NASA/CR-201741
ICASE Report No. 97-47

Graph Embedding Techniques for Bounding Condition Numbers of Incomplete Factor Preconditioners

Stephen Guattery
ICASE

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, VA

Operated by Universities Space Research Association

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

Prepared for Langley Research Center
under Contracts NAS1-97046 & NAS1-19480

September 1997

GRAPH EMBEDDING TECHNIQUES FOR BOUNDING CONDITION NUMBERS OF INCOMPLETE FACTOR PRECONDITIONERS

STEPHEN GUATTERY*

Abstract.

We extend graph embedding techniques for bounding the spectral condition number of preconditioned systems involving symmetric, irreducibly diagonally dominant M-matrices to systems where the preconditioner is not diagonally dominant. In particular, this allows us to bound the spectral condition number when the preconditioner is based on an incomplete factorization. We provide a review of previous techniques, describe our extension, and give examples both of a bound for a model problem, and of ways in which our techniques give intuitive way of looking at incomplete factor preconditioners.

Key words. incomplete Cholesky factorization, graph eigenvalues and eigenvectors, preconditioning

Subject classification. Computer Science

1. Introduction. The number of iterations required for convergence is an important measure of the performance of iterative methods such as conjugate gradient and preconditioned conjugate gradient. In most cases, this measure is difficult to determine; however, upper bounds based on the spectral condition number can be calculated.

The spectral condition number is the ratio of the largest to smallest eigenvalues of the matrix for which we are computing solutions. In preconditioned systems with matrix A and preconditioner B, this ratio is computed for the matrix $B^{-1}A$. Calculating spectral condition numbers exactly can require substantial work and storage. Hence it is desirable to find a general method that can be applied to wide range of matrices and preconditioners.

Gremban [11] used graph embeddings in bounding condition numbers for preconditioned systems where the matrix and preconditioner are symmetric, irreducibly diagonally dominant M-matrices (such matrices have positive diagonal and nonpositive off-diagonal entries). He also gave an extension that allows positive off-diagonal entries as long as the matrix remains diagonally dominant. His techniques are often easy to apply and often give good bounds. They can take advantage of the wide variety of embedding results developed in the study of networks. However, they apply only to a restricted set of matrices.

In this paper, we show how embedding techniques can be extended to handle a class of positive definite preconditioners that may not be diagonally dominant. In particular, this allows us to apply them to preconditioners formed by incomplete Cholesky factorization techniques. In addition to allowing broader application of the embedding techniques, our extension also provides some nice intuitive interpretations of incomplete factor preconditioners. We give an example at the end of this paper.

The rest of this paper is organized as follows: Section 2 covers previous work that relates graph embeddings and eigenvalue bounds; Section 3 presents the notation and terminology we use; Section 4 reviews embedding techniques for generalized Laplacians; Section 5 presents our techniques for extending these

*Institute for Computer Applications in Science and Engineering, Mail Stop 403, NASA Langley Research Center, Hampton, VA 23681-0001 (email: smg@icase.edu). This research was supported by the National Aeronautics and Space Administration under NASA Contract Nos. NAS1-97046 and NAS1-19480 while the author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23681-0001.

techniques to preconditioners based on incomplete factorizations; and Section 6 presents some examples to illustrate how these extensions can be used.

2. Previous Work. The study of the connection between Laplacian spectra (particularly with respect to λ_2, the smallest nontrivial eigenvalue of a Laplacian matrix) and properties of the associated graphs dates back to Fiedler's work in the 1970's (see, e.g., [7] and [8]).

The relationship between graph embeddings and matrix representations has been the subject of much interesting research. A large proportion of this work has been aimed at bounding the second largest eigenvalues of time-reversible Markov chains in order to bound the mixing time for random walks. The use of clique embeddings to bound eigenvalues arose in the analysis of mixing times for Markov chains by Jerrum and Sinclair [14] [17]. Further work in this direction was done by Diaconis and Strook [4] and by Sinclair [16]. Kahale [15] generalized this work in terms of methods that assign lengths to the graph edges, and showed that the best bound over all edge length assignments is the largest eigenvalue of the matrix $\Gamma^T\Gamma$, where Γ is a matrix representing the path embedding ([15] also cites unpublished work by Fill and Sokal in these directions). He also gave a semidefinite programming formulation for a model allowing fractional paths, and showed that the bound is off by at most a factor of $\log^2 n$. He showed this gap is tight; he also noted that the results can be applied to bounding λ_2 of a Laplacian from below.

Guattery, Leighton, and Miller [12] presented a lower bound technique for λ_2 of a Laplacian. It assigns priorities to paths in the embedding, and uses these to compute congestions of edges in the the original graph with respect to the embedding. Summing the congestions along the edges in a path gives the path congestion; the lower bound is a function of the reciprocal of the maximum path congestion taken over all paths. For the clique case, they showed that this method is the dual of the method presented in [15]; the best lower bounds produced by these methods are the same. They also showed how to apply their method in the Dirichlet boundary case by using star embeddings. In the clique case, they show that using uniform priorities for any tree T gives a lower bound that is within a factor proportional to the logarithm of the diameter of the tree.

Guattery and Miller [13] have shown that the bounds discussed in the previous paragraphs are not tight because of the problem representation. By incorporating edge directions into the embeddings, they were able to show that there exists an embedding (dubbed the current flow embedding) for which there is an exact relationship between the largest eigenvalue of $\Gamma^T\Gamma$ and the smallest nontrivial Laplacian eigenvalue. This is true both for clique and star embedding cases.

Gremban [11] has shown how to use embeddings to generate support numbers, which also provide bounds on the largest and smallest generalized eigenvalues (and hence the spectral condition number) of preconditioned linear systems involving a generalized definition of Laplacians. This work is reviewed in Section 4 below. He also defined the support tree preconditioner, and used the support number bounds to prove properties about the quality of these preconditioners. Gremban, Miller, and Zagha have evaluated the performance of these techniques [10].

3. Notation and Terminology.

3.1. Matrices. All matrices considered in this paper are real matrices. We use capital letters (e.g. A) to represent matrices and bold lower case letters to represent vectors (e.g., $\mathbf{x}$).

A matrix A is *diagonally dominant* if all diagonal entries are positive ($a_{ii} > 0$), and for every row i, $a_{ii} \geq \sum_{j \neq i} |a_{ij}|$. If the inequality in the second condition is strict for all rows, then the matrix is *strictly diagonally dominant*. If A is irreducible and the second condition is strict for at least one row i, then the

matrix is *irreducibly diagonally dominant.*

3.2. Generalized Laplacian Matrices and Graphs. DEFINITION 3.1. *A is a* generalized Laplacian matrix *if and only if:*

- *A is symmetric;*
- *all diagonal entries* $a_{ii} > 0$*;*
- *A is diagonally dominant.*

To insure positive definiteness, we will assume that the matrix is irreducibly diagonally dominant. Note that if the matrix is positive definite and reducible, we can break the problem into smaller pieces of the desired form.

Generalized Laplacians correspond to graphs with positive edge weights according to the following rules:

- Each row (or column) corresponds to a vertex.
- Nonzero off-diagonal entries correspond to edges. That is, for $i \neq j$, if $a_{ij} \neq 0$, then there is an edge between vertices v_i and v_j with weight $-a_{ij}$.
- The diagonal entry a_{ii} is the sum of the weights of the edges incident to vertex v_i. If a diagonal entry is greater than the sum of the incident edge weights, there is an additional edge from that vertex to an implicit zero-valued boundary vertex. While this vertex is implicit with respect to the matrix, we will represent it explicitly with respect to the graph.

When necessary, we use the following notation to relate graphs and matrices: For a Laplacian A, the associated graph is $G(A)$. The generalized Laplacian of a graph G is denoted $L(G)$.

The following property is a useful consequence of interpreting a Laplacian as a graph: Let A be a Laplacian with associated graph G. Recall that $-a_{ij}$ is the weight of edge (v_i, v_j). For all $\mathbf{x}$,

$$\mathbf{x}^T A \mathbf{x} = \sum_{(v_i,v_j)\in E(G)} -a_{ij}(x_i - x_j)^2. \tag{3.1}$$

(Edges represented by surpluses on the diagonal are included in this sum by using 0 for the value at the (implicit) boundary vertex.)

3.3. Graph Embeddings. For a graph G, we will use the notation $V(G)$ to represent the set of vertices of G, and $E(G)$ to represent the set of edges of G.

An *embedding* of H into G is a collection Γ of path subgraphs of G such that for each edge $(v_i, v_j) \in E(H)$, the embedding contains a simple path γ_{ij} from v_i to v_j in G. For full generality, we will allow fractional paths in our embeddings: i.e., an edge $(v_i, v_j) \in E(H)$ can be associated with a finite collection of simple paths from v_i to v_j in G; each such path has a positive fractional factor associated with it such that these factors add up to 1. If a path γ includes edge e, we say that γ is incident to e. The *weight of a path* w_γ is the weight of the corresponding edge in H. In the case of fractional paths, the weight is scaled by the corresponding fractional weight.

The *congestion* c_e of edge $e \in E(G)$ is the sum of the weights of the paths incident on e. (In the unweighted case, this is just the number of paths that include e.) The *congestion of the embedding* is the maximum edge congestion taken over all edges in G.

The *dilation* of an edge f in H is the length (the number of edges) in f's path γ_f in the embedding. The dilation of the embedding is the maximum dilation taken over all edges in H.

We note that when we represent matrices by graphs for embedding purposes, the graph representations (and hence the embeddings) include the (implicit) edges to the boundary indicated by surpluses on the

matrix diagonals. In such cases, the boundary is represented by a single vertex; the weights of the edges to the boundary are the diagonal surpluses.

3.4. Incomplete Factorizations. Incomplete factorization techniques provide one way of constructing preconditioners for sparse systems (see the survey by Chan and van der Vorst [3] for a good overview of incomplete factorization techniques). Since we are dealing with symmetric matrices, we will focus on incomplete Cholesky factorizations.

To keep work and storage small, incomplete factorization methods limit the amount of fill allowed in the factor. One way to do this is to specify the entries where fill is allowed. There are many ways to do this, e.g., by level of the fill produced. In general, we can specify a set of allowed fill positions. We represent this set by a 0-1 matrix S. By assumption, S includes all nonzero entries in A.

For symmetric matrix A, the incomplete Cholesky factorization algorithm produces a lower triangular factor L. For A a symmetric irreducibly diagonally dominant M-matrix, L has a positive diagonal and nonpositive off-diagonals. The preconditioner $B = LL^T$ disagrees with A only at level(0) fill positions of S.

The *error matrix* $R = B - A$ is symmetric with a zero diagonal. When A is a nonsingular generalized Laplacian, it is also an M-matrix, and off-diagonal entries are nonnegative. Since we assume the factorization is incomplete, there must be positive entries in R. It is easy to see that R is indefinite. Let vector d be the product of R and the vector with all entries 1, and let D be the diagonal matrix with the entries of d on the diagonal. We can write R as the sum of a positive semidefinite matrix $R^+ = (R + D)/2$ and a negative semidefinite matrix $R^- = (R - D)/2$.

4. Graph Embedding Techniques for Generalized Laplacians.

4.1. Preconditioned Conjugate Gradient and the Spectral Condition Number. Consider the problem of solving $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$ where A is a positive definite generalized Laplacian. This can be done using the conjugate gradient method. The number of iterations required for convergence depends on the distribution of A's eigenvalues, and can be complicated to compute (see e.g. [2]). However, an upper bound on the convergence rate can be computed from the spectral condition number $\kappa(A) = \lambda_{max}(A)/\lambda_{min}(A)$. The rate of convergence can often be improved by applying a preconditioner B and using the preconditioned conjugate gradient algorithm. In this case, the rate of convergence depends on the spectrum of $B^{-1}A$ and can be bounded above as a function of $\kappa(B^{-1}A) = \lambda_{max}(B^{-1}A)/\lambda_{min}(B^{-1}A)$. We assume that the preconditioner B is symmetric positive definite in the discussion below.

The problem of computing $\kappa(B^{-1}A)$ can be defined in terms of the *generalized eigenvalue problem*, which involves finding all scalars λ for which there exists an $\mathbf{x} \neq 0$ such that $A\mathbf{x} = \lambda B\mathbf{x}$. Since by assumption B is nonsingular, we have

$$A\,\mathbf{x} = \lambda B\,\mathbf{x} \quad \leftrightarrow \quad B^{-1}A\,\mathbf{x} = \lambda\,\mathbf{x}. \tag{4.1}$$

The ordered pair (A, B) is called a *matrix pencil*, and we denote an eigenvalue of the pencil by $\lambda(A, B)$.

In the next section we present a lemma that provides the basis for computing upper bounds on $\lambda_{max}(A, B)$, and hence on $\lambda_{max}(B^{-1}A)$. We note that since A and B are by assumption positive definite,

$$\frac{1}{\lambda_{min}(B^{-1}A)} = \lambda_{max}((B^{-1}A)^{-1}) = \lambda_{max}(A^{-1}B). \tag{4.2}$$

Thus, we can apply upper bounds to both terms in the definition of $\kappa(B^{-1}A)$.

4.2. The Support Lemma.

Let A be a symmetric matrix, B be a symmetric positive definite matrix, and let τ be a real number. Axelsson [1] gives the following lemma:

LEMMA 4.1. *If $\tau B - A$ is positive semidefinite, then $\lambda_{max}(B^{-1}A) \leq \tau$.*

Proof. Let $\mathbf{u}$ be an eigenvector of $\lambda_{max}(B^{-1}A)$. By (4.1), $A\,\mathbf{u} = \lambda B\,\mathbf{u}$. Starting from the assumption that $\tau B - A$ is positive semidefinite, we can deduce the following:

$$0 \leq \mathbf{u}^T(\tau B - A)\mathbf{u} = (\tau - \lambda)\mathbf{u}^T B\,\mathbf{u}.$$

Since B is positive definite, this implies that $\tau - \lambda \geq 0$. □

Gremban [11] refers to this as the *Support Lemma*. He defines *support* as follows: The *support* $\sigma(A, B)$ of matrix B for matrix A is

$$\min\{\tau : \tau B - A \text{ is positive semidefinite}\}$$

He uses support to find upper bounds on the spectral condition number of preconditioned systems using generalized Laplacians. Note that, by the Support Lemma, $\lambda_{max}(B^{-1}A) \leq \sigma(A, B)$, and $1/\lambda_{min}(B^{-1}A) = \lambda_{max}(A^{-1}B) \leq \sigma(B, A)$. Thus

$$\kappa(B^{-1}A) = \frac{\lambda_{max}(B^{-1}A)}{\lambda_{min}(B^{-1}A)} \leq \sigma(A, B) \cdot \sigma(B, A).$$

Gremban's method works by decomposing A into k pieces $A_1, A_2, \ldots, A_k$ such that $\sum_{i=1}^{k} A_i = A$. Likewise, suppose B can be decomposed into k positive semidefinite pieces. Assume that we have a set $\{\tau_1, \tau_2, \ldots, \tau_k\}$ such that $\tau_i B_i - A_i$ is positive semidefinite for all i. Let $\tau^* = \max_i \tau_i$. Then $\tau^* B_i - A_i$ is positive semidefinite for all i, and

$$\sum_{i=1}^{k}(\tau^* B_i - A_i) = \tau^* \sum_{i=1}^{k} B_i - \sum_{i=1}^{k} A_i = \tau^* B - A.$$

By linearity, $\tau^* B - A$ is positive semidefinite.

For generalized Laplacians the decompositions of A and B can be based on graph embeddings. Let k be the number of edges in $G(A)$. We will decompose $B - A$ into pieces $B_i - A_i$, where A_i is the Laplacian of the graph on $V(G(A))$ consisting of only edge $e_i \in E(G(A))$. B_i is the Laplacian of the (appropriately weighted) corresponding path in $G(B)$.

To determine the appropriate weighting of the paths, note that an edge e in $G(B)$ may show up in multiple paths in the decomposition. The edge must be divided up so that the weights of the pieces of e on various paths sum to w_e, the weight of e. Let c_e be the congestion of edge e in B; let w_f be the weight of edge f in A. Assume that the path for f includes e. The amount of weight from w_e assigned to the path associated with f is

$$\frac{w_f}{c_e}.$$

(In the unweighted case, this will be just $\frac{1}{c_e}$.)

Note that if there are edges in $G(B)$ that do not occur in any path, they can be separated out into a component that can support an empty component of A. Thus they do not affect the rest of the calculation.

The decomposition given above reduces the problem of finding a τ that is an upper bound of $\sigma(A, B)$ to the problem of computing τ_i's for a number of problems that consist of supporting an edge with a path.

4.3. The Path Problem and Electrical Circuits. Now consider the problem of a path γ in $G(B)$ supporting an edge e in $G(A)$. Assume the path length (the dilation for edge e) is j, and let w_e be the weight of e. For simplicity of notation, we reindex the vertices from 1 to $j+1$ according to their order along the path. Let c_i be the congestion on edge $(i, i+1)$ of the path. The weight of edge i in the path is w_e/c_i.

Using (3.1), we can state the path problem as follows: Choose τ_γ such that

$$\tau_\gamma \sum_{i=1}^{j} \frac{w_e}{c_i}(x_i - x_{i+1})^2 \geq w_e(x_1 - x_{j+1})^2,$$

or, cancelling the common factor w_e,

$$\tau_\gamma \sum_{i=1}^{j} \frac{1}{c_i}(x_i - x_{i+1})^2 \geq (x_1 - x_{j+1})^2.$$

The path problem looks like the power problem in a series resistive circuit[1]. In particular, the entries in $\mathbf{x}$ correspond to voltages at path nodes, and each value $\frac{1}{c_i}$ corresponds to the conductance between nodes i and $i+1$. Since conductances are reciprocals of resistances, the congestion of an edge can be thought of as its resistance. Thus, we can restate the path problem as follows: Given voltages at the ends of the circuit, what voltages at the internal nodes produce the minimum power dissipation?

We construct a series resistive circuit corresponding to path γ as follows: For edge i on the path, assign a resistor with resistance $r_i = c_i$. Define the *path resistance* r_γ as the sum of the resistances on the path:

$$r_\gamma = \sum_{i=1}^{j} r_i = \sum_{i=1}^{j} c_i.$$

The following theorem is well-known (see e.g. [5]):

THEOREM 4.2. *For any* $\mathbf{x}$*,*

$$r_\gamma \sum_{i=1}^{j} \frac{1}{c_i}(x_i - x_{i+1})^2 \geq (x_1 - x_{j+1})^2.$$

Proof. We can rewrite the left-hand side of the inequality in terms of congestions as follows:

$$\sum_{i=1}^{j} c_i \sum_{i=1}^{j} \frac{1}{c_i}(x_i - x_{i+1})^2.$$

Rewriting slightly and applying Cauchy-Schwarz gives the following:

$$\sum_{i=1}^{j} (\sqrt{c_i})^2 \sum_{i=1}^{j} \frac{1}{(\sqrt{c_i})^2}(x_i - x_{i+1})^2 \geq \left(\sum_{i=1}^{j} \sqrt{c_i} \frac{1}{\sqrt{c_i}}(x_i - x_{i+1}) \right)^2 = (x_1 - x_{j+1})^2.$$

The last inequality follows because the sum telescopes. This proves the theorem. □

Thus, for any path problem, it is sufficient to set $\tau_\gamma = r_\gamma$, the sum of the congestions along the path. The maximum path resistance taken over all paths in Γ is denoted $r_{\max}$. The corresponding path is called the *critical path*. By the partitioning argument in Section 4.2, setting $\tau^* = r_{\max}$ insures that τ^* is an upper bound on $\lambda_{max}(B^{-1}A)$.

Gremban suggests an easier way of computing a sufficiently large τ based on the following two facts:

[1] Laplacians can be used to represent resistive circuits, where the off-diagonal entries represent conductances between nodes in the circuit.

- For any path γ in the embedding, setting all congestions equal to the maximum congestion on the path only increases τ_γ, which thus remains an upper bound.
- If all congestions on the path have value c, then the path resistance of γ (and hence τ_γ) is the product of the path length times c.

Thus, the path resistance of any path is less than the product of the path's length times the maximum congestion on any path edge. This product is bounded above for any path by the product of the embedding's congestion times its dilation.

Though the product of the congestion of the embedding times dilation can be greater than the value of τ derived by looking at all paths, in many interesting cases the difference is not more than a constant factor. Since this product is often much easier to compute, it is a useful simplification.

4.4. An Extension for Positive Off-Diagonal Entries. Gremban gives an extension that allows the embedding techniques to be used for diagonally dominant symmetric matrices that have positive off-diagonal entries. We refer to such matrices as *extended Laplacians* (Gremban calls them generalized Laplacians, but this conflicts with Fiedler's definition of that term, which we use).

Note that any extended Laplacian A can be written as the sum of a diagonal matrix D equal to A's diagonal, a matrix O^- containing the negative off-diagonal entries, and a matrix O^+ containing the positive off-diagonal entries. The *expansion* A_{exp} of A has the following block form:

$$A_{\text{exp}} = \begin{bmatrix} D + O^- & -O^+ \\ -O^+ & D + O^- \end{bmatrix}.$$

The following lemma is a restatement of Lemma 7.3 from Gremban's thesis [11]. Let A be an extended Laplacian and $\mathbf{x}$ any vector.

LEMMA 4.3.

$$A\mathbf{x} = \mathbf{b} \text{ if and only if } A_{\text{exp}} \begin{bmatrix} \mathbf{x} \\ -\mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}.$$

The proof is straightforward and left to the reader.

It is easy to show that applying the support lemma to the expansions of extended Laplacians yields upper bounds on the eigenvalues of the original matrices. Let A and B be extended Laplacians.

LEMMA 4.4. *If τ is a nonnegative number such that $\tau B_{\text{exp}} - A_{\text{exp}}$ is positive semidefinite, then $\tau B - A$ is also positive semidefinite.*

Proof. Let τ be small enough that $\tau B - A$ is not positive semidefinite. Then there is vector $\mathbf{u}$ such that $\mathbf{u}^T (\tau B - A)\, \mathbf{u} < 0$. Let $\mathbf{v} = (\tau B - A)\, \mathbf{u}$; clearly $\mathbf{u}^T \mathbf{v} < 0$. By Lemma 4.3,

$$[\mathbf{u}^T,\ -\mathbf{u}^T]\, (\tau B_{\text{exp}} - A_{\text{exp}}) \begin{bmatrix} \mathbf{u} \\ -\mathbf{u} \end{bmatrix} = [\mathbf{u}^T,\ -\mathbf{u}^T] \begin{bmatrix} \mathbf{v} \\ -\mathbf{v} \end{bmatrix} < 0.$$

Thus, $\tau B_{\text{exp}} - A_{\text{exp}}$ is not positive semidefinite either. □

As a result, we can apply embedding techniques to the expansions of extended Laplacians to get upper bounds that apply to the extended Laplacians.

5. Bounds on the Condition Number for Incomplete Factorizations. The techniques from the previous section do not accommodate matrices that are not diagonally dominant, and thus are often unsuitable for preconditioners based on incomplete factorizations. More specifically, let A be a positive

definite generalized Laplacian, let L be the lower triangular incomplete factor of A, and let $B = LL^T = A+R$ be the resulting preconditioner. The matrix R adds positive off-diagonal entries without changing the diagonal. Typically this means there are zero-sum rows in A (i.e., nodes not adjacent to the boundary) whose corresponding rows in B have deficiencies on the diagonal.

Let $B = A + R$ be the preconditioner formed from the incomplete factorization of A. We want an upper bound on the condition number

$$\kappa\left((A+R)^{-1}A\right) = \frac{\lambda_{max}\left((A+R)^{-1}A\right)}{\lambda_{min}\left((A+R)^{-1}A\right)}.$$

Since both A and $A+R$ are symmetric positive definite, we can rewrite this as

$$\kappa\left((A+R)^{-1}A\right) = \lambda_{max}\left((A+R)^{-1}A\right) \cdot \lambda_{max}\left(A^{-1}(A+R)\right).$$

We first consider $\lambda_{max}\left((A+R)^{-1}A\right)$. By Lemma 4.1, any τ such that $\tau(A+R)-A$ is positive semidefinite is an upper bound. If R were positive semidefinite, $\tau = 1$ would work. However, R is indefinite, and A must support R as well as itself. This suggests splitting A into two parts: if we can find a positive $\alpha < 1$ such that $\alpha A + R$ is positive semidefinite, we can rewrite the expression in τ as follows:

$$\tau(A+R) - A = \tau(\alpha A + R) + \tau(1-\alpha)A - A.$$

If such an α exists, $\tau(\alpha A+R)$ will be positive semidefinite for any nonnegative τ, and it will suffice to find τ such that $\tau(1-\alpha)A-A$ is positive semidefinite. Thus $\tau \geq \frac{1}{1-\alpha}$ will be an upper bound on $\lambda_{max}\left((A+R)^{-1}A\right)$. The following lemma shows the existence of such an α.

LEMMA 5.1. *For A and R as defined above, there exists a positive $\alpha < 1$ such that $\alpha A + R$ is positive semidefinite.*

Proof. It is easy to see that

$$\mathbf{x}^T(A+R)\mathbf{x} \geq \lambda_{min}(A+R)\mathbf{x}^T\mathbf{x} \geq \frac{\lambda_{min}(A+R)}{\lambda_{max}(A)}\mathbf{x}^T A\mathbf{x}.$$

Since $A+R$ is positive definite, $\lambda_{min}(A+R) > 0$, and $\lambda_{min}(A+R)/\lambda_{max}(A)$ is positive. It is the case that

$$\lambda_{min}(A+R) \leq \lambda_{max}(A) + \lambda_{min}(R) < \lambda_{max}(A);$$

the first inequality follows from a well-known result (see e.g. Corollary 8.1.3 on p. 411 of Golub and Van Loan [9]), and the second from the fact that $\lambda_{min}(R)$ is negative. Thus there exists an α such that

$$0 < \alpha \leq 1 - \frac{\lambda_{min}(A+R)}{\lambda_{max}(A)} < 1.$$

□

It is often possible to calculate a reasonable α using embedding techniques. However, since R is indefinite, some manipulation is required. The expression $\alpha A + R$ can be rewritten as $\alpha A + R^+ + R^-$. Since R^+ is positive semidefinite, it suffices to find an α such that $\alpha A + R^-$ is positive semidefinite. This can be rewritten as $\alpha A - (-R^-)$; since R^- is the negative of a Laplacian, embedding techniques can now be applied.

To get an upper bound for $\lambda_{max}\left(A^{-1}(A+R)\right)$ we need to find a τ such that $\tau A - (A+R)$ is positive semidefinite. We can again rewrite in terms of R^+ and R^-:

$$\tau A - (A+R) = (\tau - 1)A - R^- - R^+.$$

Since $-R^-$ is positive semidefinite, it suffices to find τ such that $(\tau-1)A - R^+$ is positive semidefinite. This can be done using embedding techniques, though it requires using the expansions of the Laplacians of A and R^+ because R^+ has positive off-diagonal entries. Note that the expansion of A will be two copies of A, and that the expansion of R^+ will have edges that connect those two copies. The embedding of R^+ into A has to use the edges between each copy of A and the vertex representing the boundary.

6. Applications.

6.1. Natural Ordering for a Square Grid. To simplify the presentation in this section, we will use the same notation for both a matrix and its graph. For example, A will refer both to the matrix A and to $G(A)$. The type of object referred to will be clear from context.

Consider the level(0) incomplete Cholesky factorization of a unit-weight square grid in natural order, where every node in the outer perimeter of the grid is connected to a zero Dirichlet boundary. Let A be the generalized Laplacian of the grid, L be the lower triangular factor, $B = LL^T$, and $R = B - A$. Eijkhout [6] shows that the maximum entry of R is bounded above by the value $1 - \sqrt{2}/2$.

To use the embedding techniques to bound the condition number, we first note that the edges in R consist of one diagonal across each square face in the graph of A. It is obvious that we can partition A into distinct pairs of edges that form paths of length 2 between the endpoints of the edges in R.

We start with the upper bound on $\lambda_{max}\left((A+R)^{-1}A\right)$. This involves first embedding $-R^-$ into A. By the observation in the preceding paragraph, each edge in R^- is supported by a unique path of length 2 in A. The congestion on any edge in A is the weight of the edge in $-R^-$ that it has been assigned to support. Thus, the path resistance of any path in A supporting an edge e in $-R^-$ is twice the weight of e. Recalling that the edge weights in $-R^-$ are half the corresponding values in R, we see that the α needed for A to support R is the maximum weight entry from R (we denote it by $\max(R)$). The upper bound on $\lambda_{max}\left((A+R)^{-1}A\right)$ is $\tau = 1/(1-\alpha) = 1/(1-\max(R))$; for the grid in natural order, this value is $\sqrt{2}$.

Now consider the upper bound on $\lambda_{max}\left(A^{-1}(A+R)\right)$. A bit of terminology will help make the description of the embedding simpler: consider the grid laid out on a piece of paper; it has a left side and a right side. The edges of R^+ are diagonals of the squares in the grid, so each has a left end and a right end. Given our ordering, each vertex in the grid is the right end of at most one edge of R^+ and the left end of at most one.

Recall that the embedding uses two copies of the grid connected via the node representing the zero boundary; the edges of R^+ run between these copies. We use the following embedding: for each edge in R^+, we route a path from the left-end vertex in copy 1 to the boundary vertex through the closest perimeter vertex on the grid. The path is routed back to the right-end vertex in copy 2 via the perimeter vertex closest to the right end.

Since each grid vertex along the path in copy 1 is the left end of at most one edge in R^+, each contributes at most one unit to the congestion along the path. The longest path (which gives the most congestion) runs from vertices at the center of the grid. Their distance from the perimeter is no more than $n/2$, so by the observation above, the number of units of congestion summed along the path to the perimeter is at most $\sum_{i=1}^{n/2} i$. By symmetry, the sum along the path back to the right end has the same bound. The edges between the perimeter and the vertex representing the zero boundary each have congestion of at most $n/2$ units. Since the edges in R^+ have less than unit weight, the congestion must be scaled; multiplying the total units of congestion by the maximum element of R^+ gives an upper bound. Recalling that the off-diagonals in R^+ are half the size of those in R, we have $\max(R^+) = \max(R)/2$. Thus we can bound the path congestion above

by the following:

$$\frac{\max(R)}{2}\left(2\sum_{i=1}^{n/2} i + 2\cdot\frac{n}{2}\right) = \max(R)\left(\frac{n^2}{8}+\frac{3n}{4}\right).$$

Recall that the bound on the desired eigenvalue is actually larger by 1, giving the following:

$$\lambda_{max}\left(A^{-1}(A+R)\right) \le \max(R)\left(\frac{n^2}{8}+\frac{3n}{4}\right)+1.$$

Combining this with our previous result and the bound on the elements of R gives

$$\begin{aligned}\kappa\left((A+R)^{-1}A\right) &\le \left(\max(R)\left(\frac{n^2}{8}+\frac{3n}{4}\right)+1\right)/\left(1-\max(R)\right)\\ &\le \left(\sqrt{2}-1\right)\left(\frac{n^2}{8}+\frac{3n}{4}\right)+\frac{2}{2-\sqrt{2}}.\end{aligned}$$

6.2. Modified Incomplete Factorization. Embedding techniques can also be used in a qualitative way to explain why certain methods behave as they do or to examine strategies for producing preconditioners based on incomplete factorization. As an example, we will discuss intuitive ideas supporting the use of modified incomplete factorizations. In such factorizations, the preconditioner no longer agrees with the original matrix at all entries specified in S. Instead, agreement is enforced at off-diagonal points, and diagonal entries are modified so that the row sums of the preconditioner B agree with the row sums of A.

We start with an example that illustrates the benefits of enforcing consistent row sums. The preconditioner involved is not practical because computing $B^{-1}\mathbf{x}$ is difficult, but it clearly illustrates the main points. Let A be a positive definite generalized Laplacian, L be the lower triangular incomplete factor of A, $B = LL^T$, and $R = B - A$. Let D be a diagonal matrix with d_{ii} equal to the sum of the elements in row i of R. Let $B_{mod} = LL^T - D$; that is, B_{mod} is B modified so that its row sums are the same as those of A.

What is $\kappa(B_{mod}^{-1}A)$? Let $R_{mod} = B_{mod} - A$. Because of the change in the diagonal, R_{mod} is the negative of a Laplacian: $R_{mod}^{-} = R_{mod}$ and $R_{mod}^{+} = 0$. Recall that the upper bound on $\lambda_{max}\left(A^{-1}(A+R_{mod})\right)$ is τ such that $(\tau-1)A - R_{mod}^{+}$ is positive semidefinite. Since $R_{mod}^{+} = 0$, $\tau = 1$ will do. The upper bound on $\lambda_{max}\left((A+R_{mod})^{-1}A\right)$ is $\tau = 1/(1-\alpha)$, where $0 < \alpha < 1$ and $\alpha A - R_{mod}^{-}$ is positive semidefinite. The size of the off-diagonal entries in R_{mod}^{-} are twice the size of those in R^{-}, so α is twice as big in the modified case as it is in the unmodified case.

This suggests a number of interesting things. First, the decrease in the bound on $\lambda_{max}\left(A^{-1}(A+R)\right)$ is often substantial because we no longer need to deal with the expanded versions of A and R^{+}. Often the number of edges in R^{+} is sufficient to cause large congestion through the boundary vertex, and the paths from interior vertices can be relatively long. With the modifications we are considering, these factors disappear: In the case of the naturally ordered square grid with boundary as above, the decrease is from a value that is $\Theta(n^2)$ to a constant factor (1).

Second, the bound on $\lambda_{max}\left((A+R)^{-1}A\right)$ increases, though if α is small in the unmodified case, the increase is also small. For the model square grid problem, $\max(R)$ is less than 1/3. The upper bound on $\lambda_{max}\left((A+R_{mod})^{-1}A\right)$ therefore at most doubles.

These observations suggest that this (impractical) modified preconditioner can give large improvements in the spectral condition number over the unmodified version. (We note again that these are bounds on the spectral condition number, and that the spectral condition number only lets us find an upper bound on the convergence rate. However, experiments show that bounds on κ are good, and the connection to convergence rates is good for this model problem.)

The problem with this preconditioner is that it is hard to work with. In particular, we do not know of an easy way to compute $B^{-1}\mathbf{x}$, nor of a way to compute a close approximation of B that is easy to work with. However, there are modified incomplete factorization methods that do produce a preconditioner with row sums equal to those of A. The discussion above provides some intuition about why these modified incomplete Cholesky preconditioners do not give spectral condition numbers as small as $\kappa(B_{mod}^{-1}A)$. In particular, when the modified factorization algorithm encounters a fill entry (i, j) that is to be dropped (i.e., $[S]_{ij} = 0$), it reduces the diagonal entries for i and j. Because the columns in L are divided by the square roots of their diagonal entries, the entries in columns i and j are increased relative to their values in the unmodified factor. This increases the size of any fill entries produced by the entries in these columns; if any such fill entry is dropped, its corresponding entry in the error matrix of the preconditioner is likewise increased over the unmodified case. The dropped fill entries also cause decreases in subsequently ordered diagonal entries, allowing the effects to ripple through L. As a result, entries in the error matrix can become substantially larger than in the unmodified case.

In the model grid problem, the largest entries in the error matrix for modified incomplete Cholesky factorization approach 1/2, and the bound on $\lambda_{max}\left((A + R_{mod})^{-1}A\right)$ increases by more than a constant factor. Experiments suggest that the number of iterations needed for convergence increases as well.

REFERENCES

[1] O. Axelsson, *Bounds of eigenvalues of preconditioned matrices*, SIAM Journal on Matrix Analysis and Applications, 13 (1992), pp. 847–862.

[2] ——, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.

[3] T. F. Chan and H. A. van der Vorst, *Approximate and incomplete factorizations*, tech. report, Department of Math, University of Utrecht, 1994. Preprint 871.

[4] P. Diaconis and D. Stroock, *Geometric bounds for eigenvalues of Markov chains*, Annals of Applied Probability, 1 (1991), pp. 36–61.

[5] P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*, The Mathematical Association of America, Washington, D.C., 1984.

[6] V. Eijkhout, *Analysis of parallel incomplet point factorizations*, Linear Algebra and Its Applications, 154–156 (1991), pp. 723–740.

[7] M. Fiedler, *Algebraic connectivity of graphs*, Czechoslovak Mathematical Journal, 23 (1973), pp. 298–305.

[8] ——, *A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory*, Czechoslovak Mathematical Journal, 25 (1975), pp. 619–633.

[9] G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1989.

[10] K. Gremban, G. Miller, and M. Zagha, *Performance evaluation of a parallel preconditioner*, in 9th International Parallel Processing Symposium, Santa Barbara, April 1995, IEEE, pp. 65–69.

[11] K. D. Gremban, *Combinatorial Preconditioners for Sparse, Symmetric, Diagonally Dominant Linear Systems*, PhD thesis, Carnegie Mellon University, Pittsburgh, October 1996. CMU CS Tech Report CMU-CS-96-123.

[12] S. Guattery, F. T. Leighton, and G. Miller, *The Path Resistance Method for bounding λ_2 of a Laplacian*, in 8th ACM-SIAM Symposium on Discrete Algorithms, New Orleans, January 1997, ACM-SIAM. To appear.

[13] S. Guattery and G. L. Miller, *Embeddings and laplacian eigenvalues*, (1997). Unpublished Manuscript. Please contact S. Guattery for copies.

[14] M. Jerrum and A. Sinclair, *Approximating the permanent*, SIAM Journal on Computing, 18 (1989), pp. 1149–1178.

[15] N. Kahale, *A semidefinite bound for mixing rates of Markov chains*, in Proc. Fifth Integer Programming and Combinatorial Optimization Conference, 1996, pp. 190–203. Lecture Notes in Computer Science 1084.

[16] A. Sinclair, *Improved bounds for mixing rates of Markov chains and multicommodity flow*, Combinatorics, Probability and Computing, 1 (1992), pp. 351–370.

[17] A. Sinclair and M. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Information and Computation, 82 (1989), pp. 93–133.