$\|v\|_A^2 \leq \left(\sum_{k=1}^{J}(R_k R_k^{-1} c_k^{-1/2} v_k, R_k^{-1} c_k^{-1/2} v_k)\right)^{1/2} \left(\sum_{k=1}^{J}(R_k c_k^{1/2} I_k^T A v, c_k^{1/2} I_k^T A v)\right)^{1/2}$$

$$\leq \left(\frac{S_0}{\omega_0}\right)^{1/2} \|v\|_A \left(\sum_{k=1}^{J}(I_k R_k c_k I_k^T A v, A v)\right)^{1/2} = \left(\frac{S_0}{\omega_0}\right)^{1/2} \|v\|_A \ (B_1 A v, A v)^{1/2} .$$

Finally, we divide by $\|v\|_A$ and square to obtain

$$(B_1 A v, A v) \ \geq \ \frac{\omega_0}{S_0} \|v\|_A^2 > 0 \ , \quad \forall v \in \mathcal{H}, \quad v \neq 0.$$

□

Remark 7. Condition 1 is naturally satisfied for $k = 1, \ldots, J$, with $c_k = 1$, since the associated I_k and I^k are usually inclusion and orthogonal projection operators (which are natural adjoints when the inner-products are inherited from the parent space, as in domain decomposition). The fact that $I^0 = c_0 I_0^T$ needs to be established explicitly. Condition 2 requires the use of SPD subdomain solvers. The condition will hold, for example, when the subdomain solve is exact and the subdomain problem operator is SPD. (The latter is naturally satisfied by condition 1 and the full rank nature of I_k.) Finally, condition 3 is nontrivial and needs to be checked explicitly. The condition holds when the coarse space problem operator is SPD and the solve is exact. Note that variational conditions are not needed for the coarse space problem operator.

Additive multigrid. Given are the Hilbert space $\mathcal{H}$ and $J-1$ nested subspaces $I_k\mathcal{H}_k$ such that $I_1\mathcal{H}_1 \subseteq I_2\mathcal{H}_2 \subseteq \cdots \subseteq I_{J-1}\mathcal{H}_{J-1} \subseteq \mathcal{H}_J \equiv \mathcal{H}$. The operators I_k and I^k are the usual linear operators between the different spaces, as in the previous sections.

The error propagator of an additive MG method is defined explicitly:

$$E = I - BA = I - \omega(I_1R_1I^1 + I_2R_2I^2 + \cdots + I_{J-1}R_{J-1}I^{J-1} + R_J)A. \tag{15}$$

This can be thought of as the sum method analyzed earlier by taking $B_0 = \sum_{k=1}^{J-1} I_kR_kI^k$ and $B_1 = R_J$. This identification allows for the use of Lemma 14 to establish sufficient conditions to guarantee that additive MG yields an SPD preconditioner.

Theorem 4. *Sufficient conditions for symmetry and positivity of the additive multigrid operator B defined in (15) are*

1. $I^k = c_kI_k^T, \quad c_k > 0, \quad k = 1, \dots, J-1;$
2. R_J *is SPD in* $\mathcal{H}$;
3. R_k *is symmetric non-negative in* $\mathcal{H}_k, \quad k = 1, \dots, J-1.$

Proof. Symmetry of B_0 and B_1 is obvious. B_1 is positive by condition 2. Non-negativity of B_0 follows from

$$(B_0u, u) = \sum_{k=1}^{J-1}(I_kR_k(c_kI_k)^Tu, u) = \sum_{k=1}^{J-1} c_k(R_kI_k^Tu, I_k^Tu) \geq 0, \quad \forall u \in \mathcal{H}, u \neq 0.$$

□

Remark 8. Condition 1 of the theorem has to be imposed explicitly. Conditions 2 and 3 require the smoothers to be symmetric. The positivity of R_J is satisfied when the fine grid smoother is convergent, although this is not a necessary condition. The non-negativity of R_k, $k < J$, has to be checked explicitly. When the coarse problem operators A_k are SPD, this condition is satisfied, for example, when the smoothers are non-divergent. Note that variational conditions for the subspace problem operators are not required.

NUMERICAL RESULTS

The Poisson-Boltzmann equation describes the electrostatic potential of a biomolecule lying in an ionic solvent. This nonlinear elliptic equation for the dimensionless electrostatic potential $u(\mathbf{r})$ has the form

$$-\nabla \cdot (\epsilon(\mathbf{r})\nabla u(\mathbf{r})) + \bar{\kappa}^2 \sinh(u(\mathbf{r})) = \left(\frac{4\pi e_c^2}{k_BT}\right) \sum_{i=1}^{N_m} z_i\delta(\mathbf{r} - \mathbf{r}_i), \quad \mathbf{r} \in \mathbb{R}^3, \qquad u(\infty) = 0.$$

The coefficients appearing in the equation are discontinuous by orders of magnitude. The placement and magnitude of atomic charges are represented by source terms involving delta-functions. Analytical techniques are used to obtain boundary conditions on a finite domain boundary.

We will compare several MG and DD methods for a two-dimensional, linearized form of the Poisson-Boltzmann problem, modeling a molecule with three point charges. The surface of the molecule is such that the discontinuities do not align with the coarsest mesh or with

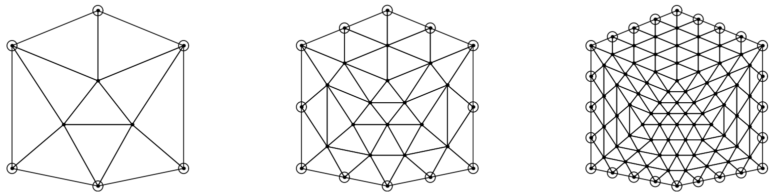

Figure 1: Example 1: Nested finite element meshes for MG.

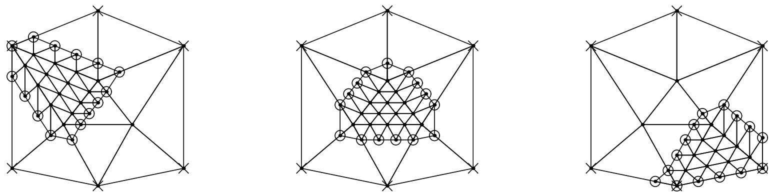

Figure 2: Example 1: Overlapping subdomains for DD.

the subdomain boundaries. Beginning with the coarse mesh shown on the left in Figure 1, we uniformly refine the initial mesh of 10 elements (9 nodes) five times, leading to a fine mesh of 2560 elements (1329 nodes). Piecewise linear finite elements, combined with one-point Gaussian quadrature, are used to discretize the problem. The three coarsest meshes used to formulate the MG methods are given in Figure 1. For the DD methods, the subdomains, corresponding to the initial coarse triangulation, are given a small overlap of one fine mesh triangle. The DD methods also employ a coarse space constructed from the initial triangulation. Figure 2 shows three overlapping subdomains overlaying the initial coarse mesh. Computed results are presented in Tables 1 to 4. Given for each experiment is the number of iterations required to satisfy the error criterion (reduction of the A-norm of the error by 10^{-10}). We report results for the unaccelerated, CG-accelerated, and Bi-CGstab-accelerated methods. The execution time differs for each method; normalized costs are tabulated in [5].

Multiplicative multigrid. The results for multiplicative V-cycle MG are presented in Table 1. Each row corresponds to a different smoothing strategy and is annotated by (ν_1, ν_2), with ν_1 pre-smoothing sweeps and ν_2 post-smoothing sweeps. An "f" indicates the use of a single forward Gauss-Seidel sweep, while a "b" denotes the use of the adjoint of the latter, i.e., a backward Gauss-Seidel sweep. Two series of results are given. For the first set, we explicitly imposed the Galerkin conditions when constructing the coarse operators. In this case, the multigrid algorithm is guaranteed to converge (cf. [5]). In the second series of tests (corresponding to the numbers in parentheses) the coarse mesh operators are constructed using standard finite element discretization. In that case, Galerkin conditions are not satisfied everywhere due to coefficient discontinuities appearing within coarse elements; hence, the MG method may diverge (DIV).

The unaccelerated MG results clearly illustrate the symmetry penalty given in Lemma 10.

Table 1: Example 1: Multiplicative MG with variational (discretized) coarse problem

ν_1	ν_2	UNACCEL		CG		Bi-CGstab	
f	0	65	(DIV)	$\gg$100	($\gg$100)	14	(16)
f	b	55	(DIV)	16	(18)	10	(15)
f	f	40	(31)	30	($\gg$100)	9	(9)
ff	0	39	(48)	$\gg$100	($\gg$100)	8	(10)
fb	0	53	(DIV)	$\gg$100	($\gg$100)	10	(11)
0	ff	39	(29)	29	($\gg$100)	8	(9)
0	fb	53	(DIV)	17	(99)	10	(12)
fb	fb	34	(27)	12	(13)	8	(8)
ff	bb	28	(18)	11	(11)	7	(7)
ff	ff	24	(15)	12	(12)	6	(6)
fff	f	24	(15)	17	(27)	6	(6)
ffff	0	25	(17)	$\gg$100	($\gg$100)	7	(6)

Table 2: Example 1: Multiplicative DD with variational (discretized) coarse problem

Accel.	subdomain solve	forw		forw/back		forw/forw	
UNACCEL	exact	40	(42)	38	(39)	20	(21)
	symmetric	279	(282)	146	(149)	140	(141)
	adjointed	–		110	(112)	102	(103)
	nonsymmetric	189	(191)	102	(104)	95	(96)
CG	exact	$\gg$500	($\gg$500)	13	(13)	20	(20)
	symmetric	140	(56)	24	(24)	29	(27)
	adjointed	–	–	21	(21)	25	(26)
	nonsymmetric	135	(83)	22	(23)	28	(28)
Bi-CGstab	exact	9	(9)	9	(9)	6	(6)
	symmetric	23	(23)	17	(16)	16	(16)
	adjointed	–	–	14	(14)	14	(13)
	nonsymmetric	19	(20)	13	(13)	13	(13)

The nonsymmetric methods are always superior to the symmetric ones (the cases (f,b), (ff,bb), and (fb,fb)). Note that minimal symmetry (ff,bb) leads to a better convergence than maximal symmetry (fb,fb). The correctness of Lemma 10 is illustrated by noting that two iterations of the (f,0) strategy are actually faster than one iteration of the (f,b) strategy; also, compare the (ff,0) strategy to the (ff,bb) one. The CG-acceleration leads to a guaranteed reduction in iteration count for the symmetric preconditioners (see Lemma 12). We observe that the unaccelerated method need not be convergent for CG to be effective. CG appears to also accelerate some non-symmetric linear methods. Yet, it seems difficult to predict failure or success beforehand in such cases. The most robust method appears to be the Bi-CGstab method. Note the tendency

Table 3: Example 1: Additive MG with variational (discretized) coarse problem

ν	UNACCEL		CG		Bi-CGstab	
f	175	(≫1000)	≫100	(≫100)	23	(52)
ff	110	(≫1000)	119	(168)	19	(43)
fb	146	(≫1000)	34	(54)	23	(49)
fff	95	(≫1000)	28	(67)	17	(37)
ffbb	100	(≫1000)	27	(47)	17	(34)
fbfb	95	(≫1000)	28	(48)	20	(43)

Table 4: Example 1: Additive DD with variational (discretized) coarse problem

subdomain solve	UNACCEL		CG		Bi-CGstab	
exact	≫1000	(≫1000)	34	(34)	25	(27)
symmetric	≫1000	(≫1000)	57	(57)	50	(49)
nonsymmetric	≫1000	(≫1000)	69	(65)	38	(41)

to favor the nonsymmetric V-cycle strategies. Overall, the fastest method proves to be the Bi-CGstab-acceleration of a (very nonsymmetric) V(1,0)-cycle.

Multiplicative domain decomposition. Results for multiplicative DD are given in Table 2. In the column "forw" the iteration counts reported were obtained with a single sweep though the subdomains on each multiplicative DD iteration. The other columns correspond to a symmetric forward/backward sweep or to two forward sweeps. Four different subdomain solvers are used: an *exact* solve, a *symmetric* method consisting of two symmetric Gauss-Seidel iterations, a *nonsymmetric* method consisting of four Gauss-Seidel iterations, and, finally, a method using four forward Gauss-Seidel iterations in the forward subdomain sweep and using their *adjoint* (i.e., four backward Gauss-Seidel iterations) in the backward subdomain sweep. The latter leads to a symmetric iteration; see Remark 2. Note that the cost of the three inexact subdomain solvers is identical.

Although apparently not as sensitive to operator symmetries as MG, the same conclusions can be drawn for DD as for MG. In particular, the symmetry penalty is seen for the pure DD results. Lemma 10 is confirmed since two iterations in the column "forw" are always more efficient than one iteration of the corresponding method in column "forw/back." The CG results indicate that using minimal symmetry (the "adjointed" column) is a more effective approach than the fully symmetric one (the "symmetric" column). The most robust acceleration is the Bi-CGstab one.

Additive multigrid. Results obtained with an additive multigrid method are reported in Table 3. The number and nature of the smoothing strategy is given in the first column of the table.

In the case of an unaccelerated additive method, the selection of a good damping parameter is crucial for convergence of the method. We did not search extensively for an optimal parameter; a selection of $\omega = 0.45$ seemed to provide good results in the case when the coarse problem was variationally defined. No ω-value leading to satisfactory convergence was found in the case when the coarse problems were obtained by discretization. In the case of CG acceleration the observed convergence behavior was completely independent of the choice of ω; see Remark 2. The symmetric methods ($\nu = fb, ffbb, fbfb$) are accelerated very well. Some of the nonsymmetric methods are accelerated too, especially when the number of smoothing steps is sufficiently large. The best method overall appears to be the Bi-CGstab acceleration of the nonsymmetric multigrid method with a single forward Gauss-Seidel sweep on each grid-level.

Additive domain decomposition. The results for additive DD are given in Table 4. The subdomain solver is either an exact solver, a symmetric solver based on two symmetric (forward/backward) Gauss-Seidel sweeps, or a nonsymmetric solver based on four forward Gauss-Seidel iterations. No value of ω was found that led to satisfactory convergence of the unaccelerated method. The CG-acceleration performs well when the linear method is symmetric and worse if nonsymmetric. Again, the best overall method is the Bi-CGstab-acceleration of the nonsymmetric additive solver.

REFERENCES

[1] S. F. Ashby, T. A. Manteuffel, and P. E. Saylor. A taxonomy for conjugate gradient methods. *SIAM J. Numer. Anal.*, 27(6):1542–1568, 1990.

[2] M. Dryja and O. B. Widlund. Towards a unified theory of domain decomposition algorithms for elliptic problems. In *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, T. F. Chan, R. Glowinski, J. Périaux, and O. B. Widlund, eds. SIAM, Philadelphia, PA, pp. 3-21, 1989.

[3] W. Hackbusch. *Iterative Solution of Large Sparse Systems of Equations.* Springer-Verlag, Berlin, Germany, 1994.

[4] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research of NBS*, 49:409–435, 1952.

[5] M. Holst and S. Vandewalle. Schwarz methods: to symmetrize or not to symmetrize. *SIAM J. Numer. Anal.* (to appear).

[6] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869, 1986.

[7] P. Sonneveld. CGS: A fast Lanczos-type solver for nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 10:36–52, 1989.

[8] H. A. van der Vorst. BI-CGSTAB: A fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 13(2):631–644, 1992.

[9] J. Xu. *Theory of Multilevel Methods.* Ph.D. thesis, Technical Report AM 48, Department of Mathematics, Penn State University, University Park, PA, July 1989.

[10] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, 1992.