# AN ALMOST OPTIMAL ALGORITHM FOR COMPUTING NONNEGATIVE RANK[*]

ANKUR MOITRA[†]

**Abstract.** Here, we give an algorithm for deciding if the nonnegative rank of a matrix $M$ of dimension $m \times n$ is at most $r$ which runs in time $(nm)^{O(r^2)}$. This is the first exact algorithm that runs in time singly exponential in $r$. This algorithm (and earlier algorithms) are built on methods for finding a solution to a system of polynomial inequalities (if one exists). Notably, the best algorithms for this task run in time exponential in the number of variables but polynomial in all of the other parameters (the number of inequalities and the maximum degree). Hence, these algorithms motivate natural *algebraic* questions whose solution have immediate *algorithmic* implications: How many variables do we need to represent the decision problem, and does $M$ have nonnegative rank at most $r$? A naive formulation uses $nr + mr$ variables and yields an algorithm that is exponential in $n$ and $m$ even for constant $r$. Arora et al. [*Proceedings of STOC*, 2012, pp. 145–162] recently reduced the number of variables to $2r^2 2^r$, and here we *exponentially* reduce the number of variables to $2r^2$ and this yields our main algorithm. In fact, the algorithm that we obtain is nearly optimal (under the exponential time hypothesis) since an algorithm that runs in time $(nm)^{o(r)}$ would yield a subexponential algorithm for 3-SAT [*Proceedings of STOC*, 2012, pp. 145–162]. Our main result is based on establishing a normal form for nonnegative matrix factorization—which in turn allows us to exploit algebraic dependence among a large collection of linear transformations with variable entries. Additionally, we also demonstrate that nonnegative rank cannot be certified by even a very large submatrix of $M$, and this property also follows from the intuition gained from viewing nonnegative rank through the lens of systems of polynomial inequalities.

**Key words.** nonnegative rank, systems of polynomial inequalities, extension complexity

**AMS subject classification.** 68Q25

**DOI.** 10.1137/140990139

## 1. Introduction.

**1.1. Background.** The nonnegative rank of a matrix is a fundamental parameter that arises throughout algorithms and complexity and admits many equivalent formulations. In particular, given a nonnegative[1] matrix $M$ of dimension $m \times n$, its nonnegative rank is the smallest $r$ for which
- $M$ can be written as the product of nonnegative matrices $A$ and $W$ which have dimension $m \times r$ and $r \times n$, respectively;
- $M$ can be written as the sum of $r$ nonnegative rank one matrices,
- there are $r$ nonnegative vectors $v_1, v_2, \ldots, v_r$ (of length $m$) such that the conic hull of $\{v_1, v_2, \ldots, v_r\}$ contains all columns in $M$.

Throughout this paper, we will denote the nonnegative rank by $rank^+(M)$ and will refer to a factorization $M = AW$ where $A$ and $W$ are nonnegative and have dimension $m \times r$ and $r \times n$, respectively, as a nonnegative matrix factorization of inner-dimension $r$.

---

[†]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139 (moitra@mit.edu). Part of this research was done while at the Institute for Advanced Study and was supported by an NSF Computing and Innovation Fellowship and NSF grant DMS-0835373.

[1]We will refer to a matrix that is entrywise nonnegative as a "nonnegative matrix."

Some of the most compelling applications of nonnegative rank are in machine learning [20, 26, 27, 2], statistics [34], combinatorics [45], and communication complexity [5, 28]. In machine learning, the benefit of requiring a matrix factorization $M = AW$ to be nonnegative is that this factorization can then be interpreted *probabilistically*. A representative application comes from the domain of topic modeling [20], where $M$ is chosen to be a so-called term-by-document matrix: the entry in row $i$, column $j$ is the frequency of occurrence of the $i$th word in the $j$th document. And computing a nonnegative matrix factorization of inner-dimension $r$ is akin to finding a collection of $r$ topics (which are each distributions on words) so that each document can be expressed as a convex combination of these $r$ topics. Nonnegative matrix factorization has found applications throughout machine learning, from topic modeling [20, 2, 3, 4] to information retrieval [43] to image segmentation [26, 27] and collaborative filtering [35]. Even this is far from an exhaustive list. We note that of particular interest in these applications are instances of this problem in which the target nonnegative rank $r$ is *small*.

In combinatorial optimization, one is often interested in expressing a polytope $P$ as the projection of a higher-dimensional polytope $Q$ which (hopefully) has much fewer facets. The minimum number of facets needed is called the *extension complexity* of $P$, and there is a rich body of literature on this subject. Yannakakis established a striking connection between extension complexity and nonnegative rank: Given the polytope $P$, one constructs the "slack matrix": the entry in row $i$, column $j$ is how slack the $i$th vertex is against the $j$th constraint. Yannakakis proved that the nonnegative rank of the slack matrix is exactly equal to the extension complexity of $P$ [45]. Fiorini et al. [14] recently used this connection and results from communication complexity to prove a remarkable lower bound, that the traveling salesman polytope has no polynomial size extended formulation. This area has seen a considerable amount of recent progress with tools from discrepancy [8, 39], information complexity [10], common information [9] and harmonic analysis [9] being used to lower bound the nonnegative rank of particular families of slack matrices.

In communication complexity, the famous log rank conjecture of Lovasz and Saks [28] asks if the log of the rank of the communication matrix and the deterministic communication complexity are polynomially related. In fact, an equivalent formulation of this problem (that follows from [5]) is that the log rank conjecture asks if the log of the rank and the log of the nonnegative rank of a Boolean matrix are polynomially related. Of crucial importance here is that the matrix in question be Boolean. For a general matrix, there is no nontrivial relationship since there are examples in which the rank is three and yet the nonnegative rank is $\Omega(\sqrt{n})$ [15]. Also in complexity theory, Nisan used nonnegative rank to prove lower bounds for noncommutative models of computation [33].

We note that nonnegative matrix factorization has also been applied to problems in biology, economics [19], and chemometrics [24] to model all sorts of processes, ranging from stimulation in the visual cortex to the dynamics of marriage. In fact, a historical curiosity is that nonnegative rank was first introduced in chemometrics, under the name of *self-modeling curve resolution*.

**1.2. Systems of polynomial inequalities.** The focus of this paper is the following.

QUESTION 1. *What is the complexity of computing the nonnegative rank?*

A priori it is not even clear that there is an algorithm that runs in any finite amount of time. But indeed, Cohen and Rothblum [12] observed that the decision

question of whether or not $rank^+(M) \leq r$ can be equivalently formulated as a system of $O(mn)$ polynomial inequalities with $mr + nr$ total variables: we can treat each entry in $A$ and each entry in $W$ as a variable, and the constraint that this be a valid nonnegative matrix factorization is exactly that $A$ and $W$ be nonnegative and that $M = AW$. The latter is a set of $mn$ degree two constraints. It is easy to see that this system of polynomial inequalities has a solution if and only if $rank^+(M) \leq r$.

Moreover, whether or not a system of polynomial inequalities has a solution is decidable. This is a quite nontrivial statement. The first algorithm is due to Tarski and Seidenberg [42, 40], and there have since been a long line of improvements to this decision procedure. The best known algorithm is due to Renegar [37], and the running time of finding a solution to a system of $p$ polynomial inequalities $f_1, f_2, \ldots, f_p$ with $k$ variables $x_1, x_2, \ldots, x_k$ and maximum degree $D$ is roughly $(Dp)^{O(k)}$. An important note is that Renegar's algorithm works in slightly more generality (and we will take advantage of this fact): one can also specify an arbitrary Boolean function $\mathbb{P} : \{-1, 0, 1\}^p \to \{0, 1\}$ that takes the sign pattern of the set of polynomials and returns a Boolean value. This in turn defines a set of feasible solutions as follows: We will be interested in

$$\bigcup_{\pi \in \{-1, 0, +1\}^k \text{s.t.} \mathbb{P}(\pi) = 1} \left\{ x \in \mathbb{R}^k \text{ s.t. } \forall_j \ sgn(f_j(x_1, x_2, \ldots, x_k)) = \pi_j \right\}.$$

In particular, the feasible solutions are all choices of values for the variables $x_1, x_2, \ldots, x_k$ so that the resulting sign pattern $sgn(f_1), sgn(f_2), \ldots, sgn(f_p)$ of the polynomials satisfies the Boolean function $\mathbb{P}$. Moreover, deciding whether this set has a solution can be solved in roughly the same amount of time (provided the Boolean function can be evaluated efficiently).

So (appealing to decision procedures for a system of polynomial inequalities) there is an algorithm for computing the nonnegative rank of a matrix that runs in a finite amount of time. Note that if the target nonnegative rank $r$ is *small* (say, three), this algorithm still runs in time exponential in $m$ and $n$. And the question of whether or not there is a faster algorithm (in particular, one which runs in polynomial time for any constant $r$) was still open. Vavasis proved that nonnegative rank is NP-hard to compute [44], but this only rules out an exact algorithm that runs in time polynomial in $n$, $m$, and $r$ (if $P \neq NP$).

The crucial observation that the reader should keep in mind throughout this paper is that the main bottleneck in finding a solution to a system of polynomial inequalities is the *number of variables*. Renegar's algorithm [37] runs in time polynomial in the number of polynomials ($p$) and the maximum degree ($D$) but runs in time exponential in the number of variables ($k$). In a technical sense, the number of variables plays an analogous role to the VC-dimension in learning theory. (This connection can be made explicit by drawing an analogy between the Milnor–Thom and Warren bounds and the Sauer–Shelah lemma. See, e.g., [1]).

Cohen and Rothblum [12] give a reduction from nonnegative rank to finding a solution to a system of polynomial inequalities that has $mr + nr$ variables, and a natural goal is to try to use fewer variables in this reduction. Using a different approach, Gillis gave an algorithm for deciding if $rank^+(M) \leq 3$ in polynomial time [16]. Arora et al. [2] do exactly this and give a reduction to a system with only $f(r) = 2r^2 2^r$ variables. This yields an exact algorithm for deciding if $rank^+(M) \leq r$ that runs in time $(nm)^{2r^2 2^r}$ which is doubly exponential in $r$ but runs in polynomial time algorithm for any fixed $r$. Furthermore Arora et al. [2] demonstrate that an exact algorithm for deciding if $rank^+(M) \leq r$ that runs in time $(nm)^{o(r)}$ would yield

a subexponential time algorithm for 3-SAT. In summary, there is an exact algorithm for deciding if $rank^+(M) \leq r$ that runs in polynomial time for any $r = O(1)$, and any algorithm must depend (at least) exponentially on $r$. However, the algorithm in [2] runs in time doubly exponential in $r$, and perhaps we could still hope for an algorithm that runs in time singly exponential in $r$. Here, we give such an algorithm; we do this by reducing the number of variables *exponentially* from $2r^2 2^r$ to $2r^2$.

And perhaps the main message in this paper is that systems of polynomial inequalities with even just a small number of variables can be remarkably expressive! We believe that this theme may find other applications: Perhaps there are other problems for which one would like to design an algorithm based on solving some appropriately chosen system of polynomial inequalities. Then in this case, reducing the number of variables can drastically improve the running time of an algorithm. Indeed, maybe this complexity measure deserves to be studied in its own right.

META QUESTION 1. *Given a decision problem, how many variables are needed to encode its answer as a system of polynomial inequalities?*

In particular, we want that the decision problem is a **YES** instance if and only if the corresponding system of polynomial inequalities has a solution.

One of the main points of the paper is that asking how many variables are needed to represent a decision problem is a basic question, can have surprisingly efficient representations, and can even lead to nearly optimal algorithms. In addition to being a basic mathematical problem in its own right, it is also at least superficially related to other algebraic measures of complexity that have been used in theoretical computer science and algebraic geometry. For example, in the sum-of-squares proof system, one starts with a collection of axioms that are themselves polynomial relations and studies the minimal degree needed to prove or refute various other statements [36, 23]. In another direction, a fundamental result in algebraic geometry is Mnev's universality theorem, which can be seen as a way to measure algebraic complexity (of oriented matroids) by studying conditions under which they can represent arbitrary semialgebraic sets [30]. In fact, this too has applications in theoretical computer science and was used by Shor to resolve the complexity of the stretchability problem [41].

**1.3. Our results.** We now state our main results: Let $M$ be a $m \times n$ nonnegative matrix and let $L$ denote the maximum bit complexity of any coefficient in $M$. We give the following algorithm. (Here, $c$ is a universal constant.)

THEOREM 1.1. *There is a $poly(n, m, L)(r4^{r+1}mn)^{cr^2}$ time algorithm for deciding if the nonnegative rank of $M$ is at most $r$. Additionally, given $\delta > 0$ (and if $rank^+(M) \leq r$), the algorithm runs in time at most $poly(n, m, L, \log \frac{1}{\delta})(r4^{r+1}mn)^{cr^2}$ and returns factors $\tilde{A}$ and $\tilde{W}$ that are entrywise close (within an additive $\delta$) to $A$ and $W$ (respectively) that are a nonnegative matrix factorization of $M$ of inner-dimension at most $r$. Furthermore, the entries of $\tilde{A}$ and $\tilde{W}$ have rational coordinates with numerators and denominators bounded in bit length by $O(L(r4^{r+1}mn)^{cr^2} + \log \frac{1}{\delta})$.*

This is the first algorithm that runs in singly exponential time as a function of $r$ and in fact is an *exponential* improvement over the previously best known algorithm due to Arora et al. [2]. Moreover, notice that the algorithm in [2] is faster than the one in [12] only if $r = O(\log(m + n))$, whereas our algorithm is in fact faster for any $r = o((mn)^{1/3})$. Our algorithm is nearly optimal (under the exponential time hypothesis), since an exact algorithm that runs in time $(nm)^{o(r)}$ would yield a subexponential time algorithm for 3-SAT [2].

We note that [2] provides an $(nm)^{cr^2}$ time algorithm for the special case in which the matrix $M$ not only has nonnegative rank at most $r$ but, furthermore, the matrix

$A$ can be assumed to be full rank. (This is called "simplicial factorization.") However, a simplicial factorization turns out to be *much* more restrictive than a general nonnegative matrix factorization. In fact, Hrubes [21] gives an explicit example of an $n \times n$ nonnegative matrix $M$ whose nonnegative rank is $\Theta(\log n)$ even though the minimum inner-dimension of any simplicial factorization of it is $n$. Thus, one should be careful about the distinction between these two problems. It is true that in many applications in machine learning, it is a priori natural to assume that $A$ is full rank; in many other applications such as in extended formulations, imposing this artificially as a constraint can drastically change the nature of the problem.

**Our techniques and a comparison.** In order to explain our techniques, it makes sense to first explain the approach of Arora et al. [2]. The starting point in earlier work is to observe that the rows of $A$ and the columns of $W$ can be obtained as an appropriate linear transformation of the rows and columns of $M$, respectively. However, when $A$ is not full rank, we may need many such linear transformations, one for each maximal linearly independent set of its columns. Thus, there could be as many as $\binom{r}{r/2}$ such linear transformations. Even so, the major complication is that when given a list of linear transformations to apply to the columns of $M$ to obtain the columns of $W$, there may be more than one viable option. One of these candidates is the true column of $W$, but if we select the wrong one (even if it is nonnegative) and insert it into $W$, there is no guarantee that there is still a compatible, nonnegative $A$ that satisfies $M = AW$. So the major issue is how to choose.

This is always handled through a structure theorem. The approach in Arora et al. [2] is substantially more complicated and is based on the notion of a *proper chain* and a *simplicial partition*, which we explain next. A proper chain is a sequence of valid nonnegative factorizations $M = AW$, $M = AW'$, $M = A'W'$. The idea is that if $A$ were fixed, we could use a simple rule for which linear transformation to apply to columns of $M$ to obtain columns of $W'$. This is how $W'$ is constructed—given $A$, for each column $M_i$ choose $W_i$ so that it is nonnegative and satisfies $M_i = AW_i$, and if there is more than one choice, choose the one whose support is lexicographically first. We can then update $A$ to $A'$ in the same manner, so that we choose $(A')^j$ so that it is nonnegative and satisfies $(A')^j W' = M^j$, and again among the viable options we choose the one whose support is lexicographically first. However, after we update $A$ to $A'$, the choices for $W'$ are not necessarily lexicographically first for $A'$ because our choices were based on $A$. So the structure theorem needs to account for the fact that the rule for how to tie break among linear transformations depends on some unspecified matrix $A$. One can then classify all such choice functions—which are called *simplicial partitions*—and brute force search over them, and then ask if any of them yields a system of polynomial constraints that is feasible.

Here, we alternately update $A$ and $W$ while maintaining the invariant that $M = AW$ is a valid nonnegative matrix factorization, until the choices for $W'$ are lexicographically first based on $A$ and the choices for $A$ are lexicographically first based on $W$. This can take exponentially many steps to converge, but it does and the important point for us is that this stronger normal form exists so when we are looking for a nonnegative matrix factorization, we can without loss of generality assume that this is the one we are looking for. This allows us to remove the use of simplicial partitions altogether, which is one of the main bottlenecks in the previous algorithm because searching over all simplicial partitions requires doubly exponential (in $r$) time.

However, there is still one major obstacle remaining to overcome, which is the number of variables. Recall that if $A$ is not full rank, then there could be as many as

$\binom{r}{r/2}$ maximal linearly independent sets of columns, each of which yields a different linear transformation. In fact, we need *all* of these to recover the columns of $W$ from the columns of $M$. (For more details, see the discussion on the cross-polytope at the end of section 2.3). Despite the fact that we need exponentially (in $r$) many linear transformations, these linear transformations are related to each other through the entries of $A$—or, more precisely, through $A^U$, where $U$ is any maximal set of linearly independent rows. We can then use Cramer's rule to express all of the entries of the linear transformations using a common set of $2r^2$ variables, through ratios of polynomials. This is the final step that allows us to reduce the number of variables in our representation from exponential in $r$ to quadratic in $r$ and obtain nearly optimal algorithms for nonnegative matrix factorization.

We also consider another basic question about the nonnegative rank of a matrix.

QUESTION 2. *Can the nonnegative rank of a matrix $M$ be certified by a small submatrix?*

Indeed—in the case of the rank—a matrix $M$ has rank at least $r$ if and only if there is an $r \times r$ submatrix of $M$ that has rank $r$. This property plays a crucial role in many applications [17], and it is natural to wonder if the nonnegative rank admits any similar characterization. As another motivation, often we are only given a subset of the entries of the matrix $M$ (for example, in the Netflix problem), and we would like to use these entries to infer properties about $M$. Yet, the nonnegative rank behaves quite differently than the rank in this regard.

THEOREM 1.2. *For any $r \in \mathbb{N}$, there is a $3rn \times 3rn$ nonnegative matrix which has nonnegative rank at least $4r$ and yet for any $< n$ rows, the corresponding submatrix has nonnegative rank at most $3r$.*

So even the submatrices consisting of a constant fraction of the rows in $M$ do not determine the nonnegative rank of $M$ even within a constant factor. This result, too, can be thought of in the language of systems of polynomial inequalities: The basic principle at play is that even though the nonnegative rank can be equivalently characterized by a system of polynomial inequalities with only $2r^2$ variables, there are systems of polynomial inequalities that are together infeasible and yet every *large* subset of the constraints is feasible. This is in stark contrast to the case of linear inequalities, for which, if the system is infeasible (and is in dimension $d$) there is a subset of just $d$ linear inequalities that is infeasible (i.e., there is a size $d$ obstruction) [29].

We remark that there is also considerable interest in computing and proving lower bounds for a related quantity called the *semidefinite rank* [18, 14]. This parameter characterizes semidefinite representability in much the same way that nonnegative rank captures representation as a linear program. Recently, Lee, Raghavendra, and Steurer [25] building on techniques in [11] gave the first exponential lower bounds for general semidefinite programs, but there are still many open questions in this area about both proving quantitatively stronger lower bounds and proving lower bounds for approximation versions of these representability questions. In another direction, it is not known whether there is an algorithm to decide if the semidefinite rank of a nonnegative matrix $M$ is at most $r$ that runs in polynomial time for $r = O(1)$ [13]. Our approach crucially uses the fact that the solution to a linear program is a rational function of the input parameters, but this is not true in the case of semidefinite programs [32] and is a major obstacle to generalizing the algorithms in this paper.

## 2. Computing the nonnegative rank.

**2.1. Stability (a normal form).** Throughout this paper, let $M$ denote an entrywise nonnegative matrix of dimension $m \times n$. We will also let $M_i$ denote the $i$th

column of $M$ and $M^j$ denote the $j$th row. Given a subset $U \subseteq [n]$, we will let $M_U$ denote the submatrix consisting of columns of $M$ from the set $U$. (And, similarly, $M^V$ is a submatrix of rows of $M$.)

DEFINITION 2.1. *$rank^+(M)$ is the smallest $r$ such that $M$ can be written as $M = AW$ where $A$ and $W$ are nonnegative and have dimension $m \times r$ and $r \times n$ respectively.*

Additionally, we will call $M = AW$ a nonnegative matrix factorization of inner-dimension $r$.

DEFINITION 2.2. *Let $cone(A) = \{\sum_i \alpha_i A_i | \forall_i : \alpha_i \geq 0\}$ be the conic hull of columns in $A$.*

*Note.* Given $A$, there is a nonnegative matrix $W$ such that $M = AW$ if and only if each column $M_i$ of $M$ is contained in $cone(A)$.

DEFINITION 2.3. *Given $A$ and a vector $v \in \mathbb{R}^m$ (recall $A$ is dimension $m \times r$), we will call a subset $S$ of columns of $A$ admissible if $v \in cone(A_S)$.*

We will make use of this definition below, where we will introduce a normal form for nonnegative matrix factorization and show that it always exists. This normal form will also make use of the *lexicographic ordering* on subsets of columns of $A$. However, we caution that the definition we will use differs from the standard definition in how it compares sets of different sizes.

DEFINITION 2.4. *Given two sets $S = \{s_1, s_2, \ldots, s_k\}$ and $T = \{t_1, t_2, \ldots, t_k\}$ of the same size so that $s_1 \leq s_2 \leq \cdots \leq s_k$ and $t_1 \leq t_2 \leq \cdots \leq t_k$, we say that $S$ is lexicographically before $T$ if there is an $i \in [k]$ and $s_j \leq t_j$ for all $j = 1, 2, \ldots, i - 1$ and $s_i < t_i$. Moreover, if $S$ and $T$ have different sizes, and $|S| < |T|$, then we define $S$ to be lexicographically before $T$.*

Let $M = AW$ be a nonnegative matrix factorization.

DEFINITION 2.5. *For each column $M_i$, let $S_i$ be the lexicographically first admissible subset (of columns of $A$) for $M_i$. Similarly, for each row $M^j$, let $T_j$ be the lexicographically first admissible subset (of rows of $W$) for $M^j$. We call $M = AW$ stable if*

1. *for each $i$, the support of $W_i$ is $S_i$,*
2. *and for each $j$, the support of $A^j$ is $T_j$.*

Next, we show that a nonnegative matrix factorization of inner-dimension $r$ can always be made stable (while preserving nonnegativity and the inner-dimension).

LEMMA 2.6. *If $M = AW$ is a nonnegative matrix factorization of inner-dimension $r$, then there is an $\tilde{A}$ and $\tilde{W}$ such that*

1. *$M = \tilde{A}\tilde{W}$, $\tilde{A}$ and $\tilde{W}$ are nonnegative and have inner-dimension $r$, and*
2. *$M = \tilde{A}\tilde{W}$ is stable.*

*Proof.* The natural approach to prove this lemma is, if $M = AW$ is not stable, update columns in $W$ or rows in $A$. The only subtle point is that if we update $A$ and $W$ at the same time to $\tilde{A}$ and $\tilde{W}$, we may not have $M = \tilde{A}\tilde{W}$. So the approach is to alternate between a $W$-updating phase and an $A$-updating phase.

In a $W$-updating phase, for each column $M_i$ let $S_i$ be the lexicographically first subset of columns of $A$ that is admissible for $M_i$. If $S_i$ is lexicographically (strictly) earlier than the support of $W_i$, we find a vector $\tilde{W}_i$ that is nonnegative, supported in $S_i$, and satisfies $M_i = A\tilde{W}_i$. If not, we set $\tilde{W}_i = W_i$. In either case, we have that $M_i = A\tilde{W}_i$ and hence $M = A\tilde{W}$. At the end of this phase, we overwrite $W$ with $\tilde{W}$. The $A$-updating phase is defined analogously, and throughout this procedure we maintain the invariant that $M = AW$ and $A$ and $W$ are nonnegative and have inner-dimension $r$. Moreover, the support of columns of $W$ and rows of $A$ are monotonically decreasing according to the lexicographical ordering, so this procedure terminates in a finite number of steps. $\square$

**2.2. Few entries determine $A$ and $W$.** Throughout this section, let $M = AW$ be a *stable* nonnegative matrix factorization. The goal in this section is to demonstrate that (given $M$), only a few entries in $A$ and $W$ are needed to determine the remaining entries. This is only a property of stable factorizations and is not guaranteed to hold for general factorizations.

Let $rank(A) = s$ and let $U \subseteq [m]$ be a set of $s$ linearly independent rows in $A$. Furthermore, let $S_1, S_2, \ldots, S_p \subseteq [r]$ be the (full) list of sets of $s$ linearly independent columns of $A^U$ (in lexicographic order). Note that $p \leq \binom{r}{s} \leq 2^r$.

DEFINITION 2.7. *The ensemble of $A$ (at $U$) is a list of linear transformations: $B_1, B_2, \ldots, B_p$ where for each $i$, $B_i$ is an $r \times s$ matrix that is zero on all rows outside the set $S_i$ and restricted to rows in $S_i$ is $(A_{S_i}^U)^{-1}$.*

Note that each submatrix $A_{S_i}^U$ is indeed invertible: $rank(A) = s$ and $U$ is a set of $s$ linearly independent rows so a set $S_i$ of columns of $A$ is linearly independent if and only if these vectors restricted to $U$ are also linearly independent. The idea behind this definition is that if a vector $x$ is supported on the set $S_i$, then taking just the entries of $Ax$ corresponding to $U$, we can recover $x = B_i A^U x$, and $B_i$ is just a zero-padding of $(A_{S_i}^U)^{-1}$ to make this work.

The main goal in this section is to show the following.

LEMMA 2.8. *For each column $M_i$, among the set of vectors $\mathcal{S} = \{B_1 M_i^U, B_2 M_i^U, \ldots, B_p M_i^U\}$, $W_i$ is the unique vector with lexicographically minimal support among all nonnegative vectors in $\mathcal{S}$.*

We will break this lemma up into two parts.

CLAIM 2.9. *$W_i$ is contained in the set $\mathcal{S}$.*

*Proof.* Let $R_i$ be the support of $W_i$. Then $R_i$ must correspond to a linearly independent set of columns of $A$—otherwise, we could find a nonnegative $\tilde{W}_i$ whose support is a strict subset of $R_i$ such that $A\tilde{W}_i = M_i$, but this would violate the condition of stability.

Because the sets of linearly independent columns of $A$ are a matroid, there is a set $S_{i'}$ of $s$ linearly independent columns of $A$ for which $R_i \subseteq S_{i'}$. Hence,

$$B_{i'} M_i^U = B_{i'}(AW_i)^U = B_{i'} A^U W_i = v.$$

However, $B_{i'}$ is zero on rows outside the set $S_{i'}$, and restricting $B_{i'} A^U$ to rows and columns in $S_{i'}$ is the $s \times s$ identity matrix. Since the support of $W_i$ is contained in $S_{i'}$, we have $W_i = v$. $\quad\Box$

We note a corollary of this lemma that will be useful later.

COROLLARY 2.10. *The support of $W_i$ corresponds to a linearly independent set of columns in $A$.*

Next, we prove the second part needed for the main result in this section.

CLAIM 2.11. *For each vector $B_{i'} M_i^U$, $AB_{i'} M_i^U = M_i$.*

*Proof.* Let $v = AB_{i'} M_i^U$. We prove this lemma in two parts: first, we prove that $v^U = M_i^U$, and then we prove the full lemma from this. Since $B_{i'}$ is zero on rows outside the set $S_{i'}$, we have

$$AB_{i'} = A_{S_{i'}} B_{i'}^{S_{i'}} = A_{S_{i'}}(A_{S_{i'}}^U)^{-1}.$$

Hence, $v^U = A_{S_{i'}}^U (A_{S_{i'}}^U)^{-1} M_i^U = M_i^U$. Consider a $j$ outside the set $U$. By the choice of $U$, the row $A^j$ can be expressed as a linear combination of rows in $A$ in the set $U$:

$$A^j = \sum_{j' \in U} \alpha_{j,j'} A^{j'}.$$

Since $AW_i = M_i$, we have $M_i^j = A^j W_i = \sum_{j' \in U} \alpha_{j,j'} A^{j'} W_i = \sum_{j' \in U} \alpha_{j,j'} M_i^{j'}$, and hence

$$
v^j = A^j B_{i'} M_i^U = \sum_{j' \in U} \alpha_{j,j'} A^{j'} B_{i'} M_i^U
$$
$$
= \sum_{j' \in U} \alpha_{j,j'} v^{j'} = \sum_{j' \in U} \alpha_{j,j'} M_i^{j'} = M_i^j. \qquad \square
$$

Now we can prove the main lemma in this section.

*Proof.* We have already shown (Claim 2.9) that $W_i$ occurs in the set $\mathcal{S}$. Consider any other *nonnegative* vector $B_{i'} M_i^U = v$. We need to show that the support of $v$ is lexicographically later than the support of $W_i$.

First, we claim that if $v \neq W_i$ then the support of $W_i$ is not the same as the support of $v$. Suppose not—i.e., $v \neq W_i$ and yet the support of $v$ and of $W_i$ are identical. (Let this set be $R$.) Indeed, $R$ must correspond to a linearly independent set of columns of $A$ (Corollary 2.10). Hence, we cannot have $A(v - W_i) = \vec{0}$ (using Claim 2.11) with $v - W_i \neq \vec{0}$ and support of $v - W_i$ contained in $R$.

So the support of $W_i$ and $v$ are not identical, and one of these must be lexicographically earlier. Suppose (for contradiction) that the support of $v$ is earlier. We know (Claim 2.9) that the support of $W_i$ is an admissible set of columns of $A$ for $M_i$. This contradicts stability (because we could update $W_i$ to $v$), and so we can conclude that the support of $W_i$ is lexicographically earlier. $\quad\square$

Let $rank(W) = t$ and let $V$ be a set of $t$ linearly independent columns of $W$. Then we can define an ensemble $C_1, C_2, \ldots, C_q$ for $W$ at $V$ analogously as we did for $A$. Similarly, we have $q \leq \binom{r}{t}$ and for all $j$, among the set

$$
\mathcal{T} = \left\{ M_V^j C_1, M_V^j C_2, \ldots, M_V^j C_q \right\}
$$

$A^j$ is the vector with lexicographically minimal support among all nonnegative vectors in $\mathcal{T}$. (This follows from the above proof by interchanging the roles of $A$ and $W$.)

**2.3. A semialgebraic set, take 1.** Our goal is to encode the question of whether or not $rank^+(M) \leq r$ as a nonemptiness problem for a semialgebraic set with a small number of variables. Recall that Renegar's algorithm allows us to specify an arbitrary Boolean function $\mathbb{P}$ on the signs of the polynomials, and as long as we can efficiently evaluate $\mathbb{P}$ we will be able to use Renegar's algorithm to determine whether this system has a solution. Our first attempt will be to choose the entries in $B_1, B_2, \ldots, B_p$ and $C_1, C_2, \ldots, C_q$ as the *variables*. Our first goal is to construct a set of polynomial constraints (using the variables) so that setting $B_1, B_2, \ldots, B_p$ and $C_1, C_2, \ldots, C_q$ to the ensembles of a stable factorization $M = AW$ is a valid solution. We then show (conversely) that any valid setting of the variables in fact yields a nonnegative matrix factorization with inner-dimension $r$.

Suppose we are given the sets $U$ and $V$ and the ensembles $B_1, B_2, \ldots, B_p$ and $C_1, C_2, \ldots, C_q$.

DEFINITION 2.12. *Let $first(\mathcal{S})$ applied to a collection of vectors output the vector with lexicographically minimal support among all nonnegative vectors in $\mathcal{S}$.*

This function can output **FAIL** if there is no nonnegative vector in $\mathcal{S}$.

CLAIM 2.13. *Set*

$$
W_i \leftarrow first(\{B_1 M_i^U, B_2 M_i^U, \ldots, B_p M_i^U\}), \tag{2.1}
$$
$$
A^j \leftarrow first(\{M_V^j C_1, M_V^j C_2, \ldots, M_V^j C_q\}). \tag{2.2}
$$

*There is an explicit Boolean function $\mathbb{P}$ that determines if for all $i$ and $j$: (1) $W_i \geq \vec{0}$, (2) $A^j \geq \vec{0}$, and (3) $A^j W_i = M_i^j$. Furthermore, $\mathbb{P}$ is a function of sign constraints on the polynomials:*

    1. *$B_{i'} M_i^U$ (for all $i, i'$),*

    2. *$M_V^j C_{j'}$ (for all $j, j'$), and*

    3. *$M_V^j C_{j'} B_{i'} M_i^U - M_i^j$ (for all $i, i', j, j'$).*

This claim is immediate, but we include a description of the Boolean function $\mathbb{P}$ for completeness

*Proof.* The Boolean function $\mathbb{P}$ will be an AND over subfunctions $\mathbb{P}_{i,j}$ defined for each $i$ and $j$: $\mathbb{P}_{i,j}$ will compute the index $i'$ and $j'$ so that $B_{i'} M_i^U$ and $M_V^j C_{j'}$ are lexicographically earliest among nonnegative vectors in the sets $\mathcal{S} = \{B_1 M_i^U, B_2 M_i^U, \ldots, B_p M_i^U\}$ and $\mathcal{T} = \{M_V^j C_1, M_V^j C_2, \ldots, M_V^j C_q\}$, respectively. This can be computed from only the signs of entries in the vectors in these sets.

Then $\mathbb{P}_{i,j}$ will check that for this $i'$ and $j'$, that $M_V^j C_{j'} B_{i'} M_i^U = M_i^j$. If there is no nonnegative vector in either $\mathcal{S}$ or $\mathcal{T}$, or there are two or more nonnegative vectors tied for lexicographically earliest support (among only nonnegative vectors), then $\mathbb{P}_{i,j}$ will output **FAIL**. □

LEMMA 2.14. *$\mathbb{P}$ will output **PASS** when $\{B_{i'}\}_{i'}$ and $\{C_{j'}\}_{j'}$ are chosen as the ensembles of a stable factorization $M = AW$.*

*Proof.* This follows immediately from Lemma 2.8. However, note that Lemma 2.8 establishes uniqueness (i.e., the vector with lexicographically earliest support among all nonnegative vectors is unique), and hence each $\mathbb{P}_{i,j}$ will not prematurely output **FAIL** for these choices of $\{B_{i'}\}_{i'}$ and $\{C_{j'}\}_{j'}$. □

Next, we prove the converse direction.

LEMMA 2.15. *If $\mathbb{P}$ outputs **PASS**, then $A$ and $W$ (as defined in (2.1) and (2.2)) are a nonnegative matrix factorization of inner-dimension $r$.*

*Proof.* We have that $W_i$ and $A^j$ are nonnegative (otherwise, $\mathbb{P}$ would have output **FAIL**) and $\mathbb{P}$ explicitly checks that $A^j W_i = M_i^j$ and hence $M = AW$. Note that $B_{i'}$ and $C_{j'}$ are $r \times s$ and $t \times r$ dimensional, so $M = AW$ does indeed have inner-dimension $r$. □

Combining Lemma 2.14 and Lemma 2.15, we have the following.

THEOREM 2.16. *$\mathbb{P}$ outputs **PASS** for some choice of $s, t, U, V, p,$ and $q$ and some setting of the variables $B_1, B_2, \ldots, B_p$ and $C_1, C_2, \ldots, C_q$ if and only if $rank^+(M) \leq r$.*

This leads to a natural approach for computing the nonnegative rank:

    1. `Guess` $s = rank(A)$, $t = rank(W)$ `(for some stable factorization` $M = AW$`)`

    2. `Guess` $U$ `and` $V$

    3. `Guess` $p \leq \binom{r}{s}$ `and` $q \leq \binom{r}{t}$

    4. `Define a semi-algebraic set where the entries of` $B_1, B_2, \ldots, B_p$ `and` $C_1, C_2, \ldots, C_q$ `are variables (using the Boolean function` $\mathbb{P}$ `in Claim 2.13)`

    5. `Run an algorithm for deciding if the semi-algebraic set is non-empty (e.g., [37])`

Currently, the best algorithms for deciding if a semialgebraic set is nonempty run in time

$$\left(\# \text{ polynomials} \times D\right)^{O(k)},$$

where $D$ is the maximum degree and $k$ is the number of *variables*. The main drawback

of the above approach is that the number of *variables* is large: There are $rsp + rtq$ variables, and indeed $p$ and $q$ can be exponential in $r$. For example, if we take the columns of $A$ to be vertices of the cross-polytope (in $r/2$ dimensions), then we do in fact need exponentially many simplices (one corresponding to each linear transformation $B_{i'}$) if we want their union to cover the convex hull of the cross-polytope and nothing else. This follows because the cross-polytope has exponentially many facets but a simplex in $r$ dimensions only has $r + 1$ facets, and each facet of the cross-polytope must be covered by some facet of one of the simplices.

Hence, the running time of the above algorithm will be doubly exponential in $r$. However, we will be able to reduce the number of variables in this semialgebraic set to polynomial in $r$. The definition of stability is somewhat delicate, but this is what allows us to get an exponential reduction in the number of variables.

**2.4. A semialgebraic set, take 2.** Here, we reduce the number of variables in the semialgebraic set *exponentially* by exploiting algebraic dependence among the matrices in the ensembles. Consider the ensemble: $B_1, B_2, \ldots, B_p$, where for each $i$, there is a linearly independent set $S_i$ of $s$ columns of $A$ and $(B_i)^{S_i} = (A^U_{S_i})^{-1}$. Recall Cramer's rule as follows.

LEMMA 2.17 (Cramer). *Let $R$ be an $s \times s$ invertible matrix. Then $(R^{-1})^j_i = det(R^{-i}_{-j})/det(R)$, where $R^{-i}_{-j}$ is the matrix $R$ with the $i$th row and the $j$th column removed.*

The key idea is that if we consider the polynomial constraints in Claim 2.13, the variables are the entries of $B_{i'}$ and there are just too many variables. But we can exploit the algebraic dependencies between the entries of $B_{i'}$ in the true solution we are looking for to reduce the number of variables. For example, consider $B_{i'} M^U_i$ which is a vector whose entries are linear functions of the variables in $B_{i'}$. Then the Boolean function $\mathbb{P}$ in Claim 2.13 is itself a function of the signs of the entries of this vector (and of the signs of other polynomials in the system of constraints). However, if we are interested in the signs of entries of this vector, some are zero just due to padding and the others that are nonzero are of the form $(A^U_{S_{i'}})^{-1} M^U_i$. But we can express the entries of the inverse as ratios of polynomials in the *new* variables corresponding to entries in $A^U$ using Cramer's rule above. We can then clear the denominator and take the signs of actual polynomials and infer signs of the ratios of polynomials, and in this way we can express the system of polynomial constraints given in Claim 2.13 for our target solution *equivalently* as a Boolean function of signs of degree at most $s^2$ polynomials in the entries of $A^U$. Similarly, in Claim 2.13 we have terms of the form

$$M^j_V C_{j'} B_{i'} M^U_i - M^j_i.$$

This is a matrix whose entries are quadratic functions of the variables in $B_{i'}$ and $C_{j'}$. Again, for our target solution we can express this as

$$\sum_{\ell \in T_{j'} \cap S_{i'}} \left( M^j_V (W^{T_{j'}}_V)^{-1} \right)_\ell \left( (A^U_{S_{i'}})^{-1} M^U_i \right)^\ell - M^j_i.$$

And now we can apply Cramer's rule again to express the entries of the inverse as ratios of polynomials and clearing the denominator. In this way, the sign of any entry in $M^j_V C_{j'} B_{i'} M^U_i - M^j_i$ can be expressed through the signs of degree at most $2r^2$ polynomials in the entries of $A^U$ and $W_V$. Applying the transformations to the set of constraints defined in Claim 2.13, we immediately get a semialgebraic set that has $rs + rt$ variables and has $r(p + q) + (p + q) + mnpq$ polynomials of degree at most $2r^2$.

Note that Lemma 2.15 still implies that if $\mathbb{P}$ outputs **PASS**, $rank^+(M) \leq r$ and a nonnegative matrix factorization of inner-dimension $r$ can be computed from the settings of the variables for the valid point in the semialgebraic set. And Lemma 2.8 still implies that this semialgebraic set is nonempty if $rank^+(M) \leq r$ (since, moreover, Lemma 2.6 implies that there is a stable factorization). Hence, the semialgebraic set has a solution if and only if $rank^+(M) \leq r$.

We can now use this reduction and known algorithms for solving systems of polynomial inequalities (as described in section 1.2) to give a nearly optimal algorithm for deciding if $M$ has nonnegative rank at most $r$. Additionally, if $rank^+(M) \leq r$, we can also compute the corresponding nonnegative factors $A$ and $W$ to within an additive $\delta$ (at the expense of an extra factor $\log \frac{1}{\delta}$ in the running time). In [37], Renegar gave the first algorithm for deciding if a system of polynomial inequalities has a solution that runs in time exponential in the number of variables. We note that in [38], Renegar extended this algorithm to also return a $\delta$-approximate solution to an algebraic formula, and this is the algorithm that we will use to actually compute the factors $A$ and $W$. We also note that these algorithms only assume access to an oracle to the Boolean function $\mathbb{P}$, and our function $\mathbb{P}$ is computable in polynomial time. Let $L$ denote the maximum bit complexity of any coefficient in $M$. Then applying the algorithms in [37, 38, 6] with our reduction, we obtain the following.

THEOREM 2.18. *There is a $poly(n, m, L)(r4^{r+1}mn)^{cr^2}$ time algorithm for deciding if the nonnegative rank of $M$ is at most $r$. Additionally, given $\delta > 0$ (and if $rank^+(M) \leq r$), the algorithm runs in time $poly(n, m, L, \log \frac{1}{\delta})(r4^{r+1}mn)^{cr^2}$ and returns factors $\tilde{A}$ and $\tilde{W}$ that are entrywise close (within an additive $\delta$) to $A$ and $W$ (respectively) that are a nonnegative matrix factorization of $M$ of inner-dimension at most $r$. Furthermore, the entries of $\tilde{A}$ and $\tilde{W}$ have rational coordinates with numerators and denominators bounded in bit length by $O(L(r4^{r+1}mn)^{cr^2} + \log \frac{1}{\delta})$.*

Alternatively, in the Blum–Shub–Smale (BSS) model [7], one can instead use the algorithm in [37, 6] to decide if $rank^+(M) \leq r$ and the running time of this algorithm is $poly(n, m) + (r4^{r+1}mn)^{cr^2}$.

We emphasize that the above algorithm is based on answering a purely algebraic question: How many variables are needed (in a system of polynomial inequalities) to encode the question whether $M$ has nonnegative rank at most $r$? We obtain an exponential improvement on the number of variables, over the results in [2], and this coupled with algorithms for computing a solution to a system of polynomial inequalities has an immediate algorithmic implication. The algorithm we obtain here is in fact nearly optimal under the exponential time hypothesis of Impagliazzo and Paturi [22], since Arora et al. [2] showed that an algorithm that decides if $rank^+(M) \leq r$ in $(nm)^{o(r)}$ time would imply a subexponential time algorithm for 3-SAT.

**3. Fragile instances of nonnegative rank.** An important property of the rank of a matrix is that if a given matrix $M$ has rank $r$, there is an $r \times r$ submatrix of $M$ that also has rank $r$. Hence, rank admits a small certificate that serves as proof that a matrix does indeed have rank at least $r$, and this fact plays a crucial role in many applications.

Here, we give highly fragile instances of nonnegative rank: There is a (nonnegative) matrix $N$ of dimension $n \times n$ with $rank^+(N) = 4r$, yet for *any* submatrix $N'$ of at most $\frac{n}{3r}$ columns of $N$, $rank^+(N') \leq 3r$. To put this result in context, consider a system of *linear* inequalities in $d$ dimensions that is infeasible. A basic result in discrete geometry [29] is that there is a subset of at most $d + 1$ of the linear inequalities that is infeasible. In section 2.4, we gave a system of *polynomial* inequalities in $2r^2$

dimensions that has a solution if and only if $rank^+(M) \leq r$. One might hope that this system is infeasible if and only if there is a small subset of the inequalities that alone is infeasible and that this would yield a subset of (say) the columns of $M$ that "proves" that $rank^+(M) > r$. Yet this is not the case, and systems of polynomial inequalities do not have the "Helly property" [29]. (Indeed, their individual constraints do not necessarily correspond to convex regions.) We will proceed by constructing a lower-dimensional fragile instance $M$ and then using it to construct a larger, block diagonal matrix $N$.

To give fragile instances of nonnegative rank, we will make use of a series of reductions of Vavasis [44] and a particular gadget in Arora et al. [2]. In fact, we make use of a crucial property of the reduction in [44] from nonnegative rank to the intermediate simplex problem—in a sense, that rows of $M$ are mapped to points and columns of $M$ are mapped to constraints when reducing to the intermediate simplex problem. We will only be interested in the intermediate simplex problem in two dimensions.

DEFINITION 3.1. *An instance of the intermediate polygon problem is a polygon $P \subset \mathbb{R}^2$ and a set $S \subset P$ of $|S| = n$ points. The goal is to find a triangle $T$ with $S \subset T \subset P$, in which case we call this a **YES** instance and otherwise call it a **NO** instance.*

Our goal is to construct an explicit instance of this problem that is a **NO** instance and yet restricting to *any* set $S' \subset S$ of at most $\frac{n}{3}$ points is a **YES** instance, and we accomplish this latter task by noticing that a particular gadget used in [2] (with a slight modification) has exactly this property. We will then be able to use this instance of the intermediate simplex problem as a gadget to construct fragile instances of nonnegative rank.

We will begin with some simple geometric lemmas and definitions.

DEFINITION 3.2. *Let $C_d = \{(x, y) | x^2 + y^2 \leq d\}$, and we will write $C$ for $C_1$. Let $o$ denote the origin.*

DEFINITION 3.3. *Let $E$ be the set of all equilateral triangles $T \subset C$ where the vertices of $T$ are on the boundary of $C$.*

In our arguments, we will also make use of the (largest) inner circle $c$ that is contained in all triangles in $E$. Equivalently, this circle is the intersection of all triangles in $E$.

DEFINITION 3.4. *Let $c = \cap_{T \in E} T = C_d$ where $d$ is defined as follows: (for an arbitrary $T \in E$), $d$ is the minimum distance from the boundary of $T$ to the origin.*

In particular, $c = C_{1/4}$, although we will not need this.

In our instance of the intermediate polygon problem, the inner polygon will be formed as an intersection of $n$ triangles $T$ each in the set $E$. The common intersection of these triangles will contain $c$, and next we prove that in fact any triangle (contained in $C$) that contains $c$ must in fact be equilateral. This will help us reason about what sorts of triangles can make our instance a **YES** instance. The following two lemmas are proved in [2], but we include the proofs here for completeness.

LEMMA 3.5 (see [2]). *Any arbitrary triangle $T$ with $c \subset T \subset C$ must be in the set $E$.*

*Proof.* Consider a triangle $T$ with $c \subset T \subset C$. Then let $e_1, e_2$, and $e_3$ be the three edges of $T$ and let $\theta_1, \theta_2$, and $\theta_3$ be the *viewing angle* from the origin $o$, namely, $\theta_i$ is the angle formed by $\langle a_i, o, b_i \rangle$, where $a_i$ and $b_i$ are the endpoints of $e_i$.

Since $o \in T$, we have that $\theta_1 + \theta_2 + \theta_3 = 2\pi$. Consider an edge $e_i$. We will prove, by contradiction, that $e_i \cap c$ must contain exactly one point (i.e., $e_i$ must be tangent to the circle $c$). Suppose not—since $c \subset T$, we must have that $e_i \cap c = \emptyset$. Then let

$\ell$ be the line parallel to $e_i$ that is tangent to $c$. Let $e_i'$ be the intersection of $\ell$ with $C$. The viewing angle $\theta_i'$ of $e_i'$ is strictly larger than $\theta_i$, yet the intersection of any line $\ell$ tangent to $c$ with $C$ has viewing angle exactly $\frac{2\pi}{3}$, and hence we conclude that $\theta_1 + \theta_2 + \theta_3 < 2\pi$, which is a contradiction.

So each $e_i$ is tangent to $c$, and in fact we can use a similar argument to conclude that each $e_i$ must be exactly the intersection of a line $\ell$ tangent to $c$ with $C$. (Otherwise, again we would have that $\theta_1 + \theta_2 + \theta_3 < 2\pi$.)

Hence, we conclude that each edge of $T$ has the same length, and each endpoint is on the boundary of $C$ so $T \in E$. $\quad\square$

Throughout the remainder of this section, consider any finite set $T_1, T_2, \ldots, T_n \in E$ of equilateral triangles, and let $S$ be the vertices of $\cap_{i=1}^n T_i$.

LEMMA 3.6 (see [2]). *Let $T$ be a triangle with $S \subset T \subset C$. Then $T \in \{T_1, T_2, \ldots, T_n\}$.*

*Proof.* Clearly, we have that $conv(S) \subset T$ since $T$ is convex, and we also have that $c \subset conv(S) = \cap_{i=1}^n T_i$. So by Lemma 3.5, we can conclude that $T$ must be in $E$. Suppose that $T \notin \{T_1, T_2, \ldots, T_n\}$.

Let $\{p_1, p_2, p_3\} = T \cap c$ (i.e., these are the three points on the boundary of $T$ closest to the origin). Similarly, for each $T_i$ let $\{p_1^i, p_2^i, p_3^i\} = T_i \cap c$. Then $\{p_1, p_2, p_3\}$ is a rotation (by $< \frac{2\pi}{3}$) of $\{p_1^i, p_2^i, p_3^i\}$, and hence $\{p_1, p_2, p_3\}$ are each strictly in the interior of $T_i$.

Hence, $\{p_1, p_2, p_3\}$ are on the boundary of $conv(S) \cap T$ but not on the boundary of $conv(S)$, so $T$ cannot contain $conv(S)$. $\quad\square$

LEMMA 3.7. *For each edge $e_j$ of a triangle $T_i$, $|e_j \cap S| = 2$ and, furthermore, for each $s \in S$, $s$ intersects the edges of exactly two (distinct) triangles in $\{T_1, T_2, \ldots, T_n\}$.*

*Proof.* Each edge of $conv(S)$ is by definition a subsegment of some unique edge $e_j$ of some triangle in $\{T_1, T_2, \ldots, T_n\}$. All we need to show is that to each edge $e_j$ (of some triangle in $\{T_1, T_2, \ldots, T_n\}$) we can find an edge of $conv(S)$ which is a subsegment of $e_j$:

Let $p_j$ be the closest point on $e_j$ to the origin. As we argued in Lemma 3.6, for all other triangles, $p_j$ is strictly in the interior. So the ray from the origin to $p_j$ hits the segment $e_j$ first (out of all edges of all triangles in $E$). Hence, $p_j$ is on the boundary of $conv(S)$, but only one edge (namely $e_j$) contains $p_j$ so the edge of $conv(S)$ that contains $p_j$ is a subsegment of $e_j$, as desired. $\quad\square$

COROLLARY 3.8. $|S| = 3n$.

As we noted, the gadget that we use here is a slight modification of the one in [2]—and the modification that we need involves rescaling.

DEFINITION 3.9. *For each triangle $T \in E$, define $T^{(1-\epsilon)}$ as the scaling down of $T$ such that the vertices of $T^{(1-\epsilon)}$ are on the boundary of $C_{1-\epsilon}$.*

This rescaling is precisely what ensures that the original instance is a **NO** instance, but as we will see, if $\epsilon$ is sufficiently small, then *every* small subset of $S$ is a **YES** instance. Note that for now the outer bounding region is not a polygon but a circle. We will work with this for the time being and fix it later by restricting to a polygon inside of $C_{1-\epsilon}$.

DEFINITION 3.10. *Let $S_i$ be the subset of $S$ that is in $T_i^{(1-\epsilon)}$.*

CLAIM 3.11. *If $\epsilon$ is sufficiently small, then $S_i = S - (\text{boundary}(T_i) \cap S)$.*

*Proof.* Recall that $conv(S) = \cap_{i=1}^n T_i$. Consider an edge $e_j$ of $T_i$. Using Lemma 3.7, we have $|e_j \cap S| = 2$. Let $H_1$ and $H_2$ be the hyperplanes whose boundary contains $e_j$ and $e_j^{(1-\epsilon)}$, respectively. Then we can choose $\epsilon$ small enough such that the re-

gion strictly between $H_1$ and $H_2$ does not contain any points in $S$, in which case $S_i = S - T_i \cap S$.   □

So consider the following instance of the intermediate polygon problem:

- Let $P = conv(\cup_i T_i^{(1-\epsilon)})$,
- and let $Q = \cap_{i=1}^n T_i$.

Moreover, $P$ is specified by its edges, and $S$ is a set of points so we choose it to be the vertices of $Q$.

CLAIM 3.12. $(P, S)$ is a **NO** instance.

*Proof.* $P \subset C_{1-\epsilon}$ by the definition of $T_i^{(1-\epsilon)}$, and using Lemma 3.5, any triangle $T$ contained in $C$ that contains $S$ must be in the set $E$; and since any triangle in $E$ has its vertices on the boundary of $C$, we conclude that $T$ is not contained in $C_{1-\epsilon}$ and hence $(P, S)$ is indeed unsatisfiable.   □

LEMMA 3.13. *For any $S' \subset S$ with $|S'| < n/2$, $(P, S')$ is a **YES** instance.*

*Proof.* Using Lemma 3.7, each $s \in S$ intersects exactly two edges of triangles in $\{T_1, T_2, \ldots, T_n\}$, so if $|S'| < n/2$, there must be a triangle $T_i$ for which $T_i \cap S' = \emptyset$.

Consider $T_i^{(1-\epsilon)}$: Using Claim 3.11, we conclude that $T_i^{(1-\epsilon)} \cap S' = S' - T_i \cap S' = S'$. And we have that $T_i^{(1-\epsilon)} \subset C_{1-\epsilon}$, so $(P, S')$ is indeed satisfiable.   □

We use the following lemma from Vavasis.

LEMMA 3.14 (see [44]). *Let $rank(M) = r$, and let $M = UV$, where $U$ and $V$ have $r$ columns and rows, respectively. Then $M$ has $rank^+(M) = r$ if and only if there is an invertible $r \times r$ matrix $Q$ such that $UQ^{-1}$ and $QV$ are both nonnegative.*

Vavasis gives a reduction from nonnegative rank to the intermediate simplex problem [44], but here we give a slight modification of this reduction that will make our exposition easier.

Consider the plane $F = \{(x, y, z) | x + y + z = 1\}$. Affinely map $P$ to this plane so that $P$ is contained in the nonnegative orthant (scale down $P$, if need be), and let the conic hull of vectors in $P$ and the origin be denoted by the cone $\mathcal{C}$.

Let $\mathcal{C} = \{\vec{v} | Av \geq 0\}$ and set the rows of $U$ to be the vertices of $F \cap \mathcal{C}$ and let $V = A^T$. Note that the vertices of $F \cap \mathcal{C}$ are just the three-dimensional coordinates corresponding to the vertices of $P$. Note that $UV$ is a nonnegative matrix, since each vertex of $F \cap \mathcal{C}$ is contained in the cone $\mathcal{C}$.

LEMMA 3.15. *There is an invertible $r \times r$ matrix $Q$ such that $UQ^{-1}$ and $QV$ are both nonnegative if and only if $(P, S)$ is a **YES** instance.*

*Proof.* Suppose $(P, S)$ is a **YES** instance. Let the rows of $Q$ be the three-dimensional coordinates of the vertices of the triangle $T$ (i.e., these are the vectors on the plane $F$). These points are in the cone $\mathcal{C}$, so $QV$ is nonnegative. Furthermore, $S \subset T$ so each row of $U$ is in the convex hull of rows of $Q$ and $UQ^{-1}$ is nonnegative.

Conversely, consider an invertible $Q$ for which $UQ^{-1}$ and $QV$ are both nonnegative. For each row in $Q$, let $p_i$ be the intersection of the ray through the origin and the row in $Q$ with $F$. $p_i \in \mathcal{C}$, so the associated *two-dimensional* point is in $P$. Furthermore, each row of $U$ is in the conic hull of $\{p_1, p_2, p_3\}$, and each $p_i$ and each row in $U$ has nonnegative entries and the sum of the entries is one. Hence, each $p_i$ and each row in $U$ has unit $\ell_1$ norm. So each row of $U$ is in the convex hull of $\{p_1, p_2, p_3\}$, and so the associated two-dimensional triangle contains $S$.   □

Note that in this reduction, rows of $M = UV$ are mapped one-to-one to points in $S$ and columns of $M$ are mapped one-to-one to facets in $P$. Hence, $(U, V)$ is a **NO** instance, but any set of $< n$ rows of $U$ is a **YES** instance.

So $M = UV$ is a nonnegative matrix of dimension $3n \times 3n$ with nonnegative rank $\geq 4$ and yet any submatrix of $< n$ rows has nonnegative rank $\leq 3$. We can boost this

construction as follows. Let $N$ be a $3rn \times 3rn$ matrix which is block diagonal, and has $M$ along the diagonal.

CLAIM 3.16. $rank^+(N) = rank^+(M)$.

*Proof.* Let $M^{(1)}, M^{(2)}, \ldots, M^{(r)}$ be the copies of $M$ along the diagonal. Consider a nonnegative matrix factorization $N = AW$ of minimum inner-dimension. We can write $N = \sum_{i=1}^{rank^+(N)} A_i W_i^T$. Then each rank one matrix $A_i W_i^T$ must have its support entirely contained in some $M^{(j)}$. And for any $M^{(j)}$, the set of $i$ where the support of $A_i W_i^T$ is contained the support of $M^{(j)}$ yields a nonnegative matrix factorization for $M^{(j)}$. This implies the claim. $\quad\square$

Using this claim, we have the following.

THEOREM 3.17. *For any $r \in \mathbb{N}$, there is a $3rn \times 3rn$ nonnegative matrix $N$ which has nonnegative rank at least $4r$ and yet for any $< n$ rows, the corresponding submatrix has nonnegative rank at most $3r$.*

An interesting open question is to characterize the family of matrices for which nonnegative rank *can* be certified by a small submatrix, since in many applications is quite natural to assume that the input matrices satisfy these conditions.

**Discussion.** The natural open question is to close the remaining algorithmic gap between the upper and lower bounds for computing the nonnegative rank. In fact, we believe that it is the lower bound that can be improved.

OPEN QUESTION 1. *Does computing the nonnegative rank require time $(nm)^{o(r^2)}$ under the exponential time hypothesis* [22]?

In particular, the construction in Arora et al. [2] uses only low-dimensional gadgets, and in the target factorization $A^U$ has only $O(1)$ nonzeros in each row. Thus, the approach we presented here would be able to use $O(r)$ variables instead of $O(r^2)$. However we believe that it should be possible to construct more powerful gadgets in higher dimensions beyond merely constructing disjoint low-dimensional gadgets and stitching them together, and this could in principle lead to better lower bounds.

REFERENCES

[1] N. ALON, *Tools from higher algebra*, in Handbook of Combinatorics, MIT Press, Cambridge, MA, 1996, pp. 1749–1783.

[2] S. ARORA, R. GE, R. KANNAN, AND A. MOITRA, *Computing a nonnegative matrix factorization–provably*, in Proceedings of STOC, 2012, pp. 145–162.

[3] S. ARORA, R. GE, AND A. MOITRA, *Learning topic models—going beyond SVD*, in Proceedings of FOCS, 2012, pp. 1–10.

[4] S. ARORA, R. GE, Y. HALPERN, D. MIMNO, A. MOITRA, D. SONTAG, Y. WU, AND M. ZHU, *A practical algorithm for topic modeling with provable guarantees*, in Proceedings of ICML, 2013, pp. 280–288.

[5] A. AHO, J. ULLMAN, AND M. YANNAKAKIS, *On notions of information transfer in VLSI circuits*, in Proceedings of STOC, 1983, pp. 133–139.

[6] S. BASU, R. POLLACK, AND M. ROY, *On the combinatorial and algebraic complexity of quantifier elimination*, J. ACM, 1996, pp. 1002–1045.

[7] L. BLUM, F. CUCKER, M. SHUB, AND S. SMALE, *Complexity of Real Computations*, Springer Verlag, New York, 1998.

[8] G. BRAUN, S. FIORINI, S. POKUTTA, AND D. STEURER, *Approximation limits of linear programs (beyond hierarchies)*, in Proceedings of FOCS, 2012, pp. 480–489.

[9] G. BRAUN AND S. POKUTTA, *Common information and unique disjointness*, in Proceedings of FOCS, 2013, pp. 688–697.

[10] M. BRAVERMAN AND A. MOITRA, *An information complexity approach to extended formula-tions*, in Proceedings of STOC, 2013, pp. 161–170.

[11] S. CHAN, J. LEE, P. RAGHAVENDRA, AND D. STEURER, *Approximate constraint satisfaction requires large LP relaxations*, in Proceedings of FOCS, 2013, pp. 360–369.

[12] J. COHEN AND U. ROTHBLUM, *Nonnegative ranks, decompositions and factorizations of non-negative matices*, Linear Algebra Appl., 1993, pp. 149–168.

[13] H. FAWZI, J. GOUVEIA, P. PARRILO, R. ROBINSON, AND R. THOMAS, *Positive semidefinite rank*, preprint, Math. Program. Ser. B, 153 (2015), pp. 133–177.

[14] S. FIORINI, S. MASSAR, S. POKUTTA, H. TIWARY, AND R. DE WOLF, *Linear vs semidefinite extended formulations: Exponential separations and strong lower bounds*, in Proceedings of STOC, 2012, pp. 95–106.

[15] S. FIORINI, T. ROTHVOSS, AND H. TIWARY, *Extended formulations for polygons*, Discrete Comput. Geom., 48 (2011), pp. 658–668.

[16] N. GILLIS, *Nonnegative Matrix Factorization: Complexity, Algorithms and Applications*, Ph.D. thesis, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium, 2011.

[17] G. GOLUB AND C. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.

[18] J. GOUVEIA, P. PARRILO, AND R. THOMAS, *Lifts of convex sets and cone factorizations*, Math. Oper. Res., 2013, pp. 248–264.

[19] L. HENRY, *Schémas de nuptialité: Déséquilibre des sexes et célibat*, Population, 24 (1969), pp. 457–486.

[20] T. HOFMANN, *Probabilistic latent semantic analysis*, in Proceedings of UAI, 1999, pp. 289–296.

[21] P. HRUBES, *On the nonnegative rank of distance matrices*, Inform. Process. Lett., 2012, pp. 457–461.

[22] R. IMPAGLIAZZO AND R. PATURI, *On the complexity of k-SAT*, J. Comput. System Sci., 62 (2001), pp. 367–375.

[23] J. LASSERRE, *Global optimization with polynomials and the problem of moments*, SIAM J. Optim., 11 (2001), pp. 796–817.

[24] W. LAWTON AND E. SYLVESTRE, *Self modeling curve resolution*, Technometrics, 13 (1971), pp. 617–633.

[25] J. LEE, P. RAGHAVENDRA, AND D. STEURER, *Lower bounds on the size of semidefinite pro-gramming relaxations*, in Proceedings of STOC, 2015, pp. 567–576.

[26] D. LEE AND H. SEUNG, *Learning the parts of objects by non-negative matrix factorization*, Nature, 1999, pp. 788–791.

[27] D. LEE AND H. SEUNG, *Algorithms for non-negative matrix factorization*, NIPS, 2000, pp. 556–562.

[28] L. LOVÁSZ AND M. SAKS, *Communication complexity and combinatorial lattice theory*, J. Comput. System Sci., 1993, pp. 322–349.

[29] J. MATOUSEK, *Lectures on Discrete Geometry*, Springer, New York, 2002.

[30] N. MNĚV, *The universality theorems on the classification problem of configuration varieties and convex polytopes varieties*, in Topology and Geometry, Springer, New York, 1988, pp. 527–544.

[31] A. MOITRA, *An almost optimal algorithm for computing nonnegative rank*, in Proceedings of SODA, 2013, pp. 1454–1464.

[32] J. NIE, K. RANESTAD, AND B. STURMFELS, *The algebraic degree of semidefinite programming*, Math. Program., 2010, pp. 379–405.

[33] N. NISAN, *Lower bounds for non-commutative computation (extended abstract)*, in Proceedings of STOC, 1991, pp. 410–418.

[34] P. PAATERO AND U. TAPPER, *Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values*, Environmetrics, 1994, pp. 111–126.

[35] C. PAPADIMITRIOU, P. RAGHAVAN, H. TAMAKI, AND S. VEMPALA, *Latent semantic indexing: A probabilistic analysis*, J. Comput. System Sci., 2000, pp. 217–235.

[36] P. PARRILO, *Structured Semidefinite Programs and Semialgebraic Geometry Method in Ro-bustness and Optimization*, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 2000.

[37] J. RENEGAR, *On the computational complexity and geometry of the first-order theory of the reals*, J. Symbolic Comput., 1992, pp. 255–352.

[38] J. RENEGAR, *On the computational complexity of approximating solutions for real algebraic formulae*, SIAM J. Comput., 1992, pp. 1008–1025.

[39] T. ROTHVOSS, *The matching polytope has exponential extension complexity*, in Proceedings of STOC, 2014, pp. 263–272.

[40] A. SEIDENBERG, *A new decision method for elementary algebra*, Ann. of Math, 60 (1954), pp. 365–374.

[41] P. SHOR, *Stretchabilty of pseudolines is NP-hard*, in Applied Geometry and Discrete Mathematics: The Victor Klee Festschrift, ACM, New York, 1991, pp. 531–554.

[42] A. TARSKI, *A Decision Method for Elementary Algebra and Geometry*, University of California Press, Berkeley, CA, 1951.

[43] W. XU AND X. LIU, AND Y. GONG, *Document clustering based on non-negative matrix factorization*, in Proceedings of SIGIR, 2003, pp. 267–273.

[44] S. VAVASIS, *On the complexity of nonnegative matrix factorization*, SIAM J. Optim., 2009, pp. 1364–1377.

[45] M. YANNAKAKIS, *Expressing combinatorial optimization problems by linear programs*, J. Comput. System Sci., 1991, pp. 441–466.