On Restricted Nonnegative Matrix Factorization*†

Dmitry Chistikov‡[1], **Stefan Kiefer**§[2], **Ines Marušić**[3], **Mahsa Shirmohammadi**[4], **and James Worrell**[5]

1 **Max Planck Institute for Software Systems (MPI-SWS), Kaiserslautern and Saarbrücken, Germany**¶
dch@mpi-sws.org
2 **University of Oxford, Oxford, United Kingdom**
Stefan.Kiefer@cs.ox.ac.uk
3 **University of Oxford, Oxford, United Kingdom**
Ines.Marusic@cs.ox.ac.uk
4 **University of Oxford, Oxford, United Kingdom**
Mahsa.Shirmohammadi@cs.ox.ac.uk
5 **University of Oxford, Oxford, United Kingdom**
James.Worrell@cs.ox.ac.uk

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. Restricted NMF requires in addition that the column spaces of M and W coincide. Finding the minimal inner dimension d is known to be NP-hard, both for NMF and restricted NMF. We show that restricted NMF is closely related to a question about the nature of minimal probabilistic automata, posed by Paz in his seminal 1971 textbook. We use this connection to answer Paz's question negatively, thus falsifying a positive answer claimed in 1974.

Furthermore, we investigate whether a rational matrix M always has a restricted NMF of minimal inner dimension whose factors W and H are also rational. We show that this holds for matrices M of rank at most 3 and we exhibit a rank-4 matrix for which W and H require irrational entries.

1998 ACM Subject Classification F.1.1 Models of Computation, F.2.1 Numerical Algorithms and Problems

Keywords and phrases nonnegative matrix factorization, nonnegative rank, probabilistic automata, labelled Markov chains, minimization

Digital Object Identifier 10.4230/LIPIcs.ICALP.2016.103

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 matrices W and H are also nonnegative. The smallest inner dimension of any such factorization is called the *nonnegative rank* of M, written $\mathrm{rank}_+(M)$.

* A full version of the paper is available at http://arxiv.org/abs/1605.07061.
† Kiefer, Marušić, Shirmohammadi, and Worrell gratefully acknowledge the support of the EPSRC.
‡ Chistikov was sponsored in part by the ERC Synergy award ImPACT.
§ Kiefer is supported by a University Research Fellowship of the Royal Society.
¶ Present address: Department of Computer Science, University of Oxford, UK.

In machine learning, NMF was popularized by the seminal work of Lee and Seung [14] as a tool for finding features in facial-image databases. Since then, NMF has found a broad range of applications – including document clustering, topic modelling, computer vision, recommender systems, bioinformatics, and acoustic signal processing [5, 4, 7, 19, 21, 22]. In applications, matrix M can typically be seen as a matrix of data points: each column of M corresponds to a data point and each row to a feature. Then, computing a nonnegative factorization $M = W \cdot H$ corresponds to expressing the data points (columns of M) as convex combinations of latent factors (columns of W), i.e., as linear combinations of latent factors with nonnegative coefficients (columns of H).

From a computational perspective, perhaps the most basic problem concerning NMF is whether a given nonnegative matrix of rational numbers M admits an NMF with inner dimension at most a given number k. Formally, the *NMF problem* asks whether $\mathrm{rank}_+(M) \le k$. In practical applications, various heuristics and local-search algorithms are used to compute an approximate nonnegative factorization, but little is known in terms of their theoretical guarantees. The NMF problem under the *separability* assumption of Donoho and Stodden [9] is tractable: an NMF $M = W \cdot H$ is called *separable* if every column of W is also a column of M. In 2012, Arora et al. [2] showed that it is decidable in polynomial time whether a given matrix admits a separable NMF with a given inner dimension. Further progress was made recently, with several efficient algorithms for computing near-separable NMFs [13, 12].

Vavasis [20] showed that the problem of deciding whether the rank of a nonnegative matrix is equal to its nonnegative rank is NP-hard. This result implies that generalizations of this problem, such as the aforementioned NMF problem, the problem of computing the factors W, H (in both exact and approximate versions), and nonnegative rank determination, are also NP-hard. It is not known whether any of these problems are in NP.

Vavasis [20] notes that the difficulty in proving membership in NP lies in the fact that a certificate for a positive answer to the NMF problem seems to require the sought factors: a pair of nonnegative matrices W, H such that $M = W \cdot H$. Related to this, Cohen and Rothblum [8] posed the question of whether, given a nonnegative matrix of rational numbers M, there always exists an NMF $M = W \cdot H$ of inner dimension equal to $\mathrm{rank}_+(M)$ such that both W and H are also matrices of rational numbers. A natural route to proving membership of the NMF problem in NP would be to give a positive answer to the question of Cohen and Rothblum (as well as a polynomial bound on the bit-length of the factors W and H). However, the question remains open. Currently the best complexity bound for the NMF problem is membership in PSPACE, which is obtained by translation into the existential theory of real-closed fields [2]. Such a translation shows that one can always choose the entries of W and H to be algebraic numbers.

In this work, we focus on the so-called *restricted* NMF (RNMF) problem, introduced by Gillis and Glineur [11]. The RNMF problem is defined as the NMF problem, except that the column spaces of M and W are required to coincide. (Note that for any NMF, the column space of M is a subspace of the column space of W.) This problem has a natural geometric interpretation as the *nested polytope problem (NPP)*: the problem of finding a minimum-vertex polytope nested between two given convex polytopes. In more detail, for a rank-r matrix M, finding an RNMF with inner dimension d is known to correspond exactly to finding a nested polytope with d vertices in an $(r-1)$-dimensional NPP.

Our contributions are as follows.

1. We establish a tight connection between NMF and the coverability relation in labelled Markov chains (LMCs). The latter notion was introduced by Paz [15]. Loosely speaking, an LMC $\mathcal{M}'$ *covers* an LMC $\mathcal{M}$ if for any initial distribution over the states of $\mathcal{M}$ there is

an initial distribution over the states of $\mathcal{M}'$ such that $\mathcal{M}$ and $\mathcal{M}'$ are equivalent. In 1971, Paz [15] asked a question about the nature of minimal covering LMCs. The question was supposedly answered positively in 1974 [3]. However, we show that the correct answer is negative, thus falsifying the claim in [3]. Instrumental to our counterexample is the observation that restricted nonnegative rank and nonnegative rank can be different. (Indeed, the wrong claims in [3] seem to implicitly rely on the opposite assumption, although the notions of NMF and RNMF had not yet been developed.)

2. We show that the RNMF problem for matrices M of rank 3 or less can be solved in polynomial time. In fact, we show that there is always a *rational* NMF of M with inner dimension $\text{rank}_+(M)$, and that it can be computed in polynomial time in the Turing model of computation. This improves a result in [11] where the RNMF problem is shown to be solvable in polynomial time assuming a RAM model with unit-cost arithmetic. Both our algorithm and the one in [11] exploit the connection to the 2-dimensional NPP, allowing us to take advantage of a geometric algorithm by Aggarwal et al. [1]. We need to adapt the algorithm in [1] to ensure that the occurring numbers are rational and can be computed in polynomial time in the Turing model of computation.
3. We exhibit a rank-4 matrix that has an RNMF with inner dimension 5 but *no rational* RNMF with inner dimension 5. We construct this matrix via a particular instance of the 3-dimensional NPP, again taking advantage of the geometric interpretation of RNMF. Our result answers the RNMF variant of Cohen and Rothblum's question in [8] negatively. The original (NMF) variant remains open.

Detailed proofs of all results can be found in the full version of this paper.

2 Nonnegative Matrix Factorization

Let $\mathbb{N}$ and $\mathbb{N}_0$ denote the set of all positive and nonnegative integers, respectively. For every $n \in \mathbb{N}$, we write $[n]$ for the set $\{1, 2, \ldots, n\}$ and write I_n for the identity matrix of order n. For any ordered field $\mathbb{F}$, we denote by $\mathbb{F}_+$ the set of all its nonnegative elements. For any vector v, we write v_i for its i^{th} entry. A vector v is called *stochastic* if its entries are nonnegative real numbers that sum up to one. For every $i \in [n]$, we write e_i for the i^{th} coordinate vector in $\mathbb{R}^n$. We write $\mathbf{1}^{(n)}$ for the n-dimensional column vector with all ones. We omit the superscript if it is understood from the context.

For any matrix M, we write M_i for its i^{th} row, M^j for its j^{th} column, and $M_{i,j}$ for its $(i, j)^{\text{th}}$ entry. The *column space* (resp., *row space*) of M, written $\text{Col}(M)$ (resp., $\text{Row}(M)$), is the vector space spanned by the columns (resp., rows) 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 the element sum of each of its columns (resp., rows) is one.

2.1 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. Note that $\text{Col}(M) \subseteq \text{Col}(W)$. We refer to d as the *inner dimension* of the NMF, and hence refer to NMF $M = W \cdot H$ as being *d-dimensional*. The *nonnegative rank over* $\mathbb{F}$ of M is the smallest number $d \in \mathbb{N}_0$ such that there exists a d-dimensional NMF over $\mathbb{F}$ of M. An equivalent characterization [8] of the nonnegative rank over $\mathbb{F}$ of M is as the smallest number of rank-1 matrices in $\mathbb{F}_+^{n\times m}$ such that M is equal to their sum. The nonnegative rank over $\mathbb{R}$ will henceforth simply be called nonnegative rank, and will be denoted by $\mathrm{rank}_+(M)$. For any nonnegative matrix $M \in \mathbb{R}_+^{n\times m}$, it is easy to see that $\mathrm{rank}(M) \le \mathrm{rank}_+(M) \le \min\{n, m\}$.

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 column-stochastic matrix M' with equal nonnegative rank. Similarly, if $M = W \cdot H$ then after removing 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 column-stochastic. If M is column-stochastic then $\mathbf{1}^\top = \mathbf{1}^\top M = \mathbf{1}^\top WD\cdot D^{-1}H = \mathbf{1}^\top D^{-1}H$, hence $D^{-1}H$ is column-stochastic as well. Thus, without loss of generality one can consider NMFs of column-stochastic matrices into column-stochastic matrices [8, Theorem 3.2].

NMF problem: Given a matrix $M \in \mathbb{Q}_+^{n\times m}$ and $k \in \mathbb{N}$, is $\mathrm{rank}_+(M) \le k$?

The NMF problem is NP-hard, even for $k = \mathrm{rank}(M)$ (see [20]). On the other hand, it is reducible to the existential theory of the reals, hence by [6, 16] it is in PSPACE.

For a matrix $M \in \mathbb{Q}_+^{n\times m}$, its nonnegative rank over $\mathbb{Q}$ is clearly at least $\mathrm{rank}_+(M)$. While those ranks are equal if $\mathrm{rank}(M) \le 2$, a longstanding open question by Cohen and Rothblum asks whether they are always equal [8]. In other words, it is conceivable that there exists a rational matrix $M \in \mathbb{Q}_+^{n\times m}$ with $\mathrm{rank}_+(M) = d$ that has no *rational* NMF with inner dimension d. Recently, Shitov [17] exhibited a nonnegative matrix (with irrational entries) whose nonnegative rank over a subfield of $\mathbb{R}$ is different from its nonnegative rank over $\mathbb{R}$.

2.2 Restricted Nonnegative Rank

For all matrices $M \in \mathbb{F}_+^{n\times m}$, an NMF $M = W \cdot H$ is called *restricted NMF (RNMF)* [11] if $\mathrm{rank}(M) = \mathrm{rank}(W)$. As we know $\mathrm{Col}(M) \subseteq \mathrm{Col}(W)$ holds for all NMF instances, the condition $\mathrm{rank}(M) = \mathrm{rank}(W)$ is then equivalent to $\mathrm{Col}(M) = \mathrm{Col}(W)$. The *restricted nonnegative rank over* $\mathbb{F}$ of M is the smallest number $d \in \mathbb{N}_0$ such that there exists a d-dimensional restricted nonnegative factorization over $\mathbb{F}$ of M. Unless indicated otherwise, henceforth we will assume $\mathbb{F} = \mathbb{R}$ when speaking of the restricted nonnegative rank of M, and denote it by $\mathrm{rrank}_+(M)$.

RNMF problem: Given a matrix $M \in \mathbb{Q}_+^{n\times m}$ and $k \in \mathbb{N}$, is $\mathrm{rrank}_+(M) \le k$?

We have the following basic properties.

▶ **Lemma 1** ([11])**.** *Let* $M \in \mathbb{R}_+^{n\times m}$. *Then* $\mathrm{rank}(M) \le \mathrm{rank}_+(M) \le \mathrm{rrank}_+(M) \le m$. *Moreover, if* $\mathrm{rank}(M) = \mathrm{rank}_+(M)$ *then* $\mathrm{rank}(M) = \mathrm{rrank}_+(M)$.

Thus, with the above-mentioned NP-hardness result, it follows that the RNMF problem is also NP-hard and in PSPACE.

For a matrix $M \in \mathbb{Q}_+^{n\times m}$, its restricted nonnegative rank over $\mathbb{Q}$ is clearly at least $\mathrm{rrank}_+(M)$. As with nonnegative rank, in general it is not known whether the restricted nonnegative ranks of M over $\mathbb{R}$ and over $\mathbb{Q}$ are equal. By [8, Theorem 4.1] and Lemma 1, this is true when $\mathrm{rank}(M) \le 2$.

RNMF has the following geometric interpretation. For a dimension $\ell \in \mathbb{N}$, the *convex combination* of a set $\{v_1, \ldots, v_m\} \subset \mathbb{R}^\ell$ is a point $\lambda_1 v_1 + \cdots + \lambda_m v_m$ where $(\lambda_1, \ldots, \lambda_m)$ is a stochastic vector. The *convex hull* of $\{v_1, \ldots, v_m\}$, written as $\mathrm{conv}\{v_1, \ldots, v_m\}$, is the set of all convex combinations of $\{v_1, \ldots, v_m\}$. We call $\mathrm{conv}\{v_1, \ldots, v_m\}$ a *polytope spanned by* $v_1, \ldots, v_m$. A *polyhedron* is a set $\{\, x \in \mathbb{R}^\ell \mid Ax + b \ge 0 \,\}$ with $A \in \mathbb{R}^{n\times\ell}$ and $b \in \mathbb{R}^n$. A set is a polytope if and only if it is a bounded polyhedron. A polytope is *full-dimensional* (i.e., has volume) if the matrix $(A \;\; b) \in \mathbb{R}^{n\times(\ell+1)}$ has rank $\ell + 1$.

Nested polytope problem (NPP): Given $r, n \in \mathbb{N}$, let $A \in \mathbb{Q}^{n\times(r-1)}$ and $b \in \mathbb{Q}^n$ be such that $P = \{\, x \in \mathbb{R}^{r-1} \mid Ax + b \geq 0 \,\}$ is a full-dimensional polytope. Let $S \subseteq P$ be a full-dimensional polytope described by spanning points. The *nested polytope problem (NPP)* asks, given A, b, S and a number $k \in \mathbb{N}$, whether there exist k points that span a polytope Q with $S \subseteq Q \subseteq P$. Such a polytope Q is called *nested* between P and S.

The following proposition appears as Theorem 1 in [11].

▶ **Proposition 2.** *The RNMF problem and the NPP are interreducible in polynomial time.*

More specifically, the reductions are as follows.

1. Given a nonnegative matrix $M \in \mathbb{Q}_+^{n\times m}$ of rank r, one can compute in polynomial time $A \in \mathbb{Q}^{n\times(r-1)}$ and $b \in \mathbb{Q}^n$ such that $P = \{\, x \in \mathbb{R}^{r-1} \mid Ax + b \geq 0 \,\}$ is a full-dimensional polytope, and m rational points that span a full-dimensional polytope $S \subseteq P$ such that
   - **(a)** any d-dimensional RNMF (rational or irrational) of M determines d points that span a polytope Q with $S \subseteq Q \subseteq P$, and
   - **(b)** any d points (rational or irrational) that span a polytope Q with $S \subseteq Q \subseteq P$ determine a d-dimensional RNMF of M.
2. Let $A \in \mathbb{Q}^{n\times(r-1)}$ and $b \in \mathbb{Q}^n$ such that $P = \{\, x \in \mathbb{R}^{r-1} \mid Ax+b \geq 0 \,\}$ is a full-dimensional polytope. Let $S \subseteq P$ be a full-dimensional polytope spanned by $s_1, \ldots, s_m \in \mathbb{Q}^{r-1}$. Then matrix $M \in \mathbb{Q}^{n\times m}$ with $M^i = As_i + b$ for $i \in [m]$ satisfies (a) and (b).

Importantly, the correspondences (a) and (b) preserve rationality. In the full version we detail the reduction from point 2 above, thereby filling in a small gap in the proof of [11].

▶ **Example 3** ([11, Example 1])**.** Using the geometric interpretation of restricted nonnegative rank it follows easily that, in general, we may have $\mathrm{rank}(M) < \mathrm{rank}_+(M) < \mathrm{rrank}_+(M)$. Let *3D-cube* NPP be the NPP instance where the inner and outer polytope are the standard 3D cube, i.e., $P = S = \{\, x \in \mathbb{R}^3 \mid x_i \in [0,1],\ 1 \leq i \leq 3 \,\}$. The only nested polytope is $Q = P$. The corresponding restricted NMF problem consists of the following matrix $M \in \mathbb{R}_+^{6\times 8}$:

$$M = \begin{pmatrix} 0&0&0&0&1&1&1&1 \\ 1&1&1&1&0&0&0&0 \\ 0&0&1&1&0&0&1&1 \\ 1&1&0&0&1&1&0&0 \\ 0&1&0&1&0&1&0&1 \\ 1&0&1&0&1&0&1&0 \end{pmatrix}.$$

We have $\mathrm{rrank}_+(M) = 8$ and $\mathrm{rank}(M) = 4$. Since $\mathrm{rank}_+(M)$ is bounded above by the number of rows in M, we have $\mathrm{rank}_+(M) \leq 6$. It is shown in [11] that $\mathrm{rank}_+(M) = 6$.

3 Coverability of Labelled Markov Chains

In this section, we establish a connection between RNMF and the coverability relation for labelled Markov chains. We thereby answer an open question posed in 1971 by Paz [15] about the nature of minimal covering labelled Markov chains.

A *labelled 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 $\mu : \Sigma \to [0,1]^{n\times n}$ specifies the transition matrices and is such that $\sum_{\sigma\in\Sigma} \mu(\sigma)$ is a row-stochastic matrix. The intuitive behaviour of the LMC $\mathcal{M}$ is as follows: When $\mathcal{M}$ is in state $i \in [n]$, it emits label σ and moves to state j, with probability $\mu(\sigma)_{i,j}$.

We extend the function μ to words by defining $\mu(\varepsilon) := I_n$ and $\mu(\sigma_1 \ldots \sigma_k) := \mu(\sigma_1) \cdots \mu(\sigma_k)$ for all $k \in \mathbb{N}$, and all $\sigma_1, \ldots, \sigma_k \in \Sigma$. Observe that $\mu(xy) = \mu(x) \cdot \mu(y)$ for all words $x, y \in \Sigma^*$. We view $\mu(w)$ for a word $w \in \Sigma^*$ as follows: if $\mathcal{M}$ is in state $i \in [n]$, it emits w and moves to state j in $|w|$ steps, with probability $\mu(w)_{i,j}$.

For $i \in [n]$ and $w \in \Sigma^*$, we write $pr_i^{\mathcal{M}}(w) := e_i^\top \cdot \mu(w) \cdot \mathbf{1}^{(n)}$ for the probability that, starting in state i, $\mathcal{M}$ emits word w. For example, in Figure 1 we have $pr_0^{\mathcal{M}}(a_1 b_1) = \frac{1}{12}$. More generally, for a given *initial distribution* π on the set of states $[n]$ (viewed as a stochastic row vector), we write $pr_\pi^{\mathcal{M}}(w) := \pi \cdot \mu(w) \cdot \mathbf{1}^{(n)}$ for the probability that $\mathcal{M}$ emits word w starting from state distribution π.

We say that an LMC $\mathcal{M}$ is *covered by* an LMC $\mathcal{M}'$, written as $\mathcal{M}' \geq \mathcal{M}$, if for every initial distribution π for $\mathcal{M}$ there exists a distribution π' for $\mathcal{M}'$ such that $pr_\pi^{\mathcal{M}}(w) = pr_{\pi'}^{\mathcal{M}'}(w)$ for all words $w \in \Sigma^*$.

The *backward matrix* of $\mathcal{M}$ is a matrix $\mathrm{Back}\,\mathcal{M} \in \mathbb{R}_+^{[n]\times\Sigma^*}$ where $(\mathrm{Back}\,\mathcal{M})_{i,w} = pr_i^{\mathcal{M}}(w)$ for every $i \in [n]$ and $w \in \Sigma^*$. The *rank* of $\mathcal{M}$ is defined by $\mathrm{rank}(\mathcal{M}) = \mathrm{rank}(\mathrm{Back}\,\mathcal{M})$. (Matrix $\mathrm{Back}\,\mathcal{M}$ is infinite, but since it has n rows, its rank is at most n.) It follows easily from the definition (see also [15, Theorem 3.1]) that $\mathcal{M}' \geq \mathcal{M}$ if and only if there exists a row-stochastic matrix A such that $A \cdot \mathrm{Back}\,\mathcal{M}' = \mathrm{Back}\,\mathcal{M}$.

LMCs can be seen as a special case of stochastic sequential machines, a class of probabilistic automata introduced and studied by Paz [15]. More specifically, they are stochastic sequential machines with a singleton input alphabet and Σ as output alphabet. In his seminal 1971 textbook on probabilistic automata [15], Paz asks the following question:

▶ **Question 4** (Paz [15], p. 38). *If an n-state LMC $\mathcal{M}$ is covered by an n'-state LMC $\mathcal{M}'$ where $n' < n$, is $\mathcal{M}$ necessarily covered by some n^*-state LMC $\mathcal{M}^*$, where $n^* < n$, such that $\mathcal{M}^*$ and $\mathcal{M}$ have the same rank?*

In 1974, a positive answer to this question was claimed [3, Theorem 13]. In fact, the author of [3] makes a stronger claim, namely that the answer to Question 4 is yes, even if the inequality $n^* < n$ in Question 4 is replaced by $n^* \leq n'$. To the contrary, we show:

▶ **Theorem 5.** *The answer to Question 4 is negative.*

Theorem 5 falsifies the claim in [3]. In the full version we discuss in detail the mistake in [3]. To prove Theorem 5 we establish a tight connection between NMF and LMC coverability:

▶ **Proposition 6.** *Given a nonnegative matrix $M \in \mathbb{Q}_+^{n\times m}$ of rank r, one can compute in polynomial time an LMC $\mathcal{M} = (m+2, \Sigma, \mu)$ of rank $r+2$ such that for all $d \in \mathbb{N}$:*

(a) *any d-dimensional NMF $M = W \cdot H$ determines an LMC $\mathcal{M}' = (d+2, \Sigma, \mu')$ with $\mathcal{M}' \geq \mathcal{M}$ and $\mathrm{rank}(\mathcal{M}') = \mathrm{rank}(W) + 2$, and*

(b) *any LMC $\mathcal{M}' = (d+2, \Sigma, \mu')$ with $\mathcal{M}' \geq \mathcal{M}$ determines a d-dimensional NMF $M = W \cdot H$ with $\mathrm{rank}(\mathcal{M}') = \mathrm{rank}(W) + 2$.*

In particular, for all $d \in \mathbb{N}$ the inequality $\mathrm{rrank}_+(M) \leq d$ holds if and only if $\mathcal{M}$ is covered by some $(d+2)$-state LMC $\mathcal{M}'$ such that $\mathcal{M}'$ and $\mathcal{M}$ have the same rank.

Assuming Proposition 6 we can prove Theorem 5:

Proof of Theorem 5. Let $M \in \{0,1\}^{6\times 8}$ be the matrix from Example 3. Let $\mathcal{M} = (10, \Sigma, \mu)$ be the associated LMC from Proposition 6. Since $M = I_6 \cdot M$ is an NMF with inner dimension 6, by Proposition 6 (a) there is an LMC $\mathcal{M}' = (8, \Sigma, \mu')$ with $\mathcal{M}' \geq \mathcal{M}$. Towards a contradiction, suppose the answer to Question 4 were yes. Then $\mathcal{M}$ is also covered by some n^*-state LMC $\mathcal{M}^*$, where $n^* \leq 9$, such that $\mathcal{M}^*$ and $\mathcal{M}$ have the same rank. The last sentence of Proposition 6 then implies that $\mathrm{rrank}_+(M) \leq 7$. But this contradicts the equality $\mathrm{rrank}_+(M) = 8$ from Example 3. Hence, the answer to Question 4 is no. ◀

To prove Proposition 6 we adapt a reduction from NMF to the trace-refinement problem in Markov decision processes [10].

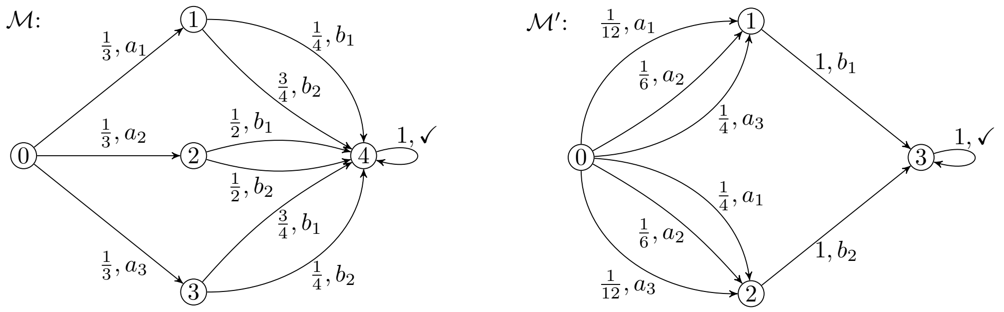

Figure 1 LMC $\mathcal{M}$ is constructed from matrix $M = \begin{pmatrix} 1/4 & 1/2 & 3/4 \\ 3/4 & 1/2 & 1/4 \end{pmatrix}$ whereas LMC $\mathcal{M}'$ is obtained by NMF $M = I_2 \cdot M$.

Proof sketch of Proposition 6. Let $M \in \mathbb{Q}_+^{n \times m}$ be a nonnegative matrix of rank r. As argued in Section 2.1, without loss of generality we may assume that M is column-stochastic and consider factorizations of M into column-stochastic matrices only.

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

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

and all other entries of $\mu(a_i)$, $\mu(b_j)$, and $\mu(\checkmark)$ are 0. See Figure 1 for an example. We have:

$$\operatorname{Back}\mathcal{M} = \begin{array}{c} \begin{array}{cccccccccccc} \varepsilon & b_1 & \cdots & b_n & \checkmark & a_i & a_ib_j & b_1\checkmark & \cdots & b_n\checkmark & \checkmark^2 & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \frac{1}{m} & \frac{1}{m}M_{j,i} & 0 & \cdots & 0 & 0 & \cdots \\ 1 & M_{1,1} & \cdots & M_{n,1} & 0 & 0 & 0 & M_{1,1} & \cdots & M_{n,1} & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots \\ 1 & M_{1,m} & \cdots & M_{n,m} & 0 & 0 & 0 & M_{1,m} & \cdots & M_{n,m} & 0 & \cdots \\ 1 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots \end{pmatrix} \end{array}.$$

Thus $\operatorname{rank}(\mathcal{M}) = \operatorname{rank}(\operatorname{Back}\mathcal{M}) \geq \operatorname{rank}(M) + 2$. The first $n + 2$ columns (indexed by $\varepsilon, b_1, \ldots, b_n, \checkmark$) in $\operatorname{Back}\mathcal{M}$ span $\operatorname{Col}(\operatorname{Back}\mathcal{M})$. Therefore, $\operatorname{rank}(\mathcal{M}) = \operatorname{rank}(M) + 2 = r + 2$.

For $d \in \mathbb{N}$, let $M = W \cdot H$ for some column-stochastic matrices $W \in \mathbb{R}_+^{n \times d}$ and $H \in \mathbb{R}_+^{d \times m}$. Define an LMC $\mathcal{M}' = (d + 2, \Sigma, \mu')$ where the states are $\{0, 1, \ldots, d, d + 1\}$. The function μ', for all $i \in [m]$, $j \in [n]$, and $l \in [d]$, is defined by:

$$\mu'(a_i)_{0,l} = \tfrac{1}{m}H_{l,i}, \qquad \mu'(b_j)_{l,d+1} = W_{j,l}, \qquad \mu'(\checkmark)_{d+1,d+1} = 1,$$

and all other entries of $\mu'(a_i)$, $\mu'(b_j)$, and $\mu'(\checkmark)$ are 0. From the NMF $M = W \cdot H$ it follows that we can factor $\operatorname{Back}\mathcal{M}$ as follows:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & H_{1,1} & \cdots & H_{d,1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & H_{1,m} & \cdots & H_{d,m} & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \cdot \begin{array}{c} \begin{array}{cccccccccc} \varepsilon & b_1 & \cdots & b_n & \checkmark & a_i & a_ib_j & b_j\checkmark & \checkmark^2 & \cdots \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \frac{1}{m} & \frac{1}{m}M_{j,i} & 0 & 0 & \cdots \\ 1 & W_{1,1} & \cdots & W_{n,1} & 0 & 0 & 0 & W_{j,1} & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ 1 & W_{1,d} & \cdots & W_{n,d} & 0 & 0 & 0 & W_{j,d} & 0 & \cdots \\ 1 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 & 1 & \cdots \end{pmatrix} \end{array}$$

where the left factor is row-stochastic (as H is column-stochastic), and the right factor equals $\operatorname{Back}\mathcal{M}'$. It follows that $\mathcal{M}' \geq \mathcal{M}$. ◀

4 Restricted NMF of Rank-3 Matrices

In this section we consider rational matrices of rank at most 3. We show that for such matrices the restricted nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ are equal and we give a polynomial-time algorithm that computes a minimal-dimension RNMF over $\mathbb{Q}$.

▶ **Theorem 7.** *Given a matrix $M \in \mathbb{Q}_+^{n\times m}$ where $\mathrm{rank}(M) \le 3$, there is a rational RNMF of M with inner dimension $\mathrm{rrank}_+(M)$ and it can be computed in polynomial time in the Turing model of computation.*

Using reduction 1 of Proposition 2, we can reduce in polynomial time the RNMF problem for rank-3 matrices to the 2-dimensional NPP, i.e., the nested polygon problem in the plane. As noted in Section 2.2, the correspondence between restricted nonnegative factorizations and nested polygons preserves rationality. Thus to prove Theorem 7 it suffices to prove:

▶ **Theorem 8.** *Given polygons $S \subseteq P \subseteq \mathbb{R}^2$ with rational vertices, there exists a minimum-vertex polygon Q nested between P and S that also has rational vertices. Moreover there is an algorithm that, given P and S, computes such a polygon in polynomial time in the Turing machine model.*

In fact, Aggarwal et al. [1] give an algorithm for the 2-dimensional NPP and prove that it runs in polynomial time in the RAM model with unit-cost arithmetic. However, they freely use trigonometric functions and do not address the rationality of the output of the algorithm nor its complexity in the Turing model. To prove Theorem 8 we show that, by adopting a suitable representation of the vertices of a nested polygon, the algorithm in [1] can be adapted so that it runs in polynomial time in the Turing model. We furthermore use this representation to prove that the minimum-vertex nested polygon identified by the resulting algorithm has rational vertices.

The remainder of the section is devoted to the proof of Theorem 8. We first recall some terminology from [1] and describe their algorithm.

A *supporting line segment* is a directed line segment, with its initial and final points on the boundary of the outer polygon P, that touches the inner polygon S on its left. A nested polygon with vertices on the boundary of P is said to be *supporting* if all but at most one of its edges are supporting line segments. A polygon nested between P and S is called *minimal* if it has the minimum number of vertices among all polygons nested between P and S. It is shown in [1, Lemma 4] that there is always a supporting polygon that is also minimal, and the algorithm given therein outputs such a polygon.

Let k be the number of vertices of a minimal nested polygon. Given a vertex v on the boundary of P, there is a uniquely defined supporting polygon Q_v with at most $k+1$ vertices. To determine Q_v one computes the supporting line segments $v_1v_2, \ldots, v_kv_{k+1}$, where $v_1 = v$; see Figure 2. Then Q_v is either the polygon with vertices $v_1, \ldots, v_k$ or the polygon with vertices $v_1, \ldots, v_{k+1}$. In the first case, Q_v is minimal. The idea behind the algorithm of [1] is to search along the boundary of P for an initial vertex v such that Q_v is minimal.

As a central ingredient for our proof of Theorem 8, we choose a convenient representation of the vertices of supporting polygons. To this end, we assume that the edges of P are oriented counter-clockwise, and we represent a vertex v on an edge pq of P by the unique $\lambda \in [0, 1]$ such that $v = (1-\lambda)p + \lambda q$. We call this the *convex representation* of v.

Similar to [1], we associate with each supporting line segment uv a *ray function* r, such that if λ is the convex representation of u then $r(\lambda)$ is the convex representation of v. The same ray function applies for supporting line segments $u'v'$ with u' in a small enough interval containing u.

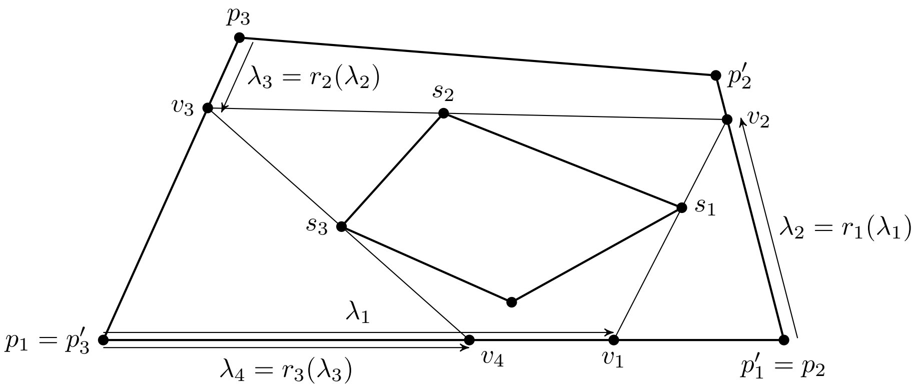

Figure 2 Supporting polygon Q_{v_1}. For every $i \in [3]$, vertex v_i lies on edge $p_i p_i'$ of P, and s_i is the point where the supporting line segment $v_i v_{i+1}$ touches the inner polygon S on its left.

In the following, when we say polynomial time, we mean polynomial time in the Turing model. To obtain a polynomial time bound, the key lemma is as follows:

▶ **Lemma 9.** *Consider bounded polygons $S \subseteq P \subseteq \mathbb{R}^2$ whose vertices are rational and of bit-length L. Then the ray function associated with a supporting line segment uv has the form $r(\lambda) = \frac{a\lambda+b}{c\lambda+d}$, where coefficients a, b, c, d are rational numbers with bit-length $O(L)$ that can be computed in polynomial time.*

Suppose that $v_1v_2, \ldots, v_kv_{k+1}$ is a sequence of k supporting line segments, with corresponding ray functions $r_1, \ldots, r_k$. Then $v_1, \ldots, v_k$ are the vertices of a minimal supporting polygon if and only if $(r_k \circ \ldots \circ r_1)(\lambda) \geq \lambda$, where λ is the convex representation of v_1.

It follows from [1] that, for each edge of P, one can compute in polynomial time a partition $\mathcal{I}$ of $[0, 1]$ into intervals with rational endpoints such that if λ_1, λ_2 are in the same interval $I \in \mathcal{I}$ then the points with convex representation λ_1 and λ_2 are associated with the same sequence of ray functions $r_1, \ldots, r_k$. Using Lemma 9 we can, for each interval $I \in \mathcal{I}$, compute these ray functions in polynomial time. Define the *slack function* $s(\lambda) = (r_k \circ \ldots \circ r_1)(\lambda) - \lambda$. In fact, this function has the form $s(\lambda) = \frac{a\lambda+b}{c\lambda+d} - \lambda$ for rational numbers a, b, c, d that are also computable in polynomial time. Then it is straightforward to check whether $s(\lambda) \geq 0$ has a solution $\lambda \in I$.

Next we show that if such a solution exists, then there exists a rational solution, which, moreover, can be computed in polynomial time. To this end, let $\lambda^* \in I$ be such that $s(\lambda^*) \geq 0$. If λ^* is on the boundary of I, then $\lambda^* \in \mathbb{Q}$. If λ^* is not on the boundary and is not an isolated solution, then there exists a rational solution in its neighbourhood. Lastly, let λ^* be an isolated solution not on the boundary. Then, λ^* is a root of both s and its derivative s'. For every $\lambda \in I$, we have

$$(c\lambda + d) \cdot s(\lambda) = a\lambda + b - \lambda \cdot (c\lambda + d).$$

Taking the derivative of the above equation with respect to λ, we get

$$c \cdot s(\lambda) + (c\lambda + d) \cdot s'(\lambda) = a - d - 2c\lambda. \tag{1}$$

Since $s(\lambda^*) = s'(\lambda^*) = 0$, from (1) we get $0 = a - d - 2c\lambda^*$. Note that $c \neq 0$ since otherwise $s \equiv 0$. Therefore, $\lambda^* = \frac{a-d}{2c} \in \mathbb{Q}$.

It follows that the vertex v represented by λ^* has rational coordinates computable in polynomial time. By computing $(r_i \circ \ldots \circ r_1)(\lambda^*)$ for $i \in [k]$, we can compute in polynomial time the convex representation of all vertices of the supporting polygon Q_v. Observe, in particular, that all vertices are rational. Hence we have proved Theorem 8.

5 Restricted NMF Requires Irrationality

Here we show that the restricted nonnegative ranks over $\mathbb{R}$ and $\mathbb{Q}$ are, in general, different.

▶ **Theorem 10.** *Let*

$$M = \begin{pmatrix} 1/8 & 1/2 & 17/22 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 3/4 & 7/12 \\ 3/4 & 3/4 & 3/11 & 2 & 1/2 & 1/6 \\ 1/4 & 1/4 & 8/11 & 1/4 & 19/8 & 55/24 \\ 1/2 & 1/8 & 1/11 & 1/8 & 15/16 & 17/16 \\ 11/16 & 5/16 & 7/44 & 1/16 & 7/32 & 43/96 \end{pmatrix} \in \mathbb{Q}_+^{6\times 6}.$$

The restricted nonnegative rank of M over $\mathbb{R}$ is 5*. The restricted nonnegative rank of M over $\mathbb{Q}$ is* 6*.*

Proof. Matrix M has an NMF $M = W \cdot H$ with inner dimension 5 with W, H as follows:

$$W = \begin{pmatrix} 0 & \frac{3+\sqrt{2}}{14} & \frac{11+\sqrt{2}}{14} & 0 & 0 \\ 0 & 0 & 0 & \frac{12-2\sqrt{2}}{17} & \frac{5}{7}+\frac{\sqrt{2}}{14} \\ 2-\sqrt{2} & 1 & \frac{3-\sqrt{2}}{7} & \frac{26+7\sqrt{2}}{17} & 0 \\ -1+\sqrt{2} & 0 & \frac{4+\sqrt{2}}{7} & \frac{21-12\sqrt{2}}{17} & \frac{39}{14}+\frac{5\sqrt{2}}{28} \\ \frac{\sqrt{2}}{2} & \frac{2}{7}-\frac{\sqrt{2}}{14} & 0 & \frac{7-4\sqrt{2}}{17} & \frac{33}{28}+\frac{\sqrt{2}}{56} \\ \frac{1}{2}+\frac{\sqrt{2}}{4} & \frac{15}{28}-\frac{\sqrt{2}}{14} & \frac{3-\sqrt{2}}{28} & 0 & \frac{3}{8}-\frac{\sqrt{2}}{16} \end{pmatrix},$$

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

Since $\operatorname{rank}(M) = \operatorname{rank}(W) = 4$, the NMF $M = W \cdot H$ is restricted. This RNMF has been obtained by reducing, according to Proposition 2, an NPP instance, which we now describe.

Figure 3 shows the outer 3-dimensional polytope P with 6 faces. The polytope P is the intersection of the following half-spaces: $y \geq 0$ (blue), $z \geq 0$ (brown), $x \geq 0$ (pink), $-x + \frac{5}{2}z + 1 \geq 0$ (yellow), $-\frac{1}{2}x - y + \frac{1}{4}z + 1 \geq 0$ (green), $-\frac{1}{4}x - y - \frac{7}{8}z + 1 \geq 0$ (transparent front). The figure also indicates an interior polytope S spanned by 6 points (black dots): $s_1 = (\frac{3}{4}, \frac{1}{8}, 0)^\top$, $s_2 = (\frac{3}{4}, \frac{1}{2}, 0)^\top$, $s_3 = (\frac{3}{11}, \frac{17}{22}, 0)^\top$, $s_4 = (2, 0, \frac{1}{2})^\top$, $s_5 = (\frac{1}{2}, 0, \frac{3}{4})^\top$, $s_6 = (\frac{1}{6}, 0, \frac{7}{12})^\top$. In the following we discuss possible locations of 5 points q_1, q_2, q_3, q_4, q_5 that span a nested polytope Q. Since s_1, s_2, s_3 all lie on the (brown) face on the xy-plane, but not on a common line, at least 3 of the q_i must lie on the xy-plane. A similar statement holds for s_4, s_5, s_6 and the xz-plane. So at least one q_i, say q_1, must lie on the x-axis.

Suppose another q_i, say q_2, lies on the x-axis. Without loss of generality we can take $q_1 = (0,0,0)^\top$ and $q_2 = (1,0,0)^\top$, as all points in P on the x-axis are enclosed by these q_1, q_2.

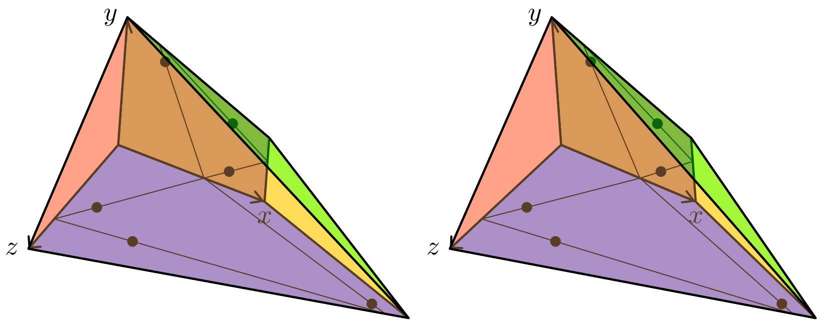

Figure 3 Instance of the nested polytope problem. The two images show orthogonal projections of a 3-dimensional outer polytope P. The black dots indicate 6 inner points (3 on the brown xy-face, and 3 on the blue xz-face) that span the interior polytope S. The two triangles on the xy-face and on the xz-face indicate the (unique) location of 5 points that span the nested polytope Q. 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., [18]. See the full version for a "cross-eyed" variant.

Figure 4 provides a detailed view of the xy-plane. To enclose s_2, some $q \in \{q_3, q_4, q_5\}$ must also lie on the xy-plane and to the right of the line that connects $q_2 = (1,0,0)^\top$ and s_2. To enclose s_3, some $q' \in \{q_3, q_4, q_5\}$ must also lie on the xy-plane and to the left of the line that connects $q_1 = (0,0,0)^\top$ and s_3. If q and q' were identical then they would lie outside P – a contradiction. Hence 4 points (namely, q_1, q_2, q, q') are on the xy-plane. This leaves only one point, say q'', that is not on the xy-plane. To enclose s_4 (see the corresponding figure in the full version), point q'' must lie on the xz-plane and must lie to the right of the line that connects $q_2 = (1,0,0)^\top$ and s_4. To enclose s_6, point q'' must lie to the left of the line that connects $q_1 = (0,0,0)^\top$ and s_6. Hence q'' lies outside P – a contradiction.

Hence we have shown that only one point, say q_1, lies on the x-axis, and two points besides q_1, say q_2, q_3, lie on the xy-plane, and two points besides q_1, say q_4, q_5, lie on the xz-plane. Figure 4 indicates a possible location (q_1^*, q_2^*, q_3^*) of q_1, q_2, q_3. The figure illustrates that the x-coordinate of q_1^* must be at least $2 - \sqrt{2}$.

Figure 5 illustrates how to prove the same fact more formally, using the concept of a slack function (see Section 4): The slack function $s(\lambda)$ for the interval containing $2 - \sqrt{2}$ has a zero at $\lambda = 2 - \sqrt{2}$, with a sign change from negative to positive. An inspection of the intervals (of the partition $\mathcal{I}$ from Section 4) to the "left" of $2 - \sqrt{2}$ reveals that none of the corresponding slack functions $\tilde{s}$ satisfies $\tilde{s}(\lambda) \geq 0$ for $\lambda < 2 - \sqrt{2}$. Similarly, the x-coordinate of q_1^* must be *at most* $2 - \sqrt{2}$, see corresponding figures in the full version. Hence $q_1^* = (2 - \sqrt{2}, 0, 0)^\top$ is necessary. This uniquely (up to permutations) determines q_2^*, q_3^* and similarly the locations q_4^*, q_5^* of q_4, q_5. With the reduction from Proposition 2 this NPP solution determines the RNMF of M mentioned at the beginning of the proof. Since there is no 4-point solution of the NPP instance, we have $\mathrm{rrank}_+(M) = 5$. (Since $\mathrm{rank}(M) = 4$, Lemma 1 implies $\mathrm{rank}_+(M) = 5$.) Since there is no 5-point rational solution of the NPP instance, the restricted nonnegative rank of M over $\mathbb{Q}$ is 6. ◀

6 Conclusion and Future Work

We have shown that an optimal *restricted* nonnegative factorization of a rational matrix may require factors that have irrational entries. An outstanding open problem is whether the same holds for general nonnegative factorizations. An answer to this question will likely shed light on the issue of whether the nonnegative rank can be computed in NP.

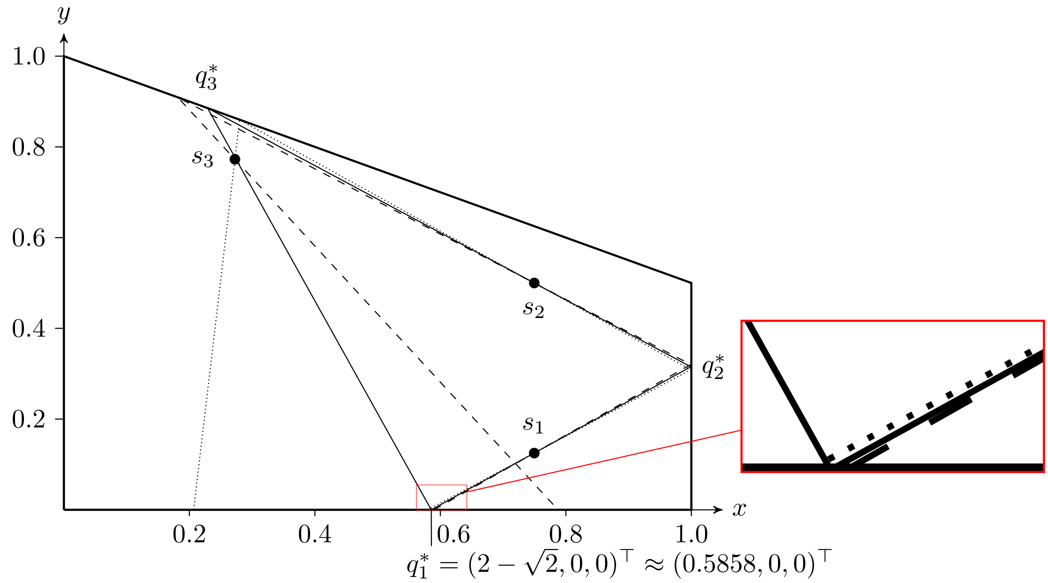

Figure 4 Detailed view of the xy-plane. The outer quadrilateral is one of 6 faces of P, the brown face in Figure 3. The points s_1, s_2, s_3 are among the 6 points that span the inner polytope S. The points q_1^*, q_2^*, q_3^* are among the 5 points that span the nested polytope Q. The area around q_1^* is zoomed in on the right-hand side. The picture illustrates that q_1^* cannot be moved left on the x-axis without increasing the number of vertices of the nested polytope: A dotted ray from a point slightly to the left of q_1^* is drawn through s_1. Its intersection with the line $x = 1$ is slightly below q_2^*. Following the algorithm of [1], the dotted ray is continued in a similar fashion, "wrapping around" s_2 and s_3, and ending on the x-axis at around $x \approx 0.2$, far left of the starting point. On the other hand, the dashed line illustrates that q_1^* could be moved right (considering only this face).

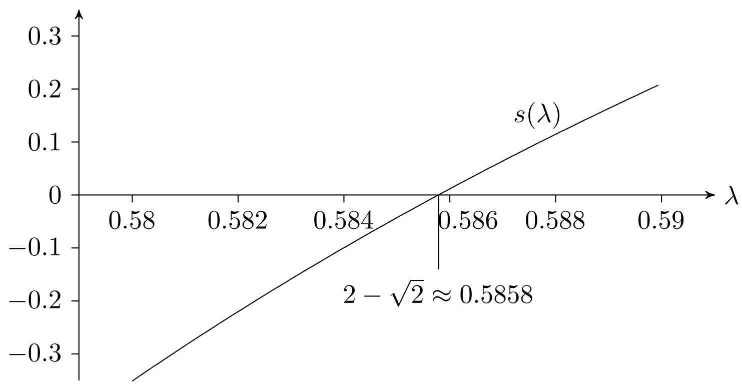

Figure 5 The slack function $s(\lambda) = \frac{52\lambda - 30}{15\lambda - 8} - \lambda$ corresponding to Figure 4. When $s(\lambda) < 0$, there is no nested triangle with vertex $(\lambda, 0, 0)$.

Another contribution of the paper has been to develop connections between nonnegative matrix factorization and probabilistic automata, thereby answering an old question concerning the latter. Pursuing this connection, and closely related to the above-mentioned open problem, one can ask whether, given a probabilistic automaton with rational transition probabilities, one can always find a minimal equivalent probabilistic automaton that also has rational transition probabilities.

Acknowledgements. The authors would like to thank Michael Benedikt for stimulating discussions, and anonymous referees for their helpful suggestions.

References

1. A. Aggarwal, H. Booth, J. O'Rourke, S. Suri, and C. K. Yap. Finding minimal convex nested polygons. *Information and Computation*, 83(1):98–110, 1989.
2. 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)*, pages 145–162, 2012.
3. F. Bancilhon. A geometric model for stochastic automata. *IEEE Trans. Computers*, 23(12):1290–1299, 1974.
4. M. W. Berry, N. Gillis, and F. Glineur. Document classification using nonnegative matrix factorization and underapproximation. In *International Symposium on Circuits and Systems (ISCAS)*, pages 2782–2785. IEEE, 2009.
5. S. S. Bucak and B. Günsel. Video content representation by incremental non-negative matrix factorization. In *Proceedings of the International Conference on Image Processing (ICIP)*, pages 113–116. IEEE, 2007.
6. J. Canny. Some algebraic and geometric computations in PSPACE. In *Proceedings of the 20th Annual Symposium on Theory of Computing (STOC)*, pages 460–467, 1988.
7. A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley Publishing, 2009.
8. J. E. Cohen and U. G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra and its Applications*, 190:149–168, 1993.
9. 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)*, pages 1141–1148, 2003.
10. 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)*, volume 9634 of *Lecture Notes in Computer Science*, pages 303–318. Springer, 2016.
11. N. Gillis and F. Glineur. On the geometric interpretation of the nonnegative rank. *Linear Algebra and its Applications*, 437(11):2685–2712, 2012.
12. N. Gillis and S. A. Vavasis. Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization. *SIAM Journal on Optimization*, 25(1):677–698, 2015.
13. A. Kumar, V. Sindhwani, and P. Kambadur. Fast conical hull algorithms for near-separable non-negative matrix factorization. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, page 231–239, 2013.
14. D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
15. A. Paz. *Introduction to probabilistic automata*. Academic Press, New York, 1971.

16 J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. Parts I–III. *Journal of Symbolic Computation*, 13(3):255–352, 1992.

17 Y. Shitov. Nonnegative rank depends on the field. Technical report, arxiv.org, 2015. Available at `http://arxiv.org/abs/1505.01893`.

18 D. Simanek. How to view 3D without glasses. `https://www.lhup.edu/~dsimanek/3d/view3d.htm`. Online, accessed in April 2016.

19 E. Tjioe, M. W. Berry, and R. Homayouni. Discovering gene functional relationships using FAUN (feature annotation using nonnegative matrix factorization). *BMC Bioinformatics*, 11(S-6):S14, 2010.

20 S. A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009.

21 T. Yokota, R. Zdunek, A. Cichocki, and Y. Yamashita. Smooth nonnegative matrix and tensor factorizations for robust multi-way data analysis. *Signal Processing*, 113:234–249, 2015.

22 S. Zhang, W. Wang, J. Ford, and F. Makedon. Learning from incomplete ratings using nonnegative matrix factorization. In *Proceedings of the 6th SIAM International Conference on Data Mining*, pages 549–553. SIAM, 2006.