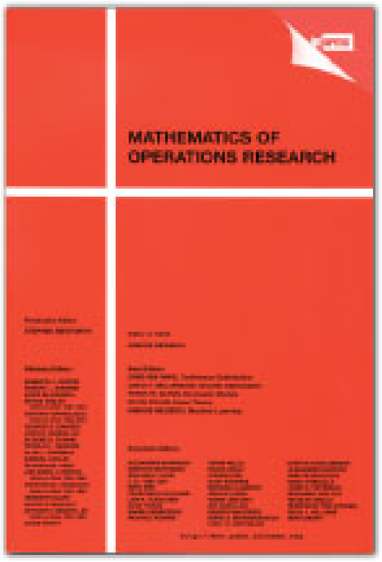

Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Exploiting Symmetries in SDP-Relaxations for Polynomial Optimization

Cordian Riener, Thorsten Theobald, Lina Jansson Andrén, Jean B. Lasserre,

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 38, No. 1, February 2013, pp. 122–141
ISSN 0364-765X (print) | ISSN 1526-5471 (online)

http://dx.doi.org/10.1287/moor.1120.0558

Exploiting Symmetries in SDP-Relaxations for Polynomial Optimization

Cordian Riener
Fachbereich Mathematik und Statistik, Universität Konstanz, 78457 Konstanz, Germany,
cordian.riener@uni-konstanz.de

Thorsten Theobald
FB 12 – Institut für Mathematik, Goethe-Universität, 60054 Frankfurt am Main, Germany,
theobald@math.uni-frankfurt.de

Lina Jansson Andrén
Department of Mathematics and Math. Statistics, Umeå Universitet, 901 87 Umeå, Sweden,
lina.andren@math.umu.se

Jean B. Lasserre
LAAS-CNRS and Institute of Mathematics, 31077 Toulouse Cedex 4, France,
lasserre@laas.fr

In this paper we study various approaches for exploiting symmetries in polynomial optimization problems within the framework of semidefinite programming relaxations. Our special focus is on constrained problems especially when the symmetric group is acting on the variables. In particular, we investigate the concept of block decomposition within the framework of constrained polynomial optimization problems, show how the degree principle for the symmetric group can be computationally exploited, and also propose some methods to efficiently compute the geometric quotient.

Key words: polynomial optimization; semidefinite programming; semidefinite relaxation symmetry; symmetric group; constrained optimization
MSC2000 subject classification: Primary: 90C22, 90C26, 14P05, 05E10
OR/MS subject classification: Primary: programming; secondary: nonlinear
History: Received March 2, 2011; revised June 26, 2012. Published online in *Articles in Advance* October 24, 2012.

1. Introduction. Solving or even computing lower bounds in constrained polynomial optimization is a difficult problem with important practical applications. In recent years, results of real algebraic geometry on the representations of positive polynomials have permitted us to define a hierarchy of semidefinite relaxations (SDP-relaxations) of these problems, which provides a monotonically nondecreasing sequence of lower bounds converging to the global minimum. See, e.g., Lasserre [21], Parrilo [27], or the survey by Laurent [26], and many references therein. However, the size of the resulting SDPs grows quickly with the problem size; typically, for an optimization problem in n variables the SDP-relaxation of order k in the hierarchy involves $O(n^{2k})$ variables and linear matrix inequalities (LMIs) of size $O(n^k)$. Therefore, and in view of the present status of SDP solvers, the applicability of the basic methodology is limited to small- or medium-sized problems unless some specific characteristics are taken into account.

One way to reduce this size limitation is to exploit *symmetries* when present in the problem definition. In the present paper, which has a foundational character, we consider the polynomial optimization problem

$$\begin{aligned} f^* = \inf \quad & f(x) \\ \text{s.t.} \quad & g_1(x) \geq 0, \dots, g_m(x) \geq 0, \end{aligned} \tag{1}$$

where $f, g_1, \dots, g_m \in \mathbb{R}[X_1, \dots, X_n]$. We assume that the polynomials are invariant by the action of a finite subgroup G of the group $\mathrm{GL}_n(\mathbb{R})$, i.e., $f(\sigma^{-1}(x)) = f(x)$ and $g_j(\sigma^{-1}(x)) = g_j(x)$ for all $\sigma \in G$, and all $j = 1, \dots, m$.

A major theoretical contribution to a systematic study of symmetries in real algebraic geometry was provided by Procesi and Schwarz [28] who gave a semi-algebraic description of the geometric quotient of a semi-algebraic set invariant under a group G (Bröcker [6]). For the special case of the symmetric group, Timofte [38] provided a very useful criterion for the nonnegativity of a polynomial. Cimprič et al. [7] studied foundational aspects of the dual problem of moments.

The systematic study of block diagonalizations of SDPs was initiated by Gatermann and Parrilo [14] (in the context of symmetries) and by Schrijver [33, 34] (in the general framework of matrix $*$-algebras). Building on this, de Klerk et al. [11] have provided a general method (the $*$-representation) to handle symmetries of any semidefinite program (see also Gijswijt [15], Kanno et al. [20], and Laurent [24]). Excellent reviews of the above are the surveys by Bachoc et al. [2] and Vallentin [39].

Contribution. In the present paper, we advance these lines of research in several ways:

1. We provide a systematic treatment of the block diagonalization in the setting of Lasserre's relaxation which is concerned with *constrained* optimization. Instead of considering a general SDP framework, we rather focus attention on the specific SDPs coming from the relaxation scheme defined in Lasserre [21]. Indeed, the symmetries on the original variables of the optimization problem induce specific additional symmetry structure on the moment and localizing matrices of the SDP-relaxation. To this end we suggest that a symmetry-adapted version of the relaxation scheme can be defined directly using an appropriate basis for the moments and derive symmetric versions of Putinar's theorems (see Theorems 3.2 and 3.5). We study a possible basis (generalized Specht polynomials as defined in §4.1) in detail for the case of the symmetric group $\mathcal{S}_n$. In this situation we show that for k fixed, the number and sizes of the LMIs in the SDP-relaxation of order k are bounded by a constant that does *not* depend on the number n of variables (Theorem 4.7). As a direct consequence, we can state some symmetric versions of representation theorems for sums of squares, in particular for the "Hilbert cases" (Theorem 4.10 and Corollaries 4.11–4.13).

2. We show how the so-called degree principle (Riener [30], Timofte [38]) can be used to transform an $\mathcal{S}_n$-invariant optimization problem into a set of lower-dimensional problems and that in some cases the resulting relaxation scheme converges finitely (Theorem 5.4). This gives a sum-of-squares-based criterion to certify nonnegativity of an $\mathcal{S}_n$-symmetric polynomial of degree 4 (Theorem 5.5).

3. We show how the geometric quotient viewpoint naturally leads to a polynomial matrix inequality (PMI) problem. For certain power sum problems (generalizing a situation studied by Brandenberg and Theobald [5]), we discuss how this leads to lower and upper bounds which can be computed quite simply (Theorems 6.6 and 6.7).

Our techniques enlarge the techniques for handling constrained optimization problems with symmetries. We feel that it is worth presenting them in a common context. We focus to a large extent on the case of the group $\mathcal{S}_n$. This is for several reasons. First, the problems that motivated the research leading to this paper came from this setting. Second, it turns out that in the situation of symmetric polynomials the complexity of the optimization problem as a function of the number of variables can be dramatically reduced with all the techniques we provide. Moreover, the symmetric group serves as a rich example to demonstrate general principles, and we remark that many combinatorial optimization problems can be put into the form of maximizing a given linear form on the orbit of a vector in a representation of the symmetric group (see Barvinok and Vershik [4] and Barvinok [3]). Whereas (on the SDP level) the general framework of block diagonalization is already well understood, the degree principle still awaits its generalization for other groups.

The authors are aware that certain algebraic techniques used in this paper might not be very familiar to optimizers in general, which may induce doubts about any real systematic implementation in some fully automized software, at least in the near future. However, there are also reasons to be more optimistic in view of the growing interest in semidefinite relaxations for polynomial optimization, and their current limitation to problems of modest size only, if no sparsity or symmetry is taken into account.

The paper is structured as follows. In §2, we give a short introduction to the SDP relaxation scheme and to PMIs. Furthermore, we introduce some representation theoretical notions, with special focus on the symmetric group $\mathcal{S}_n$. In §3 we give a systematic treatment of how invariance by a finite group can be exploited in the relaxation scheme introduced in Lasserre [21]. Section 4 is devoted to a study of optimization with symmetric polynomials. We give a detailed construction of the related moment matrices. From the constructions we then deduce representation statements for symmetric positive polynomials. In §5 we show how it is possible to use the degree principle to break some of the symmetry and thereby construct a family of lower-dimensional problems, which can be used to solve the original optimization problem.

Finally, in §6 we show how optimization problems described by invariant polynomials can be treated in the orbit space. As a direct application of this procedure we can show how to calculate bounds for a specific class of problems.

2. Preliminaries.

Let $\mathbb{R}[X]$ be the ring of polynomials in the variables $X = (X_1, \ldots, X_n)$, and let $\mathbb{R}[X]_{\le k}$ be the subset of polynomials of degree at most k. In the following subsections, we recall Lasserre's relaxation scheme for polynomial optimization, polynomial matrix inequalities (PMIs), and some basic concepts of representation theory.

2.1. Lasserre's method.

Given polynomials $f, g_1, \ldots, g_m \in \mathbb{R}[X]$, consider the general optimization problem of the form

$$f^* = \inf f(x) \qquad \text{subject to } g_1(x) \ge 0, \ldots, g_m(x) \ge 0.$$

Its feasible set $K \subseteq \mathbb{R}^n$ is the basic closed semialgebraic set

$$K := \{x \in \mathbb{R}^n \colon g_j(x) \geq 0,\ j = 1, \ldots, m\}. \tag{2}$$

Lasserre [21] has introduced the following hierarchy of semidefinite relaxations (see also Lasserre [22] and Laurent [26]). For reasons described below we will need the following technical assumption:

Assumption 2.1. *The feasible set K defined in (2) is compact, and there exists a polynomial $u \in \mathbb{R}[X]$ such that the level set $\{x \in \mathbb{R}^n \colon u(x) \geq 0\}$ is compact and u has the representation*

$$u = u_0 + \sum_{j=1}^{m} u_j g_j \tag{3}$$

for some sums of squares polynomials $u_0, u_1, \ldots, u_m \in \mathbb{R}[X]$.

Assumption 2.1 holds if, e.g., for some $j \in \{1, \ldots, m\}$, the level set $\{x \in \mathbb{R}^n \colon g_j(x) \geq 0\}$ is compact, or if K is compact and all the g_j's are affine (in which case K is a polytope). In particular, Assumption 2.1 holds if and only if for some $N \in \mathbb{N}$, the polynomial $N - \sum_{i=1}^n X_i^2$ can be written in the form (3); equivalently, this polynomial belongs to the quadratic module generated by the g_j's. For a comprehensive discussion of the last condition see Schweighofer [32, Theorem 1]. Notice that under Assumption 2.1, K is compact and thus the infimum f^* is attained on K.

The idea is to convexify the problem by considering the equivalent formulation

$$f^* = \min_{x \in K} f(x) = \min_{\mu \in \mathcal{P}(K)} \int f \, d\mu, \tag{4}$$

where $\mathcal{P}(K)$ denotes the set of all probability measures μ supported on the set K. These measures are characterized by the following theorem.

Theorem 2.2 (Putinar [29]). *Suppose Assumption 2.1 holds for the set K. A linear map $L \colon \mathbb{R}[X] \to \mathbb{R}$ is the integration with respect to a probability measure μ on K, i.e.,*

$$\exists \mu \in \mathcal{P}(K) \ \ \forall p \in \mathbb{R}[X] \qquad L(p) = \int p \, d\mu,$$

if and only if $L(1) = 1$ and $L(s_0 + \sum_{j=1}^m s_j g_j) \geq 0$ for any sum of squares polynomials $s_0, \ldots, s_m \in \mathbb{R}[X]$.

Setting $g_0 := 1$, the condition in Putinar's result is satisfied if and only if the bilinear forms $\mathcal{L}_{g_0}, \ldots, \mathcal{L}_{g_m}$ defined by

$$\begin{aligned} \mathcal{L}_{g_j} \colon \mathbb{R}[X] \times \mathbb{R}[X] &\to \mathbb{R}, \\ (p, q) &\mapsto L(p \cdot q \cdot g_j), \end{aligned}$$

are positive semidefinite (psd). With this characterization we can restate (4) as

$$f^* = \min\{L(f) \colon L \colon \mathbb{R}[X] \to \mathbb{R} \text{ linear},\ L(1) = 1 \text{ and each } \mathcal{L}_{g_j} \text{ is psd}\}. \tag{5}$$

Now fix any basis $\mathcal{B}$ of the vector space $\mathbb{R}[X]$ with $1 \in \mathcal{B}$ (for example the monomial basis X^α). For any linear map $L \colon \mathbb{R}[X] \to \mathbb{R}$ with $L(1) = 1$, setting $y_b = L(b)$ for $b \in \mathcal{B}$ identifies L with an infinite series $y = (y_b)_{b \in \mathcal{B}}$ of real numbers indexed by the elements of $\mathcal{B}$. The infinite-dimensional *moment matrix* M associated to y is indexed by $\mathcal{B}$ and given by

$$M(y)_{u,v} := L(u \cdot v), \quad u, v \in \mathcal{B}.$$

Furthermore, for each g_j, define in an analogous manner the *localizing matrix* $M(g_j y)$ by

$$M(g_j y)_{u,v} := L(u \cdot v \cdot g_j), \quad u, v \in \mathcal{B}.$$

Under Assumption 2.1, a given sequence y comes from some measure μ supported on K if and only if the moment matrix as well as the localizing matrices are psd. For practical applications of this approach, truncated versions of (5) have to be considered: Let $k \geq k_0 := \max\{\lceil \deg f/2 \rceil, \lceil \deg g_1/2 \rceil, \ldots, \lceil \deg g_m/2 \rceil\}$. Define the finite-dimensional matrix M_k by considering only rows and columns indexed by elements in $\mathcal{B}$ of degree at most k, and consider the hierarchy of semidefinite relaxations

$$Q_k\colon \quad \begin{aligned} &\inf_y L(f)\\ &M_k(y) \succeq 0,\\ &M_{k-\lceil \deg g_j/2\rceil}(g_j y) \succeq 0, \quad 1 \le j \le m,\\ &y_1 = 1, \end{aligned} \tag{6}$$

with optimal value denoted by $\inf Q_k$ (and $\min Q_k$ if the infimum is attained).

Although each of the relaxation values might not be optimal for the original problem, one has the following convergence result.

Proposition 2.3 (Lasserre [21]). *Let Assumption 2.1 hold and consider the hierarchy of SDP-relaxations $(Q_k)_{k\ge k_0}$ defined in (6). Then the sequence $(\inf Q_k)_{k\ge k_0}$ is monotonically nondecreasing and converges to f^*; that is, $\inf Q_k \uparrow f^*$ as $k \to \infty$.*

Although there are sufficient conditions to decide whether an optimal value has been reached after a certain iteration (see, for example, Henrion and Lasserre [16] and Lasserre [22]), in general only in some situations can finite convergence be guaranteed:

Proposition 2.4 (Laurent [25]). *Let $f, g_1, \ldots, g_m \in \mathbb{R}[X]$ and consider the problem*

$$\inf_{x\in\mathbb{R}^n} \{f(x)\colon g_1(x) = \cdots = g_m(x) = 0\}. \tag{7}$$

If the ideal generated by $g_1, \ldots, g_m$ is zero-dimensional, then the Lasserre relaxation scheme of (7) has finite convergence; i.e., there is an $l \ge k_0$ such that $\inf Q_l = f^*$.

2.2. Polynomial matrix inequalities. An interesting case of a polynomial optimization problem, which will be relevant for some of our approaches, arises when dealing with positive semidefiniteness of a matrix whose entries are polynomials.

Let S_m denote the set of real symmetric $m\times m$-matrices. A *polynomial matrix inequality* (*PMI*) optimization problem is an optimization problem of the form

$$\begin{aligned} f^* = \inf_{x\in\mathbb{R}^n} \quad & f(x)\\ \text{s.t.} \quad & G(x) \succeq 0, \end{aligned}$$

where $f \in \mathbb{R}[X]$ and $G(X)$ is a symmetric $m\times m$-matrix whose entries $G_{ij}(X)$ are polynomials in X.

By considering the psd condition on $G(x)$ as polynomial constraints and using the approach from §2.1, one would have to deal with polynomials of large degree. Even if all $G_{ij}(X)$ are linear, for example, the polynomial inequalities one needs to consider are of degree m. This high degree could make it hard even to explicitly calculate the first possible relaxation.

To overcome this problem, SDP hierarchies were proposed in Henrion and Lasserre [17] and Hol and Scherer [18] that take into account the semidefiniteness of a polynomial matrix. The basic idea is to generalize the standard approach in a suitable way by defining a localizing matrix for the matrix $G(X)$. This (infinite) matrix consists of blocks that are indexed by the elements of a basis $\mathcal{B}$ of $\mathbb{R}[X]$, the entries of each block are indexed with the entries of $G(X)$, and the entry i, j of the block corresponding to $u, v \in \mathcal{B}$ is

$$M(Gy)^{u,v}_{i,j} := L(u\cdot v\cdot G_{ij}(X)), \quad u, v\in\mathcal{B}, \quad i, j \in \{1, \ldots, m\}.$$

Setting $d := \max\{\lceil \deg G_{ij}(X)/2\rceil\}$ and $k \ge k_0 := \max\{\lceil \deg f/2\rceil, d\}$, one can define a relaxation

$$Q_k\colon \quad \begin{aligned} &\inf_y L(f)\\ &M_k(y) \succeq 0,\\ &M_{k-d}(G\,y) \succeq 0, \end{aligned} \tag{8}$$

with the truncated matrix M_k of M. To guarantee the convergence of this relaxation one needs to assume the Putinar condition viewed in this setting:

Assumption 2.5. *Suppose that there is $u \in \mathbb{R}[X]$ such that the level set $\{x \in \mathbb{R}^n: u(x) \geq 0\}$ is compact and u has the representation*

$$u = u_0 + \langle R(X), G(X) \rangle \tag{9}$$

for some sum of squares polynomials $u_0 \in \mathbb{R}[X]$ and a symmetric sum of squares matrix $R(X) \in \mathbb{R}[X]^{m \times m}$.

Then the following convergence statement holds (Henrion and Lasserre [17]).

Proposition 2.6. *If $G(X)$ meets the Assumption 2.5, then the sequence $(\inf Q_k)_{k \geq k_0}$ is monotonically non-decreasing and converges to f^*.*

2.3. Linear representation theory. We collect some notions from linear representation theory. As standard reference see Serre [35]. For our purposes, we always assume that G is a finite group and that $\mathbb{K}$ is either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers.

A *representation of* G is a finite-dimensional vector space V over $\mathbb{K}$ together with a group homomorphism $\rho: G \to \mathrm{GL}(V)$ into the set $\mathrm{GL}(V)$ of invertible linear transformations of V. If a basis for V is chosen, then the representation can be expressed as a group homomorphism into the group $\mathrm{GL}_n(\mathbb{K})$, where $n := \dim V$. This is known as a *matrix representation*. The action of G turns V into a *G-module*, and in fact, the notion of a representation of G and the notion of a G-module are equivalent and can be identified. Two representations (V, ρ) and (V', ρ') of the same group G are *equivalent* if there is a linear isomorphism $\phi: V \to V'$ such that $\rho'(\sigma) = \phi\rho(\sigma)\phi^{-1}$ for all $\sigma \in G$.

Example 2.7. (1) The one-dimensional representation $(\mathrm{id}, \mathbb{K})$ (i.e., $V = \mathbb{K}$ and group action $\sigma(v) = v$ for all $\sigma \in G$ and $v \in \mathbb{K}$) is called the *trivial representation*.

(2) Take any set S on which G operates and set $V = \bigoplus_{s \in S} \mathbb{C}e_s$ with formal symbols e_s $(s \in S)$. Then the obvious action of G on V defined via $\sigma(e_s) = e_{\sigma(s)}$ turns V into a G-module. In the special case when $S = G$ this is called the *regular representation*.

If there is a proper G-submodule W of V (i.e., a G-invariant proper subspace W of V), then the representation (ρ, V) is called *reducible*. If, however, the only G-invariant subspaces are V and $\{0\}$, then (ρ, V) is called *irreducible*. For a finite group G, any representation (V, ρ) decomposes as a direct sum of irreducible representations, $\rho = \bigoplus_{i=1}^h m_i \rho_i$ with multiplicities m_i. This induces an *isotypic decomposition* of the G-module V as a direct sum $V = \bigoplus_{i=1}^h V_i$ with *isotypic components* $V_i = \bigoplus_{j \in J_i} W_{ij}$, finite index sets J_i with $|J_i| = m_i$, and, for fixed i, pairwise isomorphic, irreducible components W_{ij} (where each component W_{ij} is associated with the irreducible representation ρ_i). Whereas the decomposition in isotypic components is unique, the decomposition of the V_i is in general not unique.

The following basic but fundamental result of representation theory will be very useful (see, e.g., Serre [35, Proposition 4]).

Lemma 2.8 (Schur's Lemma). *Let (ρ_1, V) and (ρ_2, W) be two irreducible representations of a group G (over $\mathbb{R}$ or $\mathbb{C}$). Then every G-homomorphism $V \to W$ is either zero or an isomorphism. Moreover, if two complex irreducible representations (ρ_1, V) and (ρ_2, W) are isomorphic, then the vector space of all G-homomorphisms $V \to W$ has dimension* 1.

When working with real representations a little precaution is necessary and it can be useful to pass to the complexification of the real representation. For a real irreducible representation (ρ, V) the following three types are distinguished (see Serre [35, Section 13.2]): If the complexification $V \otimes \mathbb{C}$ is also irreducible (type **I**), then statements such as the second part of Lemma 2.8 directly transfer from $V \otimes \mathbb{C}$ to V. However, it may occur that a real irreducible representation (ρ, V) becomes reducible when passing to the complexification. In this case, $V \otimes \mathbb{C}$ will decompose into two complex-conjugate irreducible G-submodules V_1 and V_2. Since V_1 and V_2 are complex conjugates we can "virtually" keep track of this decomposition by decomposing $V \otimes \mathbb{C}$ as $V_1 + V_2 \oplus (1/\mathrm{i})(V_1 - V_2)$, where i is the imaginary unit. This is a real basis of V, which respects the decomposition of $V \otimes \mathbb{C}$. Either the G-submodules V_1 and V_2 are nonisomorphic (type **II**) or they are isomorphic (type **III**).

Let V be a real G-module. If the complexification $V \otimes \mathbb{C}$ has the isotypic decomposition $V \otimes \mathbb{C} = V_1 \oplus \cdots \oplus V_{2l} \oplus V_{2l+1} \oplus \cdots \oplus V_h$, where each pair (V_{2j-1}, V_{2j}) is complex conjugate $(1 \leq j \leq l)$ and $V_{2l+1}, \ldots, V_h$ are real, the decomposition

$$V = (V_1 + V_2) \oplus \frac{1}{\mathrm{i}}(V_1 - V_2) \oplus \cdots \oplus (V_{2l-1} + V_{2l}) \oplus \frac{1}{\mathrm{i}}(V_{2l-1} - V_{2l}) \oplus V_{2l+1} \oplus \cdots \oplus V_h \tag{10}$$

is called a *real decomposition*.

2.4. Symmetric group. An important special case is when G is the symmetric group $\mathcal{S}_n$ on n variables. We collect some well-known facts on the irreducible representations of $\mathcal{S}_n$. For a general reference we refer to Sagan [31].

For $n \geq 1$, a *partition* λ of n (written $\lambda \vdash n$) is a sequence of weakly decreasing positive integers $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ with $\sum_{i=1}^{l} \lambda_i = n$. For two partitions $\lambda, \mu \vdash n$, we write $\lambda \trianglerighteq \mu$ if $\lambda_1 + \cdots + \lambda_i \geq \mu_1 + \cdots + \mu_i$ for all i. A *Young tableau* for $\lambda \vdash n$ consists of l rows, with λ_i entries in the ith row. Each entry is an element in $\{1, \ldots, n\}$, and each of these numbers occurs exactly once. A *standard Young tableau* is a Young tableau in which all rows and columns are increasing.

Example 2.9. For the partition $\lambda = (4, 3, 1, 1, 1) \vdash 10$, an example of a Young tableau is

$$\begin{array}{|c|c|c|c|}\hline 1 & 3 & 4 & 6 \\ \hline 5 & 7 & 8 & \multicolumn{1}{c}{} \\ \cline{1-3} 9 & \multicolumn{3}{c}{} \\ \cline{1-1} 2 & \multicolumn{3}{c}{} \\ \cline{1-1} 10 & \multicolumn{3}{c}{} \\ \cline{1-1} \end{array} .$$

An element $\sigma \in \mathcal{S}_n$ acts on a Young tableau by replacing each entry by its image under σ. Two Young tableaux t_1 and t_2 are called *row equivalent* if the corresponding rows of the two tableaux contain the same numbers. The classes of equivalent Young tableaux are called *tabloids*, and the class of a tableau t is denoted by $\{t\}$. Let $\{t\}$ be a λ-tabloid. The action of $\mathcal{S}_n$ gives rise to an $\mathcal{S}_n$-module:

Definition 2.10. Suppose $\lambda \vdash n$. The *permutation module M^λ corresponding to* λ is the $\mathcal{S}_n$-module defined by $M^\lambda = \operatorname{span}\{\{t_1\}, \ldots, \{t_l\}\}$, where $\{t_1\}, \ldots, \{t_l\}$ is a complete list of λ-tabloids.

Example 2.11. If $\lambda = (1, 1, \ldots, 1) \vdash n$, then M^λ is isomorphic to the regular representation of $\mathcal{S}_n$ from Example 2.7. In case $\lambda = (2, 1)$ a complete list of λ-tabloids is given by the representatives

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \qquad \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} . \tag{11}$$

Let t be a Young tableau for $\lambda \vdash n$, and let $\mathcal{C}_1, \ldots, \mathcal{C}_\nu$ be the columns of t. The group $\mathrm{CStab}_t = \mathcal{S}_{\mathcal{C}_1} \times \mathcal{S}_{\mathcal{C}_2} \times \cdots \times \mathcal{S}_{\mathcal{C}_\nu}$ (where $\mathcal{S}_{\mathcal{C}_i}$ is the symmetric group on $\mathcal{C}_i$) is called the *column stabilizer* of t.

The irreducible representations of the symmetric group $\mathcal{S}_n$ are in one-to-one correspondence with the partitions of n, and they are given by the Specht modules, as explained in the following.

For $\lambda \vdash n$, the *polytabloid associated with* t is defined by

$$e_t = \sum_{\sigma \in \mathrm{CStab}_t} \operatorname{sgn}(\sigma)\sigma\{t\}, \tag{12}$$

where $\operatorname{sgn}(\sigma)$ denotes the signum of σ. Then for a partition $\lambda \vdash n$, the *Specht module* S^λ is the submodule of the permutation module M^λ spanned by the polytabloids $\{e_t\colon t \text{ Young tableau for } \lambda \vdash n\}$. The dimension of S^λ is given by the number of standard Young tableaux for $\lambda \vdash n$.

Example 2.12. For $n \geq 2$, we have the decomposition into irreducible components $M^{(n-1,1)} = S^{(n)} \oplus S^{(n-1,1)}$. Namely, since the one-dimensional subspace spanned by the sum $t_1 + \cdots + t_n$ is closed under the action of $\mathcal{S}_n$, we have a copy of the *trivial* representation (which is isomorphic to the Specht module $S^{(n)}$) as an irreducible component in $M^{(n-1,1)}$. Moreover, since the tabloids in (11) are completely determined by the entry in the second row, we have identified a copy of the $(n-1)$-dimensional Specht module $S^{(n-1,1)}$ in $M^{(n-1,1)}$. Indeed, the permutation module $M^{(n-1,1)}$ decomposes as $M^{(n-1,1)} = S^{(n)} \oplus S^{(n-1,1)}$.

The decomposition of the module M^λ for a general partition $\lambda \vdash n$ will be of special interest for us. It can be described in a rather combinatorial way as follows:

Definition 2.13. (1) A generalized *Young tableau* of *shape* λ is a Young tableau T for λ such that the entries are replaced by any n-tuple of natural numbers. The *content* of T is the sequence μ such that μ_i is equal to the number of i's in T.

(2) A generalized Young tableau is called *semistandard* if its rows weakly increase and its columns strictly increase.

(3) For $\lambda, \mu \vdash n$ the *Kostka number* $K_{\lambda\mu}$ is defined as the number of semistandard Young tableaux of shape λ and content μ.

The significance of these definitions lies in the following statement which originates from Young's work:

PROPOSITION 2.14. *For a partition $\mu \vdash n$, the permutation module M^μ can be decomposed as*

$$M^\mu = \bigoplus_{\lambda \trianglerighteq \mu} K_{\lambda\mu} S^\lambda.$$

3. Symmetry-adapted relaxation. As a first possibility to exploit symmetries in the framework of polynomial optimization, we consider the relaxation scheme introduced in §2.1. Throughout the section, let f, $g_1, \ldots, g_m \in \mathbb{R}[X]$, and K be the feasible set (2).

The relaxation scheme yields a sequence of semidefinite programs. While it is one possibility to exploit the symmetries on the level of the SDP, the focus in this paper is to approach it on the level on top of this, i.e., on the level of the polynomials. To the best of our knowledge, in the framework of the relaxation scheme, no detailed investigation has been made on the use of other polynomial bases (different from the standard monomial basis). Here, the formulation of the relaxation scheme in the symmetry-adapted basis will yield us symmetry-adapted versions of Putinar's Theorems (Theorems 3.2 and 3.5) and a symmetry-adapted relaxation scheme that converges (Theorem 3.7).

In the following, assume that G is a finite group, although most of the results could be generalized to compact groups. We start by considering linear group actions of G on $\mathbb{R}^n$: For a set $S \subseteq \mathbb{R}^n$ and $\sigma \in G$, let $\sigma(S) = \{x \in \mathbb{R}^n \colon \sigma^{-1}(x) \in S\}$. By setting $p^\sigma(x) := p(\sigma^{-1}(x))$ for any polynomial $p \in \mathbb{R}[X]$ and $x \in \mathbb{R}^n$, G induces a group action on $\mathbb{R}[X]$. If $p^\sigma = p$ for all $\sigma \in G$, then the polynomial p is called *G-invariant*.

Following §2, we now consider the probability measures $\mathcal{P}(S)$ supported on a set S. For $\mu \in \mathcal{P}(S)$, $\sigma \in G$, and $\sigma^{-1}(S) \subseteq S$, define μ^σ by $\mu^\sigma(B) = \mu(\sigma^{-1}(B))$ for any Borel set $B \subseteq S$. A measure $\mu \in \mathcal{P}(S)$ is said to be *G-invariant* if $\mu = \mu^\sigma$ for all $\sigma \in G$, and the subset of all G-invariant probability measures on S is denoted by $\mathcal{P}(S)^G$. For a comprehensive foundational treatment of invariant measures we refer to Cimprič et al. [7]. Here, we mainly need the subsequent simple connection, where for a set $S \subseteq \mathbb{R}^n$ we define $S^G = \bigcap_{\sigma \in G} \sigma(S)$. A set S is called *G-invariant* if $S = S^G$. Note, however, that a G-invariant feasible set K does not necessarily require that any of its defining polynomials g_j be G-invariant.

LEMMA 3.1. *Let the feasible set $K \subseteq \mathbb{R}^n$ be G-invariant. If $h \in \mathbb{R}[X]$ is G-invariant, then*

$$\inf_{x \in K} h(x) = \inf_{\mu \in \mathcal{P}^G(K)} \int_K h \, d\mu.$$

PROOF. We have

$$h^* := \inf_{x \in K} h(x) = \inf_{\mu \in \mathcal{P}(K)} \int h \, d\mu \le \inf_{\mu \in \mathcal{P}(K^G)} \int h \, d\mu, \quad \text{as } \mathcal{P}(K)^G \subseteq \mathcal{P}(K).$$

Let (x_k) be a minimizing sequence in K such that $h(x_k) \to h^*$ as $k \to \infty$. To each x_k we can define a Dirac measure μ_k supported in x_k. Now this gives a converging sequence $(\int h(x)\, d\mu_k)$. The measure $\mu_k^* := (1/|G|) \sum_{\sigma \in G} \mu_k^\sigma$ is contained in $\mathcal{P}(K)^G$ for every k. Since h is G-invariant, $\int h(x)\, d\mu_k^* = h(x_k)$, which in turn implies $\int h(x)\, d\mu_k^* \to h^* \le \inf_{\mu \in \mathcal{P}(K)^G} \int f \, d\mu$, and so $h^* = \inf_{\mu \in \mathcal{P}(K)^G} \int f \, d\mu$. $\square$

So, to find the infimum of a G-invariant function f on a G-invariant set K we only have to consider the invariant measures supported on K. Hence to make a relaxation scheme for this setting similar to the one presented in §2.1, it suffices to consider *G-linear* maps $L^G \colon \mathbb{R}[X] \to \mathbb{R}$, i.e., linear maps with $L^G(f) = L^G(f^\sigma)$ for all $f \in \mathbb{R}[X]$ and $\sigma \in G$. In analogy to Putinar's Theorem 2.2 we can also characterize them in terms of bilinear forms.

THEOREM 3.2. *Let $g_1, \ldots, g_m \in \mathbb{R}[X]$ be G-invariant, and assume that the feasible set K satisfies Assumption 2.1. Setting $g_0 := 1$, a G-linear map $L^G \colon \mathbb{R}[X] \to \mathbb{R}$ is the integration with respect to a G-invariant measure on K if and only if the bilinear forms*

$$\begin{aligned} \mathcal{L}_{g_j}^G \colon \mathbb{R}[X] \times \mathbb{R}[X] &\to \mathbb{R} \\ (p, q) &\mapsto L^G\left(\frac{1}{|G|} \sum_{\sigma \in G} (p \cdot q)^\sigma \cdot g_j \right) \end{aligned} \tag{13}$$

are psd for all $0 \le j \le m$.

Note that the definition of $\mathcal{L}_{g_j}^G$ only uses the values of L^G on the invariant ring $\mathbb{R}[X]^G \subseteq \mathbb{R}[X]$. Indeed, any G-linear map $L^G \colon \mathbb{R}[X] \to \mathbb{R}$ is already completely determined by its values on $\mathbb{R}[X]^G$, since for any $p \in \mathbb{R}[X]$ we have $L(p) = (1/|G|) \sum_{\sigma \in G} L^G(p^\sigma) = L^G((1/|G|) \sum_{\sigma \in G} p^\sigma)$.

Proof. *Only if part.* Let $L_\mu\colon \mathbb{R}[X] \to \mathbb{R}$ be the linear functional associated with a G-invariant measure μ, $L_\mu(f) := \int_K f\,d\mu$. Then

$$\begin{aligned} L_\mu\left(\frac{1}{|G|}\sum_{\sigma\in G}(f^2)^\sigma \cdot g_j\right) &= L_\mu\left(\frac{1}{|G|}\sum_{\sigma\in G}(f^2\cdot g_j)^\sigma\right) = \frac{1}{|G|}\sum_{\sigma\in G}\int_K (f^2\cdot g_j)^\sigma\,d\mu \\ &= \frac{1}{|G|}\sum_{\sigma\in G}\int_K (f^2\cdot g_j)\,d\mu^\sigma \\ &= \frac{1}{|G|}\sum_{\sigma\in G}\int_K (f^2\cdot g_j)\,d\mu \quad \text{as } \mu^\sigma = \mu \\ &\geq 0 \quad \text{as } g_j \geq 0 \text{ on } K. \end{aligned}$$

If part. Conversely, let L^G be as in the statement of the theorem. Then for all $h \in \mathbb{R}[X]$,

$$\begin{aligned} 0 \leq \mathscr{L}^G_{g_j}(h^2) = L^G\left(\frac{1}{|G|}\sum_{\sigma\in G}(h^2)^\sigma g_j\right) &= L^G\left(\frac{1}{|G|}\sum_{\sigma\in G}(h^2 g_j)^\sigma\right) \\ &= \frac{1}{|G|}\sum_{\sigma\in G} L^G((h^2 g_j)^\sigma) \\ &= \frac{1}{|G|}\sum_{\sigma\in G} L^G(h^2 g_j) \quad \text{as } L^G \text{ is } G\text{-invariant.} \end{aligned}$$

Since this holds for all $h \in \mathbb{R}[X]$ and all $j = 1, \ldots, m$, by Putinar's Theorem L^G is the integration with respect to some measure μ on K. In addition, for every $h \in \mathbb{R}[X]$ and every $\sigma \in G$,

$$\int_K h\,d\mu = L^G(h) = L^G(h^\sigma) = \int_K h^\sigma\,d\mu = \int_K h\,d\mu^\sigma,$$

and so as K is compact, $\mu = \mu^\sigma$ for all $\sigma \in G$; i.e., μ is G-invariant. $\square$

So the G-invariant optimization problem (1) can be rephrased as

$$p^* = \inf\{L^G(p)\colon L^G \text{ a } G\text{-linear map } \mathbb{R}[X]^G \to \mathbb{R}, L^G(1) = 1 \text{ and each } \mathscr{L}^G_{g_j} \text{is psd}\}. \tag{14}$$

The reformulation already gives a first computational advantage compared to the usual approach. By the definition of $\mathscr{L}^G_{g_j}$ in (13) we can phrase the moment matrix in terms of variables that are merely indexed by a basis $\mathscr{B}$ of the invariant ring $\mathbb{R}[X]^G$. This reduction on the number of moment variables is illustrated in the following example.

Example 3.3. Let C_4 denote the cyclic group of order four that operates on $\mathbb{R}^4$ by cyclically permuting the coordinates. The space of C_4-invariant polynomials of degree at most 2 is spanned by $b_0 := 1$, $b_1 := (1/4)(x_1 + x_2 + x_3 + x_4)$, $b_2 := (1/4)(x_1^2 + x_2^2 + x_3^2 + x_4^2)$, $b_3 := (1/4)(x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1)$, $b_4 := (1/2)(x_1x_3 + x_2x_4)$. By identifying $y_i = L^G(b_i)$ with moment variables $y_0, \ldots, y_4$, the (truncated) psd condition (13) for polynomials of degree at most 2 can be stated as a matrix psd condition in only four variables, $y_1, \ldots, y_4$,

$$\begin{pmatrix} 1 & y_1 & y_1 & y_1 & y_1 \\ y_1 & y_2 & y_3 & y_4 & y_3 \\ y_1 & y_3 & y_2 & y_3 & y_4 \\ y_1 & y_4 & y_3 & y_2 & y_3 \\ y_1 & y_3 & y_4 & y_3 & y_2 \end{pmatrix} \succeq 0.$$

In addition to this reduction of the number of variables, the structure of the moment matrix approach can be simplified based on representation theory. To apply the methods from §2.3, observe that for any $k \geq 0$, the subset (of $\mathbb{R}[X]$) of polynomials of total degree at most k is finite-dimensional and thus can be viewed as a real G-module. As a consequence of this exhaustion process for $\mathbb{R}[X]$, there exists a complex decomposition of the form

$$\mathbb{R}[X] \otimes C = \bigoplus_{i=1}^{h} V_i = \bigoplus_{i=1}^{h}\bigoplus_{j\in J_i} W_{ij}, \tag{15}$$

with complex irreducible components W_{ij}, and corresponding real decomposition of the form (10),

$$\mathbb{R}[X] = (V_1 + V_2) \oplus \frac{1}{\mathrm{i}}(V_1 - V_2) \oplus \cdots \oplus (V_{2l-1} + V_{2l}) \oplus \frac{1}{\mathrm{i}}(V_{2l-1} - V_{2l}) \oplus V_{2l+1} \oplus \cdots \oplus V_h, \tag{16}$$

where i is the imaginary unit. Note that the index sets $J_1, \ldots, J_h$ may be infinite now. The component with respect to the trivial irreducible representation is the invariant ring $\mathbb{R}[X]^G$, and the elements of the other isotypic components are called *semi-invariants*.

Fix an $i \in \{1, \ldots, h\}$. Since the W_{ij} are pairwise isomorphic, we can choose G-isomorphisms $\phi_j^i\colon W_{i1} \to W_{ij}$, $j \in J_i$. Furthermore, we can assume that W_{i1} is generated by a basis $\{s_{1,u}^i\colon 1 \le u \le \dim W_{ij}\} \subseteq \mathbb{C}[X]$, such that any $s_{1,u}^i$ is in the G-orbit of $s_{1,1}^i$. The basis vector $s_{1,1}^i$ transfers to W_{ij} by the G-isomorphism ϕ_j^i via $s_{j,1}^i := \phi_j^i(s_{1,1}^i)$, and thus by this isomorphism ϕ_j^i the whole basis $\{s_{1,u}^i\colon 1 \le u \le \dim W_{i1}\}$ of W_{i1} transfers to a whole basis $\{s_{j,u}^i\colon 1 \le u \le \dim W_{ij}\}$ of W_{ij}. Set $\mathscr{S}^i = \{s_{j,1}^i\colon j \in J_i\}$.

Using the bookkeeping techniques from §2.3 (describing the transition from (15) to (16)), the set $\mathscr{S}^i \subseteq \mathbb{C}[X]$ can be transformed into a subset of $\mathbb{R}[X]$ by distinguishing the types I, II, and III. Therefore, without loss of generality we can assume $\mathscr{S}^i$ to be real.

Theorem 3.4. *Let $g_1, \ldots, g_m \in \mathbb{R}[X]$ be G-invariant, assume that the feasible set K satisfies Assumption 2.1, and set $g_0 := 1$. A G-linear map $L^G\colon \mathbb{R}[X] \to \mathbb{R}$ is the integration with respect to a G-invariant measure μ on K if and only if for all $i \in \{1, \ldots, h\}$ and all $j \in \{1, \ldots, m\}$ the bilinear map $\mathscr{L}_{g_j}^G$ from (13) restricted to $\mathscr{S}^i$ $(\subseteq \mathbb{R}[X])$ is positive semidefinite.*

Proof. By Theorem 3.2, L^G is the integration w.r.t. a measure μ supported on K if and only if all the bilinear forms $\mathscr{L}_{g_j}^G$ from (13) are psd. Clearly, the latter condition implies that all the restrictions to the $\mathscr{S}^i$ are psd.

Conversely, to keep indices simple, fix one of the polynomials $g_0, \ldots, g_m$ and call it g for short. By assumption, the restriction of $\mathscr{L}_g^G$ to any $\mathscr{S}^i$ is psd, $1 \le i \le h$. We show that this already implies that $\mathscr{L}_g^G$ is psd. To this end, we first show that for distinct real isotypic components V_l and V_k we have $L^G(p_l \cdot p_k \cdot g) = 0$ for all $p_l \in V_l$ and $p_k \in V_k$. Indeed, each p_l defines a linear map

$$\chi_{p_l}\colon V_k \to \mathbb{R}, \qquad q \mapsto \mathscr{L}_g^G(p_l, q).$$

Hence, the application $p_l \mapsto \chi_{p_l}$ gives rise to a G-homomorphism from V_l to the dual space V_k^* (and thus to a G-homomorphism from V_l to V_k.) But by Schur's Lemma 2.8 this has to be the zero map, and thus $\mathscr{L}_g^G(p_l, q) = 0$ for all $q \in V_k$.

Consequently it suffices to show that for every $i \in \{1, \ldots, h\}$ the positive semidefiniteness on $\mathscr{S}^i$ implies already positive semidefiniteness on V_i. We first assume that V_i comes from a complex irreducible representation of type **I**. Consider a pair W_{ij}, W_{ik} in the decomposition (15), where we allow $j = k$. Similar to the above reasoning, Schur's Lemma can be applied by considering a linear map $\psi_{j,k}\colon W_{ij} \to W_{ik}$ defined by its images on the basis vectors,

$$\psi_{j,k}(s_{j,u}^i) := \sum_v \mathscr{L}_g^G(s_{j,u}^i, s_{k,v}^i) \cdot s_{k,v}^i, \qquad 1 \le u \le \dim W_{ij}.$$

Since by Schur's Lemma a G-isomorphism from W_{ij} to W_{ik} is unique up to a scalar multiplication, we can write this map as a composition $\psi_{j,k} = c_{jk} \cdot \phi_k^i \circ (\phi_j^i)^{-1}$, where the constant c_{jk} can be chosen real. This implies

$$\mathscr{L}_g^G(s_{j,u}^i, s_{k,v}^i) = \delta_{uv} c_{jk},$$

where δ_{uv} denotes Kronecker's delta function. Since any $f \in V_i$ can be written in the form $f = \sum_j \sum_u \alpha_{j,u} s_{j,u}^i$ with $\alpha_{j,u} \in \mathbb{R}$, the G-invariance of $\mathscr{L}_g^G$ gives

$$\mathscr{L}_g^G(f, f) = \mathscr{L}_g^G\left(\sum_j \sum_u \alpha_{j,u} s_{j,u}^i, \sum_j \sum_u \alpha_{j,u} s_{j,u}^i\right) = \sum_u \mathscr{L}_g^G\left(\sum_j \alpha_{j,u} s_{j,1}^i, \sum_j \alpha_{j,u} s_{j,1}^i\right),$$

which is nonnegative since $\mathscr{L}_g^G$ is psd on $\mathscr{S}^i$. For the other types (**II** and **III**) the result is implied in the same way by taking additionally into account the bookkeeping techniques explained in §2.3 (see also Example 3.9 below). Therefore the forms $\mathscr{L}_g^G$ are psd on $\mathbb{R}[X]$ if and only if the restrictions to each $\mathscr{S}^i$ are psd and the statement follows with Theorem 3.2. □

We also record the following symmetric version of Putinar's Positivstellensatz, which follows from dualizing Theorem 3.4.

THEOREM 3.5. *Let $f, g_1, \ldots, g_m \in \mathbb{R}[X]$ be G-invariant, and let the feasible set K satisfy Assumption 2.1. If f is strictly positive on K, then f can be written in the form*

$$f = \rho^G\left(\sum_{i=1}^{h} q_0^i + \sum_{j=1}^{m} g_j \sum_{i=1}^{h} q_j^i\right),$$

where q_j^i is a sum of squares of polynomial in $\mathcal{S}^i$, $1 \le i \le h$, $1 \le j \le m$, and $\rho^G(p) = (1/|G|)\sum_{\sigma\in G} p^\sigma$.

Putting all this together we obtain the following symmetry-adapted relaxation scheme. For every $k \in \mathbb{N}$ let $\mathcal{B}_{2k}$ be a basis of the set $\mathbb{R}[X]^G_{\le 2k}$ of G-invariant polynomials of degree at most k, with $1 \in \mathcal{B}_{2k}$. Any G-invariant map $L^G: \mathbb{R}[X]^G_{\le 2k} \to \mathbb{R}$ can be identified with a finite sequence $y = (y_b)_{b\in\mathcal{B}_{2k}}$ by setting $y_b := L^G(b)$ for $b \in \mathcal{B}_{2k}$. With regard to the real decomposition of $\mathbb{R}[X]_{\le k}$ coming from the truncation of (16), construct the sets $\mathcal{S}_k^i = \{s_1^i, s_2^i, \ldots, s_{\eta_i}^i\} \subseteq \mathcal{S}^i$ of the basis elements of $\mathcal{S}^i$ of degree at most k. Then we define the *symmetry-adapted moment matrix* $M_k^G(y)$ by

$$M_k^G(y) := \bigoplus_{i=1}^{h} M_{k,i}^G(y), \quad \text{where } M_{k,i}^G(y)_{v,w} := \mathcal{L}_{g_0}^G(s_v^i, s_w^i) = \mathcal{L}_1^G(s_v^i, s_w^i). \tag{17}$$

The entries of $M_k^G(y)$ are linear combinations of the elements of y and hence indexed by elements in $\mathcal{B}_{2k}$. Defining the *symmetry-adapted localizing matrices* in a similar manner, we obtain the symmetry-adapted relaxation for $k \ge k_0 := \max\{\lceil \deg f/2\rceil, \lceil \deg g_1/2\rceil, \ldots, \lceil \deg g_m/2\rceil\}$,

$$Q_k^G: \quad \begin{array}{c} \inf_y \, L^G(f) \\ M_k^G(y) \succeq 0, \\ M^G_{k-\lceil \deg g_j/2\rceil}(g_j y) \succeq 0, \quad 1 \le j \le m, \\ y_1 = 1, \end{array} \tag{18}$$

with optimal value denoted by $\inf Q_k^G$ (and $\min Q_k^G$ if the infimum is attained).

REMARK 3.6. *Computational aspects:* The symmetry-adapted setting defined above can give a significant reduction compared to the original relaxation scheme. Indeed the number of variables involved equals the size of $\mathcal{B}_{2k}$. Furthermore, the symmetry-adapted moment matrix is block diagonal with blocks of sizes $\eta_1, \ldots, \eta_h$.

If the irreducible representations of a given group are known, then isotypic decompositions and therefore the basis polynomials s_j^i can be algorithmically computed using projections (see Serre [35, Proposition. 8]).

With regard to determining the irreducible components, there are theoretical methods whose computational complexity is bounded by a polynomial in $|G|$ (see Babai and Rónyai [1]). For practical purposes, the size $|G|$ of the group might not be the appropriate measure of complexity, and indeed, there are practical methods (see Dabbaghian-Abdoly [10] and Dixon [13]) which are implemented in the group-theoretic software GAP (The GAP Group [37]).

Note that all these pre-computations have to be done only once for a specific group and relaxation order.

In this setting, Proposition 2.3 can be reformulated as follows.

THEOREM 3.7. *Let $f, g_1, \ldots, g_m \in \mathbb{R}[X]$ be G-invariant, let Assumption 2.1 hold for the feasible set K, and let $(Q_k^G)_{k\ge k_0}$ be the hierarchy (18) of symmetry-adapted SDP-relaxations. Then $(\inf Q_k^G)_{k\ge k_0}$ is a monotonically nondecreasing sequence that converges to f^*.*

PROOF. As $\mathcal{P}(K)^G \subseteq \mathcal{P}(K)$ one has $\inf Q_k^G \ge \inf Q_k$ for all $k \ge k_0$. In addition, for any measure μ on K we let $\mu^\# = (1/|G|)\sum_{\sigma\in G}\mu^\sigma$. As K is G-invariant, $\mu^\#$ is also supported on K. This proves that $\inf Q_k^G \le f^*$ for all $k \ge k_0$, and so, $\inf Q_k \le \inf Q_k^G \le f^*$ for all $k \ge k_0$. Combining the latter with Proposition 2.3 yields the desired result. □

REMARK 3.8. If (to be more general than the setting considered in this article) not all g_j are G-invariant but the set K is G-invariant, or even if just the set of optimal values is invariant, it is still possible to only look at invariant moments. However, the above block structure will in general only apply for the moment matrix and the localizing matrices for the G-invariant polynomials. Note, however, that the variables in the localizing matrices still correspond to a basis for the space of G-invariants.

The general setting presented in this section leads to the question of handling the hierarchy of symmetry-adapted bases for the hierarchy of vector spaces of polynomials. Before studying in detail the symmetric group in the next section, as a warm up, it is worth revisiting Example 3.3.

Example 3.9. We continue Example 3.3. Since the cyclic group C_4 is abelian, all the irreducible representations are one-dimensional and correspond to the fourth roots of unity. Consider the symmetry-adapted relaxation for $k=1$. Using the methods from Remark 3.6 we find $\mathbb{R}[X]_1 \otimes \mathbb{C} = \bigoplus_{l=1}^4 V_l$, where $V_l = \operatorname{span}\{\sum_{j=1}^4 \omega_l^j X_j\}$ and $\omega_1, \ldots, \omega_4$ denote the four roots of unity. Thus, with regard to the symmetry-adapted moment matrix (17), we obtain $\mathcal{S}_1^{2l-1} = \operatorname{span}\{\sum_{j=1}^4 (\omega_l^j + \bar{\omega}_l{}^j) X_j\}$ and $\mathcal{S}_1^{2l} = \operatorname{span}\{(1/\mathrm{i}) \sum_{j=1}^4 (\omega_l^j - \bar{\omega}_l{}^j) X_j\}$, where $1 \le l \le 2$. With the notation of Example 3.3 the truncated symmetry-adapted moment matrix is

$$\begin{pmatrix} 1 & 2y_1 & 0 & 0 & 0 \\ 2y_1 & y_2+2y_3+y_4 & 0 & 0 & 0 \\ 0 & 0 & y_2-y_4 & 0 & 0 \\ 0 & 0 & 0 & y_2-2y_3+y_4 & 0 \\ 0 & 0 & 0 & 0 & y_2-y_4 \end{pmatrix}.$$

We get four diagonal blocks (of which three are elementary), and so we end up with a 2×2 semidefiniteness constraint instead of a 5×5 one if symmetry is not exploited.

4. Optimizing with symmetric polynomials. In this section, we provide several techniques to exploit symmetries for the symmetric group $\mathcal{S}_n$. While the representation theory of the symmetric group is a classical topic (as reviewed in §2.4), it yields some interesting (even somewhat surprising) results in our setting.

First, in §4.1 we discuss the symmetry-adapted relaxation for the symmetric group. By realizing the irreducible components in a suitable basis of polynomials (generalized Specht polynomials as defined below), the moment matrix can be characterized rather explicitly (Theorem 4.6). We derive some concrete representation theorems as corollaries for symmetric polynomials in §4.2.

4.1. Moment matrices for the symmetric group. Recall from the preliminaries that the irreducible representations of $\mathcal{S}_n$ are in natural bijection with the partitions of n. To construct a suitable generalized moment matrix we will need a graded decomposition of the vector space $\mathbb{R}[X]$ into $\mathcal{S}_n$-irreducible components. A classical construction of Specht gives a realization of the Specht modules as polynomials (see Specht [36]):

For $\lambda \vdash n$ let t be a λ-tableau. To t we associate the monomial $X^t := \prod_{i=1}^n X_i^{l(i)-1}$, where $l(i)$ is the index of the row of t containing i. Denote by $\mathcal{C}_1, \ldots, \mathcal{C}_\nu$ the columns of t and by $\mathcal{C}_j(i)$ the element in the ith row of the column $\mathcal{C}_j$. Then we associate to each column $\mathcal{C}_j$ a Vandermonde determinant

$$\mathrm{Van}_{\mathcal{C}_j} := \det \begin{pmatrix} X^0_{\mathcal{C}_j(1)} & \cdots & X^0_{\mathcal{C}_j(r_j)} \\ \vdots & \ddots & \vdots \\ X^{r_j-1}_{\mathcal{C}_j(1)} & \cdots & X^{r_j-1}_{\mathcal{C}_j(r_j)} \end{pmatrix} = \prod_{1\le i<l\le r_j} (X_{\mathcal{C}_j(l)} - X_{\mathcal{C}_j(i)}),$$

where r_j denotes the number of rows of $\mathcal{C}_j$.

The *Specht polynomial* s_t associated to t is defined as

$$s_t := \prod_{j=1}^{\nu} \mathrm{Van}_{\mathcal{C}_j} = \sum_{\sigma \in \mathrm{CStab}_t} \operatorname{sgn}(\sigma)\sigma(X^t),$$

where CStab_t is the column stabilizer of t as introduced in §2.4. The polynomials $\{s_t\colon t$ standard Young tableau to $\lambda\}$ are called the *Specht polynomials* associated to λ.

Note that for any λ-tabloid $\{t\}$ the monomial X^t is well defined, and the mapping $\{t\} \mapsto X^t$ is an $\mathcal{S}_n$-invariant mapping. Thus $\mathcal{S}_n$ operates on s_t in the same way as on the polytabloid e_t. This observation implies (see Specht [36]):

Lemma 4.1. *For any partition $\lambda \vdash n$, the Specht polynomials associated to λ span an $\mathcal{S}_n$-submodule of $\mathbb{R}[X]$ which is isomorphic to the Specht module S^λ.*

While Lemma 4.1 already gives a realization of the Specht modules in terms of polynomials, for the symmetry-adapted moment matrix we need to generalize this construction to realize these modules in terms of polynomials with prescribed exponent vectors. In the following, let $n \in \mathbb{N}$ and $\beta := (\beta_1, \ldots, \beta_n)$ be an n-tuple of nonnegative integers, and set $\mathbb{R}\{X^\beta\} := \operatorname{span}\{\sigma(X^\beta) \colon \sigma \in \mathcal{S}_n\}$. By construction, $\mathbb{R}\{X^\beta\}$ is closed under the action of $\mathcal{S}_n$ and therefore has the structure of an $\mathcal{S}_n$-module.

Denote by $\operatorname{wt}(\beta) = \sum_{i=1}^n \beta_i$ the *weight* of β. Let $b_1, \ldots, b_\ell$ be the distinct components of β (called the *parts* of β), ordered (decreasingly) according to the multiplicity of the occurrence in β. Furthermore, let $I_j = \{i \in \{1, \ldots, n\} \colon \beta_i = b_j\}$, $1 \le j \le \ell$, and note that the sets $I_1, \ldots, I_\ell$ define a partition of $\{1, \ldots, n\}$. Setting $\mu_j := |I_j|$, the vector $\mu = (\mu_1, \ldots, \mu_\ell)$ consists of monotonically decreasing components and thus defines a partition of n. We call $\mu \vdash n$ the *shape* of β.

Lemma 4.2. *For $\beta \in \mathbb{N}_0^n$, the $\mathcal{S}_n$-module $\mathbb{R}\{X^\beta\}$ is isomorphic to the permutation module M^μ, where μ is the shape of β.*

Proof. We construct an isomorphism from $\mathbb{R}\{X^\beta\}$ to M^μ by defining it on the basis elements.

First observe that $\mathbb{R}\{X^\beta\} = \operatorname{span}\{X^\gamma \colon \gamma \text{ permutation of } \beta\}$. For a fixed permutation γ of β, consider the set partition $I_1, \ldots, I_\ell$ associated to γ, and map X^γ to the μ-tabloid with rows $I_1, \ldots, I_\ell$. Since this mapping commutes with the action of $\mathcal{S}_n$, and since by Definition 2.10 the $\mathcal{S}_n$-module M^μ is spanned by the set of all μ-tabloids, the statement follows. $\square$

Now let $\lambda \vdash n$ be another partition of n. To construct the realizations of the Specht module S^μ as submodules of $\mathbb{R}\{X^\beta\}$, we look at pairs (t, T), where t is a fixed λ-tableau and T is a generalized Young tableau with shape λ and content μ. For each pair we construct a monomial $X^{(t,T)} \in \mathbb{R}\{X^\beta\}$ from its parts $b_1, \ldots, b_\ell$ in the following way. As before, let $\mathcal{C}_1, \ldots, \mathcal{C}_\nu$ be the columns of t, and denote by $T(i, j)$ the element in the ith row and jth column of T. Then define

$$X^{(t,T)} := \prod_{(i,j)} X_{\mathcal{C}_j(i)}^{b_{T(i,j)}},$$

and associate to each column $\mathcal{C}_j$ a polynomial

$$\operatorname{Van}_{\mathcal{C}_j, T} := \det \begin{pmatrix} X_{\mathcal{C}_j(1)}^{b_{T(1,j)}} & \ldots & X_{\mathcal{C}_j(k)}^{b_{T(1,j)}} \\ \vdots & \ddots & \vdots \\ X_{\mathcal{C}_j(1)}^{b_{T(k,j)}} & \ldots & X_{\mathcal{C}_j(k)}^{b_{T(k,j)}} \end{pmatrix}. \tag{19}$$

As in Specht's construction we form the product polynomial $s_{(t,T)} := \prod_{j=1}^{\nu} \operatorname{Van}_{\mathcal{C}_j, T}$, and set (by summation over the row equivalence class $\{T\}$ of T) $S_{(t,T)} := \sum_{S \in \{T\}} s_{(t,S)}$.

Lemma 4.3. *Let $\lambda \vdash n$, $\beta \in \mathbb{N}_0^n$, and μ be the shape of β. Furthermore, let t be a λ-tableau and T be a generalized Young tableau with shape λ and content μ. The $\mathcal{S}_n$-submodule $\mathbb{R}\{S_{(t,T)}\}$ of $\mathbb{R}\{X^\beta\}$ generated by the generalized Specht polynomial $S_{(t,T)}$ is isomorphic to the Specht module S^λ.*

Proof. By Lemma 4.2 we can follow Young's decomposition of M^μ. Therefore we associate to every T with shape λ and content μ an $\mathcal{S}_n$-homomorphism $\Theta_{T,t} \colon M^\lambda \to \mathbb{R}\{X^\beta\}$, which maps the given λ-tableau $\{t\}$ to $\sum_{S \in \{T\}} X^{(t,S)}$ and which naturally extends by the cyclic structure of M^λ.

Now it suffices to show that the image of $S^\lambda \subseteq M^\lambda$ under $\Theta_{T,t}$ coincides with $\mathbb{R}\{S_{(t,T)}\}$. The image of the polytabloid e_t from (12) is

$$\Theta_{T,t}(e_t) = \Theta_{T,t}\left(\sum_{\sigma \in \mathrm{CStab}_t} \operatorname{sgn}(\sigma)\sigma\{t\}\right) = \sum_{\sigma \in \mathrm{CStab}_t} \Theta_{T,t}(\operatorname{sgn}(\sigma)\sigma\{t\}).$$

Since by the Leibniz expansion of (19) we have $s_{(t,T)} = \sum_{\sigma \in \mathrm{CStab}_t} \operatorname{sgn}(\sigma)\sigma(X^{(t,T)})$, it follows that $\Theta_{T,t}(e_t) = S_{(t,T)}$, which proves the claim. $\square$

Remark 4.4. Note the following connection of the generalized Specht polynomials to the classical Schur polynomials. For a nonnegative vector $\lambda = (\lambda_1, \ldots, \lambda_l)$ the generalized Vandermonde determinant

$$a_\lambda := \det\big((X_i^{\lambda_j})_{1 \le i, j \le l}\big) \tag{20}$$

(as a polynomial in $X_1, \ldots, X_l$) is called the *alternant* of λ. Moreover, for a partition λ of length l the *Schur function* s_λ is defined by $s_\lambda = a_{\lambda+\delta}/a_\delta$, where $\delta := (l-1, l-2, \ldots, 1, 0) \in \mathbb{Z}^l$. It is well known that s_λ is a symmetric polynomial in $X_1, \ldots, X_l$ (also called a *Schur polynomial*). Hence, the alternant (20) can be written as

$$a_\lambda = s_{\lambda-\delta} \cdot a_\delta. \tag{21}$$

Now the polynomials $\operatorname{Van}_{\mathcal{C}_j, T}$ defined above can be seen as the alternant associated to the numbers $(b_{T(1,j)}, \ldots, b_{T(k,j)})$ and thus by (21) as the product of a Schur polynomial with a classical Vandermonde determinant.

Let $\mathcal{T}^0_{\lambda,\mu}$ denote the set of semistandard generalized Young tableaux of shape λ and content μ. To conclude we can summarize the above considerations.

Theorem 4.5. *Let $\beta \in \mathbb{N}_0^n$ of weight d and shape $\mu = \mu(\beta)$. Then we have*

$$\mathbb{R}\{X^\beta\} = \bigoplus_{\lambda \trianglerighteq \mu} \bigoplus_{T \in \mathcal{T}^0_{\lambda,\mu}} \mathbb{R}\{S_{(t_\lambda, T)}\},$$

where t_λ denotes the unique λ-tableau with increasing rows and columns. The multiplicity of the Specht modules S^λ in this $\mathcal{S}_n$-module is equal to the Kostka number $K_{\lambda\mu}$.

Proof. By Lemma 4.2 the $\mathcal{S}_n$-module $\mathbb{R}\{X^\beta\}$ is isomorphic to M^μ. By Proposition 2.14, the multiplicity of S^λ in M^μ is $K_{\lambda\mu}$, corresponding to the decomposition $M^\mu = \bigoplus_{\lambda \trianglerighteq \mu} \bigoplus_{T \in \mathcal{T}^0_{\lambda,\mu}} S^\lambda$ in terms of semistandard Young tableaux. Now for $T \in \mathcal{T}^0_{\lambda,\mu}$ the $\mathcal{S}_n$-homomorphism $\Theta_{T,t}$ constructed in the proof of Lemma 4.3 maps e_{t_λ} to $S_{(t_\lambda, T)}$ and thus turns this decomposition into the decomposition of $\mathbb{R}\{X^\beta\}$. $\square$

Based on these results, we can construct the symmetry-adapted moment matrix of order k. By (17) the blocks are labeled by partitions of n. To construct the block for a fixed λ consider the various $\beta = (\beta_1, \dots, \beta_n)$ with $\mathrm{wt}(\beta) = k$ and shape λ. For a given d, let $c(\mu, d)$ be the number of $\beta \in \mathbb{N}_0^n$ with $\mathrm{wt}(\beta) = d$ which have shape μ. The decomposition from Theorem 4.5 translates into the moment setting as follows.

Corollary 4.6. *For $k \in \mathbb{N}$, the kth symmetry-adapted moment matrix $M_k^G(y)$ is of the form*

$$M_k^G(y) = \bigoplus_{\lambda \vdash n} M_{k,\lambda}^G(y).$$

Each of the blocks $M_{k,\lambda}^G(y)$ consists of κ_λ rows and columns, where κ_λ equals $\sum_{d=0}^k c(\mu, d) K_{\lambda\mu}$.

Proof. The distinct irreducible representations are indexed by partitions of n. Therefore by Remark 3.6 the number of rows (and columns) of the block of $M_{k,\lambda}^G(y)$ corresponding to the irreducible component S^λ equals the number of submodules of $\mathbb{R}[X]_{\le k}$ isomorphic to S^λ. As we have

$$\mathbb{R}[X]_{\le k} = \bigoplus_{d=0}^{k}, \bigoplus_{\beta \in \mathbb{N}_0^n,\, \mathrm{wt}(\beta)=d} \mathbb{R}\{X^\beta\}, \tag{22}$$

this number is $\sum_{d=0}^k c(\beta, d) K_{\lambda\mu(\beta)}$ by Theorem 4.5. $\square$

Although the above corollary is stated only in terms of the symmetry-adapted moment matrix, it naturally also translates to the localizing matrices. The only difference concerning the size of the localizing matrices is reflected in a slightly different calculation compared to κ_λ defined above. In the case of the localizing matrix associated to a polynomial g the summation will only go up to $k - \lceil \deg g/2 \rceil$.

We obtain the following remarkable consequence.

Theorem 4.7. *For all $n \ge 2k$ the symmetry-adapted moment matrix of order k has the same structure, i.e., the same number and sizes of blocks and variables. In particular, up to the computation of the block decomposition the complexity of the question if a symmetric polynomial of degree $2k$ in n variables is a sum of squares is only depending on k.*

Proof. First observe that by Remark 3.6 the number of variables equals the dimension of the $\mathbb{R}$-vector space of symmetric polynomials of degree at most $2k$. Therefore it corresponds to the number of n-partitions of $2k$, which is just the number of partitions of $2k$ for all $n \ge 2k$. So we see that the number of variables does not increase in n once $n \ge 2k$.

Now set $n_0 = 2k$ and let l be the number of partitions of k, $\beta^{(1)}, \dots, \beta^{(l)} \in \mathbb{N}_0^{n_0}$ the distinct exponent vectors modulo permutation with $\mathrm{wt}(b^{(i)}) = k$, $1 \le k \le l$, and $\lambda^{(i)} \vdash n_0$ be the shape of $\beta^{(i)}$. The rest of the proposition follows if we can show that for every $n \ge n_0$ there exist partitions $\tilde{\lambda}^{(1)}, \dots, \tilde{\lambda}^{(m)}$ of n such that $\kappa_{\tilde{\lambda}^{(i)}} = \kappa_{\lambda^{(i)}}$ for all $1 \le i \le m$ and $\kappa_{\tilde{\lambda}} = 0$ for all other $\tilde{\lambda} \vdash n$.

First note that $\tilde{\beta}^{(i)}$ are exponent vectors that come from $\beta^{(i)}$ by adding $n - n_0$ zeros. As $n \ge n_0 \ge 2k$ this implies that the possible $\tilde{\lambda}^{(i)}$ are of the form $\tilde{\lambda}^{(i)} := (\lambda_1^{(i)} + n - n_0, \lambda_2^{(i)}, \dots, \lambda_t^{(i)})$. Since $K_{\lambda\mu} = 0$ whenever $\mu \ntrianglelefteq \lambda$ we conclude that the possible $\tilde{\mu}$ we have to consider are of the form $\tilde{\mu} := (\mu_1 + n - n_0, \mu_2, \dots, \mu_t)$ for one $\mu \ge \lambda^{(i)}$. But in this setting we have $K_{\lambda\mu} = K_{\tilde{\lambda}\tilde{\mu}}$ and the statement follows. $\square$

Example 4.8. We illustrate the techniques for a small example with $n = 3$ and $k = 2$. The moment variables are indexed by partitions of the numbers 1, 2, 3, 4, with three parts, i.e., $y_1, y_2, y_3, y_4, y_{11}, y_{22}, y_{21}, y_{111}, y_{211}$. The irreducible components are indexed by the partitions $\lambda \vdash (3)$; thus $\lambda \in \{(3), (2,1), (1,1,1)\}$. The β we have to take into account are $(0,0,0)$, $(1,0,0)$, $(2,0,0)$, $(1,1,0)$ with shape $\mu^{(1)}=(3)$, $\mu^{(2)}=(2,1)$, $\mu^{(3)}=(2,1)$, $\mu^{(4)}=(2,1)$. The semistandard generalized Young tableaux with shape μ and content $\lambda\in\{\lambda^{(1)},\dots,\lambda^{(4)}\}$ from Lemma 4.3 are:

- For $\mu=(3)$:

$$\begin{array}{|c|c|c|}\hline 1&1&1\\\hline\end{array}\,,\quad \begin{array}{|c|c|c|}\hline 1&1&2\\\hline\end{array}\,.$$

- For $\mu=(2,1)$:

$$\begin{array}{|c|c|}\hline 1&1\\\hline 2\\\cline{1-1}\end{array}\,.$$

- For $\mu=(1,1,1)$ there is no generalized semistandard Young tableau corresponding to the above $\lambda^{(i)}$.

For $\mu=(3)$, Corollary 4.6 yields a 4×4 block, with basis polynomials

$$\{1,\, X_1+X_2+X_3,\, X_1^2+X_2^2+X_3^2,\, X_1X_2+X_1X_3+X_2X_3\}.$$

Thus

$$M_{(3)}:=\begin{pmatrix}1 & 3y_1 & 3y_2 & 3y_{11}\\ 3y_1 & 3y_2+6y_{11} & 3y_3+6y_{21} & 6y_{21}+3y_{111}\\ 3y_2 & 3y_3+6y_{21} & 3y_4+6y_{22} & 6y_{31}+3y_{211}\\ 3y_{11} & 6y_{21}+3y_{111} & 6y_{31}+3y_{211} & 3y_{22}+6y_{211}\end{pmatrix}.$$

For $\mu=(2,1)$ we obtain a 3×3 block, with basis polynomials

$$\begin{aligned}&\{X_3-X_1+X_3-X_2,\,(X_3-X_1)(X_3+X_1)+(X_3-X_2)(X_3+X_2),\,(X_3-X_1)X_2+(X_3-X_2)X_1\}\\ &\quad=\{2X_3-X_2-X_1,\,2X_3^2-X_2^2-X_1^2,\,-2X_1X_2+X_2X_3+X_3X_1\}.\end{aligned}$$

Thus

$$M_{(2,1)}=\begin{pmatrix}6y_2-6y_{11} & 6y_3-6y_{21} & 6y_{21}-6y_{111}\\ 6y_3-6y_{21} & 6y_4-6y_{22} & -6y_{211}+6y_{31}\\ 6y_{21}-6y_{111} & -6y_{211}+6y_{31} & 6y_{22}-6y_{211}\end{pmatrix}.$$

Hence we only have to check semidefiniteness of a 4×4 matrix and a 3×3 matrix, instead of semidefiniteness of a single 10×10 moment matrix.

Remark 4.9. We remark that the techniques presented above also provide the tools for some explicitly stated open issues in the study of unconstrained optimization of symmetric polynomials in Gatermann and Parrilo [14, p. 124] (who—mentioning the lack of explicit formulas for the isotypic components—refer to the study of examples and asymptotics).

4.2. Sums of squares representations for symmetric polynomials. From the dual point of view, the results presented in §4.1 imply the following sums of squares decomposition theorem:

Theorem 4.10. *Let $p\in\mathbb{R}[X_1,\dots,X_n]$ be symmetric and homogeneous of degree $2d$. If p is a sum of squares, then*

$$p\in\sum_{\beta}\sum_{\lambda\trianglerighteq\mu(\beta)}\sum_{T\in\mathcal{T}^0_{\lambda,\mu(\beta)}}\Sigma(\mathbb{R}\{S_{(t_\lambda,T)}\})^2,$$

where t_λ denotes the unique λ-tableau with increasing rows and columns, β runs over the nonnegative partitions of d with n parts, and $\Sigma(\mathbb{R}\{S_{(t_\lambda,T)}\})^2$ denotes the sums of squares of polynomials in the $\mathcal{S}_n$-module $\mathbb{R}\{S_{(t_\lambda,T)}\}$.

Proof. The statement follows from dualizing (22) in connection with Theorem 4.5. □

Hilbert's Theorem (see, e.g., Laurent [26]) characterizes the cases where the notion of nonnegativity coincides with the sums of squares representability. For these cases we obtain the following corollaries which are specialized (and simplified) versions of SOS decompositions in the symmetric case.

Corollary 4.11. *Let $p\in\mathbb{R}[X_1,X_2]$ be a symmetric homogeneous form of degree $2d$. If p is nonnegative, then*

$$p\in\sum_{\substack{\alpha_1,\alpha_2\in\mathbb{N}_0\\ \alpha_1+\alpha_2=d}}\bigl\{\Sigma(\mathbb{R}\{X_1^{\alpha_1}X_2^{\alpha_2}+X_1^{\alpha_2}X_2^{\alpha_1}\})^2+\Sigma(\mathbb{R}\{X_1^{\alpha_1}X_2^{\alpha_2}-X_1^{\alpha_2}X_2^{\alpha_1}\})^2\bigr\}.$$

Corollary 4.12. *Let $p \in \mathbb{R}[X_1, \ldots, X_n]$ be a symmetric and homogeneous quadratic form. If p is nonnegative, then p can be written in the form*

$$p = \alpha(X_1 + \cdots + X_n)^2 + \beta \sum_{i<j}(X_j - X_i)^2 = (\alpha + (n-1)\beta)\sum_i X_i^2 + 2(\alpha - \beta)\sum_{i<j} X_i X_j,$$

with some coefficients $\alpha, \beta \geq 0$.

Corollary 4.13. *Let $p \in \mathbb{R}[X_1, X_2, X_3]$ be a symmetric and homogeneous polynomial of degree 4. If p is nonnegative, then p can be written in the form*

$$p = (\alpha + 2\delta)M_4 + (2\alpha + 2\varepsilon + \gamma - \delta)M_{22} + (\beta - \omega)M_{31} + (\beta + 2\gamma + 2\omega - 2\varepsilon)M_{211},$$

where $M_4 = \sum_i X_i^4$, $M_{22} = \sum_{i\neq j} X_i^2 X_j^2$, $M_{31} = \sum_{i\neq j} X_i^3 X_j$, and $M_{211} = \sum_{i\neq j\neq k} X_i^2 X_j X_k$, such that $\alpha, \gamma, \delta, \varepsilon \geq 0$ and $\alpha\gamma \geq \beta^2$ and $\delta\varepsilon \geq \omega^2$.

5. Using the degree principle. For symmetric polynomials, another possible strategy is to use the degree principle originally introduced by Timofte [38, Corollary 2.1] and recently refined by Riener [30, Theorem 4.5], apparently not well known in the optimization community. Using this principle, we show how to reduce an SDP-relaxation to a family of lower-dimensional relaxations. In some situations, this strategy might be preferable to the group-machinery developed in the above sections. In the special case of symmetric polynomials of degree 4 it reduces the nonnegativity problem to an SOS problem (and thus to a semidefinite feasibility problem); see Theorem 5.5.

Proposition 5.1 (Riener [30, Theorem 4.5]). *Let $f, g_1, \ldots, g_m \in \mathbb{R}[X]$ be symmetric and $K = \{x \in \mathbb{R}^n\colon g_1(x) \geq 0, \ldots, g_m(x) \geq 0\}$. Setting $r := \max\{2, \lfloor(\deg f)/2\rfloor, \deg g_1, \ldots, \deg g_m\}$ we have*

$$\inf_{x\in K} f(x) = \inf_{x\in K\cap A_r} f(x),$$

where A_r denotes the set of points in $\mathbb{R}^n$ with at most r distinct components.

For n, $r \in \mathbb{N}$, a vector $\omega = (\omega_1, \ldots, \omega_r)$ of positive, nonincreasing integers with $n = \omega_1 + \cdots + \omega_r$ is called an *r-partition* of n. Let Ω denote the set of all possible r-partitions of n. Then for each $\omega \in \Omega$, set

$$f^\omega := f(\underbrace{T_1, \ldots, T_1}_{\omega_1}, \underbrace{T_2, \ldots, T_2}_{\omega_2}, \ldots, \underbrace{T_r, \ldots, T_r}_{\omega_r}) \in \mathbb{R}[T_1, \ldots, T_r].$$

Similarly, let $K^\omega := \{t \in \mathbb{R}^r\colon g_1^\omega(t) \geq 0, \ldots, g_m^\omega(t) \geq 0\}$. With this notation we can transform the original optimization problem in n variables into a set of new optimization problems that involve only r variables,

$$\inf_{x\in K} f(x) = \min_{\omega\in\Omega} \inf_{t\in K^\omega} f^\omega(t). \tag{23}$$

Now one can apply the usual relaxation scheme to each of the above r-dimensional problems separately. For each $\omega \in \Omega$ let Q_k^ω be the kth relaxation (6) of $\min_{t\in K^\omega} f^\omega(t)$, with optimal value denoted by $\inf Q_k^\omega$. Putting these ideas together we obtain:

Theorem 5.2. *Let f, $g_1, \ldots, g_m \in \mathbb{R}[X]$ be G-symmetric, and let the feasible set K satisfy Assumption 2.1. Let $r := \max\{2, \lfloor(\deg f)/2\rfloor, \deg g_1, \ldots, \deg g_m\}$, and Ω be the set of r-partitions of n. Then the sequence $(\inf_{\omega\in\Omega}(\inf Q_k^\omega))_k$ converges to $\inf_{x\in K} f(x)$ for $k \to \infty$.*

Proof. By Theorem 5.1 there is an r-partition $\omega \in \Omega$ of n with $\min_{x\in K} f(x) = \min_{t\in K^\omega} f^\omega(t)$. It suffices to show that K^ω also meets Assumption 2.1. Since K meets Assumption 2.1, there is $u \in \mathbb{R}[X]$ with $u = u_0 + \sum_{j=1}^m u_j g_j$ for some sum of squares polynomials $u_0, \ldots, u_m$ such that the level set of u is compact. This representation carries over to u^ω which also has a compact level set. $\square$

Remark 5.3. At first sight it might not look profitable to replace one initial problem by a family of new problems. However, note that for fixed $r \in \mathbb{N}$ the number of r-partitions of any n is bounded by $(n+r)^r$. On the other hand a polynomial optimization problem in n variables yields a moment matrix of size $O(n^k)$ in the kth relaxation step of Lasserre's scheme. In view of the polynomial bound (for fixed r) on the number of r-partitions, it is therefore profitable to use the degree principle-based relaxation.

The process of building the r-dimensional problems can be related to breaking the symmetries as the resulting problems will in general no longer be invariant under a symmetric group $\mathcal{S}_r$. However, as dimensions drop there are situations where we will get finite convergence.

THEOREM 5.4. *Let $f, g_1, \ldots, g_m \in \mathbb{R}[X]$ be symmetric with $\deg g_j \le m$, $1 \le j \le m$, and $\deg f \le 2m$. Furthermore, assume that the variety $V(g_1, \ldots, g_m) \subseteq \mathbb{C}^n$ has codimension m. Then the relaxation sequence $(\inf_{\omega\in\Omega}(Q_k^\omega))_k$ converges to $\inf_{x\in V(g_1,\ldots,g_m)} f$ in finitely many steps, where Ω is defined as in Theorem 5.2.*

PROOF. It suffices to show that each of the varieties $V^\omega := V(g_1^\omega, \ldots, g_m^\omega)$ is zero-dimensional and then Proposition 2.4 gives the announced statement. To see that these varieties contain only finitely many points we proceed as follows:

It is classically known (see, for example, Cox et al. [9, Section 7.1]) that every symmetric polynomial g of degree d in n variables can be uniquely written as a polynomial in the first d power sum polynomials $p_1(X), \ldots, p_d(X)$, where $p_i(X) = \sum_{j=1}^n X_j^i$.

Let $\gamma_1 \ldots, \gamma_m \in \mathbb{R}[Z_1, \ldots, Z_m] \subseteq \mathbb{R}[Z_1, \ldots, Z_n]$ be polynomials such that $\gamma_i(p_1(X), \ldots, p_m(X)) = g_i(X)$. The surjective map

$$\pi\colon \mathbb{C}^n \to \mathbb{C}^n, \qquad x \mapsto (p_1(x), \ldots, p_n(x)),$$

establishes that $\mathbb{C}^n//\mathcal{S}_n$ (the so-called orbit space; cf. §6) is in fact isomorphic to $\mathbb{C}^n$.

As the variety $V(g_1, \ldots, g_m)$ is $\mathcal{S}_n$-invariant, its image in the quotient $\mathbb{C}^n//\mathcal{S}_n$ is given by $\tilde{V} := \{z \in \mathbb{C}^n\colon \gamma_i(z) = 0,\ 1 \le i \le m\}$. Since $\mathcal{S}_n$ is a finite group, the codimension of $\tilde{V}$ is also m. But this implies that $\tilde{V} \cap \{z \in \mathbb{C}^n\colon z_{m+1} = \cdots = z_n = 0\}$ is zero-dimensional. Therefore there are just finitely many $z := (z_1, \ldots, z_m)$ with $\gamma_i(z) = 0$ for $1 \le i \le m$.

Now let $\omega = (\omega_1, \ldots, \omega_r)$ be any r-partition of n and consider $V^\omega := V(g_1^\omega, \ldots, g_m^\omega) \subseteq \mathbb{C}^m$. Setting $\tilde{p}_i := \sum_{j=1}^m \omega_j T_j^i$, we get $g_i^\omega = \gamma_i(\tilde{p}_1, \ldots, \tilde{p}_m)$. For the points $y \in V^\omega$ we have $\tilde{p}_1(y) = z_1, \ldots, \tilde{p}_m(y) = z_m$ for one of the finitely many $z = (z_1, \ldots, z_m)$, with $\gamma_i(z) = 0$ for $1 \le i \le m$. Thus there are just finitely many points in V^ω. $\square$

Closely related to the question of finite convergence is the description of polynomials that are positive but not sums of squares. By Hilbert's Theorem, every nonnegative ternary quartic polynomial is a sum of squares. For quartics in more than three variables this is not true in general, not even for symmetric polynomials (see Example 5.6 below). However, for symmetric polynomials of degree 4, deciding the nonnegativity can be reduced to an SOS problem and thus to a semidefinite optimization problem.

THEOREM 5.5. *Let $f \in \mathbb{R}[X]$ be a symmetric polynomial of degree 4, and let Ω be the set of 2-partitions of n. Then f is nonnegative if and only if for all $\omega \in \Omega$ the polynomial f^ω is a sum of squares.*

PROOF. As f is of degree 4, for any $\omega \in \Omega$ the polynomial f^ω is of degree 4 in two variables. Hence, by Hilbert's Theorem f^ω is nonnegative if and only if it is a sum of squares. $\square$

EXAMPLE 5.6. Choi and Lam [8] have shown that the homogeneous polynomial of degree 4,

$$f = \sum_{i\neq j} X_i^2 X_j^2 + \sum_{i\neq j} X_i^2 X_j X_k - 4X_1X_2X_3X_4,$$

in four variables is nonnegative, but not a sum of squares. By Theorem 5.5, the nonnegativity of f is equivalent to the property that the following two homogeneous polynomials in two variables are sums of squares. We find

$$f_1 = X_2^4 + 4X_2^2X_4^2 + X_4^4 + 2X_4^3X_2 = (X_2^2)^2 + (X_4^2 + X_2X_4)^2 + (\sqrt{3}X_2X_4)^2,$$
$$f_2 = 4X_2^4 + 6X_2^2X_4^2 - 2X_2^3X_4 = \left(2X_2^2 - \frac{1}{2}X_2X_4\right)^2 + \left(\frac{1}{2}\sqrt{23}X_2X_4\right)^2,$$

which proves that f is indeed nonnegative.

6. PMI-relaxations via the geometric quotient.

In this section, we study another possibility to exploit symmetries. Namely, we want to exploit the fact that to any solution of an invariant optimization problem every point in its orbit is also optimal. Using invariant theory and a result in real algebraic geometry it is possible to characterize the space of all orbits. This orbit space approach leads very naturally to polynomial matrix inequalities (PMI). The main advantage of this approach is that, in some cases, this can decrease the degrees of the polynomials strongly. We will demonstrate this phenomenon in certain cases (such as power sum problems) where we obtain lower bounds and sometimes even upper bounds for a minimization problem by a very simple SDP relaxation.

6.1. The general setup. We consider the general G-invariant optimization problem (1), where $G \subseteq \mathrm{GL}_n(\mathbb{R})$ is a finite group. Denote the orbit of $x \in \mathbb{R}^n$ by $G(x) := \{\sigma(x): \sigma \in G\}$. The union of all orbits (with the induced topology) is called the *orbit space* $\mathbb{R}^n//G$ of G. To characterize the orbit space, let $\pi_1, \ldots, \pi_l$ be generators of the invariant ring of G (*fundamental invariants*). The projection

$$\begin{aligned} \pi: \mathbb{R}^n &\to \mathbb{R}^n//G \subseteq \mathbb{R}^l \\ x &\mapsto (\pi_1(x), \ldots, \pi_l(x)) \end{aligned}$$

defines an embedding of the orbit space into $\mathbb{R}^l$. In contrast to the complex case (see Cox et al. [9]) this map is not surjective in general. We highlight this phenomenon with the following example:

Example 6.1. Let $G = D_4$ be the dihedral group acting on $\mathbb{R}^2$. Fundamental invariants that generate $\mathbb{C}[X, Y]^{D_4}$ are given by $f_1 = x^2 + y^2$ and $f_2 = x^2y^2$ (for general methods to compute fundamental invariants we refer to Derksen and Kemper [12]). As f_1 and f_2 are, in fact, algebraically independent, we find that $\mathbb{C}^n//D_4 \simeq \mathbb{C}^2$. In the complex setting every solution to the linear system $z_1 = \alpha_1, z_2 = \alpha_2$ for some $(\alpha_1, \alpha_2) \in \mathbb{C}^2$ will give rise to an orbit of solutions of $f_1(x, y) = \alpha_1,\ f_2(x, y) = \alpha_2$. But since for optimization purposes we are interested in real solutions we have to restrict the map π to $\mathbb{R}^2$. Obviously, the image $\pi(\mathbb{R}^2)$ is contained in $\mathbb{R}^2$. On the other hand, as $f_1(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$, we have $\pi^{-1}((-1, 0)) \notin \mathbb{R}^2$. Hence, the restricted map $\pi_{|\mathbb{R}^2}$ is not surjective.

Therefore, to describe the image of $\mathbb{R}^n$ under π we need to add further constraints. In Example 6.1 for instance, the property $f_1(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^n$ implies $\pi(\mathbb{R}^2) \subseteq \{(z_1, z_2) \in \mathbb{R}^2: z_1 \geq 0\}$. Thus it seems promising to add such positivity constraints to characterize the image $\pi(\mathbb{R}^n)$ as a semialgebraic subset of $\mathbb{R}^n$. This is indeed possible and the characterization has been done by Procesi and Schwarz [28], who have determined polynomial inequalities that have to be taken additionally into account in order to characterize the embedding of $\mathbb{R}^n//G$ into the coordinate variety of $\mathbb{R}[X]^G$ of G (see also Bröcker [6]). We outline this briefly:

First, note that there exists a G-invariant inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}[X]$. For a polynomial p, the differential dp is defined by $dp = \sum_{j=1}^n (\partial p/\partial x_j) dx_j$. Then, carrying over the inner product to the differentials yields $\langle dp, dq \rangle = \sum_{j=1}^n \langle \partial p/\partial x_j, \partial q/\partial x_j \rangle$. The inner products $\langle d\pi_i, d\pi_j \rangle$ $(i, j \in \{1, \ldots, l\})$ are G-invariant. Hence, entry of the symmetric matrix

$$J = (\langle d\pi_i, d\pi_j \rangle)_{1 \leq i,\, j \leq l} \tag{24}$$

is G-invariant and can therefore be expressed in terms of $\pi_1, \ldots, \pi_l$.

Proposition 6.2 (Procesi and Schwarz [28]). *Let $G \subseteq \mathrm{GL}_n(\mathbb{R})$ be a finite group and $\pi = (\pi_1, \ldots, \pi_l)$ be fundamental invariants of G. Then the orbit space is given by polynomial inequalities,*

$$\mathbb{R}^n//G = \pi(\mathbb{R}^n) = \{z \in \mathbb{R}^l: J(z) \succeq 0,\, z \in V(I)\},$$

where $I \subseteq \mathbb{R}[z_1, \ldots, z_l]$ is the ideal of algebraic relations among $\pi_1, \ldots, \pi_l$.

Example 6.3. Continuing Example 6.1, we have $\partial f_1/\partial x = 2x$, $\partial f_1/\partial y = 2y$, $\partial f_2/\partial x = 2xy^2$, $\partial f_2/\partial y = 2x^2y$. Expressed in the original variables x, y, and in the fundamental invariants f_1, f_2, respectively, the matrix J from (24) is

$$J = \begin{pmatrix} 4(x^2+y^2) & 8x^2y^2 \\ 8x^2y^2 & 4(x^2y^4 + y^2x^4) \end{pmatrix} = \begin{pmatrix} 4f_1 & 8f_2 \\ 8f_2 & 4f_1f_2 \end{pmatrix}.$$

With the principle minors of J (which are $4f_1$, $4f_1f_2$, and $4f_1 \cdot 4f_1f_2 - (8f_2)^2$), the orbit space is given by

$$\mathbb{R}^2//D_4 = \{(z_1, z_2) \in \mathbb{R}^2: 4z_1 \geq 0, 4z_1z_2 \geq 0, 16z_1^2z_2 - (8z_2)^2 \geq 0\}.$$

Let $\tilde{f}$ and $\tilde{g}_1, \ldots, \tilde{g}_m$ be the expressions for f and $g_1, \ldots, g_m$ in the fundamental invariants. By Proposition 6.2, the G-symmetric optimization problem (1) can be equivalently expressed in the orbit space:

$$\begin{aligned} \inf \quad & \tilde{f}(z) \\ \text{s.t.} \quad & z \in V(I), \\ & \tilde{g}_j(z) \geq 0, \qquad 1 \leq j \leq m, \\ & J(z) \succeq 0. \end{aligned} \tag{25}$$

This is a PMI (as introduced in §2.2) and one can use the techniques introduced there to derive an SDP relaxation scheme. Let $s_1(z), \dots, s_r(z)$ be generators for the algebraic relations among $\pi_1, \dots, \pi_l$. Then (8) yields the hierarchy $Q_k^{//}$ of SDP relaxations

$$Q_k^{//}: \quad \begin{aligned} &\inf_y L(f) \\ &M_k(y) \succeq 0, \\ &M_{k-d}(Jy) \succeq 0, \\ &M_{k-\lceil \deg \tilde{g}_j/2\rceil}(\tilde{g}_j\, y) \succeq 0, \quad 1 \le j \le m, \\ &M_{k-\lceil \deg s_j/2\rceil}(s_j\, y) = 0, \quad 1 \le j \le r, \end{aligned} \tag{26}$$

where $k \ge k_0 := \max\{\lceil \deg \tilde{f}/2\rceil, \lceil \deg \tilde{g}/2\rceil, \lceil \deg s_j/2\rceil, d\}$ and $d = \max\{\lceil \deg J_{ij}(Z)/2\rceil\}$.

Theorem 6.4. *Let $f, g_1, \dots, g_m$ be G-invariant. If the PMI in (26) meets Assumption 2.5, then the sequence $(\inf Q_k^{//})_{k\ge k_0}$ is monotonically nondecreasing and converges to f^*.*

Proof. By Proposition 6.2 the problem described by f and $g_1, \dots, g_m$ is equivalent to (25). Now we can conclude with Proposition 2.6. □

Remark 6.5. It would be very interesting to characterize the situations where (6.2) meets Assumption 2.5 in terms of the original set K.

6.2. Lower and upper bounds for power sum problems. For constrained polynomial optimization problems described by power sums, the PMI become particularly simple. We will use the first relaxation of the sequence (26) in order to derive bounds for a particular class of problems.

Let $n, m, q \in \mathbb{N}$ with $q \ge m$, $m \le n+1$, and given some vector $\gamma \in \mathbb{R}^{m-1}$, consider the symmetric global optimization problem

$$\mathrm{P}_{nmq}\colon \ \min \sum_{i=1}^n x_i^q \qquad \text{s.t.} \quad \sum_{i=1}^n x_i^j = \gamma_k, \quad j = 1, \dots, m-1, \tag{27}$$

with optimal value denoted $\min \mathrm{P}_{nmq}$. Here, we provide upper and lower bounds for P_{nmq}.

Choose the fundamental invariants $\pi_j = (1/j)s_j$ $(1 \le j \le n)$, where $s_j := \sum_{i=1}^n x_i^j$ denotes the power sum of order j. Then the matrix $J(z)$ specializes to the Hankel matrix $H_n(s) = (s_{i+j-2})_{1\le i,\, j\le n}$.

We can exploit the double occurrence of power sums: within the optimization problem and within the Hankel matrix. Namely, for $m \le n+1$ and $m \le q \le 2n-2$, consider the following semidefinite optimization problem:

$$\mathrm{L}_{nmq} = \min_s \{s_q \mid H_n(s) \succeq 0;\ s_0 = n;\ s_j = \gamma_j,\ j = 1, \dots, m-1\}. \tag{28}$$

Theorem 6.6. *Let $n, m, q \in \mathbb{N}$ with $m \le n+1$, $m \le q \le 2n-2$, and let P_{nmq} be as in (27). Then one obtains the following lower bounds on $\min \mathrm{P}_{nmq}$.*

(a) $\min \mathrm{P}_{nmq} \ge \mathrm{L}_{nmq}$.

(b) *If $q = m = 2r$ for some r, write*

$$H_{r+1}(s) = \left(\begin{array}{c|c} H_r(\gamma) & u_r(\gamma) \\ \hline u_r^T(\gamma) & s_{2r} \end{array}\right); \qquad u_r(\gamma)^T = (\gamma_r, \dots, \gamma_{2r-1}),$$

with $\gamma_0 = n$. Then $\min \mathrm{P}_{nmq} \ge u_r(\gamma)^T H_r(\gamma) u_r(\gamma)$.

Proof. (a) Consider the equivalent formulation to (27) in the form (25). Since $J(z)$ is a Hankel matrix, every solution to this PMI is feasible for (28).

(b) In case $q = m = 2r < 2n$, we observe $r < n$ and

$$H_n(s) = \left(\begin{array}{c|c} H_{r+1}(s) & U(s) \\ \hline U^T(s) & V(s) \end{array}\right),$$

for some suitable (possibly empty) matrices $U(s) \in \mathbb{R}^{(r+1)\times(n-r-1)}$, $V(s) \in \mathbb{R}^{(n-r-1)\times(n-r-1)}$. Therefore, $H_n(s) \succeq 0$ implies $H_{r+1}(s) \succeq 0$, and the final result follows from Schur's complement applied to the Hankel matrix $H_{r+1}(s)$. □

In certain cases, we can complement this lower bound for problem (1) by an upper bound. The idea is to consider potential solutions $x \in \mathbb{R}^n$ of P_{nmq} with at most m nonzero components.

Consider the monic polynomial $p = X^m + \sum_{k=0}^{m-1} p_j X^j \in \mathbb{R}[X]$, and let $x_1, \ldots, x_m$ be the m roots (counting multiplicities) of p. A necessary and sufficient condition for all roots of p to be real is that $H_m(s) \succeq 0$, where $H_m(s)$ is the Hankel matrix with $s_0 = m$.

When $q \leq 2m-2$, we investigate the following SDP problem

$$\mathrm{U}_{nmq} = \min_s \{ s_q \mid H_m(s) \succeq 0;\, s_0 = m;\, s_j = \gamma_j,\ j = 1, \ldots, m-1 \}, \tag{29}$$

which is the same as (28) except that we now have a Hankel matrix $H_m(s)$ of dimension m instead of $H_n(s)$ of dimension n.

It is well known that the Newton sums $s_k = \sum_{j=1}^m X_j^k$, $k \geq 0$, of p are known polynomials in its coefficients $\{p_j\}$, and conversely, the coefficients p_j of p are polynomials in the s_j; i.e., we can write $s_j = P_j(p_0, \ldots, p_{m-1})$, $j \geq 0$, for some polynomials $P_j \in \mathbb{R}[p_0, \ldots, p_{m-1}]$.

In fact, the s_i and the p_j are related by Newton's identities:

$$\begin{aligned} s_k + p_{m-1}s_{k-1} + \cdots + p_0 s_{k-m} &= 0 \quad (k \geq m),\\ s_k + p_{m-1}s_{k-1} + \cdots + p_{m-k+1}s_1 &= -kp_{m-k} \quad (1 \leq k < m). \end{aligned}$$

If one knows s_j for all $j = 1, \ldots, m-1$, then one may compute the p_j for all $j = 1, \ldots, m-1$, and therefore, we can choose as the unknown of our problem the variable p_0 (the only (constant) coefficient of p that we do not know), and write

$$s_j = P_j(p_0, \ldots, p_{m-1}) = Q_j(p_0), \quad j = m, m+1, \ldots,$$

for some known polynomials $Q_j \in \mathbb{R}[p_0]$. We claim that Q_j is affine whenever $j \leq 2m-1$. Indeed, this follows from

$$\begin{aligned} s_m &= -s_0p_0 - s_1p_1 - \cdots - s_{m-1}p_{m-1},\\ s_{m+1} &= -s_1p_0 - \cdots - s_{m-1}p_{m-2} - s_m p_{m-1}\\ &= -s_1p_0 - \cdots - s_{m-1}p_{m-2} + p_{m-1}(s_0p_0 + s_1p_1 + \cdots + s_{m-1}p_{m-1})\\ &= -p_0(s_1 - s_0p_{m-1}) - p_1(s_2 - p_{m-1}s_1) - \cdots - p_{m-1}(s_m - p_{m-1}s_{m-1}),\\ s_{m+2} &= -s_2p_0 - \cdots - s_{m-1}p_{m-3} - s_m p_{m-2} - s_{m+1}p_{m-1}\\ &= -p_0(s_2 - p_{m-2}s_0 + p_{m-1}s_1 - s_0 p_{m-1}^2) - \cdots,\\ s_{m+3} &= -p_0(s_3 - s_0p_{m-3} + \cdots) - \cdots. \end{aligned}$$

Therefore, with $q \leq 2m-2$, the SDP problem (29) reads

$$\mathrm{U}_{nmq} = \min_{p_0} \{ Q_q(p_0) \colon H_m(s) \succeq 0 \}, \tag{30}$$

where $s_0 = m$ and all the entries s_j of $H_m(s)$ are replaced by their affine expression $Q_j(p_0)$ whenever $m \leq j \leq 2m-2$. This is an SDP with the single variable p_0 only.

Theorem 6.7. *Let n, m, $q \in \mathbb{N}$ with $m \leq n$ and $q \leq 2m-2$. Let P_{nmq} be as in (27) and let U_{nmq} be as in (30). Then $\min \mathrm{P}_{nmq} \leq \mathrm{U}_{nmq}$.*

In addition, if P_{nmq} has an optimal solution $x^* \in \mathbb{R}^n$ with at most m nonzero entries, then $\min \mathrm{P}_{nmq} = \mathrm{U}_{nmq}$ and so P_{nmq} has the equivalent convex formulation (30).

Proof. Let p_0 be an optimal solution of the SDP (30), and consider the monic polynomial $p \in \mathbb{R}[X]$ of degree m which satisfies the Newton identities with $s_j = \gamma_j$, $j = 1, \ldots, m-1$. The vector $x = (x_1, \ldots, x_m)$ of all its roots (counting multiplicities) is real because $H_m(s) \succeq 0$; i.e., its Hankel matrix $H_m(s)$ formed with its Newton sums s_j, $j = 1, \ldots, 2m-2$ (and $s_0 = m$), is positive semidefinite. Let $x^* = (x, 0, \ldots, 0) \in \mathbb{R}^n$. By definition of the Newton sums of p, one has $\sum_{i=1}^n (x_i^*)^k = \sum_{i=1}^m x_i^k = \gamma_k$, $k = 1, \ldots, m-1$, which shows that x^* is feasible for P_{nmq}. Therefore, $\mathrm{U}_{nmq} = s_q \geq \min \mathrm{P}_{nmq}$, the desired result. $\square$

Acknowledgments. The authors are very grateful for the comments of two anonymous referees which helped to improve the presentation. An earlier preprint version of this paper received attention and is referenced, e.g., in the surveys (Bachoc et al. [2], Laurent [26], Vallentin [39]).

References

[1] Babai L, Rónyai L (1990) Computing irreducible representations of finite groups. *Math. Comp.* 55(192):705–722.
[2] Bachoc C, Gijswijt DC, Schrijver A, Vallentin F (2012) Invariant semidefinite programs. Anjos MF, Lasserre JB, eds. *Handbook on Semidefinite, Conic and Polynomial Optimization* (Springer, New York), 219–270.
[3] Barvinok AI (1992) Combinatorial complexity of orbits in representations of the symmetric group. *Representation Theory and Dynamical Systems*, Vol. 9, Adv. Soviet Math. (Amer. Math. Soc., Providence, RI), 161–182.
[4] Barvinok AI, Vershik AM (1989) The representation theory methods in combinatorial optimization problems. *Soviet J. Comput. Systems Sci.* 27(5):1–7.
[5] Brandenberg R, Theobald T (2006) Radii minimal projections of polytopes and constrained optimization of symmetric polynomials. *Adv. Geom.* 6(1):71–83.
[6] Bröcker L (1998) On symmetric semialgebraic sets and orbit spaces. *Banach Center Publ.* 44:37–50.
[7] Cimprič J, Kuhlmann S, Scheiderer C (2009) Sums of squares and moment problems in equivariant situations. *Trans. Amer. Math. Soc* 361(2):735–765.
[8] Choi MD, Lam TY (1978) Extremal positive semidefinite forms. *Math. Ann.* 231(1):1–18.
[9] Cox D, Little J, O'Shea D (2007) *Ideals, Varieties, and Algorithms* (Springer, New York).
[10] Dabbaghian-Abdoly V (2005) An algorithm for constructing representations of finite groups. *J. Symb. Comp.* 39(6):671–688.
[11] de Klerk E, Pasechnik DV, Schrijver A (2007) Reduction of symmetric semidefinite programs using the regular $*$-representation. *Math. Program., Ser. B* 109(2–3):613–624.
[12] Derksen H, Kemper G (2002) *Computational Invariant Theory* (Springer, New York).
[13] Dixon JD (1993) Constructing representations of finite groups. *Groups and Computation*, Vol. 11, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. (Amer. Math. Soc., Providence, RI), 105–112.
[14] Gatermann K, Parrilo P (2004) Symmetry groups, semidefinite programs, and sums of squares. *J. Pure Appl. Algebra* 192(1–3):95–128.
[15] Gijswijt D (2005) *Matrix Algebras and Semidefinite Programming Techniques for Codes.* Ph.D. dissertation, Universiteit van Amsterdam.
[16] Henrion D, Lasserre JB (2005) Detecting global optimality and extracting solutions in GloptiPoly. Henrion D. Garulli A. eds. *Positive Polynomials in Control*, Vol. 312, Lecture Notes on Control and Information Sciences. (Springer-Verlag, Berlin), 293–310.
[17] Henrion D, Lasserre JB (2006) Convergent relaxations of polynomial matrix inequalities and static output feedback. *IEEE Trans. Autom. Control* 51(2):192–202.
[18] Hol, CWJ Scherer CW (2004) Sum of squares relaxations for polynomial semidefinite programming. *Proc. Symp. Mathematical Theory of Networks and Systems* (*MTNS*)*, Leuven, Belgium*.
[19] James G, Kerber A (1981) *The Representation Theory of the Symmetric Group* (Addison-Wesley, Reading, MA).
[20] Kanno Y, Ohsaki M, Murota M, Katoh N (2001) Group symmetry in interior-point methods for semidefinite programming. *Optim. Engrg.* 2(3):293–320.
[21] Lasserre JB (2001) Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* 11(3):796–817.
[22] Lasserre JB (2010) *Moments, Positive Polynomials and Their Applications* (Imperial College Press, London).
[23] Laurent M (2005) Revisiting two theorems of Curto and Fialkow on moment matrices. *Proc. Amer. Math. Soc.* 133:2965–2976.
[24] Laurent M (2007) Strengthened semidefinite programming bounds for codes. *Math. Program., Ser. B* 109(2–3):239–261.
[25] Laurent M (2007) Semidefinite representations for finite varieties. *Math. Program., Ser. A* 109(1):1–26.
[26] Laurent M (2009) Sums of squares, moment matrices and optimization over polynomials. Putinar M, Sullivant S, eds. *Emerging Applications of Algebraic Geometry*, Vol. 149, IMA Volumes in Mathematics and its Applications. (Springer, New York), 157–270.
[27] Parrilo PA (2003) Semidefinite programming relaxations for semialgebraic problems. *Math. Program., Ser. B* 96(2):293–320.
[28] Procesi C, Schwarz G (1985) Inequalities defining orbit spaces. *Invent. Math.* 81(3):539–554.
[29] Putinar M (1993) Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.* 42(3):969–984.
[30] Riener C (2012) On the degree and half degree principle for symmetric polynomials. *J. Pure Appl. Algebra* 216(4):850–856.
[31] Sagan B (2001) *The Symmetric Group*, Vol. 203, Graduate Texts in Mathematics. (Springer, New York).
[32] Schweighofer M (2005) Optimization of polynomials on compact semialgebraic sets. *SIAM J. Optim.* 15(3):805–825.
[33] Schrijver A (1979) A comparison of the Delsarte and Lovász bounds. *IEEE Trans. Inform. Theory* 25(4):425–429.
[34] Schrijver A (2005) New code upper bounds from the Terwilliger algebra and semidefinite programming. *IEEE Trans. Inform. Theory* 51(8):2859–2866.
[35] Serre J-P (2001) *Linear Representations of Finite Groups*, Vol. 42, Graduate Texts in Mathematics. (Springer, New York).
[36] Specht W (1937) Zur Darstellungstheorie der symmetrischen Gruppe. *Math. Z.* 42(1):774–779.
[37] The GAP Group (2008) *GAP—Groups, Algorithms, and Programming*. http://www.gap-system.org
[38] Timofte V (2003) On the positivity of symmetric polynomial functions. Part I: General results. *J. Math. Anal. Appl.* 284(1):174–190.
[39] Vallentin F (2009) Symmetry in semidefinite programs. *Linear Algebra Appl.* 430(1):360–369.