Electronic Colloquium on Computational Complexity, Report No. 88 (2011)

How much commutativity is needed to prove polynomial identities?

Pavel Hrubeš

June 7, 2011

Abstract

Let f be a non-commutative polynomial such that $f = 0$ if we assume that the variables in f commute. Let $\mathrm{Q}(f)$ be the smallest k such that there exist polynomials $g_1, g_1', g_2, g_2', \ldots, g_k, g_k'$ with

$$f \in I([g_1, g_1'], [g_2, g_2'], \ldots, [g_k, g_k']),$$

where $[g, h] = gh - hg$. Then $\mathrm{Q}(f) \leq \binom{n}{2}$, where n is the number of variables of f. We show that there exists a polynomial f with $\mathrm{Q}(f) = \Omega(n^2)$. We pose the problem of constructing such an f explicitly, pointing out that the solution may have applications to complexity of proofs.

1 Introduction

In this note, we address the following question. Assume that we have a polynomial f in n mutually non-commuting variables which has the property that f is zero when the variables are assumed to commute. What is the smallest number of commutativity axioms of the form $gh = hg$ one needs to use to make f vanish? More precisely: what is the smallest k such that f lies in the ideal generated by k polynomials of the form $[g, h]$, with $[g, h] = gh - hg$? We denote this k by $\mathrm{Q}(f)$ and call it the *commutative complexity* of f. The simplest example is the polynomial $f = xy - yx$. As a non-commutative polynomial f is non-zero, but it vanishes when the variables commute, and lies in the ideal of $[x, y]$. Another example is the polynomial $zxy + xyz - 2yzx$. Clearly, f lies in the ideal generated by $[x, y], [x, z], [y, z]$, which shows that $\mathrm{Q}(f) \leq 3$. But f can also be written as $[z, xy] - 2[y, zx]$ which implies that $\mathrm{Q}(f) \leq 2$. In general, f vanishes whenever we assume that all the variables of f commute and hence $\mathrm{Q}(f) \leq \binom{n}{2}$. However, the last example illustrates the fact that the commutative complexity of f can be reduced if more sophisticated commutativity axioms are employed.

The scenario we have in mind is that f is presented by an arithmetic formula, or a circuit. Imagine that we have an arithmetic formula F– that is, a syntactic expression such as $(x-y)(x+y)-xx+yy$. Such a formula is intended to compute a *commutative* polynomial. In the example when $F = (x-y)(x+y)-xx-yy$, it

ISSN 1433-8092

computes the polynomial 0. On the other hand, since F is a syntactic expression, it can also be interpreted as computing a *non-commutative* polynomial $\hat{F}$. In this case, $\hat{F} = xy - yx$, which is a non-trivial non-commutative polynomial. This allows to associate with a formula F a non-commutative polynomial which is non-zero despite the fact that F is trivial when interpreted commutatively. The quantity $\mathrm{Q}(\hat{F})$ is one possible way how to use the properties of $\hat{F}$ to argue about the formula F.

The intended meaning of $\mathrm{Q}(\hat{F})$ is to capture the number of commutativity axioms needed to prove the equation $F = 0$. Its intended application is to serve as a lower bound on lengths of proofs of polynomial identities. For specific systems for proving polynomial identities (see [3, 6] and Section 4), one can show that a good enough lower bound on the commutative complexity implies new lower bounds on lengths of proofs. This approach could potentially work not only for systems proving polynomial identities, but also for propositional proof systems such as extended Frege. Extended Frege is one the most fundamental propositional proof systems (see [2, 7]). To give superpolynomial lower-bounds on the size of extended Frege proofs is one of the main open problems of proof complexity. However, even proving superlinear lower bounds has been an unanswered challenge. In the perfect world where Q has been completely understood, one can proceed to modify Q so that it is a lower bound to extended Frege proofs. This way, one may hope to achieve up to a quadratic lower bound.

Our main result is to show that there exists a polynomial f in n variables such that $\mathrm{Q}(f) \geq \Omega(n^2)$. This qualitatively matches the generic upper-bound $\mathrm{Q}(f) \leq \binom{n}{2}$ and shows that Q is a non-trivial complexity measure. The main drawback of our proof is that it gives no indication how to construct such an f explicitly. Most importantly, the result is not sufficient for the purported proof complexity applications. We are thus lead to two open questions. The first is to construct an explicit non-commutative polynomial with a superlinear commutative complexity, in terms of its number of variables. The second is to find a non-commutative polynomial with a superlinear commutative complexity, in terms of its *circuit size*. For the first question, we will present several candidates. The second appears much more intricate, but it is this one which would be interesting in the proof complexity context.

2 Commutative complexity of non-commutative polynomials

We will be interested in *non-commutative polynomials* over a field $\mathbb{F}$. A non-commutative polynomial is a formal sum of products of variables and field elements, where we assume that the variables do not multiplicatively commute. That is, $xy \neq yx$ whenever $x \neq y$ but the variables commute with elements of $\mathbb{F}$. A non-commutative polynomial can be uniquely written as a finite sum $\sum_j c_j \alpha_j$, where $c_j \in \mathbb{F}$ and α_j is a *monomial* - a product of variables. Since most polynomials will be non-commutative, a "polynomial" will mean

'"non-commutative polynomial", unless stated otherwise. Non-commutative polynomials over a given field form a non-commutative ring. For polynomials $g_1, \ldots, g_k$, $I(g_1, \ldots, g_k)$ will denote the two-sided ideal generated by $g_1, \ldots, g_k$ in this ring.

Let f be a non-commutative polynomial in variables $x_1, \ldots, x_n$. By f^c we mean the same polynomial in which the variables are allowed to commute. That is, if $f = \sum_j c_j \alpha_j$, then f^c is the commutative polynomial $\sum_j c_j \alpha_j$. Note that it is possible that $f^c = 0$ while $f \neq 0$, as in the example $f = x_1x_2 - x_2x_1$. We will be interested precisely in such polynomials[1]. Let

$$[g, h] := gh - hg$$

be the commutator of g, h. Then the condition $f^c = 0$ is equivalent to the assumption that f lies in the ideal generated by all the polynomials $[x_i, x_j], i < j$. In other words

$$f \in I([x_i, x_j];\, i < j \in \{1, \ldots, n\})\,.$$

Let $\mathrm{Q}(f)$ denote the smallest k such that there exist polynomials $g_1, g_1', g_2, g_2', \ldots, g_k, g_k'$ with

$$f \in I([g_1, g_1'], [g_2, g_2'], \ldots, [g_k, g_k'])\,.$$

We will call $\mathrm{Q}(f)$ the *commutative complexity* of f. One can think of $\mathrm{Q}(f)$ as the smallest number of commutativity assumptions $g.g' = g'.g$ one needs to make in order to prove that $f^c = 0$.

We already know that

$$\mathrm{Q}(f) \leq \binom{n}{2}\,,$$

where n is the number of variables of f. It is easy to present a polynomial with $\mathrm{Q}(f) \geq n$. We now want to show that there exists a polynomial such that $\mathrm{Q}(f) = \Omega(n^2)$.

For a set of polynomials $f_1, \ldots f_m$ such that $f_1^c, \ldots, f_m^c = 0$, let us define $\mathrm{Q}(f_1, \ldots, f_m)$ as expected: it is the smallest k such that there exist polynomials $g_1, g_1', g_2, g_2', \ldots, g_k, g_k'$ with

$$f_1, \ldots, f_m \in I([g_1, g_1'], [g_2, g_2'], \ldots, [g_k, g_k'])\,.$$

An easy observation is that

$$\mathrm{Q}([x_i, x_j];\, i < j \in \{1, \ldots, n\}) = \binom{n}{2}\,.$$

This fact, however, is not very useful since the number of the polynomials $[x_i, x_j]$ is itself quadratic. Instead, we will show that there exist n polynomials $f_1, \ldots, f_n$ such that $\mathrm{Q}(f_1, \ldots, f_n) \sim n^2$.

[1]The set $\{f; f^c = 0\}$ is an ideal in the ring of non-commutative polynomials and is sometimes called the *commutator ideal*.

Lemma 1. *There exist polynomials $f_1, \ldots, f_n$ in n variables such that $f_1^c, \ldots, f_n^c = 0$ and $\mathrm{Q}(f_1, \ldots, f_n) = \Omega(n^2)$. Moreover, the polynomials are of degree two and have coefficients from $\{0, 1, -1\}$.*

Proof. This is mainly a standard counting argument. Consider n-tuples of polynomials $f_1, \ldots f_n$ with each f_i of the form

$$f_i = \sum_{j<k \in \{1,\ldots,n\}} c_{i,j,k}[x_j, x_k]\,, \text{ with } c_{i,j,k} \in \{0, 1\}\,.$$

Then $f_i^c = 0$ and different choices of the coefficients $c_{i,j,k}$ give distinct polynomials. Let p be a natural number such that we have $\mathrm{Q}(f_1, \ldots f_n) \leq p$ for every such n-tuple.

First, show the following: for every $f_1, \ldots, f_n$ as above there exist homogeneous linear polynomials $g_1, g_1', \ldots, g_p, g_p'$ such that $f_1, \ldots, f_n$ are linear combinations of $[g_1, g_1'], \ldots, [g_p, g_p']$. That is, for every $i \in \{1, \ldots, n\}$, there exist $a_{i,1}, \ldots, a_{i,p} \in \mathbb{F}$ such that

$$f_i = \sum_{j \in \{1,\ldots,p\}} a_{i,j}[g_j, g_j']\,. \tag{1}$$

This holds because f_i are homogeneous degree-two polynomials. For let $h^{(j)}$ denote the j-homogeneous part of h. Then $f_i = f_i^{(2)}$ and for every g, g', we have $[g, g']^{(0)} = [g, g']^{(1)} = 0$ and $[g, g']^{(2)} = [g^{(1)}, g'^{(1)}]$. Hence $(u[g, g']v)^{(2)} = u^{(0)}[g^{(1)}, g'^{(1)}]v^{(0)} = a[g^{(1)}, g'^{(1)}]$, with $a \in \mathbb{F}$. If we assume that $f_1, \ldots, f_n \in I([g_1, g_1'], \ldots, [g_p, g_p'])$, we have

$$f_i = \sum_j \sum_{k \in \{1,\ldots,p\}} u_{k,j}[g_k, g_k']v_{k,j}\,,$$

where $u_{k,j}, v_{k,j}$ are some polynomials. Hence $f_i = f_i^{(2)} = \sum_{k,j} a_{k,j}[g_k^{(1)}, g_k'^{(1)}]$ with $a_{k,j} \in \mathbb{F}$, and so $f_1, \ldots, f_n$ are linear combinations of $[g_1^{(1)}, g_1'^{(1)}], \ldots, [g_p^{(1)}, g_p'^{(1)}]$.

If the field $\mathbb{F}$ is finite, (1) already gives that p is roughly n^2. For let $q := n\binom{n}{2}$ be the number of the coefficients $c_{i,j,k}$.Then $f_1, \ldots, f_n$ are given by 2^q different choices of the coefficients. By (1), $f_1, \ldots, f_n$ is determined by the linear functions $g_1, g_1', \ldots, g_p, g_p'$ and the coefficients $a_{i,j}$, with $i \in \{1, \ldots, n\}, j \in \{1, \ldots, p\}$. The linear functions are given by the list of their $2np$ coefficients and, altogether, $f_1, \ldots, f_n$ are determined by $3np$ choices of elements of $\mathbb{F}$. This gives $2^q \leq |\mathbb{F}|^{3np}$ and so $p = \Omega(n^2)$. In general, assuming nothing about the field size, consider $f_1, \ldots, f_n$ as a $0,1$-vector in $\mathbb{F}^q$, and the list $g_k, g_k', a_{i,k}$, with $i \in \{1, \ldots, n\}, k \in \{1, \ldots, p\}$, as a vector in $\mathbb{F}^{3np}$. Then (1) gives a degree-three polynomial map $\mu : \mathbb{F}^{3np} \to \mathbb{F}^q$ such that every $0,1$-vector in $\mathbb{F}^q$ is in the range of μ. By [9] or [4], this implies that $c^{3np} \geq 2^q$ for a constant $c > 1$. Hence $p \geq c'q/n = c'\binom{n}{2} = \Omega(n^2)$. □

Next, we would like to use the polynomials $f_1, \ldots, f_n$ to obtain a single polynomial f with $\mathrm{Q}(f) \sim n^2$. The obvious choice is to introduce fresh variables

$z_1, \ldots, z_n$ and let $f := z_1 f_1 + \ldots z_n f_n$. The next lemma shows that that $\mathrm{Q}(f)$ and $\mathrm{Q}(f_1, \ldots, f_n)$ differ at most by a constant.

Lemma 2. *Let $f_1, \ldots, f_n$ be polynomials not containing the variables $z_1, \ldots, z_n$ such that $f_1^c, \ldots, f_n^c = 0$. Let $f := z_1 f_1 + \ldots z_n f_n$. Then*

$$\mathrm{Q}(f) \geq \frac{1}{3}\mathrm{Q}(f_1, \ldots, f_n)\,.$$

Proof. Assume that $\mathrm{Q}(f) = k$. We would like to show that there exist $3k$ polynomials g_i, h_i which do not depend on the variables $Z = \{z_1, \ldots, z_n\}$ such that $f \in I([g_1, h_1], \ldots, [g_{3k}, h_{3k}])$. Since g_i, h_i do not depend on Z, this implies that also $f_1, \ldots, f_n \in I([g_1, h_1], \ldots, [g_{3k}, h_{3k}])$ and therefore $\mathrm{Q}(f_1, \ldots, f_n) \leq 3k$.

In order to show this, we will define a map $\langle\cdot\rangle$ which transforms a polynomial g to a polynomial $\langle g\rangle$, with $\langle\cdot\rangle$ having the following properties:

(*i*). $\langle f\rangle = f$,

(*ii*). $\langle\cdot\rangle$ is linear, in the sense that $\langle ag + bh\rangle = a\langle g\rangle + b\langle h\rangle$ for any polynomials g, h and $a, b \in \mathbb{F}$.

(*iii*). For every polynomials g, h and $u_1, \ldots, u_m, v_1, \ldots, v_m$ there exist polynomials $g_1, g_2, g_3, h_1, h_2, h_3$ not depending on Z such that

$$\big\langle \sum_j u_j [g, h] v_j \big\rangle \in I([g_1, h_1], [g_2, h_2], [g_3, h_3])\,.$$

The existence of such a map implies the Lemma, for if $\mathrm{Q}(f) = k$ then f can be written as

$$f = \sum_{i \in \{1,\ldots,k\}} (\sum_j u_{i,j} [g_i, h_i] v_{i,j})\,.$$

By (i) and (ii), this gives

$$f = \sum_{i \in \{1,\ldots,k\}} \big\langle \sum_j u_{i,j} [g_i, h_i] v_{i,j} \big\rangle\,.$$

By (iii), $\langle \sum_j u_j [g_i, h_i] v_j\rangle \in I([g_{i,e}, h_{i,e}];\, e \in \{1, 2, 3\})$, where the latter polynomials do not depend on Z. Altogether $f \in I([g_{i,e}, h_{i,e}];\, e \in \{1, 2, 3\},\, i \in \{1, \ldots, k\})$, where $g_{i,e}, h_{i,e}$ do not depend on Z.

Let us now construct $\langle\cdot\rangle$. For a monomial α, define its Z-degree as the number of occurrences of variables from Z in α. If α has Z-degree 1, it can be uniquely written as $\alpha = \beta z \gamma$ where $z \in Z$ and β, γ do not depend on Z. Then let

$$\big\langle \alpha \big\rangle = \big\langle \beta z \gamma \big\rangle := z \gamma \beta\,.$$

If α has Z-degree not equal to 1 (that is, α either does not depend on Z, or contains more than one occurrence of variables from Z), let $\langle\alpha\rangle := 0$. For a polynomial $g = \sum_j c_j \alpha_j$, let $\langle g\rangle := \sum_j c_j \langle\alpha_j\rangle$. This guarantees that $\langle\cdot\rangle$ is linear.

(i) is also satisfied: if $z \in Z$ and γ does not depend on Z then $\lambda(z\gamma) = z\gamma$, and the polynomial f is a sum of monomials of this form.

It remains to show that $\langle\cdot\rangle$ satisfies (iii). First, it is easy to see that if g, h do not depend on Z then

$$\langle \sum_j u_j[g,h]v_j\rangle \in I([g,h]). \tag{2}$$

Second, assume that the Z-degree of every monomial in g is 1 and that h, u_j, v_j do not depend on Z. Then we claim that

$$\langle \sum_j u_j[g,h]v_j\rangle \in I([h, \sum_j u_j v_j])\,. \tag{3}$$

If $\alpha_1, \alpha_2, \gamma, \omega_1, \omega_2$ are monomials not depending on Z and $z \in Z$, we have

$$\begin{aligned}\big\langle \omega_1[\alpha_1 z\alpha_2, \gamma]\omega_2\big\rangle &= \big\langle \omega_1(\alpha_1 z\alpha_2\gamma - \gamma\alpha_1 z\alpha_2)\omega_2\big\rangle \\ &= \big\langle \omega_1\alpha_1 z\alpha_2\gamma\omega_2\big\rangle - \big\langle \omega_1\gamma\alpha_1 z\alpha_2\omega_2\big\rangle \\ &= z\alpha_2\gamma\omega_2\omega_1\alpha_1 - z\alpha_2\omega_2\omega_1\gamma\alpha_1 \\ &= z\alpha_2[\gamma, \omega_2\omega_1]\alpha_1\end{aligned}$$

By linearity of $\langle\cdot\rangle$ and bilinearity of the commutator $[\cdot,\cdot]$ this gives that

$$\big\langle u[\alpha_1 z\alpha_2, h]v\big\rangle = z\alpha_2[h, uv]\alpha_1$$

whenever u, v, h are polynomials not depending on Z. To finish the proof of (3), write g as $g = \sum_i \alpha_i z_{e(i)}\alpha_i'$ where $z_{e(i)} \in Z$ and α_i, α_i' do not depend on Z. Then

$$\begin{aligned}\big\langle \sum_j u_j[g,h]v_j\big\rangle = \big\langle \sum_j u_j[\sum_i \alpha_i z_{e(i)}\alpha_i', h]v_j\big\rangle = \sum_j\sum_i \big\langle u_j[\alpha_i z_{e(i)}\alpha_i', h]v_j\big\rangle = \\ = \sum_j\sum_i z_{e(i)}\alpha_i'[h, u_j v_j])\alpha_i = \sum_i z_{e(i)}\alpha_i'[h, \sum_j u_j v_j]\alpha_i\,.\end{aligned}$$

The final term lies in $I([h, \sum_j u_j v_j)$ as required in (3).

Any polynomial h can be written as $h^{\{0\}} + h^{\{1\}} + h^{\{>1\}}$, where $h^{\{0\}}$ does not depend on Z, every monomial in $h^{\{1\}}$ has Z-degree 1, and in $h^{\{>1\}}$ Z-degree bigger than 1. Let us have a general polynomial $\sum_j u_j[g,h]v_j$ as in (iii). Since $\langle\alpha\rangle = 0$ whenever α has Z-degree bigger than one, we can write

$$\begin{aligned}\big\langle \sum_j u_j[g,h]v_j\big\rangle = \big\langle \sum_j u_j[g^{\{0\}}, h^{\{0\}}]v_j\big\rangle + \big\langle \sum_j u_j^{\{0\}}[g^{\{1\}}, h^{\{0\}}]v_j^{\{0\}}\big\rangle + \\ + \big\langle \sum_j u_j^{\{0\}}[g^{\{0\}}, h^{\{1\}}]v_j^{\{0\}}\big\rangle\,.\end{aligned}$$

By (2), the first term is in $I([g^{\{0\}}, h^{\{0\}}])$. By (3}, the second term is in $I([h^{\{0\}}, \sum_j u_j^{\{0\}} v_j^{\{0\}}])$ and the third is in $I([g^{\{0\}}, \sum_j u_j^{\{0\}} v_j^{\{0\}}])$. This completes the proof of the condition (iii) and hence the proof of the lemma. □

Lemma 1 and Lemma 2 directly imply the following:

Theorem 3. *There exists a polynomial f in n variables such that $f^c = 0$ and $\mathrm{Q}(f) = \Omega(n^2)$. Moreover, f is of degree three and has coefficients from $\{0, 1, -1\}$.*

The statement of Lemma 2 is reminiscent of Baur-Strassen's algorithm for computing partial derivatives of a (commutative) polynomial. see[1]. This shows that, commutatively, the complexity of computing $z_1 f_1 + \dots z_n f_n$ is at least the complexity of computing $f_1, f_2, \dots, f_n$. However, the proof of Lemma 2 is quite different. We should also warn the reader that the lemma is a bit more intricate than it may appear. One may suspect that in order to have, for example, $z[x, y] \in I([g_1, g_1'], [g_2, g_2'], \dots)$ the ideal must contain $[x, y]$. However, note that $z[x, y] = [zx, y] - [z, y]x$ and so $z[x, y] \in I([zx, y], [z, y])$. For this reason, we fall short of proving the following generalization:

Let f_{ij}, $i \in \{1, \dots, p\}$, $j \in \{1, \dots, n\}$, be a set of np polynomials with $f_{ij}^c = 0$. Let $f_i := z_1 f_{i1} + \dots + z_n f_{in}$, where $z_1, \dots, z_n$ are fresh variables. Then

$$\mathrm{Q}(f_1, \dots f_p) \geq c.\mathrm{Q}(f_{ij};\, i \in \{1, \dots, p\},\, j \in \{1, \dots, n\})\,,$$

for a constant $c > 0$.

This version would be strong enough to give directly an explicit f with $\mathrm{Q}(f) = \Omega(n^2)$. For if f_{ij}, $i, j \in \{1, \dots, n\}$ are polynomials not depending on $z_1, \dots, z_n$, $y_1, \dots, y_n$, we obtain that $\mathrm{Q}(\sum_{i,j} z_i y_j f_{ij}) \geq c^2 \mathrm{Q}(f_{ij};\, i, j \in \{1, \dots, n\})$. Hence $\mathrm{Q}(\sum_{i,j} z_i y_j [x_i, x_j]) \geq c^2 \mathrm{Q}([x_i, x_j];\, i < j) = c^2 \binom{n}{2}$. This polynomial will be discussed later.

3 Questions

The main drawback of Theorem 3 is that it proves existence of f without constructing it explicitly. The first question therefore is:

Problem 1. Construct an explicit polynomial f in n variables such that $f^c = 0$ and $\mathrm{Q}(f)$ is superlinear (ideally, $\Omega(n^2)$).

Let us give some comments. First, if f has degree two then $\mathrm{Q}(f) \leq n$. Hence the candidate should be of degree at least three. In view of Theorem 3, looking for a degree-three polynomial seems to be the simplest choice. However, this has a different obstacle:

Proposition 4. *Assume that the underlying field has characteristic zero. Let f be a homogeneous polynomial of degree d such that $f^c = 0$ and*

$$f = \sum_{i \in \{1, \dots, m\}} g_{i,1} g_{i,2} \dots g_{i,d}\,,$$

where $g_{i,j}$ are linear. Then $\mathrm{Q}(f) \leq \binom{d}{2} m$.

Proof. Let I be the ideal generated by $[g_{i,j_1}, g_{i,j_2}]$ with $i \in \{1,\ldots,m\}$ and $j_1 < j_2 \in \{1,\ldots,d\}$. We want to show that $f \in I$. Let $f = \sum_{i_1,\ldots,i_d} a_{i_1\ldots i_d} x_{i_1} \ldots x_{i_d}$. For a permutation σ of $\{1,\ldots,d\}$, define $f^\sigma := \sum_{i_1,\ldots,i_d} a_{i_1\ldots i_d} x_{i_{\sigma(1)}} \ldots x_{i_{\sigma(d)}}$. The assumption that $f^c = 0$ implies

$$\sum_\sigma f^\sigma = 0\,,$$

where the summation ranges over all permutations of $\{1,\ldots,d\}$. On the other hand, we have

$$f^\sigma = \sum_{i\in\{1,\ldots,m\}} g_{i,\sigma(1)} g_{i,\sigma(2)} \cdots g_{i,\sigma(d)}\,,$$

which implies that $f - f^\sigma \in I$. Hence $\sum_\sigma (f - f^\sigma) \in I$. But $\sum_\sigma (f - f^\sigma) = \sum_\sigma f - \sum_\sigma f^\sigma = d!f$. Hence $d!f \in I$ and $f \in I$. □

If $d = 3$, the proposition tells us that f cannot be written as a sum of less than $\mathrm{Q}(f)/3$ products of linear forms. The latter problem corresponds to the rank of a dimension three tensor. In other words, if we construct a polynomial with superlinear $\mathrm{Q}(f)$ then we have also constructed a tensor of superlinear rank. To construct such a tensor is a well-known open question, which is probably more interesting than our Problem 1.

For degree-four polynomials the situation is slightly better. Motivated by Lemma 2, one may guess that the polynomial $\sum_{i,j} z_i y_j [x_i, x_j]$ is the right candidate for a solution to Problem 1. This polynomial has many variants, such as $\sum_{i,j} [x_i, x_j]^2$, but we shall focus on the following one:

$$C_n := \sum_{i,j\in\{1,\ldots,n\}} z_i x_j [x_i, x_j]\,.$$

We can show that $\mathrm{Q}(C_n)$ is superlinear assuming a conjecture from circuit complexity. C_n is related to the polynomial

$$\mathrm{I}D_n := \sum_{i,j\in\{1,\ldots,n\}} x_i x_j x_i x_j\,,$$

which was investigated in [5]. There, the authors hoped to prove that ID requires superlinear non-commutative arithmetic circuits, showing that this would imply an exponential lower bound on the circuit size of the non-commutative permanent. The lower-bound approach in [5] can be rephrased as follows. Define $S(n)$ as the smallest k so that there exist homogeneous degree-two polynomials $g_1,\ldots,g_k$ so that $\mathrm{I}D_n \in I(g_1,\ldots,g_k)$. Then $S(n)$ is a lower-bound to the circuit complexity of $\mathrm{I}D_n$. Moreover, the authors gatherered some evidence that $S(n)$ is superlinear and, optimistically, of the order $n^{2-o(1)}$. Let us show that if such an approach can be successful then C_n is a solution to Problem 1:

Proposition 5. *$\mathrm{Q}(C_n) \geq c.S(n)$, for a constant $c > 0$.*

Proof. For $i \in \{1,\ldots,n\}$ let $f_i = \sum_{j\in\{1,\ldots,n\}} x_j[x_i,x_j]$. By Lemma 2 it is sufficient to show that $\mathrm{Q}(f_1,\ldots,f_n) \geq cS(n)$. So assume that

$$f_1,\ldots,f_n \in I([g_1,g_1'],\ldots,[g_k,g_k'])\,.$$

Since $f_1,\ldots,f_n$ are homogeneous polynomials of degree three, we can assume that for every i, g_i, g_i' are homogeneous polynomials such that the sum of their degrees is at most three (this may cost a constant factor). The degree of each g_i, g_i' is at least one for otherwise $[g_i,g_i'] = 0$. This means that either both g_i, g_i' have degree 1, or one of them has degree 2. In both cases there exists a homogeneous degree-two polynomial h_i such that $[g_i,g_i'] \in I(h_i)$. (In the former case take $h_i := [g_i,g_i']$, in the latter set h_i is the degree two polynomial from $\{g_i,g_i'\}$). Hence we have

$$f_1,\ldots,f_n \in I(h_1,\ldots,h_k)\,,$$

with $h_1,\ldots,h_k$ homogeneous of degree two. Note that

$$f_i = \sum_j (x_jx_ix_j - x_j^2x_i) = \sum_j x_jx_ix_j - (\sum_j x_j^2)x_i\,.$$

Denoting $v_i := \sum_j x_jx_ix_j$, this means that

$$v_1,\ldots,v_n \in I(h_1,\ldots,h_k,\sum_j x_j^2)\,.$$

Since $\mathrm{I}D_n = \sum_i x_iv_i$, we also have $\mathrm{I}D_n \in I(h_1,\ldots,h_k,\sum_j x_j^2)$ and so $S(n) \leq k+1$. □

For polynomials of unbounded degree, the range of apparently good candidates is unlimited. A guess would be $\prod_{i,j}([x_i,x_j]+1)-1$, or $\sum_\sigma \mathrm{sgn}(\sigma)x_{\sigma(1)\ldots x_{\sigma(n)}}$. Moreover, note that any correct identity between commutative polynomials gives rise to a potential candidate. For example, take $\det(XY) = \det(X)\det(Y)$. This identity holds when the variables commute but is false in the non-commutative setting. This means that $f = \det(XY) - \det(X)\det(Y)$ is a non-trivial non-commutative polynomial, $f^c = 0$, and one may expect that $\mathrm{Q}(f)$ is fairly large. The problem here is not the lack of candidates, but the lack of proof techniques.

Since Problem 1 already appears difficult, one hesitates to present the more challenging Problem 2. However, this problem would have interesting applications:

Problem 2. Find a polynomial f such that $f^c = 0$, f can be computed by a non-commutative arithmetic circuit of size s (or better, a formula) and $\mathrm{Q}(f)$ is superlinear in s

For a background on non-commutative arithmetic circuits, see [8, 5]. The main difference between Problems 1 and 2 is that here we want a superlinear bound in

terms of the circuit size of s. Since $\mathrm{Q}(f) \leq \binom{n}{2}$, such an f must be computable by a circuit of subquadratic size. This severely restricts the range of candidates. Proposition 4 indicates that f should not have too small depth-three circuit and, more importantly, that Problem 2 has *no* solution for degree $d = 3$ – for the m in Proposition 4 is also a lower bound to the circuit size. Proposition 5 is even more embarrasing. The circuit size of C_n is at least $S(n)$ and so the assumption which guarantees that $\mathrm{Q}(C_n)$ is large also implies that the circuit complexity of C_n is large. It may happen that Problem 2 does not have a solution for a rather banal reason: it may turn out that non-commutative polynomials computable be subquadratic circuits are extremely simple objects, simple enough to have linear Q.

Finally, let us address a detail that we omitted in the Introduction. There we claimed that a formula or a circuit can be interpreted as a non-commutative formula or a circuit, which is not entirely true. In the usual definition of a commutative circuit C, there is no distinction between multiplication from left or right. Hence such a circuit implicitly assumes commutativity relations $gh = hg$, where g, h are inputs of a product gate in the circuit. We may arbitrarily fix the order of multiplication in C to obtain a proper non-commutative circuit C'. Every such circuit may compute a different non-commutative polynomial, from some family $\mathcal{F}$. However, all polynomials in $\mathcal{F}$ are equivalent modulo the ideal generated by the commutativity conditions imposed by the gates of C. The number of such conditions is at most the size of the circuit C. Hence $|\mathrm{Q}(f_1) - \mathrm{Q}(f_2)|$ is at most the size of C, for every $f_1, f_2 \in \mathcal{F}$, and so if one polynomial from $\mathcal{F}$ has superlinear commutative complexity then the same holds for every polynomial in $\mathcal{F}$.

4 Proof complexity

The commutative complexity of f corresponds to the number of commutativity axioms $gg' = g'g$ one needs to use in order to prove that $f^c = 0$. It is this interpretation which makes $\mathrm{Q}(f)$ relevant to the study of complexity of proofs. The first connection is immediate. In [3] and [6], there were introduced proof systems proving polynomial identities. In [6], the authors focus on two dominant systems, called *arithmetic Frege* and *arithmetic circuit Frege*. Arithmetic Frege proves equations of the form $F = G$, with F, G arithmetic formulas, and where $F = G$ is provable iff F, G compute the same *commutative* polynomial. The inferences in arithmetic Frege are syntactic operations on formulas, such as $F(G+H) = FG+FH$, $FG = GF$, ..., corresponding to the defining axioms of commutative rings. One can as well consider just equations of the form $F = 0$, asserting that F computes the zero polynomial. The main open question in this context is to show that there exist correct equations $F = 0$ which require long arithmetic Frege proofs. That is, we want to find an arithmetic formula F of size s such that F computes the zero polynomial but every proof of $F = 0$ in arithmetic Frege must have size superpolynomial in s. The truth is that the best known lower bound is $\Omega(s^2)$ on the proof size, and $\Omega(s)$ on the number of

proof-lines (i.e., the number of inferences)[2]. Arithmetic circuit Frege is similar to arithmetic Frege except that it uses circuits instead of formulas. There, the best lower-bound on either the size or number of inferences is only linear.

One option, how to obtain a superlinear bound, is to count the number of commutativity axioms needed to prove $F = 0$. Recall that F is an arithmetic formula such that F, when interpreted commutatively, computes the zero polynomial. However, when we interpret F as a non-commutative circuit then F computes a non-commutative polynomial $\hat{F}$. This non-commutative polynomial is in general non-zero but satisfies $\hat{F}^c = 0$. The following is immediate from the definition of the proof systems mentioned:

Observation 1. Any arithmetic Frege or arithmetic circuit Frege proof of $F = 0$ has at least $\mathrm{Q}(\hat{F})$ proof-lines.

This means that a positive solution to Problem 2 implies a superlinear lower bound on proof-size. More exactly, to obtain a lower bound on arithmetic Frege, we need Q to be superlinear in terms of the formula size, and for circuit Frege, it is sufficient in terms of the circuit size.

Propositional proof systems

A similar line of reasoning can be applied to lower bounds on propositional Frege or extended Frege proofs. We refer the reader to [2, 7] for background on propositional proof systems. Propositional proof systems operate with Boolean formulas and prove Boolean tautologies. We find it convenient to view a Boolean formula as an arithmetic formula over the field $GF(2)$. A formula F is a tautology if $F = 0$ for any $0,1$-assignment to the variables of F. The challenging problem is to find a tautology of size s which requires superpolynomial size proof in Frege or extended Frege. The best know lower bounds on size of Frege proofs is quadratic, and linear when one counts the number of inferences. For extended Frege, the lower bound is linear in both cases.

When we think of a Boolean formula F as computing a *commutative* polynomial f, the requirement that F is a tautology is equivalent to the assumption that f lies in the ideal generated by $x_1^2 + x_1, \ldots, x_n^2 + x_n$ for the variables in F. We may again again view F as computing a *non-commutative* polynomial $\hat{F}$. Here, the assumption that F is a tautology is no longer equivalent to $\hat{F} \in I(x_1^2 + x_1, \ldots, x_n^2 + x_n)$. This is witnessed by the identity

$$(g + h)^2 + (g + h) = g^2 + g + h^2 + h + [g, h]\,, \tag{4}$$

where we write $g, h = x_1, x_2$. The tautology $(x_1 + x_2)^2 + (x_1 + x_2)$ does not lie in the ideal generated by $x_1^2 + x_1, x_2^2 + x_2$ but rather in the ideal $I(x_1^2 + x_1, x_2^2 + x_2, [x_1, x_2])$ – which illustrates the relevance of the commutativity axiom in the Boolean setting. The condition that F is a tautology can be stated in two different ways:

[2]Those bounds follow from bounds on propositional Frege and Extended Frege, mentioned below.

(*i*). There exist $g_1, \ldots, g_k$ such that $\hat{F} \in I(g_1^2 + g_1, \ldots, g_k^2 + g_k)$.

(*ii*). There exist $g_1, g_1' \ldots, g_k, g_k'$ such that $\hat{F} \in I([g_1, g_1'], \ldots, [g_k, g_k'], x_1^2 + x_1, \ldots, x_n^2 + x_n)$, where $x_1, \ldots x_n$ are the variables of F.

Let us denote the smallest k such that (i) holds, by $R(\hat{F})$, and the smallest k such that (ii) holds by $\mathrm{Q}^*(\hat{F})$. As before, we can see that $\mathrm{Q}^*(\hat{F}) \leq \binom{n}{2}$. Exploiting the identity (4) shows that $R(\hat{F}) \leq 3\mathrm{Q}^\star(\hat{F}) + n$ and also $R(\hat{F}) \leq \binom{n}{2} + n$. We see no obvious reason why either of the quantities should be linear in terms of the formula size of F. Hence we could potentially obtain a superlinear lower bounds as follows:

Observation 2. The number of proof-lines in an extended Frege proof of a tautology F is at least $\Omega(R(\hat{F}))$ and its size is at least $\Omega(\mathrm{Q}^*(\hat{F}))$.

References

[1] W. Baur and V. Strassen. The complexity of partial derivatives. Theoretical Computer Science, 22:317330, 1983.

[2] S. Cook and R. A. Reckhow. The relative efficiency of propositional proof systems. Journal of Symbolic logic, 44, 36-50, 1979.

[3] P. Hrubeš and I. Tzameret. Proof complexity of polynomial identities. *CCC '09: Proceedings of the 2009 24th Annual IEEE Conference on Computational Complexity*, 41–51, 2009.

[4] P. Hrubeš and Amir Yehudayoff. Arithmetic complexity in algebraic extensions. To appear in TOC.

[5] P. Hrubeš, A. Wigderson and A. Yehudayoff. Non-commutative circuits and the sum of squares problem. *Proceedings of the 42nd ACM symposium on Theory of computing*, 667-676, 2010.

[6] P. Hrubeš and I. Tzameret. Proving properties of the determinant. A manuscript.

[7] Jan Krajíček. Bounded arithmetic, propositional logic, and complexity theory. Cambridge University Press, USA, 1995.

[8] N. Nisan. Lower bounds for non-commutative computation. Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, pages 410418, 1991.

[9] H. E. Warren. Lower bounds for approximations by nonlinear manifolds. Trans. AMS 133, pages 167–178, 1968.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de