On Rationality of Nonnegative Matrix Factorization

Dmitry Chistikov*† Stefan Kiefer† Ines Marušić† Mahsa Shirmohammadi†
James Worrell†

Abstract

Nonnegative matrix factorization (NMF) is the problem of decomposing a given nonnegative $n \times m$ matrix M into a product of a nonnegative $n \times d$ matrix W and a nonnegative $d \times m$ matrix H. NMF has a wide variety of applications, including bioinformatics, chemometrics, communication complexity, machine learning, polyhedral combinatorics, among many others. A longstanding open question, posed by Cohen and Rothblum in 1993, is whether every rational matrix M has an NMF with minimal d whose factors W and H are also rational. We answer this question negatively, by exhibiting a matrix M for which W and H require irrational entries.

As an application of this result, we show that state minimization of labeled Markov chains can require the introduction of irrational transition probabilities.

We complement these irrationality results with an NP-complete version of NMF for which rational numbers suffice.

1 Introduction

Nonnegative matrix factorization (NMF) is the task of factoring a matrix of nonnegative real numbers M (henceforth a *nonnegative* matrix) as a product $M = W \cdot H$ such that the matrices W and H are also nonnegative. As well as being a natural problem in its own right [37, 11], NMF has found many applications in various domains, including machine learning, combinatorics, and communication complexity; see, e.g., [39, 22, 28, 41] and the references therein. In topic modeling [5], NMF corresponds to reconstructing documents in a given corpus as convex combinations of (i.e., distributions on) a small number of topics, where each topic is a distribution on words.

For an NMF $M = W \cdot H$, the number of columns in W is called the *inner dimension*. The smallest inner dimension of any NMF of M is called the *nonnegative rank (over the reals)* of M; an early reference is the paper by Gregory and Pullman [25]. Similarly, the *nonnegative rank of M over the rationals* is defined as the smallest inner dimension of an NMF $M = W \cdot H$ with matrices W, H that have only *rational* entries. Cohen and Rothblum [11] posed the following problem in 1993:

> "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."

In this paper, we solve the above problem by providing an example of a rational matrix M that has different nonnegative ranks over $\mathbb{R}$ and over $\mathbb{Q}$. In other words, any NMF $M = W \cdot H$ with minimal inner dimension has irrational entries in W and H. Our counterexample is almost optimal inasmuch as M has rank 4, whereas for matrices of rank at most 2 the nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ coincide [11]. This irrationality phenomenon carries over to other optimization problems: for example, as a corollary of our main result we show that minimizing labeled Markov chains may require irrational numbers as transition probabilities.

The above rationality problem of Cohen and Rothblum was reiterated by Vavasis [38] in the context of the computational complexity of computing the nonnegative rank of a matrix. Having shown NP-hardness of deciding whether the nonnegative rank of a matrix is upper-bounded by a given number, Vavasis asks whether this problem (and various related problems concerning NMF) is in NP [38, Section 5]. While our negative resolution of the Cohen–Rothblum problem does not exclude NP membership, it does rule out a hypothetical "simple" argument for membership in NP, wherein a certificate is an NMF with rational entries of small bit size.

Deciding whether a given matrix has nonnegative rank at most k is easily seen to be reducible to the decision problem for the existential theory of real closed fields and therefore belongs to PSPACE (see, e.g., [7]). Beyond this generic upper bound, the problem has been attacked from many different angles. Here we highlight the results of Moitra [28], who found semi-algebraic descriptions of the sets of matrices of nonnegative rank at most r in which the number of variables is $O(r^2)$, and Arora et al. [4], who identified several variants of the problem that are efficiently solvable. One such variant that has been widely studied, *separable NMF* [14], asks for factorizations $M = W \cdot H$ with the following unambiguity property: for every column of W there is a row whose only nonzero entry lies in that column. In

*Max Planck Institute for Software Systems (MPI-SWS).
†University of Oxford.

terms of topic modeling, this means that every topic has an *anchor word*, occurring in no other topic. If a matrix M has such a factorization then there is a rational one that can moreover be found in polynomial time [4].

Complementing our irrationality result, we introduce a different restriction of NMF, also with a natural interpretation in terms of unambiguity: we require that every column of M (document in the corpus) be reconstructed as a convex combination of columns of W without common nonzero entries (non-overlapping topics). We show that the decision problem for this NMF variant is NP-complete, and that among such factorizations there is always a rational one of minimal inner dimension.

Summary. Our main result (Section 4) is the resolution of the Cohen–Rothblum problem: nonnegative ranks over the reals and rationals may differ. As a corollary, in Section 5 we show that minimizing labeled Markov chains may require irrational numbers. Finally, in Section 6 we define an NP-complete variant of nonnegative rank, called unambiguous rank, which is the same over the reals and rationals.

Full details of proofs of Section 4 can be found in [9].

2 Related Work and Main Ideas

Rationality of solutions to optimization problems is a central topic in computer science, in part due to its connection to the computational complexity of finding optimal solutions. For various computational problems associated with partially observable Markov decision processes [40], Nash equilibria [18], concurrent stochastic games [17], and Shapley's stochastic games [33], it has been shown that irrational numbers may be needed to express optimal solutions. In conjunction with this, complexity-theoretic lower bounds have been derived, e.g., hardness for the SQRT-SUM problem.

In the last few years, there has been progress towards resolving the Cohen–Rothblum problem. It was already known to Cohen and Rothblum [11] that the nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ coincide for matrices of rank at most 2. (Note that the usual ranks over $\mathbb{R}$ and $\mathbb{Q}$ coincide for all rational matrices.) In 2015, Kubjas et al. [26, Corollary 4.6] extended this result to matrices of nonnegative rank (over $\mathbb{R}$) at most 3. On the other hand, Shitov [34] proved that the nonnegative rank of a matrix can indeed depend on the underlying field: he exhibited a nonnegative matrix with irrational entries whose nonnegative rank over a subfield of $\mathbb{R}$ is different from its nonnegative rank over $\mathbb{R}$. Independently and concurrently with our work [9], Shitov [35] has proposed a rational matrix whose nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ are different.

Recently, the authors of this paper obtained a related result on the restricted nonnegative rank, a notion introduced and studied by Gillis and Glineur [23]. The *restricted nonnegative rank* of a nonnegative matrix M is the smallest inner dimension of an NMF $M = W \cdot H$ such that the columns of W span the same vector space as the columns of M (note that for any NMF $M = W \cdot H$ the column space of W contains the one of M). In [10], we exhibited a rational matrix whose *restricted* nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ differ.

Our approach in the present paper draws on the ideas from [10] and extends them in a substantial way. It has long been known [11] that NMF can be interpreted geometrically as finding a set of vectors (columns of W) inside a unit simplex whose convex hull contains a given set of points (columns of M). The common part of [10] and the present paper finds, in the *restricted* setting (i.e., where the columns of W and M span the same vector space), a rigid structure that exhibits irrationality. More specifically, we construct a rational matrix M' whose *restricted* NMF $M' = W \cdot H'$ with minimal inner dimension is *unique* (up to permutation and rescaling of columns of W) and has irrational entries in the factors W and H'. If we knew that this factorization were unique among *all* NMFs, we would be done. This, however, requires ruling out several classes of other hypothetical factorizations of the matrix.

Towards this goal, one might want to take advantage of the work on uniqueness properties of NMF, studied, for instance, by Thomas [37], Laurberg et al. [27], and Gillis [21]. Here we pursue a different strategy. We show that for a larger matrix $M = \begin{pmatrix} M' & W_\varepsilon \end{pmatrix}$, where W_ε is a nonnegative rational matrix which is entry-wise close to W, the only NMF (restricted or otherwise) of the same inner dimension has the same left factor W—thus extending the uniqueness property to the *non-restricted* setting.

The guiding idea behind our construction of M is that by including all columns of W into the set of columns of M we can exclude certain "undesirable" factorizations, thereby ensuring that M has no rational NMF. We show this by a constraint propagation argument. Unfortunately for this construction, the matrix W itself has irrational entries. However, we show that we can instead take any nonnegative *rational* matrix W_ε within a sufficiently small neighbourhood of W, and the undesirable factorizations will still be excluded. In the text we describe such a neighbourhood explicitly and pick a specific rational matrix W_ε from it, thus obtaining the matrix M of the above form and proving the main result of the paper.

Conceptually, the existence of a suitable matrix W_ε can be understood in terms of *upper semi-continuity* of the nonnegative rank over $\mathbb{R}$, proved by Bocci et al. [6]. By this property, if a matrix M has nonnegative rank r over $\mathbb{R}$, then all nonnegative matrices in a sufficiently small neighborhood of M have nonnegative rank r or greater (over $\mathbb{R}$). In the same manner, our proof extends the non-existence of undesirable factorizations from the matrix W to W_ε.

3 Preliminaries

Given an ordered field $\mathbb{F}$, we denote by $\mathbb{F}_+$ the set of its nonnegative elements. Given a vector $v \in \mathbb{F}^n$, we write v_i for its ith entry. Such a vector v is called *pseudo-stochastic* if its entries sum up to one. A pseudo-stochastic vector v is called *stochastic* if its entries are also nonnegative.

For any matrix M, we write $M_{i,:}$ for its ith row, $M_{:,j}$ for its jth column, and $M_{i,j}$ for its (i,j)th entry. Given non-empty subsets I and J of the rows and columns of M, respectively, we write $M_{I,J}$ for the submatrix $(M_{i,j})_{i\in I, j\in J}$ of M.

A matrix is called *nonnegative* (resp., *zero* or *rational*) if so are all its entries. A nonnegative matrix is *column-stochastic* (resp., *row-stochastic*) if its columns (rows) are stochastic. Column-stochastic matrices will henceforth simply be called stochastic matrices.

3.1 Nonnegative Rank.

Let $\mathbb{F}$ be an ordered field, such as the reals $\mathbb{R}$ or the rationals $\mathbb{Q}$. Given a nonnegative matrix $M \in \mathbb{F}_+^{n\times m}$, a *nonnegative matrix factorization (NMF) over* $\mathbb{F}$ of M is any representation of the form $M = W \cdot H$ where $W \in \mathbb{F}_+^{n\times d}$ and $H \in \mathbb{F}_+^{d\times m}$ are nonnegative matrices. We refer to d as the *inner dimension* of the NMF, and hence refer to the NMF $M = W \cdot H$ as being *d-dimensional*.

The *nonnegative rank over* $\mathbb{F}$ of M is the smallest nonnegative integer d such that there exists a d-dimensional NMF over $\mathbb{F}$ of M. An equivalent characterization [11] of the nonnegative rank of M over $\mathbb{F}$ is as the smallest number d such that M is equal to the sum of d rank-1 matrices in $\mathbb{F}_+^{n\times m}$. It is easy to see that

$$\mathrm{rank}(M) \le \mathrm{rank}_+(M) \le \min\{n, m\}$$

where $\mathrm{rank}(M)$ and $\mathrm{rank}_+(M)$ denote the rank and the nonnegative rank, respectively.

Given a nonzero matrix $M \in \mathbb{F}_+^{n\times m}$, by removing the zero columns of M and dividing each remaining column by the sum of its elements, we obtain a stochastic matrix with equal nonnegative rank. Similarly, if $M = W \cdot H$ then after removing the zero columns in W and multiplying with a suitable diagonal matrix D, we get

$$M = W \cdot H = WD \cdot D^{-1}H$$

where WD is stochastic. If M is stochastic then (writing $\mathbf{1}$ for a row vector of ones) we have

$$\mathbf{1} = \mathbf{1}M = \mathbf{1}WD \cdot D^{-1}H = \mathbf{1}D^{-1}H,$$

hence $D^{-1}H$ is stochastic as well. Thus, without loss of generality one can consider NMFs $M = W \cdot H$ in which M, W, and H are all stochastic matrices [11, Theorem 3.2]. In such a case, we will call the factorization $M = W \cdot H$ *stochastic*. An NMF $M = W \cdot H$ is called *rational* if matrices W and H are rational.

In this paper, our main goal is to compare the respective nonnegative ranks of rational matrices M over $\mathbb{R}$ and $\mathbb{Q}$. The nonnegative rank over $\mathbb{R}$ will henceforth simply be called nonnegative rank.

3.2 Nested Polygons.

In this paper, all polygons are assumed to be convex. Given two polygons, $\mathcal{R} \subseteq \mathcal{P} \subseteq \mathbb{R}^2$, a polygon $\mathcal{Q} \subseteq \mathbb{R}^2$ is said to be *nested between* $\mathcal{R}$ *and* $\mathcal{P}$ if $\mathcal{R} \subseteq \mathcal{Q} \subseteq \mathcal{P}$. Such a polygon is said to be *minimal* if it has the minimum number of vertices among all polygons nested between $\mathcal{R}$ and $\mathcal{P}$. In this subsection we recall from [2] a standardized form for minimal nested polygons, which will play an important role in the subsequent development.

Fix two polygons $\mathcal{R}$ and $\mathcal{P}$, with $\mathcal{R} \subseteq \mathcal{P}$. A *supporting line segment* is a directed line segment that has both endpoints on the boundary of the outer polygon $\mathcal{P}$ and touches the inner polygon $\mathcal{R}$ on its left. A nested polygon with vertices $q_1, \ldots, q_k$, listed in anti-clockwise order, is said to be *supporting* if the directed line segments $q_1q_2, q_2q_3, \ldots, q_{k-1}q_k$ are all supporting. Such a polygon is uniquely determined by the vertex q_1 (see [2, Section 2]) and is henceforth denoted by $\mathcal{S}_{q_1}$. It is shown in [2] that one can always find a nested polygon that is both minimal and supporting. More specifically, from [2, Lemma 4] we have:

LEMMA 3.1. *Consider a minimal nested polygon with vertices $q_1, \ldots, q_k$, listed in anti-clockwise order, where q_1 lies on the boundary of $\mathcal{P}$. The supporting polygon $\mathcal{S}_{q_1}$ is also minimal.*

4 NMF Requires Irrational Numbers

In this section we give the main contribution of this paper: a rational matrix whose respective nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ are different.

THEOREM 4.1. *Let* $M = \begin{pmatrix} M' & W_\varepsilon \end{pmatrix} \in \mathbb{Q}_+^{6\times 11}$, *where* *matrices $M' \in \mathbb{Q}_+^{6\times 6}$ and $W_\varepsilon \in \mathbb{Q}_+^{6\times 5}$ are as follows:*

$$M' = \begin{pmatrix} \frac{5}{44} & \frac{5}{11} & \frac{85}{121} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{11} & \frac{3}{11} & \frac{7}{33} \\ \frac{1}{11} & \frac{1}{44} & \frac{2}{121} & \frac{1}{44} & \frac{15}{88} & \frac{17}{88} \\ \frac{1}{44} & \frac{1}{44} & \frac{8}{121} & \frac{1}{44} & \frac{19}{88} & \frac{5}{24} \\ \frac{3}{11} & \frac{3}{11} & \frac{12}{121} & \frac{8}{11} & \frac{2}{11} & \frac{2}{33} \\ \frac{1}{2} & \frac{5}{22} & \frac{14}{121} & \frac{1}{22} & \frac{7}{44} & \frac{43}{132} \end{pmatrix},$$

$$W_\varepsilon = \begin{pmatrix} 0 & \frac{133}{165} & \frac{640}{2233} & 0 & 0 \\ \frac{1}{111540} & 0 & 0 & \frac{17209}{58047} & \frac{997}{5082} \\ \frac{114721}{892320} & \frac{1}{146850} & \frac{17}{506} & \frac{385}{1759} & \frac{2921}{203280} \\ \frac{47}{1248} & \frac{413}{5874} & \frac{1}{102718} & \frac{2915}{10554} & \frac{4381}{203280} \\ \frac{36}{169} & \frac{22}{267} & \frac{18674}{51359} & \frac{1}{116094} & \frac{3252}{4235} \\ \frac{276953}{446160} & \frac{1009}{24475} & \frac{16239}{51359} & \frac{1100}{5277} & \frac{1}{101640} \end{pmatrix}.$$

The nonnegative rank of M over $\mathbb{R}$ is 5, while the nonnegative rank of M over $\mathbb{Q}$ is 6.

In [10] the authors of this paper exhibited a matrix (which can be obtained from M' by rescaling and permuting rows and columns) whose respective *restricted* nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ (cf. Section 2) are different. Theorem 4.1 refers to general nonnegative rank.

In the rest of this section we give an overview of the proof of Theorem 4.1. Full details are given in [9].

Note that matrix M is stochastic. As argued in Section 3.1, we only need to consider stochastic NMFs of M in order to determine its nonnegative rank.

The trivial factorization $M = I \cdot M$, with I the 6×6 identity matrix, shows that the nonnegative rank of M is at most 6 over both $\mathbb{Q}$ and $\mathbb{R}$. Next we show that M has nonnegative rank at most 5 over $\mathbb{R}$. Indeed matrix M' has a 5-dimensional NMF $M' = W \cdot H'$, where matrix W is:

$$\begin{pmatrix} 0 & \frac{5}{7} + \frac{5\sqrt{2}}{77} & \frac{15+5\sqrt{2}}{77} & 0 & 0 \\ 0 & 0 & 0 & \frac{20+2\sqrt{2}}{77} & \frac{48-8\sqrt{2}}{187} \\ \frac{\sqrt{2}}{11} & 0 & \frac{4-\sqrt{2}}{77} & \frac{3}{14} + \frac{\sqrt{2}}{308} & \frac{14-8\sqrt{2}}{187} \\ \frac{-1+\sqrt{2}}{11} & \frac{4+\sqrt{2}}{77} & 0 & \frac{39}{154} + \frac{5\sqrt{2}}{308} & \frac{21-12\sqrt{2}}{187} \\ \frac{8-4\sqrt{2}}{11} & \frac{12-4\sqrt{2}}{77} & \frac{4}{11} & 0 & \frac{104+28\sqrt{2}}{187} \\ \frac{4+2\sqrt{2}}{11} & \frac{6-2\sqrt{2}}{77} & \frac{30-4\sqrt{2}}{77} & \frac{3}{11} - \frac{\sqrt{2}}{22} & 0 \end{pmatrix}$$

and matrix H' is:

$$\begin{pmatrix} \frac{1+\sqrt{2}}{4} & 0 & \frac{\sqrt{2}}{11} & \frac{1}{4} - \frac{\sqrt{2}}{8} & 0 & \frac{1}{6} + \frac{\sqrt{2}}{12} \\ 0 & \frac{1}{2} - \frac{\sqrt{2}}{8} & 1 - \frac{\sqrt{2}}{11} & 0 & 0 & 0 \\ \frac{3-\sqrt{2}}{4} & \frac{1}{2} + \frac{\sqrt{2}}{8} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{21}{34} + \frac{7\sqrt{2}}{68} & \frac{5}{6} - \frac{\sqrt{2}}{12} \\ 0 & 0 & 0 & \frac{3}{4} + \frac{\sqrt{2}}{8} & \frac{13}{34} - \frac{7\sqrt{2}}{68} & 0 \end{pmatrix}$$

Moreover, matrix W_ε has a 5-dimensional NMF $W_\varepsilon = W \cdot H_\varepsilon$ with H_ε given in Figure 1. Thus matrix M has the following stochastic 5-dimensional NMF:

$$M = W \cdot \begin{pmatrix} H' & H_\varepsilon \end{pmatrix}. \tag{4.1}$$

Notice that the matrix W_ε, which corresponds to the last 5 columns of M, can be seen as a perturbation of the left factor W in (4.1). Intuitively, we have included W_ε in M in order to rule out 5-dimensional NMFs of M other than (4.1).

Factorization (4.1) shows that M has nonnegative rank at most 5 over $\mathbb{R}$. Crucially the entries of the factors W, H', and H_ε are irrational, and the core of the proof of Theorem 4.1 consists in showing that M has no 5-dimensional NMF over $\mathbb{Q}$.

To this end, we will classify NMFs $M = L \cdot R$ of inner dimension 4 and 5 into four types according to the "zero pattern" in the first two rows of the left factor L (illustrated in Figure 2 for inner dimension 5). In Lemma 4.1 we formally define these four NMF types, and show (using the occurrence of the zero entries in M) that any NMF of M of inner dimension at most 5 matches one of these four types. As argued in Section 3.1, it suffices to consider stochastic NMFs of M.

LEMMA 4.1. *Let M be the matrix from Theorem 4.1. Let $M = L \cdot R$ be a stochastic and at most 5-dimensional NMF. Consider the following notation:*

- *k is the number of columns in L whose first and second coordinates are 0,*
- *k_1 is the number of columns in L whose first coordinate is strictly positive and second coordinate is 0, and*
- *k_2 is the number of columns in L whose second coordinate is strictly positive and first coordinate is 0.*

Then, the NMF $M = L \cdot R$ has one of the following four types*:*

1) *$k = 1$, $k_1 = 2$, $k_2 = 2$;*
2) *$k = 2$, $k_1 = 1$;*
3) *$k = 2$, $k_2 = 1$;*
4) *$k = 3$, $k_1 = 1$, $k_2 = 1$.*

Proof. The NMF $M = L \cdot R$ corresponds to representing each column of M as a convex combination of the columns of L, with the coefficients of the convex combination specified by the entries of the right factor R. As L has at most 5 columns,

$$k + k_1 + k_2 \le 5. \tag{4.2}$$

$$\begin{pmatrix} \frac{30419}{40560}+\frac{28679\sqrt{2}}{162240} & \frac{-2728}{46725}+\frac{5791\sqrt{2}}{140175} & \frac{2741}{98049}-\frac{642\sqrt{2}}{32683} & \frac{-689}{10554}+\frac{15595\sqrt{2}}{337728} & \frac{389}{1848}-\frac{5501\sqrt{2}}{36960} \\ 0 & \frac{163318}{140175}-\frac{7277\sqrt{2}}{62300} & \frac{5958}{32683}-\frac{50543\sqrt{2}}{392196} & 0 & 0 \\ 0 & \frac{-2137}{20025}+\frac{6047\sqrt{2}}{80100} & \frac{11062}{14007}+\frac{8321\sqrt{2}}{56028} & 0 & 0 \\ \frac{7443}{8840}-\frac{51313\sqrt{2}}{86190} & 0 & 0 & \frac{148897}{179418}+\frac{172627\sqrt{2}}{1435344} & \frac{-1741}{26180}+\frac{1847\sqrt{2}}{39270} \\ \frac{-408157}{689520}+\frac{1154473\sqrt{2}}{2758080} & 0 & 0 & \frac{7039}{29903}-\frac{318541\sqrt{2}}{1913792} & \frac{134461}{157080}+\frac{1163\sqrt{2}}{11424} \end{pmatrix}$$

Figure 1: Matrix H_ε.

Since the columns $M_{:,1}, M_{:,2}, M_{:,3}$ are linearly independent, matrix L has at least three columns whose second coordinate is 0. Likewise, since the columns $M_{:,4}, M_{:,5}, M_{:,6}$ are linearly independent, L has at least three columns whose first coordinate is 0. That is,

$$k + k_1 \geq 3 \quad \text{and} \quad k + k_2 \geq 3. \tag{4.3}$$

Together with (4.2), this implies that

$$2k \geq 6 - k_1 - k_2 \geq 1 + k$$

and therefore $k \geq 1$.

The columns $M_{:,1}, M_{:,2}, M_{:,3}$ have first coordinate strictly positive and second coordinate 0, while the columns $M_{:,4}, M_{:,5}, M_{:,6}$ have second coordinate strictly positive and first coordinate 0. Therefore, in order for these columns to be covered by the columns of L, we need to have:

$$k_1 \geq 1 \quad \text{and} \quad k_2 \geq 1. \tag{4.4}$$

Together with (4.2), this implies that

$$k \leq 5 - k_1 - k_2 \leq 3.$$

We conclude that $k \in \{1, 2, 3\}$. The result follows from inequalities (4.2), (4.3), and (4.4). □

We observe from Lemma 4.1 that any type-1 or type-4 NMF of M has inner dimension 5, while any type-2 or type-3 NMF of M has inner dimension 4 or 5.

In the rest of this section, we give separate arguments excluding rational NMFs of M of each of the four types.[1] These arguments are sketched below, with full details for all four types in [9]. Our arguments, in particular, show that M has no 4-dimensional NMF, rational or not, thus showing that M has nonnegative rank exactly 5 over $\mathbb{R}$, and concluding the proof of Theorem 4.1.

[1] In fact one can show [9] that factorization (4.1), which has type 1, is the unique at most 5-dimensional NMF of M over $\mathbb{R}$.

4.1 Geometry behind the Proof of Theorem 4.1.

To rule out 5-dimensional rational NMFs of M of types 1, 2, and 3, we employ geometric arguments concerning nested polygons in the plane (see Sections 4.2 and 4.3). These arguments rely on a geometric interpretation of the specific NMF $M = W \cdot \begin{pmatrix} H' & H_\varepsilon \end{pmatrix}$ that was given in (4.1). More precisely, we define a polytope $\mathcal{P} \subseteq \mathbb{R}^3$, shown in Figure 3, such that each of the columns of M and W can be associated with points in $\mathcal{P}$, with the points corresponding to the columns of M lying in the convex hull of those corresponding to the columns of W. This construction ultimately relies on a general geometric interpretation of NMF in terms of nested polytopes, introduced in [23].

4.1.1 Nested Polytopes.

To set up this geometric connection, observe that matrix M is stochastic and has rank 4, and hence the columns of M affinely span a 3-dimensional affine subspace $\mathcal{V} \subseteq \mathbb{R}^6$. All vectors in $\mathcal{V}$ are pseudo-stochastic, i.e., their entries sum to 1. The stochastic vectors in $\mathcal{V}$ (equivalently, the nonnegative vectors in $\mathcal{V}$) form a 3-dimensional polytope, say $\mathcal{P}' \subseteq \mathcal{V}$. Clearly we have $M_{:,i} \in \mathcal{P}'$ for each $i \in \{1, \ldots, 11\}$.

We will now fix a particular parameterization of $\mathcal{V}$ and $\mathcal{P}'$; that is, we define an injective affine function $f : \mathbb{R}^3 \to \mathbb{R}^6$ and a polytope $\mathcal{P} \subseteq \mathbb{R}^3$ such that

$$f(\mathbb{R}^3) = \mathcal{V} \qquad \text{and} \qquad f(\mathcal{P}) = \mathcal{P}'. \tag{4.5}$$

Specifically, define $f(x) = Cx + d$, where

$$C = \frac{1}{11} \cdot \begin{pmatrix} 0 & 10 & 0 \\ 0 & 0 & 4 \\ -1 & -2 & 1/2 \\ -1 & 0 & 5/2 \\ 4 & 0 & 0 \\ -2 & -8 & -7 \end{pmatrix} \quad \text{and} \quad d = \frac{1}{11} \cdot \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 8 \end{pmatrix}.$$

Furthermore, write $\mathcal{P} = \{\, x \in \mathbb{R}^3 \mid f(x) \geq 0 \,\}$.

$$
\begin{array}{cccc}
\text{type 1} & \text{type 2} & \text{type 3} & \text{type 4} \\
\begin{pmatrix} 0 & + & + & 0 & 0 \\ 0 & 0 & 0 & + & + \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} &
\begin{pmatrix} 0 & 0 & + & \cdot & \cdot \\ 0 & 0 & 0 & + & + \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} &
\begin{pmatrix} 0 & 0 & 0 & + & + \\ 0 & 0 & + & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} &
\begin{pmatrix} + & 0 & 0 & 0 & 0 \\ 0 & + & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}
\end{array}
$$

Figure 2: In any 5-dimensional NMF $M = L \cdot R$, matrix L matches one of the four patterns above, up to a permutation of its columns. Here $+$ denotes any strictly positive number.

Now, defining

$$
r_1 = \begin{pmatrix} 3/4 \\ 1/8 \\ 0 \end{pmatrix},\; r_2 = \begin{pmatrix} 3/4 \\ 1/2 \\ 0 \end{pmatrix},\; r_3 = \begin{pmatrix} 3/11 \\ 17/22 \\ 0 \end{pmatrix},
$$

$$
r_4 = \begin{pmatrix} 2 \\ 0 \\ 1/2 \end{pmatrix},\; r_5 = \begin{pmatrix} 1/2 \\ 0 \\ 3/4 \end{pmatrix},\; r_6 = \begin{pmatrix} 1/6 \\ 0 \\ 7/12 \end{pmatrix},
$$

we have that $f(r_i) = M'_{:,i}$ for each $i \in \{1, 2, \ldots, 6\}$. Moreover, defining

$$
q_1^\varepsilon = \begin{pmatrix} \frac{99}{169} \\ 0 \\ \frac{1}{40560} \end{pmatrix},\; q_2^\varepsilon = \begin{pmatrix} \frac{121}{534} \\ \frac{133}{150} \\ 0 \end{pmatrix},\; q_3^\varepsilon = \begin{pmatrix} \frac{9337}{9338} \\ \frac{64}{203} \\ 0 \end{pmatrix},
$$

$$
q_4^\varepsilon = \begin{pmatrix} \frac{1}{42216} \\ 0 \\ \frac{17209}{21108} \end{pmatrix},\; q_5^\varepsilon = \begin{pmatrix} \frac{813}{385} \\ 0 \\ \frac{997}{1848} \end{pmatrix},
$$

we have that $f(q_i^\varepsilon) = (W_\varepsilon)_{:,i}$ for each $i \in \{1, 2, \ldots, 5\}$. Thus, all the columns of matrix M lie in the image of f. It immediately follows that $f(\mathbb{R}^3) = \mathcal{V}$, and thence that $f(\mathcal{P}) = \mathcal{P}'$.

Moreover we have that $r_i \in \mathcal{P}$, as $f(r_i) = M'_{:,i} \in \mathcal{P}'$ for all $i \in \{1, 2, \ldots, 6\}$. Likewise we have $q_i^\varepsilon \in \mathcal{P}$, as $f(q_i^\varepsilon) = (W_\varepsilon)_{:,i} \in \mathcal{P}'$ for all $i \in \{1, 2, \ldots, 5\}$.

Figure 3 depicts $\mathcal{P}$, which has 6 faces corresponding to the inequalities of the system $Cx + d \geq 0$. In more detail, $\mathcal{P}$ is the intersection of the following half-spaces: $y \geq 0$ (blue), $z \geq 0$ (brown), $-\frac{1}{2}x - y + \frac{1}{4}z + 1 \geq 0$ (green), $-x + \frac{5}{2}z + 1 \geq 0$ (yellow), $x \geq 0$ (pink), $-\frac{1}{4}x - y - \frac{7}{8}z + 1 \geq 0$ (transparent front). The figure also shows the position of the points $r_1, \ldots, r_6$ (black dots). [2]

Next we observe that the columns of matrix W in (4.1) are also in $\mathcal{P}'$. Indeed, defining

$$
q_1^* = \begin{pmatrix} 2 - \sqrt{2} \\ 0 \\ 0 \end{pmatrix},\; q_2^* = \begin{pmatrix} \frac{3-\sqrt{2}}{7} \\ \frac{11+\sqrt{2}}{14} \\ 0 \end{pmatrix},\; q_3^* = \begin{pmatrix} 1 \\ \frac{3+\sqrt{2}}{14} \\ 0 \end{pmatrix},
$$

$$
q_4^* = \begin{pmatrix} 0 \\ 0 \\ \frac{10+\sqrt{2}}{14} \end{pmatrix},\; q_5^* = \begin{pmatrix} \frac{26+7\sqrt{2}}{17} \\ 0 \\ \frac{12-2\sqrt{2}}{17} \end{pmatrix},
$$

we have $f(q_i^*) = W_{:,i} \in \mathcal{P}'$ and hence $q_i^* \in \mathcal{P}$ for each $i \in \{1, 2, \ldots, 5\}$. That is, the NMF $M = W \cdot \begin{pmatrix} H' & H_\varepsilon \end{pmatrix}$ in (4.1) is restricted: the columns of M and the columns of W span the same vector space.

Applying the inverse of the map f column-wise to the NMFs $M' = W \cdot H'$ and $W_\varepsilon = W \cdot H_\varepsilon$, we obtain respectively

$$
\begin{pmatrix} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 \end{pmatrix} = \begin{pmatrix} q_1^* & q_2^* & q_3^* & q_4^* & q_5^* \end{pmatrix} \cdot H' \tag{4.6}
$$

and

$$
\begin{pmatrix} q_1^\varepsilon & q_2^\varepsilon & q_3^\varepsilon & q_4^\varepsilon & q_5^\varepsilon \end{pmatrix} = \begin{pmatrix} q_1^* & q_2^* & q_3^* & q_4^* & q_5^* \end{pmatrix} \cdot H_\varepsilon\,. \tag{4.7}
$$

Recall that the matrix H' is stochastic; hence equations (4.6) and (4.7) imply that the points r_i and q_i^ε are contained in the convex hull of the points q_i^*. In Figure 3, points q_1^*, q_2^*, q_3^* are the vertices of the triangle on the brown xy-face, while points q_1^*, q_4^*, q_5^* are the vertices of the triangle on the blue xz-face. The former triangle contains r_1, r_2, r_3, while the latter triangle contains r_4, r_5, r_6. Points $q_1^\varepsilon, \ldots, q_5^\varepsilon$ (not shown in Figure 3) are close to $q_1^*, \ldots, q_5^*$, with $q_2^\varepsilon, q_3^\varepsilon$ lying in the interior of the triangle on the xy-face and $q_1^\varepsilon, q_4^\varepsilon, q_5^\varepsilon$ lying in the interior of the triangle on the xz-face.

It is important to note that when we exclude certain NMFs $M = L \cdot R$ in the following subsections, we cannot a priori assume that the columns of L are in $\mathcal{V}$.

4.1.2 Nested Polygons.

In this subsection, we focus on the two faces of polytope $\mathcal{P}$ that contain the interior points $r_1, \ldots, r_6$, respectively called $\mathcal{P}_0$ and $\mathcal{P}_1$.

[2] In [10], the authors of this paper used the same polytope $\mathcal{P}$ and the same points $r_1, \ldots, r_6$ (see Figure 3) to prove a related result about the *restricted* nonnegative rank.

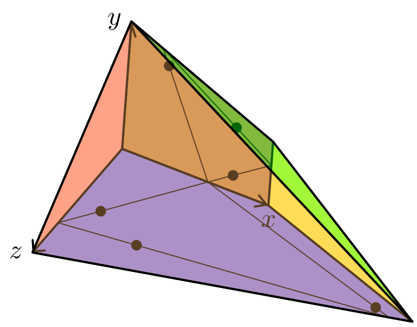

Figure 3: The two images show orthogonal projections of the 3-dimensional polytope $\mathcal{P}$. The black dots indicate 6 interior points: 3 points (r_1, r_2, r_3) on the brown xy-face, and 3 points (r_4, r_5, r_6) on the blue xz-face. The two slightly different projections are designed to create a 3-dimensional impression using stereoscopy. The "parallel-eye" technique should be used, see, e.g., [36].

Let us write $\mathcal{V}_0 \subseteq \mathbb{R}^6$ for the affine span of column vectors $M_{:,1}, M_{:,2}, M_{:,3}$. We can also characterize $\mathcal{V}_0$ as the image of the xy-plane in $\mathbb{R}^3$ under the map $f : \mathbb{R}^3 \to \mathbb{R}^6$. Indeed, we have that $f(r_1) = M_{:,1}$, $f(r_2) = M_{:,2}$, and $f(r_3) = M_{:,3}$. Thus the image of the xy-plane under f is a 2-dimensional affine space that includes $\mathcal{V}_0$ and hence is equal to $\mathcal{V}_0$. Define the polygon $\mathcal{P}_0 \subseteq \mathbb{R}^3$ by

$$\mathcal{P}_0 = \{(x, y, 0)^\top : (x, y, 0)^\top \in \mathcal{P}\}.$$

Then, f restricts to a bijection between $\mathcal{P}_0$ and the set of nonnegative vectors in $\mathcal{V}_0$. We have the following lemma:

Lemma 4.2. *Let $\mathcal{R} \subseteq \mathcal{P}_0$ be the triangle with vertices r_1, r_2, r_3 (see Figure 4). If $q_1 = (u, 0, 0)^\top$, where $0 \leq u < 2 - \sqrt{2}$, then the supporting polygon $\mathcal{S}_{q_1}$ has more than 3 vertices.*

Let us write $\mathcal{V}_1 \subseteq \mathbb{R}^6$ for the affine span of column vectors $M_{:,4}, M_{:,5}, M_{:,6}$. We can also characterize $\mathcal{V}_1$ as the image of the xz-plane in $\mathbb{R}^3$ under the map $f : \mathbb{R}^3 \to \mathbb{R}^6$. Indeed, we have that $f(r_4) = M_{:,4}$, $f(r_5) = M_{:,5}$, and $f(r_6) = M_{:,6}$. Thus the image of the xz-plane under f is a 2-dimensional affine space that includes $\mathcal{V}_1$ and hence is equal to $\mathcal{V}_1$. Define the polygon $\mathcal{P}_1 \subseteq \mathbb{R}^3$ by

$$\mathcal{P}_1 = \{(x, 0, z)^\top : (x, 0, z)^\top \in \mathcal{P}\}.$$

Then, f restricts to a bijection between $\mathcal{P}_1$ and the set of nonnegative vectors in $\mathcal{V}_1$. We have the following lemma:

Lemma 4.3. *Let $\mathcal{R} \subseteq \mathcal{P}_1$ be the triangle with vertices r_4, r_5, r_6 (see Figure 5). If $q_1 = (u, 0, 0)^\top$, where $2 - \sqrt{2} < u \leq 1$, then the supporting polygon $\mathcal{S}_{q_1}$ has more than 3 vertices.*

4.2 Ruling out Rational Type-1 NMFs. In this subsection, we argue that any type-1 NMF of M requires irrational numbers (our argument will, in fact, only depend on the matrix M' and not on W_ε).

Fix a type-1 NMF $M = L \cdot R$, i.e., such that L matches the type-1 zero pattern in Figure 2. Considering the respective zero patterns of M and L it is clear that: (i) $M_{:,1}, M_{:,2}, M_{:,3}$ all lie in the convex hull of $L_{:,1}, L_{:,2}, L_{:,3}$, and (ii) $M_{:,4}, M_{:,5}, M_{:,6}$ all lie in the convex hull of $L_{:,1}, L_{:,4}, L_{:,5}$. At a high level, our proof strategy is to use facts (i) and (ii) to respectively derive two inequalities which together force $L = W$ for W the left factor in (4.1).

The affine span of column vectors $L_{:,1}, L_{:,2}, L_{:,3}$ includes $\mathcal{V}_0$ and has dimension at most two, and hence is equal to $\mathcal{V}_0$. In particular, $L_{:,1}, L_{:,2}, L_{:,3}$ must all lie in $\mathcal{V}_0$. Since $L_{:,1}, L_{:,2}, L_{:,3}$ are moreover nonnegative, there are uniquely defined points $q_1, q_2, q_3 \in \mathcal{P}_0$ such that $f(q_i) = L_{:,i}$ for $i \in \{1, 2, 3\}$. Since NMF $M = L \cdot R$ is stochastic, the convex hull of q_1, q_2, q_3 includes r_1, r_2, r_3, i.e., triangle $q_1q_2q_3$ is nested between triangle $r_1r_2r_3$ and polygon $\mathcal{P}_0$. Since $L_{:,1}$ has 0 in its first two coordinates, by inspecting the definition of the map f we see that $q_1 = (u, 0, 0)^\top$ for some $u \in [0, 1]$. By Lemma 3.1 it follows that the supporting polygon $\mathcal{S}_{q_1}$ has three vertices. Hence Lemma 4.2 implies $u \geq 2-\sqrt{2}$.

Considering the polygon $\mathcal{P}_1$, we have $q_1 \in \mathcal{P}_1$ (recall that $f(q_1) = L_{:,1}$). Arguing as in the case of $\mathcal{P}_0$, there are uniquely defined points $q_4, q_5 \in \mathcal{P}_1$ such that $f(q_i) = L_{:,i}$ for $i \in \{4, 5\}$. Similarly as before, triangle $q_1q_4q_5$ is nested between triangle $r_4r_5r_6$ and polygon $\mathcal{P}_1$. Then Lemmas 3.1 and 4.3 imply that $u \leq 2-\sqrt{2}$. Thus, $u = 2 - \sqrt{2}$. This means that $q_1 = (2 - \sqrt{2}, 0, 0)^\top = q_1^*$. Hence

$$L_{:,1} = f(q_1) = f(q_1^*) = W_{:,1}.$$

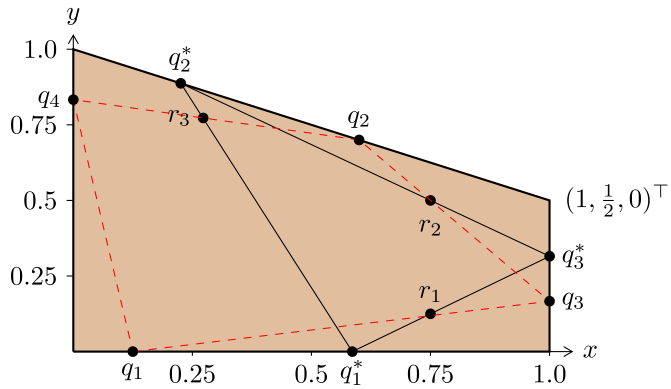

Figure 4: The outer polygon is $\mathcal{P}_0$ (after identifying the xy-plane in $\mathbb{R}^3$ with $\mathbb{R}^2$). The dashed quadrilateral shows the supporting polygon $\mathcal{S}_{q_1}$ in case $q_1 = (\frac{1}{8}, 0, 0)^\top$. The triangle with solid boundary is the supporting polygon $\mathcal{S}_{q_1^*}$, where $q_1^* = (2 - \sqrt{2}, 0, 0)^\top$.

Thus, matrix L is not rational.

We remark that this argument could be strengthened [9] to show that any type-1 NMF of M coincides with the one given in (4.1), up to a permutation of the columns of W and the rows of $\begin{pmatrix} H' & H_\varepsilon \end{pmatrix}$.

4.3 Ruling out Type-2 and Type-3 NMFs. In this subsection, we rule out type-2 NMFs of matrix M; type-3 NMFs of M are ruled out analogously. We use similar geometric arguments as in Section 4.2.

Towards a contradiction, suppose there exists a type-2 NMF $M = L \cdot R$. Then, up to some permutation of its columns, L matches the pattern

$$\begin{pmatrix} 0 & 0 & + & \cdot & \cdot \\ 0 & 0 & 0 & + & + \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 & + & \cdot \\ 0 & 0 & 0 & + \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

according to whether the inner dimension of the NMF $M = L \cdot R$ is 5 or 4, respectively. It is clear that, in either case, $M_{:,1}, M_{:,2}, M_{:,3}$ all lie in the convex hull of $L_{:,1}, L_{:,2}, L_{:,3}$.

Let us consider again the affine space $\mathcal{V}_0 \subseteq \mathbb{R}^6$ and the polygon $\mathcal{P}_0 \subseteq \mathbb{R}^3$ from Section 4.1.2. Recall that $\mathcal{P}_0$ is visualized in Figure 4. By reasoning analogously as in Section 4.2, there are uniquely defined points $q_1, q_2, q_3 \in \mathcal{P}_0$ such that $f(q_i) = L_{:,i}$ for every $i \in \{1, 2, 3\}$. It follows that the convex hull of q_1, q_2, q_3 includes r_1, r_2, r_3. Since columns $L_{:,1}$ and $L_{:,2}$ have 0 in their first two rows, inspecting the definition of the map f, we see that q_1 and q_2 lie on the x-axis in $\mathbb{R}^3$. Define $\hat{q}_1 = (0, 0, 0)^\top$ and $\hat{q}_2 = (1, 0, 0)^\top$. The triangle $\hat{q}_1 \hat{q}_2 q_3$ contains the triangle $q_1 q_2 q_3$, hence the convex hull of $\hat{q}_1, \hat{q}_2, q_3$ also includes vertices r_1, r_2, r_3. It follows that the vertices $\hat{q}_1, r_3, q_3$ are in anti-clockwise order, and that the vertices $\hat{q}_2, q_3, r_2$ are in anti-clockwise order. By inspecting the location of these points (see Figure 4), one can see that q_3 is then outside $\mathcal{P}_0$, a contradiction.

4.4 Ruling out Type-4 NMFs. Recall that we have already excluded NMFs of the matrix $M = \begin{pmatrix} M' & W_\varepsilon \end{pmatrix}$ of types 1, 2, and 3; in this subsection, we exclude NMFs of type 4. In fact, the previous subsections prove the stronger result that there is no NMF of types 1, 2, 3 for the matrix M' alone. Here we spell out the role of W_ε, effectively explaining why the matrix M is defined the way it is.

Observe that adding to M' new columns from the convex hull of the columns of W shrinks the set of possible nonnegative factorizations. Given this, our goal is to find a matrix satisfying the following desiderata:

- its entries are rational,
- its columns belong to the convex hull of the columns of W, and
- it has no type-4 NMF.

The first two items ensure that M, while being rational, admits a nonnegative factorization with left factor W, ensuring that the nonnegative rank of M over $\mathbb{R}$ is (at most) 5. The third condition, combined with the arguments from the previous subsections, ensures that the nonnegative rank of M over $\mathbb{Q}$ is 6.

It turns out that the matrix W, while manifestly failing the first requirement, satisfies the last two items

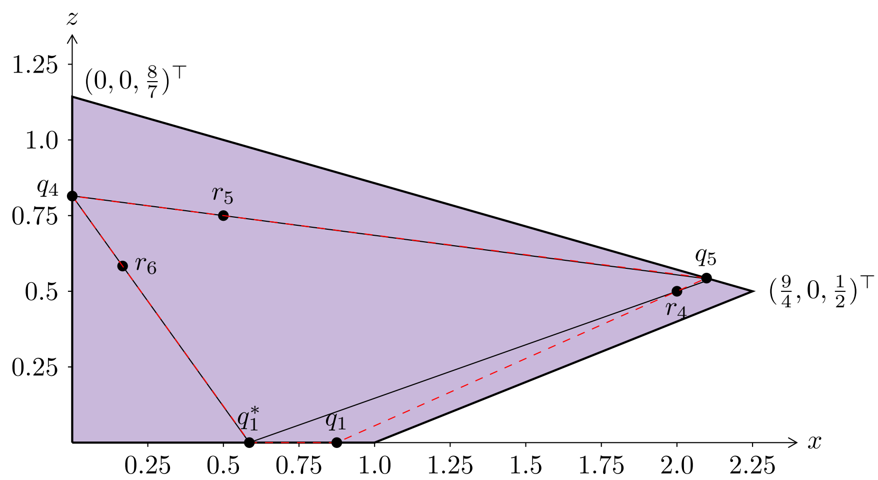

Figure 5: The outer polygon is $\mathcal{P}_1$ (after identifying the xz-plane in $\mathbb{R}^3$ with $\mathbb{R}^2$). Writing $q_1 = (\frac{7}{8}, 0, 0)^\top$, the quadrilateral with dashed boundary shows the supporting polygon $\mathcal{S}_{q_1}$. The triangle with solid boundary is the supporting polygon $\mathcal{S}_{q_1^*}$, where $q_1^* = (2-\sqrt{2}, 0, 0)^\top$.

above:

Claim 4.1. *The matrix W and, therefore, the matrix $\overline{M}$ have no type-4 NMF.*

Motivated by this observation, we arrive at the main technical result of this section, a strengthening of Claim 4.1 showing that no matrix in a suitably small neighborhood of W admits a type-4 NMF:

Lemma 4.4. *For all stochastic matrices $\widetilde{W} \in \mathbb{R}_+^{6\times 5}$ satisfying the entry-wise constraints given in Figure 6, there exists no type-4 NMF $\widetilde{W} = L \cdot R$.*

In particular, the constraints of Lemma 4.4 are satisfied by the matrix W_ε from Theorem 4.1. Therefore, the matrix M has no type-4 NMF. Since we have already exhibited an NMF $W_\varepsilon = W \cdot H_\varepsilon$, the matrix W_ε satisfies all three desiderata, thus concluding the proof.

Proof of Lemma 4.4: outline. The idea of the proof is to derive a contradiction from the assumption that there exists a stochastic matrix $\widetilde{W} \in \mathbb{R}_+^{6\times 5}$ that satisfies all constraints in Figure 6 and has a type-4 NMF $\widetilde{W} = L \cdot R$, i.e., such that L matches the type-4 zero pattern in Figure 2. To this end, we use constraint propagation to successively derive lower and upper bounds for various entries of the matrices L and R until we reach a contradiction. To give a flavor of the argument, we derive one of the required bounds below. The complete argument can be found in [9].

Specifically, towards a contradiction, assume that there exists some stochastic matrix $\widetilde{W} \in \mathbb{R}_+^{6\times 5}$ that satisfies all constraints in Figure 6 and has a type-4 NMF $\widetilde{W} = L \cdot R$. Observe that all columns of $\widetilde{W}$ lie in the convex hull of the columns of L since $\widetilde{W}_{:,j} = L \cdot R_{:,j}$ for each $j \in \{1,\ldots,5\}$. Let us consider the 4th column $\widetilde{W}_{:,4}$. Since $L_{:,2}$ is the only column of L with a strictly positive second coordinate, we have $\widetilde{W}_{2,4} = L_{2,2} \cdot R_{2,4}$. We thus obtain the constraint

$$0.29 \le \widetilde{W}_{2,4} = L_{2,2} \cdot R_{2,4} \le R_{2,4}.$$

Proceeding in this manner, at length we arrive at the statement of the lemma.

With Lemma 4.4 in hand, we observe that the decimal expansion of W_ε can be rounded as in Figure 7. Since matrix W_ε satisfies the system of constraints in Figure 6, we conclude that W_ε and hence M has no type-4 factorization.

A non-constructive alternative. Instead of deducing the result of this subsection from Lemma 4.4, one can alternatively rely on its weaker form, Claim 4.1, and give a non-constructive proof of the existence of an appropriate W_ε (satisfying the three desiderata above) via a topological argument that we sketch below. However, we emphasize that we do not know how to prove Claim 4.1 without following the arguments that prove Lemma 4.4.

We first employ the geometric constructions of Subsection 4.1 to argue that any neighbourhood of the matrix W contains a rational matrix that factors through W (i.e., whose columns belong to the convex hull of the columns of W). Indeed, consider the set $\mathcal{W}$ of all stochastic real matrices of size 6×5 that have a stochastic NMF with left factor W. Observe that $\mathcal{W}$ can

$$\begin{pmatrix} 0 & 0.8 \leq \cdot & 0.286 \leq \ \cdot \ \leq 0.287 & 0 & 0 \\ \cdot \leq \varepsilon & 0 & 0 & 0.29 \leq \cdot & 0.196 \leq \cdot \\ & \cdot \leq \varepsilon & 0.0335 \leq \cdot & 0.21 \leq \cdot & \cdot \leq 0.015 \\ & 0.07 \leq \cdot & \cdot \leq \varepsilon & 0.27 \leq \cdot & \cdot \leq 0.022 \\ & & & \cdot \leq \varepsilon & 0.767 \leq \cdot \\ 0.62 \leq \cdot & & \cdot \leq 0.32 & \cdot \leq 0.21 & \cdot \leq \varepsilon \end{pmatrix}$$

Figure 6: Entry-wise constraints, where $\varepsilon = 10^{-5}$.

$$W_\varepsilon \approx \begin{pmatrix} 0 & 0.81 & 0.2866 & 0 & 0 \\ 0.9 \cdot 10^{-5} & 0 & 0 & 0.296 & 0.1962 \\ 0.1 & 0.7 \cdot 10^{-5} & 0.03360 & 0.219 & 0.0144 \\ 0.04 & 0.0703 & 0.97 \cdot 10^{-5} & 0.276 & 0.0216 \\ 0.2 & 0.08 & 0.4 & 0.9 \cdot 10^{-5} & 0.7679 \\ 0.621 & 0.04 & 0.316 & 0.208 & 0.98 \cdot 10^{-5} \end{pmatrix}$$

Figure 7: Matrix W_ε with entries rounded off.

be characterized as the set of matrices whose columns lie in the image under f of a full-dimensional set in $\mathbb{R}^3$, namely of the convex hull $C = \mathrm{conv}\{q_1, \ldots, q_5\}$. Since the map f is specified by matrices with rational coefficients, it immediately follows that the set of rational matrices is dense in $\mathcal{W}$, as we wished to prove.

Now suppose for the sake of contradiction that every rational matrix in $\mathcal{W}$ has a type-4 NMF. By the above, for all $\delta > 0$ there exists a nonnegative matrix W_δ in the δ-neighbourhood of W which factors through W and has a type-4 NMF $W_\delta = L_\delta \cdot R_\delta$. By compactness, there exists a subsequence of matrices W_δ with decreasing δ such that the corresponding sequences L_δ and R_δ converge. Taking the limit, we arrive at the equality $W = L \cdot R$ where the right-hand side is also a type-4 NMF—which contradicts Claim 4.1. This completes the proof. (Note that Lemma 4.4 contains a constructive version of this argument.)

It is worth mentioning that this reasoning follows similar lines as the *upper semi-continuity* argument for nonnegative rank [6]: the nonnegative rank of any (rational or irrational) matrix W_ε which is entry-wise close enough to W can only be larger than or equal to that of W.

5 Labeled Markov Chain Minimization Requires Irrationality

Our solution to the Cohen–Rothblum problem can be applied to the analogous question for labeled Markov chains (LMCs):

> Given an LMC with rational transition probabilities, is there always a minimal equivalent LMC with rational transition probabilities?

In this section, we answer this question negatively.

LMCs are a fundamental model, variants of which occur under different names in the literature including (generative) probabilistic automata [1] and hidden Markov chains [24]. These models are widely employed in fields such as speech [31] and gesture [8] recognition, signal processing [12], climate modeling [3], and computational biology [16], e.g., sequence analysis [15].

Let us first introduce some notation that we will use in this section. Let $\mathbb{N}$ and $\mathbb{N}_0$ denote the set of all positive and nonnegative integers, respectively. For any $n \in \mathbb{N}$, we write $[n]$ for the set $\{1, 2, \ldots, n\}$ and I_n for the identity matrix of order n. We write $e_i = (0, \ldots, 1, \ldots, 0)$ for a coordinate row vector with 1 in its ith coordinate. We write $\mathbf{1}$ for a column vector of ones.

Given a finite alphabet Σ, we write Σ^* for the set of all words over Σ. The length of a word $w \in \Sigma^*$ is denoted by $|w|$. For any $n \in \mathbb{N}_0$, we write Σ^n for the set of all words in Σ^* of length n. We write λ for the empty word, i.e., the word of length 0. Given two words $x, y \in \Sigma^*$, we denote their concatenation by xy.

Formally, a *labeled Markov chain* (*LMC*) is a tuple $\mathcal{M} = (n, \Sigma, \mu)$ where $n \in \mathbb{N}$ is the number of states, Σ is a finite alphabet of *labels*, and function $\mu : \Sigma \to [0,1]^{n\times n}$ is such that $\sum_{\sigma\in\Sigma} \mu(\sigma)$ is a row-stochastic matrix. For each label σ, every entry $\mu(\sigma)_{i,j}$ is called a *transition probability*.

The intuitive behavior of an LMC $\mathcal{M}$ is as follows: At the start, $\mathcal{M}$ is in a specific initial state. At any time point after that $\mathcal{M}$ is in some state i, and at the next time step $\mathcal{M}$ moves to a state j emitting label σ, with probability $\mu(\sigma)_{i,j}$. For any $k \in \mathbb{N}_0$, after k time steps $\mathcal{M}$ is in some state j having generated a word of length k.

We extend the function μ to words by defining $\mu(\lambda) := I_n$ and $\mu(\sigma_1 \dots \sigma_k) := \mu(\sigma_1) \cdots \mu(\sigma_k)$ for all $k \in \mathbb{N}$ and all $\sigma_1, \dots, \sigma_k \in \Sigma$. It is easy to see that $\mu(xy) = \mu(x)\cdot\mu(y)$ for all words $x, y \in \Sigma^*$. Hence for any word $w \in \Sigma^*$: if $\mathcal{M}$ is in state i, it emits w and moves to state j in $|w|$ time steps, with probability $\mu(w)_{i,j}$.

Formally, the LMC $\mathcal{M}$ defines a function $\|\mathcal{M}\| : \Sigma^* \to [0,1]$ such that for every $w \in \Sigma^*$,

$$\|\mathcal{M}\|(w) := \sum_{j=1}^{n} e_1 \cdot \mu(w)_{1,j} = e_1 \cdot \mu(w) \cdot \mathbf{1}.$$

The value $\|\mathcal{M}\|(w)$ corresponds to the probability that $\mathcal{M}$ generates the word w in $|w|$ time steps. Hence for any $k \in \mathbb{N}_0$, the restriction of $\|\mathcal{M}\|$ to Σ^k is a probability distribution. For example, in Figure 8 we have $\|\mathcal{M}\|(a_1b_1) = \frac{1}{12} = \|\mathcal{M}'\|(a_1b_1)$.

Given two LMCs $\mathcal{M}$, $\mathcal{M}'$ over an alphabet Σ, we say that $\mathcal{M}$ is *equivalent to* $\mathcal{M}'$ if for any word $w \in \Sigma^*$, $\mathcal{M}$ and $\mathcal{M}'$ generate w (in $|w|$ time steps) with equal probability, i.e., $\|\mathcal{M}\|(w) = \|\mathcal{M}'\|(w)$. An LMC is *minimal* if no equivalent LMC has fewer states.

To construct the reduction in Proposition 5.1, we adapt reductions from NMF to the trace-refinement problem in Markov decision processes [19] and to LMC coverability [10].

PROPOSITION 5.1. *Given a nonnegative matrix $M \in \mathbb{Q}_+^{n\times m}$, one can compute in polynomial time an LMC $\mathcal{M} = (m+2, \Sigma, \mu)$ with rational transition probabilities such that for all $d \in \mathbb{N}$:*

(*i*) *any d-dimensional NMF $M = W \cdot H$ determines an LMC $\mathcal{M}' = (d+2, \Sigma, \mu')$ which is equivalent to $\mathcal{M}$, and*

(*ii*) *any LMC $\mathcal{M}' = (d+2, \Sigma, \mu')$ which is equivalent to $\mathcal{M}$ determines a d-dimensional NMF $M = W \cdot H$.*

In both (*i*) *and* (*ii*), *the NMF $M = W \cdot H$ is rational if and only if the LMC $\mathcal{M}'$ has rational transition probabilities.*

Proof. Let $M \in \mathbb{Q}_+^{n\times m}$. As argued in Section 3.1, without loss of generality we may assume that M is stochastic and consider factorizations of M into stochastic matrices only.

We define an LMC $\mathcal{M} = (m+2, \Sigma, \mu)$ as follows: The states are $0, 1, \dots, m, m+1$, with 0 being the initial state. The alphabet is $\Sigma = \{a_1, \dots, a_m\} \cup \{b_1, \dots, b_n\} \cup \{\checkmark\}$, and the function μ, for all $i \in [m]$ and all $j \in [n]$, is defined by:

$$\begin{aligned} \mu(a_i)_{0,i} &= \tfrac{1}{m}, \\ \mu(b_j)_{i,m+1} &= (M^\top)_{i,j} = M_{j,i}, \\ \mu(\checkmark)_{m+1,m+1} &= 1, \end{aligned}$$

and all other entries of $\mu(a_i)$, $\mu(b_j)$, and $\mu(\checkmark)$ are 0. See Figure 8 for an example. It is easy to see that for every $i \in [m]$, $j \in [n]$, $p \in \mathbb{N}_0$:

$$\|\mathcal{M}\|(w) = \begin{cases} 1 & w = \lambda \\ \frac{1}{m} & w = a_i \\ \frac{1}{m} M_{j,i} & w = a_i b_j \checkmark^p \\ 0 & \text{otherwise,} \end{cases}$$

where $\checkmark^p$ denotes the p-fold concatenation of the symbol $\checkmark$ by itself.

We first prove (*i*). Let $M = W \cdot H$ where $W \in \mathbb{R}_+^{n\times d}$ and $H \in \mathbb{R}_+^{d\times m}$ are stochastic matrices. Define an LMC $\mathcal{M}' = (d+2, \Sigma, \mu')$ where the states are $\{0, 1, \dots, d, d+1\}$ with 0 being the initial state. The function μ', for all $i \in [m]$, $j \in [n]$, and $l \in [d]$, is defined by:

$$\begin{aligned} \mu'(a_i)_{0,l} &= \frac{1}{m} H_{l,i}, \\ \mu'(b_j)_{l,d+1} &= W_{j,l}, \\ \mu'(\checkmark)_{d+1,d+1} &= 1, \end{aligned}$$

and all other entries of $\mu'(a_i)$, $\mu'(b_j)$, and $\mu'(\checkmark)$ are 0. See Figure 8 for an example.

We have that $\|\mathcal{M}'\|(\lambda) = 1 = \|\mathcal{M}\|(\lambda)$. Moreover, for every $i \in [m]$, $j \in [n]$, and $p \in \mathbb{N}_0$ we have:

$$\begin{aligned} \|\mathcal{M}'\|(a_i) &= e_0 \cdot \mu'(a_i) \cdot \mathbf{1} \\ &= \sum_{l\in[d]} \frac{1}{m} H_{l,i} \\ &= \frac{1}{m} = \|\mathcal{M}\|(a_i) \end{aligned}$$

and

$$\|\mathcal{M}'\|(a_i b_j \checkmark^p) = e_0 \cdot \mu'(a_i) \cdot \mu'(b_j) \cdot \mu'(\checkmark) \cdot \mathbf{1}$$

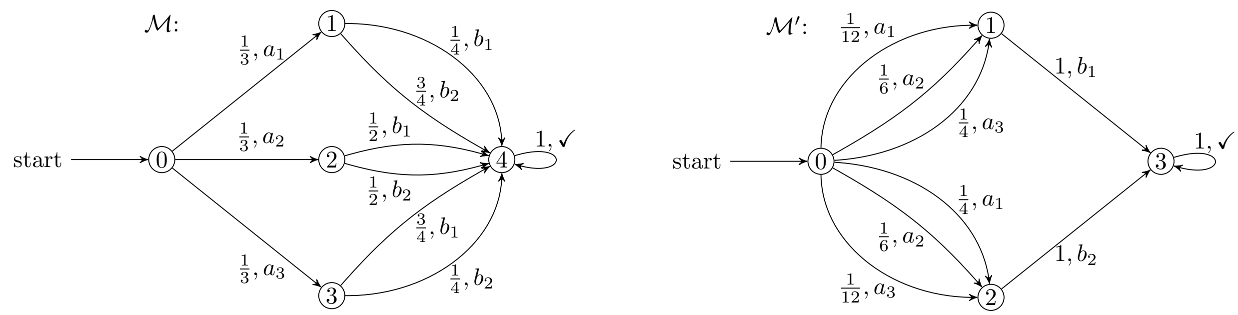

Figure 8: LMC $\mathcal{M}$ is constructed from the matrix $M = \begin{pmatrix} 1/4 & 1/2 & 3/4 \\ 3/4 & 1/2 & 1/4 \end{pmatrix}$, whereas LMC $\mathcal{M}'$ is obtained by NMF $M = I_2 \cdot M$.

$$= \sum_{l \in [d]} \frac{1}{m} H_{l,i} \cdot W_{j,l}$$
$$= \frac{1}{m} M_{j,i} = \|\mathcal{M}\|(a_i b_j \checkmark^p).$$

For every $w \in \Sigma^* \setminus \{\, \lambda, a_i, a_i b_j \checkmark^p \mid i \in [m], j \in [n], p \in \mathbb{N}_0 \,\}$ we have $\|\mathcal{M}'\|(w) = 0 = \|\mathcal{M}\|(w)$. Hence, $\mathcal{M}'$ is equivalent to $\mathcal{M}$. This proves (i).

We now prove (ii). Suppose that $\mathcal{M}$ has an equivalent LMC $\mathcal{M}' = (d+2, \Sigma, \mu')$. Without loss of generality, let the states of $\mathcal{M}'$ be $0, 1, \dots, d, d+1$ with 0 being the initial state. Let us define a matrix $H \in \mathbb{R}_+^{\{0,1,\dots,d,d+1\} \times [m]}$ where

$$(H_{:,i})^\top = m \cdot \mu'(a_i)_{0,:}$$

for all $i \in [m]$, and a matrix $W \in \mathbb{R}_+^{[n] \times \{0,1,\dots,d,d+1\}}$ where

$$(W_{j,:})^\top = \mu'(b_j) \cdot \mathbf{1}$$

for all $j \in [n]$. Since $\mathcal{M}$ and $\mathcal{M}'$ are equivalent, for every $i \in [m]$ and $j \in [n]$ we have:

$$\begin{aligned} \frac{1}{m} M_{j,i} &= \|\mathcal{M}\|(a_i b_j) \\ &= \|\mathcal{M}'\|(a_i b_j) \\ &= \mu'(a_i)_{0,:} \cdot \mu'(b_j) \cdot \mathbf{1} = \frac{1}{m} (H_{:,i})^\top \cdot (W_{j,:})^\top. \end{aligned}$$

This implies that $M = W \cdot H$. Assuming M is a nonzero matrix, there exist $i \in [m]$ and $j \in [n]$ such that

$$\frac{1}{m} M_{j,i} = \mu'(a_i)_{0,:} \cdot \mu'(b_j) \cdot \mathbf{1} > 0.$$

Since matrices $\mu'(a_i)$, $\mu'(b_j)$ are nonnegative, there exist $k, k' \in \{0, 1, \dots, d, d+1\}$ such that

$$\mu'(a_i)_{0,k} \cdot \mu'(b_j)_{k,k'} \cdot \mathbf{1}_{k'} = \mu'(a_i)_{0,k} \cdot \mu'(b_j)_{k,k'} > 0.$$

Without loss of generality, we may assume that $k' = d+1$. Then,

$$\mu'(a_i)_{0,k} \cdot \mu'(b_j)_{k,d+1} > 0. \tag{5.8}$$

The following lemma shows that the $(d+2)$-dimensional NMF $M = W \cdot H$ is, essentially, d-dimensional since one can "ignore" the first and last row of W and column of H.

Lemma 5.1. *It holds that*

$$W \cdot H = W_{[n],[d]} \cdot H_{[d],[m]}.$$

Proof. Take any $l_1 \in [n]$ and $l_2 \in [m]$. We need to show that $(W \cdot H)_{l_1,l_2} = (W_{[n],[d]} \cdot H_{[d],[m]})_{l_1,l_2}$. Since

$$(W \cdot H)_{l_1,l_2} = \sum_{l=0}^{d+1} W_{l_1,l} \cdot H_{l,l_2},$$

it suffices to show that

$$W_{l_1,0} \cdot H_{0,l_2} + W_{l_1,d+1} \cdot H_{d+1,l_2} = 0.$$

To this end, in the following we show that $W_{l_1,d+1} = 0$ and $H_{0,l_2} = 0$.

Towards a contradiction, suppose that $W_{l_1,d+1} > 0$. Since $W_{l_1,d+1} = \mu'(b_{l_1})_{d+1,:} \cdot \mathbf{1}$, by (5.8) we now have

$$\begin{aligned} &\|\mathcal{M}'\|(a_i b_j b_{l_1}) \\ &\geq \mu'(a_i)_{0,k} \cdot \mu'(b_j)_{k,d+1} \cdot \mu'(b_{l_1})_{d+1,:} \cdot \mathbf{1} > 0, \end{aligned}$$

which is a contradiction with $\mathcal{M}'$ being equivalent to $\mathcal{M}$ since $\|\mathcal{M}\|(a_i b_j b_{l_1}) = 0$. We conclude that $W_{l_1,d+1} = 0$.

Towards a contradiction, suppose that $H_{0,l_2} > 0$. Since $H_{0,l_2} = m \cdot \mu'(a_{l_2})_{0,0}$, by (5.8) this implies that

$$\begin{aligned} &\|\mathcal{M}'\|(a_{l_2} a_i b_j) \\ &\geq \mu'(a_{l_2})_{0,0} \cdot \mu'(a_i)_{0,k} \cdot \mu'(b_j)_{k,d+1} > 0, \end{aligned}$$

which is a contradiction with $\mathcal{M}'$ being equivalent to $\mathcal{M}$ since $\|\mathcal{M}\|(a_{l_2} a_i b_j) = 0$. We conclude that $H_{0,l_2} = 0$, which completes the proof. $\square$

By Lemma 5.1, we now have

$$M = W \cdot H = W_{[n],[d]} \cdot H_{[d],[m]},$$

which completes the proof of (ii). □

Now, we can give a negative answer to the question from the beginning of this section:

COROLLARY 5.1. *There exists an LMC with rational transition probabilities such that there is no minimal equivalent LMC with rational transition probabilities.*

Proof. Let $M \in \mathbb{Q}_+^{6\times 11}$ be the matrix from Theorem 4.1. Since $\mathrm{rank}_+(M) = 5$, by Proposition 5.1 one can compute in polynomial time an LMC $\mathcal{M} = (13, \Sigma, \mu)$ such that any minimal LMC equivalent to $\mathcal{M}$ has 7 states. Towards a contradiction, assume there is an LMC $\mathcal{M}' = (7, \Sigma, \mu')$ equivalent to $\mathcal{M}$ with rational transition probabilities. Then by Proposition 5.1, $\mathcal{M}'$ determines a 5-dimensional rational NMF $M = W \cdot H$. This is a contradiction since the nonnegative rank of M over $\mathbb{Q}$ is 6 by Theorem 4.1. □

As an immediate corollary of Proposition 5.1, one can relate the computational complexity of the following two problems:

- The *NMF problem* is: Given a matrix $M \in \mathbb{Q}_+^{n\times m}$ and $d \in \mathbb{N}$, is $\mathrm{rank}_+(M) \leq d$?
- The *LMC minimization problem* is: Given an LMC $\mathcal{M}$ with rational transition probabilities and $d \in \mathbb{N}$, does there exist an LMC $\mathcal{M}'$ with d states such that $\mathcal{M}$ and $\mathcal{M}'$ are equivalent?

COROLLARY 5.2. *There is a polynomial-time reduction from the NMF problem to the LMC minimization problem.*

Our NMF-hardness result can be adapted easily to (reactive) probabilistic automata [30]. Since the NMF problem is NP-hard [38], minimizing both probabilistic automata and LMCs is NP-hard. In contrast, minimization of weighted automata (which may have negative transition weights) can be done efficiently over $\mathbb{Q}$ [32]. Hence, the nonnegativity constraint makes minimization fundamentally more difficult.

6 Rationality of Unambiguous Nonnegative Rank

Consider an NMF $M = W \cdot H$, where W has dimension $n \times d$ and H has dimension $d \times m$. We call such an NMF *unambiguous* if each column of M is obtained as a nonnegative combination of orthogonal columns of W: formally, for every column $j \in \{1, \ldots, m\}$ and every two distinct rows $k, k' \in \{1, \ldots, d\}$, if $H_{k,j}, H_{k',j} \neq 0$ then $W_{:,k}$ and $W_{:,k'}$ are orthogonal column vectors. This condition can equivalently be formulated as:

$$W_{i,k} H_{k,j} W_{i,k'} H_{k',j} = 0 \quad \text{for all } i, j \text{ and } k \neq k'. \tag{6.9}$$

We recall the definition of the *Frobenius inner product*: for matrices $M, N \in \mathbb{R}^{n\times m}$, this is the usual inner product when the matrices are considered as vectors in the nm-dimensional real vector space: $\langle M, N\rangle = \sum_{i=1}^n \sum_{j=1}^m M_{ij} N_{ij}$. A sequence of matrices $M_1, \ldots, M_k$ forms an *orthogonal system* if $\langle M_s, M_t\rangle = 0$ for all matrices M_s and M_t with $s \neq t$.

LEMMA 6.1. *For any nonnegative matrix M, the following statements are equivalent:*

(U1) *there exists an expansion $M = M_1 + \ldots + M_d$ where the matrices $M_1, \ldots, M_d$ all have rank at most 1, are nonnegative, and form an orthogonal system;*

(U2) *there exists an unambiguous NMF $M = W \cdot H$ of inner dimension at most d.*

Proof. We recall that an equivalent characterization [11] of the nonnegative rank of M is as the smallest number d such that M is equal to the sum of d rank-1 matrices in $\mathbb{R}_+^{n\times m}$. Indeed, a nonzero $n \times m$ matrix has rank 1 just in case it can be written as the product $uv^\top$ of an n-dimensional column vector u and an m-dimension row vector $v^\top$. Writing an NMF of inner dimension d in the form

$$M = W \cdot H = \sum_{s=1}^{d} W_{:,s} \cdot H_{s,:},$$

it follows that such a factorization is equivalent to a decomposition of M as the sum of d rank-1 matrices in $\mathbb{R}_+^{n\times m}$.

Suppose that $M = W \cdot H$ is an unambiguous NMF of inner dimension d. Then we have that $M = M_1 + \ldots + M_d$, where $M_s := W_{:,s} \cdot H_{s,:}$ is a nonnegative rank-1 matrix for each $s \in \{1, \ldots, d\}$. The unambiguity condition (6.9) is equivalent to the condition that $M_1, \ldots, M_d$ form an orthogonal system of matrices with respect to the Frobenius inner product. Conversely, given a decomposition $M = M_1 + \ldots + M_d$ as in (U1), we obtain an unambiguous factorization $M = W \cdot H$ by selecting nonnegative matrices W and H subject to the condition $M_s := W_{:,s} \cdot H_{s,:}$ for each $s \in \{1, \ldots, d\}$. □

There is an analogy between property (U1) and the following characterizations of rank and nonnegative rank. The rank of a matrix $M \in \mathbb{R}^{n\times m}$ is the smallest number d such that M has an expansion as in (U1) where all matrices M_s, $s \in \{1, \ldots, d\}$, have rank at most 1. If $M \in \mathbb{R}_+^{n \times m}$, its nonnegative rank is the smallest number d such that M has an expansion as in (U1) where all matrices M_s have rank at most 1 and are nonnegative [11]. The property (U1) additionally requires that matrices M_s have no common nonzero positions. From the combinatorial perspective, such an expansion corresponds to partitioning the set of nonzero entries of M into a collection of disjoint *combinatorial rectangles* (these are Cartesian products of the form $I \times J$ with $I \subseteq \{1, \ldots, n\}$ and $J \subseteq \{1, \ldots, m\}$), each of which forms the support of a rank-1 submatrix of M.

Based on this analogy, we define the *unambiguous nonnegative rank* of a matrix $M \in \mathbb{R}_+^{n \times m}$, denoted $\mathrm{rank}_\perp M$, as the smallest d such that M satisfies property (U1).

For a 0-1 matrix M, Yannakakis [41] defines the unambiguous communication complexity of M to be the logarithm of what we have here called the unambiguous nonnegative rank. A more vivid interpretation arises from considering applications in topic modeling [5]: an unambiguous NMF $M = W \cdot H$ corresponds to representing documents (columns of M) as distributions on non-overlapping topics (columns of W without common nonzero entries). Here unambiguity means that, even though every word (row index i) can be associated with more than one topic, its use in each individual document is restricted to some particular topic.

Theorem 6.1. *The following statements hold:*

1. *If $M \in \mathbb{Q}_+^{n \times m}$ and $\mathrm{rank}_\perp M \le k$, then M has an unambiguous NMF $M = W \cdot H$ of inner dimension at most k where all entries of W and H are rational.*

2. *It is* NP*-complete to decide, given $M \in \mathbb{Q}_+^{n \times m}$ and $k \in \mathbb{N}$, whether $\mathrm{rank}_\perp M \le k$.*

Proof. Given a bipartite graph G with two parts $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_m\}$, let M be the 0-1 adjacency matrix of G where $M_{i,j}$ is 1 if and only if there is an edge between u_i and v_j. In any decomposition of M as the sum of an orthogonal system of nonnegative matrices $M = M_1 + \ldots + M_d$, each summand must itself be a 0-1 matrix and, having rank 1, be the adjacency matrix of a subgraph of M induced by a biclique. Thus, such a decomposition is nothing but a partition of the edges of G into d pairwise disjoint bicliques. It immediately follows that deciding whether $\mathrm{rank}_\perp M \le k$ is NP-hard (see [20]). On the other hand, it is clear that in any disjoint sum $M = M_1 + \ldots + M_d$ the nonzero entries of the right-hand rank-1 matrices are all entries of M. Thus, unambiguous rank is always witnessed by a factorization into rational matrices and is computable in NP. $\square$

Theorem 6.1 complements the irrationality result of Theorem 4.1.

7 Conclusions

In this paper we have solved the Cohen–Rothblum problem, showing that nonnegative ranks over $\mathbb{R}$ and over $\mathbb{Q}$ may differ. More precisely, our construction applies to matrices of rank 4 and greater. It was already known to Cohen and Rothblum [11] that nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ coincide for matrices of rank at most 2, and Kubjas et al. [26] showed that this also holds for matrices of nonnegative rank (over $\mathbb{R}$) at most 3. The remaining open question is whether nonnegative ranks over $\mathbb{R}$ and over $\mathbb{Q}$ differ for rank-3 matrices whose nonnegative rank (over $\mathbb{R}$) is at least 4—or whether our example is optimal in this sense.

As our results show that the nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ are different functions, the computability question emerges. While it has long been known that the nonnegative rank over $\mathbb{R}$ is computable in PSPACE (see, e.g., Cohen and Rothblum [11]), this is not known for the nonnegative rank over $\mathbb{Q}$. The problem of computing the latter is reducible to the decision problem for the existential theory of the field of rational numbers, whose decidability is a long-standing and very prominent open question [29].

Finally, we would like to point out that the complexity of the following geometric problem closely linked to NMF, the *nested polytopes* problem, is not fully known. This problem asks, given an ordered field $\mathbb{F}$ and polytopes $\mathcal{S} \subseteq \mathcal{T}$ in $\mathbb{F}^n$, whether there exists a *simple* polytope $\mathcal{N}$ such that $\mathcal{S} \subseteq \mathcal{N} \subseteq \mathcal{T}$ (cf. Gillis and Glineur [23]). The definition of "simple" can be either "having at most k vertices," or "having at most k facets," or a combination of both. For $\mathbb{F} = \mathbb{R}$, minimizing the number of vertices or, dually, facets, is known to require irrational numbers [10] even in the case of full-dimensional $\mathcal{S}$. While for some representations of the polytopes such questions are known to be NP-hard (see, e.g., Das and Goodrich [13]), their precise complexity is not known in general.

Acknowledgements

The authors would like to thank Vladimir Lysikov and Vladimir Shiryaev for stimulating discussions. Dmitry Chistikov was sponsored in part by the ERC Synergy award ImPACT and is supported by the ERC grant AVS-ISS (648701). Stefan Kiefer is supported by a University Research Fellowship of the Royal Society. Stefan Kiefer, Ines Marušić, Mahsa Shirmohammadi, and James Worrell gratefully acknowledge the support of the EPSRC.

References

[1] N. Abe and M. K. Warmuth, *On the computational complexity of approximating distributions by probabilistic automata*, Machine Learning, 9 (1992), pp. 205–260.

[2] A. Aggarwal, H. Booth, J. O'Rourke, S. Suri, and C. K. Yap, *Finding minimal convex nested polygons*, Information and Computation, 83 (1989), pp. 98–110.

[3] P. Ailliot, C. Thompson, and P. Thomson, *Space-time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions*, Journal of the Royal Statistical Society, 58 (2009), pp. 405–426.

[4] S. Arora, R. Ge, R. Kannan, and A. Moitra, *Computing a nonnegative matrix factorization - provably*, in Proceedings of the 44th Symposium on Theory of Computing (STOC), 2012, pp. 145–162.

[5] S. Arora, R. Ge, and A. Moitra, *Learning topic models - going beyond SVD*, in 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2012, pp. 1–10.

[6] C. Bocci, E. Carlini, and F. Rapallo, *Perturbation of matrices and nonnegative rank with a view toward statistical models*, SIAM J. Matrix Analysis Applications, 32 (2011), pp. 1500–1512.

[7] J. Canny, *Some algebraic and geometric computations in PSPACE*, in Proceedings of the 20th Annual Symposium on Theory of Computing (STOC), 1988, pp. 460–467.

[8] F.-S. Chen, C.-M. Fu, and C.-L. Huang, *Hand gesture recognition using a real-time tracking method and hidden Markov models*, Image and Vision Computing, 21 (2003), pp. 745–758.

[9] D. Chistikov, S. Kiefer, I. Marušić, M. Shirmohammadi, and J. Worrell, *Nonnegative matrix factorization requires irrationality*, tech. report, arxiv.org, 2016. Available at `http://arxiv.org/abs/1605.06848`.

[10] D. Chistikov, S. Kiefer, I. Marušić, M. Shirmohammadi, and J. Worrell, *On restricted nonnegative matrix factorization*, in Proceedings of the 43rd International Colloquium on Automata, Languages and Programming (ICALP), 2016.

[11] J. E. Cohen and U. G. Rothblum, *Nonnegative ranks, decompositions, and factorizations of nonnegative matrices*, Linear Algebra and its Applications, 190 (1993), pp. 149–168.

[12] M. Crouse, R. Nowak, and R. Baraniuk, *Wavelet-based statistical signal processing using hidden Markov models*, IEEE Transactions on Signal Processing, 46 (1998), pp. 886–902.

[13] G. Das and M. T. Goodrich, *On the complexity of approximating and illuminating three-dimensional convex polyhedra*, in Algorithms and Data Structures: 4th International Workshop (WADS), Springer Berlin Heidelberg, 1995, pp. 74–85.

[14] D. L. Donoho and V. Stodden, *When does non-negative matrix factorization give a correct decomposition into parts?*, in Advances in Neural Information Processing Systems (NIPS), 2003, pp. 1141–1148.

[15] R. Durbin, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, 1998.

[16] S. Eddy, *What is a hidden Markov model?*, Nature Biotechnology, 22 (2004), pp. 1315–1316.

[17] K. Etessami and M. Yannakakis, *Recursive concurrent stochastic games*, Logical Methods in Computer Science, 4 (2008).

[18] K. Etessami and M. Yannakakis, *On the complexity of Nash equilibria and other fixed points*, SIAM J. Comput., 39 (2010), pp. 2531–2597.

[19] N. Fijalkow, S. Kiefer, and M. Shirmohammadi, *Trace refinement in labelled Markov decision processes*, in Proceedings of the 19th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS), vol. 9634 of Lecture Notes in Computer Science, Springer, 2016, pp. 303–318.

[20] H. Fleischner, E. Mujuni, D. Paulusma, and S. Szeider, *Covering graphs with few complete bipartite subgraphs*, Theor. Comput. Sci., 410 (2009), pp. 2045–2053.

[21] N. Gillis, *Sparse and unique nonnegative matrix factorization through data preprocessing*, Journal of Machine Learning Research, 13 (2012), pp. 3349–3386, `http://dl.acm.org/citation.cfm?id=2503349`.

[22] N. Gillis, *The why and how of nonnegative matrix factorization*, in Regularization, Optimization, Kernels, and Support Vector Machines, M. S. J.A.K. Suykens and A. Argyriou, eds., Machine Learning and Pattern Recognition Series, Chapman & Hall/CRC, 2014, pp. 257–291. Available at `http://arxiv.org/abs/1401.5226`.

[23] N. Gillis and F. Glineur, *On the geometric interpretation of the nonnegative rank*, Linear Algebra and its Applications, 437 (2012), pp. 2685–2712.

[24] D. Gillman and M. Sipser, *Inference and minimization of hidden Markov chains*, in Proceedings of the 7th Annual ACM Conference on Computational Learning Theory (COLT), 1994, pp. 147–158.

[25] D. A. Gregory and N. J. Pullman, *Semiring rank: Boolean rank and nonnegative rank factorizations*, J. Combin. Inform. System Sci., 8 (1983), pp. 223–233.

[26] K. Kubjas, E. Robeva, and B. Sturmfels, *Fixed points of the EM algorithm and nonnegative rank boundaries*, The Annals of Statistics, 43 (2015), pp. 422–461.

[27] H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, *Theorems on positive data: On the uniqueness of NMF*, Comp. Int. and Neurosc., 2008 (2008), `http://dx.doi.org/10.1155/2008/764206`.

[28] A. Moitra, *An almost optimal algorithm for computing nonnegative rank*, SIAM Journal on Computing, 45 (2016), pp. 156–173.
[29] B. Poonen, *Hilbert's tenth problem over rings of number-theoretic interest*, Note from the lecture at the Arizona Winter School on "Number Theory and Logic", (2003).
[30] M. O. Rabin, *Probabilistic automata*, Information and Control, 6 (1963), pp. 230–245.
[31] L. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proceedings of the IEEE, 77 (1989), pp. 257–286.
[32] M. P. Schützenberger, *On the definition of a family of automata*, Information and Control, 4 (1961), pp. 245–270.
[33] L. S. Shapley, *Stochastic games*, Proceedings of the national academy of sciences, 39 (1953), pp. 1095–1100.
[34] Y. Shitov, *Nonnegative rank depends on the field*, tech. report, arxiv.org, 2015. Available at `http://arxiv.org/abs/1505.01893`.
[35] Y. Shitov, *Nonnegative rank depends on the field II*, tech. report, arxiv.org, 2016. Available at `http://arxiv.org/abs/1605.07173`.
[36] D. Simanek, *How to view 3D without glasses*. `https://www.lhup.edu/~dsimanek/3d/view3d.htm`. Online, accessed in April 2016.
[37] L. B. Thomas, *Rank factorization of nonnegative matrices (by A. Berman and R.J. Plemmons)*, SIAM Rev., 16 (1974), pp. 393–394.
[38] S. A. Vavasis, *On the complexity of nonnegative matrix factorization*, SIAM Journal on Optimization, 20 (2009), pp. 1364–1377.
[39] S. Venkatasubramanian, *Computational geometry column 55: New developments in nonnegative matrix factorization*, SIGACT News, 44 (2013), pp. 70–78.
[40] N. Vlassis, M. L. Littman, and D. Barber, *On the computational complexity of stochastic controller optimization in POMDPs*, ACM Transactions on Computation Theory (TOCT), 4 (2012), p. 12.
[41] M. Yannakakis, *Expressing combinatorial optimization problems by linear programs*, J. Comput. Syst. Sci., 43 (1991), pp. 441–466.