

Algebraic multigrid for systems of elliptic boundary-value problems

Barry Lee 

Department of Mathematics, Southern Methodist University, Dallas, Texas, USA

Correspondence

Barry Lee, Department of Mathematics, Southern Methodist University, Dallas, TX.

Email: barryl@smu.edu

Summary

This article develops an algebraic multigrid (AMG) method for solving systems of elliptic boundary-value problems. It is well known that multigrid for systems of elliptic equations faces many challenges that do not arise for most scalar equations. These challenges include strong intervariable couplings, multidimensional and possibly large near-nullspaces, analytically unknown near-nullspaces, delicate selection of coarse degrees of freedom (CDOFs), and complex construction of intergrid operators. In this article, we consider only the selection of CDOFs and the construction of the interpolation operator. The selection is an extension of the Ruge–Stuben algorithm using a new strength of connection measure taken between nodal degrees of freedom, that is, between all degrees of freedom located at a gridpoint to all degrees of freedom at another gridpoint. This measure is based on a local correlation matrix generated for a set of smoothed test vectors derived from a relaxation-based procedure. With this measure, selection of the CDOFs is then determined by the number of strongly correlated connections at each node, with the selection processed by a Ruge–Stuben coloring scheme. Having selected the CDOFs, the interpolation operator is constructed using a bootstrap AMG (BAMG) procedure. We apply the BAMG procedure either over the smoothed test vectors to obtain an intervariable interpolation scheme or over the like-variable components of the smoothed test vectors to obtain an intravariation interpolation scheme. Moreover, comparing the correlation measured between the intravariation couplings with the correlation between all couplings, a mixed intravariation and intervariable interpolation scheme is developed. We further examine an indirect BAMG method that explicitly uses the coefficients of the system operator in constructing the interpolation weights. Finally, based on a weak approximation criterion, we consider a simple scheme to adapt the order of the interpolation (i.e., adapt the caliber or maximum number of coarse-grid points that a fine-grid point can interpolate from) over the computational domain.

KEYWORDS

bootstrap and adaptive multigrid, correlation affinity measure, multigrid, Z Fisher transformation

MOS SUBJECT CLASSIFICATION

65M55; 65M70; 65N55; 65R20; 65Z05

1 | INTRODUCTION

We are interested in developing a multigrid method for the system of elliptic partial differential equations (PDEs)

$$\mathcal{L}u = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \dots & \mathcal{L}_{1n} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \dots & \mathcal{L}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{L}_{n1} & \dots & \dots & \mathcal{L}_{nn} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \quad (1)$$

defined on a smooth spatial domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, and where each \mathcal{L}_{ij} is a partial differential operator of order at most 2 (higher-order operators can be reduced to second- or lower-order operators by introducing auxiliary variables). Together with (1) are boundary conditions that give us a system boundary-value problem (BVP), which we will take to be well posed.

Developing a scalable and robust multigrid method for these systems is nontrivial. Issues include strong intervariable coupling, multidimensional near-nullspace, inadequate revealing of the near-nullspace components through relaxation, delicate construction of the interpolation operators even when computed near-nullspace components are used (e.g., in an adaptive multigrid framework¹), and others. Furthermore, additional concerns arise for algebraic multigrid (AMG), such as the selection of coarse degrees of freedom (DOFs), which may lead to crossvariable, or intervariable, interpolation where variables of different physical quantities interpolate to each other (see References²⁻¹⁰ for research on AMG for systems of PDEs, particularly for the equations of elasticity). This selection will involve a strength-of-connection or affinity measure to determine which DOFs should be retained on the coarse level in order to achieve good coarse-grid approximation using as few DOFs as possible. For scalar PDEs with discretizations leading to M-matrices, the concept of strong connections can be used. For systems of PDEs, the M-matrix property is generally lost and other measures must be developed (e.g., determinants of nodal submatrices, an option in the HYPRE software package,¹¹ and for other measures, see Reference⁷). In this article, we examine a correlation-based measure first introduced in Reference¹² for unstructured graph Laplacians, but in this article extended to systems of PDEs. One advantage of a correlation-based measure is that it exposes dependencies (i.e., sensitivities) not just between individual DOFs but also between subsets of DOFs. These subsets may include all the DOFs within a pathlength distance from a given node, or particularly for systems of PDEs, the DOFs of all variable types located at a collection of gridpoints. In this article, we consider the correlation between all the variable types located at each pair of neighboring nodes. Assuming that the fine-grid matrix is *nodally ordered* in the sense that lexicographically streaming over the gridpoints, all the variable DOFs at a gridpoint are ordered continuously before the DOFs at the next gridpoint, the pairs of neighboring nodes can be easily determined from the adjacency matrix of the system. Hence, minimal grid information will be required in this scheme.

Once the CDOFs are selected, the interpolation operator will be constructed using a bootstrap algebraic multigrid (BAMG) procedure that involves solving local least-squares problems to determine the interpolation weights.^{13,14} Since the coarsening is performed nodally, there will be flexibility in constructing intravariation or intervariable interpolation, or even a combination of the two that depends on how strongly the variables locally correlate.

The produced methods are nodal-based or point-based.^{7,9,15,16} Thus, they can handle strongly coupled systems of PDEs better than the unknown-based approach, which coarsens each variable separately leading to separate hierarchies for the variables.^{9,15} The major difference between the methods of this article and other nodal-based approaches is in the selection of the CDOFs and the construction of the interpolation operator. Unlike other nodal-based approaches, an adaptive process is used to achieve this selection and operator construction. This process can be applied to more general systems of elliptic PDEs, but this comes at a large computational cost. Thus, the schemes of this article are more expensive than other nodal-based methods. Nevertheless, the main goals of this article are to demonstrate that the statistical correlations of relaxed vectors can expose dependencies in the DOFs and thus they can be used to determine the CDOFs and to explore and extend some of the state-of-the-art AMG techniques for scalar PDEs^{13,14} to systems of PDEs.

The article proceeds as follow. In Section 2, we review multigrid for scalar PDEs and describe issues that complicate the development of multigrid for systems of PDEs. We consider some approximation measures that can guide the construction of the interpolation operator and some relaxation-based and BAMG techniques for constructing the interpolation operators. Although these measures and relaxation-based/BAMG techniques have led to the development of new schemes for scalar PDEs, their use in developing schemes for systems of PDEs have not been as systematically explored. In Section 3, we examine the original correlation-based measure for graph Laplacians and its extensions to systems of PDEs. Unlike the original method, which constructs a projected volume of the coarse agglomerate for each DOF to determine

the candidate CDOFs, we use correlation matrix measures to determine which *nodal* DOFs strongly influences the most number of *nodal* DOFs to determine the coarse *nodal* DOFs. This nodal coarsening, which retains all the variable types located at a node, allows the nodal structure to be preserved on all levels, and thus preserves some of the PDE structure. To determine whether a node is strongly or weakly connected to a given node, rather than using the arithmetic average of the correlations at all the neighboring nodes, an average will be obtained by first applying a Z Fisher transformation to the correlations.^{15,17,18} This transformation provides a variance stabilization in the sense that the set of transformed correlations converges to a normal distribution faster than the original sample set of correlations,¹⁵ that is, fewer samples are needed. Having determined the coarse DOFs, in Section 4, we construct the values of the interpolation operator using a BAMG least-squares approach. We consider intervariable and intravariation interpolation by appropriately forming the local least-squares problem and a mixed intravariation/intervariable interpolation. Moreover, guided by a weak approximation measure, we examine an adaptive interpolation caliber scheme that increases the caliber where the weak approximation bound is likely to be locally large. Finally, we examine an indirect BAMG (iBAMG) scheme that explicitly uses the coefficients of the system operator.¹⁴ In Section 5, we give some numerical results to illustrate the performance of the new multigrid procedures.

2 | MULTIGRID

One of the most efficient methods for numerically solving scalar elliptic PDEs is multigrid. Multigrid achieves its efficiency by using a hierarchy of grids, where the computation on the coarser grids costs only a fraction of the effort of computing on the original grid. By carefully designing the grid-level computation to handle only solution/error components on the scale of the level, the goal of the grid-level computation is to resolve only grid-scale features. The solution/error components are thus handled levelwise.

To achieve this efficiency and to produce a scalable multigrid method, the complementary smoothing/coarse-grid correction principle should hold, that is, what cannot be eliminated by relaxation must be eliminated through the coarse-grid correction. The purpose of smoothing (e.g., a few sweeps of Gauss–Seidel or weighted Jacobi iteration) is to “smooth” out the error in an approximate solution. In the geometric multigrid setting, for structured-grid discretizations of scalar elliptic PDEs, the smooth errors often correspond to geometrically smooth errors; in the AMG setting, these errors correspond to the algebraic near-nullspace of the matrix operator (i.e., eigenvectors corresponding to the smaller eigenvalues). In fact, for some of the basic relaxation schemes, these problematic smooth errors correspond to the near-nullspace components of the differential operator.

Turning to the complementary coarse-grid correction, this process refers to the updating of the fine-grid approximation with the solution of the coarse-grid problem. The ideal situation is to have the coarse-grid correction resolves the smooth error components. Such coarse-grid correction can be accomplished by carefully designing the intergrid interpolation operators to have the property that these errors are in their range. Assuming that these errors are the near-nullspace components of the differential operator, the interpolation operators are formed to capture error components that give small energy norm. For scalar elliptic equations, this has been well analyzed, for example, if \mathcal{L} is a scalar self-adjoint operator, then the interpolation operator P is designed so that discrete approximations to components u that satisfy

$$\frac{|(\mathcal{L}u, u)|}{\|u\|^2} \ll 1$$

are well approximated in its range. For scalar elliptic operators, one near-nullspace component is the constant $u = 1$. Let L_h denote the discretized operator, and assume a structured-grid, geometric multigrid setting. Then using the constant vector, an operator-dependent P can be constructed by locally solving

$$L_h P 1 = 0 \quad (2)$$

at the fine gridpoints. This may involve operator collapsing.¹⁹ However, in the AMG setting, even before the interpolation operator can be constructed, the CDOFs must be determined. The CDOFs must be chosen to permit the formation of accurate coarse-grid problems. Or, viewed differently, weighted sums of the error at selected CDOFs must be able to accurately represent the error at fine DOFs that are not sufficiently handled by the smoother. To select these CDOFs, a measure based on the operator is needed. For scalar elliptic PDEs with discretizations leading to M-matrices, this measure

is given by the strength of connections between the DOFs, which is determined from the coefficients of L_h . For example, in the classical Ruge–Stuben coarsening approach,⁹ the following strength of connection measure is used: DOF j is said to strongly influence DOF i if

$$|L_{h,ij}| \geq \theta \max_{k \neq i} |L_{h,ik}| \quad \text{for } 0 < \theta \leq 1. \quad (3)$$

The coarse nodes are essentially selected to be the DOFs that strongly influence the most number of fine DOFs (i.e., DOFs that have not been selected to be coarse nodes) and form a maximal independent set. With this selection and taking the constant function a priori as the near-nullspace, the coefficients of P can be constructed to ensure that this near-nullspace of L_h can be accurately represented.^{9,10}

When L_h is not an M-matrix, theoretical measures that are expressed in terms of the intergrid and relaxation operators may be used to guide the selection of CDOFs and the construction of the intergrid operator. For example, two measures^{3,20} are

$$M_1(Q, u_h) = \frac{\|(I - Q)u_h\|_{D_h}}{\|u_h\|_{L_h}} \quad M_2(Q, u_h) = \frac{(X(I - Q)u_h, (I - Q)u_h)}{\|u_h\|_{L_h}^2}, \quad (4)$$

where D_h is the diagonal of L_h , $Q = P\hat{R}$ for an operator \hat{R} that defines the coarse DOFs and with $\hat{R}P = I$, the coarse identity, and X is a symmetric and positive-definite operator that defines the relaxation. Because Q is a projection onto the range of P , M_1 quantifies how well a given interpolation operator P for a set of given coarse DOFs defined through \hat{R} can approximate u_h . Measure M_2 further includes the effect of relaxation in the approximation. If u_h is a good approximation to the near-nullspace component, then these measures can be immediately used to guide the construction of P .

Unfortunately, the near-nullspace components often are not analytically available even for general scalar PDEs. This implies that the discretized forms of them are not immediately available to select the CDOFs and form the interpolation operator. In this case, rough approximations to the near-nullspace can be generated by applying a number of sweeps of an appropriate relaxation scheme on a set of random vectors. From the complementary principle, algebraically smooth error components that are poorly handled by the smoother must first be *representable* on the coarser level. Thus, what remains after smoothing can themselves expose candidates for the CDOFs, especially small sets of them that can accurately represent the poorly handled error. This is the basis of the BAMG, adaptive multigrid schemes, and other relaxation-based schemes.^{1,13,21,22} For example, consider the BAMG approach. Let $\{v^{(\alpha)}\}_{\alpha=1}^s$ denote the set of relaxed vectors referred to as the test vectors, and let Ω^0 be the set of DOFs to be partitioned into F (fine) and C (coarse) DOFs with $\Omega^0 = C \cup F$ and $C \cap F = \emptyset$. Set C can be determined from $\{v^{(\alpha)}\}_{\alpha=1}^s$ through a strength-of-connection measure such as the correlation between the DOFs. Assuming that this partitioning has been performed, for $i \in F$ with C_i denoting a subset of CDOFs that is used to interpolate to i , the interpolation weights for DOF i are determined by solving the local least-squares problem

$$\{w_{ij} | j \in C_i\} = \arg \min_{w_{ij}} \sum_{\alpha=1}^s \left(v_i^{(\alpha)} - \sum_{j \in C_i} w_{ij} v_j^{(\alpha)} \right)^2. \quad (5)$$

Weights $\{w_{ij}\}_{j \in C_i}$ are the coefficients of the i th row of P . This is the most basic least-squares formulation for determining the interpolation weights. To further conform to measure $M_1(Q, u_h)$, a weighted least-squares formulation should be used, with the weights being the reciprocal of the L_h norm of the test vectors.^{3,13} This weighted formulation, which is used in the algorithms of this article (formulations (21)–(25)), provides heavier weighting to the algebraically smooth test vectors.

Turning to system (1), issues arise that do not occur for scalar PDEs. Even in the simpler case when (1) is self-adjoint, the relaxation process is complicated by the intervariable couplings in the system. In the scalar case, the relaxation locally updates the DOFs using the spatial dependency described through the stencil patterns and coefficients of the discretization. For systems of PDEs, the relaxation now has to handle both the spatial and intervariable dependencies. Achieving this can require a special ordering of the variable updates reflected through the strengths of the intervariable couplings.²³ However, even with this ordering, it is unclear how many relaxation sweeps are needed to sufficiently expose the unknown near-nullspace components.

Another issue in developing multigrid for (1) is the multidimensionality of the near-nullspace, whose analytic form is generally unknown. These components can have a spatial global structure and/or a spatial local structure where the

ellipticity degrades. Moreover, assuming that with a sufficient number of relaxation sweeps some of these components are exposed, choices arise on how to choose the CDOFs and how to form the interpolation operator. In the ideal case, when system (1) is diagonally dominant, or H^1 equivalent (i.e., there exists positive constants c_1, c_2 such that $c_1 \|v\|_1^2 \leq (\mathcal{L}v, v) \leq c_2 \|v\|_1^2$ for all $v \in H^1$), this dramatically simplifies. Now the near-nullspace can be approximated with the padded vectors

$$u_i = (0, \dots, 0, u_i, 0, \dots, 0)^t,$$

where u_i is the near-nullspace component of \mathcal{L}_{ii} . The CDOFs can be selected separately for each of the variables, and an appropriate interpolation operator is

$$P = \text{diag}(P_{11}, P_{22}, \dots, P_{nn}), \quad (6)$$

where P_{ii} is the interpolation operator based on scalar operator \mathcal{L}_{ii} and constructed using operator-collapsing or AMG techniques to ensure that the padded near-nullspace components are in the range of P .^{5,7,9,10,24,25} This block-diagonal interpolation leads to variables interpolating to like variables, that is, intravariation interpolation. With this interpolation, the coarse-grid operator can be formed using the Galerkin coarse-grid procedure

$$\mathcal{L}^{l+1} = P^t \mathcal{L}^l P, \quad (7)$$

where l and $(l + 1)$ are, respectively, the fine and coarse levels.

When system (1) is not diagonally dominant, the near-nullspace will not have the above zero-padded form, and the CDOFs must be carefully chosen using an appropriate strength-of-connection measure. Nevertheless, block-diagonal interpolation (6) can still be applied to form the coarse-grid operator. Alternatively, the full-matrix interpolation

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}, \quad (8)$$

or its nodally ordered form can be used (see Reference⁷ for a nodally ordered form of this). This will produce intervariable interpolation since the off-diagonal blocks lead to crosscoupling in the variables. A Galerkin coarsening with this operator produces the coarse-grid operator. This choice of interpolation may appear to be counterintuitive since the components of the coarse-grid operator will correspond to PDE operators that are very different from the fine-grid component operators, for example, component $\mathcal{L}_{ij}^{l+1} = \sum_{s=1}^n \sum_{t=1}^n P_{st}^t \mathcal{L}_{st}^l P_{ij}$, which combines all the component operators of the system, is generally characteristically different from \mathcal{L}_{ij}^l . However, viewed algebraically, the coarse-grid operator is an approximate projection of the fine-grid operator onto the space spanned by the near-nullspace vectors. Thus, the coarse-grid problem may be viewed as a transformed system. However, this viewpoint does not make the multigrid construction any easier.

3 | CORRELATION-BASED MEASURES

To select the CDOFs for systems of PDEs, a general computable measure that exposes the strength-of-connection/affinity in the DOFs is needed. One quantity that reveals this strength/affinity for *general* systems is the correlation between the DOFs. In particular, suppose $\{v^{(\alpha)}\}_{\alpha=1}^s$ is a collection of smoothed test vectors obtained by applying a number of relaxation sweeps to an initial set of random vectors. We can assume that the smoothed vectors are themselves random, although colored by the relaxation, so that they form a *sample set* for the random test vectors. Given the statistical nature of this sample set, the correlation is the natural quantity that can reveal the affinity between the DOFs.

One correlation-based measure for selecting the CDOFs is explored in Reference¹² for graph Laplacian systems (see also the recent article,⁸ which uses a correlation measure for selecting the CDOFs for PDEs). With i, j being two arbitrary vector components of vector $v^{(\alpha)}$ and the component inner product defined as

$$(v_i, v_j) = \sum_{\alpha=1}^s v_i^{(\alpha)} v_j^{(\alpha)},$$

the measure is

$$c_{ij} = \frac{|(v_i, v_j)|^2}{(v_i, v_i)(v_j, v_j)}. \quad (9)$$

If the mean of the sample set is zero, then c_{ij} is indeed the correlation between DOFs i and j . With this measure and a given threshold, i and j are said to have close affinity if the measure is greater than this threshold. DOFs that are close have a better chance of being “aggregated” together, with one of them selected as the representative CDOF or *seed* of the aggregate. In fact, this affinity measure can be used to determine the CDOFs and the grouping of FDOFs with each CDOF in a similar manner to how the strength-of-connection measure is used to achieve these goals in the Ruge–Stuben method. Specifically, in a similar manner to how the strength-of-connection is used to “fractionally aggregated” the FDOFs (i.e., fractionally in the sense that each FDOF can interpolate from several specified coarse nodes), the affinity measure can be used to fractionally aggregate the FDOFs with the coarse seeds. This will lead to an overlapping partitioning of the DOFs since a FDOF can be part of several aggregates.

The actual partitioning of the DOFs in Reference ¹² is performed by associating to each DOF a *projected volume* of the resulting aggregate if that DOF was selected as a seed. This projected volume explicitly uses the affinity measure. Specifically, for DOF i , let $N(i)$ be its neighborhood (i.e., $N(i) = \{j | L_{h,ij} \neq 0\}$) and let π_i be its volume, which is initially set to 1. The projected volume v_i for i is defined as

$$v_i = \pi_i + \sum_{j \in N(i)} \pi_j \frac{c_{ij}}{\sum_{k \in N(j)} c_{jk}}. \quad (10)$$

All DOFs that have projected volumes greater than a threshold factor (e.g., 1.5) of the average projected volume become CDOFs. This is the first pass for determining C , the set of CDOFs. The second pass can convert a FDOF into a CDOF if its affinity measure to other FDOFs is relatively large or if its weighted graph connections in L_h with FDOFs are relatively large, that is, for some threshold Q and for $i \in F$, if

$$\frac{\sum_{j \in (C \cap N(i))} c_{ij}}{\sum_{j \in N(i)} c_{ij}} \leq Q \quad \text{or} \quad \frac{\sum_{j \in (C \cap N(i))} L_{h,ij}}{\sum_{j \in N(i)} L_{h,ij}} \leq Q, \quad (11)$$

then move i to C . Interpreting, the first inequality holds if the c_{ij} 's to its neighboring FDOFs, $N(i) \setminus (C \cap N(i))$, are large compared with the affinity measure to its neighboring CDOFs, $(C \cap N(i))$. The second inequality has a similar interpretation but in terms of the graph connections.

Since the correlation exposes the dependency between the DOFs, the correlation measure is suitable for systems of PDEs, especially for exposing intervariable and intravariation DOF dependencies. However, a direct application of the above strategy does not utilize nor preserve any of the PDE properties of the problem. Since the scheme is agnostic to the variable type, the nodal structure is lost on the coarser levels, crossvariable interpolation generally will arise, and some of the variable types can vanish on the coarser levels (see Figure 1, which illustrates a coarse aggregate consisting of different variable types at different gridpoints). To avoid these issues and to provide the flexibility to produce intravariation interpolation or to limit the intervariable interpolation to within a radius of a FDOF, we generalize the correlation measure to nodal degrees of freedom, that is, all variable DOFs located at a node. We also use the actual correlation involving the sample mean and denote it by $\text{corr}(\cdot, \cdot)$, that is, for quantities x_l^q and x_m^r

$$\text{corr}(x_l^q, x_m^r) = \frac{\sum_{\alpha=1}^S (x_l^{q,(\alpha)} - \bar{x}_l^q) (x_m^{r,(\alpha)} - \bar{x}_m^r)}{\sqrt{\left(\sum_{\alpha=1}^S (x_l^{q,(\alpha)} - \bar{x}_l^q)^2 \right) \left(\sum_{\alpha=1}^S (x_m^{r,(\alpha)} - \bar{x}_m^r)^2 \right)}},$$

where \bar{x}_*^* is the mean of x_*^* . For ease of presentation, assume that there are only two variable types so that we can denote the multivariable test vector as $\begin{pmatrix} v^1 \\ v^2 \end{pmatrix}^{(\alpha)}$. Let I, J denote two nodes on the finest level, and consider the following correlation matrix defined over the test vectors:

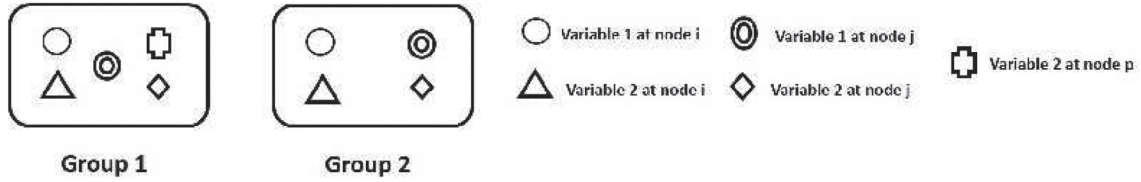


FIGURE 1 Possible coarse aggregates generated by vector-component measure (9) and a nodal-based correlation measure. Aggregate Group 1, produced by vector-component measure (9), generally consists of different variable types at different spatial nodes. Aggregate Group 2, produced by a nodal-based measure, consists of all variable types at the same spatial nodes

$$\text{corr} \left(\begin{pmatrix} v_I^1 \\ v_I^2 \end{pmatrix}, \begin{pmatrix} v_J^1 \\ v_J^2 \end{pmatrix} \right) := \begin{pmatrix} \text{corr}(v_I^1, v_J^1) & \text{corr}(v_I^1, v_J^2) & \text{corr}(v_I^2, v_J^1) & \text{corr}(v_I^2, v_J^2) \\ \text{corr}(v_J^1, v_I^1) & \text{corr}(v_J^1, v_I^2) & \text{corr}(v_J^2, v_I^1) & \text{corr}(v_J^2, v_I^2) \\ \text{corr}(v_I^1, v_J^1) & \text{corr}(v_I^1, v_J^2) & \text{corr}(v_I^2, v_J^1) & \text{corr}(v_I^2, v_J^2) \\ \text{corr}(v_J^1, v_I^1) & \text{corr}(v_J^1, v_I^2) & \text{corr}(v_J^2, v_I^1) & \text{corr}(v_J^2, v_I^2) \end{pmatrix}. \quad (12)$$

Corresponding to each I is a set of correlation matrices, one matrix for each J within a selected pathlength distance from I . The size of this set is then dependent on the selected pathlength distance. Now, taking the measure to be the Frobenius norm of this matrix, that is,

$$C_{II} = \left\| \text{corr} \left(\begin{pmatrix} v_I^1 \\ v_I^2 \end{pmatrix}, \begin{pmatrix} v_J^1 \\ v_J^2 \end{pmatrix} \right) \right\|_F, \quad (13)$$

both spatial and variable crosscorrelations contribute to this value. Alternatively, since the diagonal of matrix (12) is the identity, a modified measure that better exposes the relevant contributions is the Frobenius norm of the off-diagonal matrix of (12).

With either measures, the nodal CDOFs can be determined using the projected volume procedure, or with a strength-of-correlation and coloring scheme similar to the Ruge–Stuben coarsening method. In this article, we consider the second method.

The first step in the coloring scheme is to quantify strongly correlated connections. One way to do this is to take the arithmetic average of the C_{II} 's over all the graph connections within the selected pathlength distance to node I , and to classify node K as strongly influencing I if C_{IK} is greater than a threshold factor of this average. However, notice that even for a scalar system, arithmetically averaging a set of correlations may not accurately produce an average correlation, that is, since a correlation is the cosine of the angle between two vectors, correlations are not additive quantities and thus are not suitable for arithmetic averaging. In fact, a by-product of arithmetic averaging, which was observed experimentally, is a complication in the threshold-parameter tuning, that is, the range of parameter values to explore can be large. To tighten this range, since the sample data are assumed to be approximately bivariate normally distributed, a statistical Z Fisher transformation can be applied to the correlations.^{15,17,18} This transformation produces a variance stabilization that results in the transformed correlations converging to a normal distribution faster than the original set of correlations.¹⁵ Let r denote the correlation between two quantities. The Z value of r is

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r}. \quad (14)$$

These Z values are additive quantities, and thus, an arithmetic averaged \bar{Z} can be computed and an average correlation can be determined by inverting (14) back to r :

$$\bar{r} = \frac{e^{2\bar{Z}} - 1}{e^{2\bar{Z}} + 1}. \quad (15)$$

With this averaging, nodes that are connected to I that have correlations larger than some threshold of this average are classified as strongly influencing I .

To generalize this to systems of PDEs, a Z Fisher transformation is first applied to each matrix element of the correlation matrices associated with node I . This will lead to an average correlation matrix for this node. Taking its Frobenius

norm and comparing it with the C_{IJ} 's, the nodes that strongly influence I can be determined. For example, denoting \bar{C}_I to be the Frobenius norm of this average correlation matrix for node I , then node J is considered to be strongly influencing I if

$$C_{IJ} \geq \theta \bar{C}_I, \quad (16)$$

where θ is a threshold parameter. A standard Ruge–Stuben coloring scheme, which colors a node based on the number of nodes it strongly influences, can be applied to determine the nodal CDOFs: Let F, C, U, G, S_I^t , respectively, denote the FNODEs, CNODEs, unmarked nodes, total grid points, and nodes that I strongly influences. Then

CDOF Coloring Algorithm

Set $F := \emptyset, C := \emptyset, U := G$

1. For $I \in U$, $\lambda_I := |S_I^t \cap U| + 2 |S_I^t \cap F|$
2. If $\lambda_I \neq 0$ then
 - (a) pick $I \in U$ with maximum λ_I and set $C := C \cup \{I\}, U := U \setminus \{I\}$
 - (b) for all $J \in S_I^t \cap U$, set $F := F \cup \{J\}, U := U \setminus \{J\}$
else break
3. Go to 1.

Determining the CDOFs depends on the pathlength radius that correlation matrix (12) is constructed for. For anisotropic problems, a pathlength 2 or larger can better determine the CDOFs that follow the anisotropy. However, for systems of PDEs, using even pathlength 2 can lead to a relatively large number of correlation matrices, which results in a costly coarsening procedure. An alternative approach is to first determine the strong direct connections (i.e., strong pathlength 1 connections) to I , and then construct correlations only for pathlength 2 nodes that are directly connected to these strongly connected nodes. This is still a complex procedure. Fortunately, it was observed that correlation matrices using pathlength 1 can sufficiently capture anisotropies.

Another observation is that correlation matrices involving only the intravariation connections can be constructed. Comparing the Frobenius norm of one of these matrices with its corresponding full correlation matrix can help determine whether intravariation or intervariation interpolation should be used. In fact, since this comparison can be performed at each node, a mixed intravariation/intervariation interpolation can be performed over the grid. For example, taking (12) and the intravariation correlation

$$\widehat{\text{corr}} \left(\begin{pmatrix} v_I^1 \\ v_I^2 \end{pmatrix}, \begin{pmatrix} v_J^1 \\ v_J^2 \end{pmatrix} \right) := \begin{pmatrix} \text{corr}(v_I^1, v_I^1) & \text{corr}(v_I^1, v_J^1) & 0 & 0 \\ \text{corr}(v_J^1, v_I^1) & \text{corr}(v_J^1, v_J^1) & 0 & 0 \\ 0 & 0 & \text{corr}(v_I^2, v_I^2) & \text{corr}(v_I^2, v_J^2) \\ 0 & 0 & \text{corr}(v_J^2, v_I^2) & \text{corr}(v_J^2, v_J^2) \end{pmatrix}, \quad (17)$$

the ratio

$$\frac{\left\| \widehat{\text{corr}} \left(\begin{pmatrix} v_I^1 \\ v_I^2 \end{pmatrix}, \begin{pmatrix} v_J^1 \\ v_J^2 \end{pmatrix} \right) \right\|_F}{\left\| \text{corr} \left(\begin{pmatrix} v_I^1 \\ v_I^2 \end{pmatrix}, \begin{pmatrix} v_J^1 \\ v_J^2 \end{pmatrix} \right) \right\|_F} \quad (18)$$

can be used to determine whether intravariation or intervariation interpolation should be used at node I . Specifically, if ratio (18) is approximately 1, then intravariation interpolation can be used at I ; otherwise if (18) is small, intervariation interpolation is used.

4 | BAMG WEIGHTED LEAST-SQUARES INTERPOLATION

With the selection of the CDOFs performed, a BAMG weighted least-squares interpolation can be applied to construct P . This will be achieved using a modified form of local least-squares problem (5). To show this, we first note the close connection between (5) and the global measure $M_1(Q, u_h)$ of (4). The unweighted form of (5) simplifies the illustration of this connection. Reordering the DOFs so that all the FDOFs are ranked before the CDOFs, and taking $\hat{R} = [0 \ I]$, where 0 and I are, respectively, the $(|\text{CDOFs}| \times |\text{FDOFs}|)$ zero matrix and $(|\text{CDOFs}| \times |\text{CDOFs}|)$ identity matrix, problem (5) can be viewed as approximately constructing row i of a P that minimizes a local M_1 -related measure. To be precise, consider the global-grid problem

$$\arg \min_P \left[\max_{u_h} M_1(P [0 \ I], u_h) \right] = \arg \min_P \left[\max_{u_h} \frac{\|(I - P [0 \ I])u_h\|_{D_h}}{\|u_h\|_{L_h}} \right], \quad (19)$$

whose solution is an ideal interpolation operator P . The i th row of $(I - P [0 \ I])u_h$ is

$$[(I - P [0 \ I])u_h]_i = u_{h,i} - \sum_{j \in C_i} w_{ij} u_{h,j}. \quad (20)$$

Now consider modifying (19) to consist of only its numerator and with the D_h norm replaced with the standard l_2 norm. Then (5) can be viewed as a procedure for constructing an approximation to the i th row of P for this modified problem, with the min-max problem minimized over the set of test vectors.

The ideal interpolation described through (19) must interpolate the near-nullspace components accurately, that is, the maximum occurs when u_h is a near-nullspace component of L_h , which are roughly exposed in the test vectors. Assuming that the components of D_h are positive, a local weighted least-squares problem that better aligns with (19) is

$$\{w_{ij} | j \in C_i\} = \arg \min_{w_{ij}} \sum_{\alpha=1}^s \frac{D_{h,i}}{\|v^{(\alpha)}\|_{L_h}^2} \left(v_i^{(\alpha)} - \sum_{j \in C_i} w_{ij} v_j^{(\alpha)} \right)^2. \quad (21)$$

Of course, this minimization process is independent of $D_{h,i}$. However, this value can be used as an indicator for adapting the caliber of interpolation. At locations i where $D_{h,i}$ is substantial larger than at other locations, the caliber of interpolation is increased. This will allow a uniform distribution of the interpolation error over the grid.

Problem (21) can be extended to systems of PDEs to allow intravariation and intervariation interpolation. Since a nodal coarsening is performed, for intervariation interpolation, the only change in (21) is in the indicator for caliber adaptation. For example, for node I , $D_{h,I}$ can be the Frobenius norm of the submatrix of \mathcal{L}_h involving the rows and columns corresponding to node I . The least-squares problem will be

$$\{w_{IJ} | J \in C_I\} = \arg \min_{w_{IJ}} \sum_{\alpha=1}^s \frac{D_{h,I}}{\|v^{(\alpha)}\|_{\mathcal{L}_h}^2} \left(v_I^{(\alpha)} - \sum_{J \in C_I} w_{IJ} v_J^{(\alpha)} \right)^2. \quad (22)$$

For intravariation interpolation, the elements of block P_{ll} of (6) are computed using a least-squares problem involving only the variable l subvectors of the test vectors. Let $v^{(\alpha,l)}$ be the variable l subvector of $v^{(\alpha)}$ and $\mathcal{L}_{h,ll}$ denote the (variable l)-to-(variable l) submatrix of \mathcal{L}_h . Then the least-squares problem will be

$$\{w_{IJ}^l | J \in C_I\} = \arg \min_{w_{IJ}^l} \sum_{\alpha=1}^s \frac{D_{h,I}}{\|v^{(\alpha,l)}\|_{\mathcal{L}_{h,ll}}^2} \left(v_I^{(\alpha,l)} - \sum_{J \in C_I} w_{IJ}^l v_J^{(\alpha,l)} \right)^2. \quad (23)$$

Here, the nodal $D_{h,I}$ is used, implying that the same caliber adaptation will be applied to each variable at a given node.

4.1 | iBAMG Scheme

The interpolation weights generated by the BAMG procedures *implicitly* depend on the operator coefficients through the smoothed test vectors. The indirect BAMG (iBAMG) technique of Reference ¹⁴ constructs the weights using an *explicit*

dependence on these coefficients. Consider a fine node I . This explicit dependence is achieved by an indirect interpolation that approximates some of the FDOF connections to I with CDOF connections to I . We extend this technique to systems of PDEs. After a local relaxation update for this node, the residual of the homogeneous system is

$$r_I = \mathcal{L}_{II}v_I + \mathcal{L}_{IF_I}v_{F_I} + \mathcal{L}_{IC_I}v_{C_I} \approx 0, \quad (24)$$

where the \mathcal{L}_{I*} 's are block submatrices consisting of the rows and columns of \mathcal{L} corresponding to nodes I and $*$, and where F_I and C_I are the FDOF and CDOF direct neighbors of node I . For simplicity, we assume that F_I consists of only strong connections. The first step of one variant of iBAMG is to approximate $\mathcal{L}_{IF_I}v_{F_I}$ in terms of the v_{C_I} 's. This is done using a least-squares procedure such as

$$\{\tilde{w}_{IJ}|J \in C_I\} = \arg \min_{\tilde{w}_{IJ}} \sum_{\alpha=1}^s \frac{D_{h,I}}{\|v^{(\alpha)}\|_{\mathcal{L}_h}^2} \left(\mathcal{L}_{IF_I}v_{F_I}^{(\alpha)} - \sum_{J \in C_I} \tilde{w}_{IJ}v_J^{(\alpha)} \right)^2. \quad (25)$$

Using this in (24), the interpolation weights for node I are

$$w_{IJ} = \mathcal{L}_{II}^{-1} \left(-\tilde{w}_{IJ} - \mathcal{L}_{IC_I} \right), \quad J \in C_I. \quad (26)$$

Another variant of iBAMG, which aligns more with the original Ruge–Stuben approach, is to approximate each v_{F_I} in terms of the v_{C_I} 's by solving a set of $|F_I|$ least-squares problems, for example,

$$\{\tilde{w}_{F_I J}|J \in C_I\} = \arg \min_{\tilde{w}_{F_I J}} \sum_{\alpha=1}^s \frac{D_{h,I}}{\|v^{(\alpha)}\|_{\mathcal{L}_h}^2} \left(v_{F_I}^{(\alpha)} - \sum_{J \in C_I} \tilde{w}_{F_I J}v_J^{(\alpha)} \right)^2. \quad (27)$$

Once this is done, using (24) leads to the weights

$$w_{IJ} = \mathcal{L}_{II}^{-1} \left(-\mathcal{L}_{IF_I}[\tilde{w}]_{F_I C_I} - \mathcal{L}_{IC_I} \right), \quad (28)$$

where $[\tilde{w}]_{F_I C_I}$ is the matrix of indirect interpolation values obtained by solving the $|F_I|$ least-squares problems. However, since this approach is rather expensive, we consider only the first approach.

4.2 | Number of relaxation sweeps and test vectors

In the literature, determining the number of relaxation sweeps and test vectors in (i)BAMG is an unresolved problem even for scalar PDE problems. For these scalar problems, most of this literature assumes that the grid hierarchy is given, for example, by a geometric coarsening. With this hierarchy, ≈ 10 relaxation sweeps and 5–10 test vectors are often used in the BAMG procedures.^{13,14} In this article, since statistical correlations are employed; a larger number of test vectors is needed to capture the statistical DOF dependencies, and hence to determine the CDOFS. However, not all of these test vectors are needed in the construction of the interpolation operator, that is, the least-squares problems can be defined for only a subset of the $v^{(\alpha)}$'s. Nevertheless, requiring more test vectors means that the BAMG method is more costly for systems of PDEs, which should not be surprising. Moreover, more relaxation sweeps are generally needed to expose the near-nullspace for systems of PDEs. However, rather than using ≈ 100 sweeps, which is often needed to sufficiently capture the global nature of the near-nullspace, only $O(10)$ sweeps will be used. Justification for this relaxation reduction can be extracted from the theory of optimal interpolation described originally in Reference²⁰ for the Richardson iteration and later in Reference²⁶ for more general iterations. In Reference²⁰, it was shown that for an $(m \times m)$ real symmetric positive definite matrix with orthonormal eigenvectors $\{p_i\}$ and eigenvalues $\{\lambda_i\}$, $\lambda_1 \leq \dots \leq \lambda_m$, if \hat{R} is chosen such that

$$\hat{R}^t = [p_1, \dots, p_c], \quad c < m \quad (29)$$

then $P = \hat{R}^t$ is the optimal interpolation in the sense that it is the arg min of

$$\arg \min_P \left[\max_{u_h} M_2(P\hat{R}, u_h) \right]. \quad (30)$$

Furthermore, it was shown that

$$\min_P \left[\max_{u_h} M_2(P\hat{R}, u_h) \right] = \frac{\lambda_n}{\lambda_c}, \quad (31)$$

which leads to a two-grid convergence bound of $\left(1 - \frac{\lambda_c}{\lambda_n}\right)$. \hat{R} then defines the CDOFS to be the coordinates of a vector with respect to the eigenvectors $\{p_1, \dots, p_c\}$, and the coarse-grid operator obtained by a Galerkin coarsening is a Galerkin projection of the fine-grid operator onto the span of these eigenvectors. The CDOFs should be selected to ensure that these can be accurately accomplished for a reasonable number of eigenvectors. However, applying a lot of relaxation sweeps to the test vectors generally produces only an accurate approximation to p_1 , leading to a CDOF selection that accurately represents mainly this eigenvector. The range of P is then approximately the span of only p_1 and hence, the resulting two-grid method is approximately a 1-vector deflation scheme. Applying a few sweeps will produce a less accurate approximation to p_1 , but will enrich the test vectors with components in more of the algebraically smooth eigenvectors. Selecting the coarse nodes based on the correlations of these smoothed test vectors will lead to a better representation of more of these smooth eigenvectors.

5 | NUMERICAL EXPERIMENTS

In this section, we examine whether this correlation matrix method can detect the appropriate CDOFs and whether the combination of this CDOF selection and the BAMG interpolation construction can lead to an effective multigrid algorithm for systems of PDEs. Since the prototype software was implemented in MATLAB, we consider only systems in 2-d.

Before looking at some challenging systems of PDEs, we consider the approach's performance on a system of Laplace operators defined in the unit square and with homogeneous Dirichlet conditions:

$$\mathcal{L}u = \begin{bmatrix} -2\Delta & \Delta & 0 & \dots & 0 \\ \Delta & -2\Delta & \Delta & & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \Delta & -2\Delta & \Delta \\ 0 & \dots & \dots & \Delta & -2\Delta \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = 0 \quad (32)$$

(Although a Kronecker product preconditioner method¹⁶ can be used, we are interested in a stand-alone AMG solver for tackling this system.) The system was discretized using finite elements with bilinear polynomials on a uniform grid. We use the correlation/Ruge–Stuben coloring scheme to determine the CDOFs and consider BAMG with intravariabe, intervariable, and mixed-variable interpolation, and iBAMG with intravariabe and intervariable interpolation, which are respectively denoted by iBAMG1 and iBAMG2. For the mixed-variable interpolation, denoted by “mixed,” intervariable interpolation is used on all nodes except for nodes, where the intravariabe correlation is dominant in the correlation matrix (see Equation (18)), for which intravariabe interpolation is used. The strength-of-correlation threshold θ in (16) was set to 0.75, the number of smoothing sweeps and test vectors were set to 20 and 50, and the caliber of interpolation was set to 8. Pointwise Gauss–Seidel was used since nodal Gauss–Seidel was observed to produce similar results. To measure the convergence rate, 30 V(1,1) multigrid cycles were applied to a homogeneous problem with a random initial guess, and the rate was defined as $\frac{\|u_h^{30}\|}{\|u_h^{29}\|}$. This measure was selected because the rate between the last two iterations gave the slowest rate over the 30 cycles, that is, best reflected the spectral radius. Table 1 contains the results. For $n = 1$, a scalar Laplace equation, we obtain fair rates although not as good as a standard AMG method. This is not usual since standard AMG's strength-of-connection measure and interpolation construction explicitly use the coefficients of the operator. For $n = 2, 3$, we see that the new method achieves reasonable rates for the intravariabe BAMG schemes, while the intervariable schemes do not. This can be explained by the structure of the coarser-level operators: when the component operators have the same differential form, intravariabe interpolation with the Galerkin construction (7) of the coarse-grid

h	$n = 1$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.23	—	—	0.31	—
1/50	0.26	—	—	0.22	—
1/100	0.23	—	—	0.26	—
h	$n = 2$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.21	0.45	0.34	0.21	0.32
1/50	0.23	0.74	0.55	0.18	0.48
1/100	0.34	0.74	0.80	0.20	0.35
h	$n = 3$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.39	0.94	0.87	0.37	0.79
1/50	0.39	0.93	0.90	0.38	0.87
1/100	0.47	0.93	0.91	0.39	0.81

Note: For $n = 1$, a scalar Laplace equation, only the intravariante BAMG and iBAMG methods are applicable.

TABLE 1 Multigrid rates for coupled systems of Laplacians using the different BAMG schemes

operator preserves the PDE structure while intervariable interpolation generally does not.¹⁶ This also explains the poor performance of the mixed-variable interpolation since intervariable interpolation was used on a substantial number of nodes.

We now consider the following more difficult systems:

Problem 1: elasticity

$$\mu \Delta u + (\lambda + \mu) \nabla \nabla \cdot u = 0,$$

which has the rigid body modes as near-nullspace components, and for large λ -to- μ ratios, has a large near-nullspace. That is, we have $\frac{\lambda}{\mu} = \frac{2\nu}{1-2\nu}$, where ν is the Poisson ratio of the material. Hence, if this ratio is large, $\nu \rightarrow 0.5$ and we are approaching the incompressible limit, where problematic divergence-free modes arise. Since large near-nullspace issues are out of the scope of this article, we will not consider these scenarios. Nevertheless, even away from the incompressible limit, developing an AMG solver for these equations is difficult. Indeed, these elasticity equations have been the impetus for much of the AMG developments over the past two decades.^{2-4,7,8} The difficulty arises from the multidimensional near-nullspace (i.e., rigid body modes) and the non-M matrix property of the discretized system.

Problem 2: coupled anisotropic diffusion

$$\begin{bmatrix} \mathcal{L}_{\epsilon_1 \theta_1} & \alpha \partial_{xy} \\ \alpha \partial_{yx} & \mathcal{L}_{\epsilon_2 \theta_2} \end{bmatrix} u = 0,$$

where $\mathcal{L}_{\epsilon_i \theta_i} = (\cos^2 \theta_i + \epsilon_i \sin^2 \theta_i) \partial_{xx} + (1 - \epsilon_i) \sin(2\theta_i) + (\sin^2 \theta_i + \epsilon_i \cos^2 \theta_i) \partial_{yy}$. By taking $\epsilon_i \ll 1$ and $\theta_i \neq \frac{k\pi}{2}$ for integer k , $\mathcal{L}_{\epsilon_i \theta_i}$ will lead to an anisotropic, nongrid aligned diffusion operator. We note that scalar anisotropic, nongrid aligned diffusion problems are known to be challenging for AMG (e.g., in Reference²⁷, for $\epsilon_i = 0.01$, $\theta_i = \frac{\pi}{4}$, an accurate V(1,1) two-grid rate of ≈ 0.66 was computationally obtained).

Problem 3: first-order system least-squares (FOSLS) for neutron transport

$$\left[\sigma_t I - \frac{1}{5\sigma_t} \Delta - \left(\frac{\sigma_t}{3} + \frac{1}{15\sigma_t} \right) \nabla \nabla \cdot \right] u = 0$$

where σ_t is the total cross-section of the medium that the transport is occurring in. This is the notorious two-two block of the FOSLS PN formulation of the Boltzmann transport equation, which describes the transport of a stream of neutrons/photons as these particles collide with the nuclei of the medium.^{28,29} When the cross-section is large, we see that the grad-div component of the operator dominates, which implies a large near-nullspace.

For each of these problems, we assume homogeneous Dirichlet boundary conditions and a finite element discretization with bilinear polynomials.

CDOF selection: We consider the elasticity and anisotropic diffusion problems to illustrate the CDOF selection, particularly the algorithm's ability to trace anisotropies. To obtain easy visualization of the CDOFs, the size of the finest mesh was 50^2 . For Problem 1, the chosen domains are a unit square and a beam of size $[0, 10] \times [0, 1]$, which leads to a stretched grid. For the unit square, the Lamé constants (λ, μ) were set to $(1, 1)$, $(10, 1)$, which correspond to Poisson's ratios of 0.25 and 0.4545, respectively. For the beam, (λ, μ) were set to $(1, 1)$, $(5, 1)$. Finally, the strength-of-correlation threshold was set to ≈ 1 for the nonstretched grid cases and ≈ 1.6 for the stretched grid case. The larger threshold value will ensure that only the most highly correlated nodes will be selected as strongly influencing nodes, which for the stretched grid case will ensure tracking of anisotropies. For Problem 2, only the unit square was considered. For this problem, we took $\epsilon_1 = \epsilon_2 = \alpha = 0.01$, ranged the angles θ_1, θ_2 over a set of values, and set the threshold parameter to 1.1. We note that these problems are nontrivial for multigrid.

Figure 2 illustrates the CDOF selection for the first four coarser levels for Problem 1. From top to bottom and left to right, the plots are for the unit square with $(\lambda, \mu) = (1, 1)$, unit square with $(\lambda, \mu) = (10, 1)$, beam with $(\lambda, \mu) = (1, 1)$,

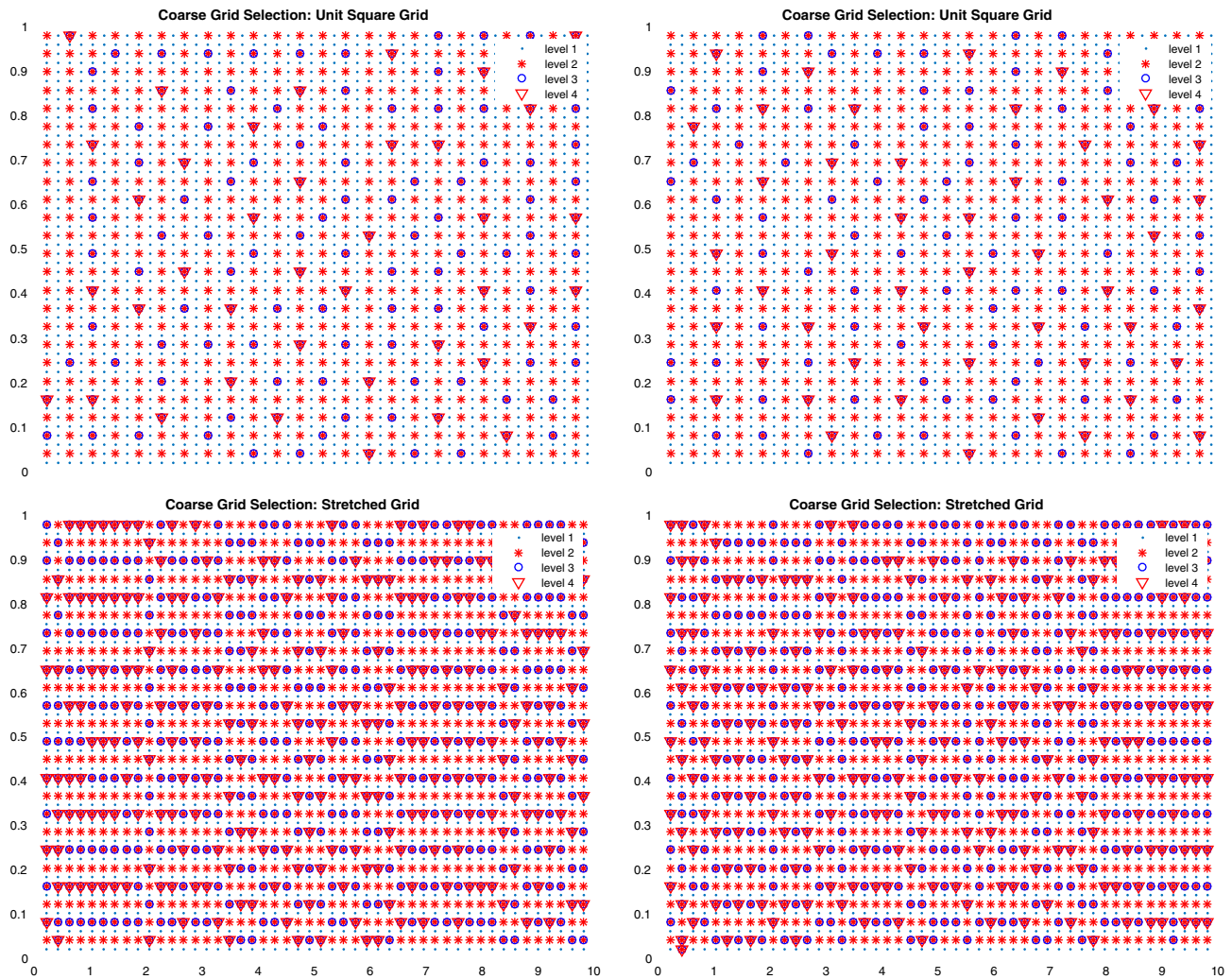


FIGURE 2 CDOF selection for the elasticity problem: $(\lambda, \mu) = (1, 1)$, $(10, 1)$ on the unit square and $(\lambda, \mu) = (1, 1)$, $(5, 1)$ on the beam $[0, 10] \times [0, 1]$ using 20 smoothing sweeps on 80 test vectors. From top to bottom and left to right, the plots are for the unit square with $(\lambda, \mu) = (1, 1)$ then $(10, 1)$, and then for the beam with $(\lambda, \mu) = (1, 1)$, $(5, 1)$

and beam with $(\lambda, \mu) = (5, 1)$. For the square domain, 20 smoothing sweeps on 80 test vectors were used; for the beam, three smoothing sweeps on 80 test vectors were used. We observed that a full coarsening is appropriately obtained in the unit square scenarios, and a coarse grid that follows the anisotropy is appropriately obtained in the beam scenarios. Figure 3 illustrates the CDOF selection for the rotated anisotropic diffusion problem for angles $(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, -\frac{\pi}{4}\right), \left(-\frac{\pi}{4}, 0\right), \left(-\frac{\pi}{4}, \frac{\pi}{4}\right), \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$. Fifteen smoothing sweeps on 80 test vectors were applied. When θ_1 aligns with θ_2 , the coarse grids are clearly appropriately chosen. When these angles are not aligned, the grids are suitably chosen, for example, for the extreme case $\theta_1 = -\theta_2$, a standard coarsening, which does not biased toward one of the anisotropies, is obtained.

For comparison, we consider the CDOF selection when an arithmetic average is used to compute the average correlation and when a projected volume approach is used instead of the correlation/Ruge–Stuben coloring scheme. We consider only the rotated anisotropic diffusion problem. Because the arithmetic average produced approximately the same CDOF selection for 15 smoothing sweeps on 80 test vectors, we consider only 40 test vectors. Since a Z Fisher transform stabilizes the variance in the sample correlations, the CDOF selection for the Z transform should be more appropriate. Figures 4 and 5 show the results. Except for angles $\left(-\frac{\pi}{4}, 0\right)$, the Z transform method produces roughly the same CDOFs when 40 and 80 test vectors were used. Using arithmetic averaging, we observe denser grids on the coarser levels for the $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$ case. Finally, for a projected volume approach with a second pass given by the left condition of (11) with $Q = 0.5$,

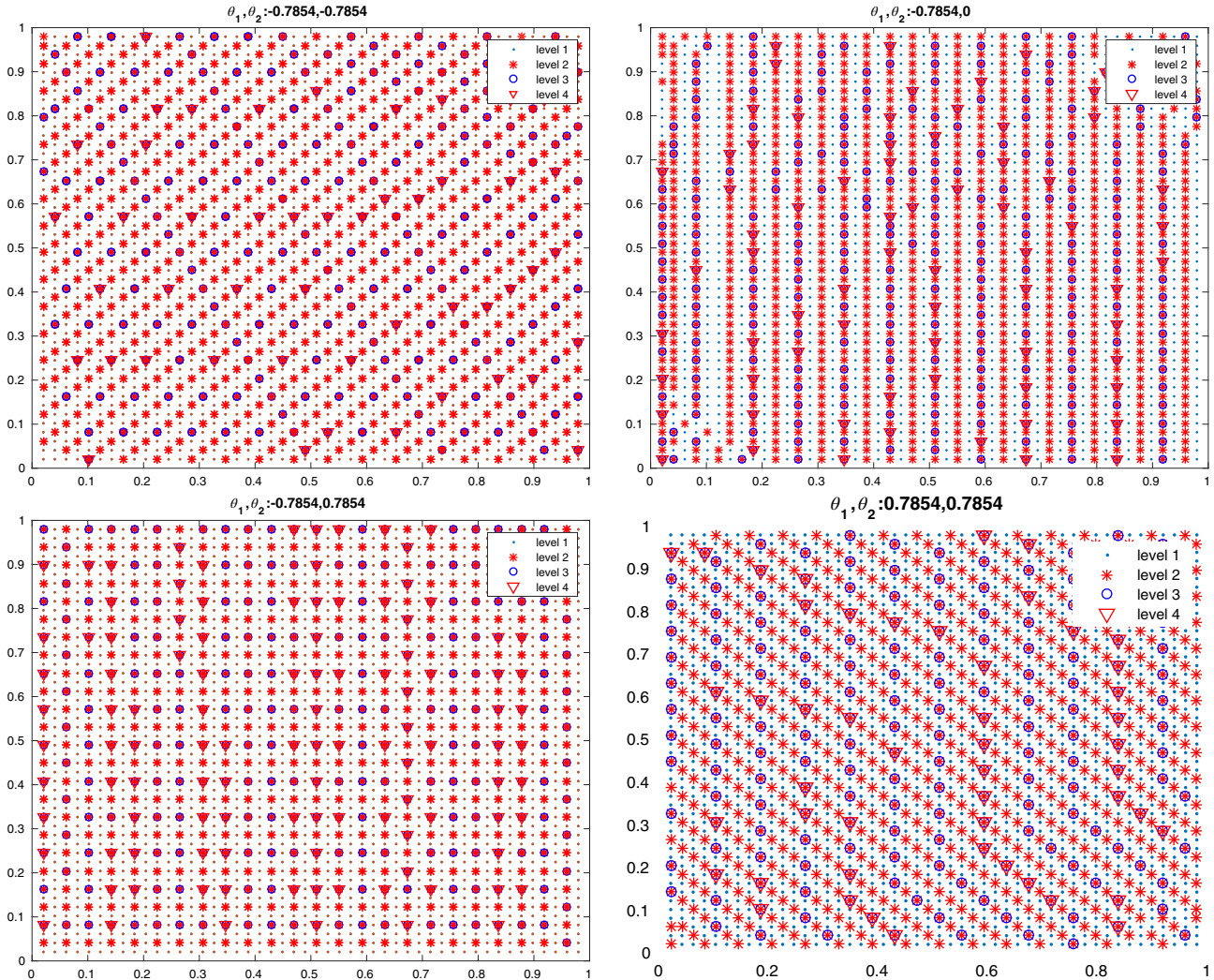


FIGURE 3 CDOF selection for the rotated anisotropic diffusion with angles $(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, -\frac{\pi}{4}\right), \left(-\frac{\pi}{4}, 0\right), \left(-\frac{\pi}{4}, \frac{\pi}{4}\right), \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$ using Z Fisher transformation of the correlations and 15 smoothing sweeps on 80 test vectors

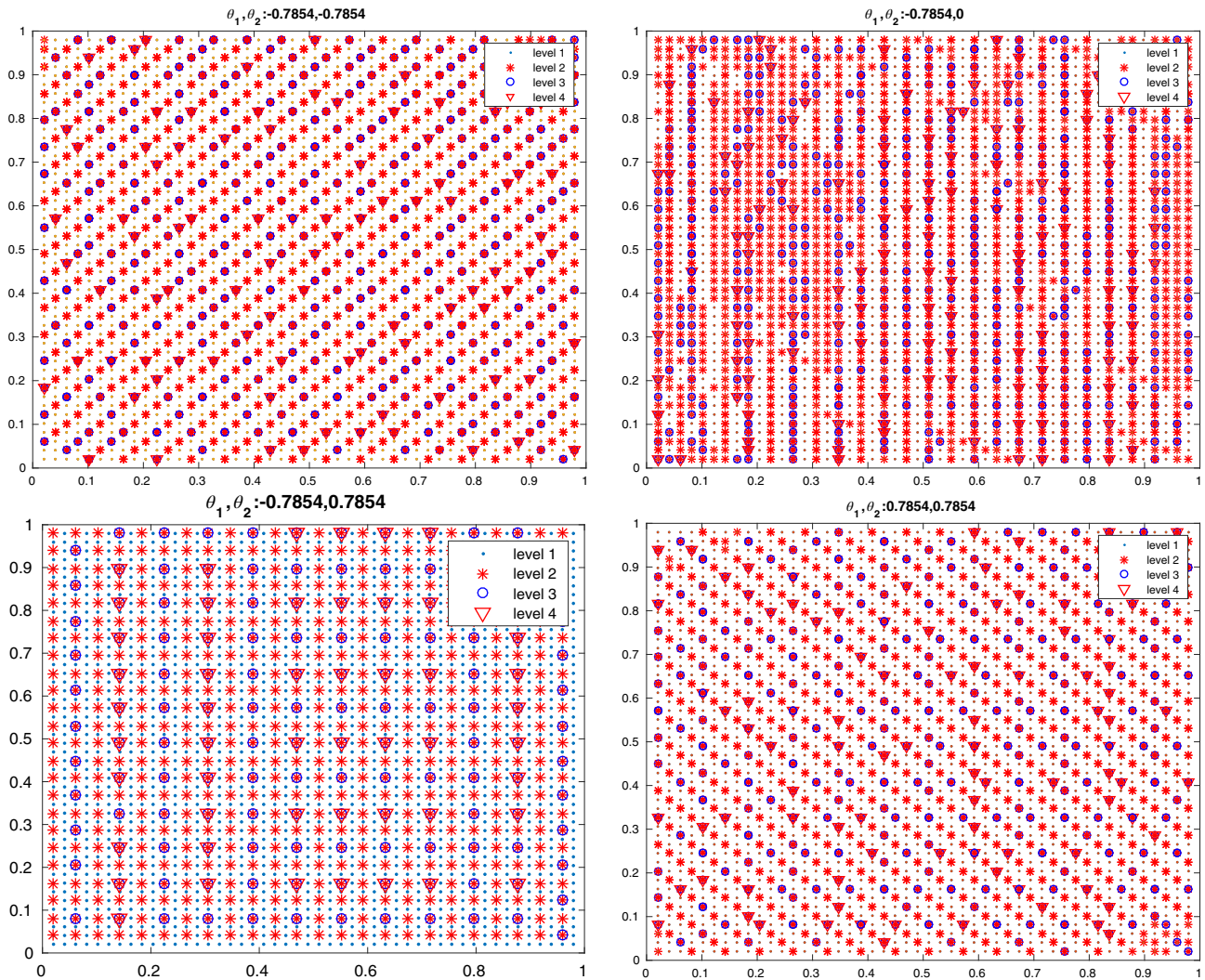


FIGURE 4 CDOF selection for the rotated anisotropic diffusion with angles $(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, -\frac{\pi}{4}\right), \left(-\frac{\pi}{4}, 0\right), \left(-\frac{\pi}{4}, \frac{\pi}{4}\right), \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$ using the Z Fisher transformation of the correlations and 15 smoothing sweeps on 40 test vectors

we obtain the CDOF selection given in Figure 6 for $\left(-\frac{\pi}{4}, -\frac{\pi}{4}\right), \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$. Again 15 smoothing sweeps on 80 test vectors were employed. Tracking of the anisotropies is not obtained, which follows because the projected volume smears away any anisotropy detected by the correlations (i.e., the projected volume for node I involves a weighted average of all the correlations between I and $J \in N(I)$).

Two-grid and multigrid performance: We now consider the two-grid and multigrid performance for the three test problems. Fifty to 80 test vectors were used for each problem, but only 30 were used in the construction of the interpolation operators, and only $O(10)$ pointwise Gauss–Seidel smoothing sweeps were applied to these vectors (nodal Gauss–Seidel generated similar results). As in the earlier experiments, 30 V(1,1) multigrid and two-grid cycles were applied to a random initial guess and the convergence rate is measured as $\frac{\|u_h^{30}\|}{\|u_h^{29}\|}$. Since the CDOFs are reasonably selected for the second level, the computed two-grid rates will be the target rate for the multigrid schemes. We note that the multigrid rates may be slightly better than the two-grid rates. This is due to the random features of the algorithm (the two-grid and multigrid runs were conducted separately).

Problem 1: elasticity The strength-of-correlation threshold was set to 0.85 and 1.8 in the square and beam scenarios, respectively. Also, the number of smoothing sweeps that are applied to the test vectors were 20 and 5 for the square and beam scenarios, respectively, and caliber 3 interpolation was used (caliber 5 at points where $\|D_{h,I}\|$ is large). Table 2 tabulates the multigrid and two-grid rates with the two-grid rates in parenthesis. We observe that for the square domain with $(\lambda, \mu) = (1, 1)$, the intravariabe BAMG and iBAMG schemes perform the best with the multigrid rates slightly

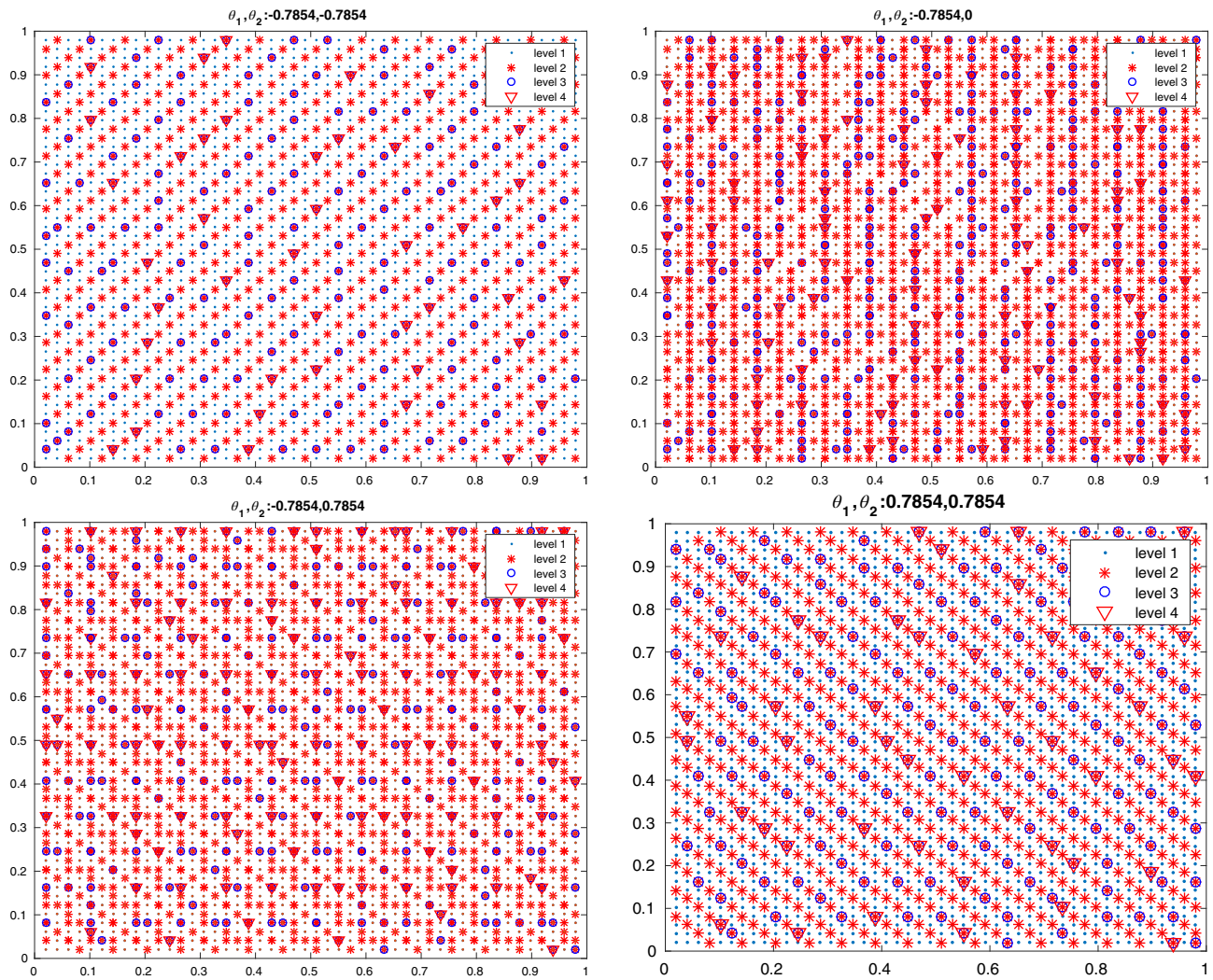


FIGURE 5 CDOF selection for the rotated anisotropic diffusion with angles $(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, -\frac{\pi}{4}\right), \left(-\frac{\pi}{4}, 0\right), \left(-\frac{\pi}{4}, \frac{\pi}{4}\right), \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$ using arithmetic averaging of the correlations using the same threshold parameters as in (3) and 15 smoothing sweeps on 40 test vectors

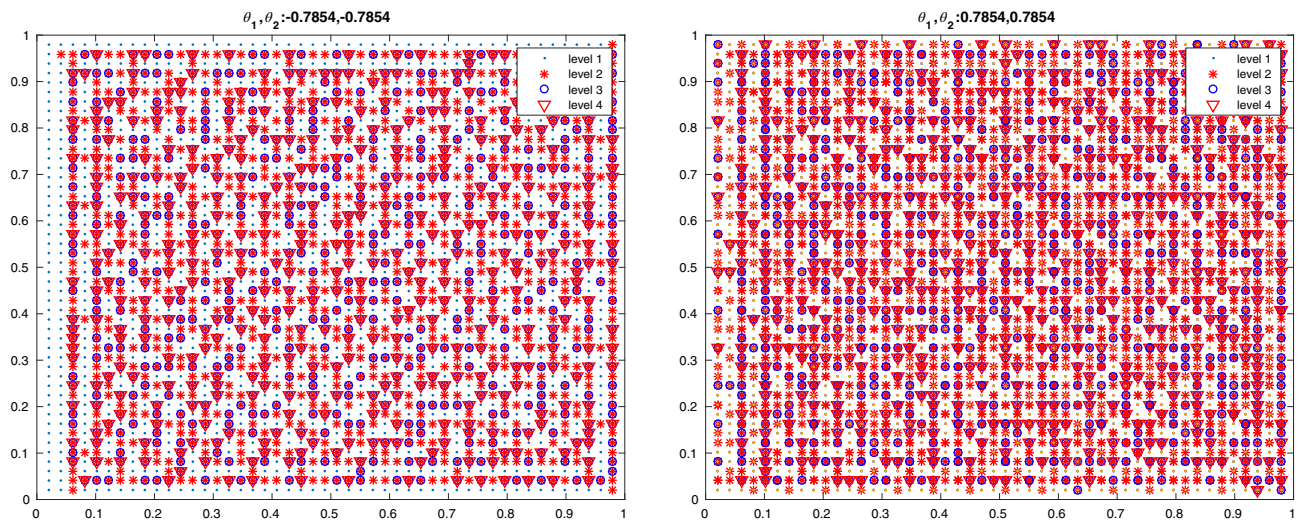


FIGURE 6 CDOF selection for the rotated anisotropic diffusion with angles $(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, -\frac{\pi}{4}\right), \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$ using projected volumes with 15 smoothing sweeps on 80 test vectors

TABLE 2 Multigrid and two-grid (in parenthesis) rates for Problem 1 scenarios using different types of interpolation

Problem 1:					
square $(\lambda, \mu) = (1, 1)$					
h	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.33 (0.28)	0.34 (0.27)	0.42 (0.27)	0.36 (0.26)	0.31 (0.29)
1/50	0.29 (0.27)	0.39 (0.23)	0.44 (0.24)	0.35 (0.26)	0.34 (0.22)
1/100	0.39 (0.30)	0.37 (0.27)	0.54 (0.30)	0.36 (0.30)	0.40 (0.27)
h	Square $(\lambda, \mu) = (10, 1)$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.70 (0.66)	0.77 (0.61)	0.69 (0.64)	0.67 (0.65)	0.73 (0.61)
1/50	0.75 (0.66)	0.82 (0.65)	0.76 (0.65)	0.71 (0.65)	0.80 (0.65)
1/100	0.78 (0.66)	0.83 (0.65)	0.79 (0.66)	0.78 (0.66)	0.79 (0.66)
h	Beam $(\lambda, \mu) = (1, 1)$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.31 (0.22)	0.25 (0.22)	0.28 (0.21)	0.31 (0.14)	0.18 (0.15)
1/50	0.76 (0.31)	0.71 (0.33)	0.86 (0.86)	0.68 (0.16)	0.45 (0.19)
1/100	0.76 (0.36)	0.74 (0.37)	0.92 (0.91)	0.76 (0.25)	0.68 (0.25)
h	Beam $(\lambda, \mu) = (5, 1)$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.77 (0.25)	0.58 (0.33)	0.85 (0.85)	0.57 (0.09)	0.73 (0.16)
1/50	0.87 (0.36)	0.84 (0.34)	0.91 (0.91)	0.84 (0.19)	0.83 (0.20)
1/100	0.87 (0.44)	0.91 (0.41)	0.94 (0.94)	0.90 (0.25)	0.86 (0.25)

TABLE 3 Multigrid rates for the beam scenarios using a larger strength-of-correlation threshold on the coarser levels

Beam $(\lambda, \mu) = (1, 1)$					
h	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/50	0.33	0.35	0.83	0.13	0.22
1/100	0.41	0.45	0.92	0.25	0.25
h	Beam $(\lambda, \mu) = (5, 1)$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/50	0.41	0.45	0.89	0.23	0.21
1/100	0.52	0.52	0.95	0.27	0.25

larger than the two-grid rates. For the square domain with $(\lambda, \mu) = (10, 1)$, which is more difficult, we again see that intravariabele BAMG and iBAMG schemes perform the best with an acceptable increase in the multigrid rates over the two-grid rates. The results, however, are different for the beam scenarios. Now the multigrid rates substantially slower than the two-grid rates for all methods except for the mixed interpolation scheme, which already has a poor two-grid rate. Except for the mixed interpolation scheme, the disparity between the two-grid and multigrid rates indicates that the coarser grids/systems may have been poorly selected/constructed. This is verified by the improved results in Table 3, which were obtained using a larger strength-of-correlation threshold of 2.5–3.0 on the coarser levels. As for the mixed-variable interpolation, we now see that its performance does not reflect the performance of either the intravariabele and intervariable performance. This indicates that mixing the interpolation can lead to poor coarse-grid approximations, as illustrated by the poor two-grid rates in the beam scenarios. Further analysis of this will be conducted in the future.

We note that the performance of these methods were about the same whether the near-nullspace components were included or not in the set of test vectors. Moreover, to illustrate that applying a lot of smoothing to the test vectors can lead

Square $(\lambda, \mu) = (1, 1)$					
#Sweeps	Intra	Inter	Mixed	iBAMG1	iBAMG2
20	0.29	0.39	0.44	0.35	0.34
200	0.45	0.41	0.62	0.41	0.38

TABLE 4 Multigrid rates using 20 and 200 smoothing sweeps on the test vectors

to slower convergence, we consider the square scenario with $(\lambda, \mu) = (1, 1)$ for $h = \frac{1}{50}$ and take 200 smoothing sweeps. Table 4 shows that the convergence degrades with additional sweeps.

Problem 2: anisotropic diffusion The strength-of-correlation threshold was set to 1.1 in all cases except $(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$, which was set to 0.7. The number of test vectors and smoothing sweeps applied on them were 80 and 15, and the interpolation caliber was 3. Table 5 illustrates the convergence rates. For $\theta_1 = \theta_2$, the multigrid rates is sufficiently close to the two-grid rates for these difficult problems (recall that for the scalar anisotropic diffusion with $\theta_1 = \frac{\pi}{4}$, the AMG V(1,1) two-grid²⁷ rate is about 0.66), with iBAMG1 performing slightly better than the others. However, when $\theta_1 \neq \theta_2$, issues arise. This is not surprising since now a good CDOF selection is difficult to achieve because of the opposing anisotropies in the diagonal operators. We also note that the performance of the mixed-variable interpolation scheme is reflective of the performance of the intervariable interpolation scheme. Additional development/analysis of this will be conducted in the future.

Problem 3: FOSLS for neutron transport The strength-of-correlation threshold was set to 0.85 in all cases, the number of test vectors and smoothing sweeps applied to them were 50 and 20, and the interpolation caliber was 3. Table 6 tabulates the convergence rates. We see that as σ_i increases, the convergence does degrade because of the increasing dominance of the grad-div operator. We also see a disparity between the multigrid and two-grid rates. Tuning the strength-of-correlation threshold parameter may ameliorate some of this. Finally, again the performance of the mixed-variable interpolation scheme is reflective of the performance of the intervariable interpolation scheme.

Problem 2: $(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, -\frac{\pi}{4}\right)$					
h	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.42 (0.26)	0.38(0.30)	0.45(0.31)	0.40 (0.24)	0.42 (0.28)
1/50	0.46 (0.40)	0.50(0.49)	0.46(0.40)	0.30 (0.27)	0.36 (0.32)
1/100	0.73 (0.73)	0.78(0.74)	0.74(0.71)	0.62 (0.57)	0.75 (0.63)
h	$(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, 0\right)$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.83 (0.75)	0.86 (0.84)	0.86 (0.85)	0.83 (0.82)	0.86 (0.80)
1/50	0.83 (0.78)	0.84 (0.73)	0.87 (0.84)	0.80 (0.80)	0.85 (0.76)
1/100	0.93 (0.88)	0.92 (0.87)	0.89 (0.87)	0.91 (0.91)	0.91 (0.87)
h	$(\theta_1, \theta_2) = \left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.59 (0.34)	0.59 (0.34)	0.57 (0.33)	0.58 (0.34)	0.62 (0.36)
1/50	0.73 (0.50)	0.72 (0.53)	0.70 (0.49)	0.66 (0.44)	0.69 (0.51)
1/100	0.85 (0.75)	0.89 (0.76)	0.86 (0.76)	0.88 (0.72)	0.89 (0.76)
h	$(\theta_1, \theta_2) = \left(\frac{\pi}{4}, \frac{\pi}{4}\right)$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.42 (0.27)	0.45 (0.29)	0.38 (0.26)	0.39 (0.21)	0.37 (0.28)
1/50	0.48 (0.44)	0.58 (0.48)	0.53 (0.42)	0.35 (0.30)	0.38 (0.37)
1/100	0.77 (0.71)	0.79 (0.75)	0.76 (0.71)	0.66 (0.62)	0.72 (0.62)

TABLE 5 Multigrid and two-grid (in parenthesis) rates for Problem 2 scenarios using different types of interpolation

TABLE 6 Multigrid and two-grid (in parenthesis) rates for Problem 3 scenarios using different types of interpolation

h	Problem 3: $\sigma_t = 1$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.36 (0.35)	0.31 (0.30)	0.34 (0.31)	0.37 (0.31)	0.33 (0.28)
1/50	0.35 (0.27)	0.39 (0.23)	0.46 (0.29)	0.38 (0.28)	0.35 (0.25)
1/100	0.42 (0.30)	0.46 (0.27)	0.40 (0.31)	0.33 (0.30)	0.44 (0.25)
h	$\sigma_t = 2$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.62 (0.59)	0.60 (0.56)	0.62 (0.56)	0.60 (0.59)	0.56 (0.58)
1/50	0.59 (0.53)	0.65 (0.52)	0.57 (0.53)	0.62 (0.54)	0.65 (0.52)
1/100	0.71 (0.54)	0.77 (0.53)	0.70 (0.54)	0.66 (0.53)	0.73 (0.54)
h	$\sigma_t = 5$				
	Intra	Inter	Mixed	iBAMG1	iBAMG2
1/25	0.82 (0.80)	0.86 (0.83)	0.83 (0.81)	0.82 (0.82)	0.83 (0.83)
1/50	0.91 (0.89)	0.91 (0.89)	0.90 (0.90)	0.91 (0.89)	0.91 (0.88)
1/100	0.92 (0.92)	0.92 (0.92)	0.93 (0.91)	0.92 (0.92)	0.92 (0.91)

Summary of numerical results Summarizing the numerical experiments, we see that the correlation-based measure can determine the appropriate CDOFs, even when there are anisotropies. However, the threshold parameter for defining strongly correlated connections may have to be tuned further on the coarser levels, as illustrated in the elasticity problems defined on beams. Moreover, it is observed that intravariation interpolation, whether through BAMG or iBAMG, is sufficient for most of the tested problems, and that iBAMG may not dramatically improve the efficiency (of course, additional testing on a larger selection of problems is needed to determine when iBAMG can lead to better convergence). Furthermore, locally mixing the intravariation and intervariability interpolation can lead to complex coarse-grid operators. This was unexpected and will require more analysis to explain. Finally, given that these schemes are nodal-based and the examined systems of PDEs are strongly coupled, these numerical results are likely better than results obtained with standard unknown-based approaches, although not as good as results for other nodal-based schemes that are designed for specific PDEs (e.g., Reference ⁷ for elasticity).

6 | CONCLUSION

This article is an initial attempt to develop an AMG method for solving general systems of elliptic PDEs. This attempt involves extensions of some of the more recent developments of AMG techniques for scalar PDEs. In this article, we examined a nodal coarsening scheme based on local correlation matrices evaluated for a set of smoothed test vectors. The Frobenius norms of these matrices provide a measure between the nodal DOFs, a measure that reflects the strength of intravariation and intervariability coupling in the DOFs. With a Ruge–Stuben coloring, the CDOFs are reasonably chosen even for anisotropic problems, and with this selection of CDOFs, the interpolation weights are generated using extensions of the BAMG and iBAMG schemes for systems of PDEs. Numerical experiments were performed to demonstrate the performance of the resulting AMG method for some difficult systems. Although the results are encouraging, further development of this method is needed. In particular, it is recognized that the algorithms in this article are computationally expensive, due to their adaptive and statistical nature. However, one of the goals of this article is only to demonstrate that the correlation matrices formed from the relaxed test vectors can expose dependencies in the nodal degrees of freedom, and hence can be used to select the coarse degrees of freedom. This correlation procedure can be applied to general systems of PDEs, particularly for systems where analytic forms of the near-nullspace components are not available. Moreover, an immediate practical tool that this procedure provides is a measure on whether intravariation interpolation is sufficient in a multigrid solver/preconditioner for a given system. Taking the correlation of the test vectors on the finest level can reveal whether the degree of freedoms correlate most strongly through the intravariation couplings. If they do, then an

efficient Ruge–Stuben AMG can be applied to the diagonal operators, and if only a preconditioner is desired, each diagonal operator can be coarsened separately with the intervariable coupling handled only on the finest level. Furthermore, for less general systems of PDEs, the correlation procedure can be applied to a small set of vectors (e.g., near-nullspace components and their coarsened forms) to determine the coarse degrees of freedom. The interpolation operators then can be constructed using existing efficient AMG procedure such as the system PDE technique of Reference ⁷. This, together with improvement in the scalability of the algorithms presented in this article, will be examined in the future.

ACKNOWLEDGEMENT

The author would like to thank two anonymous referees for their comments and suggestions that improved this article.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ORCID

Barry Lee  <https://orcid.org/0000-0003-1948-7384>

REFERENCES

1. Brezina M, Falgout RD, MacLachlan S, Manteuffel T, McCormick S, Ruge J. Adaptive algebraic multigrid. *SIAM J Sci Comput.* 2006;27:1261–1286.
2. Baker AH, Kolev TV, Yang UM. Improving algebraic multigrid interpolation operators for linear elasticity problems. *Numer Lin Alg Apps.* 2009;17:495–517.
3. Brezina M, Cleary AJ, Falgout RD, et al. Algebraic multigrid based on element interpolation (AMGe). *SIAM J Sci Comput.* 2000;22:1570–1592.
4. Chartier T, Falgout RD, Henson VE, et al. Spectral AMGe (ρ AMGe). *SIAM J Sci Comput.* 2003;25:1–26.
5. Clees T, Stuben K. Algebraic multigrid for industrial semiconductor device simulation. *Challenges in scientific computing- CISC, Berlin, Germany: Springer, 2002; p. 110–130.*
6. Fullenbach T, Stuben K. Algebraic multigrid for selected PDE systems. Paper presented at: Proceedings of the 4th European Conference on Elliptic and Parabolic Problems, Rolduc and Gaeta, 2001. World Scientific; 2002:399–410; London.
7. Griebel M, Oeltz D, Schweitzer MA. An algebraic multigrid method for linear elasticity. *SIAM J Sci Comput.* 2003;25:385–407.
8. Paludetto-Magri VA, Franceschini A, Janna C. A novel algebraic multigrid approach based on adaptive smoothing and prolongation for ill-conditioned systems. *SIAM J Sci Comput.* 2019;41:A190–A219.
9. Ruge JW, Stuben K. Algebraic multigrid (AMG). *Multigrid methods frontiers in applied mathematics.* Philadelphia, PA: S F McCormick, SIAM, 1987.
10. Trottenberg U, Oosterlee CW, Schuller A. *Multigrid.* London, Great Britain: Academic Press, 2001.
11. <https://computation.llnl.gov/casc/hypre/software.html>.
12. Livne OE, Brandt A. Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver. *SIAM J Sci Comput.* 2012;34:B499–B522.
13. Brandt A, Brannick J, Kahl K, Livshitz I. Bootstrap AMG. *SIAM J Sci Comput.* 2011;32:612–632.
14. Manteuffel T, McCormick S, Park M, Ruge J. Operator-based interpolation for bootstrap algebraic multigrid. *Numer Lin Alg Appl.* 2010;17:519–537.
15. Fouladi RT, Marani SK, Steiger JH. Moments of the fisher transformation: Applications using small samples. Paper presented at: Proceedings of the Joint Statistical Meeting; 2002:1032–1037.
16. Lee B. Parallel preconditioners and multigrid solvers for stochastic polynomial chaos discretizations of the diffusion equation at the large scale. *Numer Linear Alg Appl.* 2015;23:5–36. <https://doi.org/10.1002/nla.2000>.
17. Fisher RA. On the “probable error” of a coefficient of correlation deduced from a small sample. *Metro.* 1921;1:1–32.
18. Bulmer MG. *Principles of statistics.* New York, NY: Dover Publications, 1979.
19. Alcouffe RE, Brandt A, Dendy JE, Painter JW. The multi-grid method for the diffusion equation with strongly discontinuous coefficients. *SIAM J Sci Stat Comput.* 1981;2:430–454.
20. Falgout RD, Vassilevski PS. On generalizing the AMG framework. *SIAM J Numer Anal.* 2004;42:1669–1693.
21. Livne OE. Coarsening by compatible relaxation. *Numer Lin Alg Apps.* 2004;11:205–227.
22. Ron D, Safro I, Brandt A. Relaxation-based coarsening and multiscale graph organization. *Multiscale Model Simul.* 2011;9:407–423.
23. Yavneh I. A method for devising efficient multigrid smoothers for complicated PDE systems. *SIAM J Sci Comput.* 1993;14:1437–1463.
24. Dendy JE Jr. Black box multigrid for systems. *Appl Math Comput.* 1986;19:57–74.
25. Dendy JE Jr. Semicoarsening multigrid for systems. *Electron Trans Numer Anal.* 1997;6:97–105.
26. Xu J, Zikatanov L. Algebraic multigrid methods. *Acta Numerica.* 2017;26:591–721.
27. MacLachlan SP, Olson LN. Theoretical bounds for algebraic multigrid performance: Review and analysis. *Numer Lin Alg Apps.* 2014;21:194–220.

28. Brown PN, Lee B, Manteuffel TA. A moment-parity multigrid preconditioner for the first-order system least-squares formulation of the Boltzmann transport equation. *SIAM J Sci Comput.* 2003;25:513–533.
29. Chang B, Lee B. Multigrid algorithms for solving the multi-group, anisotropic scattering Boltzmann equation using first-order system least-squares methodology. *Electron Trans Numer Anal.* 2003;15:132–151.

How to cite this article: Lee B. Algebraic multigrid for systems of elliptic boundary-value problems. *Numer Linear Algebra Appl.* 2021;28:e2303. <https://doi.org/10.1002/nla.2303>