Trade-Offs Between Size and Degree in Polynomial Calculus

Guillaume Lagarde
LaBRI, Bordeaux, France
https://guillaume-lagarde.github.io/
guillaume.lagarde@labri.fr

Jakob Nordström
University of Copenhagen, Denmark
KTH Royal Institute of Technology, Stockholm, Sweden
https://www.csc.kth.se/~jakobn/
jn@di.ku.dk

Dmitry Sokolov
Lund University, Sweden
University of Copenhagen, Denmark
https://www.csc.kth.se/~sokolovd
sokolov.dmt@gmail.com

Joseph Swernofsky
KTH Royal Institute of Technology, Stockholm, Sweden
https://www.kth.se/profile/josephsw/
josephsw@kth.se

Abstract

Building on [Clegg et al. '96], [Impagliazzo et al. '99] established that if an unsatisfiable k-CNF formula over n variables has a refutation of size S in the polynomial calculus resolution proof system, then this formula also has a refutation of degree $k + O(\sqrt{n \log S})$. The proof of this works by converting a small-size refutation into a small-degree one, but at the expense of increasing the proof size exponentially. This raises the question of whether it is possible to achieve both small size and small degree in the same refutation, or whether the exponential blow-up is inherent. Using and extending ideas from [Thapen '16], who studied the analogous question for the resolution proof system, we prove that a strong size-degree trade-off is necessary.

2012 ACM Subject Classification Theory of computation → Proof complexity

Keywords and phrases proof complexity, polynomial calculus, polynomial calculus resolution, PCR, size-degree trade-off, resolution, colored polynomial local search

Digital Object Identifier 10.4230/LIPIcs.ITCS.2020.72

Acknowledgements The first, second, and fourth authors were funded by the Knut and Alice Wallenberg grant KAW 2016.0066 and the third author by the Knut and Alice Wallenberg grant KAW 2016.0433. In addition, the second author was supported by the Swedish Research Council grants 621-2012-5645 and 2016-00782.

1 Introduction

The main task of proof complexity is to quantify the amount of different resources required to prove that some given formula is unsatisfiable. The particular resources examined depend on the proof system under study, but it is believed that no proof system can have proofs that are both efficiently verifiable and short – that is, polynomial in the size of the given formula. Establishing such an impossibility result in full generality is equivalent to proving

NP $\neq$ coNP, and hence this goal currently seems out of reach. Current research in proof complexity instead focuses on studying more limited, concrete methods of reasoning, and on proving lower bounds for these methods.

In this paper we consider **polynomial calculus resolution (PCR)**. This is a more powerful proof system than what is arguably the most well studied system in all of proof complexity, namely **resolution** [12], but is therefore also less well understood. PCR, introduced by Clegg et al. [15] and Alekhnovich and Razborov [1] can be seen as a dynamic version of Hilbert's Nullstellensatz [6]. In order to refute a CNF formula in PCR, the clauses in the formula are translated into multilinear polynomials, and the proof of unsatisfiability, or **refutation**, consists essentially of certifying that the polynomial 1 is a member of the ideal generated by these polynomials.

Two important masures for PCR refutations are the **size** (the number of monomials in the refutation when all polynomials are expanded out as linear combinations of monomials) and the **degree** (the maximal monomial degree in the refutation). Impagliazzo et al. [24] showed a strong connection between these two measures: if a k-CNF formula over n variables admits a PCR refutation of size S then there is also a refutation of degree $k + O(\sqrt{n \log S})$. This result, which is known by [21] to be tight, plays a crucial role in almost all known size lower bounds for PCR. By proving strong enough degree lower bounds one can also obtain size lower bounds, and techniques for establishing lower bounds on degree have been developed in, e.g., [2, 21, 20, 26]. An interesting aspect of [24], however, is that the small-size refutation is not the same as the small-degree one. Instead, the transformation from small size to small degree increases the proof size exponentially. It is natural to ask whether this exponential blow-up is nescessary.

A similar question arises also in the context of the resolution proof system, where the measures of interest are length and width. Building on [24], Ben-Sasson and Wigderson [10] showed that every small-length resolution refutation can be transformed into a small-width refutation with the same parameters as in [24]. Again, this bound is again known to be tight [13], and the conversion increases the length exponentially. For resolution, Thapen [29] proved that such a blow-up cannot be avoided in the worst case. In this work, we show that this holds true for PCR as well.[1] More precisely we prove the following theorem.

▶ **Theorem 1.** *For any $\epsilon > 0$ and c large enough, there is a CNF formula φ with $\theta(c^{1+\epsilon})$ variables, of size $\theta(c^{1+\epsilon})$ and degree $O(\log c)$ such that*

- *φ has a PCR refutation of size polynomial in c and degree $c + O(\log c)$,*
- *φ has a PCR refutation of degree $O(c^{\epsilon})$,*
- *any PCR refutation of degree strictly less than $c - 1$ has size $\exp\big(c^{\Omega(\epsilon)}\big)$.*

In particular, this implies that any PCR refutation of the formula φ in degree $O(\log c) + O(\sqrt{c^{1+o(1)} \log c}) \ll c$, as obtained by the size-degree bound in [24], must have size $\exp\big(c^{\Omega(\epsilon)}\big)$.

1.1 Related work

The study of connections between different complexity measures has received a fair amount of attention in proof complexity. We give a brief overview below, referring the reader to the book [25] or the upcoming survey chapter [14] for more details.

[1] Note that this is more precise than just saying that there is a trade-off between length and width, or between size and degree, so that minimizing the latter measure must lead to an increase in the former. It is easy to show that there are trade-offs between length/size and width/degree as observed in [27], but such trade-offs are very far from being strong enough to show that the blow-ups in [10, 24] is necessary.

We have already mentioned connections between size and degree in PCR [24] and between length and width in resolution [10]. Let us also mention that in [3] a beautiful relation is exhibited between width and **space** in resolution, showing that width provides a lower bound on space, and that a similar, but weaker, result was recently proved in [19] for PCR.

When studying trade-offs, we are asking whether two different complexity measures can be minimized at the same time, or whether optimizing one measure must lead to an increase in the other in the worst case. This question was first raised in the context of proof complexity by Ben-Sasson [8], who proved trade-offs between space and width in resolution. This result was later extended to size-degree trade-offs in PCR [7], and the result for resolution was extended to a wider regime of parameters in [11].

Trade-offs between length/size and space have been shown for resolution in [5, 9] and for PCR in [7]. There are even some size-space trade-offs for stronger proof systems such as **cutting planes** [16] in [18, 22, 23], but since this proof system will not be relevant for the current paper we will not discuss it further.

For length versus width in resolution, or size versus degree in PCR, it is not too hard to show that there are trade-offs [27], but to prove that the length blow-up in [10] is necessary requires the stronger result in [29]. Recently, Razborov [28] established another length-width trade-off showing that low-width *tree-like* resolution refutations of certain k-CNF formulas must have doubly exponential size,

For Nullstellensatz, which as noted above can be viewed as a weaker version of PCR, strong trade-offs between size and degree were shown in [17]. That paper states as an open problem whether similarly strong results could be established for PCR, and this is the problem that we resolve here.

1.2 Organization of the paper

We start with preliminaries in Section 2 where we define the relevant proof systems and discuss some basic properties of ideals in polynomial rings. In Section 3, we present the family of CNF formulas that we will consider, which we call *safe colored polynomial local search* formulas. Section 4 is devoted to a description of the overall proof strategy. Section 5 develops the machinery used to prove PCR degree lower bounds together with the adaptations needed to obtain our results. Section 6, finally, contains some concluding remarks. We refer to the upcoming full-length version of the paper for the details missing in this extended abstract.

2 Preliminaries

Throughout this paper, we let $\mathbb{F}$ denote a fixed but arbitrary field. For a an integer, we denote by $[a]$ the set $\{0, 1, \ldots, a-1\}$. Given a graph H and a vertex $v \in H$, we write $N(v)$ for the set of neighbours of v in the graph H.

2.1 Resolution and Polynomial calculus

We use x to denote boolean variables ranging over $\{0, 1\}$, and write $\bar{x}$ to denote the negation of x. A **literal** is either a variable x or its negation $\bar{x}$. A **clause** C is a disjunction of literals, such as $x \vee \bar{y}$. The **width** of a clause is the number of literals in it. A **CNF formula** F is a conjunction of clauses, and F is a **k-CNF formulas** if all clauses in it have width at most k. Clauses and formulas are considered as sets, so that there are no repetitions and the ordering is immaterial.

The **resolution** proof system has a single derivation rule $\frac{C \vee x \quad D \vee \bar{x}}{C \vee D}$, called the **resolution rule**, for any clauses C, D and any variable x. In this proof system, a **derivation** of a clause D from a CNF formula $F = C_1 \wedge \cdots \wedge C_m$ is a sequence of clauses such that each clause is either an original clause C_i or the conclusion of the resolution rule applied to previously derived clauses. A **resolution refutation** of F is a derivation of the empty clause containing no literals from F. The **length** of a derivation, or refutation, is the length of the sequence of clauses, and the width is the maximum width of any clause appearing in the sequence.

Moving on to polynomial calculus, we let X denote a set of algebraic variables containing x and $\bar{x}$ for all boolean variables, where x and $\bar{x}$ are considered to be distinct, and consider polynomials in the ring $\mathbb{F}[X]$. We note that in the context of algebraic proof systems it is natural to think of 0 as **True** and 1 as **False**, and so we will use this convention in what follows.

The **PCR** proof system contains the following axioms:

- **boolean axioms:** $x^2 - x$ for all variables x;
- **complementary axioms:** $x + \bar{x} - 1$ for all variables x.

It also contains the following derivation rules:

- **linear combination:** $\frac{p \quad q}{\alpha p + \beta q}$ for any $\alpha, \beta \in \mathbb{F}, p, q \in \mathbb{F}[X]$;
- **multiplication:** $\frac{p}{xp}$ for any $p \in \mathbb{F}[X]$ and $x \in X$.

A polynomial f is **derivable** from a set of polynomials $g_1, \ldots, g_k$ (written $g_1, \ldots, g_k \vdash f$) if there is a sequence of polynomials such that each polynomial is either an axiom, an original polynomial g_i, or the conclusion of a derivation rule applied to previously derived polynomials. We refer to such a sequence of polynomials as a **PCR derivation**. A **PCR refutation** of a set of polynomials $g_1, \ldots, g_k$ is a derivation of the polynomial 1 from $g_1, \ldots, g_k$.

▶ **Example 2.** Over the variables x, y, z, if we are given xz and $y\bar{z}$ we can derive xy. This is a simulation, in PCR, of the resolution rule deriving $x \vee y$ from $x \vee z$ and $y \vee \bar{z}$.

$$\dfrac{\dfrac{\dfrac{y\bar{z}}{xy\bar{z}} \quad \dfrac{z + \bar{z} - 1}{xyz + xy\bar{z} - xy}}{xy - xyz} \quad \dfrac{xz}{xyz}}{xy}$$

A set I of polynomials in $\mathbb{F}[X]$ is an **ideal** if I is closed under linear combination and multiplication by any polynomial in $\mathbb{F}[X]$. Given a set of polynomials $S = \{g_1, \ldots, g_k\}$, the **ideal generated by** S is defined as the smallest ideal I_S that contains the set S.

Observe that, by definition, $g_1, \ldots, g_k \vdash f$ is equivalent to saying that f is in the ideal generated by $g_1, \ldots, g_k$ together with all boolean and complementary axioms. Intuitively, a PCR refutation is a certificate that the system of polynomial equations $\{g_1 = 0, \ldots, g_k = 0\}$ has no boolean solutions. If the polynomials have no common boolean solution then there always exists such a certificate, since it can be shown that 1 lies in the ideal generated by the polynomials arising in the system together with the polynomials from the boolean and complementary axioms. In other words, the PCR proof system is *sound* and *complete*.

The **size** of a PCR refutation is the total number of non-zero monomials (counted with repetition) that appear in the derivation when all polynomials are expanded out as linear combinations of monomials. The **degree** of a PCR refutation is the maximal degree of a non-zero monomial that appears in the derivation.

There is a standard **translation** $tr(\cdot)$ from clauses to monomials defined by induction in the following way:

- $tr(x) = x$;
- $tr(\neg x) = \bar{x}$;
- $tr(C \vee D) = tr(C) \cdot tr(D)$.

Using $tr(\cdot)$, any k-CNF formula F with s clauses can be converted into a set of s polynomials of degree at most k by applying the translation to all clauses in F. Denote this set by $tr(F)$. It is not hard to see that F is satisfiable if and only if the set of polynomials $tr(F)$ have a common boolean root. Therefore F is an unsatisfiable CNF formula if and only if there is a PCR refutation of the set of polynomials $tr(F)$. Furthermore, a straightforward generalization of Example 2 shows that a resolution refutation in length ℓ and width w can be converted into a PCR refutation in size $O(\ell w)$ and degree $w+1$.

A **restriction** is a partial assignment of the variables, that is a function $\rho : X \to X \cup \mathbb{F}$ such that the value $\rho(x)$ is either x or a constant from $\mathbb{F}$. For a polynomial p, we denote by $p \restriction \rho$ the polynomial p in which any variable x is replaced by $\rho(x)$.

2.2 Reduction over ideals

Two polynomials p, q are said to be **equivalent modulo an ideal** I, written $p \sim_I q$, if $p - q \in I$. This is an equivalence relation. For any polynomial p we fix a special representative of the equivalence class $[p]$ that we call the **reduction of** p **modulo** I and write as $R_I(p)$. If an ideal I is generated by a set of polynomials S, we abuse notation slightly and write $R_S(p)$ for $R_I(p)$.

To define the representative, we fix any order $\prec$ on the polynomials that respects inclusion as follows:

1. Firstly, for two monomials m_1 and m_2 we have $m_1 \prec m_2$ whenever m_1 is a submonomial of m_2.
2. Secondly, we extend this in an arbitrary way to a total order on monomials.
3. Finally, we order polynomials in the following way. To any polynomial p we associate a sequence s_p consisting of the non-zero monomials of p, sorted in decreasing order with respect to $\prec$. The polynomials are then compared by comparing lexicographically their associated sequences. For example, if $m_1 \prec m_2 \prec m_3$, then the polynomial m_2 is strictly smaller than the polynomial $m_2 + m_1$, which is itself strictly smaller than the polynomial m_3.

The representative $R_I(p)$ is then defined as $\min(\{q \in [p]\})$. We now observe two easy but useful properties of $\prec$.

▶ **Fact 3.** *For any restriction ρ, we have that $p \restriction \rho \preceq p$.*

Proof. To see this, observe that any monomial in $p \restriction \rho$ is either a monomial or a submononial from p; this can only decrease p. ◀

▶ **Fact 4.** *If I_1, I_2 are two ideals such that $I_1 \subseteq I_2$, then*

$$R_{I_2}(p \times R_{I_1}(q)) = R_{I_2}(p \times q).$$

Proof. By definition, $q = R_{I_1}(q) + h$ for some $h \in I_1$. Therefore

$$\begin{aligned} R_{I_2}(p \times q) &= R_{I_2}(p \times (R_{I_1}(q) + h)) \\ &= R_{I_2}(p \times (R_{I_1}(q)) + R_{I_2}(p \times h) \end{aligned}$$

by linearity. To conclude, observe that $p \times h \in I_1$ so is reduced to zero modulo I_1, and hence is reduced to zero modulo I_2 since $I_1 \subseteq I_2$. ◀

3 The formula

In this section, we describe the construction of the family of formulas for which we establish our size-degree trade-off result. We also argue why they have small-size refutations. Our formulas encode a modified version of the **colored polynomial local search (CPLS)** principle, and are very similar to the ones used in [29] to prove trade-offs between length and width in resolution, but there are two main differences. First, we add an *extra color* that we call the *safe color* that plays a role slightly different from the normal ones. The reason for adding this color is purely technical – we were not able to make the PCR machinery works without it. Second, instead of using a grid graph where any two consecutive layers form a complete bipartite graph, we restrict the edges between two consecutive layers to be a well-chosen expander graph. Since this makes the formula more constrained than that in [29], it makes our lower bound slightly stronger.

3.1 The modified CPLS formula

Let a, b, c be positive integers. We work only with graphs $H = (V, E)$ of the following form:

- The set V of vertices are given by $\{(i, x), i \in [a], x \in [b]\}$. We say a vertex (i, x) is the vertex x on layer i.
- Edge appear only between consecutive layers. I.e., if $((i, x), (i', x'))$ is an edge then $i' = i + 1$.

To any such graph H we associate a formula $CPLS^H(a, b, c)$ (or simply $CPLS(a, b, c)$ if H is clear from the context). Intuitively, the formula will give a set of colors (which is a subset of $[c]$) to each node in H according to the following rules.

1. Node $(0, 0)$ gets no color.
2. From every node u there is some special neighbor v on the next layer. If v gets a color then u gets a color. Specifically, there is a **safe color** (corresponding to the first color from $[c]$) and either u gets the safe color or it gets all of v's colors.
3. Every node on the bottom layer gets some color.

Of course, this means there is a path of special neighbors from $(0, 0)$ to the bottom layer. The last node must get some color, so we can trace backward and see $(0, 0)$ gets some color, a contradiction. Hence $CPLS(a, b, c)$ is unsatisfiable. Now we state this formally. We use $\Theta(abc)$ variables:

- $G(u, y)$ for each $u \in V$ and $y \in [c]$. This says whether y is set at u.
- $f(u, v)$ for each $u \in V$ and $v \in N(u)$. This says whether v is a special neighbor of u.
- $h(u)_j$ for each $u = (a - 1, x)$ and $j \in [\log c]$. The function $h(u)$ identifies a color that must be present at u, but is encoded in binary.

Now we can restate the intuitive rules above as formal axioms:

1. for each color $y \in [c]$,
   $\neg G((0, 0), y)$;
2. for each node $u \in V$
   (a) for each neighbor $v \in N(u)$, and color $y \in [c]$,
   $(f(u, v) \wedge G(v, y)) \to G(u, y) \vee G(u, 0)$;
   (b) if u is not on the bottom layer,
   $\bigvee_{v \in N(u)} f(u, v)$;
3. for each node $u \in V$ on the last layer and color $y \in [c]$,
   $(h(u) = y) \to G(u, y)$.

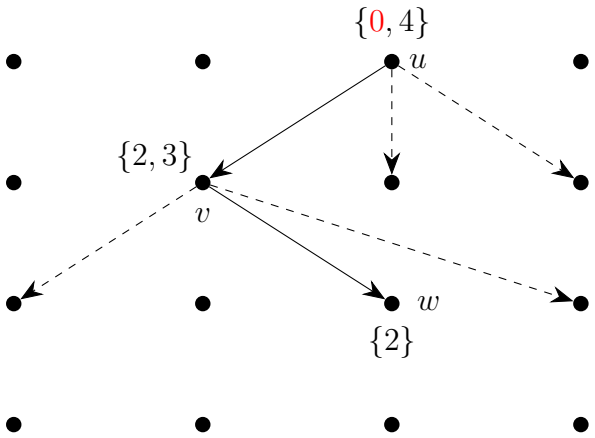

Figure 1 Nodes u and v have three possible neighbors represented by the arrows. The two arrows that are not dotted represent the actual values of f; for these arrows, axioms 2a are respected. Indeed the axioms $P_{v,w,\cdot}$ are satisfied because the colors at v is a superset of the colors at w; the axioms $P_{u,v,\cdot}$ are satisfied because u contains the safe color 0.

See Figure 1 for an illustration of the formula. We refer to instances of axioms 2a, 2b, and 3 as $P_{u,v,y}, Q_u$, and $T_{u,y}$ respectively.

▶ **Definition 5.** *An (s,e)-**expander** is a bipartite graph (V_L, V_R, E) such that for any subset of "left" vertices $S \subseteq V_L$ such that $|S| \leq s$, the following holds: $|N(S)| \geq e|S|$. We recall that $N(S)$ is the neighborhood of S.*

We will use $a = b = c^\epsilon$ in this paper. Let H_i be H restricted to the i^{th} layer. That is $H_i \coloneqq H \cap \{(i,x),(i+1,y) \mid x,y \in [b]\}$. We need H_i to be an $(a^{7/8}, e)$-expander, where any constant $e > 1$ suffices. We also need H_i to have vertex degrees bounded above and below by some constants. For simplicity we assume it has constant degree d.

▶ Remark 6. Note that it is also possible to ask each H_i to be a full bipartite graph. Because the vertex degree then grows with a, we need to change the formula by encoding the neighbors of any vertex in "binary", as in Thapen's paper [29], in order to avoid the width becoming too large (due to type 2b axioms).

3.2 Short and narrow refutations

We give a refutation of $CPLS^H(a,b,c)$ which is of small size. While this refutation can be translated into a refutation of small degree, by a result from [24], we additionally give a much lower degree refutation. The rest of the paper will consist of proving that small size and small degree cannot be obtained simultaneously. The proof of the next two propositions can be found in the appendix of the full version and are very similar to the one from Thapen [29].

▶ **Proposition 7.** *For any graph H, $CPLS^H(a,b,c)$ has a resolution refutation Π of length $O(ab^2c)$ and width $c + \log b + 1$.*

As discussed in Section 2, this gives a PCR refutation of size $O(a^2b^4c^2)$ and degree $c+\log b+2$.

▶ **Proposition 8.** *For any graph H, $CPLS^H(a,b,c)$ has a resolution refutation Π of width $a(\mathrm{d}+1) + \log c$.*

4 Restricting the formula and refutation

The strategy we are going to use is as follows: we suppose for contradiction that we have a refutation Π of the formula $CPLS$ which is both of small size and small degree. Then we will apply a random restriction ρ to both $CPLS$ and Π, which is a partial assignment to the variables. At this point, we have in hand a refutation $\Pi \restriction \rho$ of the restricted formula $CPLS \restriction \rho$. By the assumption that Π is both of small size and small degree, we show that $\Pi \restriction \rho$ enjoys some nice properties; we call such a refutation a **beautiful** refutation. We conclude by showing that $CPLS \restriction \rho$ cannot have any beautiful refutation, yielding the desired contradiction.

We start with the definition of ρ.

4.1 The restriction

Let $p = a^{-3/4}$ and $w = a^{7/8}$. The restriction ρ is randomly set in the following way.

- With probability p, for any vertex u, we set independently for all $y \in [1, c-1]$ the value of $G(u, y)$ to True or False with probability $1/2$; if this happens for a vertex u, we say that **the colors at u are set**. Moreover, if such a vertex is at the bottom layer, then we select one color y that has been set to True during the process and we set $h(u) = y$.
- If we set the colors at u then we also give the safe color to this vertex, i.e., we set $G(u, 0) =$ True. Otherwise we set $G(u, 0) =$ False.
- If we set the colors at u then with probability $\frac{1}{2}$ we also set $f(u, v) =$ True for a uniformly random chosen vertex $v \in N(u)$ and we set all others values $f(u, \cdot)$ to zero.

For the rest of the paper, we write this restriction ρ.

▶ Remark 9. Note that for any u where the colors are set, u gets the safe color. Hence we satisfy all type 2a axioms mentioning $f(u, \cdot)$ and we can treat the vertex u as if it is removed from the graph. For any node v with an edge (v, u) in the graph this reduces the (out-)degree of v by 1. We must maintain a positive degree for vertices outside the last layer. If vertices have degree at least 3 then the chance of reducing any vertex degree to 0 is $o(1)$. Hence our random choice of ρ works with high probability in this case. It is still a $CPLS$ formula (over a smaller graph) and we can now reason over this new graph.

4.2 Beautiful properties after restriction

In this section, we prove that applying the above restriction ρ to a small sized refutation Π gives a refutation $\Pi \restriction \rho$ of $CPLS \restriction \rho$ enjoying useful properties.

▶ **Definition 10.** *A term t **touches** vertex $u \in H$ iff t contains at least one of the following:*

- *$G(u, y)$ for some $y \in [c] \setminus \{0\}$;*
- *$f(u, v)$ for some $v \in N(u)$;*
- *$h(u)_j$ for some $j \in [\log c]$.*

*The **vertex-degree** of a term t is the number of vertices it touches.*

See Figure 2 for an example.

▶ **Definition 11.** *We say that a term t is **almost beautiful** if:*

- *the vertex-degree of t is at most $w + 1$;*
- *it does not contain any $G(\cdot, 0)$ variable;*
- *it touches at most $c - 2$ different colors.*

*If the vertex-degree of t is at most w then we say it is **beautiful**. By extension, a derivation Π is beautiful if every term that appears in Π is beautiful.*

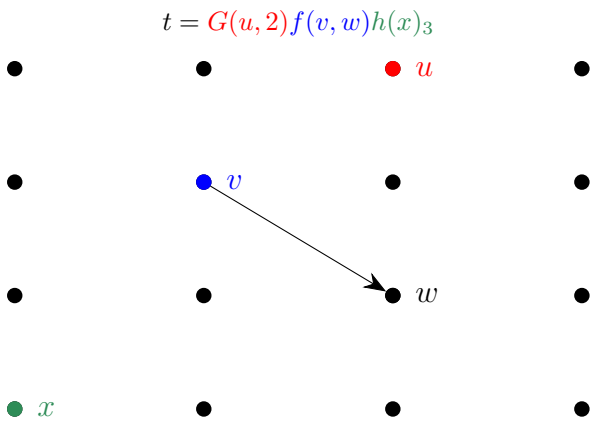

Figure 2 The term t touches three vertices that are the ones colored in the figure.

▶ **Proposition 12.** *There is an ϵ so that if $|\Pi| = s \leq 2^{c^\epsilon}$ then with probability $1 - s \cdot 2^{-c^\epsilon}$ over the choice of ρ, every term in $\Pi \restriction \rho$ is beautiful.*

Proof. Consider a term t and suppose it touches at least w distinct vertices u. We show that except with exponentially small probability ρ sets t to 0.

Suppose we set the colors at u, which happens with probability p. If t contains $G(u,y)$ for $0 < y$ then with probability $\frac{1}{2}$ we set $G(u,y) = 0$. If t contains $h(u)_j$ then because $h(u)$ is uniformly random we set $h(u)_j = 0$ with probability $\frac{1}{2}$.

We also set the arrow at u with probability $p/2$. There are d neighbors of u so we set $f(u,v)$ to True with probability $\frac{1}{d}$ and False with probability $\frac{d-1}{d}$ Hence if t contains $f(u,v)$ then we set t to 0 with probability at least $\frac{p}{2d}$.

Hence, for each u, if t touches u then with probability at least $\frac{p}{2d}$ we set t to 0. For distinct u these events are independent so the probability of not setting t to 0 is at most:

$$(1 - \frac{p}{2d})^w < e^{-\frac{1}{2d}a^{1/8}}.$$

To conclude the proof, observe that $t \restriction \rho$ does not contain any $G(\cdot, 0)$ variable since the restriction sets the value $G(u,0)$ to True of False for any vertex u. ◀

5 PCR machinery

In this section, we ask the order on monomials to have the additional property that it respects the vertex-degree: if a monomial m_1 has a vertex-degree smaller than a monomial m_2, then $m_1 \prec m_2$.

5.1 R operator

We follow a strategy similar in spirit to the technique developed by Alekhnovich and Razborov [2]. The idea is to define a linear map $R : \mathbb{F}[X] \to \mathbb{F}[X]$ that acts as a witness that the polynomial 1 is not reachable from the axioms by any beautiful derivation. Such a linear map was constructed by Alekhnovich and Razborov to separate the polynomial 1 from polynomials reachable by small degree derivations. It has been applied to Tseitin tautologies, among other formulas. Note that Mikša and Nordström [26] gave some sufficient (and quite general) conditions to prove existence of such operators. Unfortunately, we can not use directly these results since they deal with degree and we therefore need to adapt them to handle our notion of beautifulness. More precisely, we aim to prove the following theorem.

▶ **Theorem 13.** *There is a linear map $R : \mathbb{F}[X] \to \mathbb{F}[X]$ such that:*

1. $R(A) = 0$ *for any axiom A;*
2. $R(1) \neq 0$*;*
3. $R(xt) = R(xR(t))$ *whenever xt is a beautiful monomial.*

Theorem 13 is sufficient to prove our main theorem, since any polynomial derived by a beautiful derivation is mapped to zero under the R operator, leaving the polynomial 1 unreachable.

The operator R that we will use is defined on monomials and then extended linearly to general polynomials. To do that, we associate to any monomial t a set of vertices, written $\mathrm{Supp}(t)$, and the value $R(t)$ is defined to be the reduction of t under the ideal generated by axioms associated to $\mathrm{Supp}(t)$. Again, we abuse notation slightly and write this reduced polynomial as $R_{\mathrm{Supp}(t)}(t)$.

More formally, to a given set A of vertices, we identify the following set of axioms:

$$\begin{aligned}
&\{Q_u \mid u \in A\}\\
&\cup \{P_{u,v,y} \mid u \in A, v \in N(u), y \in [c]\}\\
&\cup \{T_{u,y} \mid u \in A, y \in [c]\}\\
&\cup \{\neg G((0,0), y) \mid y \in [c]\}\\
&\cup \text{all boolean axioms}\\
&\cup \text{all complementary axioms.}
\end{aligned}$$

This set contains some axioms that are related to set A as well as the axioms saying that $(0, 0)$ is not colored. $R_A(t)$ is then interpreted as the reduction of t under the ideal generated by the axioms above.

5.2 Closure/support

We first associate to any term t a set of vertices called its *support* and written $\mathrm{Supp}(t)$. The motivation here is that only axioms associated with certain vertices could have been meaningfully used to derive t. Any axiom that mentions a variable of t certainly could be used, but axioms associated with a node u that is in turn associated with a variable in t may be used, and even axioms at some nearby nodes in H may be used.

The support is defined using the following closure operation:

▶ **Definition 14.** *Given a bipartite graph with partition (V_L, V_R), and $S \subseteq V_R$ a subset of "right" vertices. We say that $\mathrm{Cl}(S) \coloneqq \{v \in V_L \mid N(v) \subseteq S\}$ is the* ***closure*** *of S.*

The following property of the closure will be useful.

▶ **Lemma 15.** *Let $G = (V_L, V_R)$ be an (s, e)-expander. If $S, S' \subseteq V_R$ such that $|S|, |S'| < e\frac{s}{2}$ then:*

- $|\mathrm{Cl}(S)| \leq \frac{|S|}{e}$*;*
- $\mathrm{Cl}(S) \cup \mathrm{Cl}(S') \subseteq \mathrm{Cl}(S \cup S')$*;*
- $\forall V \subseteq (V_L \setminus \mathrm{Cl}(S))$ *such that* $|V| \leq \frac{s}{2}$ *we have* $N(V) \not\subseteq S$*.*

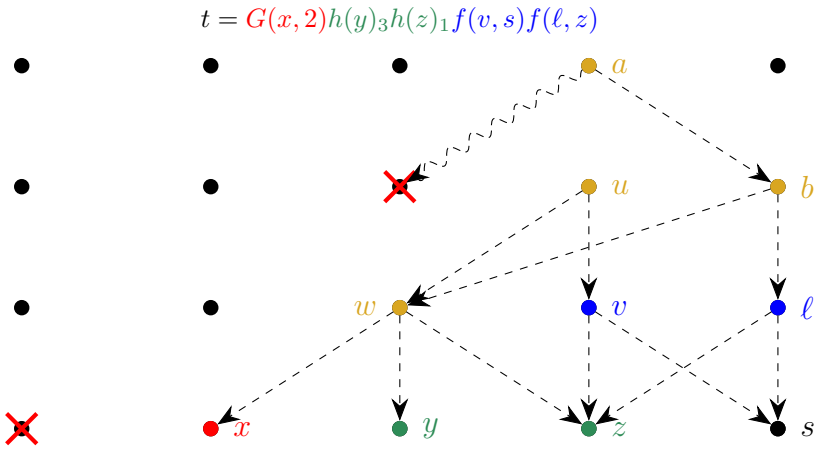

Figure 3 Example of the support of a term t. Color notation:

- nodes for which t contains f variables: $\{v, \ell\}$;
- nodes for which t contains G variables: $\{x\}$;
- nodes for which t contains h variables: $\{y, z\}$;
- nodes that are not touched by t but included in $\mathrm{Supp}(t)$: $\{a, b, u, w\}$;
- nodes that are banished from the graph after application of the restriction from Section 4.

The proof is straightforward and appears in the full version. We can now define the **support** (and **extended support**) associated with a term t. We do this by a downward induction on i over the layers of the graph H as follows:

$$\begin{aligned} b_i &:= \{(i,x) \mid \exists v. f((i,x),v) \text{ or } h((i,x)) \text{ appears in } t\} \\ c_i &:= \{(i,x) \mid \exists y. G((i,x),y) \text{ appears in } t\} \\ d_{i-1} &:= b_{i-1} \cup \mathrm{Cl}(d_i \cup c_i) \\ \mathrm{Supp} &:= \bigcup_{i \in [a]} d_i \\ \mathrm{ExSupp} &:= \bigcup_{i \in [a]} (d_i \cup c_i) \end{aligned}$$

▶ Remark 16. For any terms t, t' it holds that

$$\mathrm{Supp}(t) \cup \mathrm{Supp}(t') \subseteq \mathrm{Supp}(t \cup t').$$

Proof. Follows from Lemma 15, third item. ◀

▶ **Lemma 17.** *If t touches at most ℓ vertices where $\ell \leq 2w$, then $|\mathrm{Supp}(t)| \leq |\mathrm{ExSupp}(t)| = O(\ell)$.*

The proof uses a simple induction and telescoping sum and appears in the full version.

5.3 Proof of Theorem 13

The following propositions are sufficient to prove that the R operator fulfills the desired properties.

▶ **Proposition 18.** *For any variable x, any monomial t such that xt is beautiful, and any $t' \in R(t)$, we have $R_{\mathrm{Supp}(xt')}(xt') = R_{\mathrm{Supp}(xt)}(xt')$.*

▶ **Proposition 19.** *For any almost beautiful monomial t and any $A \subseteq H$ with $|A| \leq O(w)$ we have:*

$$R_{\mathrm{Supp}(t)\cup A}(t) = R_{\mathrm{Supp}(t)}(t).$$

With these propositions in hand, let us now prove Theorem 13

Proof of Theorem 13. We start with the first item, that $R(A) = 0$ for any axiom A. Consider some axiom A and note that it touches a set of vertices $H_A \subseteq H$ of constant size. Let $\sum t_i$ be a polynomial representation of A. By Remark 16, $\bigcup \mathrm{Supp}(t_i) \subseteq \mathrm{Supp}(\bigcup t_i)$ and by Proposition 19, $R(\sum t_i) = \sum R_{\mathrm{Supp}(t_i)}(t_i) = \sum R_{\mathrm{Supp}(\bigcup t_i)\cup H_A}(t_i) = 0$.

For the second item, that $R(1) \neq 0$, observe that since $\mathrm{Supp}(1) = \emptyset$ the set of axioms associated to it is simply:

$$\{\neg G((0,0), y) \mid y \in [c]\}$$
$$\cup \text{ all boolean axioms}$$
$$\cup \text{ all complementary axioms.}$$

This is a satisfiable set of polynomial equations and thus the polynomial 1 is irreducible over the ideal generated by $\mathrm{Supp}(1)$.

Let us now prove the third item, that $R(xt) = R(xR(t))$ whenever xt is beautiful. Consider a beautiful term xt.

$$\begin{aligned}
R(xR(t)) &= \sum_{t' \in R(t)} R(xt') && \text{by linearity} \\
&= \sum_{t' \in R(t)} R_{\mathrm{Supp}(xt')}(xt') && \text{by definition of } R(\cdot) \\
&= \sum_{t' \in R(t)} R_{\mathrm{Supp}(xt)}(xt') && \text{using Proposition 18} \\
&= R_{\mathrm{Supp}(xt)}(xR(t)) && \text{by linearity} \\
&= R_{\mathrm{Supp}(xt)}(xR_{\mathrm{Supp}(t)}(t)) && \text{by definition of } R(\cdot)
\end{aligned}$$

Observe that since $\mathrm{Supp}(t) \subseteq \mathrm{Supp}(xt)$, we can remove the righthand R operator using Fact 4. We then get

$$\begin{aligned}
&= R_{\mathrm{Supp}(xt)}(xt) \\
&= R(xt)
\end{aligned}$$

by definition of $R(\cdot)$. ◀

5.4 Proof of Proposition 19

Proof of Proposition 19. We assume for sake of contradiction that the proposition is false and let A be a minimal witness of this. Without loss of generality $A \cap \mathrm{Supp}(t) = \emptyset$ and A is nonempty. We find A' with $|A'| < |A|$ where the lemma also fails.

Let $r \coloneqq R_{\mathrm{Supp}(t)\cup A}(t)$. We have an equation $t = r + \sum q_i p_i$ where p_i is one of the axiom that correspond to $\mathrm{Supp}(t) \cup A$. By assumption, we also have that $r \prec R_{\mathrm{Supp}(t)}(t)$. We want to construct an assignment α that:

- sets p_i to 0 for at least one $p_i \in A$;
- leaves t unchanged;
- for any p_i either satisfies it or leaves it unchanged.

If such an assignment exists then $t \restriction \alpha = r \restriction \alpha + \sum (q_i \restriction \alpha)(p_i \restriction \alpha)$ and thus $t = r \restriction \alpha + \sum q'_i (p_i \restriction \alpha)$ with $r \restriction \alpha \preceq r$. Since α sets at least one p_i, this shows we can reduce t using only $\mathrm{Supp}(t) \cup (A \setminus \{p_i\})$, as desired.

First observe that if there is a vertex $u \in A$ with neighbor $v \in N(u) \setminus (A \cup \mathrm{ExSupp}(t))$ then we can:

- set the arrow out of u to point to v ($\alpha(f(u,v)) = \mathrm{True}$);
- set all other arrows out of u to False ($\forall v' \in N(u) \setminus v . \alpha(f(u,v')) = \mathrm{False}$);
- set all colors to False at v ($\forall y \in [c] . \alpha(G(v,y)) = \mathrm{False}$).

This satisfies all $P_{u,\cdot,\cdot}$ and Q_u axioms at u and leaves axioms at other vertices in $A \cup \mathrm{Supp}(t)$ untouched. Also $v \notin \mathrm{ExSupp}(t)$ and hence t is also untouched.

The only way we can fail to find such a u and v is if $N(A) \subseteq A \cup \mathrm{ExSupp}(t)$. Note that this implies A contains a vertex in the last layer. Indeed, consider the largest $i \in [a]$ that contains some vertices from A. If $i < a - 1$ then for any vertex $u \in A$ on this layer there is at least one neighbor $v \notin \mathrm{ExSupp}(t)$ because otherwise by definition of closure u should be included in the $\mathrm{Supp}(t)$. Since we picked the largest possible i we also know $v \notin A$.

If A contains a vertex on the last layer then we consider some layer i such that axioms from $\mathrm{ExSupp}(t) \cup A$ do not touch any vertex on layers $i, i+1$ and let

$$B \coloneqq \{(i', j) \mid i' > i, j \in [b]\}.$$

B is the set of all vertices between this layer and the bottom layer. Since $|A| + |\mathrm{Supp}(t)| = O(w)$ we know such a row exists. We pick some color y not mentioned by t, set $G(u,y) = \mathrm{True}$ for all $u \in B$, and set $h((a-1,x)) = y$ for any $(a-1,x) \in B \setminus \mathrm{Supp}(t)$.

Furthermore, for any $u \in B \cap A$ that is not on the last layer we know by the second item of Lemma 15 that u has a neighbor $z \notin \mathrm{ExSupp}(t)$. We set f so it points u to z and set $G(z,y') = \mathrm{False}$ for all other colors. Here we use the fact that $z \notin \mathrm{ExSupp}(t)$ and if we set colors for z the term t remains unchanged. That is, $f(u,v) = \mathrm{False}$ for all $v \in N(u) \setminus \{z\}$ and $f(u,z) = \mathrm{True}$. If there were no such z then u would be included in $\mathrm{Supp}(t)$ and hence would not be in A.

By setting colors in this way we satisfy all axioms for all $u \in B \cap A$. At $u \in B \cap \mathrm{Supp}(t)$ we only set a single color to True and satisfy the corresponding axiom. We can always pick a color y because $|t| < c$ and we touch no variable in t. Finally, we satisfy the Q_u and $T_{u,y}$ axioms anywhere they are touched by setting h and f. ◀

The proof of Proposition 18 uses similar ideas to that of Proposition 19 and appears in the full version.

6 Concluding remarks

In this paper, we have shown that although it is known from [24] that size provides an upper bound on degree in polynomial calculus resolution (PCR) – in the sense that if a set of degree-k polynomials have a PCR refutation in size S, then there is also a refutation in degree $k + O(\sqrt{n \log S})$ – it is not possible in general to construct PCR refutations that get even remotely close to these two upper bounds simultaneously. This extends the analogous trade-off result in [29] for the weaker resolution proof system.

We would like to remark that while our trade-off result is currently stated for formulas of logarithmic width – in order to avoid extra technical complications – it is also possible to obtain the result for constant-width formulas. This involves converting the wide clauses in our formulas to 3-CNF using the standard approach with extension variables. Since the upper bounds for our formulas also work in resolution, this strengthens [29]. We refer to the upcoming full-length version of the paper for the details.

As mentioned in the introduction, in addition to generalizing [29] from resolution to PCR, our main theorem is also a counterpart of the size-degree trade-offs for the weaker algebraic system Nullstellensatz established in [17]. In contrast to our result, however, the Nullstellensatz trade-offs only hold for the version of the proof system without formal variables for negated literals. It would be desirable to obtain more robust trade-off theorems that also applies to Nullstellensatz with separate formal variables for positive and negative literals.

Another, even more interesting, question would be to investigate stronger proof systems. Very recently, Atserias and Hakoniemi [4] showed that if a system of degree-k polynomial constraints over n boolean variables has a **Sums-of-Squares (SOS)** refutation of size S, then it also has a refutation of degree at most $O(k + \sqrt{n \log S})$. They also proved a similar statement for the stronger **Positivstellensatz** proof system, and an analogous result for the weaker system **Sherali-Adams** also follows from the proofs in the paper. However, for these proof systems the question of whether the conversion from small size to small degree can be achieved without blowing up the size remains open, and it is not even known if the upper bound on degree in terms of size in [4] is tight.

References

1. Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. Space Complexity in Propositional Calculus. *SIAM J. Comput.*, 31(4):1184–1211, 2002. doi:10.1137/S0097539700366735.
2. Michael Alekhnovich and Alexander A. Razborov. Lower Bounds for Polynomial Calculus: Non-Binomial Case. *Proceedings of the Steklov Institute of Mathematics*, 242:18–35, 2003. Available at http://people.cs.uchicago.edu/~razborov/files/misha.pdf. Preliminary version in *FOCS '01*.
3. Albert Atserias and Víctor Dalmau. A combinatorial characterization of resolution width. *J. Comput. Syst. Sci.*, 74(3):323–334, 2008. doi:10.1016/j.jcss.2007.06.025.
4. Albert Atserias and Tuomas Hakoniemi. Size-Degree Trade-Offs for Sums-of-Squares and Positivstellensatz Proofs. In *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA.*, pages 24:1–24:20, 2019. doi:10.4230/LIPIcs.CCC.2019.24.
5. Paul Beame, Chris Beck, and Russell Impagliazzo. Time-Space Tradeoffs in Resolution: Superpolynomial Lower Bounds for Superlinear Space. *SIAM Journal on Computing*, 45(4):1612–1645, August 2016. Preliminary version in *STOC '12*. doi:10.1137/130914085.
6. Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. Lower Bounds on Hilbert's Nullstellensatz and Propositional Proofs. *Proceedings of the London Mathematical Society*, s3-73(1):1–26, 1996. doi:10.1112/plms/s3-73.1.1.

7 Chris Beck, Jakob Nordström, and Bangsheng Tang. Some Trade-off Results for Polynomial Calculus. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC '13)*, pages 813–822, May 2013. doi:10.1145/2488608.2488711.

8 Eli Ben-Sasson. Size Space Tradeoffs for Resolution. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC '02)*, pages 457–464, May 2002. doi:10.1145/509907.509975.

9 Eli Ben-Sasson and Jakob Nordström. Understanding Space in Proof Complexity: Separations and Trade-offs via Substitutions. In *Proceedings of the 2nd Symposium on Innovations in Computer Science (ICS '11)*, pages 401–416, January 2011. arXiv:1008.1789.

10 Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow - resolution made simple. *J. ACM*, 48(2):149–169, 2001. doi:10.1145/375827.375835.

11 Christoph Berkholz and Jakob Nordström. Supercritical Space-Width Trade-offs for Resolution. In *Proceedings of the 43rd International Colloquium on Automata, Languages and Programming (ICALP '16)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 57:1–57:14, July 2016. doi:10.4230/LIPIcs.ICALP.2016.57.

12 Archie Blake. *Canonical Expressions in Boolean Algebra*. PhD thesis, University of Chicago, 1937. doi:10.2307/2267634.

13 María Luisa Bonet and Nicola Galesi. Optimality of Size-Width Tradeoffs for Resolution. *Computational Complexity*, 10(4):261–276, December 2001. Preliminary version in *FOCS '99*. doi:10.1007/s000370100000.

14 Sam Buss and Jakob Nordström. Proof Complexity and SAT Solving. In Armin Biere, Marijn J. H. Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability*. IOS Press, 2020. Chapter to appear in the 2nd edition. Draft version available at https://www.math.ucsd.edu/~sbuss/ResearchWeb/ProofComplexitySAT/ProofComplexityChapter.pdf.

15 Matthew Clegg, Jeff Edmonds, and Russell Impagliazzo. Using the Groebner Basis Algorithm to Find Proofs of Unsatisfiability. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996*, pages 174–183, 1996. doi:10.1145/237814.237860.

16 William Cook, Collette Rene Coullard, and György Turán. On the Complexity of Cutting-Plane Proofs. *Discrete Applied Mathematics*, 18(1):25–38, November 1987. doi:10.1016/0166-218X(87)90039-4.

17 Susanna F. de Rezende, Jakob Nordström, Or Meir, and Robert Robere. Nullstellensatz size-degree trade-offs from reversible pebbling. In *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA.*, pages 18:1–18:16, 2019. doi:10.4230/LIPIcs.CCC.2019.18.

18 Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. How Limited Interaction Hinders Real Communication (and What It Means for Proof and Circuit Complexity). In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS '16)*, pages 295–304, October 2016. doi:10.1109/FOCS.2016.40.

19 Nicola Galesi, Leszek Aleksander Kolodziejczyk, and Neil Thapen. Polynomial calculus space and resolution width. *Electronic Colloquium on Computational Complexity (ECCC)*, 26:52, 2019. URL: https://eccc.weizmann.ac.il/report/2019/052.

20 Nicola Galesi and Massimo Lauria. On the Automatizability of Polynomial Calculus. *Theory of Computing Systems*, 47(2):491–506, August 2010. doi:10.1007/s00224-009-9195-5.

21 Nicola Galesi and Massimo Lauria. Optimality of Size-Degree Trade-offs for Polynomial Calculus. *ACM Transactions on Computational Logic*, 12(1):4:1–4:22, November 2010. doi:10.1.1.703.6976.

22 Mika Göös and Toniann Pitassi. Communication Lower Bounds via Critical Block Sensitivity. *SIAM J. Comput.*, 47(5):1778–1806, 2018. doi:10.1137/16M1082007.

23 Trinh Huynh and Jakob Nordström. On the Virtue of Succinct Proofs: Amplifying Communication Complexity Hardness to Time-Space Trade-offs in Proof Complexity. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 233–248, 2012. doi:10.1145/2213977.2214000.

24 Russell Impagliazzo, Pavel Pudlák, and Jirí Sgall. Lower Bounds for the Polynomial Calculus and the Gröbner Basis Algorithm. *Computational Complexity*, 8(2):127–144, 1999. doi:10.1007/s000370050024.

25 Jan Krajíček. *Proof Complexity*, volume 170 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, March 2019. doi:10.1017/9781108242066.

26 Mladen Mikša and Jakob Nordström. A Generalized Method for Proving Polynomial Calculus Degree Lower Bounds. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 467–487, 2015. doi:10.4230/LIPIcs.CCC.2015.467.

27 Jakob Nordström. A Simplified Way of Proving Trade-off Results for Resolution. *Information Processing Letters*, 109(18):1030–1035, August 2009. Preliminary version in ECCC report TR07-114, 2007. doi:10.1016/j.ipl.2009.06.006.

28 Alexander A. Razborov. A New Kind of Tradeoffs in Propositional Proof Complexity. *J. ACM*, 63(2):16:1–16:14, 2016. doi:10.1145/2858790.

29 Neil Thapen. A Tradeoff Between Length and Width in Resolution. *Theory of Computing*, 12(1):1–14, 2016. doi:10.4086/toc.2016.v012a005.