

Bringing physics into the coarse-grid selection: Approximate diffusion distance/effective resistance measures for network analysis and algebraic multigrid for graph Laplacians and systems of elliptic partial differential equations

Barry Lee 

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

Correspondence

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

Email: barryl@smu.edu

Funding information

National Science Foundation

Abstract

In a recent paper, the author examined a correlation affinity measure for selecting the coarse degrees of freedom (CDOFs) or coarse nodes (C nodes) in systems of elliptic partial differential equations (PDEs). This measure was applied to a set of relaxed vectors, which exposed the near-nullspace components of the PDE operator. Selecting the CDOFs using this affinity measure and constructing the interpolation operators using a least-squares procedure, an algebraic multigrid (AMG) method was developed. However, there are several noted issues with this AMG solver. First, to capture strong anisotropies, a large number of test vectors may be needed; and second, the solver's performance can be sensitive to the initial set of random test vectors. Both issues reflect the sensitive statistical nature of the measure. In this article, we derive several other statistical measures that ameliorate these issues and lead to better AMG performance. These measures are related to a Markov process, which the PDE itself may model. Specifically, the measures are based on the diffusion distance/effective resistance for such process, and hence, these measures incorporate physics into the CDOF selection. Moreover, because the diffusion distance/effective resistance can be used to analyze graph networks, these measures also provide a very economical scheme for analyzing large-scale networks. In this article, the derivations of these measures are given, and numerical experiments for analyzing networks and for AMG performance on weighted-graph Laplacians and systems of elliptic boundary-value problems are presented.

KEYWORDS

auto-correlation, bootstrap multigrid, correlation, diffusion distance, effective resistance, graph Laplacians, multigrid, systems of elliptic partial differential equations, variance

1 | INTRODUCTION

We are interested in developing multigrid methods for weighted-graph Laplacians arising from sparse network models, and for systems of elliptic partial differential equations (PDEs). Moreover, as a by-product of this development, we are additionally interested in applying components of the developed method to analyze large-scale graph networks. For discrete weighted-graph Laplacians, the systems have the form

$$\mathbf{L}\mathbf{u} = \mathbf{f}, \quad (1)$$

where

$$\mathbf{L}_{ij} = \begin{cases} -w_{ij} & i \neq j \\ \sum_{j \neq i} w_{ij} & i = j \end{cases} \quad (2)$$

for positive weights w_{ij} with $w_{ji} = w_{ij}$, and with the equations defined on a graph $G = \{V, E, w\}$ where V, E, w are respectively the vertices, edges, and weights of the graph. For systems of PDEs, the equations in continuum have the form

$$\mathcal{L}\mathbf{u} = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} & \cdots & \mathcal{L}_{1n} \\ \mathcal{L}_{21} & \mathcal{L}_{22} & \cdots & \mathcal{L}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{L}_{n1} & \cdots & \cdots & \mathcal{L}_{nn} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}, \quad (3)$$

which are defined on a smooth spatial domain $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, and with each $\mathcal{L}_{ij}$ being a partial differential operator of order at most 2 (higher-order operators can be reduced to lower-order operators by introducing auxiliary variables). Together with (3) are boundary conditions that give us a system boundary-value problem (BVP), which we will assume to be well-posed. We also assume that the component variables u_i in the discretization of the system are defined nodally in the sense that all the u_i 's exist on each node of the grid (i.e., discretized on a non-staggered grid). Moreover, we will consider only multigrid coarsenings that preserve this nodal structure on each level of the multilevel hierarchy. That is, we will consider only nodal-based or point-based coarsenings.¹⁻⁴

For both types of systems, a key procedure in developing a multigrid method is the selection of CDOFs/C nodes¹ particularly a selection that leads to coarse grids that align with anisotropies in systems of PDEs. There are also other procedures that must be carefully designed to handle issues that can prevent efficient multigrid processing, some of which will be examined in this article (e.g., construction of the interpolation operator).

For discrete weighted-graph Laplacian systems, one challenge is coarsening highly unstructured graphs with nodes that dramatically vary in their degrees (i.e., the number of edges emanating from each node highly varies). This challenge generally does not arise in discretizations of PDEs. For graph Laplacians, very successful multigrid methods that handle this challenge have been developed in, for example, References 5–10 (see also References 11–13 for adaptive algebraic multigrid (AMG) methods for Markov chains involving graph Laplacians). The developed schemes in this article provide an alternative AMG approach that fuses multigrid with some of the trending techniques for analyzing weighted-graph Laplacians in data analysis.¹⁴⁻¹⁶

For PDEs, especially systems of PDEs, issues that are more intrinsic to the PDE structure have to be addressed in the multigrid construction. These issues include strong inter-variable couplings (i.e., u_i is strongly coupled to u_j for $i \neq j$) that not only can hamper the relaxation process but also can lead to inaccurate coarse-grid operators; and multidimensional near-nullspaces that may require complex relaxation schemes and delicate construction of the interpolation operators. Moreover, the inter-variable couplings may contribute to the coarse degree of freedom (CDOF) selection, and inter-variable interpolation, where variables of different physical quantities interpolate to each other, may be required.³

This article extends the results of Reference 3, which developed a nodal-based AMG method for systems of PDEs. This method uses a relaxation approach for determining the coarse nodes,¹⁷⁻²² and a bootstrap AMG (BAMG) least-squares procedure for constructing the interpolation operators.¹⁷ In this relaxation approach, the homogeneous system $\mathcal{L}\mathbf{u} = \mathbf{0}$ is relaxed starting from a set of random vectors, the so-called test vectors. To select the coarse nodes, a correlation matrix measure is applied to the relaxed test vectors to determine the affinity between each pair of nodes. This affinity measure

generalizes the scalar correlation measure introduced in Reference 21 for graph Laplacian systems. For systems of PDEs, this matrix measure is more appropriate than agnostically applying the scalar correlation to the discrete system since the matrix measure acknowledges nodal features. Unfortunately, although this correlation matrix technique can produce coarse grids that capture the anisotropies in PDEs, there are two major issues with this method. First, due to the statistical properties of the correlation, a substantially larger number of test vectors is generally needed for systems of PDEs than is required for scalar PDEs, particularly in the presence of anisotropies. This issue can be slightly alleviated by applying a Z Fisher transformation to the correlation, as was done in Reference 3. Second, the performance of the AMG solver can be sensitive to the initial set of random test vectors. In this article, we develop several affinity measures that ameliorate these issues.

Before continuing our discussion on the new affinity measures, we briefly summarize some additional literature on closely related BAMG and relevant AMG coarsening schemes applied to some of the scalar PDE examples to be considered later. Of special interest are two recent papers by Kahl and his collaborators^{23,24} that specifically take a statistical viewpoint in the coarsening procedure, as taken in this article. These two papers illustrate some algorithmic advantages in using statistical quantities other than the correlation in the coarsening procedure. In Reference 24, a least angle regression (LARS) approach is used to determine the strength of connection between variables in a group rather than just a pair of variables, which is most often done. Determining the strength over groups of variables can both reduce the computational cost and also reveal complex connections that are not easily exposed by sequentially comparing pairs of variables at a time. The strength of connections are defined through the regression coefficients, and after repeatedly applying this LARS approach to generate improved candidate coarse variables, the coarse variables are determined and the corresponding regression coefficients are used in the interpolation. Alternatively, in Reference 23, a conditional probability distribution (Gaussian) formula drives the coarsening approach, with realizable approximations of this formula constructed using local sample mean and covariance estimators and local semi-variogram estimators. The selection of the coarse variables and the interpolation weights are guided by a best linear unbiased predictor principle, a standard procedure used in statistics. The potential of this method comes from the low computational complexity of this method in that only a few test vectors may be needed when a semi-variogram estimator is used. Since a goal of this article is to base the coarse-node selection in relation to accurately approximating the physical diffusion process, this article considers only strength of connections between pairs of nodes at a time and will not exploit surrogate semi-variogram estimators. These advanced techniques will be examined in the future.

Other literature on BAMG and relevant AMG coarsening schemes for scalar PDEs are References 25 and 26. In Reference 26 introduces a smoothed aggregation scheme that effects long-distance interpolation suitable for non-grid-aligned anisotropic diffusion. This long-distance interpolation is achieved by increasing the sparsity pattern of the interpolation in the direction of the strong connections, which in turn is obtained via powers of the sparse strength-of-connection matrix. Long-distance interpolation is also used in Reference 25. Here, strength of connection between each pair of variables is given by an algebraic distance define through a local least-square interpolation measure. This algebraic distance is used in a compatible relaxation and a least-squares procedure to select the coarse variables and to construct the interpolation operator. Long-distance interpolation is obtained by extending the neighborhood of each fine-grid variable for its interpolatory set, with this extended neighborhood via powers of the adjacency matrix. In this article, we also perform long-distance interpolation and affinity measures computed using extended neighborhoods of each node. We refer to these distances as pathlengths and the physical interpretation relates to the diffusivity over the pathlength distance.

The new measures of this article are based on the diffusion distance and a closely related quantity, the effective resistance. These two quantities reveal how information diffuses through the domain/network²⁷⁻³³ (see also monographs^{34,35} for other relevant network concepts related to the diffusion distance). They have, in fact, been used to analyze and “coarsen” graphs (e.g., to perform graph clustering³⁶). Ironically, AMG solvers have been used to compute estimates to these quantities for network analysis, which is opposed to our goal of using estimates of these quantities to develop an AMG solver. For example, in References 37 and 38, two similar strategies were developed for approximating the diffusion distance/effective resistance for all pairs of nodes in large-scale networks. Unfortunately, to approximate the distances/resistances between all pairs of nodes or just pairs of neighboring nodes, a large number of Laplacian systems have to be solved, albeit the same solver can be used for each system. In Reference 37, the LEAN AMG method of Reference 21 was used, and in Reference 38, a unsmoothed aggregation method^{39,40} with the edge coarsening method of Reference 10 was used. We note that these strategies for approximating the diffusion distance between all pairs of nodes explicitly rely on the positivity of the weights in the Laplacian. Hence, these strategies cannot be applied to problems that have a few negative weights, that is, near graph Laplacians that arise, for example, in power grid models.⁴¹

The affinity measures of this article provide rough approximations to the diffusion distance/effective resistance without actually solving Laplacian systems. The accuracy of these approximations is sufficient for relevant analysis of graph networks. These measures also can be computed for near graph Laplacians with a few negative weights, and if more accurate approximations of the diffusion distance/effective resistance are required, an AMG solver can be constructed utilizing these measures in an adaptive AMG fashion.¹⁹

But computing approximations to the diffusion distance/effective resistance for network analysis is not the central goal of this article. This goal is to develop measures for selecting the C nodes for AMG. Most strategies for selecting these coarse quantities are based on either a strength-of-connection/affinity given by the matrix entries,^{4,39,42} the rate of convergence of a basic relaxation scheme on subsets of nodes,²⁰ or the DOF dependency defined through a correlation or distance measure.^{3,21,22,41} All of these techniques do not provide an immediate physical interpretation. However, because the measures of this article will be based on the diffusion distance/effective resistance, a physical interpretation is now available. Specifically, the measures will reflect a node's ability to diffuse information, since the diffusion distance/effective resistance between any pair of nodes measures the global or "steady state" diffusion between these nodes. Significant is the fact that the diffusion distance/effective resistance can be used to extract a special set of nodes, the best spreader nodes,³³ which have small effective resistances to all nodes of the network (i.e., there is low resistance to the flow of information between the nodes of this set and the rest of the nodes in the network). If this set is sufficiently large, a good description of the global diffusion in the problem can be represented through this set. Hence, to capture the physics on the coarse grids, it is reasonable to expect the selected C nodes be good spreaders. Hence, the C node selection can be based on this spreader feature, and such a selection can be achieved through affinity measures that reveal this feature. However, rather than have these measures expose the global spread/diffusivity at each node, we will design them to expose the local diffusivity. The reasoning behind this is that diffusivity over just a few pathlengths from a node provides a good gauge on the node's ability to diffuse information over larger distances since the latter is obtained by successively propagating the local diffusion (in the discrete case, successively applying local random walks). In fact, since the global/steady-state random walks can forget the paths taken between pairs of nodes, a local process can actually provide finer details on the diffusion. Further, monitoring the local diffusion will lead to a more scalable algorithm and will align with the multigrid philosophy of probing grid-scale features. Indeed, to capture this local diffusion feature, a relaxation-based approach on a set of test vectors will be applied. Several sweeps of the relaxation scheme can unveil the local diffusivity, which then is made observable through these measures. Moreover, since this relaxation approach can be used to detect the "diffusive" nature in systems of PDEs, extensions of these measures can be developed to select the C nodes for systems of PDEs.

Once the C nodes are selected, the interpolation operator can be constructed. For network weighted-graph Laplacians, this can be affordably constructed using the entries of $\mathbf{L}$. For PDEs, a BAMG local least-squares procedure¹⁷ can be used. In particular, for systems of PDEs, since the coarsening is performed nodally, there is flexibility in choosing intra- or inter-variable interpolation. For intra-variable interpolation, the weights also can be computed using the entries of the diagonal $\mathcal{L}_{ii}$'s. However, the effectiveness of the solver using the resulting interpolation may not be as good as applying the least-squares approach, for example, when there are positive and negative off-diagonal entries in the $\mathcal{L}_{ii}$'s such as in stretched grid discretizations.

The article proceeds as follow. In Section 2, we review multigrid for scalar and systems of PDEs. This will include a review on the relaxation-based procedure for generating a set of relaxed vectors, the strength-of-connection and correlation-based measures for determining the C nodes, and the BAMG least-squares procedure for constructing the interpolation weights. In Section 3, we review the diffusion distance/effective resistance in data analysis and weighted-graph Laplacians, and the computational cost for approximating these *global* measures. In Section 4, using different expressions for the diffusion distance/effective resistance, we derive the new *local* affinity measures for weighted-graph Laplacians. We provide some numerical examples illustrating the effectiveness of these measures for analyzing graph networks, and illustrating the performance of an AMG that utilizes these measures for solving weighted-graph Laplacian systems. In Section 5, we develop extensions of these measures for scalar and systems of elliptic equations. Lastly, in Section 6, we give numerical results illustrating the performance of AMG using these new measures for solving scalar and systems of elliptic equations.

We close this section by mentioning that the intent of this article is to present several affinity measures based on heuristics, rather than developing multigrid theory for the resulting AMG derived from these measures, such as the accuracy of interpolation or two-grid convergence theory like in References 43 and 44. These measures are based on statistical quantities that can be used to approximate the diffusion distance/effective resistance, which have physical meanings. They are also an alternative to the correlation measure, which has a relation to the accuracy of the interpolation error for the special case when caliber 1 interpolation is used (i.e., each fine-grid points interpolates from only 1 coarse-grid point).²¹

For higher-order interpolation, these new measures can lead to different views on the relationship between the selection of coarse-grid nodes and the interpolation error (see References 23 and 24 for two other approaches for coarsening based on statistical quantities other than the correlation, which lead to algorithmic advantages as we have summarized earlier).

2 | MULTIGRID

It is well-known that multigrid is one of the most efficient methods to numerically solve scalar elliptic PDEs. Multigrid achieves its efficiency by using a hierarchy of grids, where the computation on the coarser grids costs only a fraction of the effort for computing on the original, finest 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 by the coarse-grid correction. The purpose of smoothing (e.g., a few sweeps of Gauss–Seidel, weighted Jacobi, or even a Krylov 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 system operator, that is, eigenvectors associated with the smaller eigenvalues. To be precise, as shown in the recent paper⁴³ and initiated earlier in papers,^{40,44} these errors correspond to the eigenvectors associated with the smaller eigenvalues of the generalized eigenvalue problem $\mathbf{L}\boldsymbol{\eta} = \lambda\tilde{\mathbf{M}}\boldsymbol{\eta}$, where $\tilde{\mathbf{M}}^{-1}$ is the applied (symmetrized) relaxation scheme. However, in this article, we will follow earlier literature, taking the near-nullspace to refer to eigenvectors associated with the smaller eigenvalues of the system operator. When polynomial relaxation schemes are used, there is no discrepancy between the two, and when other relaxation schemes are used, explicit knowledge of the system operator’s near-nullspace components can be used to construct efficient multigrid methods (e.g., see References 45 and 46 where multigrid methods are developed for the Helmholtz equation using the planewave near-nullspace components).

The complementary coarse-grid correction procedure must eliminate the near-nullspace, which are slowly attenuated by the smoother. To achieve this, the coarse-grid problem must be carefully constructed. This involves selecting suitable CDOFs and then constructing the interpolation operator. The selected coarse CDOFs must permit an accurate approximation of the near-nullspace components at all fine nodes. A measure that reflects the dependency of the DOFs, and hence used to guide the aggregation of DOFs that can be represented by a smaller number of DOFs (i.e., the CDOFs), must be designed. For scalar elliptic PDEs with discretizations leading to $\mathbf{M}$ matrices, this measure is given by the strength of connection between the DOFs, and is determined from the entries of $\mathbf{L}$. For example, in the classical Ruge-Stuben coarsening approach,⁴ the following strength-of-connection measure is used: DOF $\mathbf{u}_j$ is said to strongly influence DOF $\mathbf{u}_i$ if

$$|\mathbf{L}_{ij}| \geq \theta \max_{k \neq i} |\mathbf{L}_{ik}| \quad \text{for } 0 < \theta \leq 1. \quad (4)$$

The CDOFs 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 CDOFs) and form a maximal independent set. Consider a scalar diffusion operator $[-\nabla \cdot a(\mathbf{x})\nabla]$ discretized by standard finite-difference on a regular grid and with much higher diffusivity between nodes j and i than other neighboring nodes of i . Measure (4) would be able to detect this high diffusivity between j and i as a strong influence because the diffusivity is reflected through $a(\mathbf{x})$ and captured in $\mathbf{L}_{ij}$. Hence, physical reasoning can be associated with this strength of connection measure. We will formalize this reasoning using more physical measures in this article.

Once the selection of the CDOFs has been performed, the interpolation can be constructed using the entries of $\mathbf{L}$. This interpolation operator should be formed such that the near-nullspace is approximately in its range. For scalar elliptic PDEs with discretizations leading to $\mathbf{M}$ matrices, a very successful interpolation formula was developed by pioneer AMG researchers.⁴ Specifically, associating the nodes with the DOFs located on them (e.g., node i with $\mathbf{u}_i$), 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 $C \cap F = \emptyset$; let $N(i)$ be the neighborhood of node i (i.e., $N(i) = \{j \mid \mathbf{L}_{ij} \neq 0\}$); let C_i be a subset of CDOFs that interpolates to a fine-node i ; and let F_i^s and F_i^w respectively be the set of neighboring F nodes that strongly influence and weakly influence node i . Then the interpolation weight from

CDOF j to FDOF i is given by

$$w_{ij} = - \frac{\mathbf{L}_{ij} + \sum_{m \in F_i^s} \left(\frac{\mathbf{L}_{im} \mathbf{L}_{mj}}{\sum_{k \in C_i} \mathbf{L}_{mk}} \right)}{\mathbf{L}_{ii} + \sum_{r \in F_i^w} \mathbf{L}_{ir}}. \quad (5)$$

References 4,42, and 39. w_{ij} is the ij 'th entry of interpolation operator $\mathbf{P}$.

Although strength of connection (4) and interpolation (5) have been highly effective for scalar diffusion equations that are discretized into $\mathbf{M}$ matrices, their effectiveness can degrade for other discretizations and for other types of PDEs. During the past few decades, there has been active research in developing more general CDOF selection procedures and interpolation schemes to handle more challenging systems.^{17,19,20,22,40,44,47,48} Guided by the complementary smoothing/coarse-grid correction principle, these methods are usually based on the near-nullspace components of the system operator, and thus, having access to these components is essential. Good approximations to them can be determined by analyzing the PDE, disregarding the boundary conditions for simplicity. Moreover, whether analytic expressions of these components are available or not, refined or rough approximations can be obtained using the relaxation. Specifically, when analytic expressions are available, discrete representations of the near-nullspace components can be refined by applying several sweeps of relaxation on them; when analytic expressions are not available, rough approximations to these components can be obtained by applying relaxation to a set of random vectors. Since the relaxed vectors provide good profiles of the near-nullspace components, these vectors can expose candidates for the CDOFs, especially small sets of them that can accurately represent the full poorly attenuated vectors. This is the basis of the BAMG, adaptive multigrid schemes, and other relaxation-based schemes.^{17,19,20,22} For example, consider the BAMG approach. Let $\{\mathbf{v}^{(\alpha)}\}_{\alpha=1}^s$ denote the set of relaxed vectors, that is, the relaxed test vectors. The set of CDOFs can be determined from $\{\mathbf{v}^{(\alpha)}\}_{\alpha=1}^s$ by applying a strength of connection measure to these vectors. For example, in Reference 21, a correlation-like measure is used in determining the CDOFs for graph Laplacian systems. Specifically, with i, j being two arbitrary vector components of vector $\mathbf{v}^{(\alpha)}$ and the component inner product defined as

$$(\mathbf{v}_i, \mathbf{v}_j) = \sum_{\alpha=1}^s \mathbf{v}_i^{(\alpha)} \mathbf{v}_j^{(\alpha)},$$

the measure is

$$c_{ij} = \frac{|\langle \mathbf{v}_i, \mathbf{v}_j \rangle|^2}{(\mathbf{v}_i, \mathbf{v}_i)(\mathbf{v}_j, \mathbf{v}_j)}. \quad (6)$$

If the mean of the sample set is zero, then c_{ij} is indeed the correlation between DOFs i and j . With (6), i and j are said to have close affinity if the measure is greater than a given 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. The actual procedure for determining the CDOFs associates to each DOF a *projected volume* that predicts how large the aggregate will be if this DOF were selected as a seed: for DOF i , let π_i be its volume, which is initially set to 1. The projected volume v_i for i is defined as

$$v_i = \pi_i + \sum_{j \in N(i)} \pi_j \frac{c_{ij}}{\sum_{k \in N(j)} c_{jk}}. \quad (7)$$

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

$$\frac{\sum_{j \in (C \cap N(i))} c_{ij}}{\sum_{j \in N(i)} c_{ij}} \leq Q \quad \text{or} \quad \frac{\sum_{j \in (C \cap N(i))} \mathbf{L}_{ij}}{\sum_{j \in N(i)} \mathbf{L}_{ij}} \leq Q, \quad (8)$$

then i becomes a CDOF.

With the DOFs partitioned into C and F subsets, the interpolation operator can be formed. For weighted-graph Laplacians, the interpolation weights can be computed using

$$w_{ij} = \frac{\mathbf{L}_{ij}}{\sum_{k \in C_i} \mathbf{L}_{ik}}. \quad (9)$$

Alternatively, especially for PDEs with near-nullspace components having complex algebraic structures, a BAMG approach can be used to construct these weights by solving local least-squares problems: for $i \in F$, the local least-squares problem is

$$\{w_{ij} | j \in C_i\} = \arg \min_{w_{ij}} \sum_{\alpha=1}^s \frac{1}{\|\mathbf{v}^{(\alpha)}\|_{\mathbf{L}}^2} \left(\mathbf{v}_i^{(\alpha)} - \sum_{j \in C_i} w_{ij} \mathbf{v}_j^{(\alpha)} \right)^2 \quad (10)$$

for unknowns $\{w_{ij} | j \in C_i\}$.^{3,44,47}

The above techniques can be used to select the CDOFs and construct $\mathbf{P}$ for scalar PDEs or graph Laplacians. Complications arise for systems of PDEs though. Because of cross-variable couplings in systems of PDEs, intra-variable (i.e., like variable) and inter-variable (i.e., cross variable) procedures must be considered. In Reference 3, correlation measures that expose the intra- and inter-variable dependencies in the DOFs were introduced. The discretizations and coarsenings are assumed to be nodal based. With l and m indexing the nodes and q and r the variables, the correlation for quantities $\mathbf{a}_l^q$ and $\mathbf{a}_m^r$

$$\text{corr}(\mathbf{a}_l^q, \mathbf{a}_m^r) = \frac{\sum_{\alpha=1}^s (\mathbf{a}_l^{q,(\alpha)} - \bar{\mathbf{a}}_l^q) (\mathbf{a}_m^{r,(\alpha)} - \bar{\mathbf{a}}_m^r)}{\sqrt{\left(\sum_{\alpha=1}^s (\mathbf{a}_l^{q,(\alpha)} - \bar{\mathbf{a}}_l^q)^2 \right) \left(\sum_{\alpha=1}^s (\mathbf{a}_m^{r,(\alpha)} - \bar{\mathbf{a}}_m^r)^2 \right)}},$$

where $\bar{\mathbf{a}}_l^q$ is the mean of $\{\mathbf{a}_l^{q,(\alpha)}\}_{\alpha=1}^s$. A correlation-like matrix then can be defined at each pair of nodal points. For example, for a two-variable system with test vector $\begin{pmatrix} \mathbf{v}_I^1 \\ \mathbf{v}_I^2 \end{pmatrix}^{(\alpha)}$, this matrix for nodes I, J is

$$\begin{aligned} & \text{corr} \left(\begin{pmatrix} \mathbf{v}_I^1 \\ \mathbf{v}_I^2 \end{pmatrix}, \begin{pmatrix} \mathbf{v}_J^1 \\ \mathbf{v}_J^2 \end{pmatrix} \right) \\ & := \begin{pmatrix} \text{corr}(\mathbf{v}_I^1, \mathbf{v}_I^1) & \text{corr}(\mathbf{v}_I^1, \mathbf{v}_J^1) & \text{corr}(\mathbf{v}_I^1, \mathbf{v}_I^2) & \text{corr}(\mathbf{v}_I^1, \mathbf{v}_J^2) \\ \text{corr}(\mathbf{v}_J^1, \mathbf{v}_I^1) & \text{corr}(\mathbf{v}_J^1, \mathbf{v}_J^1) & \text{corr}(\mathbf{v}_J^1, \mathbf{v}_I^2) & \text{corr}(\mathbf{v}_J^1, \mathbf{v}_J^2) \\ \text{corr}(\mathbf{v}_I^2, \mathbf{v}_I^1) & \text{corr}(\mathbf{v}_I^2, \mathbf{v}_J^1) & \text{corr}(\mathbf{v}_I^2, \mathbf{v}_I^2) & \text{corr}(\mathbf{v}_I^2, \mathbf{v}_J^2) \\ \text{corr}(\mathbf{v}_J^2, \mathbf{v}_I^1) & \text{corr}(\mathbf{v}_J^2, \mathbf{v}_J^1) & \text{corr}(\mathbf{v}_J^2, \mathbf{v}_I^2) & \text{corr}(\mathbf{v}_J^2, \mathbf{v}_J^2) \end{pmatrix}, \end{aligned} \quad (11)$$

and the measure is the Frobenius norm of this matrix, that is,

$$C_{IJ} = \left\| \text{corr} \left(\begin{pmatrix} \mathbf{v}_I^1 \\ \mathbf{v}_I^2 \end{pmatrix}, \begin{pmatrix} \mathbf{v}_J^1 \\ \mathbf{v}_J^2 \end{pmatrix} \right) \right\|_F. \quad (12)$$

Matrix (11) contains inter-variable correlations. To involve only the intra-variable correlations, the inter-variable correlations located in the off-diagonals blocks simply need to be zeroed off. Finally, rather than using a projected volume approach for selecting the coarse nodes, the C_{IJ} 's are used as a strength-of-connection measure in a coloring algorithm: letting $\bar{C}_I$ denote the Frobenius norm of an average correlation matrix for node I , node J is considered to strongly influence I if

$$C_{IJ} \geq \theta \bar{C}_I, \quad (13)$$

where θ is a threshold parameter. A standard Ruge-Stuben coloring scheme that uses the number of nodes each node strongly influences can be applied to determine the nodal CDOFs: Let F, C, U, G, S_I^t respectively denote the fine nodes, C nodes, unmarked nodes, total nodes, 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.

Extensions to least-squares interpolation (10) for systems of PDEs were also presented in Reference 3. For inter-variable interpolation, the weighted least-squares problems have the same form but now with test vectors defined for all the variable types, $\mathbf{v} = (\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^n)^t$. For intra-variable interpolation, the variable l block of $\mathbf{P}$ are computed with a least-squares problem involving only the variable l subvectors of the test vectors. Specifically, let $\mathbf{v}^{(\alpha, l)}$ be the variable l subvector of $\mathbf{v}^{(\alpha)}$ and $\mathcal{L}_{ll}$ denote the (variable l)-to-(variable l) submatrix of $\mathcal{L}$. Then the least-squares problem is

$$\{w_{IJ}^l | J \in C_I\} = \arg \min_{w_{IJ}^l} \sum_{\alpha=1}^s \frac{1}{\|\mathbf{v}^{(\alpha, l)}\|_{\mathcal{L}_{ll}}^2} \left(\mathbf{v}_I^{(\alpha, l)} - \sum_{J \in C_I} w_{IJ}^l \mathbf{v}_J^{(\alpha, l)} \right)^2. \quad (14)$$

The weights obtained using the above interpolation methods depend on the selection of C nodes, which in turn is conditioned on the affinity measure and the relaxed test vectors. Hence, practical issues that must be considered are the sensitivity of the measure to the initial set of random test vectors, and, closely related, the number of test vectors required to stabilize this sensitivity. The relevancy of these issues is that high sensitivity can lead to poor robustness in the constructed AMG solver. That is, high sensitivity can result in fluctuations in the solver performance for simulations that conduct separate coarsening procedures, that is, perform separate AMG setup phases. For the above correlation measures, such sensitivity is a reflection of slow convergence of the sample correlation. Unfortunately, since the number of samples required to obtain good accuracy to the correlation can vary widely from problem to problem, these measures can suffer from this sensitivity issue. Hence, we now consider measures based on the diffusion distance/effective resistance which are not only less sensitive to the initial set of random vectors but also provide a physical basis to the CDOF selection.

3 | DIFFUSION DISTANCE/EFFECTIVE RESISTANCE

The diffusion distance was introduced as a tool for extracting low-dimensional structures in data sets.^{14-16,28} The idea is to view the data as evolving under a heat flow on a manifold $\mathcal{M}$ and eventually settling into a low-dimensional organization of the data. To measure the “similarities” between data points at any step of the evolution, the diffusion distance between a pair of points is defined to be the probability of transitioning from one point to the other, transversing all available paths between the pair at that evolutionary step. To be precise, for any $\tau \geq 0$, let $k_\tau(\cdot, \cdot)$ be a kernel defined on $\mathcal{M} \times \mathcal{M}$ which provides a prior notion of the affinity/similarity between the data points at the start of the propagation, $\tau = 0$. There are several requirements on k_τ :

- k_τ is symmetric and non-negative on $\mathcal{M} \times \mathcal{M}$ for any $\tau \geq 0$,
- k_τ satisfies the semi-group property

$$\int_{\mathcal{M}} k_{\tau_1}(x, u) k_{\tau_2}(u, y) du = k_{\tau_1 + \tau_2}(x, y) \quad \forall x, y \in \mathcal{M}, \quad \forall \tau_1, \tau_2 \geq 0,$$

- k_τ satisfies

$$\int_{\mathcal{M}} k_\tau(x, y) dx = 1 \quad \forall y \in \mathcal{M}, \quad \forall \tau \geq 0.$$

Also, let

$$T_\tau g(x) := \int_{\mathcal{M}} k_\tau(x, y) g(y) dy.$$

In the finite-dimensional scenario, $\mathcal{M} = \mathfrak{R}^m$ and k_0 is $\mathbf{A}^t$ for an $(m \times m)$ matrix $\mathbf{A}$ with non-negative entries a_{ij} . $\mathbf{A}^t$ is a Markov process with its column sum equal to 1, that is, the third property on k_τ . Also, the matrix-vector product $\mathbf{A}^t \mathbf{g}$ is the analogue of $T_\tau g(x)$.

Returning to the continuum scenario, the diffusion distance $d_\tau(x, y)$ at time τ between any pair x, y in $\mathcal{M}$ is given by

$$\begin{aligned} [d_\tau(x, y)]^2 &= (T_\tau(\delta_x - \delta_y), T_\tau(\delta_x - \delta_y)) \\ &= (T_\tau \delta_x, T_\tau \delta_x) + (T_\tau \delta_y, T_\tau \delta_y) - 2(T_\tau \delta_x, T_\tau \delta_y) \\ &= \|k_\tau(\cdot, x)\|_2^2 + \|k_\tau(\cdot, y)\|_2^2 - 2(k_\tau(\cdot, x), k_\tau(\cdot, y)), \end{aligned} \quad (15)$$

where δ_z is the $L^1(\mathcal{M})$ delta function satisfying

$$T_\tau \delta_z(w) = \int_{\mathcal{M}} k_\tau(w, u) \delta_z(u) du = k_\tau(w, z),$$

and $(\cdot, \cdot)$ and $\|\cdot\|_2$ are respectively the $L^2(\mathcal{M})$ inner product and norm. Interpreting (15), $(\delta_x - \delta_y)$ corresponds to taking a unit source at x and unit sink at y , $T_\tau(\delta_x - \delta_y)$ corresponds to the heat flow at time τ given the source $(\delta_x - \delta_y)$, and applying the $L^2(\mathcal{M})$ inner product accounts for transversing all possible paths between x and y in the flow at time τ .

In the discrete setting, which is more relevant to this article, each step in the propagation corresponds to a multiplication with $\mathbf{A}^t$. With the data points indexed as i and j , the initial step is

$$\begin{aligned} [d_0(i, j)]^2 &= (\mathbf{A}^t(\mathbf{e}_i - \mathbf{e}_j), \mathbf{A}^t(\mathbf{e}_i - \mathbf{e}_j)) \\ &= (\mathbf{A}^t \mathbf{e}_i, \mathbf{A}^t \mathbf{e}_i) + (\mathbf{A}^t \mathbf{e}_j, \mathbf{A}^t \mathbf{e}_j) - 2(\mathbf{A}^t \mathbf{e}_i, \mathbf{A}^t \mathbf{e}_j) \\ &= [\mathbf{A} \mathbf{A}^t]_{ii} + [\mathbf{A} \mathbf{A}^t]_{jj} - 2[\mathbf{A} \mathbf{A}^t]_{ij}, \end{aligned} \quad (16)$$

where $\mathbf{e}_i$ is the i 'th canonical basis vector. At step $\tau = n - 1$, we have

$$[d_{n-1}(i, j)]^2 = \left[\mathbf{A}^n (\mathbf{A}^t)^n \right]_{ii} + \left[\mathbf{A}^n (\mathbf{A}^t)^n \right]_{jj} - 2 \left[\mathbf{A}^n (\mathbf{A}^t)^n \right]_{ij}. \quad (17)$$

Now a question that arises is whether a steady-state case is obtained when $n \rightarrow \infty$. In Reference 38, it is shown that if $\mathbf{L}$ is a normalized graph Laplacian with $\mathbf{A}$ defined by

$$\mathbf{A} = \mathbf{I} - \mathbf{L},$$

then the sequence converges if $\mathbf{A}$ is row stochastic and irreducible. The limit is the effective resistance. Specifically, denoting the effective resistance between i and j as $r_{1,ij}$, we have

$$\begin{aligned} r_{1,ij}^2 &= (\mathbf{e}_i - \mathbf{e}_j)^t \mathbf{L}^\dagger (\mathbf{e}_i - \mathbf{e}_j) \\ &= \mathbf{L}_{ii}^\dagger + \mathbf{L}_{jj}^\dagger - 2\mathbf{L}_{ij}^\dagger, \end{aligned} \quad (18)$$

where $\mathbf{L}^\dagger$ is the pseudo-inverse of $\mathbf{L}$. (In existing literature, $\mathbf{L}$ is often an unnormalized graph Laplacian.) Given this relationship between the diffusion distance and effective resistance, the latter describes the global or steady-state diffusion

of the problem, with a smaller effective resistance implying better diffusion between the points. A smaller resistance means that there are many paths between the nodes so that if some of the edges are disconnected, the nodes remain connected through other paths.

We also have the total effective resistance of the network defined by

$$r_G := \sum_{i=1}^m \sum_{j>i}^m r_{1,ij}, \quad (19)$$

and the best spreader defined as nodes i satisfying

$$\min_i \sum_{j=1}^m r_{1,ij}. \quad (20)$$

In the analysis of networks (e.g., power grids), resistance r_G is a good indicator on how robust the network is when edges and nodes are deleted, that is, does the network system remain operational as nodes and edges are removed? Also, the best spreaders are nodes where there is low resistance to the diffusion to other nodes.

Note that the diffusion distance is defined as the $L^2(\mathcal{M})$ norm of the action of T_τ on $L^1(\mathcal{M})$ -normalized functions. This inconsistency in norms can lead to scaling issues. In Reference 28, an example is given to illustrate the scaling issue in the diffusion distance for kernel $k_\tau(x, y) = \frac{e^{-|x-y|^2/4\tau}}{(4\pi\tau)^{-n/2}}$ (the fundamental solution for the heat equation in $\mathfrak{R}^n$). Thus, they suggest $L^2(\mathcal{M})$ -normalized sources and sinks of the form

$$\psi_z = \frac{\delta_z}{||T_\tau \delta_z||_2}$$

in the continuum setting, and the $l^2(\mathfrak{R}^m)$ -normalized sources and sinks of the form

$$\hat{\mathbf{e}}_i = \frac{\mathbf{e}_i}{||(\mathbf{A}^t)^n \mathbf{e}_i||}$$

in the discrete setting. With these changes the new distances are

$$[\hat{d}_\tau(x, y)]^2 = 2 - 2 \frac{(k_\tau(\cdot, x), k_\tau(\cdot, y))}{||k_\tau(\cdot, x)||_2 ||k_\tau(\cdot, y)||_2}, \quad (21)$$

$$[\hat{d}_{n-1}(i, j)]^2 = 2 - 2 \frac{[\mathbf{A}^n (\mathbf{A}^t)^n]_{ij}}{\sqrt{[\mathbf{A}^n (\mathbf{A}^t)^n]_{ii} [\mathbf{A}^n (\mathbf{A}^t)^n]_{jj}}}. \quad (22)$$

Noting that $\mathbf{L}_{ii}^\dagger$ is non-negative for weighted graph Laplacians,³³ we also introduce scaled sources and sinks of the form $\hat{\mathbf{e}}_i = \frac{\mathbf{e}_i}{\sqrt{\mathbf{L}_{ii}^\dagger}}$ to get the effective resistance

$$r_{2,ij}^2 = 2 - 2 \frac{\mathbf{L}_{ij}^\dagger}{\sqrt{\mathbf{L}_{ii}^\dagger} \sqrt{\mathbf{L}_{jj}^\dagger}}, \quad (23)$$

when $\mathbf{L}_{ii}^\dagger$ and $\mathbf{L}_{jj}^\dagger$ are positive.

The diffusion distance and effective resistance are physics-based measures describing the affinity between nodes in a network. Also, the best spreaders are significant to the overall diffusion process in the network. The best spreader nodes should be included in a coarsened graph so that an accurate description of the diffusion process in the full network is retained. Thus, it is reasonable to require these nodes be retained on coarser levels in a multilevel or coarse-graining method. However, determining these spreaders entails computing the effective resistance, and this computation can be more expensive than performing this coarsening task. Consider the cost of computing the resistance. First, if only the effective resistance $r_{1,ij}$ is needed, then the Laplacian system

$$\mathbf{L}\mathbf{r}_1 = \mathbf{e}_i - \mathbf{e}_j$$

must be solved. An appropriate AMG solver can be constructed to optimally achieve this, but then it is not cost effective in determining the coarse nodes using the resistances, that is, in an adaptive AMG procedure.¹⁹ In fact, to determine the best spreaders, the resistances between all pairs of nodes are needed. One of the most efficient schemes to approximate these uses the Johnson-Lindenstrauss lemma.⁴⁹ To describe this approach, note that because the weights w_{ij} are assumed positive, the Laplacian can be decomposed as

$$\mathbf{L} = \mathbf{B}^t \mathbf{W} \mathbf{B} = \mathbf{B}^t \mathbf{W}^{1/2} \mathbf{W}^{1/2} \mathbf{B},$$

where $\mathbf{W} \in \mathfrak{R}^{|E| \times |E|}$ is the diagonal matrix containing the weights of the Laplacian, and $\mathbf{B} \in \mathfrak{R}^{|E| \times |V|}$ is the signed edge-vertex incidence matrix

$$\mathbf{B}_{ev} = \begin{cases} 1 & v \text{ is the head of } e \\ -1 & v \text{ is the tail of } e \\ 0 & \text{otherwise.} \end{cases}$$

Clearly,

$$r_{1,ij}^2 = \|\mathbf{W}^{1/2} \mathbf{B} \mathbf{L}^\dagger (\mathbf{e}_i - \mathbf{e}_j)\|_2^2. \quad (24)$$

Now, the Johnson-Lindenstrauss lemma states that for a random matrix $\mathbf{Q} \in \mathfrak{R}^{\kappa \times |E|}$

$$\mathbf{Q}_{ij} = \begin{cases} \frac{1}{\sqrt{\kappa}} & \text{probability } 1/2 \\ -\frac{1}{\sqrt{\kappa}} & \text{probability } 1/2, \end{cases}$$

with $\kappa \geq \frac{24 \log |V|}{\epsilon^2}$, we have

$$(1 - \epsilon) r_{1,ij}^2 \leq \|\mathbf{Q} \mathbf{W}^{1/2} \mathbf{B} \mathbf{L}^\dagger (\mathbf{e}_i - \mathbf{e}_j)\|_2^2 \leq (1 + \epsilon) r_{1,ij}^2$$

with probability of at least $(1 - 1/|V|)$.^{29,38} Thus, an approximation to the effective resistance matrix $\left[r_{1,ij}^2 \right]_{i,j=1}^{|V|}$ can be constructed by

1. form $\mathbf{Y} = \mathbf{Q} \mathbf{W}^{1/2} \mathbf{B} \in \mathfrak{R}^{\kappa \times |V|}$, which can be done in $O(\kappa |E|) = O(|E|/\epsilon^2)$ operations since $\mathbf{B}$ has only 2 entries per row;
2. for $i = 1, \dots, \kappa$, with $\mathbf{y}_i$ denoting the transpose of the i 'th row of $\mathbf{Y}$, construct the i 'th row of matrix $\mathbf{Z} \in \mathfrak{R}^{\kappa \times |V|}$ by solving

$$\mathbf{L} \mathbf{z}_i = \mathbf{y}_i,$$

3. compute the approximation of $r_{1,ij}^2$ with $\|\mathbf{Z}_{\cdot,i} - \mathbf{Z}_{\cdot,j}\|_2^2$.

The major cost in this algorithm is solving the κ Laplacian systems in step 2. Although the same solver can be used for each system, this algorithm is impractical for determining the coarse nodes, and the cost can be rather excessive for other applications where the effective resistance is relevant. One such application is the “stability” analysis of networks, where nodes that lead to major disconnections in the graph are sought, that is, removing these nodes will “destabilize” the network. A reasonable assumption is that these nodes are best spreaders-physically, if the best spreaders are removed, then the diffusivity of the problem is degraded, which can be taken as an indicator that a destabilization of the network has occurred. To determine good candidates for the best spreaders, a more efficient method for approximating the effective resistances is desired.

The effective resistance may also provide an “imprecise” description of the diffusion in the network. It was shown in References 30 and 31 that for some random graphs, with nodes i and j sufficiently far apart,

$$r_{1,ij}^2 \approx \left(\frac{1}{\deg(i)} + \frac{1}{\deg(j)} \right), \quad (25)$$

where $\deg(l)$ is the degree of node l . This shows that the topology of the network does not participate in $r_{1,ij}^2$ since the intermediate paths in the random walk do not contribute in (25). Intuitively, this can be seen in a large network, where the intermediate paths are transversed in the Markov process. The steady-state diffusion between nodes i and j then depends only on the immediate edges/paths emanating from i and j , that is, on $\deg(i)$ and $\deg(j)$. On the other hand, when the nodes are close, more details of the diffusion process are retained, and hence, $r_{1,ij}^2$ can more accurately reflect the diffusion between these nodes.

4 | NEW AFFINITY MEASURES

As brought out earlier, the diffusivity over a few pathlengths from a node provides a good gauge on the diffusion around that node. This, with the fact that the effective resistance can be reliable for nodes that are close to each other, prods us to an approach based on local diffusion. Consider a diffusion process in a local patch centered at node i , and picture the diffusion from i to its neighboring nodes in the patch. Physically, diffusion means that there will be strong dependency between the values of $\mathbf{u}$ at these neighboring nodes and its value at i . Hence, a measure of this dependency can be used to estimate the strength of the local diffusion. One way to measure this is to take the correlation of these values over an ensemble of diffusion states. In particular, the relaxation approach of References 22 and 21 can be used: relaxing using pointwise Gauss–Seidel, or even more complex relaxation schemes, on an initial set of random test vectors simulates the diffusion process and taking the local correlation of these relaxed vectors gives an estimate of the diffusivity in this patch. However, as we have seen, slow convergence of the sample correlation can lead to sensitivity issues. Thus, instead of the correlation, we develop measures based on the effective resistance, using effective resistance expressions (18) and (23).

To this end, consider the spectral decomposition of $\mathbf{L}^\dagger$:

$$\mathbf{L}^\dagger = \sum_{p=1} \frac{1}{\lambda_p} \boldsymbol{\eta}_p \boldsymbol{\eta}_p^t, \quad (26)$$

where the λ_p 's are the positive eigenvalues in ascending order and the $\boldsymbol{\eta}_p$'s are their corresponding eigenvectors. A crude estimate to $\mathbf{L}^\dagger$ is the first term of the sum, and a rough approximation to $\boldsymbol{\eta}_1$ is any one of the relaxed vectors $\{\mathbf{v}^{(\alpha)}\}_{\alpha=1}^s$. Hence,

$$\lambda_1 r_{1,ij}^2 \propto \left(\mathbf{v}_i^{(\alpha)} \mathbf{v}_i^{(\alpha)} + \mathbf{v}_j^{(\alpha)} \mathbf{v}_j^{(\alpha)} - 2 \mathbf{v}_i^{(\alpha)} \mathbf{v}_j^{(\alpha)} \right) \quad (27)$$

for $\alpha \in \{1, 2, \dots, s\}$. Since λ_1 is independent of i, j , we can take the righthand side of (27) to reflect the strength of the diffusion between nodes i and j . Moreover, since a set of test vectors is available, it would be beneficial to utilize the whole set, which can lead to a more stable procedure. Thus, assume that the mean, $\bar{\mathbf{v}}$, of the relaxed vectors is $\mathbf{0}$. This is a reasonable assumption since the initial set of random test vectors has mean $\mathbf{0}$, these vectors were obtained by relaxing on the homogeneous equation, and the relaxation operator is linear. (In the actually computation, the sample mean of the relaxed vectors can be used instead.) We have

$$\begin{aligned} [\tilde{r}_{1,ij}]^2 &:= \frac{1}{s} \sum_{\alpha=1}^s \left(\mathbf{v}_i^{(\alpha)} \mathbf{v}_i^{(\alpha)} + \mathbf{v}_j^{(\alpha)} \mathbf{v}_j^{(\alpha)} - 2 \mathbf{v}_i^{(\alpha)} \mathbf{v}_j^{(\alpha)} \right) \\ &= \frac{1}{s} \sum_{\alpha=1}^s (\mathbf{v}_i^{(\alpha)} - \mathbf{0})^2 + \frac{1}{s} \sum_{\alpha=1}^s (\mathbf{v}_j^{(\alpha)} - \mathbf{0})^2 - \frac{2}{s} \sum_{\alpha=1}^s (\mathbf{v}_i^{(\alpha)} - \mathbf{0})(\mathbf{v}_j^{(\alpha)} - \mathbf{0}) \\ &= \frac{1}{s} \sum_{\alpha=1}^s (\mathbf{v}_i^{(\alpha)} - \bar{\mathbf{v}}_i)^2 + \frac{1}{s} \sum_{\alpha=1}^s (\mathbf{v}_j^{(\alpha)} - \bar{\mathbf{v}}_j)^2 - \frac{2}{s} \sum_{\alpha=1}^s (\mathbf{v}_i^{(\alpha)} - \bar{\mathbf{v}}_i)(\mathbf{v}_j^{(\alpha)} - \bar{\mathbf{v}}_j) \\ &= \text{var}(\mathbf{v}_i) + \text{var}(\mathbf{v}_j) - 2\text{cov}(\mathbf{v}_i, \mathbf{v}_j), \end{aligned} \quad (28)$$

where $\text{var}(\cdot)$ and $\text{cov}(\cdot, \cdot)$ are the sample variance and sample covariance of the data, that is, for data sets $\{b_1^{(\alpha)}\}_{\alpha=1}^s$ and $\{b_2^{(\alpha)}\}_{\alpha=1}^s$,

$$\text{var}(b_1) = \frac{1}{s} \sum_{\alpha=1}^s (b_1^{(\alpha)} - \bar{b}_1)^2 \quad \text{cov}(b_1, b_2) = \frac{1}{s} \sum_{\alpha=1}^s (b_1^{(\alpha)} - \bar{b}_1)(b_2^{(\alpha)} - \bar{b}_2).$$

Finally, using the identity

$$\text{var}(b_1 - b_2) = \text{var}(b_1) + \text{var}(b_2) - 2\text{cov}(b_1, b_2),$$

we have

$$[\tilde{r}_{1,ij}]^2 = \text{var}(\mathbf{v}_i - \mathbf{v}_j), \quad (29)$$

and hence, we define the measure

$$c_{1,ij} = \frac{1}{\sqrt{\text{var}(\mathbf{v}_i - \mathbf{v}_j)}} \quad \text{and} \quad c_{1,ii} = 0. \quad (30)$$

Since the variance generally requires less samples to converge than is required in the correlation, the sensitivity issues will be ameliorated for this new measure.

Another measure can be similarly developed using effective resistance expression (23). This gives

$$\begin{aligned} [\tilde{r}_{2,ij}]^2 &= 2 - 2 \sqrt{\left[\frac{\text{cov}(\mathbf{v}_i, \mathbf{v}_j)}{\sqrt{\text{var}(\mathbf{v}_i)} \sqrt{\text{var}(\mathbf{v}_j)}} \right]^2} \\ &= 2 - 2|\text{corr}(\mathbf{v}_i, \mathbf{v}_j)|, \end{aligned} \quad (31)$$

and the measure is

$$c_{2,ij} = \begin{cases} \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_i, \mathbf{v}_j)|}} & |\text{corr}(\mathbf{v}_i, \mathbf{v}_j)| \neq 1 \\ \text{large value} & |\text{corr}(\mathbf{v}_i, \mathbf{v}_j)| = 1 \end{cases} \quad \text{and} \quad c_{2,ii} = 0, \quad (32)$$

which is well-defined because $|\text{corr}(\mathbf{v}_i, \mathbf{v}_j)| \leq 1$. For our computation, when $|\text{corr}(\mathbf{v}_i, \mathbf{v}_j)| = 1$, we fix the large value to 10^5 . It can also be set to be several orders of magnitude larger than the average of $\{c_{2,ij}\}_{|\text{corr}(\mathbf{v}_i, \mathbf{v}_j)| \neq 1}$, for j 's in a neighborhood of i . Note that although this measure involves the correlation, numerical experiments indicate that the nonlinear transformation of the correlation in (32) can ameliorate the sensitivity issues.

Lastly, to further reduce the number of test vectors, a measure can be designed using a sample auto-correlation of the relaxed vectors. Recall that given a times series sequence of multivariate data $\{\mathbf{b}^{(\tau)}\}_{\tau=1}^m$, the sample covariance for offset h is defined by

$$\mathbf{\Gamma}(h) = \frac{1}{m} \sum_{\tau=1}^{m-h} (\mathbf{b}^{(\tau+h)} - \bar{\mathbf{b}})(\mathbf{b}^{(\tau)} - \bar{\mathbf{b}})^t, \quad (33)$$

where $\bar{\mathbf{b}}$ is the time average $\bar{\mathbf{b}} = \frac{1}{m} \sum_{\tau=1}^m \mathbf{b}^{(\tau)}$ of the sequence. $\mathbf{\Gamma}(h)$ is a square matrix with its size equal to the number of components in $\mathbf{b}$, and its ij 'th entry is given by

$$\gamma_{ij}(h) = \text{cov}(\mathbf{b}_i^{(\tau+h)}, \mathbf{b}_j^{(\tau)}).$$

The sample auto-correlation for $\{\mathbf{b}^{(\tau)}\}_{\tau=1}^m$ is the matrix with ij 'th entry

$$\text{autocorr}\left(\mathbf{b}_i^{(\tau+h)}, \mathbf{b}_j^{(\tau)}\right) := \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)}\sqrt{\gamma_{jj}(0)}}. \quad (34)$$

Note that because of the offset, this matrix is not symmetric.

A sample auto-correlation can be applied to any one of the test vectors, with the time series being the sequence of relaxation iterates starting after a “spin-up” of the initial test vector is performed, that is, after a few relaxation sweeps are performed to attenuate the randomness in the initial vector. Thus, taking h to be a small positive number, this auto-correlation gives an average of how the nodal values of $\mathbf{v}^{(\alpha)}$ auto-correlate through the relaxation sweeps, the simulated diffusion process. Using effective resistance expression (23), we consider

$$c_{3,ij}^{(\alpha)} = \frac{1}{\left|2 - 2 \text{autocorr}\left(\mathbf{v}_i^{(\alpha, \tau+h)}, \mathbf{v}_j^{(\alpha, \tau)}\right)\right|}, \quad (35)$$

and introduce the measure

$$c_{3,ij} = \sqrt{\frac{1}{s} \sum_{\alpha=1}^s c_{3,ij}^{(\alpha)}}. \quad (36)$$

The anticipation is that fewer test vectors will be needed since the node dependency is gleaned more thoroughly in each relaxed vector.

Now with these measures, the local diffusivity within a neighborhood of each node can be approximated. Specifically, for each $i \in V$, measure (30), (32), or (36) is computed between i and any node within a given pathlength distance to it. The computed measures then are used in a strength-of-connection procedure to determine the directions of strong diffusivity for each $i \in V$. To select the C nodes, a coloring scheme can be used: nodes that have strong diffusivity (i.e., strong influence) to the most number of nodes within the neighborhood are selected as C nodes. Alternatively, the total diffusivity at each node can be used: with each node summing up its measures to all nodes within a pathlength-distance neighborhood, nodes having larger sums are selected as C nodes. For measures (30) and (32), because they relate to the effective resistance, this approach corresponds to choosing good spreaders for the C nodes. This strategy is particularly appropriate for graphs with highly varying node degrees (see Sections 4.1 and 4.2).

4.1 | Network analysis experiments

Measures (30), (32), and (36) provide efficient means for estimating the diffusion process in a network. Thus, we examine their ability to detect the nodes that are most relevant to the vulnerability/resiliency of a network, vulnerability in the sense that the system operation (e.g., flow/distribution of electric power) define on the network degrades when these nodes are removed, that is, the so-called percolation problem in network analysis.³⁴ Given the physical significance of the best spreaders, the best spreaders are the obvious choice. To assess the relevancy of a set of nodes to the network, we consider the number of components the network is broken up into when the set is removed, the size of the largest or so-called giant component of the severed network,² and most importantly, the total effective resistance of the giant component (a large total effective resistance indicates that the surviving major sub-network is itself vulnerable to instabilities, which increases the relevancy of the set of removed nodes because the decomposed system is now closely unoperable). For comparison, these quantities are also computed using the exact effective resistance and the Johnson-Lindenstrauss approximation. For these two schemes, the best spreaders are determined using (20) with a post-processing step: (20) is used first to select l_2 nodes that give the smallest sums, and then from these nodes, l_1 nodes ($l_2 > l_1$) with the most number of connections satisfying $r_{1,ij} < \theta \bar{r}_{1,i}$ are selected as the candidate spreaders. Here, $\bar{r}_{1,i}$ is the average of node i 's effective resistance with all other nodes, and the post-processing step ensures that nodes with high and “balanced” diffusivity are selected. (Note that a coloring scheme by itself should not be applied to V to select the spreaders because nodes that poorly fail (20) can still strongly influence many nodes.)

TABLE 1 Network analysis for the BA and WS network models using the exact effective resistance and Johnson-Lindenstrauss (JL) approximation with $\kappa = 200$ samples.

# Spreaders	Scenario	Method	# Comps	Size of common Lg Comp	R_G Lg Comp
5	BA	$L^\dagger$	84	424/159	225,706/32,306
		JL	145	159/332	32,306/170,697
	WS	$L^\dagger$	1	995	156,176
		JL	1	995	155,920
10	BA	$L^\dagger$	129	339/159	138,480/32,306
		JL	176	159/332	32,306/170,697
	WS	$L^\dagger$	1	990	157,005
		JL	1	990	156,631

Note: Candidate spreader nodes are removed from the network, and the resiliency/vulnerability of the network is assessed by the number of components the network is broken up into, the size of the most commonly selected largest component(s), and the total effective resistance of the largest component(s).

For measures (30), (32), and (36), we use an analogue of (20) to select the nodes:

$$\max_i \sum_j c_{l,ij}, \quad (37)$$

with the summation taken over nodes j within a given pathlength distance from i . A similar post-processing procedure will be used to select the nodes: given a set of l_2 nodes that generate the largest sums $\sum_j c_{l,ij}$, choose the l_1 nodes of this set that have the most number of connections satisfying $c_{l,ij} > \theta \bar{c}_{l,i}$, where $\bar{c}_{l,i}$ is the average of $c_{l,ij}$ in i 's pathlength neighborhood.

We examine networks coming from realistic DC power grid models. Although power grids sometimes have been discussed as small-world networks,⁵⁰ there remains controversy on whether this is indeed the case.⁵¹ Thus, before examining the realistic DC models, we consider the Barabasi-Albert (BA) and Watts-Strogatz (WS) models,^{32,34,35} which respectively represent scale-free (power-law) and small-world networks. We consider models consisting of 1000 nodes, with the WS network having 4000 edges and an average node degree of 8. We also fix the networks for all the simulations so that the same random graphs are used in all the methods. Table 1 provides results for the BA and WS networks using the exact effective resistance and the Johnson-Lindenstrauss (JL) approximation ($\kappa = 200$ samples) to select 5 and 10 spreaders that are removed from the network. Given the statistical nature of the JL approximation, 10 simulations for this method are performed and the average number of components the original network is broken up into and the size(s) of the most commonly selected giant component are tabulated. The second largest giant component for the BA model generated by the exact effective resistance approach is also recorded. From Table 1, we see that the JL approximation is able to accurately mimic the results obtained by the exact effective resistance for the WS model, but was only able to accurately capture the second largest component for the BA model.

Table 2 illustrates the results for the BA and WS models using measures (30) and (32), and measure (36). Given the statistical aspect of the test vector construction, 10 simulations are again performed. But whereas the statistical nature of the Johnson-Lindenstrauss method affects the approximation to the pseudo-inverse, and indirectly the selection of nodes, the statistical nature in the test vector procedure leads to an ensemble of snapshots of the diffusion process via the relaxation. Each snapshot leads to a different set of relevant nodes, and hence, giving a more comprehensive assessment of the nodes through the different observations of the diffusion process.

For measures (30) and (32), 20 test vectors with 5 relaxation sweeps are used. For (36) only 3 test vectors with 15 Gauss-Seidel relaxation sweeps are used, with a few of these sweeps used to spin up the sequence of iterates to attenuate strong randomness in the initial test vectors. The number of relaxation sweeps is kept low in order to expose the diffusivity over a large range of frequencies (i.e., since an AMG solver is not being constructed, the algebraically smooth frequencies are not pertinent in these experiments). Finally, for all three measures, the affinity is taken over pathlength distances of 5 and 3 for the BA and WS models, respectively. We see that all three measures were able to mimic the results of the JL and exact effective resistance performance for the WS model. For the BA model, measures (30) and (32) were able to accurately mimic the results of the JLT method, that is, accurately obtaining the second largest giant component and roughly producing the same number of components as the JLT method. Measure (36), moreover, was able to produce giant

TABLE 2 Network analysis using measures (32), (30), and (36) with pathlength distances 5 and 3 for the BA and WS graph models, respectively, to determine the spreaders.

# Spreaders	Scenario	Measure	# Comps	Size of common Lg Comp	R_G Lg Comp
5	BA (pathln 5)	(32)	146	159	32,306
		(30)	148	159	32,306
		(36)	86.8	332/401/159	170,697/213,227/32,306
	WS (pathln 3)	(32)	1	995	155,888
		(30)	1	995	155,310
		(36)	1	995	155,815
10	BA	(32)	154	159	32,306
		(30)	161	159	32,306
		(36)	93.9	332/159	170,697/32,306
	WS	(32)	1	990	156,456
		(30)	1	990	155,549
		(36)	1	990	156,423

Note: The resiliency/vulnerability of the network is assessed by the number of components the network is broken up into, the size of the most commonly selected largest component(s), and the total effective resistance of largest component(s). 5 smoothings and 20 test vectors are used for measures (32) and (30), and 15 smoothings with offset $h = 1$ and 3 test vectors for measure (36).

components of sizes and R_G 's more consistent to the ones produced by the exact effective resistance. Overall, from these scale-free and small-world examples, measures (30) and (32), and measure (36) provide a promising efficient technique to detect nodes that are most relevant to the vulnerability/resiliency of a network without having to solve graph Laplacian systems nor having to form $\mathbf{L}^\dagger$.

We turn to the realistic DC power grid models, with the experiments conducted as in the BA/WS models. Table 3 gives the results for the exact effective resistance and JL approximation. For the JL approximation, we now consider the average size of the largest giant components and their R_G values over the 10 simulations. Perusing over this table, we see again that more components are formed as more candidate spreaders are removed. Moreover, the total effective resistance of the giant component is larger than the full network itself, indicating these nodes are highly relevant to the network's vulnerability/resiliency. Overall, both of these methods are able to detect relevant nodes, but at the expense of forming/approximating the pseudo-inverse of the graph Laplacians.

Tables 4 and 5 illustrate the results using measures (30) and (32), and measure (36). For measures (30) and (32) we again use 20 test vectors with 5 relaxation sweeps per vector; for (36) we now take 1 or 3 test vectors with 10 Gauss–Seidel relaxation sweeps per vector, with a few of these sweeps used to spin up the sequence of iterates to attenuate strong randomness in the initial test vectors. A pathlength distance of 5 is used for all scenarios and measures (a distance of 8 is also used on the Texan power grid for measure (36) to illustrate that improvement can be obtained by varying the pathlength since we would expect a longer pathlength to more accurately reveal the global diffusivity). With the exception of the Texan grid, all three measures decompose the original networks into a fair number of components that is somewhat reflective of the exact effective resistance and JL approximation, although measure (32) often generates less components. The sizes of the giant components produced by the measures are overall comparable to the sizes of the giant components produced by the exact effective resistance and Johnson-Lindenstrauss approximation. However, the total effective resistances for the surviving giant sub-networks produced by the three measures are almost always larger than their counterparts produced by the exact effective resistance and Johnson-Lindenstrauss approximation. Moreover, we see that measure (30) is effective using only 1 test vector.

These results demonstrate that the local diffusivity, based on *a posteriori* observations via the relaxation and quantified with these measures, can select nodes that are highly relevant to the vulnerability/resiliency of the networks. Moreover, since the local diffusivity approach produces giant components with larger total effective resistances than the ones produced by a direct effective resistance approach, the vulnerability/resiliency of a network can be assessed more accurately from local diffusion interactions than from an analytic, steady-state description given by the effective resistance.

TABLE 3 Network analysis using the exact effective resistance and Johnson-Lindenstrauss (JL) approximation with $\kappa = 200$ samples.

# Spreaders	Scenario	Method	# Comps	Size Lg Comp	R_G Lg Comp	R_G full graph
5	European (1354 nodes)	$\mathbf{L}^\dagger$	3	1323	50,472	48,230
		JL	3.2	1328	50,821	48,092
	French (1888 nodes)	$\mathbf{L}^\dagger$	5	1877	203,966	198,461
		JL	5.6	1871	202,022	198,533
	Polish (3374 nodes)	$\mathbf{L}^\dagger$	1	3369	858,408	829,741
		JL	1.1	3369	859,652	826,801
	Texan (2000 nodes)	$\mathbf{L}^\dagger$	12	1984	271,524	248,725
		JL	11.7	1984	252,835	248,145
10	European	$\mathbf{L}^\dagger$	4	1291	60,565	48,230
		JL	5.2	1289	54,522	47,962
	French	$\mathbf{L}^\dagger$	5	1874	212,973	198,461
		JL	7.1	1850	206,235	197,717
	Polish	$\mathbf{L}^\dagger$	1	3364	892,862	829,741
		JL	2.3	3361	888,143	827,759
	Texan	$\mathbf{L}^\dagger$	29	1962	271,646	248,725
		JL	21.9	1969	264,009	248,741

Note: Candidate spreader nodes are removed from the network, and the resiliency/vulnerability of the network is assessed by the number of components the network is broken up into, the size of the largest component, and the total effective resistance of the largest component.

4.2 | Weighted-graph Laplacian experiments

Recall that a set of nodes that preserves an accurate description of the global diffusion of the problem is a reasonable choice for the coarse nodes. Hence, we select the coarse nodes to be good spreaders, which is measured by the effective resistance or an approximation to it. Specifically, measures (30), (32), and (36) will be used to determine coarse nodes with low effective resistance, that is, we consider using these measures to construct an AMG solver for weighted-graph Laplacians. We assume that the nodes of the graph can have large variations in their degrees (i.e., large variations in the number of non-zero entries in the rows of the graph Laplacian). Because some nodes can have high diffusivity but low degrees, a coloring scheme will not be used to select the C nodes—a coloring scheme will be partial to high degree nodes since they have a better chance to strongly influence more nodes. Instead, nodes with higher local diffusivity will be selected as C nodes. To describe the procedure, let

$$c_{l,i} = \sum_j c_{l,ij} \quad l = 1, 2, 3$$

denote the local diffusivity of node i using affinity measure c_l . Here, the summation is performed over j 's within a given pathlength distance from i . Then node i is taken to be a C node if

$$c_{l,i} > \theta \bar{c}_l,$$

where θ is a given threshold and $\bar{c}_l$ is the average of the global set $\{c_{l,i}\}_{i=1}^{|V|}$. This is the first phase in constructing C. The second phase moves any nodes in F to C that have sufficiently stronger affinity or matrix connections to other F nodes than to its neighboring C nodes. This criterion is described by inequalities (8).

Since an AMG solver is now being constructed, the above c_l measures are applied to relaxed vectors that sufficiently approximate the near-nullspace. Thus, in comparison to the network analysis experiments, more relaxation sweeps must be applied to the initial set of random test vectors, which means that the above procedure considers the local diffusivity of

TABLE 4 Network analysis using measures (32) and (30) with pathlength distance 5 to determine the spreaders.

# Spreaders	Scenario	Measure	# Comps	Size Lg Comp	R_G Lg Comp
5	European	(32)	1.0	1349	51,633
		(30)	4.0	1312	53,719
	French	(32)	1.6	1882	216,415
		(30)	8.0	1871	211,741
	Polish	(32)	1.0	3369	854,431
		(30)	1.3	3368	859,709
	Texan	(32)	2.0	1994	251,176
		(30)	8.1	1988	248,226
10	European	(32)	2.0	1322	55,759
		(30)	4.6	1295	54,939
	French	(32)	7.7	1867	226,291
		(30)	9.4	1864	223,512
	Polish	(32)	4.9	3357	878,935
		(30)	1.6	3363	909,734
	Texan	(32)	5.4	1986	254,571
		(30)	9.7	1981	250,212

Note: The resiliency/vulnerability of the network is assessed by the number of components the network is broken up into, the size of the largest component, and the total effective resistance of largest component. 5 smoothings and 20 test vectors are used.

TABLE 5 Network analysis using measure (36) with pathlength distances 5 and 8 (only Texan) to determine the spreader nodes.

# Spreaders	Scenario	Pathlength	# Comps		Size Lg Comp		R_G Lg Comp	
			1 TV	3 TVs	1 TV	3 TVs	1 TV	3 TVs
5	European	5	2.1	1.8	1347	1348	52,196	52,715
	French	5	2.5	3.2	1881	1879	214,008	210,630
	Polish	5	1.0	1.0	3369	3369	857,342	855,361
	Texan	5	2.2	1.0	1994	1995	250,679	249,523
		8	4.6	3.1	1991	1993	263,715	264,453
10	European	5	6.9	6.7	1308	1323	54,202	55,226
	French	5	9.4	9.4	1863	1863	223,392	223,676
	Polish	5	1.9	1.5	3363	3364	901,238	898,585
	Texan	5	5.7	3.5	1985	1988	253,113	253,806
		8	13.7	17.0	1977	1974	270,450	268,230

Note: The resiliency/vulnerability of the network is assessed by the number of components the network is broken up into, the size of the largest component, and the total effective resistance of largest component. 10 smoothings with offset $h = 1$, and 1 test vector and 3 test vectors are used.

TABLE 6 Convergence rates for 30 AMG $V(1, 1)$ cycles applied to weighted-graph Laplacians arising from the BA and WS graph models and DC power grid models.

Scenario	Measure	Convergence rate	
		Pathlength 1	Pathlength 3
BA (4000 nodes)	(32)	0.48	0.17
	(30)	0.55	0.18
	(36)	0.52	0.29
WS (4000 nodes)	(32)	0.08	0.07
	(30)	0.08	0.08
	(36)	0.09	0.09
European	(32)	0.58	0.46
	(30)	0.53	0.41
	(36)	0.41	0.44
French	(32)	0.54	0.52
	(30)	0.62	0.47
	(36)	0.46	0.38
Polish	(32)	0.59	0.54
	(30)	0.42	0.42
	(36)	0.55	0.54
Texan	(32)	0.34	0.33
	(30)	0.33	0.33
	(36)	0.37	0.38

Note: Measures (32) and (30) use 20 test vectors relaxed 20 sweeps. Measure (36) uses 1 test vector relaxed 20 sweeps, with 10 of these sweeps used to spin-up the iterates for the auto-correlation. Caliber 2 interpolation is used in the BA and WS models, and caliber 3 in the DC models.

just the algebraically smooth components. In the following experiments, 20 Gauss–Seidel sweeps are applied to each test vector, and 20 test vectors are used for measures (30) and (32), and 1 for the auto-correlation measure (36). For measure (36), an offset $h = 1$ and a spin-up of 10 Gauss–Seidel sweeps are used.

As for the interpolation and coarse-grid operator, $\mathbf{P}$ is constructed according to (9) and the coarse-grid operator according to the Galerkin triple matrix product $\mathbf{P}'\mathbf{L}\mathbf{P}$. Further, $\mathbf{P}$ is a β caliber interpolation (i.e., each fine-grid value interpolates from at most β coarse-grid values) but with the β C nodes restricted to the immediate neighborhood of the fine-grid node. This prevents the coarse-grid operator from becoming too dense.

The experiments involve the weighted-graph Laplacians from the BA and WS networks with 4000 nodes and the DC power grid models. To assess the solver's efficiency, the convergence rate is taken over thirty $V(1, 1)$ cycles applied to a homogeneous system with a random initial guess. The rate is defined as $\frac{\|\mathbf{u}^{30}\|}{\|\mathbf{u}^{29}\|}$ and averaged over 10 simulation runs. Table 6 tabulates results using pathlengths 1 and 3 in the affinity measures, and calibers 2 and 3 interpolation for the BS/WS models and DC models, respectively. We see some dependency on the pathlength distance, but overall the rates are decent. In particular, measure (36)'s performance is impressive since it required only 1 test vector, with its relaxation iterates carefully gleaned through their auto-correlation.

5 | MEASURES FOR AMG COARSENING OF SYSTEMS OF PDES

Although the effective resistance is defined for weighted-graph Laplacians, an extension to invertible scalar diffusion operators readily follows by replacing the pseudo-inverse with the inverse in (18). In fact, the discretized operators for these problems are weighted-graph Laplacians with most of the nodes having roughly the same degree and each node

having a self loop. Because the nodes have roughly the same degree, a coloring scheme can be used to select the C nodes, that is, the coloring is now more impartial to the graph structure. Utilizing (30), (32), or (36) as the affinity measure in the coloring, the local diffusivity is used to determine the C nodes.

Unfortunately, extensions to systems of elliptic PDEs is more difficult. With respect to the effective resistance, rather than using the individual entries of the inverse, we consider

$$\mathbf{r}_{1,I}^2 := [\mathbf{L}^{-1}]_{II} + [\mathbf{L}^{-1}]_{JJ} - 2[\mathbf{L}^{-1}]_{IJ},$$

which consists of submatrices of $\mathbf{L}^{-1}$ involving the rows and columns corresponding to all variables at nodes I and J . The matrix $\mathbf{r}_{1,I}^2$ integrates the nodal structure of the discretization (i.e., intra-variable and inter-variable connections between I and J), and is obtained by an appropriate ordering of the unknowns and taking $\mathbf{e}_I, \mathbf{e}_J$ in (18) to be the submatrices of the identity consisting of the columns corresponding to all variables at nodes I, J . Needless to say, $\mathbf{r}_{1,I}^2$ is actually not needed since only the local affinity measures will be used. Like in the correlation measure of Reference 3, these measures will involve local matrices that incorporate the cross-variable coupling. Assuming an n -variable system, an analogue of variance measure (30) is

$$C_{1,IJ} = \begin{cases} \left\| \begin{pmatrix} \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^1 - \mathbf{v}_J^1)}} & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^1 - \mathbf{v}_J^2)}} & \cdots & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^1 - \mathbf{v}_J^n)}} \\ \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^2 - \mathbf{v}_J^1)}} & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^2 - \mathbf{v}_J^2)}} & \cdots & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^2 - \mathbf{v}_J^n)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^n - \mathbf{v}_J^1)}} & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^n - \mathbf{v}_J^2)}} & \cdots & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^n - \mathbf{v}_J^n)}} \end{pmatrix} \right\|_F & I \neq J \\ \left\| \begin{pmatrix} 0 & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^1 - \mathbf{v}_I^2)}} & \cdots & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^1 - \mathbf{v}_I^n)}} \\ \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^2 - \mathbf{v}_I^1)}} & 0 & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^2 - \mathbf{v}_I^3)}} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^n - \mathbf{v}_I^1)}} & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^n - \mathbf{v}_I^2)}} & \cdots & 0 \end{pmatrix} \right\|_F & I = J, \end{cases} \quad (38)$$

where the Frobenius norm is applied to incorporate contributions from all variable couplings between nodes I and J , and from cross-variable couplings at node I to itself. With I^l and J^m respectively indexing variable l at node I and variable m at node J , when both $I \neq J$ and $l \neq m$, the entries of (38) correspond to formula (28) evaluated at $I^l J^m$, that is, $\tilde{r}_{I^l J^m}^2$. This formula is evaluated separately at each $I^l J^m$. Moreover, note that unlike in measure (12), where the cross-variable couplings within nodes I and J contribute to the measure between nodes I and J when $I \neq J$ (i.e., C_{IJ} for $I \neq J$ includes contributions from $\text{corr}(\mathbf{v}_I^l, \mathbf{v}_I^m)$ and $\text{corr}(\mathbf{v}_J^l, \mathbf{v}_J^m)$), measure (38) *correctly* includes only couplings across I and J when $I \neq J$.

An analogue to measure (32) is

$$C_{2,IJ} = \begin{cases} \left\| \begin{pmatrix} \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^1, \mathbf{v}_J^1)|}} & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^1, \mathbf{v}_J^2)|}} & \cdots & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^1, \mathbf{v}_J^n)|}} \\ \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^2, \mathbf{v}_J^1)|}} & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^2, \mathbf{v}_J^2)|}} & \cdots & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^2, \mathbf{v}_J^n)|}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^n, \mathbf{v}_J^1)|}} & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^n, \mathbf{v}_J^2)|}} & \cdots & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^n, \mathbf{v}_J^n)|}} \end{pmatrix} \right\|_F & I \neq J \\ \left\| \begin{pmatrix} 0 & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^1, \mathbf{v}_I^2)|}} & \cdots & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^1, \mathbf{v}_I^n)|}} \\ \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^2, \mathbf{v}_I^1)|}} & 0 & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^2, \mathbf{v}_I^3)|}} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^n, \mathbf{v}_I^1)|}} & \frac{1}{\sqrt{2-2|\text{corr}(\mathbf{v}_I^n, \mathbf{v}_I^2)|}} & \cdots & 0 \end{pmatrix} \right\|_F & I = J. \end{cases} \quad (39)$$

For (36), with $\mathbf{v}_I^{(\alpha, l, \tau+h)}$ denoting the variable l component of the α 'th test vector at relaxation iterate $\tau + h$, we use

$$[\mathbf{r}_{3,IJ}^{(\alpha)}]^2 = \begin{cases} \left(\begin{array}{ccc} \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,1,\tau+h)}, \mathbf{v}_J^{(\alpha,1,\tau)})|} & \cdots & \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,1,\tau+h)}, \mathbf{v}_J^{(\alpha,n,\tau)})|} \\ \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,2,\tau+h)}, \mathbf{v}_J^{(\alpha,1,\tau)})|} & \cdots & \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,2,\tau+h)}, \mathbf{v}_J^{(\alpha,n,\tau)})|} \\ \vdots & & \vdots \\ \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,n,\tau+h)}, \mathbf{v}_J^{(\alpha,1,\tau)})|} & \cdots & \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,n,\tau+h)}, \mathbf{v}_J^{(\alpha,n,\tau)})|} \end{array} \right)_F & I \neq J \\ \left(\begin{array}{ccc} \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,1,\tau+h)}, \mathbf{v}_I^{(\alpha,1,\tau)})|} & \cdots & \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,1,\tau+h)}, \mathbf{v}_I^{(\alpha,n,\tau)})|} \\ \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,2,\tau+h)}, \mathbf{v}_I^{(\alpha,1,\tau)})|} & \cdots & \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,2,\tau+h)}, \mathbf{v}_I^{(\alpha,n,\tau)})|} \\ \vdots & & \vdots \\ \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,n,\tau+h)}, \mathbf{v}_I^{(\alpha,1,\tau)})|} & \cdots & \frac{1}{|2-2\text{autocorr}(\mathbf{v}_I^{(\alpha,n,\tau+h)}, \mathbf{v}_I^{(\alpha,n,\tau)})|} \end{array} \right)_F & I = J, \end{cases}$$

and take

$$C_{3,IJ} = \sqrt{\frac{1}{s} \sum_{\alpha=1}^s [\mathbf{r}_{3,IJ}^{(\alpha)}]^2}. \quad (40)$$

The above measures were derived by applying the measures of Section 4 at each $I^l J^m$ coupling separately. We can also derive a measure that takes all the couplings between nodes I and J together. To this end, let $\mathbf{v}_I^{(\alpha)}$ be the subvector of test vector $\mathbf{v}^{(\alpha)}$ at node I . Then as in the derivation of (28), we have

$$\begin{aligned} & \frac{1}{s} \sum_{\alpha=1}^s \left[\mathbf{v}_I^{(\alpha)} \left(\mathbf{v}_I^{(\alpha)} \right)^t + \mathbf{v}_J^{(\alpha)} \left(\mathbf{v}_J^{(\alpha)} \right)^t - \mathbf{v}_I^{(\alpha)} \left(\mathbf{v}_J^{(\alpha)} \right)^t - \mathbf{v}_J^{(\alpha)} \left(\mathbf{v}_I^{(\alpha)} \right)^t \right] \\ &= \frac{1}{s} \sum_{\alpha=1}^s \left[\left(\mathbf{v}_I^{(\alpha)} - \mathbf{0} \right) \left(\mathbf{v}_I^{(\alpha)} - \mathbf{0} \right)^t + \left(\mathbf{v}_J^{(\alpha)} - \mathbf{0} \right) \left(\mathbf{v}_J^{(\alpha)} - \mathbf{0} \right)^t - \left(\mathbf{v}_I^{(\alpha)} - \mathbf{0} \right) \left(\mathbf{v}_J^{(\alpha)} - \mathbf{0} \right)^t - \left(\mathbf{v}_J^{(\alpha)} - \mathbf{0} \right) \left(\mathbf{v}_I^{(\alpha)} - \mathbf{0} \right)^t \right] \\ &= \text{cov}(\mathbf{v}_I, \mathbf{v}_I) + \text{cov}(\mathbf{v}_J, \mathbf{v}_J) - \text{cov}(\mathbf{v}_I, \mathbf{v}_J) - \text{cov}(\mathbf{v}_J, \mathbf{v}_I) \\ &:= \mathbf{C}, \end{aligned}$$

where each $\text{cov}(\cdot, \cdot)$ is an $(n \times n)$ multivariate covariance matrix. For example,

$$\text{cov}(\mathbf{v}_I, \mathbf{v}_I) = [\text{cov}(\mathbf{v}_I^l, \mathbf{v}_I^m)]_{l,m=1}^n,$$

where entries $\text{cov}(\mathbf{v}_I^l, \mathbf{v}_I^m)$ are scalar covariances. Expanding out each $(n \times n)$ multivariate covariance matrix, the lm 'th entry of the sum is

$$\text{cov}(\mathbf{v}_I^l, \mathbf{v}_I^m) + \text{cov}(\mathbf{v}_J^l, \mathbf{v}_J^m) - \text{cov}(\mathbf{v}_I^l, \mathbf{v}_J^m) - \text{cov}(\mathbf{v}_J^l, \mathbf{v}_I^m) = \text{cov}(\mathbf{v}_I^l - \mathbf{v}_J^l, \mathbf{v}_I^m - \mathbf{v}_J^m).$$

Hence,

$$\mathbf{C} = \begin{pmatrix} \text{var}(\mathbf{v}_I^1 - \mathbf{v}_J^1) & \text{cov}(\mathbf{v}_I^1 - \mathbf{v}_J^1, \mathbf{v}_I^2 - \mathbf{v}_J^2) & \cdots & \text{cov}(\mathbf{v}_I^1 - \mathbf{v}_J^1, \mathbf{v}_I^n - \mathbf{v}_J^n) \\ \text{cov}(\mathbf{v}_I^2 - \mathbf{v}_J^2, \mathbf{v}_I^1 - \mathbf{v}_J^1) & \text{var}(\mathbf{v}_I^2 - \mathbf{v}_J^2) & \cdots & \text{cov}(\mathbf{v}_I^2 - \mathbf{v}_J^2, \mathbf{v}_I^n - \mathbf{v}_J^n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\mathbf{v}_I^n - \mathbf{v}_J^n, \mathbf{v}_I^1 - \mathbf{v}_J^1) & \text{cov}(\mathbf{v}_I^n - \mathbf{v}_J^n, \mathbf{v}_I^2 - \mathbf{v}_J^2) & \cdots & \text{var}(\mathbf{v}_I^n - \mathbf{v}_J^n) \end{pmatrix}. \quad (41)$$

From (28), the diagonal entries of (41) approximate the effective resistances of the like-variables between nodes I and J . The off-diagonal entries expose the cross-variable coupling, although their connection to the effective resistance is not obvious. Lastly, instead of taking (41) as the measure, we define the measure

$$C_{4,IJ} = \left\| \begin{pmatrix} \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^1 - \mathbf{v}_J^1)}} & \frac{1}{\sqrt{1 - |\text{corr}(\mathbf{v}_I^1 - \mathbf{v}_J^1, \mathbf{v}_I^2 - \mathbf{v}_J^2)|}} & \cdots & \frac{1}{\sqrt{1 - |\text{corr}(\mathbf{v}_I^1 - \mathbf{v}_J^1, \mathbf{v}_I^n - \mathbf{v}_J^n)|}} \\ \frac{1}{\sqrt{1 - |\text{corr}(\mathbf{v}_I^2 - \mathbf{v}_J^2, \mathbf{v}_I^1 - \mathbf{v}_J^1)|}} & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^2 - \mathbf{v}_J^2)}} & \cdots & \frac{1}{\sqrt{1 - |\text{corr}(\mathbf{v}_I^2 - \mathbf{v}_J^2, \mathbf{v}_I^n - \mathbf{v}_J^n)|}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{1 - |\text{corr}(\mathbf{v}_I^n - \mathbf{v}_J^n, \mathbf{v}_I^1 - \mathbf{v}_J^1)|}} & \frac{1}{\sqrt{1 - |\text{corr}(\mathbf{v}_I^n - \mathbf{v}_J^n, \mathbf{v}_I^2 - \mathbf{v}_J^2)|}} & \cdots & \frac{1}{\sqrt{\text{var}(\mathbf{v}_I^n - \mathbf{v}_J^n)}} \end{pmatrix} \right\|_F, \quad (42)$$

so that a large $C_{4,IJ}$ means strong affinity between I and J .

Measure $C_{4,IJ}$ can be applied directly. However, because of possible disparity in the magnitudes of the variance and correlation entries, a two-stage measure can be utilized instead. In the first stage, nodes that have strong affinity to I are determined by comparing the Frobenius norm of the diagonal of the matrix in (42). In the second stage, nodes are possibly added to this set by comparing the Frobenius norm of the off-diagonal of this matrix. For systems of PDEs with weak cross-variable couplings, the first stage itself can be sufficient.

With the C nodes selected using a coloring scheme with any of the above measures, the interpolation weights can be computed with the intra-variable least-squares scheme (14) or an analogous inter-variable scheme. The weights also can be constructed with the indirect BAMG (iBAMG) technique of Reference 48 extended to systems of PDEs. Details on this extension can be found in Reference 3.

6 | NUMERICAL EXPERIMENTS

Since an AMG solver is being constructed, a sufficient number of relaxation sweeps must be performed on the initial set of random test vectors to expose the near-nullspace components. With these relaxed vectors, measures (30), (32), and (36) are applied to scalar PDEs, and measures (38), (39), and (42) to systems of PDEs. All the PDEs examined are defined on a unit square with homogenous Dirichlet boundary conditions and discretized with bilinear finite elements on uniform tessellations. Since the prototype software was implemented in Matlab, only spatial 2-d examples are considered. As for the relaxation, pointwise or point-based Gauss–Seidel is employed on the scalar and systems of PDEs. For systems of PDEs, a point-based, rather than a nodal or blockwise relaxation, is chosen in order to more appropriately simulate the diffusion process because the local solves in the nodal/blockwise schemes would describe a strong physical requirement on the diffusion process. Needless to say, these more complex schemes can still be used to simulate the diffusion. In the point-based Gauss–Seidel scheme, each DOF located at a node is updated separately before moving onto the DOFs on the next node. Although multiple updates can be performed on the DOFs at a node before moving onto the next node, in the experiments, only 1 update is performed, that is, pointwise Gauss–Seidel is applied to nodally ordered systems.

Because the chosen discretization leads to graphs with most nodes having equal degree, the coloring scheme of Section 2 is used to select the C nodes. As for interpolation, we use the iBAMG scheme, but employ only intra-variable interpolation for systems of PDEs since inter-variable interpolation was shown in Reference 3 to offer little improvement for the PDEs examined. Again, thirty V(1,1) multigrid cycles are applied to problems defined with homogeneous right-hand sides and random initial guesses. The convergence rate of the solver is defined as $\frac{\|\mathbf{u}_h^{30}\|}{\|\mathbf{u}_h^{29}\|}$, averaged over ten simulations.

We also stress that very little effort was performed in adjusting parameters for optimal convergence, for example, adjusting the number of test vectors and number of relaxation sweeps per vector, the caliber of interpolation, and the strong-affinity threshold parameter θ in (13). More thorough parameter tuning can lead to better convergence. However, parameter tuning is not considered in this article.

6.1 | Scalar PDEs

The scalar PDEs we consider are the Laplace equation $-\nabla \cdot \nabla u(x, y) = 0$, the random diffusion equation $-\nabla \cdot D(x, y) \nabla u(x, y) = 0$ where coefficient $D(x, y)$ is normally distributed with mean 1 and standard deviation 0.05, and the rotated anisotropic diffusion equation

TABLE 7 Two-grid rates for 30 $V(2, 2)$ cycles applied to discretized scalar PDEs.

Scenario	# nodes	2-grid rates: iBAMG		2-grid rates: (9)/iBAMG
		Measure (30)	Measure (32)	Measure (36)-1TV/5TV
Laplace	20^2	0.09	0.07	0.21/0.16
	40^2	0.13	0.12	0.24/0.24
	80^2	0.17	0.22	0.30/0.34
Rand. Diff.	20^2	0.09	0.08	0.25/0.16
	40^2	0.16	0.10	0.28/0.24
	80^2	0.17	0.20	0.36/0.36
Rot. Diff. $\frac{\pi}{4}$	20^2	0.15	0.08	0.66/0.43
	40^2	0.20	0.16	0.86/0.73
	80^2	0.28	0.16	0.88/0.86
Rot. Diff. $-\frac{\pi}{4}$	20^2	0.16	0.11	0.69/0.36
	40^2	0.38	0.16	0.89/0.72
	80^2	0.46	0.17	0.86/0.85
Rot. Diff. $\frac{\pi}{8}$	20^2	0.31	0.15	0.73/0.68
	40^2	0.38	0.49	0.86/0.82
	80^2	0.44	0.66	0.88/0.91

Note: Measures (32) and (30) used 20 test vectors with 30 relaxation sweeps, and caliber 2 interpolation. Measure (36) used 1 or 5 test vectors with 50 relaxation sweeps of which 20 of these sweeps are used in the spin-up, and caliber 2 interpolation.

$$[(\cos^2\phi + \epsilon \sin^2\phi)\partial_{xx} + (1 - \epsilon)\sin(2\phi)\partial_{xy} + (\sin^2\phi + \epsilon \cos^2\phi)\partial_{yy}]u(x, y) = 0$$

with rotation ϕ . For the third PDE, when $\epsilon \ll 1$ and $\theta \neq \frac{k\pi}{2}$ for integer k , the anisotropy is not grid aligned, which leads to challenging systems for multigrid (for $\epsilon = 0.01$ and $\phi = \frac{\pi}{4}$, an accurate $V(1,1)$ two-grid rate of ≈ 0.66 was computationally obtained in Reference 52). We take $\epsilon = 0.0001$. We first consider two-grid rates. For each equation, 20 test vectors with 30 pointwise Gauss–Seidel sweeps on each vector are used in measures (30) and (32). (The number of test vectors is a large reduction in the amount used in the original correlation measure of Reference 3, that is, 80 test vectors with 15 smoothing sweeps were used there.) For auto-correlation measure (36), only 1 and 5 test vectors are used with 50 smoothing sweeps, of which 20 of these sweeps are used to spin-up each test vector. For the 1 test-vector case, operator-dependent interpolation (9) is applied; for the 5 test-vector case, iBAMG interpolation is applied. Lastly, a pathlength-distance neighborhood of 1 is used in each measure for the Laplace and random diffusion equations. For the anisotropic equation, we take pathlength 1 for determining the strength of connection for measure (30) and extend the pathlength to 2 for the interpolatory neighboring set for each fine-grid node^{25,26}; and for (32), we take pathlength 2 for determining the strength of connection and then extend the pathlength to 4 for the interpolatory neighboring set. For measure (36), a pathlength of 1 was retained for the interpolatory neighborhood since extending this radius did not improve the performance for the anisotropic problem.

Table 7 tabulates the two-grid $V(2, 2)$ rates (as in Reference 25) for the AMG schemes constructed using the different measures using caliber 2 interpolation. These rates were relatively insensitive to the initial set of random test vectors. We also see that “normalized” measure (32) gives the best overall performance. Unfortunately, although the rates for the auto-correlation measure (36) when applied to Laplace and random diffusion equations are decent, the rates are poor for the anisotropic problem and increasing the number of test vectors did not improve these rates. Given the decent rates for the weighted graph Laplacian problems in Section 4.2 and for the Laplace and random diffusion equations here, an immediate conjecture is that the anisotropy leads to *colored* randomness in the relaxed test vectors and their iterates. This colored randomness implies a biasing, which may be affecting the auto-correlation more substantially than the statistical variance and correlation. A thorough examination of this will be conducted in the future.

TABLE 8 Multigrid rates for 30 $W(2, 2)$ cycles applied to discretized scalar PDEs.

Scenario	# nodes	MG $W(2,2)/V(2,2)$ rates: iBAMG		MG $W(2,2)/V(2,2)$ rates: (9)/iBAMG Measure (36)-1TV/5TV
		Measure (30)	Measure (32)	
Laplace	20^2	0.10	0.12	0.37/0.13
	40^2	0.17	0.20	0.51/0.23
	80^2	0.21	0.26	0.51/0.33
Rand. Diff.	20^2	0.12	0.09	0.27/0.15
	40^2	0.21	0.21	0.37/0.29
	80^2	0.24	0.25	0.43/0.44
Rot. Diff. $\frac{\pi}{4}$	20^2	0.17	0.14	0.55/0.45
	40^2	0.26	0.20	0.83/0.83
	80^2	0.48	0.24	0.91/0.89
Rot. Diff. $-\frac{\pi}{4}$	20^2	0.21	0.16	0.65/0.40
	40^2	0.43	0.21	0.84/0.79
	80^2	0.62	0.24	0.92/0.90
Rot. Diff. $\frac{\pi}{8}$	20^2	0.33	0.19	0.75/0.73
	40^2	0.43	0.49	0.88/0.87
	80^2	0.61	0.69	0.92/0.92

Note: Measures (32) and (30) used 20 test vectors with 40 relaxation sweeps, and caliber 2 interpolation. Measure (36) used 1 or 5 test vectors with 50 relaxation sweeps of which 20 of these sweeps are used in the spin-up, and caliber 2 interpolation.

We turn to multigrid rates. $V(2, 2)$ cycles are applied to the Laplace and random diffusion equations, but $W(2, 2)$ cycles are now applied to the anisotropic equation, as in Reference 25. The setup is the same as in the two-grid simulations but now 40 Gauss–Seidel sweeps are performed on the test vectors when applying measures (30) and (32). Table 8 provides the results. We see that again (32) gives the best overall performance, and that the multilevel rates degrade from their two-grid rates, particularly for the anisotropic problems. Such scaling issue has been observed in the past, and can be improved by integrating a multilevel eigensolver into the BAMG procedure to produce more accurate test vectors to the near-nullspace components.^{17,25} Furthermore, for non-grid aligned anisotropies, as observed in Reference 25, scaling can be achieved when these W-cycle solvers are used as preconditioners for a Krylov iteration. Lastly, we see that the auto-correlation measure leads to decent multigrid rates only for the Laplace and random diffusion equations.

6.2 | Systems of PDEs

We consider both strongly and mildly cross-variable coupled systems of PDEs. Because of the poor performance in the auto-correlation measure for scalar PDEs, measure (40) will not be considered for the systems of PDEs.

Problem 1: Strongly coupled Laplacian system: Before looking at more interesting systems, we consider systems of strongly coupled Laplace operators:

$$\mathcal{L}\mathbf{u} = \begin{bmatrix} -2\Delta & \Delta & 0 & \cdots & 0 \\ \Delta & -2\Delta & \Delta & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \Delta & -2\Delta & \Delta \\ 0 & \cdots & \cdots & \Delta & -2\Delta \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \mathbf{0}. \quad (43)$$

TABLE 9 Convergence rates for 30 AMG $V(1, 1)$ cycles applied to discretized systems of strongly coupled Laplacians.

Laplacian Sys	# nodes	$V(1, 1)$ Conv rates: Cal.2/Cal.4			
		Meas (38)	Meas (39)	Meas (42)	Meas (42) diag
(2×2)	20^2	0.29/0.16	0.33/0.21	0.33/0.16	0.32/0.15
	40^2	0.35/0.17	0.39/0.17	0.36/0.17	0.38/0.17
	80^2	0.41/0.17	0.47/0.17	0.39/0.17	0.41/0.17
(3×3)	20^2	0.48/0.38	0.49/0.38	0.48/0.38	0.48/0.37
	40^2	0.52/0.38	0.55/0.38	0.53/0.39	0.53/0.38
	80^2	0.54/0.38	0.59/0.41	0.55/0.38	0.55/0.38

Note: For measures (38), (39), and (42) 20 test vectors are used with 20 relaxation sweeps, and and calibers 2 and 4 iBAMG interpolation.

(Although a Kronecker product preconditioner method can be used,⁵³ we are interested in the new affinity measures' ability to construct effective AMG solvers for these systems.) Measures (38), (39), (42), and the diagonal of (42) are applied with 20 test vectors and 20 point-based Gauss–Seidel sweeps per vector, and the strength-of-affinity computation is performed with pathlength distance 1. The interpolatory set is also kept to a neighborhood of pathlength 1 about a fine-grid node. Table 9 gives the results. The rates are similar for each measure, and are again similar to the results obtained using the correlation measure of Reference 3 but utilizing less test vectors (50 test vectors with 20 smoothing sweeps were used in Reference 3). It is a bit surprising that the diagonal of measure (42) performed well. This may be due to the similar structure of the diagonal and off-diagonal component operators in (43), that is, this similarity permits the diagonal of (42) to be reflective of the whole measure. Overall, the rates have a slight dependency on the problem size when caliber 2 interpolation is used, but scaling is gained when the caliber is increased.

Problem 2: Elasticity: A relevant system of PDEs is the displacement formulation of linear elasticity:

$$\mu \Delta \mathbf{u} + (\lambda + \mu) \nabla \nabla \cdot \mathbf{u} = \mathbf{0}, \quad (44)$$

where $\mathbf{u}$ is the displacement and λ, μ are the Lamé constants of the material. The rigid body modes are the near-nullspace components of this system, and for large λ -to- μ ratios, the system has a high-dimensional 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 the multi-scale divergence-free near-nullspace components arise. Since large near-nullspace issues are out of the scope of this article, we will not consider this limiting scenario. However, in addition to taking a unit square domain, we also consider an elongated beam of size (10×1) . Keeping the number of nodes in the x and y directions equal, the discretization leads to grid-aligned anisotropy, which makes the discretized systems more difficult for AMG.

Again, we use measures (38), (39), (42), and the diagonal of (42) to determine the affinity between the nodes. For the square domain scenarios, 20 test vectors with 20 point-based Gauss–Seidel sweeps per vector are used in each measure; for the beam scenarios, 20 test vectors with 10 Gauss–Seidel sweeps are used in the measures. For the square, a pathlength of 1 is used to determine the affinity and to circumscribe the interpolatory set. For the beam, for measures (38), (42), and the diagonal of (42), a pathlength of 1 is used to determine the affinity but a pathlength of 2 is used to circumscribe the interpolatory set; and for measure (39), a pathlength of 2 is used to determine the affinity but a pathlength of 4 is used to circumscribe the interpolatory set. Table 10 summarizes the results for $W(2, 2)$ multigrid cycles with caliber 2 iBAMG interpolation. These rates are better than those using correlation measure of Reference 3, and required substantially less test vectors (80 test vectors were used there) and with the performance less sensitive to the initial set of random test vectors. Increasing the radius of the interpolatory set to effect long range interpolation to handle the grid anisotropy improved the convergence in the beam scenarios, particularly for measures (39), (42), and the diagonal of (42). Although the rates for (38) are not as good as the other measures, they are still impressive although slightly dependent on the problem size. For the square scenarios, all the measures performed equally well and we observe more difficulty in handling the larger λ -to- μ ratio case (there is now no grid anisotropy to ameliorate the strong cross-variable, cross-derivative coupling in the second term of (44)). Overall, the rates are slightly dependent on the problem size but this issue can be removed by integrating a multilevel eigensolver in the BAMG procedure or by simply wrapping these multigrid solvers with a Krylov iteration.

TABLE 10 Convergence rates for 30 AMG multigrid $W(2, 2)$ cycles applied to a finite element discretization of (44).

Scenario	# nodes	Multigrid $W(2, 2)$ Conv rate: iBAMG			
		Meas (38)	Meas (39)	Meas (42)	Meas (42) diag
Square: $(\lambda, \mu) = (1, 1)$	20^2	0.12	0.13	0.15	0.14
	40^2	0.20	0.21	0.20	0.20
	80^2	0.26	0.26	0.26	0.26
Square: $(\lambda, \mu) = (10, 1)$	20^2	0.51	0.48	0.49	0.49
	40^2	0.60	0.58	0.59	0.59
	80^2	0.64	0.64	0.62	0.61
Beam: $(\lambda, \mu) = (1, 1)$	20^2	0.21	0.05	0.05	0.06
	40^2	0.23	0.05	0.05	0.07
	80^2	0.28	0.06	0.07	0.08
Beam: $(\lambda, \mu) = (10, 1)$	20^2	0.13	0.07	0.07	0.07
	40^2	0.14	0.08	0.08	0.08
	80^2	0.17	0.10	0.10	0.10

Note: For each scenario, 20 test vectors with 10 (for the beam) or 20 (for the square) point-based Gauss–Seidel sweeps per vector and caliber 2 iBAMG interpolation are used in each measure.

TABLE 11 Convergence rates for 30 AMG $W(2, 2)$ cycles applied to the coupled rotated anisotropic diffusion system (45).

Scenario: (ϕ_1, ϕ_2)	# nodes	$W(2, 2)$ Conv rate: iBAMG			
		Meas (38)	Meas (39)	Meas (42)	Meas (42) diag
$(-\frac{\pi}{4}, -\frac{\pi}{4})$	20^2	0.21	0.18	0.11	0.12
	40^2	0.36	0.20	0.28	0.27
	80^2	0.49	0.22	0.29	0.37
$(-\frac{\pi}{4}, -\frac{\pi}{6})$	20^2	0.34	0.33	0.29	0.29
	40^2	0.44	0.42	0.42	0.40
	80^2	0.65	0.47	0.58	0.54
$(-\frac{\pi}{4}, \frac{\pi}{4})$	20^2	0.67	0.72	0.71	0.66
	40^2	0.82	0.83	0.84	0.81
	80^2	0.87	0.84	0.91	0.86

Note: Each measure used 20 test vectors with 20 relaxation sweeps per vector and caliber 2 interpolation.

Problem 3: Coupled anisotropic diffusion: An extension of the scalar rotated anisotropic diffusion problem is the system

$$\begin{bmatrix} \mathcal{L}_{\epsilon_1 \phi_1} & \alpha \partial_{xy} \\ \alpha \partial_{yx} & \mathcal{L}_{\epsilon_2 \phi_2} \end{bmatrix} \mathbf{u} = \mathbf{0}, \quad (45)$$

where $\mathcal{L}_{\epsilon_i \phi_i} = (\cos^2 \phi_i + \epsilon_i \sin^2 \phi_i) \partial_{xx} + (1 - \epsilon_i) \sin(2\phi_i) \partial_{xy} + (\sin^2 \phi_i + \epsilon_i \cos^2 \phi_i) \partial_{yy}$ with rotation ϕ_i . As we saw earlier, the diagonal component operators themselves can be challenging for multigrid. Now the operators are coupled, which will lead to additional challenges especially when the rotations in the diagonal operators are not the same. Such non-alignment can confuse the C-node selection because the different directions of anisotropies can disorient the detection of the nodal affinities. For the cross-variable coupling, we take $\alpha = 0.0001$ in system (45). Several rotation pairs are considered, and for each scenario, 20 test vectors with 20 point-based Gauss–Seidel sweeps per vector are employed in each measure. Furthermore, the pathlength settings for the measures are the same as in the elasticity experiments.

Table 11 tabulates the results for multigrid $W(2, 2)$ cycles. When $\phi_1 = \phi_2 = -\frac{\pi}{4}$, we see that the measures performed as well as its corresponding scalar problem with $\phi = -\frac{\pi}{4}$ using measures (30) and (32). The performance for all the measures degraded when diagonal operator rotations do not align, as illustrated when $\phi_1 = -\frac{\pi}{4}$ and $\phi_2 = -\frac{\pi}{6}$ and more so for the larger non-alignment $(\phi_1, \phi_2) = \left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$. This is to be expected since the non-aligned anisotropies are disorienting the measures detection of the nodal affinities. Because of this, there are scaling issues, which will require future investigation to resolve.

7 | CONCLUSION

This article examines some new affinity measures for selecting the CDOFs/C nodes. They can be utilized in the efficient analysis of networks and in the construction of AMG solvers for weighted graph Laplacians and systems of elliptic PDEs. These measures are related to the diffusion distance/effective resistance of a Markov process, and thus they bring physics into the coarse-grid selection procedure. These measures can construct AMG solvers that have better performance than the solvers constructed with the original correlation measure of Reference 3. In particular, compared to this earlier measure, less number of test vectors are needed to provide comparable or better convergence rates, and the performance behavior of AMG is less sensitive to the initial set of random test vectors. The results are encouraging, both for large-scale network analysis and for AMG development for PDEs.

ACKNOWLEDGMENTS

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

CONFLICT OF INTEREST STATEMENT

This study does not have any conflicts to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ENDNOTES

¹We will interchangeably use DOF and node to indicate either quantity.

²Although this assessment measure is often used in graph analysis, it can be misleading. For example, a smaller giant component may indicate high vulnerability than a larger giant component, that is, a more detrimental decomposition of the network has occurred. But if the system operation in this giant component is relatively stable, then the vulnerability may not be as bad as this measure indicates it to be.

ORCID

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

REFERENCES

- Fullenbach T, Stuben K. Algebraic multigrid for selected PDE systems. Proceedings of the 4th European Conference on Elliptic and Parabolic Problems, Rolduc and Gaeta, 2001. London: World Scientific; 2002. p. 399–410.
- Griebel M, Oeltz D, Schweitzer MA. An algebraic multigrid method for linear elasticity. *SIAM J Sci Comput*. 2003;25:385–407.
- Lee B. Algebraic multigrid for systems of elliptic boundary-value problems. *Numer Linear Algebra Appl*. 2019;28:e2303. <https://doi.org/10.1002/nla.2303>
- Ruge JW, Stuben K. Algebraic multigrid (AMG). In: McCormick SF, editor. Multigrid methods, Frontiers in applied mathematics. Philadelphia: SIAM; 1987.
- Barker AT, Gelever SV, Lee CS, Osborne SV, Vassilevski PS. Multilevel spectral coarsening for graph Laplacian problems with application to reservoir simulation. *SIAM J Sci Comput*. 2021;42:A2737–65.
- Brannick J, Chen Y, Kraus J, Zikatanov L. Algebraic multilevel preconditioners for the graph Laplacian based on matching in graphs. *SIAM J Numer Anal*. 2013;51:1805–27.
- D'Ambra P, Cuttillo L, Vassilevski PS. Bootstrap AMG for spectral clustering. *Comput Math Methods*. 2019;1:e1020.
- Napov A, Notay Y. An efficient multigrid method for graph Laplacian systems. *Electron Trans Numer Anal*. 2016;45:201–18.
- Napov A, Notay Y. An efficient multigrid method for graph Laplacian systems II: robust aggregation. *SIAM J Sci Comput*. 2017;39:S379–S403.
- Urschel JC, Xu J, Hu X, Zikatanov LT. A Cascadic multigrid algorithm for computing the Fiedler vector of graph Laplacians. *J Comput Math*. 2015;33:209–26.

11. Bolten M, Brandt A, Brannick J, Frommer A, Kahl K, Livshits I. A bootstrap algebraic multilevel method for Markov chains. *SIAM J Sci Comput.* 2011;33:3425–46.
12. de Sterck H, Manteuffel TA, McCormick SF, Miller K, Pearson J, Ruge J, et al. Smoothed aggregation multigrid for Markov chains. *SIAM J Sci Comput.* 2010;32:40–61.
13. de Sterck H, Manteuffel TA, McCormick SF, Miller K, Ruge J, Sanders G. Algebraic multigrid for Markov chains. *SIAM J Sci Comput.* 2010;32:544–62.
14. Belkin M, Niyogi P. Towards a theoretical foundation for Laplacian-based manifold methods. *J Comput Syst Sci.* 2008;74:1289–308.
15. Coifman R, Lafon S. Diffusion maps. *Appl Comput Harmon Anal.* 2006;20:5–30.
16. Nadler B, Lafon S, Coifman RR, Kevrekidis IG. Diffusion maps, spectral clustering and the reaction coordinates of dynamical systems. *Appl Comput Harmon Anal.* 2006;21:113–27.
17. Brandt A, Brannick J, Kahl K, Livshitz I. Bootstrap AMG. *SIAM J Sci Comput.* 2011;32:612–32.
18. Brannick JJ, Falgout RD. Compatible relaxation and coarsening in algebraic multigrid. *SIAM J Sci Comput.* 2010;32:1393–416.
19. Brezina M, Falgout RD, MacLachlan S, Manteuffel T, McCormick S, Ruge J. Adaptive algebraic multigrid. *SIAM J Sci Comput.* 2006;27:1261–86.
20. Livne OE. Coarsening by compatible relaxation. *Numer Linear Algebra Appl.* 2004;11:205–27.
21. Livne OE, Brandt A. Lean algebraic multigrid (LAMG): fast graph Laplacian linear solver. *SIAM J Sci Comput.* 2012;34:B499–B522.
22. Ron D, Safro I, Brandt A. Relaxation-based coarsening AMD multiscale graph organization. *Multiscale Model Simul.* 2011;9:407–23.
23. Gottschalk J, Kahl K. Coarsening in algebraic multigrid using Gaussian processes. *Electron Trans Numer Anal.* 2021;54:514–33.
24. Kahl K, Rottmann M. Least angle regression coarsening in bootstrap algebraic multigrid. *SIAM J Sci Comput.* 2018;40:A3928–54.
25. Brandt A, Brannick J, Kahl K, Livshitz I. Algebraic distance for anisotropic diffusion problems: multilevel results. *Electron Trans Numer Anal.* 2015;44:472–96.
26. Schroder JB. Smoothed aggregation solvers for anisotropic diffusion. *Numer Linear Algebra Appl.* 2012;19:296–312.
27. Ellens W, Spieksma FM, Van Mieghem P, Jamakovic A, Kooij RE. Effective graph resistance. *Numer Linear Algebra Appl.* 2011;435:2491–506.
28. Goldberg MJ, Kim S. Some remarks on diffusion distances. *J Appl Math.* 2010;2010:464815. <https://doi.org/10.1155/2010/464815>
29. Spielman DA, Srivastava N. Graph sparsification by effective resistance. *SIAM J Comput.* 2011;40:1913–26.
30. von Luxburg U, Radl A, Hein M. Getting lost in space: large sample analysis of the resistance distance. In: Lafferty J, Williams C, Shawe-Taylor J, Zemel R, Culotta A, editors. *Advances in neural information processing systems*. Volume 10. New York: Curran Associates, Inc.; 2010.
31. von Luxburg U, Radl A, Hein M. Hitting and commute times in large random neighborhood graphs. *J Mach Learn Res.* 2014;14:1751–98.
32. Van Mieghem P. *Graph spectra for complex networks*. New York: Cambridge University Press; 2011.
33. Van Mieghem P, Devriendt K, Cetinay H. Pseudoinverse of the Laplacian and best spreader node in a network. *Phys Rev E.* 2017;96:032311-1–032311-22.
34. Newman M. *Networks*. 2nd ed. Oxford, UK: Oxford University Press; 2018.
35. Van Mieghem P. *Performance analysis of complex networks and systems*. Cambridge, UK: Cambridge University Press; 2014.
36. Alev VL, Anari N, Lau LC, Gharan SO. Graph clustering using effective resistance. 9th Innovations in Theoretical Computing Science Conference. Saarbrücken: Dagstuhl Publishing; 2018.
37. Bergamini E, Wegner M, Lukarski D, Meyerhenke H. Estimating current-flow closeness centrality with a multigrid Laplacian solver. *Internet Math.* 2016;12:281–314.
38. Lin J, Cowen LJ, Hescott B, Hu X. Computing the diffusion state distance on graphs via algebraic multigrid and random projections. *Numer Linear Algebra Appl.* 2018;25:e2156. <https://doi.org/10.1002/nla.2156>
39. Trottenberg U, Oosterlee CW, Schuller A. *Multigrid*. London: Academic Press; 2001.
40. Xu J, Zikatanov L. Algebraic multigrid methods. *Acta Numer.* 2017;26:591–721.
41. Lee B, Batista EP. Algebraic multigrid for the nonlinear powerflow equations. *Numer Linear Algebra Appl.* 2020;28:e2347. <https://doi.org/10.1002/nla.2347>
42. Briggs WL, Henson VE, McCormick SF. *A multigrid tutorial*. 2nd ed. Philadelphia: SIAM; 2000.
43. Brannick J, Cao F, Kahl K, Falgout RD, Hu X. Optimal interpolation and compatible relaxation in classical algebraic multigrid. *SIAM J Sci Comput.* 2018;40:A1473–93.
44. Falgout RD, Vassilevski PS. On generalizing the AMG framework. *SIAM J Numer Anal.* 2004;42:1669–93.
45. Brandt A, Livshitz I. Wave-ray multigrid method for solving standing wave equations. *Electron Trans Numer Anal.* 1997;6:162–81.
46. Lee B, Manteuffel TA, McCormick SF, Ruge J. First-order system least squares for Helmholtz equations. *SIAM J Sci Comput.* 2000;21:1927–49.
47. Brezina M, Cleary AJ, Falgout RD, Henson VE, Jones JE, Manteuffel TA, et al. Algebraic multigrid based on element interpolation (AMGe). *SIAM J Sci Comput.* 2000;22:1570–92.
48. Manteuffel T, McCormick S, Park M, Ruge J. Operator-based interpolation for bootstrap algebraic multigrid. *Numer Linear Algebra Appl.* 2010;17:519–37.
49. Johnson WB, Lindenstrauss J. Extensions of Lipschitz mappings into a Hilbert space. *Contemp Math.* 1984;26:189–206.
50. Mei S, Zhang X, Cao M. Complex small-world power grids. *Power grid complexity*. Berlin: Springer; 2011. p. 161–78.
51. Hartmann B, Sugar V. Searching for small-world and scale-free behaviour in long-term historical data of a real-world power grid. *Sci Rep.* 2021;11:6575. <https://doi.org/10.1038/s41598-021-86103-7>

52. MacLachlan SP, Olson LN. Theoretical bounds for algebraic multigrid performance: review and analysis. *Numer Linear Algebra Appl.* 2014;21:194–220.
53. Lee B. Parallel preconditioners and multigrid solvers for stochastic polynomial chaos discretizations of the diffusion equation at the large scale. *Numer Linear Algebra Appl.* 2015;23:5–36. <https://doi.org/10.1002/nla.2000>

How to cite this article: Lee B. Bringing physics into the coarse-grid selection: Approximate diffusion distance/effective resistance measures for network analysis and algebraic multigrid for graph Laplacians and systems of elliptic partial differential equations. *Numer Linear Algebra Appl.* 2023;e2539. <https://doi.org/10.1002/nla.2539>