Extended Formulations, Nonnegative Factorizations, and Randomized Communication Protocols

Yuri Faenza[1,*], Samuel Fiorini[2,**], Roland Grappe[3,***], and Hans Raj Tiwary[2,†]

[1] Dipartimento di Matematica Pura e Applicata, Università di Padova, Via Trieste 63, 35121 Padova, Italy
faenza@math.unipd.it

[2] Département de Mathématique, Université Libre de Bruxelles CP 216, Boulevard du Triomphe, 1050 Brussels, Belgium
{sfiorini,htiwary}@ulb.ac.be

[3] Laboratoire d'Informatique de Paris-Nord, UMR CNRS 7030, Institut Galilée - Université Paris-Nord, Avenue Jean-Baptiste Clément, 93430 Villetaneuse, France
roland.grappe@lipn.univ-paris13.fr

Abstract. We show that the binary logarithm of the nonnegative rank of a nonnegative matrix is, up to small constants, equal to the minimum complexity of a randomized communication protocol computing the matrix in expectation. We use this connection to prove new conditional lower bounds on the sizes of extended formulations, in particular, for perfect matching polytopes.

Keywords: Extended formulations, Nonnegative rank, Communication protocols.

1 Introduction

Extended formulations are a powerful tool for minimizing linear or, more generally, convex functions over polyhedra (see, e.g., Ziegler [19] for background on polyhedra and polytopes). Consider a polyhedron P in $\mathbb{R}^d$ and a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$, that has to be minimized over P. If a small size linear description of P is known, then minimizing φ over P can be done efficiently using an interior point algorithm, or the simplex algorithm if φ is linear and theoretical efficiency

[*] Supported by the Progetto di Eccellenza 2008–2009 of the Fondazione Cassa di Risparmio di Padova e Rovigo.

[**] Partially supported by the *Actions de Recherche Concertées* (ARC) fund of the French community of Belgium.

[***] Partially supported by the Progetto di Eccellenza 2008–2009 of the Fondazione Cassa di Risparmio di Padova e Rovigo.

[†] Postdoctoral Researcher of the *Fonds National de la Recherche Scientifique* (F.R.S.–FNRS).

A.R. Mahjoub et al. (Eds.): ISCO 2012, LNCS 7422, pp. 129–140, 2012.

is not required. However, P can potentially have many facets. Or worse: it can be that no explicit complete linear description of P is known. This does not necessarily make the optimization problem at hand difficult, since the existence of an efficient algorithm solving the separation problem for P implies that optimizing over P can be done efficiently (see [9]). However, this result uses the ellipsoid algorithm, which is useless practically.

Now suppose there exists a polyhedron Q in a higher dimensional space $\mathbb{R}^e$ such that P is the image of Q under a linear projection $\pi : \mathbb{R}^e \to \mathbb{R}^d$. The polyhedron Q together with π define an *extended formulation*, or *extension* of P. Minimizing φ over P amounts to minimizing $\varphi \circ \pi$ over Q. If Q has few facets, then we can resort to an interior point algorithm or the simplex algorithm to solve the optimization problem. Of course, one should also take into account the size of the coefficients in the linear description of Q and in the matrix of π, but we will ignore this here. The success of extended formulations is due to the fact that a moderate increase in dimension can result in a dramatic decrease in the number of facets. As we will see later in the paper, P may have exponentially many facets, while Q only polynomially many (see also the recent surveys by Conforti et al. [4] and by Kaibel [10] for other examples).

We define the *size* of an extension Q as the number of facets of Q, and the *extension complexity* of a polyhedron P as the minimum size of any extension of P. Following [7], we denote this by $\mathrm{xc}(P)$. The extension complexity of a polyhedron is a far better measure of how "complex" a polyhedron is than, for instance, its number of facets or its number of vertices and extreme rays.

Because we mainly consider polytopes, we assume from now on that P is bounded, that is, P is a polytope. This is not a major restriction. So consider a polytope P in $\mathbb{R}^d$ with m facets and n vertices. Let $h_1, \ldots, h_m$ be m affine functions on $\mathbb{R}^d$ such that $h_1(x) \geqslant 0, \ldots, h_m(x) \geqslant 0$ are the facet-defining inequalities of P. Let also $v_1, \ldots, v_n$ denote the vertices of P. The *slack matrix* of P is the nonnegative $m \times n$ matrix $S = S(P) = (s_{ij})$ with $s_{ij} = h_i(v_j)$.

A *rank-r nonnegative factorization* of a nonnegative matrix M is an expression of M as a product $M = AB$ where A and B are nonnegative matrices with r columns and r rows, respectively. The *nonnegative rank* of M, denoted by $\mathrm{rank}_+(M)$, is the minimum natural r such that M admits a rank-r nonnegative factorization [3]. Observe that the nonnegative rank of M can also be defined as the minimum natural r such that M is the sum of r nonnegative rank one matrices. In a seminal paper, Yannakakis [18] proved, among other things, that the extension complexity of a polytope is precisely the nonnegative rank of its slack matrix (see also [7]).

Theorem 1. *For all polytopes P, $\mathrm{xc}(P) = \mathrm{rank}_+(S(P))$.*

Before going on, we sketch the proof of half of the theorem. Assuming $P = \{x \in \mathbb{R}^d : Ex \leqslant g\}$, consider a rank r nonnegative factorization $S(P) = FV$ of the slack matrix of P. Then it can be shown that $Q := \{(x, y) \in \mathbb{R}^{d+r} : Ex + Fy = g,\ y \geqslant 0\}$ is an extension of P. Notice that Q has at most r facets, and r extra variables. Taking $r = \mathrm{rank}_+(S(P))$ implies $\mathrm{xc}(P) \leqslant \mathrm{rank}_+(S(P))$. Moreover, since P is a polytope, one can also assume that Q is bounded, as shown by the following lemma, whose proof we postpone to the journal version of the paper.

Lemma 2. *Let $P = \{x \in \mathbb{R}^d : Ex \leqslant g\}$ be a polytope, let $S(P) = FV$ be a rank-r nonnegative factorization of the slack matrix of P with $r := \operatorname{rank}_+(S(P))$, and let $Q := \{(x, y) \in \mathbb{R}^{d+r} : Ex + Fy = g,\ y \geqslant 0\}$. Then Q is bounded.*

In the work of Yannakakis [18] also appeared a connection between extended formulations and communication complexity (the book of Kushilevitz and Nisan [13] is a standard reference on communication complexity). Every deterministic communication protocol computing a nonnegative matrix M (traditionally M is a binary matrix) yields a nonnegative factorization of M, and thus an extended formulation. Indeed, such a protocol defines a partition of the matrix into submatrices whose entries are all equal. Notice that the rows and columns of such a "monochromatic" submatrix are not necessarily consecutive. Each submatrix yields a nonnegative rank one matrix, and the sum of the resulting matrices is precisely M. The rank of this nonnegative factorization of M is at most 2^c, where c is the complexity of the protocol. When M is the slack matrix of a polytope P, we obtain an extension of P.

Notably, Yannakakis [18] used this connection to obtain a quasipolynomial size extended formulation for the stable set polytope of a n-vertex perfect graph from a deterministic communication protocol computing the corresponding slack matrix with polylogarithmic complexity.

The aim of this paper is to prove new results on extended formulations by tightening the connection between extended formulations, nonnegative factorizations and communication complexity. In Section 2, we define the different polytopes considered here, and describe their facets and vertices. In Section 3, we discuss deterministic and randomized communication protocols and define what it means for a randomized communication protocol with private randomness and nonnegative outputs to compute a given nonnegative matrix M in expectation. Then we prove in Sections 4 and 5 that the minimum complexity of a randomized protocol with nonnegative outputs computing M in expectation is, up to small additive constants, the binary logarithm of the nonnegative rank of M. This is done in two parts. Let c denote the minimum complexity of a randomized protocol computing M in expectation, and let $r := \operatorname{rank}_+(M)$. First, in Section 4, we prove the inequality $\lg r \leqslant c$. (Throughout this paper, lg denotes the binary logarithm.) Second, in Section 5, we prove the converse inequality $c \leqslant \lg r + O(1)$. The two inequalities together imply $\lg r = c + \Theta(1)$. By Theorem 1, we obtain a new characterization of the extension complexity of polytopes. The above results were recently[1] generalized to a correspondence between *semidefinite* extended formulations of a polytope P and *quantum* one-way protocols computing the slack matrix $S(P)$ in expectation, see Fiorini et al. [8]. In the context of that work, this SDP-quantum correspondence eventually led

[1] We point out that the present paper was completed roughly one year before [8] but was only submitted one month later.

to superpolynomial lower bounds on the extension complexity of the cut, stable set and TSP polytopes [8].

Finally, in Section 6, we use this characterization to prove new results on extended formulations of perfect matching polytopes, a prominent family of polytopes for which the extension complexity is unknown. Yannakakis [18] proved that every *symmetric* extension of the perfect matching polytope of the complete graph K^n has exponential size (we do not formally define *symmetric* here; the interested reader may refer to [18]). Here, we show roughly that there is a tradeoff between the amount of randomness used by an extension of the perfect matching polytope K^n, regarded as a randomized protocol, and the size of this extension. In particular, we prove that if the protocol detects non-zero entries of the slack matrix with constant probability, then the extension has exponential size. A similar result holds for the spanning tree polytope of K^n.

2 Polytopes Relevant to This Work

Now we describe briefly various families of polytopes relevant to this paper. For a more detailed presentation of those we refer the reader to Schrijver [16].

Let I be a finite ground set. The *characteristic vector* of a subset $J \subseteq I$ is the vector $\chi^J \in \mathbb{R}^I$ such that, for each $i \in I$, $\chi^J_i = \begin{cases} 1 & \text{if } i \in J \\ 0 & \text{if } i \notin J \end{cases}$. For $x \in \mathbb{R}^I$, we let $x(J) := \sum_{i \in J} x_i$. Throughout this section, $G = (V, E)$ denotes a (simple, undirected) graph. For $U \subseteq V$, we denote the edges of the subgraph induced by U as $E[U]$. The *cut* defined by U, denoted as $\delta(U)$, is the set of edges of G exactly one of whose endpoints is in U. In this paper, we will often take G to be the *complete graph* K^n with vertex set $V(K^n) = [n] := \{1, \dots, n\}$ and edge set $E(K^n) = \{ij : i, j \in [n], i \neq j\}$.

A *spanning tree* of G is a tree T (i.e. a simple, connected graph without cycles) whose set of vertices and edges respectively satisfy $V(T) = V$ and $E(T) \subseteq E$. The *spanning tree polytope* $\mathrm{P}_{\text{spanning tree}}(G)$ of G is the convex hull of the characteristic vectors of the spanning trees of G. Edmonds [6] showed that this polytope admits the following linear description (see also [16], page 861):

$$\begin{aligned} x(E[U]) &\leqslant |U| - 1 && \text{for nonempty } U \subseteq V, \\ x(E) &= |V| - 1, \\ x_e &\geqslant 0 && \text{for } e \in E. \end{aligned}$$

A *perfect matching* of G is set of edges $M \subseteq E$ such that every vertex of G is incident to exactly one edge in M. The *perfect matching polytope* $\mathrm{P}_{\text{perfect matching}}(G)$ of the graph G is the convex hull of the characteristic vectors of the perfect matchings of G. Edmonds [5] showed that the perfect matching polytope of G can be described as follows (see also [16], page 438):

$$\begin{aligned} x(\delta(U)) &\geqslant 1 && \text{for } U \subseteq V \text{ with } |U| \text{ odd}, \\ x(\delta(v)) &= 1 && \text{for } v \in V, \\ x_e &\geqslant 0 && \text{for } e \in E. \end{aligned}$$

3 Communication Complexity

Let X, Y and Z be arbitrary finite sets with $Z \subseteq \mathbb{R}_+$, and let $f : X \times Y \to Z$ be a function. Suppose that there are two players Alice and Bob who wish to compute $f(x, y)$ for some inputs $x \in X$ and $y \in Y$. Alice knows only x and Bob only y. Hence, they must exchange information to be able to compute the value of $f(x, y)$, even though each player possesses unlimited computational power.

The communication is carried out as a protocol that is agreed on beforehand by Alice and Bob, on the sole basis of the function f. At each step of the protocol, one of the player has the token. He/she sends a bit to the other, that depends only on his/her input and on previously exchanged bits. The transmitted bit determines which player has the token in the next step. This is repeated until the value of f on (x, y) is known to both players. The minimum number of bits exchanged between the players in the worst case to be able to evaluate f by any protocol is called the *communication complexity* of f.

In this section we describe deterministic protocols briefly and then randomized protocols (with private random bits). In the literature, a randomized protocol is said to compute a function f if for all inputs $(x, y) \in X \times Y$ the protocol outputs the correct value, namely $f(x, y)$, with high probability. Here we consider a new notion of computation where the value output by the protocol on input (x, y) has to equal $f(x, y)$ in expectation. For a thorough description of deterministic as well as randomized protocols (with the usual notion of computation) we refer the reader to the book by Kushilevitz and Nisan [13].

3.1 Deterministic Protocols

A protocol is best viewed as a rooted binary tree where each internal node is marked either Alice or Bob. The leaves have values associated with them. An execution of the protocol on a particular input is a root-to-leaf path in the tree. At a node owned by Alice, following the path to the left subtree corresponds to Alice sending a zero to Bob and taking the right subtree corresponds to Alice sending a one to Bob; and similarly for nodes owned by Bob. In case the protocol is deterministic, to each input $(x, y) \in X \times Y$ corresponds a unique path from the root to one of the leaves, and the value at that leaf is $f(x, y)$. Thus none of the players use any randomness to decide which bits to send to the other player.

More formally, we define a *deterministic protocol* as a rooted binary tree with some extra information attached to its nodes. Each internal node has a *type*, which is either X or Y. To each node v of type X is attached a function $p_v : X \to \{0, 1\}$; to each node v of type Y is attached a function $q_v : Y \to \{0, 1\}$; and to each leaf v is attached a number $\lambda_v \in \mathbb{R}_+$, called the *value* of that leaf. An *execution* of the protocol on input $(x, y) \in X \times Y$ is a root-to-leaf path that starts at the root and descends to a leaf. At any internal node v of type X the execution follows the edge to the left child if $p_v(x) = 0$ and to the right child if $p_v(x) = 1$. Similarly, at any internal node v of type Y, the execution follows the edge to the left child if $q_v(y) = 0$ and to the right child if $q_v(x) = 1$. The *value of the execution* is the value of the leaf attained by the execution. A deterministic protocol is said to *compute* the function f if for each input pair (x, y) the value of the execution of the protocol is exactly $f(x, y)$. The *complexity* of a protocol is the height of the corresponding tree.

These formal definitions capture the informal ones given above. Observe that the nodes of type X are assigned to Alice, and those of type Y to Bob. Observe also that Alice and Bob have unlimited resources for performing their part of the computation. It is only the communication between the two players that is accounted for. When presenting a protocol, we shall often say that one of the two players sends an integer k rather than a binary value. This should be interpreted as the player sending the binary encoding of k or, in terms of the tree, as a sequence of $\lceil \log k \rceil$ nodes of the same type. Similarly, when we say that a node has $k \in Z_+$ children, and account for $\lceil \log k \rceil$ in the height of the tree.

Given an ordering $x_1, \ldots, x_m$ of the elements of X, and $y_1, \ldots, y_n$ of the elements of Y, we can visualize the function $f : X \times Y \to Z$ as a $m \times n$ nonnegative matrix $M = M(f) = (m_{ij})$ such that $m_{ij} = f(x_i, y_j)$ for all $(i, j) \in [m] \times [n]$. Now consider a deterministic protocol computing f. Each of its leaves v determines a subset of rows $R = R_v$ and columns $C = C_v$ such that any input (x_i, y_j), the execution of the protocol on (x_i, y_j) ends at leaf v if and only if $i \in R$ and $j \in C$, that is, $(i, j) \in R \times C$. On each of the inputs (x_i, y_j) with $(i, j) \in R \times C$, the function f evaluates to same value, namely the value at leaf v. The set $R \times C$ is called a *rectangle*. When v varies among the leaves of the protocol, the rectangles $R_v \times C_v$ form a partition of $[m] \times [n]$. It is easy to see that such a partition can be used to write M as a sum of nonnegative rank one matrices, one for each leaf. In fact, for each leaf v, define a $m \times n$ matrix M_v whose entry in the ith row and jth column is given by $f(x_i, y_j)$ if $i \in R_v$ and $j \in C_v$, and 0 otherwise. Thus the support of M_v is $R_v \times C_v$. Each of these matrices has rank at most one and we have that $M = \sum_{v \in L} M_v$, where L denotes the set of leaves of the protocol.

If M is the slack matrix of a polytope P, it follows from Theorem 1 that P has an extension of size at most $|L| \leqslant 2^c$, where c is the complexity of the protocol. This was first observed by Yannakakis [18]. He proved the existence of a $n^{O(\log n)}$ size extension for the stable set polytope of a n-vertex perfect graph by giving a $O(\log^2 n)$ complexity deterministic protocol for computing its slack matrix.

3.2 Randomized Protocols

Randomized protocols are similar to deterministic ones except the players are allowed to use random bits to decide what to send to the other player. As mentioned earlier, the notion of computation "in expectation" that we define here differs from the usual notion of computation "with high probability".

Let X and Y be finite sets. A *randomized protocol with private random bits and nonnegative outputs* (or shortly, a *randomized protocol*) is a rooted binary tree with some extra information attached to the nodes. Each internal node has a *type*, which is either X or Y. To each node v of type X is attached a function $p_v : X \to [0, 1]$; to each node v of type Y is attached a function $q_v : Y \to [0, 1]$;

and to each leaf v is attached a nonnegative number $\lambda_v \in \mathbb{R}_+$, called the *value* of that leaf. The functions p_v and q_v define *transition probabilities*.

An *execution* of the protocol on input $(x, y) \in X \times Y$ is a random root-to-leaf path that starts at the root and descends to the left child of an internal node v with probability $p_v(x)$ if v is of type X and $q_v(y)$ if v is of type Y, and to the right child of v with the complementary probability $1 - p_v(x)$ if v is of type X and $1 - q_v(y)$ if v is of type Y. The *value* of the execution is the value of the leaf attained by the execution.

For each fixed input $(x, y) \in X \times Y$, the value of an execution on input (x, y) is a random variable. We say that the protocol *computes* a function $f : X \times Y \to \mathbb{R}_+$ *in expectation* if the expectation of this random variable on each $(x, y) \in X \times Y$ is precisely $f(x, y)$. The *complexity* of a protocol is the height of the corresponding tree. Note that one player, say Alice, choosing and sending a number in $\{1, \dots, k\}$ with some probability function depending on x can be modeled within this framework with a tree of height $\lceil \log k \rceil$. As observed in Section 3.1, we can regard a function $f : X \times Y \to \mathbb{R}_+$ as a nonnegative matrix $M = M(f)$ with $m = |X|$ rows and $n = |Y|$ columns. Below, as is natural, we will not make a distinction between these two types of objects.

4 Factorizations from Protocols

Theorem 3. *If there exists a randomized protocol of complexity c computing a matrix $M \in \mathbb{R}_+^{X \times Y}$ in expectation, then the nonnegative rank of M is at most 2^c.*

Proof. Each node v of the protocol has a corresponding *traversal probability matrix* $P_v \in \mathbb{R}_+^{X \times Y}$ such that, for all inputs $(x, y) \in X \times Y$, the entry $P_v(x, y)$ is the probability that an execution on input (x, y) goes through node v. We claim that P_v is always a rank one matrix.

We prove this by induction on the depth of a node, starting from the root. When v is the root, P_v is an all-one matrix because all executions start at the root. Thus $P_v = \mathbf{1}\mathbf{1}^T$ is a rank one matrix in this case.

Next, consider a node u of depth at least one and its parent v. Without loss of generality, we assume that v is of type X, that is, v is assigned to Alice. Assume that $P_v = pq^T$ for some nonnegative vectors $p \in \mathbb{R}^X$ and $q \in \mathbb{R}^Y$. Then we have $P_u = p'q^T$ where $p'(x) = p(x)p_v(x)$ for $x \in X$ in case u is the left child of v, and $p'(x) = p(x)(1 - p_v(x))$ for $x \in X$ in case u is the right child of v. This proves the claim.

Finally, let L be the set of all leaves of the protocol and λ_v be the value at leaf v. Because the protocol computes M in expectation, for all inputs $(x, y) \in X \times Y$ we have $M(x, y) = \sum_{v \in L} \lambda_v P_v(x, y)$. Thus, $M = \sum_{v \in L} \lambda_v P_v$. Since the claim holds, each term in this last sum is a nonnegative rank one matrix. The theorem follows. □

Recall that the polytopes considered in this paper have some facet-defining inequalities enforcing nonnegativity of the variables along with other facet-defining inequalities. The next lemma (whose simple proof we skip) will allow us to ignore the rows corresponding to nonnegativity inequalities, and focus on the non-trivial parts of the slack matrices considered here.

Lemma 4. *Let $P \subseteq \mathbb{R}^d_+$ be a polytope and let $S'(P)$ denote the submatrix of $S(P)$ obtained by deleting the rows corresponding to nonnegativity inequalities. If there is a complexity c randomized protocol for computing $S'(P)$ in expectation, then there is a complexity $1 + \max\{c, \lceil \lg d \rceil\}$ randomized protocol for computing $S(P)$ in expectation.*

For the protocols constructed here, we always have $c \geqslant \lceil \lg d \rceil$. Because of Lemma 4, we can thus ignore the nonnegativity inequalities without blowing up the size of any extension by more than a factor of two. Moreover, in terms of lower bounds, it is always safe to ignore inequalities because the nonnegative rank of a matrix cannot increase when rows are deleted. We conclude this section with two examples: the first one is a reinterpretation of a well-known $O(n^3)$ size extended formulation for the spanning tree polytopes due to Martin [14]. The second one concerns the perfect matching polytopes and is implicit in Kaibel et al. [11].

Example 1. *Let P denote the spanning tree polytope of the complete graph K^n, see Section 2. The (non-trivial part of the) slack matrix of P has one column per spanning tree T and one row per proper nonempty subset U of vertices. The slack of T with respect to the inequality that corresponds to U is the number of connected components of the subgraph of T induced by U (denoted by $T[U]$ below) minus one.*

In terms of the corresponding communication problem, Alice has a proper nonempty set U and Bob a spanning tree T. Together, they wish to compute the slack of the pair (U, T). Alice sends the name of some (arbitrarily chosen) vertex u in U. Then Bob picks an edge e of T uniformly at random and sends to Alice the endpoints v and w of e as an ordered pair of vertices (v, w), where the order is chosen in such a way that w is on the unique path from v to u in the tree. That is, she makes sure that the directed edge (v, w) "points" towards the root u. Then Alice checks that $v \in U$ and $w \notin U$, in which case she outputs $n - 1$; otherwise she outputs 0.

The resulting randomized protocol is clearly of complexity $3 \lg n + O(1)$. Moreover, it computes the slack matrix in expectation because for each connected component of $T[U]$ distinct from that which contains u, there is exactly one directed edge (v, w) that will lead Alice to output a non-zero value. Since she outputs $(n - 1)$ in this case, the expected value of the protocol on pair (U, T) is $(n-1) \cdot (k-1)/(n-1) = k-1$, where k is the number of connected components of $T[U]$. The corresponding extended formulation has size $O(n^3)$.

For the next example, we will need the fact that one can cover the complete graph K^n with $k = O(2^{n/2}\text{poly}(n))$ balanced complete bipartite graphs $G_1,\ldots,$ G_k in such a way that every perfect matching of K^n is a perfect matching of at least one of the G_i's. Given a matching M of K^n and $X \subseteq [n]$, we say that X

is *compatible* with M if all the edges of M have exactly one end in X. We say that $X \subseteq [n]$ is an $(n/2)$*-subset* of $[n]$ if $|X| = n/2$. We defer the proof of the following lemma to the journal version of the paper.

Lemma 5. *Let n be an even positive integer. There exists a collection of $k = O(2^{n/2}\sqrt{n}\ln n)$ $(n/2)-$subsets $X_1,\ldots,$ X_k of $[n]$ such that for every perfect matching M of K^n at least one of the subsets X_i is compatible with M.*

Example 2. *Assume that n is even and let P denote the perfect matching polytope of the complete graph K^n with vertex set $[n]$, see Section 2. The (non-trivial part of the) slack matrix of P has one column per perfect matching M, and its rows correspond to odd sets $U \subseteq [n]$. The entry for a pair (U, M) is $|\delta(U) \cap M| - 1$ (recall that $\delta(U)$ denotes the set of edges that have one endpoint in U and the other endpoint in $\overline{U}$, the complement of U).*

We describe a randomized protocol for computing the slack matrix in expectation, of complexity at most $(1/2 + \varepsilon)n$, where $\varepsilon > 0$ can be made as small as desired by taking n large. First, Bob finds an $(n/2)$-subset $X \subseteq [n]$ that is compatible with his matching M, and tells the name of this subset to Alice, see Lemma 5. Then Alice checks which of X and $\overline{X}$ contains the least number of vertices of her odd set U. Without loss of generality, assume it is X. If $U \cap X = \varnothing$ then, because $U \subseteq \bar{X}$ and X is compatible with M, Alice can correctly infer that the slack is $|U| - 1$, and outputs this number. Otherwise, she picks a vertex u of $U \cap X$ uniformly at random and send its name to Bob. He replies by sending the name of u', the mate of u in the matching M. Alice then checks whether u' is in U or not. If u' is not in U, then she outputs $|U| - 1$. Otherwise u' is in U, and she outputs $|U| - 1 - 2|U \cap X|$. Telling the name of X can be done in at most $n/2 + \lg\sqrt{n} + \lg\lg n + O(1)$ bits, see Lemma 5. The extra amount of communication is $2\lg n + O(1)$ bits. In total, at most $(1/2 + \varepsilon)n$ bits are exchanged, for n sufficiently large ($\varepsilon > 0$ can be chosen arbitrarily). One easily checks that the expected value output by Alice (in the case $U \cap X \neq \varnothing$) is $|\delta(U) \cap M| - 1$, hence we conclude that the protocol correctly computes the slack matrix of the perfect matching polytope.

The resulting extension has size at most $2^{(1/2+\varepsilon)n} \leqslant (1.42)^n$, whereas the main result of Yannakakis [18] gives a lower bound of $\binom{n}{n/4} \geqslant (1.74)^n$ for the size of any symmetric *extension. (The two previous inequalities hold for sufficiently small $\varepsilon > 0$ and sufficiently large n.)*

5 Protocols from Factorizations

Theorem 6. *If the nonnegative rank of matrix $M \in \mathbb{R}_+^{m\times n}$ has a rank r nonnegative factorization, then there exists a randomized protocol computing M in expectation, whose complexity is at most $\lg r + O(1)$.*

Proof. Let $A \in \mathbb{R}_+^{m\times r}$ and $B \in \mathbb{R}_+^{r\times n}$ be nonnegative matrices such that $M = AB$. Let $\varDelta$ denote the maximum row sum of A. Thus, $M = (A/\varDelta)(\varDelta B)$. Let $\widehat{A}$ denote the $m \times (r + 1)$ matrix obtained from $A/\varDelta$ by appending a column whose entries are chosen so that each row-sum of $\widehat{A}$ is precisely 1. Thus $\widehat{A}$ is row-stochastic. Let $\widehat{B}$ denote the $(r+1) \times n$ matrix obtained from ΔB by appending a zero row. Notice that $M = \widehat{A}\widehat{B}$.

The protocol is as follows: Alice knows a row index i, and Bob knows a column index j. Together they want to compute $M(i,j)$ in expectation, by exchanging as few bits as possible. They proceed as follows: Alice selects a column index $k \in [r+1]$ according to the probabilities found in row i of matrix $\widehat{A}$, sends this index to Bob, and Bob outputs the entry of $\widehat{B}$ in row k and column j. This randomized protocol computes the matrix M in expectation. Indeed, the expected value on input (i,j) is $\sum_{k=1}^{r+1} \widehat{A}(i,k)\widehat{B}(k,j) = M(i,j)$. the number of bits exchanged is $\lceil \lg(r+1) \rceil$, thus the complexity is at most $\lg r + O(1)$. □

6 New Lower Bound for Perfect Matching Polytopes

We have seen that every extension of a polytope P corresponds to a randomized protocol computing its slack matrix $S(P)$ in expectation and vice-versa. Now we show in particular that for the perfect matching polytope if we restrict ourselves only to those extensions that can determine with a constant probability whether or not an entry in the slack matrix is zero (e.g. deterministic protocols), then every extension has an exponential size.

6.1 A Reduction from the Set Disjointness Problem

The *set disjointness problem* is the following communication problem: Alice and Bob each are given a subset of $[n]$. They wish to determine whether the two subsets intersect or not. In other words, Alice and Bob have to compute the *set disjointness function* DISJ defined by $\mathrm{DISJ}(A,B) = 1$ if A and B are disjoint subsets of $[n]$, and $\mathrm{DISJ}(A,B) = 0$ if A and B are non-disjoint subsets of $[n]$. It is known that any randomized protocol that computes the disjointness function *with high probability* (that is, the probability that the value output by the protocol is correct is, for each input, bounded from below by a constant strictly greater than 0) has $\Omega(n)$ complexity, see, e.g., Kushilevitz and Nisan [13], Babai et. al. [1], Kalyanasundaram and Schnitger [12], and Razborov [15].

To each matrix $M \in \mathbb{R}_+^{X \times Y}$, we associate the following communication problem, that we call the *support problem*: Alice is given a row x of M and Bob a column y of M. They wish to determine whether $M(x,y) = 0$ or $M(x,y) > 0$. In the first case, they output 0 and in the second case they output 1.

Lemma 7. *There is a reduction from the set disjointness problem for subsets of $[n]$ to the support problem for the slack matrix of the perfect matching polytope for perfect matchings of K^ℓ, where $\ell \leqslant 3n+8$, that uses $O(1)$ extra communication.*

Proof. Let A and B be the sets respectively given to Alice and Bob. After sending 1 bit to Alice, Bob and Alice can make sure that both B and its complement $[n] - B$ contain an even number of elements (eventually adding dummy elements to the initial ground set $[n]$). Let $k \leqslant n+2$ denote the number of elements currently in the ground set, and let $\ell := 3k + 2 \leqslant 3n + 8$. We define an odd set U and a perfect matching M as follows. First, we let

$$U := \{i : i \in A\} \cup \{i + k : i \in A\} \cup \{3k + 1\}.$$

Second, M is obtained by adding matching edges to the partial matching $\{\{i, i + k\} : i \in [k] - B\} \cup \{\{i + k, i + 2k\} : i \in B\} \cup \{\{3k + 1, 3k + 2\}\}$ in such a way that each of the extra edges matches two consecutive unmatched vertices both in $\{i : i \in [k]\}$ or both in $\{i + 2k : i \in [k]\}$. It can be easily verified that A and B are disjoint if and only if the slack for (U, M) is zero. The theorem follows. □

6.2 The New Lower Bound

Theorem 8. *Consider an extended formulation for the perfect matching polytope of K^n and a corresponding randomized protocol computing the slack matrix of this polytope in expectation. If the probability that the protocol outputs a non-zero value, given a pair (U, M) with positive slack, is at least $p(n)$, then the protocol has complexity $\Omega(np(n))$ and the extended formulation has size $2^{\Omega(np(n))}$.*

Proof. Let c be the complexity of the randomized protocol computing the (non-trivial part of the) slack matrix of the perfect matching polytope of K^n in expectation. From this protocol, we obtain a new randomized protocol for the corresponding support problem by $\lceil 1/p(n) \rceil$ independent executions of the given protocol, and outputting 1 if at least one of the executions led to a non-zero value or 0 otherwise. The new protocol is such that, for all pairs (U, M) with a positive slack, the probability of outputting a zero value is at most

$$(1 - p(n))^{\frac{1}{p(n)}} \leqslant \frac{1}{\mathrm{e}},$$

where e is the Euler's number. Thus, there is constant probability that the value returned by the algorithm is positive. This gives a randomized protocol of complexity $O(c/p(n))$ the outputs the correct solution to the support problem for the slack matrix of the perfect matching polytope with high probability. The theorem follows directly from Lemma 7 and from the fact that the set disjointness problem has randomized communication complexity $\Omega(n)$. □

A statement analogous to Theorem 8 holds for the spanning tree polytope of K^n as well, even though for this polytope an extended formulation of size $O(n^3)$ exists. We defer details to the journal version of the paper.

Theorem 9. *Consider an extended formulation for the spanning tree polytope of K^n and a corresponding protocol computing the slack matrix of this polytope. If the probability that the protocol outputs a non-zero value, given a pair (U, T) with positive slack, is at least $p(n)$, then the protocol has complexity $\Omega(np(n))$ and the extended formulation has size $2^{\Omega(np(n))}$.*

Acknowledgements. The authors thank Sebastian Pokutta, Ronald de Wolf and three anonymous referees for their useful feedback.

References

1. Babai, L., Frankl, P., Simon, J.: Complexity classes in communication complexity theory. In: 27th Annual Symposium on Foundations of Computer Science (FOCS 1986), pp. 337–347. IEEE Computer Society Press, Toronto (1986)
2. Chvátal, V.: On certain polytopes associated with graphs. J. Comb. Theory B 18, 138–154 (1975)
3. Cohen, J.E., Rothblum, U.G.: Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. Linear Algebra Appl. 190, 149–168 (1993)
4. Conforti, M., Cornuéjols, G., Zambelli, G.: Extended formulations in combinatorial optimization. 4OR 8(1), 1–48 (2010)
5. Edmonds, J.: Maximum matching and a polyhedron with 0, 1 vertices. J. Res. Nat. Bur. Stand 69B, 125–130 (1965)
6. Edmonds, J.: Matroids and the greedy algorithm. Math. Program. 1, 127–136 (1971)
7. Fiorini, S., Kaibel, V., Pashkovich, K., Theis, D.O.: Combinatorial bounds on nonnegative rank and extended formulations, arXiv:1111.0444
8. Fiorini, S., Massar, S., Pokutta, S., Tiwary, H.R., de Wolf, R.: Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds. In: Proceedings of the 44th ACM Symposium on Theory of Computing, STOC 2012 (to appear, 2012)
9. Grötschel, M., Lovász, L., Schrijver, A.: Geometric algorithms and combinatorial optimization. In: Algorithms and Combinatorics, 2nd edn., vol. 2. Springer, Berlin (1993)
10. Kaibel, V.: Extended formulations in combinatorial optimization. Optima 85, 2–7 (2011)
11. Kaibel, V., Pashkovich, K., Theis, D.O.: Symmetry Matters for the Sizes of Extended Formulations. In: Eisenbrand, F., Shepherd, F.B. (eds.) IPCO 2010. LNCS, vol. 6080, pp. 135–148. Springer, Heidelberg (2010)
12. Kalyanasundaram, B., Schnitger, G.: The probabilistic communication complexity of set intersection. SIAM J. Discr. Math. 5(4), 545–557 (1992)
13. Kushilevitz, E., Nisan, N.: Communication complexity. Cambridge University Press, Cambridge (1997)
14. Martin, R.K.: Using separation algorithms to generate mixed integer model reformulations. Oper. Res. Lett. 10(3), 119–128 (1991)
15. Razborov, A.A.: On the distributional complexity of disjointness. Theor. Comput. Sci. 106(2), 385–390 (1992)
16. Schrijver, A.: Combinatorial optimization. Polyhedra and efficiency. Algorithms and Combinatorics, vol. 24(A, B). Springer, Berlin (2003)
17. Vazirani, V.V.: Approximation algorithms. Springer (2001)
18. Yannakakis, M.: Expressing combinatorial optimization problems by linear programs. J. Comput. System Sci. 43(3), 441–466 (1991)
19. Ziegler, G.M.: Lectures on Polytopes. Graduate Texts in Mathematics, vol. 152. Springer, Berlin (1995)