Preconditioners for Singular Black Box Matrices

William J. Turner
Department of Mathematics & Computer Science
Wabash College
Crawfordsville, IN 47933 USA
turnerw@wabash.edu

ABSTRACT

This paper develops preconditioners for singular black box matrix problems. We introduce networks of arbitrary radix switches for matrices of any square dimension, and we show random full Toeplitz matrices are adequate switches for these networks. We also show a random full Toeplitz matrix to satisfy all requirements of the Kaltofen-Saunders black box matrix rank algorithm without requiring a diagonal multiplier.

Categories and Subject Descriptors

I.1.2 [**Symbolic and Algebraic Manipulation**]: Algorithms—*algebraic algorithms, analysis of algorithms*

General Terms

Algorithms, Performance, Reliability, Theory.

Keywords

Black box linear algebra, rank preconditioner, singular linear system, switching network, Toeplitz matrix, Wiedemann method.

1. INTRODUCTION

Kaltofen and Saunders [8] provide algorithms based on the Wiedemann method to compute the rank of a black box matrix and to uniformly sample the solution space of a singular linear system. Both of these algorithms rely on preconditioning a matrix to place it into a generic rank profile as defined by Kaltofen and Lobo [7]. The Kaltofen-Saunders rank algorithm also requires multiplying the resulting matrix with a diagonal matrix.

Chen *et al.* [2, Thm.3.1] show how to reduce the PreCondGen preconditioning problem of placing a matrix into a generic rank profile to the PreCondInd problem of localizing the linear independence of the system by preconditioning a matrix A to make the leading $r = \text{rank}(A)$ columns linearly independent. Solutions to the PreCondInd problem include a parameterized matrix based on rearrangeable permutation networks such as Beneš networks [12, §V] and unit upper and lower triangular Toeplitz matrices [8, Thm.2]. Other preconditioners, such as those described in Eberly and Kaltofen [4], also require the transpose of A. Chen *et al.* [2, §6] describe how to construct a PreCondInd preconditioner by generalizing a Beneš network to butterfly networks for arbitrary dimension n. Section 2.1 shows how to further generalize these switching networks to switches of arbitrary radix ρ, which decreases both the depth of the network and the number of switches required.

These arbitrary radix switching networks cannot directly precondition the matrix. Instead of switching the rows or columns of a matrix, preconditioning matrices mix them with a generic exchange matrix to achieve the desired linear independence. Section 2.2 shows preconditioners using radix-ρ switching networks and random full Toeplitz matrices as generic exchange matrices solve the PreCondInd problem over a large field with a high probability.

The generalized switching networks of arbitrary radix may offer only a limited increase in performance over previously known butterfly networks, but we believe they may provide other advantages. In particular, they provide us a way to search for preconditioning matrices to solve the PreCondInd problem. For example, we can generalize the result in Section 2.2 to Toeplitz matrices of arbitrary size. In particular, these results allow us to use a single random full Toeplitz matrix preconditioner to solve the PreCondInd problem.

Turner [11; 10, Ch.3] uses determinantal divisors to relax slightly the generic rank profile condition of the Kaltofen-Saunders rank algorithm before applying a diagonal multiplier. Instead of requiring a nonzero leading principal minor of size r, any nonzero principal $r \times r$ minor suffices. This relaxation, along with Section 2.2, means one random full Toeplitz matrix can precondition the matrix so the Kaltofen-Saunders rank algorithm can then apply the random diagonal multiplier. Section 3 shows the diagonal multiplier is superfluous and a random full Toeplitz matrix alone fulfills the requirements of the Kaltofen-Saunders rank algorithm. This preconditioner requires fewer random variables and has an improved probability of success over butterfly network preconditioners, but at the cost of a slower matrix-vector product.

2. GENERALIZED NETWORKS

Preconditioners based on Beneš networks [1] solve the PreCondInd preconditioning problem. In this section, we

ISSAC'05, July 24–27, 2005, Beijing, China.

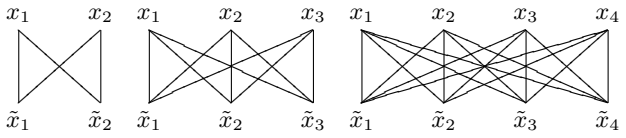

Figure 1: We may represent radix-2 (left), radix-3 (middle), and radix-4 (right) switches by the complete bipartite graphs $K_{2,2}$, $K_{3,3}$, and $K_{4,4}$, respectively. We also call radix-2 a butterfly switch.

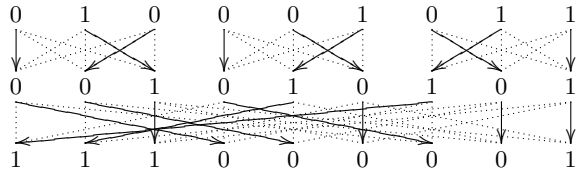

Figure 2: Radix-3 switching network. The solid arrows show how to set the switches to achieve the desired output. In this case, the network switches the four inputs designated with 1s to four contiguous outputs.

improve on previous work by Parker [9] and Chen *et al.* [2] by generalizing their butterfly networks to networks of arbitrary radix switches. We then show a Toeplitz matrix satisfies the requirements of an arbitrary radix exchange matrix.

2.1 Arbitrary Radix Switching Networks

Consider an $n \times n$ matrix over a field $\mathbb{F}$. We want to make the leading r rows or columns of the matrix linearly independent. A switching network can exchange rows or columns until the first r rows or columns are linearly independent. However, unlike Chen *et al.* [2], we want to exchange more than two rows or columns at a time to decrease the depth of the network and the number of switches required.

A radix-ρ switch generalizes a butterfly switch to allow any permutation of ρ inputs. As Figure 1 shows, we can depict a radix-ρ switch as a complete bipartite graph of 2ρ nodes. We consider the nodes across the top of the graphs as the inputs to the switch and the nodes across the bottom as the outputs. This complete bipartite graph allows the network to route each input directly to any desired output.

We want to use a radix-ρ switching network to switch any r inputs of an arbitrary number n inputs to the beginning of the network. However, as in Chen *et al.* [2, §6], we must first consider the case of switching any r inputs into any contiguous block when n is a power of ρ: $n = \rho^\ell$.

An ℓ-dimensional radix-ρ switching network is a recursive network that uses $\rho^{\ell-1}$ radix-ρ switches to merge the outputs of ρ radix-ρ subnetworks, each of dimension $\ell - 1$, such that the network merges the ith outputs of each of the subnetworks, $1 \le i \le \rho^{\ell-1}$. Figure 2 illustrates a two-dimensional radix-3 switching network with nine nodes at each level.

We begin by showing how a radix-ρ switching network can switch any r inputs into any contiguous block of outputs when n is a power of ρ. We will use this result when we consider arbitrary n.

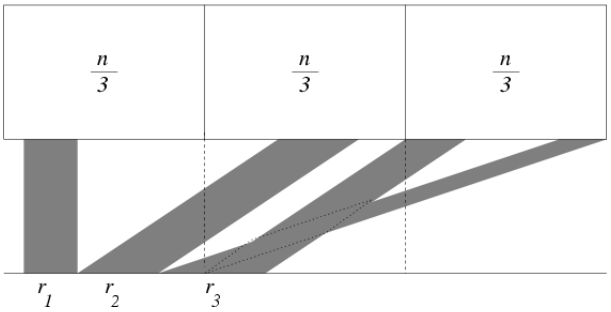

Figure 3: An example of a radix-3 network switching desired indices into position. Notice we must wrap indices around the outside of the right subnetwork to allow them to switch into the correct position.

LEMMA 2.1. *Let $n = \rho^\ell$ where $\rho \ge 2$. The ℓ-dimensional radix-ρ switching network can switch any r indices $1 \le i_1 < \cdots < i_r \le n$ into any desired contiguous block of indices. For our purposes, wrapping the block around the outside preserves contiguity. (For example, we consider the 1s in the last row of Figure 2 contiguous.)*

PROOF. Let us prove the lemma by induction. The proof is trivial for $n = 1$ because no switches are required.

Suppose the lemma holds for n/ρ. Let us divide the n nodes into ρ equal subnetworks. We can construct radix-ρ switching networks of dimension $\ell - 1$ for each of these collections of n/ρ nodes. Let $r_0 = 0$, $r_\rho = r$, and r_k be such that i_{r_k} is the last of the chosen indices before the end of the kth radix-ρ network: $i_{r_k} \le kn/\rho < i_{r_k+1}$ for $k = 1, \ldots, \rho - 1$. Because the lemma holds for n/ρ, we know the kth subnetwork can arrange the indices $i_{r_{k-1}+1}, \ldots, i_{r_k}$ into any desired contiguous blocks for $k = 1, \ldots, \rho$.

The contiguous block desired from the network is either contained in the interior of the network in indices $j, \ldots, j + r - 1$ or it wraps around the outside of the network and we can denote it by indices $1, \ldots, j-1$ and $n-r+j, \ldots, n$. We can convert this second situation into the first by instead thinking of switching the $n - r$ indices not originally chosen into the contiguous block $1 \le j, \ldots, j+n-r-1 \le n$. Thus, we need only consider the first situation.

We can use the first subnetwork to place inputs $i_1, \ldots, i_{r_1}$ so they switch into outputs $j, \ldots, j + r_1 - 1$. We can then use the second subnetwork to place inputs $i_{r_1+1}, \ldots, i_{r_2}$ so they switch into outputs $j + r_1, \ldots, j + r_2 - 1$. We can continue in this manner until we use the ρth subnetwork to place inputs $i_{r_{\rho-1}+1}, \ldots, i_r$ so they switch into outputs $j + r_{\rho-1}, \ldots, j + r - 1$. In each of these subnetworks, the indices may wrap around the outside of the subnetwork to allow them to switch into the correct position as Figure 3 demonstrates.

Using the induction hypothesis, each of these ρ subnetworks has dimension $\ell - 1$ and $n(\log_\rho(n) - 1)/\rho^2$ switches. With the addition of a level of n/ρ switches to combine these subnetworks, the final radix-ρ switching network has dimension ℓ requires a total of $s = n \log_\rho(n)/\rho$ switches. □

Lemma 2.1 says we can switch any r rows or columns of an $n \times n$ matrix into any contiguous block when $n = \rho^\ell$. Recall our original goal of not restricting n to a power of ρ. To do so, we must describe a generalized radix-ρ switching network for an arbitrary n.

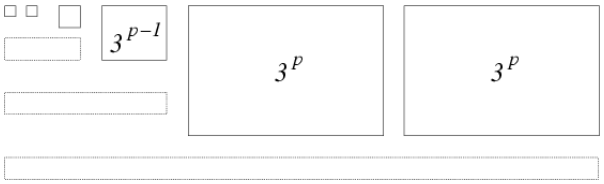

Figure 4: Generalized radix-3 switching network. The dotted boxes represent the extra layer of switches.

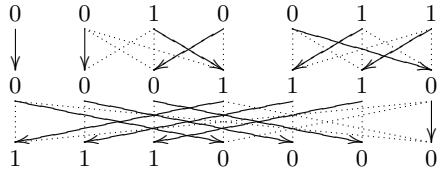

Figure 5: Generalized radix-3 switching network. The solid arrows show how to set the switches to achieve the desired output. In this case, the network switches the three inputs designated with 1s to the first three outputs.

When n is not a power of ρ, let us decompose n as a sum of powers of ρ:

$$n = \sum_{i=1}^{p} n_i \text{ where } \begin{cases} n_i = c_i\ \rho^{\ell_i} \\ c_i \in \{1, \ldots, \rho - 1\} \\ \ell_1 < \ell_2 < \cdots < \ell_p \end{cases} \tag{2.1}$$

The radix-ρ switching network is the case when $p = c_1 = 1$.

We construct radix-ρ switching networks for each of the ρ^{ℓ_i} blocks. We then build a generalized radix-ρ switching network by connecting these networks with switches recursively such that for $k = 2, \ldots, p$, the network merges the $\sum_{i=1}^{k-1} n_i$ of a generalized subnetwork with the far right nodes of each of the ρ^{ℓ_k} blocks with switches of radix $c_k + 1 \le \rho$. If $c_k > 1$, the network merges the remaining $n_k - \sum_{i=1}^{k-1} n_i$ nodes of each of the c_k blocks of size ρ^{ℓ_k} with switches of radix $c_k < \rho$. These two sets of extra switches do not interact, so they constitute only one extra layer of switches after the last level of radix-ρ switching networks.

We may think of this level of connecting switches as one set of switches of radix $c_k+1 \le \rho$ that switch the corresponding outputs of the c_k radix-ρ networks with the combined outputs of the generalized radix-ρ network for $\sum_{i=1}^{k-1} n_i$ and $\rho^k - \sum_{i=1}^{k-1} n_i$ "phantom" outputs to the left of the generalized subnetwork. In this way, we require only one additional level of switches to connect all of the subnetworks. Figure 4 shows the general organization of a generalized radix-3 network, and Figure 5 shows a specific generalized radix-3 network acting on seven inputs.

Before we consider an arbitrary network of size n, let us first expand Lemma 2.1 to include the case when $p = 1$ but $c_1 \neq 1$; in other words, $n = c\,\rho^{\ell}$ where $1 < c < \rho$. This case does not arise when $\rho = 2$.

LEMMA 2.2. *Let $n = c\,\rho^{\ell}$ where $\rho \ge 2$ and $1 \le c \le \rho - 1$. The generalized radix-ρ switching network described above can switch any r indices $1 \le i_1 < \cdots < i_r \le n$ into any desired contiguous block of indices; for our purposes, wrapping the block around the outside preserves contiguity. Furthermore, the network has a depth of $\lceil \log_\rho(n) \rceil$ and a total of no more than $\rho^{\lceil \log_\rho(n) \rceil} \lceil \log_\rho(n) \rceil / \rho$ switches of radix at most ρ. The network attains this bound only when $c = 1$ or $\ell = 0$.*

PROOF. The proof follows directly from Lemma 2.1 when $c = 1$. Otherwise, let $r_0 = 0$, $r_c = r$, and r_k be such that i_{r_k} is the last of the chosen indices before the end of the kth radix-ρ network: $i_{r_k} \le kn/\rho < i_{r_k+1}$ for $k = 1, \ldots, c-1$. As in the proof of Lemma 2.1, we need only consider the case when the desired contiguous block is contained within the interior of the network in indices $j, \ldots, j+r-1$, and we can use the kth radix-ρ subnetwork to place inputs $i_{r_{k-1}+1}, \ldots, i_{r_k}$ so they switch into outputs $j + r_{k-1}, \ldots, j + r_k - 1$. In each of these radix-ρ networks, the indices may wrap around the outside of the network to allow them to switch into the correct position as was needed in the proof of Lemma 2.1.

The number of switches needed when $c \neq 1$ is the number for the c radix-ρ networks plus ρ^{ℓ} switches to combine the networks:

$$s = \frac{c\,\rho^{\ell}}{\rho}\ell + \rho^{\ell} \le \frac{\rho\,\rho^{\ell}}{\rho}\ell + \rho^{\ell} = \frac{\rho^{\lceil \log_\rho(n) \rceil}}{\rho} \lceil \log_\rho(n) \rceil.$$

The network never attains this bound when $c \neq 1$ and $\ell \neq 0$. Furthermore, the network has depth $\ell + 1 = \lceil \log_\rho(n) \rceil$ for $c \neq 1$. □

This means the generalized radix-ρ switching network uses $\rho^{\lceil \log_\rho(n) \rceil} \lceil \log_\rho(n) \rceil / \rho$ switches when $n = \rho^{\ell}$ or $n = c$; otherwise the network uses fewer switches. Using this result, we can show the generalized radix-ρ switching network can switch any r inputs into the first block of r inputs for any n.

THEOREM 2.1. *Suppose $\rho \ge 2$. The generalized radix-ρ switching network described above can switch any r indices $1 \le i_1 < \cdots < i_r \le n$ into the contiguous block $1, 2, \ldots, r$. Furthermore, it has a depth of $\lceil \log_\rho(n) \rceil$ and a total of no more than $\rho^{\lceil \log_\rho(n) \rceil} \lceil \log_\rho(n) \rceil / \rho$ switches of radix at most ρ. The network attains this bound only when $n = \rho^{\lceil \log_\rho(n) \rceil}$ or $n < \rho$.*

PROOF. The proof follows from Lemma 2.2 if $n = c\rho^{\ell}$ where $1 \le c \le \rho - 1$. Otherwise, let us decompose n into powers of ρ (2.1). We know $p > 1$ and $n_p > \sum_{i=1}^{p-1} n_i$. We prove the first part of the theorem by induction.

Suppose the theorem is true for $\sum_{i=1}^{p-1} n_i$. Let r_1 be such that i_{r_1} is the last index in the first subnetwork:

$$i_{r_1} \le \sum_{i=1}^{r-1} n_i < i_{r_1+1}.$$

We can switch the indices $i_1, \ldots, i_{r_1}$ into the contiguous block $1, \ldots, r_1$ using a generalized butterfly network. In addition, let $r_0 = 0$, $r_{c_p+1} = r$, and r_k be such that i_{r_k} is the last of the chosen indices before the end of the $(k-1)$st radix-ρ network:

$$i_{r_k} \le (k-1)\,\rho^{\ell} + \sum_{i=1}^{r-1} n_i < i_{r_k+1}, \qquad k = 2, \ldots, c_p + 1.$$

By Lemma 2.1, the kth radix-ρ subnetwork can arrange inputs $i_{r_k+1}, \ldots, i_{r_{k+1}}$ into any desired contiguous block of

outputs, $k = 1, \ldots, c_p$. In particular, we use the kth subnetwork to place these inputs so they switch into outputs $r_k + 1, \ldots, r_{k+1}$. In each of these radix-ρ networks, the indices may wrap around the outside of the network to allow them to switch into the correct position.

The total number of butterfly switches is the number of switches for each of the subnetworks plus the number of switches to combine them. When $c_p = 1$, the network requires $\sum_{i=1}^{p-1} n_i < \rho^{\ell_p}$ switches for the combining. When $c_p \neq 1$, it requires ρ^{ℓ} switches. We may also count the switches as the sum of the number of switches for each of the n_i blocks plus the number of switches to connect these blocks:

$$s = \sum_{i=1}^{p} \frac{c_i\, \rho^{\ell_i} \ell_i}{\rho} + \sum_{i=1}^{p-1} \left(\sum_{j=1}^{i} c_j\, \rho^{\ell_j} \right) + \sum_{\substack{i=2 \\ c_i > 1}}^{p} \left(\rho^{\ell_i} - \sum_{j=1}^{i-1} c_j\, \rho^{\ell_j} \right).$$

From the decomposition of n into powers of ρ (2.1), we know $\ell_i \geq i - 1$ for $1 \leq i \leq p$, so the number of switches required when $p > 1$ is

$$s \leq \sum_{i=1}^{p} \frac{n_i}{\rho} \ell_i + \sum_{i=2}^{p} \rho^{\ell_i} < \rho^{\ell_p} \ell_p + \rho^{\ell_p} = \frac{\rho^{\lceil \log_\rho(n) \rceil} \lceil \log_\rho(n) \rceil}{\rho}.$$

The network never attains this bound if $p > 1$. However, Lemma 2.2 tells us the network does attain the bound when $p = 1$ and $c = 1$ or $\ell = 0$. These correspond to the cases $n = \rho^{\lceil \log_\rho(n) \rceil}$ and $n < \rho$.

The depth of the network when $p > 1$ is the depth of the largest subnetworks of a power of three plus one level to combine the rest of the network with these subnetworks. Thus, the complete network has a depth of $\ell_p + 1 = \lceil \log_\rho(n) \rceil$. □

2.2 Generic Exchange Matrices

Following Wiedemann [12], we can use switching networks to construct left and right PreCondInd preconditioners for an $n \times n$ matrix by embedding $\rho \times \rho$ exchange matrices into an $n \times n$ identity matrix. Each switch in the network implements a directed acyclic arithmetic circuit, which we can represent by a linear equation. For example,

$$\begin{bmatrix} \tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{1,1,k} & a_{1,2,k} & a_{1,3,k} \\ a_{2,1,k} & a_{2,2,k} & a_{2,3,k} \\ a_{3,1,k} & a_{3,2,k} & a_{3,3,k} \end{bmatrix}$$

represents the circuit for the kth radix-3 switch of a network shown in Figure 1. We embed the $\rho \times \rho$ matrix in an $n \times n$ identity matrix in the fashion of an elementary matrix that can exchange the ρ columns by replacing the principal—not necessarily leading principal—submatrix of the identity matrix composed of the desired rows and columns with the exchange matrix. We then consider the symbolic preconditioner $\mathcal{L} = \prod_{k=1}^{s} \mathcal{S}_k(\alpha_{1,1,k}, \ldots, \alpha_{\rho,\rho,k})$ where $\mathcal{S}_k$ symbolically implements the kth embedded switch in the network of s switches and $\alpha_{i,j,k}$ is a symbol. We denote symbols and symbolic preconditioners with Greek and calligraphic letters, and we denote their evaluated counterparts with corresponding roman letters.

Let A be a fixed $n \times n$ matrix of rank r. The first r columns of the symbolically preconditioned matrix $A\mathcal{L}$ are linearly independent over the field $\mathbb{F}(\alpha_{1,1,k}, \ldots, \alpha_{\rho,\rho,s})$ because we may evaluate the symbols in such a manner that the network switches r linearly independent columns to the left. We need to minimize the cost of the matrix-vector product for the preconditioning matrix L where the symbols have been evaluated at fixed random values as a black box matrix. We can use an arbitrary random $\rho \times \rho$ matrix where each entry contains a different symbol, but it requires $O(\rho^2)$ field multiplications. We now describe a Toeplitz exchange matrix that will accomplish the task while requiring only $O(\rho \log(\rho) \operatorname{loglog}(\rho))$ field operations. As in Chen *et al.* [2, §6.2], we will prove the result by induction on the levels of the generalized butterfly network.

Consider a field $\mathbb{F}$ and the symbolic Toeplitz matrix

$$\mathcal{T} = \begin{bmatrix} \alpha_\rho & \alpha_{\rho+1} & \ldots & \alpha_{2\rho-1} \\ \alpha_{\rho-1} & \alpha_\rho & \ldots & \alpha_{2\rho-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \ldots & \alpha_\rho \end{bmatrix}, \tag{2.2}$$

where $\alpha_1, \ldots, \alpha_{2\rho-1}$ are distinct indeterminates over $\mathbb{F}$. The (i,j)th entry of $\mathcal{T}$ is $\mathcal{T}_{[i,j]} = \alpha_{\rho+j-i}$. Let $\mathcal{T}_{[i_1,\ldots,i_m;j_1,\ldots,j_m]}$ denote the submatrix of $\mathcal{T}$ formed by the intersection of rows $i_1, \ldots, i_m$ and columns $j_1, \ldots, j_m$, so $\det(\mathcal{T}_{[i_1,\ldots,i_m;j_1,\ldots,j_m]})$ denotes the corresponding minor.

When employed as a generic exchange matrix embedded in an $n \times n$ identity matrix to obtain the symbolic preconditioner $\widehat{\mathcal{T}}$, this symbolic Toeplitz matrix will interact with only ρ of the n columns of A. Let us denote the submatrix of A formed by these ρ columns and all n rows as A'. Then, the symbolic preconditioning $A\widehat{\mathcal{T}}$ only changes these columns, and the matrix equation $A'\mathcal{T}$ captures the entire preconditioning.

The submatrix A' affected by $\mathcal{T}$ includes a subset of r' of the r linearly independent columns of A where $0 \leq r' \leq r$. We must show these r' columns remain linearly independent as they pass through this level of the network. Then, if we assign unique indeterminates to each symbolic Toeplitz exchange matrix, then the network must preserve the linear independence of the full r columns of A in each level. Thus, the first r columns of $A(\prod_{k=1}^{s} \widehat{\mathcal{T}}_k(\alpha_{1,k}, \ldots, \alpha_{2\rho-1,k}))$ are linearly independent over $\mathbb{F}(\alpha_{1,k}, \ldots, \alpha_{2\rho-1,s})$.

To show we can move the linear independence of the desired r' columns through a single exchange matrix we examine the leading term of the resulting preconditioned matrix $A'\mathcal{T}$. Using the lexicographic monomial ordering with $\alpha_1 \prec \alpha_2 \prec \cdots \prec \alpha_{2n-1}$ [3, Def. 2.2.3], the leading term of the minor the intersection of rows $i_1, \ldots, i_m$ and columns $j_1, \ldots, j_m$ of $\mathcal{T}$ is

$$\operatorname{lt}\left(\det\left(\mathcal{T}_{[i_1,\ldots,i_m;j_1,\ldots,j_m]}\right)\right) = (-1)^{\lfloor m/2 \rfloor} \prod_{k=1}^{m} \alpha_{n+j_{m+1-k}-i_k}. \tag{2.3}$$

The kth largest index of α in this leading term, which corresponds to the kth most significant α in the leading term, is $n + j_{m+1-k} - i_k$. Given the row indices $i_1, \ldots, i_m$, the leading term of the minor uniquely determines the column indices $j_1, \ldots, j_m$, and vice versa.

The Binet-Cauchy formula tells us $\det((A'\mathcal{T})_{[\mathscr{I},\mathscr{J}]})$ is a sum of products of minors of A' and $\mathcal{T}$:

$$\det((A'\mathcal{T})_{[\mathscr{I},\mathscr{J}]}) = \sum_{\substack{\mathscr{K} = \{k_1,\ldots,k_m\} \\ 1 \leq k_1 < \cdots < k_m \leq n}} \det(A'_{[\mathscr{I},\mathscr{K}]}) \det(\mathcal{T}_{[\mathscr{K},\mathscr{J}]}).$$

(See, for example, Gantmacher [5, §2.5].) Because the submatrices $\mathcal{T}_{[\mathscr{K},\mathscr{J}]}$ are all homogeneous of degree one, their determinants are all homogeneous of degree m. The minors $\det(\mathcal{T}_{[\mathscr{K},\mathscr{J}]})$ are linearly independent because each has a unique leading term under the lexicographic monomial ordering with $\alpha_1 \prec \alpha_2 \prec \cdots \prec \alpha_{2\rho-1}$. There exists at least one index set $\mathscr{K}$ such that $\det(A'_{[\mathscr{I},\mathscr{K}]}) \neq 0$ because $\mathscr{I}$ is a set of indices of linearly independent columns of A. The minor $\det((A'\mathcal{T})_{[\mathscr{I},\mathscr{J}]})$ is a nonzero sum of linearly independent polynomials, so it cannot be zero. This means every minor of the symbolically preconditioned matrix $A'\mathcal{T}$ involving only linearly independent columns of A' must be nonzero. If the matrix A' has rank at least r', then every set of $m \leq r'$ columns of the symbolically preconditioned matrix $A'\mathcal{T}$ must be linearly independent.

Thus we can move the linear independence around within one switch. By using different indeterminates α_i for each symbolic Toeplitz matrix, we ensure linearly independent columns remain linearly independent at each level of the generalized radix-ρ switching network.

For the kth switch in the generalized radix-ρ switching network, we can embed the $\rho \times \rho$ symbolic Toeplitz matrix $\mathcal{T}_k$ into an $n \times n$ identity matrix as was done by Chen *et al.* [2, §6] to obtain the matrix $\hat{\mathcal{T}}_k$. By randomly choosing a value a_i for each indeterminate α_i, we convert the symbolic Toeplitz matrix $\mathcal{T}_k$ into the Toeplitz matrix T_k and the symbolic exchange matrix $\hat{\mathcal{T}}_k$ into the randomly chosen exchange matrix $\hat{T}_k$. We prove the following counterpart to Theorem 6.3 in Chen *et al.* [2].

THEOREM 2.2. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ have r linearly independent columns, let s be the number of switches in the generalized radix-ρ switching network from Theorem 2.1, and let S be a finite subset of $\mathbb{F}$. Let N be the number of random numbers required in this network. If $a_1, \ldots, a_N$ are randomly chosen uniformly and independently from S, then the first r columns of $A(\prod_{k=1}^{s} \widehat{T}_k)$ are linearly independent with probability at least*

$$1 - \frac{r\lceil \log_\rho(n)\rceil}{|S|} \geq 1 - \frac{n\lceil \log_\rho(n)\rceil}{|S|}.$$

PROOF. The network requires $N < 2\ \rho^{\lceil \log_\rho(n)\rceil}\lceil \log_\rho(n)\rceil$ random numbers because the radix-ρ switching network uses $s \leq \rho^{\lceil \log_\rho(n)\rceil}\lceil \log_\rho(n)\rceil/\rho$ switches of radix at most ρ by Theorem 2.1 and each switch uses no more than $2\rho - 1$ random numbers.

Let $\tilde{A} = A(\prod_{k=1}^{s} \widehat{\mathcal{T}}_k)$. Because each of the $\lceil \log_\rho(n)\rceil$ levels in the generalized radix-ρ switching network increases the degree of the polynomials by one, the columns of $\tilde{A}$ are vectors of polynomials in $\alpha_1, \ldots, \alpha_N$ of degree $\lceil \log_\rho(n)\rceil$, and the determinant of an $r \times r$ submatrix of this preconditioned matrix is a polynomial of degree $r\lceil \log_\rho(n)\rceil$.

Because A has r linearly independent columns, we can designate the generalized radix-ρ switching network of Theorem 2.1 to switch these columns to the first r columns of the preconditioned matrix $\tilde{A}$. There is an $r \times r$ submatrix of the first r columns of $\tilde{A}$ whose determinant is not identically zero. Because this determinant is a polynomial of degree $r\lceil \log_\rho(n)\rceil$, the Schwartz-Zippel lemma tells us $(a_1, \ldots, a_N)$ is a root of it with probability at most $r\lceil \log_\rho(n)\rceil/|S|$. Therefore, the corresponding $r \times r$ submatrix of $A(\prod_{k=1}^{s} \widehat{T}_k)$ whose determinant is not zero with probability no greater than $1 - r\lceil \log_\rho(n)\rceil/|S|$. □

The $\log_\rho(n)$ term in Theorem 2.2 means a preconditioner based on a generalized radix-ρ switching network offers some improvement as ρ increases. However, the number of random values required—and thus the size of the random set S from which they are chosen—increases.

3. RANK PRECONDITIONER

In Turner [11; 10, Ch.3], we consider a preconditioned matrix $A\mathcal{D}$, where $\mathcal{D}$ is a symbolic diagonal matrix. We show every invariant factor of the characteristic matrix $\lambda I - A\mathcal{D}$ of size m such that $\operatorname{rank}(A) < m \leq n$ is divisible by λ, which gives an upper bound for the degree of λ in the minimal polynomial $f^{A\mathcal{D}}$. To find a lower bound, we rely on the proof of Lemma 4.1 in Chen *et al.* [2] to show the characteristic polynomial of $A\mathcal{D}$ has no repeated factors except λ. The invariant factors other than the largest, which is the minimal polynomial $f^{A\mathcal{D}}$, are all powers of λ. We then use the existence of a nonzero principal minor to prove a lower bound for the λ-degree equal to the upper bound.

We now use a similar technique of examining the invariant factors of the characteristic matrix to show a random full Toeplitz matrix preconditioner satisfies the requirements of the Kaltofen-Saunders rank algorithm without requiring a diagonal multiplier. We begin by again using the invariant factors to prove an upper bound on the λ-degree of the minimal polynomial of the symbolically preconditioned matrix.

LEMMA 3.1. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ have rank r, let $s \in \mathbb{Z}_{>0}$, and let $\mathcal{B}_1, \mathcal{B}_2 \in \mathbb{F}[\alpha_1, \alpha_2, \ldots, \alpha_s]^{n\times n}$, where $\alpha_1, \ldots, \alpha_s$ are distinct indeterminates over $\mathbb{F}$. Let λ be an indeterminate distinct from $\alpha_1, \ldots, \alpha_s$. Then, for all k with $r + 1 \leq k \leq n$, λ divides the kth invariant factor of $\lambda I - \mathcal{B}_1 A \mathcal{B}_2$. Furthermore, the degree of λ in the minimal polynomial of $\mathcal{B}_1 A \mathcal{B}_2$ is bounded by $\deg_\lambda(f^{\mathcal{B}_1 A \mathcal{B}_2}) \leq r + 1$.*

PROOF. The first part follows from Turner [10, Lem.3.5] and the fact $\operatorname{rank}(\mathcal{B}_1 A \mathcal{B}_2) \leq \operatorname{rank}(A)$ by Sylvester's inequality [5, p.66].

The characteristic polynomial of $\mathcal{B}_1 A \mathcal{B}_2$ has λ-degree n and is the product of the invariant factors of the characteristic matrix, $s_k(\lambda I - \mathcal{B}_1 A \mathcal{B}_2)$ where $1 \leq k \leq n$. Because the minimal polynomial of $\mathcal{B}_1 A \mathcal{B}_2$ is the largest of these invariant factors, we know it has λ-degree

$$\deg_\lambda(f^{\mathcal{B}_1 A \mathcal{B}_2}) = n - \sum_{k=1}^{n-1} \deg_\lambda(s_k(\lambda I - \mathcal{B}_1 A \mathcal{B}_2)).$$

However, because λ divides at least $n-r-1$ of these invariant factors, $\deg_\lambda(f^{\mathcal{B}_1 A \mathcal{B}_2}) \leq n - (n - r - 1) = r + 1$. □

We must show the lower bound $\deg_\lambda(f^{\tilde{A}}) \geq r+1$ without relying on a diagonal multiplier and Chen *et al.* [2]. In order to use the leading term of the characteristic polynomial to show $\det(\lambda I - \tilde{A})$ has no repeated factors except λ, we shall again turn to the uniqueness of the leading terms of the minors of the symbolic Toeplitz matrix $\mathcal{T}$. This time we shall use the elementary symmetric functions of the eigenvalues of $A\mathcal{T}$ to find the leading term of the characteristic polynomial of $A\mathcal{T}$. First we must limit the row and column indices allowed for a nonzero minor of the original matrix A.

Recall every matrix $A \in \mathbb{F}^{m\times m}$ is row-equivalent to a row-reduced echelon matrix [6, Thm.5]. Let us denote the columns of A corresponding to the columns of the corresponding row-reduced echelon matrix containing the leading nonzero entries of the rows as the pivot columns of A.

LEMMA 3.2. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ have rank r and pivot column indices $\mathscr{P} = \{p_1, \ldots, p_r\}$ where $1 \le p_1 < \cdots < p_r \le n$. Furthermore, let $\mathscr{Q} = \{q_1, \ldots, q_r\}$ where $1 \le q_1 < \cdots < q_r \le n$ be the indices of the last r linearly independent rows of A. If $\mathscr{I} = \{i_1, \ldots, i_m\}$ where $1 \le i_1 < \cdots < i_m \le n$ and $\mathscr{J} = \{j_1, \ldots, j_m\}$ where $1 \le j_1 < \cdots < j_m \le n$ are sets of $m \le r$ row and column indices, respectively, of A with $\det(A_{[\mathscr{I},\mathscr{J}]}) \neq 0$, then $j_\ell \ge p_\ell$ and $i_{m+1-\ell} \le q_{r+1-\ell}$ for all $1 \le \ell \le m$.*

PROOF. The rows indexed by the set $\mathscr{I}$ must be linearly independent for the minor $A_{[\mathscr{I},\mathscr{J}]}$ to be nonzero. The last m linearly independent rows of A are the rows index by $\{q_{r+1-m}, \ldots, q_r\}$. Thus, $i_{m+1-\ell} \le q_{r+1-\ell}$ for all $1 \le \ell \le m$.

Because every matrix A is row-equivalent to a row-reduced echelon matrix R_A, there exists an invertible matrix B_A such that $A = B_A R_A$. The Binet-Cauchy formula says the minor $A_{[\mathscr{I},\mathscr{J}]}$ of the matrix A is a sum of products of minors of B_A and R_A:

$$\det(A_{[\mathscr{I},\mathscr{J}]}) = \sum_{\substack{\mathscr{K}=\{k_1,\ldots,k_m\}\\ 1\le k_1<\cdots<k_m\le n}} \det(B_{A[\mathscr{I},\mathscr{K}]}) \det(R_{A[\mathscr{K},\mathscr{J}]}).$$

Suppose there exists an index ℓ such that $j_\ell < p_\ell$, and let $\widehat{\mathscr{J}}$ and $\mathscr{N}$ be the sets of indices $\widehat{\mathscr{J}} = \{j_1, \ldots, j_\ell\}$ and $\mathscr{N} = \{1, \ldots, n\}$. At most, the first first $\ell - 1$ entries of each column $R_{A[j_1]}, \ldots, R_{A[j_\ell]}$ of the row reduced echelon form R_A of A are nonzero. Thus, the submatrix $R_{A_{[\mathscr{N},\widehat{\mathscr{J}}]}}$ is not of full rank. By adding more columns, the submatrix $R_{A[\mathscr{N},\mathscr{J}]}$ must also not have full rank. Thus, $\det(R_{A[\mathscr{K},\mathscr{J}]}) = 0$ for every index set $\mathscr{K}$. Therefore, $\det(A_{[\mathscr{I},\mathscr{J}]}) = 0$, which contradicts the original assumptions. □

We can use this result to bound the leading term of any nonzero principal minor of the symbolically preconditioned matrix $A\mathcal{T}$.

THEOREM 3.1. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ have rank r, and let $\mathcal{T} \in \mathbb{F}[\alpha_1, \ldots, \alpha_{2n-1}]^{n\times n}$ be the symbolic Toeplitz matrix (2.2) where $\alpha_1, \ldots, \alpha_{2n-1}$ are distinct indeterminates over $\mathbb{F}$. Furthermore, let $\mathscr{P} = \{p_1, \ldots, p_r\}$ where $1 \le p_1 < \cdots < p_r \le n$ be the pivot column indices of A, and let $\mathscr{Q} = \{q_1, \ldots, q_r\}$ where $1 \le q_1 < \cdots < q_r \le n$ be the indices of the last r linearly independent rows of A. If $\mathscr{I} = \{i_1, \ldots, i_m\}$ where $1 \le i_1 < \cdots < i_m \le n$ is any set of $m \le r$ indices, then the leading term of the principal minor $\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})$ under the lexicographic monomial ordering with $\alpha_1 \prec \alpha_2 \prec \cdots \prec \alpha_{2n-1}$ is bounded by*

$$\begin{aligned}\operatorname{lt}\left(\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})\right) &\preceq \prod_{k=1}^{m} \alpha_{n+i_{r+1-k}-p_k} \\ &\preceq \prod_{k=1}^{m} \alpha_{n+q_{r+1-k}-p_k}.\end{aligned}$$

PROOF. The Binet-Cauchy formula says each minor of the matrix $A\mathcal{T}$ is a sum of products of minors of A and $\mathcal{T}$:

$$\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]}) = \sum_{\substack{\mathscr{J}=\{j_1,\ldots,j_m\}\\ 1\le j_1<\cdots<j_m\le n}} A_{[\mathscr{I},\mathscr{J}]}\mathcal{T}_{[\mathscr{J},\mathscr{I}]}. \quad (3.1)$$

Lemma 3.2 says if $A_{[\mathscr{I},\mathscr{J}]} \neq 0$, then $j_k \ge p_k$ and $i_{m+1-k} \le q_{r+1-k}$ for $1 \le k \le m$, which means

$$\alpha_{n+i_{m+1-k}-j_k} \preceq \alpha_{n+i_{m+1-k}-p_k} \preceq \alpha_{n+q_{r+1-k}-p_k}$$

for $1 \le k \le m$. Thus, for every index set $\mathscr{J}$ such that $A_{[\mathscr{I},\mathscr{J}]} \neq 0$,

$$\operatorname{lt}\left(\det(\mathcal{T}_{[\mathscr{J},\mathscr{I}]})\right) \preceq \prod_{k=1}^{m} \alpha_{n+i_{m+1-k}-p_k} \preceq \prod_{k=1}^{m} \alpha_{n+q_{r+1-k}-p_k}$$

by equation 2.3. The bound on the leading term of the principal minor then follows from the Binet-Cauchy formula (3.1). □

We can use this bound on the leading term of the nonzero principal minors of $A\mathcal{T}$ to bound the leading term of the characteristic polynomial $\lambda I - A\mathcal{T}$. To find this leading term, we need to find the leading terms of the $r \times r$ principal minors of $A\mathcal{T}$. We begin by using one particular nonzero minor of A to show the leading term attains its bound.

LEMMA 3.3. *Let $\mathbb{F}$ be a field, and let $A \in \mathbb{F}^{n\times n}$ have rank r and pivot column indices $\mathscr{P} = \{p_1, \ldots, p_r\}$ where $1 \le p_1 < \cdots < p_r \le n$. If $\mathscr{I} = \{i_1, \ldots, i_m\}$ where $1 \le i_1 < \cdots < i_m \le n$ is any index set of r linearly independent rows of A, then $\det(A_{[\mathscr{I},\mathscr{P}]}) \neq 0$.*

PROOF. The columns $A_{[p_1]}, \ldots, A_{[p_r]}$ of A form a basis for the column space of A, which means there exists a matrix $B \in \mathbb{F}^{r\times n}$ such that $A = A_{[\mathscr{N},\mathscr{P}]}B$ where $\mathscr{N}$ is the set of indices $\mathscr{N} = \{1, \ldots, n\}$. Considering only the rows of A indexed by the set $\mathscr{I}$, we obtain $A_{[\mathscr{I},\mathscr{N}]} = A_{[\mathscr{I},\mathscr{P}]}B$. By Sylvester's inequality [5, p.66],

$$\operatorname{rank}(A_{[\mathscr{I},\mathscr{N}]}) \le \operatorname{rank}(A_{[\mathscr{I},\mathscr{P}]}) \le r.$$

However, the rows of A indexed by the set $\mathscr{I}$ are linearly independent, so $\operatorname{rank}(A_{[\mathscr{I},\mathscr{N}]}) = r$. Thus, $\operatorname{rank}(A_{[\mathscr{I},\mathscr{P}]}) = r$ and $\det(A_{[\mathscr{I},\mathscr{P}]}) \neq 0$. □

We can now determine the leading term of any nonzero $r \times r$ principal minor of $A\mathcal{T}$.

THEOREM 3.2. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ have rank r and pivot column indices $\mathscr{P} = \{p_1, \ldots, p_r\}$ with $1 \le p_1 < \cdots < p_r \le n$, and let $\mathcal{T} \in \mathbb{F}[\alpha_1, \ldots, \alpha_{2n-1}]^{n\times n}$ be the symbolic Toeplitz matrix (2.2) where $\alpha_1, \ldots, \alpha_{2n-1}$ are distinct indeterminates over $\mathbb{F}$. If $\mathscr{I}$ where $1 \le i_1 < \cdots < i_m \le n$ is any index set of r rows of A, then either the principal minor $\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})$ is zero or its leading term under the lexicographic monomial ordering with $\alpha_1 \prec \alpha_2 \prec \cdots \prec \alpha_{2n-1}$ is*

$$\begin{aligned}&\operatorname{lt}\left(\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})\right) \\ &\qquad = (-1)^{\lfloor r/2 \rfloor} \det(A_{[\mathscr{I},\mathscr{P}]}) \prod_{k=1}^{r} \alpha_{n+i_{r+1-k}-p_k}.\end{aligned}$$

PROOF. If the rows of A indexed by $\mathscr{I}$ are not linearly independent, then $\det(A_{[\mathscr{I},\mathscr{J}]}) = 0$ for every set of r column indices $\mathscr{J}$. The Binet-Cauchy formula (3.1) then says the principal minor $\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})$ is zero.

Suppose the rows of A indexed by $\mathscr{I}$ are linearly independent and $\det(A_{[\mathscr{I},\mathscr{J}]}) \neq 0$ for the index set $\mathscr{J} = \{j_1, \ldots, j_r\}$ where $1 \le j_1 < \cdots < j_r \le n$. Lemma 3.2 says $j_k \ge p_k$ for $1 \le k \le r$, and thus

$$\operatorname{lt}\left(\det(\mathcal{T}_{[\mathscr{J},\mathscr{I}]}\right) \preceq \prod_{k=1}^{r} \alpha_{n+i_{r+1-k}-p_k}.$$

If $\mathscr{J} = \mathscr{P}$, then

$$\operatorname{lt}\left(\det(\mathcal{T}_{[\mathscr{P},\mathscr{I}]})\right) = (-1)^{\lfloor r/2 \rfloor} \prod_{k=1}^{r} \alpha_{n+i_{r+1-k}-p_k}$$

by equation 2.3. On the other hand, if $\mathscr{J} \neq \mathscr{P}$, then $j_k > p_k$ for some k, and $\operatorname{lt}(\det(\mathcal{T}_{[\mathscr{J},\mathscr{I}]})) \prec \operatorname{lt}(\det(\mathcal{T}_{[\mathscr{P},\mathscr{I}]}))$. The leading term of the principal minor then follows from the Binet-Cauchy formula (3.1). □

We use our previous results about the leading terms of the minors of $A\mathcal{T}$ to find the leading terms of the elementary symmetric functions of the eigenvalues of $A\mathcal{T}$. Recall the mth elementary symmetric function of the eigenvalues of $A\mathcal{T}$ is the sum of the $m \times m$ principal minors of $A\mathcal{T}$:

$$E_m(A\mathcal{T}) = \sum_{\substack{\mathscr{I}=\{i_1,\ldots,i_m\} \\ 1\leq i_1<\cdots<i_m\leq n}} \det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]}). \tag{3.2}$$

LEMMA 3.4. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ have rank r, and let $\mathcal{T} \in \mathbb{F}[\alpha_1, \ldots, \alpha_{2n-1}]^{n\times n}$ be the symbolic Toeplitz matrix* (2.2) *where $\alpha_1, \ldots, \alpha_{2n-1}$ are distinct indeterminates over $\mathbb{F}$. Furthermore, let $\mathscr{P} = \{p_1, \ldots, p_r\}$ where $1 \leq p_1 < \cdots < p_r \leq n$ be the pivot column indices of A, and let $\mathscr{Q} = \{q_1, \ldots, q_r\}$ where $1 \leq q_1 < \cdots < q_r \leq n$ be the indices of the last r linearly independent rows of A. The leading term of the rth elementary symmetric function of the eigenvalues of $A\mathcal{T}$ under the lexicographic monomial ordering with $\alpha_1 \prec \alpha_2 \prec \cdots \prec \alpha_{2n-1}$ is*

$$\operatorname{lt}\left(E_r(A\mathcal{T})\right) = (-1)^{\lfloor r/2 \rfloor} A_{[\mathscr{Q},\mathscr{P}]} \prod_{k=1}^{r} \alpha_{n+q_{r+1-k}-p_k}.$$

Otherwise, $\operatorname{lt}(E_m(A\mathcal{T})) \preceq \prod_{k=1}^{m} \alpha_{n+q_{r+1-k}-p_k}$ *if $m < r$ and $E_r(A\mathcal{T}) = 0$ if $m > r$.*

PROOF. Because $\operatorname{rank}(A\mathcal{T}) \leq \operatorname{rank}(A) = r$ by Sylvester's inequality [5, p.66], we know every minor of $A\mathcal{T}$ of size greater than r is zero. Thus, $E_m(A\mathcal{T}) = 0$ for $m > r$.

Otherwise, let $\mathscr{I}$ be the index set $\mathscr{I} = \{i_1, \ldots, i_m\}$ where $1 \leq i_1 < \cdots < i_m \leq n$. If $m < r$, Theorem 3.1 says

$$\operatorname{lt}\left(\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})\right) \preceq \prod_{k=1}^{m} \alpha_{n+q_{r+1-k}-p_k},$$

which means, by the definition of the elementary symmetric function (3.2), $\operatorname{lt}(E_m(A\mathcal{T})) \preceq \prod_{k=1}^{m} \alpha_{n+q_{r+1-k}-p_k}$.

If $m = r$, Theorem 3.2 says

$$\operatorname{lt}\left(\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})\right) = (-1)^{\lfloor r/2 \rfloor} \det(A_{[\mathscr{I},\mathscr{P}]}) \prod_{k=1}^{r} \alpha_{n+i_{r+1-k}-p_k}.$$

If $\mathscr{I} = \mathscr{Q}$, then

$$\operatorname{lt}\left(\det((A\mathcal{T})_{[\mathscr{Q},\mathscr{Q}]})\right) = (-1)^{\lfloor r/2 \rfloor} A_{[\mathscr{Q},\mathscr{P}]} \prod_{k=1}^{r} \alpha_{n+q_{r+1-k}-p_k}.$$

If $\mathscr{I} \neq \mathscr{Q}$, then $i_k > p_k$ for some k by Lemma 3.2, and

$$\operatorname{lt}\left(\det((A\mathcal{T})_{[\mathscr{I},\mathscr{I}]})\right) \prec \operatorname{lt}\left(\det((A\mathcal{T})_{[\mathscr{Q},\mathscr{Q}]})\right).$$

Thus, the leading term of the rth elementary symmetric function is $\operatorname{lt}(E_r(A\mathcal{T})) = \operatorname{lt}(\det((A\mathcal{T})_{[\mathscr{Q},\mathscr{Q}]}))$. □

We can finally state the leading term of $\det(\lambda I - A\mathcal{T})$. We will use this leading term to show the characteristic polynomial has no repeated factors other than λ.

THEOREM 3.3. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ have rank r, and let $\mathcal{T} \in \mathbb{F}[\alpha_1, \ldots, \alpha_{2n-1}]^{n\times n}$ be the symbolic Toeplitz matrix* (2.2) *where $\alpha_1, \ldots, \alpha_{2n-1}$ are distinct indeterminates over $\mathbb{F}$. Furthermore, let $\mathscr{P} = \{p_1, \ldots, p_r\}$ where $1 \leq p_1 < \cdots < p_r \leq n$ be the pivot column indices of A, and let $\mathscr{Q} = \{q_1, \ldots, q_r\}$ where $1 \leq q_1 < \cdots < q_r \leq n$ be the indices of the last r linearly independent rows of A. If λ is an indeterminate distinct from $\alpha_1, \ldots, \alpha_{2n-1}$, then λ^{n-r} divides* $\det(\lambda I - A\mathcal{T})$ *and the leading term of the characteristic polynomial* $\det(\lambda I - A\mathcal{T})$ *under the lexicographic monomial ordering with $\lambda \prec \alpha_1 \prec \alpha_2 \prec \cdots \prec \alpha_{2n-1}$ is*

$$(-1)^{\lfloor 3r/2 \rfloor} A_{[\mathscr{Q},\mathscr{P}]} \lambda^{n-r} \prod_{k=1}^{r} \alpha_{n+p_{m+1-k}-q_k}.$$

PROOF. The matrix $A\mathcal{T}$ has characteristic polynomial

$$\det(\lambda I - A\mathcal{T}) = \sum_{m=0}^{n} (-1)^m E_m(A\mathcal{T}) \lambda^{n-m}. \tag{3.3}$$

Lemma 3.4 says $E_m(A\mathcal{T}) = 0$ for $m > r$ and λ^{n-r} divides the characteristic polynomial. The lemma also says

$$\operatorname{lt}\left(E_r(A\mathcal{T})\right) = (-1)^{\lfloor r/2 \rfloor} A_{[\mathscr{Q},\mathscr{P}]} \prod_{k=1}^{r} \alpha_{n+q_{r+1-k}-p_k},$$

which means

$$\operatorname{lt}\left((-1)^r E_r(A\mathcal{T})\lambda^{n-r}\right) = (-1)^{\lfloor 3r/2 \rfloor} A_{[\mathscr{Q},\mathscr{P}]} \lambda^{n-r} \prod_{k=1}^{r} \alpha_{n+q_{r+1-k}-p_k}.$$

If $m < r$, then $\lambda \prec \alpha_{n+q_{r+1-k}-p_k}$ for all k means

$$\operatorname{lt}\left((-1)^m E_m(A\mathcal{T})\lambda^{n-m}\right) \preceq \lambda^{n-m} \prod_{k=1}^{m} \alpha_{n+q_{r+1-k}-p_k} \prec \operatorname{lt}\left((-1)^r E_r(A\mathcal{T})\lambda^{n-r}\right).$$

and the proof follows from equation 3.3. □

Notice this leading term contains the smallest degree of λ possible and is at most linear in each of the α_k. This is the result we desired to show the characteristic polynomial has no repeated factors other than λ.

THEOREM 3.4. *Let $\mathbb{F}$ be a field, let $A \in \mathbb{F}^{n\times n}$ have rank r with $r \leq n-1$, and let $\mathcal{T} \in \mathbb{F}[\alpha_1, \ldots, \alpha_{2n-1}]^{n\times n}$ be the symbolic Toeplitz matrix* (2.2) *where $\alpha_1, \ldots, \alpha_{2n-1}$ are distinct indeterminates over $\mathbb{F}$. Then, the matrix $A\mathcal{T}$ has characteristic polynomial $\det(\lambda I - A\mathcal{T}) = \lambda^{n-r} g$ and minimal polynomial $f^{A\mathcal{T}} = \lambda\, g$ where g is squarefree, λ does not divide g, and $\deg_\lambda(f^{A\mathcal{T}}) = r + 1$.*

PROOF. Because λ^{n-r} divides the characteristic polynomial by Theorem 3.3, we know $\det(\lambda I - A\mathcal{T}) = \lambda^{n-r} g$. By the same theorem, λ^{n-r+1} does not divide the leading term of the characteristic polynomial, so we know the leading term of g is constant in λ and at most linear in each α_k. Therefore, λ does not divide g. λ.

Suppose g has a repeated factor h. Then, h^2 divides g, and $(\operatorname{lt}(h))^2 = \operatorname{lt}(h^2)$ must divide the leading term of g. This means $\operatorname{lt}(h) = 1$, and thus g is squarefree.

Because $\det(\lambda I - A\mathcal{T})$ has no repeated factors other than λ, all the invariant factors of $\det(\lambda I - A\mathcal{T})$ except the largest are either 1 or a power of λ, so $f^{A\mathcal{T}} = \lambda^p g$ where $p \geq 1$. Because $\deg_\lambda(\lambda I - A\mathcal{T}) = n$, we know $\deg_\lambda(g) = r$, which means $\deg_\lambda(f^{A\mathcal{T}}) \geq r+1$. The proof then follows from the upper bound on $\deg_\lambda(f^{A\mathcal{T}})$ given by Lemma 3.1 where $\mathcal{B}_1 = I$ and $\mathcal{B}_2 = \mathcal{T}$. □

This means we can use the symbolic preconditioner $A\mathcal{T}$ in the Kaltofen-Saunders rank algorithm. The algorithm would be deterministic, but it would also be exponential in the size of the input matrix. We can use the Schwartz-Zippel Lemma [13] to convert this symbolic preconditioner to a randomized preconditioner that will make the Kaltofen-Saunders rank algorithm probabilistic and polynomial time.

THEOREM 3.5. *Let $\mathbb{F}$ be a field, $A \in \mathbb{F}^{n\times n}$ have rank r with $r \leq n-1$, and S be a finite subset of $\mathbb{F}$. If*

$$T = \begin{bmatrix} a_n & a_{n+1} & \dots & a_{2n-1} \\ a_{n-1} & a_n & \dots & a_{2n-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & a_n \end{bmatrix} \in \mathbb{F}^{n\times n}, \tag{3.4}$$

where $a_1, \dots, a_{2n-1}$ are chosen uniformly and independently from S, then the matrix AT has characteristic polynomial $\det(\lambda I - AT) = \lambda^{n-r} g(\lambda)$ and minimal polynomial $f^{AT} = \lambda\, g(\lambda)$ where $g(0) \neq 0$ and $\deg(f^{AT}) = r+1$, all with probability at least

$$1 - \frac{r(r+1)}{2|S|} \geq 1 - \frac{n(n-1)}{2|S|}.$$

PROOF. Suppose $|\mathbb{F}| > r(r+1)/2$; otherwise the result is trivial. By Theorem 3.4, if $\mathcal{T} \in \mathbb{F}[\alpha_1, \dots, \alpha_{2n-1}]^{n\times n}$ is the symbolic Toeplitz matrix (2.2) where $\alpha_1, \dots, \alpha_{2n-1}$ are distinct indeterminates over $\mathbb{F}$, then

$$\det(A\mathcal{T}) = \lambda^{n-r} g(\lambda, \alpha_1, \dots, \alpha_{2n-1})$$

and $f^{A\mathcal{T}} = \lambda\, g$ where g is not divisible by λ and $f^{A\mathcal{T}}$ has degree $r+1$ in lambda. Let $y \in \mathbb{F}^n$ be a vector such that $y, (A\mathcal{T})y, \dots, (A\mathcal{T})^r y$ are linearly independent. There is a $(r+1)\times(r+1)$ submatrix of the matrix with these vectors as its columns whose determinant is a nonzero polynomial in $\alpha_1, \dots, \alpha_{2n-1}$. This polynomial has total degree at most $r(r+1)/2$ in the indeterminates $\alpha_1, \dots, \alpha_{2n-1}$. If values $a_1, \dots, a_{2n-1}$ are chosen uniformly and independently from S for the indeterminates $\alpha_1, \dots, \alpha_{2n-1}$, the determinant is a nonzero element of $\mathbb{F}$ with probability at least $1 - r(r+1)/(2|S|)$ by the Schwartz-Zippel Lemma. The vectors $y, (AT)y, \dots, (AT)^r y$ are linearly independent for our randomly chosen Toeplitz matrix (3.4) and the kth invariant factor of $\lambda I - AT$ is obtained from the kth invariant factor of $\lambda I - A\mathcal{T}$ by replacing the indeterminates $\alpha_1, \dots, \alpha_{2n-1}$ with the values $a_1, \dots, a_{2n-1}$, respectively. □

Not only does AT satisfy the requirements of the Kaltofen-Saunders rank algorithm, but TA does as well. The proof follows from the uniqueness of the Smith form to show the matrices $\lambda I - A^T T$ and $\lambda I - TA$ have the same Smith form.

This is an improvement in probability of success over requiring a diagonal multiplier. For comparison, the Kaltofen-Saunders rank algorithm using two butterfly network preconditioners and a diagonal multiplier or two unit Toeplitz matrices and a diagonal multiplier have probabilities of success of at least

$$\left(1 - \frac{r\lceil \log_2(n)\rceil}{|S|}\right)^2 \left(1 - \frac{r(r+1)}{2|S|}\right)$$

and

$$1 - \frac{3r(r+1)}{2|S|},$$

respectively. On the other hand, Toeplitz matrices have slower matrix-vector products than butterfly network preconditioners, requiring $O(n\log(n)\,\mathrm{loglog}(n))$ field operations instead of only $O(n\log(n))$).

4. REFERENCES

[1] BENEŠ, V. E. Permutation groups, complexes, and rearrangeable connecting networks. *Bell System Tech. J. 43* (1964), 1641–1656.

[2] CHEN, L., EBERLY, W., KALTOFEN, E., SAUNDERS, B. D., TURNER, W. J., AND VILLARD, G. Efficient matrix preconditioners for black box linear algebra. *Linear Algebra Appl. 343-344* (Mar. 2002), 119–146. Special issue on *Infinite Systems of Linear Equations Finitely Specified*, edited by P. Dewilde, V. Olshevsky and A. H. Sayed.

[3] COX, D., LITTLE, J., AND O'SHEA, D. *Ideals, Varieties, and Algorithms: An Introduction to Computational Geometry and Commutative Algebra.* Springer Verlag, Heidelberg, Germany, 1996.

[4] EBERLY, W., AND KALTOFEN, E. On randomized Lanczos algorithms. In *Proc. ISSAC '97* (New York, 1997), W. W. Küchlin, Ed., ACM Press, pp. 176–183.

[5] GANTMACHER, F. R. *Matrix Theory*, vol. I. AMS Chelsea Pub., Providence, 1977.

[6] HOFFMAN, K., AND KUNZE, R. *Linear Algebra*, 2nd ed. Prentice Hall, Upper Saddle River, NJ, 1971.

[7] KALTOFEN, E., AND LOBO, A. On rank properties of toeplitz matrices over finite fields. In *Proc. ISSAC '96* (New York, 1996), Y. N. Lakshman, Ed., ACM Press, pp. 241–249.

[8] KALTOFEN, E., AND SAUNDERS, B. D. On Wiedemann's method of solving sparse linear systems. In *Proc. AAECC-9* (Heidelberg, Germany, 1991), H. F. Mattson, T. Mora, and T. R. N. Rao, Eds., vol. 539 of *Lect. Notes Comput. Sci.*, Springer Verlag, pp. 29–38.

[9] PARKER, D. S. Random butterfly transformations with applications in computational linear algebra. Technical Report CSD-950023, UCLA, 1995.

[10] TURNER, W. J. *Black Box Linear Algebra with the LinBox Library.* PhD thesis, N. Carolina State Univ., Raleigh, North Carolina, May 2002.

[11] TURNER, W. J. Determinantal divisors and matrix preconditioners. Submitted to J. Symbolic Comput., June 2003.

[12] WIEDEMANN, D. H. Solving sparse linear equations over finite fields. *IEEE Trans. Inform. Theory IT-32*, 1 (1986), 54–62.

[13] ZIPPEL, R. Probabilistic algorithms for sparse polynomials. In *Proc EUROSAM '79* (Heidelberg, Germany, 1979), E. W. Ng, Ed., vol. 72 of *Lect. Notes Comput. Sci.*, Springer Verlag, pp. 216–226.