SYMMETRY REDUCTION IN AM/GM-BASED OPTIMIZATION

PHILIPPE MOUSTROU, HELEN NAUMANN, CORDIAN RIENER, THORSTEN THEOBALD, AND HUGUES VERDURE

Abstract. The arithmetic mean/geometric mean-inequality (AM/GM-inequality) facilitates classes of non-negativity certificates and of relaxation techniques for polynomials and, more generally, for exponential sums. Here, we present a first systematic study of the AM/GM-based techniques in the presence of symmetries under the linear action of a finite group. We prove a symmetry-adapted representation theorem and develop techniques to reduce the size of the resulting relative entropy programs. We study in more detail the complexity gain in the case of the symmetric group. In this setup, we can show in particular certain stabilization results. We exhibit several sequences of examples in growing dimensions where the size of the reduced problem stabilizes. Finally, we provide some numerical results, emphasizing the computational speed-up.

1. Introduction

Deciding whether a real function only takes non-negative values is a fundamental question in real algebraic geometry. Non-negativity certificates and optimization approaches are tightly related to each other by observing that the infimum f^* of a function $f : \mathbb{R}^n \to \mathbb{R}$ can be expressed as the largest $\lambda \in \mathbb{R}$ for which $f - \lambda$ is non-negative on $\mathbb{R}^n$:

$$f^* = \inf\{f(x) \ : \ x \in \mathbb{R}^n\} \ = \ \sup\{\lambda \in \mathbb{R} \ : \ f - \lambda \text{ is non-negative on } \mathbb{R}^n\}.$$

Both in the context of polynomials and in the broader context of exponential sums, the last years have seen strong interest in non-negativity certificates and optimization techniques based on the arithmetic mean/geometric mean-inequality (AM/GM inequality). More precisely, an exponential sum (or *signomial*) supported on a finite subset $\mathcal{T} \subset \mathbb{R}^n$ is a linear combination $\sum_{\alpha\in\mathcal{T}} c_\alpha \exp(\langle \alpha, x\rangle)$ with real coefficients c_α. In particular cases, the non-negativity of the real function defined by an exponential sum can be decided via the arithmetic-geometric mean inequality. For example, for support points $\alpha_0, \ldots, \alpha_m \in \mathbb{R}^n$ and coefficients $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^n_+$ satisfying $\sum_{i=1}^m \lambda_i = 1$ and $\sum_{i=1}^m \lambda_i \alpha_i = \alpha_0$, the exponential sum

$$\sum_{i=1}^m \lambda_i \exp(\langle \alpha_i, x\rangle) - \exp(\langle \alpha_0, x\rangle)$$

is non-negative on $\mathbb{R}^n$ as a consequence of the weighted arithmetic-geometric mean inequality, namely $\sum_{i=1}^m \lambda_i \exp(\langle \alpha_i, x\rangle) \geqslant \prod_{i=1}^m (\exp(\langle \alpha_i, x\rangle))^{\lambda_i}$. Clearly, sums of such exponential

Date: December 9, 2021.

2010 *Mathematics Subject Classification.* 14P05, 20C30, 90C30.

Key words and phrases. Positive functions, SAGE certificates, Symmetry reduction, Symmetric group, Relative entropy programming.

sums are non-negative as well. Note that exponential sums can be seen as a generalization of polynomials: when $\mathcal{T} \subset \mathbb{N}^n$, the transformation $x_i = \ln y_i$ gives polynomial functions $y \mapsto \sum_{\alpha \in \mathcal{T}} c_\alpha y^\alpha$ on $\mathbb{R}^n_{>0}$.

These AM/GM-based certificates appear to be particularly useful in sparse settings. In the specialized situation of polynomials, they can be seen as an alternative to non-negativity certificates based on sums of squares. The ideas of these approaches go back to Reznick [30] and have been recently brought back into the focus of the developments by Pantea, Koeppl, and Craciun [28], Chandrasekaran and Shah [7] ("*SAGE*" cone: *sums of arithmetic-geometric exponentials*) and Iliman and de Wolff [18] ("*SONC*" cone: *sums of non-negative circuit polynomials*), see also [21] for a generalized, uniform framework. The AM/GM certificates can be effectively obtained by relative entropy programming (see [7, 8]), and in restricted settings these relative entropy programs become geometric programs [19]. These techniques have been extended to cover constrained situations, prominently by the work of Murray, Chandrasekaran and Wierman based on partial dualization [25]. This method can also be approached from sublinear circuits, see [26]. Furthermore, in the setting of polynomials, the AM/GM-based approaches can be combined with sums of squares [20]. Other recent approaches to sparse polynomials besides the ones based on the AM/GM inequality can be found in the sparse moment hierarchies [38, 39] and in the works exploiting correlative sparsity [22], [36]. Term sparsity is related to sign-symmetries and it is possible to combine correlative and term sparsity [40].

From an algebraic point of view, a problem is *symmetric* when it is invariant under some group action. Symmetries are ubiquitous in the context of polynomials and optimization, since they manifest both in the problem formulation and the solution set. This often allows to reduce the complexity of the corresponding algorithmic questions. Regarding the set of solutions, it was observed by Terquem as early as in 1840 that a symmetric polynomial does not always have a fully symmetric minimizer (see also Waterhouse's survey [41]). However, in many instances, the set of minimizers contains highly symmetric points (see [15, 23, 31, 35]). With respect to problem formulations, symmetry reduction has provided essential advances in many situations (see, for example, [3, 4, 9, 11]), especially in the context of sums of squares (see [2, 6, 10, 16, 17, 29, 32]).

The current paper starts with the question to which extent symmetries can be exploited in AM/GM-based optimization assuming that the problem affords symmetries. We provide a first systematic study of the AM/GM-based approaches in G-invariant situations under the action of a group G. Our focus is on symmetry-adapted representation theorems, and algorithmic symmetry reduction techniques.

Our contributions. 1. We prove a symmetry-adapted decomposition theorem and develop a symmetry-adapted relative entropy formulation of the cone of SAGE exponentials in a general G-invariant setting.

2. This adaption reduces the size of the resulting relative entropy programs or geometric programs, see Theorem 3.1, Theorem 4.1 and Corollary 4.3. As revealed by these statements, the gain depends on the orbit structure of the group action.

3. In the case of the symmetric group, we use combinatorial aspects of the representation theory of the symmetric group in order to measure the size of the resulting relative

entropy program. In particular, we identify situations in which the size of the symmetry adapted relative entropy program stabilizes with respect to the number of variables, see Theorem 5.2.

4. We evaluate the structural results in the paper in terms of computations. In situations with strong symmetry structure, the number of variables and the number of equations and inequalities becomes substantially smaller. Accordingly, the interior-point solvers underlying the computation of SAGE bounds then show strong reductions of computation time. In various cases, the symmetry-adapted computation succeeds when the conventional SAGE computation fails.

We mostly concentrate on the unconstrained optimization, but the techniques can generally also be extended to the constrained case. See, for example, Corollary 4.5. Constrained versions of the SAGE techniques are still rather recent and practical implementations in an early stage; see the recent work [13] for converging hierarchies and their implementation.

The paper is structured as follows. After collecting relevant notions and concepts in Section 2, we provide in Section 3 a specific way of writing sums of arithmetic-geometric exponentials in the presence of a group symmetry. In Section 4, we study how to characterize and to decide whether a G-symmetric exponential sum is contained in the SAGE cone with reduced relative entropy programs. The case of the symmetric group is treated in Section 5, while Section 6 provides experimental results of an implementation of the symmetry reduction techniques. We conclude the paper in Section 7.

Acknowledgement. The authors gratefully acknowledge partial support through the project "Real Algebraic Geometry and Optimization" jointly funded by the German Academic Exchange Service DAAD and the Research Council of Norway RCN, through the Tromsø Research Foundation grant agreement 17matteCR as well as through the project Pure Mathematics in Norway funded by the Trond Mohn Foundation and by the Tromsø Research Foundation. Thanks also to Riley Murray as well as to the anonymous referees for helpful comments.

2. Preliminaries

Throughout the article, we use the notation $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$. For a finite subset $\mathcal{T} \subset \mathbb{R}^n$, let $\mathbb{R}^{\mathcal{T}}$ be the set of $|\mathcal{T}|$-tuples whose components are indexed by the set $\mathcal{T}$. We denote by $\langle \cdot, \cdot \rangle$ the standard Euclidean inner product in $\mathbb{R}^n$.

The SAGE cone. For a given non-empty finite set $\mathcal{T}$, we consider exponential sums supported on $\mathcal{T}$ as defined in the Introduction. For finite $\mathcal{T} \subset \mathbb{R}^n$, the SAGE cone $C_{\mathrm{SAGE}}(\mathcal{T})$ is defined as

$$C_{\mathrm{SAGE}}(\mathcal{T}) := \sum_{\beta \in \mathcal{T}} C_{\mathrm{AGE}}(\mathcal{T} \setminus \{\beta\}, \beta),$$

where for $\mathcal{A} := \mathcal{T} \setminus \{\beta\}$

$$C_{\mathrm{AGE}}(\mathcal{A}, \beta) := \left\{ f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^{\langle \alpha, x \rangle} + c_\beta e^{\langle \beta, x \rangle} \ : \ c_\alpha \geqslant 0 \text{ for } \alpha \in \mathcal{A},\ c_\beta \in \mathbb{R},\ f \geqslant 0 \text{ on } \mathbb{R}^n \right\}$$

denotes the non-negative exponential sums which may only have a negative coefficient in the term indexed by β (see [7]). The elements in these cones are called *SAGE signomials* and *AGE signomials*, respectively. The cone $C_{\mathrm{SAGE}}(\mathcal{T})$ is a closed convex cone in $\mathbb{R}^{\mathcal{T}}$ (see [21, Proposition 2.10]).

Membership in this convex cone can be decided in terms of relative entropy programming. For a finite set $\emptyset \neq \mathcal{A} \subset \mathbb{R}^n$, denote by $D : \mathbb{R}^{\mathcal{A}}_{>0} \times \mathbb{R}^{\mathcal{A}}_{>0} \to \mathbb{R}$,

$$D(\nu, \gamma) \ = \ \sum_{\alpha \in \mathcal{A}} \nu_\alpha \ln \left(\frac{\nu_\alpha}{\gamma_\alpha} \right)$$

the *relative entropy function*, which can be extended to $\mathbb{R}^{\mathcal{A}}_{+} \times \mathbb{R}^{\mathcal{A}}_{+} \to \mathbb{R} \cup \{\infty\}$ via the conventions $0 \cdot \ln \frac{0}{y} = 0$ for $y \geqslant 0$ and $y \cdot \ln \frac{y}{0} = \infty$ for $y > 0$. To decide membership of a given signomial f supported on $\mathcal{T}$ in the SAGE cone, assume that f is written in the form

$$f = \sum_{\alpha \in \mathcal{A}} c_\alpha \exp(\langle \alpha, x \rangle) + \sum_{\beta \in \mathcal{B}} c_\beta \exp(\langle \beta, x \rangle)$$

with $c_\alpha > 0$ for $\alpha \in \mathcal{A}$ and $c_\beta < 0$ for $\beta \in \mathcal{B}$. In this notation, the overall support set of f is $\mathcal{T} = \mathcal{A} \cup \mathcal{B}$. Accordingly, for disjoint sets $\emptyset \neq \mathcal{A} \subset \mathbb{R}^n$ and $\mathcal{B} \subset \mathbb{R}^n$, it is convenient to denote by

$$C_{\mathrm{SAGE}}(\mathcal{A}, \mathcal{B}) := \sum_{\beta \in \mathcal{B}} C_{\mathrm{AGE}}(\mathcal{A} \cup \mathcal{B} \setminus \{\beta\}, \beta) \tag{2.1}$$

the *signed SAGE cone*, which allows negative coefficients only in a certain subset $\mathcal{B}$ of the support $\mathcal{A} \cup \mathcal{B}$. This is a common notation in optimization viewpoints [12, 14, 19, 24, 25].

Proposition 2.1 ([24])**.** *A signomial f belongs to $C_{\mathrm{SAGE}}(\mathcal{A}, \mathcal{B})$ if and only if for every $\beta \in \mathcal{B}$ there exist $c^{(\beta)} \in \mathbb{R}^{\mathcal{A}}_{+}$ and $\nu^{(\beta)} \in \mathbb{R}^{\mathcal{A}}_{+}$ such that*

$$\begin{array}{rcll} \sum_{\alpha \in \mathcal{A}} \nu^{(\beta)}_\alpha \alpha & = & (\sum_{\alpha \in \mathcal{A}} \nu^{(\beta)}_\alpha) \beta & \text{for } \beta \in \mathcal{B}, \\ D(\nu^{(\beta)}, e \cdot c^{(\beta)}) & \leqslant & c_\beta & \text{for } \beta \in \mathcal{B}, \\ \sum_{\beta \in \mathcal{B}} c^{(\beta)}_\alpha & \leqslant & c_\alpha & \text{for } \alpha \in \mathcal{A}. \end{array}$$

Note that this proposition reflects the statement of Murray, Chandrasekaran and Wierman [24] that every SAGE signomial can be decomposed into AGE signomials in such a way that every term with a negative coefficient only appears in a single AGE signomial.

Optimizing over the SAGE cone. Since the SAGE cone is contained in the cone of non-negative signomials, relaxing to the SAGE cone gives an approximation of the global infimum f^* of a signomial f supported on $\mathcal{T}$:

$$f^{\mathrm{SAGE}} := \sup\{\lambda \in \mathbb{R} \ : \ f - \lambda \in C_{\mathrm{SAGE}}(\mathcal{T})\}$$

satisfying $f^{\mathrm{SAGE}} \leqslant f^*$.

Constrained versions. While many aspects of this article are devoted to the unconstrained situation, we briefly collect the extension of SAGE certificates to the constrained situation. Let K be a convex and closed subset of $\mathbb{R}^n$. For a convex set $K \subset \mathbb{R}^n$ and a non-empty finite set $\mathcal{T} \subset \mathbb{R}^n$, the K-SAGE cone $C_K(\mathcal{T})$ is defined (see [25]) as

$$C_K(\mathcal{T}) := \sum_{\beta \in \mathcal{T}} C_K(\mathcal{T} \setminus \{\beta\}, \beta),$$

where for $\mathcal{A} := \mathcal{T} \setminus \{\beta\}$,

$$C_K(\mathcal{A}, \beta) := \Big\{ f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^{\langle \alpha, x \rangle} + c_\beta e^{\langle \beta, x \rangle} \ : \ c_\alpha \geqslant 0 \text{ for } \alpha \in \mathcal{A},\ c_\beta \in \mathbb{R},\ f \geqslant 0 \text{ on } K \Big\}.$$

Moreover, (2.1) can be generalized by defining, for disjoint sets $\emptyset \neq \mathcal{A} \subset \mathbb{R}^n$ and $\mathcal{B} \subset \mathbb{R}^n$, the *signed K-SAGE cone*

$$C_K(\mathcal{A}, \mathcal{B}) := \sum_{\beta \in \mathcal{B}} C_K(\mathcal{A}, \beta).$$

This is the set of K-SAGE signomials, where negative coefficients are only possible in a certain subset $\mathcal{B}$ of the support $\mathcal{A} \cup \mathcal{B}$. The following decomposition result holds.

Theorem 2.2 ([25], Corollary 5)**.** *If $f \in C_K(\mathcal{A}, \mathcal{B})$ with $c_\alpha > 0$ for all $\alpha \in \mathcal{A}$ and $c_\beta < 0$ for all $\beta \in \mathcal{B} \neq \emptyset$, then there exist K-AGE signomials $f_\beta \in C_K(\mathcal{A}, \beta)$ for $\beta \in \mathcal{B}$ such that $f = \sum_{\beta \in \mathcal{B}} f_\beta$.*

For the constrained approach, a similar result to Proposition 2.1 is known.

Proposition 2.3 ([25])**.** *$f \in C_K(\mathcal{A}, \mathcal{B})$ if and only if for every $\beta \in \mathcal{B}$ there exist $c^{(\beta)} \in \mathbb{R}_+^{\mathcal{A}}$ and $\nu^{(\beta)} \in \mathbb{R}_+^{\mathcal{A}}$ such that*

$$\begin{aligned} D(\nu^{(\beta)}, e \cdot c^{(\beta)}) + \sup_{x \in K} \langle - \sum_{\alpha \in \mathcal{A}} \nu_\alpha^{(\beta)} (\alpha - \beta), x \rangle \ &\leqslant\ c_\beta \quad \textit{for } \beta \in \mathcal{B}, \\ \sum_{\beta \in \mathcal{B}} c_\alpha^{(\beta)} \ &\leqslant\ c_\alpha \quad \textit{for } \alpha \in \mathcal{A}. \end{aligned}$$

3. Orbit decompositions of symmetric exponential sums

In this section, we provide a structural result on the decomposition of symmetric SAGE exponentials as sums of orbits of (non-symmetric) AGE exponentials.

Let G be a finite group acting linearly on $\mathbb{R}^n$ on the left, namely we have a group homomorphism

$$\begin{array}{cccc} \varphi & : & G & \to & \mathrm{GL}_n(\mathbb{R}) \\ & & \sigma & \mapsto & \varphi(\sigma) \end{array}.$$

For $\sigma \in G$ and $x \in \mathbb{R}^n$, we denote by $\sigma \cdot x$ the image of x through $\varphi(\sigma)$. In order to get a left action on the set of functions defined on $\mathbb{R}^n$, we need to take

$$(\sigma * f)(x) = f(\sigma^{-1} \cdot x) = f(\varphi(\sigma^{-1})(x)). \tag{3.1}$$

For a signomial $f(x) = \sum_\alpha c_\alpha \exp(\langle \alpha, x\rangle)$, we see an exponent vector α as an element of the dual space. Then, the dual action of G on the exponent vectors is given by

$$\sigma \perp \alpha := \varphi(\sigma^{-1})^{\#}(\alpha),$$

where $A^{\#}$ denotes the adjoint operator of A. Note that this is a left action as well. Therefore, even if the exponents and the variables lie in isomorphic spaces, the actions of G on these spaces are different and dual to each other, and satisfy

$$\langle \alpha, \sigma \cdot x\rangle = \langle \alpha, \varphi(\sigma)(x)\rangle = \langle \varphi(\sigma)^{\#}(\alpha), x\rangle = \langle \sigma^{-1} \perp \alpha, x\rangle$$

and furthermore, for a signomial f,

$$(\sigma * f)(x) = f(\sigma^{-1} \cdot x) = \sum_\alpha c_\alpha \exp(\langle \alpha, \sigma^{-1} \cdot x\rangle) = \sum_\alpha c_\alpha \exp(\langle \sigma \perp \alpha, x\rangle). \tag{3.2}$$

From now on, in order to keep notation as light as possible, with a slight abuse of notation, we write $\sigma(x) = \sigma \cdot x$ for the action on the variables, $\sigma f = \sigma * f$ for the action on functions, and $\sigma(\alpha) = \sigma \perp \alpha$ for the dual action. Even if the actions are different, the context should clarify the correspondence.

For a set $\mathcal{S} \subset \mathbb{R}^n$ of exponent vectors, the *orbit of* $\mathcal{S}$ under G is

$$G \cdot \mathcal{S} = \{\sigma(s) : s \in \mathcal{S}, \sigma \in G\}.$$

We call a subset $\hat{\mathcal{S}} \subset \mathcal{S}$ *a set of orbit representatives for* $\mathcal{S}$ if $\hat{\mathcal{S}}$ is an inclusion-minimal set with $(G \cdot \hat{\mathcal{S}}) = \mathcal{S}$. Moreover, let $\operatorname{Stab} \beta := \{\sigma \in G \ : \ \sigma(\beta) = \beta\}$ denote the *stabilizer* of an exponent vector β.

In the following statements, we consider G-invariant signomials f. It is convenient to write f here in the form

$$f = \sum_{\alpha \in \mathcal{A}} c_\alpha \exp(\langle \alpha, x\rangle) + \sum_{\beta \in \mathcal{B}} c_\beta \exp(\langle \beta, x\rangle) \tag{3.3}$$

with $c_\alpha > 0$ for $\alpha \in \mathcal{A}$ and $c_\beta < 0$ for $\beta \in \mathcal{B}$. As already mentioned in connection with the definition of the signed SAGE cone in (2.1), the overall support set of f is $\mathcal{A} \cup \mathcal{B}$.

The following theorem shows a natural decomposition of a G-invariant signomial f by means of a set of orbit representatives $\hat{\mathcal{B}}$ of $\mathcal{B}$. For every representative $\hat{\beta} \in \hat{\mathcal{B}}$, it is not necessary to take into account the action of all permutations $\sigma \in G$, but it suffices to consider the possibly smaller set $G/\operatorname{Stab}(\hat{\beta})$.

Theorem 3.1. *Let $K \subset \mathbb{R}^n$ be convex and G-invariant, let f be a G-invariant signomial of the form (3.3) and $\hat{\mathcal{B}}$ be a set of orbit representatives for $\mathcal{B}$. Then $f \in C_K(\mathcal{A}, \mathcal{B})$ if and only if for every $\hat{\beta} \in \hat{\mathcal{B}}$, there exists a K-AGE signomial $h_{\hat{\beta}} \in C_K(\mathcal{A}, \hat{\beta})$ such that*

$$f = \sum_{\hat{\beta} \in \hat{\mathcal{B}}} \sum_{\rho \in G/\operatorname{Stab}(\hat{\beta})} \rho h_{\hat{\beta}}. \tag{3.4}$$

The functions $h_{\hat{\beta}}$ can be chosen to be invariant under the action of $\operatorname{Stab}(\hat{\beta})$.

Here, $\rho \in G/\operatorname{Stab}(\hat{\beta})$ shortly denotes that ρ runs over a set of representatives of the left quotient space $G/\operatorname{Stab}(\hat{\beta})$, which is defined through the left cosets $\{\sigma \operatorname{Stab}(\hat{\beta}) : \sigma \in G\}$. We will also use the right quotient space, denoted by $\operatorname{Stab}(\hat{\beta})\backslash G$, further below. To illustrate the theorem, we give an example.

Example 3.2. Let $K := \mathbb{R}^3$ and $G := \mathcal{S}_3$ be the symmetric group on three variables. We consider the signomial

$$f \ = \ e^{6x_1} + e^{6x_2} + e^{6x_3} + e^{x_1+x_2+x_3} - \delta(e^{x_1+2x_2+2x_3} + e^{2x_1+x_2+2x_3} + e^{2x_1+2x_2+x_3})$$

with some constant $\delta \in \mathbb{R}$. Following the notation of Theorem 3.1, let $\mathcal{B} = \{(1,2,2)^T, (2,1,2)^T, (2,2,1)^T\}$, and choose $\hat{\beta} = (1,2,2)^T$ as representative of the single $\mathcal{S}_3$-orbit in $\mathcal{B}$. This gives $\operatorname{Stab}(\hat{\beta}) = \{\mathrm{id}, (2,3)\}$. The left cosets are $\{\mathrm{id}, (2,3)\}$, $\{(1,2,3), (1,2)\}$ as well as $\{(1,3,2), (1,3)\}$, so that we can choose representatives to write $\mathcal{S}_3/\operatorname{Stab}(\hat{\beta}) = \{\mathrm{id}, (1,2,3), (1,3,2)\}$. By Theorem 3.1, the signomial f is SAGE if and only if there exist $a, b, c, d \geqslant 0$ such that

$$h_{\hat{\beta}} = ae^{x_1+x_2+x_3} + be^{6x_1} + ce^{6x_2} + de^{6x_3} - \delta e^{x_1+2x_2+2x_3}$$

is an AGE signomial, invariant under the action of $\operatorname{Stab}(\hat{\beta})$ and satisfies condition (3.4), that is,

$$\begin{aligned} f = {} & 3ae^{x_1+x_2+x_3} + (b+c+d)e^{6x_1} + (b+c+d)e^{6x_2} + (b+c+d)e^{6x_3} \\ & - \delta(e^{x_1+2x_2+2x_3} + e^{2x_1+x_2+2x_3} + e^{2x_1+2x_2+x_3}). \end{aligned}$$

This implies $3a = 1$, $c = d$ and $b + c + d = 1$. With this decomposition, it can be shown that the maximal choice for δ is $\delta = \sqrt[3]{\frac{9}{4}}$, which occurs when $a = \frac{1}{3}$, $b = \frac{1}{6}$ and $c = d = \frac{5}{12}$.

Proof. Since it is clear that a signomial f of the form (3.4) is non-negative, we only have to show the converse direction. Let $f \in C_K(\mathcal{A}, \mathcal{B})$. By Theorem 2.2, there exist K-AGE signomials $f_\beta \in C_K(\mathcal{A}, \beta)$ for $\beta \in \mathcal{B}$, such that $f = \sum_{\beta \in \mathcal{B}} f_\beta$. The G-invariance of f gives

$$f = \frac{1}{|G|} \sum_{\sigma \in G} \sigma f = \frac{1}{|G|} \sum_{\sigma \in G} \sum_{\beta \in \mathcal{B}} \sigma f_\beta. \tag{3.5}$$

The idea is to group in this sum all the σf_β that have the same "possibly negative" term. According to (3.2), the possibly negative term of σf_β is given by $\sigma(\beta)$. For any $\beta \in \mathcal{B}$, the signomial

$$h_\beta = \frac{1}{|G|} \sum_{\sigma \in G} \sigma f_{\sigma^{-1}(\beta)}$$

is a sum of K-AGE signomials in $C_K(\mathcal{A}, \beta)$, hence it is contained in $C_K(\mathcal{A}, \beta)$ as well. Moreover, (3.5) can be expressed as

$$f = \frac{1}{|G|} \sum_{\sigma \in G} \sum_{\beta \in \mathcal{B}} \sigma f_\beta = \frac{1}{|G|} \sum_{\sigma \in G} \sum_{\gamma \in \mathcal{B}} \sigma f_{\sigma^{-1}(\gamma)} = \sum_{\gamma \in \mathcal{B}} h_\gamma.$$

Let $\beta \in \mathcal{B}$ and $\hat{\beta} \in \hat{\mathcal{B}}$ be the representative of its orbit in $\hat{\mathcal{B}}$. If $\sigma, \tau \in G$ are such that $\sigma(\hat{\beta}) = \tau(\hat{\beta}) = \beta$, then $\tau^{-1}\sigma \in \operatorname{Stab}(\hat{\beta})$ and $\tau = \sigma$ in $G/\operatorname{Stab}(\hat{\beta})$. Hence,

$$f = \sum_{\hat{\beta} \in \hat{\mathcal{B}}} \sum_{\rho \in G/\operatorname{Stab}\hat{\beta}} h_{\rho(\hat{\beta})}. \tag{3.6}$$

Now observe that $h_{\rho(\beta)} = \rho h_\beta$ for every $\beta \in \mathcal{B}$ and $\rho \in G$, because

$$|G| \rho h_\beta = \sum_{\sigma \in G} \rho\sigma f_{\sigma^{-1}(\beta)} = \sum_{\tau \in G} \tau f_{\tau^{-1}\rho(\beta)} = |G| h_{\rho(\beta)}. \tag{3.7}$$

Substituting (3.7) into (3.6) gives $f = \sum_{\hat{\beta} \in \hat{\mathcal{B}}} \sum_{\rho \in G/\operatorname{Stab}\hat{\beta}} \rho h_{\hat{\beta}}$ as desired. Moreover, the $\operatorname{Stab}(\hat{\beta})$-invariance of $h_{\hat{\beta}}$ for $\hat{\beta} \in \hat{\mathcal{B}}$ follows from (3.7). □

Remark 3.3. Note that the previous results extend naturally to compact/reductive groups, since they mainly rely on the existence of a Reynolds operator. For the sake of simplicity, we presented them for finite groups, where the Reynolds operator corresponds to a finite average over the group.

4. Symmetry reduction in relative entropy programming

Building upon the previous decomposition theorem, we provide a symmetry-adapted relative entropy formulation for containment in the SAGE cone.

Theorem 4.1. *Let $\hat{\mathcal{B}}$ be a set of orbit representatives for $\mathcal{B}$. A G-invariant signomial f of the form (3.3) is contained in $C_{\mathrm{SAGE}}(\mathcal{A}, \mathcal{B})$ if and only if for every $\hat{\beta} \in \hat{\mathcal{B}}$ there exist $c^{(\hat{\beta})} \in \mathbb{R}_+^{\mathcal{A}}$ and $\nu^{(\hat{\beta})} \in \mathbb{R}_+^{\mathcal{A}}$, invariant under the action of $\operatorname{Stab}(\hat{\beta})$, such that*

$$\sum_{\alpha \in \mathcal{A}} \nu_\alpha^{(\hat{\beta})} (\alpha - \hat{\beta}) \;=\; 0 \quad \textit{for every } \hat{\beta} \in \hat{\mathcal{B}}, \tag{4.1}$$

$$D(\nu^{(\hat{\beta})}, e \cdot c^{(\hat{\beta})}) \;\leqslant\; c_{\hat{\beta}} \quad \textit{for every } \hat{\beta} \in \hat{\mathcal{B}}, \tag{4.2}$$

$$\sum_{\hat{\beta} \in \hat{\mathcal{B}}} \sum_{\sigma \in \operatorname{Stab}(\hat{\beta}) \backslash G} c_{\sigma(\alpha)}^{(\hat{\beta})} \;\leqslant\; c_\alpha \quad \textit{for every } \alpha \in \mathcal{A}. \tag{4.3}$$

Remark 4.2. The right coset condition (4.3) can equivalently be expressed in terms of the left cosets,

$$\sum_{\hat{\beta} \in \hat{\mathcal{B}}} \sum_{\sigma \in G/\operatorname{Stab}\hat{\beta}} c_{\sigma^{-1}(\alpha)}^{(\hat{\beta})} \;\leqslant\; c_\alpha \quad \text{for every } \alpha \in \mathcal{A}.$$

Namely, if $\beta \in \mathcal{B}$, $\hat{\beta} \in \hat{\mathcal{B}}$ and $\sigma, \tau \in G$ are such that $\sigma^{-1}(\hat{\beta}) = \tau^{-1}(\hat{\beta}) = \beta$, then $\tau\sigma^{-1} \in \operatorname{Stab}(\hat{\beta})$ and $\tau = \sigma$ in the right quotient space $\operatorname{Stab}(\hat{\beta}) \backslash G$.

Proof of Theorem 4.1. If f is G-symmetric, then, by Theorem 3.1, there exist $\operatorname{Stab}(\hat{\beta})$-invariant AGE signomials $h_{\hat{\beta}} \in C_{\mathrm{SAGE}}(\mathcal{A}, \hat{\beta})$ for every $\hat{\beta} \in \hat{\mathcal{B}}$ such that

$$f = \sum_{\hat{\beta} \in \hat{\mathcal{B}}} \sum_{\rho \in G/\operatorname{Stab}(\hat{\beta})} \rho h_{\hat{\beta}}.$$

Writing $h_{\hat{\beta}}$ in the form

$$h_{\hat{\beta}} = \sum_{\alpha \in \mathcal{A}} c_\alpha^{(\hat{\beta})} \exp(\langle \alpha, x \rangle) + c_{\hat{\beta}} \exp(\langle \hat{\beta}, x \rangle)$$

with coefficients $c_\alpha^{(\hat{\beta})}$ and $c_{\hat{\beta}}$ for $\alpha \in \mathcal{A}$ and $\hat{\beta} \in \hat{\mathcal{B}}$, the two conditions (4.1) and (4.2) follow from the property $h_{\hat{\beta}} \in C_{\mathrm{SAGE}}(\mathcal{A}, \hat{\beta})$. For (4.3), we observe that for $\alpha \in \mathcal{A}$, the coefficient of $\exp(\langle \alpha, x \rangle)$ in $\rho h_{\hat{\beta}}$ is $c^{(\hat{\beta})}_{\rho^{-1}(\alpha)}$. We obtain inequality (4.3), even with equality, by setting $\sigma := \rho^{-1}$ and summing over $\hat{\beta} \in \hat{\mathcal{B}}$ and over $\sigma \in \mathrm{Stab}(\hat{\beta}) \backslash G$, following Remark 4.2. Moreover, the $\mathrm{Stab}(\hat{\beta})$-invariance of $h_{\hat{\beta}}$ implies the $\mathrm{Stab}(\hat{\beta})$-invariance of $c^{(\hat{\beta})}$. In order to make $\nu^{(\hat{\beta})}$ invariant under $\mathrm{Stab}(\hat{\beta})$, we can replace it with

$$\mu_\alpha^{(\hat{\beta})} = \frac{1}{|\,\mathrm{Stab}(\hat{\beta})|} \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \nu^{(\hat{\beta})}_{\sigma(\alpha)}.$$

Obviously, this has no influence on (4.3). For (4.1), we have

$$\begin{aligned} |\,\mathrm{Stab}(\hat{\beta})| \sum_{\alpha \in \mathcal{A}} \mu_\alpha^{(\hat{\beta})} (\alpha - \hat{\beta}) &= \sum_{\alpha \in \mathcal{A}} \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \nu^{(\hat{\beta})}_{\sigma(\alpha)} (\alpha - \hat{\beta}) \\ &= \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \sigma^{-1} \sum_{\alpha \in \mathcal{A}} \nu^{(\hat{\beta})}_{\sigma(\alpha)} (\sigma(\alpha) - \sigma(\hat{\beta})) \\ &= \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \sigma^{-1} \sum_{\alpha \in \mathcal{A}} \nu^{(\hat{\beta})}_{\alpha} (\alpha - \hat{\beta}) = 0. \end{aligned}$$

Finally, for (4.2), using $c_\alpha^{(\hat{\beta})} = c^{(\hat{\beta})}_{\sigma(\alpha)}$ for $\sigma \in \mathrm{Stab}(\hat{\beta})$ and applying Jensen's inequality on the convex function $x \mapsto x \ln x$ gives, for all $\alpha \in \mathcal{A}$,

$$\begin{aligned} \mu_\alpha^{(\hat{\beta})} \ln \frac{\mu_\alpha^{(\hat{\beta})}}{c_\alpha^{(\hat{\beta})}} &= \left(\frac{1}{|\,\mathrm{Stab}(\hat{\beta})|} \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \nu^{(\hat{\beta})}_{\sigma(\alpha)} \right) \ln \frac{\frac{1}{|\,\mathrm{Stab}(\hat{\beta})|} \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \nu^{(\hat{\beta})}_{\sigma(\alpha)}}{c_\alpha^{(\hat{\beta})}} \\ &= c_\alpha^{(\hat{\beta})} \left(\frac{\sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \nu^{(\hat{\beta})}_{\sigma(\alpha)} / c^{(\hat{\beta})}_{\sigma(\alpha)}}{|\,\mathrm{Stab}(\hat{\beta})|} \ln \frac{\sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \nu^{(\hat{\beta})}_{\sigma(\alpha)} / c^{(\hat{\beta})}_{\sigma(\alpha)}}{|\,\mathrm{Stab}(\hat{\beta})|} \right) \\ &\leqslant c_\alpha^{(\hat{\beta})} \left(\frac{1}{|\,\mathrm{Stab}(\hat{\beta})|} \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \frac{\nu^{(\hat{\beta})}_{\sigma(\alpha)}}{c^{(\hat{\beta})}_{\sigma(\alpha)}} \ln \frac{\nu^{(\hat{\beta})}_{\sigma(\alpha)}}{c^{(\hat{\beta})}_{\sigma(\alpha)}} \right). \end{aligned}$$

Using again the $\mathrm{Stab}(\hat{\beta})$-invariance of $c^{(\hat{\beta})}$ and the precondition then yields

$$\sum_{\alpha \in \mathcal{A}} \mu_\alpha^{(\hat{\beta})} \ln \frac{\mu_\alpha^{(\hat{\beta})}}{e c_\alpha^{(\hat{\beta})}} \leqslant \frac{1}{|\,\mathrm{Stab}(\hat{\beta})|} \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} \sum_{\alpha \in \mathcal{A}} \nu^{(\hat{\beta})}_{\sigma(\alpha)} \ln \frac{\nu^{(\hat{\beta})}_{\sigma(\alpha)}}{e c^{(\hat{\beta})}_{\sigma(\alpha)}} \leqslant \frac{1}{|\,\mathrm{Stab}(\hat{\beta})|} \sum_{\sigma \in \mathrm{Stab}(\hat{\beta})} c_{\hat{\beta}} = c_{\hat{\beta}}.$$

Conversely, assume that $c^{(\hat{\beta})}$ and $\nu^{(\hat{\beta})}$, invariant under the action of $\mathrm{Stab}(\hat{\beta})$, satisfy (4.1)–(4.3). Let $\beta \in \mathcal{B}$ and $\hat{\beta} \in \hat{\mathcal{B}}$ be the representative of its orbit in $\hat{\mathcal{B}}$. If $\sigma, \tau \in G$ are such that $\sigma(\beta) = \tau(\beta) = \hat{\beta}$, then $\tau\sigma^{-1} \in \mathrm{Stab}(\hat{\beta})$ and $\tau = \sigma$ in $\mathrm{Stab}(\hat{\beta})\backslash G$. Since $c^{(\hat{\beta})}$ and $\nu^{(\hat{\beta})}$ are invariant under $\mathrm{Stab}(\hat{\beta})$, we have

$$c^{(\hat{\beta})}_{\tau(\alpha)} = c^{(\hat{\beta})}_{\sigma(\alpha)}, \quad \nu^{(\hat{\beta})}_{\tau(\alpha)} = \nu^{(\hat{\beta})}_{\sigma(\alpha)} \quad \text{for } \alpha \in \mathcal{A}.$$

Thus we can define

$$c^{(\beta)}_{\alpha} = c^{(\hat{\beta})}_{\sigma(\alpha)}, \quad \nu^{(\beta)}_{\alpha} = \nu^{(\hat{\beta})}_{\sigma(\alpha)} \quad \text{for } \alpha \in \mathcal{A},$$

which is independent of σ such that $\sigma(\beta) = \hat{\beta}$. As a consequence, if $\tau \in \mathrm{Stab}(\hat{\beta})\backslash G$, then $c^{(\tau^{-1}(\hat{\beta}))}_{\alpha} = c^{(\hat{\beta})}_{\tau(\alpha)}$ is well defined.

To see that the first conditions of Proposition 2.1 are satisfied, let $\beta \in \mathcal{B}$ and $\sigma \in G$ such that $\sigma(\beta) = \hat{\beta}$. Then

$$\begin{aligned}\sum_{\alpha\in\mathcal{A}} \nu^{(\beta)}_{\alpha}(\alpha - \beta) &= \sum_{\alpha\in\mathcal{A}} \nu^{(\hat{\beta})}_{\sigma(\alpha)}(\alpha - \sigma^{-1}(\hat{\beta})) \\ &= \sigma^{-1}\sum_{\alpha\in\mathcal{A}} \nu^{(\hat{\beta})}_{\sigma(\alpha)}(\sigma(\alpha) - \hat{\beta}) = \sigma^{-1}\sum_{\alpha\in\mathcal{A}} \nu^{(\hat{\beta})}_{\alpha}(\alpha - \hat{\beta}) = 0\end{aligned}$$

$$\text{and} \quad D(\nu^{(\beta)}, ec^{(\beta)}) = D(\nu^{(\hat{\beta})}, ec^{(\hat{\beta})}) \leqslant c_{\hat{\beta}} = c_{\beta}.$$

For the third condition of Proposition 2.1, we obtain

$$\sum_{\beta\in\mathcal{B}} c^{(\beta)}_{\alpha} = \sum_{\hat{\beta}\in\hat{\mathcal{B}}} \sum_{\tau\in\mathrm{Stab}(\hat{\beta})\backslash G} c^{(\tau^{-1}(\hat{\beta}))}_{\alpha} = \sum_{\hat{\beta}\in\hat{\mathcal{B}}} \sum_{\tau\in\mathrm{Stab}(\hat{\beta})\backslash G} c^{(\hat{\beta})}_{\tau(\alpha)} \leqslant c_{\alpha},$$

which altogether shows that $f \in C_{\mathrm{SAGE}}(\mathcal{A}, \mathcal{B})$.

□

The following consequence of Theorem 4.1 further reduces the number of variables in the relative entropy program, since a certain number of $c^{(\hat{\beta})}_{\alpha}$ and $\nu^{(\hat{\beta})}_{\alpha}$ are actually equal, and we can take each $c^{(\hat{\beta})}, \nu^{(\hat{\beta})}$ in the ground set $\mathbb{R}_{+}^{\mathcal{A}/\mathrm{Stab}(\hat{\beta})}$.

Corollary 4.3. *Let $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ be a set of orbit representatives for $\mathcal{A}$ and $\mathcal{B}$. A G-invariant signomial f of the form (3.3) is contained in $C_{\mathrm{SAGE}}(\mathcal{A}, \mathcal{B})$ if and only if for every $\hat{\beta} \in \hat{\mathcal{B}}$ there exist $c^{(\hat{\beta})} \in \mathbb{R}_{+}^{\mathcal{A}/\mathrm{Stab}(\hat{\beta})}$ and $\nu^{(\hat{\beta})} \in \mathbb{R}_{+}^{\mathcal{A}/\mathrm{Stab}(\hat{\beta})}$ such that*

$$\sum_{\alpha\in\mathcal{A}/\mathrm{Stab}(\hat{\beta})} \nu^{(\hat{\beta})}_{\alpha} \sum_{\alpha'\in\mathrm{Stab}(\hat{\beta})\cdot\alpha} (\alpha' - \hat{\beta}) \;=\; 0 \quad \text{for every } \hat{\beta} \in \hat{\mathcal{B}}, \tag{4.4}$$

$$\sum_{\alpha\in\mathcal{A}/\mathrm{Stab}(\hat{\beta})} \left|\mathrm{Stab}(\hat{\beta})\cdot\alpha\right| \nu^{(\hat{\beta})}_{\alpha} \ln\frac{\nu^{(\hat{\beta})}_{\alpha}}{ec^{(\hat{\beta})}_{\alpha}} \;\leqslant\; c_{\hat{\beta}} \quad \text{for every } \hat{\beta} \in \hat{\mathcal{B}}, \tag{4.5}$$

$$\sum_{\hat{\beta}\in\hat{\mathcal{B}}} \frac{|\mathrm{Stab}(\alpha)|}{|\mathrm{Stab}(\hat{\beta})|} \sum_{\gamma\in(G\cdot\alpha)/\mathrm{Stab}(\hat{\beta})} \left|\mathrm{Stab}(\hat{\beta})\cdot\gamma\right| c^{(\hat{\beta})}_{\gamma} \;\leqslant\; c_{\alpha} \quad \text{for every } \alpha \in \hat{\mathcal{A}}. \tag{4.6}$$

Proof. For (4.4) and (4.5), equivalence to their versions in Theorem 4.1 is straightforward to check. For (4.6), equivalence to (4.3) follows by observing that for every $\alpha \in \mathcal{A}$

$$\sum_{\sigma \in \operatorname{Stab}(\hat{\beta})\backslash G} c^{(\hat{\beta})}_{\sigma(\alpha)} = \sum_{\sigma \in \operatorname{Stab}(\hat{\beta})\backslash G} \frac{1}{|\operatorname{Stab}(\hat{\beta})|} \sum_{\tau \in \operatorname{Stab}(\hat{\beta})} c^{(\hat{\beta})}_{\tau(\sigma(\alpha))} = \frac{1}{|\operatorname{Stab}(\hat{\beta})|} \sum_{\rho \in G} c^{(\hat{\beta})}_{\rho(\alpha)}$$
$$= \frac{|\operatorname{Stab}(\alpha)|}{|\operatorname{Stab}(\hat{\beta})|} \sum_{\gamma \in G\cdot\alpha} c^{(\hat{\beta})}_{\gamma} = \frac{|\operatorname{Stab}(\alpha)|}{|\operatorname{Stab}(\hat{\beta})|} \sum_{\gamma \in (G\cdot\alpha)/\operatorname{Stab}(\hat{\beta})} \left|\operatorname{Stab}(\hat{\beta}) \cdot \gamma\right| c^{(\hat{\beta})}_{\gamma},$$

and the last expression only depends on the orbit $G \cdot \alpha$ rather than on α itself. ☐

Remark 4.4. Note that we cannot simply assume $c^{(\beta)}_{\alpha} = c^{(\beta)}_{\alpha'}$ for some $\alpha' \in G \cdot \alpha$ and, similarly, we cannot simply assume $\nu^{(\beta)}_{\alpha} = \nu^{(\beta)}_{\alpha'}$ for some $\alpha' \in G\cdot\alpha$, for instance due to (2.1). Namely, if an element β lies in $\operatorname{conv} \mathcal{A}$ with barycentric coordinates λ, say $\beta = \sum_{\alpha\in\mathcal{A}} \lambda_\alpha \alpha$, then for any $\sigma \in G$, we have

$$\sigma(\beta) = \sigma\left(\sum_{\alpha\in\mathcal{A}} \lambda_\alpha \alpha\right) = \sum_{\alpha\in\mathcal{A}} \sigma(\lambda_\alpha \alpha) = \sum_{\alpha\in\mathcal{A}} \lambda_\alpha \sigma(\alpha)$$

rather than $\sigma(\beta) = \sigma(\sum_{\alpha\in\mathcal{A}} \lambda_\alpha \alpha) = \sum_{\alpha\in\mathcal{A}} \lambda_{\sigma(\alpha)} \sigma(\alpha)$. Of course, this caveat does not occur whenever there is a single inner term.

For symmetric constraint sets K, a constrained version of Theorem 4.1 (and similarly, of Corollary 4.3) can be given as well. The proof is similar.

Corollary 4.5. *Let $K \subset \mathbb{R}^n$ be convex and G-invariant. A G-invariant signomial f of the form* (3.3) *is contained in $C_K(\mathcal{A}, \mathcal{B})$ if and only if for every $\hat{\beta} \in \hat{\mathcal{B}}$ there exist $c^{(\hat{\beta})} \in \mathbb{R}^{\mathcal{A}}_+$ and $\nu^{(\hat{\beta})} \in \mathbb{R}^{\mathcal{A}}_+$ such that*

$$\begin{aligned} D(\nu^{(\hat{\beta})}, e \cdot c^{(\hat{\beta})}) + \sup_{x\in K} \langle \big(-\sum_{\alpha\in\mathcal{A}} \nu^{(\hat{\beta})}_{\alpha} (\alpha - \hat{\beta}) \big), x \rangle &\leqslant c_{\hat{\beta}} \quad \text{for every } \hat{\beta} \in \hat{\mathcal{B}}, \\ \sum_{\hat{\beta}\in\hat{\mathcal{B}}} \sum_{\sigma\in \operatorname{Stab}\hat{\beta}\backslash G} c^{(\hat{\beta})}_{\sigma(\alpha)} &\leqslant c_\alpha \quad \text{for every } \alpha \in \mathcal{A}. \end{aligned}$$

To close this section we discuss the resulting complexity reduction:

Note that the initial relative entropy formulation which does not take the symmetry into consideration will involve $2|\mathcal{B}||\mathcal{A}|$ variables. Furthermore, since every vector equality in (4.4) brings n scalar equalities, it will consist of $|\mathcal{B}|n + |\mathcal{B}| + |\mathcal{A}|$ (in)equalities.

In contrast, let us analyze the number of variables and constraints involved in the relative entropy program in Corollary 4.3. Observe that $\mathcal{A}/\operatorname{Stab}(\hat{\beta})$ is the disjoint union of the $G\cdot\hat{\alpha}/\operatorname{Stab}(\hat{\beta})$ where $\hat{\alpha}$ runs through $\hat{\mathcal{A}}$. It follows that for every pair $\hat{\beta} \in \hat{\mathcal{B}}, \hat{\alpha} \in \hat{\mathcal{A}}$, we have exactly $2|(G \cdot \hat{\alpha})/\operatorname{Stab}(\hat{\beta})|$ variables $c^{(\hat{\beta})}_{\gamma}$ and $\nu^{(\hat{\beta})}_{\gamma}$.

By definition, $|(G \cdot \hat{\alpha})/\operatorname{Stab}(\hat{\beta})|$ is the number of $\operatorname{Stab}(\hat{\beta})$-orbits in $G \cdot \hat{\alpha}$. Since $G \cdot \hat{\alpha}$ is in bijection with $\operatorname{Stab}\hat{\alpha}\backslash G$ we get a bijection between $(G \cdot \hat{\alpha})/\operatorname{Stab}(\hat{\beta})$ and the set of double cosets $\operatorname{Stab}(\hat{\alpha})\backslash G/\operatorname{Stab}(\hat{\beta})$. Therefore, the number of orbits in question equals

$|\operatorname{Stab}(\hat\alpha)\backslash G/\operatorname{Stab}(\hat\beta)|$, satisfying, according to Burnside's Lemma (see for instance [34, Lemma 7.24.5]):

$$|\operatorname{Stab}(\hat\alpha)\backslash G/\operatorname{Stab}(\hat\beta)| = \frac{1}{|\operatorname{Stab}(\hat\alpha)||\operatorname{Stab}(\hat\beta)|}\sum_{\substack{\sigma\in\operatorname{Stab}(\hat\alpha)\\ \tau\in\operatorname{Stab}(\hat\beta)}} |G^{\sigma,\tau}|,$$

where $|G^{\sigma,\tau}|$ is the number of elements of G fixed under the action of (σ,τ). From another point of view, this number can be interpreted in terms of representation theory as follows: It is given by the inner product of the two characters corresponding to the representations induced respectively by the trivial representations of $\operatorname{Stab}(\hat\alpha)$ and $\operatorname{Stab}(\hat\beta)$ on G (see [34, Exercise 7.77.a.] for more details).

Furthermore, (4.4) amounts to $|\hat{\mathcal{A}}|+|\hat{\mathcal{B}}|$ inequalities, together with one vector equality for every element of $\hat{\mathcal{B}}$. We observe that for a given $\hat\beta$, this vector is invariant by $\operatorname{Stab}(\hat\beta)$ and therefore is contained in $(\mathbb{R}^n)^{\operatorname{Stab}(\hat\beta)}$, the subspace of $\mathbb{R}^n$ of points fixed by $\operatorname{Stab}(\hat\beta)$. Thus, by projecting onto this subspace the number of resulting equations reduces to $\dim\left((\mathbb{R}^n)^{\operatorname{Stab}(\hat\beta)}\right)$. As a conclusion, we obtain:

Theorem 4.6. *Let $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ be a set of orbit representatives for $\mathcal{A}$ and $\mathcal{B}$. For $\hat\alpha\in\hat{\mathcal{A}}$, $\hat\beta\in\hat{\mathcal{B}}$, denote by ${}_{\hat\alpha}G_{\hat\beta}$ the cardinality $|\operatorname{Stab}(\hat\alpha)\backslash G/\operatorname{Stab}(\hat\beta)|$, and by $n_{\hat\beta}$ the dimension of the fixed subspace $(\mathbb{R}^n)^{\operatorname{Stab}(\hat\beta)}$. Then, the relative entropy program in Corollary 4.3 consists of*

$$2\sum_{\substack{\hat\alpha\in\hat{\mathcal{A}}\\ \hat\beta\in\hat{\mathcal{B}}}} {}_{\hat\alpha}G_{\hat\beta}\ \textit{variables},\quad \sum_{\hat\beta\in\hat{\mathcal{B}}} n_{\hat\beta}\ \textit{scalar equalities, and}\quad |\hat{\mathcal{A}}|+|\hat{\mathcal{B}}|\ \textit{inequalities.}$$

The next section will make the theorem more concrete in the special case of the symmetric group.

5. The case of the symmetric group

In this section, we focus our attention to the case of the symmetric group $\mathcal{S}_n$ acting on $\mathbb{R}^n$ by permutation of the coordinates: for $\sigma\in\mathcal{S}_n, x\in\mathbb{R}^n$,

$$\sigma(x) = (x_{\sigma^{-1}(1)},\ldots,x_{\sigma^{-1}(n)}).$$

Note that because the action is orthogonal, the dual action on the exponent vectors is the same. Optimization problems invariant under this action can arise in different contexts naturally, for example, in the context of graph homomorphisms ([5]). This action is very natural, and the theory of representation of the symmetric group is very well understood, and affords strong connections with combinatorics. This connection has been successfully used in several instances to reduce the sizes of optimization problems. In particular, it was shown in [32, Theorem 4.7] (see also [10, Theorem 3.21]) that the size of a semidefinite program which certifies if a given symmetric polynomial is a sum of squares is stabilizing once the number of variables is big enough. Similar to these results, we show in Theorem 5.2 an analogous result of stabilization in the AM/GM setup. This result mainly stems from the fact that the cardinalities appearing in Theorem 4.6 have a combinatorial interpretation in the context of symmetric group actions.

From now on, we use the symbols λ, μ to denote partitions. This should cause no confusion to the use of these symbols in other contexts in earlier sections.

First, up to permutation, every $\alpha \in \mathbb{R}^n$ is of the form

$$\alpha = (\underbrace{\alpha_1, \ldots, \alpha_1}_{\lambda_1}, \underbrace{\alpha_2, \ldots, \alpha_2}_{\lambda_2}, \ldots, \underbrace{\alpha_k, \ldots, \alpha_k}_{\lambda_k}),$$

with $\lambda_1 \geqslant \lambda_2 \ldots \geqslant \lambda_k > 0$ and thus its stabilizer $\operatorname{Stab}(\alpha)$ is, up to conjugation, of the form

$$\mathcal{S}_{\lambda_1} \times \cdots \times \mathcal{S}_{\lambda_k},$$

so that $|\operatorname{Stab}(\alpha)| = \lambda_1! \cdots \lambda_k!$. The corresponding partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ of n is called the *orbit type* $\Lambda(\alpha)$ of α. We then denote by $\operatorname{len}(\alpha)$ the *length* of this partition, namely $\operatorname{len}(\alpha) = k$. Consequently, for $\hat{\beta} \in \hat{\mathcal{B}}$, the dimension $n_{\hat{\beta}}$ of the fixed subspace $(\mathbb{R}^n)^{\operatorname{Stab}(\hat{\beta})}$ is precisely $\operatorname{len}(\hat{\beta})$.

Furthermore, let $\alpha \in \hat{\mathcal{A}}$ of orbit type $\Lambda(\alpha) = (\lambda_1, \ldots, \lambda_k)$, and $\beta \in \hat{\mathcal{B}}$ of orbit type $\Lambda(\beta) = (\mu_1, \ldots, \mu_\ell)$. Then, the interpretation of ${}_{\hat{\alpha}}(\mathcal{S}_n)_{\hat{\beta}}$ as the inner product of characters gives a combinatorial understanding of the number ${}_{\hat{\alpha}}(\mathcal{S}_n)_{\hat{\beta}} = |\operatorname{Stab}(\hat{\alpha})\backslash\mathcal{S}_n/\operatorname{Stab}(\hat{\beta})|$: it is given by the number $N_{\Lambda(\alpha),\Lambda(\beta)} = |\mathcal{M}_{\Lambda(\alpha),\Lambda(\beta)}|$, where $\mathcal{M}_{\Lambda(\alpha),\Lambda(\beta)}$ is the set of matrices of size $k \times \ell$ with non-negative integer coefficients such that, for $1 \leqslant i \leqslant k$ the elements of the ith row sum up to λ_i, and for $1 \leqslant j \leqslant \ell$ the elements of the jth column sum up to μ_j. This quantity can be alternatively computed by using the so-called *Kostka numbers* defined for pairs of partitions. More precisely, we have

$${}_{\hat{\alpha}}(\mathcal{S}_n)_{\hat{\beta}} = N_{\Lambda(\alpha),\Lambda(\beta)} = \sum_{\mu} K_{\mu,\Lambda(\alpha)} K_{\mu,\Lambda(\beta)},$$

where μ runs through the partitions of n. For more details about these interpretations, see [34, Chapter 7], in particular Corollary 7.12.3 therein.

Now we illustrate the potential gain of this reduction, already in a very small example:

Example 5.1. Consider the support set $\{\alpha_0, \ldots, \alpha_7\} = \{(0,0,0)^T, (7,0,0)^T, (0,7,0)^T, (0,0,7)^T, (1,1,2)^T, (1,2,1)^T, (2,1,1)^T, (2,2,2)^T\}$ and let $G := \mathcal{S}_3$ be the symmetric group on three elements. In order to avoid too heavy notation, we will write $c_j^{(i)}$ instead of $c_{\alpha_j}^{(\alpha_i)}$ and $\nu_j^{(i)}$ instead of $\nu_{\alpha_j}^{(\alpha_i)}$. Consider a signomial

$$f(x_1, x_2, x_3) = \sum_{i=0}^{7} c_i e^{\langle \alpha_i, (x_1, x_2, x_3) \rangle},$$

with $c_0, c_1, c_2, c_3 > 0$ and $c_4, c_5, c_6, c_7 < 0$, i.e., set $\mathcal{A} = \{\alpha_0, \ldots, \alpha_3\}$, $\mathcal{B} = \{\alpha_4, \ldots, \alpha_7\}$. Then $\hat{\mathcal{A}} = \{\alpha_0, \alpha_1\}$ and $\hat{\mathcal{B}} = \{\alpha_4, \alpha_7\}$ are sets of orbit representatives. The corresponding partitions are $\Lambda(\alpha_0) = \Lambda(\alpha_7) = (3)$, and $\Lambda(\alpha_1) = \Lambda(\alpha_4) = (2,1)$. Then, by Corollary 4.3,

$f \in C_{\text{SAGE}}(\mathcal{A}, \mathcal{B})$ if and only there exist $c^{(4)} = (c_0^{(4)}, c_1^{(4)}, c_3^{(4)})$, $\nu^{(4)} = (\nu_0^{(4)}, \nu_1^{(4)}, \nu_3^{(4)})$, $c^{(7)} = (c_0^{(7)}, c_1^{(7)})$ and $\nu^{(7)} = (\nu_0^{(7)}, \nu_1^{(7)})$ satisfying the conditions

$$\begin{aligned}
\nu_0^{(4)}(\alpha_0 - \alpha_4) + \nu_1^{(4)}(\alpha_1 + \alpha_2 - 2\alpha_4) + \nu_3^{(4)}(\alpha_3 - \alpha_4) &= 0, \\
\nu_0^{(7)}(\alpha_0 - \alpha_7) + \nu_1^{(7)}(\alpha_1 + \alpha_2 + \alpha_3 - 3\alpha_7) &= 0, \\
\nu_0^{(4)} \ln \frac{\nu_0^{(4)}}{c_0^{(4)}} + 2\nu_1^{(4)} \ln \frac{\nu_1^{(4)}}{c_1^{(4)}} + \nu_3^{(4)} \ln \frac{\nu_3^{(4)}}{c_3^{(4)}} &\leqslant c_4, \\
\nu_0^{(7)} \ln \frac{\nu_0^{(7)}}{c_0^{(7)}} + 3\nu_1^{(7)} \ln \frac{\nu_1^{(7)}}{c_1^{(7)}} &\leqslant c_7, \\
3c_0^{(4)} + c_0^{(7)} &\leqslant c_0, \\
2c_1^{(4)} + c_3^{(4)} + c_1^{(7)} &\leqslant c_1.
\end{aligned}$$

Note that here $\text{len}(\alpha_4) = 2$ and $\text{len}(\alpha_7) = 1$ so that the two vectorial equations bring together $2 + 1$ scalar equations. In total, we get $2(1 + 1 + 2 + 1) = 10$ variables and $2 + 1 + 2 + 2 = 7$ linear constraints, against $2 \cdot 4 \cdot 4 = 32$ variables and $4 \cdot 3 + 4 + 4 = 20$ linear constraints forgetting about symmetries.

In the context of the symmetric group it is natural to consider situations in which the number of variables grows. Such situations were studied for example in the context of sums of squares relaxations. Several examples were observed in which the complexity of the symmetry-adapted semi-definite program is independent of the number of variables. Analogously, we now describe natural sequences of signomials where the size of the relative entropy program stabilizes when the number of variables is large enough.

Fix $n_0 \in \mathbb{N}$ and start with a signomial f_{n_0} in n_0 variables, represented by the orbit representatives of the exponent vectors $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$, as well as the corresponding coefficients. For each of these exponents α, we denote by $\tilde{\Lambda}(\alpha)$ the orbit type of α *where we forget about the* 0 *entries*. For instance, when $n_0 = 3$, $\hat{\mathcal{A}} = \{\hat{\alpha}\} = \{(1, 1, 2)\}$ and $\hat{\mathcal{B}} = \{\hat{\beta}\} = \{(0, 0, 1)\}$, then $\tilde{\Lambda}(\hat{\alpha}) = (2, 1)$ while $\tilde{\Lambda}(\hat{\beta}) = (1)$. Note that these sequences do not have to be partitions of n, we therefore introduce

$$\text{wt}(\alpha) = \sum_{\lambda \in \tilde{\Lambda}(\alpha)} \lambda,$$

counting the number of non-zero coordinates of α, and refer to it as the *weight* of α. Hence $\text{wt}(1, 1, 2) = 3$, while $\text{wt}(0, 0, 1) = 1$. Now, for every $n \geqslant n_0$, we can see α as an exponent in $\mathbb{R}^n$, by adding $n - n_0$ zeroes. This procedure does not affect $\tilde{\Lambda}(\alpha)$ and $\text{wt}(\alpha)$. In this way, we can define for every $n > n_0$, the unique $\mathcal{S}_n$-invariant signomial f_n whose support is made of the $\mathcal{S}_n$-orbits of $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ with the corresponding coefficients. Clearly, in this situation, the number of constraints C_n in Corollary 4.3 does not depend on n, since it only involves $|\hat{\mathcal{B}}|$, $|\hat{\mathcal{A}}|$, and the length of the elements in $\hat{\mathcal{B}}$, which does not change when $n \geqslant n_0 + 1$. In this framework, a similar phenomenon holds for the number of variables:

Theorem 5.2 (*Stabilization Theorem*)**.** *Let $n_0 \in \mathbb{N}$, and $\hat{\mathcal{A}}$, $\hat{\mathcal{B}}$ be finite orbit representatives of exponent vectors in $\mathbb{R}^{n_0}$. Consider, for $n \geqslant n_0$, the signomial f_n previously defined,* *and denote by V_n the number of variables in the symmetry-adapted relative entropy program in Corollary 4.3. Let*

$$m = \max\{\mathrm{wt}(\alpha)\ : \alpha \in \hat{\mathcal{A}} \cup \hat{\mathcal{B}}\}.$$

Then, for every $n \geqslant 2m$, $V_n = V_{2m}$.

Proof. We shall show by induction that for every $n \geqslant 2m$, $V_n = V_{2m}$. The initial step being obvious, assume $n > 2m$. The definition of m ensures that for $n \geqslant 2m$, for every α in $\hat{\mathcal{A}} \cup \hat{\mathcal{B}}$, the coordinate occurring the most in α is 0, and therefore

$$\Lambda(\alpha) = (n - \mathrm{wt}(\alpha), \lambda_1, \dots, \lambda_k),$$

where $(\lambda_1, \dots, \lambda_k) = \tilde{\Lambda}(\alpha)$. Remember that the number of variables is given by

$$V_n = 2 \sum_{\hat{\alpha} \in \hat{\mathcal{A}}, \hat{\beta} \in \hat{\mathcal{B}}} N^n_{\Lambda(\hat{\alpha}),\Lambda(\hat{\beta})},$$

where, if $\Lambda(\hat{\alpha}) = (n - wt(\hat{\alpha}), \lambda_1, \dots, \lambda_k)$ and $\Lambda(\hat{\beta}) = (n - \mathrm{wt}(\hat{\beta}), \mu_1, \dots, \mu_\ell)$, the quantity $N^n_{\Lambda(\hat{\alpha}),\Lambda(\hat{\beta})}$ counts the number of matrices of size $(k+1) \times (\ell+1)$ with non-negative integer coefficients of the form

$$\begin{array}{c} \begin{array}{ccccc} n - \mathrm{wt}(\hat{\beta}) & \mu_1 & \mu_2 & \dots & \mu_\ell \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \dots & \cdot \end{bmatrix} \begin{array}{c} n - \mathrm{wt}(\hat{\alpha}) \\ \lambda_1 \\ \vdots \\ \lambda_k \end{array} \end{array} \tag{5.1}$$

where the labels give the sum of the coefficients in the corresponding row/column. Since $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ keep the same number of elements, we only need to show that for every $\hat{\alpha} \in \hat{\mathcal{A}}$, $\hat{\beta} \in \hat{\mathcal{B}}$, we have $N^n_{\Lambda(\hat{\alpha}),\Lambda(\hat{\beta})} = N^{n-1}_{\Lambda(\hat{\alpha}),\Lambda(\hat{\beta})}$.

If we start with a matrix of the form

$$\begin{array}{c} \begin{array}{ccccc} n - 1 - \mathrm{wt}(\hat{\beta}) & \mu_1 & \mu_2 & \dots & \mu_\ell \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \dots & \cdot \end{bmatrix} \begin{array}{c} n - 1 - \mathrm{wt}(\hat{\alpha}) \\ \lambda_1 \\ \vdots \\ \lambda_k \end{array} \end{array}, \tag{5.2}$$

adding 1 to the top left coefficients provides a matrix of the form (5.1), which proves $N^n_{\Lambda(\hat{\alpha}),\Lambda(\hat{\beta})} \geqslant N^{n-1}_{\Lambda(\hat{\alpha}),\Lambda(\hat{\beta})}$.

In order to show the reverse inequality, we claim that the top left coefficient in (5.1) cannot be 0. Indeed, if it is 0, then the sum of the coefficients in the first row is at most

$$\mu_1 + \mu_2 + \cdots + \mu_\ell = \mathrm{wt}(\hat{\beta}).$$

This implies $n - \mathrm{wt}(\hat{\alpha}) \leqslant \mathrm{wt}(\hat{\beta})$, which forces $n \leqslant 2m$ and gives a contradiction.

Now, if the top left coefficient is a positive integer, subtracting 1 to this coefficient provides a matrix of the form (5.2), which proves $N^{n}_{\Lambda(\hat\alpha),\Lambda(\hat\beta)} \leqslant N^{n-1}_{\Lambda(\hat\alpha),\Lambda(\hat\beta)}$, and hence $V_n = V_{n-1}$. □

Building on this, we can actually show that for a large class of problems we have a stabilization.

Theorem 5.3. *Let $k, l, w \in \mathbb{N}$ be fixed. Then for every integer $n \geqslant 2w$ and every $\mathcal{S}_n$-invariant signomial $f \in \mathcal{C}(\mathcal{A},\mathcal{B})$ with $|\hat{\mathcal{A}}| \leqslant k$, $|\hat{\mathcal{B}}| \leqslant l$, and*

$$\max_{\hat\gamma \in \hat{\mathcal{A}}\cup\hat{\mathcal{B}}} \mathrm{wt}(\hat\gamma) \leqslant w,$$

the number of constraints and the number of variables of the symmetry adapted program are bounded by constants only depending on k, l and w:

$$C_n \leqslant k + l + l(w+1) \quad \text{and} \quad V_n \leqslant 2lku(w),$$

where $u(w) = \displaystyle\sum_{i=0}^{w} \binom{w}{i}^2 i!\,.$

Proof. Let us begin with the number of constraints. This follows from Theorem 5.2, because $|\hat{\mathcal{A}}| \leqslant k$, $|\hat{\mathcal{B}}| \leqslant l$ and $n_{\hat\beta} \leqslant w+1$, since $\mathrm{wt}(\hat\beta) \leqslant w$. As in the previous proof, we have $\Lambda(\hat\alpha)_1 = n - \mathrm{wt}(\hat\alpha)$ and similarly for $\hat\beta$. For the number of variables, we will show that

$$N_{\Lambda(\hat\alpha),\Lambda(\hat\beta)} \leqslant N_{(n-w,1^w),(n-w,1^w)} = u(w) \tag{5.3}$$

for every $\hat\alpha, \hat\beta$ satisfying the conditions of the theorem. This will be done in two steps. First, we show that if $\lambda = (\lambda_1, \ldots, \lambda_t, 1)$ is a partition, and $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_{t-1}, \lambda_t + 1)$, then for every partition μ, we have

$$N_{\lambda',\mu} \geqslant N_{\lambda,\mu} \quad \text{and} \quad N_{\mu,\lambda'} \geqslant N_{\mu,\lambda}.$$

Indeed, there is a surjection from the set $\mathcal{M}_{\lambda',\mu}$ onto $\mathcal{M}_{\lambda,\mu}$. Namely let $(x_1, \ldots, x_k)$ denote the t-th line of an element in $\mathcal{M}_{\lambda,\mu}$. Let s be such that $x_s > 0$. Replacing the t-th line of this element by $(x_1, \ldots, x_{s-1}, x_s - 1, x_{s+1}, \ldots x_k)$ and inserting $(0, \ldots, 0, 1, 0, \ldots, 0)$ as the $(t+1)$-th line we get an element in $\mathcal{N}_{\lambda',\mu}$. By applying this procedure recursively for rows and columns we get the inequality in (5.3).

To show the equality in (5.3), observe that the top-left element of $N_{(n-w,1^w),(n-w,1^w)}$has to be an integer k between $n-2w$ and $n-w$. For every such choice, we have to distribute $n-w-k$ ones in the first row and first column. This gives $\binom{w}{n-w-k}^2$ possibilities. Restricted to the $w \times w$ lower right submatrix, these selected lines and columns contain only 0. For each of these possibilities, after removing these chosen lines and columns, we get an $(n-k) \times (n-k)$ matrix which contains exactly one 1 per line and column. There are $(n-k)!$ such matrices. By a change of the index variable, we get the desired result:

$$V_n = 2 \sum_{\substack{\hat\alpha \in \hat{\mathcal{A}} \\ \hat\beta \in \hat{\mathcal{B}}}} {}_{\hat\alpha}(\mathcal{S}_n)_{\hat\beta} = 2 \sum_{\substack{\hat\alpha \in \hat{\mathcal{A}} \\ \hat\beta \in \hat{\mathcal{B}}}} N_{\Lambda(\alpha),\Lambda(\beta)} \leqslant 2lku(w).$$

□

We conclude this section by giving explicit estimates on the signomials where $|\hat{\mathcal{B}}| = 1$, and $\hat{\mathcal{A}} = \{0, \hat{\alpha}\}$. We have chosen four different classes of examples that show the influence of the sizes of the orbits on the numbers of variables and constraints. These classes represent extremal situations, namely when the orbits are either very large or very small. In these situations, we can actually compute the exact number of variables and constraints in both cases according to the previous discussions. Note that the last case falls into the framework of Theorem 5.2, where $\mathrm{wt}(\hat{\alpha}) = \mathrm{wt}(\hat{\beta}) = 1$. There is also a stabilization in the first sequence: when $\mathrm{len}(\hat{\beta}) = 1$, for every $\hat{\alpha} \in \hat{\mathcal{A}}$, the number of variables ${}_{\hat{\alpha}}(\mathcal{S}_n)_{\hat{\beta}}$ is equal to 1. The subsequent table summarizes our analysis. Specific signomials realizing the cases are given in Examples 6.1–6.4 in the next section.

		Standard method		**Symmetric method**		
$\lvert\mathcal{S}_n \cdot \hat{\beta}\rvert$	$\lvert\mathcal{S}_n \cdot \hat{\alpha}\rvert$	V_n	C_n	V_n	C_n	Example
1	$n!$	$2n! + 3$	$n! + n + 2$	5	4	6.1
$n!$	n	$2(n+1)n! + 1$	$(n+1)(n!+1)$	$2n + 3$	$n + 3$	6.2
$n!$	$n!$	$2(n!+1)n! + 1$	$n!(n+2) + 1$	$2n! + 3$	$n + 3$	6.3
n	n	$2n(n+1) + 1$	$(n+1)^2$	7	5	6.4

TABLE 1. Comparison of the parameters when $\hat{\mathcal{A}} = \{0, \hat{\alpha}\}$ and $\hat{\mathcal{B}} = \{\hat{\beta}\}$.

6. Numerical experiments

To illustrate the previous considerations, we present in this section classes of examples that spotlight the computational gains by the comparison of calculation times in the case of the symmetric group. For these computations, we used the MOSEK solver and Python 3.7 on an Intel(R) Xeon(R) Platinum 8168 CPU with 2.7 GHz and 768 GB of RAM under CentOS Linux release 7.9.2009. Keeping the previous notation, for the standard method, that is the method that does not exploit the symmetries, the input consists of $\mathcal{A}$, $\mathcal{B}$ as well as the coefficients, while for the symmetry-adapted version (Corollary 4.3), the input is $\hat{\mathcal{A}}$, $\hat{\mathcal{B}}$ and the coefficients. This difference of input is mainly due to practical considerations and does not in itself influence the comparison of the time used by the solver. When both methods give an answer, the bounds coincide.

In all the tables in the sequel, dim is the dimension, V_n and C_n are the number of variables and constraints of the program, while t_s and t_r denote the solver time and the overall running time (including the building of the optimization program) in seconds. These examples confirm that symmetry reduction can drastically decrease computation complexity, which results in faster computation and the possibility of solving larger problems.

The first four examples give numerical results for each of the classes discussed in Table 1. We tried to choose the coefficients in a way that avoids numerical issues, namely preventing the bound from being either too small or too large. We conducted the experiments until the number of inner or outer coefficients was too large to deal with in the standard program. The program using the symmetry reduction can even successfully carry out the experiments in dimensions beyond the ones listed in the tables. When further increasing dimensions, one has to be careful though, because some of the coefficients in the symmetry reduced constraints may endure a combinatorial explosion, which might cause some numerical issues.

Example 6.1. Consider first the signomial

$$f_n^{(1)} = n \sum_{\sigma \in \mathcal{S}_n} \sigma \exp(\langle \alpha, x \rangle) - n \exp(\langle \beta, x \rangle),$$

where $\beta = (1, \ldots, 1)$ and $\alpha = (1, 2, \ldots, n)$. The numerical results are shown in Table 2.

		Standard method				Symmetric method			
dim	bound	V_n	C_n	t_s	t_r	V_n	C_n	t_s	t_r
2	-0.0741	7	6	0.0458	0.0472	5	4	0.0479	0.0496
3	-0.1250	15	11	0.0666	0.0689	5	4	0.0459	0.0471
4	-0.1566	51	30	0.0690	0.0732	5	4	0.0679	0.0692
5	-0.1757	243	127	0.2005	0.2386	5	4	0.0641	0.0654
6	-0.1868	1443	728	1.045	1.174	5	4	0.0270	0.0280
7	-0.1929	10083	5049	14.84	15.67	5	4	0.0268	0.0278
8	-0.1956	80643	40330	236.2	242.7	5	4	0.0274	0.0283
9	-0.1962	725763	362891	24 774	24 837	5	4	0.0322	0.0332

TABLE 2. Numerical results for $f_n^{(1)}$.

Example 6.2. Consider now the signomial

$$f_n^{(2)} = \sum_{i=1}^{n} \exp(n^2 x_i) - \frac{1}{(n-1)!} \sum_{\sigma \in \mathcal{S}_n} \sigma \exp(\langle \beta, x \rangle),$$

where $\beta = (1, 2, \ldots, n)$ (and $\alpha = (n^2, 0, \ldots, 0)$). The numerical results are shown in Table 3.

		Standard method				Symmetric method			
dim	bound	V_n	C_n	t_s	t_r	V_n	C_n	t_s	t_r
2	-0.2109	13	9	0.0568	0.0585	7	5	0.0583	0.060
3	-0.4444	49	28	0.0893	0.0938	9	6	0.0438	0.046
4	-0.6853	241	125	0.2670	0.2830	11	7	0.0679	0.0717
5	-0.9295	1441	726	1.097	1.183	13	8	0.1028	0.1084
6	-1.176	10081	5047	8.276	8.866	15	9	0.1399	0.1605
7	-1.423	80641	40328	179.3	191.3	17	10	0.0781	0.0857
8	-1.670	725761	362889	16 709	17 744	19	11	0.1020	0.1111

TABLE 3. Numerical results for $f_n^{(2)}$.

Example 6.3. Next, we consider the case where both orbits are of maximal size. Let

$$f_n^{(3)} = \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \sigma \exp(\langle \alpha, x \rangle) - \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \sigma \exp(\langle \beta, x \rangle),$$

where $\beta = (1, 2, \ldots, n)$ and $\alpha = (2, 8, \ldots, 2n^2)$.

The numerical results are shown in Table 4.

		Standard method				Symmetric method			
dim	bound	V_n	C_n	t_s	t_r	V_n	C_n	t_s	t_r
2	-0.4178	13	9	0.0852	0.0873	7	5	0.0741	0.0766
3	-0.5162	85	31	0.0674	0.0751	15	6	0.0598	0.0629
4	-0.5824	1201	145	0.3074	0.3466	51	7	0.2249	0.2405
5	-0.6305	29041	841	12.81	13.51	243	8	0.7791	0.8642
6	-0.6675	1038241	5761	14 649	14 688	1443	9	4.303	4.818

TABLE 4. Numerical results for $f_n^{(3)}$.

Example 6.4. Finally, we consider the case where both orbits are small. Let

$$f_n^{(4)} = \sum_{i=1}^{n} \exp(n^2 x_i) - \sum_{i=1}^{n} \exp((n-1)(x_1 + \cdots + x_n) + x_i),$$

($\beta = (n, n-1, n-1, \ldots, n-1)$ and $\alpha = (n^2, 0, \ldots, 0)$). The numerical results are shown in Table 5.

		Standard method				Symmetric method			
dim	bound	V_n	C_n	t_s	t_r	V_n	C_n	t_s	t_r
10	-0.3468	221	121	0.2621	0.2772	7	5	0.0543	0.0552
20	-0.3580	841	441	0.2712	0.2924	7	5	0.0381	0.0392
30	-0.3615	1861	961	0.5142	0.5575	7	5	0.0375	0.0386
40	-0.3631	3281	1681	1.123	1.189	7	5	0.0381	0.0394
50	-0.3641	5101	2601	2.356	2.466	7	5	0.0371	0.0385
100	-0.3660	20201	10201	46.22	46.63	7	5	0.0365	0.0384
150	-0.3667	45301	22801	303.2	304.3	7	5	0.0454	0.0481
200	-0.3670	80401	40401	1530	1532	7	5	0.0565	0.0590
250	-0.3671	125501	63001	5358	5361	7	5	0.0335	0.0362
300	-0.3673	180601	90601	12 990	12 995	7	5	0.0476	0.0509
350	-0.3674	245701	123201	30 219	30 226	7	5	0.0513	0.0551

TABLE 5. Numerical results for $f_n^{(4)}$.

Example 6.5. Next, we give an example where $\mathcal{A}$ and $\mathcal{B}$ consist of two orbits each:

$$\hat{\mathcal{A}} = \{(n^2, 0, \ldots, 0), (1, 4, \ldots, n^2)\} \text{ and } \hat{\mathcal{B}} = \{(1, \ldots, 1), (1, 2, \ldots, n)\}.$$

In this case, we are still able to compute the number of constraints and the number of variables. With the standard approach,

$$V_n = 2(n! + n + 1)(n! + 1) + 1, \quad C_n = (n! + 1)(n + 2) + n,$$

while using symmetries,

$$V_n = 2n! + 2n + 9, \quad C_n = n + 6.$$

Table 6 shows the numerical results for the signomials

$$g_n = \frac{1}{n}\sum_{i=1}^{n} \exp(n^2 x_i) + \frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n} \sigma \exp(\langle\alpha, x\rangle) - \exp(x_1 + \cdots + x_n) - \frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n} \sigma \exp(\langle\beta, x\rangle)$$

for $\alpha = (1, 4, \ldots, n^2)$ and $\beta = (1, 2, \ldots, n)$.

		Standard method				**Symmetric method**			
dim	bound	V_n	C_n	t_s	t_r	V_n	C_n	t_s	t_r
2	-0.4311	31	14	0.1003	0.1027	17	8	0.1290	0.1326
3	-0.6643	141	38	0.1232	0.1291	27	9	0.1285	0.1375
4	-0.8070	1451	154	0.3944	0.4390	65	10	0.2275	0.2470
5	-0.9009	30493	852	14.88	15.55	259	11	0.6945	0.7689
6	-0.9660	1048335	5774	14 187	14 220	1461	12	3.832	4.281

Table 6. Numerical results for g_n.

Example 6.6. Finally, we consider the symmetric dwarfed signomial. The dwarfed polynomial is available on the accompanying web page of [33, Example 4.5]. This is a polynomial in seven variables, of total degree six, and degree four in each variable. We symmetrize it by applying the Reynolds operator with respect to the full symmetric group $\mathcal{S}_7$. The associated signomial (the symmetrized dwarfed signomial) has 113 non-zero coefficients. We have

$$\hat{\mathcal{A}} = \{(4, 2, 0, 0, 0, 0, 0), (4, 0, 0, 0, 0, 0, 0), (2, 2, 2, 0, 0, 0, 0)\}$$

and

$$\hat{\mathcal{B}} = \{(2, 2, 0, 0, 0, 0, 0), (2, 0, 0, 0, 0, 0, 0)\}.$$

The symmetry reduced program needs 0.191 seconds to find the bound, while the standard program needs 0.912 seconds.

7. Conclusion and open questions

We have developed techniques to exploit symmetries in AM/GM-based optimization and confirmed their benefit in terms of computational results. In particular, in the case of symmetric signomials, we showed that both theoretically as well as practically our orbit reduction allow for substantial computational gains. This motivates a theoretical study of the strength of the AM/GM bounds in this framework. In particular, it encourages the comparison of the symmetric SAGE cones with respect to the cone of symmetric non-negative signomials.

The orbit decomposition in Theorem 3.1 can also be used for SAGE and SONC decompositions obtained by second-order cone representations, which were studied [1, 27, 37]. It would be interesting to investigate symmetry exploitation in that context in further detail.

References

[1] G. Averkov. Optimal size of linear matrix inequalities in semidefinite approaches to polynomial optimization. *SIAM J. Appl. Algebra and Geometry*, 3(1):128–151, 2019.

[2] C. Bachoc, D. C. Gijswijt, A. Schrijver, and F. Vallentin. Invariant semidefinite programs. In *Handbook on Semidefinite, Conic and Polynomial Optimization*, pages 219–269. Springer, 2012.

[3] C. Bachoc and F. Vallentin. New upper bounds for kissing numbers from semidefinite programming. *J. Amer. Math. Soc.*, 21(3):909–924, 2008.

[4] E. R. Bazan and E. Hubert. Multivariate interpolation: Preserving and exploiting symmetry. *J. Symb. Comp.* 107:1–22, 2021.

[5] G. Blekherman, A. Raymond, M. Singh, and R. Thomas. Simple graph density inequalities with no sum of squares proofs. *Combinatorica*, 40(4):455–471, 2020.

[6] G. Blekherman and C. Riener. Symmetric non-negative forms and sums of squares. *Discrete Comput. Geom.* 65:764–799, 2021.

[7] V. Chandrasekaran and P. Shah. Relative entropy relaxations for signomial optimization. *SIAM J. Optim.*, 26(2):1147–1173, 2016.

[8] V. Chandrasekaran and P. Shah. Relative entropy optimization and its applications. *Math. Program., Ser. A*, 161(1-2):1–32, 2017.

[9] E. de Klerk and R. Sotirov. Exploiting group symmetry in semidefinite programming relaxations of the quadratic assignment problem. *Math. Program.*, 122(2):225, 2010.

[10] S. Debus and C. Riener. Reflection groups and cones of sums of squares. Preprint, arXiv:2011.09997, 2020.

[11] C. Dobre and J. Vera. Exploiting symmetry in copositive programs via semidefinite hierarchies. *Math. Program.*, 151(2):659–680, 2015.

[12] M. Dressler, J. Heuer, H. Naumann, and T. de Wolff. Global optimization via the dual SONC cone and linear programming. In *Proc. 45th International Symposium on Symbolic and Algebraic Computation*, pages 138–145, 2020.

[13] M. Dressler and R. Murray. Algebraic perspectives on signomial optimization. Preprint, arXiv:2107.00345, 2021.

[14] M. Dressler, S. Iliman, and T. de Wolff. An approach to constrained polynomial optimization via nonnegative circuit polynomials and geometric programming. *J. Symb. Comp.*, 91:149–172, 2019.

[15] T. Friedl, C. Riener, and R. Sanyal. Reflection groups, reflection arrangements, and invariant real varieties. *Proc. Amer. Math. Soc.*, 146(3):1031–1045, 2018.

[16] K. Gatermann and P. A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *J. Pure & Applied Algebra*, 192(1-3):95–128, 2004.

[17] A. Heaton, S. Hoşten, and I. Shankar. Symmetry adapted Gram spectrahedra. *SIAM J. Applied Algebra & Geometry*, 5:140–164, 2021.

[18] S. Iliman and T. de Wolff. Amoebas, nonnegative polynomials and sums of squares supported on circuits. *Res. Math. Sci.*, 3(paper no. 9), 2016.

[19] S. Iliman and T. de Wolff. Lower bounds for polynomials with simplex Newton polytopes based on geometric programming. *SIAM J. Optim.*, 26(2):1128–1146, 2016.

[20] O. Karaca, G. Darivianakis, P. Beuchat, A. Georghiou, and J. Lygeros. The REPOP toolbox: Tackling polynomial optimization using relative entropy relaxations. In *20th IFAC World Congress, IFAC PapersOnLine*, volume 50(1), pages 11652–11657. Elsevier, 2017.

[21] L. Katthän, H. Naumann, and T. Theobald. A unified framework of SAGE and SONC polynomials and its duality theory. *Math. Comput.*, 90:1297–1322, 2021.

[22] J. B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM J. Optim.*, 17:822–843, 2006
[23] P. Moustrou, C. Riener, and H. Verdure. Symmetric ideals, Specht polynomials and solutions to symmetric systems of equations. *J. Symb. Comp.*, 107:106–121, 2021.
[24] R. Murray, V. Chandrasekaran, and A. Wierman. Newton polytopes and relative entropy optimization. *Found. Comput. Math.*, 21:1703–1737, 2021.
[25] R. Murray, V. Chandrasekaran, and A. Wierman. Signomial and polynomial optimization via relative entropy and partial dualization. *Math. Program. Comput.*, 13:257–295, 2021.
[26] R. Murray, H. Naumann, and T. Theobald. Sublinear circuits and the constrained signomial optimization problem. Preprint, arXiv:2006.06811, 2020.
[27] H. Naumann and T. Theobald. The $\mathcal{S}$-cone and a primal-dual view on second-order representability. *Beiträge Algebra Geom.*, 62:229–249, 2021.
[28] C. Pantea, H. Koeppl, and G. Craciun. Global injectivity and multiple equilibria in uni- and bi-molecular reaction networks. *Discrete and Continuous Dynamical Systems - Series B*, 17(6):2153–2170, May 2012.
[29] A. Raymond, J. Saunderson, M. Singh, and R. R. Thomas. Symmetric sums of squares over k-subset hypercubes. *Math. Program., Ser. A*, 167(2):315–354, 2018.
[30] B. Reznick. Forms derived from the arithmetic-geometric inequality. *Math. Annalen*, 283(3):431–464, 1989.
[31] C. Riener. On the degree and half-degree principle for symmetric polynomials. *J. Pure & Applied Algebra*, 216(4):850–856, 2012.
[32] C. Riener, T. Theobald, L. Jansson-Andrén, and J. B. Lasserre. Exploiting symmetries in SDP-relaxations for polynomial optimization. *Math. Oper. Res.*, 38(1):122–141, 2013.
[33] H. Seidler and T. de Wolff. An experimental comparison of SONC and SOS certificates for unconstrained optimization. Preprint, arXiv:1808.08431, 2018. Accompanying Web page `https://www3.math.tu-berlin.de/combi/RAAGConOpt/comparison_paper/`.
[34] R. P. Stanley. *Enumerative Combinatorics*, volume 2. Cambridge University Press, 1999.
[35] V. Timofte. On the positivity of symmetric polynomial functions. Part I: General results. *J. Math. Analysis and Applications*, 284(1):174–190, 2003.
[36] H. Waki, S. Kim, M. Kojima, and M. Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optim.*, 17:218–242, 2006
[37] J. Wang and V. Magron. A second order cone characterization for sums of nonnegative circuits. In *Proc. 45th International Symposium on Symbolic and Algebraic Computation*, pages 450–457, 2020.
[38] J. Wang, V. Magron, and J. B. Lasserre. Chordal-TSSOS: a moment-SOS hierarchy that exploits term sparsity with chordal extension. *SIAM J. Optim.*, 31:114–141, 2021.
[39] J. Wang, V. Magron, and J. B. Lasserre. TSSOS: a moment-SOS hierarchy that exploits term sparsity. *SIAM J. Optim.*, 31:1–29, 2021.
[40] J. Wang, V. Magron, J. B. Lasserre, N. H. A. Mai. CS-TSSOS: Correlative and term sparsity for large-scale polynomial optimization. Preprint, arXiv:2005.02828, 2020.
[41] W. C. Waterhouse. Do symmetric problems have symmetric solutions? *Amer. Math. Monthly*, 90(6):378–387, 1983.

Philippe Moustrou: Institut de Mathématiques de Toulouse, UMR 5219, UT2J, 31058 Toulouse, France

Helen Naumann, Thorsten Theobald: FB 12 – Institut für Mathematik, Goethe-Universität, Postfach 11 19 32, 60054 Frankfurt am Main, Germany

Cordian Riener, Hugues Verdure: Department of Mathematics and Statistics, UiT – The Arctic University of Norway, 9037 Tromsø, Norway