Nonnegative Ranks, Decompositions, and Factorizations of Nonnegative Matrices

Joel E. Cohen
The Rockefeller University
1230 York Avenue, Box 20
New York, New York 10021-6399

and

Uriel G. Rothblum
Faculty of Industrial Engineering and Management
Technion—Israel Institute of Technology
Haifa 32000, Israel
and
RUTCOR—Rutgers Center for Operations Research
Rutgers University
P.O.B. 5062
New Brunswick, New Jersey 08903-5062

Submitted by Hans Schneider

ABSTRACT

The *nonnegative rank* of a nonnegative matrix is the smallest number of nonnegative rank-one matrices into which the matrix can be decomposed additively. Such decompositions are useful in diverse scientific disciplines. We obtain characterizations and bounds and show that the nonnegative rank can be computed exactly over the reals by a finite algorithm.

1. INTRODUCTION

Consider matrices all elements of which belong to a given ordered field G such as the rational numbers Q or the real numbers R. A rank-one matrix can be written as xy^T, where x and y are column vectors over G and ^{T}

655 Avenue of the Americas, New York, NY 10010 0024-3795/93/$6.00

denotes the transpose operator. Given a matrix A over G, the central problem of this paper is to find the smallest number q of rank-one nonnegative matrices over G such that A equals their sum. Following Gregory and Pullman (1983), we call this (smallest) number q the *nonnegative rank* of A. The problem of obtaining such decompositions of nonnegative matrices, or corresponding approximations, arises in a variety of scientific contexts like demography, quantum mechanics, combinatorial optimization, complexity theory, probability, and statistics; see Section 6 for details.

In a special symmetric version of our problem, Berman and Hershkowitz (1987) define a matrix A to be *completely positive* if there exists q nonnegative column vectors $b_1, \dots, b_q$ such that $A = \Sigma_{i=1}^{q} b_i b_i^T$, and they call the smallest such q the *factorization index* of A. We do not deal with this problem here.

The organization of the remainder of this paper is as follows. In Section 2, we obtain a number of characterizations and bounds of the nonnegative rank of nonnegative matrices; some resemble standard facts about regular ranks. In Section 3 we establish relationships with bivariate probability matrices, stochastic matrices, and the geometry of polytopes. In Section 4 we examine matrices with rank 2 or less. Using Tarski's principle and a quantifier elimination algorithm due to Renegar, we show in Section 5 that the nonnegative rank can be computed (finitely) over the reals, and we describe the complexity of the calculation. Finally, applications are discussed in Section 6.

2. NONNEGATIVE RANKS OF NONNEGATIVE MATRICES

Let G be a given ordered field such as the rationals Q and the reals R. We use the standard notation for operations in ordered fields. For a matrix A in $G^{m \times n}$, for $i = 1, \dots, m$ and for $j = 1, \dots, n$, let A_i denote the ith *row* of A, let A^j denote the jth *column* of A, and let A_i^j denote the ijth *element* of A. So, $A_i^j = (A_i)^j = (A^j)_i$. A matrix A is called *nonnegative*, written $A \geqslant 0$, if all of its elements are nonnegative.

Let A be a nonnegative matrix in $G^{m \times n}$. We define the *nonnegative column rank* of A, denoted c-rank$_+(A)$, as the smallest nonnegative integer q for which there exist nonnegative (column) vectors $v^1, v^2, \dots, v^q$ in G^m such that each column of A has a representation as a linear combination with nonnegative coefficients of $v^1, v^2, \dots, v^q$ (following standard convention, we define the empty sum as zero). The *nonnegative row rank* of A, denoted r-rank$_+(A)$, is defined as the nonnegative column rank of A^T, the *transpose* of A.

LEMMA 2.1. *Let A be a nonnegative matrix in $G^{m\times n}$, and let q be a nonnegative integer. Then the following are equivalent:*

(a) $q \geqslant$ c-rank$_+(A)$,
(b) $q \geqslant$ r-rank$_+(A)$,
(c) *there exist two nonnegative matrices $V \in G^{m\times q}$ and $U \in G^{q\times n}$ such that $A = VU$, and*
(d) *there exist nonnegative vectors $v^1, v^2, \dots, v^q$ in G^m and vectors $u^1, u^2, \dots, u^q$ in G^n such that $A = \sum_{t=1}^{q} v^t(u^t)^T$.*

Proof. The case where $q = 0$ is trivial; thus, assume that $q \geqslant 1$.

(a) $\Rightarrow$ (c): Assume that (a) holds. By possibly augmenting the spanning vectors in an arbitrary way we have that there exist vectors $v^1, v^2, \dots, v^q$ in G^m such that each column of A is a linear combination with nonnegative coefficients of these vectors. For $j = 1, 2, \dots, n$, let u^j be the vector whose coordinates are the nonnegative coefficients corresponding to A^j, i.e., $A^j = \sum_{t=1}^{u}(u^j)_t v^t$. Let $V \in G^{m\times q}$ and $U \in G^{q\times n}$ be the matrices whose columns are $v^1, v^2, \dots, v^q$ and $u^1, u^2, \dots, u^n$, respectively. Then V and U are nonnegative, and for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$,

$$(VU)_i^j = \sum_{t=1}^{q} V_i^t U_t^j = \sum_{t=1}^{q} (u^j)_t (v^t)_i = \left(\sum_{t=1}^{q} (u^j)_t v^t\right)_i = (A^j)_i = A_i^j.$$

So $VU = A$ and (c) follows.

(c) $\Rightarrow$ (a): Suppose that (c) holds and $V \in G^{m\times q}$ and $U \in G^{q\times n}$ are nonnegative matrices such that $A = VU$. Then, for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$,

$$A_i^j = (VU)_i^j = \sum_{t=1}^{q} V_i^t U_t^j = \left(\sum_{t=1}^{q} U_t^j V^t\right)_i,$$

i.e., $A^j = \sum_{t=1}^{q} U_t^j V^t$. So each column of A is a linear combination with nonnegative coefficients of the nonnegative vectors $V^1, V^2, \dots, V^q$.

The equivalence (b) $\Leftrightarrow$ (c) follows by applying the established equivalence of (a) and (c) to A^T and from the fact that $A = VU$ if and only if $A^T = U^T V^T$. Finally, the equivalence (c) $\Leftrightarrow$ (d) follows from the fact that if $V \in G^{m\times q}$ and $U \in G^{q\times n}$, then $VU = \sum_{t=1}^{q} V^t U_t$. ■

COROLLARY 2.2. *Let A be a nonnegative matrix in $G^{m\times n}$ and let q be a nonnegative integer. Then the following are equivalent:*

(a) $q = \text{c-rank}_+(A)$,
(b) $q = \text{r-rank}_+(A)$,
(c) *q is the smallest integer for which there exist two nonnegative matrices $V \in G^{m\times q}$ and $U \in G^{q\times n}$ such that $A = VU$, and*
(d) *q is the smallest integer for which there exist nonnegative vectors $v^1, v^2, \dots, v^q$ in G^m and vectors $u^1, u^2, \dots, u^q$ in G^n such that $A = \sum_{t=1}^{q} v^t(u^t)^T$.*

Given a nonnegative matrix A in $G^{m\times n}$, we define the *nonnegative rank* of A, denoted $\text{rank}_+(A)$, as the integer q for which the four equivalent conditions of the above corollary apply. Gregory and Pullman (1983) use (c), in the context of semirings, to define the nonnegative rank; further, they observe the equivalence of the four conditions of Corollary 2.2.

When we rely on condition (c), we refer to a representation of a nonnegative matrix A of the form $A = VU$, where V and U are nonnegative matrices, as a *nonnegative factorization* of A. When we rely on condition (d), we refer to a representation of the nonnegative matrix A of the form $A = \sum_{t=1}^{q} v^t(u^t)^T$, where $v^1, v^2, \dots, v^q$ and $u^1, u^2, \dots, u^q$ are nonnegative vectors, as a *nonnegative rank-one decomposition* of A.

Herbert Robbins (private communications) gave easily computable lower and upper bounds on the nonnegative rank of a matrix; see also Gregory and Pullman (1983).

LEMMA 2.3 (H. Robbins). *Let A be a nonnegative matrix in $G^{m\times n}$. Then* $\text{rank}(A) \leqslant \text{rank}_+(A) \leqslant \min(m, n)$.

Let $A \in G^{m\times n}$ be nonnegative. We call two entries A_i^j and A_k^p of A *independent* if $A_i^j A_k^p > 0$ and $A_i^p A_k^j = 0$. The following observation by an anonymous referee provides lower bounds on the nonnegative rank. It is easily verified from the characterization of nonnegative ranks via nonnegative rank-one decompositions.

LEMMA 2.4. *Let $A \in G^{m\times n}$ be nonnegative. If A contains a set of q pairwise independent entries, then* $\text{rank}_+(A) \geqslant q$.

The above definition of independence is stronger than the one given by Gregory and Pullman (1983, p. 225), and the inequality of Lemma 2.4 is in the reverse direction from the one in their Lemma 1.2.

It is shown in Section 4 that if a matrix has fewer than four rows or fewer than four columns, then its rank and nonnegative rank coincide. H. Robbins (private communications) constructed the following 4×4 real matrix whose rank differs from its nonnegative rank.

EXAMPLE (H. Robbins). Let

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

It is straightforward to determine that $\operatorname{rank}(A) = 3$. Also, Lemma 2.3 implies that $\operatorname{rank}_+(A) \leqslant 4$. Finally, as the entries A_1^1, A_2^3, A_3^2, and A_4^4 are pairwise independent, Lemma 2.4 implies that $\operatorname{rank}_+(A) \geqslant 4$.

We next give bounds on the nonnegative ranks of outcomes of some matrix operations.

LEMMA 2.5. *Let A and B be nonnegative matrices in $G^{m \times n}$. Then*

(a) $\operatorname{rank}_+(A) = \operatorname{rank}_+(A^T)$, *and*
(b) $\operatorname{rank}_+(A + B) \leqslant \operatorname{rank}_+(A) + \operatorname{rank}_+(B)$.

Proof. Parts (a) and (b) are immediate from the characterizations of the nonnegative rank via nonnegative factorization and nonnegative rank-one decomposition, respectively. ∎

LEMMA 2.6. *Let A and B be nonnegative matrices in $G^{m \times s}$ and $G^{s \times n}$, respectively. Then*

$$\operatorname{rank}_+(AB) \leqslant \min\{\operatorname{rank}_+(A), \operatorname{rank}_+(B)\}.$$

Proof. Let $\operatorname{rank}_+(A) = q$ and $\operatorname{rank}_+(B) = q'$. Then there exist factorizations of A and B, respectively, of the form $A = UV$ and $B = U'V'$ where $U \in G^{m \times q}$, $V \in G^{q \times s}$, $U' \in G^{s \times q'}$, and $V' \in G^{q' \times n}$ are nonnegative matrices. Then the factorizations $AB = U(VU'V') = (UVU')V'$ show that $\operatorname{rank}_+(AB) \leqslant \min\{q, q'\} = \min\{\operatorname{rank}_+(A), \operatorname{rank}_+(B)\}$. ∎

LEMMA 2.7. *Let* $A(1)$ *and* $A(2)$ *be nonnegative matrices in* $G^{m\times n(1)}$ *and* $G^{m\times n(2)}$, *respectively, and consider the matrix* $A = [A(1), A(2)] \in G^{m\times[n(1)+n(2)]}$. *Then*

$$\max\{\mathrm{rank}_+[A(1)], \mathrm{rank}_+[A(2)]\} \leqslant \mathrm{rank}_+(A) \leqslant \mathrm{rank}_+[A(1)] + \mathrm{rank}_+[A(2)].$$

Corresponding inequalities holds for "row partitions."

A *positive diagonal scaling* of a matrix $A \in G^{m\times n}$ is a matrix $B \in G^{m\times n}$ having a representation $B = DAE$ where D and E are, respectively, $m \times m$ and $n \times n$ diagonal matrices having positive diagonal elements.

LEMMA 2.8. *Let A be a nonnegative matrix in* $G^{m\times n}$, *and let B be a positive diagonal scaling of A. Then* $\mathrm{rank}_+(A) = \mathrm{rank}_+(B)$.

Proof. Suppose B has the representation $B = DAE$ where D and E are, respectively, $m \times m$ and $n \times n$ diagonal matrices having positive diagonal elements. As $B = (DA)E$, Lemma 2.6 implies that $\mathrm{rank}_+(B) \leqslant \mathrm{rank}_+(DA) \leqslant \mathrm{rank}_+(A)$. Also, as $A = D^{-1}BE^{-1}$, a symmetric argument shows the reverse inequality $\mathrm{rank}_+(A) \leqslant \mathrm{rank}_+(B)$. ∎

LEMMA 2.9. *Let* $D \in G^{n\times n}$ *be a nonnegative diagonal matrix. Then* $\mathrm{rank}_+(D) = \mathrm{rank}(D)$.

Proof. By Lemma 2.3, $\mathrm{rank}_+(D) \geqslant \mathrm{rank}(D)$. Next, for $t = 1, \ldots, m$ let e^t be the tth unit vector in G^n. Then $D = \sum_{\{i:\, D_i^i > 0\}} D_i^i(e^i)(e^i)^T$, and this decomposition shows that $\mathrm{rank}_+(D) \leqslant \mathrm{rank}(D)$. ∎

A nonnegative matrix A is defined to be *row-allowable* if each row of A contains at least one positive element. A nonnegative matrix A is defined to be *nondegenerate* if both A and A^T are row-allowable.

LEMMA 2.10. *Let* $A \in G^{m\times n}$ *be a nonnegative, row-allowable (respectively, nondegenerate) matrix with* $q \equiv \mathrm{rank}_+(A)$. *If* $A = VU$ *where* $V \in G^{m\times q}$ *and* $U \in G^{q\times m}$ *are nonnegative matrices, then V and U are both row-allowable (nondegenerate).*

3. REDUCTION TO BIVARIATE PROBABILITY MATRICES AND STOCHASTIC MATRICES

A *probability vector* is a nonnegative vector with element sum 1. A nonnegative matrix is called a *bivariate probability matrix* if its element sum is 1. A nonnegative matrix is called *row-stochastic* (*column-stochastic*) if the element sum of each of its rows (columns) is 1. For each positive integer k, let $e^{(k)}$ be the vector in G^k in which all coordinates are 1. Then $r \in G^k$ is a probability vector if and only if $r \geqslant 0$ and $r^T e^{(k)} = 1$. Also, a nonnegative matrix $P \in G^{m \times n}$ is a bivariate probability matrix, a row-stochastic matrix, or a column-stochastic matrix if and only if $e^{(m)T} P e^{(n)} = 1$, $P e^{(n)} = e^{(m)}$, or $[e^{(m)}]^T P = [e^{(n)}]^T$, respectively. A bivariate probability matrix is called *independent* if $P = rs^T$ for some probability vectors $r \in G^m$ and $s \in G^n$.

Consider a nonnegative nonzero matrix $A \in G^{m \times n}$. By dividing the matrix A by the sum of its elements, we obtain a nonnegative matrix P which is a bivariate probability matrix. Further, as P is a scaling of A, Lemma 2.8 assures that $\operatorname{rank}_+(P) = \operatorname{rank}_+(A)$. Also, by dropping the zero rows of A and by dividing each remaining row by its element sum we obtain a row-stochastic matrix S. As dropping zero rows of a matrix and the division by rows by positive scalars preserve the nonnegative rank (Lemma 2.8), we have that $\operatorname{rank}_+(S) = \operatorname{rank}_+(A)$. Finally, by dropping the zero columns of A and dividing each of the remaining columns by its element sum, we obtain a column-stochastic matrix T with $\operatorname{rank}_+(T) = \operatorname{rank}_+(A)$. So, when examining nonnegative ranks, we can restrict attention to any of the following classes: bivariate probability matrices, row-stochastic matrices, or column-stochastic matrices.

The next two results provide useful modifications of the decomposition and factorization criteria of nonnegative ranks which apply, respectively, to bivariate probability matrices and to row- and column-stochastic matrices.

THEOREM 3.1. *Let $P \in G^{m \times n}$ be a bivariate probability matrix. Then* $\operatorname{rank}_+(A)$ *is the smallest nonnegative integer p such that P can be expressed as a convex combination of p independent bivariate probability matrices.*

Proof. Let $q \equiv \operatorname{rank}_+(P)$, and let p be the smallest nonnegative integer such that P can be expressed as a convex combination of p independent bivariate probability matrices. Then, trivially, $q \leqslant p$. To see the reverse inequality, consider a representation of P as $\sum_{t=1}^{q} v^t (u^t)^T$ where $v^1, v^2, \dots, v^q$ and $u^1, u^2, \dots, u^q$ are nonnegative vectors in G^m and G^n, respectively. Then none of the vectors $v^1, v^2, \ldots, v^q, u^1, u^2, \ldots, u^q$ is the zero vector and

$$1 = \sum_{i=1}^{m} \sum_{j=1}^{n} P_i^j = [e^{(m)}]^T P [e^{(n)}] = \sum_{t=1}^{q} \left\{[e^{(m)}]^T v^t\right\}\left\{(u^t)^T [e^{(n)}]\right\}.$$

For $t = 1, \ldots, q$, let $\beta_t \equiv \{[e^{(m)}]^T v^t\}\{(u^t)^T[e^{(n)}]\}$, $v'^t \equiv v^t/\{(v^t)^T[e^{(n)}]\}$, and $u'^t \equiv u^t/[e^{(m)}]^T u^t$. Then the representation $P = \sum_{t=1}^{q} v^t(u^t)^T = \sum_{t=1}^{q} \beta_t(v'^t)(u'^t)^T$ proves that $p \leqslant \text{rank}_+(P) = q$. ■

THEOREM 3.2. *Let $P \in G^{m \times n}$ be a row-stochastic (column-stochastic) matrix. Then* $\text{rank}_+(P)$ *is the smallest nonnegative integer q such that there exist row-stochastic (column-stochastic) matrices $R \in G^{m \times q}$ and $S \in G^{q \times n}$ where $P = RS$.*

Proof. We consider only the case where P is row-stochastic. Let p be the smallest nonnegative integer such that $P = RS$ for some row-stochastic matrices $R \in G^{m \times p}$ and $S \in G^{p \times n}$. Also, let $q \equiv \text{rank}_+(P)$. Then, trivially, $q \leqslant p$. To see the converse inequality, let $V \in G^{m \times q}$ and $U \in G^{q \times n}$ be nonnegative matrices for which $P = VU$. By the minimality of q, U has no zero rows. Consider the diagonal matrix $D \in G^{q \times q}$ whose diagonal elements are the corresponding row sums of U. Then D is nonsingular and $D^{-1}U$ is row-stochastic. Further, as P is row-stochastic, we have that $e^{(m)} = Pe^{(n)} = VUe^{(n)} = (VD)(D^{-1}U)e^{(n)} = (VD)e^{(q)}$, showing that VD is row-stochastic. So, $R \equiv VD \in G^{m \times q}$ and $S \equiv D^{-1}U \in G^{q \times n}$ are row-stochastic matrices that satisfy $P = RS$. Hence, $p \leqslant \text{rank}_+(P) = q$. ■

Let $P \in G^{m \times n}$ be a column-stochastic matrix. The problem of finding column-stochastic matrices $R \in G^{m \times q}$ and $S \in G^{q \times n}$ where $P = RS$ amounts to finding q probability vectors in G^m, say $r^1, r^2, \ldots, r^q$ (corresponding to the columns of R), and n probability vectors in G^q, say $s^1, s^2, \ldots, s^n$ (corresponding to the columns of S), such that for $j = 1, 2, \ldots, n$, $P^j = \sum_{t=1}^{q}(s^j)_t r^t$, i.e., P^j is a convex combination of $r^1, r^2, \ldots, r^q$. So the problem of computing the nonnegative rank and a corresponding nonnegative factorization of an $m \times n$ column-stochastic matrix P reduces to finding the smallest number of probability vectors in G^m such that each of the probability vectors $P^1, P^2, \ldots, P^n$ can be expressed as a convex combination of these vectors. As the set of probability vectors in G^m is the convex hull of the m unit vectors, the problem can be formulated in the following broader geometric perspective:

INSCRIBING-POLYTOPE PROBLEM. Given vectors $a^1, a^2, \ldots, a^n, b^1, b^2, \ldots, b^p$ in G^m, where each of the vectors $b^1, b^2, \ldots, b^p$ is contained in the convex hull of $a^1, a^2, \ldots, a^n$, find the smallest number of vectors $c^1, c^2, \ldots, c^q$ such that $\text{conv}\{a^1, a^2, \ldots, a^n\} \supseteq \text{conv}\{c^1, c^2, \ldots, c^q\} \supseteq \text{conv}\{b^1, b^2, \ldots, b^p\}$. Further, find the representation of each of the b^t's as a convex combination of the c^t's.

4. COMPUTING THE NONNEGATIVE RANK AND CORRESPONDING FACTORIZATIONS WHEN THE RANK IS 2 OR LESS

THEOREM 4.1. *Let* $A \in G^{m \times n}$ *be a nonnegative matrix with* $\text{rank}(A) \leqslant 2$. *Then* $\text{rank}_+(A) = \text{rank}(A)$.

Proof. If $\text{rank}(A) = 0$, A is the zero matrix and the asserted equality is trivial.

Next assume that $\text{rank}(A) = 1$. Then $A = vu^T$ where $v \in G^m$ and $u \in G^n$. So, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, we have $A_i^j = v_i u_j$, and the nonnegativity of A implies that $A_i^j = |A_i^j| = |v_i|\,|u_j|$. So $A = |v|\,|u|^T$ for the vectors $|v| \in G^m$ and $|u| \in G^n$ whose coordinates are the absolute values of the coordinates of v and u, respectively.

Finally, assume that $\text{rank}(A) = 2$. As discussed in Section 3, we may assume that A is column-stochastic. So the columns of A are probability vectors. As $\text{rank}(A) = 2$, there exist two vectors in G^m such that each column of A is a linear combination of these vectors. After possible rearrangement of the columns of A, standard results from linear algebra show that we may assume that these vectors are A^1 and A^2. For $j = 1, \ldots, n$, let A^j have the representation $A^j = \alpha_j A^1 + \beta_j A^2$. As A^1 and A^2 are probability vectors,

$$1 = [e^{(m)}]^T A^j = \alpha_j [e^{(m)}]^T A^1 + \beta_j [e^{(m)}]^T A^2 = \alpha_j + \beta_j,$$

implying that $\alpha_j = 1 - \beta_j$. Now, let $\beta^* \equiv \max\{\beta_j : j = 1, 2, \ldots, n\}$, let $\beta_* \equiv \min\{\beta_j : j = 1, 2, \ldots, n\}$, and let i^* and i_* be the corresponding indices. Then

$$A^{i^*} = (1 - \beta^*) A^1 + \beta^* A^2 \quad \text{and} \quad A^{i_*} = (1 - \beta_*) A^1 + \beta_* A^2.$$

Further, for $j = 1, 2, \ldots, n$, $\beta_* \leqslant \beta_j \leqslant \beta^*$, implying that there exists a scalar δ_j with $0 \leqslant \delta_j \leqslant 1$ such that $\beta_j = (1 - \delta_j)\beta_* + \delta_j \beta^*$. Thus,

$$\begin{aligned} A^j &= (1 - \beta_j) A^1 + \beta_j A^2 \\ &= \{1 - [(1 - \delta_j)\beta_* + \delta_j \beta^*]\} A^1 + [(1 - \delta_j)\beta_* + \delta_j \beta^*] A^2 \\ &= \{(1 - \delta_j)(1 - \beta_*) + \delta_j(1 - \beta^*)\} A^1 + [(1 - \delta_j)\beta_* + \delta_j \beta^*] A^2 \\ &= (1 - \delta_j)[(1 - \beta_*) A^1 + \beta_* A^2] + \delta_j[(1 - \beta^*) A^1 + \beta^* A^2] \\ &= (1 - \delta_j) A^{i_*} + \delta_j A^{i^*}. \end{aligned}$$

As A^{i_*} and A^{i^*} are columns of the original matrix A, they are nonnegative. So each column of A is a convex combination of these two nonnegative vectors, and therefore $\text{rank}_+(A) \leqslant 2 = \text{rank}(A)$. The reverse inequality is given in Lemma 2.3. ■

As the rank of a matrix is easily computable by standard algorithms (e.g., Gaussian elimination), matrices having rank two or less are easy to identify. Theorem 4.1 shows that the nonnegative rank of such matrices equals their rank; hence, it is easy to determine. Further, our proof of Theorem 4.1 provides an efficient computational method for determining the corresponding nonnegative factorization (over arbitrary ordered fields) for matrices in this restrictive class.

COROLLARY 4.2. *Let $A \in G^{m \times n}$ be a nonnegative matrix. If either $m \in \{1, 2, 3\}$ or $n \in \{1, 2, 3\}$, then* $\text{rank}_+(A) = \text{rank}(A)$.

Proof. If $\min\{m, n\} \leqslant 2$, Theorem 4.1 applies. The only remaining case is when $m = n = 3$. if $\text{rank}(A) = 2$, Theorem 4.1 again applies. If $\text{rank}(A) = 3 = m = n$, Lemma 2.3 applies. ■

5. COMPUTING THE NONNEGATIVE RANK OVER THE REALS

In this section we describe a procedure for computing the nonnegative rank of an arbitrary nonnegative matrix over the reals. The procedure employs a quantifier elimination algorithm for first-order formulae over the reals.

Let $A \in R^{m \times n}$ be a nonnegative matrix. Lemma 2.3 shows that rank $(A) \leqslant \text{rank}_+(A) \leqslant \min\{m, n\}$. Hence, to compute the nonnegative rank of A it suffices to check for $q = \text{rank}(A), \text{rank}(A) + 1, \dots, \min\{m, n\}$ whether $\text{rank}_+(A) \leqslant q$. By using a bisection procedure, the number of required tests is of the order $\log\{\min\{m, n\} - \text{rank}(A)\}$.

From condition (c) of Lemma 2.1, deciding whether $\text{rank}_+(A) \leqslant q$ for a positive integer q is equivalent to checking whether there exist nonnegative matrices $V \in R^{m \times q}$ and $U \in R^{q \times n}$ such that $VU = A$, i.e., testing the feasibility of the (nonlinear) system given by

$$\sum_{k=1}^{q} x_{ik} y_{kj} = A_i^j, \tag{5.1}$$

$$x_{ik} \geqslant 0 \qquad \text{for} \quad i = 1, \dots, m \quad \text{and} \quad k = 1, \dots, q, \tag{5.2}$$

and

$$y_{kj} \geqslant 0 \qquad \text{for} \quad k = 1, \dots, q \quad \text{and} \quad j = 1, \dots, n. \tag{5.3}$$

The variables in this system are the x_{ik}'s and y_{kj}'s. The A_i^j's constitute the data.

The solution of a nonlinear system like (5.1)–(5.3) is not easy. Still, as (5.1)–(5.3) is a system of polynomial equations and inequalities over the reals, it is possible to eliminate the existential quantifiers that assert feasibility of the system and obtain an (alternative) set of polynomial equations and inequalities in the data that will characterize the feasibility of (5.1)–(5.3). Tarski (1951), Seidenberg (1954), Cohen (1969), Collins (1969), and Renegar (1992) show that when a problem is presented by polynomial equations and inequalities that are tied by connectors like $\vee$ ("or"), $\wedge$ ("and"), $\rightarrow$ ("implies"), $\leftrightarrow$ ("is equivalent"), and $\neg$ ("negation") and by quantifiers like $\forall$ (the universal quantifier "for all") or $\exists$ (the existential quantifier "for some"), it is possible to obtain an equivalent problem which has no quantifiers. The operations that are required to eliminate the quantifiers are restricted to the five elementary operations of ordered fields—additions, subtractions, multiplications, divisions and comparisons—and to the evaluation of Boolean functions. Renegar (1992) surveys the results about the complexity of quantifier elimination over the reals and introduces a new method for quantifier elimination.

We next summarize the complexity of the quantifier elimination algorithm of Renegar (1992). We then apply his results to the problem of determining the solvability of (5.1)–(5.3). Consider a first-order formula over the reals having the form

$$\left(Q_1 x^{[1]} \in R^{n_1}\right) \cdots \left(Q_\omega x^{[\omega]} \in R^{n_\omega}\right) P\left(y, x^{[1]}, \ldots, x^{[\omega]}\right), \tag{5.4}$$

where each Q_k is one of the two quantifiers $\exists$ ("there exists") or $\forall$ ("for all"), where $y = (y_1, \ldots, y_{n_0})$ are free (unquantified) variables and where $P(y, x^{[1]}, \ldots, x^{[\omega]})$ is a quantifier-free Boolean formula. The atomic predicates of $P(y, x^{[1]}, \ldots, x^{[\omega]})$ are assumed to have the form $g_i(y, x^{[1]}, \ldots, x^{[\omega]}) \, \Delta_i \, 0$, $i = 1, \ldots, M$, where $g_i : \prod_{k=0}^{\omega} R^{n_k} \to R$ is a polynomial of degree at most $d \geqslant 2$ and Δ_i is any one of the standard relations $\geqslant$, $>$, $=$, $\neq$, $\leqslant$, and $<$, and $P(\cdot)$ is determined by a Boolean function $\mathbb{P} : \{0, 1\}^M \to \{0, 1\}$ and a function $B : R^{n_0} \times \prod_{k=1}^{\omega} R^{n_k} \to \{0, 1\}$, where for $y \in R^{n_0}$ and $x \in \prod_{k=1}^{\omega} R^{n_k}$, $P(y, x) \equiv \mathbb{P}[B(y, x)]$, and for $i = 1, \ldots, M$,

$$B(y, x)_i \equiv \begin{cases} 1 & \text{if } g_i(y, x) \, \Delta_i \, 0, \\ 0 & \text{otherwise.} \end{cases}$$

The quantifier elimination algorithm of Renegar (1992) requires at most $(Md)^{2^{O(\omega)}\prod_k n_k}$ multiplications and additions, and at most $(Md)^{O(\sum_k n_k)}$ calls to $\mathbb{P}$. The method requires no divisions. The method can be implemented in parallel, requiring at most $[2^{\omega}(\prod_k n_k)\log(Md)]^{O(1)}$ additions and multiplications on each of $(Md)^{2^{O(\omega)}\prod_k n_k}$ processors that execute such operations and $(Md)^{O(\sum_k n_k)}$ parallel calls to $\mathbb{P}$ by $N(Md)^{O(\sum_k n_k)}$ corresponding processors, where each call to $\mathbb{P}$ is executed by N parallel processors (where N is any positive integer).

When restricted to formulae involving only polynomials with integer coefficients of bit length at most L, the algorithm becomes a bit-model quantifier elimination method requiring at most $L(\log L)(\log\log L)$ $(Md)^{2^{O(\omega)}\prod_k n_k}$ sequential bit operations and $(Md)^{O(\sum_k n_k)}$ calls to $\mathbb{P}$. When implemented in parallel the algorithm requires at most $(\log L)[2^{\omega}(\prod_k n_k)$ $\log(Md)]^{O(1)}$ sequential bit operations on each of $L^2(Md)^{2^{O(\omega)}\prod_k n_k}$ processors that execute such operations and $(Md)^{O(\sum_k n_k)}$ parallel calls to $\mathbb{P}$ by $N(Md)^{O(\sum_k n_k)}$ corresponding processors, where each call to $\mathbb{P}$ is executed by N parallel processors (where N is any positive integer).

The quantifier elimination method constructs a quantifier-free formula of the following simple form:

$$\bigvee_{i=1}^{I} \bigwedge_{j=1}^{J_i} h_{ij}(y) \, \Delta_{ij} \, 0,$$

where $I \leqslant (Md)^{2^{O(\omega)}\Pi_k n_k}$, where $J_i \leqslant (Md)^{2^{O(\omega)}\Pi_k n_k}$ for each $i = 1, \ldots, I$, where the degree of each of the polynomials h_{ij} is at most $(Md)^{2^{O(\omega)}\Pi_k n_k}$, and where each Δ_{ij} is one of the standard relations $\geqslant$, $>$, $=$, $\neq$, $\leqslant$, and $<$. If the coefficients of each g_i are all integers of bit length L, the coefficients of each of the polynomials h_{ij} will be integers of bit length at most $(L + n_0)(Md)^{2^{O(\omega)}\Pi_k n_k}$.

The system (5.1)–(5.3) has $(m + n)q$ quantified variables (the x_{ij}'s and y_{ij}'s) and mn unquantified variables (the A_i^j's). Further, (5.1) consists of mn polynomial equations where each polynomial has degree two, and (5.2)–(5.3) consist of $q(m + n)$ nonnegativity constraints which are polynomial inequalities where the degree of the corresponding polynomial is one. Testing for the feasibility of (5.1)–(5.3) can be cast via a formula of the type given in (5.4) with

$$\begin{aligned} \omega &= 1, \\ n_0 &= mn, \\ n_1 &= q(m + n) \leqslant 2mn, \\ M &= mn + q(m + n) \leqslant 3mn, \\ d &= 2, \end{aligned}$$

where $\mathbb{P} : \{0, 1\}^{mn + q(m+n)} \to \{0, 1\}$ is the Boolean function defined for $u \in \{0, 1\}^{mn + q(m+n)}$ by

$$\mathbb{P}(u) = \begin{cases} 1 & \text{if} \quad u_i = 1 \text{ for all } 1 \leqslant i \leqslant mn + q(m + n), \\ 0 & \text{otherwise.} \end{cases}$$

The evaluation of $\mathbb{P}$ is simple, as it will not always require the evaluation of all the coordinates of its argument. Further, the evaluation of $\mathbb{P}$ can be executed in parallel by $mn + q(m + n) \leqslant 3mn$ parallel univariate Boolean processors.

As usual, the notation $O(\cdot)$ denotes an arbitrary real-valued function over the reals (or over the integers) such that for some positive number K, $|O(x)| \leqslant K|x|$ for every x in its domain.

Renegar's complexity results yield the following theorem:

THEOREM 5.1. *Let $n \geqslant 1$. There is an algorithm for determining the feasibility of* (5.1)–(5.3) *that requires at most*

$$(6mn)^{2^{O(1)}m^2n^2} \text{ multiplications and additions}$$

and

$$(6mn)^{O(mn)} \text{ calls to } \mathbb{P}.$$

The method requires no divisions. The method can be implemented in parallel, requiring at most $[4m^2n^2 \log(6mn)]^{O(1)}$ *additions and multiplications on each of* $(6mn)^{2^{O(1)}m^2n^2}$ *processors that execute such operations and* $3mn(6mn)^{O(mn)}$ *univariate Boolean processors.*

When restricted to formulae involving only polynomials with integer coefficients of bit length at most L, the algorithm becomes a bit-model quantifier elimination method requiring at most

$$L(\log L)(\log\log L)(6mn)^{2^{O(1)}m^2n^2} \text{ sequential bit operations}$$

and

$$(6mn)^{O(mn)} \text{ calls to } \mathbb{P}.$$

When implemented in parallel the algorithm requires at most $(\log L)[4m^2n^2 \log(6mn)]^{O(1)}$ *sequential bit operations on each of* $L^2(6mn)^{2^{O(1)}m^2n^2}$ *processors that execute such operations and* $3mn(6mn)^{O(mn)}$ *univariate Boolean processors.*

The quantifier elimination method constructs a quantifier-free formula

$$\bigvee_{i=1}^{I} \bigwedge_{j=1}^{J_i} [h_{ij}(y)\, \Delta_{ij}\, 0],$$

where

$$I \leqslant (6mn)^{2^{O(1)}m^2n^2},$$

$$J_i \leqslant (6mn)^{2^{O(1)}m^2n^2},$$

$h_{ij}: R^{m\times n} \to R$ *is a polynomial whose degree is at most* $(6mn)^{2^{O(1)}m^2n^2}$, *and* Δ_{ij} *is one of the standard relations* $\geqslant$, $>$, $=$, $\neq$, $\leqslant$, *and* $<$. *If the coefficients of each* g_i *are all integers of bit length L, the coefficients of each of the polynomials* h_{ij} *will be integers of bit length at most* $L(6mn)^{2^{O(1)}m^2n^2}$.

The complexity of an algorithm that relies on Renegar's method will be obtained by multiplying the complexity asserted in Theorem 5.1 by

$\log[\min\{m, n\} - \operatorname{rank}(A)]$, the number of times that the feasibility of (5.1)–(5.3) has to be tested in order to compute $\operatorname{rank}_+(A)$.

The solvability of (5.1)–(5.3) depends on the ordered field over which the question is posed, not just on the ordered field which contains the A_i^j's. For example, when the A_i^j's are rational numbers, the question about the feasibility of (5.1)–(5.3) when the x_{ij}'s and the y_{ij}'s are required to be rationals and the related question when the x_{ij}'s and y_{ij}'s can be arbitrary real numbers could have different answers. Renegar's algorithm, like all other known methods for eliminating quantifiers, is applicable over the reals and not over the rationals. When all the elements of the given matrix are rationals, these methods determine feasibility of (5.1)–(5.3) over the reals, but not necessarily over the rationals.

How sensitive is a question to the ordered field over which it is asked? Any condition (as in Corollary 2.2) used to define the nonnegative rank of a matrix requires the specification of the ordered field over which it is applied. We pose an open problem:

PROBLEM. Show that the nonnegative ranks of a rational matrix over the reals and over the rationals coincide, or provide an example where the two ranks are different.

Section 4 shows that if $\operatorname{rank}(A) = 2$, the nonnegative ranks of A over all ordered fields coincide.

An ordered field G is called *real closed* [e.g., Jacobson (1964, pp. 273–277)] if every positive element of G has a square root and any polynomial of odd degree with coefficients in G has a root in G. For example, the reals are a real closed field. Also, every ordered field has an extension which is real closed; see Jacobson (1964, p. 285).

Seidenberg (1954) observed that the quantifier elimination method of Tarski (1951) applies to all real closed fields. He concluded that two formulae are equivalent over the reals if and only if they are equivalent over all real closed fields. It follows from this observation, known as *Tarski's principle*, that every quantifier elimination method over the reals, including Renegar's algorithm, applies to all real closed fields. So Renegar's algorithm can be used to compute nonnegative ranks of matrices over any real closed field with the complexity stated in Theorem 5.1. Of course, the results about bit operations are not applicable if the given matrix contains noninteger coefficients. Eaves and Rothblum (1989) discuss the use of an algorithm for solving a problem over one real closed field to solve corresponding problems over arbitrary real closed fields.

6. APPLICATIONS

The problem of expressing a nonnegative matrix, exactly or approximately, as a sum of nonnegative matrices of rank 1 arises in several sciences. We list several examples.

In quantum mechanics, Suppes and Zanotti (1981) propose that if X and Y are two correlated random variables, then a random variable Z "explains" the correlation between X and Y if, conditional on Z, X and Y are independent. Suppose X takes m distinct values $i = 1, \dots, m$, Y takes n distinct values $j = 1, \dots, n$, and Z takes p values $k = 1, \dots, p$. Let $P \in R^{m \times n}$ represent the joint probability distribution of X and Y, and let $x \in R^m$, $y \in R^n$, and $z \in R^p$ represent, respectively, the probability distributions of X, Y, and Z, i.e., for relevant values of i, j, and k, $P_i^j = \text{Prob}\{X = i$ and $Y = j\}$, $x_i = \text{Prob}\{X = i\} = \sum_{j=1}^n P_i^j$, $y_j = \text{Prob}\{Y = j\} = \sum_{i=1}^m P_i^j$, and $z_k = \text{Prob}\{Z = k\}$. Also, for $k = 1, \dots, p$, let $x^{(k)} \in R^m$ and $y^{(k)} \in R^n$ be, respectively, the conditional probability vectors of X and Y when $\{Z = k\}$, i.e., $x_i^{(k)} = \text{Prob}\{X = i \mid Z = k\}$ and $y_j^{(k)} = \text{Prob}\{Y = j \mid Z = k\}$. The X and Y are conditionally independent given Z if for all relevant values of i, j, and k,

$$\text{Prob}\{X = i \text{ and } Y = j | Z = k\} = \text{Prob}\{X = i | Z = k\} \times \text{Prob}\{Y = j | Z = k\} = x_i^{(k)} y_j^{(k)},$$

in which case

$$P_i^j = \sum_{k=1}^p \text{Prob}\{Z = k\} \text{Prob}\{X = i \text{ and } Y = j | Z = k\} = \sum_{k=1}^p z_k \left\{ x^{(k)} [y^{(k)}]^T \right\}_i^j,$$

i.e.,

$$P = \sum_{k=1}^p z_k \left\{ x^{(k)} [y^{(k)}]^T \right\}. \tag{6.1}$$

Alternatively, assume that $P \in R^{m \times n}$ is a matrix having a representation as in (6.1), where z is a probability vector in R^p, $x^{(1)}, \dots, x^{(p)}$ are probability vectors in R^m, and $y^{(1)}, \dots, y^{(p)}$ are probability vectors in R^n. We show that in this case there exist random variables X, Y, and Z such that Z takes p values, the joint distribution of X and Y is given by P, and X and Y are conditionally independent given Z. Consider the sample space $\Omega \equiv \{1, \ldots, p\} \times \{1, \ldots, m\} \times \{1, \ldots, n\}$ with the probability distribution having $\text{Prob}\{(i, j, k)\} = z_k x_i^{(k)} y_j^{(k)}$. This definition yields a probability distribution as

$$\sum_{k=1}^{p} \sum_{i=1}^{m} \sum_{j=1}^{n} z_k x_i^{(k)} y_j^{(k)} = \sum_{k=1}^{p} z_k = 1.$$

Let the random variables X, Y, and Z be defined on Ω by

$$X(i,j,k) = i, \qquad Y(i,j,k) = j, \quad \text{and} \quad Z(i,j,k) = k.$$

Then, for $k = 1, \ldots, p$, $\Pr(Z = k) = \sum_{i=1}^{m} \sum_{j=1}^{n} z_k x_i^{(k)} y_j^{(k)} = z_k$, and for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, (6.1) implies that

$$\text{Prob}\{X = i \text{ and } Y = j\} = \sum_{k=1}^{p} \text{Prob}\{(i,j,k)\} = \sum_{k=1}^{p} z_k x_i^{(k)} y_j^{(k)} = P_i^j,$$

i.e., the joint distribution of X and Y is represented by P; further, if $z_k \neq 0$,

$$\text{Prob}\{X = i \text{ and } Y = j | Z = k\} = x_i^{(k)} y_j^{(k)}$$

$$= \text{Prob}\{X = i | Z = k\} \, \text{Prob}\{Y = j | Z = k\},$$

i.e., X and Y are conditionally independent given Z.

We have seen that an $m \times n$ nonnegative matrix P represents the joint probability distribution of a pair of random variables which are conditionally independent given a random variable taking p values, if and only if P has a representation as in (6.1) where z is a probability vector in R^p, $x^{(1)}, \ldots, x^{(p)}$ are probability vectors in R^m, and $y^{(1)}, \ldots, y^{(p)}$ are probability vectors in R^n. By Theorem 3.1, $\text{rank}_+(P)$ is the smallest integer p for which the latter can be accomplished. In the terminology of Suppes and Zanotti (1981), $\text{rank}_+(P)$ is the smallest support of a "hidden" random variable which explains the correlation between a pair of random variables whose joint distribution is represented by P.

In the demography of marriage [Henry (1969a, 1969b, 1972), Saboulin (1985)] men are classified into a finite number m of age categories, e.g., under age 25, 25–29, 30–34, etc., and women are similarly classified into a finite number of n of age categories. An $m \times n$ matrix A tabulates the ages of grooms and brides on marriage licenses issued during a certain time period, where A_i^j is the number of marriages between grooms in age category i and brides in age category j. Henry proposed that a marriage matrix is a sum of "panmictic components," which are nonnegative matrices each of rank 1. The rationale for this proposal is that a marriageable man aged 30, say, is surely indifferent to the exact age of a potential bride, over some range of age; other things being equal, he does not care if she is 28 years and 30 days or 28 years and 31 days old, and he may not care whether she is 28 or 29 years old, or even whether she is 22 or 32 years old. Hence, over some range he is willing to pick a bride of a given age in proportion to the frequency with which brides of that age are present in the marriageable population. Similarly, a marriageable woman is willing to pick a groom regardless of age, over some range of age. This indifference with respect to age implies that interactions between male and female age classes behave like the chemical law of mass action. Within a certain range of ages for brides and a certain range of ages for grooms, the number of marriages of men in category i with women in category j should be proportional to the product of the number of marriageable men in category i and the number of marriageable women in category j. Within these age ranges, for a certain circle of marriageable men and women, the marriage matrix should therefore be of rank one. This rank-one matrix for a restricted range of ages is called a *panmictic component*. A decomposition of the matrix A into a sum that has the smallest number of rank-one components defines the nonnegative rank as in Section 2.

In statistics, Herbert Robbins (personal communication on 10 September 1988) considered a problem arising in the analysis of contingency tables. A contingency table is an $m \times n$ matrix A in which element A_i^j is the number of occurrences of events of type (i, j). The question Robbins addressed was how to measure dependence within a contingency table when the occurrence of i is not independent of the occurrence of j. According to Robbins (10 September 1988), thirteen years previously, while talking to Michael Rabin at Yorktown Heights, New York, it occurred to Robbins to measure the amount of dependence in a contingency table A by seeing how closely A could be approximated, first, by a single nonnegative matrix of rank 1 (where an exact approximation is obtainable in the case of independence), then by a sum of two nonnegative matrices of rank 1, then by a sum of three nonnegative matrices of rank 1, and so on. The number of nonnegative summands of rank 1 required to approximate a given nonnegative matrix A "satisfactorily" measures the amount of dependence in A. Robbins (personal communication on 28 January 1985) began thinking about this problem in 1936, as a result of his work on tensor products in differential geometry. Breiman (1991) computed corresponding approximations that minimize the expected squared residuals.

Levin (1985) developed an algorithm that solves the following related problem: given a positive matrix A, find the maximal positive matrix B of rank 1 such that $B \leqslant A$. Here "maximal" means that if C is any positive matrix of rank 1 such that $C \leqslant A$, then $C \leqslant B$. So far, Levin has proved that his algorithm converges to the globally maximal B, and he gives B in closed form when $m = 2$.

Hayashi (1982, p. 78), apparently independently, considered the bilinear model $A_i^j = \sum_{t=1}^q x_i^t y_j^t$, where A_i^j is measured, and showed that it leads to a kind of principal-component analysis. He pointed out that Harshman (1970) considered an extension to trivariate distributions $A(i, j, k) = \sum_{t=1}^q x_i^t y_j^t z_k^t$. From this extension, a further generalization to measurements A that are dependent on n dimensions $i_1, i_2, \dots, i_n$ is obvious, and was developed independently by Breiman (1991).

Yannakakis (1988, p. 226) used linear programming to express combinatorial optimization problems. He showed that it is possible to express a polytope in R^N with m facets and n vertices by a linear program where the number of variables plus the number of constraints is bounded by the product of a (universal) constant and $\operatorname{rank}_+(A) + N$, where A is a certain nonnegative $m \times n$ matrix that depends on the representation of the polytope, its facets, and its vertices.

In algebraic complexity theory, Nisan (1991, p. 416) showed that for a homogeneous algebraic function f of degree d on N variables and for $k \in \{0, 1, \dots, d\}$, a certain quantity, called the k-monotone algebraic branching program complexity, is exactly $\operatorname{rank}_+(A)$, when A is a nonnegative $m \times n$ matrix, $m = N^k$, $n = N^{d-k}$, and the elements of A are the (nonnegative) coefficients of the monomials in the expansion of f.

We thank Alon Orlitsky for referring us to the work of Yannakakis (1988) and Nisan (1991), and Herbert Robbins and Bruce Levin for helpful comments. Also, we thank the referee for referring us to the work of Gregory and Pullman (1983) and for many helpful comments. The work of Joel E. Cohen was supported in part by U.S. National Science Foundation grant BSR 87-05047, and by the hospitality of Mr. and Mrs. William T. Golden. The work of Uriel G. Rothblum was partially supported by Office of Naval Research grant N00014-92-J1142 and Air Force grants AFOSR-89-0008 and AFOSR-90-0512 to Rutgers University.

REFERENCES

Berman, A., and Hershkowitz, D. 1987. Combinatorial results on completely positive matrices, *Linear Algebra Appl.* 95:111–125.

Berman, A., and Plemmons, R. 1979. *Nonnegative Matrices in the Social Sciences*, Academic, New York.

Breiman, L. 1991. The Π method for estimating multivariate functions from noisy data, *Technometrics* 33:125–143.

Cohen, P. J. 1969. Decision procedures for real and p-adic fields, *Comm. Pure Appl. Math.* 22:131–151.

Collins, G. E. 1969. Quantifier elimination for real closed fields by cylindrical algebraic decomposition, in *Lecture Notes in Comput. Sci.* 33, Springer-Verlag, 515–532.

Eaves, B. C., and Rothblum, U. G. 1989. A theory on extending algorithms for parametric problems, *Math. Oper. Res.* 14:502–533.

Gregory, D. A., and Pullman, N. J. 1983. Semiring rank: Boolean rank and nonnegative rank factorization, *J. Combin. Inform. System Sci.* 3:223–233.

Harshman, R. A. 1970. Foundations of the PARAFAC Procedure: Models and Conditions for an 'Explanatory' Multi-Modal Factor Analysis, UCLA Working Paper in Phonetics, 16.

Hayashi, C. 1982. An algorithm for the solution of PARAFAC model and analysis of the data by the international rice adaptation experiments, *Bull. Biometric Soc. Japan* 3:77–91.

Henry, L. 1969a. Schémas de nuptialité: Déséquilibre des sexes et célibat, *Population (Paris)* 24:457–486.

Henry, L. 1969b. Schémas de nuptialité: Déséquilibre des sexes et âge au mariage, *Population (Paris)* 24:1067–1122.

Henry, L. 1972. Nuptiality, *Theoret. Population Biol.* 3:135–152.

Jacobson, N. 1964. *Lectures in Abstract Algebra III*, Van Nostrand, Princeton, N.J.

Lancaster, P., and Tismenetsky, M. 1985. *The Theory of Matrices with Applications*, 2nd ed., Academic, Orlando, Fla.

Levin, B. 1985. On Calculating Maximum Rank-One Underapproximations for Positive Arrays, Technical Report B-48, Div. of Biostatistics, Columbia Univ., New York.

Nisan, N. 1991. Lower bounds for non-commutative computation, in *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, pp. 410–418.

Renegar, J. 1992. On the computational complexity and geometry of the first-order theory of the reals, parts I, II, III, *J. Symbolic Logic* 13:255–352.

Saboulin, M. de. 1985. Remarques sur le célibat engendré par les déséquilibres de la répartition par sexe, presented at Union Internationale pour l'Etude Scientifique de la Population, Florence, Séance F. 22.

Seidenberg, A. 1954. A new decision method for elementary algebra, *Ann. of Math.* 60:365–374.

Suppes, P., and Zanotti, M. 1981. When are probabilistic explanations possible?, *Synthese* 48:191–199.

Tarski, A. 1951. *A Decision Method for Elementary Algebra and Geometry*, 2nd ed., rev., Univ. of California Press, Berkeley.

Yannakakis, M. 1988. Expressing combinatorial optimization problems by linear programs, in *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pp. 223–228.

Received 1 October 1991; final manuscript accepted 8 September 1991