# Smoothed Aggregation Solvers for Anisotropic Diffusion

Jacob B. Schroder[1]

[1] *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, L-561, Livermore, CA 94551, U.S.A.*

## SUMMARY

A smoothed-aggregation-based algebraic multigrid solver for anisotropic diffusion problems is presented. Algebraic multigrid (AMG) is a popular and effective method for solving sparse linear systems that arise from discretizing partial differential equations. However, while AMG was designed for elliptic problems, the case of non-grid-aligned anisotropic diffusion is not adequately addressed by existing methods. To achieve scalable performance, it is shown that neither new coarsening nor new relaxation strategies are necessary. Instead, a novel smoothed aggregation approach is developed that combines long-distance interpolation, coarse-grid injection, and an energy-minimization strategy that finds the interpolation weights. Previously developed theory by Falgout and Vassilevski is used to discern that existing coarsening strategies are sufficient, but that existing interpolation methods are not. In particular, an interpolation quality measure tracks "closeness" to the ideal interpolant and guides the interpolation sparsity pattern choice. While the interpolation quality measure is computable for only small model problems, an inexact, but computable, measure is proposed for larger problems. This paper concludes with encouraging numerical results that also potentially show broad applicability, e.g., for linear elasticity.
Copyright © 2011 John Wiley & Sons, Ltd.

## 1. INTRODUCTION

Algebraic multigrid (AMG), e.g., classical AMG [1, 2] and smoothed aggregation (SA) [3, 4], is an efficient solution method for the large, sparse systems that arise from discretizing partial differential equations (PDEs). Classical multigrid methods were developed for symmetric positive-definite (SPD) systems that arise from elliptic PDEs; however, not all elliptic PDEs are adequately addressed by existing multigrid methods. For example, non-grid-aligned anisotropic diffusion is known to be problematic for even simple model problems, such as 2D rotated anisotropic diffusion defined by

$$-(c^2 + \epsilon s^2)u_{xx} - 2(1-\epsilon)cs\, u_{xy} - (\epsilon c^2 + s^2)u_{yy} = f, \tag{1}$$

where $\epsilon = 0.001$, $c = \cos(\theta)$, $s = \sin(\theta)$, and $\theta$ is the angle of rotation. When (1) is discretized with bilinear (Q1) finite elements on a regular grid of the unit box, the results using SA are disappointing because scalability is lost for the non-grid-aligned case (cf. Table IV), thus, motivating the search

---

for a better solver. Additionally, as shown in Section 5, classical SA deteriorates further for more real world applications. The model problems for this paper use this Q1 discretization for (1) and three representative angles ($\theta \in \{0, 3\pi/16, 4\pi/16\}$) that expose the basic nature of multigrid for non-grid-aligned anisotropic diffusion.

One possibility for improvement is more robust coarsening strategies [5, 6]. For the grid-aligned case, it is well-known [7, 8] that a robust strategy semi-coarsens in the direction of strong diffusion and yields scalability (e.g., when $\theta = 0$ in Table IV). These advanced coarsening strategies target scalability by also intelligently coarsening in the non-grid-aligned case. However, this has proved to be only a partial fix, as evidenced in Table IV where [6] is used. Additionally, as shown in Section 2.2, AMG theory [9] indicates that these existing robust coarsening strategies are sufficient for the linear discretizations considered here.

Another robust coarsening strategy is compatible relaxation (CR) [10, 11], which has been applied to anisotropic diffusion problems while being coupled to energy-minimizing interpolation strategies [12, 13]. However, these two approaches, while offering improvement, also leave scalability for the non-grid-aligned case unresolved.

Geometric multigrid methods offer another path to scalability for some anisotropic problems. One such possibility is an approach [14] for non-elliptic and singular perturbation problems, which uses a related, but geometric, strategy to that used here. In both cases, the weak approximation property [15], which roughly stipulates that the accuracy of interpolation for a fine-grid eigenmode must be proportional to the corresponding fine-grid eigenvalue, is quantified and used to pinpoint that interpolation must be improved. The approach [14] goes further by quantifying how aliasing algebraically oscillatory modes on the fine grid to algebraically smooth modes on the coarse grid is detrimental. Additionally, the detrimental effects of characteristic modes that are significantly smoother in the direction of strong diffusion than in the direction of weak diffusion are identified and mitigated.

Another geometric possibility is line relaxation, and in the algebraic setting, it is also possible to construct "lines" based on a strength-of-connection concept. Additionally, the work [16] shows improved performance for model non-grid-aligned anisotropic diffusion problems. The geometric-based strength-of-connection concept, associated residual transfer operators, and coarse-grid operators that match the amount of viscosity on the fine grid are shown to be effective. In particular, one iteration of the geometric full multigrid cycle exhibits textbook efficiency for a fixed grid size by reducing the algebraic error to the level of discretization error. However, because the focus of this paper is on purely algebraic methods, and because of the difficulty of effectively parallelizing line smoothers, these options are not considered. Yet another geometric option is AMGe [17], which exhibits improved performance for anisotropic problems, but requires availability of element stiffness matrices, which is not presumed here.

Thus, the current state does not provide an algebraic and scalable scheme for non-grid-aligned anisotropic diffusion. Given the broad need for such solvers (e.g., Poisson operators on stretched meshes [18] and computational fluid dynamics [19, 20]), this paper explores a promising new direction. Scalable multigrid relies on the complementary relationship between relaxation and interpolation. Therefore, the lack of scalability must be the result of some combination of inadequate relaxation, coarsening, or interpolation. Given that existing AMG theory indicates (cf. Section 3) that existing advanced coarsening strategies are sufficient, improving coarsening is not pursued. Additionally, since line relaxation, as mentioned above, is impractical, this paper instead focuses on improving interpolation, with the following specific contributions. A measure [9] used here tracks "closeness" to the ideal interpolant and quantifies the poor quality of existing interpolation methods. This measure is then used to guide the choice for the interpolation sparsity pattern. While the interpolation quality measure is only computable for small model problems, an inexact, but computable, measure is proposed for general problems. The result is a novel long-distance interpolation approach, coupled with coarse-grid injection, where the interpolation weights are determined by the energy-minimizing prolongation smoother [13]. Once the ideal interpolant is adequately approximated, the numerical results indicate that improved interpolation is sufficient for

scalability when using linear discretizations. In other words, neither relaxation nor coarsening need improvement.

In Section 2, a brief overview of SA is given. In Section 3, it is shown that existing coarsening strategies are sufficient. In Section 4, the new interpolation strategy is motivated and then specified. In Section 5, supporting numerical results are given. In Section 6, future work is discussed and, in Section 7, concluding remarks are made.

## 2. SMOOTHED AGGREGATION OVERVIEW

An overview of SA is now given. The presentation follows standard SA, but uses newer techniques such as the energy-minimizing prolongation smoother [13] and the strength measure [6]. Also, a root node perspective is employed to induce coarse-level injection.

Generally speaking, multigrid automatically constructs a hierarchy of coarse sets of degrees-of-freedom and interpolation operators. When relaxation and interpolation are complementary, i.e., when interpolation is accurate for modes that are slow to be reduced by relaxation, rapid convergence occurs. The low energy (or algebraically smooth) error left by relaxation is accurately interpolated to the next coarser level, where it is effectively reduced. Optimality occurs when this procedure is made recursive, with a coarsest level of trivial size.

Here, standard V-cycles are used for the recursive solve phase. First, pre-smoothing (e.g., Gauss-Seidel or weighted Jacobi) occurs on the finest level to reduce high energy error. The residual equation is then interpolated to the first coarse level and solved recursively. The recursive process halts once the coarsest level has been reached, and a direct solve is done. The fine-level solution is then updated using the interpolated error correction from the coarse level. Finally, a step of post-smoothing occurs.

The SA setup phase is outlined in Algorithm 1. Level 0 is the finest level; $P_\ell$ denotes prolongation from level $\ell + 1$ to level $\ell$; $P_\ell^*$ denotes restriction from level $\ell$ to level $\ell + 1$; and $A_\ell$ is the operator on level $\ell$. Subscripts are dropped when the discussion is invariant to the level. SA begins with a set of user-provided near-null-space modes, $B_0$, which are typically null-space modes of the governing PDE without boundary conditions, and are, thus, only rough representations of the true near-null-space of the linear system. For example, $B_0$ is the constant for diffusion and the rigid-body-modes for elasticity. Therefore, $B$ is pre-smoothed in Algorithm 1 for improvement, especially near the domain boundaries.

Next, the coarse-grid is constructed by aggregation, which corresponds to a non-overlapping covering, $\mathcal{N}$, of the matrix graph, i.e., each vertex is assigned to an aggregate, typically through a greedy procedure. The graph of $A$ may be used, but it is more robust to use the graph of a strength-of-connection matrix $S$. The matrix $S$ is based on $A$, but reflects the removal of entries $(i, j)$, between which algebraically smooth error cannot be well-approximated with $B$ [6]. The $(i, j)$ that are removed are dubbed weak connections and those that remain are dubbed strong connections.

Aggregation is then used to induce a tentative prolongation, $T$, that exactly interpolates $B$ with $T_\ell B_{\ell+1} = B_\ell$. This is achieved by injecting $B$ into the aggregation pattern, $\mathcal{N}$, i.e.,

$$T_\ell = \begin{bmatrix} Q^{(1)} & 0 & \cdots & \\ 0 & Q^{(2)} & 0 & \cdots \\ & & \ddots & \\ & \cdots & 0 & Q^{(k)} \end{bmatrix}, \quad B_{\ell+1} = \begin{bmatrix} R^{(1)} \\ R^{(2)} \\ \vdots \\ R^{(k)} \end{bmatrix}, \quad (2)$$

where $k$ is the total number of aggregates, $B^{(i)}$ is the submatrix of $B_\ell$ obtained by taking only rows assigned to the $i^{th}$ aggregate and $B^{(i)} = Q^{(i)} R^{(i)}$ via a QR factorization. Thus, $T$ is block diagonal, where the blocks correspond to only those degrees-of-freedom within an aggregate.

Prolongation smoothing then produces $P$ by improving $T$ with the goal of making span$(P)$ more accurate for low energy modes. As a result, the sparsity pattern of $P$ is typically expanded from the block diagonal form in (2). Classical SA uses one iteration of weighted Jacobi, i.e., $P = (I - \omega D^{-1} A) T$. Here, a more sophisticated energy-minimization procedure is used to smooth

$T$, and is outlined in Section 2.1. Finally, Galerkin coarsening is used to construct the operator for the next coarse level.

---

**Algorithm 1:** `sa_setup`$(A_0, B_0)$

---

1  **for** $\ell = 0$ **to** $\ell_{max}$ **do**
2      smooth $\beta$ times on $A_\ell B_\ell = \mathbf{0}$
3      $S_\ell = $ `strength`$(A_\ell)$
4      $\mathcal{N}_\ell = $ `aggregate`$(S_\ell)$
5      $T_\ell, B_{\ell+1} = $ `inject_modes`$(\mathcal{N}_\ell, B_\ell)$
6      $P_\ell = $ `energy_min`$(A_\ell, T_\ell)$
7      $A_{\ell+1} = P_\ell^* A_\ell P_\ell$
8  **return** $A_0, \ldots, A_\ell, P_0, \ldots, P_{\ell-1}$

---

### 2.1. F/C-Based Energy-Minimization Prolongation Smoothing

This section outlines the energy-minimizing prolongation smoother used here. To begin, the root node perspective is presented, which induces an $F/C$-style $P$. This perspective views each aggregate as being centered around a node, and then constructs a $P$ that injects from a coarse-level node to this fine-level root node. The root node of an aggregate is the seed node chosen during the greedy aggregation algorithm when constructing that aggregate. The root nodes and non-root nodes are analogous to the $C$-points and $F$-points, respectively, from classical AMG. Such a $P$ is easily constructed. First, take $T$ and scale each column so that the entry corresponding to each root node row is one. Second, to preserve exact interpolation, $T_\ell B_{\ell+1} = B_\ell$, $B_{\ell+1}$ must be scaled accordingly to equal injection [2] of $B_\ell$. Finally, the smoothing of $T$ only occurs for non-root node rows (i.e. $F$-rows), thus, yielding a $P$ that can be row-permuted to the $F/C$-style sparsity pattern, where $I$ denotes coarse-grid injection,

$$P = \begin{bmatrix} P_F \\ I \end{bmatrix}. \tag{3}$$

A brief description of the energy-minimizing prolongation smoother, as used here, is now given. For more details, see [13]. Since A is SPD, projected CG is used to compute

$$\operatorname*{argmin}_{P \in V} \sum_j \|P_{(j)}\|_A^2, \tag{4}$$

where $P_{(j)}$ is the $j^{th}$ column and $V$ is a Krylov subspace over which two constraints are enforced: $P_\ell B_{\ell+1} = B_\ell$ and a sparsity pattern constraint. $\check{V}$ has the form

$$\{X(I \otimes A)\check{T}, (X(I \otimes A))^2\check{T}, \ldots\}, \tag{5}$$

where the ˇ operator denotes column-wise conversion of a matrix to a column-vector, e.g., $\check{T} = [T_{(1)}^T \ T_{(2)}^T \ T_{(3)}^T \ldots]^T$ and $T_{(j)}$ denotes the $j^{th}$ column of $T$. The identity matrix, $I$, has size equal to the coarse-grid size and the projection operator, $X$, enforces the constraints.

The projection, $X$, enforces the near-null-space preservation constraint, $P_\ell B_{\ell+1} = B_\ell$, as follows. The initial interpolation guess, $T$, satisfies the near-null-space constraint by construction (see equation (2)). Thus, the purpose of $X$ is to ensure that this property is not perturbed, i.e., that all subsequent updates, $U \in V$, to $T$ satisfy $U_\ell B_{\ell+1} = 0$. To accomplish this, the action of $X$ when applied to some $U$ essentially right multiplies only the allowed nonzero portion of the $i^{th}$ row by the $\ell_2$-projection $(I - B^{(i)}((B^{(i)})^* B^{(i)})^{-1}(B^{(i)})^*)$. The submatrix $B^{(i)}$ is obtained from $B_{\ell+1}$ by taking only the rows assigned to the allowed nonzero pattern of this $i^{th}$ row. The importance of enforcing this accurate interpolation of certain known low-energy mode(s), most commonly the

---

[2]It is assumed here that $B_0$ has only one column. For a discussion of the multiple column case, see Section 6.2.

constant, is a common feature of multigrid interpolation, e.g., classical AMG, and classical SA, and is similarly important here for the overall success of energy-minimization.

The projection, $X$, enforces the sparsity pattern constraint by zeroing out those entries not present in the graph of

$$\begin{bmatrix} (|S|^d|T|)_F \\ I \end{bmatrix}, \tag{6}$$

where $d \in \{1, 2, \ldots\}$, $|\cdot|$ denotes an element-wise absolute value and $(\cdot)_F$ represents restriction to $F$-rows. By increasing $d$, the sparsity pattern grows in the direction of strong connections and coarse-grid injection is preserved. This flexibility for larger sparsity patterns is crucial to the success of the proposed approach. Typically, $2d$ iterations of the energy-minimization process are taken, with the complexity of the whole process on the order of classical prolongation smoothing [21], $P = p_d(A)T$, where $p_d(A)$ is a $d^{th}$ degree polynomial in $A$ constructed to reduce energy, e.g., Chebyshev.

An important difference with classical prolongation smoothing is that the near-null-space modes are exactly incorporated into span($P$). On account of this, the near-null-space modes are pre-smoothed in Algorithm 1, so that inaccuracies, especially near the boundaries, are alleviated. For many problems, this pre-smoothing of the near-null-space is unnecessary, but note that for the case of the Helmholtz problem [22] and some 3D elasticity problems, pre-relaxing $B$ is beneficial.

## 2.2. Benefits of the $F/C$-Based Approach

The $F/C$-based approach has some attractive properties. The identity block in (3) guarantees that the columns of $P$ are linearly independent, which is important during the energy-minimization process. Moreover, the ideal $P$ (cf. [9]) for (3) is given by

$$P_{ideal} = \begin{bmatrix} -A_{FF}^{-1}A_{FC} \\ I \end{bmatrix}, \tag{7}$$

where $A_{FF}$ corresponds to $F$-rows and $F$-columns of $A$, and $A_{FC}$ corresponds to $F$-rows but $C$-columns. Intuitively, $P_{ideal}$ has "zero" energy at $F$-points because $AP_{ideal}$ is zero at all $F$-rows.

Equation (7) also importantly indicates that the $F/C$ approach yields an energy-minimization process that is better conditioned than for the non-$F/C$ approach. With the $F/C$ approach, (7) implies that the energy-minimization process essentially solves $A_{FF}P_F = -A_{FC}$ column-wise, subject to the constraints. On the other hand, the non-$F/C$ approach induces an energy-minimization process that essentially solves $AP = 0$ column-wise, subject to the constraints. Since $A_{FF}$ is generally better conditioned than $A$, it is clear that the $F/C$ approach yields a better conditioned energy-minimization problem. Moreover, as discussed next in equation (8), acceptable $F/C$ splittings are defined by requiring that the induced $A_{FF}$ is well-conditioned.

Last, we consider two useful theoretical measures from [9] that apply only in the $F/C$ setting. For a more detailed discussion, see [9]. Define $M$ to be the preconditioner for $A$ provided by a simple relaxation method, e.g., the diagonal of $A$ in the case of Jacobi, with $M_{FF}$ representing restriction to $F$-rows and $F$-columns. Then the quantity

$$\rho_{cr} = \left\| I - M_{FF}^{-1}A_{FF} \right\|_{A_{FF}} \tag{8}$$

measures the quality of the coarse grid (i.e., the $F/C$ splitting) by measuring the convergence of $F$-point relaxation. Compatible relaxation (CR) [10, 11] methods are based on estimating $\rho_{cr}$ and other similar quantities. Small $\rho_{cr}$ values, e.g., in the range 0.1–0.6, are desirable and indicate a good $F/C$ splitting and a well-conditioned $A_{FF}$.

Another quantity measures interpolation quality. Let $Q = P\tilde{R}$ be an oblique projector, where $P$ is the multigrid prolongation operator of the form (3) and $\tilde{R}$ is injection at root nodes, so that $\tilde{R}P = I$. Then the quantity $\eta$, which is defined by the smallest such value that satisfies

$$\langle AQe, Qe \rangle \le \eta \langle Ae, e \rangle, \quad \forall e, \tag{9}$$

is bounded below by 1 and measures the quality of $P$. Intuitively, an $\eta \approx 1$ implies that an algebraically smooth fine-grid eigenmode is represented on the coarse grid in proportion to the corresponding fine-grid eigenvalue.

When combined such that $\rho_{cr}$ is small and $\eta$ is order one (e.g., 2 or 3), AMG theory [9] informs us that the resulting two-level method is effective. As a multilevel heuristic, ensuring a good two-level method is often sufficient and, thus, we use $\rho_{cr}$ and $\eta$ to next determine that interpolation, not coarsening, must be improved.

## 3. SUITABLE COARSENING FOR NON-GRID-ALIGNED ANISOTROPIES

In this section, we argue that existing advanced coarsening strategies are sufficient for scalable multigrid performance. It is well-known [7, 8] that semi-coarsening for the case of (nearly) grid-aligned anisotropic diffusion produces scalable multigrid solvers. For the case of non-grid-aligned anisotropic diffusion, previous results are less clear, but we argue here that aggregates that generally track the direction of the anisotropy, but that do not semi-coarsen, are sufficient. Existing advanced coarsening strategies [5, 6] already accomplish such aggregation. Here, the strength measure [6] is used with an appropriate drop-tolerance of $4.0$, unless otherwise noted.

As an example, consider the model problem case of $\theta = 3\pi/16$ with corresponding generic strength stencil shown in Figure 1. Solid lines represent a strong connection. Degree-of-freedom $i$ is
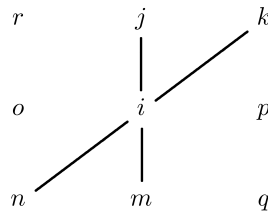


Figure 1. Generic strength-of-connection stencil for $\theta = 3\pi/16$.

strongly connected in two directions, vertically and diagonally from the lower-left to the upper-right. Two typical aggregates for this case are shown in Figure 2 as the shaded regions, which contain the square nodal locations. The two large black squares are the root-nodes of each aggregation, with the smaller squares represnting the "$F$-point" members of each aggregate. Essentially, the "smearing" in the aggregate shapes reflects the "smearing" in the matrix stencil.
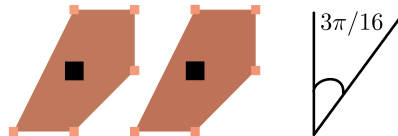


Figure 2. Two typical aggregates for $\theta = 3\pi/16$.

We now use $\rho_{cr}$ to quantify the fact that this aggregation strategy is sufficient. Table I depicts $\rho_{cr}$ values on the finest level for the model problems discretized on a $61 \times 61$ grid, with $\rho_{cr}$ values on coarse levels being similar. We immediately notice that, for the rotated cases, the $\rho_{cr}$ value is actually smaller, thus, indicating a better $F/C$ splitting than for the grid-aligned case, which is known to produce a scalable multigrid solver. Thus, we conclude that existing strength-of-connection measures are sufficient for scalable multigrid, at least when considering bilinear (or similar) discretizations.

| $\theta$ | 0 | $3\pi/16$ | $4\pi/16$ |
|---|---|---|---|
| $\rho_{cr}$ | 0.44 | 0.27 | 0.29 |

Table I. Exact values for $\rho_{cr}$, model rotated anisotropic diffusion.

## 4. SUITABLE INTERPOLATION FOR NON-GRID-ALIGNED ANISOTROPIES

In this section, an improved strategy for constructing $P$ for non-grid-aligned anisotropies is motivated. The other possible means of improving SA, better coarsening or relaxation, are not explored, for the reasons previously discussed. Key to the proposed strategy are long-distance interpolation stencils, which are defined as expanding the sparsity pattern for $P$ with $d > 1$ (typically, $d \in \{2, 3, 4, 5\}$).

As an example, Figure 3 visually depicts for $\theta = 3\pi/16$ how the sparsity pattern expands for a generic aggregate in the center of the domain. The large completely black square denotes the root node of the aggregate, and for the $d = 2$ and 3 cases, the large, but hollow, black squares represent other root nodes excluded from the sparsity pattern. The nonzero pattern is depicted by the shaded area. When $d = 0$, the nonzero pattern represents that of the tentative prolongator, $T$. When $d = 1$, the sparsity pattern is that achieved for standard filtered prolongation smoothing. When $d > 1$, the sparsity pattern uses long-distance connections.
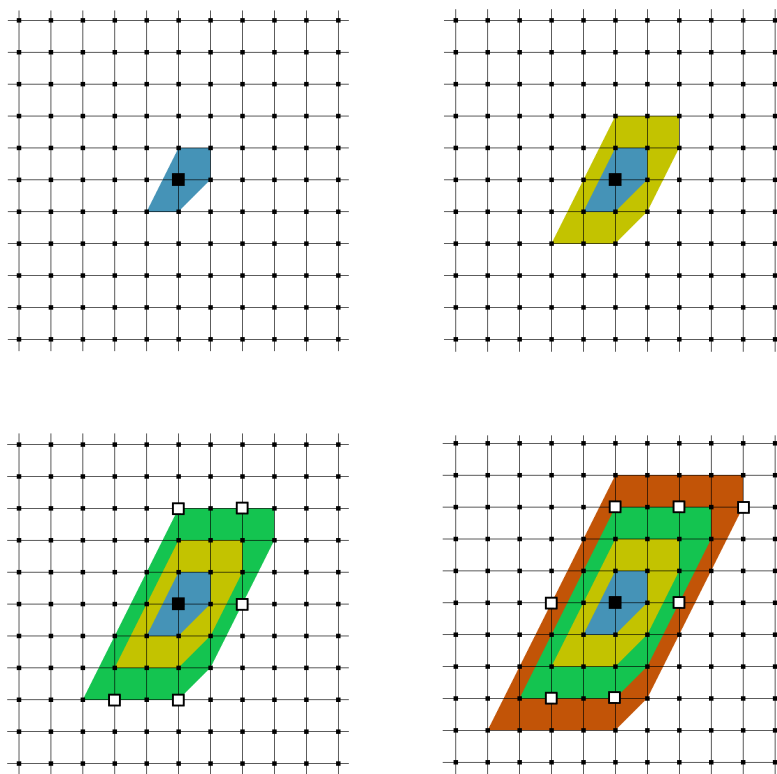


Figure 3. Letting $\theta = 3\pi/16$, sparsity pattern for generic aggregate, from left to right $d = 0, 1, 2, 3$.

The need for long-distance interpolation is now quantified. Table II shows $\eta$ values on the finest level for the model problems discretized on a $61 \times 61$ grid over various values of $d$. As $d$ increases for the non-grid-aligned cases, the $\eta$ values decrease until they reach the goal of being order 1 and then stagnate (remember that the optimal $\eta$ is 1). Once this occurs, a scalable SA solver is achieved, which is denoted by the gray shading. *Scalability depends on the level of approximation of $P_{ideal}$ by P, regardless of $\theta$.* If standard weighted Jacobi is used to smooth $P$, the results are similar to $d = 1$. Since $d = 4$ or 5 is the most robust value and is effective for each $\theta$, it becomes the standard choice in the results section.

The $\eta$ values on coarse levels are similar to Table II, except for $\theta = 3\pi/16$, where $\eta$ stagnates at approximately 16 on the first coarse level. This may be an indicator why the convergence factor changes from 0.12 to 0.15 for this angle in Table V.

| $\theta$ / $d$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| $3\pi/16$ | 130 | 21 | 4.2 | 1.1 | 1.1 |
| $4\pi/16$ | 140 | 92 | 28 | 1.7 | 1.7 |

Table II. Exact values for $\eta$, model rotated anisotropic diffusion.

In conclusion, this strategy is similar to that employed by sparse approximate inverses (SPAI) [23], where the allowed nonzero pattern is expanded in the direction of large coefficients in order to improve the sparse approximation. Here, the sparse approximation is for $P_F \approx -A_{FF}^{-1}A_{FC}$. Additionally, the use of $\eta$ to evaluate the closeness of $P$ to $P_{ideal}$, and, thus, expand the sparsity pattern until an acceptable level of closeness is achieved, is a specific contribution of this paper.

### 4.1. Visual Motivation

To further the motivation, columns of $P$ are visually compared for isotropic diffusion, $\theta = 3\pi/16$, and $\theta = 0$ over various prolongation settings. The column of $P$ that corresponds to the aggregate with root node located in the center of the domain is plotted and normalized such that the largest entry is 1. The small negative entries that exist when $\theta = 3\pi/16$ for $P_{ideal}$ and $d = 4$ have also been dropped for plotting purposes. The four different prolongation settings examined are $P_{ideal}$, the energy-minimizing prolongation smoother for both $d = 1$ and 4, and the energy-minimizing prolongation smoother *without* coarse-grid injection for $d = 4$, i.e., the identity block in (6) is replaced by the sparsity pattern defined by the $C$-rows, $(|S|^d|T|)_C$. The plotted columns of $P_{ideal}$ are dense, although this is difficult to visually discern because the plotted values approach zero as distance from the center point increases. This is especially true for the case of $\theta = 0$.

Figure 4 demonstrates that the columns of $P_{ideal}$ become less local for the rotated anisotropic case, lending support to expanding the sparsity pattern of $P$. Figure 5 demonstrates that $d = 4$ produces a $P$ that visually has the character of $P_{ideal}$. There is a constant-like relationship in strong directions, but not in the weak directions. Figure 6 depicts $d = 1$, where the nonzeros of $P$ are similar to the corresponding locations for $d = 4$, thus, indicating that the expanded sparsity pattern of $d = 4$ is a key reason for its superior performance. Figure 7 depicts $d = 4$ *without* coarse-grid injection. Here, the resulting columns of $P$ no longer resemble the corresponding columns of $P_{ideal}$. Experiments also confirm that not enforcing coarse-grid injection results in deteriorated SA performance.
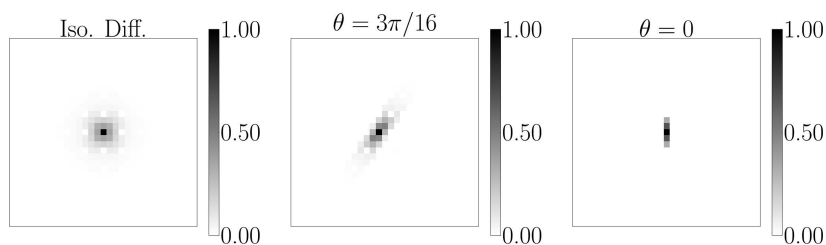


Figure 4. Normalized columns of $P_{ideal}$ for generic aggregate.

### 4.2. Complexity

A primary concern about the proposed approach is the complexity, where attention must be paid to which aggregation or $F/C$ splitting technique is used. For instance, geometric semi-coarsening by three for the angle of $\theta = 3\pi/16$, while using the strength-of-connection scheme utilized in Figure 3, leads to unacceptably high complexities because of the slow coarsening coupled with a relatively dense $P$. However, this type of situation is generally avoided by SA. For SA, the coarsening rate and the coarse matrix stencil size are always connected through the strength-of-connection matrix
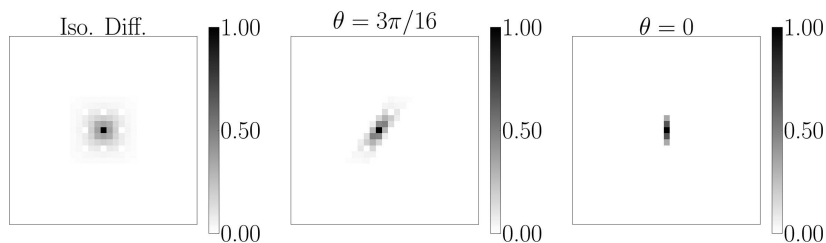
Figure 5. Normalized columns of $P$ for generic aggregate, $d = 4$.
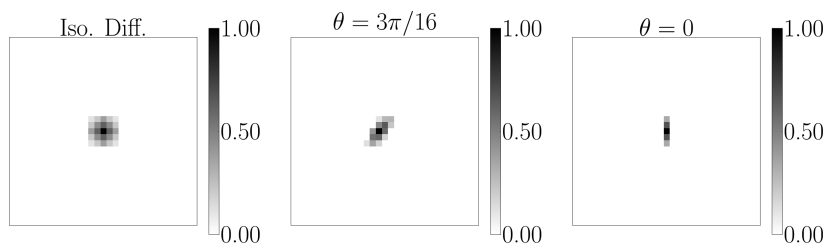


Figure 6. Normalized columns of $P$ for generic aggregate, $d = 1$.
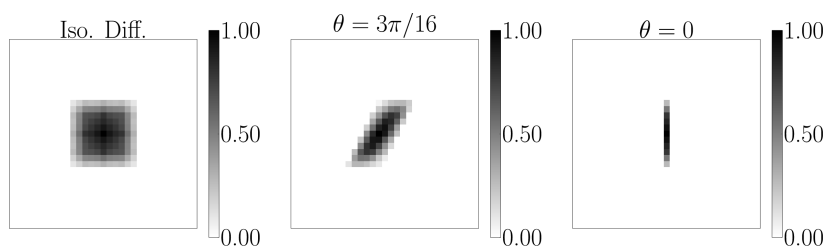


Figure 7. Normalized columns of $P$ for generic aggregate, $d = 4$ *without* Coarse Grid Injection.

(cf. Algorithm 1). The denser coarse matrix generated by the proposed approach typically generates a denser strength-of-connection matrix on that coarse level, which in turn accelerates coarsening (see Table III, where for levels 4 and 5, the coarsening rate has increased above the standard semi-coarsening rate of three for $\theta = 0$).

More generally, the coarse stencil size *increase* is balanced by the coarsening rate. To explain this concept, we examine $\theta = 0$, where semi-coarsening by three occurs for the first three levels, but the stencil of $P$ has been increased with $d = 4$. For instance, on level 0, a typical column of $P$ has 9 nonzeros (for $d = 1$, a typical column has only 5 nonzeros). We are interested in measuring the impact of this denser $P$ on complexity, and use as our metric operator complexity, i.e., the total number of nonzeros for all matrices in the multilevel hierarchy, relative to the number of nonzeros for the finest-level matrix. This quantity both indicates the storage cost and the cost of applying one V-cycle, relative to the fine grid. In particular, we connect operator complexity to the increase in the coarse matrix stencil that is induced by this denser $P$.

Define $\beta_\ell$ to be the average stencil size on level $\ell$ and $n_\ell$ to be the number of degrees-of-freedom on level $\ell$. Then, $n_\ell \beta_\ell$ is the number of nonzeros on level $\ell$ and the operator complexity, $\alpha$, is

$$\alpha = 1 + \frac{1}{\beta_0 n_0} \sum_{\ell=1}^{\ell_{max}} \beta_\ell n_\ell. \tag{10a}$$

For $\theta = 0$, we have semi-coarsening by 3 and a 9-point fine grid stencil, yielding

$$\alpha = 1 + \frac{1}{9n_0} \sum_{\ell=1}^{\ell_{max}} \beta_\ell \frac{n_0}{3^\ell} \tag{10b}$$

$$= 1 + \sum_{\ell=1}^{\ell_{max}} \left( \frac{(\beta_\ell)^{1/\ell}}{3^{2/\ell+1}} \right)^\ell. \tag{10c}$$

If the summation term from (10c), call it $\kappa_\ell$, is bounded by a $\hat{\kappa} < 1$, then by a geometric series argument,

$$\alpha \leq \frac{1}{1 - \hat{\kappa}}. \tag{11}$$

The size of $\kappa_\ell$ indicates if the coarse stencil size is too large at that level, i.e., values consistently near 0.9 or 1.0 indicate unmanageable operator complexities. Additionally, the change in $\kappa_\ell$ from level to level communicates if the coarse stencil size is increasing too fast relative to the coarsening rate, i.e., if $\kappa_\ell$ trends towards 1, then the coarse stencil size is increasing too fast.

In Table III, we compute $\kappa_\ell$ and observe that stencil growth is indeed slow enough to be balanced by the coarsening rate and result in a moderate $\alpha$. In fact, if one uses the table to estimate $\hat{\kappa} \approx 0.6$, then (11) implies $\alpha = 2.5$, which is roughly the operator complexity observed in practice (cf. Table V). A similar analysis can be done for $\theta = 3\pi/16$ and $4\pi/16$, where $\beta_\ell$ is typically larger, e.g., 75, but the coarsening rate is commensurately faster.

For additional clarification, we also present in Table III for each level the number of degrees-of-freedom and the percentage of the total nonzeros. This data also indicates that coarse grid storage costs are increased for this approach, but not prohibitively.

|            | $n_\ell$   | Nonzeros  | $\beta_\ell$ | $\kappa_\ell$ |
|------------|-----------|-----------|--------------|---------------|
| Level $\ell$ 0 | 1 046 529 | 45%       | 9            |               |
| 1          | 348 843   | 35%       | 21           | 0.78          |
| 2          | 116 622   | 14%       | 27           | 0.58          |
| 3          | 38 874    | 5%        | 26           | 0.48          |
| 4          | 4 095     | 1%        | 53           | 0.52          |
| 5          | 365       | $\ll 0\%$ | 34           | 0.43          |

Table III. SA coarsening statistics, $\theta = 0$.

Overall, the proposed approach adds to complexity, but not in an intractable way. The coarse grid stencil growth does not outstrip the coarsening rate, thus, resulting in higher, but not prohibitive, operator complexities for the model problems. Additionally, larger coarse grid stencils eventually induce more aggressive coarsening on coarse levels, as shown in Table III for levels 4 and 5. We remark, however, that larger coarse grid stencils are problematic in some settings, most notably in parallel.

## 5. RESULTS

Numerical evidence is now provided that the proposed strategy yields a robust solver. V(1,1)-cycles with symmetric Gauss-Seidel and a maximum coarsest-grid size of 500 degrees-of-freedom are used as a preconditioning in PCG with a relative residual tolerance of $10^{-8}$. $P$ is computed with $d = 4$ or 5. The near-null-space mode is $B_0 = \mathbf{1}$. All experiments are carried out in the software package PyAMG [24].

First, results for classical SA applied to the model problem (1) are shown in Table IV. The strength-of-connection drop-tolerance has been tuned for each selected $\theta$ and filtered prolongation smoothing has been used, making the results a best case scenario for classical SA. The quantity $\gamma$

is the residual reduction factor, i.e., the factor by which the residual is reduced each PCG step. "Op. Comp." is the operator complexity and "Work" refers to the work per digit-of-accuracy, e.g., 6 work units indicates work equal to 6 finest-level matrix-vector products in order to reduce the residual by one order of magnitude. Scalability only occurs for (nearly) grid-aligned anisotropy (e.g., when $\theta = 0$). Scalability is achieved when work does not increase with the matrix size. The three regular grids examined are $511 \times 511$, $1023 \times 1023$ and $2047 \times 2047$.

| $\theta$ | 0 | | | $3\pi/16$ | | | $4\pi/16$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Grid | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| $\gamma$ | 0.05 | 0.05 | 0.05 | 0.49 | 0.53 | 0.55 | 0.35 | 0.40 | 0.45 |
| Op. Comp. | 1.5 | 1.5 | 1.5 | 1.4 | 1.4 | 1.4 | 1.8 | 1.8 | 1.8 |
| Work | 5 | 5 | 5 | 18 | 20 | 22 | 16 | 18 | 21 |

Table IV. Classical SA results, model rotated anisotropic diffusion.

Table V depicts the results for the model problems using the proposed approach. The operator complexity is reported to reassure the reader, as indicated in Section 4.2, that the large coarse-grid stencils do not create an unusable method. For the problems considered here, the complexity is tolerably high, i.e., 2–2.5. Overall, the results indicate that $\gamma$ and work are (nearly) scalable for all $\theta$.

| $\theta$ | 0 | | | $3\pi/16$ | | | $4\pi/16$ | | |
|---|---|---|---|---|---|---|---|---|---|
| Grid | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| $\gamma$ | 0.05 | 0.05 | 0.05 | 0.12 | 0.14 | 0.15 | 0.20 | 0.20 | 0.20 |
| Op. Comp. | 2.2 | 2.2 | 2.2 | 2.4 | 2.4 | 2.4 | 2.1 | 2.1 | 2.1 |
| Work | 7 | 7 | 7 | 10 | 11 | 11 | 13 | 13 | 13 |

Table V. SA results, model rotated anisotropic diffusion.

Table VI shows results for 2D isotropic diffusion discretized with linear triangular elements on a stretched mesh. The mesh is an unstructured triangulation of the unit circle that, later, has the $y$-coordinates stretched by a factor of 100, thus, yielding non-grid-aligned anisotropic behavior. The stretching of various meshes is depicted in Figure 8. The matrix dimensions for each grid size are $13\,373$, $53\,069$, $217\,143$, and $869\,261$. Table entries are presented as $a\,(b)$, representing quantities when using energy-minimization and standard weighted Jacobi, respectively. Scalability is observed, while standard prolongation smoothing struggles.

| grid # | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\gamma$ | 0.27 (0.55) | 0.30 (0.64) | 0.35 (0.71) | 0.34 (0.78) |
| Work | 16 (22) | 19 (29) | 22 (39) | 21 (53) |
| Op. Comp. | 2.4 (1.4) | 2.5 (1.5) | 2.5 (1.5) | 2.5 (1.4) |

Table VI. SA results, isotropic diffusion on stretched unit circle, entries refer to experiments with the proposed $P$ and weighted Jacobi, respectively.

Table VII depicts the results for 1D radiation transport mapped to spherical coordinates, which yields the 2D problem

$$\mathbf{b} \cdot \nabla u + \sigma u = f, \tag{12}$$

where $\mathbf{b}(x,y) = [xy,\, 1 - y^2]$ and $1/\sigma$ represents the mean free path between collisions. Here, the domain is $[0,1] \times [-1,1]$ and $\sigma = 10^{-6}$, indicating almost no collisions. The discretization used is Q1 finite elements with FOSLS (first-order system least squares) [25]. FOSLS yields a matrix that has the dominant differential term of rotated anisotropic diffusion defined by $-\nabla \cdot [\mathbf{b}^T \mathbf{b}] \nabla$,
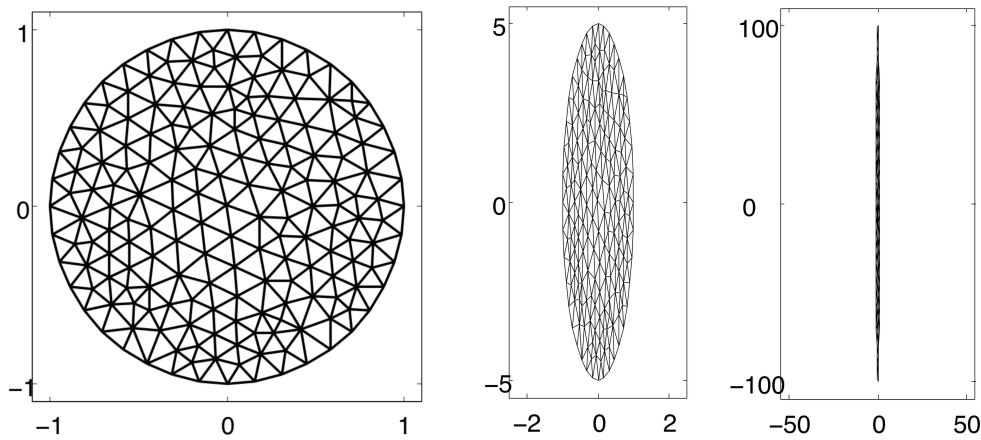
Figure 8. Stretched circular meshes, from left to right, stretch factors are 1, 5 and 100.

in addition to some lower-order terms. Four grid sizes are examined: $256 \times 512$, $512 \times 1024$, $1024 \times 2048$, and $2048 \times 4096$. Given the difficulty of this problem, W-cycles are used to enhance scalability. Standard prolongation smoothing again struggles, while the proposed approach is (nearly) scalable.

| grid # | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\gamma$ | 0.40 (0.64) | 0.42 (0.70) | 0.44 (0.76) | 0.45 (0.81) |
| Work | 71 (45) | 77 (58) | 83 (74) | 84 (93) |
| Op. Comp. | 2.5 (1.4) | 2.4 (1.4) | 2.4 (1.4) | 2.4 (1.4) |

Table VII. SA results, radiation transport, entries refer to experiments
with the proposed $P$ and weighted Jacobi, respectively.

Figure 9 depicts the lines of strong diffusion for the transport problem, which curve through the domain in a U-shaped pattern. Next, Figure 10 shows how aggregation follows the direction of this anisotropy. The aggregation is depicted over only the lower half of the domain, but the upper half is a mirror image. The aggregates are depicted as geometric shapes (lines or polygons), while the nodal locations are small squares. The blow-up of the aggregate shows a 3D plot of that aggregate's corresponding column in $P$, which curves according to the direction of the anisotropy.

While the proposed solver enjoys good performance for linear discretizations, discretizing the transport problem with biquadratics exposes some limitations. For instance, Table VIII depicts such results. The proposed solver outperforms classical prolongation smoothing, but does not exhibit scalable performance.

| grid # | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\gamma$ | 0.70 (0.84) | 0.75 (0.88) | 0.79 (0.91) | 0.83 (0.94) |
| Work | 111 (110) | 146 (153) | 178 (208) | 220 (269) |
| Op. Comp. | 2.0 (1.3) | 1.9 (1.3) | 1.9 (1.3) | 1.9 (1.3) |

Table VIII. SA results, radiation transport, biquadratic discretization, entries refer
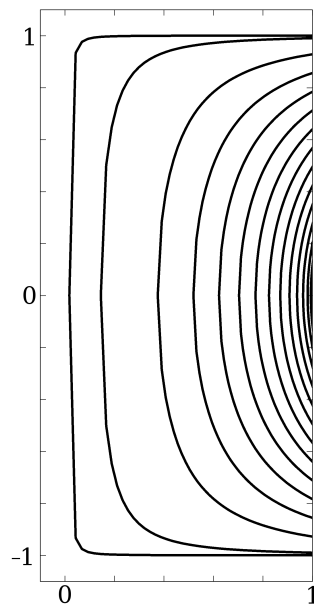to experiments with the proposed $P$ and weighted Jacobi, respectively.

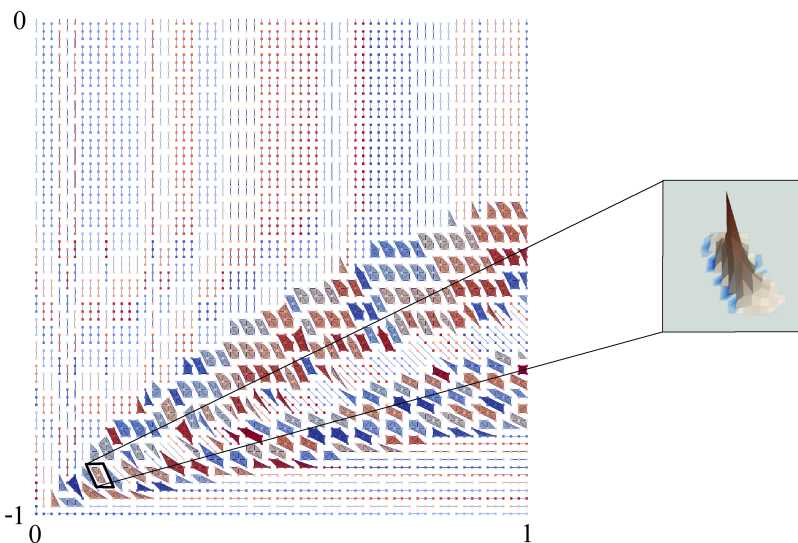Figure 9. Lines of strong diffusion for transport problem.



Figure 10. Aggregates for transport problem, with blow-up of selected column of $P$.

## 6. FUTURE WORK

### 6.1. A Computable $\eta$

A drawback of this approach is that exact computation of $\eta$ requires dense eigenvalue decompositions. For anisotropic diffusion, this is surmountable, because we examine small, simple model problems that exhibit the desired behavior. These model problems, then, give the insight needed to construct a robust solver for real world applications. However, this is not always the case, which leads us to develop a computable approximation $\hat{\eta}$.

Let $Q$ be the oblique projector from before. Then, the definition of $\eta$ from equation (9) is equivalent to the generalized eigenvalue problem

$$\eta = \lambda_{max} \quad \text{for} \quad Q^T A Q e = \lambda A e \tag{13a}$$

$$A^{-1} Q^T A Q e = \lambda e. \tag{13b}$$

By substituting in the approximation to $A^{-1}$ given by one or two multigrid V-cycles (call it $\hat{A}^{-1}$), we arrive at the approximation, $\hat{\eta}$, given by

$$\hat{\eta} = \lambda_{max} \quad \text{for} \quad \hat{A}^{-1} Q^T A Q e = \lambda e. \tag{13c}$$

Importantly, $\hat{\eta}$ is the maximum eigenvalue of an operator whose action is efficiently represented as a sequence of matrix-vector products, followed by application of one or two V-cycles. Thus, an iterative procedure, such as Arnoldi, is used to approximate $\hat{\eta}$.

Experimentally, $\hat{\eta}$ agrees with $\eta$ for 2–3 decimal places for the model problems. Additionally, $\hat{\eta}$ behaves as we expect for the transport problem, i.e., it is reduced as $d$ is increased, until finally stagnating at a value that is order one. Table IX presents $\hat{\eta}$ values for the first three levels in the hierarchy for the $256 \times 512$ grid. Interestingly, the data indicates that reducing $\hat{\eta}$ on all levels is

| $d$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Level 0 | 3.4 | 2.4 | 2.2 | 2.2 | 2.1 |
| 1 | 15 | 5.9 | 1.8 | 2.0 | 1.7 |
| 2 | 25 | 2.6 | 2.6 | 1.7 | 1.4 |

Table IX. Values for $\hat{\eta}$ for the transport problem.

important. It is not until $\hat{\eta}$ is small on the first three levels that SA performance shows scalability. Computationally, $\hat{\eta}$ is a feasible diagnostic tool, but it is still an expensive one. For the results in Table IX, 50–100 Arnoldi iterations were required.

In addition to the experimental evidence, there is also a heuristic argument as to why using a V-cycle approximation to $A^{-1}$ is reasonable. Equation (13c) is equivalent to the eigenvalue problem for the transpose,

$$\hat{\eta} = \lambda_{max} \quad \text{for} \quad Q^T A Q \hat{A}^{-1} e = \lambda e. \tag{14a}$$

Then, using the fact that the projector $QQ = Q$, we arrive at

$$(Q^T A Q)(Q \hat{A}^{-1}) e = \lambda e. \tag{14b}$$

The action of $\hat{A}^{-1}$ is being projected onto $\text{span}(Q)$, meaning that $\hat{A}^{-1}$ must be accurate for those modes lying in $\text{span}(P)$. This is, of course, precisely what a 2-level multigrid cycle does by providing an energy optimal coarse-grid correction in $\text{span}(P)$. Yet, since the V-cycle used here is multilevel, the coarse-grid correction is not generally optimal and we expect some degradation in approximating $\eta$. However, experimental evidence suggests that this degradation is small for the problems considered here.

Use of $\hat{\eta}$, in conjunction with existing computable approximations to $\rho_{cr}$ available through CR, allow the solver design strategy of this paper to be applied to general problems. The quality of interpolation is gauged with $\hat{\eta}$, while the quality of the $F/C$ splitting is gauged with $\rho_{cr}$ approximations obtained with CR. While the cost of CR is problem dependent, it is important to note that this cost can be high, especially if the $F/C$ splitting produces a poorly conditioned $A_{FF}$ block. The focus of future work is on further reducing the computational cost of $\hat{\eta}$ and also on verifying the utility of $\hat{\eta}$ for a broader range of test problems.

### 6.2. Allowing for Arbitrarily Many Vectors in $B$

A potential drawback of this approach is that, when using coarse-grid injection, the number of functions fitted over each aggregate during the construction of $T$ (see (2)) must equal the supernode size [3]. For 3D elasticity and diffusion, supernodes are typically of size three and one, respectively, thus, allowing for three and one functions over each aggregate, respectively.

This restriction comes about in the following way. The QR procedure described in (2) effectively sets the number of degrees-of-freedom per node on the coarse level to equal the number of near-null-space modes. This means, however, that there is not necessarily a one-to-one correspondence between coarse- and fine-grid degrees-of-freedom. Thus, when the number of near-null-space modes is not equal to the number of degrees-of-freedom per supernode on the fine grid, then it is not possible to simply normalize columns so that a root node entry has a value of one. In other words, some columns do not correspond to a unique root node entry.

Classical SA has no such restriction on the number of near-null-space modes. For instance, with 3D elasticity, the six rigid-body-modes are typically fit over each aggregate, i.e., three displacements and three rotations. However, the proposed approach automatically compensates for this potential drawback. To show this, an elasticity experiment is run with $B_0$ equal to the three displacements. Then, $e$ is computed, which measures how well span($P$) approximates the rotational rigid-body-modes. Let

$$e = \sum_{i=1,2,3} \frac{\|(I - P(P^T P)^{-1} P^T)\hat{B}_{(i)}\|}{\|\hat{B}_{(i)}\|_A}, \tag{15}$$

where $\hat{B}_{(i)}$ is the $i^{th}$ rotation and $P$ is taken from the finest level. Thus, $e$ represents how well span($P$) approximates $\hat{B}$ in terms of the weak approximation property.

The test problem is isotropic linearized elasticity defined by

$$- \operatorname{div}\left(\lambda \operatorname{tr}\left(\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right)/2\right) I + \mu\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right)\right) = f, \tag{16}$$

where $\lambda$ and $\mu$ are the Lamé parameters, $I$ is the identity matrix and tr() is the trace function. The GetFem++ package [26] is used to discretize a tripod with a downward force applied to it, using tetrahedra and linear basis functions. An example coarse mesh is depicted in Figure 11. $P$ has dimensions $16\,341 \times 1\,347$ and the experiment is run with various $d$ and, for comparison, weighted Jacobi. Table X depicts these results.
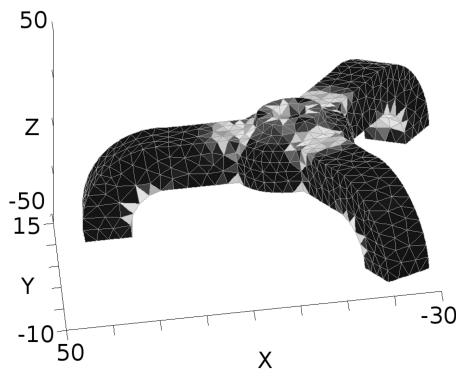


Figure 11. Example coarse tripod mesh.

As $d$ increases, both $\gamma$ and $e$ improve, but at the cost of a higher operator complexity. For comparison, use of classical SA and all six rigid-body-modes in $B_0$ yields a solver with $\gamma = 0.15$,

---

[3] A common SA approach when multiple variables exist at each spatial location for the original matrix problem is supernodes [4]. Supernodes cause aggregation to happen in two stages at the finest level. First, supernodes are formed by aggregating degrees-of-freedom at the same spatial location together. Second, strength-of-connection information is used to aggregate the supernodes.

| Prol. Smoother: | w-Jacobi | $d=1$ | $d=2$ | $d=3$ | $d=4$ | $d=5$ |
|---|---|---|---|---|---|---|
| $e$ | 1.1e-1 | 6.4e-2 | 4.0e-2 | 2.7e-2 | 2.0e-2 | 1.7e-2 |
| $\gamma$ | 0.62 | 0.45 | 0.40 | 0.30 | 0.17 | 0.16 |
| Work | 23 | 13 | 14 | 12 | 9 | 10 |
| Op. Comp. | 1.2 | 1.2 | 1.4 | 1.6 | 1.8 | 2.0 |

Table X. SA results and error in span($P$) for rotations.

an operator complexity of 1.8, and a work per digit-of-accuracy of 8. Additionally, there is numerical evidence that $d \in \{3, 4, 5\}$ yields scalable results for this elasticity problem. However, more experiments are needed to confirm this result for more general problems.

Moreover, an extension of the energy-minimizing prolongation smoother is planned so that the rotations are exactly incorporated into span($P$) through a modification of the constraints. So long as there are enough nonzeros in each row of $P$, this goal is achievable.

## 7. CONCLUSIONS

A novel interpolation strategy of coarse-grid injection coupled with energy-minimization and long-distance interpolation stencils is developed. For the non-grid-aligned anisotropic diffusion problems examined with a linear discretization, the resulting solvers appear to be (nearly) scalable and are the first such *algebraic* solvers known to the author. Moreover, this work indicates that neither new coarsening nor new relaxation strategies are necessary.

Established theory for $F/C$-style AMG provides the tools ($\eta$ and $\rho_{cr}$) to design the proposed solver, by guiding the sparsity pattern choice for $P$ and assuring us that available methods for obtaining an $F/C$ splitting are sufficient. Additionally, the proposed computable approximation, $\hat{\eta}$, appears to have practical uses and to predict scalability.

## REFERENCES

1. Brandt A, McCormick SF, Ruge JW. Algebraic multigrid (AMG) for sparse matrix equations. *Sparsity and Its Applications*, Evans DJ (ed.). Cambridge Univ. Press: Cambridge, 1984; 257–284.
2. Ruge JW, Stüben K. Algebraic multigrid (AMG). *Multigrid Methods*, McCormick SF (ed.). Frontiers Applied Mathematics, SIAM: Philadelphia, 1987; 73–130.
3. Vaněk P, Mandel J, Brezina M. Algebraic multigrid based on smoothed aggregation for second and fourth order problems. *Computing* 1996; **56**:179–196.
4. Vaněk P, Brezina M, Mandel J. Convergence of algebraic multigrid based on smoothed aggregation. *Numerische Mathematik* 2001; **88**(3):559–579.
5. Brannick J, Brezina M, MacLachlan S, Manteuffel T, McCormick S, Ruge J. An energy-based AMG coarsening strategy. *Numerical Linear Algebra with Applications* 2006; **13**(2-3):133–148.
6. Olson LN, Schroder JB, Tuminaro RS. A new perspective on strength measures in algebraic multigrid. *Numerical Linear Algebra with Applications* 2010; **17**(4):713–733.
7. Trottenberg U, Oosterlee C, Schüller A. *Multigrid*. Academic Press: London, UK, 2001.
8. Briggs WL, Henson VE, McCormick SF. *A multigrid tutorial*. 2nd edn., SIAM: Philadelphia, PA, USA, 2000.
9. Falgout RD, Vassilevski PS. On generalizing the algebraic multigrid framework. *SIAM Journal on Numerical Analysis* 2004; **42**(4):1669–1693.
10. Brandt A. General highly accurate algebraic coarsening. *Electronic Transactions on Numerical Analysis* 2000; **10**:1–20.
11. Livne OE. Coarsening by compatible relaxation. *Numerical Linear Algebra with Applications* 2004; **11**(2-3):205–227.
12. Brannick J, Zikatanov L. Algebraic multigrid methods based on compatible relaxation and energy minimization. *Domain Decomposition Methods in Science and Engineering XVI*, *Lecture Notes in Computational Science and Engineering*, vol. 55, Widlund O, Keyes DE (eds.). Springer: Berlin, 2007; 15–26.
13. Olson LN, Schroder JB, Tuminaro RS. A general interpolation strategy for algebraic multigrid using energy-minimization. *SIAM Journal on Scientific Computing* 2011; **33**:966–991.
14. Yavneh I. Coarse-grid correction for nonelliptic and singular perturbation problems. *SIAM Journal on Scientific Computing* 1998; **19**:1682–1699.
15. Bramble JH, Pasciak JE, Wang J, Xu J. Convergence estimates for multigrid algorithms without regularity assumptions. *Mathematics of Computation* 1991; **57**(195):23–45.

16. Diskin B. Multigrid algorithm with conditional coarsening for non-aligned sonic flow. *Electronic Transactions on Numerical Analysis* December 1997; **6**:106–119.
17. Jones JE, Vassilevski PS. AMGe based on element agglomeration. *SIAM Journal on Scientific Computing* 2001; **23**(1):109–133.
18. Morano E, Mavriplis DJ, Venkatakrishnan V. Coarsening strategies for unstructured multigrid techniques with application to anisotropic problems. *Seventh Copper Mountain Conference on Multigrid Methods*, Melson N, Manteuffel T, McCormick S, Douglas C (eds.), NASA Conference Publication 3339, 1996; 591–606.
19. Mavriplis DJ. Multigrid strategies for viscous flow solvers on anisotropic unstructured meshes. *Journal of Computational Physics* 1998; **145**(1):141–165.
20. Oosterlee CW. A GMRES-based plane smoother in multigrid to solve 3D anisotropic fluid flow problems. *Journal of Computational Physics* 1997; **130**(1):41–53.
21. Brezina M, Vaněk P. A black-box iterative solver based on a two-level Schwarz method. *Computing* 1999; **63**:233–263.
22. Olson LN, Schroder JB. Smoothed aggregation for Helmholtz problems. *Numerical Linear Algebra with Applications* 2010; **17**(2-3):361–386.
23. Grote M, Huckle T. Sparse approximate inverses. *SIAM Journal on Scientific Computing* 1997; **18**:838–853.
24. Bell WN, Olson LN, Schroder J. PyAMG: Algebraic multigrid solvers in Python v2.0. http://www.pyamg.org 2011. Release 1.0.
25. Sterck HD, Manteuffel TA, McCormick SF, Olson L. Least-squares finite element methods and algebraic multigrid solvers for linear hyperbolic PDEs. *SIAM Journal on Scientific Computing* 2004; **26**(1):31–54.
26. Renard Y. The GetFem++ Project. http://download.gna.org/getfem/doc/getfem_project.pdf.