OBTAINING BOUNDS ON THE TWO NORM OF A MATRIX FROM THE SPLITTING LEMMA*

DORON CHEN†, JOHN R. GILBERT‡, AND SIVAN TOLEDO†

Abstract. The splitting lemma is one of the two main tools of *support theory*, a framework for bounding the condition number of definite and semidefinite preconditioned linear systems. The splitting lemma allows the analysis of a complicated system to be partitioned into analyses of simpler systems. The other tool is the symmetric-product-support lemma, which provides an explicit spectral bound on a preconditioned matrix.

The symmetric-product-support lemma shows that under suitable conditions on the null spaces of A and B, the finite eigenvalues of the pencil (A, B) are bounded by $\|W\|_2^2$, where $U = VW$, $A = UU^T$, and $B = VV^T$. To apply the lemma, one has to construct a W satisfying these conditions, and to bound its 2-norm.

In this paper we show that in all its existing applications, the splitting lemma can be viewed as a mechanism to bound $\|W\|_2^2$ for a given W. We also show that this bound is sometimes tighter than other easily-computed bounds on $\|W\|_2^2$, such as $\|W\|_F^2$ and $\|W\|_1 \|W\|_\infty$. The paper shows that certain regular splittings have useful algebraic and combinatorial interpretations. In particular, we derive six separate algebraic bounds on the 2-norm of a real matrix; to the best of our knowledge, these bounds are new.

Key words. matrix norm bounds, two-norm, norm bounds for sparse matrices, splitting lemma, support theory, support preconditioning

AMS subject classifications. 15A60, 65F10, 65F35, 65F50

1. Introduction.

Support theory [1, 5] is a set of tools used to bound the condition numbers of preconditioned systems. Support theory employs two devices to bound the support of a preconditioner: one is the splitting lemma [1, 9], and the other is the symmetric product support lemma [5]. In this paper we compare bounds which arise from these two tools, and introduce new bounds.

Conjugate Gradient (CG) is a common iterative algorithm for solving symmetric positive-definite linear systems $Ax = b$. Given a symmetric positive-definite matrix A, the number of iterations needed to reduce the norm of the residual by a constant factor is bounded by the spectral condition number of A. The spectral condition number $\kappa(A)$ is the ratio of the extreme eigenvalues of A, $\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$. The convergence of CG, as well as of many other iterative solvers, can often be improved by use of a preconditioner B. When using a preconditioner, the number of iterations needed for convergence is bounded by the condition number of the preconditioned system, which is the ratio of the extreme *finite eigenvalues* of the matrix pencil (A, B), defined as follows.

DEFINITION 1.1. *The number λ is a finite generalized eigenvalue of the matrix pencil (A, B) if there exists a vector x such that $Ax = \lambda Bx$ and $Bx \neq 0$.*

Support theory [1, 5] is a framework for bounding the condition number of definite and semidefinite preconditioned linear systems. In early support-theory papers [1, 9], three main tools were used: the support lemma, the splitting lemma and the congestion-dilation lemma. The support lemma showed how to bound the finite eigenvalues of (A, B) in terms of a number $\sigma(A, B)$ called the *support* of (A, B).

*This research was supported in part by an IBM Faculty Partnership Award, by grants 572/00 and 848/04 from the Israel Science Foundation (founded by the Israel Academy of Sciences and Humanities), and by grant 2002261 from the United-States-Israel Binational Science Foundation.

†School of Computer Science, Tel-Aviv University.

‡University of California, Santa Barbara.

The splitting lemma shows that $\sigma(A, B) \leq \max_i \sigma(A_i, B_i)$, where $A = \sum_i A_i$ and $B = \sum_i B_i$. The congestion-dilation lemma showed how to directly bound $\sigma(A_i, B_i)$ when A_i and B_i are particularly simple: when the graph of A_i consists of a single edge, and the graph of B_i is a simple path between that edge's endpoints[1]. In these early papers all the matrices involved had to be diagonally dominant, but that is irrelevant for our paper. In essence, the splitting lemma allowed a complex problem to be broken into simple parts, and the congestion-dilation lemma allowed each part to be analyzed.

Boman and Hendrickson later presented support theory in a completely algebraic framework, which does not refer to graphs, paths, and so on [5]. Their framework still used the support lemma, but they replaced much of the rest with a single powerful lemma, the symmetric-product-support lemma. This lemma shows that under suitable conditions on the null spaces of A and B, the finite eigenvalues of the pencil (A, B) are bounded by $\|W\|_2^2$, where $U = VW$, $A = UU^T$, and $B = VV^T$. To apply the lemma, one has to construct a W satisfying these conditions, and to bound its 2-norm. They also show that the bounds that were previously derived by the splitting and congestion-dilation lemmas can be directly obtained by applying their new lemma together with the norm bound $\|W\|_2^2 \leq \|W\|_1 \|W\|_\infty$. It seemed that the splitting lemma was no longer useful.

However, recent results by Spielman and Teng again used the splitting lemma [14, 16]. What, then, is the role of the splitting lemma in the Boman-Hendrickson symmetric-product-support framework? This paper shows that in all its existing applications [1, 9, 14, 16], the splitting lemma can be viewed as a mechanism to bound $\|W\|_2^2$ for a given W. We also show that this bound is sometimes tighter than other easily-computed bounds on $\|W\|_2^2$, such as $\|W\|_F^2$ and $\|W\|_1 \|W\|_\infty$.

We also show that certain regular splittings have useful combinatorial interpretations. These interpretations can be exploited to construct and analyze graph algorithms for constructing preconditioners, such as the algorithms in [3, 9, 14, 16, 17]. In particular, one of these interpretations was used, with a different proof, in [14].

Path embeddings have also been used to bound the smallest nonzero eigenvalue of Laplacian matrices. To do so, one embeds the complete graph in the target graph. Our bounds apply to embeddings of arbitrary graphs, so they are more general. However, special cases of some of our bounds have already been discovered in the more restricted case [7, 10, 11, 12, 13].

This paper is organized as follows. The next section describes the basic results of support theory. Section 3 proves the splitting lemma and shows that the symmetric-product-support lemma implies it. Section 4 describes our main technical tools, orthonormal stretchings and fractional splittings. Section 5 proposes two splitting heuristics and shows that they lead to new algebraic and combinatorial bounds on the 2-norm of a matrix. Section 6 shows two additional bounds on the 2-norm. Section 7 quantifies the behavior of each one of the new norm bounds on an example. In particular, the example shows that the different bounds can be asymptotically different, some tight and some loose. Section 8 presents our conclusions.

2. Background. This section provides key definitions and known lemmas that we use in the rest of the paper. We start with the definition of support and with the support lemma.

[1]The graph G_A of an n-by-n symmetric matrix A is a weighted graph $G = (V, E, w)$, where $V = \{1, 2, \ldots, n\}$, $E = \{(i, j) \colon A_{ij} \neq 0\}$, and the weight of an edge $w(i, j)$ is $w(i, j) = -A_{ij}$.

2.1. Support. DEFINITION 2.1. *The support $\sigma(A, B)$ of a matrix pencil (A, B) is the smallest number τ such that $\tau B - A$ is positive semidefinite. If there is no such number, we take $\sigma(A, B) = \infty$.*

The importance of support numbers stems from the following lemma:

LEMMA 2.2. *(Support Lemma [9]) If λ is a finite generalized eigenvalue of (A, B) and B is positive semidefinite and $\mathrm{null}(A) \subseteq \mathrm{null}(B)$, then $\lambda \leq \sigma(A, B)$. When $\sigma(A, B)$ is finite, the bound is tight.*

Next, we state the key result in Boman and Hendrickson's support framework.

LEMMA 2.3. *(Symmetric-product-support lemma [5]) Suppose $U \in \mathbb{R}^{n \times k}$ is in the range of $V \in \mathbb{R}^{n \times p}$. Then*

$$\sigma\left(UU^T, VV^T\right) = \min_W \|W\|_2^2 \ \textit{subject to}\ VW = U \ .$$

2.2. Combinatorial Interpretations of Support Bounds. Lemma 2.3 is often used after factoring the n-by-n coefficient matrix A into $A = UU^T$, where U is n-by-m and the preconditioner B into $B = VV^T$, where V is n-by-k (note that there are no special conditions on V and U; they need not be triangular or orthogonal). Typically, the columns of U are *edge vectors* [3], i.e. each column of U corresponds to one off-diagonal in the matrix A. Similarly, each column of V corresponds to one off-diagonal of B. This particular factorization is used when A and B are symmetric diagonally-dominant (SDD) matrices. A matrix A that can be decomposed into $A = UU^T$ where U has at most two nonzeros per column is called a *factor-width-2* matrix; the properties of such matrices have been explored in [4, 6]. When A is factor-width-2 and has nonpositive offdiagonal entries, it is often called a weighted *Laplacian* matrix.

When A is decomposed into $A = UU^T$ in this way, every column of U corresponds to an edge in the graph G_A of A or to a vertex in G_A, and similarly for $B = VV^T$. In this case, any matrix W satisfying $U = VW$ can be seen as an *embedding* of G_A in G_B. Suppose that column j of U is an edge vector in G_A (otherwise it is a vertex vector). Then $U_{:,j} = VW_{:,j}$ (we use Matlab notation, in which a colon represents all the possible indices). The nonzero entries in $W_{:,j}$ specify a set of edge and vertex vectors in V. We say that the edge in G_A is embedded into this set of edges and vertices in G_B. In some support preconditioners, the embedding is always of edges into simple paths and vertices into single vertices [1, 9, 17]. In other support preconditioners, some edges are embedded into up to two cycles and up to two paths [3].

In fact, the analysis of the preconditioner usually goes in the other direction. One first shows that given A and B, there exists a "good" embedding of G_A into G_B. Then, from this embedding, one shows how to construct W. Finally, some bound on $\|W\|_2^2$ is proven, and this bounds the finite spectrum of the preconditioned system. Common bounds on $\|W\|_2^2$ that have been used in support preconditioners are $\|W\|_2^2 \leq \|W\|_1 \|W\|_\infty$, which has been used implicitly in [1, 3, 9, 17], and $\|W\|_2^2 \leq \|W\|_F^2$, which is used in [2]. In this setup, a good embedding is one that leads to a small norm bound, that is, to a small value for $\|W\|_1 \|W\|_\infty$ or $\|W\|_F^2$.

When W is an embedding of edges into simple paths and of vertices into vertices, the two bounds $\|W\|_2^2 \leq \|W\|_1 \|W\|_\infty$ and $\|W\|_2^2 \leq \|W\|_F^2$ have useful combinatorial interpretations. The first can be interpreted as product of the worst dilation of a path times the worst congestion through an edge of G_B. Here the dilation of a path π between the endpoints of an edge $e \in G_A$ is defined to be $\sum_{e' \in \pi} |W_{e',e}|$. The congestion through an edge $e' \in G_B$ is defined to be $\sum_{e\colon e' \in \pi(e)} |W_{e',e}|$, where $\pi(e)$

is the path that embeds e. The bound $\|W\|_2^2 \leq \|W\|_F^2$ can be interpreted as the sum of all the dilations of all the paths, but with a different definition for dilation, $\sum_{e' \in \pi} W_{e',e}^2$.

2.3. The Splitting Lemma. We now state formally the Splitting Lemma, which is the focus of this paper.

LEMMA 2.4. *(The Splitting Lemma) Let $A = A_1 + A_2 + \ldots + A_q$ and let $B = B_1 + B_2 + \ldots + B_q$. If all A_i and B_i are symmetric positive semidefinite, and if for each i, A_i is in the range of B_i, then $\sigma(A, B) \leq \max_i \sigma(A_i, B_i)$.*

Typically, this lemma is used by decomposing A into a sum of rank-1 matrices, each corresponding to one off-diagonal, and by decomposing B into path matrices, matrices that can be symmetrically permuted to a tridiagonal form, and which have only one nonzero irreducible block.

In the rest of this paper, we focus on symmetric positive semidefinite matrices, but we do not assume that they are diagonally-dominant (unless specified otherwise).

3. The Symmetric Product Support Lemma Implies the Splitting Lemma. What is the relationship of the splitting lemma to the symmetric-product-support lemma? In this section we begin the study of this question. This section shows that the splitting lemma is weaker, in the sense that the symmetric-product-support lemma implies the splitting lemma. The following proof proves Lemma 2.4 using a straightforward application of the symmetric-product support lemma.

Proof. Let $A_i = U_i U_i^T$ and $B_i = V_i V_i^T$ be arbitrary symmetric-product decompositions of A_i and B_i. Such a decomposition always exists, given our assumption that both matrices are symmetric positive semidefinite. For example, we can use the scaled eigenvectors of A_i as the columns of U_i, where the scaling is by the square root of the corresponding eigenvalue, and similarly for B_i. Let U_S be the concatenation of $U_1, U_2, \ldots, U_q$ and V_S be the concatenation of $V_1, V_2, \ldots, V_q$. That is,

$$U_S = \begin{pmatrix} U_1 & U_2 & U_3 & \cdots & U_q \end{pmatrix},$$

and similarly for V_S. Then $U_S U_S^T = \sum_i U_i U_i^T = \sum_i A_i = A$, and $V_S V_S^T = B$.

By the assumption that A_i is in the range of B_i, the factor U_i must be in the range of V_i. Therefore, there exists a W_i such that $U_i = V_i W_i$.

Let $\hat{W}_i$ be the minimizer of $\min_{W_i} \|W_i\|_2$ subject to $U_i = V_i W_i$. By the Symmetric-Product Support lemma, $\sigma(A_i, B_i) = \left\|\hat{W}_i\right\|_2^2$.

Let

$$W = \begin{pmatrix} \hat{W}_1 & & & \\ & \hat{W}_2 & & \\ & & \ddots & \\ & & & \hat{W}_q \end{pmatrix}.$$

We claim that $V_S W = U_S$.

$$
\begin{aligned}
V_S W &= \begin{pmatrix} V_1 & V_2 & V_3 & \cdots & V_q \end{pmatrix} \begin{pmatrix} \hat{W}_1 & & & \\ & \hat{W}_2 & & \\ & & \ddots & \\ & & & \hat{W}_q \end{pmatrix} \\
&= \begin{pmatrix} V_1\hat{W}_1 & V_2\hat{W}_2 & V_3\hat{W}_3 & \cdots & V_q\hat{W}_q \end{pmatrix} \\
&= \begin{pmatrix} U_1 & U_2 & U_3 & \cdots & U_q \end{pmatrix} \\
&= U_S
\end{aligned}
$$

The norm of W is equal to $\max_i \left\|\hat{W}_i\right\|$, so by the Symmetric-Product Support Lemma it follows that $\sigma(A,B) \le \|W\|_2^2 = \max_i \left\|\hat{W}_i\right\|_2^2 = \max_i \sigma(A_i, B_i)$. ◻

4. Splitting and Stretching. In this section we show a deeper connection between splitting and the symmetric-product-support lemma. We begin by defining an operation called orthonormal stretching, which allows us to obtain one symmetric-product-support triplet from another. We then show that an important class of splittings, the one which has been used almost exclusively in applications, can be interpreted as an orthonormal stretching. That is, splitting is usually a way to obtain one symmetric-product-support triplet from another, and in particular, to obtain a triplet in which computing $\|W\|_2$ is easy.

4.1. Orthonormal Stretching. The *orthonormal stretching* of a symmetric-product-support triplet (U,V,W) is a pair of matrices $(S,\tilde{W})$: a k-by-$\tilde{k}$ matrix S with orthonormal rows ($SS^T = I$), and a matrix $\tilde{W}$ such that $W = S\tilde{W}$.

Why is stretching important for support theory? Because, as the next lemma shows, the triplet $(U, VS, \tilde{W})$ is also a symmetric-product-support triplet, because various norms of $\tilde{W}$ bound the corresponding norms of W and because $\sigma(UU^T, VV^T) \le \left\|\tilde{W}\right\|_2^2$.

Therefore, orthonormal stretching is useful when it allows us to take a symmetric-product-support triplet (U,V,W), for which bounding the norm of W is difficult, and obtain a new triplet $(U, VS, \tilde{W})$, for which bounding the norm of $\tilde{W}$ is easier. The norm of $\tilde{W}$ still bounds $\sigma(UU^T, VV^T)$, which in turn bounds the spectrum of the preconditioned linear system.

LEMMA 4.1. *Let $U = VW$, and let S and $\tilde{W}$ be an orthonormal stretching of (U,V,W), so $W = S\tilde{W}$. Then, using the notation $\tilde{V} = VS$, the following hold:*

1. $\tilde{V}\tilde{V}^T = VV^T$
2. $V = \tilde{V}S^T$
3. $U = \tilde{V}\tilde{W}$
4. $\|W\|_2 \le \left\|\tilde{W}\right\|_2$
5. $\|W\|_F \le \left\|\tilde{W}\right\|_F$
6. $\|W\|_\infty \le \sqrt{\tilde{k}}\left\|\tilde{W}\right\|_\infty$*, where $\tilde{k}$ is the number of columns in S.*
7. $\|W\|_1 \le \sqrt{\tilde{k}}\left\|\tilde{W}\right\|_1$

Proof. Most of the claims are nearly trivial.

1. $\tilde{V}\tilde{V}^T = (VS)(S^TV^T) = V(SS^T)V = VV^T$

2. $V = VI = V(SS^T) = (VS)S^T = \tilde{V}S^T$
3. $U = VW = V(S\tilde{W}) = (VS)\tilde{W} = \tilde{V}\tilde{W}$
4. $\|W\|_2 = \left\|S\tilde{W}\right\|_2 \le \|S\|_2 \left\|\tilde{W}\right\|_2 = \left\|\tilde{W}\right\|_2$, because $\|S\|_2 = 1$.
5. To show that $\|W\|_F \le \left\|\tilde{W}\right\|_F$, we need only compare each column of W to the corresponding column in $\tilde{W}$. Let w_j ($\tilde{w}_j$) be column j of W (of $\tilde{W}$). Then $w_j = S\tilde{w}_j$, so $\|w_j\|_2 = \|S\tilde{w}_j\|_2 \le \|S\|_2 \|\tilde{w}_j\|_2 = \|\tilde{w}_j\|_2$. Since the norm of each column of $\tilde{W}$ is greater or equal to the norm of the corresponding column of W, $\|W\|_F \le \left\|\tilde{W}\right\|_F$.
6. Since $W = S\tilde{W}$, we have $\|W\|_\infty = \left\|S\tilde{W}\right\|_\infty \le \|S\|_\infty \left\|\tilde{W}\right\|_\infty$. We prove the claim by bounding the ∞-norm of S, $\|S\|_\infty = \max_i \sum_{j=1}^{\tilde{k}} |S_{ij}|$. Each row in S is a size-$\tilde{k}$ vector with unit norm. Let $\mathrm{sum}(v)$ be the sum of the absolute values of the entries of a vector v. It is easy to show that the maximum of $\mathrm{sum}(v)$, over all the vectors v with norm 1, is obtained when all the entries of v are equal. The sum of the entries, for the maximal vector, is the square root of the size of the vector. Therefore, for any row i of S we have $\sum_{j=1}^{\tilde{k}} |S_{ij}| \le \sqrt{\tilde{k}}$, so $\|S\|_\infty \le \sqrt{\tilde{k}}$, which proves the claim.
7. Let S' be a completion of S to a $\tilde{k}$-by-$\tilde{k}$ orthonormal matrix. Then $\|S\|_1 \le \|S'\|_1$ because

$$\|S\|_1 = \max_j \sum_{i=1}^{k} |S_{ij}| = \max_j \sum_{i=1}^{k} \left|S'_{ij}\right| \le \max_j \sum_{i=1}^{\tilde{k}} \left|S'_{ij}\right| = \|S'\|_1 \ .$$

An equivalent argument to that of claim 6 shows that $\|S'\|_1 \le \tilde{k}$, which proves the claim.

□

But how do we find a useful orthonormal stretching $(S, \tilde{W})$, a stretching for which the norm of $\tilde{W}$ is easy to bound? The next part of this section shows that in many cases, splitting can be interpreted as such a stretching.

4.2. Orthonormal Stretching Via Fractional Splitting. In this section we explain the connection between the orthonormal stretching and splitting.

DEFINITION 4.2. *A* splitting set *for an n-by-m matrix U and an n-by-k matrix V is a set $D_1, D_2, \ldots, D_m$ of k-row matrices satisfying*

- $\sum_{j=1}^{m} D_j D_j^T = I$, *and*
- *for each j, $U_{:,j}$ is in the range of VD_j.*

LEMMA 4.3. *Let $D_1, D_2, \ldots, D_m$ be a splitting set for U and V. Then $A_j = (U_{:,j})(U_{:,j})^T$ and $B_j = (VD_j)(VD_j)^T$ is a splitting of $A = UU^T$ and $B = VV^T$ in the sense of the splitting lemma. That is, $A = \sum_{i=1}^{m} A_i$, $B = \sum_{i=1}^{m} B_i$, and each A_i is in the range of B_i.*

Proof. Clearly $A = \sum_{j=1}^{m} A_j$. The sum of the B_j's satisfies

$$\begin{aligned}
\sum_{j=1}^{m} B_i &= \sum_{j=1}^{m} (VD_j)(VD_j)^T \\
&= \sum_{j=1}^{m} VD_jD_j^TV^T \\
&= V\left(\sum_{j=1}^{m} D_jD_j^T\right)V^T \\
&= VV^T \\
&= B\ .
\end{aligned}$$

Since for each j, $U_{:,j}$ is in the range of VD_j, each A_j is in the range B_j. □

A splitting set can be difficult to construct, due to the second condition in its definition. But a W satisfying $U = VW$ offers an opportunity to create a special family of splitting sets.

Definition 4.4. *Let U be an n-by-m matrix, V an n-by-k matrix, and W a k-by-m matrix such that $U = VW$. A* fractional splitting set *for U, V, and W is a set of k-by-k diagonal matrices $D_1, D_2, \ldots, D_m$ satisfying the following conditions.*

- *The indices of the nonzero diagonal entries in D_j is the set $\{i\colon W_{i,j} \neq 0\}$.*
- $\sum_{j=1}^{m} D_jD_j^T = I$.

Lemma 4.5. *A fractional splitting set is a splitting set.*

Proof. We need to show that for each j, $U_{:,j}$ is in the range of VD_j. Let D_j^+ be the Moore-Penrose pseudo-inverse of D_j. Since D_j is diagonal, D_j^+ is also diagonal, with

$$\left(D_j^+\right)_{ii} = \begin{cases} (D_j)_{ii}^{-1} & (D_j)_{ii} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

(see, for instance, [8, Section 5.5.4]). The matrix $D_jD_j^+$ is diagonal with zeros and ones on the diagonal, with ones in positions that correspond to nonzeros in $W_{:,j}$. Therefore, $D_jD_j^+W_{:,j} = W_{:,j}$, so $VD_jD_j^+W_{:,j} = VW_{:,j}$. Therefore,

$$\begin{aligned}
VD_j\left(D_j^+W_{:,j}\right) &= VW_{:,j} \\
&= U_{:,j}\ ,
\end{aligned}$$

which proves the claim. □

We now show that a fractional splitting set defines not only a splitting of A and B, but also an orthonormal stretching of V and W. We begin by showing how to derive S from the D_j's.

Lemma 4.6. *Let $D_1, D_2, \ldots, D_m$ be a splitting set. Then the concatenation S of the D_j's, $S = \begin{pmatrix} D_1 & D_2 & D_3 & \cdots & D_m \end{pmatrix}$ has orthonormal rows (the concatenation matrix S consists of the columns of D_1 followed by the columns of D_2, and so on).*

Proof. $SS^T = \sum_{j=1}^{m} D_jD_j^T = I$. □

Clearly, the proof of the previous lemma only relies on one of the two conditions that splitting sets must satisfy.

Next, we show how to construct $\tilde{W}$. The example in the beginning of Section 5 illustrates this construction.

LEMMA 4.7. *Let $D_1, D_2, \ldots, D_m$ be a fractional splitting set for some U, V, and W, and let S be defined as in Lemma 4.6. Let*

$$\tilde{W}_{:,j} = \left(\; 0\,; \quad 0\,; \quad \cdots \quad 0\,; \quad D_j^{+} W_{:,j}\,; \quad 0\,; \quad \cdots \quad 0\,; \quad 0 \; \right) ,$$

where 0 denotes the k-by-1 zero vector. (We use the Matlab notation: a semicolon denotes stacking blocks, so $(A \; ; \; B) = (A^T \;\; B^T)^T$.) That is, in the first column of $\tilde{W}$ the first k elements are $D_1^{+} W_{:,1}$ and the rest are zeros. The second column of $\tilde{W}$ starts with k zeros, then the elements of $D_2^{+} W_{:,2}$, followed by zeros, and so on. Then $W = S\tilde{W}$.

Proof. We prove the lemma column by column,

$$\begin{aligned}
\left(S\tilde{W}\right)_{:,j} &= S\tilde{W}_{:,j} \\
&= \left(\; D_1 \quad D_2 \quad \cdots \quad D_{j-1} \quad D_j \quad D_{j+1} \quad \cdots \quad D_{m-1} \quad D_m \; \right) \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ D_j^{+} W_{:,j} \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \\
&= \left(\sum_{i \neq j} D_i \cdot 0 \right) + D_j D_j^{+} W_{:,j} \\
&= 0 + W_{:,j} \; .
\end{aligned}$$

∎

An important benefit of using a fractional splitting to define an orthonormal stretching $(S, \tilde{W})$ is that the 2-norm of $\tilde{W}$ is easy to compute.

LEMMA 4.8. *Let $(S, \tilde{W})$ be an orthonormal stretching defined by a fractional splitting as in Lemmas 4.6 and 4.7. Then*

$$\left\| \tilde{W} \right\|_2 = \max_{j=1}^{m} \left\| \tilde{W}_{:,j} \right\|_2 \; .$$

Proof. By the construction of $\tilde{W}$ given in Lemma 4.7 it is clear that its columns are orthogonal. ∎

In general, splittings and symmetric products are not isomorphic. A splitting $A = \sum A_i$ and $B = \sum B_i$ does not define symmetric products $A = UU^T$ and $B = VV^T$, not even implicitly. Also, a symmetric-product-support triplet does not define a splitting. But we have shown that an important class of splittings does define an orthonormal stretching, a way to get one symmetric-product-support triplet from another.

In most of the applications of the splitting lemma [9, 1, 14, 16], there is also a symmetric-product representation of A and B, a representation using edge and vertex vectors, and an implicit W. Furthermore, in these applications, the splitting of A and B can almost always be interpreted as a fractional splitting set $D_1, \ldots, D_m$

of the symmetric-product factors U and V. In all of these cases, the splitting can be interpreted as an orthonormal stretching of a symmetric-product-support triplet.

Before we conclude this section, we show that for orthonormal stretchings derived from fractional splitting sets, one of the norm-bounds on $\tilde{W}$ can be tightened.

LEMMA 4.9. *Let $(S, \tilde{W})$ be an orthonormal stretching of (U, V, W), derived from a fractional splitting set, as defined in Lemmas 4.6 and 4.7. Then* $\|W\|_1 \le \left\|\tilde{W}\right\|_1$.

Proof. We show that $\|S\|_1 \le 1$. By definition, $\|S\|_1 = \max_j \sum_{i=1}^{k} |S_{ij}|$. By the construction of S in Lemma 4.6, each column of S is a column of one of the D_j's. Each column of D_j has exactly one nonzero. Because $\sum_j D_j D_j^T = I$, that nonzero must be no larger than 1 in absolute value. Therefore, each column of S has exactly one nonzero no larger than 1 in absolute value, which proves the claim that $\|S\|_1 \le 1$. Therefore $\|W\|_1 = \left\|S\tilde{W}\right\|_1 \le \|S\|_1 \left\|\tilde{W}\right\|_1 \le \left\|\tilde{W}\right\|_1$. ∎

Note that whenever each column of S has a single nonzero, $\|W\|_1 \le \left\|\tilde{W}\right\|_1$, even if $\tilde{W}$ was not derived from a fractional splitting set. In particular, the 1-norm bound given in the previous lemma may hold even when the columns of $\tilde{W}$ are not orthogonal. When $(S, \tilde{W})$ are obtained from a fractional splitting set, the bound $\|W\|_2^2 \le \|W\|_1 \|W\|_\infty \le \sqrt{k} \left\|\tilde{W}\right\|_1 \left\|\tilde{W}\right\|_\infty$ is not particularly useful, because we can directly compute $\left\|\tilde{W}\right\|_2$. But in more general cases this bound may be useful.

5. How to Split. The choice of D_j's in a fractional splitting can have a profound influence on how close $\left\|\tilde{W}\right\|_2$ is to $\|W\|_2$. We use a fractional splitting because $\left\|\tilde{W}\right\|_2$ is easy to compute and it bounds $\|W\|_2$. In this section we show that a poor choice of D_j's can lead to $\left\|\tilde{W}\right\|_2$ being so large that it teaches us nothing about $\|W\|_2$. We also suggest two simple and efficient heuristics to find splittings with a small $\left\|\tilde{W}\right\|$. From one of these heuristics we obtain two combinatorial bounds on support preconditioners; one of these bounds was already suggested in a more general form by Spielman and Teng [14] using an entirely different proof, and the other is new.

5.1. An Example. We first show that if the choice of splitting is poor, then the resulting norm bound is useless. Let

$$A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}, U = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & -1 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}, V = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

To complete U and V to a symmetric-product-support triplet, we use

$$W = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

The 2-norm of W is $\|W\|_2 = \sqrt{3}$.

We now split A and B using the following fractional support set,

$$D_1 = \begin{pmatrix} \epsilon & 0 \\ 0 & 0 \end{pmatrix}, D_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix}, D_3 = \begin{pmatrix} \sqrt{1-\epsilon^2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix},$$

where $\epsilon > 0$ is small. Therefore,

$$S = \begin{pmatrix} D_1 & D_2 & D_3 \end{pmatrix} = \begin{pmatrix} \epsilon & 0 & 0 & 0 & \sqrt{1-\epsilon^2} & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix} .$$

We now construct $\tilde{W}$ according to lemma 4.7, starting with the pseudo-inverses,

$$D_1^+ = \begin{pmatrix} 1/\epsilon & 0 \\ 0 & 0 \end{pmatrix} , D_2^+ = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2} \end{pmatrix} , D_3^+ = \begin{pmatrix} 1/\sqrt{1-\epsilon^2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} ,$$

so

$$\tilde{W} = \begin{pmatrix} 1/\epsilon & & \\ 0 & & \\ & 0 & \\ & \sqrt{2} & \\ & & 1/\sqrt{1-\epsilon^2} \\ & & \sqrt{2} \end{pmatrix} .$$

For small ϵ, $\left\|\tilde{W}\right\|_2 = 1/\epsilon$ is arbitrarily large, so it is not a useful bound on $\|W\|_2 = \sqrt{3}$.

Clearly, $\epsilon = 1/\sqrt{2}$ is a better choice than a small ϵ, yielding a $\left\|\tilde{W}\right\|_2 = 2$, still not completely tight, but better. In this case, a fractional splitting can actually achieve $\left\|\tilde{W}\right\|_2 = \|W\|_2 = \sqrt{3}$. Let

$$S = \begin{pmatrix} \sqrt{1/3} & 0 & 0 & 0 & \sqrt{2/3} & 0 \\ 0 & 0 & 0 & \sqrt{1/3} & 0 & \sqrt{2/3} \end{pmatrix} ,$$

so

$$\tilde{W} = \begin{pmatrix} \sqrt{3} & & \\ 0 & & \\ & 0 & \\ & \sqrt{3} & \\ & & \sqrt{3/2} \\ & & \sqrt{3/2} \end{pmatrix} = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{3/2} \\ 0 & 0 & \sqrt{3/2} \end{pmatrix} .$$

(We show $\tilde{W}$ twice, with and without all the zeros, to emphasize the structure of its columns.) Still, the ϵ-example shows that a poor splitting yields useless bounds.

5.2. The Rowwise Heuristic . When $W = S\tilde{W}$, row i in W is a linear combination of a set of rows in $\tilde{W}$, where the coefficients of the linear combinations come from row i in S. When $(S, \tilde{W})$ is an orthonormal stretching derived from a fractional splitting set, the row sets that combine to form rows of W are disjoint. This is a consequence of the fact that columns in S have no more than a single nonzero. Therefore such stretchings map disjoint sets of rows of $\tilde{W}$ to the rows of W.

The first heuristic that we propose finds a fractional splitting that ensures that the 2-norm of each nonzero row of $\tilde{W}$ that maps onto row i in W is exactly $\|W_{i,:}\|_2$. This ensures, in a heuristic way, that $\tilde{W}$ is not too large.

Here is another way to interpret this heuristic. Under an orthonormal stretching derived from a fractional splitting set, each nonzero in W is mapped into a nonzero in

$\tilde{W}$. The rowwise heuristic ensures that all the nonzeros in $\tilde{W}$ that map to nonzeros in row i of W have the same magnitude.

LEMMA 5.1. *Let us define m diagonal matrices, such that the (i,i) value in the j-th matrix is*

$$(D_j)_{i,i} = \frac{W_{i,j}}{\|W_{i,:}\|_2} .$$

Then the D_j's are a fractional splitting set for W.

Proof. Clearly, the indices of the nonzero diagonal entries in D_j is the set $\{i \colon W_{i,j} \neq 0\}$.

We need to show that $\sum_{j=1}^{m} D_j D_j^T = I$. A sum of diagonal matrices is also diagonal. Therefore, $\sum_{j=1}^{m} D_j D_j^T = \sum_{j=1}^{m} D_j^2$ is diagonal. We need only show that each diagonal entry in $\sum_{j=1}^{m} D_j^2$ is 1. By definition, the (i,i) entry in D_j is $W_{i,j}/\|W_{i,:}\|_2$. The (i,i) entry in $\sum_{j=1}^{m} D_j^2$ is

$$\sum_{j=1}^{m} \frac{W_{i,j}^2}{\|W_{i,:}\|_2^2} = \frac{1}{\|W_{i,:}\|_2^2} \cdot \sum_{j=1}^{m} W_{i,j}^2 = \frac{1}{\|W_{i,:}\|_2^2} \cdot \|W_{i,:}\|_2^2 = 1 .$$

□

We now prove that this splitting preserves the 2-norm of rows in W. We need the following notation: $\eta(i,j) = (j-1)\cdot k + i$ (for $1 \le i \le k$ and $1 \le j \le m$). Matrix S is a concatenation of the D_j matrices. Therefore, S is a concatenation of m k-by-k matrices. $\eta(i,j)$ is the index of the column in S corresponding to the i-th column in D_j.

LEMMA 5.2. *Let $\tilde{W}$ be an orthonormal stretching of W derived using the rowwise fractional-splitting heuristic. If $W_{i,j} \neq 0$, then $\left\|\tilde{W}_{\eta(i,j),:}\right\|_2 = \left|\left(D_j^+ W_{:,j}\right)_i\right| = \|W_{i,:}\|_2$. Therefore, the norm of each nonzero row in $\tilde{W}$ is the norm of some row in W, and for each row in W, there is at least one row with the same norm in $\tilde{W}$.*

Proof. Rows in $\tilde{W}$ have at most a single nonzero, when $\tilde{W}$ is constructed from a fractional splitting. Therefore, for each nonzero row, the norm of a row is the absolute value of the single nonzero element in that row (the norm of a zero row is zero). All nonzeros in column j of $\tilde{W}$ are entries of $D_j^+ W_{:,j}$. Therefore, for each $1 \le i \le k$ and $1 \le j \le m$, $\left\|\tilde{W}_{\eta(i,j),:}\right\|_2 = \left|\left(D_j^+ W_{:,j}\right)_i\right|$. If $W_{i,j} \neq 0$, then $(D_j)_{i,i} = W_{i,j}/\|W_{i,:}\|_2 \neq 0$. Therefore $\left(D_j^+\right)_{i,i} = \|W_{i,:}\|_2 / W_{i,j}$ and thus:

$$\begin{aligned}\left(D_j^+ W_{:,j}\right)_i &= \left(D_j^+\right)_{i,i} \cdot W_{i,j} \\ &= \frac{\|W_{i,:}\|_2}{W_{i,j}} \cdot W_{i,j} \\ &= \|W_{i,:}\|_2 .\end{aligned}$$

This proves the lemma, because

$$\left\|\tilde{W}_{\eta(i,j),:}\right\|_2 = \left|(D_j^+ W_{:,j})_i\right| = \|W_{i,:}\|_2 .$$

□

We now prove a new bound on $\|W\|_2$, a bound which we later show has a useful combinatorial interpretation.

LEMMA 5.3. *Let W be a k-by-m real matrix. Then*

$$\|W\|_2^2 \le \max_j \sum_{i\colon W_{i,j}\neq 0} \|W_{i,:}\|_2^2 = \max_j \sum_{i\colon W_{i,j}\neq 0} \sum_{c=1}^{m} W_{i,c}^2 \ .$$

Proof. We stretch W orthonormally using the rowwise fractional-splitting heuristic. Then $\|W\|_2^2 \le \left\|\tilde{W}\right\|_2^2$, and the columns of $\tilde{W}$ are orthogonal, so

$$\left\|\tilde{W}\right\|_2^2 = \max_j \left\|\tilde{W}_{:,j}\right\|_2^2 \ .$$

All that remains to show is that $\left\|\tilde{W}_{:,j}\right\|_2^2 = \sum_{i\colon W_{i,j}\neq 0} \sum_{c=1}^{m} W_{i,c}^2$.

$$\begin{aligned}\left\|\tilde{W}_{:,j}\right\|_2^2 &= \left\|D_j^{+} W_{:,j}\right\|_2^2 \\ &= \sum_{i=1}^{k} \left(D_j^{+} W_{:,j}\right)_i^2 \\ &= \sum_{i\colon W_{i,j}\neq 0} \|W_{i,:}\|_2^2 \ .\end{aligned}$$

The last equality is by lemma 5.2. □

LEMMA 5.4. *Let W be a k-by-m real matrix. Then*

$$\|W\|_2^2 \le \max_i \sum_{j\colon W_{i,j}\neq 0} \|W_{:,j}\|_2^2 = \max_i \sum_{j\colon W_{i,j}\neq 0} \sum_{r=1}^{k} W_{r,j}^2 \ .$$

Proof. The previous lemma, applied to W^T, proves the claim, since $\left\|W^T\right\|_2 = \|W\|_2$. □

5.3. A Combinatorial Interpretation and the Spielman-Teng Bound. Given a symmetric-product-support triplet (U, V, W), a column of W can be viewed as an embedding of a column of U into the columns of V, since $U_{:,j} = VW_{:,j}$. The nonzero elements in the column of W specify a *generalized path* in V that supports the column in U. When U and V have at most two nonzeros per column, (that is, UU^T and VV^T have factor-width 2), they can be viewed as the weighted incidence matrices of G_{UU^T} and G_{VV^T}. In that case, a generalized path defined by a column of W is, indeed, an edge set, although this edge set may not form a simple path. We now define the dilation of a path that supports a column of U, and the congestion caused by the paths that utilize a column of V.

DEFINITION 5.5. *Let (U, V, W) be a symmetric-product-support triplet. We say that column i in V* supports *column j in U if $W_{i,j} \neq 0$. The* 2-dilation *of a column j in U is*

$$dil_2(U, V, W, j) = dil_2(j) = \|W_{:,j}\|_2^2 \ .$$

The 2-congestion *of a column i in V is*

$$cong_2(U, V, W, i) = cong_2(i) = \|W_{i,:}\|_2^2 \ .$$

These definitions, together with Lemmas 5.3 and 5.4, give the following results.

LEMMA 5.6. *Let (U, V, W) be a symmetric-product-support triplet. Then*

$$\|W\|_2^2 \leq \max_j \sum_{i \text{ supports } j} cong_2(i) \ .$$

The next lemma is a special case of the Spielman-Teng Support Theorem [14, Theorem 2.1]. Their proof technique, however, is different. The result stated in Lemma 5.6 is, to the best of our knowledge, new.

LEMMA 5.7. *Let (U, V, W) be a symmetric-product-support triplet. Then*

$$\|W\|_2^2 \leq \max_i \sum_{i \text{ supports } j} dil_2(j) \ .$$

5.4. The Frobenius Heuristic. Another approach to fractionally splitting W is to minimize the Frobenius norm of $\tilde{W}$. The Frobenius heuristic defines m diagonal matrices, such that the (i, i) value in the j-th matrix is

$$(D_j)_{i,i} = \frac{\sqrt{|W_{i,j}|}}{\sqrt{\sum_{c=1}^{m} |W_{i,c}|}} \ .$$

LEMMA 5.8. *The preceding definition of the D_j's defines a fractional splitting set.*

Proof. As in Lemma 5.1, we need to prove that $\sum_{j=1}^{m} D_j D_j^T = I$. Matrix $\sum_{j=1}^{m} D_j D_j^T = \sum_{j=1}^{m} D_j^2$, being the sum of diagonal matrices, is also diagonal. We need only show that each diagonal entry in $\sum_{j=1}^{m} D_j^2$ is 1. By definition, the (i, i) entry in D_j is $\sqrt{|W_{i,j}|}/\sqrt{\sum_{c=1}^{m} |W_{i,c}|}$. The (i, i) entry in $\sum_{j=1}^{m} D_j^2$ is

$$\sum_{j=1}^{m} \frac{|W_{i,j}|}{\sum_{c=1}^{m} |W_{i,c}|} = \frac{1}{\sum_{c=1}^{m} |W_{i,c}|} \cdot \sum_{j=1}^{m} |W_{i,j}| = 1 \ .$$

$\square$

LEMMA 5.9. *The Frobenius heuristic minimizes $\left\|\tilde{W}\right\|_F$ over all fractional splittings of W.*

Proof. We prove the lemma in two steps. We first show that the minimization problem can be broken up into k separate problems, each involving one row of S and one row of W. We then show that the Frobenius heuristic minimizes the contribution of each row to $\left\|\tilde{W}\right\|_F$, and is, hence, an optimal Frobenius-norm minimization strategy.

The nonzero elements of $\tilde{W}$ are the nonzero elements of the vectors $D_j^+ W_{:,j}$ for $j = 1, \ldots, m$. The i-th element of the vector $D_j^+ W_{:,j}$ is $\left(D_j^+\right)_{i,i} W_{i,j}$. If $W_{i,j} = 0$ then $\left(D_j^+\right)_{i,i} W_{i,j} = 0$, otherwise $\left(D_j^+\right)_{i,i} W_{i,j} = W_{i,j} / \left(D_j\right)_{i,i}$.

Therefore, the Frobenius norm of $\tilde{W}$ is

$$\left\|\tilde{W}\right\|_F^2 = \sum_{j=1}^{m} \sum_{i: W_{i,j} \neq 0} \left(\frac{W_{i,j}}{\left(D_j\right)_{i,i}} \right)^2 \ .$$

In this double sum, the outer summation is over columns of $\tilde{W}$, and inner summation is over the nonzeros in a particular column. Each nonzero of W appears exactly once in the summation. We can change the order of summation so that we sum over rows of W,

$$\left\|\tilde{W}\right\|_F^2 = \sum_{i=1}^{k} \sum_{\substack{j=1 \\ W_{i,j}\neq 0}}^{m} \left(\frac{W_{i,j}}{(D_j)_{i,i}} \right)^2 . \tag{5.1}$$

To minimize the Frobenius norm, we minimize this sum subject to the constraints

$$\sum_{\substack{j=1 \\ W_{i,j}\neq 0}}^{m} (D_j)_{i,i}^2 = 1 \text{ for all } i = 1, \ldots, k .$$

Since we have a separate constraint for each one of the inner sums in Equation 5.1 (for each row of W), the global minimum of the Frobenius norm is achieved when each one of the inner sums is minimized.

We now turn to the second part of the proof, showing that the heuristic does minimize each inner sum. The inner sum minimization is equivalent to finding the vector $(x_1, \ldots, x_m)$ that minimizes $\sum_{i=1}^{m} (c_i/x_i)^2$ subject to $\sum_{i=1}^{m} x_i^2 = 1$. The vector c corresponds to the nonzero elements in the ith row of W and the vector x to the corresponding elements of S. We prove by induction on m that the minimum is $\left(\sum c_i\right)^2$ and that it is achieved at

$$x_i = \frac{\sqrt{|c_i|}}{\sqrt{\sum_{j=1}^{m} |c_j|}} .$$

The inductive claim is actually slightly stronger. We prove that when the constraint is replaced by $\sum_{i=1}^{m} x_i^2 = z$ for some $z > 0$, the minimum is $z^{-1}\left(\sum c_i\right)^2$ and that it is achieved at

$$x_i = \frac{\sqrt{z\,|c_i|}}{\sqrt{\sum_{j=1}^{m} |c_j|}} .$$

For $m = 1$ the only choice for x_1 is $x_1 = \sqrt{z}$ and it is easy to verify that the claim holds. Assume that the claim holds for $m-1$. For any value of $0 < x_m < \sqrt{z}$, the minimum of the sum $\sum_{i=1}^{m-1} (c_i/x_i)^2$ subject to $\sum_{i=1}^{m-1} x_i^2 = z - x_m^2$ is, by the inductive claim, $(z - x_m^2)^{-1} \left(\sum_{i=1}^{m-1} c_i\right)^2$. The total minimization problem, then, is to minimize

$$f(x_m) = \frac{\left(\sum_{i=1}^{m-1} c_i\right)^2}{(z - x_m^2)} + \frac{c_m^2}{x_m^2}$$

subject to $0 < x_m < \sqrt{z}$. The derivative of this objective function with respect to x_m^2 is

$$\frac{\partial f}{\partial (x_m^2)} = \frac{\left(\sum_{i=1}^{m-1} c_i\right)^2}{(z - x_m^2)^2} - \frac{c_m^2}{(x_m^2)^2} .$$

It is easy to verify that the derivative vanishes at

$$x_m = \frac{\sqrt{z\,|c_m|}}{\sqrt{\sum_{j=1}^{m} |c_j|}} \ .$$

Clearly, this value of x_m satisfies $0 < x_m < \sqrt{z}$, so it solves the constrained minimization problem. Given this value of x_m, we have

$$(z - x_m^2) = z \left(\frac{\sum_{j=1}^{m-1} |c_j|}{\sum_{j=1}^{m} |c_j|} \right) \ .$$

By induction, for $i < m$, the optimal value of x_i under the constraint $\sum_{i=1}^{m-1} x_i^2 = z - x_m^2$ is achieved at

$$x_i = \frac{\sqrt{(z - x_m^2)\,|c_i|}}{\sqrt{\sum_{j=1}^{m-1} |c_j|}} = \frac{\sqrt{z\,|c_i|}}{\sqrt{\sum_{j=1}^{m} |c_j|}} \ .$$

This concludes the inductive claim and the entire proof. □

An alternative way to prove the second part of the proof, proposed by Dan Spielman, uses Lagrange multipliers.

Proof. The proof that the minimization problem can be broken into k independent subproblems is the same as in the first proof. We now show that $x_i = \sqrt{|c_i| / \sum_j |c_j|}$ is a minimizer of $\sum_{j=1}^{m} (c_j/x_j)^2$ subject to $\sum_{j=1}^{m} x_j^2 = 1$.

Let $f(x_1, \ldots, x_m, \lambda) = \sum_{j=1}^{m} \left(\frac{c_j}{x_j}\right)^2 + \lambda \left(\sum_{j=1}^{m} x_j^2 - 1\right)$. The minimizer satisfies

$$\begin{array}{rcl} 0 &=& \frac{\partial f}{\partial x_i} &=& -2 \cdot \frac{c_i^2}{x_i^3} + 2\lambda x_i \ , \\ 0 &=& \frac{\partial f}{\partial \lambda} &=& \sum_{j=1}^{m} x_j^2 - 1 \ . \end{array}$$

It follows that $c_i^2 = \lambda x_i^4$, therefore $x_i^2 = |c_i| / \sqrt{\lambda}$. Since $\sum_{j=1}^{m} x_j^2 = 1$ it follows that

$$\begin{array}{rcl} \sum_{j=1}^{m} \frac{|c_j|}{\sqrt{\lambda}} &=& 1 \\ \frac{1}{\sqrt{\lambda}} \sum_{j=1}^{m} |c_j| &=& 1 \\ \sqrt{\lambda} &=& \sum_{j=1}^{m} |c_j| \ . \end{array}$$

Since $x_i^2 = |c_i| / \sqrt{\lambda}$ it follows that

$$x_i^2 = \frac{|c_i|}{\sum_{j=1}^{m} |c_j|} \ ,$$

and thus

$$x_i = \frac{\sqrt{|c_i|}}{\sqrt{\sum_{j=1}^{m} |c_j|}} \ .$$

□

Like the rowwise heuristic, the Frobenius heuristic also produces new algebraic bounds on $\|W\|_2$. These bounds and their proofs were discovered by Dan Spielman [15]. Before we state and prove the bounds, we prove an auxiliary result.

LEMMA 5.10. *[15] Let $\tilde{W}$ be an orthonormal stretching of W derived using the Frobenius fractional-splitting heuristic. If $W_{i,j} \neq 0$, then* $\left\|\tilde{W}_{\eta(i,j),:}\right\|_2 = \left|\left(D_j^+ W_{:,j}\right)_i\right| = \sqrt{|W_{i,j}|} \cdot \sqrt{\sum_{c=1}^m |W_{i,c}|}$.

Proof. As in lemma 5.2, for each $1 \leq i \leq k$ and $1 \leq j \leq m$, $\left\|\tilde{W}_{\eta(i,j),:}\right\|_2 = \left|\left(D_j^+ W_{:,j}\right)_i\right|$. If $W_{i,j} \neq 0$, then $\left(D_j\right)_{i,i} = \sqrt{|W_{i,j}|}/\sqrt{\sum_{c=1}^m |W_{i,c}|} \neq 0$. Therefore $\left(D_j^+\right)_{i,i} = \sqrt{\sum_{c=1}^m |W_{i,c}|}/\sqrt{|W_{i,j}|}$ and thus:

$$\begin{aligned}
\left(D_j^+ W_{:,j}\right)_i &= \left(D_j^+\right)_{i,i} \cdot W_{i,j} \\
&= \frac{\sqrt{\sum_{c=1}^m |W_{i,c}|}}{\sqrt{|W_{i,j}|}} \cdot W_{i,j} \\
&= \sqrt{|W_{i,j}|} \cdot \sqrt{\sum_{c=1}^m |W_{i,c}|} \,.
\end{aligned}$$

This proves the lemma, because each row of $\tilde{W}$ has at most a single nonzero, so

$$\left\|\tilde{W}_{\eta(i,j),:}\right\|_2 = \left|(D_j^+ W_{:,j})_i\right| = \sqrt{|W_{i,j}|} \cdot \sqrt{\sum_{c=1}^m |W_{i,c}|} \,.$$

□

We now state and prove new bounds on $\|W\|_2$.

LEMMA 5.11. *[15] Let W be a k-by-m matrix. Then*

$$\|W\|_2^2 \leq \max_j \sum_{i:W_{i,j}\neq 0} |W_{i,j}| \cdot \left(\sum_{c=1}^m |W_{i,c}|\right) \,.$$

Proof. We stretch W orthonormally using the Frobenius fractional-splitting heuristic. Then $\|W\|_2^2 \leq \left\|\tilde{W}\right\|_2^2$, and the columns of $\tilde{W}$ are orthogonal, so

$$\left\|\tilde{W}\right\|_2^2 = \max_j \left\|\tilde{W}_{:,j}\right\|_2^2 \,.$$

All that remains to show is that $\left\|\tilde{W}_{:,j}\right\|_2^2 = \sum_{i:W_{i,j}\neq 0} |W_{i,j}| \cdot (\sum_{c=1}^m |W_{i,c}|)$. We have

$$\begin{aligned}
\left\|\tilde{W}_{:,j}\right\|_2^2 &= \left\|D_j^+ W_{:,j}\right\|_2^2 \\
&= \sum_{i=1}^k \left(D_j^+ W_{:,j}\right)_i^2 \\
&= \sum_{i=1}^k |W_{i,j}| \cdot \left(\sum_{c=1}^m |W_{i,c}|\right) \\
&= \sum_{i:W_{i,j}\neq 0} |W_{i,j}| \cdot \left(\sum_{c=1}^m |W_{i,c}|\right) \,.
\end{aligned}$$

The equality of the second and third lines is by lemma 5.10. ∎

LEMMA 5.12. *[15] Let W be a k-by-m matrix. Then*

$$\|W\|_2^2 \le \max_i \sum_{j:W_{i,j}\neq 0} |W_{i,j}| \cdot \left(\sum_{r=1}^{k} |W_{r,j}|\right) .$$

Proof. The previous lemma, applied to W^T, proves the claim, since $\left\|W^T\right\|_2 = \|W\|_2$. ∎

The bounds in lemmas 5.3 and 5.11 are structurally similar. Both bound $\|W\|_2^2$ using an expression of the form

$$\|W\|_2^2 \le \max_j \sum_{i:W_{i,j}\neq 0} \sum_{c=1}^{m} g\left(W_{i,j}, W_{i,c}\right) .$$

In lemma 5.3 we have $g\left(W_{i,j}, W_{i,c}\right) = W_{i,c}^2$ and in lemma 5.11 we have $g\left(W_{i,j}, W_{i,c}\right) = |W_{i,j}| \cdot |W_{i,c}|$. In both cases the maximum is over sums of functions of the same nonzero elements of W. A similar relationship exists between lemmas 5.4 and 5.12.

We note that there exist matrices W for which applying the Frobenius heuristic gives a smaller $\left\|\tilde{W}\right\|_2$ than the rowwise heuristic, and that there are matrices for which the rowwise heuristic gives a smaller $\left\|\tilde{W}\right\|_2$. In general, neither of the two is an optimal 2-norm minimization strategy.

6. Gram Bounds on the Two Norm. In this section we suggest two additional bounds on the 2-norm of W. In one particular case, these two bounds are equivalent to the bounds proved in Lemmas 5.6 and 5.7.

LEMMA 6.1. *For any matrix W,*

$$\|W\|_2^2 \le \left\|WW^T\right\|_1 = \left\|WW^T\right\|_\infty .$$

Proof. For all matrices W, $\|W\|_2^2 = \left\|WW^T\right\|_2$. For any matrix A, we have $\|A\|_2^2 \le \|A\|_1 \|A\|_\infty$. In particular, $\left\|WW^T\right\|_2^2 \le \left\|WW^T\right\|_1 \left\|WW^T\right\|_\infty$. Since WW^T is symmetric, $\left\|WW^T\right\|_1 = \left\|WW^T\right\|_\infty$. This concludes the proof. ∎

Similarly,

LEMMA 6.2. *For any matrix W,*

$$\|W\|_2^2 \le \left\|W^TW\right\|_1 = \left\|W^TW\right\|_\infty .$$

Consider the case where A and B are symmetric and diagonally-dominant matrices, and the weights of all the edges in A's and B's underlying graphs are 1. Given an embedding of the edges of A into simple paths in B, all the entries of W are either 0, 1 or -1. In this case, the dilation of an edge is exactly the length of its supporting path. It is easy to see that, in this case, the $\left\|WW^T\right\|_1$ bound is the same as the bound given in lemma 5.7. Similarly, the $\left\|W^TW\right\|_1$ bound is the same as the bound in lemma 5.6.

In more complex cases, however, the two norm bounds given in this section are not equivalent to the bounds in Lemmas 5.6 and 5.7.

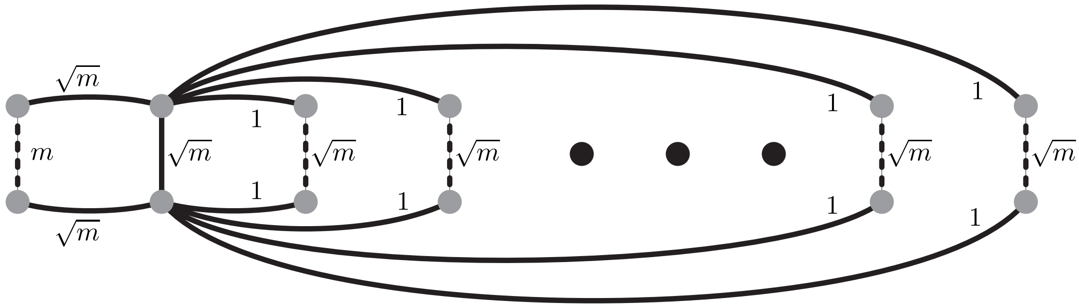

FIG. 7.1. *A weighted graph G_{UU^T} with $2m+4$ vertices ($m+2$ on the top and $m+2$ on the bottom). The dashed edges are in G_{UU^T} but not in G_{VV^T}. The edge weights given are the nonzero coefficients of the corresponding columns of U and V; for example, the edge with weight m corresponds to a column $(m,-m,0,\ldots,0)^T$ in U.*

7. An Example. The example that we present in this section shows that the new norm bound given in Lemma 5.6 is sometimes asymptotically tighter than all the other norm bounds that we are aware of. In this example, Lemma 5.6 tightly bounds the two norm, while all the other bounds are asymptotically loose. In particular, the $\|W\|_F$-norm bound, the $\|W\|_1\|W\|_\infty$-norm bound, the $\left\|WW^T\right\|_1$-norm bound and its equivalents, and the bound in Lemma 5.4 are all loose.

Consider the 1-by-2 block matrix $W=(W'|I)$, where I is the $(2m+3)$-by-$(2m+3)$ identity for some m, and where W' is the $(2m+3)$-by-$(m+1)$ matrix

$$W' = \begin{pmatrix} \sqrt{m} & 1 & \cdots & 1 \\ \sqrt{m} & & & \\ & \sqrt{m} & & \\ & & \ddots & \\ & & & \sqrt{m} \\ \sqrt{m} & & & \\ & \sqrt{m} & & \\ & & \ddots & \\ & & & \sqrt{m} \end{pmatrix}.$$

The matrix W corresponds to an embedding of the edges of the graph shown in Figure 7.1 onto paths in the same graph, but without the dashed edges. Because the graph without the dashed edges is a tree, each edge in the original graph is supported by exactly one simple path.

We can prove the following norm bounds on W. We omit the proofs.

- $\|W\|_2^2 = 4m+1$.
- $\|W\|_F^2 = 2m^2+6m+3 = \Theta(m^2)$.
- $\|W\|_1\|W\|_\infty = (\sqrt{m}+m+1)(3\sqrt{m}) = \Theta(m^{1.5})$.
- $\left\|W^TW\right\|_1 = 3m+(m+3)\sqrt{m} = \Theta(m^{1.5})$.
- $\left\|WW^T\right\|_1 = 4m+2m\sqrt{m}+1 = \Theta(m^{1.5})$.
- $\max_j \sum_{i:\,W_{i,j}\neq 0} \|W_{i,:}\|_2^2 = 4m+3 = \Theta(m)$.
- $\max_i \sum_{j:\,W_{i,j}\neq 0} \|W_{:,j}\|_2^2 = \|W'\|_F^2+1 = 2m^2+4m+1 = \Theta(m^2)$.
- $\max_j \sum_{i:\,W_{i,j}\neq 0} |W_{i,j}|\cdot\left(\sum_{c=1}^m |W_{i,c}|\right) = \sqrt{m}(\sqrt{m}+m+1)+2\sqrt{m}(\sqrt{m}+1) = \Theta(m^{1.5})$.

- $\max_i \sum_{j:W_{i,j}\neq 0} |W_{i,j}| \cdot \left(\sum_{r=1}^{k} |W_{r,j}|\right) = \sqrt{m} \cdot 3\sqrt{m} + m(1 + 2\sqrt{m}) + 1 = \Theta(m^{1.5})$.

For large m, none of these bounds on the 2-norm are tight, except for one, $4m + 3$, which is not only asymptotically tight, but is off by only a small additive constant.

8. Conclusions. We have shown that applying the splitting lemma to the analysis of support-graph preconditioners can be viewed as a mechanism to bound the norm of a matrix W. The mechanism works by orthonormally stretching W into a larger matrix $\tilde{W}$ whose 2-norm bounds that of W but is easier to compute.

In doing so, we have unified the "old-style" support theory, in which the analysis of a preconditioner usually starts by splitting, and the "new-style" support theory, which relies on the symmetric-product-support lemma, usually without splitting.

We also presented six new bounds on the 2-norm of the matrix, given in Lemmas 5.3, 5.4, 5.11, 5.12, 6.1, and 6.2. One of the four was already given by Spielman and Teng, but not in the form of a norm bound. Four of the new bounds have useful combinatorial interpretations. Special cases of some of our new bounds were previously used to bound the smallest nonzero eigenvalue of Laplacian matrices [7, 10, 11, 12, 13].

Viewing splitting as a way of bounding $\|W\|_2$ using $\left\|\tilde{W}\right\|_2$ leads to systematic splitting strategies that aim to minimize some other norm of $\tilde{W}$. We propose two such strategies in this paper; one is a heuristic which preserves in $\tilde{W}$ the 2-norm of rows of W, and another which minimizes the Frobenius norm of $\tilde{W}$. Both are analytically and computationally simple.

We have also shown that one of the new bounds can be asymptotically tighter than all the other norm bounds that we are aware of. The problem of ranking the bounds by tightness, or showing that they cannot be ranked, remains open.

Acknowledgements. Thanks to the three anonymous referees for numerous corrections, suggestions and comments. Lemmas 5.11 and 5.12 were discovered by one of the referees, who turned out to be Dan Spielman; he also suggested the alternative proof of Lemma 5.9.

REFERENCES

[1] M. Bern, J. R. Gilbert, B. Hendrickson, N. Nguyen, and S. Toledo, *Support-graph preconditioners*. Submitted to SIAM Journal on Matrix Analysis and Applications, 29 pages, Jan. 2001.

[2] E. Boman and B. Hendrickson, *On spanning tree preconditioners*. Manuscript, Sandia National Laboratories, 2001.

[3] E. G. Boman, D. Chen, B. Hendrickson, and S. Toledo, *Maximum weight basis preconditioners*. To appear in the *Journal on Numerical Linear Algebra*, 29 pages, June 2001.

[4] E. G. Boman, D. Chen, O. Parekh, and S. Toledo, *On the factor-width and symmetric H-matrices*. Submitted to Linear Algebra and its Applications, 2004.

[5] E. G. Boman and B. Hendrickson, *Support theory for preconditioning*, SIAM J. of Matrix Analysis, 25 (2004), pp. 694–717.

[6] D. Chen and S. Toledo, *Combinatorial charaterization of the null spaces of symmetric H-matrices*. To Appear in Linear Algebra and its Applications, 2004.

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

[8] G. H. Golub and C. F. V. Loan, *Matrix Computations*, Johns Hopkins University Press, 3rd ed., 1996.

[9] K. D. Gremban, *Combinatorial Preconditioners for Sparse, Symmetric, Diagonally Dominant Linear Systems*, PhD thesis, School of Computer Science, Carnegie Mellon University, Oct. 1996. Technical Report CMU-CS-96-123.

[10] S. Guattery, T. Leighton, and G. L. Miller, *The path resistance method for bounding the smallest nontrivial eigenvalue of a Laplacian*, Combinatorics, Probability, and Computing, 8 (1999), pp. 441–460.

[11] S. Guattery and G. L. Miller, *Graph embeddings and Laplacian eigenvalues*, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 703–723.

[12] N. Kahale, *A semidefinite bound for mixing rates of Markov chains*, Random Structures and Algorithms, 11 (1997), pp. 299–313.

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

[14] D. Spielman and S.-H. Teng, *Solving sparse, symmetric, diagonally-dominant linear systems in time* $o(m^{1.31})$, in Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science, Cambridge, MA, Oct. 2003, pp. 416–427.

[15] D. A. Spielman, *Private communication.* Mar. 2005.

[16] D. A. Spielman and S.-H. Teng, *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, in Proceedings of the 36th annual ACM symposium on Theory of computing, 2004, pp. 81–90.

[17] P. M. Vaidya, *Solving linear equations with symmetric diagonally dominant matrices by constructing good preconditioners.* Unpublished manuscript. A talk based on the manuscript was presented at the IMA Workshop on Graph Theory and Sparse Matrix Computation, October 1991, Minneapolis.