Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Algebraic multilevel preconditioners for the graph Laplacian based on matching in graphs

J. Brannick, Y. Chen, J. Kraus, L. Zikatanov

RICAM-Report 2012-28

ALGEBRAIC MULTILEVEL PRECONDITIONERS FOR THE GRAPH LAPLACIAN BASED ON MATCHING IN GRAPHS

J. BRANNICK, Y. CHEN, J. KRAUS, AND L. ZIKATANOV

Abstract. This paper presents estimates of the convergence rate and complexity of an algebraic multilevel preconditioner based on piecewise constant coarse vector spaces applied to the graph Laplacian. A bound is derived on the energy norm of the projection operator onto any piecewise constant vector space, which results in an estimate of the two-level convergence rate where the coarse level graph is obtained by matching. The two-level convergence of the method is then used to establish the convergence of an Algebraic Multilevel Iteration that uses the two-level scheme recursively. On structured grids, the method is proven to have convergence rate $\approx (1 - 1/\log n)$ and $O(n \log n)$ complexity for each cycle, where n denotes the number of unknowns in the given problem. Numerical results of the algorithm applied to various graph Laplacians are reported. It is also shown that all the theoretical estimates derived for matching can be generalized to the case of aggregates containing more than two vertices.

1. Introduction

In this paper we analyze an aggregation based Algebraic Multigrid (AMG) algorithm for graph Laplacians. A typical AMG algorithm has two phases: (1) a setup phase to construct a nested sequence of coarse spaces; (2) a solve phase which uses the multilevel hierarchy to compute the solution. Two well known and popular approaches for an AMG setup are classical AMG [3] and (smoothed) aggregation AMG [20, 25, 8, 26, 14], which are distinguished by the type of coarse variables used in the construction of AMG interpolation. In an aggregation-based AMG algorithm, the setup phase partitions the "fine grid" variables into disjoint sets, called aggregates. Then, a column (or several columns as in [26]) of interpolation is associated to each aggregate, which has nonzero entries only for the unknowns belonging to this aggregate.

Our focus is on the convergence analysis of a class of aggregation-type AMG methods with multilevel hierarchies constructed via a pair-wise aggregation, or matching. The aim is to analyze the matching AMLI solver for the graph Laplacian in detail as an important and crucial step in gaining an in-depth understanding of a multilevel solvers for general graphs. We first demonstrate that a two-level method based on such aggregation (or even more general types of aggregations) for the graph Laplacian on a general graph is uniformly convergent and, thus, can be used within an Algebraic Multilevel Iteration (AMLI) [1, 27] as a preconditioner in the conjugate gradient iteration to obtain a nearly optimal solver. A noteworthy feature of the approach is its simplicity, which makes it possible to analyze the convergence and complexity of the method with few assumptions and without relying on geometric information.

The idea of aggregating unknowns to coarsen a system of discretized partial differential equations dates back to work by Leont'ev in 1959 [17]. Simon and Ando developed a related technique for aggregating dynamic systems in 1961 [20] and a two-grid aggregation-based scheme was considered in the context of solving Markov chain systems by Takahashi in 1975 [23]. Multilevel hierarchies

Date: November 11, 2012.

1991 *Mathematics Subject Classification.* 65N30, 65N15.

This work was supported in part by the National Science Foundation, DMS-0810982, OCI-0749202 and by the Austrian Science Fund, Grant P22989-N18.

based on pair-wise aggregation (matching) have been used in graph partitioning algorithms [13, 12], and the numerical solution of convection diffusion problems [14].

Efficient multilevel graph Laplacian solvers are important in numerous application areas, including finite element and finite difference discretizations of elliptic partial differential equations (PDE), data mining, clustering in images, analysis of network graphs. Recently, notable developments in the algorithm design, which provide fast and efficient solvers for graph Laplacians include Lean Algebraic Multigrid (LAMG), of Brandt and Livne [18], combinatorial multigrid and multilevel preconditioners, see [16, 15]. For a recent overview of graph Laplacian preconditioners we refer to Spielman [21].

The aggregation-based methods have been studied extensively since the 90s and numerous algorithms and theoretical results have followed [8, 26, 14]. Vanék introduced an extension of these methods known as smoothed aggregation multigrid in which smoothing steps are applied to the columns of the aggregation-based interpolation operator to accelerate two-level convergence and a modification of this two-level algorithm with over-correction is presented in [25]. A multilevel smoothed aggregation algorithm and its convergence analysis are found in [24] and, in [27], an improved convergence theory of the method is presented. The latter theory is then extended to allow for aggressive coarsening, provided an appropriate polynomial smoother is used [7]. Further generalizations known as adaptive smoothed aggregation [6] as well as aggregating graph nodes using algebraic distances [19, 10, 2, 18] have been developed and studied. Variants of the above approaches continue to be developed for use in scientific computing and have been developed for higher order partial differential equations [26], convection diffusion problems [14], Markov chains [22, 2], and the Dirac equation in quantum chromodynamics [4].

The remainder of the paper is organized as follows. In Section 2, we introduce the graph Laplacian problem and discuss some of its applications. In Section 3, we introduce a graph matching algorithm and demonstrate that the energy norm of the ℓ^2-projection onto the coarse space is a key quantity in deriving convergence and complexity estimates of the method. Additionally, we introduce an approach computing an approximation of the energy norm of this projection operator. In Section 4, we present an analysis of the two-level method for the graph Laplacian. In Section 5, we consider the convergence and complexity of the resulting AMLI method, and in Section 6 we provide numerical results and address some practical issues of the method.

2. Problem formulation and notation

Consider an undirected unweighted connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ denotes the set of vertices and $\mathcal{E}$ denotes the set of edges of $\mathcal{G}$. We denote the cardinality of a finite set $\mathcal{X}$ by $|\mathcal{X}|$ and we set $n = |\mathcal{V}|$. By $(\cdot,\cdot)$ we denote the inner product in $\ell^2(\mathbb{R}^n)$. The superscript T denotes the adjoint with respect to this inner product. Next, we define the discrete gradient operator associated with G, $B : \mathbb{R}^n \mapsto \mathbb{R}^{|\mathcal{E}|}$ by

$$(B\boldsymbol{u})_k = u_i - u_j, \quad k = (i,j) \in \mathcal{E}, \quad i < j. \tag{2.1}$$

Here, $k = (i,j) \in \mathcal{E}$ denotes the edge that connects vertices i and j, and u_i and u_j are the i-th and j-th coordinate of the vector $\boldsymbol{u}$, respectively. The *graph Laplacian* $A : \mathbb{R}^n \mapsto \mathbb{R}^n$ is then defined as

$$\begin{aligned} &A = B^T B, \quad \text{or equivalently,} \\ &(A\boldsymbol{u}, \boldsymbol{v}) = (B\boldsymbol{u}, B\boldsymbol{v}) \quad \text{for all} \quad \boldsymbol{u} \in \mathbb{R}^n, \quad \boldsymbol{v} \in \mathbb{R}^n. \end{aligned} \tag{2.2}$$

Clearly, A is symmetric and positive semi-definite and its kernel is the space spanned by the constant vector $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^n$. These properties can also be verified using the matrix form of A,

which in the canonical Euclidean basis in $\mathbb{R}^n$ is

$$(A)_{ij} = \begin{cases} d_i & i = j; \\ -1 & i \neq j, (i,j) \in \mathcal{E}; \\ 0 & i \neq j, (i,j) \notin \mathcal{E}; \end{cases}$$

where d_i is the degree (the number of neighbors) of the i-th vertex. From the definition of A is immediate to check that

$$(A\boldsymbol{u}, \boldsymbol{v}) = \sum_{k=(i,j)\in\mathcal{E}} (u_i - u_j)(v_i - v_j).$$

With A we associate the following linear system: Given a vector $\boldsymbol{f} \in \mathbb{R}^n$, satisfying $(\boldsymbol{f}, \mathbf{1}) = 0$, find $\boldsymbol{u} \in \mathbb{R}^n$, such that

$$A\boldsymbol{u} = \boldsymbol{f}, \quad \text{and} \quad (\boldsymbol{u}, \mathbf{1}) = 0. \tag{2.3}$$

For brevity, we kept the analysis confined mostly to the case of unweighted graphs. It is however possible and straightforward to generalize our results to weighted graphs with positive (or even negative weights). The theory developed here for multilevel aggregation solvers applied to graph Laplacians should provide insights on how to design a solver for more general weighted graph Laplacians which have applications to anisotropic diffusion problems, or numerical models involving non-PDE graphs. We next consider some of the changes that occur if weights are introduced, or, as in PDE discretizations, one considers more general positive definite matrices A.

Weighted graphs. Assume that the graph is weighted and the k-th edge is assigned a weight w_k, then the corresponding bilinear form of A is

$$(A\boldsymbol{u}, \boldsymbol{v}) = \sum_{k=(i,j)\in\mathcal{E}} w_k(u_i - u_j)(v_i - v_j).$$

Define $D : \mathbb{R}^{|\mathcal{E}|} \mapsto \mathbb{R}^{|\mathcal{E}|}$ as a diagonal matrix whose k-th diagonal entry is equal to w_k, then the matrix A can be decomposed as

$$A = B^T D B.$$

Finite element and finite difference discretizations of elliptic PDEs with Neumann boundary conditions results in such weighted graph Laplacians.

More general positive definite matrices. Here we show how solution of systems with more general positive definite $(n \times n)$ matrices can be cast in terms of solution of systems with graph Laplacians in $\mathbb{R}^{n+1}$. Assume that A is positive definite, $A = A_s + A_t$, where both A_s and A_t are positive semidefinite. Further, assume that the only vector in the null space of A_s is $\mathbf{1}$.

Remark 2.1. *For example, such matrices A are the obtained from discretization of second order scalar elliptic PDEs. The simplest case is probably the five point finite difference discretization of the Laplace operator on a rectangular grid. In such case A_s corresponds to all the interior nodes and A_t is a diagonal matrix with nonzero diagonal elements corresponding to the nodes near the boundary (rather* next to *the boundary).*

Denoting for clarity the constant vector in $\mathbb{R}^n$ by $\mathbf{1}_n$ and setting $\mathbf{1}_{n+1} = (\mathbf{1}_n, 1)^T$, we show next that solving $A\boldsymbol{u} = \boldsymbol{f}$ is equivalent to solving $\widetilde{A}\boldsymbol{U} = \boldsymbol{F}$, with

$$\widetilde{A} = \begin{pmatrix} A_s + A_t & -A_t\mathbf{1}_n \\ -(A_t\mathbf{1}_n)^T & (A_t\mathbf{1}_n, \mathbf{1}_n) \end{pmatrix} \quad \text{and} \quad \boldsymbol{F} = \begin{pmatrix} \boldsymbol{f} \\ -(\boldsymbol{f}, \mathbf{1}_n) \end{pmatrix}$$

Clearly, if A_s is a weighted graph Laplacian, then the augmented linear system is also a graph Laplacian and our algorithm and its analysis readily apply to many of the discretizations of scalar elliptic PDEs.

To show the equivalence, we note that, from the definitions of $\widetilde{A}$ and $\boldsymbol{F}$ immediately have

$$(\boldsymbol{F}, \mathbf{1}_{n+1}) = 0, \quad \text{and} \quad \widetilde{A}\mathbf{1}_{n+1} = \mathbf{0} \in \mathbb{R}^{n+1}.$$

Further, let $\boldsymbol{U} = (\widetilde{\boldsymbol{u}}, \alpha)^T \in \mathbb{R}^{n+1}$, $\widetilde{\boldsymbol{u}} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ be the unique solution to

$$\widetilde{A}\boldsymbol{U} = \boldsymbol{F}, \quad (\boldsymbol{U}, \mathbf{1}_{n+1}) = 0. \tag{2.4}$$

Then the vector $\boldsymbol{u} = (\widetilde{\boldsymbol{u}} - \alpha\mathbf{1}_n)$ is the solution to $A\boldsymbol{u} = \boldsymbol{f}$.

The proof is simple, and follows from the definitions and our assumptions on A, A_s and A_t. We have

$$\begin{aligned}(A_s + A_t)\boldsymbol{u} &= (A_s + A_t)(\widetilde{\boldsymbol{u}} - \alpha\mathbf{1}_n) = (A_s + A_t)\widetilde{\boldsymbol{u}} - \alpha(A_s + A_t)\mathbf{1}_n \\ &= (A_s + A_t)\widetilde{\boldsymbol{u}} - \alpha A_t\mathbf{1}_n = \boldsymbol{f}.\end{aligned}$$

In the last step we have used that $A_s\mathbf{1}_n = 0$, and the first component of $\widetilde{A}\boldsymbol{U} = F$.

It is also straightforward to check that the converse statement is also true: if $\boldsymbol{u}$ is solution to $A\boldsymbol{u} = \boldsymbol{f}$ then we have that the unique solution of (2.4) is

$$\boldsymbol{U} = \begin{pmatrix}\boldsymbol{u} \\ 0\end{pmatrix} - \frac{(\boldsymbol{u}, \mathbf{1}_n)}{n+1}\begin{pmatrix}\mathbf{1}_n \\ 1\end{pmatrix}.$$

Thus we have shown that solving one of the problems, $A\boldsymbol{u} = \boldsymbol{f}$, or, $\widetilde{A}\boldsymbol{U} = \boldsymbol{F}$, gives the solution to the other.

3. Space decomposition based on matching

We begin with an outline of the basic notions related to aggregation using matching. We consider and construct an edge-space projection which commutes with the discrete gradient and this property provides one of the key ingredients in the convergence analysis.

3.1. Subspaces by graph partitioning and graph matching.

A graph partitioning of $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a set of subgraphs $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$, each with set of vertices $\mathcal{V}_i$ and set of edges $\mathcal{E}_i$ and such that

$$\cup_i \mathcal{V}_i = \mathcal{V}, \quad \mathcal{V}_i \cap \mathcal{V}_j = \emptyset, \quad i \neq j.$$

Without loss of generality we assume that all the subgraphs are non empty and connected. One of the simplest examples of such a graph partitioning is a matching, i.e, a collection (subset $\mathcal{M}$) of edges in $\mathcal{E}$ such that no two edges in $\mathcal{M}$ are incident.

For a given graph partitioning, subspaces of $V = \mathbb{R}^{|\mathcal{V}|}$ are defined as

$$V_c = \{\boldsymbol{v} \in V |\ (\boldsymbol{v})_j = (\boldsymbol{v})_k \text{ if } j \in \mathcal{V}_i \text{ and } k \in \mathcal{V}_i, \quad \forall i\}.$$

Note that each vertex in $\mathcal{G}$ corresponds to a connected subgraph S of $\mathcal{G}$ and every vertex of $\mathcal{G}$ belongs to exactly one such component. The vectors from V_c are constants on these connected subgraphs. Of importance is the ℓ^2-orthogonal projection on V_c, which is denoted by Q, and defined as follows:

$$(Q\boldsymbol{v})_i = \frac{1}{|\mathcal{V}_k|}\sum_{j \in \mathcal{V}_k} v_j, \quad \forall i \in \mathcal{V}_k. \tag{3.1}$$

Given a graph partitioning, the coarse graph $\mathcal{G}_c = \{\mathcal{V}_c, \mathcal{E}_c\}$ is defined by assuming that all vertices in each of the subgraphs from the partitioning form an equivalence class, and that $\mathcal{V}_c$ and $\mathcal{E}_c$ are the quotient set of $\mathcal{V}$ and $\mathcal{E}$ under this equivalence relation. That is, any vertex in $\mathcal{V}_c$ corresponds to a subgraph in the partitioning of $\mathcal{G}$, and the edge (i, j) exists in $\mathcal{E}_c$ if and only if the i-th and

j-th subgraphs are connected in the graph $\mathcal{G}$. Figure 1 is an example of a matching in a graph and the resulting coarse graph.

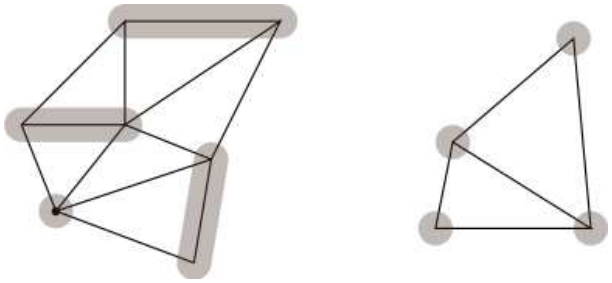


Figure 1: Matching $\mathcal{M}$ in a graph $\mathcal{G}$ (left) and the coarse graph $\mathcal{G}_c$ (right)

As mentioned above, the reason to focus on matching is that it simplifies the computation of several key quantities used in the upcoming estimates derived for a perfect matching and it is possible to show that a matching which is not perfect can be analyzed in a similar way.

3.2. **Commutative diagram.** Let B be the discrete gradient of a graph Laplacian A, as defined in (2.1) and Q be defined as in (3.1). Assume that there exists an operator Π_k such that the following commutative diagram holds true:

$$\begin{array}{ccc} \mathbb{R}^{|\mathcal{V}|} & \xrightarrow{B} & \mathbb{R}^{|\mathcal{E}|} \\ {\scriptstyle Q}\downarrow & & \downarrow{\scriptstyle \Pi} \\ V_c & \xrightarrow[B]{} & \mathbb{R}^{|\mathcal{E}|} \end{array}$$

The proof of this assumption is provided later on. From the commutative relation $BQ = \Pi B$ it follows that

$$|Q\boldsymbol{v}|_A^2 = \|BQ\boldsymbol{v}\|^2 = \|\Pi B\boldsymbol{v}\|^2 \leq \|\Pi\|^2 |\boldsymbol{v}|_A^2. \tag{3.2}$$

Thus, an estimate on the A-semi-norm of Q amounts to an estimate of the ℓ^2-norm of Π. In the next subsection, an explicit form of Π is constructed and an estimate of its ℓ^2-norm is derived.

Remark 3.1. *In the more general case of a weighted graph Laplacian, i.e., assuming that the weight matrix $D \neq I$, the bound on the norm $|Q|_A$ becomes*

$$|Q\boldsymbol{v}|_A^2 = (DBQ\boldsymbol{v}, BQ\boldsymbol{v}) = (D\Pi B\boldsymbol{v}, \Pi B\boldsymbol{v}) \leq \|D^{1/2}\Pi_k D^{-1/2}\|^2 |\boldsymbol{v}|_A^2,$$

where D can have also negative weights, which results in a matrix $D^{1/2}\Pi_k D^{-1/2}$ that is complex-valued. A detailed analysis in such a setting and the application of this idea to anisotropic diffusion problems are discussed in [5].

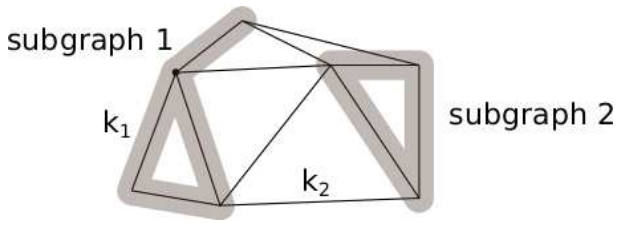
3.3. **Construction of Π in case of piece-wise constant spaces.** Here, we proceed with an explicit construction and ℓ^2-norm estimate of the operator Π.

For any graph partitioning in which the subgraphs are connected, a given edge belongs to the set of "internal edges", whose vertices belong to the same subgraph, or to the set of "external edges", whose vertices belong to two distinct subgraphs. For example, let $\mathcal{G}_1$ and $\mathcal{G}_2$ denote the subgraphs 1 and 2 in Fig. 2, then k_1 is an internal edge and k_2 is an external edge.

Since the vector $Q\boldsymbol{v}$ has the same value on the two endpoints of the edge k_1, we have that $(BQ\boldsymbol{v})_{k_1} = 0$. Accordingly, all entries in $(\Pi)_{k_1}$, the k_1-th row of Π, are set to zero:

$$(\Pi B\boldsymbol{v})_{k_1} = (\Pi)_{k_1} B\boldsymbol{v} = 0.$$

Figure 2: Connected components and the construction of Π_k



For the external edge k_2, it follows that $(\Pi)_{k_2}$ satisfies

$$(\Pi)_{k_2}(B\boldsymbol{v}) = (BQ\boldsymbol{v})_{k_2} = \frac{1}{|\mathcal{V}_1|}\sum_{i_1\in\mathcal{V}_1} v_{i_1} - \frac{1}{|\mathcal{V}_2|}\sum_{i_2\in\mathcal{V}_2} v_{i_2}, \tag{3.3}$$

for every $\boldsymbol{v}$. The following Lemma is useful in computing explicitly the entries of $(\Pi)_{k_2}$.

Lemma 3.2. *Let $A:\mathbb{R}^n\mapsto\mathbb{R}^n$ be a positive semidefinite operator and let $\{\chi_i\}_{i=1}^n$ be a basis of $\mathbb{R}^n$. Assume that the null space of A is one dimensional, namely there exists a nonzero vector $\boldsymbol{s}$ such that $\mathrm{Ker}(A)=\mathrm{span}(\boldsymbol{s})$, and for every integer $1\le i\le n$ we have $(\chi_i,\boldsymbol{s})=1$. We then have:*

(i) *For any i, the operator $\widetilde{A}:\mathbb{R}^n\mapsto\mathbb{R}^n$ with $\widetilde{A}\boldsymbol{u}=(A\boldsymbol{u}+(\chi_i,\boldsymbol{u})\chi_i)$ is invertible.*

(ii) *The following identity holds for all $\boldsymbol{u}\in\mathbb{R}^n$:*

$$\frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\boldsymbol{u},\boldsymbol{s}) - (\boldsymbol{u},\chi_i) = \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\widetilde{A}^{-1}\boldsymbol{s},A\boldsymbol{u}).$$

Proof. To establish (i) it suffices to show that $\widetilde{A}\boldsymbol{v}=0$ implies $\boldsymbol{v}=0$. Assuming that $\widetilde{A}\boldsymbol{v}=0$ for some $\boldsymbol{v}\in\mathbb{R}^n$ it follows that

$$0=(\widetilde{A}\boldsymbol{v},\boldsymbol{v})=(A\boldsymbol{v},\boldsymbol{v})+(\chi_i,\boldsymbol{v})^2.$$

Note that both terms on the right side of the above identity are nonnegative and, hence, their sum can be zero if and only if both terms are zero. Since A is positive semidefinite by assumption with one dimensional null space, from $(A\boldsymbol{v},\boldsymbol{v})=0$ we conclude that $\boldsymbol{v}=\alpha\boldsymbol{s}$ for some $\alpha\in\mathbb{R}$. For the second term we have that $0=(\chi_i,\boldsymbol{v})^2=\alpha^2(\chi_i,\boldsymbol{s})^2$, and since $(\chi_i,\boldsymbol{s})\neq 0$ for all i, it follows that $\alpha=0$ and hence $\boldsymbol{v}=0$. This proves (i).

Now, applying (i) the result (ii) follows:

$$\begin{aligned}
\frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\widetilde{A}^{-1}\boldsymbol{s},A\boldsymbol{u}) &= \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\widetilde{A}^{-1}\boldsymbol{s},A\boldsymbol{u}+(\chi_i,\boldsymbol{u})\chi_i) - \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\widetilde{A}^{-1}\boldsymbol{s},(\chi_i,\boldsymbol{u})\chi_i)\\
&= \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\widetilde{A}^{-1}\boldsymbol{s},\widetilde{A}\boldsymbol{u}) - \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\chi_i,\boldsymbol{u})(\boldsymbol{s},\widetilde{A}^{-1}\chi_i)\\
&= \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\boldsymbol{s},\boldsymbol{u}) - \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\chi_i,\boldsymbol{u})(\boldsymbol{s},\widetilde{A}^{-1}\chi_i)\\
&= \frac{1}{(\boldsymbol{s},\boldsymbol{s})}(\boldsymbol{u},\boldsymbol{s}) - (\boldsymbol{u},\chi_i).
\end{aligned}$$

Here, the equality $\widetilde{A}\boldsymbol{s}=A\boldsymbol{s}+(\chi_i,\boldsymbol{s})\chi_i=\chi_i$ was used, implying that $\widetilde{A}^{-1}\chi_i=\boldsymbol{s}$. □

Remark 3.3. *A special case is obtained by setting $\boldsymbol{s}=\mathbf{1}$ and $\chi_i=\boldsymbol{e}_i$, where $\boldsymbol{e}_i$ denotes the i-th Euclidean canonical basis vector. Then, it follows that*

$$u_i = \langle\boldsymbol{u}\rangle - \frac{1}{n}((A+\boldsymbol{e}_i\boldsymbol{e}_i^T)^{-1}\mathbf{1},A\boldsymbol{u}) \tag{3.4}$$

in which $\langle\boldsymbol{u}\rangle := \frac{1}{n}\sum_{i=1}^n u_i$ denotes the average value of $\boldsymbol{u}$.

Next, denote by B_m the restriction of B to a subgraph $\mathcal{G}_m$, and set $A_m := B_m^T B_m$. Then the l-th component u_l of $\boldsymbol{u}$ satisfies

$$u_l = \langle \boldsymbol{u} \rangle_m + \frac{1}{|\mathcal{V}_m|}(B_m(A_m + \boldsymbol{e}_l \boldsymbol{e}_l^T)^{-1}\mathbf{1}_m, B_m \mathcal{I}_m \boldsymbol{u}), \tag{3.5}$$

where $l = 1 \dots |\mathcal{V}_m|$ are the local indices of the vertex set $\mathcal{V}_m$. The operator $\langle \cdot \rangle_m$ and the term $\mathbf{1}_m$ are the averaging operator and the constant vector restricted to the subgraph $\mathcal{G}_m$, and $\mathcal{I}_m : \mathbb{R}^{|\mathcal{V}|} \mapsto \mathbb{R}^{|\mathcal{V}_m|}$ maps the global edge indices to the local edge indices.

Applying this formula to $\mathcal{G}_i$ and $\mathcal{G}_j$ gives the row of the operator Π for the edge k that connects $\mathcal{G}_i$ and $\mathcal{G}_j$ as follows

$$(\Pi)_k = C_i^T \mathcal{I}_i^T + \boldsymbol{e}_k^T - C_j^T \mathcal{I}_j^T. \tag{3.6}$$

Here C_i is given by

$$C_i = \frac{1}{|\mathcal{V}_i|} B_i (A_i + \boldsymbol{e}_i \boldsymbol{e}_i^T)^{-1} \mathbf{1}_i$$

which then makes the summation in (3.6) valid. The vector C_j is defined in a similar way.

Assume that the global indices of the vertices in $\mathcal{G}_i$ and $\mathcal{G}_j$ are ordered consecutively as decreasing integers starting at $k-1$ and increasing integers starting at $k+1$, respectively. Then, the k-th row of Π can be expressed as

$$(\Pi)_k = \big[0, \dots, 0, C_i^T, 1, C_j^T, 0, \dots, 0\big] \tag{3.7}$$

where the number 1 is in the k-th position. Note that from (3.5) it follows that the property (3.3) holds for this construction of Π, since by definition of Q we have

$$\begin{aligned}
(BQ\boldsymbol{v})_k &= \langle \boldsymbol{v} \rangle_i - \langle \boldsymbol{v} \rangle_j \\
&= v_i - v_j + \frac{1}{|\mathcal{V}_i|}(B_i(A_i + \boldsymbol{e}_i \boldsymbol{e}_i^T)^{-1}\mathbf{1}_i, B_i \mathcal{I}_i \boldsymbol{v}) - \frac{1}{|\mathcal{V}_j|}(B_j(A_j + \boldsymbol{e}_j \boldsymbol{e}_j^T)^{-1}\mathbf{1}_j, B_j \mathcal{I}_j \boldsymbol{v}) \\
&= (\boldsymbol{e}_k, B\boldsymbol{v}) + (C_i, B_i \mathcal{I}_i \boldsymbol{v}) + (C_j, B_j \mathcal{I}_j \boldsymbol{v}) \\
&= (\Pi)_k B \boldsymbol{v},
\end{aligned}$$

where $k = (i, j)$, and i and j, both in local indices, are the incident vertices of k.

4. A two-level method

In this section, the ℓ^2-orthogonal projection given in (3.1) based on a matching $\mathcal{M}$ is proven to be stable assuming the maximum degree of the graph $\mathcal{G}$ is bounded. Then, a two-level preconditioner is derived and the condition number of the system preconditioned by this two-level method is proven to be uniformly bounded (under the same assumption).

4.1. **Two-level stability.** The construction of Π for a matching $\mathcal{M}$ proceeds as follows. First, note that all rows of Π that correspond to an edge $k = (i, j) \in \mathcal{M}$ are identically zero. On the other hand, if the edge $k = (i, j) \notin \mathcal{M}$, then it is an external edge and, thus, by (3.7), the k-th row of Π is

$$\begin{aligned}
(\Pi)_k &= \Bigg[0, \dots, 0, \quad \frac{1}{2}(1, -1)\left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}^T\right)^{-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\
&\qquad 1, \quad -\frac{1}{2}(1, -1)\left(\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}^T\right)^{-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad 0, \dots, 0\Bigg] \\
&= \big[0, \dots, 0, \frac{1}{2}, 1, -\frac{1}{2}, 0, \dots, 0\big].
\end{aligned}$$

Hence,

$$(\Pi)_{kl} = \begin{cases} 1 & k \notin \mathcal{M} \text{ and } l = k; \\ \pm\frac{1}{2} & k \notin \mathcal{M}, l \in \mathcal{M} \text{ and } l \cap k \neq \emptyset; \\ 0 & \text{elsewhere.} \end{cases} \tag{4.1}$$

The alternative way of describing the entries in Π is by showing that,

$$(\Pi)_{kl} = \begin{cases} 1 & l \notin \mathcal{M} \text{ and } k = l; \\ \pm\frac{1}{2} & l \in \mathcal{M}, k \notin \mathcal{M} \text{ and } k \cap l \neq \emptyset; \\ 0 & \text{elsewhere.} \end{cases} \tag{4.2}$$

Formula (4.1) implies that, the k-th row of Π can be a zero row if $k \in \mathcal{M}$, or a row with 3 non-zero entries if $k \notin \mathcal{M}$, which results in

$$\|\Pi\|_\infty = \max_k \sum_l |\Pi_{kl}| = 1 + |\pm 1/2| + |\pm 1/2| = 2.$$

Formula (4.2) implies that, the l-th column of Π can have exactly 1 non-zero entry if $l \notin \mathcal{M}$, or s non-zeros entries whose values are $\pm 1/2$ if $l \in \mathcal{M}$. Here s is the number of edges satisfying $k \notin \mathcal{M}$ and $k \cap l \neq \emptyset$ for any given $l \in \mathcal{M}$, thus is bounded by $2d-2$, where d is the maximum degree of the graph, since an edges can have at most $2d-2$ neighboring edges. This leads to

$$\|\Pi\|_1 = \max_l \sum_k |\Pi_{kl}| = \max\big(1, (2d-2)|\pm 1/2|\big) = \max\big(1, d-1\big).$$

On a graph whose maximal degree is larger or equal to 2, the estimates on the infinity norm and ℓ^1-norm of Π result in the following estimate on $\rho(\Pi\Pi^T)$:

$$\rho(\Pi\Pi^T) = \|\Pi\|_2^2 \leq \|\Pi\|_1 \|\Pi\|_\infty = 2d-2. \tag{4.3}$$

Remark 4.1. *Applying Gerschgorin's theorem directly to the matrix $\Pi\Pi^T$ leads to the sharper estimate: $\rho(\Pi\Pi^T) \leq d$.*

Formula (4.3) implies directly the following lemma.

Lemma 4.2. *On any graph whose maximum degree is 2 (e.g. such graph is a path), the operator Π defined in* (4.1) *satisfies $\Pi_k(B\boldsymbol{v}) = (BQ\boldsymbol{v})_k$ and the following estimate holds*

$$|Q|_A^2 \leq \|\Pi\|_2^2 \leq 2d-2 = 2.$$

Numerical tests show that this is a sharp estimate on the semi-norm of $|Q|_A$, leading to reliable AMG methods and fast convergence.

4.2. **A two-level preconditioner.** Here, using an estimate of the stability of the matching projection (i.e, the norm $|Q|_A$, where Q is defined via the matching) two-level convergence is established. Assume that for a graph Laplacian $A : \mathbb{R}^n \mapsto \mathbb{R}^n$ a perfect matching is given and consider the $n \times n/2$ matrix P whose k-th column is given by

$$(P)_k = \boldsymbol{e}_{i_k} + \boldsymbol{e}_{j_k}, \tag{4.4}$$

where $k = 1, ..., n/2$ and (i_k, j_k) is the k-th edge in $\mathcal{M}$. Further, define Q to be the ℓ^2-projection from $\mathbb{R}^n$ to $\{Pv | v \in \mathbb{R}^{n/2}\}$, i.e.,

$$Q = P(P^TP)^{-1}P^T.$$

Similar to the definition of P, define Y as the $n \times n/2$ matrix whose columns are given by

$$(Y)_k = \boldsymbol{e}_{i_k} - \boldsymbol{e}_{j_k}, \tag{4.5}$$

where $k = 1, ..., n/2$ and (i_k, j_k) is the k-th edge in $\mathcal{M}$. Then, the matrix (Y, P) is orthogonal and the columns of Y and P form a hierarchical bases, which can be used to relate the two-level method to a block factorization as follows.

Given A, P, and Y, define

$$\widehat{A} = (Y, P)^T A (Y, P) = \begin{pmatrix} Y^T AY & Y^T AP \\ P^T AY & P^T AP \end{pmatrix}.$$

A direct calculation then shows that

$$\widehat{A} = L \begin{pmatrix} Y^T AY & 0 \\ 0 & S \end{pmatrix} L^T, \tag{4.6}$$

where

$$S = P^T AP - P^T AY (Y^T AY)^{-1} Y^T AP \tag{4.7}$$

is the Schur complement and

$$L = \begin{pmatrix} I & 0 \\ P^T AY (Y^T AY)^{-1} & I \end{pmatrix}. \tag{4.8}$$

Next, define $\mathcal{G}_c$ as the unweighted coarse graph and denote by A_c the graph Laplacian of $\mathcal{G}_c$. In contrast to most of the existing AMG methods, here $A_c \neq P^T AP$, except in special cases, e.g., for 1 dimensional problems. Let, σ be a positive constant such that

$$\sigma = \sup_{\boldsymbol{v}:(\boldsymbol{v},\mathbf{1})=0} \frac{(AP\boldsymbol{v}, P\boldsymbol{v})}{(A_c\boldsymbol{v}, \boldsymbol{v})}. \tag{4.9}$$

Then, the fact that all weights in the graph corresponding to $P^T AP$ are larger than or equal to one implies $(AP\boldsymbol{v}, P\boldsymbol{v}) \geq (A_c\boldsymbol{v}, \boldsymbol{v}), \forall \boldsymbol{v}$, and

$$\frac{(\sigma A_c\boldsymbol{v}, \boldsymbol{v})}{(AP\boldsymbol{v}, P\boldsymbol{v})} \in [1, \sigma], \quad \forall \boldsymbol{v} : (\boldsymbol{v}, \mathbf{1}) = 0.$$

Consider the two-level preconditioner $\widehat{G}$ which uses the coarse graph Laplacian A_c by

$$\widehat{G} = L \begin{pmatrix} Y^T AY & 0 \\ 0 & \sigma A_c \end{pmatrix} L^T. \tag{4.10}$$

Let M be a preconditioner for $Y^T AY$, and D be a preconditioner for the graph Laplacian A_c. Then, a two-level preconditioner $\widehat{B}$ is defined by

$$\widehat{B} = \widetilde{L} \begin{pmatrix} M(M + M^T - Y^T AY)^{-1} M^T & 0 \\ 0 & \sigma D \end{pmatrix} \widetilde{L}^T, \tag{4.11}$$

where

$$\widetilde{L} = \begin{pmatrix} I & 0 \\ P^T AY M^{-1} & I \end{pmatrix}.$$

As observed in [11] and [27], this gives a block matrix representation of the two-level method

$$\begin{aligned} I - (Y, P)\widehat{G}^{\dagger}(Y, P)^T A &= (I - Y(Y^T AY)^{-1} Y^T A)(I - P(\sigma A_c)^{\dagger} P^T A)(I - Y(Y^T AY)^{-1} Y^T A) \\ I - (Y, P)\widehat{B}^{\dagger}(Y, P)^T A &= (I - YM^{-T} Y^T A)(I - P(\sigma D)^{\dagger} P^T A)(I - YM^{-1} Y^T A), \end{aligned}$$

where the pseudo-inverse operator denoted by † is used since the graph Laplacian is semi-definite. The convergence of the two-level method can now be estimated by comparing $\widehat{A}$ and the preconditioner $\widehat{B}$.

The remainder of this section is dedicated to establishing a spectral equivalence between $\widehat{A}$ and $\widehat{B}$ for the two-level matching algorithm. The proof uses the following Lemma.

Lemma 4.3. *For any $\boldsymbol{x} \in \mathbb{R}^{n/2}$ the Schur complement S as given in (4.7) satisfies*

$$(S\boldsymbol{x}, \boldsymbol{x}) = \inf_{\boldsymbol{w}} \big(A(Y\boldsymbol{w} + P\boldsymbol{x}), (Y\boldsymbol{w} + P\boldsymbol{x})\big). \tag{4.12}$$

Proof. Note that

$$\begin{aligned} \big(AY(Y^TAY)^{-1}Y^TAP\boldsymbol{x}, P\boldsymbol{x}\big) &= \big(AY(Y^TAY)^{-1}Y^TAP\boldsymbol{x}, Y(Y^TAY)^{-1}Y^TAP\boldsymbol{x}\big) \\ &= \|Y(Y^TAY)^{-1}Y^TAP\boldsymbol{x}\|_A^2, \end{aligned}$$

because here, $Y(Y^TAY)^{-1}Y^TAP\boldsymbol{x}$ is the A orthogonal projection of $P\boldsymbol{x}$ onto the space spanned by the columns of Y and, thus, minimizes the distance (in A norm) between $P\boldsymbol{x}$ and this space. Hence,

$$\begin{aligned} (S\boldsymbol{x}, \boldsymbol{x}) &= \|P\boldsymbol{x}\|_A^2 - \|Y(Y^TAY)^{-1}Y^TAP\boldsymbol{x}\|_A^2 \\ &= \inf_{\boldsymbol{w}} \big(A(Y\boldsymbol{w} + P\boldsymbol{x}), (Y\boldsymbol{w} + P\boldsymbol{x})\big) \end{aligned}$$

□

Let $\widehat{\mathbf{1}}$ be a vector satisfying $(Y, P)\widehat{\mathbf{1}} = \mathbf{1}$, then the following lemma now holds.

Lemma 4.4. *Let $c_g = \sigma|Q|_A^2$, where σ is defined as in (4.9). Then for any $\boldsymbol{v}$, such that $(\boldsymbol{v}, \widehat{\mathbf{1}}) = 0$, we have*

$$\frac{(\widehat{G}\boldsymbol{v}, \boldsymbol{v})}{(\widehat{A}\boldsymbol{v}, \boldsymbol{v})} \in [1, c_g] \tag{4.13}$$

where $\widehat{A}$ and $\widehat{G}$ are defined in (4.6) and (4.10), respectively.

Proof. First we have

$$(AP\boldsymbol{x}, P\boldsymbol{x}) \geq \inf_{\boldsymbol{w}} \big(A(Y\boldsymbol{w} + P\boldsymbol{x}), (Y\boldsymbol{w} + P\boldsymbol{x})\big).$$

Next, using Lemma 4.3, we find

$$\begin{aligned} \frac{(AP\boldsymbol{x}, P\boldsymbol{x})}{(S\boldsymbol{x}, \boldsymbol{x})} &= \frac{(AP\boldsymbol{x}, P\boldsymbol{x})}{\inf_{\boldsymbol{w}} \big(A(Y\boldsymbol{w} + P\boldsymbol{x}), (Y\boldsymbol{w} + P\boldsymbol{x})\big)} \\ &= \sup_{\boldsymbol{w}} \frac{(AP\boldsymbol{x}, P\boldsymbol{x})}{\big(A(Y\boldsymbol{w} + P\boldsymbol{x}), (Y\boldsymbol{w} + P\boldsymbol{x})\big)} \\ &= \sup_{\boldsymbol{u} = Y\boldsymbol{w} + P\boldsymbol{x}} \frac{(AQ\boldsymbol{u}, Q\boldsymbol{u})}{(A\boldsymbol{u}, \boldsymbol{u})} \leq \sup_{\boldsymbol{u}} \frac{(AQ\boldsymbol{u}, Q\boldsymbol{u})}{(A\boldsymbol{u}, \boldsymbol{u})} = |Q|_A^2. \end{aligned}$$

Note that the only difference between the preconditioners $\widehat{G}$ and $\widehat{A}$ is that the former matrix uses σA_c, whereas the latter uses S to define the 2-2 block. The spectral equivalence relation between the operators σA_c and S can be derived from

$$\begin{aligned} \inf_{\boldsymbol{u}} \frac{\sigma(A_c\boldsymbol{u}, \boldsymbol{u})}{(AP\boldsymbol{u}, P\boldsymbol{u})} \inf_{\boldsymbol{v}} \frac{(AP\boldsymbol{v}, P\boldsymbol{v})}{(S\boldsymbol{v}, \boldsymbol{v})} &\leq \frac{\sigma(A_c\boldsymbol{w}, \boldsymbol{w})}{(S\boldsymbol{w}, \boldsymbol{w})} \\ &\leq \sup_{\boldsymbol{u}} \frac{\sigma(A_c\boldsymbol{u}, \boldsymbol{u})}{(AP\boldsymbol{u}, P\boldsymbol{u})} \sup_{\boldsymbol{v}} \frac{(AP\boldsymbol{v}, P\boldsymbol{v})}{(S\boldsymbol{v}, \boldsymbol{v})}, \quad \forall \boldsymbol{w} : (\boldsymbol{w}, \mathbf{1}) = 0, \end{aligned}$$

which implies

$$\frac{\sigma(A_c\boldsymbol{w}, \boldsymbol{w})}{(S\boldsymbol{w}, \boldsymbol{w})} \in [1, \sigma|Q|_A^2], \qquad \forall \boldsymbol{w} : (\boldsymbol{w}, \mathbf{1}) = 0.$$

Hence, for any $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $(\boldsymbol{y}, \mathbf{1}) = 0$ we have

$$\frac{\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^T \begin{pmatrix} Y^TAY & 0 \\ 0 & \sigma A_c\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}}{\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^T \begin{pmatrix} Y^TAY & 0 \\ 0 & S\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}} = \frac{(AY\boldsymbol{x}, Y\boldsymbol{x}) + \sigma(A_c\boldsymbol{y}, \boldsymbol{y})}{(AY\boldsymbol{x}, Y\boldsymbol{x}) + (S\boldsymbol{y}, \boldsymbol{y})} \in [1, \sigma|Q|_A^2],$$

which shows that (4.13) holds for all $\boldsymbol{v}$ such that $(\boldsymbol{v}, \widehat{\mathbf{1}}) = 0$ since L given by (4.8) is nonsingular. □

Since the application of the two-level preconditioner $\widehat{G}$ requires exact solves with Y^TAY and the graph Laplacian A_c, the convergence rate of a method that uses $\widehat{B}$ which is defined by replacing these exact solves with approximate ones is of interest. Combining Lemma 4.4 and the two-level convergence estimate (Theorem 4.2 in [11]) yields the following result.

Theorem 4.5. *If the preconditioners M and D are spectrally equivalent to Y^TAY and A_c such that*

$$\frac{\left((M^T + M - Y^TAY)^{-1}M\boldsymbol{u}, M\boldsymbol{u}\right)}{(AY\boldsymbol{u}, Y\boldsymbol{u})} \in [1, \kappa_s] \quad \textit{and} \quad \frac{(D\boldsymbol{w}, \boldsymbol{w})}{(A_c\boldsymbol{w}, \boldsymbol{w})} \in [1, \eta], \quad \forall \boldsymbol{u}, \boldsymbol{w} : (\boldsymbol{w}, \mathbf{1}) = 0,$$

then

$$\frac{(\widehat{B}\boldsymbol{v}, \boldsymbol{v})}{(\widehat{A}\boldsymbol{v}, \boldsymbol{v})} \in [1, (\kappa_s + \sigma\eta - 1)|Q|_A^2], \qquad \forall \boldsymbol{v} : (\boldsymbol{v}, \widehat{\mathbf{1}}) = 0. \tag{4.14}$$

Note that this estimate reduces to (4.13) when $M = Y^TAY$ and $D = A_c$.

4.3. **Convergence estimate for matching.** In the following we will show the sharpness of the estimate provided by Theorem 4.5 for the case when the graph Laplacian corresponds to a structured grid, and the coarse space is given by aligned matching.

Define an m-dimensional hypercubic grid and the related graph Laplacian $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ such that the following conditions are satisfied.

(1) A vertex $i_{\boldsymbol{v}} \in \mathcal{V}$ corresponds to a vector $\boldsymbol{v} \in \mathbb{R}^m$, and $(\boldsymbol{v}, \boldsymbol{e}_j) \in [1, 2, \ldots, s_j]$, $j = 1, 2, \ldots, m$. Here e_j is an Euclidean basis and $s_1, s_2, \ldots, s_m$ are given positive integers that represent the numbers of vertices along all dimensions.
(2) An edge $k = (i_{\boldsymbol{u}}, i_{\boldsymbol{v}})$ is in the edge set $\mathcal{E}$ if and only if $\boldsymbol{u} - \boldsymbol{v} = \pm\boldsymbol{e}_j$ and $j \in [1, 2, \ldots, m]$.[1]

Then we have the following estimate for the energy norm $|Q|_A$.

Lemma 4.6. *Let $\mathcal{G}$ be an m-dimensional hypercubic grid and $k \in [1, 2, \ldots, m]$ a fixed dimension. Assume that s_k is an even number. The matching along the k-th dimension is defined as*

$$\mathcal{M} = \{l = (i_{\boldsymbol{v}}, i_{\boldsymbol{v}+\boldsymbol{e}_k}) | v \in \mathcal{V}, \textit{ and } (v, e_k) \textit{ is an odd number } \}.$$

Let Q be the ℓ^2-projection defined in (3.1) *resulting from the matching $\mathcal{M}$. Then $|Q|_A \leq 2$.*

Proof. Define the set Ω be the collection of all edges along the k-th dimension, as

$$\Omega = \{l = (i_{\boldsymbol{u}}, i_{\boldsymbol{v}}) | \boldsymbol{v} - \boldsymbol{u} = \boldsymbol{e}_k\}.$$

Also define $\overline{\Omega} = \mathcal{E} \setminus \Omega$ and the graph Laplacians A_Ω and $A_{\overline{\Omega}}$, derived from Ω and $\overline{\Omega}$ respectively.

The graphs in the set Ω are paths, whose maximum degree is 2, and $\mathcal{M} \subset \Omega$ is a matching on these paths. Therefore by Lemma 4.2 it is true that

$$(A_\Omega Q\boldsymbol{u}, Q\boldsymbol{u}) \leq 2(A_\Omega \boldsymbol{u}, \boldsymbol{u}). \tag{4.15}$$

[1]This paper deals with undirected graphs only, the notation $k = (i_{\boldsymbol{u}}, i_{\boldsymbol{v}})$, however, can also be used for edges in directed graphs.

On the other hand, the matching is aligned on the set $\overline{\Omega}$, meaning that any two matched pairs are connected through 0 or 2 edges in $\overline{\Omega}$, thus the edges in set $\overline{\Omega}$ can be subdivided into many sets of edges of the same type, one of which is shown in Fig. 3. Notice that in this figure, the edges (i,k)

Figure 3: Matching $\mathcal{M}$ on a subset of $\overline{\Omega}$

and (j,l) are in $\overline{\Omega}$, while (i,j) and (k,l) are in $\mathcal{M}$. Using the definition of Q, the energy norm of Q is estimated on the the subset of $\overline{\Omega}$ indicated by Fig. 3, by

$$2\left(\frac{u_i+u_j}{2}-\frac{u_k+u_l}{2}\right)^2=\frac{1}{2}\left((u_i-u_k)+(u_j-u_l)\right)^2\leq(u_i-u_k)^2+(u_j-u_l)^2.$$

This implies that

$$(A_{\overline{\Omega}}Q\boldsymbol{u},Q\boldsymbol{u})\leq(A_{\overline{\Omega}}\boldsymbol{u},\boldsymbol{u}). \tag{4.16}$$

Combining (4.15) and (4.16) results in $(AQ\boldsymbol{u},Q\boldsymbol{u})\leq 2(A\boldsymbol{u},\boldsymbol{u})$, which proves that $|Q|_A\leq 2$. □

Remark 4.7. *A similar estimate follows for aligned partitionings consisting of line segments of size m. Namely, in this case it can be shown that $|Q|_A^2\leq m$ holds. Comparing this result with the result from Lemma 4.6, however, already suggests that using a more shape regular partitioning rather than one consisting of lines is more appropriate since this results in smaller values of the semi-norm $|Q|_A$. These estimates also suggest the construction of AMLI methods where certain Chebyshev polynomials are used to stabilize the condition number in the multilevel setting.*

A bound on the constant κ_s follows by using the fact that Y^TAY is well conditioned and its condition number depends on the degree of the graph, but not on its size.

Lemma 4.8. *Let $\mathcal{M}$ be the perfect matching in a graph whose maximum degree is d, and let S be defined as in* (4.5)*, then we have*

$$\frac{(AY\boldsymbol{w},Y\boldsymbol{w})}{(\boldsymbol{w},\boldsymbol{w})}\in[4,2d],\quad\forall\boldsymbol{w}\neq 0.$$

Proof. The A-norm of the vector $Y\boldsymbol{w}$ is computed by definition:

$$(AY\boldsymbol{w},Y\boldsymbol{w})\geq\sum_{k=(i,j)\in\mathcal{M}}\left((Y\boldsymbol{w})_i-(Y\boldsymbol{w})_j\right)^2=\sum_{k=(i,j)\in\mathcal{M}}\left((Y\boldsymbol{w})_i+(Y\boldsymbol{w})_i\right)^2=4\boldsymbol{w}^T\boldsymbol{w}.$$

We also have

$$\rho(Y^TAY)\leq\|Y^TAY\|_1\leq\|Y^T\|_1\|A\|_\infty\|Y\|_\infty=2d.$$ □

From the Lemma it follows that for any $\epsilon>0$ there exists a smoother M such that the bound on the constant κ_s in Theorem 4.5 is

$$\kappa_s\leq 1+\epsilon.$$

This result in turn implies that an efficient solver for Y^TAY can be constructed by applying a few conjugate gradient iterations with an overall cost that is linear with respect to the size of Y^TAY.

The constant σ in (4.9) can be estimated by checking the weights of the weighted graph Laplacian P^TAP. Taking any two distinct subgraphs (edges) in the matching, say the k-th and l-th such that $k\neq l$, it follows that the corresponding entry $(P^TAP)_{kl}$ is equal to the number of exterior

edges that connect these subgraphs. For an aligned matching in a fixed direction in a hypercubic grid, these weights are bounded by 2. For any general graph A, the weights in P^TAP are bounded by 4, since there are at most 4 distinct edges that connect to any other 2 distinct edges. Then, denoting by A_c the *unweighted* graph Laplacian on the graph defined by P^TAP, and noting that all off-diagonal entries of A_c are equal to -1, it follows that

$$\sigma = \begin{cases} 2 & \text{for an aligned matching on a hypercubic grid of any dimension;} \\ 4 & \text{for a given matching on any graph.} \end{cases}$$

Remark 4.9. *These estimates can be generalized to other subgraph partitionings in a similar way. As an example, consider again a graph for a hypercubic grid of any dimension. Then, for line aggregates of size m (aligned with the grid) the following estimates hold*

$$|Q|_A^2 \le m, \quad \kappa_s \le 1+\epsilon, \quad \eta = 1, \quad \sigma \le m.$$

Such estimates give insight into the design of a nearly optimal multilevel method. Moreover, the bounds are sharp enough, namely, the corresponding multilevel method can be proven to have convergence rate $\approx (1 - 1/\log n)$ and $O(n \log n)$ complexity.

5. Algebraic multilevel iteration (AMLI) based on matching

In this section, a multilevel method that uses recursively the two-level matching methods from Section 4.2 in combination with a polynomial stabilization, also known as Algebraic Multilevel Iteration (AMLI) cycle is analyzed. The aim is to construct a method with low (nearly optimal) computational complexity and (nearly) optimal convergence rate.

5.1. **Multilevel hierarchy.** Assume that $A_J = A$ is an $n \times n$ graph Laplacian matrix where $n = 2^J$. For $k = 1, \ldots, J$ define the matching $\mathcal{M}_k$ and the prolongation operator P_k according to (4.4), then compute the graph Laplacian A_k of the coarse graph $\mathcal{G}_k$. Recall that $A_{k-1} \neq P_k^T A_k P_k$. The index k starts at 1 because the analysis is simpler if the coarsest graph has more than 1 vertex. Also, define Y_k and L_k for A_k as in (4.5) and (4.8), and let the two-level preconditioner $\widehat{G}_k$ on each level k be given by

$$\widehat{G}_k = L_k \begin{pmatrix} Y_k^T A_k Y_k & 0 \\ 0 & \sigma A_{k-1} \end{pmatrix} L_k^T, \qquad k = 2, \ldots, J.$$

Then an AMLI preconditioner is defined recursively by

$$\begin{aligned} B_1^{-1} &= A_1^{\dagger}, \\ \widehat{B}_k^{-1} &= L_k^{-T} \begin{pmatrix} (Y_k^T A_k Y_k)^{-1} & 0 \\ 0 & \sigma^{-1} B_{k-1}^{-1} q_{k-1}(A_{k-1} B_{k-1}^{-1}) \end{pmatrix} L_k^{-1}, \quad k = 2, \ldots, J, \\ B_k^{-1} &= (Y_k, P_k)^T \widehat{B}_k^{-1} (Y_k, P_k), \qquad k = 2, \ldots, J, \end{aligned}$$

where $q_k(t)$ is a polynomial that determines a special coarse level correction on the k-th level.

In the remainder of this section, sufficient conditions for guaranteeing the spectral equivalence between the multilevel preconditioner B_J, as defined above, and the graph Laplacian A are derived. We first prove two auxiliary results, which are needed in the analysis thereafter.

Proposition 5.1. *Let $A : V \mapsto V$ and $G : V \mapsto V$ be symmetric positive semidefinite operators on a finite dimensional real Hilbert space V. Suppose that the following spectral equivalence holds:*

$$c_0(A\boldsymbol{v}, \boldsymbol{v}) \le (G\boldsymbol{v}, \boldsymbol{v}) \le c_1(A\boldsymbol{v}, \boldsymbol{v}), \quad c_0 > 0, \quad c_1 > 0. \tag{5.1}$$

Then, we also have that

$$c_1^{-1}(A^{\dagger}\boldsymbol{v}, \boldsymbol{v}) \le (G^{\dagger}\boldsymbol{v}, \boldsymbol{v}) \le c_0^{-1}(A^{\dagger}\boldsymbol{v}, \boldsymbol{v}). \tag{5.2}$$

Proof. Observe that the spectral equivalence given in (5.1) implies that A and G have the same null-space (and also the same range, because they are symmetric). Also, note that, if $\boldsymbol{v}$ is in this null space, then (5.2) trivially holds. Thus, without loss of generality, we restrict our considerations below to the range of G and A.

After change of variables $\boldsymbol{w} = A^{1/2}\boldsymbol{v}$ from the upper bound in (5.1) we may conclude that

$$\frac{\|G^{1/2}\big(A^\dagger\big)^{1/2}\boldsymbol{w}\|^2}{\|\boldsymbol{w}\|^2} \leq c_1, \quad \text{and hence,} \quad \|G^{1/2}\big(A^\dagger\big)^{1/2}\|^2 \leq c_1.$$

Since $G^{1/2}\big(A^\dagger\big)^{1/2} = \Big(\big(A^\dagger\big)^{1/2}G^{1/2}\Big)^T$, we obtain that $\|G^{1/2}\big(A^\dagger\big)^{1/2}\| = \|\big(A^\dagger\big)^{1/2}G^{1/2}\|$. Using this identity, the estimate above, we have for all $\boldsymbol{u}$ and all $\boldsymbol{w} = [G^\dagger]^{1/2}\boldsymbol{u}$:

$$c_1 \geq \|\big(A^\dagger\big)^{1/2}G^{1/2}\|^2 \geq \frac{\|\big(A^\dagger\big)^{1/2}G^{1/2}\boldsymbol{w}\|^2}{\|\boldsymbol{w}\|^2}, \quad \text{and hence,} \quad c_1 \geq \frac{\|\big(A^\dagger\big)^{1/2}\boldsymbol{u}\|^2}{\|\big(G^\dagger\big)^{1/2}\boldsymbol{u}\|^2}.$$

The estimate above implies that $c_1^{-1}(A^\dagger\boldsymbol{u},\boldsymbol{u}) \leq (G^\dagger\boldsymbol{u},\boldsymbol{u})$ and thus the lower bound in (5.2). The upper bound follows from repeating basically the same steps with the roles of G and A interchanged. □

The elementary results in the next proposition are used later in the proof of Lemma 5.5.

Proposition 5.2. *Let $\theta \in [0,1]$ and define $q(t;\theta) = \dfrac{4}{\theta+1}(1 - \dfrac{t}{\theta+1})$ and $\widetilde{q}(t;\theta) = tq(t;\theta)$. Then,*

(i) $\max\limits_{t\in[\theta,1]} \widetilde{q}(t;\theta) = 1$;

(ii) $\min\limits_{t\in[\theta,1]} \widetilde{q}(t;\theta) = \widetilde{q}(\theta;\theta) = \widetilde{q}(1;\theta)$;

(iii) $\dfrac{d\widetilde{q}(1;\theta)}{d\theta} \geq 0$ *(monotonicity).*

Proof. The proofs of (i) and (ii) follow from the identity $\widetilde{q}(t;\theta) = 1 - \big(2t/(\theta+1) - 1\big)^2$. The proof of (iii) is also straightforward and follows from the fact that $\theta \in [0,1]$ and hence

$$\frac{d\widetilde{q}(1;\theta)}{d\theta} = \frac{4}{(\theta+1)^2}\left(\frac{2}{\theta+1} - 1\right) \geq 0.$$ □

Next we derive estimates for the growth of the terms in a sequence, recursively defined using $\widetilde{q}(1;\theta)$, which we use later to bound the convergence rate.

Proposition 5.3. *Let, $1 \leq c \leq 4$ be a given constant. Further let $q(t;\theta) = \dfrac{4}{\theta+1}(1 - \dfrac{t}{\theta+1})$ and $\widetilde{q}(t;\theta) = tq(t;\theta)$ (as in Proposition 5.2). Define*

$$\theta_1 = 1; \qquad \theta_{k+1} = \frac{1}{c}\widetilde{q}(1;\theta_k), \quad \text{for} \quad k = 1,2,\ldots \tag{5.3}$$

Then, the following relations are true for $k = 1,2,\ldots$:

(i) $\dfrac{2}{\sqrt{c}} - 1 \leq \theta_{k+1} \leq \theta_k \leq 1$;

(ii) $\theta_k \geq \max\left\{\dfrac{2}{\sqrt{c}} - 1, \dfrac{1}{2k + \ln k}\right\}$.

Proof. The first item (i) follows from algebraic manipulations and the estimates given in Proposition 5.2. To show that $\theta_{k+1} \leq \theta_k$, we assume that $\theta_k \geq 2/\sqrt{c} - 1$ (which is certainly true for

$k=1$). To prove that $\theta_{k+1} \geq 2/\sqrt{c}-1$ we observer that from $\theta_k \geq 2/\sqrt{c}-1$, the monotonicity property (iii) in Proposition 5.2 implies that

$$\theta_{k+1} = \frac{1}{c}\tilde{q}(1;\theta_k) \geq \frac{1}{c}\tilde{q}\left(1;\frac{2}{\sqrt{c}}-1\right) = \frac{2}{\sqrt{c}}-1.$$

Using again that $\theta_k \geq 2/\sqrt{c}-1$ gives also that

$$\theta_{k+1}-\theta_k = \frac{\theta_k}{(\theta_k+1)^2}\left(\frac{4}{c}-(\theta_k+1)^2\right) \leq 0.$$

The proof of the second item (ii) is a bit more involved. We prove this item by deriving an upper bound on $\zeta_k = \frac{1}{\theta_k}$. Observe that, from the recurrence relation for θ_k we have

$$\zeta_{k+1} = \frac{c}{4}(\zeta_k + 2 + \frac{1}{\zeta_k}), \qquad \zeta_1 = 1. \tag{5.4}$$

We first show that the sequence above is growing fastest for $c=4$. Indeed, let

$$s_{k+1} = s_k + 2 + \frac{1}{s_k}, \qquad s_1 = 1. \tag{5.5}$$

A standard induction argument shows that

$$\zeta_k \leq s_k, \qquad \text{and} \qquad 2k-1 \leq s_k, \qquad \forall k.$$

Expanding s_k by using the recurrence formula (5.5), and using the estimate $\sum_{i=1}^{k-1}\frac{1}{2i-1} \leq \ln k + 1$ one obtains

$$s_k = s_1 + 2(k-1) + \sum_{i=1}^{k-1}\frac{1}{s_i} \leq 1 + 2(k-1) + \sum_{i=1}^{k-1}\frac{1}{2i-1} \leq 2k + \ln k,$$

which provides an upper bound for ζ_k. Hence $1/(2k+\ln k)$ is a lower bound for θ_k. □

The following Lemma provides a spectral equivalence relation between $\widehat{G}_k^{\dagger}$ and $\widehat{B}_k^{-1}$.

Lemma 5.4. *If $\lambda_1 \leq \lambda(B_k^{-1}A_k) \leq \lambda_2$ and $tq_k(t) > 0$ for $\lambda_1 \leq t \leq \lambda_2$, then*

$$\min\{1, \min_{\lambda_1\leq t\leq\lambda_2} tq_k(t)\} \leq \frac{(\widehat{B}_{k+1}^{-1}\boldsymbol{v},\boldsymbol{v})}{(\widehat{G}_{k+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})} \leq \max\{1, \max_{\lambda_1\leq t\leq\lambda_2} tq_k(t)\}, \tag{5.6}$$
$$\forall \boldsymbol{v} : (\boldsymbol{v},\widehat{\mathbf{1}}) = 0, \quad k = 1,\ldots,J-1.$$

Proof. For any vector $\boldsymbol{v}$,

$$\frac{\big(q_k(A_kB_k^{-1})\boldsymbol{v},B_k^{-1}\boldsymbol{v}\big)}{(A_k^{\dagger}\boldsymbol{v},\boldsymbol{v})} = \frac{\big(q_k(A_k^{\frac{1}{2}}B_k^{-1}A_k^{\frac{1}{2}})(A_k^{\frac{1}{2}})^{\dagger}\boldsymbol{v},A_k^{\frac{1}{2}}B_k^{-1}A_k^{\frac{1}{2}}(A_k^{\frac{1}{2}})^{\dagger}\boldsymbol{v}\big)}{(A_k^{\dagger}\boldsymbol{v},\boldsymbol{v})} = \frac{\big(q_k(Z)\boldsymbol{w},Z\boldsymbol{w}\big)}{(\boldsymbol{w},\boldsymbol{w})},$$

where $\boldsymbol{w} = (A_k^{\frac{1}{2}})^{\dagger}\boldsymbol{v}$ and $Z = A_k^{\frac{1}{2}}B_k^{-1}A_k^{\frac{1}{2}}$. Further, since Z has the same eigenvalues as $B_k^{-1}A_k$, we conclude that

$$\min_{\lambda_1\leq t\leq\lambda_2} tq_k(t) \leq \frac{\big(q_k(A_kB_k^{-1})\boldsymbol{v},B_k^{-1}\boldsymbol{v}\big)}{(A_k^{\dagger}\boldsymbol{v},\boldsymbol{v})} \leq \max_{\lambda_1\leq t\leq\lambda_2} tq_k(t).$$

This implies that for any $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\begin{aligned}
&\frac{\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^T \begin{pmatrix}(Y_{k+1}^T A_{k+1} Y_{k+1})^{-1} & 0\\ 0 & \sigma^{-1} B_k^{-1} q_k(A_k B_k^{-1})\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}}{\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}^T \begin{pmatrix}(Y_{k+1}^T A_{k+1} Y_{k+1})^{-1} & 0\\ 0 & \sigma^{-1} A_k^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}}\\
= \;& \frac{\big((Y_{k+1}^T A_{k+1} Y_{k+1})^{-1}\boldsymbol{x},\boldsymbol{x}\big) + \sigma^{-1}(B_k^{-1} q(A_k B_k^{-1})\boldsymbol{y},\boldsymbol{y})}{\big((Y_{k+1}^T A_{k+1} Y_{k+1})^{-1}\boldsymbol{x},\boldsymbol{x}\big) + \sigma^{-1}(A_k^{-1}\boldsymbol{y},\boldsymbol{y})}\\
\in \;& \Big[\min\{1, \min_{\lambda_1\le t\le\lambda_2} tq(t)\}, \max\{1, \max_{\lambda_1\le t\le\lambda_2} tq(t)\}\Big],
\end{aligned}$$

and, hence, by using the definition of $\widehat{G}_k$ and $\widehat{B}_k^{-1}$, it follows that

$$\frac{(\widehat{B}_{k+1}^{-1}\boldsymbol{v},\boldsymbol{v})}{(\widehat{G}_{k+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})} \in \Big[\min\{1, \min_{\lambda_1\le t\le\lambda_2} tq(t)\}, \max\{1, \max_{\lambda_1\le t\le\lambda_2} tq(t)\}\Big]. \tag{5.7}$$

□

Combining the above lemma with Lemma (4.4) the spectral equivalence between B_k^{-1} and $A_k^{\dagger}$, $k = 1, \dots, J$ follows and is shown in the next Lemma.

Lemma 5.5. *Assume that the two level preconditioner G_k satisfies*

$$(\widehat{A}_k\boldsymbol{v},\boldsymbol{v}) \le (\widehat{G}_k\boldsymbol{v},\boldsymbol{v}) \le c_g(\widehat{A}_k\boldsymbol{v},\boldsymbol{v}), \quad \forall \boldsymbol{v} \text{ and } k = 2,\dots,J. \tag{5.8}$$

with constant c_g, such that $1 \le c_g \le 4$. Define

$$q_k(t) = q(t,\theta_k), \tag{5.9}$$

where θ_k are defined as

$$\theta_1 = 1; \qquad \theta_{k+1} = \frac{1}{c_g}\widetilde{q}(1;\theta_k) = \frac{t}{c_g}q_k(1).$$

Then, the following inequalities hold for all $\boldsymbol{v} : (\boldsymbol{v},\mathbf{1}) = 0$ and $k = 1,\dots,J$.

$$\theta_k \le \frac{(B_k^{-1}\boldsymbol{v},\boldsymbol{v})}{(A_k^{\dagger}\boldsymbol{v},\boldsymbol{v})} \le 1, \tag{5.10}$$

$$\max\left\{\frac{2}{\sqrt{c_g}} - 1, \frac{1}{2k + \ln k}\right\} \le \frac{(B_k^{-1}\boldsymbol{v},\boldsymbol{v})}{(A_k^{\dagger}\boldsymbol{v},\boldsymbol{v})}. \tag{5.11}$$

Proof. We give a proof of (5.10) by induction. Clearly, for $k = 1$, $B_1^{-1} = A_1^{\dagger}$, and hence, (5.10) holds. We assume that the inequalities (5.10) hold for $k = l$ and we aim to prove them for $k = l+1$. For all $\boldsymbol{v}$ such that $(\boldsymbol{v},\mathbf{1}) = 0$ we have

$$\frac{(\widehat{B}_{l+1}^{-1}\boldsymbol{v},\boldsymbol{v})}{(\widehat{A}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})} = \frac{(\widehat{G}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})}{(\widehat{A}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})}\frac{(\widehat{B}_{l+1}^{-1}\boldsymbol{v},\boldsymbol{v})}{(\widehat{G}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})}$$

Then, from (5.8), Proposition 5.1 and Lemma 5.4 (applied in that order) it follows that

$$\frac{1}{c_g} \le \frac{(\widehat{G}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})}{(\widehat{A}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})} \le 1, \quad \text{and} \quad \min\{1, \min_{t\in[\theta_k,1]} tq_k(t)\} \le \frac{(\widehat{B}_{l+1}^{-1}\boldsymbol{v},\boldsymbol{v})}{(\widehat{G}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})} \le \max\{1, \max_{t\in[\theta_k,1]} tq_k(t)\}.$$

Next, by Proposition 5.2 and Proposition 5.3 we find that

$$\theta_{l+1} = \frac{1}{c_g}\min\{1, \min_{t\in[\theta_k,1]} tq_l(t)\} \le \frac{(\widehat{B}_{l+1}^{-1}\boldsymbol{v},\boldsymbol{v})}{(\widehat{A}_{l+1}^{\dagger}\boldsymbol{v},\boldsymbol{v})} \le \max\{1, \max_{t\in[\theta_k,1]} tq_l(t)\} = 1.$$

Finally, from the definition of B_k^{-1} and $A_k^\dagger$ in terms of $\widehat{B}_k^{-1}$ and $\widehat{A}_k^\dagger$, it immediately follows that

$$\theta_k \le \frac{(B_k^{-1}\boldsymbol{v},\boldsymbol{v})}{(A_k^\dagger\boldsymbol{v},\boldsymbol{v})} = \frac{\big(\widehat{B}_k^{-1}(Y,P)\boldsymbol{v},(Y,P)\boldsymbol{v}\big)}{\big(\widehat{A}_k^\dagger(Y,P)\boldsymbol{v},(Y,P)\boldsymbol{v}\big)} \le 1, \quad (\boldsymbol{v},\mathbf{1}) = 0. \tag{5.12}$$

The proof of (5.11) follows from item (ii) in Proposition 5.3. □

The spectrum estimate (5.10) suggests that, B_J^{-1} can be used as a preconditioner in the a conjugate gradient method for solving a linear system whose coefficient matrix is A_J. It also leads to the following convergence estimate of a power method.

Theorem 5.6. *Assume that there is a constant c_g such that $1 \le c_g \le 4$ and $(\widehat{A}_k\boldsymbol{v},\boldsymbol{v}) \le (\widehat{G}_k\boldsymbol{v},\boldsymbol{v}) \le c_g(\widehat{A}_k\boldsymbol{v},\boldsymbol{v})$ for all $\boldsymbol{v}$ and $k = 2,\dots,J$. Then*

$$\rho\big((I-\Pi_{\mathbf{1}})(I-B_k^{-1}A_k)\big) \le \min\left\{\frac{2\sqrt{c_g}-2}{\sqrt{c_g}}, \frac{2k+\ln k-1}{2k+\ln k}\right\} < 1,$$

where $\Pi_{\mathbf{1}}$ is the ℓ^2-projection to the space of constant vectors.

Proof. The proof is a direct application of the results in Lemma 5.5. □

Note that the estimates above do not require that c_g is strictly less than 4. However, if $c_g < 4$ then Theorem 5.6 guarantees uniform convergence (independent of the number of levels J), whereas for $c_g = 4$ the convergence rate can only be estimated by $\rho \approx (1 - 1/\log n)$.

A generalization of this estimate can be obtained by assuming that $c_g \le m^2$ for an integer m, in which case there exists a polynomial $q(t)$ of order $m-1$ such that the following spectral equivalence relation can be shown

$$\frac{m^2-c_g}{(m^2-1)c_g} \le \frac{(B_k^{-1}\boldsymbol{v},\boldsymbol{v})}{(A_k^\dagger\boldsymbol{v},\boldsymbol{v})} \le 1, \quad \forall \boldsymbol{v} : (\boldsymbol{v},\mathbf{1}) = 0 \text{ and } k = 1,\dots,J.$$

This implies that the power method with AMLI preconditioner, employing the polynomial $q(t)$ on all levels, has a bounded convergence rate, i.e.,

$$\rho\big((I-\Pi_{\mathbf{1}})(I-B_J^{-1}A)\big) \le \frac{m^2(c_g-1)}{c_g(m^2-1)}.$$

For a matching on a hypercubic grid, as discussed above, the constant c_g approaches 4 asymptotically. This suggests to find the best possible AMLI polynomials for the condition $c_g = 4$, and to test how the AMLI convergence rate relates to the number of levels.

Remark 5.7. *The nearly optimal convergence rate can also be proven for the AMLI methods when the coarse partitioning consists of paths of m vertices where $m > 2$.*

6. Numerical results

In the previous section, the convergence rate of the two-level matching method was used to establish the convergence of the matching-based AMLI method. Here, a numerical implementation that is strictly a translation of this theoretical analysis is considered. Then, a simplified and more efficient variant of the method is developed and tested.

To study the effectiveness of the algorithm and the sharpness of the theoretical estimates of its performance derived in the previous section, the method is applied as a preconditioner to the conjugate gradient iteration. In all tests, the stopping criterion for the PCG solver is an error reduction in A-norm by a factor 10^{10}. The average convergence rate, r_a, and the convergence rates computed by the condition number estimates obtained from the Lanczos algorithm and the AMLI polynomial, denoted by r_e and r_k, respectively, are reported. To reduce the effects of randomness in the numerical results, for each combination of testing parameters, the PCG method is run for five

n	k	r_k	r_e	r_a
128	13.9	0.58	0.56	0.54
256	16.0	0.60	0.59	0.55
512	18.0	0.62	0.58	0.57
1024	20.1	0.64	0.60	0.60
2048	22.1	0.65	0.61	0.61

(a) Square domain with n^2 unknowns

n	k	r_k	r_e	r_a
128	13.9	0.58	0.56	0.56
256	16.0	0.60	0.57	0.59
512	18.0	0.62	0.57	0.58
1024	20.1	0.64	0.59	0.59
2048	22.1	0.65	0.60	0.61

(b) L-shaped domain with $(3/4)n^2$ unknowns

Table 6.1: Results of the AMLI preconditioned CG method applied to the graph Laplacians defined on 2D grids.

n	k	r_k	r_e	r_a
16	16.0	0.60	0.55	0.55
32	20.1	0.64	0.59	0.59
64	24.2	0.66	0.62	0.62
128	28.2	0.68	0.64	0.64

(a) Cubic domain with n^3 unknowns

n	k	r_k	r_e	r_a
16	16.0	0.60	0.55	0.54
32	20.1	0.64	0.59	0.59
64	24.2	0.66	0.62	0.62
128	28.2	0.68	0.64	0.64

(b) Fichera domain with $(7/8)n^3$ unknowns

Table 6.2: Results of the ordinary AMLI preconditioned CG method applied to the graph Laplacians defined on 3D grids.

right hand sides computed by random left hand sides, and the convergence estimate that represents the worst case is reported.

6.1. **An exact implementation of the AMLI method.** As a first test the matching AMLI solver is applied to the graph Laplacian corresponding to 2- and 3-dimensional structured grids on convex and non-convex domains. The Fichera corner domain consists of a twice-unit cube centered at the origin with a single octant removed, see, e.g., [9].

The coarsening is obtained by applying matching only in a single direction on each level until the coarsest level is 1-dimensional, which is then solved using an LU factorization. The AMLI polynomial $q_k(t)$ on the k-th level is determined by the theoretically estimated condition number, given by the recursive formula (5.4). The system $Y_k^T A_k Y_k$ is solved exactly by an LU factorization on smaller grids or CG iteration down to 10^{-6} relative residual on larger grids of the hierarchy.

Such an AMLI method, which is designed to have all assumptions in Theorem 5.6 satisfied, is named "ordinary AMLI method." The results are reported in Table 6.1 and 6.2 and confirm that the actual convergence rate of the method, r_a, and the estimates r_e and r_k match, which indicates that the condition number grows logarithmically with respect to the problem size.

6.2. **Modified AMLI solver for matching.** Next, a more practical variant of the matching AMLI preconditioner is developed. First, the exact $Y_k^T A_k Y_k$ solvers are replaced by Richardson iterations with weights computed using the ℓ^1-induced norm of these matrices, instead of the common choice of their largest eigenvalues.

n	k	r_k	r_e	r_a
128	13.0	0.57	0.59	0.54
256	15.0	0.59	0.62	0.58
512	17.0	0.61	0.64	0.59
1024	19.0	0.63	0.65	0.63
2048	21.0	0.64	0.65	0.65

(a) Square domain with n^2 unknowns

n	k	r_k	r_e	r_a
128	13.0	0.57	0.58	0.56
256	15.0	0.59	0.60	0.56
512	17.0	0.61	0.62	0.57
1024	19.0	0.63	0.64	0.62
2048	21.0	0.64	0.69	0.67

(b) L-shaped domain with $(3/4)n^2$ unknowns

Table 6.3: Results of the modified AMLI preconditioned CG method applied to the graph Laplacians defined on 2D grids.

n	k	r_k	r_e	r_a
16	15.0	0.59	0.50	0.42
32	19.0	0.63	0.54	0.49
64	23.0	0.65	0.57	0.52
128	27.0	0.68	0.59	0.56

(a) Cubic domain with n^3 unknowns

n	k	r_k	r_e	r_a
16	15.0	0.59	0.49	0.49
32	19.0	0.63	0.54	0.50
64	23.0	0.65	0.57	0.56
128	27.0	0.68	0.57	0.60

(b) Fichera domain with $(7/8)n^3$ unknowns

Table 6.4: Results of the modified AMLI preconditioned CG method applied to the graph Laplacians defined on 3D grids.

The lower order term $\ln J = \ln \log_2 n$ that appears in Proposition 5.3 (ii) is neglected, since it is smaller than $2J$ and bounded by 4 for $n \leq 2^{50}$. Another modification to the scheme is the choice of the scaling σ in Lemma 4.4 away from 2. Numerical results suggest that $\sigma = 2 - 1/(2 \log_2 N)$, where N is the number of vertices of the graph, is usually a better scaling than the estimated bound $\sigma = 2$ used in the analysis. We use this choice for the structured mesh problems and for the unstructured problems the scaling is computed through a numerical method.

In Table 6.3 and 6.4, the convergence rate estimates of this approach applied to the same structured problems are reported. Although some of the assumptions of the theory are violated by the method, its performance is similar to that of the approach considered in the previous tests.

Remark 6.1. *A more practical strategy is to use a numerical procedure, e.g., a Lanczos algorithm with an AMLI preconditioner on the k-th level, to estimate the smallest eigenvalue of $B_k^{-1}A_k$, which is then used to determine the AMLI polynomial on the $(k+1)$-th level. Numerical tests indicate that such a strategy improves the convergence rate of the AMLI method, especially for 3- and higher-dimensional problems. However, the setup phase becomes also more expensive as compared to the case in which the polynomials are constructed as suggested in Proposition 5.3.*

6.3. **Unstructured grids.** Finally, tests of this AMLI preconditioned conjugate gradient method applied to the graph Laplacian defined on more general graphs, coming from unstructured meshes resulting from triangulations of a 2-dimensional grid on a square domain, or a 3-dimensional grid on a cubic domain, are considered. The unstructured grid is generated by perturbing grid points of a

n	k	r_k	r_e	r_a
128	16.0	0.60	0.70	0.58
256	18.0	0.62	0.72	0.54
512	20.1	0.64	0.74	0.63
1024	22.1	0.65	0.75	0.65
2048	24.2	0.66	0.76	0.67

(a) Ordinary AMLI

n	k	r_k	r_e	r_a
128	15.0	0.59	0.70	0.70
256	17.0	0.61	0.71	0.70
512	19.0	0.63	0.72	0.72
1024	21.0	0.64	0.73	0.73
2048	23.0	0.65	0.75	0.75

(b) Modified AMLI

Table 6.5: Results of the CG method preconditioned by variants of the matching AMLI methods applied to the graph Laplacian defined on 2D unstructured grids of size n^2.

structured grid by a random vector of length $h/2$, where h is the mesh size of the original structured grid, followed by a Delaunay triangulation. Then, a random matching is applied recursively to generate a multilevel hierarchy with $(\log_2 N)/2$ levels. The 3-dimensional unstructured grids are generated in a similar way and the multilevel hierarchy is constructed accordingly by the random matching algorithm.

The results of these tests are reported in Table 6.5 and 6.6. The systems with the blocks $Y_k^T A_k Y_k$ arising in the application of the two-level preconditioners are solved via an inner CG iteration; This is feasible since these matrices are proven well conditioned even for unstructured grids. The recursive formula (5.9) is used to derive the polynomials used in the AMLI cycles, and the scaling constants are computed using

$$\sigma_k = \max_{i \neq j} \frac{(P_k^T A_k P_k)_{ij}}{(A_{k+1})_{ij}},$$

which ensures that the upper bound in (5.10) is always 1. This in turn guarantees that the AMLI method, as a preconditioner for the CG method, is always positive semi-definite since the polynomials, constructed according to (5.9), are negative when $t > 1$. Assuming that the scaling constant σ_k is smaller than the value suggested above, then there exists a vector $\boldsymbol{v}$ such that

$$(G_k^\dagger \boldsymbol{v}, \boldsymbol{v}) > (A_k^\dagger \boldsymbol{v}, \boldsymbol{v}),$$

which makes it possible that $(B_k^{-1}\boldsymbol{v}, \boldsymbol{v}) > (A_k^\dagger \boldsymbol{v}, \boldsymbol{v})$. Note that if the matrix $B_k^{-1} q(A_k B_k^{-1})$ is indefinite the preconditioner B_{k+1} will be indefinite as well.

For the results on the right of Table 6.5 and 6.6, the solution of the systems with the blocks $Y_k^T A_k Y_k$ is replaced by one Richardson iteration each, and the AMLI polynomials are constructed based on (5.9) without taking into account the lower order term $\ln k$. The asymptotic convergence rates are close to those for the *ordinary AMLI* method.

7. Conclusions

An algebraic formula for estimating the convergence rate of an aggregation-based two level method is derived, and it is shown that the formula can be used to obtain sharp estimates of the convergence rates in the special case where matching is used. With the use of geometric information, a sharp bound of the two-level method is derived. The nearly optimal convergence and complexity of the multilevel method that uses AMLI cycles is also established. The reported numerical tests illustrate the sharpness of the theoretical estimates. Moreover all the theoretical results can be generalized to aggregates of general size and, hence, can be used to study an approach which combines aggressive aggregation with AMLI cycles, which should result in a fast and memory

n	k	r_k	r_e	r_a
16	18.0	0.62	0.65	0.48
32	22.1	0.65	0.67	0.55
64	26.2	0.67	0.70	0.62
128	30.3	0.69	0.74	0.60

(a) Ordinary AMLI

n	k	r_k	r_e	r_a
16	17.0	0.61	0.59	0.55
32	21.0	0.64	0.63	0.58
64	25.0	0.67	0.65	0.62
128	29.0	0.69	0.67	0.65

(b) Modified AMLI

Table 6.6: Results of the CG method preconditioned by variants of the matching AMLI methods applied to the graph Laplacian defined on 3D unstructured grids of size n^3.

efficient solver for graph Laplacians. Development and analysis of such a scheme and one that uses more general smoothers are subject of on-going research.

References

[1] O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods. I. *Numer. Math.*, 56(2-3):157–177, 1989.

[2] A. Brandt, J. Brannick, M. Bolten, A. Frommer, K. Kahl, and I. Livshits. Bootstrap AMG for Markov chains. *SIAM J. Sci. Comp.*, 33:3425–3446, 2011.

[3] A. Brandt, S. McCormick, and J. W. Ruge. Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations. Technical report, Colorado State University, Fort Collins, Colorado, 1983.

[4] J. Brannick, R. Brower, M. Clark, J. Osborn, and C. Rebbi. Adaptive multigrid algorithm for lattice QCD. *Phys. Rev. Lett.*, 100, 2008.

[5] J. Brannick, Y. Chen, and L. Zikatanov. An algebraic multilevel method for anisotropic elliptic equations based on subgraph matching. *Numer. Linear Algebra Appl.*, 19:279–295, 2012.

[6] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge. Adaptive smoothed aggregation (αSA). *SIAM J. Sci. Comp.*, 25(6):1896–1920, 2004.

[7] M. Brezina, P.Vaněk, and P. Vassilevski. An improved convergence analysis of smoothed aggregation algebraic multigrid. *Journal of Numerical Linear Algebra and Applications*, n/a:n/a, 2010. accepted for publication (March 2, 2011).

[8] Marian Brezina, Caroline I. Heberton, Jan Mandel, and Petr Vaněk. An iterative method with convergence rate chosen a priori. UCD/CCM Report 140, Center for Computational Mathematics, University of Colorado at Denver, February 1999. Presented at the 1998 Copper Mountain Conference on Iterative Methods.

[9] P. Brune, M. Knepley, and L. R. Scott. Unstructured geometric multigrid in two and three dimensions on complex and graded meshes. *SIAM Journal for Scientific Computation*, 2012. Accepted.

[10] Jie Chen and Ilya Safro. Algebraic distance on graphs. *SIAM J. Sci. Comput.*, 33(6):3468–3490, 2011.

[11] Robert D. Falgout, Panayot S. Vassilevski, and Ludmil T. Zikatanov. On two-grid convergence estimates. *Numer. Linear Algebra Appl.*, 12(5-6):471–494, 2005.

[12] George Karypis and Vipin Kumar. A parallel algorithm for multilevel graph partitioning and sparse matrix ordering. *Journal of Parallel and Distributed Computing*, 48:71–85, 1998.

[13] George Karypis and Vipin Kumar. Parallel multilevel k-way partitioning scheme for irregular graphs. *SIAM Review*, 41(2):278–300, 1999.

[14] H. Kim, J. Xu, and L. Zikatanov. A multigrid method based on graph matching for convection-diffusion equations. *Numer. Linear Algebra Appl.*, 10:181–195, 2003.

[15] I. Koutis. *Combinatorial and algebraic tools for optimal multilevel algorithms*. PhD thesis, Pittsburgh, PA 15213, 2007.

[16] I. Koutis, G. L. Miller, and D. Tolliver. Combinatorial preconditioners and multilevel solvers for problems in computer vision and image processing. In *Proceedings of the 5th International Symposium on Advances in Visual Computing: Part I, ISVC '09*, pages 1067–1078, Berlin, Heidelberg, 2009. Springer-Verlag.

[17] A. F. Leont'ev. Sequences of linear aggregates formed from solutions of differential equations. *Mat. Sb. (N.S.)*, 48(90):129–136, 1959.

[18] O. Livne and A. Brandt. Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver. *SIAM Journal on Scientific Computing*, 34(4):B499–B522, 2012.

[19] Dorit Ron, Ilya Safro, and Achi Brandt. Relaxation-based coarsening and multiscale graph organization. *Multiscale Model. Simul.*, 9(1):407–423, 2011.
[20] H.A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29:111–138, 1961.
[21] D.A. Spielman. Algorithms, graph theory, and linear equations in Laplacian matrices. In R. Bhatia, editor, *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010)*, volume 4, pages 2698–2722. World Scientific, 2010.
[22] H. De Sterck, T. Manteuffel, S. McCormick, Q. Nguyen, and J. Ruge. Multilevel adaptive aggregation for Markov chains with application to web ranking. *SIAM J. Sci. Comput.*, 30:2235–2262, 2008.
[23] Yukio Takahashi. A lumping method for numerical calculations of stationary distributions of markov chains. *Research Report B-18, Department of Information Sciences, Tokyo*, 23, 1975.
[24] P. Vaněk, M. Brezina, and J. Mandel. Convergence of algebraic multigrid based on smoothed aggregation. *Numer. Math.*, 88:559–579, 2001.
[25] P. Vaněk. Acceleration of algebraic multigrid by smoothed transfer operators. *Applications of Mathematics*, 37(1):13–28, 1992.
[26] P. Vaněk, J. Mandel, and M. Brezina. Algebraic multigrid by smoothed aggregation for second and fourth order elliptic problems. *Computing*, 56:179–196, 1996.
[27] Panayot S. Vassilevski. *Multilevel block factorization preconditioners*. Springer, New York, 2008. Matrix-based analysis and algorithms for solving finite element equations.

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA
E-mail address: brannick@psu.edu

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA
E-mail address: chen_y@math.psu.edu

Johann Radon Institute, Austrian Academy of Sciences, Altenberger Str. 69, 4040 Linz, Austria
E-mail address: johannes.kraus@oeaw.ac.at

Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA
E-mail address: ludmil@psu.edu