



MATRIX RANK CERTIFICATION*

B. DAVID SAUNDERS[†], ARNE STORJOHANN[‡], AND GILLES VILLARD[§]

Abstract. Randomized algorithms are given for computing the rank of a matrix over a field of characteristic zero with conjugation operator. The matrix is treated as a black box. Only the capability to compute matrix \times column-vector and row-vector \times matrix products is used. The methods are exact, sometimes called seminumeric. They are appropriate for example for matrices with integer or rational entries. The rank algorithms are probabilistic of the Las Vegas type; the correctness of the result is guaranteed.

Key words. Matrix rank, Minimal polynomial, Black box matrix, Seminumeric computation, Exact arithmetic, Randomized algorithms, Las Vegas algorithms.

AMS subject classifications. 15A03, 65F50, 68W30.

1. Introduction. The rank of a matrix A over a field F can be computed using an elimination method. However, this may be excessively costly in time and/or space. Iterative “black box” methods, also called “matrix-free” methods, are an alternative to using elimination.

Several randomized black box algorithms of the Monte Carlo kind—probably correct (not certified) and always fast—for computing the rank have been developed [6, 11, 16]. For A an $n \times n$ matrix of rank r they require $O(r)$ matrix-vector and vector-matrix products involving A and $O(nr)$ additional operations over F [6, Theorem 6.2]. Note that the cost of a matrix-vector product may be much less than n^2 field operations for a sparse or structured matrix. Also, the black box methods require space for only $O(n)$ additional field elements beyond the matrix storage, whereas elimination usually requires $O(n^2)$. This improvement in space complexity is an important consideration for large sparse matrices in practice. The Monte Carlo black box methods depend on random preconditioners and random vectors. In the likely event that these random choices produce preconditioners and projection vectors with the desired properties, the rank is correctly computed. However, it was not known how to incorporate detection of wrong answers into the black box approach.

The methods presented here can be used to remove the possibility of an erroneous result in the case when F is a field of characteristic zero with conjugation operator, which we assume throughout the rest of the paper. For instance, F is an ordered field or a subfield of the complex numbers.

We give two randomized black box algorithms of the Las Vegas kind where the answer is always correct and computed quickly with controllably high probability.

* Received by the editors 10 August 2001. Accepted for publication 18 January 2004. Handling Editor: Ludwig Elsner.

[†] Dept. of Computer and Information Sci., University of Delaware, Newark, DE 19716, USA (saunders@udel.edu). Generously supported by the NSF, grant CCR-9712362.

[‡] School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada (astorjoh@scg.uwaterloo.ca). Generously supported by the NSERC.

[§] CNRS - Laboratoire LIP, ENSL 46, Allée d’Italie, 69364 Lyon Cedex 07, France (Gilles.Villard@ens-lyon.fr).

These algorithms essentially provide output certificates for the approaches in [6, 11]. They require the computation of the trace of ADA^* or $(ADA^*)^2$ where D is a diagonal matrix and A^* the Hermitian transpose of A . For instance, when $A \in F^{n \times m}$ is a black box the trace computation costs n or $2n$ multiplications of A by a vector, the same number of multiplications of A^* by a vector, plus $O(nm)$ field operations. In addition to the trace computation, the algorithms use an expected number of $O(r)$ multiplications of A by a vector, the same number of multiplications of A^* by a vector and, respectively, $O((m+n)r)$ and $O((m+n \log r)r)$ additional operations in F . The first algorithm, presented in Section 2, is based on minimal polynomial computation using Wiedemann's algorithm [16]. The second algorithm, presented in Section 3, gives a rank certificate based on the Lanczos approach. We apply these algorithms to derive new bit-complexity estimates for computing the rank of integer matrices in Section 4 and give concluding remarks in Section 5.

2. Rank Certificate using Trace. The problem is to compute the rank of a given matrix $A \in F^{n \times m}$. We will reduce this problem to that of computing the minimal polynomial of a square matrix $B \in F^{n \times n}$ that has the same rank as A and possesses in addition the following properties:

- a** B is diagonalizable, that is, the Jordan form of B can be written as $\text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$, where r is the rank of B , and the λ_* are the nonzero eigenvalues of B in the appropriate extension field.
- b** B is positive semi-definite, that is, $\lambda_i > 0$, $1 \leq i \leq r$.

The matrix B will be constructed from A as in the following fact and proposition. We use A^* to mean the Hermitian transpose of A , the transpose of A with entries conjugated.

FACT 1. *Let $A \in F^{n \times m}$ be given. Let D be an $m \times m$ diagonal matrix with positive real entries from F , so that D can be expressed as EE^* for a diagonal matrix E in the algebraic closure of F . Then $B = ADA^*$ has the same rank as A and possesses properties **a** and **b**.*

Similar preconditioned forms such as DAA^*D or DAA^* are discussed in [4, 15]. The form ADA^* has the additional property that, when applied over a field of positive characteristic, the rank is likely preserved [13] (with some exceptions [6]).

PROPOSITION 1. *Let B be as in Fact 1. Let the diagonal entries in D be chosen uniformly and randomly from a subset of $F \setminus \{0\}$ with cardinality s . With probability at least $1 - n(n-1)/(2s)$ the minimal polynomial of B is $x \prod_{i=1}^r (x - \lambda_i)$ when B is singular ($r < n$) and $\prod_{i=1}^n (x - \lambda_i)$ otherwise ($r = n$).*

Proof. We use the techniques of [4, Section 4]. We first show that with $\mathcal{D} = \text{diag}(y_1, \dots, y_m)$ a diagonal matrix whose diagonal entries are indeterminates, the characteristic polynomial $c(x) = x^n + c_1 x^{n-1} + \dots + c_{n-1} x + c_n$ of $\mathcal{B} = ADA^*$ has no repeated factor other than x . Each coefficient c_i is a sum of $i \times i$ minors of \mathcal{B} or is zero. Since an $i \times i$ minor of \mathcal{B} is a linear combination over F of $i \times i$ minors of \mathcal{D} , each c_i is either homogeneous of degree i in y_1, \dots, y_m or is zero. Hence $c(x)$ is homogeneous of degree n in y_1, \dots, y_m and x . In addition, $c(x)$ is at most linear in each y_j since each $i \times i$ minor of \mathcal{D} is. Now, if $c(x)$ has a repeated factor $\bar{c}(x)$ so that $\bar{c}^2(x)$ is a factor of $c(x)$, then no indeterminate y_j can occur in $\bar{c}(x)$ since otherwise $\bar{c}^2(x)$ and



$c(x)$ would not be linear in y_j . Thus the repeated factor $\bar{c}(x)$ is a polynomial in $F[x]$, is homogeneous since any factor of $c(x)$ is and must be a monomial in x .

Together with property **a**, the fact that the nonzero eigenvalues of $\mathcal{B} = ADA^*$ are distinct gives that the minimal polynomial of \mathcal{B} has degree $\rho = r + 1$ if \mathcal{B} is singular and $\rho = r = n$ otherwise. Let u be a vector in F^n such that $u, Bu, B^2u, \dots, B^{\rho-1}u$ are linearly independent. There is a $\rho \times \rho$ submatrix of the matrix with these vectors as its columns whose determinant is a nonzero polynomial of total degree $\rho(\rho-1)/2 \leq n(n-1)/2$ in the indeterminates y_1, \dots, y_m . By the Schwartz-Zippel Lemma [5, 14, 17] if we evaluate this polynomial, each variable y_1, \dots, y_m chosen uniformly and randomly from a subset of F of size s , then the probability that the result is nonzero is at least $1 - n(n-1)/(2s)$. Hence the minimal polynomial of B also has degree ρ with probability at least $1 - n(n-1)/(2s)$. Since B is diagonalizable (property **a**) the assertion of the proposition follows. \square

Proposition 1 gives that the minimal polynomial of B is $xh(x)$ or $h(x)$ where $h(x) = \prod_{i=1}^r (x - \lambda_i)$, $\lambda_i \neq 0$, with high probability if the cardinality s is large enough. In this case the rank of B and hence of A can be recovered as $\deg h(x)$.

The minimal polynomial of B can be computed in an output sensitive fashion (with respect to r) by adapting Wiedemann's approach [16]. Choose a random vector $u \in F^{n \times 1}$ and, for increasing i , iteratively apply the Berlekamp-Massey algorithm to the scalar sequence prefix $u^*u, u^*Bu, u^*B^2u, \dots, u^*B^{2i}u$. Stop for the first i such that this sequence prefix has an annihilator of degree i . Because we are assuming F has characteristic zero, this annihilator will be the minimal polynomial of the vector sequence u, Bu, B^2u, \dots , and thus be a factor of the minimal polynomial of B . In particular, the matrix $K_u = [u, Bu, B^2u \dots, B^{i-1}u]$ is rank deficient if and only if the Hankel matrix $H_u = K_u^*K_u$ is singular.

LEMMA 2.1. [16] *Let $B \in F^{n \times n}$ be as in Fact 1. There exists a Monte Carlo probabilistic algorithm that recovers the minimal polynomial of B using $O(r)$ matrix-vector products involving B plus additional $O(nr)$ field operations, r the rank of B . The output will always be a monic factor of the minimal polynomial of B .*

Suppose B possesses property **a** with nonzero eigenvalues $\lambda_1, \dots, \lambda_r$. Then r is the rank of B . Let $g(x) = x^q + g_1x^{q-1} + \dots + g_q$ ($g_q \neq 0$) be such that the minimal polynomial of B is equal to $g(x)$ or $xg(x)$. Let $f(x) = x^p + f_1x^{p-1} + \dots + f_p$ ($f_p \neq 0$) be a monic factor of $g(x)$. Thus $f(x)|g(x)|h(x)$ where $h(x) = \prod_{i=1}^r (x - \lambda_i) = x^r + h_1x^{r-1} + (\text{lower order terms})$. Then $p \leq q \leq r$ and, up to reordering of the λ_i , we have $f_1 = -(\lambda_1 + \dots + \lambda_p)$, $g_1 = -(\lambda_1 + \dots + \lambda_q)$ and $h_1 = -(\lambda_1 + \dots + \lambda_r)$. Now suppose that B possesses also property **b**. Then $f_1 = g_1$ if and only if $p = q$. Similarly, $g_1 = h_1$ if and only if $q = r$. Using the fact that $\lambda_1 + \dots + \lambda_r = \text{trace}(B)$ we get the following result:

LEMMA 2.2. *Let B possess property **a** and **b**. Let $f(x) = x^p + f_1x^{p-1} + \dots + f_p$ ($f_p \neq 0$) be a monic factor of the minimal polynomial of B . Then $-f_1 = \text{trace}(B)$ if and only if p is the rank of B .*

We can now give our first algorithm for rank.

Algorithm Rank-certificate-using-trace

Input: $A \in F^{n \times m}$.

Output: rank A or “failed”.

1. Choose D to be an $m \times m$ diagonal matrix with entries chosen uniformly and randomly from a subset of F with cardinality $2n^2$.
Let $B := ADA^*$.
2. Compute a monic factor of the minimal polynomial of B which is, with probability at least $3/4$, the minimal polynomial.
3. Express the factor as $f(x)$ or $xf(x)$ where
 $f(x) = x^p + f_1x^{p-1} + \dots + f_p$ with $f_p \neq 0$.
4. If $-f_1 = \text{trace}(B)$ return p otherwise return “failed” (or start over).

Repetition of algorithm Rank-certificate-using-trace is required with probability less than $n(n-1)/(4n^2)+1/4 < 1/2$. Note also that a matrix-vector product involving B requires one product of A by a vector, one product of A^* by a vector and $O(m)$ additional field operations. We get the following result as a corollary to all of the above.

PROPOSITION 2. *Let $A \in F^{n \times m}$ be of rank r . The Las Vegas algorithm Rank-certificate-using-trace works as announced using the trace of ADA^* , $O(r)$ further multiplications of A by a vector, the same number of multiplications of A^* by a vector, and $O((n+m)r)$ additional field operations.*

For instance, in the black box case the trace of B may be computed from the diagonal entries of $A \cdot (DA^*)$ in n multiplications of A and of A^* by vectors plus $O(nm)$ operations.

3. Rank Certificate using Orthogonalization. Our second rank certificate, à la Lanczos, is based on vector norms. It can be used for instance in combination with Monte Carlo black box algorithms for the rank such as the one based on the Lanczos approach in [6]. We assume that $A \in F^{n \times m}$ has presumed rank r . We will use the same preconditioning as in Section 2 and thus consider $B = ADA^* \in F^{n \times n}$ for a random diagonal matrix D . Given a basis u_1, \dots, u_r of the (presumed) range space $\mathcal{V} \subseteq F^n$ of B , to certify that the rank of B is r can be done by showing that all the column vectors b_1, \dots, b_n of B are in \mathcal{V} . For F a field as specified, we may equivalently show that the projections $\bar{b}_i = b_i - \sum_{j=1}^r \gamma_{i,j} u_j$ (for appropriate coefficients $\gamma_{i,j}$ in F) of the b_i 's onto \mathcal{V}^\perp are zero. This also turns out to be equivalent to certify that $\tau_i = \langle b_i, \bar{b}_i \rangle = 0$, $1 \leq i \leq n$, or

$$(3.1) \quad \sum_{i=1}^n \tau_i = \sum_{i=1}^n \langle b_i, \bar{b}_i \rangle = 0$$

since the dot products must be nonnegative.

For computing the \bar{b}_i 's and the τ_i 's we introduce K_u , an $n \times r$ matrix whose columns form a Krylov basis of the (presumed) range space of B . Such a matrix can be computed from a random vector $v \in F^n$ and $u = Bv$ which is therefore a random vector in the range space of B . If r is the actual rank of A then Proposition 1 gives



that with high probability the minimal polynomial of B has degree r (when A has full row rank) or $r + 1$. Hence we know from [16, section VI] (see also Lemma 2.1) or from [10, section 2] that with high probability $K_u = [u, Bu, B^2u \dots, B^{r-1}u]$ has rank r . The matrix $H_u = K_u^*K_u$ which is a square Hankel matrix of dimension r is thus invertible with high probability. The b_i 's are projected onto \mathcal{V}^\perp using the matrix $P \in F^{r \times n}$ such that:

$$(3.2) \quad H_u P = K_u^* K_u P = K_u^* B$$

or equivalently, such that $K_u^*(B - K_u P) = 0$. Taking \mathcal{V} equal to the range space of K_u the columns of $B - K_u P$ are the \tilde{b}_i 's, and we see that the test dot products of (3.1) are the diagonal entries of

$$(3.3) \quad B(B - K_u P).$$

The rank certification thus amounts to the following. Identity (3.2) gives that the matrix P can be computed as $H_u^{-1}(K_u^*B)$. The construction of the Krylov matrix K_u and of K_u^*B require $O(r)$ products of B by vectors. The matrix H_u is computed in $O(nr)$ and since it is a Hankel matrix one may check its invertibility in $O(r \log^2 r)$ and compute the product $H_u^{-1}(K_u^*B)$ in $O(rn \log r)$ operations in F [3, 2] (see also [1, sections 2.5-2.7]). The computation of the diagonal entries of (3.3) then needs the trace of B^2 , $O(r)$ products of B by vectors to get the matrix BK_u and $O(nr)$ operations to get the diagonal entries of $BK_u P$.

Algorithm Rank-certificate-using-orthogonalizations

Input: $A \in F^{n \times m}$; r , the presumed rank of A .

Output: rank A or “failed”.

1. Let $B := ADA^*$. * *Random preconditioning* *
2. Choose a random vector v . Let $u := Bv$.
3. Apply B iteratively to compute K_u and $K'_u = BK_u$.
4. If $\det H_u = \det K_u^* K_u = 0$ then return “failed” (or start over).
 Otherwise use a Hankel solver for $P := H_u^{-1}(K'_u)^*$.
5. Apply B to compute $B_{i,i}^2$, $1 \leq i \leq n$.
6. Let $\tau_i = B_{i,i}^2 - \langle (K'_u)_{i,\cdot}, P_{\cdot,i} \rangle$, $1 \leq i \leq n$.
7. If $\sum_{i=1}^n \tau_i = 0$ then return r otherwise return “failed” (or start over).

As in Section 2 we can fix the random choice of D and v in order to achieve a probability of repetition of the algorithm smaller than $1/2$. Hence if r is the actual rank of A , the algorithm will certify the value with a probability at least $1/2$. If the input r is not the rank then the algorithm will always fail. Indeed, if r is too small then some column of B , say the j -th one, will not belong to the range space of K_u and will lead to $\tau_j \neq 0$. If r is larger than the rank then H_u will be singular. Note that when r linearly independent columns of B are known in advance only $n - r$ dot products (for the remaining columns) need to be tested.

PROPOSITION 3. *Let $A \in F^{n \times m}$. The Las Vegas algorithm Rank-certificate-using-orthogonalizations works as announced using the trace of $(ADA^*)^2$, $O(r)$ fur-*

ther multiplications of A by a vector, the same number of multiplications of A^* by a vector and $O((m + n \log r)r)$ additional field operations.

In the black box case the trace is computed using $2n$ multiplications of A and of A^* by vectors plus $O(nm)$ operations in F .

The term $O(nr \log r)$ in the cost of this second certificate differs by a logarithmic factor from the corresponding term $O(nr)$ in Section 2. Despite its larger asymptotical cost for larger r , we have presented this second rank certificate for possible insights in finding a certificate over any field. Also, although we have in mind exact (symbolic) computation here, its potentially greater stability properties may be relevant in some contexts. We may also notice that the two certificates are closely related to each other: the test $\text{trace}(B) + f_1 = 0$ may be compared to the test $\sum_{i=1}^n \tau_i = 0$.

4. Integer matrix rank. The rank of an integer matrix $A \in \mathbf{Z}^{n \times m}$ can be computed with high probability as the rank of A modulo a randomly chosen prime number p . If p is chosen in a sufficiently large set with respect to n and to the infinity norm $\|A\|$ of A , this leads to a Monte Carlo algorithm for computing the rank of A using $O(n^3 \log \log \|A\| + n^2 \log \|A\|)$ bit operations. That cost is without utilizing sub-cubic matrix multiplication algorithms and with p chosen in a set of primes having $O(\log n + \log \log \|A\|)$ bits (see, e.g., [7, Section 3.2] for this latter choice). However this approach may produce wrong answers.

If the rank is to be certified, the lengths of the integers involved in the computation further affect the running time of the algorithms and an extra factor $O(n)$ occurs in the complexity estimates.

Algorithm Rank-certificate-using-trace can be adapted to the case of a dense integer matrix $A \in \mathbf{Z}^{n \times m}$ and improve the complexity exponent for certifying the rank. Construct $B = ADA^T \in \mathbf{Z}^{n \times n}$ as in Fact 1 and Proposition 1. The baby-step/giant-step approach of Kaltofen [9, 12] can be used to construct a monic factor of the minimal polynomial of B (which will with high probability be the minimal polynomial of B) using an expected number of $O(n^{3.5} \log^2 \|A\|)$ bit operations. Our rank certificate can also be applied for certifying the rank of sparse or structured integer matrices using the black box approach. Indeed the techniques we have presented can be combined with Chinese remaindering (following [7, Corollary 3.10] for instance). In both the dense and the black box case, one needs to certify the fact that the computed monic polynomial is a factor of the minimal polynomial over \mathbf{Z} , this can be accomplished using the techniques of [8, Lemma 2.4].

5. Conclusions. We have provided two algorithms of Las Vegas type for exact computation of the rank of a matrix over a field of characteristic zero.

We note that in the black box model the certificates both add $O(n)$ matrix-vector products to the $O(r)$ products needed for the rank itself. This extra cost for smaller r is for computing the trace of the conditioned matrix. Under the same assumptions it would be interesting to know whether the rank can be certified in only $O(r)$ applications of the black box.

For a number of applications it would be desirable to efficiently certify the rank of a matrix over a field with positive characteristic, in particular over a finite field. Our methods do not work in this setting, the essential problem being the existence of

self-orthogonal vectors. However, the probability estimates for the Monte Carlo rank algorithms typically require random choice from a set whose size is a small multiple n^2 . When $n > 2^{16}$ or $n > 2^{32}$, for example, this can force modular methods to choose large finite fields requiring multiple computer words to store each individual field element and requiring relatively expensive arithmetic costs. In practice, the rank is correctly found, even when the random values are from a much smaller set, say of size $O(n)$. The algorithms of this paper can be used over finite fields as heuristics to strengthen confidence in the result. For instance, naively, one would suppose that if the trace corresponds to the first coefficient of the purported minimal polynomial of a preconditioned matrix, it is a strong indicator that the polynomial is in fact the minimal polynomial. Still we have no argument to quantify the probability here.

In addition to this use as a heuristic, it may be hoped that one or the other of the two presented algorithms will provide insight useful for solving the open problem for any field.

REFERENCES

- [1] Dario Bini and Victor Pan. *Polynomial and matrix computations*. Birkhäuser, Boston, 1994.
- [2] R.R. Bitmead and B.D.O. Anderson. Asymptotically fast solution of Toeplitz and related systems of linear equations. *Linear Algebra and its Applications*, 34:103–116, 1980.
- [3] R.P. Brent, F.G. Gustavson, and D.Y.Y. Yun. Fast solution of Toeplitz systems of equations and computation of Padé approximations. *Journal of Algorithms*, 1:259–295, 1980.
- [4] L. Chen, W. Eberly, E. Kaltofen, B.D. Saunders, W.J. Turner, and G. Villard. Efficient matrix preconditioners for black box linear algebra. *Linear Algebra and its Applications*, 343/344:119–146, 2002.
- [5] R.A. DeMillo and R.J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193–195, 1978.
- [6] W. Eberly and E. Kaltofen. On randomized Lanczos algorithms. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation* (Maui, Hawaii), 176–183. ACM Press, New York, 1997.
- [7] Mark Giesbrecht. Fast computation of the Smith form of a sparse integer matrix. *Computational Complexity*, 10:41–69, 2001.
- [8] Mark Giesbrecht and Arne Storjohann. Computing rational forms of integer matrices. *Journal of Symbolic Computation*, 34(3):157–172, 2002.
- [9] E. Kaltofen. On computing determinants of matrices without divisions. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation* (Berkeley, California), 342–349. ACM Press, New York, 1992.
- [10] E. Kaltofen and V. Pan. Processor efficient parallel solution of linear systems over an abstract field. In *Proceedings of the 3rd Annual ACM Symposium on Parallel Algorithms and Architecture (SPAA'91)*, 180–191. ACM Press, New York, 1991.
- [11] E. Kaltofen and B.D. Saunders. On Wiedemann's method of solving sparse linear systems. In *Proceedings of Applied Algebraic Algorithms and Error Correcting Codes* (New Orleans, LA), 29–38. *Lecture Notes in Computer Science*, Vol. 539, Springer, Berlin, 1991.
- [12] Erich Kaltofen. An output-sensitive variant of the baby steps/giant steps determinant algorithm. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation* (Lille, France), 138–144. ACM Press, New York, 2002.
- [13] B. LaMacchia and A. Odlyzko. Solving large sparse linear systems over finite fields. In *Advances in Cryptology*, (Santa Barbara, CA), 109–133. *Lecture Notes in Computer Science*, Vol. 537, Springer, Berlin, 1990.

- [14] J.T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the Association for Computing Machinery*, 27:701–717, 1980.
- [15] William J. Turner. *Black box linear algebra with the LINBOX library*. PhD thesis, North Carolina State Univ., Raleigh, North Carolina, August 2002 (193 pages).
- [16] D. Wiedemann. Solving sparse linear equations over finite fields. *IEEE Transactions on Information Theory*, IT-32:54–62, 1986.
- [17] R. Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation* (Marseille, France), 216–226. *Lecture Notes in Computer Science*, Vol. 72, Springer, Berlin, 1979.