A Non-Linear Lower Bound for Constant Depth Arithmetical Circuits via the Discrete Uncertainty Principle

Maurice J. Jansen Kenneth W.Regan*

November 10, 2006

Abstract

We prove a non-linear lower bound on the size of a bounded depth bilinear arithmetical circuit computing the circular convolution mapping in case the input vectors are of prime length. For this proof we utilize a strengthing of the Donoho-Stark uncertainty principle [DS89], as given by Tao [Tao05], and a combinatorial lemma by Raz and Shpilka [RS03].A new proof is given of the Donoho-Stark uncertainty principle.

Keywords. Computational complexity, arithmetical circuits, lower bounds, constant depth bilinear circuits.

1 Introduction

One of the central mysteries in arithmetic circuit complexity over infinite fields F is the computational power conferred by the ability to use "for free" constants of arbitrary magnitude and/or precision from F. Morgenstern makes the argument that most algorithms used in practice only use constants of "reasonably" bounded magnitude [Mor73]. A possible exception is perhaps formed by algorithms with constants obtained via derandomization procedures or polynomial interpolation.

In any case, if we restrict circuits to have their scalars to be of constant bounded magnitude it *does* become easier to prove lower bounds. For example, we have the volumetric lower bounds of Morgenstern for *bounded coefficient linear* circuits [Mor73], and Raz gives a tight $\Omega(N \log N)$ lower bound in the *bounded coefficient bilinear model* for the mapping defined by multiplication of two $n \times n$ matrices, where $N = n^2$ [Raz02]. Bürgisser and Lotz, building on the work of Raz, prove a tight $\Omega(n \log n)$ lower bound for the convolution of two n-vectors of variables [BL02].

Considering linear and bilinear circuits, in case constants are unrestricted however, no non-linear lower bounds are known. The question being, whether this is just perceptual due to a current lack of lower bound techniques, or whether there is a real loss in computational power when restricting scalar magnitudes. Although we do not know of any non-linear lower bounds for unrestricted linear and bilinear circuits, what is known are size-depth tradeoffs. Namely, for linear circuits there are the results by Pudlak [Pud94], and for bilinear circuits of constant depth Raz and Shpilka prove a non-linear lower bound for the matrix multiplication mapping [RS03].

In this paper, building on the work of [RS03], we prove a size-depth tradeoff for the circular convolution mapping that was considered in [BL02]. We do this utilizing a discrete variant of the

*Part of this work by both authors was supported by NSF Grant CCR-9821040. Corresponding author: Dept. of CSE at UB, 201 Bell Hall, Buffalo, NY 14260-2000; (716) 645-3180 x114, fax 645-3464; regan@cse.buffalo.edu

Heisenberg uncertainty principle. This principle from quantum mechanics is widely known, even to the extent of having had a cultural impact, and expresses the inherent impossibility of simultaneously knowing, to arbitrary precision, certain complementary observables in nature. For example, one cannot simultaneously, through measurement, determine both the position and velocity of some given elementary particle to arbitrary precision. Mathematically, the main issue is that a function and its continuous Fourier transform cannot be simultaneously arbitrarily narrowly "localized". Donoho and Stark considered this phenomena under various measures of localization for the *discrete* Fourier transform [DS89]. To obtain our result we will use a strengthening of the Donoho-Stark uncertainty principle for the prime case due to Tao [Tao05].

The rest of this paper is organized as follows. First, we give the necessary mathematical preliminaries, including a new proof of the Donoho-Stark principle, in Section 2. Then Section 3 contains the lower bound. Finally, we mention some open problems and conclusions in Section 4.

2 Preliminaries

We define the discrete Fourier transform matrix DFT_n by

$$(DFT_n)_{ij} = \omega^{ij},$$

where $\omega = e^{2\pi i/n}$. Its unitary version we denote by F_n:

$$F_n = \frac{DFT_n}{\sqrt{n}}.$$

The conjugate transpose of a matrix A will be denoted by A^*.

Definition 2.1. The **cyclic convolution** $x \circ y$ of two n-vectors $x = (x_0, x_1, \ldots, x_{n-1})^T$ and $y = (y_0, y_1, \ldots, y_{n-1})^T$ is the n-vector $(z_0, \ldots, z_{n-1})^T$ with components

$$z_k = \sum_{i+j \equiv k \bmod n} x_i y_j$$

for $0 \leq k < n$.

For example, for $n = 5$, we get

$$x \circ y = \begin{pmatrix} x_0y_0 + x_4y_1 + x_3y_2 + x_2y_3 + x_1y_4 \\ x_1y_0 + x_0y_1 + x_4y_2 + x_3y_3 + x_2y_4 \\ x_2y_0 + x_1y_1 + x_0y_2 + x_4y_3 + x_3y_4 \\ x_3y_0 + x_2y_1 + x_1y_2 + x_0y_3 + x_4y_4 \\ x_4y_0 + x_3y_1 + x_2y_2 + x_1y_3 + x_0y_4 \end{pmatrix}.$$

When fixing $x = a = (a_0, \ldots, a_{n-1})^T$, the induced map on y is computed by the circulant matrix $Circ(a)$, which we define by

$$Circ(a) = \begin{pmatrix} a_0 & a_{n-1} & \cdots & a_2 & a_1 \\ a_1 & a_0 & \cdots & a_3 & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n-2} & a_{n-3} & \cdots & a_0 & a_{n-1} \\ a_{n-1} & a_{n-2} & \cdots & a_1 & a_0 \end{pmatrix}.$$

That is, we have that

$$x \circ y = Circ(x)y = Circ(y)x.$$

Convolution can be computed using the Fourier transform, according to the following folklore result:

Theorem 2.1 (The Convolution Theorem) *For any $a \in \mathbf{C}^n$,*

$$Circ(a) = F_n diag(DFT_n a) F_n^*.$$

In the above, for a vector $x = (x_1, x_2, \ldots, x_n)^T$,

$$\mathrm{diag}(x) = \begin{pmatrix} x_1 & 0 & \cdots & 0 & 0 \\ 0 & x_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \\ 0 & 0 & \cdots & x_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & x_n \end{pmatrix}.$$

2.1 Discrete Uncertainty Principles

We begin with an alternative proof, which is new to our knowledge, of the Donoho-Stark discrete uncertainty principle.

Definition 2.2. For an n-vector f, define the *support of f* to be the set $\mathrm{supp}(f) = \{i : f_i \neq 0\}$.

The size of the support of a vector f is a crude measure of the amount of localization of a vector. Analogous to the Heisenberg uncertainty principle, we can prove that for this measure a vector f and its Fourier transform $\hat{f}$ cannot both be arbitrarily narrowly localized. More precisely, we have the following theorem:

Theorem 2.2 ([DS89]) *For any n-vector $f \neq 0$,*

$$|\mathrm{supp}(f)| \cdot |\mathrm{supp}(\hat{f})| \geq n, \tag{1}$$

where $\hat{f} = F_n f$ is the discrete Fourier transform of f.

Proof. Consider an arbitrary Fourier transform pair $(f, \hat{f})$ with $\hat{f} = F_n f$ and $f \neq 0$. Since

$$Circ(f) = \sqrt{n} F_n^* \mathrm{diag}(\hat{f}) F_n,$$

we have that

$$\mathrm{supp}(\hat{f}) = rank(Circ(f)).$$

Let R be the maximum number of zeroes following a non-zero entry in f (in the cyclic sense). Then $R \geq \frac{n}{|\mathrm{supp}(f)|} - 1$.

Namely, if this were not the case, then imagine partitioning the entries of f as follows: Start at an arbitrary nonzero position. Set $i = 1$. If there are no other zero positions then B_i equals this position. Otherwise, let B_i be this position together with all the zero positions that follow it (in the cyclic sense). Repeat this process for the next i. We obtain this way $B_1, B_2, \ldots, B_{|\mathrm{supp}(f)|}$ that partition all n entries of f. By the above then, for each i, $|B_i| \leq R + 1 < \frac{n}{|\mathrm{supp}(f)|}$. So

$$|\bigcup_i B_i| < |\mathrm{supp}(f)| \cdot \frac{n}{|\mathrm{supp}(f)|} = n.$$

This is a contradiction, because $B_1, B_2, \ldots, B_{|\mathrm{supp}(f)|}$ partition the n entries of f.

The above implies the first $R+1$ rows of $Circ(f)$ are independent, because they contain a square submatrix that is upper triangular (modulo cylic shifts). Hence $rank(Circ(f)) \geq R+1 \geq \frac{n}{|\mathrm{supp}(f)|}$. □

Interestingly enough, divisibility properties of n play an important role in the analysis. For example, Tao showed that, in case n is prime, the inequality (1) can be significantly improved. The proof relies on the well-known fact that for prime p the discrete Fourier transform matrix DFT_p is regular.

Definition 2.3. An $n \times n$ marix A is called *regular* if any square submatrix of A is non-singular.

Theorem 2.3 *For prime p, DFT_p is a regular matrix.*

The first proof of this fact is attributed to Chebotarëv, who proved it in 1926 (see [SJ96]). Although typical proofs of this fact are field theoretic in nature, Tao gives a proof by elementary means. Once one has established this fact the following can be proved quite readily:

Theorem 2.4 ([Tao05]) *For prime p, for any nonzero p-vector f and its Fourier transform $\hat{f} = F_p f$ we have that*

$$|\mathrm{supp}(f)| + |\mathrm{supp}(\hat{f})| \geq p+1.$$

Proof. Let $k = p - |\mathrm{supp}(\hat{f})|$. There are k zeroes in $\hat{f}$. Let $I \subseteq \{0, 1, \ldots, p-1\}$ be the indices of these zeroes. Suppose $|\mathrm{supp}(f)| \leq k$. Let $J \subseteq \{0, 1, \ldots, p-1\}$ be a set of size k that contains all indices of non-zero entries of f. In the following $DFT^p_{I,J}$ denotes the minor of DFT_p with rows I and columns J. We have that

$$(DFT^p_{I,J})f_J = (DFT_p f)_I = 0,$$

but $f_J \neq 0$ since $f \neq 0$. This is a contradiction since $DFT^p_{I,J}$ is non-singular. Hence $|\mathrm{supp}(f)| > k = p - |\mathrm{supp}(\hat{f})|$. □

Actually, in the above proof we only used the fact that DFT_p is a regular matrix, so more generally we have:

Theorem 2.5 *Let A be an $n \times n$ regular matrix and consider pairs $(f, \hat{f} := Af)$ where $f \neq 0$. Then*

$$|\mathrm{supp}(f)| + |\mathrm{supp}(\hat{f})| \geq n+1.$$

2.2 Slow Growing Functions

Following [Pud94, RS03] we define:

Definition 2.4. For a function $f : \mathbf{N} \to \mathbf{N}$, define $f^{(i)}$ to be the composition of f with itself i times:

1. $f^{(0)}$ is the identity function,
2. $f^{(i)} = f \circ f^{(i-1)}$, for $i > 0$.

Futhermore, for f such that $f(n) < n$, for all $n > 0$, define

$$f^*(n) = \min\{i : f^{(i)} \leq 1\}$$

The following set of extremely slow-growing functions $\lambda_d(n)$ will be used to express the lower bounds. Each $\lambda_d(n)$ is a monotone increasing function tending to infinity.

Definition 2.5 ([Pud94, RS03]). Let

1. $\lambda_1(n) = \lfloor \sqrt{n} \rfloor$,
2. $\lambda_2(n) = \lceil \log n \rceil$,
3. $\lambda_d(n) = \lambda^*_{d-2}(n)$, for $d > 2$.

For a directed acyclic graph G, V_G denotes the set of all nodes, I_G those with in-degree 0, and O_G those with out-degree 0. The depth of G is the length in edges of the longest path from I_G to O_G. Raz and Shpilka prove the following combinatorial lemma:

Lemma 2.6 ([RS03]) *For any $0 < \epsilon < \frac{1}{400}$ and any layered directed acyclic graph G of depth d with more than n vertices and less than $\epsilon \cdot n \cdot \lambda_d(n)$ edges, the following is satisfied:*

For some k with $\sqrt{n} \le k = o(n)$, there exist subsets $I \subset I_G$, $O \subset O_G$, and $V \subset V_G$ for which $|I|, |O| \le 5\epsilon \cdot d \cdot n$ and $|V| = k$, and such that the total number of directed paths from $I_G \backslash I$ to $O_G \backslash O$ that do not pass through nodes in V is at most $\epsilon \cdot \frac{n^2}{k}$.

2.3 Circuits for Circular Convolution

We consider bounded depth bilinear arithmetical circuits with arbitrary fan-in and fan-out as in [RS03]. Constants on the wires are assumed to be from the complex field $\mathbf{C}$, and our circuits are assumed to be layered. More precisely, there are two disjoint sets of inputs, say x and y variables. Separately for each of these inputs there are layered linear circuits, i.e. consisting only of addition gates, computing linear forms in x and y. Then there is a single layer of multiplication gates multiplying the computed linear forms. Finally, there is a layered linear circuit taking the output of the multiplication gates as input. The circuit computes formal polynomials in x and y variables in the obvious manner, with constants on the wires intended as scalar multiplication. We will give lower bounds on the number of edges present in the circuit below the multiplication gates.

Definition 2.6. For a bounded depth bilinear circuit C we define its size $s(C)$ to be the number of edges in the circuit between the multiplication gates and the outputs, and define by its depth $d(C)$ to be the length of a longest path in edges from a multiplication gate to an output.

Note that Cooley and Tukey [CT65] give $O(n \log n)$ size, $O(\log n)$ depth linear circuits that compute DFT_n. So using theorem 2.1, we obtain $O(n \log n)$ size bilinear circuits for computing circular convolution. These circuits have complex coefficients on the wires of norm 1. Burgisser and Lotz prove that this is optimal for circuits that have their constants restricted to be of norm $O(1)$ [BL02]. See [Jan06] for some generalization of this result.

3 Lower Bounds for Cyclic Convolution

We begin with the following easy proposition:

Proposition 3.1 *Any bilinear circuit of depth 1 computing circular convolution $x^T Circ(y)$ has size $s(C) \ge n^2$.*

Proof. A circuit of depth 1 has a very simple structure. There are some number r of multiplication gates M_r computing products $M_r = L_r(x)R_r(y)$, where $L_r(x)$ and $R_r(y)$ are linear forms. Then there is one layer of output gates, each gate computing summation over some set of input multiplication gates.

We will argue that each output gate must be connected to at least n multiplication gates. For purpose of contradiction suppose that this is not the case. Say some output gate O_i takes input from $< n$ multiplication gates. Consider the subspace of dimension at least 1 defined by equations $L_j(x) = 0$, for each multiplication gate j attached to output O_i. We can select a non-zero vector a from this space such that for any assigment $y = b$,

$$(a^T Circ(b))_i = 0.$$

This yields a contradiction, for example we can take b^T to be equal to a^* shifted by i, then $(a^T Circ(b))_i = ||a||_2^2$, which is non-zero, since a is a non-zero vector. $\square$

We now state and prove our main result.

Theorem 3.2 *There exists $\epsilon > 0$ such that if p is a prime number, any layered bilinear circuit with inputs $x = (x_0, x_1, \ldots, x_{p-1})$ and $y = (y_0, y_1, \ldots, y_{p-1})$ of depth d computing cyclic convolution $x^T Circ(y)$ has size $s(C) \geq \epsilon p \lambda_d(p)$.*

Proof. Consider the circuit computing

$$x^T Circ(y) = x^T F_p \text{diag}(DFT_p(y)) F_p^*.$$

We first apply substitutions $x^T := x^T F_p^*$ and $y = \frac{1}{n} DFT_P^* y$ at the inputs. This does not alter the circuit below the multiplication gates, but now we have a circuit computing

$$x^T \text{diag}(y) F_p^*.$$

Let G be the directed acyclic graph of depth d given by the part of circuit below the multiplication gates. The set I_G is the collection of multiplication gates $M_i = L_i(x)R_i(y)$, where $L_i(x)$ and $R_i(y)$ are linear forms. Take $O_G = \{1, 2, \ldots, p\}$ to be the set of outputs of the circuit. Let $\epsilon > 0$ be some small enough constant to be determined later. Trivially G has at least p vertices. Suppose that G has strictly fewer than $\epsilon p \cdot \lambda_d(p)$ edges. Lemma 2.6 applies, and we obtain sets $I \subset I_G$, $O \subset O_G$ and $V \subset V_G$ such that

1. $|I|, |O| \leq 5\epsilon dp$,
2. $|V| = k$, with $\sqrt{n} \geq k = o(p)$, and
3. the total number of directed paths from $I_G \backslash I$ to $O_G \backslash O$ that do not pass through nodes in V is at most $\epsilon \frac{p^2}{k}$.

For each output node $i \in O_G \backslash O$, define $P(i)$ to be the number of multiplication gates in $I_G \backslash I$ for which there exists a directed path that bypasses V and reaches node i. Let R be a set of $r = 10k$ output gates with lowest $P(i)$ values. By averaging we get that

$$\sum_{r \in R} P(r) \leq \frac{r}{|O_G \backslash O|} \sum_{r \in O_G \backslash O} P(r) \leq \frac{r}{p - 5\epsilon dp} \cdot \frac{\epsilon p^2}{k} = \frac{10 \epsilon p}{1 - 5\epsilon d}.$$

Let I' be the set of all multiplication gates in $I_G \backslash I$ for which there exist directed paths to nodes in R that bypass V. We can conclude that

$$|I'| \leq \frac{10\epsilon p}{1 - 5\epsilon d}.$$

Define a linear subspace W by the set of equations

$$R_i(y) = 0 \text{ for all } i \in I \cup I'.$$

For any fixed substitution for $y \in W$ the resulting circuit has all of the gates computing linear function in the x variables. Relative to a fixed choice for y, define linear subspace W_y by equations $g_v(x) = 0$ for all $v \in V$, where $g_v(x)$ denotes the linear form computed at gate v. Note that $\dim(W) \geq p - 5\epsilon dp - \frac{10\epsilon p}{1-5\epsilon d}$ and $\dim(W_y) \geq p - k$, for each y. Now we have arranged that for each $y \in W$, and each $x \in W_y$,

$$(x^T diag(y) F_p^*)_r = 0, \tag{2}$$

for each $r \in R$.

In order to reach a contradiction, we will now argue that it is possible to select $y \in W$ and $x \in W_y$ such that some output in R is non-zero.

First of all, fix a vector $y \in W$ that has at most $5\epsilon dp + \frac{10\epsilon p}{1-5\epsilon d}$ zeroes: this can be done because $\dim(W) \geq p - 5\epsilon dp - \frac{10\epsilon p}{1-5\epsilon d}$. Let A be the set of indices i for which $y_i = 0$. Let $m = |A|$. Let W_y' be a subspace of W_y of dimension 1 obtained by adding equations to the defining set of W_y as follows. For the first stage add $x_i = 0$ for each $i \in A$. In a second stage, start adding equations that require $x_i = 0$ for $i \notin A$, until the dimension has been cut down to 1. Since we are starting out with a space of dimension $p - k$, after the first stage, the dimension will be cut down to at most $p - k - m$, so we will be able to add $x_i = 0$ in the second stage for at least $p - k - m - 1$. many i with $i \notin A$. Provided ϵ is small enough, since $k = o(n)$, $k + m$ will be less than a small fraction of p, so we are guaranteed that we can indeed complete this process still leaving a subspace of non-trivial dimension. Select an arbitrary non-zero vector x from W_y'. Observe that of the $p - m$ indices i not in A, x_i is non-zero for at most $k + 1$ entries, and that x_i is zero for all $i \in A$. So x_i is zero for each i for which $y_i = 0$. Since x itself is a nonzero vector there must be some place i where x_i and y_i are both nonzero.

Let $f = x^T diag(y)$ and $\hat{f} = f F_P^*$. We thus conclude that f is a non-zero vector, but that $|\text{supp}(f)| \leq k + 1$.

By the discrete uncertainty principle for cyclic groups of prime order [Tao05], stated in Theorem 2.4, we have that

$$\text{supp}(f) + \text{supp}(\hat{f}) \geq p + 1.$$

Hence the output vector of the circuit $\hat{f}$ is non-zero in at least $p + 1 - (k + 1) = p - k$ places. Since R is of size $10k$, by the pigeonhole principle, there must be some output in R that is non-zero. This is in contradiction with equation (2). □

4 Conclusion

The first obvious question is of course whether we can drop the primality assumption and still obtain the conclusion of Theorem 3.2. Unfortunately, without the assumption of primality, the uncertainty principle weakens beyond the point of being usable in our argument. It is also not obvious for *circular* convolution how to reduce the non-prime case to the prime case by means doing a padding/embedding-type argument, which would be clear for non-circular convolution.

A second more open ended question is the role that the various uncertainty principles, continuous or discrete, can play in circuit lower bound arguments. In [Jan06] this question is probed some further for the bounded-coefficient bilinear model. Perhaps the uncertainty principle can also play a role in quantum circuit lower bounds.

Finally, there still is the central open problem of obtaining any kind of non-linear lower bound for (unrestricted) linear circuits. However, no notable progress has been made on this question for over 35 years.

References

[BL02] P. Bürgisser and M. Lotz. Lower bounds on the bounded coefficient complexity of bilinear maps. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 659–668, 2002.

[CT65] J.W. Cooley and J.W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

[DS89] D.L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49:906–931, 1989.

[Jan06] Maurice J. Jansen. *Lower Bound Frontiers in Arithmetical Circuit Complexity*. PhD thesis, University at Buffalo, 2006.

[Mor73] J. Morgenstern. Note on a lower bound of the linear complexity of the fast Fourier transform. *J. Assn. Comp. Mach.*, 20:305–306, 1973.

[Pud94] P. Pudlák. Communication in bounded-depth circuits. *Combinatorica*, 14:203–216, 1994.

[Raz02] R. Raz. On the complexity of matrix product. In *Proc. 34th Annual ACM Symposium on the Theory of Computing*, pages 144–151, 2002. Also ECCC TR 12, 2002.

[RS03] R. Raz and A. Shpilka. Lower bounds for matrix product, in arbitrary circuits with bounded gates. *SIAM J. Comput.*, 32:488–513, 2003.

[SJ96] P. Stevenhagen and H.W. Lenstra Jr. Chebotarëv and his density theorem. *Mathematical Intelligencer*, 18:26–37, 1996.

[Tao05] T. Tao. An uncertainty principle for cyclic groups of prime order. *Mathematical Research Letters*, 12:121–127, 2005.