Finite convergence of sum-of-squares hierarchies for the stability number of a graph

Monique Laurent *

Luis Felipe Vargas †

March 9, 2021

Abstract

We investigate a hierarchy of semidefinite bounds $\vartheta^{(r)}(G)$ for the stability number $\alpha(G)$ of a graph G, based on its copositive programming formulation and introduced by de Klerk and Pasechnik [*SIAM J. Optim.* 12 (2002), pp.875–892], who conjectured convergence to $\alpha(G)$ in $r = \alpha(G) - 1$ steps. Even the weaker conjecture claiming finite convergence is still open. We establish links between this hierarchy and sum-of-squares hierarchies based on the Motzkin-Straus formulation of $\alpha(G)$, which we use to show finite convergence when G is acritical, i.e., when $\alpha(G \setminus e) = \alpha(G)$ for all edges e of G. This relies, in particular, on understanding the structure of the minimizers of Motzkin-Straus formulation and showing that their number is finite precisely when G is acritical. As a byproduct we show that deciding whether a standard quadratic program has finitely many minimizers does not admit a polynomial-time algorithm unless P=NP.

***K*eywords**

Stable set problem, α-critical graph, polynomial optimization, Lasserre hierarchy, sum-of-squares polynomial, finite convergence, copositive programming, standard quadratic programming, semidefinite programming, Motzkin-Straus formulation.

1 Introduction

Given a graph $G = (V, E)$, its *stability number* $\alpha(G)$ is defined as the largest cardinality of a stable set in G. Computing the stability number of a graph is a central problem in combinatorial optimization, well-known to be NP-hard [14]. Many approaches based, in particular, on semidefinite programming have been developed for constructing good relaxations. A starting point to define hierarchies of approximations for the stability number is the following formulation by Motzkin and Straus [30], which expresses $\alpha(G)$ via quadratic optimization over the standard simplex Δ_n:

$$\frac{1}{\alpha(G)} = \min\{x^T(A_G + I)x : x \in \Delta_n\}, \tag{M-S}$$

where $\Delta_n = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i = 1\}$ and A_G is the adjacency matrix of G. Based on (M-S), de Klerk and Pasechnik [10] proposed the following reformulation:

$$\alpha(G) = \min\{t : x^T(t(I + A_G) - J)x \geq 0 \quad \text{for all } x \in \mathbb{R}^n_+\}, \tag{1.1}$$

which boils down to linear optimization over the copositive cone

$$\text{COP}_n := \{M \in \mathcal{S}^n : x^T M x \geq 0 \; \forall x \in \mathbb{R}^n_+\}.$$

Indeed, $\alpha(G)$ equals the smallest scalar t for which the matrix $M_{G,t} := t(I + A_G) - J$ is copositive, i.e., belongs to COP_n. For $x \in \mathbb{R}^n$ set $x^{\circ 2} := (x_1^2, \ldots, x_n^2)$ and for a matrix $M \in \mathcal{S}^n$ define the polynomials

$$p_M(x) = x^T M x \quad \text{and} \quad P_M(x) = p_M(x^{\circ 2}) = (x^{\circ 2})^T M x^{\circ 2}. \tag{1.2}$$

* Centrum Wiskunde & Informatica (CWI), Amsterdam, and Tilburg University. monique.laurent@cwi.nl

† Centrum Wiskunde & Informatica (CWI), Amsterdam. luis.vargas@cwi.nl

This work is supported by the European Union's Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie Actions Grant Agreement No. 813211 (POEMA).

Then M is copositive precisely when the polynomial p_M is nonnegative over $\mathbb{R}^n_+$ or, equivalently, when P_M is nonnegative over $\mathbb{R}^n$. Based on this observation, Parrilo [34] introduced the following two subcones of $\mathcal{S}^n$:

$$\mathcal{C}_n^{(r)} = \Big\{M : \Big(\sum_{i=1}^n x_i\Big)^r p_M(x) \in \mathbb{R}_+[x]\Big\}, \quad \mathcal{K}_n^{(r)} = \Big\{M : \Big(\sum_{i=1}^n x_i^2\Big)^r P_M(x) \in \Sigma\Big\}, \tag{1.3}$$

which provide sufficient conditions for matrix copositivity: $\mathcal{C}_n^{(r)} \subseteq \mathcal{K}_n^{(r)} \subseteq \text{COP}_n$ for any $r \geq 0$. Here $\mathbb{R}_+[x]$ is the set of polynomials with nonnegative coefficients and Σ denotes the set of sum-of-squares polynomials. De Klerk and Pasechnik [10] used these two cones to define the following parameters:

$$\zeta^{(r)}(G) = \min\{t : t(I + A_G) - J \in \mathcal{C}_n^{(r)}\}, \tag{1.4}$$

$$\vartheta^{(r)}(G) = \min\{t : t(I + A_G) - J \in \mathcal{K}_n^{(r)}\}, \tag{1.5}$$

which provide upper bounds on the stability number: $\alpha(G) \leq \vartheta^{(r)}(G) \leq \zeta^{(r)}(G)$. It is known that the program (1.4) is feasible, i.e., $\zeta^{(r)}(G) < \infty$, if and only if $r \geq \alpha(G) - 1$ and also that $\zeta^{(r)}(G) < \alpha(G) + 1$, i.e., $\lfloor \zeta^{(r)}(G) \rfloor = \alpha(G)$, if and only if $r \geq \alpha(G)^2 - 1$ [10, 42]. On the other hand, the parameter $\vartheta^{(r)}(G)$ provides a nontrivial bound already at order $r = 0$. Indeed, as shown in [10], the parameter $\vartheta^{(0)}(G)$ coincides with $\vartheta'(G)$, the strengthening of the theta number $\vartheta(G)$ by Lovász [24], proposed in [41]. Recall that

$$\vartheta(G) = \max\{\langle J, X\rangle : \text{Tr}(X) = 1, X_{ij} = 0 \ (\{i,j\} \in E), X \succeq 0\},$$

and $\vartheta'(G)$ is obtained by adding the nonnegativity constraint $X \geq 0$ to the above program. As is well-known we have

$$\alpha(G) \leq \vartheta'(G) \leq \vartheta(G) \leq \chi(\overline{G}), \tag{1.6}$$

where $\chi(\overline{G})$ denotes the coloring number of $\overline{G}$ (the complementary graph of G), i.e., the smallest number of cliques of G needed to cover V.

Hence one can find $\alpha(G)$, *after rounding*, in $(\alpha(G))^2$ steps of the hierarchy $\zeta^{(r)}(G)$ or $\vartheta^{(r)}(G)$. It is known that the linear bound $\zeta^{(r)}(G)$ is *never exact*: if G is not the complete graph then $\zeta^{(r)}(G) > \alpha(G)$ for all r [42]. On the other hand, de Klerk and Pasechnik [10] conjecture that rounding is not necessary for the semidefinite parameter $\vartheta^{(r)}(G)$ and moreover that $\alpha(G)$ steps suffice to reach convergence.

Conjecture 1 (De Klerk and Pasechnik [10])**.** *For any graph G we have:* $\vartheta^{(\alpha(G)-1)}(G) = \alpha(G)$.

In fact, it is not even known whether finite convergence holds at some step, so also the following weaker conjecture is still open in general.

Conjecture 2. *For any graph G we have:* $\vartheta^{(r)}(G) = \alpha(G)$ *for some* $r \in \mathbb{N}$.

Let us call the smallest integer r for which $\vartheta^{(r)}(G) = \alpha(G)$ the *ϑ-rank* (or, simply, the *rank*) of G, denoted as ϑ-rank(G). Then Conjecture 2 asks whether the rank is finite for all graphs, while Conjecture 1 asks whether ϑ-rank$(G) \leq \alpha(G) - 1$.

We recap some of the known results on these conjectures. In view of (1.6), if $\alpha(G) = \chi(\overline{G})$ then $\vartheta^{(0)}(G) = \alpha(G)$ and thus G has ϑ-rank 0; this holds, e.g., for perfect graphs [24]. Every graph satisfying $\vartheta(G) = \alpha(G)$ also has ϑ-rank 0; this is the case, e.g, for the Petersen graph and, more generally, for Kneser graphs [23]. It is known that odd cycles and wheels have ϑ-rank 1 and thus satisfy Conjecture 1 [10]. Conjecture 1 has been shown to hold for all graphs with $\alpha(G) \leq 8$ in [12] (see also [42] for the case $\alpha(G) \leq 6$), but the general case is still wide open. Note that the conjectured bound $\alpha(G) - 1$ on ϑ-rank(G) is tight. As a first example, the cycle C_5 has $\alpha(C_5) = 2$ and ϑ-rank$(C_5) = 1$. As a second example, the complement of the icosahedron has $\alpha(G) = 3$ and ϑ-rank$(G) = 2$; indeed, ϑ-rank$(G) \geq 2$ as $\vartheta^{(1)}(G) = 1 + \sqrt{5} > 3$ [10], and ϑ-rank$(G) \leq 2$ as Conjecture 1 holds when $\alpha(G) = 3$.

In this paper we want to further investigate the above conjectures.

Links to other hierarchies of Lasserre type

Our approach is to relate the bounds $\vartheta^{(r)}(G)$ to other bounds that can be obtained by applying the Lasserre hierarchy to the polynomial optimization problem (M-S). For this consider the polynomials

$$f_G(x) = x^T(I + A_G)x \quad \text{and} \quad F_G(x) = f_G(x^{\circ 2}) = (x^{\circ 2})^T(I + A_G)x^{\circ 2}.$$

That is, $f_G = p_M$ and $F_G = P_M$ for the matrix $M = I + A_G$ (recall (1.2)). Yet another reformulation of (M-S) is that $\alpha(G)$ can also be obtained via polynomial optimization over the unit sphere:

$$\frac{1}{\alpha(G)} = \min\Big\{F_G(x) : x \in \mathbb{R}^n, \sum_{i=1}^n x_i^2 = 1\Big\}. \qquad \text{(M-S-Sphere)}$$

Now one can obtain bounds on $\alpha(G)$ by applying the sum-of-squares approach of Lasserre [17] to any of the two formulations (M-S) and (M-S-Sphere). First we recall some notation. Given polynomials $g_0 = 1, g_1, \ldots, g_m \in \mathbb{R}[x]$ and $r \in \mathbb{N}$ define the sets

$$\mathcal{M}(g_1, \ldots, g_m)_r = \Big\{\sum_{j=0}^m \sigma_j g_j : \sigma_j \in \Sigma, \deg(\sigma_j g_j) \le 2r\Big\}, \tag{1.7}$$

$$\mathcal{T}(g_1, \ldots, g_m)_r = \mathcal{M}\Big(\prod_{j\in J} g_j : J \subseteq [m]\Big)_r, \tag{1.8}$$

known, respectively, as the quadratic module and the preordering generated by the g_j's, truncated at degree $2r$. In addition, given polynomials $h_1, \ldots, h_k \in \mathbb{R}[x]$, the set

$$\langle h_1, \ldots, h_k\rangle_r = \Big\{\sum_{i=1}^k u_i h_i : u_i \in \mathbb{R}[x], \deg(u_i h_i) \le r\Big\} \tag{1.9}$$

is the ideal generated by the h_i's, truncated at degree r. Throughout, $\mathbb{R}[x]_r$ denotes the set of polynomials with degree at most r and we set $\Sigma_r = \Sigma \cap \mathbb{R}[x]_{2r}$, which consists of all polynmials of the form $\sum_i p_i^2$ for some $p_i \in \mathbb{R}[x]_r$. Corresponding to problems (M-S) and (M-S-Sphere) we now define the parameters

$$f_G^{(r)} = \sup\Big\{\lambda : f_G - \lambda \in \mathcal{M}(x_1, \ldots, x_n)_r + \Big\langle 1 - \sum_{i=1}^n x_i\Big\rangle_{2r}\Big\}, \tag{1.10}$$

$$f_{G,po}^{(r)} = \sup\Big\{\lambda : f_G - \lambda \in \mathcal{T}(x_1, \ldots, x_n)_r + \Big\langle 1 - \sum_{i=1}^n x_i\Big\rangle_{2r}\Big\}, \tag{1.11}$$

$$F_G^{(r)} = \sup\Big\{\lambda : F_G - \lambda \in \Sigma_r + \Big\langle 1 - \sum_{i=1}^n x_i^2\Big\rangle_{2r}\Big\}, \tag{1.12}$$

which clearly satisfy $1/\alpha(G) \ge f_{G,po}^{(r)} \ge f_G^{(r)}$, $1/\alpha(G) \ge F_G^{(r)}$ and $F_G^{(2r)} \ge f_G^{(r)}$ for any $r \in \mathbb{N}$. We will establish further links, also to the parameters $\vartheta^{(r)}(G)$. In particular, we show that the approach based on approximating the copositive cone by the cones $\mathcal{K}_n^{(2r)}$ (as in (1.5)) and the approach based on using the preordering truncated at degree $r+1$ (as in (1.11)) are equivalent: for any $r \ge 0$ we have

$$\frac{1}{\alpha(G)} \ge \frac{1}{\vartheta^{(2r)}(G)} = F_G^{(2r+2)} = f_{G,po}^{(r+1)} \ge f_G^{(r+1)}. \tag{1.13}$$

We say that *finite convergence* holds for the parameter $f_G^{(r)}$ if $f_G^{(r)} = 1/\alpha(G)$ for some $r \in \mathbb{N}$, and analogously for the other parameters. Based on the inequalities (1.13) we see that finite convergence for the parameters $f_G^{(r)}$ implies finite convergence for the other parameters, and thus in particular for $\vartheta^{(r)}(G)$, which would settle Conjecture 2.

Role of critical edges

Our first main result is showing finite convergence of the bounds $f_G^{(r)}$ for the class of *acritical graphs.* Recall that an edge e of G is said to be *critical* if $\alpha(G\backslash e) = \alpha(G) + 1$. The graph G is called α*-critical* (or, simply, *critical*) when all its edges are critical, and *acritical* when G does not have any critical edge. For example, odd cycles are α-critical while even cycles are acritical. Critical edges and critical graphs have been studied in the literature; see, e.g. [25]. It turns out that the notion of critical edges plays a central role in the study of the finite convergence of the above hierarchies of bounds.

On the one hand, it can be easily observed that deleting noncritical edges can only increase the ϑ-rank. Indeed, if $\alpha(G \setminus e) = \alpha(G)$ then $M_G - M_{G\setminus e} = \alpha(G)(A_G - A_{G\setminus e})$ is entry-wise nonnegative and thus belongs to $\mathcal{K}^{(0)} \subseteq \mathcal{K}^{(r)}$. Hence, $M_{G\setminus e} \in \mathcal{K}_n^{(r)}$ implies $M_G \in \mathcal{K}_n^{(r)}$, which shows ϑ-rank$(G) \le \vartheta$-rank$(G \setminus e)$. Hence, after iteratively deleting noncritical edges, we obtain a subgraph H of G which is critical with $\alpha(H) = \alpha(G)$ and

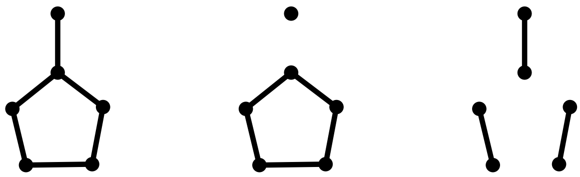

Figure 1: Graph G (left), graph H_1 (middle), graph H_2 (right)

satisfies: ϑ-rank$(G) \leq \vartheta$-rank(H). As shown in Example 1.1 below this inequality can be strict. Therefore, finite convergence of the parameters $\vartheta^{(r)}(G)$ (or $f^{(r)}_{G,po}, F^{(r)}_G$) for the class of *critical* graphs implies the same property for general graphs. Summarizing, it would suffice to show Conjectures 1 and 2 for the class of *critical* graphs.

On the other hand, we can show finite convergence of the parameters $f^{(r)}_G$ for the class of *acritical* graphs (see Theorem 5.1) and thus Conjecture 2 holds for acritical graphs (see Corollary 5.2).

It turns out that critical edges also play a crucial role in the analysis of the graphs with ϑ-rank 0. In [22] we can indeed characterize the *critical* graphs with ϑ-rank 0, namely, as those that be covered by $\alpha(G)$ cliques, i.e., such that $\alpha(G) = \chi(\overline{G})$. In addition, we show that the problem of deciding whether a graph has ϑ-rank 0 can be algorithmically reduced to the same question restricted to the class of acritical graphs.

Example 1.1. *Consider the graph G in Figure 1, obtained by adding one pending node to the cycle C_5. Then, $\alpha(G) = 3 = \overline{\chi}(G)$ and thus ϑ-rank$(G) = 0$. Note that G has two critical subgraphs H_1 and H_2 with $\alpha(H_1) = \alpha(H_2) = 3$, shown in Figure 1: H_1 is C_5 with an isolated node, which has ϑ-rank$(H_1) = 1$ (see, e.g., [10]), while H_2 consists of three independent edges with ϑ-rank$(H_2) = 0$ (since $\alpha(H_2) = \overline{\chi}(H_2) = 3$).*

Number of global minimizers and finite convergence

A main reason why critical edges play a role in the study of finite convergence comes from the fact that problem (M-S) has infinitely many global minimizers when G has critical edges. Indeed, next to the global minimizers arising from the maximum stable sets (of the form $\chi^S/\alpha(G)$ with S stable of size $\alpha(G)$), also some special convex combinations of them are global minimizers when G has critical edges; see Proposition 4.3. Our approach to prove finite convergence of the bounds $f^{(r)}_G$ is to apply a result by Nie [31] (itself based on the so-called Boundary Hessian Condition of Marshall [27]), which requires to check whether the classical sufficient optimality conditions hold at all global minimizers of (M-S). These conditions imply in particular that the problem must have finitely many minimizers, which explains why we can only apply it to acritical graphs.

There is an easy remedy to force having finitely many minimizers, simply by perturbing the Motzkin-Straus formulation (M-S). Indeed, if we replace the adjacency matrix A_G by $(1+\epsilon)A_G$ for any $\epsilon > 0$, then the corresponding quadratic program still has optimal value $1/\alpha(G)$, but now the only global minimizers are those arising from the maximum stable sets. To get this property it would in fact suffice to perturb the adjacency matrix at the positions corresponding to the critical edges of G. For the hierarchies of parameters obtained via this perturbed formulation we can show the finite convergence property, see Theorem 5.8. However, since we do not know a bound on the order of convergence, which *does not depend on* ϵ, it remains unclear how this can be used to derive the finite convergence of the original (unperturbed) parameters.

Nevertheless, as a byproduct of our analysis of the minimizers of the (perturbed) Motzkin-Straus formulation, we can show NP-hardness of the problem of deciding whether a standard quadratic optimization problem has finitely many global minimizers. The key idea is to reduce it to the problem of testing critical edges which is itself NP-hard, see Section 6.

Links to related literature

Given a graph G define the poynomial $Q_G(x) = (x^{\circ 2})^T(\alpha(G)(I + A_G) - J)x^{\circ 2}$, which is an even form (i.e., a homogeneous polynomial, with all variables appearing with an even degree) with degree 4. As Q_G is nonnegative on $\mathbb{R}^n$, by Artin's theorem, it can be written as a sum of squares of rational functions: $Q_G = \sum_{i=1}^m p_i^2/h^2$ for some $p_i, h \in \mathbb{R}[x]$. Then, what Conjecture 2 claims is that the denominator h^2 can be chosen to be of the form $(\sum_i x_i^2)^r$ for some $r \in \mathbb{N}$. Note that if Q_G would be strictly positive (i.e., vanish only at the origin) then this claim would follow from a result of Pólya [35] (see also Reznick [37]). However, the polynomial Q_G is not strictly positive, since any global minimizer of problem (M-S) provides a nonzero root of Q_G lying in Δ_n. On the positive side, Scheiderer [39] shows that if Q is an arbitrary form in three variables that is nonnegative on $\mathbb{R}^3$ then it is indeed true that $(\sum_{i=1}^3 x_i^2)^r Q \in \Sigma$ for some $r \in \mathbb{N}$. On the negative side, for any $n \geq 4$, there are examples of n-variate nonnegative polynomials Q for which $(\sum_i x_i^2)^r Q \notin \Sigma$ for all $r \in \mathbb{N}$; such Q can be chosen to be an even form of degree 4 for $n \geq 7$ (following arguments in [11]). So Conjecture 2 claims a rather remarkable property for the class of forms Q_G (and Conjecture 1 claims an even stronger property). In this paper we will show that Conjecture 2 holds when the graph G is acritical, which corresponds to the case when the polynomial Q_G has finitely many zeros in the simplex Δ_n. We will in fact show this property for a larger class of degree 4 even forms (see Section 5.2).

Our approach relies on considering the Lasserre hierarchy (1.10) for problem (M-S) and using the fact that its finite convergence implies finite convergence of the hierarchy $\vartheta^{(r)}(G)$ (in view of (1.13)). The goal is thus to show finite convergence of Lasserre hierarchy (1.10) or, equivalently, that the polynomial $f_G - 1/\alpha(G) = x^T(I + A_G)x - 1/\alpha(G)$ belongs to the quadratic module $\mathcal{M}(x_1, \ldots, x_n, \pm(1 - \sum_i x_i))$. The question of identifying sufficient conditions for finite convergence of Lasserre hierarchy applied to a polynomial optimization problem has been much studied in the literature; see, in particular, the works by Scheiderer [38, 39], Marshall [27, 28, 29], Kriel and Schweighofer [15, 16], Nie [31], and references therein. Assume f is a polynomial nonnegative on a basic closed semialgebraic set K defined by polynomial (in)equalities, whose associated quadratic module $\mathcal{M}$ is Archimedean. Marshall [29, Theorem 1.3] gives a set of algebraic conditions on the zeros of the polynomial f in the set K, known as the *Boundary Hessian Condition* (BHC), that ensures that f belongs to the quadratic module $\mathcal{M}$. Nie [31] shows that (BHC) holds if the natural sufficient optimality conditions hold at all the global minimizers of f over K and thus Lasserre hierarchy has finite convergence (see Theorem 2.3 below). Note that a restriction to the application of these results is that these optimality conditions (and (BHC)) can hold only when the number of global minimizers is finite. Since these conditions depend on the optimization problem, one faces the same issues also when using the (richer) preordering instead of the quadratic module. We make this remark in view of the equivalent reformulation of the parameters $\vartheta^{(r)}(G)$ in terms of the preordering-based hierarchy $f^{(r)}_{G,po}$ mentioned in (1.13). Let us also mention that while the result in [29] does not come with a degree bound for the order of the relaxation where finite convergence takes place, such a degree bound is given in [15]. However the results in [15] require (among others) the additionnal restriction that the finitely many global minimizers should all lie in the interior of the set K, which is not the case for problem (M-S), neither for its perturbations introduced in the paper. Finally, there are other results that show finite convergence of the Lasserre hierarchy, for instance, under some convexity assumptions (see [8, 19]), or when the semi-algebraic set K is finite (see Nie [32]), or when the description of the set K is enriched with various additional polynomial constraints (e.g., arising from KKT conditions) (see, e.g., [13, 33] and further references therein).

There is also interest in the literature in understanding when the first level of Lasserre hierarchy (also known as the Shor relaxation or the basic semidefinite relaxation) is exact when applied to quadratic optimization problems (see, e.g., the recent papers [7, 43] and further references therein). For standard quadratic programs, where one wants to minimize a quadratic form $p_M(x) = x^T M x$ over Δ_n, we characterize the set of matrices M for which the first level relaxation is exact. Moreover, we show that this holds precisely when the first level relaxation is feasible (see Lemma 3.1). In the special case of problem (M-S), when $M = I + A_G$, the first level relaxation gives the parameter $f_G^{(1)}$, which will be shown to be exact (i.e., equal to $1/\alpha(G)$) precisely when the graph G is a disjoint union of cliques (see Lemma 3.13). One can also use the preordering instead of the quadratic module and ask when the corresponding first level relaxation is exact. For problem (M-S) this amounts to asking when $f^{(1)}_{G,po} = 1/\alpha(G)$ or, equivalently (in view of (1.13)), when $\vartheta^{(0)}(G) = \alpha(G)$. Characterizing these graphs seems difficult in general, but, when restricting to critical graphs, $\vartheta^{(0)}(G) = \alpha(G)$ if and only if G can be covered by $\alpha(G)$ cliques (see [22]).

Finally let us point out that the hierarchies considered in this paper are all based on continuous formulations of the stability number. Alternatively, one can formulate $\alpha(G)$ as the maximum value of $\sum_{i \in V} x_i$ taken over all x in the discrete cube $\{0,1\}^n$ that satisfy the edge constraints $x_i + x_j \leq 1$ for all $\{i,j\} \in E$. One can model the binary variables by the quadratic constraints $x_i^2 = x_i$ ($i \in [n]$) and apply the Lasserre/Parrilo approach, which provides a hierarchy of bounds, known to converge to $\alpha(G)$ in finitely many steps, in fact in $\alpha(G)$ steps [18, 20]. When adding suitable nonnegativity conditions one gets parameters $\text{las}_r(G)$ satisfying $\alpha(G) \leq \text{las}_r(G) \leq \vartheta^{(r-1)}(G)$, with $\text{las}_1(G) = \vartheta^{(0)}(G)$ [12]. Hence, what Conjecture 1 claims is that the continuous copositive-based hierarchy $\vartheta^{(r)}(G)$ has the same finite convergence behaviour as the discrete Lasserre hierarchy. As observed above this question is also relevant to several other interesting aspects of real algebraic geometry.

Organization of the paper

The paper is organized as follows. In Section 2, we recall the classical optimality conditions in nonlinear programming and their use to show finite convergence of the Lasserre hierarchy for polynomial optimization. In Section 3 we link several sum-of-squares approximation hierarchies for standard quadratic programs and we discuss some questions about the feasibily and exactness of these relaxations and their application to the Motzkin-Straus formulation (M-S). Section 4 is focused on the study of the minimizers of problem (M-S), where, in particular, we prove that the problem has finitely many minimizers precisely when the graph is acritical. In Section 5 we apply the previous results to show finite convergence of the semidefinite hierarchy $\vartheta^{(r)}(G)$ for acritical graphs. In addition, we propose perturbed hierarchies for the stability number and we give some facts and open questions about them. In Section 6, we investigate the complexity of the problem of deciding whether a standard quadratic program has finitely many minimizers.

Notation

Notation about polynomials will be given in Section 2, but here we group some notation about graphs and matrices that is used throughout the paper. Given a graph $G = (V = [n], E)$, a set $S \subseteq V$ is stable if it does not contain an edge, and $\alpha(G)$ is the maximum cardinality of a stable set. A set $C \subseteq V$ is a clique if any two distinct vertices in C are adjacent, and $\chi(\overline{G})$ denotes the minimum number of cliques whose union is V. For a set $S \subseteq V$ and a vertex $j \in V \setminus S$, we set $\deg_S(j) = |N_S(j)|$, where $N_S(j) = \{i \in S : \{i, j\} \in E\}$ denotes the set of vertices in S adjacent to j. Given two sets $S, T \subseteq V$ we set $N_S(T) = \{i \in S : \{i, j\} \in E \text{ for some } j \in T\}$. An edge $e \in E$ is critical if $\alpha(G \setminus e) = \alpha(G) + 1$, G is called critical if all its edges are critical and G is called acritical if none of its edges is critical. Observe that G is acritical precisely when $\deg_S(j) \geq 2$ for every stable set S with $|S| = \alpha(G)$ and every $j \in V \setminus S$. For a subset $U \subseteq V$, $G[U]$ denotes the induced subgraph, with vertex set U and edges the pairs $\{i, j\} \in E$ that are contained in U. For a vector $x \in \mathbb{R}^n$ we let $\text{Supp}(x) = \{i \in [n] : x_i \neq 0\}$ denote the support of x. In addition, $e = (1, \ldots, 1)^T$ denotes the all-ones vector, $\{e_1, \ldots, e_n\}$ denotes the standard unit basis of $\mathbb{R}^n$, $I \in \mathcal{S}^n$ denotes the identity matrix and $J = ee^T \in \mathcal{S}^n$ the all-ones matrix. We also use the symbols J_n and $J_{n,m}$ to denote the all-ones matrix of size $n \times n$ and $n \times m$, respectively.

2 Preliminaries on polynomial optimization

Given polynomials f, g_j for $j \in [m]$, and h_i for $i \in [k]$, consider the polynomial optimization problem:

$$f_{\min} = \inf\{f(x) : g_j \geq 0 \ (j \in [m]), h_i(x) = 0 \ (i \in [k])\} = \inf\{f(x) : x \in K\}, \tag{P}$$

setting $K = \{x \in \mathbb{R}^n : g_j(x) \geq 0 \ (j \in [m]), h_i(x) = 0 \ (i \in [k])\}$. A well-known approach for solving problem (P) is the Lasserre-Parrilo approach, which is based on using positivity certificates arising from suitable sums of squares representations for polynomials that are nonnegative over the feasible set K. Such positivity certificates arise by considering the (truncated) quadratic module, preordering and ideal introduced in relations (1.7), (1.8) and (1.9). Set $g = (g_1, \ldots, g_m)$ and $h = (h_1, \ldots, h_k)$ for a short-hand, and $\mathcal{M}(g) = \bigcup_{r \geq 0} \mathcal{M}(g)_r$, $\langle h \rangle = \bigcup_{r \geq 0} \langle h \rangle_r$. Then $\mathcal{M}(g) + \langle h \rangle$ is said to be *Archimedean* if the polynomial $R^2 - \sum_{i=1}^n x_i^2$ belongs to $\mathcal{M}(g) + \langle h \rangle$ for some $R \in \mathbb{R}$. Note this implies that K is compact. The following results by Schmüdgen [40] and Putinar [36] play a central role in polynomial optimization.

Theorem 2.1. *Assume the feasible region K of* (P) *is compact. Then any polynomial that is strictly positive on K belongs to $\mathcal{T}(g) + \langle h \rangle$ (Schmüdgen [40]). If in addition $\mathcal{M}(g) + \langle h \rangle$ is Archimedean, then any polynomial that is strictly positive on K belongs to $\mathcal{M}(g) + \langle h \rangle$ (Putinar [36]).*

Using the truncated quadratic module and preordering leads to the parameters:

$$f^{(r)} := \sup\{\lambda : f - \lambda \in \mathcal{M}(g)_r + \langle h \rangle_{2r}\}, \tag{2.1}$$

$$f^{(r)}_{po} := \sup\{\lambda : f - \lambda \in \mathcal{T}(g)_r + \langle h \rangle_{2r}\}, \tag{2.2}$$

to which we will refer as the *Lasserre hierarchy* (or the *sum-of-squares hierarchy*), sometimes adding the adjective 'preordering-based' when referring to $f^{(r)}_{po}$. Clearly we have $f^{(r)} \leq f^{(r)}_{po} \leq f_{\min}$, $f^{(r)} \leq f^{(r+1)}$ and $f^{(r)}_{po} \leq f^{(r+1)}_{po}$ for all r. As a direct application of Theorem 2.1, the parameters $f^{(r)}_{po}$ converge asymptotically to $f_{\min}$ when K is compact, while the (possibly weaker) parameters $f^{(r)}$ also converge asymptotically to $f_{\min}$ under the Archimedean condition. We are interested in problems for which the Lasserre hierarchy has *finite* convergence. We say the parameters $f^{(r)}$ have *finite convergence* if $f^{(r)} = f_{\min}$ for some $r \in \mathbb{N}$; analogously for the parameters $f^{(r)}_{po}$.

In order to prove finite convergence of the Lasserre hierarchy for some special classes of polynomial optimization problems, we will use a result of Nie [31], which relies on the optimality conditions for nonlinear optimization. So we start with a quick recap on these optimality conditions, which we state here for problem (P) though they hold in a more general setting (see, e.g., [4]).

Let u be a local minimizer of problem (P) and let $J(u) = \{j \in [m] : g_j(u) = 0\}$ be the index set of the active inequality constraints at u. We say that the *constraint qualification condition (CQC)* holds at u if the gradients of the active constraints at u are linearly independent:

$$\text{The vectors in } \{\nabla g_j(u) : j \in J(u)\} \cup \{\nabla h_i(u) : i \in [k]\} \text{ are linearly independent.} \tag{CQC}$$

If (CQC) holds at u then there exist Lagrange multipliers $\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_m \in \mathbb{R}$ satisfying

$$\nabla f(u) = \sum_{i=1}^{k} \lambda_i \nabla h_i(u) + \sum_{j=1}^{m} \mu_j \nabla g_j(u), \tag{FOOC}$$

$$\mu_1 g_1(u) = 0, \ldots, \mu_m g_m(u) = 0, \mu_1 \geq 0, \ldots \mu_m \geq 0. \tag{CC}$$

The condition (FOOC) is known as the *first order optimality condition* and (CC) as the *complementarity condition*. If it holds that

$$\mu_j > 0 \text{ for every } j \in J(u), \quad \mu_j = 0 \text{ for } j \in [m] \setminus J(u), \tag{SCC}$$

then we say that the *strict complementarity condition* (SCC) holds at u. Define the Lagrangian function

$$L(x) = f(x) - \sum_{i=1}^{k} \lambda_i h_i(x) - \sum_{j \in J(u)} \mu_j g_j(x).$$

Another necessary condition for u to be a local minimizer is the *second order necessity condition* (SONC):

$$v^T \nabla^2 L(u) v \geq 0 \text{ for all } v \in G(u)^\perp, \tag{SONC}$$

where $G(u)$ is the matrix with rows the gradients of the active constraints at u and $G(u)^\perp$ is its kernel:

$$G(u)^\perp = \{x \in \mathbb{R}^n : x^T \nabla g_j(u) = 0 \text{ for all } j \in J(u) \text{ and } x^T \nabla h_i(u) = 0 \text{ for all } i \in [k]\}.$$

If it holds that

$$v^T \nabla^2 L(u) v > 0 \text{ for all } 0 \neq v \in G(u)^\perp, \tag{SOSC}$$

then we say that the *second order sufficiency condition* (SOSC) holds at u. The relations between these optimality conditions and the local minimizers are summarized in the following classical result.

Theorem 2.2 (see, e.g., [4]). *Let u be a feasible solution of problem* (P).

(i) *Assume u is a local minimizer of* (P) *and (CQC) holds at u. Then the conditions* (FOOC), (CC) *and* (SONC) *hold at u.*

(ii) *Assume that (FOOC), (SCC) and (SOSC) hold at u. Then u is a strict local minimizer of* (P).

The relation between the optimality conditions for problem (P) and finite convergence of the parameters $f^{(r)}$ is given by the following result of Nie [31].

Theorem 2.3 (Nie [31]). *Consider problem* (P) *and the parameters $f^{(r)}$ from (2.1). Assume that the Archimedean condition holds, i.e., $R^2 - \sum_{i=1}^n x_i^2 \in \mathcal{M}(g) + \langle h \rangle$ for some $R \in \mathbb{R}$, and that the constraint qualification* (CQC), *strict complementary* (SCC) *and second order sufficency* (SOSC) *conditions hold at every global minimizer of* (P). *Then Lasserre's hierarchy $f^{(r)}$ has finite convergence, i.e., we have $f^{(r)} = f_{\min}$ for some $r \in \mathbb{N}$.*

Note that, under the assumptions of Theorem 2.3, all global minimizers of (P) are *strict* minimizers (by Theorem 2.2 (ii)) and thus (P) has *finitely many* global minimizers.

3 Links between the various hierarchies

In this section we prove relation (1.13), which establishes links between the various hierarchies of bounds $\vartheta^{(r)}(G)$, $f_G^{(r)}$ $f_{G,po}^{(r)}$ and $F_G^{(r)}$ from relations (1.5), (1.10), (1.11) and (1.12). We start with establishing these links in the more general setting of standard quadratic programs.

3.1 Links between the hierarchies for standard quadratic programs

Given a symmetric matrix $M \in \mathcal{S}^n$, recall the polynomials $p_M(x) = x^T Mx$ and $P_M(x) = p_M(x^{\circ 2})$ from (1.2). We consider the following *standard quadratic optimization problem*:

$$p_{\min} = \min \Big\{ p_M(x) : x \in \Delta_n \Big\}, \tag{3.1}$$

which can be equivalently reformulated as optimization over the unit sphere:

$$p_{\min} = \min \Big\{ P_M(x) : x \in \mathbb{R}^n, \sum_{i=1}^n x_i^2 = 1 \Big\}. \tag{3.2}$$

In analogy to definitions (1.10), (1.11) and (1.12) we can define the corresponding sum-of-squares hierarchies for both problems (3.1) and (3.2), and the preordering-based hierarchy for the simplex formulation (3.1), leading to the parameters

$$p_M^{(r)} = \max \Big\{ \lambda : p_M - \lambda \in \mathcal{M}(x_1, x_2, \dots, x_n)_r + \Big\langle \sum_{i=1}^n x_i - 1 \Big\rangle_{2r} \Big\}, \tag{3.3}$$

$$p_{M,po}^{(r)} = \max \Big\{ \lambda : p_M - \lambda \in \mathcal{T}(x_1, x_2, \dots, x_n)_r + \Big\langle \sum_{i=1}^n x_i - 1 \Big\rangle_{2r} \Big\}, \tag{3.4}$$

$$P_M^{(r)} = \max \Big\{ \lambda : P_M - \lambda \in \Sigma_r + \Big\langle \sum_{i=1}^n x_i^2 - 1 \Big\rangle_{2r} \Big\} \tag{3.5}$$

for any integer $r \geq 1$. Observe that we are in the Archimedean setting and that the above programs are feasible for any $r \geq 2$. To see this one can use the following identities: for any $i \in [n]$,

$$1 - x_i = 1 - \sum_{k=1}^n x_k + \sum_{k \in [n] \setminus \{i\}} x_k, \quad 1 - x_i^2 = \frac{(1 + x_i)^2}{2}(1 - x_i) + \frac{(1 - x_i)^2}{2}(1 + x_i).$$

This implies $n - \sum_i x_i^2 \in \mathcal{M}(x_1, \dots, x_n)_2 + \langle 1 - \sum_i x_i \rangle_4$, thus showing the Archimedean condition holds. We next verify feasibility of the programs. If $M \succeq 0$ then the polynomial p_M belongs to Σ_1 and thus the programs defining $p_M^{(1)}, p_{M,po}^{(1)}, P_M^{(2)}$ are feasible. Otherwise, $\mu := \lambda_{\min}(M) < 0$ and $p_M(x) - n\mu = x^T(M - \mu I)x - \mu(n - \sum_i x_i^2)$, which shows feasibility of the programs defining $p_M^{(r)}, p_{M,po}^{(r)}, P_M^{(r)}$ for $r \geq 2$. In addition note that $p_{M,po}^{(1)}$ is finite when M is entry-wise nonnegative. Observe also that the optimum is attained in the above programs since the search region for $p - \lambda$ is a closed set (see [26]).

Now, we characterize the set of matrices M for which the program (3.3) is feasible at order $r = 1$. Moreover, we prove that in that case the program is exact, i.e., $p_M^{(1)} = p_{\min}$.

Lemma 3.1. *Given a symmetric matrix $M \in \mathcal{S}^n$, the following assertions are equivalent.*

(i) *The program (3.3) is feasible for $r = 1$, i.e., $p_M^{(1)}$ is finite.*

(ii) *There exist $\lambda \in \mathbb{R}$ and $a \in \mathbb{R}^n_+$ such that $M - \lambda J - (ae^T + ea^T)/2 \succeq 0$.*

(iii) $p_M^{(1)} = p_{\min}$.

Proof. We first prove (i) $\iff$ (ii). Assume program (3.3) is feasible, i.e., there exist $\lambda \in \mathbb{R}$, $a \in \mathbb{R}^n_+$, $Q \succeq 0$ and $u(x) \in \mathbb{R}[x]$ such that

$$x^T Mx - \lambda = x^T Qx + a^T x + (e^T x - 1)u(x).$$

Then there exists $v(x) \in \mathbb{R}[x]$ such that

$$x^T Mx - \lambda(e^T x)^2 = x^T Qx + (a^T x)(e^T x) + (e^T x - 1)v(x).$$

Hence the quadratic polynomial $x^T(M - \lambda J - Q - (ae^T + ea^T)/2)x$ vanishes on $\{x : e^T x = 1\}$ and thus on $\mathbb{R}^n$, which implies $M - \lambda J - Q - (ae^T + ea^T)/2 = 0$ and thus (ii) holds. The argument can be clearly reversed, which shows the equivalence of (i) and (ii).

As (iii) implies (i) it suffices now to show (ii) $\Longrightarrow$ (iii). By the above argument, if (ii) holds then we have

$$p_M^{(1)} = \sup\{\lambda : \lambda \in \mathbb{R}, a \in \mathbb{R}^n_+,\ M - \lambda J - (ae^T + ea^T)/2 \succeq 0\}. \tag{3.6}$$

Define the matrices $A_i = (e_i e^T + e e_i^T)/2$ for $i \in [n]$. Then the dual program of (3.6) reads

$$\inf\{\langle M, X\rangle : \langle J, X\rangle = 1, \langle A_i, X\rangle \geq 0\ (i \in [n]), X \succeq 0\}. \tag{3.7}$$

As program (3.7) is strictly feasible and bounded from below by $p_M^{(1)}$, strong duality holds and the optimum value of (3.7) is equal to $p_M^{(1)}$. We now show that $p_{\min} \leq p_M^{(1)}$. For this let X be feasible for (3.7) and define the vector $x = Xe$. Then $x \in \Delta_n$ since $x_i = \langle A_i, X\rangle \geq 0$ for all $i \in [n]$, and $e^T x = \langle J, X\rangle = 1$, which implies $x^T M x \geq p_{\min}$. In addition, we have $X - xx^T \succeq 0$, which follows from the fact that

$$\begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0,$$

(as $X \succeq 0$, $x = Xe$ and $e^T X e = 1$). Consider also a feasible solution (λ, a) to (3.6), so that $M - \lambda J - \sum_{i=1}^n a_i A_i \succeq 0$. Then we have $\langle M - \lambda J - \sum_i a_i A_i, X - xx^T\rangle \geq 0$ which, combined with $\langle J, X - xx^T\rangle = 0$ and $\langle A_i, X - xx^T\rangle = 0$ for all $i \in [n]$, implies that $\langle M, X\rangle \geq x^T M x \geq p_{\min}$ and thus $p_M^{(1)} \geq p_{\min}$, as desired. □

Here is an immediate consequence of the reformulation of the parameter $p_M^{(1)}$ given in (3.6), that we will need later.

Lemma 3.2. *Assume that the program (3.6) defining $p_M^{(1)}$ is feasible, i.e., $M = \lambda J + Q + (ae^T + ea^T)/2$ for some $\lambda \in \mathbb{R}$, $Q \succeq 0$ and $a \in \mathbb{R}^n_+$. Then, for any $i \neq j \in [n]$, we have $M_{ii} + M_{jj} - 2M_{ij} = Q_{ii} + Q_{jj} - 2Q_{ij} \geq 0$. In addition, if $M_{ii} + M_{jj} - 2M_{ij} = 0$ then $Q(e_i - e_j) = 0$.*

Proof. Direct verification. □

Alternatively, following [6, 10], problem (3.1) can be reformulated as a copositive program:

$$p_{\min} = \max\Big\{\lambda : M - \lambda J \in \text{COP}_n\Big\}. \tag{3.8}$$

By replacing the cone COP_n by its subcone $\mathcal{K}_n^{(r)}$ we now obtain the following lower bound for $p_{\min}$:

$$\Theta_M^{(r)} := \max\Big\{\lambda : M - \lambda J \in \mathcal{K}_n^{(r)}\Big\} \tag{3.9}$$

for any integer $r \geq 0$. Note that $\lambda = \min_{i,j} M_{ij}$ provides a feasible solution for (3.9) since then $M - \lambda J$ belongs to $\mathcal{K}_n^{(0)}$. We begin with the following easy relationships among the above parameters:

Lemma 3.3. *For all $r \geq 1$ we have: $\max\{p_M^{(r)}, p_{M,po}^{(r)}, P_M^{(r)}, \Theta_M^{(r)}\} \leq p_{\min}$ and $p_M^{(r)} \leq \min\{P_M^{(2r)}, p_{M,po}^{(r)}\}$.*

Proof. That all parameters are lower bounds for $p_{\min}$ follows from their definition, the inequality $p_M^{(r)} \leq p_{M,po}^{(r)}$ follows from the inclusion $\mathcal{M}(x_1, \ldots, x_n)_r \subseteq \mathcal{T}(x_1, \ldots, x_n)_r$ and, for the inequality $p_M^{(r)} \leq P_M^{(2r)}$, note that $p_M - \lambda \in \mathcal{M}(x_1, \ldots, x_n)_r + \langle 1 - \sum_i x_i\rangle_{2r}$ implies $P_M - \lambda \in \Sigma_{2r} + \langle 1 - \sum_i x_i^2\rangle_{4r}$. □

Following [9] we can now relate the bounds in (3.5) and (3.9). For this we use the following result from [9] (see Proposition 2 and Lemma 1 there).

Theorem 3.4 (de Klerk et al. [9])**.** *Let q be a form of even degree $2d \geq 2$. For any $r \in \mathbb{N}$ we have:*

$$q(x)\left(\sum_{i=1}^n x_i^2\right)^r \in \Sigma_{r+d} \Longleftrightarrow q \in \Sigma_{r+d} + \left\langle 1 - \sum_{i=1}^n x_i^2 \right\rangle_{2(r+d)}.$$

Lemma 3.5. *For any $M \in \mathcal{S}^n$ and $r \geq 0$, we have: $\Theta_M^{(r)} = P_M^{(r+2)}$.*

Proof. By definition, $\Theta_M^{(r)}$ is the largest λ for which the matrix $M - \lambda J$ belongs to the cone $\mathcal{K}_n^{(r)}$ or, equivalently, the polynomial $(\sum_i x_i^2)^r (P_M(x) - \lambda(\sum_i x_i^2)^2)$ belongs to Σ_{r+2}. In view of Theorem 3.4 this is equivalent to requiring that $P_M - \lambda(\sum_i x_i^2)^2$ belongs to $\Sigma_{r+2} + \langle 1 - \sum_i x_i^2 \rangle_{2(r+2)}$. Now observe that

$$P_M(x) - \lambda = P_M(x) - \lambda\Big(\sum_i x_i^2\Big)^2 + \lambda\Big(\Big(\sum_i x_i^2\Big)^2 - 1\Big), \tag{3.10}$$

where $(\sum_i x_i^2)^2 - 1 = (\sum_i x_i^2 - 1)(\sum_i x_i^2 + 1) \in \langle 1 - \sum_i x_i^2 \rangle_{2(r+2)}$. From this we obtain the desired identity $\Theta_M^{(r)} = P_M^{(r+2)}$. $\square$

Next we relate the preordering-based bound $p_{M,po}^{(r)}$ (for the simplex formulation) and the Lasserre bound $P_M^{(r)}$ (for the sphere formulation). For this we use the following result of [44].

Theorem 3.6 (Pena et al. [44])**.** *Let q be a homogeneous polynomial of degree d and define the polynomial $Q(x) = q(x^{\circ 2})$. Then, $Q \in \Sigma_d$ if and only if q can be decomposed as*

$$q(x) = \sum_{\substack{I \subseteq [n] \\ |I| \le d, |I| \equiv d \pmod 2}} \sigma_I(x) \prod_{i \in I} x_i, \tag{3.11}$$

where σ_I is a homogeneous polynomial with degree at most $d - |I|$ and $\sigma_I \in \Sigma$.

As an application we recall the characterization for the cone $\mathcal{K}_n^{(0)}$, consisting of the matrices M for which the polynomial $P_M(x) = (x^{\circ 2})^T M x^{\circ 2}$ is a sum of squares.

Proposition 3.7 (Parrilo [34])**.** *A matrix M belongs to $\mathcal{K}_n^{(0)}$ if and only if there exist matrices $P \succeq 0$ and $N \ge 0$ such that $M = P + N$, where we may assume without loss of generality that $N_{ii} = 0$ for all $i \in [n]$.*

Lemma 3.8. *For any $M \in \mathcal{S}^n$ and $r \ge 0$, we have: $p_{M,po}^{(r)} = P_M^{(2r)}$.*

Proof. First, assume λ is feasible for $p_{M,po}^{(r)}$, i.e.,

$$p_M(x) - \lambda = \sum_{\substack{I \subseteq [n] \\ |I| \le r, |I| \equiv r \pmod 2}} \sigma_I(x) \prod_{i \in I} x_i + u(x)\left(1 - \sum_{i=1}^n x_i\right),$$

where $\sigma_I \in \Sigma$ is a form of degree at most $2r - |I|$ and $\deg(u) \le 2r - 1$. Replacing throughout x by $x^{\circ 2}$ we obtain a decomposition of $P_M - \lambda$ in $\Sigma_{2r} + \langle 1 - \sum_i x_i^2 \rangle_{4r}$, which shows that $P_M^{(2r)} \ge p_{M,po}^{(r)}$. We now show the reverse inequality. For this assume λ is feasible for $P_M^{(2r)}$, i.e.,

$$P_M(x) - \lambda = \sigma(x) + \Big(1 - \sum x_i^2\Big) u(x),$$

where $\sigma \in \Sigma_{2r}$ and $\deg(u) \le 4r - 2$. Hence, using (3.10), the homogeneous polynomial $P_M(x) - \lambda(\sum_i x_i^2 - 1)^2$ belongs to $\Sigma_{2r} + \langle 1 - \sum_i x_i^2 \rangle_{4r}$. Applying Theorem 3.4 to it we can conclude that

$$\Big(\sum_{i=1}^n x_i^2\Big)^{2r-2} \Big(P_M(x) - \lambda\Big(\sum_{i=1}^n x_i^2\Big)^2\Big) \in \Sigma_{2r}.$$

Since this is a homogeneous polynomial in $x^{\circ 2}$ we can apply Theorem 3.6 to it and conclude that

$$\Big(\sum_{i=1}^n x_i\Big)^{2r-2} \Big(p_M(x) - \lambda\Big(\sum_{i=1}^n x_i\Big)^2\Big) = \sum_{\substack{I \subseteq [n] \\ |I| \le 2r, |I| \equiv 2r \pmod 2}} \sigma_I(x) \prod_{i \in I} x_i, \tag{3.12}$$

where $\sigma_I \in \Sigma$ has degree at most $2r - |I|$. Notice that

$$\Big(\sum_{i=1}^n x_i\Big)^{2r-2} = \Big(1 - 1 + \sum_{i=1}^n x_i\Big)^{2r-2} = 1 + h(x)\Big(1 - \sum_{i=1}^n x_i\Big),$$

for some $h \in \mathbb{R}[x]_{2r-3}$. Using this observation, (3.12) implies

$$p_M(x) - \lambda\Big(\sum_{i=1}^n x_i\Big)^2 \in \mathcal{T}(x_1,\dots,x_n)_r + \Big\langle 1 - \sum_{i=1}^n x_i \Big\rangle_{2r}.$$

Using again (3.10) we obtain

$$p_M(x) - \lambda \in \mathcal{T}(x_1,\dots,x_n)_r + \Big\langle 1 - \sum_{i=1}^n x_i \Big\rangle_{2r},$$

which shows $p^{(r)}_{M,po} \geq P^{(2r)}_M$. □

As a direct consequence of Lemmas 3.3, 3.5 and 3.8 we obtain the following links among the above parameters.

Corollary 3.9. *For any $M \in \mathcal{S}^n$ and $r \geq 0$ we have:*

$$p_{\min} \geq P^{(2r+2)}_M = \Theta^{(2r)}_M = p^{(r+1)}_{M,po} \geq p^{(r+1)}_M. \tag{3.13}$$

Remark 3.10. *In view of the formulation (3.6) for the parameter $p^{(1)}_M$, the difference with the parameter $p^{(1)}_{M,po} = P^{(2)}_M = \Theta^{(0)}_M$ lies in the fact that, while for $p^{(1)}_M$ we search for a decomposition $M = \lambda J + Q + (ea^T + ae^T)/2 \succeq 0$ with $Q \succeq 0$ and $a \in \mathbb{R}^n_+$, in the definition of $\Theta^{(0)}_M$ we search for a decomposition $M = \lambda J + Q + N \succeq 0$ with $Q \succeq 0$, but now N can be an arbitrary entry-wise nonnegative matrix.*

3.2 Application to the stable set problem

Here we apply the results in Section 3.1 to the formulation of the stability number $\alpha(G)$ via the Motzkin-Straus formulation (M-S), the special instance of standard quadratic program where we select the matrix $M = I + A_G$. As in the introduction we set $f_G = p_M$, $F_G = P_M$ and $f_{G,po} = p_{M,po}$ for this matrix $M = I + A_G$. As a direct application of Corollary 3.9, we obtain

$$\frac{1}{\alpha(G)} \geq F^{(2r+2)}_G = f^{(r+1)}_{G,po} \geq f^{(r+1)}_G. \tag{3.14}$$

It remains to link the parameters $\vartheta^{(r)}(G)$ and $\Theta^{(r)}_M$ for the matrix $M = I + A_G$.

Lemma 3.11. *For any graph G and $r \geq 0$, we have: $\Theta^{(r)}_{I+A_G} = \dfrac{1}{\vartheta^{(r)}(G)}$.*

Proof. Directly from the definitions of $\vartheta^{(r)}(G)$ in (1.5) and $\Theta^{(r)}_{I+A_G}$ in (3.9). □

Combining (3.14) and Lemma 3.11 we obtain the inequalities claimed in (1.13), which we repeat here for convenience.

Corollary 3.12. *For any graph G and $r \geq 0$ we have*

$$\frac{1}{\alpha(G)} \geq \frac{1}{\vartheta^{(2r)}} = F^{(2r+2)}_G = f^{(r+1)}_{G,po} \geq f^{(r+1)}_G.$$

We now use the result of Lemma 3.1 to characterize when the parameter $f^{(1)}_G$ is feasible (and thus exact).

Lemma 3.13. *For any graph G, the parameter $f^{(1)}_G$ is finite or, equivalently, $f^{(1)}_G = 1/\alpha(G)$, if and only if G is a disjoint union of cliques.*

Proof. We use Lemma 3.1 applied to the matrix $M = I + A_G$. First, assume $M = \lambda J + Q + (ae^T + ea^T)/2$ for some $\lambda \in \mathbb{R}$, $Q \succeq 0$ and $a \in \mathbb{R}^n_+$, we show that G is a disjoint union of cliques. For this it suffices to show that $\{1,2\},\{1,3\} \in E$ implies $\{2,3\} \in E$. This follows easily using Lemma 3.2. Indeed, if $\{1,2\},\{1,3\} \in E$ then we have $M_{11} + M_{22} - 2M_{12} = 0$ and thus $Q(e_1 - e_2) = 0$ and, in the same way, $Q(e_1 - e_3) = 0$. This implies $Q(e_2 - e_3) = 0$ and thus $M_{22} + M_{33} - 2M_{23} = 0$, i.e., $\{2,3\} \in E$.
Conversely, assume G is a disjoint union of cliques, say $V = C_1 \cup \ldots \cup C_k$ where $k = \alpha(G)$ and each C_i is a clique of G. We show that $p^{(1)}_M = p_{\min}$. For this note that, for any $x \in \Delta_n$, we have

$$x^T(I + A_G)x = \sum_{i=1}^k \Big(\sum_{j \in C_i} x_j\Big)^2 \geq \frac{1}{k} = \frac{1}{\alpha(G)},$$

where we use Cauchy-Schwartz inequality for the inequality. This shows $p^{(1)}_M \geq p_{\min}$ and thus equality holds. □

In Section 5 we will investigate finite convergence of the hierarchy $f_G^{(r)}$, which, in view of Corollary 3.12, directly implies finite convergence of the hierarchy $\vartheta^{(r)}(G)$. For this we will use Theorem 2.3 that requires to understand the structure of the global minimizers of problem (M-S), which is what we do in the next section.

4 Minimizers of the Motzkin-Straus formulation

In this section we prove some properties of the (global and local) minimizers of the Motzkin-Straus formulation (M-S) for the stability number $\alpha(G)$. In particular, we characterize the global minimizers and we prove that their number is finite if and only if the graph G has no critical edges (see Proposition 4.3 and Corollary 4.4). As mentioned earlier the property of having finitely many minimizers is indeed important in the analysis of finite convergence of the Lasserre hierarchy.

We start with a useful property of the local minimizers for a class of standard quadratic programs.

Lemma 4.1. *Consider the standard quadratic program*

$$p_{\min} = \min\{p_M(x) = x^T M x : x \in \Delta_n\}, \tag{4.1}$$

where M is a matrix of the form

$$M = \begin{pmatrix} 1 & 1 & a_1^T \\ 1 & 1 & a_2^T \\ a_1 & a_2 & M_0 \end{pmatrix}, \tag{4.2}$$

i.e., with $M_{11} = M_{12} = M_{22} = 1$, $a_1, a_2 \in \mathbb{R}^{n-2}$ and $M_0 \in \mathcal{S}^{n-2}$. Assume x is a local minimizer of (4.1) with $x_1, x_2 > 0$ and define the vectors $\tilde{x} = x + x_2 e_1 - x_2 e_2$ and $\overline{x} = x - x_1 e_1 + x_1 e_2 \in \Delta_n$. Then, for any scalar $t \in [0,1]$, we have $p_M(t\tilde{x} + (1-t)\overline{x}) = p_M(x)$ and thus, in particular, $p_M(\tilde{x}) = p_M(\overline{x}) = p_M(x)$. In addition, if x is a strict local minimizer of (4.1) then $x_1 x_2 = 0$.

Proof. Set $z := (x_3, x_4, \dots, x_n) \in \mathbb{R}^{n-2}$ and consider the optimization problem obtained from (4.1), where we let the first two variables free and fix the remaining $n-2$ variables equal to z:

$$\min\{(y_1 + y_2)^2 + 2y_1 a_1^T z + 2y_2 a_2^T z + z^T M_0 z : y_1, y_2 \geq 0,\ y_1 + y_2 = 1 - e^T z\}. \tag{4.3}$$

Since $y_1 + y_2 = 1 - e^T z$ is constant, problem (4.3) is a linear program and therefore it has an optimal solution with $y_1 = 0$ or $y_2 = 0$. Assume, e.g., that $a_1^T z \leq a_2^T z$, then $(y_1 = 1 - e^T z, y_2 = 0)$ is an optimal solution of (4.3), with objective value $p_M(\tilde{x})$. In addition, we have $p_M(\tilde{x}) \leq p_M(x)$ since (x_1, x_2) is feasible for (4.3). We claim that $p_M(\tilde{x}) = p_M(x)$. For this assume $p_M(\tilde{x}) < p_M(x)$ or, equivalently, $a_1^T z < a_2^T z$. Then, for any $t \in (0,1)$, one can easily verify that $p_M(tx + (1-t)\tilde{x}) < p_M(x)$. Since this holds for any t close to 1, we contradict the assumption that x is a local minimum of (4.1). Therefore, equality $p_M(\tilde{x}) = p_M(x)$ holds, which implies $a_1^T z = a_2^T z$. In turn this also implies that $p_M(x) = p_M(\overline{x})$ and thus $p_M(x) = p_M(t\tilde{x} + (1-t)\overline{x})$ for any $t \in [0,1]$.
Finally, assume x is a strict local minimizer and $x_1, x_2 > 0$. As $p_M(x) = p_M(t\tilde{x} + (1-t)\overline{x})$, where $t\tilde{x} + (1-t)\overline{x}$ tends to x when t tends to $x_1/(x_1 + x_2)$, we get a contradiction with the fact that x is a strict local minimizer. Thus $x_1 x_2 = 0$ holds. ☐

In what follows we consider a graph $G = (V = [n], E)$ and the corresponding Motzkin-Straus problem (M-S). We now show some structural results for the (strict) local and global minimizers of (M-S).

Lemma 4.2. *Let $x \in \Delta_n$ and assume its support $S = \text{Supp}(x)$ is a stable set of G. If x is a local minimizer of problem (M-S) then $x = \chi^S/|S|$ and S is a maximal stable set. In particular, if x is a global minimizer then $x = \chi^S/\alpha(G)$ where S has cardinality $\alpha(G)$.*

Proof. First we show that $x = \chi^S/|S|$. For this assume for contradiction that $x_i > x_j$ for some $i \neq j \in S$. Consider the vector $\tilde{x} = x + \epsilon(e_j - e_i) \in \Delta_n$ where $0 < \epsilon < x_i - x_j$. Then we have

$$f_G(\tilde{x}) = f_G(x) - 2\epsilon(x_i - x_j - \epsilon) < f_G(x),$$

which contradicts the fact that x is a local minimizer. This shows $x = \chi^S/|S|$. Then $f_G(x) = \frac{1}{|S|}$ and thus $|S| = \alpha(G)$ if x is a global minimizer, which shows the last claim of the lemma. Now we show that S is a maximal stable set. Assume for contradiction that there exists $j \notin S$ such that $S \cup \{j\}$ is stable and consider the vector $\tilde{x} = x + \epsilon(e_j - e_i) \in \Delta_n$ for some $i \in S$ and $0 < \epsilon < x_i$. Then we have

$$f_G(\tilde{x}) = f_G(x) - 2\epsilon(x_i - \epsilon) < f_G(x),$$

contradicting again that x is a local minimizer. Thus S is a maximal stable set. ☐

We will characterize the (strict) local minimizers whose support is a stable set in Proposition 4.5 and Proposition 4.7 below. First we characterize the global minimizers of problem (M-S).

Proposition 4.3. *Let $x \in \Delta_n$ with support $S = \text{Supp}(x)$ and let $C_1, \ldots, C_k$ denote the connected components of the graph $G[S]$. Then x is a global minimizer of problem (M-S) if and only if $k = \alpha(G)$, the sets $C_1, \ldots, C_k$ are cliques of G and $\sum_{i \in C_h} x_i = 1/k$ for all $h \in [k]$. In that case all the edges of $G[S]$ are critical. Therefore, if G has no critical edges then the global minimizers are the vectors of the form $x = \chi^S/\alpha(G)$, where S is a stable set of size $\alpha(G)$.*

Proof. We first show the 'only if' part. Let x be a global minimizer. Pick nodes $a_1 \in C_1, \ldots a_k \in C_k$ in the different connected components of $G[S]$. Then the set $I = \{a_1, \ldots, a_k\}$ is a stable set of G. Define the vector $y \in \Delta_n$, with entries $y_{a_h} = \sum_{i \in C_h} x_i$ for $h \in [k]$ and $y_i = 0$ for all remaining vertices $i \in V \setminus I$. By applying iteratively Lemma 4.1 (with the matrix $M = I + A_G$, using the edges in a spanning tree in each connected component C_h) we obtain that $f_G(y) = f_G(x)$. Therefore, y is a global minimizer whose support is a stable set and thus, by Lemma 4.2, we obtain $k = \alpha(G)$ and $\sum_{i \in C_h} x_i = y_h = \frac{1}{k}$ for all $h \in [k]$. It is clear that each component (say) C_1 is a clique. Indeed, if $i \neq j \in C_1$ are not adjacent then the set $\{i, j\} \cup \{a_2, \ldots, a_k\}$ would be a stable set of size $\alpha(G) + 1$. Moreover, the edge $\{i, j\}$ is critical since both sets $\{i, a_2, a_3, \ldots, a_k\}$ and $\{j, a_2, a_3, \ldots, a_k\}$ are stable sets of size $\alpha(G)$.

We now show 'if part'. For this assume $k = \alpha(G)$, each component C_h of $G[S]$ is a clique and $\sum_{i \in C_h} x_i = 1/k$ for $h \in [k]$. Then we have

$$f_G(x) = \sum_{h=1}^{k} \Big(\sum_{i \in C_h} x_i \Big)^2 = \frac{1}{k} = \frac{1}{\alpha(G)},$$

which shows x is a global minimizer.

The final claim about the global minimizers when G has no critical edges follows as a direct consequence of the above characterization. □

Corollary 4.4. *The following conditions are equivalent:*

(i) *The graph G has no critical edges.*

(ii) *The number of global minimizers of problem (M-S) is finite (and equal to the number of maximum stable sets).*

Proof. The implication (i) $\Longrightarrow$ (ii) follows directly from Proposition 4.3.
(ii) $\Longrightarrow$ (i): Assume for contradiction that $\{1, 2\}$ is a critical edge of G, we construct infinitely many global minimizers. Indeed, as $\{1, 2\}$ is a critical edge there exists a set $I \subseteq V$ such that $I \cup \{1\}$ and $I \cup \{2\}$ are stable sets of size $\alpha(G)$. Then, for any $t \in [0, 1]$, $f_G(x) = 1/\alpha(G)$ for the vector $x = (t\chi^{I \cup \{1\}} + (1-t)\chi^{I \cup \{2\}})/\alpha(G)$, which thus gives infinitely many global minimizers. □

In Lemma 4.2 we saw that if $\chi^S/|S|$ is a local minimizer then S is a maximal stable set, i.e., $\deg_S(j) \geq 1$ for all $j \in V \setminus S$. We now sharpen this result and show that the stronger condition '$\deg_S(j) \geq 2$ for all $j \in V \setminus S$' characterizes the *strict* local minimizers.

Proposition 4.5. *Let G be a graph and let $u \in \Delta_n$. The following assertions are equivalent.*

(i) *u is a strict local minimizer of problem (M-S).*

(ii) $\deg_S(j) \geq 2$ *for all $j \in V \setminus S$ and $u = \chi^S/|S|$, where we set $S = \text{Supp}(u)$.*

(iii) *The optimality conditions (FOOC), (SCC) and (SOSC) hold at u.*

Proof. We first prove (i) ⇒ (ii). Assume u is a strict local minimizer with support $S = \text{Supp}(u)$ and set $k = |S|$. By the second part of Lemma 4.1, we know that S is a stable set and, by Lemma 4.2, $u = \chi^S/|S|$ and S is a maximal stable set. Assume there exists $j \notin S$ with $\deg_S(j) = 1$, and let $i \in S$ be the (unique) neighbor of j in S. Then the set $\tilde{S} = (S \setminus \{i\}) \cup \{j\}$ is a stable set of size k. Consider the vectors $\tilde{u} = \chi^{\tilde{S}}/k$ and $z = tu + (1-t)\tilde{u}$ for $t \in (0, 1)$. Then $z_i = t/k$, $z_j = (1-t)/k$, $z_v = 1/k$ for all $v \in S \setminus \{i\}$, $z_v = 0$ otherwise, and we have

$$f_G(z) = \sum_{v \in S \cup \{j\}} z_v^2 + 2z_i z_j = \frac{t^2}{k^2} + \frac{(1-t)^2}{k^2} + \frac{k-1}{k^2} + 2\frac{t(1-t)}{k^2} = \frac{1}{k},$$

which contradicts that u is a strict local minimizer when considering t close to 1. Hence one must have $\deg_S(j) \geq 2$ for all $j \notin S$.

Now we prove (ii) $\Rightarrow$ (iii). Assume $u = \chi^S/k$, where S is a stable set with $k = |S|$ and $\deg_S(j) \geq 2$ for all $j \notin S$. Consider the polynomials $g_i(x) = x_i$ for $i \in [n]$, $h(x) = \sum_{i=1}^n x_i - 1$, so that the feasible region of problem (M-S) is defined by the constraints $g_i(x) \geq 0$ for $i \in [n]$, and $h(x) = 0$. The active constraints at u are the constraints $g_i(x) \geq 0$ for $i \notin S$, and $h(x) = 0$. Hence $J(u) = V \setminus S$. Clearly, (CQC) holds at u since the gradients of the active constraints at u are the vectors e and e_i for $i \in V \setminus S$, which are linearly independent (as $S \neq \emptyset$). Next note that

$$\frac{\partial f_G}{\partial x_i}(u) = \begin{cases} \frac{2}{k} & \text{if } i \in S, \\ \frac{2}{k}\deg_S(i) & \text{if } i \notin S. \end{cases} \tag{4.4}$$

The first order optimality condition reads

$$\nabla f_G(u) = \lambda e + \sum_{i \notin S} \mu_i e_i, \quad \text{ where } \lambda \in \mathbb{R},\ \mu_i \geq 0 \text{ for } i \in V \setminus S.$$

Looking at coordinate $i \in S$ we obtain $\lambda = 2/k$ and then, for each coordinate $i \notin S$, we obtain $\mu_i = 2(\deg_S(i) - 1)/k$. Since, by assumption, $\deg_S(i) \geq 2$ for $i \notin S$ it follows that $\mu_i > 0$ and thus (FOOC) and (SCC) hold at u. Finally, we check that also (SOSC) holds. For this consider the Lagrangian function

$$L(x) = f_G(x) - \lambda\Big(\sum_{i=1}^n x_i - 1\Big) - \sum_{i \in J(u)} \mu_i x_i = f_G(x) - \frac{2}{k}\Big(\sum_{i=1}^n x_i - 1\Big) - \sum_{i \notin S}(\deg_S(i) - 1)x_i.$$

As all the constraints are linear we have $\nabla^2 L(x) = \nabla^2 f_G(x) = 2(A_G + I)$. Consider a vector $0 \neq v \in G(u)^\perp$. Then $v_i = 0$ for $i \notin S$, therefore $v^T \nabla^2 L(u) v = 2\sum_{i \in S} v_i^2 > 0$ since $v \neq 0$. So (SOSC) holds at u.
Finally, the implication (iii) $\Rightarrow$ (i) follows from Theorem 2.2 (ii). $\square$

Example 4.6. *Given an integer $r \geq 3$ consider disjoint sets $V_2, \ldots, V_r$ with $|V_i| = i$ for $2 \leq i \leq r$. Let G be the complete $(r-1)$-partite graph with vertex set $V = V_2 \cup \ldots \cup V_r$. Then, by Proposition 4.5, each vector χ^{V_i}/i is a strict local minimizer of problem (M-S) for graph G, while χ^{V_r}/r is the only global minimizer.*

Now, we characterize the local minimizers of (M-S) whose support is a stable set of G.

Proposition 4.7. *Let S be a stable set of G. For $i \in S$ let $N_1(i) = \{j \in V \setminus S : N_S(i) = \{i\}\}$ be the set of vertices having only the vertex i as neighbour in S. Then $x = \chi^S/|S|$ is a local minimizer of (M-S) if and only if S is a maximal stable set and $N_1(i)$ is a clique for all $i \in S$.*

Proof. First, we show the "only if" part. Assume that $x = \chi^S/|S|$ is a local minimizer. Then, by Lemma 4.2, S is a maximal stable set. Assume that $N_1(i)$ is not a clique for some $i \in S$. Then there exist $j, k \in N_1(i)$ such that $\{j, k\} \notin E$. Consider the vector $\tilde{x} = x - \epsilon x_i + \frac{\epsilon}{2}x_j + \frac{\epsilon}{2}x_k$ for ϵ close to 0. Then we have

$$f_G(\tilde{x}) = \sum_{v \in S \setminus \{i\}} x_v^2 + (x_i - \epsilon)^2 + \frac{\epsilon^2}{2} + 2(x_i - \epsilon)\epsilon = f_G(x) - \frac{\epsilon^2}{2} < f_G(x),$$

contradicting that x is a local minimizer.

We now show the "if part". Define the sets $T = \{i \in V : \deg_S(i) = 1\}$ and $R = \{i \in V : \deg_S(i) \geq 2\}$, which partition the set $V \setminus S$. Let $\tilde{x} = x + \epsilon \in \Delta_n$ with $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n) \in \mathbb{R}^n$, $\epsilon_i \geq 0$ for $i \in T \cup R$ and $\sum_{i \in V} \epsilon_i = 0$. Then

$$f_G(\tilde{x}) = f_G(x) + \epsilon^T (A_G + I)\epsilon + 2x^T (A_G + I)\epsilon.$$

We claim that $f_G(x) \leq f_G(\tilde{x})$, i.e.,

$$\epsilon^T (A_G + I)\epsilon + 2x^T (A_G + I)\epsilon \geq 0 \quad \text{ whenever } \quad ||\epsilon||_\infty \leq 1/|S|^2. \tag{4.5}$$

We use the following notation: for a subset $I \subseteq V$, $\epsilon_I \in \mathbb{R}^n$ has entries $(\epsilon_I)_i = \epsilon_i$ for $i \in I$ and $(\epsilon_I)_i = 0$ for $i \in V \setminus I$. We bound the first term in (4.5):

$$\epsilon^T (A_G + I)\epsilon \geq \epsilon_{S \cup T}^T (A_G + I)\epsilon_{S \cup T} + 2\epsilon_R^T (A_G + I)\epsilon_S + \epsilon_R^T (A_G + I)\epsilon_R.$$

By assumption $S \cup T$ is a disjoint union of cliques, hence the submatrix of $A_G + I$ indexed by $S \cup T$ is positive semidefinite and thus $\epsilon_{S \cup T}^T (A_G + I)\epsilon_{S \cup T} \geq 0$. Also, we have $\epsilon_R^T (A_G + I)\epsilon_R \geq 0$ because $\epsilon_j \geq 0$ for $j \in R$. Hence we obtain

$$\epsilon^T (A_G + I)\epsilon \geq 2\epsilon_R^T (I + A_G)\epsilon_S = 2\sum_{j \in R} \epsilon_j \sum_{i \in N_S(j)} \epsilon_i \geq -\frac{2}{|S|}\sum_{j \in R} \epsilon_j,$$

where we use the fact that $\epsilon_i \geq -1/|S|^2$ for $i \in S$ and $\epsilon_j \geq 0$ for $j \in R$. Now, we bound the second term in (4.5):

$$\begin{aligned} 2x^T(A_G+I)\epsilon &= \tfrac{2}{|S|}\Big(\textstyle\sum_{i\in S}\epsilon_i + \sum_{j\in V}\deg_S(j)\epsilon_j\Big) \\ &\geq \tfrac{2}{|S|}\Big(\textstyle\sum_{i\in S}\epsilon_i + \sum_{j\in T}\epsilon_j + 2\sum_{j\in R}\epsilon_j\Big) \\ &= \tfrac{2}{|S|}\textstyle\sum_{j\in R}\epsilon_j, \end{aligned}$$

where we have used the fact that $\deg_S(j) \geq 2$ for $j \in R$, $\deg_S(j) = 1$ for $j \in T$, and $\sum_{i\in V}\epsilon_i = 0$. Combining these two inequalities, we obtain $f_G(x) \leq f_G(\tilde{x})$ as desired. □

Example 4.8. *As an illustration consider the graph G on $V = [5]$ with edges $\{1,3\}, \{2,4\}$ and $\{2,5\}$, so $\alpha(G) = 3$. Consider the vector $x = \chi^S/2$, where $S = \{1,2\}$ is a maximal stable set, but $N_1(2) = \{4,5\}$ is not a clique. Then, as shown in the above lemma, x is not a local minimum. Indeed, setting $z = \chi^{\{3,4,5\}}/3$, we have $f_G((1-t)x + tz) - f_G(x) = -t^2/6 \leq 0$ for all $t \in [0,1]$. On the other hand, $f_G(1/2+\epsilon, 1/2-\epsilon, 0, 0, 0) = 1/2 + 2\epsilon^2 > f_G(x)$ for all small $\epsilon > 0$. Hence, x is not a local minimizer, nor a local maximizer.*

5 Finite convergence and perturbed hierarchies

In this section we give a partial positive answer to Conjecture 2 and show that the de Klerk-Pasechnik hierarchy $\vartheta^{(r)}(\cdot)$ has finite convergence for the class of acritical graphs. Our approach relies on proving finite convergence of the (weaker) Lasserre hierarchy applied to problem (M-S) for acritical graphs. In addition, we propose a perturbed formulation for the stability number for which the corresponding hierarchy has finite convergence for all graphs.

5.1 Finite convergence of Lasserre hierarchy for the Motzkin-Straus formulation

In Section 4 we proved that the set of global minimizers of problem (M-S) is finite if and only if the graph G is acritical or, equivalently, every global minimizer is a strict minimizer. In addition, as a direct application of Proposition 4.5, we have shown that this holds if and only if the classical optimality conditions hold at all the global minimizers. Hence, in summary, if G has a critical edge then we cannot apply Theorem 2.3 since problem (M-S) has infinitely many minimizers. On the other hand, if G is acritical then we may directly apply Theorem 2.3 and conclude that the Lasserre hierarchy $f_G^{(r)}$ corresponding to problem (M-S) has finite convergence. So the following holds.

Theorem 5.1. *Let G be a graph with no critical edges. Then, $f_G^{(r)} = 1/\alpha(G)$ for some $r \in \mathbb{N}$.*

Corollary 5.2. *Let G be a graph with no critical edges. Then, $\vartheta^{(r)}(G) = \alpha(G)$ for some $r \in \mathbb{N}$.*

Proof. This follows directly from Theorem 5.1 and Corollary 3.12. □

Hence, for problem (M-S), having finitely many global minimizers implies finite convergence of the Lasserre hierarchy. We now recall a known example which shows that this does not hold for general polynomial optimization problems.

Example 5.3. *(See, e.g., [21, Example 6.19]). Consider the problem of minimizing a polynomial p over the unit ball in $\mathbb{R}^n$. Assume p is homogeneous, $p(x) > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$, and p is not a sum of squares of polynomials. Then the minimum of p over the unit ball is $p_{\min} = 0$ and the origin is the unique global minimizer. However it is known that the corresponding Lasserre hierarchy does not have finite convergence, see Example 6.19 in [21] for details. The main reason is that a decomposition of the form $p = s_0 + s_1(1 - \sum_i x_i^2)$ with $s_0, s_1 \in \Sigma$ would imply $p \in \Sigma$. For the polynomial p one may, for instance, consider a perturbation of the Motzkin form: $p_\epsilon = x_1^4x_2^2 + x_1^2x_2^4 - 3x_1^2x_2^2x_3^2 + x_3^6 + \epsilon(x_1^6 + x_2^6 + x_3^6)$, selecting $\epsilon > 0$ such that $p_\epsilon \notin \Sigma$.*

5.2 Perturbed Motzkin-Straus formulation and hierarchies

The above finite convergence result relies on Theorem 2.3 that can only be applied to problems with finitely many optimal solutions, which holds for problem (M-S) only for acritical graphs. We now propose an alternative formulation for $\alpha(G)$, which is a perturbation of problem (M-S) designed in such a way that the number of global minimizers becomes finite, thus allowing us to prove finite convergence of the corresponding (perturbed) hierarchies for any graph G.

Given a scalar $\epsilon > 0$, consider the following perturbation of problem (M-S):

$$\min\{x^T((1+\epsilon)A_G + I)x : x \in \Delta_n\}. \qquad \text{(M-S-eps)}$$

So, for the perturbed matrix $M = I + (1+\epsilon)A_G$, we have $M_{ij} = M_{ji} > M_{ii} = M_{jj} = 1$ for any edge $\{i,j\}$ of G. First, we show a useful property for the global minimizers of any standard quadratic program (4.1) whose matrix M has this property.

Lemma 5.4. *Given a scalar $t > 1$ consider the standard quadratic program (4.1), where the matrix M is of the form*

$$M = \begin{pmatrix} 1 & t & a_1^T \\ t & 1 & a_2^T \\ a_1 & a_2 & M_0 \end{pmatrix}, \tag{5.1}$$

where $a_1, a_2 \in \mathbb{R}^{n-2}$ and $M_0 \in \mathcal{S}^{n-2}$. Then $u_1 u_2 = 0$ holds for every global minimizer u of (4.1).

Proof. Set $z = (u_3, u_4, \ldots, u_n)$ and assume for contradiction that $u_1, u_2 > 0$. Consider the feasible points $\tilde{u} = (u_1 + u_2, 0, z)$ and $\overline{u} = (0, u_1 + u_2, z)$. Then, for the polynomial $P_M(x) = x^T M x$, we have

$$\begin{aligned} p_M(\tilde{u}) &= (u_1+u_2)^2 + 2(u_1+u_2)z^T a_1 + z^T M_0 z, \\ p_M(\overline{u}) &= (u_1+u_2)^2 + 2(u_1+u_2)z^T a_2 + z^T M_0 z, \\ p_M(x) &= u_1^2 + u_2^2 + 2tu_1u_2 + 2u_1 z^T a_1 + 2u_2 z^T a_2 + z^T M_0 z. \end{aligned}$$

By assumption, u is a global minimizer of p_M over Δ_n, thus $p_M(\tilde{u}), p_M(\overline{u}) \geq p_M(u)$. This implies, respectively, $2u_2(z^T a_1 - z^T a_2), 2u_1(z^T a_2 - z^T a_1) \geq 2u_1u_2(t-1) > 0$ since $u_1, u_2 > 0$ and $t > 1$, and thus $z^T a_1 - z^T a_2 > 0$ and $z^T a_2 - z^T a_1 > 0$, a contradiction. ☐

Lemma 5.5. *For any graph G the optimal value of problem (M-S-eps) is $1/\alpha(G)$ and the global minimizers are the vectors of the form $u = \chi^S/\alpha(G)$, where S is a stable set of G with size $\alpha(G)$.*

Proof. If S is a stable set of size $\alpha(G)$, then $u^T((1+\epsilon)A_G + I)u = 1/\alpha(G)$ for $u = \chi^S/\alpha(G) \in \Delta_n$, which shows the optimal value of (M-S-eps) is at most $1/\alpha(G)$. On the other hand, for any $x \in \Delta_n$, $x^T((1+\epsilon)A_G + I)x \geq x^T(A_G + I)x \geq 1/\alpha(G)$, and thus the optimal value of (M-S-eps) is equal to $1/\alpha(G)$ and every global minimizer u of (M-S-eps) is also a global minimizer of (M-S). By Lemma 5.4 the support of any global minimizer u of (M-S-eps) must be a stable set S and, as u is also a global minimzer of (M-S), it follows from Lemma 4.2 that S has size $\alpha(G)$ and $u = \chi^S/\alpha(G)$. ☐

Remark 5.6. *A first observation is that Lemma 5.4 still holds if we use different perturbations ϵ for the edges, since the only property of the matrix $M = I + (1+\epsilon)A_G$ appearing in (M-S-eps) that we used is the fact that $M_{ij} = M_{ji} > M_{ii} = M_{jj} = 1$ for the edges $\{i,j\}$ of G.*

A second observation is that Lemma 5.4 also holds if we only perturb the entries corresponding to the critical *edges of G. Indeed, let $G_c = (V, E_c)$, where E_c denotes the set of critical edges of G, and consider the variation of (M-S-eps) where we use the matrix $M = I + A_G + \epsilon A_{G_c}$ (instead of $I + A_G + \epsilon A_G$). Then the optimum value is still equal to $1/\alpha(G)$. Indeed, if u is a global minimizer with support S then, by Lemma 5.4, the only edges that can be contained in S are the non-critical of G. On the other hand, as u is also a global minimizer of (M-S), by Corollary 4.4, any edge contained in S must be a critical edge. It therefore follows that S must be a stable set and $|S| = \alpha(G)$.*

Again, we can reformulate (M-S-eps) as a polynomial optimization problem over the sphere:

$$\frac{1}{\alpha(G)} = \min\{(x^{\circ 2})^T((1+\epsilon)A_G + I)x^{\circ 2} : x_1^2 + x_2^2 + \cdots + x_n^2 = 1\}. \tag{M-S-Sphere-eps}$$

For convenience define the polynomials

$$f_{G,\epsilon}(x) = x^T((1+\epsilon)A_G + I)x = f_G(x) + \epsilon x^T A_G x \quad \text{and} \quad F_{G,\epsilon}(x) = f_{G,\epsilon}(x^{\circ 2}).$$

We can also define the Lasserre hierarchies for the stability number based on the formulations (M-S-eps) and (M-S-Sphere-eps):

$$f_{G,\epsilon}^{(r)} = \max\Big\{\lambda : x^T((1+\epsilon)A_G + I)x - \lambda \in \mathcal{M}\Big(x_1, x_2, \ldots, x_n\Big)_r + \Big\langle \sum_{i=1}^n x_i - 1 \Big\rangle_{2r}\Big\}, \tag{5.2}$$

$$f_{G,po,\epsilon}^{(r)} = \max\Big\{\lambda : x^T((1+\epsilon)A_G + I)x - \lambda \in \mathcal{T}\Big(x_1, x_2, \ldots, x_n\Big)_r + \Big\langle \sum_{i=1}^n x_i - 1 \Big\rangle_{2r}\Big\}, \tag{5.3}$$

$$F_{G,\epsilon}^{(r)} = \max\Big\{\lambda : x^{\circ 2^T}((1+\epsilon)A_G + I)x^{\circ 2} - \lambda \in \Sigma_r + \Big\langle \sum_{i=1}^n x_i^2 - 1 \Big\rangle_{2r}\Big\}, \tag{5.4}$$

as analogues of (1.10), (1.11) and (1.12). We also have the corresponding copositive programming formulation:

$$\alpha(G) = \min\{t : t((1+\epsilon)A_G + I) - J \in \mathrm{COP}_n\} \tag{5.5}$$

for the stability number and the associated ϵ-perturbed linear and semidefinite hierarchies:

$$\zeta_\epsilon^{(r)}(G) = \min\{t : t((1+\epsilon)A_G + I) - J \in \mathcal{C}_n^{(r)}\}, \tag{5.6}$$

$$\vartheta_\epsilon^{(r)}(G) = \min\{t : t((1+\epsilon)A_G + I) - J \in \mathcal{K}_n^{(r)}\}, \tag{5.7}$$

in analogy to (1.4) and (1.5).

From the discussion in Section 3.1, the parameters $f_{G,po,\epsilon}^{(r)}$ $(r \geq 1)$, $f_{G,\epsilon}^{(r)}$, $F_{G,po,\epsilon}^{(r)}$ $(r \geq 2)$ and $\vartheta_\epsilon(G)$ $(r \geq 0)$ are finite. In addition, as a direct application of Lemma 3.2, if G is not the empty graph then the program (5.2) is infeasible at order $r = 1$, i.e., $f_{G,\epsilon}^{(1)} = -\infty$ for any $\epsilon > 0$.

As an application of Corollary 3.9, we have the following analogue of Corollary 3.12 linking the above hierarchies.

Lemma 5.7. *Let G be a graph and $\epsilon > 0$. Then, for any $r \geq 0$, we have*

$$\frac{1}{\alpha(G)} \geq \frac{1}{\vartheta_\epsilon^{(2r)}(G)} = F_{G,\epsilon}^{(2r+2)} = f_{G,po,\epsilon}^{(r+1)} \geq f_{G,\epsilon}^{(r+1)}.$$

We now show finite convergence of these hierarchies for any graph G.

Theorem 5.8. *Let G be a graph and $\epsilon > 0$. Then, we have $f_{G,\epsilon}^{(r)} = \frac{1}{\alpha(G)}$ for some $r \in \mathbb{N}$.*

Proof. We make again use of Theorem 2.3. Let u be a global minimizer of (M-S-eps), we show that the conditions (FOOC), (SCC), (SOSC) hold at u. By Lemma 5.5, $u = \chi^S/\alpha(G)$, where S is a stable set of size $\alpha(G)$. As the constraints of (M-S-eps) are the same as those of (M-S) we can follow the proof of the 'if part' of Proposition 4.5, where the gradient of the objective function now reads

$$\frac{\partial f_{G,\epsilon}}{\partial x_i}(u) = \begin{cases} \frac{2}{\alpha(G)} & \text{if } i \in S, \\ \frac{2(1+\epsilon)}{\alpha(G)} \deg_S(i) & \text{if } i \notin S. \end{cases}$$

Writing $\nabla f_{G,\epsilon}(u) = \lambda \nabla h(u) + \sum_{i \in V\setminus S} \mu_i \nabla g_i(u)$, where $h(x) = \sum_i x_i - 1$ and $g_i(x) = x_i$, we obtain that $\lambda = \frac{2}{\alpha(G)}$ and $\mu_i = \frac{2}{\alpha}((1+\epsilon)\deg_S(i) - 1) \geq \frac{2}{\alpha(G)}\epsilon > 0$ for $i \notin S$, since $\deg_S(i) \geq 1$. Hence, strict complementarity (SCC) holds. Finally, we check (SOSC). For the Lagrangian function $L(x) = f_{G,\epsilon}(x) - \lambda h(x) - \sum_{i \notin S} \mu_i g_i(x)$, we have $\nabla^2 L(x) = \nabla^2 f_{G,\epsilon}(x) = 2(A_G + I)$. Now, take $v \in G(u)^\perp \setminus \{0\}$, so $v_i = 0$ for $i \in S$ and thus $v_i \neq 0$ for some $i \in V \setminus S$. Then $v^T(A_G + I)v = \sum_{i \notin S} v_i^2 > 0$, which shows (SOSC) holds and thus concludes the proof. ☐

Corollary 5.9. *Let G be a graph and $\epsilon > 0$. Then there exists $r \geq 0$ such that $\vartheta_\epsilon^r(G) = \alpha(G)$.*

We conclude this section with some observations on the role of the perturbation parameter ϵ in the different hierarchies. Clearly we have $\zeta_\epsilon^{(r)}(G) \leq \zeta^{(r)}(G)$ and $\vartheta_\epsilon^{(r)}(G) \leq \vartheta^{(r)}(G)$ for any $r \geq 0$ and any $\epsilon > 0$. We first show that the perturbation parameter ϵ in fact plays no role for the linear hierarchy $\zeta_\epsilon^{(r)}(G)$.

Theorem 5.10. *For all $r \in \mathbb{N}$ and $\epsilon > 0$, we have $\zeta_\epsilon^{(r)}(G) = \zeta^{(r)}(G)$.*

Proof. The inequality $\zeta_\epsilon^{(r)}(G) \leq \zeta^{(r)}(G)$ is clear, so we show the reverse inequality. For this assume the matrix $M_{t,\epsilon} := t(I + (1+\epsilon)A_G) - J$ belongs to the cone $\mathcal{C}_n^{(r)}$, we show that also the matrix $M_t := t(I + A_G) - J$ belongs to $\mathcal{C}_n^{(r)}$, which implies $\zeta^{(r)}(G) \leq \zeta_\epsilon^{(r)}(G)$, as desired. By assumption, $M_{t,\epsilon} \in \mathcal{C}_n^{(r)}$ means that the polynomial $(\sum_i x_i)^r p_{M_{t,\epsilon}}(x)$ has nonnegative coefficients. (Recall the notation from (1.2)). Following [5], for any matrix M and $r \in \mathbb{N}$ we have

$$\Big(\sum_i x_i\Big)^r p_M(x) = \Big(\sum_i x_i\Big)^r x^T M x = \sum_{\beta \in I(n,r+2)} \frac{r!}{\beta!} c_\beta x^{2\beta},$$

where $c_\beta := \beta^T M \beta - \beta^T \mathrm{diag}(M)$ and $\mathrm{diag}(M) \in \mathbb{R}^n$ is the vector with entries M_{ii} for $i \in [n]$. Therefore, for the matrix $M = M_{t,\epsilon}$, the polynomial $(\sum_i x_i)^r p_M(x)$ has nonnegative coefficients if and only if $c_\beta \geq 0$ for all

$\beta \in I(n, r+2)$. We will now prove that the propery of having $c_\beta \geq 0$ for all $\beta \in I(n, r+2)$ is independent on ϵ. For this let $\beta \in I(n, r+2)$. We have

$$\begin{aligned} c_\beta &= \beta^T M \beta - \beta^T \mathrm{diag}(M) \\ &= t \sum_{i=1}^n \beta_i^2 + t(1+\epsilon)\beta^T A_G \beta - \left(\sum_{i=1}^n \beta_i \right)^2 - (t-1) \sum_{i=1}^n \beta_i \\ &= t \left(\sum_{i=1}^n \beta_i^2 + (1+\epsilon)\beta^T A_G \beta - (r+2) \right) - (r+2)(r+1). \end{aligned}$$

Therefore, $c_\beta \geq 0$ for all $\beta \in I(n, r+2)$ if and only if $t \geq \frac{(r+1)(r+2)}{\beta^* - (r+2)}$, where we set

$$\beta^* = \min\{ f_{G,\epsilon}(\beta) = \beta^T (I + (1+\epsilon)A_G)\beta : \beta \in I(n, r+2)\}. \tag{5.8}$$

We now observe that the optimum value of the program (5.8) is attained at some $\beta \in I(n, r+2)$ whose support is stable. For this let β be a minimizer of (5.8) and assume $\beta_1, \beta_2 > 0$ where $\{1, 2\}$ is an edge of G; we use the usual argument of shifting weights to create another minimizer whose support does not contain the edge $\{1, 2\}$. For this note that the matrix $I + (1+\epsilon)A_G$ has the form (5.1) with $t = 1 + \epsilon$. Set $z = (\beta_3, \ldots, \beta_n)$, say $z^T a_1 \leq z^T a_2$ and consider the new vector $\tilde{\beta} = (\beta_1 + \beta_2, 0, z) \in I(n, r+2)$, so that $f_{G,\epsilon}(\tilde{\beta}) \geq f_{G,\epsilon}(\beta)$. On the other hand, we have

$$f_{G,\epsilon}(\tilde{\beta}) - f_{G,\epsilon}(\beta) = -2\epsilon \beta_1 \beta_2 - 2\beta_2 (z^T a_2 - z^T a_1) \leq 0,$$

which implies $f_{G,\epsilon}(\tilde{\beta}) = f_{G,\epsilon}(\beta)$, as desired.

So we have shown that the optimum value of (5.8) does in fact not involve the parameter ϵ. Therefore, if the polynomial $(\sum_i x_i)^r p_{M_{t,\epsilon}}(x)$ has nonnegative coefficients then also the polynomial $(\sum_i x_i)^r p_{M_t}(x)$ has nonnegative coefficients. This shows that $\zeta^{(r)}(G) \leq \zeta_\epsilon^{(r)}(G)$. $\square$

For the semidefinite hierarchy $\vartheta_\epsilon^{(r)}(G)$ we can only prove that the first level of the hierarchy is independent on ϵ.

Lemma 5.11. *For any $\epsilon > 0$ we have $\vartheta_\epsilon^{(0)}(G) = \vartheta^{(0)}(G)$.*

Proof. The inequality $\vartheta_\epsilon^{(0)}(G) \leq \vartheta^{(0)}(G)$ is clear, so we show the reverse inequality. For this let t be feasible for $\vartheta_\epsilon^{(0)}(G)$, we show that t is also feasible for $\vartheta^{(r)}(G)$. By assumption, the matrix $t((1+\epsilon)A_G + I) - J$ belongs to $\mathcal{K}^{(0)}$, i.e., there exists a matrix $P \succeq 0$ such that $P_{ii} = t - 1$ for any $i \in [n]$ and $P \leq t((1+\epsilon)A_G + I) - J$ (entry-wise) (recall the characterization of $\mathcal{K}^{(0)}$ in Proposition 3.7). We now claim that $P \leq t(I + A_G) - J$, which shows that t is feasible for $\vartheta^{(0)}(G)$, as desired. Indeed, if $\{i, j\}$ is an edge, then the inequality $P_{ij} \leq t - 1$ follows using the fact that $P \succeq 0$ and $P_{ii} = P_{jj} = t - 1$ and, if $\{i, j\}$ is not an edge, then $P_{ij} \leq -1$ follows from the assumption $P \leq t((1+\epsilon)A_G + I) - J$. $\square$

Question 5.12. *Is it true that, for any $\epsilon > 0$ and any $r \in \mathbb{N}$, $\vartheta_\epsilon^{(r)}(G) = \vartheta^{(r)}(G)$?*

Clearly, a positive answer to this question would imply the finite convergence of the hierarchy $\vartheta^{(r)}(G)$ and thus settle Conjecture 2.

Observe that the parameters $\vartheta_\epsilon^{(r)}$ and $F_{G,\epsilon}^{(r)}$ are monotone in ϵ:

$$0 < \epsilon_1 < \epsilon_2 \Longrightarrow \alpha(G) \leq \vartheta_{\epsilon_2}^{(r)}(G) \leq \vartheta_{\epsilon_1}^{(r)}(G) \quad \text{and} \quad F_{G,\epsilon_1}^{(r)} \leq F_{G,\epsilon_2}^{(r)} \leq \frac{1}{\alpha(G)},$$

which follows using the fact $(\epsilon_2 - \epsilon_1)A_G$ is entry-wise nonnegative. So, if we increase ϵ, we can only get improved bounds for $\alpha(G)$. On the other hand, the behaviour of the parameters $f_{G,\epsilon}^{(r)}$ is not clear as ϵ changes. In fact, the perturbed bound can be worse than the original one. For instance, $f_{G,\epsilon}^{(1)} = -\infty$ for every $\epsilon > 0$ when G is not the empty graph, while $f_G^{(1)} = 1/\alpha(G)$ when G is a disjoint union of cliques.

6 Complexity of deciding finiteness of the global minimizers

As we saw earlier, having finitely many minimizers is a property which plays an important role in the study of finite convergence of Lasserre hierarchy for polynomial optimization. This raises the question of understanding the complexity of deciding whether a polynomial optimization problem has finitely many minimizers. Here, as a byproduct of our results in the previous sections about global minimizers of standard quadratic programs, we show that unless P=NP there is no polynomial-time algorithm to decide whether a standard quadratic program has finitely many global minimizers. The complexity of several other decision problems about minimzers in polynomial optimization has been studied rencently in [1, 2]. In particular, Ahmadi and Zhang [2] show that it is strongly NP-hard to decide whether a polynomial of degree 4 has a local minimizer over $\mathbb{R}^n$; they also show that the same holds for deciding if a quadratic polynomial has a local minimizer (or a strict local minimizer) over a polyhedron. In addition they show that unless P=NP there cannot be a polynomial-time algorithm that finds a point within Euclidean distance c^n (for any constant $c \geq 0$) of a local minimizer of an n-variate quadratic polynomial over a polytope.

In this section we consider the following problem:

FINITE-MIN: Given an instance of problem (P), does it have finitely many global minimizers?

Consider first the case when (P) is a linear optimization problem, i.e., when the objective and the constraints are linear polynomials. Then the problem is convex and thus, if x_1 and x_2 are two distinct global minimizers then, for every $0 \leq t \leq 1$, the point $z = tx_1 + (1-t)x_2$ is also a global minimizer. Hence the problem has finitely many minimizers if and only if it has a unique one. Therefore, the problem of deciding whether a linear program has finitely many global minimizers is equivalent to the problem of deciding whether it has a unique optimal solution and a polynomial-time algorithm for this problem was given by Appa [3].

In the rest of the section we prove that problem FINITE-MIN is hard even when restricting to standard quadratic programs. For this, we first consider the following combinatorial problems, which we will use to prove this hardness result. Recall that given a graph $G = (V, E)$, an edge $e \in E$ is critical if $\alpha(G \setminus e) = \alpha(G) + 1$.

CRITICAL-EDGE: Given a graph $G = (V, E)$ and an edge $e \in E$, is e a critical edge of G?

STABLE-SET: Given a graph G and $k \in \mathbb{N}$, does $\alpha(G) \geq k$ hold?

The problem STABLE-SET is well-known to be NP-Complete [14]. From this we now prove that unless P=NP there is no polynomial-time algorithm to decide whether an edge is critical.

Theorem 6.1. *If there is a polynomial-time algorithm that solves the problem* CRITICAL-EDGE*, then P=NP.*

Proof. Assume that there exists a polynomial-time algorithm for CRITICAL-EDGE; we show how to use it to solve STABLE-SET. For this let $G = ([n], E)$ be an instance of STABLE-SET and order its edges as $e_1, e_2, \ldots, e_m$. Then, for each $i = 1, 2, \ldots, m$, we check whether the edge e_i is critical in the graph $G_{i-1} := G \setminus \{e_1, \ldots, e_{i-1}\}$. If the answer is yes then we have $\alpha(G_i) = \alpha(G_{i-1}) + 1$ and, otherwise, $\alpha(G_i) = \alpha(G_{i-1})$. After checking all the m edges we end up with the empty graph G_m on n nodes, with $\alpha(G_m) = n$. Let p be the number of critical edges that have been encountered while checking all the m edges. Then we have $n = \alpha(G_m) = p + \alpha(G)$ and thus $\alpha(G) = n - p$ has been computed. Hence a polynomial-time algorithm for CRITICAL-EDGE implies a polynomial-time algorithm for computing $\alpha(G)$. ☐

Using this reduction we now prove that the problem of deciding whether a standard quadratic optimization problem has finitely many optimal solutions is hard. For this, given a graph $G = ([n], E)$ and a fixed edge $e \in E$, consider the following standard quadratic program:

$$\min\ x^T(I + A_G + A_{G\setminus e})x \ \text{ subject to }\ x \geq 0,\ \sum_{i=1}^{n} x_i = 1, \tag{6.1}$$

where in the matrix defining the objective function, all edges of G get weight 2, except the selected edge e which keeps weight 1. First observe that the optimum value of (6.1) is equal to $1/\alpha(G)$; the argument is analogous to the one used for the corresponding claim in Lemma 5.5 and thus omitted.

Lemma 6.2. *The optimal value of problem (6.1) is equal to* $1/\alpha(G)$.

Theorem 6.3. *Given a graph $G = (V = [n], E)$ and an edge $e \in E$, problem (6.1) has infinitely many global minimizers if and only if e is a critical edge of G.*

Proof. Say e is the edge $\{1, 2\}$. First assume e is a critical edge, we show that (6.1) has infinitely many optimal solutions. Since e is critical, there exists $I \subseteq V$ such that both sets $I \cup \{1\}$ and $I \cup \{2\}$ are stable sets of size $\alpha(G)$. Then both vectors $\tilde{x} = \chi^{I\cup\{1\}}/\alpha(G)$ and $\bar{x} = \chi^{I\cup\{2\}}/\alpha(G)$ are optimal solutions of (6.1). Now we prove that, for every $0 < t < 1$, $x = t\tilde{x} + (1-t)\bar{x}$ is also an optimal solution. Indeed, $x_i = 1/\alpha(G)$ for $i \in I$, $x_1 = t$, $x_2 = 1-t$ and $x_j = 0$ otherwise, and the objective value of x is equal to

$$\frac{\alpha(G)-1}{\alpha(G)^2} + t^2 + (1-t)^2 + 2t(1-t) = \frac{1}{\alpha(G)}.$$

Hence problem (6.1) has infinitely many solutions if e is critical.

Conversely, assume that (6.1) has infinitely many global minimizers, we show that e is a critical edge. Let u be a global minimizer of (6.1) and $S = \mathrm{Supp}(u)$, then u is also a global minimizer of the original problem (M-S). If S is a stable set then, by Lemma 4.2, S has size $\alpha(G)$ and $u = \chi^S/\alpha(G)$ (since u is a global minimizer of (M-S)). On the other hand, if S is not stable then, in view of Lemma 5.4, we know that the only edge that can be contained in S is the edge e. As we assume that (6.1) has infinitely many global minimizers, at least one of them (say u) has its support S which contains the edge e. From this, we will now show that the edge e is critical. Note that the matrix $I + A_G + A_{G\setminus\{e\}}$ is of the form (4.2). Hence, by Lemma 4.1, we know that both points $\tilde{u} = u + u_2e_1 - u_2e_2$ and $\overline{u} = u - u_1e_1 + u_1e_2$ are optimal solutions of (6.1). Moreover, $\mathrm{Supp}(\tilde{u}) = S \setminus \{2\}$ and $\mathrm{Supp}(\overline{u}) = S \setminus \{1\}$ are stable sets, since $\{1, 2\}$ is the only edge contained in S. Therefore, as we just argued above, $|S \setminus \{1\}| = \alpha(G)$, which shows that the edge e is critical. ☐

Corollary 6.4. *If there is a polynomial-time algorithm to decide whether a standard quadratic program has finitely many global minimizers then P=NP.*

7 Concluding remarks

We have shown finite convergence of the de Klerk-Pasechnik hierarchy $\vartheta^{(r)}(G)$ for the class of acritical graphs by relating it to the sum-of-squares hierarchy (1.10) for the Motzkin-Straus formulation of $\alpha(G)$. Proving finite convergence for all graphs remains wide open. In fact, as we have observed, it would be sufficient to show finite convergence for the class of critical graphs. The hierarchy (1.10) however is weaker than the sum-of-squares hierarchy (1.11) based on using the preordering (generated by the polynomials defining the simplex Δ_n), which we have shown to be equivalent to the hierarchy $\vartheta^{(r)}(G)$. A possible approach to solve Conjecture 2 could therefore be to fully exploit this additionnal real algebraic structure. Another approach could be to use the perturbed sum-of-squares hierarchies that we have introduced and for which we could show finite convergence; such a strategy would require to be able to show degree bounds on the level of finite convergence that do not depend on the perturbation parameter.

Showing the stronger Conjecture 1, which asks whether ϑ-rank$(G) \leq \alpha(G) - 1$, seems even more challenging. The resolution in [12] for graphs with small stability number $\alpha(G) \leq 8$ required technically involved arguments. It is likely that the full resolution will need a new set of dedicated tools. As pointed out in [12], one of the difficulties lies in understanding the behaviour of the ϑ-rank under the operation of adding isolated nodes. We will further investigate this question in follow-up work [22].

While we could characterize the graphs for which the first level of the sum-of-squares hierarchy (1.10) (at order $r = 1$) is exact, the analogous question for the first level of the pre-ordering based hierarchy (1.11) is much more difficult. This is equivalent to understanding which graphs have ϑ-rank 0, a question which we will investigate in [22] and where critical edges also play a crucial role.

References

[1] A.A. Ahmadi and J. Zhang. Complexity aspects of local minima and related notions. *arXiv:2008.06148*, 2020.

[2] A.A. Ahmadi and J. Zhang. On the complexity of finding a local minimizer of a quadratic function over a polytope. *arXiv:2008.05558*, 2020.

[3] G. Appa. On the uniqueness of solutions to linear programs. *The Journal of the Operational Research Society*, 53(10):1127–1132, 2002.

[4] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[5] I. Bomze and E. de Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24:163–185, 12 2002.

[6] I.M. Bomze, M. Dür, E. de Klerk, C. Roos, and A.J. Quist. On copositive programming and standard quadratic optimiztion problems. *Journal of Global Optimization*, 18:301–320, 2000.

[7] S. Burer and Y. Ye. Exact semidefinite formulations for a class of (random and non-random) nonconvex quadratic programs. *Mathematical Programming*, 181:1–17, 2020.

[8] E. de Klerk and M. Laurent. On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems. *SIAM Journal on Optimization*, 21:824–832, 2011.

[9] E. de Klerk, M. Laurent, and P. Parrilo. *On the equivalence of algebraic approaches to the minimization of forms on the simplex*. Number 312 in LNCIS. Springer Verlag, Germany, 2005.

[10] E. de Klerk and D. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12:875–892, 04 2002.

[11] C.N. Delzell. A Constructive, Continuous Solution to Hilbert's 17th Problem, and Other Results in Semi-algebraic Geometry. PhD thesis, Stanford University, 1980.

[12] N. Gvozdenović and M. Laurent. Semidefinite bounds for the stability number of a graph via sums of squares of polynomials. *Mathematical Programming*, 110:145–173, 2007.

[13] A. Harrow, A. Natarajan, and X. Wu. An improved semidefinite programming hierarchy for testing entanglement. *Commun. Math. Physics*, 352:881–904, 2017.

[14] R. Karp. *Reducibility among combinatorial problems*, pages 85–103. Plenum Press, New York, 1972.

[15] T.-L. Kriel and M. Schweighofer. On the exactness of Lasserre relaxations and purestates over real closed fields. *Foundations of Computational Mathematics*, (19):1223–1263, 2018.

[16] T.-L. Kriel and M. Schweighofer. On the exactness of Lasserre relaxations for compact convex basic closed semialgebraic sets. *SIAM Journal on Optimization*, (28(2)):1796–1816, 2018.

[17] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.

[18] J.B. Lasserre. *An explicit exact SDP relaxation for nonlinear 0-1 programs*, volume 2081 of *Lecture Notes in Computer Science*, pages 293–303. 2001.

[19] J.B. Lasserre. Convexity in semi-algebraic geometry and polynomial optimization. *SIAM Journal on Optimization*, 19:1995–2014, 2019.

[20] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28(3):470–496, 2003.

[21] M. Laurent. *Sums of squares, moment matrices and optimization over polynomials*, volume 149 of *IMA Volumes in Mathematics and its Applications*, pages 157–270. 2009.

[22] M. Laurent and L.F. Vargas. Work in preparation. 2021.

[23] L. Lovász. Kneser's conjecture, chromatic number, and homotopy. *J. Comb. Theory, Ser. A*, 25:319–324, 1978.

[24] L Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1–7, 1979.

[25] L. Lovász and M. Plummer. *Matching Theory.* Annals of Discrete Mathematics, North-Holland, vol. 29, 1986.

[26] M. Marshall. Optimization of polynomial functions. *Canadian Mathematical Bulletin*, 46:575–587, 2003.

[27] M. Marshall. Representations of non-negative polynomials having finitely many zeros. *Annales de la Faculté des Sciences de Toulouse*, (Vol. XV (3)):599–609, 2006.

[28] M. Marshall. *Positive Polynomials and Sums of Squares*. America Mathematical Society, 2008.

[29] M. Marshall. Representation of non-negative polynomials, degree bounds and applications to optimization. *Canadian Journal of Mathematics*, (61(1)):205–221, 2009.

[30] T. Motzkin and E. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

[31] J. Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. *Mathematical Programming*, 146(1-2):97–121, 2012.

[32] J. Nie. Polynomial optimization with real varieties. *SIAM Journal on Optimization*, 23(3):1634–1646, 2013.

[33] J. Nie. Tight relaxations for polynomial optimization and Lagrange multiplier expressions. *Mathematical Programming*, 178:1–37, 2019.

[34] P.A. Parrilo. Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization. PhD thesis, California Institute of Technology, 2000.

[35] G. Pólya. Über positive Daarstellung von Polynomen. *Naturforsch. Ges. Zurich*, (73):141–145, 1928.

[36] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Ind. Univ. Math. J.*, 42:969–984, 1993.

[37] B. Reznick. Uniform denominators in Hilbert's seventeenth problem. *Mathematische Zeitschrift*, (220):75.

[38] C. Scheiderer. Distinguished representations of non-negative polynomials. *Journal of Algebra*, (289):558–573, 2005.

[39] C. Scheiderer. Sums of squares on real algebraic surfaces. *Manuscripta Mathematica*, 119:395–410, 04 2006.

[40] K. Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Math. Ann.*, 289:203–206, 1991.

[41] A. Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Trans. Inform. Theory*, 25:425–429, 1979.

[42] J. Vera, J. Pena, and L.F. Zuluaga. Computing the stability number of a graph via linear and semidefinite programming. *SIAM Journal on Optimization*, 18(1):87–105, 2007.

[43] A.L. Wang and F. Kilinç-Karzan. On the tightness of SDP relaxations of QCQPs. *Mathematical Programming*, 2021.

[44] L.F. Zuluaga, J. Vera, and J. Pena. LMI approximations for cones of positive semidefinite forms. *SIAM Journal on Optimization*, 16:1076–1091, 2006.