comput. complex. (2020) 29:6

https://doi.org/10.1007/s00037-020-00196-6

computational complexity

ON ϵ-SENSITIVE MONOTONE COMPUTATIONS

Pavel Hrubeš

Abstract. We show that strong-enough lower bounds on monotone arithmetic circuits or the nonnegative rank of a matrix imply unconditional lower bounds in arithmetic or Boolean circuit complexity. First, we show that if a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree d has an arithmetic circuit of size s then $(x_1 + \cdots + x_n + 1)^d + \epsilon f$ has a monotone arithmetic circuit of size $O(sd^2 + n \log n)$, for some $\epsilon > 0$. Second, if $f : \{0,1\}^n \to \{0,1\}$ is a Boolean function, we associate with f an explicit exponential-size matrix $M(f)$ such that the Boolean circuit size of f is at least $\Omega(\min_{\epsilon>0}(\mathrm{rk}_+(M(f) - \epsilon J)) - 2n)$, where J is the all-ones matrix and rk_+ denotes the nonnegative rank of a matrix. In fact, the quantity $\min_{\epsilon>0}(\mathrm{rk}_+(M(f) - \epsilon J))$ characterizes how hard is it to distinguish rejecting and accepting inputs of f by means of a linear program. Finally, we introduce a proof system resembling the monotone calculus of Atserias et al. (J Comput Syst Sci 65:626–638, 2002) and show that similar ϵ-sensitive lower bounds on monotone arithmetic circuits imply lower bounds on proof-size in the system.

Keywords. Arithmetic circuit complexity, Extension complexity, Nonnegative rank

Subject classification. 69Q09, 68Q17

1. Introduction

The paper investigates three topics connected by one underlying theme. We show that strong-enough monotone lower bounds imply lower bounds on arithmetic circuit size, Boolean circuit size, and in proof complexity. In contrast to similar earlier results, the unique feature of our "strong-enough" monotone lower bounds is that they

Published online: 25 July 2020 Birkhäuser

are highly discontinuous: in each case, we are required to have a hard polynomial/matrix which is ϵ-close to an easy one.

Arithmetic circuits. In arithmetic circuit complexity, the major open problem is to find an explicit polynomial which requires arithmetic circuits of superpolynomial size. See, e.g., (Bürgisser *et al.* 1997; Shpilka & Yehudayoff 2010; Valiant 1979) for more details. On the other hand, it is comparatively easy to prove such a lower bound for monotone arithmetic circuits—circuits over the reals where we allow nonnegative constants only. For example, the permanent polynomial requires monotone circuits of size $2^{\Omega(n)}$ and the central symmetric polynomial requires monotone depth $\Omega(\log^2 n)$. This can be found in the works (Valiant 1980) and (Shamir & Snir 1979), see also (Hrubeš & Yehudayoff 2011; Nisan 1991; Shpilka & Yehudayoff 2010). This creates the impression that monotone arithmetic lower bounds are easy in all cases. However, we observe that in general, monotone lower bounds are essentially as hard as unrestricted ones. Given a polynomial f of degree d, we show that if f has a small arithmetic circuit then $g_\epsilon := (x_1 + \cdots + x_n + 1)^d + \epsilon f$ has a small monotone arithmetic circuit, for some $\epsilon > 0$. In other words, a monotone lower bound on g_ϵ which works for every $\epsilon > 0$ gives an unconditional lower bound for f. This result is reminiscent of a similar result of Valiant in Boolean circuit complexity concerning the so-called slice functions (Valiant 1986; Wegener 1987). Observe that g_0 has a small monotone arithmetic circuit and hence hardness of f requires that monotone circuit size of g_ϵ displays significant discontinuity as ϵ approaches zero. Our current techniques for obtaining monotone lower bounds seem inapplicable in this situation, and an improvement of these techniques would be desirable.

Boolean circuits and linear programs. In Boolean circuit complexity, we point out that lower bounds on the nonnegative rank of a matrix imply circuit lower bounds. The nonnegative rank of a matrix, rk_+, is a quantity which has found several applications in communication complexity and extension complexity of polytopes. See, for example, the seminal paper of Yannakakis 1991 or (Fiorini *et al.* 2011; Rothvoß 2011). Given a Boolean function $f : \{0,1\}^n \to \{0,1\}$, we associate with f an explicit exponential-size matrix $M(f)$ with positive integer entries. Basically, $M(f)$ records Hamming distances between rejecting and accepting inputs of f. Given $\epsilon > 0$, $M_\epsilon(f)$ is obtained by subtracting ϵ from every entry of $M(f)$. We show that the quantity $\min_{\epsilon>0} \mathrm{rk}_+(M_\epsilon(f))$ is a lower bound on the Boolean circuit size of f. We again note that $M_0(f) = M(f)$ itself has small nonnegative rank and hence, for our estimate to be interesting, there must be a large gap between $\mathrm{rk}_+(M_0(f))$ and $\lim_{\epsilon\to 0_+} \mathrm{rk}_+(M_\epsilon(f))$. It therefore seems that understanding possible discontinuities of the nonnegative rank would give some insight into circuit complexity; we also give some results in this direction.

A similar phenomenon has been noted earlier: in (Hrubeš 2012) and (Goos *et al.* 2018), it was observed that $\mathrm{rk}_+(M_1(f))$ (the case of $\epsilon = 1$) is a lower bound on Boolean formula size. Furthermore, Goos *et al.* (2018) used this strategy to obtain new lower bounds on monotone formula size. Given the connection between nonnegative rank and extension complexity of polytopes, it is not surprising that $\mathrm{rk}_+(M_\epsilon(f))$ has a geometric interpretation: it captures how hard it is to distinguish accepting and rejecting inputs of f by means of a linear program. This intuition is used extensively in our arguments. Among other things, it implies that there exists a (non-explicit) f such that $\min_{\epsilon>0} \mathrm{rk}_+(M_\epsilon(f))$ is exponential. Therefore, at least in principle, rk_+ can yield exponential circuit lower bounds. One can compare this with the result of Razborov (1992) which says that the usual matrix rank cannot give non-trivial circuit lower bounds. In contrast, we assert here that nonnegative rank can—the price being that rk_+ is way less understood (and NP-hard to compute (Vavasis 2008)).

Proof complexity. The major open problem in proof complexity is to obtain superpolynomial lower bounds on proof-size in the Frege proof system (see, e.g., (Krajíček 1995)). Atserias *et al.* (2002) consider the so-called *monotone calculus*, MLK. In its inference rules, MLK resembles the Frege system except that it proves implications $A \to B$ where A, B are monotone Boolean formulas.

Nevertheless, it was shown in (Atserias *et al.* 2002) that this restricted system quasipolynomially simulates the full Frege system. In this paper, we introduce a proof system called *algebraic monotone calculus*, AMC. The system proves inequalities of the form $f \preceq g$, where f and g are monotone polynomials. The intended interpretation of a line $f \preceq g$ is that every monomial which appears in f with a non-zero coefficient appears also in g. The system AMC can be viewed as a weakening of the Boolean monotone calculus MLK, in the same way that an arithmetic circuit can be seen as weakening of a Boolean circuit. We will show that the size of an AMC-proof of $f \preceq g$ can be characterized by the minimum monotone circuit size of $g - \epsilon f$, overall $\epsilon > 0$. This means that monotone arithmetic lower bounds of this form imply AMC lower bounds. It is however not clear how this reflects on the complexity of MLK or Frege proofs: we do not know whether AMC is actually weaker than MLK, or whether they might simulate each other, at least on inputs of a specific form.

2. Overview of main results

2.1. Monotone arithmetic circuits. Unless stated otherwise, *an arithmetic circuit* will always be an arithmetic circuit over $\mathbb{R}$ with binary operations addition and multiplication (see, e.g., (Hrubeš & Yehudayoff 2011; Shpilka & Yehudayoff 2010) for an exact definition). The *size* of a circuit is the number of its gates. A *monotone* arithmetic circuit is one in which all the constants are nonnegative. As the first main result, we prove in Section 4:

THEOREM 2.1. *Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial of degree d which can be computed by an arithmetic circuit of size s. Then, there exists $\epsilon_0 > 0$ such for every $0 < \epsilon < \epsilon_0$, $(x_1 + \cdots + x_n + 1)^d + \epsilon f$ can be computed by a monotone arithmetic circuit of size $O(sd^2 + n \log n)$.*

In other words, in order to prove a lower bound on arithmetic circuit size of f, it is enough to prove a monotone circuit lower bound on $g_\epsilon = (\sum_i x_i + 1)^d + \epsilon f$, for $\epsilon > 0$ sufficiently small. Observe that the "universal polynomial" $U = (\sum_i x_i + 1)^d$ itself has a small monotone circuit (of size $O(n + \log d)$). Hence, the bound of Theorem 2.1 is qualitatively tight for small d: if g_ϵ can be computed by a small monotone circuit for some $\epsilon > 0$ then $f = (g_\epsilon - U)\epsilon^{-1}$ has a small arithmetic circuit as well.

Another way to interpret the result is as follows. It is well-known that every arithmetic circuit can be simulated by an arithmetic circuit with only one subtraction—see (Valiant 1980) or Lemma 4.4. That is, if f has a circuit of size s then f can be written as $f = f_+ - f_-$, where f_+, f_- are monotone polynomials of monotone circuit size $O(s)$. Then, Theorem 2.1 simply asserts that f_- can be chosen as a scalar multiple of a fixed universal polynomial independent of f. Theorem 2.1 is also reminiscent of the so-called slice functions in Boolean complexity. Recall that a Boolean function h is a slice function, if there exists a k such that h accepts on all inputs of Hamming weight $> k$, rejects on inputs of Hamming weight $< k$, and is arbitrary on inputs with weight k. Valiant (1986) has shown that for slice functions, general and monotone Boolean circuits are of essentially the same power. This resembles Theorem 2.1, because the universal polynomial U is constant on inputs with fixed $\sum x_i$.

Let us mention that the classical lower bounds (Shamir & Snir 1979; Valiant 1980) are insufficient to prove monotone lower bounds for polynomials of the form required by Theorem 2.1. These lower bounds take into account only the monomial structure of a polynomial. For example, not only that the permanent is hard but any polynomial which has the same set of monomials with non-zero coefficients is hard for monotone circuits. However, the polynomial $(1 + \sum_i x_i)^d$ contains all monomials of degree d with a non-zero coefficient. There is at least one recent lower bound which goes beyond monomial counting due to Yehudayoff (2019). However, the bound is continuous in the coefficients of the polynomial in focus.

In Section 4, we also note that Theorem 2.1 holds in several other settings—we can choose a different universal polynomial, or consider restricted algebraic models of computation.

2.2. Boolean circuits and nonnegative rank. Let $M \in \mathbb{R}^{n\times m}$ be a nonnegative matrix. The *nonnegative rank of* M, $\mathrm{rk}_+(M)$, is defined as the smallest k so that there exist nonnegative matrices $A \in \mathbb{R}^{n\times k}$, $B \in \mathbb{R}^{k\times m}$ with $M = AB$. In other words, it is the smallest k so that M can be written as a sum of k nonnegative rank-one matrices. If M contains a negative entry, we set $\mathrm{rk}_+(M) := \infty$

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Let $f^{-1}(0)$ be the set of rejecting inputs of f and $f^{-1}(1)$ the set of its accepting inputs. Based on f, we define the matrix $M(f)$ as follows. $M(f)$ is a $|f^{-1}(0)| \times |f^{-1}(1)|$ matrix whose rows are indexed by rejecting inputs and the columns by accepting inputs of f. For every $y \in f^{-1}(0)$, $x \in f^{-1}(1)$,

$$M(f)_{y,x} := h(x,y),$$

where $h(x,y)$ is the Hamming distance of x and y.

Hence, $M(f)$ is an exponential size matrix with nonnegative integer entries from $\{1,\dots,n\}$. Since every accepting input differs from every rejecting input, we emphasize that every entry of $M(f)$ is greater or equal to one. Let J be the matrix of the same dimension as M whose every entry equals 1. This means that for every $\epsilon \le 1$, $M(f) - \epsilon J$ is nonnegative.

Define $L(f)$ as the smallest number of leaves in a de Morgan formula computing f, and $C(f)$ the size of a smallest Boolean circuit computing f. In (Hrubeš 2012), it was observed

Proposition 2.2. (Hrubeš 2012) *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Then, $L(f) \ge \mathrm{rk}_+(M(f) - J)/(2n-1)$.*

This was independently noted by Goos *et al.* (2018). There, a similar strategy was applied to prove lower bounds on monotone formula size. In Section 3, we will present a similar connection with circuit size:

Theorem 2.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Then, $\min_{\epsilon>0} \mathrm{rk}_+(M(f) - \epsilon J) \le O(C(f) + n)$. Moreover, there exists a (non-explicit) f with $\min_{\epsilon>0} \mathrm{rk}_+(M(f) - \epsilon J) \ge 2^{\Omega(n)}$.*

Hence, Theorem 2.3 lower-bounds the circuit size of f in terms of the smallest nonnegative rank of $M(f) - \epsilon J$ with $\epsilon > 0$. Let us give some comments on how the nonnegative rank varies with ϵ. First, one can see (cf. Observation 3.7)

$$\mathrm{rk}_+(M(f)) \leq 2n,$$

and hence the nonnegative rank is small if $\epsilon = 0$. Since $M - \epsilon_1 J = (M - \epsilon_2 J) + (\epsilon_2 - \epsilon_1)J$ and $\mathrm{rk}_+(J) = 1$, we can see that

$$\mathrm{rk}_+(M(f) - \epsilon_1 J) \quad \leq \quad \mathrm{rk}_+(M(f) - \epsilon_2 J) + 1, \quad \text{if } \epsilon_1 \leq \epsilon_2.$$

This means that as $\epsilon > 0$ increases, $\mathrm{rk}_+(M(f) - \epsilon J)$ can decrease at most by one. Moreover, it can be shown (cf. Observation 3.8) that $\lim_{\epsilon \to 0_+} \mathrm{rk}_+(M(f) - \epsilon J)$ exists and differs from $\min_{\epsilon > 0} \mathrm{rk}_+(M(f) - \epsilon J)$ by at most one. Hence, Theorem 2.3 requires $\lim_{\epsilon \to 0_+} \mathrm{rk}_+(M(f) - \epsilon J)$ to be much larger than $\mathrm{rk}_+(M(f))$—the nonnegative rank is heavily discontinuous at $\epsilon = 0$.

Discontinuities of nonnegative rank. In Section 3.2, we give some properties of possible discontinuities of nonnegative rank. For example we, show that if M is positive then

$$\lim_{\epsilon \to 0_+} \mathrm{rk}_+(M - \epsilon J) \leq O(2^{\mathrm{rk}_+(M)})$$

Strictly positive rank rk_{++}. We also introduce the concept of *strictly positive* rank, or simply *strict rank*. This is defined as nonnegative rank, except that one needs to express M in terms of positive rank one matrices. More exactly, let $M \in \mathbb{R}^{n \times m}$ be a positive matrix. The *strict rank of M*, $\mathrm{rk}_{++}(M)$, is defined as the smallest k so that there exist positive matrices $A \in \mathbb{R}^{n \times k}$, $B \in \mathbb{R}^{k \times m}$ with $M = AB$. In Section 3.3 we will outline some properties of rk_{++}. In particular, we will see that

$$\mathrm{rk}_{++}(M) - 1 \leq \min_{\epsilon > 0} \mathrm{rk}_+(M - \epsilon J) \leq \mathrm{rk}_{++}(M) + 1$$

This also entails $\mathrm{rk}_{++}(M(f)) \leq O(C(f) + n)$.

2.2.1. Extended formulations and separation complexity. The lower bound in Theorem 2.3 is in fact a bound on a different quantity, which we call *linear separation complexity*. Informally, it captures how hard it is to distinguish rejecting and accepting inputs by means of linear programs.

Following (Braun *et al.* 2012; Rothvoß 2011; Yannakakis 1991), let us give some definitions from extension complexity of polytopes. A polyhedron is a subset of $\mathbb{R}^n$ defined by a finite set of linear inequalities. A polytope is a bounded polyhedron. An *extended formulation* of a polyhedron P is a linear system $L(x, y)$ in variables x and d new variables y

$$Ax + By \geq b,\ A_0x + B_0y = b_0 \tag{2.4}$$

such that $P = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^d\ L(x, y) \text{ holds}\}$. The *size* of the formulation is the number of inequalities in the system. Extension complexity of P, $\mathrm{xc}(P)$, is defined as the size of a smallest extended formulation of P.

Let $f : \{0, 1\}^n \to \{0, 1\}$ be a Boolean function. A polyhedron $P \subseteq \mathbb{R}^n$ will be called a *separating polyhedron for f* if

$$f^{-1}(1) \subseteq P, \text{ and } f^{-1}(0) \cap P = \emptyset.$$

We define the *linear separation complexity*, or simply *separation complexity*, $\mathrm{sep}(f)$, as the minimum extension complexity of a separating polyhedron for f. In other words, $\mathrm{sep}(f)$ is the smallest r so that there exists a linear system $L(x, y)$ with r inequalities (and any number of equalities) so that

$$f^{-1}(1) = \{x \in \{0, 1\}^n : \exists y \in \mathbb{R}^d\ L(x, y) \text{ holds}\}.$$

In Section 3, we give the following characterization of $\mathrm{sep}(f)$:

THEOREM 2.5. *Let $R := \min_{\epsilon>0} rk_+(M(f) - \epsilon J)$. Then, $R - 2n - 1 \leq sep(f) \leq R$.*

While the phrase "linear separation complexity" was introduced in (Hrubeš 2019), the same concept has appeared earlier. Valiant (1982) has observed that linear separation complexity is, up to a constant factor, a lower bound on the Boolean circuit complexity of

f. This appears again in the seminal paper of Yannakakis (1991). In the context of proof complexity, Pudlák & de Oliveira Oliveira (2017) investigated a host of related complexity measures, including *monotone* separation complexity. Recently, Atserias *et al.* (2019) gave a lower bound on $\mathrm{sep}(f)$ under an additional symmetry assumption.[1]

Observe that the smallest separating polyhedron for f is simply the polytope $P_f := \mathrm{conv}(f^{-1}(1))$, the convex hull of accepting inputs of f. The Yannakakis' paper started a fruitful direction of research into the extension complexity of 0/1-polytopes. These polytopes are well-studied. Rothvoß (2011) has shown that there exists an f such that P_f has extension complexity $2^{\Omega(n)}$. In an ensuing body of research, the same was proved for explicit functions (see, e.g., (Rothvoss 2017) and references within). However, P_f is just one of infinitely many separating polytopes for f and these results do not imply a lower bound on $\mathrm{sep}(f)$. The paper (Hrubeš 2019) has made a modest step in this direction: it was shown that there exists a (non-explicit) f with $\mathrm{sep}(f) \geq 2^{\Omega(n)}$.

Monotone separation complexity $\mathbf{sep}_+(f)$. In Section 3.4, we focus on monotone Boolean functions and give an analogy of Theorem 2.5 for monotone f and monotone separating polyhedra.

2.3. A related proof system. We define a new proof system AMC called *Algebraic Monotone Calculus.*

The lines of AMC are of the form $f \preceq g$ where f, g are *monotone* polynomials. The axioms are

$$f \preceq f \qquad (A1), \qquad\qquad a \preceq b \qquad (A2), \quad \text{if } a, b \in \mathbb{R}, a \geq 0, b > 0.$$

The rules are

$$\frac{f_1 \preceq g,\ \ g \preceq f_2}{f_1 \preceq f_2}\ \ (R1), \qquad \frac{f_1 \preceq g_1,\ \ f_2 \preceq g_2}{f_1 \circ f_2 \preceq g_1 \circ g_2}\ \ (R2), \qquad \circ \in \{+, \times\}$$

An AMC-proof of $f \preceq g$ is a sequence $f_1 \preceq g_1, \ldots f_m \preceq g_m$ such that $f_m = f$, $g_m = g$ and every line in the sequence is either an

[1]Moreover, they measure the complexity of a polytope differently: namely, as the size of the bit-representation of its defining constraints.

axiom (A1) or (A2), or has been obtained from previous lines by means of the rules (R1) or (R2). The *size* of the proof is defined as the smallest $s \geq m$ so that there exists a monotone circuit of size s which simultaneously computes the polynomials $f_1, g_1, \ldots, f_m, g_m$

Note that the axiom (A2) gives $a \preceq b$ and $b \preceq a$ for every $a, b > 0$. This amounts to taking a homomorphism of $\mathbb{R}_+$ into the Boolean semiring. The intended interpretation of lines $f \preceq g$ is as follows. For a polynomial f, let supp(f) denote the set of monomials which have a non-zero coefficient in f. For example, supp$(2xy + \sqrt{3}z) = \{xy, z\}$. We stipulate that supp$(0) = \emptyset$ and supp$(a) = \{\emptyset\}$, if $a \in \mathbb{R} \setminus \{0\}$. Then, $f \preceq g$ can be interpreted as asserting supp$(f) \subseteq$ supp(g). The system AMC is sound and complete in the following sense:

PROPOSITION 2.6. *Let f, g be monotone polynomials. Then, the following are equivalent.*

(i) supp$(f) \subseteq$ supp(g),

(ii) there exists $\epsilon > 0$ such that $g - \epsilon f$ is monotone,

(iii) there exists an AMC-proof of $f \preceq g$.

Our main result concerning AMC is:

THEOREM 2.7. *Assume that $f \preceq g$ has an AMC-proof of size s. Then, there exists $\epsilon > 0$ such that $g - \epsilon f$ has a monotone arithmetic circuit of size $O(s)$. Conversely, assume that $\epsilon > 0$ is such that f, g and $g - \epsilon f$ have monotone circuits of size $\leq s$. Then, $f \preceq g$ has an AMC-proof of size $O(s)$.*

Our motivation for introducing the system AMC is manifold. First, the system arises quite naturally in the context of Theorem 2.1. As we remark in Section 5.1, Theorem 2.1 can be seen as an application of Theorem 2.7. Second, the theorem gives a concrete approach to proving lower bounds on the AMC proof-size. Finally, AMC is related to the monotone calculus MLK from (Atserias *et al.* 2002) and, by extension, to the Frege system (cf. Section 5.2). It is possible that understanding AMC would give some insight into the Frege system.

We prove Theorem 2.7 in Section 5. There, we also explain connections between AMC and other proof systems.

Some notation. As customary, $[m] = \{1, \ldots, m\}$. For a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, $\mathrm{supp}(v) = \{i \in [n] : v_i \neq 0\}$ is the set of nonzero coordinates of v. For a matrix $M \in \mathbb{R}^{n\times m}$, $\mathrm{supp}(M)$ is defined similarly. A vector or a matrix will be called nonnegative/positive if every entry is nonnegative/positive. $J_{n,m}$ will denote the $n \times m$ matrix with every entry equal to one. We set $J_n := J_{n,n}$ and sometimes drop the subscript if the dimension is clear from context. $\mathrm{rk}(M)$ will denote the usual linear algebraic rank of a matrix.

Organization. Theorem 2.1 is proved in Section 4. There, we also present some variants of the theorem. Section 3 contains proofs of Theorems 2.3 and 2.5. Section 3.3 contains some results on the strict rank of a matrix, and Section 3.2 presents facts about discontinuities of nonnegative rank. Section 3.4 deals with monotone polyhedra and monotone Boolean functions. In Section 5, we prove Theorem 2.7 as well as discuss the system AMC in some detail.

3. Linear separation complexity

In this section, we prove Theorems 2.3 and 2.5. Recall the definition of separation complexity from Section 2.2. There, we also mentioned the following results:

Proposition 3.1. *(i) (Valiant 1982; Yannakakis 1991) For every Boolean function f, sep$(f) \leq O(C(f))$.*

(ii) (Hrubeš 2019) There exists $f : \{0,1\}^n \to \{0,1\}$ with sep$(f) \geq 2^{\Omega(n)}$.

For $x, y \in \mathbb{R}^n$, let

$$h(x,y) := \sum_{i=1}^{n} x_i(1-y_i) + \sum_{i=1}^{n}(1-x_i)y_i.$$

If x, y are Boolean vectors, $h(x,y)$ is simply their Hamming distance. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. For a parameter r, let $S_r(f)$ be the polyhedron

$$S_r(f) := \{x \in \mathbb{R}^n : \forall y \in f^{-1}(0),\ h(x,y) \geq r\}. \tag{3.2}$$

Note that

$$S_{r_1}(f) \supseteq S_{r_2}(f), \quad \text{if } r_1 \leq r_2.$$

Let $r_0 := \min_{x \in f^{-1}(1), y \in f^{-1}(0)} h(x,y)$. Then, $r_0 \geq 1$. The key property of S_r is the following:

LEMMA 3.3. *For every $0 < r \leq r_0$, $S_r(f)$ is a separating polyhedron for f. Conversely, assume that P is a separating polyhedron for f. Then, there exists $0 < \epsilon \leq r_0$ such that $P \cap [0,1]^n \subseteq S_\epsilon$.*

PROOF. If $0 < r \leq r_0$, S_r is a separating polyhedron for f: S_r contains all accepting inputs x of f since for every rejecting input y of f, $h(x,y) \geq r_0 \geq r$. $S_r(f)$ contains no rejecting input y of f since $h(y,y) = 0 < r$.

For the second part, assume that P is a separating polyhedron for f. Fix $y \in f^{-1}(0)$. We claim that $h(x,y) > 0$ for every $x \in P \cap [0,1]^n$. This is because $h(x,y) = 0$ implies $x = y$ on $x \in [0,1]^n$. Since $P \cap [0,1]^n$ is compact, there exists $\epsilon_y > 0$ such that $h(x,y) > \epsilon_y$ for every $x \in P$. Setting $\epsilon := \min_{y \in f^{-1}(0)} \epsilon_y$ shows that $P \cap [0,1]^n \subseteq S_\epsilon$. □

Following (Braun *et al.* 2012; Yannakakis 1991), we now define slack matrices. Let V be a sequence $v_1, \dots, v_{m_1}$ of points in $\mathbb{R}^n$ and $L(x)$ a system $\ell_1(x) \geq b_1, \dots, \ell_{m_2}(x) \geq b_{m_2}$ of inequalities in $\mathbb{R}^n$. The slack matrix with respect to V and $L(x)$ is the $m_2 \times m_1$ matrix T such that

$$T_{i,j} = \ell_i(v_j) - b_i.$$

Let $P_0 := \mathrm{conv}(V)$ be the convex hull of V and $P_1 := \{x \in \mathbb{R}^n :\ L(x) \text{ holds}\}$. If $P_0 \subseteq P_1$ then T is nonnegative. In (Braun *et al.* 2012), we can find:

LEMMA 3.4 (Braun *et al.* 2012). *Let $P_0 \subseteq P_1$ and T be as above. Define $xc(P_0, P_1)$ as the minimum $xc(P)$ over all polyhedra with $P_0 \subseteq P \subseteq P_1$. Then*

$$rk_+ T - 1 \leq xc(P_0, P_1) \leq rk_+ T.$$

THEOREM 2.5 (restated). *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Let $R := \min_{\epsilon>0} rk_+(M(f) - \epsilon J)$. Then, $R - 2n - 1 \leq sep(f) \leq R$.*

PROOF. We first observe that $M(f) - \epsilon J$ is a slack matrix defined by the points $\mathrm{conv}(f^{-1}(1))$ and the inequalities in (3.2) defining S_ϵ. Hence, if $M(f) - \epsilon J$ is nonnegative then S_ϵ is a separating polyhedron by Lemma 3.3. By the previous lemma, $\mathrm{sep}(f) \leq \mathrm{xc}(S_\epsilon) \leq \mathrm{rk}_+(M - \epsilon J)$ and so $\mathrm{sep}(f) \leq R$. To prove the opposite inequality, assume that P is a separating polyhedron for f with $\mathrm{xc}(P) = t$. Let $P' := P \cap [0,1]^n$. Then, $xc(P') \leq t + 2n$. By Lemma 3.3 there exists $\epsilon > 0$ such that $P' \subseteq S_\epsilon$. Since $f^{-1}(1) \subseteq P'$, Lemma 3.4 gives $\mathrm{rk}_+(M(f) - \epsilon J) \leq \mathrm{xc}(P') + 1$. Therefore, $R \leq t + 2n + 1$ and so $t \geq R - 2n - 1$. □

3.1. Consequences and comments. As a consequence of Theorem 2.5, Proposition 3.1 implies Theorem 2.3:

COROLLARY (Theorem 2.3 restated). *Let f be an n-variate Boolean function. Then, $\min_{\epsilon>0} rk_+(M(f) - \epsilon J) \leq O(C(f) + n)$. Furthermore, there exists an f such that $\min_{\epsilon>0} rk_+(M(f) - \epsilon J) \geq 2^{\Omega(n)}$.*

Moreover, we also obtain that nonnegative rank is heavily discontinuous:

COROLLARY 3.5. *For every n there exists $n \times n$ matrix M with positive integer entries such that $\mathrm{rk}_+(M) = O(\log(n))$ but for every $\epsilon > 0$, $rk_+(M - \epsilon J) \geq n^{\Omega(1)}$.*

Another implication is that $\mathrm{sep}(f)$ and $\mathrm{sep}(\neg f)$ are almost the same, a fact not entirely apparent from the definition of $\mathrm{sep}(f)$:

COROLLARY 3.6. $|sep(f) - sep(\neg f)| \leq 2n + 1$.

PROOF. Note that $M(\neg f)$ is the transposition of $M(f)$ and apply Theorem 2.5. □

Let H_n be the $2^n \times 2^n$ matrix whose rows and columns are indexed by Boolean strings of length n and $(H_n)_{x,y}$ is the Hamming distance of x and y. Note that $M(f)$ is a submatrix of H_n for every n-variate Boolean function f.

Observation 3.7. *(i)* $rk(H_n + aJ_{2^n}) = n+1$ *whenever* $a \geq 0$.

(ii) $n + \Omega(\log n) \leq rk_+(H_n) \leq 2n$.

Proof. Let V be the $2^n \times n$ matrix whose rows consist of all $0,1$-vectors of length n and let $\jmath$ the all-ones column vector of length 2^n. Then, $H_n + aJ_{2^n} = (V, J_{2^n,n} - V, \jmath) \cdot (J_{2^n,n} - V, V, a\jmath)^t$. The $2n+1$ columns of $(V, J_{2^n,n} - V, \jmath)$ lie in the linear span of the columns of V and $\jmath$ which implies that $\mathrm{rk}(aJ_{2^n} + H_n) \leq n+1$. Setting $a := 0$, this also show $\mathrm{rk}_+(H_n) \leq 2n$. To see that $\mathrm{rk}(aJ_{2^n} + H_n) \geq n+1$, consider the $(n+1) \times (n+1)$ submatrix of H_n with rows and columns indexed by strings of hamming weight 0 or 1. The matrix has full rank. The lower bound $\mathrm{rk}_+(H_n) \geq n + \Omega(\log n)$ follows by noting that H_n has zero diagonal, is non-zero everywhere else, and applying the result (deCaen *et al.* 1981) (see also (Beasley & Laffey 2009)). □

The following example shows that some dependency on n in Theorem 2.5 is inevitable.

Example. Let $f(x_1, \ldots, x_n)$ be the Boolean function such that $f(x_1, \ldots, x_n) = x_1$. Then, $\mathrm{sep}(f) = 0$, because accepting and rejecting inputs can be distinguished by the equation $x_1 = 1$ (recall that equations do not contribute to separation complexity). On the other hand, $M(f)$ is the matrix $J_{2^{n-1}} + H_{n-1}$. Hence, $\mathrm{rk}_+(M(f) - \epsilon J_{2^{n-1}}) \geq n$ for every $\epsilon > 0$.

A different example would come from $f(x_1, \ldots, x_n) := \bigvee_{i<j} x_i \wedge x_j$, which is more convincing in that f depends on all its variables.

The reason for the gap between $R-2n-1$ and R in Theorem 2.5 is the following. R is an upper bound on the extension complexity of a separating polyhedron P, whereas $R-1$ is a lower bound on extension complexity of P *intersected with* $[0,1]^n$. Indeed, the gap would disappear had we defined separation complexity differently, by explicitly requiring the separating polytope P to satisfy $P \subseteq [0,1]^n$. This would be at the cost of replacing $M(f)$ and J in Theorem 2.5 by more complicated matrices.

3.2. On continuity of nonnegative rank. We now present some facts about the behavior of $\mathrm{rk}_+(M+\epsilon N)$. We start with a general fact:

OBSERVATION 3.8. *Let M, N be matrices of the same dimension. For $n \in \mathbb{N}$, let $T_n := \{x \in \mathbb{R} : rk_+(M+xN) \le n\}$. Then, T_n is a finite union of closed intervals. In particular, if M is nonnegative and $supp(N) \subseteq supp(M)$, then $\lim_{\epsilon\to 0_+} rk_+(M+\epsilon N)$ exists and $rk_+(M) \le \lim_{\epsilon\to 0_+} rk_+(M+\epsilon N)$.*

PROOF. That T_n is a closed set follows from the definition of rk_+. That it is a finite union of intervals follows from Tarski–Seidenberg theorem (see, e.g., (Basu *et al.* 2006)). □

Recall that, by Corollary 3.5, $\mathrm{rk}_+(M-\epsilon J)$ can be much larger than $\mathrm{rk}_+(M)$, even if J is a rank-one positive matrix. In the rest of this section, we want to estimate how large the gap between $\mathrm{rk}_+(M)$ and $\lim_{\epsilon\to 0_+} \mathrm{rk}_+(M+\epsilon N)$ can be. The following lemma is interesting in its own right. It asserts that we can add to M a positive rank one matrix V so that $\mathrm{rk}_+(M+V)$ and $\mathrm{rk}(M+V)$ are virtually the same. Geometrically, this can be interpreted as saying that every finite set of points in $\mathbb{R}^n$ is contained inside a polytope with $n+1$ facets.

LEMMA 3.9. *Let $V, N \in \mathbb{R}^{m\times n}$ where V is a positive rank-one matrix. Then, for every $t \ge 0$ large enough, $rk_+(N+tV) \le rk(N)+1$.*

PROOF. Let $r := \mathrm{rk}(N)$. Fix t_0 such that $N_1 := N + t_0 V$ is nonnegative. We can write $V = v_1 v_2^t$ where $v_1 \in \mathbb{R}^m$, $v_2 \in \mathbb{R}^n$ are column vectors. Since $\mathrm{rk}(N_1) \le r+1$, we can write $N_1 = AB$ with $A \in \mathbb{R}^{n\times(r+1)}$, $B \in \mathbb{R}^{(r+1)\times n}$. Furthermore, we can choose A, B in such a way that A is nonnegative and $Aw = v_1$, where w is a positive vector. To see that, pick a basis $u_1, \dots, u_r$ of the columns of N. Let the columns of A consist of the vectors $v_1 - \delta u_1, \dots, v_1 - \delta u_r$, $v_1 + \delta \sum_{i=1}^r u_i$, where $\delta > 0$ is small enough. Then, $v_1, u_1, \dots, u_r$ lie in the span of columns of A and $A \cdot (1, \dots, 1)^t = (r+1)v_1$.

To proceed, let t_1 be such that $B + t_1 w v_2^t$ is nonnegative. Then, $N_2 := A(B + t_1 w v_2^t)$ has nonnegative rank at most $r + 1$. Furthermore,

$$N_2 = AB + t_1 A w v_2^t = N_1 + t_1 v_1 v_2^t = N + t_0 V + t_1 V,$$

and hence $\mathrm{rk}_+(N + tV) \leq r + 1$ for every t large enough. □

THEOREM 3.10. *Let M, N be matrices of the same dimension with M nonnegative and supp$(N) \subseteq$ supp(M). Then,*

(i) $\lim_{\epsilon \to 0_+} \mathrm{rk}_+(M + \epsilon N) \leq 2^{\mathrm{rk}_+(M)}(\mathrm{rk}(N) + 2)$,

(ii) $\lim_{\epsilon \to 0_+} \mathrm{rk}_+(M + \epsilon N) \leq 2^{\mathrm{rk}_+(M)+2} + \mathrm{rk}(N)$, *assuming M is positive.*

PROOF. We first note that (i) implies (ii). If M is positive, write $M + \epsilon N = (M - \epsilon^{1/2} V) + \epsilon^{1/2}(V + \epsilon^{1/2} N)$, where V is a positive rank-one matrix. Using the previous lemma and part (i), we obtain $\mathrm{rk}_+(V + \epsilon^{1/2} N) \leq \mathrm{rk}(N) + 1$ and $\mathrm{rk}_+(M - \epsilon^{1/2} V) \leq 3 \cdot 2^{\mathrm{rk}_+(M)}$, if ϵ is small enough. This gives $\lim_{\epsilon \to 0_+} \mathrm{rk}_+(M + \epsilon N) \leq 3 \cdot 2^{\mathrm{rk}_+(M)} + \mathrm{rk}(N) + 1 \leq 4 \cdot 2^{\mathrm{rk}_+(M)} + \mathrm{rk}(N)$.

Part (ii) is proved by induction on $\mathrm{rk}_+(M)$. Fix $k \in \mathbb{N}$. For $r \in \mathbb{N}$, let $g(r)$ denote the smallest s so that

$$\lim_{\epsilon \to 0_+} \mathrm{rk}_+(M + \epsilon N) \leq s,$$

for every M, N with $\mathrm{rk}_+(M) \leq r$, $\mathrm{rk}(N) \leq k$, and $\mathrm{supp}(N) \subseteq \mathrm{supp}(M)$. Lemma 3.9 implies

$$g(1) \leq k + 1. \tag{3.11}$$

We now want to bound $g(r+1)$ in terms of $g(r)$. Let $M, N \in \mathbb{R}^{n \times m}$ be such that $\mathrm{supp}(N) \subseteq \mathrm{supp}(M)$, $\mathrm{rk}(N) \leq k$, $\mathrm{rk}_+(M) = r + 1$. We have $M = V + M'$, where V is a nonnegative rank-one matrix and $\mathrm{rk}_+(M') = r$. We have $\mathrm{supp}(V) = S \times T$, where $S \subseteq [n]$, $T \subseteq [m]$. For a matrix $Q \in \mathbb{R}^{n \times m}$, and $S' \subseteq [n]$, $T' \subseteq [m]$ let

$Q(S', T') \in \mathbb{R}^{n\times m}$ be obtained by replacing every row not in S' and every column not in T' by zero in Q. Furthermore, let

$$Q(0) := Q(S,T),\ Q(1) := Q(S,[m]\setminus T),\ Q(2) := Q([n]\setminus S,[m]).$$

This guarantees that $Q(0), Q(1), Q(2)$ have disjoint support and $Q = Q_0 + Q_1 + Q_2$. Hence, $(M + \epsilon N) = \sum_{i=0}^{2}(M(i) + \epsilon N(i))$ and

$$\mathrm{rk}_+(M+\epsilon N) \le \sum_{i=0}^{2} \mathrm{rk}_+(M(i)+\epsilon N(i)).$$

By definition, $V = V(0)$, and $M(0) + \epsilon N(0)$ can be written as $M'(0) + (V + \epsilon N(0))$ where $\mathrm{supp}(N(0) \subseteq \mathrm{supp}(V)$. Using Lemma 3.9, we have

$$\mathrm{rk}_+(M(0)+\epsilon N(0)) \le r + k + 1,$$

if ϵ is small enough. Furthermore, $V(1), V(2)$ are zero matrices, and we have $M(i) + \epsilon N(i) = M'(i) + \epsilon N(i)$ for $i \in \{1, 2\}$. Since, $\mathrm{rk}_+(M'(i) \le r$ and $\mathrm{rk}(N_i) \le k$, we have

$$\mathrm{rk}_+(M(i)+\epsilon N(i)) \le g(r), \ \text{ for } i \in \{1,2\}.$$

We have therefore established that

$$g(r+1) \le r + k + 1 + 2g(r).$$

Every such recursion has a solution $g(r) \le a2^r - r - k - 2$. To meet the initial condition (3.11), it is enough to set $a := k + 2$. Hence we have proved $g(r) \le (k+2)2^r - r - k - 2 \le (k+2)2^r$, as required. □

3.3. Strict rank rk_{++}. Recall the definition of $\mathrm{rk}_{++}(M)$ in Section 2.2. We now explain that Theorem 2.5 can be stated in terms of the strict rank of $M(f)$. We also give some bounds on rk_{++} in terms of rk_+.

The following lemma gives some equivalent definitions of the strict rank:

Lemma 3.12. *Let M be an $m \times n$ positive matrix. Then, the following are equivalent:*

(i) $rk_{++}(M) \leq r$

(ii) *M can be written as $M = V_1 + \cdots + V_r$ where $V_1, \ldots, V_r$ are nonnegative rank-one matrices and V_1 is positive.*

(iii) *M can be written as $M = AB$, where $A \in \mathbb{R}^{m\times r}$ is nonnegative and $B \in \mathbb{R}^{r\times n}$ is positive.*

Proof. (i) implies (ii) and (iii) by definition of $\mathrm{rk}_{++}(M)$.

(ii) $\Longrightarrow$ (iii). If M is as in (ii) then it can be written as $M = AB$ where $A \in \mathbb{R}^{m\times r}$, $B \in \mathbb{R}^{r\times n}$ are nonnegative and, moreover, the first column of A and the first row of B are positive. Let $E_\delta := I_r + \delta v_1^t v_2$ where $v_1 = (0, 1, \ldots, 1)$, $v_2 = (1, 0, \ldots, 0)$. If $\delta > 0$ then $E_\delta B$ is positive. If δ is small enough then AE_δ^{-1} is nonnegative. Hence, $(AE_\delta^{-1})(E_\delta B)$ is as required in (iii).

(iii) $\Longrightarrow$ (i). Let A, B be as in the assumption of (iii). Let $D_\delta := I_r + \delta J_r$. If δ is small enough then $D_\delta^{-1}B$ is positive. Since M is positive, every row of A is non-zero and so AD_δ is positive. Hence $M = (AD_\delta)(D_\delta^{-1}B)$ is a positive factorization of M. □

For example, part (iii) of the proposition implies $\mathrm{rk}_{++}(M) \leq \min(m, n)$. Furthermore:

Proposition 3.13. *Let M, V be positive matrices of the same dimension such that V has rank one. Then,*

$$\lim_{\epsilon \to 0_+} rk_+(M - \epsilon V) - 1 \leq rk_{++}(M) \leq \min_{\epsilon > 0} rk_+(M - \epsilon V) + 1.$$

Proof. Let $r := \mathrm{rk}_{++}(M)$. By definition, $M = V_1 + \cdots + V_r$, where V_i are positive rank one matrices. By Lemma 3.9, $\mathrm{rk}_+(V_1 - \epsilon V) \leq 2$, for every $\epsilon > 0$ sufficiently small. This gives $\lim_{\epsilon\to 0_+}(M - \epsilon V) \leq r + 1$. Furthermore, $M = (M - \epsilon V) + \epsilon V$. Hence if $\mathrm{rk}_+(M - \epsilon V) = s$, Lemma 3.12 part (ii) gives $r \leq s + 1$. □

Using Theorem 2.5, Proposition 3.13 implies:

COROLLARY 3.14. *Let* $f : \{0,1\}^n \to \{0,1\}$ *be a Boolean function. Then,* $rk_{++}(M(f)) - 2(n+1) \le sep(f) \le rk_{++}(M(f)) + 1$.

We shall note below that the bound in Corollary 3.14 can be slightly improved. As for the connection between $\mathrm{rk}_+(M)$ and $\mathrm{rk}_{++}(M)$, we observe:

PROPOSITION 3.15. *(i)* $rk_{++}(M) \le O(2^{rk_+(M)})$*, if* M *is positive.*

(ii) For every n*, there exists* $n \times n$ *matrix* M *with positive integer entries such that* $rk_+(M) = O(\log(n))$ *but* $rk_{++}(M) \ge n^{\Omega(1)}$.

PROOF. The first part follows from Theorem 3.10 and Proposition 3.13. The second part follows from Proposition 3.13 and Corollary 3.5. □

Like nonnegative rank, strict rank can be interpreted geometrically. Let $P_0 := \mathrm{conv}(v_1, \ldots, v_{m_1})$ be a polytope in $\mathbb{R}^n$ and $P_1 \subseteq \mathbb{R}^n$ be a set defined by *strict* inequalities $\ell_1(x) > b_1, \ldots, \ell_{m_2}(x) > b_{m_2}$. Define the slack matrix $T \in \mathbb{R}^{m_2 \times m_1}$ as above, i.e., $S_{i,j} = \ell_i(v_j) - b_i$. If $P_0 \subseteq P_1$, T is a positive matrix. Without a proof, we remark that Lemma 3.4 can be stated in terms of $\mathrm{rk}_{++}(T)$:

(i). Assume $P_0 \subseteq P_1$. Define $\mathrm{xc}(P_0, P_1)$ as the minimum $\mathrm{xc}(P)$ overall polyhedra $P_0 \subseteq P \subseteq P_1$. Then, $\mathrm{rk}_{++}T - 1 \le \mathrm{xc}(P_0, P_1) \le \mathrm{rk}_{++}T$.

(ii). The proof of Theorem 2.5 could be carried out directly using $\mathrm{rk}_{++}(M(f))$ giving a slight improvement of Corollary 3.14:

$$\mathrm{rk}_{++}(M(f)) - 2n - 1 \le \mathrm{sep}(f) \le \mathrm{rk}_{++}(M(f)).$$

3.4. Monotone polyhedra. We will now focus on monotone Boolean functions. We present an analogy of Theorem 2.5 for monotone functions and monotone polyhedra. We define *monotone separation complexity of* f which captures how hard it is to distinguish accepting inputs of f from rejecting inputs by means of a linear program in which the variables of f have nonnegative coefficients. This is interesting for at least two reasons. First, in this restricted version there is a greater hope to prove unconditional lower bounds for an explicit function f. Second, such lower bounds may have applications in proof complexity, see (Pudlák & de Oliveira Oliveira 2017).

For $x, y \in \mathbb{R}^n$, we write $x \leq y$ if $y - x$ is nonnegative. A polyhedron P will be called *monotone* if for every $x \leq y \in \mathbb{R}^n$, $x \in P$ implies $y \in P$. One can see that a monotone P can be defined as $\{x \in \mathbb{R}^n : Ax \geq b\}$ where A is nonnegative. For a polyhedron $P \subseteq \mathbb{R}^n$, let

$$P^* := \{x \in \mathbb{R}^n : \exists y \in P,\, x \geq y\}$$

be the monotone closure of P. Recall that a Boolean function f is monotone if for every $x \leq y \in \{0,1\}^n$, $f(x) = 1$ implies $f(y) = 1$. Given a monotone Boolean function f, we define its *monotone separation complexity*, $\mathrm{sep}_+(f)$, as the smallest r so that there exists a polyhedron P with $\mathrm{xc}(P) = r$ such that P^* is a separating polyhedron for f. In other words, there exists a polyhedron $Q \subseteq \mathbb{R}^{n+d}$ which can be defined using r inequalities (and any number of equalities) such that

$$f^{-1}(1) = \{x \in \{0,1\}^n :\, \exists y \in \mathbb{R}^n \exists z \in \mathbb{R}^d, x \geq y, (y, z) \in Q\}.$$

We do not include the inequalities $x \geq y$ as contributing to the complexity of the system. An equally reasonable definition would be to define $\mathrm{sep}_+(f)$ as the smallest extension complexity of a monotone separating polyhedron for f. But note that $\mathrm{xc}(P^*) \leq \mathrm{xc}(P) + n$ and hence the two alternatives are related:

(*i*). If P is a monotone separating polyhedron for f then $\mathrm{sep}_+(f) \leq \mathrm{xc}(P)$,

(*ii*). There exists a monotone separating polyhedron P for f with $\mathrm{xc}(P) \leq \mathrm{sep}_+(f) + n$.

For $x, y \in \mathbb{R}^n$, let

$$h_+(x, y) := \sum_{i=1}^{n} x_i(1 - y_i).$$

If x, y are Boolean vectors, $h_+(x, y)$ equals the number of coordinates i such that $x_i = 1, y_i = 0$. Define $M_+(f)$ as the $|f^{-1}(0)| \times |f^{-1}(1)|$ matrix whose rows are indexed by rejecting inputs and columns by accepting inputs of f and, given $y \in f^{-1}(0), x \in f^{-1}(1)$,

$$M_+(f)_{y,x} := h_+(x, y).$$

We note that $\mathrm{rk}_+(M_+(f)) \leq n$. An analogy of Theorem 2.5 is the following:

THEOREM 3.16. *Let $f : \{0,1\}^n \to \{0,1\}$ be a monotone Boolean function. Then,* $\min_{\epsilon>0} rk_+(M_+(f) - \epsilon J) - 2n - 1 \leq sep_+(f) \leq \min_{\epsilon>0} rk_+(M_+(f) - \epsilon J)$.

PROOF. For a parameter r, let $S_r(f)$ be the polyhedron

$$S_r := \{x \in \mathbb{R}^n : \forall y \in f^{-1}(0),\, h_+(x, y) \geq r\}.$$

Let r_0 be the minimum $h_+(x, y)$ over $x \in f^{-1}(1)$ and $y \in f^{-1}(0)$. Then, $r_0 \geq 1$.

CLAIM. *For every $0 < r \leq r_0$, S_r is a monotone separating polyhedron for f. Conversely, assume P is a polyhedron such that P^* is separating for f. Then, $P \cap [0,1]^n \subseteq S_\epsilon$ for some $\epsilon > 0$.*

PROOF OF THE CLAIM. S_r is monotone, because it is defined by inequalities with nonnegative coefficients. If $r \leq r_0$, $S_r \supseteq f^{-1}(1)$ by the definition of r_0. If $r > 0$, S_r contains no rejecting input y, because $h_+(y, y) = 0$. For the second part, assume that P^* is a separating polyhedron for f. Fix $y \in f^{-1}(0)$. Then, for every $x \in P \cap [0,1]^n$, $h_+(x, y) > 0$. For otherwise, assume that $h_+(x, y) = 0$. Since $x \in [0,1]^n$, we then have $x \leq y$ and so $y \in P^*$—contradicting the assumption that P^* is separating. Hence there exists $\epsilon_y > 0$ such that for every $x \in P \cap [0,1]^n$, $h_+(x, y) \geq \epsilon_y$. Setting $\epsilon := \min_y \epsilon_y$ gives $P \cap [0,1]^n \subseteq S_\epsilon$. □

The rest of the proof proceeds using the Claim in the same manner as the proof of Theorem 2.5. □

3.4.1. Gap between rk_+ and rk_{++} implies a lower bound on sep_+. As in Corollary 3.14, Theorem 3.16 could be stated in terms of $\mathrm{rk}_{++}(M_+(f))$. Since $\mathrm{rk}_+(M_+(f))$ is small, this shows that any lower bound on $\mathrm{sep}_+(f)$ implies a gap between $\mathrm{rk}_{++}(M(f))$ and $\mathrm{rk}_+(M(f))$. We now note that the converse is also true: any separation between rk_{++} and rk_+ yields a lower bound on $\mathrm{sep}_+(f)$ for some f.

Let $N \in \mathbb{R}^{n\times m}$ be a positive matrix of nonnegative rank r, together with its nonnegative factorization $N = A \cdot B$, where $A \in \mathbb{R}^{n\times r}$, $B \in \mathbb{R}^{r\times n}$. With N, we associate the following monotone Boolean function $f_N : \{0,1\}^r \to \{0,1\}$. Given $x \in \{0,1\}^r$, $f_N(x) = 1$ iff there exists a column u of B with $\mathrm{supp}(u) \subseteq \mathrm{supp}(x)$. In other words f_N accepts x if there is a column u of B and $a > 0$ such that $ax \geq u$.

Observation 3.17. *$\mathit{sep}_+(f_N) \geq \mathit{rk}_{++}(N) - (4r+2)$.*

Proof. Let A_0 be the $0,1$-matrix with $\mathrm{supp}(A_0) = \mathrm{supp}(A)$ and similarly for B_0 and B. Let $N_0 := A_0 \cdot B_0$. There exists $\delta > 0$ such that $\mathrm{rk}_+(N - \delta N_0) \leq 2r$. Since $N = (N - \delta N_0) + \delta N_0$, Lemma 3.12 part (ii) implies

$$\mathrm{rk}_{++}(N) \leq 2r + \mathrm{rk}_{++}(N_0).$$

Let N_0' be the matrix obtained by removing from N_0 identical rows and columns. From definition of f_N, we obtain that N_0' is a submatrix of $M_+(f_N)$. Theorem 3.16 and Proposition 3.13 give that $\mathrm{sep}_+(f_N) \geq \mathrm{rk}_{++}(M_+(f_N)) - 2(r+1)$. This gives $\mathrm{sep}_+(f_N) \geq \mathrm{rk}_{++}(N_0') - 2(r+1) \geq \mathrm{rk}_{++}(N) - 4r - 2$. □

4. Monotone arithmetic circuits

We now want to prove Theorem 2.1. A polynomial f will be called *homogeneous*, if every monomial in f with non-zero coefficient has the same degree. If f has degree d, we can always write $f = \sum_{k=0}^{d} f^{(k)}$ where $f^{(k)}$ is homogeneous of degree k. Let C be an arithmetic circuit. Given a gate u in C, let $\hat{u}$ be the polynomial computed by u and $\deg(u) := \deg(\hat{u})$ be its degree. C will be called *homogeneous*, if for every sum gate $u_1 + u_2$ in C, we have $\deg(u_1) = \deg(u_2)$.

Theorem 2.1 relies on the following main lemma:

LEMMA 4.1. *Let C be a monotone homogeneous circuit of size s in variables $x_1, \ldots, x_n$. Then, to every gate u in C we can assign $R_u > 0$ such that for all gates*

$$R_u \cdot (\sum_{i \in [n]} x_i)^{\deg(u)} - \hat{u}$$

can be simultaneously computed by a monotone circuit of size $O(s + n \log n)$.

PROOF. Let $L := x_1 + \cdots + x_n$. By induction on the depth of u, we construct R_u as well as the circuit computing $h_u := R_u \cdot L^{\deg(u)} - \hat{u}$.

If u is an input gate, we let $R_u = u$, if u is a constant, and $R_u = 1$, if u is a variable x_j. In the former case, $h_u = 0$, and the latter, $h_u = L - x_j = \sum_{i \in [n] \setminus \{j\}} x_i$.

If $u = u_1 + u_2$ is a sum gate, let $R_u := R_{u_1} + R_{u_2}$. Since $\deg(u) = \deg(u_1) = \deg(u_2)$, this guarantees

$$h_u = h_{u_1} + h_{u_2}. \tag{4.2}$$

If $u = u_1 \times u_2$ is a product gate, let $R_u := R_{u_1} R_{u_2}$. We have $\deg(u) = \deg(u_1) + \deg(u_2)$, and hence

$$\begin{aligned} h_u &= R_{u_1} R_{u_2} L^{\deg(u)} - \hat{u}_1 \hat{u}_2 \\ &= (R_{u_1} L^{\deg(u_1)} - \hat{u}_1) R_{u_2} L^{\deg(u_2)} + \hat{u}_1 (R_{u_2} L^{\deg(u_2)} - \hat{u}_2) \\ &= h_{u_1} R_{u_2} L^{\deg(u_2)} + \hat{u}_1 h_{u_2} \end{aligned} \tag{4.3}$$

To construct the circuit, we first note that:

(*i*). $L - x_1, \ldots, L - x_n$ can be simultaneously computed by a monotone circuit of size $O(n \log n)$,

(*ii*). all powers L^k such that k is the degree of some gate in C can be simultaneously computed by a circuit of size $O(s + n)$.

The circuit in (*i*) is easily constructed recursively (doubling n at each step). For part (*ii*), remove from C all its sum gates, by replacing each sum gate by one of its inputs. Next, replace each constant in the circuit by 1 and each variable by L.

Part (*i*) means that all h_u's corresponding to input gates can be simultaneously computed by a monotone circuit of size $O(n \log n)$. Equations (4.2) and (4.3) imply that, for a non-input gate u, we can compute h_u using h_{u_1}, h_{u_2}, polynomials from (*i*) or (*ii*), plus a constant number of extra gates. This gives overall complexity $O(s + n \log n)$. □

The following lemma is quite standard and we omit the proof:

Lemma 4.4. *Assume that f has degree d and an arithmetic circuit of size s. Then, f can be written as $f = f_+ - f_-$ where f_+, f_- can both be computed by a monotone arithmetic circuit of size $O(s)$. Furthermore, the homogeneous parts $f_+^{(0)}, \ldots, f_+^{(d)}$, $f_-^{(0)}, \ldots, f_-^{(d)}$ can be simultaneously computed by a homogeneous monotone circuit of size $O(sd^2)$.*

Theorem 2.1 (restated). *Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial of degree d which can be computed by a circuit of size s. Then, there exists $\epsilon_0 > 0$ such for every $0 < \epsilon < \epsilon_0$, the polynomial $(1 + \sum_{i=0}^n x_i)^d + \epsilon f$ has a monotone circuit of size $O(sd^2 + n \log n)$.*

Proof. Write $f = f_+ - f_-$ as in the previous lemma. Then, also $f = f'_+ - f'_-$, where $f'_+ = \sum_{k=0}^d f_+^{(k)}$ and $f'_- = \sum_{k=0}^d f_-^{(k)}$. Setting $L := \sum_{i \in [n]} x_i$, Lemma 4.1 gives a monotone circuit simultaneously computing $R_k L^k - f_-^{(k)}$, $0 \leq k \leq d$, of size $O(sd^2 + n \log n)$. Let $R := \sum_{k=0}^d R_k$. Then, $R(L+1)^d - \sum_{k=0}^d R_k L^k$ has a monotone circuit of size $O(n + d)$. Hence we obtain a monotone circuit for

$$(R(L+1)^d - \sum_{k=0}^{d} R_k L^k) + (\sum_{k=0}^{d} R_k L^k - f_-^{(k)}) = R(L+1)^d - f'_-$$

of size $O(sd^2 + n \log n)$. This gives the required circuit for $R(L+1)^d + f = f'_+ + (R(L+1)^d - f'_-)$. The same holds for every $R' \geq R$. To conclude the theorem, it is enough to set $\epsilon_0 := (R)^{-1}$. □

4.1. Modifications of Theorem 2.1. The theorem can be reproduced in many shapes and forms, depending on the choice of the "universal polynomial" and the computational model one has in mind. The polynomial $(\sum_{i\in[n]} x_i+1)^d$ could be replaced by several other polynomials U—the minimum requirements being that U contains all monomials of degree $\leq d$ and that it has a small monotone circuit. In Proposition 4.8, we give an example of such an alternative choice. Moreover, the same argument applies to restricted models such as multilinear circuits or bounded-depth circuits, which we also discuss below.

Let

$$H_n^d = \sum_{0\leq k_1,\dots,k_n\leq d\colon \sum k_i\leq d} x_1^{k_1}\cdots x_n^{k_n},$$

be the complete symmetric polynomial of degree d in variables $x_1,\dots,x_n$. H_n^d contains all monomials of degree at most d in n variables with coefficient 1. It has the same set of monomials as $(\sum_{i\in[n]} x_i + 1)^d$, but H_n^d has zero-one coefficients. We note that it can recursively defined by

$$H_0^d = 1,\ H_n^d = \sum_{j=0}^{d} x_n^j \cdot H_{n-1}^{d-j},\ \text{if } n>0, \tag{4.5}$$

which shows that $H_n^0,\dots,H_n^d$ can be simultaneously computed by a monotone circuit of size $O(nd^2)$.

LEMMA 4.6. *There exists $r=r(n,d)>0$ such that $rH_n^d-(1+\sum_{i\in[n]} x_i)^d$ has a monotone circuit of size $O(nd^2)$.*

PROOF. Let

$$h_n^k(r) := rH_n^k - (1+\sum_{i\in[n]} x_i)^k.$$

We shall construct a sequence $r_0,r_1,\dots$ of positive numbers such that for every n, the polynomials $h_n^0(r_n),\dots,h_n^d(r_n)$ can be simultaneously computed by a monotone circuit of size $O(nd^2)$.

Let $r_0 := 1$. Assume we have already constructed r_{n-1}. Setting $\ell := \sum_{i\in[n-1]} x_i + 1$, (4.5) gives

$$
\begin{aligned}
h_n^k(r) &= \sum_{j=0}^{k} x_n^j \left(r H_{n-1}^{k-j} - \binom{k}{j} \ell^{k-j} \right) \\
&= \sum_{j=0}^{k} x_n^j \binom{k}{j} h_{n-1}^{k-j}(r_{n-1}) + \sum_{j=0}^{d} x_n^j H_{n-1}^{k-j} \left(r - \binom{k}{j} r_{n-1} \right). \qquad (4.7)
\end{aligned}
$$

It is now enough to set r_n large enough so that all the terms $r_n - \binom{k}{j} r_{n-1}$ are nonnegative. Then, (4.7) allows to compute $h_n^k(r_n)$ from $h_n^0(r_{n-1}), \ldots, h_n^d(r_{n-1})$ and $H_n^0, \ldots, H_n^d$ using $O(d)$ extra gates. □

The following is an analogy of Theorem 2.1:

PROPOSITION 4.8. *Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be a polynomial of degree d which can be computed by a circuit of size s. Then, there exists $\epsilon_0 > 0$ such for every $0 < \epsilon < \epsilon_0$, the polynomial $H_n^d + \epsilon f$ has a monotone circuit of size $O((s+n)d^2 + n \log n)$.*

PROOF. By Theorem 2.1, we have a small monotone circuit for $U + \epsilon f$ for every $0 < \epsilon < \epsilon_0'$, where $U := (1 + \sum_{i\in[n]} x_i)^d$. Lemma 4.6 gives a small circuit for $rH_n^d - U$ for some $r > 0$. Hence $rH_n^d + \epsilon f = rH_n^d - U + (U + \epsilon f)$ has a small circuit for every $\epsilon < \epsilon_0'$ and it is enough to set $\epsilon_0 := \epsilon_0'/R$. □

As an illustration, we present other possible variants of Theorem 2.1. We choose the examples of multilinear circuits, $\Sigma\Pi\Sigma$-circuits and high-degree computations. Recall that a *syntactically multilinear circuit* (Raz 2004) is an arithmetic circuit C such that for every product gate $u_1 \times u_2$, the sub-circuits of C rooted at u_1 and u_2 share no common variable. We define $\Sigma\Pi\Sigma$-circuit of type (m, k) to be an expression of the form

$$\sum_{i\in[m]} \prod_{i\in[k]} f_{i,j},$$

where $f_{i,j}$ are polynomials of degree at most one. We observe the following:

(*i*). Assume that f has a syntactically multilinear circuit of size s. Then, there is an $\epsilon > 0$ such that $\prod_{i\in[n]}(x_i + 1) + \epsilon f$ has a monotone circuit of size $O(sn)$.

(*ii*). Assume that f has a $\Sigma\Pi\Sigma$ circuit of type (m, k). Then, there is an $\epsilon > 0$ such that $(\sum_{i\in[n]} x_i + 1)^k + \epsilon f$ has a monotone $\Sigma\Pi\Sigma$-circuit of type $(O(mk^2), k)$.

(*iii*). Assume $f \in \mathbb{R}[x]$ is a univariate polynomial with a circuit of size s. Then, there is an $\epsilon > 0$ such that $(1 + x)^{2^s} + \epsilon f$ has a monotone circuit of size $O(s^2)$.

Observe that the bounds no longer depend on the degree of f. This is important especially in the case (*iii*), where the factor of d^2 in Theorem 2.1 would be quite meaningless. We take the liberty to omit the proofs of (*i*)–(*iii*): they proceed in a similar way as Theorem 2.1. In the case of (*ii*), one must reproduce Lemma 4.4 in bounded depth - which can be achieved by the interpolation trick of Ben-Or (cf. Shpilka & Wigderson (1999)).

4.2. A comparison between Theorems 2.1 and 2.5. Let us now explain differences and similarities between the results of this section and Section 3. In order to do this, we need to bring them to a common language. For a polynomial f, denote $\mu(f)$ as the smallest number of non-scalar multiplications needed to compute f by means of an arithmetic circuit, and similarly for $\mu_+(f)$ and a monotone arithmetic circuit. The results of this section could have been stated in terms of $\mu(f)$ instead of total circuit size: namely, if $\mu(f)$ is small then so is $\mu_+(U + \epsilon f)$, where U is a suitable universal polynomial. Furthermore, let us assume that f is a bilinear degree-two polynomial, $f = \sum_{i,j} M_{i,j} x_i y_j$, where M is a real matrix. Observe that, up to a constant factor, $\mu(f)$ equals $\mathrm{rk}(M)$ and, if M is nonnegative, $\mu_+(f) = \mathrm{rk}_+(M)$. Taking the universal polynomial U as $U := \sum_{i,j} x_i y_j$, Theorem 2.1 could be rephrased as asserting $\mu_+(U + \epsilon f) \leq O(\mu(f))$ for some $\epsilon > 0$. In the language of matrices, this means that $\mathrm{rk}_+(J + \epsilon M) \leq O(\mathrm{rk}(M))$. This is something we already know, for Lemma 3.9 gives

$$\min_{\epsilon>0} \mathrm{rk}_+(J + \epsilon M) \leq \mathrm{rk}(M) + 1. \tag{4.9}$$

In this sense, we have obtained a lower bound on $\mathrm{rk}(M)$ in terms of $\mathrm{rk}_+(J+\epsilon M)$—a rather paradoxical thing to say, since rk_+ is way harder to understand than rk. In contrast, Theorem 2.5 lower bounds $\mathrm{sep}(f)$ in terms of $\mathrm{rk}_+(M-\epsilon J)$. This is similar to the bound in (4.9), except that the roles of M and J are exchanged. But the difference is significant: in (4.9) we have a rank-one matrix J which is ϵ-perturbed by a complicated matrix M, whereas in $\mathrm{rk}_+(M-\epsilon J)$, we have a complicated matrix M which is ϵ-perturbed by a rank-one matrix J.

5. The system AMC

As an exercise on AMC proofs, we start with a lemma:

LEMMA 5.1. *(i) Assume that $f_1, \ldots, f_m$ can be simultaneously computed by a monotone circuit of size s and $u, v \in \mathbb{R}^m$ are nonnegative vectors with $supp(u) \subseteq supp(v)$. Then, $\sum_{i\in[m]} u_i f_i \preceq \sum_{i\in[m]} v_i f_i$ has an AMC-proof of size $s+O(m)$.*

(ii) Assume that $f_1 \preceq g_1, \ldots, f_m \preceq g_m$ can be proved by a proof of size s. Then, $\sum_{i\in[m]} f_i \preceq \sum_{i\in[m]} g_i$ and $\prod_{i\in[m]} f_i \preceq \prod_{i\in[m]} g_i$ have a proof of size $s+O(m)$.

PROOF. Part (ii) is obtained by applying the rule (R2) m times. For (i), first prove $u_i f_i \preceq v_i f_i$ using the axioms $f_i \preceq f_i$, $u_i \preceq v_i$ and the rule (R2). Then, we can apply part (ii). □

We now prove Proposition 2.6 and Theorem 2.7.

PROOF OF PROPOSITION 2.6. (i)≡ (ii) is obvious.

(iii) $\Longrightarrow$ (i) can be directly proved by induction on the number of lines in a AMC-proof: $\mathrm{supp}(f) \subseteq \mathrm{supp}(g)$ holds for an axiom $f \preceq g$ and the rules preserve this property.

(ii) $\Longrightarrow$ (iii). We can write $g = h + \epsilon f$ where $h := g - \epsilon f$. By Lemma 5.1 part (i), there is an AMC-proof of $f \preceq h + \epsilon f$ and hence of $f \preceq g$. □

PROOF OF THEOREM 2.7. The "converse" part has been explained in the proof of Proposition 2.6 and it remains to prove the main part of the theorem. In order to simplify the argument, it is convenient to restrict the rule (R2). We will call an application of the rule *simple*, if at least one of its assumptions is an axiom (A1). That is, we modify the rule as

$$\frac{f_1 \preceq g_1,\ f \preceq f}{f_1 \circ f \preceq g_1 \circ f},\ (\circ \in \{+, \times\}).$$

We note this does not affect proof size:

Claim. *Assume that $f \preceq g$ has a AMC-proof of size s. Then, it has an AMC-proof with only simple applications of (R2) of size $O(s)$.*

Proof. We want to derive $f_1 \circ f_2 \preceq g_1 \circ g_2$ from $f_1 \preceq g_1$ and $f_2 \preceq g_2$ by means of simple applications of (R2). To do that, we may first derive $f_1 \circ f_2 \preceq f_1 \circ g_2$ and $f_1 \circ g_2 \preceq g_1 \circ g_2$, and then apply (R1). □

Let $f_1 \preceq g_1, \ldots, f_m \preceq g_m$ be an AMC-proof of size s. By the Claim, we can assume that all the applications of (R2) in the proof are simple. For every $j \in [m]$, we will find $0 < \epsilon_j \leq 1$ such that the polynomials $\{f_i, g_i, g_i - \epsilon_i f_i : i \leq m\}$ can be simultaneously computed by a monotone circuit of size $O(s)$. Let

$$h_i(\epsilon) := g_i - \epsilon f_i.$$

If $f_i \preceq g_i$ is an axiom (A1), we set $\epsilon_i := 1$. If it is an axiom of the form (A2), fix ϵ_i so that $b - \epsilon_i a$ is nonnegative. Assume $f_i \preceq g_i$, was derived from $f_p \preceq g_p$ and $f_q \preceq g_q$ by means of (R1). Then, $g_p = f_q$ and $f_p = f_i$, $g_q = g_i$. Set $\epsilon_i := \epsilon_p \epsilon_q$. This gives

$$\begin{aligned} h_i(\epsilon_i) =& (g_i - \epsilon_q f_q) + \epsilon_q (f_q - \epsilon_p f_i) = (g_q - \epsilon_q f_q) + \epsilon_q (g_p - \epsilon_p f_p) \\ =& h_q(\epsilon_q) + \epsilon_q h_p(\epsilon_p). \end{aligned} \tag{5.2}$$

Assume $f_i \preceq g_i$, was derived from $f_p \preceq g_p$ and an axiom $f_q \preceq f_q$ by means of (R2). Then, let $\epsilon_i := \epsilon_p$. If $\circ = +$, we have

$$h_i(\epsilon_i) = h_p(\epsilon_p) + f_q(1 - \epsilon_p). \tag{5.3}$$

If $\circ = \times$, we have

$$h_i(\epsilon_i) = h_p(\epsilon_p) \cdot f_q. \tag{5.4}$$

Equations (5.2)–(5.4) give a prescription how to compute $h_i(\epsilon_i)$ from the polynomials $f_1, g_1, \dots, f_m, g_m$ and $h_1(\epsilon_1), \dots, h_{i-1}(\epsilon_{i-1})$ using a constant number of additional gates. Altogether, we have shown that $\{f_i, g_i, g_i - \epsilon_i f_i : i \leq m\}$ can be simultaneously computed by a monotone circuit of size $O(s + m)$. Since we have explicitly defined proof size so that $m \leq s$, we obtain that $g_m - \epsilon_m f_m$ has a monotone circuit of size $O(s)$. □

A comment on the power of AMC In the definition of size of AMC-proof, we care only about the circuit size of the polynomials in a line $f \preceq g$, ignoring the question whether the circuits computing f and g make sense in the rest of the proof. For example, we can derive $f \preceq f + g$ as in Lemma 4.6, but it may happen that the smallest circuit for $f + g$ will have nothing to do with f or g. In other words, we gave AMC the power to decide polynomial identities for free. For this reason, the proof of Proposition 2.6 also shows that $f \preceq g$ has an AMC-proof with a constant number of lines (regardless the complexity of f or g). If desired, the system could be made more realistic: we could require AMC to work with arithmetic circuits to begin with, and add more syntactic axioms such as $f(g_1 + g_2) \preceq f g_1 + f g_2$.

5.1. A comparison between Theorems 2.1, 2.5 and 2.7. Observe that Theorem 2.1 can be seen as a corollary of Theorem 2.7. For, assume that f is a homogeneous polynomial of degree d with monotone arithmetic circuit of size s. Then, by induction on s, we can easily construct an AMC-proof of $f \preceq (\sum_i x_i + 1)^d$ of size $O(s + n \log n)$—indeed, this is what Lemma 4.1 implicitly does. This by Theorem 2.7 gives that $(\sum_i x_i + 1)^d - \epsilon f$ has a monotone circuit of size $O(s + n \log n)$. (For non-monotone or inhomogeneous f, we then invoke Lemma 4.4.) Furthermore, the proof of Lemma 4.6 can be interpreted as constructing an AMC-proof of $(\sum_i x_i + 1)^d \preceq H_n^d$.

For comparison with Theorem 2.5, recall the definition of μ and μ_+ from Section 4.2 and the discussion therein. Furthermore, given an AMC-proof S, define the μ_+-complexity as the smallest k so that all the polynomials in S can be simultaneously computed by a monotone arithmetic with k non-scalar multiplications. Then, a lower bound on $\min_{\epsilon>0} \mathrm{rk}_+(M-\epsilon J)$, or on linear separation complexity, can be viewed as a rudimentary AMC lower bound:

OBSERVATION 5.5. *For every n, there exist degree two polynomials f, g with $\mu_+(f), \mu_+(g) \le O(\log n)$ and supp$(f) \subseteq$ supp(g) such that every AMC-proof of $f \preceq g$ has μ_+-complexity $n^{\Omega(1)}$.*

PROOF. Let M be the matrix from Corollary 3.5. Let $f := \sum_{i,j\in[n]} x_i y_j$ and $g := \sum_{i,j\in[n]} M_{i,j} x_i y_j$. Then, $\mu_+(f) = 1$, $\mu_+(g) = \mathrm{rk}_+(M) \le O(\log n)$, and $\mu_+(g - \epsilon f) = \mathrm{rk}_+(M - \epsilon J) \ge n^{\Omega(1)}$ for every $\epsilon > 0$. We leave as an exercise to show that Theorem 2.7 remains valid when measuring the μ_+-complexity: that is, if $f \preceq g$ has a proof of μ_+-complexity s then $\mu_+(g - \epsilon f) \le O(s)$. This gives the required bound. □

5.2. Connections to other proof systems. Let A be a DNF formula in variables $x_1, \ldots, x_n$. Namely,

$$A = \bigvee_{j=1}^{m} A_j, \text{ where } A_j = \bigwedge_{i\in S_j} x_i \wedge \bigwedge_{i\in \bar{S}_j} \neg x_i, \tag{5.6}$$

and $S_j, \bar{S}_j$ are some disjoint subsets of $[n]$. With A, we associate its *characteristic polynomial*, $\chi_n(A)$, as follows. $\chi_n(A)$ is in $2n$ variables $x_1^0, x_1^1, \ldots, x_n^0, x_n^1$, representing the original variables and their negations. For $\sigma \in \{0,1\}^n$, let $x^\sigma := \prod_{i=1}^n x_i^{\sigma_i}$. Then,

$$\chi_n(A) := \sum_{\sigma\in\{0,1\}^n} c_\sigma x^\sigma, \text{ where } c_\sigma := |\{j \in [m] : \sigma \text{ satisfies } A_j\}|.$$

In other words, the coefficient of x^σ is the number of terms A_j satisfied by σ. Hence, $\chi_n(A)$ is a homogeneous polynomial of degree n with integer coefficients from $\{0, \ldots, m\}$.

Setting

$$\chi_n(1) := \prod_{i\in[n]} (x_i^0 + x_i^1),$$

the definition of $\chi_n(A)$ guarantees:

Observation 5.7. *A is a tautology if and only if* $supp(\chi_n(1)) \subseteq supp(\chi_n(A))$.

Note that $\chi_n(A)$ has a monotone arithmetic circuit of size $O(nm)$: given that no A_j contains simultaneously a variable and its negation, we can write

$$\chi_n(A) = \sum_{j=1}^{m}(\prod_{i\in S_j} x_i^1 \prod_{i\in \bar{S}_j} x_i^0 \prod_{i\in([n]\setminus(S_j\cup\bar{S}_j))} (x_i^0 + x_i^1)).$$

Resolution. Recall that Resolution is a proof system designed to refute unsatisfiable CNFs, see, e.g., (Krajíček 1995) for details. More exactly, a *clause* is a set of variables or their negations. A CNF formula can be viewed as a set of clauses $\mathcal{C}$. Resolution has a single rule of inference

$$\frac{C \cup \{x\}, D \cup \{\neg x\}}{C \cup D}.$$

A resolution refutation starts from clauses in $\mathcal{C}$ and derives the empty clause by means of the resolution rule.

Proposition 5.8. *Let A be a DNF as in (5.6). Assume that* $\neg A$ *has a resolution refutation with k lines. Then,* $\chi_n(1)\preceq\chi_n(A)$ *has an AMC-proof of size* $O((m+k)n)$.

Proof. Assume that $\mathcal{C} := \neg A$ has a resolution refutation R with k lines. Without loss of generality, we can assume that no clause in R contains both a variable and its negation, and that in the resolution rule above C, D themselves do not contain x or $\neg x$.

For a clause C, let

$$\alpha_i(C) = \begin{cases} x_i^0, & x_i \in C \\ x_i^1, & \neg x_i \in C \\ x_i^0 + x_i^1, & \text{otherwise.} \end{cases}$$

Let $\alpha(C) := \prod_{i=1}^n \alpha_i(C)$. This guarantees that

$$\chi_n(A) = \sum_{j=1}^{m} \alpha(\neg A_j), \quad \chi_n(1) = \alpha(\emptyset). \tag{5.9}$$

CLAIM. *Assuming C, D do not contain x_i,*

$$\alpha(C \cup D) \preceq \alpha(C \cup x_i) + \alpha(D \cup \neg x_i)$$

has an AMC-proof of size $O(n)$.

PROOF OF THE CLAIM. By definition, $\alpha(C \cup D) = \alpha(C \cup D \cup \{x_i\}) + \alpha(C \cup D \cup \{\neg x_i\})$. Hence, it is enough to construct proofs of $\alpha(C \cup D \cup \{x_i\}) \preceq \alpha(C \cup \{x_i\})$ and $\alpha(C \cup D \cup \{\neg x_i\}) \preceq \alpha(D \cup \{\neg x_i\})$. Both these inequalities are of the form $\alpha(D_1) \preceq \alpha(D_2)$ with $D_2 \subseteq D_1$. But for every i, $\alpha_i(D_1) \preceq \alpha_i(D_2)$ has a constant size proof: either $\alpha_i(D_1) = \alpha_i(D_2)$, or $\alpha_i(D_1) = x_i^e$, $e \in \{0, 1\}$ and $\alpha_i(D_2) = x_i^0 + x_i^1$. Using Lemma 5.1 part (ii), we obtain a proof of $\prod_i \alpha_i(D_1) \preceq \prod_i \alpha_i(D_2)$ of size $O(n)$. □

Using the Claim, we can construct a proof of $\alpha(C) \preceq \chi_n(A)$ for every clause C in R. If $C = \neg A_j$ is an initial clause in $\mathcal{C}$, this follows from Lemma 5.1 and (5.9). If C was obtained by resolving clauses C', D', we use the claim to derive $\alpha(C) \preceq \chi_n(A)$ from $\alpha(C') \leq \chi_n(A)$ and $\alpha(D') \preceq \chi_n(A)$. This will give an AMC-proof of $\alpha(\emptyset) \preceq \chi_n(A)$. Altogether, the proof will have size at most $O((m + k)n)$. □

COROLLARY 5.10. *Let A be as in Proposition 5.8. Then, there exists $\epsilon > 0$ such that $\chi_n(A) - \epsilon\chi_n(1)$ has a monotone arithmetic circuit of size $O(n(k + m))$.*

We believe that Proposition 5.8 and its corollary can be improved to give monotone $\Sigma\Pi\Sigma$-circuits (as defined in Section 4.1).

Monotone calculus. Recall the monotone calculus proof system, MLK, as considered by Atserias et al. in (Atserias *et al.* 2002). In this system, one proves tautologies $A \to B$ where A, B are monotone formulas. The system starts from axioms such as $A \to A$, and derives new formulas by means of inference rules of the flavor

$$\frac{A \to B}{A \to B \vee C}, \quad \frac{A \to B,\ B \to C}{A \to C}.$$

In (Atserias *et al.* 2002), it was shown that MLK quasipolynomially simulates the Frege system. Moreover, if we allow MLK to work with Boolean circuits rather than formulas, the system polynomially simulates the Extended Frege system.

For a monotone Boolean circuit A, let A^* be the monotone arithmetic circuit obtained by replacing $\wedge, \vee$ in A by $\times, +$, respectively. Let F_A be the monotone polynomial computed by A^*. Observe that $\mathrm{supp}(F_A) \subseteq \mathrm{supp}(F_B)$ implies that $A \to B$ is a tautology. The converse is not true: for example $x \wedge y \to y$ is a tautology, whereas $\mathrm{supp}(F_{x\wedge y}) = \{xy\}$ and $\mathrm{supp}(F_x) = \{x\}$. This means there exist tautologies $A \to B$ such that $F_A \preceq F_B$ is not even provable in AMC. We remark without a proof that one can simulate MLK by AMC augmented with the Boolean axioms $f \preceq 1$ and $f \preceq f^2$. In this sense, AMC can be seen as a weakening of the monotone calculus. It is, however, an open problem whether the Boolean axioms can help in proving $\chi_n(1) \preceq \chi_n(A)$.

6. Open problems

We end by giving some open problems. The first one asks for new monotone arithmetic lower bounds. Recall that the permanent polynomial is defined as $\mathrm{perm}_n = \sum_\sigma \prod_{i=1}^n x_{i,\sigma(i)}$, where σ ranges overall permutations of $[n]$.

Open problem 1. Show that $\prod_{i=1}^n (\sum_{j=1}^n x_{i,j}) - \mathrm{perm}_n$ requires a monotone arithmetic circuit of superpolynomial size. How about $\prod_{i=1}^n (\sum_{j=1}^n x_{i,j}) + \mathrm{perm}_n$?

The next problem concerns continuity of nonnegative rank as discussed in Section 3.2.

Open problem 2. Given $M, V \in \mathbb{R}^{n\times m}$ and $z \in \mathbb{R}$, let $r(z) := \mathrm{rk}_+(M - zV)$. Assuming M, V are positive and V is a rank-one matrix, how many discontinuities can the function $r(z)$ have? How many times can $r(z)$ decrease as z increases?

The next two questions concern monotone separation complexity and strict rank from Sections 3.4 and 3.3. They are closely related due to the discussion in Section 3.4.1.

Open problem 3. Find an explicit monotone Boolean function f such that $\mathrm{sep}_+(f)$ is superpolynomial.

Open problem 4. Find an explicit positive matrix M such that $\mathrm{rk}_{++}(M)$ is superpolynomial in terms of $\mathrm{rk}_{+}(M)$.

The final problems are related to the system AMC from Section 2.3 and 5:

Open problem 5. Find a pair of monotone polynomials f, g with $\mathrm{supp}(f) \subseteq \mathrm{supp}(g)$ such that for every $\epsilon > 0$, $g - \epsilon f$ requires monotone arithmetic circuit of size superpolynomial in the monotone arithmetic circuit size of f and g.

Open problem 6. Does AMC polynomially simulate the Frege system? More exactly, assume that a DNF A has a Frege proof of size s. Is there an AMC-proof of $\chi_n(1) \preceq \chi_n(A)$ of size polynomial in s?

Acknowledgements

We thank Pavel Pudlák for inspiration and Amir Yehudayoff for, unwittingly, making the author to write the paper in the first place. Supported by the GACR grant 19-27871X.

References

A. Atserias, A. Dawar & J. Ochremiak (2019). On the power of symmetric linear programs. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*.

A. Atserias, N. Galesi & P. Pudlák (2002). Monotone simulations of non-monotone proofs. *J. of Computer and System Sciences* **65**, 626–638.

S. Basu, R. Pollack & M.F. Roy (2006). *Algorithms in real algebraic geometry*. Springer-Verlag.

L. Beasley & T. Laffey (2009). Real rank versus nonnegative rank. *Linear Algebra and its Applications* **431**(12), 2330–2335.

G. Braun, S. Fiorini, S. Pokutta & D. Steuer (2012). Approximation Limits of Linear Programs (Beyond Hierarchies). In *FOCS*, 480–489.

P. Bürgisser, M. Clausen & M. A. Shokrollahi (1997). *Algebraic complexity theory*, volume 315 of *A series of comprehensive studies in mathematics*. Springer.

D. deCaen, D. Gregory & N. Pullman (1981). The Boolean rank of zero one matrices. In *Proceedings of the Third Caribean Conference on Combinatorics and Computing*, 169–173.

Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary & Ronald de Wolf (2011). Linear vs. Semidefinite Extended Formulations: Exponential Separation and Strong Lower Bounds. *CoRR* arXiv:1111.0837 .

M. Goos, R. Jain & T. Watson (2018). Extension complexity of independent set polytopes. *SIAM J. Comput.* **47**(1), 241–269.

P. Hrubeš (2012). On the nonnegative rank of distance matrices. *Information Processing Letters* **112**(11), 457–461.

P. Hrubeš (2019). On the complexity of computing a random Boolean function over the reals. *ECCC*.

Pavel Hrubeš & Amir Yehudayoff (2011). Homogeneous formulas and symmetric polynomials. *Computational Complexity* **20**(3), 559–578.

J. Krajíček (1995). *Bounded arithmetic, propositional logic, and complexity theory*. Cambridge University Press, USA.

N. Nisan (1991). Lower bounds for non-commutative computation. In *Proceeding of the 23th STOC*, 410–418.

P. Pudlák & M. de Oliveira Oliveira (2017). Representations of monotone Boolean functions by linear programs. In *Proceedings of the 32nd Computational Complexity Conference*.

R. Raz (2004). Multi-linear formulas for Permanent and Determinant are of super-polynomial size. In *Proceeding of the 36th STOC*, 633–641.

A. Razborov (1992). On submodular complexity measures. In *Boolean functions complexity*, 76–83. Cambridge University Press.

T. Rothvoss (2017). The matching polytope has exponential extension complexity. *J. of the ACM* **64**(6).

Thomas Rothvoss (2011). Some 0/1 polytopes need exponential size extended formulations. *CoRR* arXiv:1105.0036.

E. Shamir & M. Snir (1979). On the depth complexity of formulas. *Journal Theory of Computing Systems* **13**(1), 301–322.

Amir Shpilka & Avi Wigderson (1999). Depth-3 arithmetic formulae over fields of characteristic zero. In *In CCC*, 87. IEEE Computer Society.

Amir Shpilka & Amir Yehudayoff (2010). Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science* **5**(3), 207–388.

L. G. Valiant (1979). Completeness Classes in Algebra. In *STOC*, 249–261.

L. G. Valiant (1980). Negation can be exponentially powerful. *Theoretical Computer Science* **12**, 303–314.

L. G. Valiant (1982). Reducibility by algebraic projections. *Enseign. Math.* **28**, 253–268.

L. G. Valiant (1986). Negation is powerless for Boolean slice functions. *SIAM J. Comput.* **15**(2), 531–535.

S. A. Vavasis (2008). On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization* **20**(3), 1364-1377.

I. Wegener (1987). The complexity of Boolean functions.

Mihalis Yannakakis (1991). Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences* **43**(3), 441–466.

A. Yehudayoff (2019). Separating monotone VP and VNP. In *STOC*.

Manuscript received 6 March 2019

Pavel Hrubeš
Institute of Mathematics
of ASCR, Prague Czech Republic.
pahrubes@gmail.com