Exponential Lower Bounds for Depth 3 Arithmetic Circuits in Algebras of Functions over Finite Fields

D. Grigoriev

IMR Université Rennes-1
Beaulieu 35042 Rennes France
dima@maths.univ-rennes1.fr

A. Razborov[1]

Steklov Mathematical Institute
Gubkina 8, 117966, GSP-1
Moscow, Russia
razborov@genesis.mi.ras.ru

Abstract

A depth 3 arithmetic circuit can be viewed as a sum of products of linear functions. We prove an exponential complexity lower bound on depth 3 arithmetic circuits computing some natural symmetric functions over a finite field F. Also, we study the complexity of the functions $f : D^n \to F$ for subsets $D \subset F$. In particular, we prove an exponential lower bound on the complexity of depth 3 arithmetic circuits computing some explicit functions $f : (F^*)^n \to F$ (in particular, the determinant of a matrix).

[1]Partially supported by RBRF grants 96-01-01222, 96-15-96090 and by INTAS grant 96-753.

Introduction

A depth 3 arithmetic circuit can be viewed as a sum of products of linear functions. Despite this clear structure, only a handful of lower bounds for explicit polynomials are known over infinite fields. Super-polynomial lower bounds have been proven only under the assumption that the circuits involve just (homogeneous) linear forms, rather than arbitrary linear functions, by the same token, if products in a circuit contain a bounded number of linear functions (see [7, 11]). For general circuits no bounds for depth 3 circuits are known better than the classical $\Omega(n \log n)$ bound [15, 2] for arbitrary depth circuits (observe that this bound concerns the *multiplicative* complexity, being different from the complexity measure of the number of gates at the middle level of a depth 3 arithmetic circuit (1) which we study in the present paper). Using some ideas from [7], [13] recently proved a nearly quadratic lower bound for depth 3 *formulae* computing some symmetric functions.

The situation changes significantly when our underlying field is finite, both in terms of the framework as well as in terms of approaches and results. Let us call *syntactical* the ordinary framework of algebraic complexity in which polynomials from $F[X_1, \ldots, X_n]$ are understood as formal syntactical expressions. In this framework an exponential $\exp(\Omega(n))$ lower bound on the complexity of *general* depth 3 arithmetic circuits for the determinant of $n \times n$ matrices was recently proved in [9].

An equally natural framework, also extensively studied in the literature, treats polynomials from $F[X_1, \ldots, X_n]$ as *functions* $F^n \to F$, and we call this framework *functional*. It is equivalent to working in the factor-algebra $F[X_1, \ldots, X_n]/(X_1^q - X_1, \ldots, X_n^q - X_n)$, where $q = |F|$, and every syntactical computation is also a computation in the functional framework. Respectively, obtaining lower bounds for functions is an even more difficult task, and prior to this paper exponential lower bounds were known only for the case $F = GF(2)$ [12].

Finally, in the seminal paper [14] Smolensky proposed to study also computations in the function algebra $\{0,1\}^n \to F$ for fields other than $GF(2)$, and, for obvious reasons, we call this framework *Boolean* (syntactically, this means that we impose the relations $X_i^2 = X_i$ for all variables X_i). The bulk of the research in this framework was devoted to *Boolean* circuits, i.e., to circuits composed entirely of $\{0,1\}$-valued gates. In particular, [12, 14] proved exponential lower bounds for bounded depth Boolean circuits over the basis $\{\neg, \wedge, \vee, MOD_p\}$ that make the closest approximation to arithmetic circuits in the Boolean world. Motivated by a related research in the structural complexity, [1] proposed to study in the Boolean framework also computations by *arithmetic* circuits. In particular, they showed that after taking the union over all finite fields, bounded depth arithmetic circuits capture exactly the complexity class ACC^0. So, these circuits form a natural hierarchical structure within the latter class that might be useful for understanding its power, and this feeling is further confirmed by the current paper. Prior to it, no non-trivial lower bounds were known in this model for depth 3 circuits over any field other than $GF(2)$.

Our contributions are as follows.

First, we give a short proof of an $\exp(\Omega(n))$ lower bound for depth 3 circuit in the functional framework over any finite field. More specifically, we show that every depth 3 circuits over a prime field $GF(p)$ computing suitably defined generalizations of MOD_q and MAJ must have that many gates. Then we give an easy extension to arbitrary (i.e., not necessarily prime) finite fields. This in particular gives an alternative (and much simpler) proof of exponential lower bounds over finite fields in the syntactical framework for explicitly given symmetric functions (rather than the determinant [9]).

In the Boolean framework we can prove new lower bounds only for the field $GF(3)$. Our techniques, however, are more general, and a substantial part of them can be applied to larger fields as well. In order to understand and precisely state the corresponding results, we observe that there is no intrinsic difference between functional and Boolean frameworks: these are simply the two opposite poles of more general *quasi-boolean* framework in which we study computations of functions $f : D^n \to F$, where $D \subseteq F$ is *arbitrary*. In these terms, we can prove an $\exp(\Omega(n))$ lower bound for functions $f : (F^*)^n \to F$ over *arbitrary* finite fields with at least three elements, and when $F = GF(3)$ this becomes equivalent (up to a linear transformation on variables) to the Boolean framework. In particular, this result strengthens our previous bound in the functional framework (the reason for including the latter in the paper lies in its simplicity and applicability to symmetric functions).

The results of the paper were announced in [10].

Table 1 summarizes our current knowledge about the best known lower bounds for depth 3 arithmetic circuits.

The paper is organized as follows. In Section 1 we give a short proof of our bound in the functional framework (Theorem 1.1).

Field\Framework	Boolean	Functional	Syntactical
$GF(2)$	$\exp(\Omega(n))$ [12] $\Rightarrow$		$\exp(\Omega(n))$
$GF(3)$	$\exp(\Omega(n))$ Corollary 2.3	$\exp(\Omega(n))$ Theorem 1.1 $\Rightarrow$	$\exp(\Omega(n))$
$GF(q),\ q > 3$	?	$\exp(\Omega(n))$ Theorem 1.1 $\Rightarrow$	$\exp(\Omega(n))$
infinite	?	$\Omega(n \log n)$ [15, 2]	

Table 1: Lower bounds for depth 3 arithmetic circuits

The rest of the paper is devoted to the quasi-boolean setting. In Section 2 we give some basic properties of the algebra of all functions $f : D^n \to F$, where $D \subseteq F$, and state our main result which is a combinatorial property of functions $(F^*)^n \to F$ that implies large complexity w.r.t. arithmetic depth 3 circuits (Theorem 2.2). As one of the applications of this general criterium we obtain $\exp(\Omega(\sqrt{n}))$ lower bound for the determinant and the permanent of an $n \times n$ matrix (compare with $\exp(\Omega(n))$ lower bound [9] for the determinant in the syntactical framework).

Sections 3-6 are entirely devoted to the proof of Theorem 2.2, and we hope that some of the techniques we introduce on this way might be helpful in other situations as well.

In Section 3 we introduce a slight variation of Valiant's rigidity function [16, 6] that we call *m-rigid rank* and show a lower bound on this measure in terms of more constructive *m-communication rank* also introduced in this section. The bound is shown to be tight in case of constant m (which is the only case we need in this paper).

In Section 4 we prove that a product of linear functions has only a few non-zeros on D^n, provided that this product has a large communication (or thereby, rigid as well) rank (Lemma 4.1). This allows us to approximate products of linear functions with large communication rank by a zero function, and to deal in the sequel only with the products having small communication rank.

In Section 5 we provide an approximation of a product with small communication rank by a function having some special form, and combine this with material from Section 4 into Theorem 5.1. In the partial case $D = F^*$ this results in an approximation by a sparse polynomial (Lemma 5.3). Moreover, the support (the set of monomials) of this polynomial lies in a union of few balls (w.r.t. the Hamming metric), each of a small radius.

Finally, in Section 6 we prove that if the support of a function $f : (F^*)^n \to F$ has a large coding distance, and its size is relatively small, then f can not be approximated well by a sparse polynomial of the above form. This concludes the proof of our main Theorem 2.2.

1 Exponential lower bound for depth 3 arithmetic circuits for symmetric functions over a finite field

We study depth 3 arithmetic circuits, that is representations of functions in the following form:

$$f = \sum_{1 \le \nu \le N} \prod_i L_{\nu i}, \qquad (1)$$

where $L_{\nu i} = \sum_{1 \le j \le n} \alpha_{ij}^{(\nu)} X_j + \alpha_i^{(\nu)}$ are linear functions. We call the right-hand side of (1) a depth 3 arithmetic circuit since it contains 3 layers (with unbounded fan-in) of computations: its first layer consists in computing linear functions $L_{\nu i}$, the second layer is computing their products over i, and the third one is computing the sum over ν just according the right-hand side. In this section we consider the *functional framework* in which the identity (1) is understood as the identity of functions $f : F^n \to F$ over the underlying field F. Our purpose is to give a short proof of exponential lower bounds on the complexity (in fact, on the number of gates at the middle level N) in the representations (1) for quite natural symmetric functions f over any fixed finite field F. Let us begin with the case $F = GF(p)$, where p is a fixed prime.

Viewing each element $x \in F$ as an *integer* $0 \le x \le p-1$, one can define for any prime q the generalization $MOD_{q,F} : F^n \to F$ of the corresponding Boolean function MOD_q as follows:

$$MOD_{q,F}(x_1, \dots, x_n) = \begin{cases} 1 & \text{if } \sum\limits_{1 \le j \le n} x_j \equiv 0 \pmod{q} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.1 *Provided $F = GF(p)$ and q is another prime, for every representation $MOD_{q,F}$ in the form* (1) *the lower bound $N \ge \exp(\Omega(n))$ holds.*

Proof. Similarly to [12, 14], we want to show that every small size depth 3 arithmetic circuit can be approximated by a low degree polynomial. Let us consider an individual product $\Pi = \prod_i L_i$, $L_i = \sum\limits_{1 \le j \le n} \alpha_{ij} X_j + \alpha_i$. By its *rank* $rk(\Pi)$ we mean the rank of the matrix of coefficients (α_{ij}) of the linear functions $\{L_i\}$ without their free terms. Take some threshold r (to be specified later). We treat separately the products Π with rank greater or less than r.

Case 1. $rk(\Pi) \ge r$.

In this case we have the obvious bound on the probability

$$Pr\left[x \in F^n : \prod_i L_i(x) \ne 0\right] \le \left(\frac{p-1}{p}\right)^r.$$

Case 2. $rk(\Pi) < r$.

Then we have the bound on the degree $\deg(\Pi) \le r(p-1)$. Indeed, express each L_i as a linear combination of (less than r) elements of a basis, thereupon open the parenthesis in the product and use the relation $L^p = L$ present in the functional framework.

Applying this case analysis separately to every product in (1), we have the following:

Lemma 1.2 *For every function f representable in the form* (1) *and every parameter r there exists a polynomial g with $deg(g) \le r(p-1)$ such that*

$$Pr\left[x \in F^n : f(x) \ne g(x)\right] \le \left(\frac{p-1}{p}\right)^r N.$$

Unfortunately, we do not know how to apply this lemma directly: it seems, known lower bounds on the rate of approximation of explicit functions by low degree polynomials over finite fields other than $GF(2)$ are too weak. We turn around this difficulty by the following simple trick that gives us a *direct* reduction to necessary bounds in the Boolean case [14].

Consider the Boolean cube $B^n = \{0,1\}^n \subseteq F^n$. For any vector $c = (c_1, \ldots, c_n) \in F^n$ consider the (shifted) function $MOD_{q,F}^{(c)} : B^n \to F$ defined by $MOD_{q,F}^{(c)}(x_1, \ldots, x_n) = MOD_{q,F}(x_1 + c_1, \ldots, x_n + c_n)$. Actually, for any function $h : B^n + c \to F$ one could define the shifted function $h^{(c)} : B^n \to F$ by $h^{(c)}(x) = h(x + c)$. We call c *nondegenerated* if at least $n/3$ of its coordinates $c_1, \ldots, c_n$ are distinct from $p-1$, and *degenerated* otherwise. Clearly,

$$Pr\left[c \in F^n : c \text{ is nondegenerated}\right] \geq \frac{1}{2}.$$

Suppose now that $MOD_{q,F}$ has a representation (1) with N terms, and let g be chosen in accordance with Lemma 1.2. Since any point from F^n belongs to the same number 2^n of the shifted boolean cubes $B^n + c$, there exists at least one *nondegenerated* vector $c \in F^n$ for which

$$Pr\left[x \in B^n : MOD_{q,F}^{(c)}(x) \neq g^{(c)}(x)\right] \leq 2\left(\frac{p-1}{p}\right)^r N. \qquad (2)$$

Assume w.l.o.g. that $0 \leq c_1, \ldots, c_s \leq p-2,\ s \geq n/3$. Then there exists a fixed 0-1 assignment $a_{s+1}, \ldots, a_n$ to the last $n-s$ variables such that the bound (2) is preserved, i.e.,

$$Pr\left[x \in B^s : MOD_{q,F}^{(c)}(x,a) \neq g^{(c)}(x,a)\right] \leq 2\left(\frac{p-1}{p}\right)^r N. \qquad (3)$$

Notice that $MOD_{q,F}^{(c)}(x,a) \equiv MOD_{t,q}(x)$ for some $0 \leq t < q$, where $MOD_{t,q} : B^s \to B$ is the *Boolean* function defined in [14]. We need the following numerical refinement on the main technical tool from that paper:

Lemma 1.3 (Smolensky) *Let q and p be different primes, $0 \leq t < q$ and $g(X_1, \ldots, X_n)$ be a polynomial over $GF(p)$ of degree at most d. Then $MOD_{t,q}(x_1, \ldots, x_n)$ and $g(x_1, \ldots, x_n)$ diverge in at least $\Omega\left(\sum_{i=0}^{n/2-d-1} \binom{n}{i}\right) \geq \Omega\left(\binom{n}{n/2-d-1}\right)$ points from B^n.*

[14], as well as [12] in a similar statement, put here $d \sim \sqrt{n}$ which ensures disagreement in at least a polynomial fraction of all inputs. We notice that one can achieve better results when d is linear in n. In particular, comparing (3) and Lemma 1.3, the latter with $n := s,\ g := g^{(c)}(x, a),\ d := (p-1)r$, we get in our case:

$$N \geq 2^{-s} \cdot \exp(\Omega(r)) \cdot \binom{s}{s/2 - O(r)}.$$

When $r = \epsilon s$ for a constant ϵ, $\binom{s}{s/2-\epsilon s} \geq 2^{s(H(1/2-\epsilon)-o(1))}$, where $H(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$ is the entropy function. Since $H'(1/2) = 0$, it follows that when ϵ is small enough, $N \geq \exp(\Omega(r)) \geq \exp(\Omega(n))$. Theorem 1.1 is proved.

One can introduce a symmetric function $MAJ_F : F^n \to F$ similar to the customary $MAJ : B^n \to \{0, 1\}$ and being universal for all symmetric functions. Namely, $MAJ_F(x_1, \ldots, x_n) = 1$ if $\gamma_0 \leq \gamma_1 \leq \cdots \leq \gamma_{p-1}$ where γ_i equals to the number of i among $x_1, \ldots, x_n$ for $0 \leq i \leq p-1$, otherwise $MAJ_F(x_1, \ldots, x_n) = 0$. One can show (similar to [12, Theorem 4]) that any symmetric function $f(x_1, \ldots, x_n) : F^n \to F$ could be represented as a F-linear combination of polynomially many functions of the form

$$MAJ_F(X_1, \ldots, X_n, \eta_1, \ldots, \eta_{n(p-1)})$$

for suitable $\eta_i \in F$, $1 \leq i \leq n(p-1)$. This entails the following corollary.

Corollary 1.4 *For the complexity of any depth 3 arithmetic circuit* (1) *representing MAJ_F the lower bound $N \geq \exp(\Omega(n))$ holds.*

Finally, we notice that there are many natural ways of extending the definitions of $MOD_{q,F}, MAJ_F$ to the fields $F = GF(p^\lambda),\ \lambda > 1$ such that the same method gives lower bounds also in this case. For example, we might fix an arbitrary $GF(p)$-linear retraction $\phi : F \to GF(p)$ and then let

$$MOD_{q,F}(x_1, \ldots, x_n) = MOD_{q,GF(p)}(\phi(x_1), \ldots, \phi(x_n)),$$

and similarly for MAJ_F. It is easy to see that with this definition the proofs of Theorem 1.1 and Corollary 1.4 extend to arbitrary finite fields.

2 Quasi-boolean functions over finite fields and main results

In the rest of the paper we deal with the following *quasi-boolean* setting. Let $q = p^\lambda$, $F = GF(q)$, and fix a subset $D \subseteq F$ of cardinality $d = |D| \geq 2$. We are interested in functions $f : D^n \to F$ and call them *quasi-boolean* extending the *Boolean framework* from [14] where D was just $\{0, 1\}$. Alternatively, one could also view f as a partial function on the entire space F^n.

Let $g(X) = \prod_{a \in D}(X - a)$. Then the F-algebra of all functions $f : D^n \to F$ is isomorphic to the quotient algebra

$$\mathcal{A} = F[X_1, \ldots, X_n]/(g(X_1), \ldots, g(X_n)).$$

The main purpose in the rest of the paper is to obtain lower bounds on the complexity of depth 3 arithmetic circuits (1) for certain explicit functions $f \in \mathcal{A}$, equality (1) being viewed also in the algebra $\mathcal{A}$, in the case $d = q - 1$. Before formulating our results, however, let us mention some easy properties of $\mathcal{A}$ (cf. [14], also [8]).

Lemma 2.1 *a) $\mathcal{A}$ is an algebra of principal ideals;*

b) for any $f \in \mathcal{A}$ the number of nonzeroes $|\{x \in D^n : f(x) \neq 0\}|$ coincides with the dimension of the principal ideal $(f) \subseteq \mathcal{A}$.

Clearly, monomials of the form $X_1^{u_1} \cdot \ldots \cdot X_n^{u_n}$, $0 \leq u_1, \ldots, u_n < d$, constitute a basis of $\mathcal{A}$; for an element $f \in \mathcal{A}$ we refer to its degree w.r.t. this basis. In abuse of notation we identify sometimes f with the corresponding polynomial in this basis.

Below it will be sometimes convenient to imagine a metric geometry in the space of monomials $\mathcal{M} = \{X_1^{u_1} \cdot \ldots \cdot X_n^{u_n}\}_{0 \leq u_1, \ldots, u_n < d}$ endowed with the Hamming distance ρ (being equal to the number of distinct coordinates). For an n-tuple $u = (u_1, \ldots, u_n)$ we will abbreviate the corresponding monomial $\prod_{j=1}^n X_j^{u_j}$ to X^u.

Suppose now that $D \subseteq F^*$ and, moreover, that D is a coset modulo some (cyclic) subgroup in F^*. Then the minimal polynomial $g = \prod_{a \in D}(X - a)$ of D is a binomial $q = X^d - b$ and $\mathcal{M}$ is the *multiplicative basis* in the sense that the following two properties are satisfied:

1. The set of functions $D^n \to F$ representable as $\alpha \cdot X^u$, where $\alpha \in F^*$, $X^u \in \mathcal{M}$, forms a group w.r.t. multiplication;

2. the Hamming distance is invariant under multiplication in this group: $\rho(m'm_1, m'm_2) = \rho(m_1, m_2)$.

For $\mathcal{N} \subseteq \mathcal{M}$, define its *coding distance* $R(\mathcal{N})$ as the minimum of $\rho(m_1, m_2)$ over all pairs $m_1 \neq m_2,\ m_1, m_2 \in \mathcal{N}$. Thus, $R(\mathcal{N})$ is equal to the ordinary coding distance $R(U)$ of the set $U = \{u \mid X^u \in \mathcal{N}\}$ considered as an error-correcting code in $[d]^n$. For $f \in \mathcal{A}$ represented in the basis $\mathcal{M}$, let $\mathrm{supp}\,(f)$ be the set of all monomials that occur in f with a non-zero coefficient. Our main lower bound criterium is the following theorem.

Theorem 2.2 *For every fixed finite field F there exists a positive constant $\epsilon = \epsilon(F) > 0$ such that the following holds. Let $\mathcal{N} \subseteq \mathcal{M}$ satisfy the inequality*

$$|\mathcal{N}| + n \leq \exp(\epsilon \cdot R(\mathcal{N})^3/n^2). \tag{4}$$

Then for any function $f : (F^*)^n \to F$ with $supp\,(f) = \mathcal{N}$, any depth 3 arithmetic circuit (1) *computing f has at least $|\mathcal{N}|$ gates at the middle level.*

Corollary 2.3 *There exists a positive constant $\epsilon > 0$ such that for any $U \subseteq B^n$ satisfying*

$$|U| + n \leq \exp(\epsilon \cdot R(U)^3/n^2)$$

any depth 3 arithmetic circuit of the form (1) *computing the function*

$$\sum_{u \in U} \prod_{j=1}^{n} (X_j + 1)^{u_j} : B^n \to GF(3)$$

must have at least $|U|$ gates at the middle level.

Proof. Immediate from Theorem 2.2 after the linear substitution $X_j \mapsto X_j + 1$.

Let p_0 be any prime divisor of $q - 1$, and let $U \subseteq (GF(p_0))^n$ be any explicit $GF(p_0)$-linear code with the coding distance $R = \delta_0 n$ and dimension $k = \epsilon_0 n$ for some positive constants δ_0, ϵ_0. Among examples of such codes are e.g. Justensen codes [19] and Goppa codes [18]. Let $u_1, \ldots, u_k$ be an explicit basis of U over $GF(p_0)$. Removing, if necessary, some vectors from this basis, we may assume w.l.o.g. that $\epsilon_0 < \epsilon\delta_0^3/(\log p_0)$, where ϵ is the constant from Theorem 2.2. Then Theorem 2.2 implies in particular an $\exp(\Omega(n))$ lower bound on the complexity of any depth 3 arithmetic circuit computing the function

$$f_U = \sum_{u \in U} X^{\left(\frac{q-1}{p_0}\right)u} : (F^*)^n \to F.$$

Next, since U is an $GF(p_0)$-linear space with basis $u_1, \dots, u_k$ we have

$$f_U = \sum_{\mu_1,\dots,\mu_k \in GF(p_0)} X^{\left(\frac{q-1}{p_0}\right)(\mu_1 u_1 + \dots + \mu_k u_k)} = \prod_{i=1}^{k} \left(\sum_{\mu \in GF(p_0)} X^{\mu \frac{q-1}{p_0} u_i} \right),$$

thus f_U can be obtained as a projection of the following function

$$\prod_{i=1}^{k} \sum_{\mu=1}^{p_0} \prod_{j=1}^{n(q-1)} X_{i\mu j} : F^* \to F. \tag{5}$$

Hence, the complexity of (5) w.r.t. Σ_3 arithmetic circuits is $\exp(\Omega(n))$; on the other hand, this function is represented as a Π_3 arithmetic circuit of size $O(n^2)$. This gives us a separation between these two classes.

Finally, we make use of the construction from [17] (see also [5]). Following [17] we say that a polynomial $f \in F[X_1, \dots, X_n]$ is a projection of a polynomial $g \in F[Y_1, \dots, Y_m]$ if substituting in g for each variable Y_j in a suitable way either one of the variables $X_1, \dots, X_n$ or a constant from F, one gets f. It is proved in [17] that if f is representable by an arithmetic formula of a size t then f is a projection of

- Determinant $\sum_{\pi \in S_{t+2}} (-1)^{sgn(\pi)} \prod_{1 \le i \le t+2} X_{i,\pi(i)}$;
- Permanent $\sum_{\pi \in S_{t+2}} \prod_{1 \le i \le t+2} X_{i,\pi(i)}$;
- Hamiltonian cycles polynomial $\sum_{\pi} \prod_{1 \le i \le t+2} X_{i,\pi(i)}$, where the summation is taken over all permutations π which consist of a single cycle.

Thus, $m = (t+2)^2$ in the construction of [17]. Applying this result to the function f_U in algebra $\mathcal{A}$, we conclude with the following corollary.

Corollary 2.4 *For each of the following three functions* $: (F^*)^{n^2} \to F$*:*

- *Determinant*
- *Permanent*

- *Hamiltonian cycles polynomial*

any depth 3 arithmetic circuit (1) *computing* f *must have at least* $\exp(\Omega(\sqrt{n}))$ *gates at the middle level.*

The rest of the paper is entirely devoted to the proof of Theorem 2.2. As we noted in Introduction, we will try to present as many techniques as possible for as general D as possible, and we will employ the same idea of approximation as in Section 1. For doing that we need some properties of individual products $\Pi = \prod_i L_i$ which would ensure that $\Pi(x) = 0$ holds with high probability on D^n. Clearly, the ordinary notion of rank is already not good enough: for example, members of the multiplicative basis $\mathcal{M}$ never evaluate to 0 on D^n. More generally, the same holds if L_i are arbitrary linear functions without zeros in D^n, and we begin our analysis with identifying the case when we at least know that such "unpleasant" L_i must have only a constant number of variables.

Definition 2.5 For any integer $2 \leq d \leq q$ a *linear d-sweep* $\ell s(d)$ (w.r.t. the field $F = GF(q)$) is the minimal m (provided that it does exist) such that for any m subsets $D_1, \ldots, D_m \subseteq F$, $|D_1| = \ldots = |D_m| = d$ any linear function $L(X_1, \ldots, X_m) = a_1X_1 + \cdots + a_mX_m + a$ with nonvanishing coefficients $a_1 \neq 0, \ldots, a_m \neq 0$, sweeps the entire F in the sense $a_1D_1 + \cdots + a_mD_m = F$.

Lemma 2.6 *a)* $\ell s(d)$ *is defined if and only if* $d > p^{\lambda-1}$, *in this case* $\ell s(d) \leq q - d + 1$;

b) for $q/2 < d < q$ *we have* $\ell s(d) = 2$.

Proof. a) The part only if is obvious since in case $\lambda > 1$, $d \leq p^{\lambda-1}$ one can take arbitrary $D_1 = \ldots = D_m \subseteq GF(p^{\lambda-1})$ and $a_1, \ldots, a_m, a \in GF(p^{\lambda-1})^*$. To prove the inverse, let us show that

$$|a_1D_1 + \cdots + a_{\ell+1}D_{\ell+1}| > |a_1D_1 + \cdots + a_\ell D_\ell|, \qquad (6)$$

unless already $a_1D_1 + \cdots + a_\ell D_\ell = F$. Together with $|a_1D_1| = d$, this will entail a) by induction on ℓ.

Pick arbitrarily $\alpha_0 \in D_{\ell+1}$. Since $|D_{\ell+1}| > p^{\lambda-1}$, $D_{\ell+1} - \alpha_0$ generates F as a $GF(p)$-linear space. Hence, there exist elements $\alpha_1, \ldots, \alpha_\lambda \in D_{\ell+1}$ such that λ elements $(\alpha_1 - \alpha_0), \ldots, (\alpha_\lambda - \alpha_0)$ constitute a basis of F over $GF(p)$.

We claim that $a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_i \neq a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_0$ for at least one $1 \leq i \leq \lambda$. The claim implies (6) because $|a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_0| = |a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_i| = |a_1D_1 + \cdots + a_\ell D_\ell|$.

To prove the claim, suppose the contrary. Then $a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_0 = a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_0 + a_{\ell+1}(\alpha_i - \alpha_0)$, for any $1 \leq i \leq \lambda$. Therefore, $a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_0 = a_1D_1 + \cdots + a_\ell D_\ell + a_{\ell+1}\alpha_0 + a_{\ell+1}\sum_{1\leq i\leq\lambda} b_i(\alpha_i - \alpha_0)$ for arbitrary $b_1, \ldots, b_\lambda \in GF(p)$. But since the latter sum sweeps the entire F, we get a contradiction.

b) Let $L = a_1X_1 + a_2X_2 + a$. If $|D_1|, |D_2| > q/2$, then for any $\alpha \in F$ the sets $a_1D_1 + a \subseteq F$ and $\alpha - a_2D_2 \subseteq F$ have nonempty intersection. This implies $\alpha \in a_1D_1 + a_2D_2 + a$.

3 Rigid rank and communication rank of a matrix

In this section we continue our search for assumptions on a product $\Pi = \prod_i L_i$ that ensure its vanishing almost everywhere on D^n. We already know one class of non-vanishing products of large rank: these are those products in which L_i do not represent 0 on D^n, and we know from Lemma 2.6 that L_i then must have only $O(1)$ variables each. This class of "bad" products can be clearly further extended to products of the form

$$\prod_i (L_i' + L_i''), \qquad (7)$$

where L_i' do not represent 0, and $\prod_i L_i''$ has a low rank. Our eventual goal is to show that this example encompasses already essentially *all* non-vanishing products.

The fact that a product can *not* be represented in the form (7), where L_i''s depend on few variables and $rk(\prod_i L_i'')$ is low, is clearly akin to the standard matrix rigidity function $R_A(r)$ [16, 6], and in this section we give its satisfactory description in terms of *internal* properties of the matrix.

Let $A = (a_{ij})$ be a $k \times n$ matrix over some (not necessarily finite in this section) field, and let $m \geq 0$ be an integer. For subsets $I \subseteq \{1, \ldots, k\}$, $J \subseteq \{1, \ldots, n\}$ we denote by A_{IJ} the submatrix of A formed by its rows from I and the columns from J. For $i \in \{1, \ldots, k\}$, A_{iJ} is the corresponding subrow of the ith row.

Definition 3.1 *The m-rigid rank $rrk_m(A)$ of A* is defined as the minimal possible rank of matrices B which differ from A by at most m entries in each row.

Definition 3.2 *The m-communication rank $crk_m(A)$ of A* is defined as the maximal possible number r of its rows $I \subseteq \{1, \dots, k\}$, $|I| = r$ such that there exist pairwise disjoint sets of columns $J_0, \dots, J_m$, also of cardinality r each, with the property that all submatrices A_{IJ_ℓ}, $0 \le \ell \le m$ are non-singular.

Notice that both rrk_m and crk_m are not invariant in general with respect to transposing the matrix A. Obviously, rrk_0 and crk_0 coincide with the usual rank. The connection with the standard rigidity function $R_A(r)$ (the minimal overall number of changes in A required to reduce its rank to r) is provided by the inequality

$$R_A(rrk_m(A)) \le km.$$

The term "communication" is suggested by the resemblance to the common (worst-case partition) scenario in communication complexity.

The following lemma relates the rigid and communication ranks.

Lemma 3.3 $rrk_m(A) \le (m+2)\ crk_m(A) \le (m+2)(m+1) rrk_m(A)$.

Proof of the left inequality. Choose $I, J_0, \dots, J_m$ accordingly to Definition 3.2, so that $|I| = |J_0| = \dots = |J_m| = r = crk_m(A)$. Denote $J = \{1, \dots, n\} \setminus (J_0 \cup \dots \cup J_m)$.

Take any row $\rho \in \{1, \dots, k\} \setminus I$. For every $0 \le \ell \le m$ there exists a unique linear combination $A_{\rho, J_\ell} = \sum_{i \in I} \alpha_i^{(\ell)} A_{i, J_\ell}$. Consider the set $J'_\ell \subseteq J$ consisting of all the columns $j \in J$ such that $a_{\rho j} - \sum_{i \in I} \alpha_i^{(\ell)} a_{ij} \ne 0$. Observe that for any $j \in J'_\ell$, the $(r+1) \times (r+1)$ matrix $A_{I \cup \{\rho\}, J_\ell \cup \{j\}}$ is non-singular.

We claim that $|J'_{\ell_0}| \le m$ for some $0 \le \ell_0 \le m$. Assuming the contrary, one can sequentially for $\ell = 0, \dots, m$ pick pairwise distinct $j_\ell \in J'_\ell$. Then all $(r+1) \times (r+1)$ matrices $A_{I \cup \{\rho\}, J_\ell \cup \{j_\ell\}}$, $0 \le \ell \le m$ are non-singular, that contradicts to the equality $r = crk_m(A)$.

Now take $(m+2)r$ n-dimensional vectors, among which there are r rows $A_{i, \{1, \dots, n\}}$, $i \in I$ and $(m+1)r$ unit vectors $e_j = (0, \dots, 0, 1, 0, \dots, 0)$ (where one is located at j-th position) for $j \in \bigcup_{0 \le \ell \le m} J_\ell$. To complete the proof of the left inequality in the lemma, it suffices to show that each row of A equals to a suitable linear combination of these $(m+2)r$ vectors up to at most m entries.

This is obvious for the rows $A_{i,\{1,\ldots,n\}}$, $i \in I$.

Take $\rho \in \{1,\ldots,k\} \setminus I$ and utilize the notation introduced above. Then the difference $A_{\rho,\{1,\ldots,n\}} - \sum_{i\in I} \alpha_i^{(\ell_0)} A_{i,\{1,\ldots,n\}}$ equals to a linear combination of $(m+1)r$ vectors e_j up to at most m entries from J'_{ℓ_0}.

Proof of the right inequality. Again we denote $r = crk_m(A)$ and choose $I, J_0, \ldots, J_m$ accordingly to Definition 3.2. Choose a matrix B accordingly to Definition 3.1 so that $rk(B) = rrk_m(A)$, and mark all the entries at which A and B differ. There are at most (rm) marked entries in the rows from I. Therefore, for some $0 \le \ell \le m$ at most $\frac{rm}{m+1}$ of these entries are located in A_{I,J_ℓ}. This implies $rk(B) \ge rk(B_{I,J_\ell}) \ge \frac{r}{m+1}$ and completes the proof of the right inequality.

4 Products of large communication rank vanish almost everywhere

In this section we show that products of large *communication* rank do vanish almost everywhere on D^n. In combination with Lemma 3.3, this will imply that every non-vanishing product must be representable in the form (7), where L'_i depends on few variables, and $\prod_i L''_i$ has low (ordinary) rank.

Throughout the section, we fix a product of linear functions $\Pi = \prod_i L_i$ and a subset $D \subseteq F = GF(p^\lambda)$ of cardinality $d > p^{\lambda-1}$. Π is viewed as a quasi-boolean function $\Pi : D^n \to F$ (see Section 2). Let $L_i = \sum_{1\le j\le n} a_{ij}X_j + a_i$ where $a_{ij}, a_i \in F$. By the m_1*-communication rank* $crk_{m_1}(\Pi)$ *of* Π we mean the m_1-communication rank of the matrix $A = (a_{ij})$ of its coefficients (thus, the free coefficients of L_i are excluded). Since $d > p^{\lambda-1}$, the linear d-sweep $\ell s(d)$ is defined and does not exceed $q - d + 1$ by Lemma 2.6 a); we denote it by m.

Lemma 4.1

$$Pr[x \in D^n : \Pi(x) \ne 0] \le \exp(-\Omega(crk_{m-1}(\Pi))).$$

Proof. Let $r = crk_{m-1}(\Pi)$. Selecting r rows from the matrix (a_{ij}) in accordance with Definition 3.2, we can assume w.l.o.g. that Π is the product of just r linear functions. Moreover, varying the specifications from D for all the variables corresponding to the complement $\{1, 2, \ldots, n\} \setminus (J_0 \cup \ldots \cup J_{m-1})$, we can assume w.l.o.g. that the matrix of the coefficients $A = (a_{ij})$ is of size $r \times rm$ and consists of m non-singular $r \times r$ submatrices $A^{(1)}, \ldots, A^{(m)}$. Since the bound claimed in the lemma will be proved for an arbitrary specification, this would imply the lemma by the standard averaging argument.

For each $1 \le \ell \le m$ denote by x^ℓ a random vector from D^r, and denote by $x^{\ell 1}, \ldots, x^{\ell d}$ d independent copies of x^ℓ; thus, all dm vectors $x^{\ell\sigma}$ $(1 \le \ell \le m,\ 1 \le \sigma \le d)$ are picked out independently. Our first purpose is to prove the following bound on the probability:

$$Pr[\forall\ \sigma_1, \ldots, \sigma_m \in \{1, \ldots, d\}\ \Pi(x^{1,\sigma_1}, \ldots, x^{m,\sigma_m}) \neq 0] \le \exp(-\Omega(r)). \quad (8)$$

We show by induction on $\ell \le m$ that for suitable constants $\gamma_\ell > 1,\ \delta_\ell > 0$ with probability greater than $1 - \exp(-\delta_\ell r)$ there exist $s \ge r/\gamma_\ell$ rows $I \subseteq \{1, \ldots, r\}$ in A such that for every $1 \le t \le \ell$ all d values of the vectors $A^{(t)}x^{t1}, \ldots, A^{(t)}x^{td}$ in every of these rows are pairwise distinct.

The base for $\ell = 0$ is obvious. For the inductive step from ℓ to $\ell + 1$, we will treat the newly introduced vectors $A^{(\ell+1)}x^{\ell+1,1}, \ldots, A^{(\ell+1)}x^{\ell+1,d}$ by an internal induction on $\sigma < d$. More specifically, we show, increasing σ one by one that for suitable constants $\gamma'_\sigma > 1,\ \delta'_{\ell+1,\sigma} > 0$ with probability greater than $1 - \exp(-\delta'_{\ell+1,\sigma} r)$ there exist $s' \ge s/\gamma'_\sigma$ rows $I' \subseteq I$ in $A^{(\ell+1)}$ such that all the entries of the vectors $A^{(\ell+1)}x^{\ell+1,1}, \ldots, A^{(\ell+1)}x^{\ell+1,\sigma}$ in every row from I' are pairwise distinct. Suppose that for some value of σ we already have such an I' and denote by $\rho : F^r \to F^{s'}$ the projection onto the coordinates from I'.

The number of vectors in $F^{s'}$ such that at most s'/γ (for a certain constant $\gamma > 1$) its entries differ from all the corresponding entries of the vectors $\rho A^{(\ell+1)}x^{\ell+1,1}, \ldots, \rho A^{(\ell+1)}x^{\ell+1,\sigma}$ is equal to

$$\sum_{k=0}^{s'/\gamma} \binom{s'}{k} \sigma^{s'-k}(q - \sigma)^k. \quad (9)$$

We claim that for any fixed $v \in F^{s'}$,

$$Pr\left[\rho A^{(\ell+1)}x^{\ell+1,\sigma+1} = v\right] \le d^{-s'}. \quad (10)$$

Indeed, select s' columns J of the matrix $A^{(\ell+1)}$ such that $A^{(\ell+1)}_{I',J}$ is non-singular. Varying specifications from D for all the variables not in J, and noticing that for every such specification all the $d^{s'}$ vectors $\rho A^{(\ell+1)} x^{\ell+1,\sigma+1}$ are pairwise distinct, we get (10).

Combining (9) and (10), we see that the probability that at most s'/γ (for a certain constant $\gamma > 1$) entries of the vector $\rho A^{(\ell+1)} x^{\ell+1,\sigma+1}$ differ from all the corresponding entries of the vectors $\rho A^{(\ell+1)} x^{\ell+1,1}, \ldots, \rho A^{(\ell+1)} x^{\ell+1,\sigma}$, does not exceed

$$d^{-s'} \sum_{k=0}^{s'/\gamma} \left(\begin{array}{c} s' \\ k \end{array} \right) \sigma^{s'-k} (q-\sigma)^k < \left(\frac{\sigma}{d}\right)^{s'} \sum_{k=0}^{s'/\gamma} \left(\begin{array}{c} s' \\ k \end{array} \right) (q-\sigma)^k. \qquad (11)$$

Since $\sigma \leq d-1$, for large enough constant γ there exists a constant $\eta > 0$ such that the latter expression is less than $\exp(-\eta s')$.

Thus, if we set $\gamma'_{\sigma+1} = \gamma'_\sigma \gamma$ (so that $\gamma'_\sigma = \gamma^\sigma$, $\gamma_\ell = \gamma^{d\ell}$ and $s' = r/\gamma^{d\ell+\sigma}$), the probability that the entries of the vectors

$$\rho A^{(\ell+1)} x^{\ell+1,1}, \ldots, \rho A^{(\ell+1)} x^{\ell+1,\sigma+1}$$

are pairwise distinct in at least $s/\gamma'_{\sigma+1}$ rows is greater than

$$(1 - \exp(-\delta'_{\ell+1,\sigma} r))(1 - \exp(-\eta s')) \geq 1 - \exp(-\delta'_{\ell+1,\sigma+1} r)$$

for an appropriate $\delta'_{\ell+1,\sigma+1} > 0$. This completes the inner induction (on σ), to complete the external induction (on ℓ) put $\delta_{\ell+1} = \delta'_{\ell+1,d}$.

Now, to prove (8), we simply observe that if there exists a single row i in which for every $1 \leq \ell \leq m$ all d entries $A^{(\ell)} x^{\ell 1}, \ldots, A^{(\ell)} x^{\ell d}$ are pairwise distinct, then also $\exists \sigma_1, \ldots, \sigma_m \in \{1, \ldots, d\}$ such that $\Pi(x^{1,\sigma_1}, \ldots, x^{m,\sigma_m}) = 0$. Indeed, since the i-th linear function in the product Π equals to $(A^{(1)} x^1)_i + \cdots + (A^{(m)} x^m)_i + a_i$, Definition 2.5 of the linear d-sweep shows that there exist $\sigma_1, \ldots, \sigma_m \in \{1, \ldots, d\}$ such that $(A^{(1)} x^{1,\sigma_1})_i + \cdots + (A^{(m)} x^{m,\sigma_m})_i + a_i = 0$, which completes the proof of (8).

Our next goal is to prove an inequality for expectations in a rather general probabilistic-theoretical setting:

$$E\left[\prod_{\sigma_1,\ldots,\sigma_m \in \{1,\ldots,d\}^m} g(Y^{1,\sigma_1}, \ldots, Y^{m,\sigma_m})\right] \geq (E[g(Y^1, \ldots, Y^m)])^{d^m}, \qquad (12)$$

where $Y^1, \ldots, Y^m$ are independent random variables, $Y^{\ell 1}, \ldots, Y^{\ell d}$ are independent copies of Y^ℓ, $1 \leq \ell \leq m$, and g is a real-valued nonnegative function.

Taking as $Y^1, \ldots, Y^m$ the vectors $x^1, \ldots, x^m$, and as $g(Y^1, \ldots, Y^m)$ the characteristic function of the predicate $\Pi(x^1, \ldots, x^m) \neq 0$, we infer Lemma 4.1 from (12) together with (8).

To prove (12) denote

$$E\left[\prod_{\sigma_{\ell+1},\ldots,\sigma_m \in \{1,\ldots,d\}^{m-\ell}} g(Y^1, \ldots, Y^\ell, Y^{\ell+1,\sigma_{\ell+1}}, \ldots, Y^{m,\sigma_m})\right]$$

by e_ℓ. We show for every $1 \leq \ell \leq m$ that $e_{\ell-1} \geq e_\ell^d$, which will imply $e_0 \geq e_m^{d^m}$, i.e. (12).

For any (fixed for the time being) tuple

$$y = (y^1, \ldots, y^{\ell-1}, y^{\ell+1,1}, \ldots, y^{\ell+1,d}, \ldots, y^{m1}, \ldots, y^{md})$$

denote the expectation

$$E\left[\prod_{\sigma_{\ell+1},\ldots,\sigma_m \in \{1,\ldots,d\}^{m-\ell}} g(y^1, \ldots, y^{\ell-1}, Y^\ell, y^{\ell+1,\sigma_{\ell+1}}, \ldots, y^{m,\sigma_m})\right]$$

by $e_\ell(y)$. Then

$$e_\ell = E[e_\ell(Y^1, \ldots, Y^{\ell-1}, Y^{\ell+1,1}, \ldots, Y^{\ell+1,d}, \ldots Y^{m1}, \ldots, Y^{md})].$$

On the other hand we have

$$\begin{aligned} & E\left[\prod_{\sigma_\ell,\sigma_{\ell+1},\ldots,\sigma_m \in \{1,\ldots,d\}^{m-\ell+1}} g(y^1, \ldots, y^{\ell-1}, Y^{\ell,\sigma_\ell}, y^{\ell+1,\sigma_{\ell+1}}, \ldots, y^{m,\sigma_m})\right] \\ & = E\Bigg[\prod_{\sigma_\ell \in \{1,\ldots,d\}} \Bigg(\prod_{\sigma_{\ell+1},\ldots,\sigma_m \in \{1,\ldots,d\}^{m-\ell}} g(y^1, \ldots, y^{\ell-1}, \\ & \qquad Y^{\ell,\sigma_\ell}, y^{\ell+1,\sigma_{\ell+1}}, \ldots, y^{m,\sigma_m})\Bigg)\Bigg] \\ & = (e_\ell(y))^d, \end{aligned}$$

the latter equality holds since $Y^{\ell 1}, \ldots, Y^{\ell d}$ are independent. Taking the expectations against all tuples y, we get

$$e_{\ell-1} = E[(e_\ell(Y^1, \ldots, Y^{\ell-1}, Y^{\ell+1,1}, \ldots, Y^{\ell+1,d}, \ldots, Y^m, \ldots, Y^{md}))^d].$$

Due to Jensen's inequality ($E[Z^d] \geq (E[Z])^d$) we conclude that the latter expectation is greater or equal to $(E[e_\ell(Y)])^d = e_\ell^d$, which completes the proof of (12) and thereby, Lemma 4.1.

5 Approximating depth 3 arithmetic circuits by (N, r)-sparse polynomials

In Sections 2-4 we proved that every product Π that can *not* be represented in the form (7), where L'_i depend on $O(1)$ variables, and $\prod_i L''_i$ has low rank, is sufficiently well approximated by 0. In this section we complete the analysis and treat the products that *can* be represented so. Essentially, we are able to approximate them by (sums of moderate number of) the products $\prod_i L_i$ in which every L_i does not represent 0 on D^n. In particular, in the case $D = F^*$ we get an approximation by sparse polynomials of a certain special form.

More specifically, we have the following theorem.

Theorem 5.1 *Let $F = GF(p^\lambda)$, $D \subseteq F$ be such that $|D| > p^{\lambda-1}$, $\Pi = \prod_i L_i$ be a product of linear functions, and r be any threshold.*

a) There exists a function g of the form

$$g = \sum_{\nu=1}^{\exp(O(r))} g_\nu \prod_i L_{\nu i}, \tag{13}$$

where g_ν are products of at most $O(r)$ linear functions each, $L_{\nu i}$ do not have zeros in D^n, and

$$Pr\,[x \in D^n : \Pi(x) \neq g(x)] \leq \exp(-\Omega(r)). \tag{14}$$

b) If $|D| > p^\lambda/2$, then we can additionally require that $L_{\nu i}$ in (13) have the special form

$$L_{\nu i} = (X_{j_i} - z_{\nu i}), \tag{15}$$

where X_{j_i} is a variable depending only on i, and $z_{\nu i} \in F \setminus D$.

Proof. We give a complete proof of part b) (as this is the part which is really used in the next section), and then sketch how to extend the argument to prove part a).

Since $d = |D| > p^\lambda/2$, we have $\ell s(d) = 2$ by Lemma 2.6 b). If $rrk_1(\Pi) \geq r$ then, according to Lemma 3.3, $crk_1(\Pi) \geq r/3$ and, due to Lemma 4.1, $Pr\,[x \in D^n : \Pi(x) \neq 0] \leq \exp(-\Omega(r))$, so we can simply let $g = 0$. Henceforth, we assume that $rrk_1(\Pi) \leq r$. Choose r linear forms $(\phi_1, \ldots, \phi_r)$ and variables X_{j_i} such that each linear function L_i can be represented as an F-linear combination

$$L_i = \sum_{s=1}^{r} a_{is}\phi_s + b_i X_{j_i} + c_i.$$

By Lagrange interpolation,

$$\Pi = \sum_{\alpha \in F^r} \prod_{\substack{1 \le s \le r \\ \alpha'_s \in F \setminus \{\alpha_s\}}} \frac{\phi_s - \alpha'_s}{\alpha_s - \alpha'_s} \cdot \prod_i (b_i X_{j_i} + \sum_{s=1}^{r} a_{is}\alpha_s + c_i). \tag{16}$$

Let $\Pi^{(\alpha)} = \prod_{b_i \neq 0}(X_{j_i} - z_{\alpha i})$, $z_{\alpha i} = -(\sum_{s=1}^{r} a_{is}\alpha_s + c_i)/b_i$ be an individual term in this sum (up to a multiplicative constant). Let

$$R = \{X_{j_i} \mid \exists i (b_i \neq 0 \wedge z_{\alpha i} \in D)\}.$$

Case 1. $|R| < (2q \log q)r$.
Choose arbitrarily $z \notin D$, and let

$$\Pi'^{(\alpha)} = \prod_{\substack{b_i \neq 0 \\ z_{\alpha i} \notin D}} (X_{j_i} - z_{\alpha i}) \cdot \prod_{\substack{b_i \neq 0 \\ z_{\alpha i} \in D}} (X_{j_i} - z).$$

Then the products $\Pi'^{(\alpha)}$ have the form $\prod_i L_{\nu i}$ required in (13), (15). On the other hand, the term $\Pi^{(\alpha)}$ contributes $g_\alpha \Pi'^{(\alpha)}$ to the sum (16), where

$$g_\alpha = \prod_{\substack{1 \le s \le r \\ \alpha'_s \in F \setminus \{\alpha_s\}}} \frac{\phi_s - \alpha'_s}{\alpha_s - \alpha'_s} \cdot \prod_{\substack{b_i \neq 0 \\ z_{\alpha i} \in D}} \frac{X_{j_i} - z_{\alpha i}}{X_{j_i} - z}.$$

The first term $\displaystyle\prod_{\substack{1 \le s \le r \\ \alpha'_s \in F \setminus \{\alpha_s\}}} \frac{\phi_s - \alpha'_s}{\alpha_s - \alpha'_s}$ here is already a product of r linear functions. The second term $\displaystyle\prod_{\substack{b_i \neq 0 \\ z_{\alpha i} \in D}} \frac{X_{j_i} - z_{\alpha i}}{X_{j_i} - z}$ is equal to $\displaystyle\prod_{X_j \in R} g_{\alpha j}(X_j)$, where $g_{\alpha j}$ are some functions in one variable on D; we can represent them as degree $O(1)$ polynomials in one variable. This implies that $\displaystyle\prod_{X_j \in R} g_{\alpha j}(X_j)$ and, henceforth, g_α can be represented as sums of $\exp(O(r))$ products with $O(r)$ linear functions in each.

Case 2. $|R| \geq (2q\log q)r$

This means that $\Pi^{(\alpha)}$ contains at least $(2q\log q)r$ linear functions $(X_{j_i} - z_{\alpha i})$ with $z_{\alpha i} \in D$ and pairwise distinct j_i's. Hence,

$$Pr\left[x \in D^n : \Pi^{(\alpha)}(x) \neq 0\right] \leq \left(1 - \frac{1}{d}\right)^{(2q\log q)r} \leq q^{-2r}.$$

Since there are at most q^r different α's, we can safely approximate all $\Pi^{(\alpha)}$ corresponding to Case 2 by 0.

Thus, we simply take $\sum_{\{\alpha|\text{ Case }1\}} g_\alpha \cdot \Pi'^{(\alpha)}$ as the required approximation (13) which proves part b).

Proof of part a) (sketch). $m = \ell s(d)$ is defined by Lemma 2.6 a), and, for the same reasons as before, we can assume $rrk_{m-1}(\Pi) \leq r$. We again decompose Π in the form

$$\Pi = \sum_{\alpha \in F^r} \prod_{\substack{1 \leq s \leq r \\ \alpha'_s \in F\setminus\{\alpha_s\}}} \frac{\phi_s - \alpha'_s}{\alpha_s - \alpha'_s} \cdot \Pi^{(\alpha)},$$

where $\Pi^{(\alpha)} = \prod_i L_{\alpha i}$, and $L_{\alpha i}$ this time have at most $(m-1)$ variables each. Let us split $\Pi^{(\alpha)}$ in two parts $\Pi^{(\alpha)} = \Pi^{(\alpha)}_{\neq 0} \cdot \Pi^{(\alpha)}_0$, where $\Pi^{(\alpha)}_{\neq 0}$ consists of those $L_{\alpha i}$ which do not have zeros in D^n, whereas all functions from $\Pi^{(\alpha)}_0$ have there at least one zero. Let h be the maximal number of variables in a linear function from $\Pi^{(\alpha)}_0$; thus, originally we have $h \leq m-1$. We are going to reduce the value of h by applying recursively to $\Pi^{(\alpha)}_0$ a generalization of the analysis from the proof of part b).

Case 1. *There exists a set* $R = \{X_{j_1}, \ldots, X_{j_s}\}$ *of at most* $\gamma_h r$ *variables such that every linear function* $L_{\alpha i}$ *in* $\Pi^{(\alpha)}_0$ *essentially depends on at least one variable from* R. *Here* $\gamma_h > 0$ *are some absolute constants such that* $\gamma_{m-1} \gg \gamma_{m-2} \gg \ldots \gg \gamma_1$.

In this case we apply Lagrange interpolation once more and write down the representation

$$\Pi^{(\alpha)} = \sum_{\beta \in D^s} \prod_{\substack{1 \leq t \leq s \\ \beta'_t \in D\setminus\{\beta_t\}}} \frac{X_{j_t} - \beta'_t}{\beta_t - \beta'_t} \cdot \Pi^{(\alpha\beta)}$$

(valid on D^n), where $\Pi^{(\alpha\beta)}$ is obtained from $\Pi^{(\alpha)}$ by substituting $\beta_1, \ldots, \beta_s$ for the variables $X_{j_1}, \ldots, X_{j_s}$, respectively. In $L_{\alpha i} \in \Pi^{(\alpha)}_0$ this substitution decreases the number of variables (so that now it becomes at most $(h-1)$), and the images of $L_{\alpha i} \in \Pi^{\alpha}_{\neq 0}$ still do not have zeros in D^{n-s}. Thus, we have reduced in Case 1 the product $\Pi^{(\alpha)}$ to $\exp(O(\gamma_h r))$ products $\Pi^{(\alpha\beta)}$ with smaller values of h.

Case 2. *There is no set of variables R described above.*

Select among the linear functions in $\Pi^{(\alpha)}_0$ the maximal possible set

$$L_{\alpha 1}, \ldots, L_{\alpha M}$$

such that no variable occurs in two of them. We claim that $M \geq \gamma_h r/h$. Indeed, otherwise we could take as R the set of all variables occurring in these M functions: $|R| \leq \gamma_h r$ since every $L_{\alpha i}$ has at most h variables. Now, the events $L_{\alpha i}(x) = 0$ $(1 \leq i \leq M)$ are independent, and each of them occurs with probability $\Omega(1)$. Hence,

$$Pr\left[x \in D^n : \Pi^{(\alpha)}_0(x) \neq 0\right] \leq \exp(-\Omega(\gamma_h r)),$$

and at most $\exp(O((\gamma_1 + \cdots + \gamma_{h-1})r))$ such $\Pi^{(\alpha)}_0$ resulted from branching in Case 1 at previous steps. Since $\gamma_{m-1} \gg \gamma_{m-2} \gg \ldots \gg \gamma_1$, we still can approximate all of them by 0.

After applying this recursive procedure $(m-1)$ times, we completely kill our $\Pi^{(\alpha)}_0$ which completes the proof of part a).

Theorem 5.1 looks especially simple in the case $D = F^*$.

Definition 5.2 A polynomial of the form $\sum_{1 \leq \nu \leq N} g_\nu X^{u_\nu}$ where X^{u_ν} is a monomial and $\deg(g_\nu) \leq r$ is called (N, r)*-sparse.*

Lemma 5.3 *For every finite field F there exists a positive constant $\delta_0 = \delta_0(F)$ such that the following holds. For every parameter r and every function $f : (F^*)^n \to F$ computed by an arithmetic depth 3 circuit* (1) *with the complexity $N \leq \exp(\delta_0 r)$, there exists an $(N, O(r))$-sparse polynomial g such that*

$$Pr[x \in (F^*)^n : f(x) \neq g(x)] \leq \exp(-\Omega(r)). \tag{17}$$

Proof. Immediate from Theorem 5.1 b): when $D = F^*$, all $z_{\nu i}$ in (15) must be 0, and (13) simplifies to the $(1, O(r))$-sparse polynomial $g = \left(\sum_{\nu=1}^{\exp(O(r))} g_\nu\right) \cdot \prod_i X_{j_i}$. Summing these approximations over all N gates at the middle level of our circuits, we get (17) (provided δ_0 is smaller than the constant assumed in (14)).

6 Lower bound on the number of non-zeroes for (N, r)-sparse polynomials

Although we are interested mainly in the case $D = F^*$, our main argument in this section is valid for any $D \subseteq F^*$ which is a coset modulo some subgroup in F^*. Recall from Section 2 that in that case the algebra $\mathcal{A}$ of functions $D^n \to F$ has the multiplicative basis of monomials $\mathcal{M} = \{X^u\}_{0 \le u_1, \dots, u_n < d}$, $d = |D|$. One can rephrase Definition 5.2 using the (Hamming metric ρ) geometric language in $\mathcal{M}$ (see Section 2): if a polynomial g is (N, r)-sparse then supp (g) lies in a union of N balls each of radius r (centered at X^{u_ν}). Conversely, if supp (g) can be covered by N balls of radius r, then g is $(N, (d-1)r)$-sparse.

The following lemma provides a lower bound on the number of nonzeroes of an (N, r)-sparse polynomial.

Lemma 6.1 *Let $D \subseteq F^*$ be a coset modulo some subgroup in F^*, $f \in \mathcal{A}$ be a (N, r)-sparse polynomial, such that for a certain R the support supp (f) contains a monomial X^{u_0} such that $\rho(X^{u_0}, X^u) \ge R$ for any other monomial $X^u \in$ supp (f). Then*

$$Pr[x \in D^n : f(x) \neq 0] \ge \exp\left(-O\left(\frac{n}{R}\left(\frac{r^2}{R} + \log(N + n)\right)\right)\right). \qquad (18)$$

Proof. Replacing f by $f \cdot (a\, X^{u_0})^{-1}$ where $a \in F^*$ is the coefficient at the monomial X^{u_0} in f (and taking into account that $\mathcal{M}$ is a multiplicative basis of $\mathcal{A}$), we can assume w.l.o.g. that $u_0 = 1$ and that the free term of f is 1. We keep the notation X^{u_ν} $(1 \le \nu \le N)$ from Definition 5.2 for the centers of the radius r balls that cover supp (f). Note that if the ball centered at X^{u_ν} contains at least one monomial X^u from supp (f) other than 1, then

$$\rho(1, X^{u_\nu}) \ge \rho(1, X^u) - \rho(X^{u_\nu}, X^u) \ge R - r. \qquad (19)$$

Put

$$s = \left\lfloor \frac{Cn}{R}(r + (R\log(N + n))^{1/2}) \right\rfloor \qquad (20)$$

for an appropriate sufficiently large constant C which will be specified later. W.l.o.g. we can assume that $s < \frac{d-1}{2d}n$, because otherwise (18) is trivial.

Consider the sphere $S \subseteq \mathcal{M}$ of the radius $\frac{d-1}{d}n - s$ centered at 1 (w.l.o.g. we can assume that $d|n$), i.e. $S = \left\{X^u \mid \rho(1, X^u) = \frac{d-1}{d}n - s\right\}$. Since

$$|S| = (d-1)^{\frac{d-1}{d}n - s} \binom{n}{\frac{d-1}{d}n - s} \ge d^n \exp\left(-O\left(\frac{s^2}{n}\,\frac{d^2}{d-1} + \log n\right)\right),$$

we notice the bound on the probability

$$Pr[X^u \in \mathcal{M} : X^u \in S] \geq \exp\left(-O\left(\frac{s^2}{n} + \log n\right)\right),$$

and the right-hand side of the latter inequality has the same order of magnitude as the right-hand side of the desired inequality (18).

Let us view a polynomial from $\mathcal{A}$ as a row of its d^n coefficients at the monomials from $\mathcal{M}$. We supposed proved that one can pick out at least half of the elements X^u from S such that the matrix composed of the rows $X^u f$ for these $X^u \in S$ contains the unit submatrix just in the set of columns X^u. That means that the dimension of the ideal $(f) \subseteq \mathcal{A}$ is greater or equal to $|S|/2$. Then Lemma 2.1 b) would imply (18) due to the bound on $|S|$ obtained above.

We call $X^u \in S$ *remote* if

$$\rho(X^u, X^{u_\nu}) > \frac{d-1}{d}n - s + r$$

for all those ν for which the ball centered at X^{u_ν} covers at least one non-trivial monomial from $\mathrm{supp}\,(f)$. Observe that if we compose the above matrix of the rows $X^u f$ for all remote X^u then it contains the desired unit submatrix. Indeed, any monomial $X^{u'} \in \mathrm{supp}\,(f)$, $X^{u'} \neq 1$ belongs to a ball with the radius r centered at X^{u_ν} for some ν. Therefore,

$$\rho(X^u, X^{u'}) \geq \rho(X^u, X^{u_\nu}) - \rho(X^{u_\nu}, X^{u'}) > \frac{d-1}{d}n - s.$$

Hence for any $X^{u_0} \in S$ (not necessarily remote) we have

$$\begin{aligned}\rho(X^u, X^{u_0}X^{u'}) &\geq \rho(X^u, X^{u'}) - \rho(X^{u'}, X^{u_0}X^{u'}) = \rho(X^u, X^{u'}) - \rho(1, X^{u_0}) \\ &= \rho(X^u, X^{u'}) - \left(\frac{d-1}{d}n - s\right).\end{aligned}$$

Thus, $\rho(X^u, X^{u_0}X^{u'}) > 0$ which means that the row $X^{u_0}f$ can not have a non-zero entry in any column X^u for a *remote* X^u, except for appearance of an entry equal to 1 in the column X^{u_0}.

In order to justify the remaining goal, i.e., to show that at least half of the elements $X^u \in S$ are remote, it suffices to prove for every ν the bound on the probability

$$Pr\left[X^u \in S : \rho(X^u, X^{u_\nu}) \leq \frac{d-1}{d}n - s + r\right] \leq \frac{1}{2N}. \tag{21}$$

A random monomial X^u in S can be constructed in two steps. First, we choose a random $I \subseteq \{1, \ldots, n\}$ of cardinality $\frac{d-1}{d}n - s$. Then we pick a random monomial X^u in $S_I \subseteq S$, where $S_I = \{X^u : u_i \neq 0$ if and only if $i \in I\}$. Accordingly to this construction, we split the proof of (21) in two parts. Denote $I_\nu = \{i : (u_\nu)_i \neq 0\}$, and let $w = |I_\nu|$. First, we show

$$Pr\left[|I \cap I_\nu| \geq w \cdot \left(\frac{d-1}{d} - \frac{s}{2n}\right)\right] \leq \frac{1}{4N}. \tag{22}$$

Then we show that for every *individual* I such that $|I \cap I_\nu| < w \cdot \left(\frac{d-1}{d} - \frac{s}{2n}\right)$, we have

$$Pr\left[X^u \in S_I : \rho(X^u, X^{u_\nu}) \leq \frac{d-1}{d}n - s + r\right] \leq \frac{1}{4N}. \tag{23}$$

(22) and (23) will clearly imply (21).

The best way to avoid tedious calculations in proving (22) is to replace I by its Bernoulli variant $\tilde{I}$, i.e., every event $i \in \tilde{I}$ occurs with probability $\frac{d-1}{d} - \frac{s}{n}$, and these events are independent for different i. Since $E[|\tilde{I}|] = \frac{d-1}{d}n - s$, we get

$$Pr\left[|\tilde{I}| = \frac{d-1}{d}n - s\right] \geq n^{-1},$$

and all I with $|I| = \frac{d-1}{d}n - s$ are attained by $\tilde{I}$ with the same probability. Hence, for proving (22) it would suffice to prove

$$Pr\left[|\tilde{I} \cap I_\nu| \geq w \cdot \left(\frac{d-1}{d} - \frac{s}{2n}\right)\right] \leq \frac{1}{4nN}. \tag{24}$$

However, $|\tilde{I} \cap I_\nu|$ is equal to the sum of w Bernoulli variables $\xi_1, \ldots, \xi_w$ with $Pr[\xi_i = 1] = \frac{d-1}{d} - \frac{s}{n}$. We can assume $R \geq 2r$ since otherwise the lemma becomes trivial due to the presence of the term $\frac{nr^2}{R^2}$ in (18). Since $w \geq R - r$ by (19), this implies

$$w \geq R/2. \tag{25}$$

Now we simply apply Chernoff inequality [3] for estimating the probability that $|\tilde{I} \cap I_\nu|$ deviates from its expectation by at least $\frac{sw}{2n}$, and this gives us (24):

$$\begin{aligned} Pr\left[|\tilde{I} \cap I_\nu| \geq w \cdot \left(\frac{d-1}{d} - \frac{s}{2n}\right)\right] &\leq \exp\left(-\Omega\left(\frac{s^2 w}{n^2}\right)\right) \\ &\leq \exp\left(-\Omega\left(\frac{s^2 R}{n^2}\right)\right) \leq \frac{1}{4nN} \end{aligned}$$

if the constant C in (20) is large enough. Thus, (22) is also proved.

For proving (23), let us notice that

$$\begin{aligned}\rho(X^u, X^{u_\nu}) &= |I| + |I_\nu| - |I \cap I_\nu| - |\{ i \in I \cap I_\nu \mid u_i = (u_\nu)_i \}| \\ &> \frac{d-1}{d}n - s + w - w\left(\frac{d-1}{d} - \frac{s}{2n}\right) \\ &\qquad -|\{ i \in I \cap I_\nu \mid u_i = (u_\nu)_i \}| \\ &= \frac{d-1}{d}n + \frac{w}{d} - s + \frac{ws}{2n} - |\{ i \in I \cap I_\nu \mid u_i = (u_\nu)_i \}|.\end{aligned}$$

Provided $C > 4$ in (20), (25) implies $\frac{ws}{2n} \geq r$. Thus, it is sufficient to show that

$$Pr\left[X^u \in S_I : |\{ i \in I \cap I_\nu \mid u_i = (u_\nu)_i \}| \geq \frac{w}{d}\right] \leq \frac{1}{4N}. \tag{26}$$

But $|\{i \in I \cap I_\nu : u_i = (u_\nu)_i\}|$ is once more the sum of $|I \cap I_\nu| \leq w\left(\frac{d-1}{d} - \frac{s}{2n}\right)$ Bernoulli variables attaining 1 with probability $\frac{1}{d-1}$ each. Applying once more Chernoff inequality, we get (26). This also completes the proof of (23), (21) and Lemma 6.1.

Now we are ready to complete the proof of Theorem 2.2. Let $D = F^*$, $\mathcal{N} \subseteq \{X_1^{u_1} \cdot \ldots \cdot X_n^{u_n}\}_{0 \leq u_1, \ldots, u_n < q-1}$ satisfy (4), where the constant ϵ will be specified later, and $f : (F^*)^n \to F$ be a function with $\mathrm{supp}\,(f) = \mathcal{N}$ computed by a depth 3 circuit (1) with N gates at the middle level. Suppose that $N < |\mathcal{N}|$.

Let r be a parameter satisfying the restriction

$$\frac{1}{6C_1} R(\mathcal{N}) \geq r \geq \frac{1}{\delta_0} \log |\mathcal{N}|, \tag{27}$$

where δ_0 is the constant from Lemma 5.3, and C_1 is the constant assumed in the expression $O(r)$ in that lemma (the exact value of r will be specified later). Applying Lemma 5.3, we will find an $(N, C_1 r)$-sparse polynomial g such that (17) holds.

Let us look more closely at the difference $f - g$. For every ball B in our collection of N balls of radius $C_1 r$ covering $\mathrm{supp}\,(g)$ there exists at most one $X^u \in \mathrm{supp}\,(f)$ with $\rho(X^u, B) < R/3$, $R = R(\mathcal{N})$. Indeed, if we had two different monomials $X^u, X^{u'}$ with this property in $\mathrm{supp}\,(f)$, then we would also have

$$\rho(X^u, X^{u'}) \leq \rho(X^u, B) + 2C_1 r + \rho(X^{u'}, B) < R,$$

the latter inequality following from the left-hand side of (27), and that would contradict the definition of R. Since $N < |\mathcal{N}|$, by the pigeon-hole-principle there exists $X^{u_0} \in \mathrm{supp}\,(f)$ such that $\rho(X^{u_0}, X^u) \geq R/3$ for every $X^u \in \mathrm{supp}\,(g)$ and every $X^u \in \mathrm{supp}\,(f)$ other than X^{u_0}. Clearly, $f - g$ is $(2|\mathcal{N}|, C_1 r)$-sparse. Thus, we can apply to it Lemma 6.1 (with $N := 2|\mathcal{N}|$, $r := C_1 r$ and $R := R/3$) and conclude

$$Pr[x \in D^n : (f-g)(x) \neq 0] \geq \exp\left(-O\left(\frac{n}{R}\left(\frac{r^2}{R} + \log(|\mathcal{N}| + n)\right)\right)\right).$$

Comparing this with (17), we find

$$\frac{n}{R}\left(\frac{r^2}{R} + \log(|\mathcal{N} + n|)\right) \geq \delta_2 r \tag{28}$$

for some constant $\delta_2 > 0$.

We let now

$$\delta_3 = \min\left\{\frac{1}{6C_1}, \frac{\delta_2}{2}\right\}$$

and

$$r = \frac{\delta_3 R^2}{n}.$$

Our choice of r already ensures the left-hand side of (27), as well as the bound $\frac{nr^2}{R^2} \leq \frac{\delta_2 r}{2}$. Comparing this with (28), we get

$$\log(|\mathcal{N} + n|) \geq \frac{\delta_2 r R}{2n} = \frac{\delta_2 \delta_3 R^3}{2n^2}. \tag{29}$$

Thus, if we choose the constant $\epsilon = \epsilon(F)$ in Theorem 2.2 in such a way that $\epsilon < \delta_0 \delta_3$ and $\epsilon < \frac{\delta_2 \delta_3}{2}$, then (4) will also ensure the lower bound on r in (27), and lead to the contradiction with (29). This contradiction completes the proof of Theorem 2.2.

References

[1] M.Agrawal, E.Allender, S.Datta, On TC^0, AC^0, and arithmetic circuits. ECCC Report TR97-016.

[2] W. Baur, V.Strassen. The complexity of partial derivatives. Theor.Comput.Sci., 22, 1983, p.317–330.

[3] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Annals of Math. Stat., 23, 1952, p. 493–509.

[4] J. Friedman, A note on matrix rigidity. Combinatorica, 13, 1993, p. 235–239.

[5] J. von zur Gathen. Feasible arithmetic computations: Valiant's hypothesis. J. Symp. Comput., 4, 1987, p. 137–172.

[6] Д.Ю.Григорьев. Использование понятий отделенности и независимости для доказательства нижних оценок сложности схем. Записки научных семинаров Ленингр. отделения Математического института им. В.А.Стеклова Академии Наук СССР, 60, 1976, p. 38–48 (English transl.: D.Grigoriev. Using the notions of separability and independence for proving lower bounds on the circuit complexity. J. Soviet Math., vol.14, 5, 1980, p.1450–1456.)

[7] Д.Ю.Григорьев. Нижние оценки в алгебраической сложности вычислений. Записки научных семинаров Ленингр. отделения Математического института им. В.А.Стеклова Академии Наук СССР, 118, 1982, p.25–82 (English transl.: D. Grigoriev. Lower bounds in algebraic complexity. J. Soviet Math., 29, 1985, p. 1388–1425.)

[8] D. Grigoriev. Testing shift-equivalence of polynomials by deterministic, probabilistic and quantum machines. Theor. Comput. Sci., 180, 1997, p. 217–228.

[9] D. Grigoriev, M. Karpinski. An exponential lower bound for depth 3 arithmetic circuits. Proc. ACM Symp. Th. Comput., Dallas, 1998, p. 577–582.

[10] D. Grigoriev, A.Razborov. Exponential complexity lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. Proc. IEEE Symp. Found. Comput. Sci., Palo Alto, 1998, p. 269–278.

[11] N. Nisan, A. Wigderson. Lower bound on arithmetic circuits via partial derivatives. Proc. IEEE Symp. Found. Comput. Sci., 1995, p. 16–25.

[12] А. А. Разборов. Нижние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического

сложения. Математические Заметки, 41, 1987, p. 598-607. Engl. transl.: A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. Math. Notes, 41, 1987, p. 333–338.

[13] A. Shpilka, A. Wigderson. Depth-3 arithmetic formulae over fields of characteristic zero, Manuscript, 1998.

[14] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. Proc. ACM Symp. Th. Comput., 1987, p. 77–82.

[15] V. Strassen. Die Berechnungskomplexität von Elementarsymmetrischen Funktionen und von Interpolationskoeffizienten. Numer. Math., 20, 1973, p. 238–251.

[16] L. Valiant. Graph-theoretic arguments in low-level complexity. Proceedings of the 6th MFCS, Lecture Notes in Computer Science, 53, 1977, p. 162-176

[17] L. Valiant. Completeness classes in algebra. Proc. ACM Symp. Th. Comput. 1979, p. 259–261.

[18] G. van der Geer, J, van Lint. Introduction to Coding Theory and Algebraic Geometry. Birkhaüser Verlag, 1988.

[19] J. van Lint. Introduction to Coding Theory. Springer-Verlag, 1982.