Matrix Rank Certification

David Saunders, Arne Storjohann, Gilles Villard

Laboratoire de l'Informatique du Parallélisme

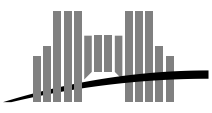
École Normale Supérieure de Lyon
Unité Mixte de Recherche CNRS-INRIA-ENS LYON n° 5668

Matrix Rank Certification

B. David Saunders, Arne Storjohann and Gilles Villard

Août 2001

Research Report N° 2001-30



École Normale Supérieure de Lyon
46 Allée d'Italie, 69364 Lyon Cedex 07, France
Téléphone : +33(0)4.72.72.80.37
Télécopieur : +33(0)4.72.72.80.80
Adresse électronique : `lip@ens-lyon.fr`

Matrix Rank Certification

B. David Saunders, Arne Storjohann and Gilles Villard

Août 2001

Abstract

Randomized algorithms are given for computing the rank of a matrix over a field of characteristic zero. The matrix is treated as a black box. Only the capability to compute matrix × column-vector and row-vector × 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.

Keywords: Linear algebra, randomized algorithms, black box matrix, matrix rank, seminumeric computation.

Résumé

Nous proposons deux algorithmes probabilistes pour le calcul du rang d'une matrice sur un corps de caractéristique zéro. La matrice est vue comme une boîte noire. Les seuls opérations où elle est impliquée sont des produits matrice × vecteur-colonne et vecteur-ligne × matrice. Les méthodes sont exactes, appropriées aux matrices entières ou rationnelles par exemple. Les algorithmes sont probabilistes de type Las Vegas c'est-à-dire que le résultat est garanti.

Mots-clés: Algèbre linéaire, algorithmes probabilistes, matrice boîte noire, matrice creuse, rang.

Matrix Rank Certification

B. David Saunders*
Dpt. of Comp. and Infor. Sc.
University of Delaware
Newark, Delaware 19716, USA
saunders@mail.eecis.udel.edu
www.cis.udel.edu/~saunders

Arne Storjohann
Dpt. of Computer Science
University of Western Ontario
London, Ontario, N6A 5B7 Canada
astorjoh@scg.math.uwaterloo.ca
www.scl.csd.uwo.ca/~storjoha

Gilles Villard
CNRS, Laboratoire LIP
Ecole Normale Supérieure de Lyon
46 Allée d'Italie
69364 Lyon Cedex France
Gilles.Villard@ens-lyon.fr
www.ens-lyon.fr/~gvillard

Abstract

Randomized algorithms are given for computing the rank of a matrix over a field of characteristic zero. The matrix is treated as a black box. Only the capability to compute matrix×column-vector and row-vector×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.

1 Introduction

The rank of an $n \times n$ 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 are an alternative to using elimination.

Several Monte Carlo black box methods for rank have been developed [5, 8]. They require $O(n)$ matrix-vector products. 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 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. The methods presented here can be used to remove the possibility of an erroroneous result in the case when F is a field of characteristic zero.

We give two algorithms. Each requires an expected number of $O(n)$ matrix-vector products and additional $O(n^2)$ field operations to compute the correct rank of A. The first algorithm, presented in Section 2, is based on minimal polynomial computation using Wiedemann's algorithm [11]. The second algorithm, presented in Section 3, is based on the Lanczos approach. Both of our algorithms require that the field be of characteristic zero.

*Generously supported by the NSF, grant CCR-9712362 (Saunders).

2 Rank Certificate using Trace

All matrices in this section are over a field F of characteristic zero with conjugation operator. The problem is to compute the rank of a given matrix A. We will reduce this problem to that of computing the minimal polynomial of a square matrix B that possesses the following properties:

- **a** B is diagonalizable, that is, the Jordan form of B can be written as $\mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$ where λ_* are the nonzero eigenvalues of B in the appropriate extension field. We will use the fact that in this case the rank of B equals r.
- **b** B is positive semi-definite, that is, $\lambda_i > 0$ for all $i, 1 \leq i \leq r$.

With high probability B should possess also property:

- **c** The minimal polynomial of B is $xh(x)$ or $h(x)$ where $h(x) = \prod_{i=1}^{r}(x - \lambda_i)$. We will use the fact that this condition holds when $\lambda_i \neq \lambda_j$ for $i \neq j$.

Such a matrix B can be constructed using the following fact and lemma. Note that we use A^* to mean the Hermitian transpose of A, the transpose of A with entries conjugated.

Fact 2.1 *Let $A \in F^{n \times m}$ be given. Let D be a $m \times m$ 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]. The ADA^* has the additional property that, when applied over a field of positive characteristic, the rank is likely preserved [9] (with some exceptions [5]).

Proposition 2.2 *Let B be as in Fact 2.1. If diagonal entries in D are chosen uniformly and randomly from a subset of $F \setminus \{0\}$ with cardinality s then B possesses in addition property* **c** *with probability at least $1 - 2n^2/s$.*

The Schwartz-Zippel Lemma [10, 12] states that if we evaluate a multivariate polynomial of total degree d, with coefficients from F, each variable chosen uniformly and randomly from a subset S of F of size s, then the probability that the result is nonzero is $> 1 - d/s$. Proposition 2.2 follows as a corollary of the Schwartz-Zippel Lemma and the next result.

Lemma 2.3 *Let A be an $n \times m$ matrix over F with $n \leq m$. Let $D = \mathrm{diag}(y_1, \ldots, y_m)$ be a diagonal matrix of indeterminants. Write the characteristic polynomial $\det(xI_n - ADA^*)$ of ADA^* as $x^{n-l}g(x)$ where $g(x) \in F[y_1, \ldots, y_m][x]$ has a nonzero constant coefficient with respect to x. Then l is the rank of A and the discriminant of g with respect to x is not the zero polynomial in $y_1, \ldots, y_m$. This discriminant will have total degree bounded by $2n^2$.*

Proof: Let r be the rank of A. Then there exists a symmetric $r \times r$ minor of ADA^* with rank r. Without loss of generality, assume the principal $r \times r$ minor has rank r. Consider the specialization of $x^{n-l}g$ if we substitute $y_{r+1} = \cdots = y_m = 0$. We get

$$\begin{aligned} \det(xI_n - A\,\mathrm{diag}(y_1, \ldots, y_r, 0, \ldots, 0)A^*) &= x^{n-r}\det(xI_r - \bar{A}\bar{D}\bar{A}^*) \\ &= x^{n-r}\bar{g} \end{aligned}$$

where $\bar{D} = \mathrm{diag}(y_1, \ldots, y_r)$ and $\bar{A}$ is the principal $r \times r$ submatrix of A. The trailing degree of $x^{n-l}g$ cannot be less than the specialization $x^{n-r}\bar{g}$. Since $\bar{A}\bar{D}\bar{A}^*$ is nonsingular we get $l \leq r$. To

see that $l \geq r$, note that the coefficient of x^{n-i} $(1 \leq i \leq n)$ in $x^{n-l}g$ is the sum of all symmetric $i \times i$ minors of the rank r matrix ADA^*. Since ADA^* has rank r, these coefficients must be zero for $i > r$.

At this point we have $\deg \bar{g} = \deg g$ where $\bar{g}$ is equal to g but with some indeterminates set to zero. Thus, to show that g is squarefree it will be sufficient to show that $\bar{g}$ is squarefree. Note that $\bar{A}\bar{D}\bar{A}^*$ is similar to $\bar{A}^*\bar{A}\bar{D}$ and that $\bar{A}^*\bar{A}$ has each principal minor nonzero. From a result of Wiedemann [11, last lemma on page 59] it follows that that the discriminant of $\det(xI_r - \bar{A}^*\bar{A}\bar{D})$ with respect to x is not identically zero. The degree bound is easy to derive. □

Lemma 2.4 *[4, Theorem 4.7] Let B be as in Fact 2.1.If diagonal entries in D are chosen uniformly and randomly from a subset of $F \setminus \{0\}$ with cardinality s then B possesses in addition property* **c** *with probability at least $1 - 2n^2/s$.*

The minimal polynomial of B can be recovered using the following result.

Lemma 2.5 *[11] Let $B \in F^{n\times n}$. There exists a Monte Carlo probabilistic algorithm that recovers the minimal polynomial of B using $O(n)$ matrix-vector products involving B plus additional $O(n^2)$ field operations. The output will always be a monic factor of the minimal polynomial of B.*

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

Lemma 2.6 *Let B possess property* **a** *and* **b**. *Let $f(x) = x^p + f_1x^{p-1} + \ldots + f_p$ $(f_p \neq 0)$ be a monic factor of the minimal polynomial of B. Then $-f_1 = \mathrm{trace}(B)$ 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. Construct B as in Fact 2.1such that B possesses property **c** with probability at least 3/4.
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 $f(x) = x^p + f_1x^{p-1} + \cdots + f_p$ with $f_p \neq 0$.
4. If $-f_1 = \mathrm{trace}(B)$ return p otherwise return "failed" (or start over).

Repetition of algorithm Rank-certificate-using-trace is required with probability less than $1 - (3/4)^2 < 1/2$. Note also that a matrix vector product involving B requires one matrix-vector and one vector-matrix product involving A plus additional n field multiplications, n the column dimension of A. We get the following result as a corollary to all of the above.

Proposition 2.7 *Let $A \in F^{n\times m}$. The Las Vegas algorithm* rank-certificate-using-trace *works as announced using $O(n)$ matrix-vector and vector-matrix products involving A plus $O(nm)$ additional field operations.*

3 Rank Certificate using Orthogonalization

Our second rank certificate is based on vector norms, rather than on an identity involving the trace of the matrix, We assume that $A \in F^{n\times m}$ has presumed rank r. We will use the same preconditioning as in section 2 to apply Lemma 2.4 and thus consider $B = ADA^*$ for a random diagonal matrix D. Given a basis $u_1, \ldots, u_r$ of the (presumed) range space $\mathcal{V}$ of B, to certify that the rank of B is r can be done by showing that all the column vectors $b_1, \ldots, b_n$ of B are in $\mathcal{V}$. For F a field as specified, we may equivalently show that the projections $\bar{b}_i = b_i - \sum_j \gamma_{j=1}^{r} u_j$ of the b_i's onto $\mathcal{V}^\perp$ are zero. It is equivalent to certify that:

$$\tau_i = \langle b_i, \bar{b}_i \rangle = 0,\ 1 \le i \le n$$

or

$$\sum_{i=1}^{n} \tau_i = \sum_{i=1}^{n} \langle b_i, \bar{b}_i \rangle = 0 \tag{1}$$

since the dot products must be positive. The orthogonalized vectors will be computed *à la* Lanczos. We introduce K_u, a $n \times r$ matrix whose columns form a Krylov basis of the (presumed) range space of B. Such a matrix can be computed from a random vector $v \in F^m$ and $u = Bv$ which is therefore a random vector in the range space of B. By Lemma 2.4, with high probability the minimum polynomial of B has degree r (when A is invertible) or $r+1$ and we know from [11, section VI] – where the minimum polynomial of a matrix is computed from a Krylov basis (see also Lemma 2.5) – or from [7, section 2], that with high probability, $K_u = [u, Bu, B^2u \ldots, B^{r-1}u]$ has rank r. The matrix $H_u = K_u^* K_u$ which is square Hankel of dimension r is thus invertible with high probability. The b_i are projected onto $\mathcal{V}^\perp$ using the matrix $P \in F^{r\times m}$ such that:

$$H_u P = K_u^* K_u P = K_u^* B$$

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 $\bar{b}_i$'s and we see that the test dot products of (1) are the diagonal entries of

$$B(B - K_u P). \tag{2}$$

The cost of the rank certification thus amounts to the following. The matrix P may be computed as $H_u^{-1}(K_u^* B)$. The construction of the Krylov matrix K_u and and of $K_u^* B$ require $O(n)$ products of B by vectors. The matrix H_u is computed in $O(n^2)$ and since it is Hankel one may check its invertibility in $O(n \log^2 n)$ and compute the product $H_u^{-1}(K_u^* B)$ in $O(n^2 \log n)$ arithmetic operations [3, 2] (see also [1, sections 2.5-2.7]). The computation of the diagonal entries of (2) then needs $O(n)$ products of B by vectors to get the diagonal entries of B^2 and to get the matrix BK_u. In $O(n^2)$ final operations the diagonal entries of BK_uP and thus the target scalar products are known. This leads to:

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 $B := ADA^*$. * *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 $B^2_{i,i}$, $1 \le i \le n$.
6. Let $\tau_i = B^2_{i,i} - \langle (K'_u)_{i,\cdot}, P_{\cdot,i}\rangle$, $1 \le i \le n$.
7. If $\sum_{i=1}^n \tau_i = 0$ then return r otherwise return "failed" (or start over).

If r is the actual rank of A, the algorithm will certify the value with a probability arbitrarily close to zero if the entries of v are chosen uniformly and independently from a subset of F containing sufficiently many elements (see Lemma 2.4 and [11, 7]). 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, H_u will be singular. The cost of the algorithm could be made rank sensitive if r linearly independent columns are known by testing only $n - r$ dot products. We have proven:

Proposition 3.1 *Let $A \in F^{n\times m}$. The Las Vegas algorithm* rank-certificate-using-orthogonalizations *works as announced using $O(n)$ matrix-vector and vector-matrix products involving A plus $O(nm + n^2 \log n)$ additional field operations.*

This second certificate is asymptotically more expensive by a log factor than the one in section 2. It is proposed for possible insights in finding a certificate for any field. Also, although we have in mind exact (symbolic) computation here, it's greater stability properties may be relevant in some contexts. We may also notice that the two certificates are related each other: the test $\mathrm{trace}(B) + g_1 = 0$ may be compared to the test $\sum_i \tau_i = 0$.

4 Conclusions

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

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 don't work in this setting, the essential problem being the existence of self-orthogonal vectors. It may be hoped that one or the other of these two algorithms will provide insight useful in solving that open problem.

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 so, 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. However we have no argument to quantify the probability here.

Algorithm Rank-certificate-using-trace can be adapted to the case of a dense integer matrix $A \in \mathbf{Z}^{n\times m}$. Construct $B = ADA^T \in \mathbf{Z}^{n\times n}$ as in Fact 2.1 and Proposition 2.2. The baby-step/giant-step approach of Kaltofen [6] 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 $O(n^{3.5}(\log||A||_2)^2)$ bit operations.

References

[1] D. Bini and V. Pan. *Polynomial and matrix computations.* Birkhäuser, 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 Appl.*, 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. Research Report RR 2001-05, LIP, ENS Lyon, France. Jan. 2001.

[5] W. Eberly and E. Kaltofen. On randomized Lanczos algorithms. In *International Symposium on Symbolic and Algebraic Computation, Maui, Hawaii, USA*, pages 176–183. ACM Press, July 1997.

[6] E. Kaltofen. On computing determinants of matrices without divisions. In *International Symposium on Symbolic and Algebraic Computation, Berkeley, California, USA*, July 1992.

[7] E. Kaltofen and V. Pan. Processor efficient parallel solution of linear systems over an abstract field. In *Proc. 3rd Annual ACM Symposium on Parallel Algorithms and Architecture.* ACM-Press, 1991.

[8] E. Kaltofen and B.D. Saunders. On Wiedemann's method of solving sparse linear systems. In *Proc. AAECC-9*, LNCS 539, Springer Verlag, pages 29–38, 1991.

[9] B. LaMacchia and A. Odlyzko. Solving large sparse linear systems over finite fields. In *Advances in Cryptology - CRYPTO '90*, volume LNCS 537, pages 109–133. Springer, 1990.

[10] J.T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27:701–717, 1980.

[11] D. Wiedemann. Solving sparse linear equations over finite fields. *IEEE Transf. Inform. Theory*, IT-32:54–62, 1986.

[12] R. Zippel. Probabilistic algorithms for sparse polynomials. In *Proc. EUROSAM 79*, pages 216–226, Marseille, 1979.