On Randomized Lanczos Algorithms

Wayne Eberly*
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada T2N 1N4
eberly@cpsc.ucalgary.ca
http://www.cpsc.ucalgary.ca/~eberly

Erich Kaltofen†
Department of Mathematics
North Carolina State University
Raleigh, North Carolina, USA 27695-8205
kaltofen@eos.ncsu.edu
http://www4.ncsu.edu/~kaltofen

Abstract

Las Vegas algorithms that are based on Lanczos's method for solving symmetric linear systems are presented and analyzed. These are compared to a similar randomized Lanczos algorithm that has been used for integer factorization, and to the (provably reliable) algorithm of Wiedemann. The analysis suggests that our Lanczos algorithms are preferable to several versions of Wiedemann's method for computations over large fields, especially for certain symmetric matrix computations.

1 Introduction

Sparse or structured systems of linear equations over fields arise in a variety of applications; for example, many methods for integer factorization require the solutions of large, sparse systems over finite fields.

Several algorithms have been proposed for this computation over the years. Until recently, the algorithm of Wiedemann [15] was the only such algorithm known to be provably efficient and reliable for computations for arbitrary fields — particularly, over small finite fields. However, other algorithms continued to be modified and applied to solve these problems. For example, LaMacchia and Odlyzko [7] (among others) have adapted Lanczos' and conjugate gradient methods in order to factor integers and compute discrete logarithms, and this motivates the present work.

The algorithms of Lanczos and Wiedemann can each be used to solve a nonsingular system of n linear equations in n unknowns over the complex numbers, by performing a linear number (in n) of multiplications of either the coefficient matrix or its transpose by a vector and by performing additional work using $\Theta(n^2)$ additional operations over the ground field; the cost of the matrix-vector multiplications generally dominates the cost of all other work. While these algorithms have essentially the same asymptotic time complexity, Wiedemann's algorithm (as originally described) requires either that certain matrix-vector products be stored for reuse (using quadratic storage space when linear space suffices for the Lanczos methods) or that these products be recomputed (so that up to n more matrix-vector products are performed).

*Research was supported in part by the Natural Sciences and Engineering Research Council of Canada, under research grant number OGP0089756.

†Research was supported in part by the Natural Sciences Foundation, under grant number CCR-9319776.

However, Wiedemann's algorithm is provably reliable for computations over arbitrary fields, while problems arise when one tries to apply Lanczos' method to solve systems of linear equations over fields of positive characteristic. In particular, the existence of vectors that are "self-orthogonal" introduces the possibility of a division by zero when the standard Lanczos algorithm is applied. "Lanczos methods with lookahead" attempt to address this problem, and either reduce or eliminate the possibility of a division by zero; algorithms of this type have been described by Coppersmith [1], Montgomery [10], and Teitelbaum [13]. These algorithms are somewhat more complicated than the standard algorithm, and may require additional storage space, additional matrix-vector multiplications, or both — so that the apparent advantage of using the Lanczos method over Wiedemann's algorithm may be reduced or eliminated altogether when lookahead is used.

Recently, Lambert [8] has applied lookahead along with some additional techniques, to produce a "Lanczos-style" algorithm that is provably as reliable as the original Wiedemann algorithm and requires almost exactly the same number of matrix-vector multiplications in the worst case (again, along with $\Theta(n^2)$ additional field operations), in either the case that linear or quadratic storage is required. Thus this is certainly competitive with the original Wiedemann algorithm. Furthermore, it can be shown that either a smaller number of matrix-vector multiplications, or a smaller amount of storage space, is needed by Lambert's new algorithm if it does not have to "look ahead" too far (see also Teitelbaum [13] for a similar argument). Thus, Lambert's algorithm may be superior to Wiedemann's (original) algorithm in some cases. However, this advantage has yet to be proved unconditionally.

LaMacchia and Odlyzko [7] have used a different approach in order to improve the reliability of Lanczos' algorithms over fields of positive characteristic: They randomize the system to be solved in a simple (and extremely inexpensive) way, rather than adopting any kind of a lookahead strategy, and their algorithm fails — one hopes, with small probability — on an attempt to divide by zero. Their experimental evidence suggests that this approach is effective, provided that one works over a field with substantially more than n elements (where, again, n is the order of the system to be solved).

We remark that "block Wiedemann" algorithms have been described and analyzed by Coppersmith [2], Kaltofen [4], and Villard [14]. These are now provably as reliable as Wiedemann's or Lambert's algorithms for computations

over finite fields of any size [14], and (with appropriately chosen parameters) they require fewer matrix-vector multiplications than either one, while using at the same time linear storage space. To our knowledge, Kaltofen's sequential version of the block Wiedemann algorithm [4, Corollary to Theorem 7] has the least number of matrix-vector products of any known algorithm. Its full analysis is at this time restricted to nonsingular matrices; see Section 8 for a more detailed discussion.

In this paper, we present randomized Lanczos algorithms, that apply the standard Lanczos algorithm (which is reviewed in Sections 2 and 3) to perturbed systems. These systems are obtained by randomizing the coefficient matrix (by a method similar to that used by LaMacchia and Odlyzko) and by randomizing the vector b as well. More precisely, we present three successively more general and expensive randomizations of the given system, and we identify systems that can be solved reliably using each of these, provided that one can choose elements uniformly and independently from a sufficiently large subset of the ground field.

The simplest randomization we consider (in Section 4) is a randomization of the vector b, through addition by a random element of the column space of A, that leaves the coefficient matrix unchanged. Our results imply that if the coefficient matrix A is square, symmetric, and has a characteristic polynomial that is a product of z^{n-r} and a squarefree polynomial of degree r that is not divisible by z, where n and r are the order and rank of A respectively, then this randomization is sufficient with high probability.

The second randomization we consider (in Section 5) includes pre- and post-multiplication of the coefficient matrix by a randomly chosen diagonal matrix, along with the above randomization of b. If the coefficient matrix A is square, symmetric, and has "generic rank profile" (or, more generally, if the rows and columns of A can be permuted in a symmetric way to achieve this), then this second randomization is also sufficient with high probability.

The "randomized Lanczos algorithms" that use these randomizations have almost the same storage requirements and use almost the same number of matrix-vector multiplications by the coefficient matrix as the standard Lanczos algorithm. These randomized algorithms appear to be more efficient (although, also more limited) than previous algorithms for which proofs of reliability are available.

The third randomization considered (in Section 6) includes a further randomization of the coefficient matrix; multiplication of a resulting coefficient matrix by a vector requires matrix-vector multiplications by both the original coefficient matrix and its transpose, so that the resulting randomized algorithm requires approximately twice as many matrix-vector products (with approximately the same storage requirements) as the standard Lanczos algorithm.

This (final) randomized algorithm can be applied to solve an arbitrary linear system $Ax = b$ with high probability, provided that the ground field is sufficiently large and that a solution for the system does exist. The asymptotic results for this algorithm are inferior to those that have already been established for the sequential versions of the block Wiedemann algorithm discussed above, when the coefficient matrix is nonsingular. However, they do suggest that the new algorithm is slightly superior to all other known algorithms, including Wiedemann's algorithm (as originally given) and Lambert's new algorithm, for sparse and structured matrix computations over large fields. This is especially so if the number of field elements that can be stored and reused during the computation is restricted to linear in the dimension of the coefficient matrix.

As mentioned above, this work was motivated by the experimental results of LaMacchia and Odlyzko. Their work concerned the use of a "part" of the randomization we describe. In particular, LaMacchia and Odlyzko employ the matrix randomization given here in Section 6 without randomizing b (as in Section 4) or employing the additional matrix randomization given here in Section 5.

Since LaMacchia and Odlyzko found that this partial randomization works well in practice, it seems natural to ask whether the additional randomizations given here are necessary. In Section 7, we show that they are required if one wants an algorithm that works as reliably as the one given in Section 6 in the general case. In particular, we present a family of linear systems over arbitrarily large fields for which LaMacchia and Odlyzko's algorithm must fail. An alternative (inexpensive) randomization is considered in this section, and proved to be insufficient, in the same way.

As noted above, this paper concludes with a comparison of Lanczos- and Wiedemann-style algorithms, in Section 8.

Additional modifications of the Lanczos method remain to be analyzed. In particular, Coppersmith's and Montgomery's algorithms are "block Lanczos algorithms" — they use multiplication of matrices by blocks of vectors (along with lookahead strategies), in order to reduce the number of operations to be performed. We are presently investigating randomized block Lanczos algorithms, and suspect that it will be possible to use randomization to replace lookahead strategies in the block algorithms for computations over large fields, as is the case for Lanczos algorithms that do not employ blocking.

While this short version of the paper presents all the randomizations that are discussed, along with bounds on the probability that they are sufficient, it omits proofs that these bounds are correct. A full version of this paper, which includes these proofs, is available at the URLs for the authors that are listed at the beginning of the paper.

2 The Standard Lanczos Algorithm

Lanczos' method [9] was developed to solve systems with real coefficients. However, as described below, it can be applied over other fields as well.

Consider now the problem of solving a system $Ax = b$ of n linear equations in n unknowns over a field F, when the coefficient matrix A is symmetric.

Let K denote the Krylov space generated by b, that is, the subspace of $\mathsf{F}^{n\times 1}$ spanned by the vectors $b, Ab, A^2b, \ldots$. The map from $K \times K$ to F given by

$$\langle x,\, y\rangle = x^t Ay \qquad \text{for } x, y \in K \tag{2.1}$$

is F-linear in both inputs x and y. Since A is symmetric, it is easy to verify that

$$\langle Ax,\, y\rangle = x^t A^2 y = \langle x,\, Ay\rangle \qquad \text{for } x, y \in K. \tag{2.2}$$

Now, let s be the dimension of K and suppose K has a basis $\omega_0, \omega_1, \ldots, \omega_{s-1}$ whose elements are orthogonal with respect to the above map,

$$\langle \omega_i,\, \omega_j\rangle = \omega_i^t A \omega_j = 0 \quad \text{if } 0 \le i, j < s \text{ and } i \ne j. \tag{2.3}$$

Furthermore, suppose that no element of this basis is self-orthogonal with respect to this map,

$$\langle \omega_i, \omega_i \rangle = \omega_i^t A \omega_i \neq 0 \qquad \text{if } 0 \leq i < s. \tag{2.4}$$

Then, every element x of K is determined by the values $\langle x, \omega_i \rangle$ for $0 \leq i < s$:

$$x = \sum_{i=0}^{s-1} \frac{\langle x, \omega_i \rangle}{\langle \omega_i, \omega_i \rangle} \omega_i. \tag{2.5}$$

This is easily proved using Equations (2.3) and (2.4), and the condition that $\omega_0, \omega_1, \ldots, \omega_{s-1}$ is a basis for K.

To solve the system $Ax = b$ by Lanczos' method, one attempts to construct a basis $\omega_0, \omega_1, \ldots, \omega_{s-1}$ for K satisfying equations (2.3) and (2.4) by setting ω_0 to be b and, for $i \geq 0$, setting ω_{i+1} to be the vector obtained by "orthogonalizing" $A\omega_i$ with respect to $\omega_0, \omega_1, \ldots, \omega_i$:

$$\omega_{i+1} = A\omega_i - \sum_{j=0}^{i} \frac{\langle A\omega_i, \omega_j \rangle}{\langle \omega_j, \omega_j \rangle} \omega_j. \tag{2.6}$$

The method fails if a nonzero vector ω_i is encountered along the way such that $\langle \omega_i, \omega_i \rangle = 0$. Provided this does not occur, it is easily checked by induction on i that $\omega_0, \omega_1, \ldots, \omega_{i-1}$ span the same subspace as $b, Ab, \ldots, A^{i-1}b$, so that (in particular) if s is the dimension of the Krylov space K then $\omega_0, \omega_1, \ldots, \omega_{s-1}$ forms a basis for K.

If it were necessary to orthogonalize $A\omega_i$ with respect to ω_j explicitly, for all $j \leq i$, then it would be necessary to compute $\langle x, y \rangle$ for $\Theta(s^2)$ pairs of vectors $x, y \in \mathsf{F}^{n \times 1}$ in order to construct the above basis for K. Fortunately, this is not the case.

Lemma 2.1. *Let $A \in \mathsf{F}^{n \times n}$ be symmetric, $b \in \mathsf{F}^{n \times 1}$, and let $\omega_1, \omega_2, \ldots, \omega_{s-1}$ be as given in Equation (2.6). If $0 \leq i < s$ and $0 \leq j \leq i - 2$ then $\langle A\omega_i, \omega_j \rangle = 0$.*

It follows that it is sufficient to orthogonalize $A\omega_i$ with respect to ω_i and ω_{i-1} in order to ensure that the resulting vector ω_{i+1} is orthogonal to each of $\omega_0, \omega_1, \ldots, \omega_i$. Thus we will set

$$\begin{aligned} \omega_{i+1} &:= A\omega_i - \frac{\langle A\omega_i, \omega_i \rangle}{\langle \omega_i, \omega_i \rangle} \omega_i - \frac{\langle A\omega_i, \omega_{i-1} \rangle}{\langle \omega_{i-1}, \omega_{i-1} \rangle} \omega_{i-1} \\ &= A\omega_i - \frac{\omega_i^t A^2 \omega_i}{\omega_i^t A \omega_i} \omega_i - \frac{\omega_i^t A^2 \omega_{i-1}}{\omega_{i-1}^t A \omega_{i-1}} \omega_{i-1}. \end{aligned}$$

We stop as soon as a vector ω_i is found such that $\langle \omega_i, \omega_i \rangle = 0$. If $\omega_i = 0$ then $i = s$ and $\omega_0, \omega_1, \ldots, \omega_{s-1}$ is a basis for K. Otherwise the standard Lanczos algorithm cannot be used to construct a basis for this space.

Suppose now that A is nonsingular or, slightly more generally, that there is a unique x in the Krylov space K such that $Ax = b$. If the algorithm did not fail on inputs A and b then, since $\langle x, \omega_i \rangle = x^t A \omega_i = b^t \omega_i$ for $0 \leq i < s$, and $\omega_0, \omega_1, \ldots, \omega_{s-1}$ is a basis for K,

$$x = \sum_{j=0}^{s-1} \frac{b^t \omega_j}{\omega_j^t A \omega_j} \omega_j. \tag{2.7}$$

If we define

$$x_i = \sum_{j=0}^{i} \frac{b^t \omega_i}{\langle \omega_j, \omega_j \rangle} \omega_j, \tag{2.8}$$

Input: $A \in \mathsf{F}^{n \times n}$, symmetric; $b \in \mathsf{F}^{n \times 1}$
Output: $x \in \mathsf{F}^{n \times 1}$ such that $Ax = b$, or failure

{ ω_{-1}, v_0 and t_{-1} are defined to ensure that ω_1 is correctly computed from ω_0 }
$\omega_{-1} := 0$; $v_0 := 0$; $t_{-1} := 1$
{ $v_{i+1} = A\omega_i$ and $t_i = \langle \omega_i, \omega_i \rangle$ for $i \geq 0$ }
$\omega_0 := b$
if $\omega_0 = 0$ then $x := 0$; return x
else
  $v_1 := A\omega_0$; $t_0 := v_1^t \omega_0$
  if $t_0 = 0$ then return failure
  else
    $x_0 := \dfrac{b^t \omega_0}{t_0} \omega_0$
  end if
end if
$i := 0$
repeat
  $\omega_{i+1} := v_{i+1} - \dfrac{v_{i+1}^t v_{i+1}}{t_i} \omega_i - \dfrac{v_{i+1}^t v_i}{t_{i-1}} \omega_{i-1}$
  if $\omega_{i+1} = 0$ then
    { The next test is not needed if A is nonsingular or if this is called by an algorithm that checks its output }
    if $Ax_i \neq b$ then
      return failure
    end if
    $x := x_i$; return x
  else
    $v_{i+2} := A\omega_{i+1}$; $t_{i+1} := \omega_{i+1}^t v_{i+2}$
    if $t_{i+1} = 0$ then return failure
    else
      $x_{i+1} := x_i + \dfrac{b^t \omega_{i+1}}{t_{i+1}} \omega_{i+1}$; $i := i + 1$
    end if
  end if
until false

Figure 1: The Standard Lanczos Algorithm

then (setting $x_{-1} = 0$),

$$x_i = x_{i-1} + \frac{b^t \omega_i}{\omega_i^t A \omega_i} \omega_i \tag{2.9}$$

and, clearly, $x = x_s$.

The standard Lanczos algorithm is given in Figure 1. Since $s \leq \min(n, r + 1)$, if A has rank r, this algorithm uses at most $\min(n, r + 1)$ multiplications of A by a vector to compute $v_1, v_2, \ldots, v_s$ from $\omega_0, \omega_1, \ldots, \omega_{s-1}$ respectively, and at most one additional multiplication by A to check its output if A is singular (and the output is not verified elsewhere). It also uses $O(s) \subseteq O(r)$ additional computations of (standard) inner products and sums of vectors, and $O(nr) \subseteq O(n^2)$ additional operations over the field F. Since it is only necessary to remember ω_{i-1}, t_{i-1}, ω_i, t_i, and x_i

in order to generate ω_{i+1}, t_{i+1} and x_{i+1}, it is clear that this algorithm uses linear storage space.

3 Correctness of the Standard Algorithm

Definition 3.1. A matrix $A \in \mathsf{F}^{n\times n}$ has *generic rank profile* if the leading $i \times i$ submatrix of A is nonsingular for every integer i such that $1 \le i \le n$.

Suppose as above that $b \in \mathsf{F}^{n\times 1}$ and s is the dimension of the Krylov space K for A and b, and let

$$H(A,b) = \begin{bmatrix} b^tAb & b^tA^2b & \cdots & b^tA^sb \\ b^tA^2b & b^tA^3b & \cdots & b^tA^{s+1}b \\ \vdots & \vdots & \ddots & \vdots \\ b^tA^sb & b^tA^{s+1}b & \cdots & b^tA^{2s-1}b \end{bmatrix}, \qquad (3.1)$$

so that the $(i,j)^{\text{th}}$ entry is $b^tA^{i+j-1}b$ for $1 \le i,j \le s$.

Lemma 3.2. *Suppose that $A \in \mathsf{F}^{n\times n}$ is symmetric and that $b \in \mathsf{F}^{n\times 1}$ is in the column space of A. Then the standard Lanczos algorithm can be used to find a basis $\omega_0, \omega_1, \ldots, \omega_{s-1}$ for K consisting of vectors satisfying Equations (2.3) and (2.4), and a vector $x \in K$ such that $Ax = b$, if and only if $H(A,b)$ has generic rank profile.*

4 Randomization of b

We now describe a condition on a symmetric matrix $A \in \mathsf{F}^{n\times n}$ that ensures that the set of vectors $b \in \mathsf{F}^{n\times 1}$, such that the system $Ax = b$ can be solved successfully using the standard Lanczos algorithm, is a dense subset of the column space of A — provided that the ground field F is sufficiently large. If A satisfies this condition and $b \in \mathsf{F}^{n\times 1}$ is in the column space of A, and if a vector d is chosen uniformly and randomly from $\mathsf{F}^{n\times 1}$ (or from $S^{n\times 1}$ for a sufficiently large finite subset S of F), then there is a reasonably high probability that the standard Lanczos algorithm can be used successfully to solve the system $A\hat{x} = (b + Ad)$. In this case it is clear that $x = \hat{x} - d$.

Let

$$H_i(A,b) = \begin{bmatrix} b^tAb & b^tA^2b & \cdots & b^tA^ib \\ b^tA^2b & b^tA^3b & \cdots & b^tA^{i+1}b \\ \vdots & \vdots & \ddots & \vdots \\ b^tA^ib & b^tA^{i+1}b & \cdots & b^tA^{2i-1}b \end{bmatrix} \in \mathsf{F}^{i\times i}$$

be a Hankel matrix with $(j,k)^{\text{th}}$ entry $b^tA^{j+k-1}b$ for $1 \le j,k \le i$, for $1 \le i \le s$, so that $H_i(A,b)$ is the leading $i \times i$ minor of the matrix $H(A,b)$ given in equation (3.1).

Lemma 4.1. *Suppose that $A \in \mathsf{F}^{n\times n}$ is symmetric with rank $r \le n$ and that $b \in \mathsf{F}^{n\times 1}$ is in the column space of A. Let $y_1, y_2, \ldots, y_n$ be indeterminates over F and let*

$$\hat{b} = b + A\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathsf{F}[y_1, y_2, \ldots, y_n]^{n\times 1}.$$

If the characteristic polynomial $f \in \mathsf{F}[z]$ of A is the product of z^{n-r} and a squarefree polynomial of degree r that is not divisible by z, then $\det H_i(A,\hat{b})$ is a nonzero polynomial in $\mathsf{F}[y_1, y_2, \ldots, y_n]$ with total degree at most $2i$ in the indeterminates $y_1, y_2, \ldots, y_n$, for $1 \le i \le r$.

Input: $A \in \mathsf{F}^{n\times n}$, symmetric; $b \in \mathsf{F}^{n\times 1}$
Output: $x \in \mathsf{F}^{n\times 1}$ such that $Ax = b$, or failure

Select elements $\gamma_1, \gamma_2, \ldots, \gamma_n$ uniformly and independently from a finite subset S of F
$\hat{b} := b + A\vec{\gamma}$, for $\vec{\gamma} = \begin{bmatrix}\gamma_1 & \gamma_2 & \cdots & \gamma_n\end{bmatrix}^t$
Apply the standard Lanczos algorithm to try to find a vector $\hat{x}$ such that $A\hat{x} = \hat{b}$
if this attempt succeeds then
  $x := \hat{x} - \vec{\gamma}$; return x
else
  return failure
end if

Figure 2: Randomization of b

The next theorem is a consequence of Lemma 4.1 and the "Schwartz-Zippel lemma" ([12], [16]).

Theorem 4.2. *Let $A \in \mathsf{F}^{n\times n}$ be symmetric with rank $r \le n$ such that the characteristic polynomial of A (in $\mathsf{F}[z]$) is the product of z^{n-r} and a squarefree polynomial of degree r that is not divisible by z. Let $b \in \mathsf{F}^{n\times 1}$ be in the column space of A, and let S be a finite subset of F.*

If $\gamma_1, \gamma_2, \ldots, \gamma_n$ are chosen uniformly and independently from S, and

$$\hat{b} = b + A\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} \in \mathsf{F}^{n\times 1},$$

then the standard Lanczos algorithm can be used successfully to find a vector $\hat{x} \in \mathsf{F}^{n\times 1}$ such that $A\hat{x} = \hat{b}$ with probability at least $1 - \frac{n(n+1)}{|S|}$.

An algorithm that "randomizes the right side" à la Kaltofen and Saunders [6] is given in Figure 2. Theorem 4.2 implies that this algorithm can be used to solve a system $Ax = b$ with high probability if the characteristic polynomial of A satisfies the given condition and if the ground field is sufficiently large. It is clear by inspection of this algorithm that it uses only one more multiplication of A by a vector than the standard Lanczos algorithm, and only $O(n)$ other additional field operations.

5 An Efficient but Limited Matrix Randomization

The condition on the symmetric matrix A that is given in Lemma 4.1 and Theorem 4.2 can be relaxed slightly, by introducing a simple randomization of the input matrix.

The next lemma is a slight generalization of a lemma given by Wiedemann [15].

Lemma 5.1. *Suppose $A \in \mathsf{F}^{n\times n}$ is symmetric with positive rank $r \le n$ and that there exists a permutation matrix $P \in \mathsf{F}^{n\times n}$ such that the leading $i \times i$ submatrix of P^tAP is nonsingular for $1 \le i \le r$. Let $y_1, y_2, \ldots, y_r$ be indetermi-*

Input: $A \in \mathsf{F}^{n\times n}$, symmetric; $b \in \mathsf{F}^{n\times 1}$
Output: $x \in \mathsf{F}^{n\times 1}$ such that $Ax = b$, or failure

Select nonzero elements $\alpha_1, \alpha_2, \ldots, \alpha_n$ uniformly and independently from a finite subset S of $\mathsf{F} \setminus \{0\}$

$$D_\alpha = \begin{bmatrix} \alpha_1 & & & 0 \\ & \alpha_2 & & \\ & & \ddots & \\ 0 & & & \alpha_n \end{bmatrix}$$

Apply the algorithm given in Figure 2 to try to find a vector $\tilde{x}$ such that $\tilde{A}\tilde{x} = \tilde{b}$, for $\tilde{A} = D_\alpha A D_\alpha$ and $\tilde{b} = D_\alpha b$, without using matrix multiplication to compute the entries of $\tilde{A}$ — so that the matrix-vector products $v_1 = D_\alpha v$, $v_2 = A v_1$, and $v_3 = D_\alpha v_2$ are computed in order to obtain $\tilde{A}v$ for any given vector v

if this attempt succeeds then
  $x := D_\alpha \tilde{x}$; return x
else
  return failure
end if

Figure 3: An Efficient but Limited Randomization

nates over F*, let*

$$\tilde{A} = \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_n \end{bmatrix} \cdot A \cdot \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_n \end{bmatrix},$$

and let $f = \det(zI_n - \tilde{A}) \in \mathsf{F}[y_1, y_2, \ldots, y_n, z]$ *(so that* f *is the characteristic polynomial of* $\tilde{A}$*). Then* f *is divisible by* z^{n-r}*, but not by* z^{n-r+1}*, and, if* $g = \frac{1}{z^{n-r}} f \in \mathsf{F}[y_1, y_2, \ldots, y_n, z]$*, then the discriminant of* g *with respect to* z *is a nonzero polynomial in* $\mathsf{F}[y_1, y_2, \ldots, y_n]$ *with total degree at most* $4nr - 2n$ *in the indeterminates* $y_1, y_2, \ldots, y_n$*.*

An algorithm employing this modification of A is given in Figure 3. This uses the same number of multiplications of A by a vector in the worst case as the algorithm given in Figure 2 (hence, at most one more than the standard Lanczos algorithm) and uses only $O(nr)$ additional operations over F if r is the rank of A.

Theorem 4.2, Lemma 5.1, and the Schwartz-Zippel lemma imply that this algorithm has a high probability of success if A is as described in the above lemma, the system $Ax = b$ has a solution, and the ground field is sufficiently large.

Theorem 5.2. *If* A *is as described in Lemma 5.1,* b *is in the column space of* A*, and the algorithm given in Figure 3 is used to try to solve the system* $Ax = b$*, then the algorithm succeeds with probability at least* $1 - \frac{5n^2 - n}{|S|}$*.*

6 A General Randomization

Now we are ready to consider the general case.

Lemma 6.1. *Suppose* $A \in \mathsf{F}^{m\times n}$ *has rank* r*, and let* $P \in \mathsf{F}^{n\times n}$ *be a permutation matrix such that the leftmost*

Input: $A \in \mathsf{F}^{m\times n}$, $b \in \mathsf{F}^{m\times 1}$
Output: $x \in \mathsf{F}^{n\times 1}$ such that $Ax = b$, or failure

Select nonzero elements $\beta_1, \beta_2, \ldots, \beta_m$ uniformly and independently from a finite subset S of $\mathsf{F} \setminus \{0\}$

$$D_\beta = \begin{bmatrix} \beta_1 & & & 0 \\ & \beta_2 & & \\ & & \ddots & \\ 0 & & & \beta_m \end{bmatrix}$$

Apply the algorithm given in Figure 3 to try to find a vector x such that $A^*x = b^*$, for $A^* = A^t D_\beta A$ and $b^* = A^t D_\beta b$, without using matrix multiplication to compute the entries of A^* — so that the matrix-vector products $v_1 = Av$, $v_2 = D_\beta v_1$, and $v_3 = A^t v_2$ are computed in order to obtain A^*v for any given vector v

if this attempt succeeds then
  { The next test is not needed if $m = n$ and A is nonsingular }
  if $Ax = b$ then
    return x
  else
    return failure
  end if
else
  return failure
end if

Figure 4: A General Randomization

r columns of the matrix AP are linearly independent. Let $x_1, x_2, \ldots, x_m$ be indeterminates over F, and let

$$A^* = A^t \cdot \begin{bmatrix} x_1 & & & 0 \\ & x_2 & & \\ & & \ddots & \\ 0 & & & x_m \end{bmatrix} \cdot A.$$

Then, for $1 \le i \le r$, the determinant of the $i \times i$ leading submatrix of $P^t A^* P$ is a nonzero polynomial with total degree at most i in the indeterminates $x_1, x_2, \ldots, x_m$.

Suppose now that $\beta_1, \beta_2, \ldots, \beta_m \in \mathsf{F}$ are nonzero, that $A^* = A^t D_\beta A$, for

$$D_\beta = \begin{bmatrix} \beta_1 & & & 0 \\ & \beta_2 & & \\ & & \ddots & \\ 0 & & & \beta_m \end{bmatrix},$$

and that A and A^* have the same rank; then they also have the same row spaces, there exists a matrix $B \in \mathsf{F}^{m\times n}$ such that $A = BA^*$, and A and A^* have the same right nullspaces as well. In this case, if $x \in \mathsf{F}^{n\times 1}$ and $b \in \mathsf{F}^{m\times 1}$ such that $Ax = b$, then $A^*x = b^*$ for $b^* = A^t D_\beta b \in \mathsf{F}^{n\times 1}$. Conversely, if b is in the column space of A, so that $Ax = b$ for some $x \in \mathsf{F}^{n\times 1}$, and if $A^*x^* = b^*$ for some vector $x^* \in \mathsf{F}^{n\times 1}$, then $A^*(x - x^*) = b^* - b^* = 0$, so that $A(x - x^*) = 0$ as well,

and $Ax^* = Ax = b$. That is, if A^* is defined as above and A and A^* have the same rank, then it suffices to solve the system $A^*x = b^*$ in order to find a vector $x \in \mathsf{F}^{n\times 1}$ such that $Ax = b$.

An algorithm that is based on this observation is given in Figure 4. This algorithm requires at most one multiplication of A^t by a vector and one multiplication of A by a vector, for each multiplication of A by a vector that is used by the algorithm given in Figure 3 (when the output is not checked), as well as one more multiplication of A^t by a vector in order to compute b^*, and one more multiplication of A by a vector to validate the output. Thus, it uses at most $\min(n+2, r+3)$ multiplications of A by a vector and the same number of multiplications of A^t by a vector, if $r \leq \min(n, m)$ is the rank of A.

Theorem 5.2, Lemma 6.1, and the Schwartz-Zippel lemma now imply the following.

Theorem 6.2. *Let S be a finite subset of F that does not include 0, let $A \in \mathsf{F}^{m\times n}$ have rank r, and let $b \in \mathsf{F}^{m\times 1}$. Given A and b as input, the algorithm shown in Figure 4 will either report failure or return a vector $x \in \mathsf{F}^{n\times 1}$ such that $Ax = b$. If b is in the column space of A (so that the system has a solution) then the probability of failure is at most $\frac{11n^2-n}{2|S|}$.*

The algorithm requires at most $\min(n+2, r+3)$ multiplications of A^t by a vector, at most the same number of multiplications of A by a vector, and $O(nr)$ additional operations over F.

If in the algorithm shown in Figure 4 the matrix A is given by a black box function for performing the matrix times vector product, the problem of computing the product of A^t with a vector arises. There is the transposition principle, known in the control theory literature as Tellegen's theorem [11], which states that a linear algorithm for A times a vector can be converted to one for A^t times a vector at no loss of efficiency in time. Space efficiency may be lost somewhat [3].

7 Inadequacy of Simpler Randomizations

An inspection of the algorithms given above will confirm that these attempt to solve a system $Ax = b$ in the general case by applying the standard Lanczos method to a system $A^*x^* = b^*$, where

$$A^* = D_\alpha A^t D_\beta A D_\alpha$$

and where D_α and D_β are nonsingular diagonal matrices whose diagonal entries are chosen uniformly and independently from a finite subset of the ground field.

LaMacchia and Odlyzko [7] report that a similar (but slightly simpler) randomization works well in practice. To solve a system $Ax = b$ for $A \in \mathsf{F}^{m\times n}$ and $b \in \mathsf{F}^{m\times 1}$, they form and solve the system $\widehat{A}x = \hat{b}$, where

$$\widehat{A} = A^t D_\alpha A$$

and where $\hat{b} = A^t D_\alpha b$. As they note, if nonzero values $\alpha_1, \alpha_2, \ldots, \alpha_n$ are chosen uniformly and independently from a sufficiently large finite subset S of $\mathsf{F} \setminus \{0\}$, then the probability is high that $\widehat{A}$ has the same rank as A, so that if $x \in \mathsf{F}^{n\times 1}$ such that $\widehat{A}x = \hat{b}$ then $Ax = b$ as well. Their experiments suggest that self-orthogonal vectors are not generally encountered when solving the system $\widehat{A}x = \hat{b}$, so that the standard Lanczos algorithm can be used.

However, this will not be true in all cases. In particular, suppose F has positive characteristic p, let $m = n = p$, and let

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \in \mathsf{F}^{p\times p},$$

so that the $(i,j)^{\text{th}}$ entry of A is 1 for $1 \leq i, j \leq p$. In this case,

$$\widehat{A} = A^t D_\alpha A = \begin{bmatrix} \alpha & \alpha & \cdots & \alpha \\ \alpha & \alpha & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & \alpha \end{bmatrix} = \alpha A \in \mathsf{F}^{p\times p}$$

for $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_p$. Now, if b is in the column space of A then

$$b = \begin{bmatrix} \beta \\ \beta \\ \vdots \\ \beta \end{bmatrix}$$

for some element β of F, and $b^t b = b^t \widehat{A} b = 0$ for every choice of $\alpha_1, \alpha_2, \ldots, \alpha_p$ and β. Thus it will not be possible to use LaMacchia and Odlyzko's randomization of the Lanczos method to solve the system $Ax = b$ in this case — and a different randomization (such as our own) is needed to establish results as general as Theorem 6.2.

If $m = n$ and A is symmetric, as in the above example, then it would generally be preferable to try to solve a system $A^\star x = b^\star$ (with the standard Lanczos method) where

$$A^\star = X^t A X \tag{7.1}$$

for some sparse or structured matrix $X \in \mathsf{F}^{n\times n}$ — so that one needs only to multiply A by a vector and perform a small number of additional operations over F in order to perform a matrix-vector multiplication involving $A^\star$, rather than having to multiply both A and A^t by vectors, as is the case for the randomization given in Section 6. Lemma 5.1 implies that the more efficient randomization (7.1) can be used if A has rank r and the leading $i \times i$ submatrix of $P^t A P$ is nonsingular for $1 \leq i \leq r$, for any integer r and some permutation matrix P. It would certainly be desirable to strengthen this lemma, by eliminating or weakening the extra conditions on A that it requires. However, these conditions cannot be eliminated completely, if the ground field has characteristic two: Suppose, in particular, that F has characteristic two, n is even, that

$$A = \begin{bmatrix} 0 & 1 & & & & & 0 \\ 1 & 0 & & & & & \\ & & 0 & 1 & & & \\ & & 1 & 0 & & & \\ & & & & \ddots & & \\ & & & & & 0 & 1 \\ 0 & & & & & 1 & 0 \end{bmatrix} \in \mathsf{F}^{n\times n},$$

and observe that $z^t Az = 0$ for every vector $z \in \mathsf{F}^{n\times 1}$. Then, $y^t A^\star y = y^t (X^t AX)y = 0$ for every matrix $X \in \mathsf{F}^{n\times n}$ and every vector $y \in \mathsf{F}^{n\times 1}$ as well, so that the standard Krylov method cannot be applied to solve *any* system $A^\star x = b^\star$ successfully, for A as above, $A^\star$ as in (7.1), and for nonzero $b^\star$.

At present, we do not know whether symmetric randomizations of the form given in (7.1) can be used reliably to solve systems over fields of larger characteristic.

8 A Comparison with the Wiedemann Approach

Suppose now that A is nonsingular, and that Wiedemann's original algorithm is applied to try to solve the system $Ax = b$. The algorithm begins with a random selection of a row vector $u \in \mathsf{F}^{1\times n}$ and computation of the minimal polynomial of the sequence

$$ub, uAb, uA^2b, \ldots \tag{8.1}$$

The algorithm succeeds, using a single row vector u, if this is the same as the minimal polynomial of the sequence of vectors

$$b, Ab, A^2b, \ldots$$

Now, suppose the dimension of the Krylov space for A and b is s (as in previous sections); in the nonsingular case, the minimal polynomial of the sequence shown in (8.1) is the same as the minimal polynomial of the sequence

$$uAb, uA^2b, uA^3b, \ldots$$

and this is the same as the minimal polynomial of the above sequence of vectors if and only if the Hankel matrix

$$H(u,b) = \begin{bmatrix} uAb & uA^2b & \cdots & uA^sb \\ uA^2b & uA^3b & \cdots & uA^{s+1}b \\ \vdots & \vdots & \ddots & \vdots \\ uA^sb & uA^{s+1}b & \cdots & uA^{2s-1}b \end{bmatrix}$$

is nonsingular (see [15], [6], and [5]). This minimal polynomial must be computed and, if matrix multiplication is not to be used, so that the matrix powers $A^2, A^3, \ldots$ are unavailable, then the number of matrix-vector products (involving A or A^t) that are used to determine this differs by a small constant from the number of matrix-vector products needed to solve this system with the algorithm in Figure 4.

However, it is either necessary to recompute the products

$$b, Ab, A^2b, \ldots \tag{8.2}$$

or to store them in order to complete this application of Wiedemann's algorithm. Thus, it appears that one must either use approximately s more matrix-vector products than one needs with the randomized Lanczos algorithm, or one must use quadratic storage space, if Wiedemann's algorithm is to be used.

If A is symmetric and P^tAP has generic rank profile for some permutation P, then the algorithm given in Figure 3 can be used to solve the system $Ax = b$ instead of the algorithm in Figure 4, reducing by approximately s the number of matrix-vector multiplications needed to solve the system. While this would seem to increase the advantage of Lanczos' method over Wiedemann's algorithm in this case, the comparison is unfair — because the same improvement can be made for Wiedemann's algorithm as well. In particular, we can "randomize the right side" vector b when Wiedemann's algorithm is used, as was done here in Section 4. Now, if we choose the row vector u deterministically as $u = b^t$ instead of choosing it randomly, then the above matrix $H(u,b)$ equals the matrix $H(A,b)$ given in equation (3.1). Since A is symmetric, this matrix (and its minimal polynomial) can be computed using approximately s fewer matrix-vector multiplications than are required in the general case. Moreover, since the application of the Lanczos method only succeeds if $H(A,b)$ has generic rank profile, the probability that this modified version of Wiedemann's algorithm succeeds is at least as high as the probability that the Lanczos method does. Thus, the apparent advantage of the Lanczos method over Wiedemann's method can be overcome, leaving the recomputation of the Krylov vectors (8.2) as the remaining disadvantage for Wiedemann's approach.

Of course, Wiedemann's method has one advantage over the Lanczos method: It is provably efficient and reliable for sparse matrix computations over small fields. To the best of our knowledge, Wiedemann's algorithm, the new algorithm of Lambert [8], and Villard's version of the block Wiedemann algorithm [14], are at this time the only algorithms for which this is the case.

Comparison with the block Wiedemann method [2, 4, 14] further complicates the situation. The first goal of blocking is to parallelize the algorithm. In addition, Kaltofen [4, Corollary to Theorem 7] shows that it is possible to solve an $n \times n$ nonsingular linear system with no more than $(1+\epsilon)n + O(1)$ sequential matrix times vector products, where ϵ is an arbitrarily small positive constant, and simultaneously $O(n^{2+o(1)})$ additional arithmetic operations and $O(n)$ auxiliary storage, where the constants implied by the big-O estimates grow as ϵ approaches 0. Hence for coefficient matrices that have a relatively expensive matrix times vector product, blocking can be used to optimize the overall sequential cost and possibly out-perform the non-blocked Wiedemann and Lanczos algorithms. Coppersmith's heuristics and Villard's analysis suggest that blocking is a means to diminish failure probabilities in the small field case. Whether any of these two advantages (fewer matrix times vector products, smaller failure probability) of the block Wiedemann algorithm in its sequential setting carry over to the block Lanczos approach [1, 10] is unknown to us.

References

[1] Coppersmith, D. Solving linear equations over GF(2); block Lanczos algorithm. *Linear Algebra Appl. 192* (1993), 33–60.

[2] Coppersmith, D. Solving homogeneous linear equations over GF(2) via block Wiedemann algorithm. *Math. Comput. 62*, 205 (1994), 333–350.

[3] Griewank, A. Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation. *Optimization Methods & Software 1* (1992), 35–54.

[4] Kaltofen, E. Analysis of Coppersmith's block Wiedemann algorithm for the parallel solution of sparse linear systems. *Math. Comp. 64* (1995), 777–806.

[5] Kaltofen, E., and Pan, V. Processor efficient parallel solution of linear systems over an abstract field. In

Proc. 3rd Ann. ACM Symp. Parallel Algor. Architectures (1991), ACM Press, pp. 180–191.

[6] Kaltofen, E., and Saunders, B. D. On Wiedemann's method of solving sparse linear systems. In *Proc. 9th Applied Algebra, Algebraic Algor., Error-Correcting Codes* (Berlin and New York, 1991), vol. 539 of *Lecture Notes in Comput. Sci.*, Springer-Verlag, pp. 29–38.

[7] LaMacchia, B. A., and Odlyzko, A. M. Solving large sparse linear systems over finite fields. In *Advances in Cryptology — CRYPTO '90* (Berlin and New York, 1990), vol. 537 of *Lecture Notes in Comput. Sci.*, Springer-Verlag, pp. 109–133.

[8] Lambert, R. *Computational Aspects of Discrete Logarithms.* PhD thesis, University of Waterloo, 1996.

[9] Lanczos, C. Solution of systems of linear equations by minimized iterations. *J. Res. Nat. Bureau of Standards 49* (1952), 33–53.

[10] Montgomery, P. L. A block Lanczos algorithm for finding dependencies over GF(2). In *Proc. EUROCRYPT '95* (Heidelberg, Germany, 1995), vol. 921 of *Springer Lecture Notes Comput. Sci.*, Springer Verlag, pp. 106–120.

[11] Penfield Jr., P., Spencer, R., and Duinker, S. *Tellegen's Theorem and Electrical Networks.* M.I.T. Press, Cambridge, MA, 1970.

[12] Schwartz, J. T. Fast probabilistic algorithms for verification of polynomial identities. *J. Assoc. Comput. Mach. 27* (1980), 701–717.

[13] Teitelbaum, J. Euclid's algorithm and the Lanczos method over finite fields. Manuscript, January 1996.

[14] Villard, G. Further analysis of Coppersmith's block Wiedemann algorithm for the solution of sparse linear systems. In *Proc. ISSAC '97* (1997).

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

[16] Zippel, R. Probabilistic algorithms for sparse polynomials. In *EUROSAM '79* (Berlin and New York, 1979), vol. 72 of *Lecture Notes in Comput. Sci.*, Springer-Verlag, pp. 216–226.

A Proofs of Results in the Body of the Paper

These appendices would not appear in conference proceedings if this paper were accepted, but are provided as an aid to reviewers.

A.1 Proofs of Results in Section 2

Proof of Lemma 2.1. If $0 \le i < s$ and $0 \le j \le i-2$, then

$$\begin{aligned}
\langle A\omega_i,\, \omega_j\rangle &= \langle \omega_i,\, A\omega_j\rangle && \text{by equation (2.2)}\\
&= \left\langle \omega_i,\, \omega_{j+1} + \sum_{h=0}^{j} \frac{\langle A\omega_j,\, \omega_h\rangle}{\langle \omega_h,\, \omega_h\rangle}\omega_h \right\rangle && \text{by equation (2.6)}\\
&= \langle \omega_i,\, \omega_{j+1}\rangle + \sum_{h=0}^{j} \frac{\langle A\omega_j,\, \omega_h\rangle}{\langle \omega_h,\, \omega_h\rangle}\langle \omega_i,\, \omega_h\rangle && \text{by linearity}\\
&= 0
\end{aligned}$$

by equation (2.3), since $j+1 < i$. □

A.2 Proof of Results in Section 3

The next lemma will be used to prove Lemma 3.2.

Lemma A.1. *Suppose $A \in \mathsf{F}^{n\times n}$ is symmetric with rank $r \le n$, and let $b \in \mathsf{F}^{n\times 1}$.*

1. *If the minimal polynomial $h \in \mathsf{F}[z]$ of the sequence $b, Ab, A^2b, \ldots$ is not divisible by z, then b is in the column space of A, and there exists a unique vector x in the Krylov space K such that $Ax = b$.*
2. *If the minimal polynomial $h \in \mathsf{F}[z]$ of the sequence $b, Ab, A^2b, \ldots$ is divisible by z, then there exists a nonzero element c of the Krylov space K such that $Ac = 0$.*

Proof. Let $h = z^s + h_{s-1}z^{s-1} + \cdots + h_1 z + h_0$ be the minimal polynomial of the above sequence, for $s \ge 0$ and for $h_{s-1}, \ldots, h_1, h_0 \in \mathsf{F}$.

If $s = 0$ then $h = 1$ (so that h is not divisible by z), $b = 0$, and, taking $x = 0 \in K$, both parts 1 and 2 of the lemma are trivially satisfied. Thus we will assume that $s > 0$ and, therefore, that $b \ne 0$.

If h is not divisible by z then $h_0 \ne 0$; set

$$x = -\frac{1}{h_0}\left(A^{s-1}b + h_{s-1}A^{s-2}b + \cdots + h_2 Ab + h_1 b\right) \in K;$$

then $Ax - b = -\frac{1}{h_0}h(A)b = 0$, so that the Krylov space does include a solution (x) for the system of linear equations $Ax = b$.

Now let $g_1, g_2 \in \mathsf{F}[z]$ and let $x_1 = g_1(A)b \in K$ and $x_2 = g_2(A)b \in K$ such that $Ax_1 = b = Ax_2$. Then clearly $A(g_1(A) - g_2(A))b = 0$, so that the polynomial $z(g_1 - g_2)$ is divisible by the minimal polynomial h of the sequence $b, Ab, A^2b, \ldots$. Since z does not divide h and is irreducible in the unique factorization domain $\mathsf{F}[z]$, z and h are relatively prime, and h must divide $g_1 - g_2$, so that $(g_1(A) - g_2(A))b = 0$ and $x_1 = x_2$. Thus there is only element x in K such that $Ax = b$, and part 1 of the lemma is correct.

Now suppose instead that z divides h. Then $\hat{h} = \frac{1}{z}h \in \mathsf{F}[z]$, and $c = \hat{h}(A)b$ belongs to K. Clearly, $\hat{h}$ does not divide h, so $c \ne 0$. Finally, $Ac = h(A)b = 0$, establishing part 2. □

Proof of Lemma 3.2. Recall that the standard Lanczos algorithm attempts to form linearly independent vectors $\omega_0, \omega_1, \ldots, \omega_{s-1}$ in the Krylov space K such that

$$\omega_i^t A\omega_i \ne 0 = \omega_i^t A\omega_j$$

for $0 \leq i < s$ and for $0 \leq j < s$ such that $j \neq i$. It is clear, by inspection of the algorithm, that if these vectors exist then there exist elements $\gamma_{i,j}$ of F, for $0 \leq j < i$, such that

$$\omega_i = A^i b + \gamma_{i,i-1} A^{i-1} b + \gamma_{i,i-2} A^{i-2} b + \cdots + \gamma_{i,0} b \quad \text{(A.1)}$$

for $0 \leq i < s$.

We now show that the above values $\omega_0, \omega_1, \ldots, \omega_{s-1}$ can be found (and that $\langle \omega_i, \omega_i \rangle$ is nonzero for all i) if and only if the Hankel matrix $H(A, b)$ has generic rank profile. For $0 \leq i < s$, let $W_i \in \mathsf{F}^{n \times (i+1)}$ be the matrix with columns $\omega_0, \omega_1, \ldots, \omega_i$ and let $B_i \in \mathsf{F}^{n \times (i+1)}$ be the matrix with columns $b, Ab, A^2b, \ldots, A^ib$ (in order from left to right in each case). Then

$$W_i = B_i X_i$$

for the upper triangular matrix

$$X_i = \begin{bmatrix} 1 & \gamma_{1,0} & \gamma_{2,0} & \cdots & \gamma_{i,0} \\ & 1 & \gamma_{2,1} & \cdots & \gamma_{i,1} \\ & & 1 & \cdots & \gamma_{i,2} \\ & & & \ddots & \vdots \\ 0 & & & & 1 \end{bmatrix} \in \mathsf{F}^{(i+1)\times(i+1)}$$

with ones on the diagonal and with $(j,k)^{\text{th}}$ entry $\gamma_{k-1,j-1}$ for $1 \leq j < k \leq i+1$. It is clear that X_i is nonsingular (and has determinant 1).

Let $t_i = \langle \omega_i, \omega_i \rangle = \omega_i^t A \omega_i$, for $0 \leq i < s$. If the Lanczos algorithm succeeds then $t_i \neq 0$ for all such i. Therefore, by orthogonality of the ω_i's (see equation (2.3)),

$$W_i^t A W_i = \begin{bmatrix} t_0 & & & 0 \\ & t_1 & & \\ & & \ddots & \\ 0 & & & t_i \end{bmatrix}$$

is a nonsingular diagonal matrix of order $i+1$. Since X_i is nonsingular, $(X_i^{-1})^t (W_i^t A W_i) X_i^{-1}$ exists and is nonsingular. However,

$$\begin{aligned} &(X_i^{-1})^t (W_i^t A W_i) X_i^{-1} \\ &= (W_i X_i^{-1})^t A (W_i X_i^{-1}) = B_i^t A B_i \\ &= \begin{bmatrix} b^tAb & b^tA^2b & \cdots & b^tA^ib \\ b^tA^2b & b^tA^3b & \cdots & b^tA^{i+1}b \\ \vdots & \vdots & \ddots & \vdots \\ b^tA^ib & b^tA^{i+1}b & \cdots & b^tA^{2i-1}b \end{bmatrix} \end{aligned}$$

is the leading $(i+1)\times(i+1)$ minor of $H(A, b)$. Thus $H(A, b)$ has generic rank profile if the standard Lanczos method can be used to construct a basis $\omega_0, \omega_1, \ldots, \omega_{s-1}$ satisfying Equations (2.3) and (2.4).

The converse is also true and can be established by using induction on i to show that if the leading $j \times j$ submatrix of $H(A, b)$ is nonsingular, for all j between 1 and i, then $\omega_0, \omega_1, \ldots, \omega_{i-1}$ exist and none of these is self-orthogonal. Indeed, if $\omega_0 = b$ as usual, then the top left entry of $H(A, b)$ is $\langle \omega_0, \omega_0 \rangle$ so that ω_0 is not self-orthogonal if the leading 1×1 minor of $H(A, b)$ is nonsingular. It follows by the relationship between $W_i^t A W_i$ and the leading $(i+1)\times(i+1)$ minor that if ω_j is not self-orthogonal for $0 \leq j < i$, then ω_i is also not self-orthogonal if the leading $(i+1) \times (i+1)$ minor is nonsingular.

Now, an inspection of the standard Lanczos algorithm establishes that it always returns failure (without finding a proposed solution x) if $H(A, b)$ does not have generic rank profile. On the other hand, if $H(A, b)$ does have generic rank profile than a "proposed solution" x is always found. This is returned if and only if it is verified that $Ax = b$ (and failure is reported otherwise). Thus, in order to complete the proof, it suffices to show that if a proposed solution is found at all, when b is in the column space of A, then $Ax = b$.

Suppose, then, that a proposed solution x is returned by the algorithm. Then $H(A, b)$ has generic rank profile, there exists a basis $\omega_0, \omega_1, \ldots, \omega_{s-1}$ for the Krylov space K satisfying Equations (2.3) and (2.4) (and, such a basis has been found by the algorithm before finding x). For $0 \leq i < s$, let $t_i = \langle \omega_i, \omega_i \rangle$; then $t_i \neq 0$. If $W \in \mathsf{F}^{n \times s}$ is the matrix with columns $\omega_0, \omega_1, \ldots, \omega_{s-1}$ then $W = W_{s-1}$ and, as noted above,

$$W^t A W = \begin{bmatrix} t_0 & & & 0 \\ & t_1 & & \\ & & \ddots & \\ 0 & & & t_{s-1} \end{bmatrix}$$

is a nonsingular matrix in $\mathsf{F}^{s \times s}$.

Now consider the minimal polynomial $h \in \mathsf{F}[z]$ for the sequence $b, Ab, A^2b, \ldots$. If $h(0) = 0$ then, by part 2 of Lemma A.1, there exists a nonzero member c of the Krylov space K such that $Ac = 0$. Since $\omega_0, \omega_1, \ldots, \omega_{s-1}$ is a basis for K there would exist elements $\gamma_0, \gamma_1, \ldots, \gamma_{s-1}$ of F, not all zero, such that

$$c = \gamma_0 \omega_0 + \gamma_1 \omega_1 + \cdots + \gamma_{s-1} \omega_{s-1}.$$

However, this could imply that

$$(W^t A W) \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{s-1} \end{bmatrix} = (W^t A) \left(W \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{s-1} \end{bmatrix} \right) = W^t A c = 0,$$

contradicting the fact that $W^t A W$ is nonsingular. Thus, $h(0) \neq 0$.

Now, part 1 of Lemma A.1 also implies that K includes a (unique) vector $\hat{x}$ such that $A\hat{x} = b$. For $0 \leq i < s$,

$$\langle \omega_i, \hat{x} \rangle = \omega_i^t A \hat{x} = \omega_i^t b.$$

It is clear by inspection of the standard Lanczos algorithm that if x is the proposed solution found by the algorithm, then

$$x = \sum_{j=0}^{s-1} \frac{b^t \omega_j}{\langle \omega_j, \omega_j \rangle} \omega_j \in K.$$

Equation (2.3) therefore implies that

$$\langle \omega_i, x \rangle = \left\langle \omega_i, \frac{b^t \omega_i}{\langle \omega_i, \omega_i \rangle} \omega_i \right\rangle = \omega_i^t b = \langle \omega_i, \hat{x} \rangle,$$

for $0 \leq i < s$. Since x and $\hat{x}$ both belong to K, and the vectors $\omega_0, \omega_1, \ldots, \omega_{s-1}$ are not self-orthogonal, it follows that $x = \hat{x}$, so that $Ax = b$ as required.

□

A.3 Proofs of Results in Section 4

The first lemma, given below, will be needed to prove Lemma 4.1 in the nonsingular case.

Lemma A.2. *Suppose $A \in \mathsf{F}^{n\times n}$ is nonsingular, symmetric, and has a squarefree characteristic polynomial. Then there exists an extension field E of F and a nonsingular matrix $Y \in \mathsf{E}^{n\times n}$ such that $Y^t = Y^{-1}$ and such that Y^tAY is a diagonal matrix.*

Proof of Lemma A.2. Let $f \in \mathsf{F}[x]$ be the characteristic polynomial and let $\widehat{\mathsf{E}} \supseteq \mathsf{F}$ be a splitting field for f. Since A is nonsingular and f is squarefree, $\widehat{\mathsf{E}}$ contains distinct nonzero elements $\zeta_1, \zeta_2, \ldots, \zeta_n$ such that

$$f = \prod_{j=1}^{n} (x - \zeta_j) \in \widehat{\mathsf{E}}[x]. \tag{A.2}$$

For $1 \le i \le n$, the matrix $A - \zeta_i I_n$ is clearly singular in $\widehat{\mathsf{E}}^{n\times n}$, so there exists a nonzero vector $\hat{x}_i \in \widehat{\mathsf{E}}^{n\times 1}$ such that $(A - \zeta_i I_n)\hat{x}_i = 0$, and so that $\hat{x}_i$ is an eigenvector for the matrix A and eigenvalue ζ_i.

The vectors $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$ are linearly independent; for, otherwise (since $\hat{x}_1$ is nonzero) there must exist an integer i between two and n such that $\hat{x}_i$ is a linear combination of $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{i-1}$, so that

$$\hat{x}_i = \lambda_1\hat{x}_1 + \lambda_2\hat{x}_2 + \cdots + \lambda_{i-1}\hat{x}_{i-1}$$

for elements $\lambda_1, \lambda_2, \ldots, \lambda_{i-1}$ of $\widehat{\mathsf{E}}$, and such that the vectors $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{i-1}$ are themselves linearly independent. However this would imply that

$$\begin{aligned}(\lambda_1\zeta_i)\hat{x}_1 + (\lambda_2\zeta_i)\hat{x}_2 + \cdots + (\lambda_{i-1}\zeta_i)\hat{x}_{i-1} \\ &= \zeta_i\hat{x}_i = A\hat{x}_i \\ &= A(\lambda_1\hat{x}_1 + \lambda_2\hat{x}_2 + \cdots + \lambda_{i-1}\hat{x}_{i-1}) \\ &= (\lambda_1\zeta_1)\hat{x}_1 + (\lambda_2\zeta_2)\hat{x}_2 + \cdots + (\lambda_{i-1}\zeta_{i-1})\hat{x}_{i-1}.\end{aligned}$$

Since $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{i-1}$ are linearly independent, it would follow that the coefficient of $\hat{x}_j$ in the first and last of these expressions are equal, so that $\lambda_j(\zeta_i - \zeta_j) = 0$. Since $\zeta_1, \zeta_2, \ldots, \zeta_i$ are distinct, this would imply that $\lambda_1 = \lambda_2 = \cdots = \lambda_{i-1} = 0$, contradicting the fact that $\hat{x}_i$ is a nonzero vector.

It is clear that $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$ is a basis for $\widehat{\mathsf{E}}^{n\times 1}$. The elements of this basis are pairwise orthogonal with respect to the standard inner product; for, otherwise, there would exist i and j such that $1 \le i, j \le n$, $i \ne j$, and such that

$$\begin{aligned}0 \ne \hat{x}_i^t\hat{x}_j &= (A^{-1}\hat{x}_i)^t(A\hat{x}_j) \\ &= (\zeta_i^{-1}\hat{x}_i)^t(\zeta_j\hat{x}_j) \\ &= (\zeta_j/\zeta_i)\hat{x}_i^t\hat{x}_j,\end{aligned}$$

contradicting the fact that the eigenvalues ζ_i and ζ_j of A are distinct.

Let $\tau_i = \hat{x}_i^t\hat{x}_i$ for $1 \le i \le n$; then $\tau_i \ne 0$ for all i (otherwise, since $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$ is a basis for $\mathsf{E}^{n\times 1}$ and $\hat{x}_i^t\hat{x}_j = 0$ for $j \ne i$, $\hat{x}_i$ would be orthogonal to all of $\mathsf{E}^{n\times 1}$). Let E be an extension of $\widehat{\mathsf{E}}$ in which each of the polynomials $z^2 - \tau_1, z^2 - \tau_2, \ldots, z^2 - \tau_n$ (in the indeterminate z) has a root. In particular, let $\sigma_i \in \mathsf{E}$ for $1 \le i \le n$ such that $\sigma_i^2 = \tau_i$. Clearly σ_i is nonzero, since τ_i is; let $Y \in \mathsf{E}^{n\times n}$ be the matrix with columns $\sigma_1^{-1}\hat{x}_1, \sigma_2^{-1}\hat{x}_2, \ldots, \sigma_n^{-1}\hat{x}_n$ (from left to right). Then it is easily checked that $Y^tY = I_n$ and (by considering the action of Y^tAY on each of the unit vectors) that

$$Y^tAY = \begin{bmatrix}\zeta_1 & & & 0 \\ & \zeta_2 & & \\ & & \ddots & \\ 0 & & & \zeta_n\end{bmatrix}$$

is a nonsingular diagonal matrix in $\mathsf{E}^{n\times n}$, as desired. □

Lemma A.3. *Suppose $A \in \mathsf{F}^{n\times n}$ is nonsingular and symmetric, and that $b \in \mathsf{F}^{n\times 1}$. Let $y_1, y_2, \ldots, y_n$ be indeterminates over F and let*

$$\hat{b} = b + A\begin{bmatrix}y_1 \\ y_2 \\ \vdots \\ y_n\end{bmatrix} \in \mathsf{F}[y_1, y_2, \ldots, y_n]^{n\times 1}.$$

If the characteristic polynomial of A is squarefree (that is, it splits into distinct linear factors over an extension of the ground field) then $\det H_i(A, \hat{b})$ is a nonzero polynomial in $\mathsf{F}[y_1, y_2, \ldots, y_n]$ with total degree at most $2i$ in the indeterminates $y_1, y_2, \ldots, y_n$, for $1 \le i \le n$.

Proof. The stated degree bound is easily proved, since each entry of $H_i(A, \hat{b})$ has total degree at most two in the indeterminates $y_1, y_2, \ldots, y_n$, and since this matrix has order i. Thus it is sufficient to show that $\det H_i(A, \hat{b})$ is a nonzero polynomial in $y_1, y_2, \ldots, y_n$ for all $i \le n$. Let

$$h_i = \det H_i(A, \hat{b}) \in \mathsf{F}[y_1, y_2, \ldots, y_n];$$

it will suffice to find an extension field E of F and elements $\gamma_1, \gamma_2, \ldots, \gamma_n$ of E, such that $h_i(\gamma_1, \gamma_2, \ldots, \gamma_n) \ne 0$ in E.

By Lemma A.2, there exists an extension $\widetilde{\mathsf{E}}$ of F and a nonsingular matrix $Y \in \widetilde{\mathsf{E}}^{n\times n}$ such that $Y^{-1} = Y^t$ and such that

$$Y^tAY = \begin{bmatrix}\zeta_1 & & & 0 \\ & \zeta_2 & & \\ & & \ddots & \\ 0 & & & \zeta_n\end{bmatrix}$$

is a nonzero diagonal matrix in $\widetilde{\mathsf{E}}^{n\times n}$. The elements $\zeta_1, \zeta_2, \ldots, \zeta_n$ are nonzero and distinct, since A is similar to the above diagonal matrix (in $\widetilde{\mathsf{E}}^{n\times n}$), is nonsingular, and has a characteristic polynomial that is squarefree.

Since Y is nonsingular, there exists a (unique) vector

$$\vec{\mu} = \begin{bmatrix}\mu_1 \\ \mu_2 \\ \vdots \\ \mu_n\end{bmatrix} \in \widetilde{\mathsf{E}}^{n\times 1}$$

for every vector

$$\vec{\lambda} = \begin{bmatrix}\lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n\end{bmatrix} \in \widetilde{\mathsf{E}}^{n\times 1}$$

such that $Y\vec{\mu} = \vec{\lambda}$. For $1 \le j \le n$,

$$\begin{aligned}\vec{\lambda}^t A^j \vec{\lambda} &= \vec{\mu}^t Y^t A^j Y \vec{\mu} \\ &= \vec{\mu}^t (Y^t A Y)^j \vec{\mu} \\ &= \zeta_1^j \mu_1^2 + \zeta_2^j \mu_2^2 + \cdots + \zeta_n^j \mu_n^2.\end{aligned}$$

This is the j^{th} entry of the vector

$$V \cdot \begin{bmatrix} \zeta_1 & & & 0 \\ & \zeta_2 & & \\ & & \ddots & \\ 0 & & & \zeta_n \end{bmatrix} \cdot \begin{bmatrix} \mu_1^2 \\ \mu_2^2 \\ \vdots \\ \mu_n^2 \end{bmatrix},$$

for

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \zeta_1 & \zeta_2 & \cdots & \zeta_n \\ \zeta_1^2 & \zeta_2^2 & \cdots & \zeta_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_1^{n-1} & \zeta_2^{n-1} & \cdots & \zeta_n^{n-1} \end{bmatrix} \in \widetilde{\mathsf{E}}^{n \times n},$$

where V is a nonsingular Vandermonde matrix of order n, since $\zeta_1, \zeta_2, \ldots, \zeta_n$ are distinct.

For $1 \le i \le n$, let $e_i \in \widetilde{\mathsf{E}}^{n \times 1}$ be the vector with i^{th} entry one and with all other entries equal to zero. Since V is nonsingular there exist elements $\eta_1, \eta_2, \ldots \eta_n$ of $\widetilde{\mathsf{E}}$ such that

$$V \cdot \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} = e_i.$$

Now let $\mathsf{E} \supseteq \widetilde{\mathsf{E}}$ be a field extension of $\widetilde{\mathsf{E}}$ (and hence of F) which includes elements $\phi_1, \phi_2, \ldots, \phi_n$ such that $\phi_j^2 = \eta_j / \zeta_j$ for $1 \le j \le n$, and let $\mu_1, \mu_2, \ldots, \mu_n \in \mathsf{E}$ such that

$$Y \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}.$$

Then, for $1 \le j \le n$,

$$\begin{aligned}\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}^t A^j \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} &= \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix}^t (Y^t A Y)^j \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix} \\ &= \phi_1^2 \zeta_1^j + \phi_2^2 \zeta_2^j + \cdots + \phi_n^2 \zeta_n^j \\ &= \eta_1 \zeta_1^{j-1} + \eta_2 \zeta_2^{j-1} + \cdots + \eta_n \zeta_n^{j-1} \\ &= \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{if } j \ne i. \end{cases}\end{aligned}$$

Finally, let $\gamma_1, \gamma_2, \ldots, \gamma_n \in \mathsf{E}$ such that $b + A\vec{\gamma} = \vec{\mu}$, for

$$\vec{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} \quad \text{and} \quad \vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix};$$

such elements $\gamma_1, \gamma_2, \ldots, \gamma_n$ do exist, since A is nonsingular. Then the matrix $H_i(A, b + A\vec{\gamma})$ is a Hankel matrix with ones on the antidiagonal and zeroes above it — so, clearly, $h_i(\gamma_1, \gamma_2, \ldots, \gamma_n) = \det H_i(A, \vec{\mu}) = \pm 1 \ne 0$, as required. □

The next two lemmas will be used to establish Lemma 4.1 in the singular case as well.

Lemma A.4. *Suppose $A \in \mathsf{F}^{n \times n}$ is symmetric with rank $r \le n$ and that the minimal polynomial of A (in $\mathsf{F}[z]$) is not divisible by z^2. Then, if $b \in \mathsf{F}^{n \times 1}$ is in the column space of A, so that the system of equations $Ax = b$ has a solution, then the minimal polynomial of the sequence $b, Ab, A^2b, \ldots$ is not divisible by z.*

Proof. This result is trivial if A is nonsingular, so we will assume that $r < n$. Then the minimal polynomial of A is $z\hat{h}$ for some polynomial $\hat{h} \in \mathsf{F}[z]$ such that $\hat{h}(0) \ne 0$. Let

$$V_0 = \{Au \mid u \in \mathsf{F}^{n \times 1}\} \subseteq \mathsf{F}^{n \times 1}$$

and

$$V_1 = \{\hat{h}(A)u \mid u \in \mathsf{F}^{n \times 1}\} \subseteq \mathsf{F}^{n \times 1}.$$

Clearly $Au \in V_0$ whenever $u \in V_0$ and $Au \in V_1$ whenever $u \in V_1$. Let $m_i \in \mathbb{Z}$ be the dimension of V_i and let $x_{i,1}, x_{i,2}, \ldots, x_{i,m_i}$ be a basis for V_i, for $i = 0$ and $i = 1$. Since z and $\hat{h}$ are relatively prime in $\mathsf{F}[z]$, there exist polynomials $f, g \in \mathsf{F}[z]$ such that $zf + \hat{h}g = 1$. Thus, if $u_0 = Af(A)u$ and $u_1 = \hat{h}(A)g(A)u$ for $u \in \mathsf{F}^{n \times 1}$ then $u_0 \in V_0$, $u_1 \in V_1$, and $u = u_1 + u_2$, so that the vectors

$$x_{0,1}, x_{0,2}, \ldots, x_{0,m_0}, x_{1,1}, x_{1,2}, \ldots, x_{1,m_1} \tag{A.3}$$

span $\mathsf{F}^{n \times 1}$. Furthermore, $V_0 \cap V_1 = \{0\}$: For, if $v \in V_0 \cap V_1$ then $Av = 0$ since $v \in V_1$, $\hat{h}(A)v = 0$, since $v \in V_0$, and thus $v = (f(A)A + g(A)\hat{h}(A))v = 0$ as well. Therefore the vectors shown in (A.3) are also linearly independent and form a basis for $\mathsf{F}^{n \times 1}$. Since V_0 is the image of A, $m_0 = r$ and $m_1 = n - r$. Let $X \in \mathsf{F}^{n \times n}$ be the nonsingular matrix whose columns are the vectors in (A.3) in order from left to right. Since $Au_0 \in V_0$ for all $u_0 \in V_0$ and $Au_1 = 0$ for all $u_1 \in V_1$,

$$X^{-1}AX = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} \tag{A.4}$$

for some matrix $B \in \mathsf{F}^{r \times r}$. Since A has rank r, B has rank r as well, so B is nonsingular.

Let $b \in \mathsf{F}^{n \times 1}$ be in the column space of A, so that there exists a vector $x \in \mathsf{F}^{n \times 1}$ such that $Ax = b$. Then, since $X^{-1}Ax = X^{-1}b$, there exists a vector $y = X^{-1}x \in \mathsf{F}^{n \times 1}$ such that $(X^{-1}AX)y = X^{-1}b$. Let $y_U, b_U \in \mathsf{F}^{r \times 1}$ and let $y_L, b_L \in \mathsf{F}^{(n-r) \times 1}$ such that

$$y = \begin{bmatrix} y_U \\ y_L \end{bmatrix} \quad \text{and} \quad X^{-1}b = \begin{bmatrix} b_U \\ b_L \end{bmatrix};$$

then by Equation (A.4), $By_U = b_U$, and $b_L = 0$. Since B is nonsingular, A has minimal polynomial $z\hat{h}$, and $\hat{h}(0) \ne 0$, it can be argued using Equation (A.4) that B has minimal polynomial $\hat{h}$. Now, since $b_L = 0$, $\hat{h}(X^{-1}AX)X^{-1}b = 0$. However, $\hat{h}(X^{-1}AX)X^{-1}b = X^{-1}\hat{h}(A)b$, so (since X^{-1} is nonsingular) $\hat{h}(A)b = 0$ as well. Thus the minimal polynomial of the sequence $b, Ab, A^2b, \ldots$ is not divisible by z (since it must divide $\hat{h}$). □

Lemma A.2 can now be generalized.

Lemma A.5. *Suppose $A \in \mathsf{F}^{n\times n}$ is symmetric with rank $r \le n$, and that the characteristic polynomial $f \in \mathsf{F}[z]$ of A is the product of z^{n-r} and a squarefree polynomial of degree r that is not divisible by z. Then the minimal polynomial of A is not divisible by z^2, and there exists an extension field E of F, a nonsingular matrix $Y \in \mathsf{E}^{n\times n}$, a nonsingular matrix $Z \in \mathsf{E}^{(n-r)\times(n-r)}$, and a nonsingular diagonal matrix $D \in \mathsf{E}^{r\times r}$ whose diagonal entries are distinct, such that*

$$Y^tY = \begin{bmatrix} I_r & 0 \\ 0 & Z \end{bmatrix}$$

and

$$Y^tAY = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}.$$

Proof. This is a consequence of Lemma A.2 if A is nonsingular. Therefore, we will assume that $r < n$.

Suppose the characteristic polynomial of A is $z^{n-r}g$, so that $g \in \mathsf{F}[z]$ is squarefree and $g(0) \neq 0$. Then the minimal polynomial of A is $z^{1+t}g$ for some integer $t \ge 0$, since the minimal polynomial divides the characteristic polynomial and is divisible by each irreducible factor of the characteristic polynomial. It is clear (by consideration of the Jordan normal form of A) that the rank of A is at least $\deg(g) + t = r + t$. Therefore, since A has rank r, $t = 0$ and A has minimal polynomial zg, which is not divisible by z^2, as claimed.

Since g is squarefree and not divisible by z, there exists an extension $\widehat{\mathsf{E}}$ of F and distinct nonzero elements $\zeta_1, \zeta_2, \ldots, \zeta_r$ of $\widehat{\mathsf{E}}$ such that

$$g = \prod_{j=1}^{r} (x - \zeta_j) \in \widehat{\mathsf{E}}[x].$$

This resembles, and will be used in the same way as, the factorization given in Equation (A.2) for the characteristic polynomial of (a nonsingular matrix) A, in the proof of Lemma A.2. In particular, there exist nonzero vectors $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_r \in \widehat{\mathsf{E}}^{n\times 1}$ such that $A\hat{x}_i = \zeta_i\hat{x}_i$ for $1 \le i \le r$. As argued in the proof of Lemma A.2 (for the corresponding vectors $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$, immediately following Equation (A.2)), the vectors $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_r$ are linearly independent. If $1 \le i, j \le r$ and $i \neq j$, then $\hat{x}_i^t\hat{x}_j = 0$: For A is symmetric, and

$$\begin{aligned} \zeta_i\hat{x}_i^t\hat{x}_j &= (\zeta_i\hat{x}_i)^t\hat{x}_j = (A\hat{x}_i)^t\hat{x}_j \\ &= \hat{x}_i^t(A\hat{x}_j) = \hat{x}_i^t(\zeta_j\hat{x}_j) = \zeta_j\hat{x}_i^t\hat{x}_j. \end{aligned}$$

Since A has rank r, its (column) nullspace has dimension $n-r$. Let $v_1, v_2, \ldots, v_{n-r}$ be a basis for this null space. Then, since the minimal polynomial of A is not divisible by z^2, A is similar to a block diagonal matrix with a nonsingular matrix and a matrix of zeroes as blocks (see Equation (A.4)), and the vectors

$$\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_r, v_1, v_2, \ldots, v_{n-r}$$

are linearly independent and form a basis for $\mathsf{F}^{n\times 1}$. Now, if $1 \le i \le r$ and $1 \le j \le n - r$ then

$$\hat{x}_i^tv_j = \frac{1}{\zeta_i}(A\hat{x}_i)^tv_j = \frac{1}{\zeta_i}\hat{x}_i^t(Av_j) = \frac{1}{\zeta_i}\hat{x}_i^t \cdot 0 = 0.$$

It follows that $\hat{x}_i^t\hat{x}_i \neq 0$ for $1 \le i \le r$: Otherwise, $\hat{x}_i$ would be nonzero but orthogonal to all of $\mathsf{F}^{n\times 1}$. Let $\tau_i = \hat{x}_i^t\hat{x}_i$, let E be an extension of $\widehat{\mathsf{E}}$ (and of F) containing elements $\sigma_1, \sigma_2, \ldots, \sigma_r$ such that $\sigma_i^2 = \tau_i$ for $1 \le i \le r$, and let $Y \in \mathsf{E}^{n\times n}$ be the matrix with columns $\sigma_1^{-1}\hat{x}_1, \sigma_2^{-1}\hat{x}_2, \ldots, \sigma_r^{-1}\hat{x}_r, v_1, v_2, \ldots, v_{n-r}$ in order from left to right. Then it easily checked that

$$Y^tY = \begin{bmatrix} I_r & 0 \\ 0 & Z \end{bmatrix}$$

for some matrix $Z \in \mathsf{E}^{(n-r)\times(n-r)}$. Since the columns of Y are linearly independent, Y, Y^t, and Y^tY are all nonsingular matrices, so that Z is nonsingular in $\mathsf{E}^{(n-r)\times(n-r)}$ as well. Finally, since the columns of Y are all eigenvectors of A it is also clear that

$$Y^tAY = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

for

$$D = \begin{bmatrix} \zeta_1 & & & 0 \\ & \zeta_2 & & \\ & & \ddots & \\ 0 & & & \zeta_r \end{bmatrix},$$

so that D is nonsingular, diagonal, and has distinct diagonal entries, as required. □

Proof of Lemma 4.1. Since this follows from Lemma A.3 when A is nonsingular, we will assume that the rank r of A is strictly less than n. As was the case for Lemma A.3, the degree bound claimed for the polynomial

$$h_i = \det H_i(A, \hat{b})$$

is easily verified, so it is sufficient to confirm that this polynomial is not identically zero for $1 \le i \le r$, and this can be established by showing that there is an extension E of F and elements $\gamma_1, \gamma_2, \ldots, \gamma_n$ of E such that $h_i(\gamma_1, \gamma_2, \ldots, \gamma_n)$ is not equal to zero.

By Lemma A.5, there exists an extension $\widetilde{\mathsf{E}}$ of F, a nonsingular matrix $Y \in \widetilde{\mathsf{E}}^{n\times n}$, a nonsingular matrix $Z \in \widetilde{\mathsf{E}}^{(n-r)\times(n-r)}$, and a nonsingular diagonal matrix $D \in \widetilde{\mathsf{E}}^{r\times r}$ with distinct diagonal entries, such that

$$Y^tY = \begin{bmatrix} I_r & 0 \\ 0 & Z \end{bmatrix}$$

and

$$Y^tAY = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}.$$

Clearly, then,

$$Y^{-1}AY = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

as well, and

$$Y^tA^jY = Y^{-1}A^jY = \begin{bmatrix} D^j & 0 \\ 0 & 0 \end{bmatrix} \quad \text{for } j \ge 1.$$

As in the proof of Lemma A.5, let $\zeta_1, \zeta_2, \ldots, \zeta_r$ be the (distinct) diagonal entries of D; then these are also the nonzero eigenvalues of A.

If $c \in \mathsf{F}^{n\times 1}$ is in the column space of A then, since the minimal polynomial of A is not divisible by z^2 (by

Lemma A.5), the minimal polynomial of the sequence $c, Ac, A^2c, \ldots$ is not divisible by z (by Lemma A.4) and

$$Y^{-1}c = Y^t c = \begin{bmatrix} c_U \\ c_L \end{bmatrix}$$

for $c_U \in \widetilde{\mathsf{E}}^{r\times 1}$ and $c_L = 0 \in \widetilde{\mathsf{E}}^{(n-r)\times 1}$.

Consequently,

$$(Y^t A^j Y)\cdot Y^t c = (Y^{-1}AY)\cdot Y^{-1}c = \begin{bmatrix} D^j c_U \\ 0 \end{bmatrix}$$

for $j \geq 0$.

Now, the proof proceeds in much the same way as for the nonsingular case. That is, for $1 \leq i \leq r$, we choose $\eta_1, \eta_2, \ldots, \eta_r \in \widetilde{\mathsf{E}}$ such that

$$V \cdot \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_r \end{bmatrix} = e_i,$$

where $e_i \in \widetilde{\mathsf{E}}^{n\times 1}$ has i^{th} entry 1 and j^{th} entry 0 for $1 \leq j \leq r$ and $j \neq i$, and where

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \zeta_1 & \zeta_2 & \cdots & \zeta_r \\ \zeta_1^2 & \zeta_2^2 & \cdots & \zeta_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ \zeta_1^{r-1} & \zeta_2^{r-1} & \cdots & \zeta_r^{r-1} \end{bmatrix} \in \widetilde{\mathsf{E}}^{r\times r}$$

is a nonsingular Vandermonde matrix of order r. We let $\mathsf{E} \supseteq \widetilde{\mathsf{E}}$ be a field extension of $\widetilde{\mathsf{E}}$ containing elements $\phi_1, \phi_2, \ldots, \phi_r$ such that $\phi_j^2 = \eta_n/\zeta_j$ for $1 \leq j \leq r$, and set $\mu_1, \mu_2, \ldots, \mu_n \in \mathsf{E}$ such that $Y\vec{\phi} = \vec{\mu}$, for

$$\vec{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathsf{E}^{n\times 1}$$

and

$$\vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_r \\ \mu_{r+1} \\ \mu_{r+2} \\ \vdots \\ \mu_n \end{bmatrix} \in \mathsf{E}^{n\times 1}.$$

Since D is nonsingular, $\vec{\phi}$ is in the column space of $Y^{-1}AY$, and $\vec{\mu}$ is in the column space of A. Since b is also in the column space of A there exist elements $\gamma_1, \gamma_2, \ldots, \gamma_n$ of E such that $b + A\vec{\gamma} = \vec{\mu}$, where

$$\vec{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} \in \mathsf{E}^{n\times 1}.$$

Then, as argued in the proof for the nonsingular case,

$$\vec{\mu}^t A^j \vec{\mu} = \begin{cases} 0 & \text{if } 1 \leq j \leq i, \\ 1 & \text{if } j = i+1. \end{cases}$$

Thus $H_i(A, b + A\vec{\gamma})$ is a Hankel matrix with ones on the antidiagonal and zeroes above it, and $h_i(\gamma_1, \gamma_2, \ldots, \gamma_n) = \det H_i(A, b + A\vec{\gamma}) = \pm 1 \neq 0$, as required. □

Proof of Theorem 4.2. By Lemma 3.2, the standard Lanczos algorithm can be applied successfully to solve the system $A\hat{x} = \hat{b}$ if and only if the Hankel matrix $H(A, \hat{b})$ has generic rank profile.

Lemma 4.1 implies that if $\gamma_1, \gamma_2, \ldots, \gamma_n$ are replaced in the above matrix by indeterminates $y_1, y_2, \ldots, y_n$ over F, respectively, then the determinant of the leading $i \times i$ submatrix of the resulting Hankel matrix is a nonzero polynomial with total degree at most $2i$ in the indeterminates $y_1, y_2, \ldots, y_n$, for $1 \leq i \leq n$. Let $h \in \mathsf{F}[y_1, y_2, \ldots, y_n]$ be the product of these polynomials; then h has total degree at most $n(n+1)$ in the indeterminates $y_1, y_2, \ldots, y_n$ and is nonzero in $\mathsf{F}[y_1, y_2, \ldots, y_n]$. Furthermore, if $\gamma_1, \gamma_2, \ldots, \gamma_n \in \mathsf{F}$ such that $h(\gamma_1, \gamma_2, \ldots, \gamma_n) \neq 0$, then the matrix $H(A, \hat{b})$ (obtained from $\gamma_1, \gamma_2, \ldots, \gamma_n$ as shown in the statement of the theorem, above) has generic rank profile.

The "Schwartz-Zippel lemma" [12], [16] now implies the claim. □

A.4 Proof of Results in Section 5

Proof of Lemma 5.1. Since A has rank r in $\mathsf{F}^{n\times n}$ it has rank r in $\mathsf{F}(y_1, y_2, \ldots, y_n)^{n\times n}$ as well, as does $\tilde{A}$. There exists a nonsingular matrix $X \in \mathsf{F}(y_1, y_2, \ldots, y_n)^{n\times n}$, whose rightmost columns form a basis for the nullspace of $\tilde{A}$, such that

$$X^{-1}\tilde{A}X = \begin{bmatrix} \tilde{A}_{1,1} & 0 \\ \tilde{A}_{2,1} & 0 \end{bmatrix}$$

for a matrix $\tilde{A}_{1,1} \in \mathsf{F}(y_1, y_2, \ldots, y_n)^{r\times r}$ and for $\tilde{A}_{1,2} \in \mathsf{F}(y_1, y_2, \ldots, y_n)^{(n-r)\times r}$. It is clear that the characteristic polynomial of $X^{-1}AX$ is the product of z^{n-r} and the characteristic polynomial of $\tilde{A}_{1,1}$. Since $\tilde{A}$ and $X^{-1}AX$ are similar, they have the same characteristic polynomial, so (since the entries of $\tilde{A}$ belong to $\mathsf{F}[y_1, y_2, \ldots, y_n]$), the characteristic polynomial of A is $z^{n-r}g$, for a polynomial $g \in \mathsf{F}[y_1, y_2, \ldots, y_n, z]$.

Let

$$g = z^r + g_{r-1}z^{r-1} + \cdots + g_1 z + g_0$$

for $g_{r-1}, \ldots, g_1, g_0 \in \mathsf{F}[y_1, y_2, \ldots, y_n]$. Then, since these are also coefficients of the characteristic polynomial of $\tilde{A}$, g_i has total degree at most $2n$ in the indeterminates $y_1, y_2, \ldots, y_n$ for $0 \leq i < r$. The discriminant of g is the determinant of a matrix, each of whose entries is either zero, one, or one of $g_0, g_1, \ldots, g_{r-1}$, and whose order is at most $2r-1$; this implies the degree bound stated in the lemma for the discriminant of g. It now remains only to argue that g is not divisible by z and that the discriminant of g is not identically zero.

We begin by considering the special case that A is nonsingular with generic rank profile, so that $r = n$ and we could set P to be the identity matrix, in the statement of the lemma. In this special case, $f = g$ and this polynomial

is not divisible by z, since A and $\tilde{A}$ are nonsingular. Wiedemann [15] shows that the discriminant of the characteristic polynomial of the matrix

$$A \cdot \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_n \end{bmatrix}$$

is not identically zero in this case, using induction on n; his argument can be applied to the above matrix $\tilde{A}$ as well. In particular, suppose that the entry of A in row i and column j is $a_{i,j}$, for $1 \le i, j \le n$. If $n = 1$ then

$$\tilde{A} = \left[a_{1,1} y^2\right],$$

$f = z - a_{1,1} y_1^2$, and the discriminant of f with respect to z is one. Suppose now that the result is correct for $(n-1) \times (n-1)$ matrices for $n > 1$. Substituting 0 for y_n in $\tilde{A}$ gives

$$A' = D_{n-1} \cdot A \cdot D_{n-1} = \begin{bmatrix} & & 0 \\ & \tilde{A}_{n-1} & \vdots \\ 0 & \cdots & 0 \end{bmatrix},$$

where

$$D_{n-1} = \begin{bmatrix} y_1 & & & & 0 \\ & y_2 & & & \\ & & \ddots & & \\ & & & y_{n-1} & \\ 0 & & & & 0 \end{bmatrix},$$

$\tilde{A}_{n-1}$ is the matrix

$$\begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_{n-1} \end{bmatrix} \cdot A_{n-1} \cdot \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_{n-1} \end{bmatrix},$$

and where A_{n-1} is the leading $(n-1) \times (n-1)$ submatrix of A. Clearly, the characteristic polynomial of A' is zg, where g is the characteristic polynomial of $\tilde{A}_{n-1}$, and (since the discriminant of zg with respect to z is the resultant of zg and $g + z(\partial g/\partial z)$) the discriminant of the characteristic polynomial of A' is the product of $(\det \tilde{A}_{n-1})^2$ and the discriminant of g. It is clear that all the leading submatrices of A_{n-1} are nonsingular, since they are also leading submatrices of A. It follows by the inductive hypothesis that g is a nonzero polynomial, as is the discriminant of the characteristic polynomial of A' with respect to z. The discriminant of the characteristic polynomial of A is therefore nonzero as well.

Next, suppose that A might be singular, but suppose again that the leading $i \times i$ submatrix of A is nonsingular for $1 \le i \le r$, so that (once again) one could choose P to be the identity matrix in the statement of the lemma. If the variables $y_{r+1}, y_{r+2}, \ldots, y_n$ are replaced with zeroes then the matrix obtained from A is

$$\tilde{A}' = \begin{bmatrix} \tilde{A}_r & 0 \\ 0 & 0 \end{bmatrix},$$

for

$$\tilde{A}_r = \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_r \end{bmatrix} \cdot A_r \cdot \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_r \end{bmatrix},$$

where A_r is the leading $i \times i$ submatrix of A.

Clearly, the characteristic polynomial of $\tilde{A}'$ is the product of z^{n-r} and the characteristic polynomial of $\tilde{A}_r$, so that the matrix $\tilde{A}_r$ has characteristic polynomial $\hat{g} = g(y_1, y_2, \ldots, y_r, 0, 0, \ldots, 0, z)$. Since A_r is nonsingular with generic rank profile (in this special case), $\hat{g}$ is not divisible by z, and it follows by the argument given above for the nonsingular case that the discriminant of $\hat{g}$ is not identically zero. This clearly implies that g is not divisible by z and has a nonzero discriminant as well, when the leading $i \times i$ submatrix of A is nonsingular and has generic rank profile.

Now, this statement is still true if the indeterminates $y_1, y_2, \ldots, y_m$ are permuted in the definition of $\tilde{A}$. Therefore if A and P are as given in the statement of the lemma then, since the leading $r \times r$ submatrix of $P^t A P$ is nonsingular and has generic rank profile, the characteristic polynomial of the matrix

$$A^\star = \left(P^t \cdot \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_n \end{bmatrix} \cdot P \right) \cdot P^t A P \cdot \left(P^t \cdot \begin{bmatrix} y_1 & & & 0 \\ & y_2 & & \\ & & \ddots & \\ 0 & & & y_n \end{bmatrix} \cdot P \right)$$

equals $z^{n-r} g$, for some polynomial $g \in \mathsf{F}[y_1, y_2, \ldots, y_n, z]$ that is not divisible by z such that the discriminant of g is nonzero. However, $A^\star = P^t A P$, so $A^\star$ is similar to A and these matrices have the same characteristic polynomial. Thus, f and g are as claimed whenever A and P are as described in the statement of the lemma. □

Proof of Theorem 5.2. An inspection of the algorithm confirms that it attempts to use the algorithm of Figure 2 to solve a linear system with coefficient matrix DAD, where D is a diagonal matrix whose diagonal entries are chosen uniformly and independently from a finite subset S of $\mathsf{F} \setminus \{0\}$.

Now, Lemma 5.1 implies that the matrix $\tilde{A}$ mentioned in that lemma has a characteristic polynomial $f = z^{n-r} g$, where $g \in \mathsf{F}[y_1, y_2, \ldots, y_n, z]$ is not divisible by z and where the discriminant of g (with respect to z) is a nonzero polynomial with total degree at most $4nr - 2n$ in $y_1, y_2, \ldots, y_n$.

Suppose $n = r$; then $f = g$. The Schwartz-Zippel lemma implies that the characteristic polynomial of DAD is squarefree (since its discriminant is nonzero) with probability at least $1 - \frac{4n^2 - 2n}{|S|}$. Since DAD is nonsingular, the constant coefficient of this polynomial is guaranteed to be nonzero — that is, it is guaranteed that this polynomial is not divisible by z.

On the other hand, if $n \neq r$ then the discriminant of the above polynomial g is a nonzero polynomial with total degree at most $4n^2 - 6n$ in $y_1, y_2, \ldots, y_n$. If we write

$$g = g_r z^r + g_{r-1} z^{r-1} + \cdots + g_1 z + g_0,$$

where $g_r, g_{r-1}, \ldots, g_1, g_0 \in \mathsf{F}[y_1, y_2, \ldots, y_n]$, then it is clear by the construction of $\hat{A}$ that g_0 is a nonzero polynomial with total degree at most $2n$ in $y_1, y_2, \ldots, y_n$. Thus, the product of g_0 and the discriminant of g is a nonzero polynomial in $\mathsf{F}[y_1, y_2, \ldots, y_n]$ with total degree at most $4n^2 - 4n < 4n^2 - 2n$ in $y_1, y_2, \ldots, y_n$, in this case.

Now, it follows by the Schwartz-Zippel lemma that the characteristic polynomial of the above matrix DAD is a polynomial, $\hat{f}$, in $\mathsf{F}[z]$, such that $\hat{f}$ is divisible by z^{n-r} but not by z^{n-r+1} and such that that $\frac{1}{z^{n-r}}\hat{f}$ is squarefree (again, since its discriminant is nonzero), again with probability at least $1 - \frac{4n^2-2n}{|S|}$.

In other words, the probability that this is *not* the case is at most $\frac{4n^2-2n}{|S|}$.

On the other hand, if this *is* the case, then Theorem 4.2 implies that the probability that the algorithm fails is at most $\frac{n^2+n}{|S|}$.

Thus, the probability that the algorithm fails (in all cases) is at most $\frac{4n^2-2n}{|S|} + \frac{n^2+n}{|S|} = \frac{5n^2-n}{|S|}$, as claimed. □

A.5 Proof of Results in Section 6

Proof of Lemma 6.1. Let $f_i \in \mathsf{F}[x_1, x_2, \ldots, x_m]$ be the determinant of the leading $i \times i$ minor of $P^t A^\star P$, for $1 \leq i \leq r$. Since each entry of $P^t A^\star P$ is an F-linear combination of the indeterminates $x_1, x_2, \ldots, x_m$, the degree bound stated for f_i is a consequence of the fact that f_i is a polynomial function of the entries of $P^t A^\star P$ with total degree i in these entries.

Set

$$D_x = \begin{bmatrix} x_1 & & & 0 \\ & x_2 & & \\ & & \ddots & \\ 0 & & & x_m \end{bmatrix},$$

so that $A^\star = A^t D_x A$ and $P^t A^\star P = (AP)^t D_x (AP)$. As well, for $I = \{1, 2, \ldots, i\}$ and for $J \subseteq \{1, 2, \ldots, m\}$, let $(AP)_{J,I}$ be the submatrix of AP including the rows with indices in J and with the first i columns, and let $(D_x)_{J,J}$ be the submatrix D_x with rows and columns whose indices are in J. Since D_x is a diagonal matrix, a formula of Cauchy-Binet can be used to show that

$$\begin{aligned} f_i &= \sum_{\substack{J=\{j_1,j_2,\ldots,j_i\}\subseteq\{1,2,\ldots,m\} \\ |J|=i}} \det(D_x)_{J,J} \cdot \left(\det(AP)_{J,I}\right)^2 \\ &= \sum_{\substack{J=\{j_1,j_2,\ldots,j_i\}\subseteq\{1,2,\ldots,m\} \\ |J|=i}} \left(\det(AP)_{J,I}\right)^2 x_{j_1} x_{j_2} \ldots x_{j_i}. \end{aligned}$$

Now, by the choice of P, the leftmost i columns of AP are linearly independent. Thus there exists a set $J \subseteq \{1, 2 \ldots, m\}$ of size i such that the $i \times i$ submatrix $(AP)_{J,I}$ is nonsingular. For $1 \leq j \leq m$, let

$$\gamma_j = \begin{cases} 1 & \text{if } j \in J, \\ 0 & \text{if } j \notin J. \end{cases}$$

Then, clearly,

$$f_i(\gamma_1, \gamma_2, \ldots, \gamma_m) = \left(\det(AP)_{I,J}\right)^2 \neq 0,$$

so f_i is nonzero as well, as claimed. □

Proof of Theorem 6.2. Since A has rank r, there exists a permutation matrix $P \in \mathsf{F}^{n \times n}$ such that the leftmost r columns of AP are linearly independent.

Now, consider the matrix $A^\star = A^t D_\beta A$ defined and used by the algorithm shown in Figure 4. It follows by Lemma 6.1 and the Schwartz-Zippel lemma that the leading $r \times r$ submatrix of $P^t A^\star P$ is nonsingular and has generic rank profile with probability at least $1 - \frac{r(r+1)}{2|S|} \geq 1 - \frac{n^2+n}{2|S|}$.

That is, this *is not* the case with probability at most $\frac{n^2+n}{2|S|}$.

On the other hand, if it *is* the case, then Theorem 5.2 implies that the probability of failure of the algorithm is at most $\frac{5n^2-n}{|S|}$.

Thus the probability of failure of the algorithm (in all cases) is at most $\frac{n^2+n}{2|S|} + \frac{5n^2-n}{|S|} = \frac{11n^2-n}{2|S|}$, as claimed. □