An Exponential Lower Bound for Depth 3 Arithmetic Circuits

Dima Grigoriev* Marek Karpinski†

Abstract

We prove the first exponential lower bound on the size of any depth 3 arithmetic circuit with unbounded fanin computing an explicit function (the *determinant*) over an arbitrary finite field. This answers an open problem of [N91] and [NW95] for the case of finite fields. We intepret here arithmetic circuits in the algebra of polynomials over the given field. The proof method involves a new argument on the rank of linear functions, and a group symmetry on polynomials vanishing at certain nonsingular matrices, and could be of independent interest.

Introduction

In this paper we are interested in a fundamental problem of computing functions by unbounded fanin arithmetic circuits of depth 3. Unlike the boolean circuits, general arithmetic circuits of depth 3 are surprisingly powerful. They can compute (via polynomial interpolation) in polynomial size any symmetric function. To date however the best lower bound known for general arithmetic circuit size was only slightly superlinear $\Omega(n \log n)$ ([S73]).

In this paper we prove the first superpolynomial (in fact exponential) size lower bound on depth 3 arithmetic unbounded fanin circuits computing an explicit function, the determinant function, over an arbitrary finite field. In this paper, we interpret the arithmetic circuits in the polynomial algebra over the given field.

The determinant function is especially interesting because of its algebraic *universality* property ([V79]) over arbitrary fields.

We refer a general reader to [L84] and [H77] for all the needed notions used in our proof.

We denote by $F = \mathbb{F}_q$ a finite field with q elements. We shall study fields for $q \geq 3$ (for $q = 2$, the boolean case, the lower bound could be derived from [R87] and [V79]).

We study the representation of $\mathrm{Det} = \mathrm{Det}_n = \sum_\sigma (-1)^{\mathrm{sgn}(\sigma)} X_{1,\sigma(1)} \cdots X_{n,\sigma(n)}$ in the polynomial algebra $F[X_{1,1}, \ldots, X_{n,n}]$ in the form of a depth 3 arithmetic circuit, or equivalently, an expansion:

$$\mathrm{Det} = \sum_{1 \leq \ell \leq N} \prod_m L_{\ell,m} \qquad (1)$$

where each $L_{\ell,m} = \sum_{i,j} a_{i,j}^{(\ell,m)} X_{i,j} + a_0^{(\ell,m)} \in F[X_{1,1}, \ldots, X_{n,n}]$ is a linear function in the variables $X_{1,1}, \ldots, X_{n,n}$. Our purpose is to prove the following exponential lower bound on the size of a representation (1). From this, the lower bound on the size of any depth 3 arithmetic unbounded fanin circuit computing the determinant follows.

Theorem For any $q \geq 3$ there is a constant $\delta > 1$ such that in a representation (1) the number of terms $N = \Omega(\delta^n)$.

*Dept. of Computer Science and Mathematics, Penn State University, University Park. Research partially supported by NSF Grant CCR-9424358. Email: dima@cse.psu.edu

†Dept. of Computer Science, University of Bonn, 53117 Bonn. Research partially supported by the International Computer Science Institute, Berkeley, DIMACS, by the DFG Grant 673/4-1, ESPRIT BR Grants 7079, 21726, and EC-US 030. Email: marek@cs.bonn.edu

Representations of the form (1), but under the restriction that $L_{\ell,m}$ are (homogeneous) linear *forms*, rather than functions, were considered in [G82] (over an arbitrary field), where lower bounds on N were established. The basic idea in [G82] was to design a linear operator on polynomials into matrices which maps a product $\prod_m L_{\ell,m}$ into a matrix of a bounded rank. This approach was also used in [R87]. Later a different method of proving lower bounds on N (again under a similar to [G82] assumption that the degree of each product $\prod_m L_{\ell,m}$, i.e. the number of linear functions in the product, is bounded) was proposed in [NW95]. The core of the method was to estimate the dimension of all the partial derivatives (up to a certain order).

On the other hand, the circuits with a bounded depth (and unbounded fanin) were studied in connection with the boolean AC^0 class and an exponential lower bound on their sizes was proved in [R87], [S87]. The methods in both papers were working just for boolean circuits, and it would be interesting to explore whether they could be extended to arbitrary finite fields (the present authors were not able to do it). These boolean methods imply, in particular, an exponential lower bound on the size of any bounded depth boolean circuit for determinant (and consequently, of any arithmetic circuit over $\mathbb{F}_2$). The representation (1) can be viewed as a depth 3 *arithmetic* circuit. In contrast to boolean circuits we interpret arithmetic circuits (1) as an identity in the polynomial algebra (vs. the algebra of functions over F, see section 1 below). Recently, Razborov [R98] was able to generalize our results to the algebra of functions over F.

An important problem remains open to get lower bounds for representations of the kind (1), or for the more general bounded depth circuits, over the arbitrary fields including zero characteristic.

The rest of the paper is devoted to the proof of the Theorem. In Section 1 we treat the representation (1) and its partial derivatives in the algebra of functions over F and partition the terms $\prod_m L_{\ell,m}$ into two groups, regarding the rank of the family of linear functions $\{L_{\ell,m}\}_m$ being greater or less J than, respectively, a certain integer (threshold). We show that the products $\prod_m L_{\ell,m}$ with a large rank vanish (and moreover with a large multiplicity) everywhere out of a small fraction of points from F^{n^2} (which could be informally viewed as "erroneous" points). For the products with a small rank we estimate from above the dimension of the set of all its derivatives (up to some order) restricted to the algebra of functions over F.

In Section 2 we study linear combinations of minors (of a fixed size) of a matrix vanishing at all the points (in other words, matrices) out of an "erroneous" set, because minors are just the derivatives of Det. Since the full linear group $GL_n(F)$ acts on linear combinations of minors, we show that a small number of shifts by means of elements from $GL_n(F)$ allow to get rid of the "erroneous" set and to obtain a linear combination of minors vanishing at all *nonsingular* matrices. Finally, we prove that it is impossible.

1 A product of linear functions in the algebra of functions over a finite field

Denote by $\mathcal{A}$ the algebra of all functions $f : F^{n^2} \to F$ which can be naturally identified with the quotient algebra

$$F[X_{1,1}, \ldots, X_{n,n}]/(\{X_{i,j}^q - X_{i,j}\}_{1 \leq i,j \leq n})$$

For any set $E \subset F^{n^2}$ of $n \times n$ matrices one can consider (as in [S87]) a quotient algebra $\mathcal{A}_E$ of $\mathcal{A}$ over the ideal of all the functions from $\mathcal{A}$ vanishing everywhere out of E. Obviously, $\dim \mathcal{A}_E = \dim_F \mathcal{A}_E = q^{n^2} - \mid E \mid$. Conversely, any quotient algebra of $\mathcal{A}$ equals to $\mathcal{A}_E$ for a suitable E (we do not use this remark). Talking about some elements from $\mathcal{A}_E$ we mean the images of the elements from $\mathcal{A}$ in the quotient algebra.

Fix a constant $\gamma > 0$ satisfying the inequality $\gamma < q^{-q/2}$. Then there exists a constant β such that

$$q^{q\gamma} < q^{\beta} < \gamma^{-s\gamma} \tag{2}$$

Introduce also a threshold

$$r = [\beta\, n] \tag{3}$$

For the time being we fix a product $\prod_m L_{\ell,m}$ (of linear functions (see (1)). By its rank r_ℓ we mean the dimension of the family of the linear functions $\{L_{\ell,m}\}_m$, in other words, the rank of the matrix of their coefficients $(a_{i,j}^{(\ell,m)}, a_0^{(\ell,m)})$ (which has n^2+1 columns, hence $r_\ell \leq n^2+1$). We treat separately two cases: when the rank r_ℓ is less or greater, respectively, than the threshold r and consider the restriction of the product along with its derivatives onto the space F^{n^2} (in other words, the points defined over F).

Large rank

Let $r_\ell \geq r$. Then the number of points from space F^{n^2} (of all $n \times n$ matrices with the entries from F), at which at most γn among the linear functions $\{L_{\ell,m}\}_m$ vanish, does not exceed

$$q^{n^2-r_\ell}((q-1)^{r_\ell} + \left(\binom{r_\ell}{1}\right)(q-1)^{r_\ell-1} + \ldots + \left(\binom{r_\ell}{\lceil \gamma n \rceil}\right)(q-1)^{r_\ell - \lceil \gamma n \rceil})$$

since one can choose a basis $\mathcal{L}_1, \ldots, \mathcal{L}_{r_\ell}$ of r_ℓ functions among $\{L_{\ell,m}\}_m$ and assign in an arbitrary way the values for $\mathcal{L}_1, \ldots, \mathcal{L}_{r_\ell}$ (among these values at most γn are zeros). A described point will play a role of an "erroneous" point at which all the derivatives of the product $\prod_m L_{\ell,m}$ of the order $\lceil \gamma n \rceil$ may not vanish. The obtained bound can be estimated from above by

$$q^{n^2-r_\ell}\left(\binom{r_\ell}{\lceil \gamma n \rceil}\right)(q-1)^{r_\ell-\lceil \gamma n \rceil}(\gamma n + 1) \tag{4}$$

since the sequence $(q-1)^{r_\ell}, (\binom{r_\ell}{1})(q-1)^{r_\ell-1}, \ldots$ increases until $(\binom{r_\ell}{\lceil r_\ell/q \rceil})(q-1)^{r_\ell - \lceil r_\ell/q \rceil}$ and beyond that decreases (taking into account (3) and the left inequality (2)).

Now we show that (4) can be estimated from above by

$$q^{n^2}\alpha^n \tag{5}$$

for a suitable $\alpha < 1$ depending on q, γ, β. Denote $r = y_0 q\gamma n$ (for an appropriate $y_0 > 1$ (see (3))) and $r_\ell = yq\gamma n$ where $y \geq y_0$. Using Stirling's formula one concludes that (4) is less (up to a factor polynomial in n) than $q^{n^2}(\frac{yq(q-1)}{(yq-1)q})^{yq\gamma n}(\frac{yq-1}{q-1})^{\gamma n}$. It suffices to check that $(\frac{yq(q-1)}{(yq-1)q})^{yq}(\frac{yq-1}{q-1}) < \alpha_1$ for any $y \geq y_0$ and a certain $\alpha_1 < 1$ depending only on q, y_0. The logarithmic derivative $q \log \frac{y(q-1)}{yq-1}$ (over y) of the left side of the latter inequality is negative for any $y > 1$, hence the left side decreases for $y \geq 1$ (for $y = 1$ it equals 1), that proves (5).

Small rank

Now let $r_\ell < r$.

Note that derivatives of all the orders (actually, we are interested just in the order $\lceil \gamma n \rceil$) of the product $\prod_m L_{\ell,m}$ lie in the F-linear hull of the products of the form $\mathcal{L}_1^{i_1} \cdots \mathcal{L}_{r_\ell}^{i_{r_\ell}}$ for all nonnegative integers $i_j, 1 \leq j \leq r_\ell$. When subsequently we restrict these derivatives onto the space F^{n^2}, thus treating them as elements from the algebra $\mathcal{A}$, they would lie in the F-linear hull of the products $\mathcal{L}_1^{i_1} \cdots \mathcal{L}_{r_\ell}^{i_{r_\ell}}$, $0 \leq i_j \leq q-1, 1 \leq j \leq r_\ell$. Therefore, the dimension of the set of these images in $\mathcal{A}$ of the derivatives is less than q^r.

The derivatives of the order $\lceil \gamma n \rceil$ of Det are exactly the minors $M_{I,J}$ of the size $(n - \lceil \gamma n \rceil) \times (n - \lceil \gamma n \rceil)$, where I, J are subsets of the sets of rows and columns, respectively, $|\ I\ |=|\ J\ |= n - \lceil \gamma n \rceil$. We take all the derivatives of the order $\lceil \gamma n \rceil$ of both sides of (1) and subsequently restrict them onto F^{n^2} (thus, treating them as elements from the algebra $\mathcal{A}$). Denote by $E \subset F^{n^2}$ the union of the ("erroneous") sets considered above for all the products from (1) of big ranks. Then the images in the quotient algebra $\mathcal{A}_E$ of taken derivatives vanish for all big rank products and we conclude with the following Lemma (making use also of (5))

Lemma 1 For any $\delta > 1$, if Det has a representation (1) with $N < \delta^n$ then the set of all minors $M = \{M_{I,J}\}_{|I|=|J|=n-\lceil\gamma n\rceil}$ has the dimension less than $\delta^n q^r$ in the quotient algebra $\mathcal{A}_E$ for an appropriate subset $E \subset F^{n^2}$ of the size $|E| \leq q^{n^2}(\delta\alpha)^n$.

Remark. The statement of the Lemma is nontrivial when δ satisfies the following inequalities

$$\delta\alpha < 1;\ \delta q^{\beta} < \gamma^{-2\gamma} \tag{6}$$

The second inequality means that the dimension of the minors from the Lemma is less than the number $(\binom{n}{\lceil\gamma n\rceil})^2$ of all $(n-\lceil\gamma n\rceil)\times(n-\lceil\gamma n\rceil)$ minors (due to the Stirling's formula and (3)). Furthermore, any small enough $\delta > 1$ satisfies (6) due to the right inequality (2), and any such δ one could use in the statement of the Theorem (see above).

Henceforth, we assume that δ satisfies (6).

2 Group symmetry on polynomials vanishing at matrices

Denote by $\mathcal{H}$ (being isomorphic to $F^{\binom{n}{\lceil\gamma n\rceil}^2}$) the F-space of all the linear combinations of the minors from M. Observe that nonzero elements of $\mathcal{H}$ are also nonzero in $\mathcal{A}$ (a stronger statement will appear below in Lemma 3), thereby one can think that $\mathcal{H} \subset \mathcal{A}$.
For any point (matrix) $a \in F^{n^2}$ denote by $H_a \in \mathcal{H}$ a hyperplane consisting of all $f \in \mathcal{H}$ such that $f(a) = 0$. Lemma 1 states actually that the codimension in $\mathcal{H}$

$$c = \operatorname{codim}\left(\bigcap_{a\notin E} H_a\right) < \delta^n q^r$$

Because of the second inequality (6) and again the Stirling's formula we get the inequality

$$\dim \mathcal{H} > c\,\eta^n \tag{7}$$

for a suitable constant $\eta > 1$.
Denote the full linear groups $G = Gl_n(F) \subset F^{n^2}$, it is well known that

$$|G| = (q^n-1)(q^n-q)\cdots(q^n-q^{n-1}) \\ \geq q^{n^2}(q-2)/(q-1)$$

(remind that $q \geq 3$). For any $g \in G$ one can consider an F-linear operator $T_g : \mathcal{H} \to \mathcal{H}$ defined for any $f \in \mathcal{H}$ and any matrix $a \in F^{n^2}$ by the formula $(T_g(f))(a) = f(ga)$ (moreover, one could define T_{g_1} by the same formula for any not necessarily nonsingular matrix $g_1 \in F^{n^2}$). The latter formula defines an operator $T_g : \mathcal{H} \to \mathcal{H}$ since the minors from M of the matrix ga are the linear combinations (with the coefficients depending only on g) of the minors from M of a. Thus, T_g provides a representation of G because $T_{g_1g_2} = T_{g_1}T_{g_2}$ (more precisely, this representation is the direct sum of $\binom{n}{\lceil\gamma n\rceil}$ copies of $\lceil\gamma n\rceil$-th wedge power of the natural representation of G on F^n).
Clearly $T_{g^{-1}}(H_a) = H_{ga}$. Consider now a plane

$$P = \bigcap_{a\in G\setminus E} H_a \subset \mathcal{H},$$

its codimension $c_1 = \operatorname{codim} P \leq c$. Also denote $E_1 = E \cap G$. So, from now on we restrict ourselves to considering just matrices from G (rather than from the whole set of matrices F^{n^2}).
Now assume that a subset $S \subset G$ satisfies the following property

$$\bigcup_{g\in S} g\,(G \setminus E_1) = G \tag{8}$$

For any $g \in G$ we have

$$T_{g^{-1}}(P) = \bigcap_{a\in G\setminus E_1} T_{g^{-1}}(H_a) = \bigcap_{b\in g(G\setminus E_1)} H_b.$$

Therefore, we get from (8) that

$$\bigcap_{b\in G} H_b \;=\; \bigcap_{g\in S} T_{g^{-1}}(P) \qquad (9)$$

Next we need the following combinatorial lemma (see e.g. [L75]).

Lemma 2([L75]) Let (V,R) be a directed (regular) graph with $|V| = m$ vertices and with the in-degree and the out-degree of each vertex both equal to d. Then there exists a subset $U \subset V$ of a size $O(\frac{m}{d}\,\log(d+1))$ such that for any vertex $v \in V$ there is a vertex $u \in U$ forming an edge $(u,v) \in R$.

Construct a directed regular graph with the set of vertices G and an edge (g_2, g_1) if and only if $g_2^{-1} g_1 \notin E_1$. Applying to this graph Lemma 2 supplies us with a set $S \subset G$ such that for any $g_1 \in G$ there is $g \in S$ satisfying $g^{-1} g_1 \notin E_1$, or equivalently $g_1 \in g(G \setminus E_1)$. Thus, S fulfills (8).
According to Lemma 2 and taking into account Lemma 1 and the first inequality (6)

$$|S| \le O\left(\frac{|G|}{|G| - q^{n^2}(\delta\,\alpha)^n} n^2 \log q\right) \le O(n^2).$$

Finally, we show that $\bigcap_{b\in G} H_b \neq 0$. Indeed, $\operatorname{codim} T_{g^{-1}}(P) = \operatorname{codim} P = c_1 \le c$ for any $g \in G$. Hence $\operatorname{codim} \bigcap_{g\in S} T_{g^{-1}}(P) \le O(|S|\;c_1) \le O(n^2\;c)$ which is less than $\dim \mathcal{H}$ because of (7). Therefore, $0 \neq \bigcap_{g\in S} T_{g^{-1}}(P) = \bigcap_{b\in G} H_b$ (see (9)). Take an arbitrary $0 \neq f \in \bigcap_{b\in G} H_b$, this means that f vanishes at all nonsingular matrices. So, to complete the proof of the Theorem (see the introduction), we need the following lemma.

Lemma 3 No multilinear polynomial $0 \neq f \in F[X_{1,1}, \ldots, X_{n,n}]$ vanishes at all nonsingular matrices (note that $q \ge 3$).

Proof of Lemma 3 goes by induction on n. The base of induction for $n = 1$ is evident. For the inductive step suppose the contrary. Some variable occurs in f, permuting the rows and the columns we can assume w.l.o.g. that $X_{n,n}$ occurs in f. Then $f = X_{n,n}\, f_1 + f_0$, where $f_1 \not\equiv 0$, f_0 are multilinear polynomials being independent from $X_{n,n}$. On the other hand, $\mathrm{Det} = X_{n,n}\, M_{n,n} + h$, where $M_{n,n}$ is $(n-1) \times (n-1)$ minor and h is independent from $X_{n,n}$.
For the time being, specify the variables $X_{k,\ell} = x_{k,\ell}^{(0)} \in F$ for all $1 \le k, \ell \le n-1$ in such a way that $M_{nn}(\{x_{k,\ell}^{(0)}\}) \neq 0$ (so far, there are many possibilities for specifying). Also we get a multilinear polynomial

$$\begin{aligned} f(\{x_{k,\ell}^{(0)}\}) &= \\ &X_{n,n} f_1(\{x_{k,\ell}^{(0)}\}) + f_0(\{x_{k,\ell}^{(0)}\}) = \\ &X_{n,n} \overline{f}_1 + \overline{f}_0, \end{aligned}$$

where $\overline{f}_1$, $\overline{f}_0$, $\in F[X_{n,1}, \ldots, X_{n,n-1}, X_{1,n}, \ldots, X_{n-1,n}]$. For any set of the values of the variables

$$\begin{aligned} X_{n,k} = x_{n,k}^{(0)} \in F,\;\; X_{k,n} &= \\ x_{k,n}^{(0)} \in F,\;\; 1 \le k \le n-1 & \qquad (10) \end{aligned}$$

there are exactly $(q-1) \ge 2$ values of $X_{n,n}$ such that Det does not vanish. Therefore, the multilinear polynomials $\overline{f}_1$, $\overline{f}_0$ vanish identically: indeed, otherwise for some values (10) a nonvanishing identically linear polynomial

$$\begin{aligned} &X_{n,n}\, \overline{f}_1(\{x_{n,k}^{(0)}, x_{k,n}^{(0)}\}_k) + \\ &\quad \overline{f}_0(\{x_{n,k}^{(0)}, x_{k,n}^{(0)}\}_k) \in F[X_{n,n}] \end{aligned}$$

would have $q-1 \ge 2$ roots.
On the other hand, there is an appropriate set of values (10) for which the substitution of these values $\tilde{f}_1 = f_1(\{x_{n,k}^{(0)}, x_{k,n}^{(0)}\}_k) \in F[X_{1,1}, \ldots, X_{n-1,n-1}]$ provides a nonvanishing identically polynomial. As we have seen above, $\tilde{f}_1$ vanishes at any nonsingular $(n-1) \times (n-1)$ matrix $\{x_{k,\ell}^{(0)}\}_{1 \le k,\ell \le n-1}$; that contradicts to the inductive hypothesis and proves Lemma 3. $\square$

3 Open Problems

An intriguing open problem remains to extend our exponential lower bound for depth 3 arithmetic circuits to arbitrary fields including characteristic zero.

Acknowledgements

We thank László Babai, Sasha Razborov, Avi Widgerson, and Andy Yao for interesting discussions on the subject of this paper.

References

[G82] D. Grigoriev, *Lower Bounds in Algebraic Complexity*, J. Soviet Math., **29** (1985), pp. 1388–1425.

[H77] R. Hartshorne, *Algebraic Geometry*, Springer Verlag, 1977.

[L75] L. Lovász, *On the Ratio of Optimal Integral and Rational Covers*, Discrete Mathematics, **13** (1975), pp. 383–390.

[N91] N. Nisan, *Lower Bound for Non-Commutative Computation*, Proc. 23rd ACM STOC (1991), pp. 410–418.

[L84] S. Lang, *Algebra (2nd Edition)*, Addison-Wesley, 1984.

[NW95] N. Nisan and A. Widgerson, *Lower Bound on Arithmetic Circuits via Partial Derivatives*, Proc. IEEE FOCS (1995), pp. 16–25.

[R87] A. Razborov, *Lower Bounds on the Size of Bounded Depth Circuits over a Complete Basis with Logical Addition*, Math. Notes, **41** (1987), pp. 333–338.

[R98] A. Razborov, *Personal communication*, 1998.

[S87] R. Smolensky, *Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity*, Proc. 19th ACM STOC (1987), pp. 77–82.

[S73] V. Strassen, *Die Berechnungskomplexität von Elementarsymmetrischen Funktionen und von Interpolationskoeffizienten*, Numer. Math. 20 (1973), pp. 238–251.

[S76] V. Strassen, *Computational Complexity over Finite Fields*, SIAM J. Comput. 5 (1976), pp. 324–331.

[V79] L. Valiant, *Completeness Classes in Algebra*, Proc. 11th ACM STOC (1979), pp. 259–261.