On the Bit Complexity of Sum-of-Squares Proofs

Ben Weitz[*]

December 6, 2016

FILES/ASBTRACT

[*] UC Berkeley, bsweitz@cs.berkeley.edu. Supported by an NSF Graduate Research Fellowship (NSF award no DGE 1106400).

1 Introduction

2 Preliminaries

For a set of real polynomials $\mathcal{P} = \{p_1, p_2, \ldots, p_m\}$, we denote their generated ideal in $\mathbb{R}[x]$ by $\langle \mathcal{P} \rangle$ or $\langle p_1, \ldots, p_m \rangle$. For an ideal I of $\mathbb{R}[x]$, its real variety is defined

$$\mathcal{V}_{\mathbb{R}}(I) := \{x \in \mathbb{R} | \forall p \in I : p(x) = 0\}.$$

For a set $S \subseteq \mathbb{R}$, its vanishing ideal is defined

$$\mathcal{I}(S) = \{p \in \mathbb{R}[x] | \forall x \in S : p(x) = 0\}.$$

We call $\mathcal{P}$ *real complete* if $\mathcal{I}(\mathcal{V}_{\mathbb{R}}(\mathcal{P})) = \langle \mathcal{P} \rangle$. For any polynomial $p \in \langle \mathcal{P} \rangle$, we say p has a degree d derivation from $\mathcal{P}$ if there exist there exist real $\lambda_1(x), \ldots, \lambda_m(x)$ such that $p(x) = \sum_i \lambda_i(x) p_i(x)$ and $\max_i \deg \lambda_i p_i \leq d$. We often refer to the λ_i as the derivation of p. Finally, if we call $\mathcal{P}$ *$k(d)$-effective* if $\mathcal{P}$ is real complete and any polynomial $p \in \langle \mathcal{P} \rangle$ of degree d has a derivation from $\mathcal{P}$ of degree $k(d)$.

Lemma 2.1. *If $\mathcal{P}$ is real complete with finite real variety S, then $\mathbb{R}[x]/\langle \mathcal{P} \rangle$ is an $\mathbb{R}$-vector space of dimension $|S|$ with inner product $\langle \overline{p}, \overline{q} \rangle = \sum_{\alpha \in S} p(\alpha) q(\alpha)$ for any $p \in \overline{p}$, $q \in \overline{q}$.*

Proof. Let $I = \langle \mathcal{P} \rangle$. Clearly $\mathbb{R}[x]/I$ is an $\mathbb{R}$-vector space, but furthermore it is easy to show that $\mathbb{R}[x]/I \equiv \mathbb{R}^S$. Consider the evaluation map $E : \mathbb{R}[x]/I \to \mathbb{R}^S$ given by $E(\overline{p}) = (\alpha \to p(\alpha))$. First, we show that this map is well-defined. If $p, q \in \overline{p}$, then $p - q \in I$. Because S is the real variety of $\mathcal{P}$, $(p - q)(\alpha) = 0$ for any $\alpha \in S$, and thus $p(\alpha) = q(\alpha)$. Next, E is clearly linear, and if you take a high enough degree polynomial it is clear that E is onto. Now if $E(\overline{p}) = 0$, then $\forall \alpha \in S : p(\alpha) = 0$. Because $\mathcal{P}$ is real complete, $p \in I$, and so $\overline{p} = \overline{0}$. The claimed inner product is simply the Euclidean inner product of $\mathbb{R}^S$ composed with the evaluation map E. □

2.1 Sum of Squares Proofs

Definition 2.2. Let $S \subseteq \mathbb{R}^n$ be finite, $\mathcal{P}$ be a generating set for $\mathcal{I}(S)$, and let $r(x)$ be a polynomial such that $r(\alpha) \geq 0$ for each $\alpha \in S$. We say that $r(x)$ has a *Sum-of-Squares proof of nonnegativity from $\mathcal{P}$* if there is a polynomial identity of the form

$$r(x) = \sum_i^m h_i^2(x) + \sum_{p \in \mathcal{P}} \lambda_p(x) p(x).$$

We say the proof has degree d if $\max\{\deg h_i^2, \deg \lambda_p p\} = d$. We will sometimes refer to the collection $\Pi_{\mathcal{P}} = \{h_i, \lambda_p | i \in [m], p \in \mathcal{P}\}$ as the proof.

We will be concerned with not just the degree of these proofs, but also their bit complexity. To this end, we define the following norms on polynomials and proofs: For $p(x) \in \mathbb{R}[x]$, we write $\|p\|$ for the absolute value of the maximum coefficient of p in the standard monomial basis, and for any collection of polynomials $\mathcal{P}$, we write $\|\mathcal{P}\| = \max_{p \in \mathcal{P}} \|p\|$.

3 Effective Nullstellensatz Yields Low Bit Complexity

In this section we prove our main theorem:

Theorem 3.1. *Let $\mathcal{P} = \{p_1, \ldots, p_m\}$ be a real complete, $k(d)$-effective set of polynomials with $S = \mathcal{V}_{\mathbb{R}}(\mathcal{P})$. Let $r(x)$ be a polynomial nonnegative on S, and assume r has a degree d sum-of-squares proof of nonnegativity $r(x) = \sum_{i=1}^{t} q_i^2 + \sum_{i=1}^{m} \lambda_i p_i$. Then r has a degree $k(d)$ sum-of-squares proof of nonnegativity such that the coefficients of every polynomial appearing in the proof are bounded by $O(\exp(\|\mathcal{P}\| + \|r\|))$.*

Proof. Our strategy is to simplify the SOS part of the given SOS-proof and move all of the potentially huge coefficients into the latter term which lies in the ideal $I = \langle \mathcal{P} \rangle$. Next, we use the fact that $\mathcal{P}$ is effective to argue that you don't need to use huge coefficients to express whatever the leftover polynomial. To that end, we want to reduce each q_i to a canonical form. Recall that since $\mathcal{P}$ is complete, $V = \mathbb{R}[x]/I$ is a real vector space with inner product $\langle p, q \rangle = \sum_{\alpha \in S} p(\alpha) q(\alpha)$. Let $V_d = (\mathbb{R}[x]/I)_d$ denote the subspace of V where each equivalence class contains a representative of degree at most d. Clearly the standard monomial basis is a spanning set for V_d, so it contains some basis for V_d. Take this basis and the defined inner product and use the Gram-Schmidt process to produce an orthonormal basis $\{v_1, \ldots, v_s\}$ for V_d. We need to bound $\|v_i(x)\|$: Note that each $\alpha \in S$ satisfies $\|\alpha\| = \text{poly}(\|\mathcal{P}\|)$, and so at each step in the Gram-Schmidt process, the inner product between two polynomials p and q is at most $\text{poly}(\|\mathcal{P}\|, \|p\|, \|q\|, n^d)$, and since there are at most n^d steps, it produces polynomials satisfying $\|v_i(x)\| \leq \text{poly}(\|cP\|, n^d)$ for each i. Define a vector of polynomials $v = [v_1, \ldots, v_s]$, and note there must exist vectors of reals $c_1, \ldots, c_t$ such that $\forall \alpha \in S : c_i^T v(\alpha) = q_i(\alpha)$, and thus $q_i^2(x) - (c_i^T v(x))^2 \in I$. Because $\mathcal{P}$ is k-effective, there must be a way to write $q_i^2(x) - (c_i^T v(x))^2$ as a polynomial combination of $\{p_1, \ldots, p_m\}$ with each term having degree bounded by $k(d)$. Thus there is a degree $k(d)$ SOS-proof of nonnegativity for r:

$$\begin{aligned} r(x) &= \sum_{i=1}^{t} (c_i^T v(x))^2 + \sum_{i=1}^{m} \lambda_i' p_i \\ &= \sum_{i=1}^{t} c_i c_i^T \cdot v(x) v(x)^T + \sum_{i=1}^{m} \lambda_i' p_i \\ &= (C \cdot v(x) v(x)^T) + \sum_{i=1}^{m} \lambda_i' p_i \end{aligned}$$

Now if we take this polynomial identity and average it over every $\alpha \in S$, we get

$$\begin{aligned} E_\alpha[r(\alpha)] &= (C \cdot E_\alpha[v(\alpha) v(\alpha)^T]) + 0 \\ &= C \cdot Id \\ &= Tr(C) \end{aligned}$$

where we used the fact that $\{v_1, \ldots, v_s\}$ are orthonormal. Clearly $E_\alpha[r(\alpha)] = \text{poly}(\|r\|)$, and since C is PSD, this implies $\|C\|_F = \text{poly}(\|r\|)$, and thus $\|v(x)^T C v(x)\| = \text{poly}(\|r\|, \|\mathcal{P}\|, n^d)$. Now it remains to bound the coefficients of all λ_i'. If we let the coefficients of polynomials σ_i be variables, it's clear that the system of equations induced by the polynomial identity

$$r(x) - (C \cdot v(x) v(x)^T) = \sum_i \sigma_i(x) p_i(x)$$

is feasible, but note that this system is *linear* in the coefficients of σ_i, there are at most $O(n^{k(d)})$ equations, and the entries are at most $\text{poly}(\|r\|, \|\mathcal{P}\|, n^d)$. By Cramer's rule, we can pick a solution σ_i^* with $\|\sigma_i^*\| \leq \exp(\text{poly}(\|r\|, \|\mathcal{P}\|, n^d))$. Thus we can replace the λ_i' with σ_i^* and achieve an SOS-proof of bounded bit complexity. □

We obtain an immediate corollary:

Corollary 3.2. *If $\mathcal{P}$ is complete and a Grobner basis for $\langle \mathcal{P} \rangle$, then any degree d SOS proof from $\mathcal{P}$ can be taken to have polynomial bit complexity.*

Proof. Every Grobner basis is d-effective, so just apply Theorem 3.1. □

Remark 3.3. In CITE ODONNEL, the author gives an example $\mathcal{P}$ and linear $r(x)$ for which r has a degree two SOS proof, but it must have exponential bit-complexity. We simply note here that the $\mathcal{P}$ he gives is $(d+2)$-effective but not d-effective, so if you're willing to pay a tiny bit in the degree you can take the SOS proofs to have polynomial bit-complexity.

With our main theorem in hand, we move on to using it to prove that many applications of the SOS algorithm remain automatizable. One might worry that Grobner bases are the *only* types of base constraints that are effective, and that this theorem will not be very useful. We dispel this notion in Section 4 by showing the natural description of the MAX-BISECTION ideal is effective, even though its Grobner basis has exponential size. Before we move on to these though, we strengthen the theorem slightly. Note that we are only allowing ourselves efficient Nullstellensatz-type proofs, even though the SOS proof system is more powerful. As an example, consider the following system: $\mathcal{P} = \{x_i^2 - 1 | i \in [n]\} \cup \{\sum_i x_i - n\}$. It's clear that the only feasible solutions to this system are $x_i = 1$ for all i. However, the polynomial $x_i - 1$ does not have a derivation in degree less than n from these constraints! However, if we allow ourselves to use SOS-proofs, one can write both $x_i - 1$ and $1 - x_i$ as a SOS modulo $\mathcal{P}$, thereby proving that $x_i - 1 \in I$. This motivates the following (obvious) lemma:

Lemma 3.4. *Let $\mathcal{P}$ and $\mathcal{Q}$ be such that $\langle \mathcal{P} \rangle = \langle \mathcal{Q} \rangle$ and let $\mathcal{Q}$ be complete and $k(d)$-effective. If every $q \in \mathcal{Q}$ of degree d can be written $q(x) = \sum_i s_i(x)^2 + \sum_i \lambda_i(x) p_i(x)$ in degree $k'(d)$ with $\|s_i\|, \|\lambda_i\| \leq poly(\|\mathcal{P}\|)$, then for any SOS proof of $r(x)$ from $\mathcal{P}$ of degree d, there is an SOS proof of $r(x)$ from $\mathcal{P}$ of degree $k(d) + k'(d)$ with coefficients bounded by $O(\exp(\|\mathcal{P}\| + \|r\|))$.*

4 CSPs With Balance Constraints

CSPs are polynomial optimization problems with very simple domains. Their solution space is completely described by the Boolean equations $\{x_i^2 - 1\}$, which are a Grobner Basis and thus d-effective, so any SOS problem on CSPs is easily automatizable by the Ellipsoid algorithm. However, one could also ask for a balance constraint when optimizing a CSP, for example to solve the MAX-BISECTION problem. In this case, the polynomial constraints are given by $\mathcal{P} = \{x_i^2 - 1 | i \in [n]\} \cup \{\sum_i x_i\}$. Now $\mathcal{P}$ is not a Grobner basis, and indeed the Grobner basis has exponential size, despite $\langle \mathcal{P} \rangle$ being generated by a very small set of polynomials. However, it turns out that these sets of polynomials are still effective, and thus the SOS algorithm is still automatizable. This proof follows a very similar path to CITE ME, due to the obvious symmetry of the constraints.

Theorem 4.1. *Let c be an integer. Then for every $n \geq c$, the set of polynomials*

$$\mathcal{P}(c,n) = \{x_i^2 - x_i | 1 \leq i \leq n\} \cup \{\sum_{i=1}^{n} x_i - c\}$$

is $(2d, \mathrm{poly}(n^d))$-effective.

We will consider only $0 \leq c \leq n/2$. The proof for $n/2 \leq c \leq n$ is symmetric. To prove this theorem, we treat polynomials of degree at most c and polynomials of degree greater than c differently. This is because PC proofs suffice for polynomials of small degree, but do not for polynomials of degree c. This is because if $S \subseteq [n]$ with $|S| > c$, then $\prod_{i \in S} x_i = 0$ by the pigeonhole principle (since in a group of $c+1$ variables, at least one of them must be 0), but the pigeonhole principle is hard to prove with PC proofs. We need to use PC$_>$ proofs for these high-degree monomials. So we prove the following two lemmas:

Lemma 4.2. *Let $r(x) \in \langle \mathcal{P}(c,n) \rangle$ be a polynomial such that $\deg r = d \leq c$. Then r has a degree d PC derivation from $\mathcal{P}(c,n)$.*

Lemma 4.3. *Let $S \subseteq [n]$ with $|S| \geq c+1$. Then $\prod_{i \in S} x_i$ has a $(2|S|, O(1))$-PC$_>$ derivation from $\mathcal{P}(c,n)$.*

To see how these two lemmas imply the theorem, first WLOG assume $r(x)$ is multilinear. Let $r(x) = r_h(x) + r_l(x)$, where r_h is all monomials of degree at least $c+1$. Since $r, r_h \in \langle \mathcal{P}(c,n) \rangle$, so too must r_l. By Lemma 4.2, r_l has a degree d PC derivation from $\mathcal{P}(c,n)$. By Lemma 4.3, each monomial in r_h has a degree $2d$ PC$_>$ derivation from $\mathcal{P}(c,n)$ where each SOS polynomial has constant-size coefficients. Since there are at most n^{d+1} high degree monomials, this implies $r(x)$ has a $(2d, \text{poly}(n^d))$-PC$_>$ derivation from $\mathcal{P}(c,n)$. It remains to prove the two lemmas:

(Proof of Lemma 4.2). We proceed by induction on c. If $c = 0$, because $\mathcal{P}(c,n)$ is feasible the only constant polynomial in $\langle \mathcal{P}(c,n) \rangle$ is the zero polynomial, which has the trivial derivation. Now consider the case of $c+1$. We proceed in two parts. First, if r is fully symmetric, we show that it has a degree d derivation. Secondly, for any polynomial $p \in \langle \mathcal{P}(c,n) \rangle$ we prove that $p - \frac{1}{n!}\sum_{\sigma \in S_n} \sigma p$ has a degree d derivation from $\mathcal{P}$, where σ acts on p by permuting the labels of the variables. Taken together, these two facts imply that r has a degree d derivation from $\mathcal{P}$.

To prove the first part, note that a symmetric polynomial r is a linear combination of the elementary symmetric polynomials, and it is clear that $e_k(x)$ can be derived by taking the polynomial $(\sum_i x_i - c)^k$, reducing it to multilinear using the boolean constraints, and then reducing by $e_l(x)$ for each $l < k$. This will result in a constant polynomial, which must be the zero polynomial since that is the only constant polynomial in $\langle \mathcal{P}(c,n) \rangle$.

To prove the second part, let σ_{ij} be the transposition of labels i and j, and consider the polynomial $r - \sigma_{ij} r$. Writing $r = r_i x_i + r_j x_j + r_{ij} x_i x_j + q_{ij}$, where none of r_i,r_j,r_{ij}, nor q_{ij} depend on x_i or x_j, we can rewrite

$$r - \sigma_{ij} r = (r_i - r_j)(x_i - x_j).$$

Now because $r - \sigma_{ij} r$ evalutes to zero on any boolean string with exactly c ones, if we set $x_i = 1$ and $x_j = 0$, we know that $r_i - r_j$ is a polynomial that must evaluate to zero on any boolean string with exactly $c-1$ ones. Because $\mathcal{P}(c-1, n-2)$ is complete and $\deg r_i - r_j = d - 1 \leq c - 1$, by the inductive hypothesis, $r_i - r_j$ has a degree $d-1$ PC derivation from $\mathcal{P}(c-1, n-2)$ (recall that since $c \leq n/2$, $c - 1 \leq (n-2)/2$). This implies that $(r_i - r_j)(x_i - x_j)$ has a degree $d-1$ PC derivation from $\mathcal{P}(c,n)$:

$$\begin{aligned}(r_i - r_j)(x_i - x_j) &= \left[\sum_{t \neq i,j} \lambda_t \cdot (x_t^2 - x_t) + \lambda \cdot \left(\sum_{t \neq i,j} x_t - (c-1)\right)\right](x_i - x_j) \\ &= \sum_t \lambda'_t \cdot (x_t^2 - x_t) + \lambda \cdot \left(\sum_{t \neq i,j} x_t - (c-1) + (x_i + x_j - 1)\right)(x_i - x_j)\end{aligned}$$

$$= \sum_t \lambda'_t \cdot (x_t^2 - x_t) + \lambda' \cdot \left(\sum_t x_t - c \right)$$

where we used the fact that $(x_i + x_j - 1)(x_i - x_j) - (x_i^2 - x_i) + (x_j^2 - x_j) = 0$. The degree of this derivation is at most d because each λ_t has degree at most $d-3$, and $\lambda'_t = \lambda_t(x_i - x_j)$, and similarly for λ. Thus the inductive hypothesis implies that $r - \sigma_{ij} r$ has a degree d derivation, and since transpositions generate the symmetric group, this implies that $r - \frac{1}{n!} \sum_{\sigma \in S_n} \sigma r$ has a degree d PC derivation from $\mathcal{P}(c, n)$. □

(Proof of Lemma 4.3). First, note that for any monomial $x_M = \prod_{i \in M} x_i$, we have $(1 - x_M)^2 = (1 - x_M) + q_1(x)$ and $x_M^2 = x_M + q_2(x)$, where $q_1(x)$ and $q_2(x)$ are generated from the boolean constraints and of degree $|M|$. Now WLOG let $S = \{1, 2, 3, \ldots, c+1\}$. If we multiply $c - \sum_{i=1}^n x_i$ by $x_1 x_2 \ldots x_c$ and reduce by the boolean constriants, we get the polynomial $-x_1 x_2 \ldots x_c(x_{c+1} + x_{c+2} + \ldots x_n)$. Now because each monomial is a square modulo the boolean constraints, this implies that

$$-x_S = \sum_{i=c+1}^n (x_1 x_2 \ldots x_c \cdot x_i)^2 - x_1 x_2 \ldots x_c (\sum_{i=1}^n x_i - c) + \sum_{i=1}^n \lambda_i (x_i^2 - x_i).$$

and as we explained above, $x_S = x_S^2 + \sum_{i=1}^n \lambda_i (x_i^2 - x_i)$. Thus x_S has a $(2|S|, 1)$-PC$_>$ derivation from $\mathcal{P}(c, n)$. □

The fact that this ideal is effective gives an example of a set of constraints which are very far from being a Grobner Basis for their ideal, yet still allow for effective derivations.

Acknowledgements

We would like to thank Prasad Raghavendra, Satish Rao, and Ben Recht for helpful discussions, and Moritz Hardt and Samuel B. Hopkins for helpful questions. We also thank several anonymous reviewers for very helpful comments.

FILES/NOISE-NNM FILES/ERRORS-APPENDIX FILES/WHITENING-APP