MOMENT METHODS IN ENERGY MINIMIZATION: NEW BOUNDS FOR RIESZ MINIMAL ENERGY PROBLEMS

DAVID DE LAAT

Abstract. We use moment methods to construct a converging hierarchy of optimization problems to lower bound the ground state energy of interacting particle systems. We approximate the infinite dimensional optimization problems in this hierarchy by block diagonal semidefinite programs. For this we develop the necessary harmonic analysis for spaces consisting of subsets of another space, and we develop symmetric sum-of-squares techniques. We compute the second step of our hierarchy for Riesz s-energy problems with five particles on the 2-dimensional unit sphere, where the $s = 1$ case known as the Thomson problem. This yields new sharp bounds (up to high precision) and suggests the second step of our hierarchy may be sharp throughout a phase transition and may be universally sharp for 5-particles on S^2. This is the first time a 4-point bound has been computed for a continuous problem.

1. Introduction

We consider the problem of finding the ground state energy of a system of interacting particles. An important example is the *Thomson problem*, where we minimize the sum

$$\sum_{1 \leq i < j \leq N} \frac{1}{\|x_i - x_j\|_2}$$

over all sets $\{x_1, \ldots, x_N\}$ of N distinct points in the unit sphere $S^2 \subseteq \mathbb{R}^3$. Here $\|x_i - x_j\|_2$ is the *chordal distance* between x_i and x_j. A simple optimality proof for the configuration consisting of three equally spaced particles on a great circle was given in 1912 [18], but for $N > 3$ we seem to require more involved techniques. In 1992, Yudin [51] introduced a beautiful method, based on earlier work for spherical codes by Delsarte, Goethals, and Seidel [16], which in addition to the $N \leq 3$ cases can be used to prove optimality for 4, 6, and 12 particles [1, 51]. Here the configurations are given by the vertices of the regular tetrahedron, octahedron, and icosahedron.

Yudin's bound is a relaxation of the above energy minimization problem. It is a simpler minimization problem whose optimal value lower bounds the ground state energy. This means that a feasible solution of the dual problem of this relaxation, which is a maximization problem in the form of an infinite dimensional linear program, provides a lower bound on the ground state energy. For $N = 2, 3, 4, 6, 12$ this bound is sharp, and the optimal dual solutions form optimality certificates; that is, the dual solutions become easy to check optimality proofs. In 2006, Cohn and Kumar [11] used this method in their proof of universal optimality of the above configurations (as well as many other configurations in higher dimensional spheres),

---

Date: 16 October 2016.

1991 *Mathematics Subject Classification.* 52C17, 90C22.

Key words and phrases. Thomson problem, Riesz s-energy, 4-point bounds, semidefinite programming, Lasserre hierarchy, harmonic analysis on spaces of subsets, invariant polynomials.

Centrum Wiskunde & Informatica (CWI), The Netherlands. Funded by Vidi grant 639.032.917 and TOP grant 617.001.351 from the Netherlands Organization for Scientific Research (NWO).

A preliminary version of this paper appeared in the author's PhD thesis [28].

where a configuration is said to be *universally optimal* if it is optimal for all *completely monotonic* (smooth, nonnegative functions whose derivatives alternate in sign) pair potentials in the squared chordal distance.

In the derivation of Yudin's bound one considers conditions on pairs of particles, and so this is called a 2-*point bound*. In 2012, Cohn and Woo [13] derived 3-point bounds for energy minimization based on earlier work by Schrijver [43] for binary codes and Bachoc and Vallentin [4] for spherical codes. They used this to prove universal optimality of the vertices of the rhombic dodecahedron in $\mathbb{RP}^2$. In [38] this is extended to k-point bounds, but here the sphere is required to be at least $k-1$ dimensional, which means this approach cannot be used to go beyond 3-point bounds for energy minimization on S^2.

In this paper (see Section 2) we construct a hierarchy $E_1, E_2, \ldots$ of increasingly strong relaxations for energy minimization. Each E_t is a minimization problem whose optimal value lower bounds the ground state energy E. To construct this hierarchy we use the moment methods developed in [30], which generalize techniques from the Lasserre hierarchy [32] in polynomial optimization to an infinite dimensional setting. We can interpret the t-th step E_t as a $\min\{2t, N\}$-point bound, and in Section 4 we prove convergence to the optimal energy in at most N steps. After symmetry reduction (see below), the first step E_1 becomes essentially the same as Yudin's bound. In Section 3 we show the derivation of the linear constraints in the optimization problems E_t. Here we derive enough constraints to ensure convergence of the hierarchy, but keep the constraint set small enough to allow for a satisfying duality theory, which is important for performing computations.

The problems E_t are infinite dimensional minimization problems where the optimization variables are Radon measures. Naturally, this implies the optimization variables in the dual maximization problems E_t^* are continuous functions. We show how to approximate the duals E_t^* by semidefinite programs that are block diagonalized into sufficiently small blocks so that it becomes possible to numerically compute the 4-point bound E_2 for interesting problems. This leads to the best known bounds for these problems, and this demonstrates the computational applicability of the moment techniques developed in [30].

In Section 5 we discuss a class of infinite dimensional conic optimization problems that occur naturally when forming moment relaxations of problems in infinitely many binary variables. The relaxations E_t fit into this framework, where each point in the container corresponds to a binary variable indicating whether this position is occupied by a particle. Similarly, the relaxations for geometric packing problems from [30] fit into this framework. To find good lower bounds we need to find good feasible solutions to the dual optimization problems, and for this we discuss duality and symmetry reduction for this more general class of problems. In particular, we give a sufficient criterion for convergence of the optimal values when we approximate the cone in the dual programs by simpler cones.

To find good feasible solutions of the dual problems E_t^* we use harmonic analysis, sum-of-squares characterizations, and semidefinite programming. These tools are also used for computing the 2 and 3-point bounds mentioned above, but for $t > 1$, our problems E_t^* are of a rather different form. In the 2 and 3-point bounds for energy minimization on the sphere, the dual variables are positive definite kernels $K \colon S^2 \times S^2 \to \mathbb{R}$. By *positive definite* we mean that the matrices $(K(x_i, x_j))_{i,j=1}^n$ are positive semidefinite for all $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in S^2$. For 2-point bounds these kernels can be assumed to be invariant under the orthogonal group $O(3)$ and for 3-point bounds under the stabilizer subgroup of a point in S^2. In the dual programs E_t^*, however, the variables are positive definite kernels $K \colon I_t \times I_t \to \mathbb{R}$, where I_t is the set of independent sets of size at most t in a graph G on $V = S^2$. Here G

is a topological packing graph, where there is an edge between vertices that are close, so that we exclude configurations that are already known to be nonoptimal. This graph G inherits the symmetry of the problem, and we may assume K to be invariant under this symmetry. This means that to compute E_t^*, we have to optimize over the cone of Γ-invariant, positive definite kernels $K\colon I_t \times I_t \to \mathbb{R}$.

We give a nonconstructive proof that for each compact, metrizable container V (in particular, $V = S^2$), this cone of Γ-invariant, positive definite kernels is equal to the closure of the union of a sequence of certain simpler inner approximating cones (see Section 6 and Appendix A). Each of these inner approximating cones can be parametrized by a finite product of positive semidefinite matrix cones. By using harmonic analysis and symmetric tensor powers we show how such a sequence can be constructed explicitly. We carry out this construction for the case where $V = S^2$ and $t = 2$. In this way we obtain an explicit sequence of optimization problems $E_{2,d}^*$, each having finite dimensional variable space, whose optimal values lower bound and converge to the optimal value of the 4-point bound $E_2^* = E_2$.

In Section 7 we use invariant theory to write the problems $E_{t,d}^*$, which have finitely many variables but infinitely many constraints, as semidefinite programs with semialgebraic constraints. Here we use *semidefinite programming* (the optimization of a linear functional over the intersection of an affine space with a cone of positive semidefinite matrices), because there exist efficient algorithms to solve these problems numerically. By a *semialgebraic constraint* we mean the requirement that a polynomial, whose coefficients depend linearly on the entries of the positive semidefinite matrix variable(s), is nonnegative on a basic closed semialgebraic set. We model these semialgebraic constraints as semidefinite constraints by using sum-of-squares characterizations from real algebraic geometry. In this way we obtain a sequence of semidefinite programs $E_{t,d,\delta}^*$, whose optimal values lower bound and converge to $E_{t,d}^*$ as the sum-of-squares degree δ goes to infinity.

A semidefinite program is said to be *block diagonalized* if we can write it as the optimization of a linear functional over the intersection of an affine space with a finite product of positive semidefinite matrix cones. Block diagonalization is important because the complexity of solving a semidefinite program depends strongly on the size of the largest block. As described above, the problems $E_{t,d}$ are already block diagonalized through the use of harmonic analysis, which exploits the symmetry of the container and pair potential. In Section 8 we use additional symmetry, the interchangeability of the particles, to derive symmetries in the semialgebraic constraints in these problems. We then give a symmetrized version of Putinar's theorem, which allows us to exploit symmetries in semialgebraic constraints. This can lead to significantly smaller block sizes, and, as we show by applying this to the problems $E_{2,d,\delta}^*$, this can lead to significant computational savings.

Although the Nth relaxation E_N is guaranteed to give the ground state energy E, the advantage of using a hierarchy is that E_t can already be sharp for much smaller values of t. For example, Yudin's bound, which is essentially equal to the symmetry reduced version of E_1^*, is sharp for the Thomson problem for $N = 2, 3, 4, 6, 12$. It would be very interesting if this pattern continues; that is, if E_2 would be sharp for several new values of N. The 3-point bound is conjectured [13] to be sharp for $N = 8$, and since E_2 is a 4-point bound it should also be sharp for $N = 8$. As a first step into investigating whether E_2 is sharp for new values of N – and to demonstrate that it is possible to compute the second step of our hierarchy – we compute $E_{2,6,6}^*$ numerically (with high precision) for the five particle case of the Thomson problem. The optimal value given by the semidefinite programming solver, consisting of 28 decimal digits, coincides with the first 28 decimal digits of the energy of the configuration consisting of the vertices of the triangular bipiramid, which is a strong indication that the bound is sharp. That is, the inequalities in

$$E_{2,6,6}^* \leq E_{2,6}^* \leq E_2^* = E_2 \leq E$$

are attained. This is the first time a 4-point bound has been computed for a continuous problem.

The case of 5 particles on S^2 is particularly interesting because it is one of the simplest mathematical models that admits a phase transition, where by a *phase transition* we mean that a slight change of the pair potential results in a discontinuous jump from one global optimum to another. The *Riesz s-energy* of a configuration $\{x_1, \ldots, x_N\} \subseteq S^2$ is given by

$$\sum_{1 \leq i < j \leq N} \frac{1}{\|x_i - x_j\|_2^s}.$$

In [35] it is conjectured that the configuration consisting of the vertices of the triangular bipyramid is optimal for $0 < s \leq s^* \approx 15.048$, and the configuration consisting of the vertices of the square pyramid (where the latitude of the base depends on the value of s) is optimal for $s \geq s^*$. Optimality has been proven for $s = 1$ and $s = 2$ by using Hessian bounds and essentially enumerating all possibilities [44], and recently, Schwartz [45] extended his result to the entire interval $0 \leq s \leq 6$ using an observation of Tumanov [48]. In this paper we are interested in finding sharp dual solutions, which can be used (see, for instance, [13]) to generate easily verifiable optimality proofs; see also the discussion at the end of Section 9. In addition to the $s = 1$ case we compute E_2 for $s = 2, \ldots, 7$, where $E_{2,6,6}^*$ is (numerically) sharp for $s = 1, \ldots, 5$ and $E_{2,6,8}^*$ is (numerically) sharp for $s = 6, 7$. Here, again, we verify that the 28 decimal digits given by the solver agree with the first 28 digits of the energy of a configuration. Since we need to increasy the parameter δ when s gets larger, we have not been able to compute other sharp bounds. Based on the above (strong) evidence we have the following conjecture.

Conjecture 1.1. *The bound E_2 is sharp for the minimal Riesz s-energy of* 5 *particles on S^2 for $s = 1, \ldots, 7$.*

In [13, Conjecture 15] it is conjectured that if there exists a completely monotonic potential function (in the squared chordal distance) for which a k-point bound is sharp for N particles on S^{n-1}, and if this function is not a polynomial, then this k-point bound is universally sharp for N particles on S^{n-1}. Here by *universally sharp* we mean that the bound gives the ground state energy for all completely monotonic pair potentials in the squared chordal distance. This means that if Conjecture 1.1 is true for at least one value of s, and if [13, Conjecture 15] is true for our version E_2 of 4-point bounds for energy minimization, then the bound E_2 is sharp for all positive s, and is thus is sharp throughout a phase transition. In fact, then the following stronger conjecture holds.

Conjecture 1.2. *The bound E_2 is universally sharp for* 5 *particles on S^2.*

Given a particle configuration on a sphere (or more generally, on a 2-point homogeneous space), there exists a beautiful sufficient criterion, based on geometric design theory, that says Yudin's bound is sharp for this configuration [11, 33]. No such criterion is known for k-point bounds with $k > 2$. Together with the newly found (numerically) sharp instances and other sharp instances for E_2 possibly still to be discovered (see Section 9), the new approach in this paper to formulating k-point bounds for energy minimization may help in formulating such a sufficient criterion. This may help in uncovering the geometric ideas behind optimality and universal optimality of particle configurations.

2. A hierarchy of relaxations for energy minimizaton

In this section we derive a hierarchy of relaxations for energy minimization. We model the space containing the particles by a compact metric space (V, d), and assume the pair potential is given by a continuous function $h\colon (0, \mathrm{diam}(V)] \to \mathbb{R}$, where we assume that $h(s) \to \infty$ as $s \downarrow 0$. This natural assumption is mainly made for convenience as it avoids having to work with multisets. We denote the number of particles in the system by N. The *ground state energy* is given by the minimizing

$$\sum_{1 \leq i < j \leq N} h(d(x_i, x_j))$$

over all sets $\{x_1, \ldots, x_N\}$ of N distinct points in V. For the Thomson problem we have $V = S^2 \subseteq \mathbb{R}^3$, $d(x, y) = \|x - y\|_2$, and $h(c) = 1/c$ (the *Coulomb energy*).

To compactify this problem, which will be important when we discuss duality, we introduce a graph that allows us to discard some configurations that are already known to be nonoptimal. Let B be an upper bound on the minimal energy. Such a number can be obtained by computing the energy of an arbitrary configuration of N distinct points. We now let G be the graph with vertex set V, where distinct vertices x and y are adjacent whenever $h(d(x, y)) > B$. This ensures that the optimal N point configurations are among the independent sets of this graph, where an *independent set* is a subset of the vertices for which no two vertices are adjacent. Alternatively we can use a result such as [27], which states that for certain energy potentials the points of an optimal N-point configuration must be at least a certain distance D apart form each other, and we can define the graph G by letting x and y be adjacent whenever $d(x, y) < D$. Since this has the potential to exclude more configurations this approach might lead to stronger relaxations.

Let I_t be the set of independent sets of cardinality at most t. We endow $I_t \setminus \{\emptyset\}$ with a topology as a subset of the quotient space V^t/q, where q maps a tuple $(x_1, \ldots, x_t)$ to the set $\{x_1, \ldots, x_t\}$. We endow I_t with the disjoint union topology by $I_t = I_t \setminus \{\emptyset\} \cup \{\emptyset\}$; that is, the set $\emptyset$ is an isolated point in I_t. Let $I_{=t}$ be the set of independent sets of cardinality t, and endow $I_{=t}$ with the topology as a subset of I_t. The graph G is an example of a compact topological packing graph as defined in [30]. A *topological packing graph* is a graph whose vertex set is a Hausdorff topological space where each clique is contained in an open clique. It follows that the sets I_t and $I_{=t}$ are compact metric spaces. From the definition of G it follows that $I_{=N}$ is nonempty.

The ground state energy can be computed as

$$E = \min_{S \in I_{=N}} \chi_S(f).$$

Here $f \in \mathcal{C}(I_N)$ is defined as

$$f(s) = \begin{cases} h(d(x, y)) & \text{if } S = \{x, y\} \text{ with } x \neq y, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\chi_S = \sum_{R \subseteq S} \delta_R,$$

where δ_R is the Dirac point measure at R (so that $\delta_R(f) = f(R)$). Here $\mathcal{C}(I_N)$ is the space of real-valued, continuous functions on I_N. The continuity of f follows from the continuity of h and the fact that $I_{=2}$ is both open and closed in I_N (see [30]). Since we minimize the continuous function $S \mapsto \chi_S(f)$ over a compact set, the minimum above is attained.

To obtain energy lower bounds we construct a hierarchy $E_1, E_2, \ldots$ of relaxations of the above problem. These are minimization problems such that for each feasible solution of E (a configuration of N particles) we can immediately construct a feasible solution of E_t having the same objective value. The problems E_t have the important feature that we can give their dual optimization problems in an explicit form, which is crucial for performing computations, and that we can prove the duality gap to be zero. In Section 4 we show that the Nth step E_N in this hierarchy gives the ground state energy, and that the extreme points of the feasible set of E_N are precisely the measures χ_S with $S \in I_{=N}$. That is, we show E_N is a *sharp* relaxation of E. The problems E_t are convex optimization problems, and we say E_N is a *convexification* of E.

Denote by $\mathcal{M}(I_{2t})$ the space of signed Radon measures on I_{2t}. Given an independent set $S \in I_{=N}$, define $\lambda_S \in \mathcal{M}(I_{2t})$ by restricting χ_S to I_{2t} if $2t \leq N$, or by extending with zeros if $2t \geq N$. In the t-th step E_t we will optimize over measures $\lambda \in \mathcal{M}(I_{2t})$ that we require to satisfy three properties that are satisfied by λ_S. The first of these properties is that λ_S is a positive measure. The second property is that

$$\lambda_S(I_{=i}) = \binom{N}{i} \quad \text{for all} \quad 0 \leq i \leq 2t,$$

where $\binom{N}{i} = 0$ for $i > N$. The third property is more subtle: The measure λ_S satisfies a moment condition. We use the moment techniques from [30] to define what we mean by this. Define the operator

$$A_t \colon \mathcal{C}(I_t \times I_t)_{\mathrm{sym}} \to \mathcal{C}(I_{2t}),\ A_t K(S) = \sum_{\substack{J,J' \in I_t \\ J \cup J' = S}} K(J, J'),$$

where $\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$ is the space of symmetric, continuous functions $I_t \times I_t \to \mathbb{R}$, which we call *symmetric kernels*. A kernel K is said to be *positive definite* if the matrices $(K(J_i, J_j))_{i,j=1}^n$ are positive semidefinite for all $n \in \mathbb{N}$ and $J_1, \ldots, J_n \in I_t$. The positive definite kernels form a convex cone which we denote by $\mathcal{C}(I_t \times I_t)_{\succeq 0}$. By the Riesz representation theorem, the topological duals of $\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$ and $\mathcal{C}(I_{2t})$ can be identified with the spaces $\mathcal{M}(I_t \times I_t)_{\mathrm{sym}}$ and $\mathcal{M}(I_{2t})$. Here $\mathcal{M}(I_t \times I_t)_{\mathrm{sym}}$ consists of the symmetric Radon measures, which are the measures μ that satisfy

$$\mu(E \times F) = \mu(F \times E) \quad \text{for all Borel sets} \quad E, F \subseteq I_t.$$

The dual cone of $\mathcal{C}(I_t \times I_t)_{\succeq 0}$ is defined by

$$\mathcal{M}(I_t \times I_t)_{\succeq 0} = \Big\{ \mu \in \mathcal{M}(I_t \times I_t)_{\mathrm{sym}} : \mu(K) \geq 0 \text{ for all } K \in \mathcal{C}(I_t \times I_t)_{\succeq 0} \Big\},$$

and we call the elements in this cone *positive definite measures*. We have the adjoint operator

$$A_t^* \colon \mathcal{M}(I_{2t}) \to \mathcal{M}(I_t \times I_t)_{\mathrm{sym}},$$

which is defined by $A_t^* \lambda(K) = \lambda(A_t K)$ for all $\lambda \in \mathcal{M}(I_{2t})$ and $K \in \mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$. We use this dual operator and the cone of positive definite measures to define the moment condition on λ:

Definition 2.1. *A measure $\lambda \in \mathcal{M}(I_{2t})$ is of positive type if*

$$A_t^* \lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0}.$$

See [30, Remark 1] for an explanation why we use the term positive type here. The measure λ_S defined above is of positive type: For each $K \in \mathcal{C}(I_t \times I_t)_{\succeq 0}$, we have

$$A_t^* \lambda_S(K) = \sum_{R \subseteq S} \sum_{\substack{J,J' \in I_t \\ J \cup J' = R}} K(J, J') = \sum_{\substack{J,J' \in S \\ |J|, |J'| \leq t}} K(J, J') \geq 0.$$

We define the t-th step in our hierarchy by optimizing over measures $\lambda \in \mathcal{M}(I_{2t})$ satisfying the three properties discussed above.

Definition 2.2. *For $t \in \mathbb{N}$, define*

$$E_t = \min\Big\{\lambda(f) : \lambda \in \mathcal{M}(I_{2t}) \text{ positive and of positive type}, \\ \lambda(I_{=i}) = \tbinom{N}{i} \text{ for } 0 \le i \le 2t\Big\}.$$

In Section 5.1 we prove strong duality, which implies the minimum here is attained. By construction, the measure λ_S is feasible for E_t, so $E_t \le E$ for all t. A similar argument shows $E_t \le E_{t+1}$ for all t. In Section 4 we prove $E_N = E$.

3. Connection to the Lasserre hierarchy

We originally derived the hierarchy $\{E_t\}$ by applying a variation of the Lasserre hierarchy from polynomial optimization to energy minimization problems where the container V is finite. We then reformulated (and possibly weakened, but still preserving convergence) the constraints in the resulting relaxations into a form that allows for a useful generalization to the case where V is infinite.

A *polynomial optimization problem* is a problem of the form

$$\inf\Big\{p(x) : x \in \mathbb{R}^n,\, g_j(x) \ge 0 \text{ for } j \in [m]\Big\},$$

where $p, g_1, \dots, g_m \in \mathbb{R}[x_1, \dots, x_n]$. In general, finding the global minimum and proving optimality of a point $x \in \mathbb{R}^n$ are both difficult problems. A powerful and popular approach of obtaining lower bounds on the global minimum is to use the Lasserre hierarchy [31], which is a sequence of increasingly strong semidefinite programming relaxations.

We are interested in *binary* polynomial optimization problems, which are of the form

$$\inf\Big\{p(x) : x \in \{0,1\}^n,\, g_j(x) \ge 0 \text{ for } j \in [m]\Big\},$$

where $[m] = \{1, \dots, m\}$. These problems are special cases of polynomial optimization problems, because we can enforce the constraints $x \in \{0,1\}^n$ by adding the polynomial constraints $x_i(1-x_i) \ge 0$ and $-x_i(1-x_i) \ge 0$ for each $i \in [n]$. If we assume the container of an energy minimization problem to be finite, say, $V = [n]$ for some $n \in \mathbb{N}$, then we can write the problem E as the binary polynomial optimization problem

$$\min\Big\{\sum_{1 \le i < j \le n} f(\{i,j\}) x_i x_j : x \in \{0,1\}^n,\, \kappa(x) = 0\Big\},$$

where $\kappa(x) = \sum_{i=1}^n x_i - N$, and where f is the pair potential as defined in Section 2.

Let G be the graph with vertex set $[n]$ and no edges. Then I_t is the set of all subsets of $[n]$ of cardinality at most t. In a binary polynomial optimization problem, we may assume all polynomials to be square free, and for such a polynomial we write

$$p(x) = \sum_{S \in I_{\deg(p)}} p_S x^S, \quad \text{where} \quad x^S = \prod_{i \in S} x_i.$$

Given an integer $t \in \mathbb{N}$ and a vector $y \in \mathbb{R}^{I_{2t}}$, define the t-th *moment matrix* $M_t(y) \in \mathbb{R}^{I_t \times I_t}$ by $M_t(y)_{J,J'} = y_{J \cup J'}$. We define the t-th *localizing matrix* with respect to a polynomial $g \in \mathbb{R}[x_1, \dots, x_n]$ to be the partial matrix $M_t^g(y)$, which has the same row and column indices as $M_t(y)$, where the (J, J')-entry is set to

$$\sum_{R \in I_{\deg(g)}} y_{J \cup J' \cup R}\, g_R$$

whenever $|J \cup J'| \le 2t - \deg(g)$, and left unspecified otherwise. By $M_t^p(y) \succeq 0$ we mean that y is a vector such that $M_t^p(y)$ can be completed to a positive semidefinite matrix (this is a semidefinite constraint on y). Using these definitions we define for $t \geq \deg(p)$ the following semidefinite programming relaxation of the binary polynomial optimization problem given above:

$$\inf\Big\{ \sum_{S \in I_{\deg(p)}} p_S y_S : y \in \mathbb{R}^{I_{2t}}_{\geq 0},\, y_\emptyset = 1,\ M_t(y) \succeq 0,\\ M_t^{g_j}(y) \succeq 0 \text{ for } j \in [m]\Big\}.$$

These relaxations were introduced by Lasserre in [32]. The only modifications we make here is that we restrict y to be nonnegative, and originally the localizing matrices $M_t^{g_j}(y)$ are defined to be full matrices indexed by $I_{t-\lceil \deg(g_j)/2 \rceil}$, but here we take them to be partial matrices indexed by I_t. These make the semidefinite programs only slightly more difficult to solve, but in some cases, such as the case of energy minimization as discussed here, it leads to much stronger bounds.

In the binary polynomial optimization formulation for energy minimization we have two polynomial constraints: $\kappa(x) \geq 0$ and $-\kappa(x) \geq 0$. So, in the relaxation we have the constraints $M_t^\kappa(y) \succeq 0$ and $-M_t^\kappa(y) = M_t^{-\kappa}(y) \succeq 0$. This reduces to $M_t^\kappa(y) = 0$; that is, all specified entries of $M_t^\kappa(y)$ are required to be zero. These constraints reduce to the linear constraints

$$N y_S = \sum_{j=1}^n y_{S \cup \{j\}} \quad \text{for all} \quad S \in I_{2t-1}.$$

So, for energy minimization, we get the relaxations

$$L_t = \inf\Big\{ \sum_{1 \leq i < j \leq n} f(\{i,j\})\, y_{\{i,j\}} : y \in \mathbb{R}^{I_{2t}}_{\geq 0},\, y_\emptyset = 1,\ M_t(y) \succeq 0,\\ N y_S = \sum_{j=1}^n y_{S \cup \{j\}} \quad \text{for} \quad S \in I_{2t-1}\Big\}.$$

The linear constraints in these problems become problematic when we want to generalize V from the finite set $[n]$ to an uncountable set. This is because in the infinite dimensional generalization we want to use measures $\lambda \in \mathcal{M}(I_{2t})$ instead of vectors $y \in \mathbb{R}^{I_{2t}}$ (because this allows for a satisfying duality theory; see Section 5.1), and these then become uncountably many "thin" constraints on λ. By thin we mean that the constraints are of the form $\lambda(E) = b$, where E is a set with empty interior, which means these constraints have no grip on the part of a measure that is zero on sets with empty interior.

In the following lemma we show these constraints imply $2t + 1$ very natural constraints on y. In particular, this lemma implies that for a feasible solution y of L_t, we have $y_S = 0$ for all $S \subseteq [n]$ with $|S| > N$. If we replace the linear constraints in L_t by these induced constraints, then we obtain the problem E_t as defined in the previous section for the case where $V = [n]$.

Lemma 3.1. *Let $t \in \mathbb{N}_0$ and $y \in \mathbb{R}^{I_{2t}}$. If*

$$y_\emptyset = 1 \quad \textit{and} \quad N y_S = \sum_{j=1}^n y_{S \cup \{j\}} \quad \textit{for all} \quad S \in I_{2t-1},$$

then

$$\sum_{S \in I_{=i}} y_S = \binom{N}{i} \quad \textit{for all} \quad 0 \leq i \leq 2t.$$

Proof. For $i = 0$ we have

$$\sum_{S \in I_{=i}} y_S = y_\emptyset = 1 = \binom{N}{i}.$$

If $\sum_{S\in I_{=i-1}} y_S = \binom{N}{i-1}$ for some $0 \le i \le 2t-1$, then

$$\begin{aligned}\sum_{S\in I_{=i}} y_S &= \frac{1}{i}\sum_{S\in I_{=i-1}}\sum_{j\in[n]\setminus S} y_{S\cup\{j\}} = \frac{1}{i}\sum_{S\in I_{=i-1}}\left(\sum_{j=1}^{n} y_{S\cup\{j\}} - |S|y_S\right)\\ &= \frac{1}{i}\sum_{S\in I_{=i-1}}\big(Ny_S-(i-1)y_S\big) = \frac{1}{i}\sum_{S\in I_{=i-1}}(N-i+1)y_S\\ &= \frac{N-i+1}{i}\sum_{S\in I_{=i-1}} y_S = \frac{N-i+1}{i}\binom{N}{i-1} = \binom{N}{i}.\end{aligned}$$

Hence, the proof follows by induction. □

4. Convergence to the ground state energy

In this section we show that the hierarchy $\{E_t\}$ converges to the optimal energy E in at most N steps. Moreover, the extreme points of the feasible set of E_N are precisely the measures χ_S with $S \in I_{=N}$. These results follow from the following proposition, whose proof follows directly from the proof of [30, Proposition 4.1].

Proposition 4.1. *For each measure $\lambda \in \mathcal{M}(I_{2t})$ there exists a unique measure $\sigma \in \mathcal{M}(I_{2t})$ such that $\lambda = \int \chi_S \, d\sigma(S)$. If λ is supported on I_t and is of positive type, that is, $A_t^*\lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0}$, then σ is a positive measure supported on I_t.*

Using this proposition we can prove the convergence result:

Proposition 4.2. *The Nth step E_N gives the optimal energy E.*

Proof. Let $\lambda \in \mathcal{M}(I_{2N})$ be feasible for E_N. We have $\lambda \ge 0$ and $\lambda(I_{=i}) = \binom{N}{i} = 0$ for $i > N$, so λ is supported on I_N. Since λ is also of positive type, by Proposition 4.1 there exists a positive measure $\sigma \in \mathcal{M}(I_N)$ such that $\lambda = \int \chi_S \, d\sigma(S)$. We have

$$1 = \tbinom{N}{0} = \lambda(\{\emptyset\}) = \int \chi_S(\{\emptyset\})\, d\sigma(S) = \int d\sigma = \sigma(I_N),$$

so σ is a probability measure. Moreover,

$$1 = \tbinom{N}{N} = \lambda(I_{=N}) = \int \chi_S(I_{=N})\, d\sigma(S) = \sigma(I_{=N}),$$

so σ is supported on $I_{=N}$. The objective value of λ is given by

$$\lambda(f) = \int \chi_S(f)\, d\sigma(S) \ge \int E \, d\sigma = E,$$

where the inequality follows since $\chi_S(f) \ge E$ for all $S \in I_{=N}$. It follows that $E_N \ge E$. Since we already known $E_N \le E$, this completes the proof. □

Using the ideas of the above proof together with the proof of [30, Proposition 4.1] it follows that the extreme points of the feasible set of E_N are precisely the measures χ_S with $S \in I_{=N}$.

5. Optimization with infinitely many binary variables

We discuss the duality theory and symmetry reduction for a more general type of problems that arise naturally when forming moment relaxations of optimization problems with infinitely many binary variables. This includes the moment relaxations for both energy minimization and packing problems. Although there are infinitely many variables, we assume that in a feasible solution only finitely many of them active (nonzero) at the same time, and active variables cannot be too close. For this we assume $G = (V, E)$ to be a compact *topological packing graph* as discussed in Section 2.

Definition 5.1. *Let G be a topological packing graph. Given integers t and m, functions $f, g_1, \ldots, g_m \in \mathcal{C}(I_{2t})$, and scalars $b_1, \ldots, b_m \in \mathbb{R}$, we define the optimization problem $H = H_{G,t}^{\inf}(h; g_1, \ldots, g_m; b_1, \ldots, b_m)$ by*

$$H = \inf \Big\{ \lambda(f) : \lambda \in \mathcal{M}(I_{2t})_{\geq 0},\, A_t^* \lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0},\, \lambda(g_j) = b_j \text{ for } j \in [m] \Big\}.$$

For energy minimization we have

$$E_t = H_{G,t}^{\min}(f; 1_{I=0}, \ldots, 1_{I=2t}; \tbinom{N}{0}, \ldots, \tbinom{N}{2t}),$$

where G and f are the graph and potential function as defined in Section 2. For packing problems in discrete geometry, the t-th step of the hierarchy from [30] is given by $H_{G,t}^{\max}(1_{I=1}; 1_{\{\emptyset\}}, 1)$, where G is the topological packing graph defining the packing problem.

5.1. **Duality.** The optimization problem H is a conic program over the cone

$$\mathcal{M}(I_t \times I_t)_{\succeq 0} \times \mathcal{M}(I_{2t})_{\geq 0},$$

where we refer to [5] for an introduction to conic programming. If we endow both $\mathcal{M}(I_t \times I_t)_{\mathrm{sym}}$ and $\mathcal{M}(I_{2t})$ with the weak* topologies, then the topological dual spaces can be identified with $\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$ and $\mathcal{C}(I_{2t})$. The tuples $(\mathcal{C}(I_{2t}), \mathcal{M}(I_{2t}))$ and $(\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}, \mathcal{M}(I_t \times I_t)_{\mathrm{sym}})$ are dual pairs, and the dual pairings

$$\langle f, \lambda \rangle = \lambda(f) = \int f(S)\, d\lambda(S) \quad \text{and} \quad \langle K, \mu \rangle = \mu(K) = \int K(J, J')\, d\mu(J, J')$$

are nondegenerate. The dual cones are then given by $\mathcal{C}(I_t \times I_t)_{\succeq 0}$ and $\mathcal{C}(I_{2t})_{\geq 0}$, and by conic duality we obtain the dual conic program

$$H^* = \sup \Big\{ \sum_{i=1}^m b_i a_i : a \in \mathbb{R}^m,\, K \in \mathcal{C}(I_t \times I_t)_{\succeq 0},\, f - \sum_{i=1}^m a_i g_i - A_t K \in \mathcal{C}(I_{2t})_{\geq 0} \Big\}.$$

By weak duality we have $H^* \leq H$. The following theorem, which is a slight generalization of the results in [30, Chapter 3], gives a sufficient condition for strong duality.

Theorem 5.2. *If H admits a feasible solution, and if the set*

$$\Big\{ \lambda \in \mathcal{M}(I_{2t})_{\geq 0} : A_t^* \lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0},\, \lambda(f) = \lambda(g_1) = \cdots = \lambda(g_m) = 0 \Big\},$$

is trivial, then strong duality holds: $H = H^*$ and the minimum in H is attained.

Proof. To show that strong duality holds we use a closed cone condition described in for instance [5]. This closed cone condition says that if H admits a feasible solution, and the cone

$$K = \big\{ (A_t^* \lambda - \mu, \lambda(g_1), \ldots, \lambda(g_m), \lambda(f)) : \lambda \in \mathcal{M}(I_{2t})_{\geq 0},\, \mu \in \mathcal{M}(I_t \times I_t)_{\succeq 0} \big\}$$

is closed in $\mathcal{M}(I_t \times I_t)_{\succeq 0} \times \mathbb{R}^m \times \mathbb{R}$, then strong duality holds: $H = H^*$ and the minimum in H is attained.

This cone K decomposes as the Minkowski difference $K = K_1 - K_2$, with

$$K_1 = \{ (A_t^* \lambda, \lambda(g_1), \ldots, \lambda(g_m), \lambda(f)) : \lambda \in \mathcal{M}(I_{2t})_{\geq 0} \}$$

and

$$K_2 = \{ (\mu, 0, 0) : \mu \in \mathcal{M}(I_t \times I_t)_{\succeq 0} \}.$$

By Klee [25] and Dieudonné [17], a sufficient condition for the cone K to be closed is if $K_1 \cap K_2 = \{0\}$, K_1 is closed and locally compact, and K_2 is closed. The first condition $K_1 \cap K_2 = \{0\}$ follows immediately from the hypothesis of the theorem. In [30] it is shown that K_1 is closed and locally compact. That K_2 is closed follows immediately from $\mathcal{M}(I_t \times I_t)_{\succeq 0}$ being closed. ☐

In [30, Lemma 3.1.5] it is shown that the set

$$\big\{\lambda \in \mathcal{M}(I_{2t})_{\geq 0} : A_t^* \lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0},\, \lambda(\{\emptyset\}) = 0\big\}$$

is trivial, which means Theorem 5.2 applies to the case where each $\lambda \in \mathcal{M}(I_{2t})_{\geq 0}$ with $\lambda(f) = \lambda(g_1) = \cdots = \lambda(g_m) = 0$ satisfies $\lambda(\{\emptyset\}) = 0$. For each t, the program E_t admits a feasible solution (see Section 2), and E_t satisfies this property by the constraint $\lambda(I_{=0}) = \binom{N}{0}$. Thus for every t strong duality holds for the pair (E_t, E_t^*).

5.2. **Symmetry reduction.** Given a compact group Γ with a continuous action on the vertex set V of a compact topological packing graph G, we say the optimization problem $H_{G,t}^{\inf}(f; g_1, \ldots, g_m; b_1, \ldots, b_m)$ is Γ-*invariant* if

(1) the adjacency relations in G are invariant under the action of Γ, so that the action extends to a continuous action on I_N given by $\gamma\emptyset = \emptyset$ and $\gamma\{x_1, \ldots, x_N\} = \{\gamma x_1, \ldots, \gamma x_N\}$;
(2) the functions $f, g_1, \ldots, g_m$ are Γ-invariant.

Using this definition, the relaxations $E_t = H_{G,t}^{\min}(f; 1_{I=0}, \ldots, 1_{I=2t}; \binom{N}{0}, \ldots, \binom{N}{2t})$ are Γ-invariant whenever the metric d of the container V is Γ-invariant.

We use this symmetry to restrict to invariant variables in both the primal and dual optimization problems. To make the best use of the symmetry we should take Γ as large as possible. For energy minimization problems this means we should take it to be the symmetry group of the metric space (V, d), which for the Thomson problem means we take $\Gamma = O(3)$.

Let $\mathcal{C}(I_{2t})^\Gamma$ be the subspace consisting of Γ-invariant functions, and $\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}^\Gamma$ the subspace of Γ-invariant symmetric kernels, and define the cones

$$\mathcal{C}(I_{2t})_{\geq 0}^\Gamma = \mathcal{C}(I_{2t})_{\geq 0} \cap \mathcal{C}(I_{2t})^\Gamma \quad \text{and} \quad \mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma = \mathcal{C}(I_t \times I_t)_{\succeq 0} \cap \mathcal{C}(I_t \times I_t)_{\mathrm{sym}}^\Gamma.$$

Given a function $f \in \mathcal{C}(I_{2t})$, we define its symmetrization $\bar{f} \in \mathcal{C}(I_{2t})^\Gamma$ by

$$\bar{f}(S) = \int_\Gamma f(\gamma S)\, d\gamma,$$

where we integrate over the normalized Haar measure of Γ. Similarly, given a kernel $K \in \mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$, we define its symmetrization $\bar{K}$ by

$$\bar{K}(J, J') = \int_\Gamma K(\gamma J, \gamma J')\, d\gamma.$$

Using these definitions we can define the symmetrizations $\bar{\lambda}$ and $\bar{\mu}$ of measures $\lambda \in \mathcal{M}(I_{2t})$ and $\mu \in \mathcal{M}(I_t \times I_t)_{\mathrm{sym}}$ by $\bar{\lambda}(f) = \lambda(\bar{f})$ and $\bar{\mu}(K) = \mu(\bar{K})$. It follows that the spaces $\mathcal{M}(I_{2t})^\Gamma$ and $\mathcal{M}(I_t \times I_t)_{\mathrm{sym}}^\Gamma$ can be identified with the topological duals of $\mathcal{C}(I_{2t})^\Gamma$ and $\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}^\Gamma$, and as in the nonsymmetrized situation these form dual pairs.

We have

$$A_t K(\gamma S) = \sum_{\substack{J, J' \in I_t \\ J \cup J' = \gamma S}} K(J, J') = \sum_{\substack{J, J' \in I_t \\ \gamma^{-1} J \cup \gamma^{-1} J' = S}} K(J, J') = \sum_{\substack{J, J' \in I_t \\ J \cup J' = S}} K(\gamma J, \gamma J'),$$

so A_t maps Γ-invariant kernels to Γ-invariant functions. This means we can view A_t as an operator from $\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}^\Gamma$ to $\mathcal{C}(I_{2t})^\Gamma$, and we can view A_t^* as an operator from $\mathcal{M}(I_{2t})^\Gamma$ to $\mathcal{M}(I_t \times I_t)_{\mathrm{sym}}^\Gamma$.

We define the symmetrization of a Γ-invariant primal problem H from Definition 5.1 by

$$H_\Gamma = \min\Big\{\lambda(f) : \lambda \in \mathcal{M}(I_{2t})_{\geq 0}^\Gamma,\, A_t^* \lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0}^\Gamma,\, \lambda(g_i) = b_i \text{ for } i \in [m]\Big\},$$

and the symmetrization of the dual program H^* by

$$H_\Gamma^* = \sup\Big\{\sum_{i=1}^m b_i a_i : a \in \mathbb{R}^m,\, K \in \mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma,\, f - \sum_{i=1}^m a_i g_i - A_t K \in \mathcal{C}(I_{2t})_{\geq 0}^\Gamma\Big\}.$$

Given feasible solutions λ and K of H and H^*, the symmetrizations $\bar{\lambda}$ and $\bar{K}$ are feasible for H_Γ and H_Γ^* and have the same objective values. This shows $H = H_\Gamma$ and $H^* = H_\Gamma^*$. As a result we have $H_\Gamma = H_\Gamma^*$, which alternatively could be shown by proving strong duality for the symmetrized problems.

In the following section we discuss the construction of a nested sequence $\{C_d\}_{d=0}^\infty$ of inner approximating cones of $\mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma$ such that the union $\cup_{d=0}^\infty C_d$ is uniformly dense in $\mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma$. This sequence is constructed in such a way that optimization over C_d is easies than optimization over the original cone, and gets more difficult as d grows. Define H_d^* to be the optimization problem H_Γ^* with the cone of invariant, positive definite kernels replaced by its inner approximation C_d. Now we give a sufficient condition for the programs H_d^* to approximate the program H_Γ^*. The following proposition applies for the case of $H = E_t$ by selecting selecting $c = 0$ and $y = -e$, where e is the all ones vector. This shows that if we let $E_{t,d}^*$ be the problem E_t^* with the cone $\mathcal{C}(I_t \times I_t)_{\succeq 0}$ replaced by C_d, then $E_{t,d}^* \to E_t$ as $d \to \infty$.

Proposition 5.3. *If there exists a scalar $c \in \mathbb{R}$ and a vector $y \in \mathbb{R}^m$ for which $cf - \sum_{i=1}^m y_i g_i$ is a strictly positive function, then $H_d^* \to H^*$ as $d \to \infty$.*

Proof. Select $c \in \mathbb{R}$ and $y \in \mathbb{R}^m$ for which $cf - \sum_{i=1}^m y_i g_i$ is a strictly positive function. Let (a, K) be a feasible solution of H_Γ^* and let $\varepsilon > 0$. Let

$$\kappa = \min_{S \in I_{2t}} \left(cf(S) - \sum_{i=1}^m y_i g_i(S) \right),$$

where the minimum is attained and strictly positive because we optimize a continuous function over a compact set. We have $f - \sum_{i=1}^m a_i g_i - A_t K \geq 0$, so, for any $\delta \geq 0$, we have

$$f + \delta c f - \sum_{i=1}^m (a_i + \delta y_i) g_i - A_t K \geq \delta \kappa,$$

and hence

$$f - \sum_{i=1}^m \frac{a_i + \delta y_i}{1 + \delta c} g_i - A_t\Big(\frac{1}{1+\delta c} K\Big) \geq \frac{\delta\kappa}{1+\delta c}.$$

Since $\bigcup_{d=0}^\infty C_d$ is uniformly dense in $\mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma$, and since A_t is a bounded operator, there exists a $d_\delta \in \mathbb{N}_0$ and a kernel $L_\delta \in C_{d_\delta}$ such that

$$\left\| A_t\Big(\frac{1}{1+\delta c} K\Big) - A_t L_\delta \right\|_\infty \leq \frac{\delta\kappa}{1+\delta c}.$$

This means that

$$f - \sum_{i=1}^m \frac{a_i + \delta y_i}{1+\delta c} g_i - A_t L_\delta \geq 0.$$

So, for all $\delta > 0$, $((a + \delta y)/(1 + \delta c), L_\delta)$ is feasible for $H_{d_\delta}^*$, and as $\delta \downarrow 0$, its objective value goes to the objective value of (a, K). This shows $H_d^* \to H^*$ as $d \to \infty$. □

6. Approximating the cone of invariant positive definite kernels

In this section we show how to approximate the cone of invariant positive definite kernels by a sequence of simpler inner approximating cones, where this sequence converges in the sense that the union of the inner approximating cones is uniformly dense, and where each of the inner approximating cones is isomorphic to a finite product of positive semidefinite matrix cones (their elements are said to be block diagonalized). In Section 6.1 we first give some background on symmetry adapted systems and zonal matrices and how these can be used to construct inner approximating cones. Then we use the results from Appendix A to show the existence of an inner approximating sequence as mentioned above for the cone $\mathcal{C}(X \times X)^{\Gamma}_{\succeq 0}$, where X is a compact metric space and Γ a compact group with a continuous action on X. In Section 6.2 we show how such a sequence can be constructed explicitly for the case where $X = I_t$. In Section 6.3 we then perform the construction explicitly for $V = S^2$, $\Gamma = O(3)$, and $t = 2$, which we later use to compute E_2^*.

6.1. **Symmetry adapted systems and zonal matrices.** In this section we show how to define the inner approximating cones by using symmetry adapted systems and zonal matrices. Since we will use representation theory, it is convenient to work over the complex numbers and first consider the cone $\mathcal{C}(X \times X; \mathbb{C})^{\Gamma}_{\succeq 0}$ of *Hermitian*, Γ-invariant, positive definite kernels. Here a Hermitian kernel $K \in \mathcal{C}(X \times X; \mathbb{C})$ is said to be *positive definite* if

$$\sum_{i,j=1}^{n} c_i \overline{c_j} K(x_i, x_j) \geq 0 \quad \text{for all} \quad n \in \mathbb{N}, x \in X^n, c \in \mathbb{C}^n.$$

We start by giving the definition of a symmetry adapted system for X. Let μ be a Radon measure on X that is strictly positive and Γ-invariant. By *strictly positive* we mean that $\mu(U) > 0$ for all open sets U in X, and by Γ-invariant we mean $\mu(\gamma U) = \mu(U)$ for all $\gamma \in \Gamma$ and all Borel sets U in X. Such a measure always exists (see Lemma A.2). We define an *orthonormal system* of X to be a set consisting of continuous, complex-valued functions on X that are orthonormal with respect to the $L^2(X, \mu; \mathbb{C})$ inner product

$$\langle f, g \rangle = \int f(x) \overline{g(x)}\, d\mu(x).$$

Such a system is said to be *complete* if its span is uniformly dense in the space $\mathcal{C}(X; \mathbb{C})$ of continuous complex-valued functions on X.

To define what it means for such a system to be symmetry adapted we need some representation theory. A *unitary representation* of Γ is a continuous group homomorphism from Γ to the group $U(\mathcal{H})$ of unitary operators on a nontrivial Hilbert space $\mathcal{H}$, where $U(\mathcal{H})$ is equipped with the weak (or strong, they are the same here) operator topology. Such a representation is said to be *irreducible* when $\mathcal{H}$ does not admit a nontrivial closed invariant subspace. Two unitary representations $\pi_1 \colon \Gamma \to U(\mathcal{H}_1)$ and $\pi_2 \colon \Gamma \to U(\mathcal{H}_2)$ are *equivalent* if there exists a unitary operator $T \colon \mathcal{H}_1 \to \mathcal{H}_2$ that is Γ-equivariant; that is, $T\pi_1(\gamma)u = \pi_2(\gamma)Tu$ for all $\gamma \in \Gamma$ and $u \in \mathcal{H}_1$. Let $\hat{\Gamma}$ be a complete set of inequivalent irreducible unitary representations of Γ, and denote the dimension of a representation $\pi \in \hat{\Gamma}$ by d_π. A particularly important example of a unitary representation is given by

$$L \colon \Gamma \to U(L^2(X, \mu; \mathbb{C})),\ L(\gamma)f(x) = f(\gamma^{-1}x).$$

A complete orthonormal system of X is said to be a *symmetry adapted system* of X if there exist numbers $0 \leq m_\pi \leq \infty$ for which we can write the system as

$$\Big\{ e_{\pi,i,j} : \pi \in \hat{\Gamma},\ i \in [m_\pi],\ j \in [d_\pi] \Big\},$$

with $\mathcal{H}_{\pi,i} = \operatorname{span}\{e_{\pi,i,1}, \ldots, e_{\pi,i,d_\pi}\}$ equivalent to π as a unitary subrepresentation of L, and where there exist Γ-equivariant unitary operators $T_{\pi,i,i'} \colon \mathcal{H}_{\pi,i} \to \mathcal{H}_{\pi,i'}$ with $e_{\pi,i',j} = T_{\pi,i,i'} e_{\pi,i,j}$ for all π, i, i', and j. In Theorem A.6 we show such a system always exists. The number m_π can be shown to be equal to the dimension of the space $\operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)$ of Γ-equivariant, continuous functions from X to $\mathcal{H}_\pi$, where $\mathcal{H}_\pi$ is the Hilbert space of the irreducible representation π, and hence does not depend on the choice of symmetry adapted system.

The spaces $H_{\pi,i}$ are pairwise orthogonal irreducible subrepresentations of L, each spanned by continuous functions, such that $\sum_{\pi\in\hat{\Gamma}}\sum_{i\in[m]} H_{\pi,i}$ is uniformly dense in $\mathcal{C}(X;\mathbb{C})$. On the other hand, when we are given a set of subspaces satisfying the above properties, then we can immediately construct a symmetry adapted system by simply selecting appropriate bases of the subspaces.

Now we show how the extreme rays of $\mathcal{C}(X\times X;\mathbb{C})_{\succeq 0}^{\Gamma}$ suggest a "simultaneous block diagonalization" of the kernels in this cone. A nonzero vector x in a cone K lies on an *extreme ray* if $x_1, x_2 \in \mathbb{R}_{\geq 0}x$ for all $x_1, x_2 \in K$ with $x = x_1 + x_2$. In Theorem A.1 we show that a kernel $K \in \mathcal{C}(X\times X;\mathbb{C})_{\succeq 0}^{\Gamma}$ lies on an extreme ray if and only if there exists an irreducible unitary representation $\pi\colon \Gamma \to U(\mathcal{H}_\pi)$ and a function $\varphi \in \mathrm{Hom}_\Gamma(X, \mathcal{H}_\pi)$ such that

$$K(x,y) = \langle \varphi(x), \varphi(y)\rangle \quad \text{for all} \quad x,y\in X.$$

In the case where $m_\pi < \infty$, we have a finite basis $\varphi_1,\ldots,\varphi_{m_\pi}$ of $\mathrm{Hom}_\Gamma(X,\mathcal{H}_\pi)$, and the map

$$A \mapsto \sum_{i,j=1}^{m_\pi} A_{i,j}\langle \varphi_i(\cdot), \varphi_j(\cdot)\rangle$$

is an isomorphism from the cone of $m_\pi \times m_\pi$ Hermitian positive semidefinite matrices to the convex hull of the extreme rays of $\mathcal{C}(X\times X;\mathbb{C})_{\succeq 0}^{\Gamma}$ that correspond to π. Moreover, if $\hat{\Gamma}$ is finite and m_π is finite for all $\pi\in\hat{\Gamma}$, this gives an isomorphism between a finite product of Hermitian positive semidefinite matrix cones and $\mathcal{C}(X\times X;\mathbb{C})_{\succeq 0}^{\Gamma}$. For our applications, however, we are particularly interested in the situation where these numbers are not finite, and hence we need to consider convergence.

Given a symmetry adapted system $\{e_{\pi,i,j}\}$ of X, we define the matrices

$$E_\pi(x)_{i,j} = e_{\pi,i,j}(x) \quad \text{for} \quad \pi\in\hat{\Gamma}, x\in X, i\in[m_\pi], j\in[d_\pi],$$

and we use this to define the *zonal matrices*

$$Z_\pi(x,y) = E_\pi(x)E_\pi(y)^* \quad \text{for} \quad \pi\in\hat{\Gamma} \quad \text{and} \quad x,y\in X,$$

where $E_\pi(y)^*$ is the conjugate transpose of $E_\pi(y)$. The *Fourier coefficients* of a kernel $K\in\mathcal{C}(X\times X;\mathbb{C})$ are defined as

$$\hat{K}(\pi) = \frac{1}{d_\pi}\iint K(x,y)Z_\pi(x,y)^*\, d\mu(x)d\mu(y), \quad \text{for} \quad \pi\in\hat{\Gamma},$$

where the matrices are integrated entrywise. The *inverse Fourier transform* reads

$$K(x,y) = \sum_{\pi\in\hat{\Gamma}}\sum_{i,i'=1}^{m_\pi} \hat{K}(\pi)_{i,i'} Z_\pi(x,y)_{i,i'},$$

where the series in general converges in L^2. The kernel K is positive definite if and only if $\hat{K}(\pi)$ is positive semidefinite for all $\pi\in\hat{\Gamma}$ (see Lemma A.7). In the special case where the action of Γ on X has finitely many orbits and K is positive definite, the above series converges absolutely-uniformly (this is an extension of Bochner's theorem [8] to the case of finitely many orbits [28]). We are interested in the situation of infinitely many orbits, where the above series in general does not converge uniformly.

For each $\pi\in\hat{\Gamma}$, let $R_{\pi,0}\subseteq R_{\pi,1}\subseteq\ldots$ be finite subsets of $[m_\pi]$ such that $\bigcup_{d=0}^\infty R_{\pi,d} = [m_\pi]$ and such that for each d, the set $R_{\pi,d}$ is empty for all but finitely many π. Let $Z_{\pi,d}$ be the finite principal submatrix of the zonal matrix Z_π containing only the rows and columns indexed by elements from $R_{\pi,d}$. Let $C_{\pi,d}$ be the cone of kernels of the form $(x,y) \mapsto \langle A, Z_{\pi,d}(x,y)^* \rangle$, where A ranges over the Hermitian positive semidefinite matrices of appropriate size, and where $\langle A, B\rangle = \mathrm{trace}(AB^*)$ is the trace inner product. We define the dth inner approximating cone C_d by the (Minkowski) sum $\sum_{\pi \in \hat{\Gamma}} C_{\pi,d}$. Then we have

$$C_0 \subseteq C_1 \subseteq \ldots \subseteq \mathcal{C}(X \times X; \mathbb{C})_{\succeq 0}^{\Gamma},$$

and in Theorem A.8 we show $\bigcup_{d=0}^{\infty} C_d$ is uniformly dense in $\mathcal{C}(X \times X; \mathbb{C})_{\succeq 0}^{\Gamma}$.

Let

$$D_d = \big\{ (K + \overline{K})/2 : K \in C_d \big\} \subseteq C_d,$$

so that $\bigcup_{d=0}^{\infty} D_d$ is uniformly dense in $\mathcal{C}(X \times X)_{\succeq 0}^{\Gamma}$. If all irreducible representations of Γ are of real type; that is, if each representation $\pi \in \hat{\Gamma}$ is unitarily equivalent to a representation $\Gamma \to O(d_\pi) \subseteq U(d_\pi)$, then there exists a symmetry adapted system of X consisting of real-valued functions. This means we can choose the symmetry adapted system in such a way that the zonal matrices are real-valued, and

$$D_d = \Big\{ \sum_{\pi \in \hat{\Gamma}} \Big\langle F_\pi, Z_{\pi,d}(\cdot,\cdot)^* \Big\rangle : F_\pi \in S_{\succeq 0}^{R_{\pi,d}} \text{ for } \pi \in \hat{\Gamma} \Big\}.$$

If $\hat{\Gamma}$ also contains representations of complex or quaternionic type (these are the two remaining possibilities), then one should construct a *real* symmetry adapted system, where π ranges over the real irreducible representations. See [23] or [46] where this is discussed for finite groups. The irreducible representations of the groups considered in this paper are all of real type.

6.2. **Harmonic analysis on subset spaces.** In this section we show how to construct the sequence $\{D_d\}$ as defined above for the special case where $X = I_t$. We give a construction in two steps: The main step is that we show how to construct a symmetry adapted system for X_t, where X_t is a structurally simpler space that contains I_t as an embedding. This yields a sequence of inner approximating cones of $\mathcal{C}(X_t \times X_t)_{\succeq 0}^{\Gamma}$. Then we restrict the domains of the kernels in these inner approximations to the smaller space $I_t \times I_t$ to obtain inner approximations of $\mathcal{C}(I_t \times I_t)_{\succeq 0}^{\Gamma}$.

Let

$$X_t = \bigcup_{i=0}^{t} V^i / S_i,$$

where S_i is the symmetric group on i elements. The set X_t obtains a topology by using the topology of V and the product, quotient, and disjoint union topologies. We define a continuous action of Γ on X_t by

$$\gamma \big\{ (x_{\sigma(1)}, \ldots, x_{\sigma(i)}) : \sigma \in S_i \big\} = \big\{ (\gamma x_{\sigma(1)}, \ldots, \gamma x_{\sigma(i)}) : \sigma \in S_i \big\}.$$

The space I_t embeds as a closed, Γ-invariant subspace into X_t by the embedding that sends $\{x_1, \ldots, x_i\} \in I_{=i}$ to $\{(x_{\sigma(1)}, \ldots, x_{\sigma(i)}) : \sigma \in S_i\}$. Notice that we could also embed I_t in $V^t/S_t \cup \{e\}$ (where the empty set maps to an additional point e), but for computational reasons it is preferable to embed I_t into X_t.

We first show that each kernel in the cone $\mathcal{C}(I_t \times I_t)_{\succeq 0}^{\Gamma}$ is the restriction to $I_t \times I_t$ of a kernel from $\mathcal{C}(X_t \times X_t)_{\succeq 0}^{\Gamma}$, which shows shows that if a sequence of inner approximations of $\mathcal{C}(X_t \times X_t)_{\succeq 0}^{\Gamma}$ has dense union, then the corresponding sequence of inner approximations of $\mathcal{C}(I_t \times I_t)_{\succeq 0}^{\Gamma}$ also has dense union.

Lemma 6.1. *Let (X,d) be a compact metric space with a continuous action of a compact group Γ, and let Y be a closed Γ-invariant subspace of X. Each kernel $K \in \mathcal{C}(Y \times Y)_{\succeq 0}^{\Gamma}$ is the restriction to $Y \times Y$ of a kernel in $\mathcal{C}(X \times X)_{\succeq 0}^{\Gamma}$.*

Proof. Let $K \in \mathcal{C}(Y \times Y)_{\succeq 0}^{\Gamma}$. By Mercer's theorem there exists a sequence of functions $\{e_i\}$ in $\mathcal{C}(Y)$ such that $\sum_{i=1}^{\infty} e_i \otimes \overline{e_i}$ converges absolutely and uniformly to K. In particular this means $\sum_{i=1}^{\infty} |e_i|^2$ converges uniformly.

Given a point $x \in X$, define

$$Y_x = \big\{ y \in Y : d(y,x) \leq d(z,x) \text{ for all } z \in X \big\}.$$

For each i we define the function $c_i \colon X \to \mathbb{R}$ by

$$c_i(x) = \min_{y \in Y_x} |e_i(y)|^2 \quad \text{for} \quad x \in X.$$

We first show c_i is lower semicontinuous. For $\kappa > 0$, let

$$Y_{x,\kappa} = Y \setminus \bigcup_{y \in Y_x} B_\kappa(y)^\circ,$$

where $B_\kappa(y)^\circ$ is the open ball of radius κ about y. The set $Y_{x,\kappa}$ is closed, so there exist a $\delta > 0$ such that $Y_z \cap Y_{x,\kappa} = \emptyset$ for all $z \in B_\delta(x)$. This means that for every $\varepsilon > 0$, there is a $\delta > 0$ such that $c_i(z) \geq c_i(x) - \varepsilon$ for all $z \in B_\delta(x)$.

Since X is a compact metric space, it is perfectly normal, which implies the existence of a function $\iota \in \mathcal{C}(X)$ such that $\iota|_Y = 1$ and $\iota|_{X \setminus Y} < 1$. By Tietze' extension theorem there exist functions $f_i \in \mathcal{C}(X; K)$ with $f_i|_Y = e_i$. Let

$$Q_i = \{ x \in X : |f_i(x)|^2 \geq c_i(x) + 2^{-i} \}.$$

The set Q_i is disjoint from Y, and it follows from c_i being lower semicontinuous that Q_i is closed and hence compact. So, $M_i = \max_{x \in Q_i} \iota(x)$ exists and is strictly smaller than 1. Let

$$B_i = \max_{x \in Q_i} \frac{|f_i(x)|}{\sqrt{c_i(x) + 1/2^i}},$$

and let k_i be an integer such that $M_i^{k_i} B_i \leq 1$. It follows that

$$|g_i|^2 \leq c_i + \frac{1}{2^i}, \quad \text{where} \quad g_i = \iota^{k_i} f_i.$$

Let $\varepsilon > 0$. Choose $N_1 \in \mathbb{N}$ such that $\sum_{i=N_1}^\infty 1/2^i \leq \varepsilon/2$. Choose $N_2 \in \mathbb{N}$ such that

$$\Big\| \sum_{i=m}^n |e_i|^2 \Big\|_\infty \leq \frac{\varepsilon}{2}$$

for all $n \geq m \geq N_2$. This is possible because $\sum_{i=1}^\infty |e_i|^2$ converges uniformly and hence Cauchy uniformly. Let $N = \max\{N_1, N_2\}$. Then,

$$\Big\| \sum_{i=m}^n g_i \otimes \overline{g_i} \Big\|_\infty \leq \Big\| \sum_{i=m}^n |g_i|^2 \Big\|_\infty \leq \Big\| \sum_{i=m}^n c_i \Big\|_\infty + \sum_{i=m}^n 1/2^i.$$

We have

$$\Big\| \sum_{i=m}^n c_i \Big\|_\infty = \sup_{x \in X} \sum_{i=m}^n \min_{y \in Y_x} |e_i(y)|^2.$$

We use the axiom of choice to select an element $y_x \in Y_x$ for each $x \in X$. Then,

$$\sup_{x \in X} \sum_{i=m}^n \min_{y \in Y_x} |e_i(y)|^2 \leq \sup_{x \in X} \sum_{i=m}^n |e_i(y_x)|^2 = \sup_{x \in Y} \sum_{i=m}^n |e_i(x)|^2 = \Big\| \sum_{i=m}^n |e_i|^2 \Big\|_\infty \leq \frac{\varepsilon}{2}.$$

and $\sum_{i=m}^n 1/2^i \leq \varepsilon/2$, so $\sum_{i=m}^n g_i \otimes \overline{g_i}$ converges uniformly Cauchy and hence uniformly. Let P be the limit function.

Define $K' \in \mathcal{C}(X \times X)_{\succeq 0}^\Gamma$ by $K'(x,y) = \int P(\gamma x, \gamma y)\, d\gamma$, where we integrate over the normalized Haar measure of Γ. Since $P|_{Y \times Y} = K$ is Γ-invariant, the restriction of K' to $Y \times Y$ equals K, which completes the proof. □

To give an explicit construction of a symmetry adapted system for X_t we use symmetric tensor powers. Given a vector space $\mathcal{V}$, denote the nth tensor power of $\mathcal{V}$ by $\mathcal{V}^{\otimes n}$; that is, $\mathcal{V}^{\otimes n} = \mathcal{V} \otimes \cdots \otimes \mathcal{V}$ (n times). Given $v_1, \ldots, v_n \in \mathcal{V}$ and $\sigma \in S_n$, let

$$(\otimes_{i=1}^n v_i)^\sigma = \otimes_{i=1}^n v_{\sigma(i)},$$

and extend this operation to $\mathcal{V}^{\otimes n}$ by linearity. Define the nth *symmetric tensor power* of $\mathcal{V}$ as

$$\mathcal{V}^{\odot n} = \Big\{ \sum_{\sigma \in S_n} w^\sigma : w \in \mathcal{V}^{\otimes n} \Big\}.$$

We have $w^\sigma = w$ for all $w \in \mathcal{V}^{\odot n}$ and $\sigma \in S_n$, and $\mathcal{V}^{\odot n} = \operatorname{span}\{v^{\otimes n} : v \in \mathcal{V}\}$ [14]. If $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces, then we equip the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ with the inner product $\langle u_1 \otimes u_2, v_1 \otimes v_2 \rangle = \langle u_1, v_1 \rangle \langle u_2, v_2 \rangle$, where we extend linearly in the first and antilinearly in the second component. We denote the Hilbert space obtained by taking the completion in the metric given by this inner product by $\mathcal{H}_1 \hat{\otimes} \mathcal{H}_2$. A symmetric tensor power $\mathcal{H}^{\odot n}$ of a Hilbert space $\mathcal{H}$ gets a metric as a subspace of $\mathcal{H}^{\hat{\otimes} n}$, and we denote the completion in this metric by $\mathcal{H}^{\hat{\odot} n}$. We have

$$\mathcal{H}^{\hat{\odot} n} = \operatorname{closure}(\operatorname{span}(\{v^{\otimes n} : v \in \mathcal{H}\})),$$

where the closure is in $\mathcal{H}^{\hat{\otimes} n}$. The *(inner) tensor product representation*

$$\pi_1 \otimes \pi_2 \colon \Gamma \to U(\mathcal{H}_1 \hat{\otimes} \mathcal{H}_2)$$

of two unitary representations $\pi_1 \colon \Gamma \to U(\mathcal{H}_1)$ and $\pi_2 \colon \Gamma \to U(\mathcal{H}_2)$ is defined by

$$(\pi_1 \otimes \pi_2)(\gamma)(v_1 \otimes v_2) = (\pi(\gamma) v_1) \otimes (\pi(\gamma) v_2),$$

and we have similar definitions for finite products and finite (symmetric) powers.

Let μ be a strictly positive Γ-invariant Radon measure on X. This defines a strictly positive, Γ-invariant Radon measure ν on X_t by

$$\nu(f) = \sum_{i=0}^t \int \cdots \int_V f(\{(x_{\sigma(1)}, \ldots, x_{\sigma(i)}) : \sigma \in S_i\}) \, d\mu(x_1) \cdots d\mu(x_i).$$

For each $0 \leq i \leq t$, we define the operator

$$L_i \colon L^2(V, \mu; \mathbb{C})^{\hat{\odot} i} \to L^2(X_t, \nu; \mathbb{C})$$

by setting

$$L_i(f^{\otimes i})(\{(x_{\sigma(1)}, \ldots, x_{\sigma(i)}) : \sigma \in S_i\}) = \prod_{k=1}^i f(x_k)$$

for $f \in L^2(V, \mu; \mathbb{C})$ and extending by linearity and continuity. These are isometric, Γ-equivariant operators with pairwise orthogonal images, such that

$$\bigoplus_{i=0}^t L_i(L^2(V, \mu; \mathbb{C})^{\hat{\odot} i}) = L^2(X_t, \nu; \mathbb{C}),$$

and

$$\sum_{i=0}^t L_i(\mathcal{C}(V; \mathbb{C})^{\odot i})$$

is uniformly dense in $\mathcal{C}(X_t; \mathbb{C})$.

We assume we have a symmetry adapted system of V. As discussed in the previous section, such a system defines a sequence $\{\mathcal{H}_k\}_{k=1}^m$ (where $1 \leq m \leq \infty$) of pairwise orthogonal, Γ-irreducible (and hence finite dimensional) subrepresentations of $L^2(V, \mu; \mathbb{C})$, such that the algebraic sum $\sum_{k=1}^m \mathcal{H}_k$ is uniformly dense in $\mathcal{C}(V; \mathbb{C})$.

The spaces

$$\bigotimes_{k=1}^{m} \mathcal{H}_k^{\odot \tau_k}, \quad \text{for} \quad \tau \in D_i = \Big\{ \tau \in \mathbb{N}_0^m : \sum_{k=1}^m \tau_k = i \Big\} \quad \text{and} \quad 0 \leq i \leq t,$$

are pairwise orthogonal, Γ-invariant subspaces of $L^2(V, \mu; \mathbb{C})^{\otimes i}$. Notice that the above tensor products are finite even if $m = \infty$, since τ_k is nonzero for at most finitely many k. For each $0 \leq i \leq t$, we will define (see below) a Γ-equivariant, unitary operator

$$T_i \colon \widehat{\bigoplus_{\tau \in D_i}} \bigotimes_{k=1}^{m} \mathcal{H}_k^{\odot \tau_k} \to L^2(V, \mu; \mathbb{C})^{\hat{\odot} i},$$

so that

$$\sum_{\tau \in D_i} T_i\Big(\bigotimes_{k=1}^{m} \mathcal{H}_k^{\odot \tau_k} \Big)$$

is uniformly dense in $\mathcal{C}(V; \mathbb{C})^{\odot i}$.

The finite dimensional spaces $\otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}$, for $\tau \in D_i$ and $0 \leq i \leq t$, decompose into Γ-irreducible representations; that is, there exist Γ-equivariant, unitary operators

$$M_\tau \colon \bigoplus_{\pi \in R_\tau} \mathcal{H}_\pi \to \bigotimes_{k=1}^{m} \mathcal{H}_k^{\odot \tau_k},$$

where $\mathcal{H}_\pi$ is the Hilbert space of the irreducible representation $\pi \in R_\tau$, and where R_τ is some finite subset of $\hat{\Gamma}$. In Section 6.3 we construct the sets R_τ and the operators M_τ explicitly for the case where $V = S^2$, $\Gamma = O(3)$, and $t = 2$.

By composing the operators defined above we can define a symmetry adapted system of X_t. For each $\pi \in \hat{\Gamma}$, let $\{e_{\pi,1}, \ldots, e_{\pi,d_\pi}\}$ be an orthonormal basis of $\mathcal{H}_\pi$. Then,

$$\Big\{ L_i(T_i(M_\tau(e_{\pi,j}))) : 0 \leq i \leq t,\, \tau \in D_i,\, \pi \in R_\tau,\, j \in [d_\pi] \Big\} \tag{1}$$

is a symmetry adapted system of X_t.

In the remainder of this section we give the precise definition of T_i and show it is a well-defined, Γ-equivariant, unitary operator. This generalizes a result from [2] from finite to infinite direct sums and from vector spaces to representations. Given $\tau \in D_i$, let A_τ be the subgroup consisting of all $\sigma \in S_i$ for which the set

$$\Big\{ \sum_{k=1}^{j-1} \tau_k + 1, \sum_{k=1}^{j-1} \tau_k + 2, \ldots, \sum_{k=1}^{j} \tau_k \Big\}$$

is invariant under the permutation σ on $[i]$ for each $j \in [m]$. Let B_τ be the left coset space of S_i modulo A_τ. Given $\tau \in D_i$, $w \in \otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}$, and $[\sigma] \in B_\tau$, the operation $w \mapsto w^\sigma$ is well-defined, and

$$\frac{1}{|B_\tau|} \sum_{[\sigma] \in B_\tau} w^\sigma \in L^2(V, \mu; \mathbb{C})^{\odot i}.$$

Moreover, if we fix an element $w_\tau \in \otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}$ for every $\tau \in D_i$, then w_τ^σ and $w_{\tau'}^{\sigma'}$ are orthogonal whenever $\tau \neq \tau'$ or $\sigma \neq \sigma'$. Hence, we can define T_i by setting

$$T_i(w) = \frac{1}{|B_\tau|} \sum_{[\sigma] \in B_\tau} w^\sigma, \quad \text{for} \quad w \in \bigotimes_{k=1}^{m} \mathcal{H}_k^{\odot \tau_k} \quad \text{and} \quad \tau \in D_i,$$

and extending by linearity and continuity.

Lemma 6.2. *For each $0 \le i \le t$, T_i is a Γ-equivariant, unitary operator, and*

$$\sum_{\tau \in D_i} T_i\Big(\bigotimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}\Big)$$

is uniformly dense in $\mathcal{C}(V;\mathbb{C})^{\odot i}$.

Proof. Given $w \in \otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}$, we have

$$\|T_i(w)\| = \left\| \frac{1}{|B_d|} \sum_{[\sigma] \in B_\tau} w^\sigma \right\| = \frac{1}{|B_\tau|} \sum_{[\sigma] \in B_\tau} \|w^\sigma\| = \|w\|,$$

so T_i is a linear isometry.

The span of the elements of the form $(\sum_{k=1}^m v_k)^{\otimes i}$, where $v_k \in \mathcal{H}_k$ for $k \in [m]$ and $v_k = 0$ for all but finitely many k, is uniformly dense in $\mathcal{C}(V;\mathbb{C})^{\odot i}$. Such an element has a preimage under T_i:

$$\begin{aligned}\Big(\sum_{k=1}^m v_k\Big)^{\otimes i} &= \sum_{k_1,\dots,k_i=1}^m \bigotimes_{j=1}^i v_{k_j} = \sum_{\{k_1,\dots,k_i\} \subseteq [m]} \sum_{\sigma \in S_i} \bigotimes_{j=1}^i v_{k_{\sigma(j)}} \\ &= \sum_{\tau \in D_i} \sum_{[\sigma] \in B_\tau} \Big(\bigotimes_{k=1}^m v_k^{\otimes \tau_k}\Big)^\sigma = T_i\Big(\sum_{\tau \in D_i} \bigotimes_{k=1}^m v_k^{\otimes \tau_k}\Big).\end{aligned}$$

So,

$$\sum_{\tau \in D_i} T_i\Big(\bigotimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}\Big)$$

is uniformly dense in $\mathcal{C}(V;\mathbb{C})^{\odot i}$. This means that T_i is an isometry whose image is dense in $L^2(V,\mu;\mathbb{C})^{\odot i}$, and T_i therefore is a unitary operator.

Since the spaces $\mathcal{H}_k$ are Γ-invariant, the spaces $\otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}$, for $\tau \in D_i$, are also Γ-invariant. So, for $w \in \otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}$, with $\tau \in D_i$, we have

$$\begin{aligned}T_i(\pi^{\otimes i}(\gamma) w) &= \frac{1}{|B_\tau|} \sum_{[\sigma] \in B_\tau} (\pi^{\otimes i}(\gamma) w)^\sigma = \frac{1}{|B_\tau|} \sum_{[\sigma] \in B_\tau} \pi^{\otimes i}(\gamma) w^\sigma \\ &= \pi^{\otimes i}(\gamma) \left(\frac{1}{|B_\tau|} \sum_{[\sigma] \in B_\tau} w^\sigma \right) = \pi^{\otimes i}(\gamma) T_i(w),\end{aligned}$$

which means that T_i is Γ-equivariant. □

6.3. **Explicit computations for the sphere.** In this section we explicitly construct the sequence $\{D_d\}$ of inner approximations (see Section 6.1) of $\mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma$, where $V = S^2$, $\Gamma = O(3)$, and $t = 2$. As explained in Section 6.2, for this we need to construct a symmetry adapted system for

$$X_2 = \bigcup_{i=0}^2 V^i / S_i.$$

Here we give all formulas explicitly so that a software implementation can be written to generate the zonal matrices.

Let $\mathcal{H}_\ell \subseteq \mathcal{C}(S^2;\mathbb{C})$ be the space of spherical harmonics of degree ℓ. A *spherical harmonic* is the restriction to S^2 of a homogeneous polynomial in $\mathbb{C}[x,y,z]$ that vanishes under the Laplacian $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$; see for instance [24]. The space $\mathcal{H}_\ell$ has dimension $2\ell + 1$. Moreover, these spaces are orthogonal and form irreducible subrepresentations of the unitary representation

$$L_s \colon SO(3) \to U(L^2(S^2,\mu;\mathbb{C})),\ L_s(\gamma) f(x) = f(\gamma^{-1} x),$$

where μ is the invariant measure on the sphere with normalization $\mu(S^2) = 1$. The subrepresentations $\mathcal{H}_\ell$ in fact form a complete set of subrepresentations: The algebraic sum of the spaces $\mathcal{H}_\ell$ is uniformly dense in $\mathcal{C}(S^2;\mathbb{C})$, and $L^2(S^2,\mu;\mathbb{C})$ decomposes as the Hilbert space direct sum of the spaces $\mathcal{H}_\ell$. Moreover, every irreducible unitary representation $SO(3)$ is equivalent to $\mathcal{H}_\ell$ for some ℓ.

The *Laplace spherical harmonics* Y_ℓ^m, for $-\ell \leq m \leq \ell$, provide an explicit set of orthonormal bases of the spaces $\mathcal{H}_\ell$. The functions Y_ℓ^m are typically defined as

$$Y_\ell^m(\vartheta,\varphi) = c_\ell^m P_\ell^m(\cos(\varphi))e^{im\vartheta},$$

where we use the spherical coordinates

$$x = \cos(\vartheta)\sin(\varphi), \quad y = \sin(\vartheta)\sin(\varphi), \quad z = \cos(\varphi).$$

Here

$$c_\ell^m = (-1)^m\sqrt{(2\ell+1)\frac{(\ell-m)!}{(\ell+m)!}}$$

is a normalization constant, and P_ℓ^m is the ℓth *associated Legendre polynomial* of order m, where both use the Condon–Shortley phase convention. We can define P_ℓ^m as

$$P_\ell^m(z) = (-1)^m(1-z^2)^{m/2}\frac{d^m}{dz^m}\left(P_\ell(z)\right)$$

where

$$P_\ell(x) = \frac{1}{2^\ell \ell!}\frac{d^\ell}{dx^\ell}(x^2-1)^\ell.$$

is the ℓth Legendre polynomial.

In cartesian coordinates Y_ℓ^m becomes

$$c_l^m P_l^m\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right)\left(\frac{x+iy}{\sqrt{x^2+y^2}}\right)^m,$$

and by using the identity $x^2+y^2+z^2=1$ as well as the above definition of the associated Legendre polynomials, we can write Y_ℓ^m as the polynomial

$$Y_\ell^m(x,y,z) = (-1)^m c_\ell^m \frac{d^m}{dx^m}\left(P_\ell(z)\right)(x+iy)^m.$$

From the definition of P_ℓ we see that when ℓ is even (odd), then every term of P_ℓ has even (odd) degree. This means we can multiply the terms in $Y_\ell^m(x,y,z)$ with appropriate powers of $x^2+y^2+z^2$ to make $Y_\ell^m(x,y,z)$ into a homogeneous polynomial of degree ℓ.

In general, an inner tensor product (see previous section) of irreducible representations is not irreducible. By the above discussion we know that a tensor product $\mathcal{H}_{\ell_1}\otimes\mathcal{H}_{\ell_2}$ must be isomorphic to a direct sum of the spaces $\mathcal{H}_\ell$. Indeed, we have

$$\mathcal{H}_{\ell_1}\otimes\mathcal{H}_{\ell_2} \simeq \mathcal{H}_{|\ell_1-\ell_2|}\oplus\cdots\oplus\mathcal{H}_{\ell_1+\ell_2},$$

where the $SO(3)$-equivariant, unitary operator

$$\Phi_{\ell_1,\ell_2}\colon \mathcal{H}_{\ell_1}\otimes\mathcal{H}_{\ell_2} \to \mathcal{H}_{|\ell_1-\ell_2|}\oplus\cdots\oplus\mathcal{H}_{\ell_1+\ell_2}$$

is rather nontrivial, but can be given explicitly by using the Clebsch–Gordan coefficients [24]. Set

$$\Phi_{\ell_1,\ell_2}(Y_{\ell_1}^{m_1}\otimes Y_{\ell_2}^{m_2}) = \sum_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2}\sum_{m=-\ell}^{\ell} C_{\ell_1,m_1,\ell_2,m_2}^{\ell,m} Y_\ell^m$$

and extend by linearity, where the Clebsch–Gordan coefficients are given by

$$C^{\ell,m}_{\ell_1,m_1,\ell_2,m_2} = \delta_{m_1+m_2=m}\Big(\frac{(2\ell+1)(\ell_1+\ell_2-\ell)!(\ell_1-\ell_2+\ell)!(-\ell_1+\ell_2-\ell)!}{(\ell_1+\ell_2+\ell+1)!}$$
$$\cdot\,(\ell_1+m_1)!(\ell_1-m_1)!(\ell_2+m_2)!(\ell_2-m_2)!(\ell+m)!(\ell-m)!\Big)^{1/2}$$
$$\cdot \sum_{\nu=-\infty}^{\infty}(-1)^\nu\Big(\nu!(\ell_1+\ell_2-\ell-\nu)!(\ell_1-m_1-\nu)!(\ell_2+m_2-\nu)!$$
$$\cdot\,(\ell-\ell_2+m_1+\nu)!(\ell-\ell_1-m_2+\nu)!\Big)^{-1}.$$

Since the Clebsch–Gordan coefficients are real numbers, it follows that

$$\Phi^{-1}_{\ell_1,\ell_2}(Y^m_\ell) = \sum_{m_1=-\ell_1}^{\ell_1}\sum_{m_2=-\ell_2}^{\ell_2} C^{\ell,m}_{\ell_1,m_1,\ell_2,m_2} Y^{m_1}_{\ell_1}\otimes Y^{m_2}_{\ell_2}.$$

For any set of scalars $\{c_m\}$ we have

$$\Phi_{\ell',\ell'}\Big(\sum_{m_1=-\ell'}^{\ell'} c_{m_1}Y^{m_1}_{\ell'}\otimes\sum_{m_1=-\ell'}^{\ell'} c_{m_1}Y^{m_1}_{\ell_1}\Big)$$
$$= \sum_{\ell=0}^{2\ell'}\sum_{m=-\ell}^{\ell}\left(\sum_{m_1,m_2=-\ell'}^{\ell'} c_{m_1}c_{m_2}C^{\ell,m}_{\ell',m_1,\ell',m_2}\right)Y^m_\ell,$$

and by using the symmetry relation

$$C^{\ell,m}_{\ell_2,m_2,\ell_1,m_1} = (-1)^{\ell_1+\ell_2-\ell}C^{\ell,m}_{\ell_1,m_1,\ell_2,m_2}$$

we obtain

$$\sum_{m_1,m_2=-\ell'}^{\ell'} c_{m_1}c_{m_2}C^{\ell,m}_{\ell',m_1,\ell',m_2} = 0$$

for all odd numbers ℓ. This shows that

$$\Phi_{\ell',\ell'}(\mathcal{H}^{\odot 2}_{\ell'}) \subseteq \mathcal{H}_0\oplus\mathcal{H}_2\oplus\cdots\oplus\mathcal{H}_{2\ell'}.$$

We have

$$\dim(\mathcal{H}^{\odot 2}_{\ell'}) = \binom{\dim(\mathcal{H}_{\ell'})+1}{2} = \binom{2\ell'+2}{2} = 2(\ell')^2+3\ell'+1$$
$$= \sum_{k=0}^{\ell'}(4k+1) = \dim(\mathcal{H}_0\oplus\mathcal{H}_2\oplus\cdots\oplus\mathcal{H}_{2\ell'}),$$

and therefore

$$\mathcal{H}^{\odot 2}_{\ell'} \simeq \mathcal{H}_0\oplus\mathcal{H}_2\oplus\cdots\oplus\mathcal{H}_{2\ell'}.$$

Let Φ_ℓ be the isomorphism $\mathcal{H}^{\odot 2}_\ell \to \mathcal{H}_0\oplus\mathcal{H}_2\oplus\cdots\oplus\mathcal{H}_{2\ell}$ defined by $\Phi_\ell = \Phi_{\ell,\ell}|_{\mathcal{H}^{\odot 2}_\ell}$.

We have seen how $L^2(S^2,\mu;\mathbb{C})$ decomposes into $SO(3)$-irreducible representations, and how tensor products and symmetric tensor powers of these irreducibles decompose into irreducibles. In the next section it will be essential that instead of the group $SO(3)$, we consider the full symmetry group $O(3)$ of S^2. The special orthogonal group $SO(3)$ forms a normal subgroup of $O(3)$. Since $\mathbb{R}^3$ is odd dimensional, the inversion operation $x\mapsto -x$ is not contained in $SO(3)$. This operation, which we denote by $-I$, generates a 2 element normal subgroup of $O(3)$, and the orthogonal group $O(3)$ is isomorphic to the direct product $\mathbb{Z}_2\times SO(3)$. Thus, for each irreducible representation $\mathcal{H}_\ell$ of $SO(3)$, we define two nonequivalent irreducible representations $\pi^p_\ell\colon O(3)\to U(\mathcal{H}^p_\ell)$, with $p=\pm 1$, where $\mathcal{H}^{+1}_\ell$ and $\mathcal{H}^{-1}_\ell$ are both isomorphic to $\mathcal{H}_\ell$ as Hilbert spaces, and where $\pi^p_\ell|_{SO(3)}$ is equivalent to

$\mathcal{H}_\ell$, but where $\pi_\ell^p(-I)f = pf$ for all f. It follows that π_ℓ^p is a subrepresentation of the unitary representation

$$L\colon O(3) \to U(L^2(S^2, \mu; \mathbb{C})),\ L(\gamma)f(x) = f(\gamma^{-1}x)$$

if and only if $p = (-1)^\ell$. For $f_1 \in \mathcal{H}_{\ell_1}^{p_1}$ and $f_2 \in \mathcal{H}_{\ell_2}^{p_2}$ we have

$$(\pi_{\ell_1}^{p_1} \otimes \pi_{\ell_2}^{p_2})(-I)(f_1 \otimes f_2) = \pi_{\ell_1}^{p_1}(-I)f_1 \otimes \pi_{\ell_2}^{p_2}(-I)f_2 = p_1p_2(f_1 \otimes f_2),$$

which implies

$$\mathcal{H}_{\ell_1}^{p_1} \otimes \mathcal{H}_{\ell_2}^{p_2} \simeq \mathcal{H}_{|\ell_1-\ell_2|}^{p_1p_2} \oplus \cdots \oplus \mathcal{H}_{\ell_1+\ell_2}^{p_1p_2} \quad \text{and} \quad (\mathcal{H}_{\ell'}^{p})^{\odot 2} \simeq \mathcal{H}_0^{+1} \oplus \mathcal{H}_2^{+1} \oplus \cdots \oplus \mathcal{H}_{2\ell'}^{+1}.$$

The operators Φ_{ℓ_1,ℓ_2} and Φ_ℓ defined above now become $O(3)$-equivariant, unitary operators

$$\Phi_{\ell_1,\ell_2}\colon \mathcal{H}_{\ell_1}^{p_1} \otimes \mathcal{H}_{\ell_2}^{p_2} \to \mathcal{H}_{|\ell_1-\ell_2|}^{p_1p_2} \oplus \cdots \oplus \mathcal{H}_{\ell_1+\ell_2}^{p_1p_2}$$

and

$$\Phi_\ell\colon (\mathcal{H}_{\ell'}^{p})^{\odot 2} \to \mathcal{H}_0^{+1} \oplus \mathcal{H}_2^{+1} \oplus \cdots \oplus \mathcal{H}_{2\ell'}^{+1}.$$

We use the operators Φ_{ℓ_1,ℓ_2} and Φ_ℓ to give an explicit definition of the operators M_τ from Section 6.2. Let e_ℓ denote the vector in $\mathbb{N}_0^\infty$ with $(e_\ell)_{\ell'} = 1$ if $\ell = \ell'$ and 0 otherwise. For $\tau \in D_0$, M_τ becomes the identity operator $\mathcal{H}_0^1 \to \mathbb{C}$. Each $\tau \in D_1$ is of the form $\tau = e_\ell$ for some ℓ, and for such a vector τ, the operator M_τ is the identity operator $\mathcal{H}_\ell^p \to \mathcal{H}_\ell^p$, where $p = (-1)^\ell$. If $\tau \in D_2$ is of the form $\tau = e_{\ell_1} + e_{\ell_2}$ with $\ell_1 \neq \ell_2$, then M_τ is given by the operator $\Phi_{\ell_1,\ell_2}^{-1}$. If $\tau \in D_2$ is of the form $2e_\ell$ for some $\ell \in \mathbb{N}_0$, then M_τ is given by Φ_ℓ^{-1}.

The representations in $\hat{\Gamma}$ can be indexed by $(\ell, p) \in \mathbb{N}_0 \times \{\pm 1\}$, and a symmetry adapted system of X_2 has the form

$$\big\{e_{(\ell,p),\tau,m} : \ell \in \mathbb{N}_0,\ p = \pm 1,\ \tau \in R_{(\ell,p)},\ -\ell \leq m \leq \ell\big\},$$

where $R_{(\ell,p)} = R_{(\ell,p)}^0 \cup R_{(\ell,p)}^1 \cup R_{(\ell,p)}^2$,

$$R_{(\ell,p)}^0 = \begin{cases} \{0\} & \text{if } \ell = 0 \text{ and } p = 1, \\ \emptyset & \text{otherwise,} \end{cases} \qquad R_{(\ell,p)}^1 = \begin{cases} \{e_\ell\} & \text{if } p = (-1)^\ell, \\ \emptyset & \text{otherwise,} \end{cases}$$

and

$$R_{(\ell,p)}^2 = \Big\{e_{\ell_1} + e_{\ell_2} : \delta_{2\nmid\ell} \leq |\ell_1 - \ell_2| \leq \ell \leq \ell_1 + \ell_2,\ (-1)^{\ell_1+\ell_2} = p\Big\},$$

where $\delta_{2\nmid\ell}$ is 1 if ℓ is odd, and 0 if ℓ is even. Here the basis element $e_{(\ell,p),\tau,m}$ can be computed as $L_i(T_i(M_\tau(Y_\ell^m)))$, where $i = \sum_{\ell'} \tau_{\ell'}$.

The rows and columns of the zonal matrices $Z_{(\ell,p)}(T, T')$ constructed using this symmetry adapted system are indexed by $R_{(\ell,p)}$. To obtain the finite dimensional inner approximating cones we need to select finite subsets $R_{(\ell,p),d} \subseteq R_{(\ell,p)}$ so that $\bigcup_{d=0}^\infty R_{(\ell,p),d} = R_{(\ell,p)}$, and for each d only finitely many sets $R_{(\ell,p),d}$ are nonempty. One particularly nice way (see below) to do this is to set

$$R_{(\ell,p),d} = \big\{\tau \in R_{(\ell,p)} : \sum_i \tau_i \leq d\big\}.$$

All irreducible representations of $O(3)$ are of real type, and the zonal matrices constructed above are all real-valued. As shown in Section 6.1, we can use these to construct the sequence $\{D_d\}$ of inner approximations of $\mathcal{C}(X_2 \times X_2)_{\succeq 0}^\Gamma$. By Lemma 6.1 we then get the desired sequence of inner approximations of $\mathcal{C}(I_2 \times I_2)_{\succeq 0}^\Gamma$.

The symmetry adapted system above is constructed in such a way that the functions

$$(x_1, \ldots, x_i) \mapsto A_2 Z_{(\ell,p)}(\cdot,\cdot)_{\tau,\tau'}(\{x_1, \ldots, x_i\}), \tag{2}$$

for $0 \leq i \leq 4$ and $\tau, \tau' \in R_{(\ell,p),d}$, are polynomials (of degree $2d$ in $3i$ variables), which will be important in the next section. As a final remark of this section we note that although the construction of the symmetry adapted systems is more involved for the cases $n > 3$ and $t > 2$, by using higher dimensional spherical harmonics one can still construct a symmetry adapted systems that have the above polynomial property.

7. Semidefinite programs with semialgebraic constraints

In this section we reduce the dual problem E_t^* for the Riesz s-energy problem on S^{n-1}, with $s \in \mathbb{N}$, to a sequence of semidefinite programs with semialgebraic constraints. Here, by a *semialgebraic constraint*, we mean the requirement that a polynomial, whose coefficients depend linearly on the entries of the positive semidefinite matrix variable(s), is nonnegative on a given basic closed semialgebraic set. A *basic closed semialgebraic set* in $\mathbb{R}^n$ is a subset that has a description of the form

$$S(g_1, \ldots, g_m) = \big\{x \in \mathbb{R}^n : g_i(x) \geq 0 \text{ for } i \in [m]\big\},$$

where $g_1, \ldots, g_m \in \mathbb{R}[x_1, \ldots, x_n]$. In Section 8 we show how these programs can be approximated by semidefinite programs and how symmetries in the semialgebraic constraints can be exploited to (further) block diagonalize the semidefinite programming formulations.

Following Section 2 and Section 5, the second step in the dual hierarchy for the Riesz s-energy problem on S^{n-1} reads

$$E_t^* = \sup\Big\{ \sum_{i=0}^{2t} \binom{N}{i} a_i : a \in \mathbb{R}^{\{0,\ldots,2t\}},\ K \in \mathcal{C}(I_t \times I_t)_{\succeq 0},\\ a_i + A_t K(S) \leq f(S) \text{ for } S \in I_{=i} \text{ and } i = 0, \ldots, 2t \Big\},$$

where I_t is the set of independent sets of cardinality at most t in a topological packing graph G on S^{n-1} as discussed in Section 2, and where

$$f(S) = \begin{cases} \|x-y\|_2^{-s} & \text{if } S = \{x, y\} \text{ with } x \neq y, \\ 0 & \text{otherwise.} \end{cases}$$

From the definition of G we obtain a $U \in (-1, 1)$ such that two vertices x and y in S^{n-1} are adjacent if and only if $x \cdot y \geq U$.

In the symmetrized version of this problem, as derived in Section 5.2, we restrict to $O(n)$-invariant kernels. In Section 6.3 we give a sequence $\{D_d\}$ of inner approximating cones of $\mathcal{C}(I_t \times I_t)_{\succeq 0}^{\Gamma}$, where $\Gamma = O(n)$. By replacing $\mathcal{C}(I_t \times I_t)_{\succeq 0}^{\Gamma}$ with D_d, we obtain the following sequence of approximations:

$$E_{t,d}^* = \sup\Big\{ \sum_{i=0}^{2t} \binom{N}{i} a_i : a \in \mathbb{R}^{\{0,\ldots,2t\}},\ F_\pi \in S_{\succeq 0}^{R_{\pi,d}} \text{ for } \pi \in \hat{\Gamma},\\ a_i + A_2 K(S) \leq f(S) \text{ for } S \in I_{=i} \text{ and } i = 0, \ldots, 2t \Big\},$$

where

$$K(J, J') = \sum_{\pi \in \hat{\Gamma}} \Big\langle F_\pi, Z_{\pi,d}(J, J)^* \Big\rangle.$$

For each d, the problem $E_{t,d}^*$ has finite dimensional variable space, and as shown in Section 5.2, we have $E_{t,d}^* \to E_t^* = E_t$ as $d \to \infty$, and, as shown in Section 4, we have $E_{N,d}^* \to E$ as $d \to \infty$.

Given $0 \leq i \leq 2t$, let $\mathbb{R}[x_1, \ldots, x_i]$ be the ring of real polynomials in ni variables, where each x_k denotes a vector of n variables. As shown in Section 6.3, there exist polynomials $p_i \in \mathbb{R}[x_1, \ldots, x_i]$ such that

$$p_i(x_1, \ldots, x_i) = a_i + A_t K(\{x_1, \ldots, x_i\}) \quad \text{for all} \quad \{x_1, \ldots, x_i\} \in I_{=i},$$

where the coefficients of p_i depend linearly on the entries of the vector a and the matrices $F_{(\ell,p)}$. For the case where $n = 3$ and $t = 2$ we have explicitly derived these polynomials in Section 6.3.

By construction, the polynomials $p_i \in \mathbb{R}[x_1, \ldots, x_i]$ are $O(n)$-invariant:

$$p_i(x_1, \ldots, x_i) = p_i(\gamma x_1, \ldots, \gamma x_i) \quad \text{for all} \quad x_1, \ldots, x_i \in S^2 \quad \text{and} \quad \gamma \in O(n).$$

So, by a theorem from invariant theory (see, for instance, [26, Theorem 10.2]) it follows that p_i can be written as a polynomial in the inner products

$$x_1 \cdot x_1, x_1 \cdot x_2, \ldots, x_i \cdot x_i.$$

On the sphere we have the identity $x_1 \cdot x_1 = \ldots = x_i \cdot x_i = 1$, so there exists a (in general non unique) polynomial $q_i \in \mathbb{R}[u_1, \ldots, u_{\binom{i}{2}}]$ such that

$$(3) \quad p_i(x_1, \ldots, x_i) = q_i(x_1 \cdot x_2, x_1 \cdot x_3, \ldots, x_{i-1} \cdot x_i) \quad \text{for all} \quad x_1, \ldots, x_i \in S^{n-1},$$

where the coefficients of q_i again depend linearly on the entries of the vector a and the matrices $F_{(\ell,p)}$. The above result is nonconstructive, and in Section 9 we show how we compute the polynomials q_i. The use of this theorem is why we need $O(n)$-invariance instead of just $SO(n)$-invariance, for otherwise the polynomials q_i would also need to depend on the determinants of the $n \times n$ matrices whose columns are given by n distinct vectors from $\{x_1, \ldots, x_i\}$, which means we would have too many variables to be able to perform computations.

The degenerate polynomials q_0 and q_1 have 0 variables; they are linear combinations of the entries of the vector a and the matrices $F_{(\ell,p)}$. The constraints $a_0 + A_t K|_{I_{=0}} \leq 0$ and $a_1 + A_t K|_{I_{=1}} \leq 0$ in $E_{t,d}^*$, where we use that $f|_{I_1} \equiv 0$, therefore reduce to the two linear constraints $q_0 \leq 0$ and $q_1 \leq 0$.

For distinct $x, y \in S^{n-1}$ we have

$$f(\{x, y\}) = \frac{1}{\|x - y\|_2^s} = \frac{1}{(2 - 2x \cdot y)^{s/2}}.$$

By using the substitution $w = \sqrt{2 - 2u}$, we can reformulate the constraint

$$a_2 + A_t K|_{I_{=2}} \leq f|_{I_{=2}}$$

in $E_{t,d}^*$ as the semialgebraic constraint

$$w^s \, q_2(1 - w^2/2) \leq 1 \quad \text{for} \quad w \in [\sqrt{2 - 2U}, 2].$$

If s is even, we can use a more efficient formulation (in terms of the degree of the polynomials), where we write the constraint $a_2 + A_t K|_{I_{=2}} \leq f|_{I_{=2}}$ in $E_{t,d}^*$ as the semialgebraic constraint

$$(2 - 2u)^{s/2} q_2(u) \leq 1 \quad \text{for} \quad u \in [-1, U].$$

The set of independent sets of cardinality i can be described as

$$I_{=i} = \Big\{ \{x_1, \ldots, x_i\} \subseteq S^{n-1} : x_k \cdot x_{k'} \leq U \text{ for } 1 \leq k < k' \leq i \Big\}.$$

So, with

$$P_i = \big\{ (x_1 \cdot x_2, x_1 \cdot x_3, \ldots, x_{i-1} \cdot x_i) : \{x_1, \ldots, x_i\} \in I_{=i} \big\},$$

a constraint of the form $a_i + A_t K|_{I=i} \leq 0$, for $i \in \{3, 4, \ldots, 2t\}$, can be written as $q_i|_{P_i} \leq 0$. This means we can write the problem $E^*_{t,d}$ as

$$
\begin{aligned}
(4) \quad E^*_{t,d} = \sup \Big\{ \sum_{i=0}^{2t} \binom{N}{i} a_i : \ & a \in \mathbb{R}^{\{0,\ldots,2t\}},\ F_\pi \in S^{R_{\pi,d}}_{\succeq 0} \text{ for } \pi \in \hat{\Gamma}, \\
& q_0 \leq 0,\ q_1 \leq 0, \\
& w^s\, q_2(1 - w^2/2) \leq 1 \text{ for } w \in [\sqrt{2-2U}, 2] \text{ if } 2 \nmid s, \\
& (2-2u)^{s/2} q_2(u) \leq 1 \text{ for } u \in [-1, U] \text{ if } 2 \mid s, \\
& q_i|_{P_i} \leq 0 \text{ for } i = 3, \ldots, 2t \Big\}.
\end{aligned}
$$

To describe P_i as a semialgebraic set we first observe that by using the Gram decomposition of a positive semidefinite matrix, it can be written as

$$
P_i = \Big\{ u \in \mathbb{R}^{\binom{i}{2}} : u_j \leq U \text{ for } j \in \big[\tbinom{i}{2}\big],\ \mathcal{E}(u) \succeq 0,\ \operatorname{rank}(\mathcal{E}(u)) \leq n \Big\},
$$

where $\mathcal{E}(u)$ is the symmetric $i \times i$-matrix with ones on the diagonal and the entries of u in the upper and lower diagonal parts. Using Sylvester's criterion for positive semidefinite matrices we obtain the semialgebraic description

$$
\begin{aligned}
(5) \qquad P_i = \Big\{ u \in \mathbb{R}^{\binom{i}{2}} : \ & u_j \leq U \text{ for } j \in [\tbinom{i}{2}], \\
& g(u) \geq 0 \text{ for } g \in G_{i,j} \text{ with } 2 \leq j \leq n, \\
& g(u) = 0 \text{ for } g \in G_{i,j} \text{ with } n+1 \leq j \leq i \Big\},
\end{aligned}
$$

where $G_{i,j}$ is the set of principal minors (the determinants of principal submatrices) of $\mathcal{E}(u) \in \mathbb{R}^{i \times i}$ of order j. This shows $E^*_{t,d}$ is a semidefinite program with semialgebraic constraints.

Here we make two observations which are important from a computational perspective in modeling the semialgebraic constraints as semidefinite constraints (see Section 8). The first observations is that P_i is compact and that the constraints $g(u) \geq 0$ for $g \in G_{i,2}$ provide an "algebraic certificate" of this compactness. As explained in the following section, this means the above description is Archimedean, so that we can apply Putinar's theorem.

The second observation concerns additional symmetry in the semialgebraic constraints. The particles in an energy minimization problem are interchangeable, and this implies that the polynomials $p_1, \ldots, p_{2t}$ are not only invariant under the group $O(n)$, but also under (some) permutations of the ni variables: We have

$$
p_i(x_1, \ldots, x_i) = p_i(x_{\sigma(1)}, \ldots, x_{\sigma(i)}) \quad \text{for all} \quad x_1, \ldots, x_i \in S^{n-1} \quad \text{and} \quad \sigma \in S_i.
$$

This implies we can choose the polynomials $q_1, \ldots, q_{2t}$ in such a way that they have additional symmetry. Let $\operatorname{Aut}^*(K_i)$ be the edge-automorphism group of the complete graph K_i on i vertices, and let $\phi_i \colon S_i \to \operatorname{Aut}^*(K_i)$ be the (not necessarily surjective) map that sends a permutation of the vertices of K_i to the corresponding permutation of the edges. If q_i is a polynomial that satisfies (3), then the polynomial

$$
(6) \qquad \overline{q}_i(u_1, \ldots, u_r) = \frac{1}{i!} \sum_{\sigma \in \phi_i(S_i)} q_i(u_{\sigma(1)}, \ldots, u_{\sigma(r)})
$$

also satisfies (3) and is invariant under the group $\phi_i(S_i)$. So, we may assume the polynomials q_i to be $\phi_i(S_i)$-invariant. Since the sets $G_{i,j}$ used to define the semialgebraic sets P_i are also invariant under this group, we say the semialgebraic constraints in the problem $E^*_{t,d}$ are $\phi_i(S_i)$-invariant.

8. Sum of squares characterizations for invariant polynomials

In this section we first give some background on how Putinar's theorem can be used to approximate a semidefinite program with semialgebraic constraints by a sequence of block diagonal semidefinite programs. Then we show how symmetry in the polynomial constraints can be used to further block diagonalize these semidefinite programming formulations into smaller blocks. We show this can lead to significant computational savings by applying this to the problems $E_{2,d}^*$ from the previous section.

The *quadratic module* generated by $g_1, \ldots, g_m \in \mathbb{R}[x_1, \ldots, x_n]$ is given by

$$M(g_1, \ldots, g_m) = \Big\{ \sum_{i=0}^m g_i s_i : s_0, \ldots, s_m \in \mathbb{R}[x_1, \ldots, x_n] \text{ SOS polynomials} \Big\},$$

where g_0 denotes the constant one polynomial, and where a sum of squares (SOS) polynomial is a polynomial of the form $\sum_k p_k^2$, with $p_1, \ldots, p_K \in \mathbb{R}[x_1, \ldots, x_n]$. Polynomials in $M(g_1, \ldots, g_m)$ are nonnegative on the *basic closed semialgebraic* set

$$S(g_1, \ldots, g_m) = \big\{ x \in \mathbb{R}^n : g_i(x) \geq 0 \text{ for } i \in [m] \big\}.$$

The usefulness of the quadratic module stems from Putinar's theorem [42], which says that under the condition that $M(g_1, \ldots, g_m)$ is Archimedean, every strictly positive polynomial on $S(g_1, \ldots, g_m)$ is contained in $M(g_1, \ldots, g_m)$. A quadratic module $M(g_1, \ldots, g_m)$ is *Archimedean* if it contains a polynomial p such that $S(p)$ is compact. Such a polynomial p provides an algebraic certificate of the compactness of $S(g_1, \ldots, g_m)$.

For $\delta \in \mathbb{N}_0$, we define the *truncated quadratic module* $M_\delta(g_1, \ldots, g_m)$ in the same way as we defined $M(g_1, \ldots, g_m)$, except now we require each s_i to have degree at most $2h_i$, where $h_i = \lfloor (\delta - \deg(g_i))/2 \rfloor$. Since higher degree terms can cancel each other out, the inclusion

$$M_\delta(g_1, \ldots, g_m) \subseteq M(g_1, \ldots, g_m) \cap \mathbb{R}[x_1, \ldots, x_n]_\delta,$$

can be strict. Here $\mathbb{R}[x_1, \ldots, x_n]_\delta$ denotes the vector space of polynomials of degree at most δ. Putinar's theorem shows that each polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ with $p(x_1, \ldots, x_n) > 0$ for all $(x_1, \ldots, x_n) \in S(g_1, \ldots, g_m)$ is contained in $M_\delta(g_1, \ldots, g_m)$ for all large enough δ, and in [39] an upper bound on the smallest δ for which this is true is given in terms of the polynomials $g_1, \ldots, g_m$, the degree of p, and the minimum of p over $S(g_1, \ldots, g_m)$.

Let $v_{h_i}(x)$ be a vector whose entries form a basis of $\mathbb{R}[x_1, \ldots, x_n]_{h_i}$. A polynomial of degree at most $2h_i$ is a sum of squares if and only if it can be written as

$$s_i(x) = v_{h_i}(x)^{\mathsf{T}} Q_i v_{h_i}(x),$$

where Q_i is a positive semidefinite matrix of size $\binom{n+h_i}{n}$. (To prove $v_{h_i}(x)^{\mathsf{T}} Q_i v_{h_i}(x)$ is a sum of squares one can use a Cholesky factorization $Q_i = R_i^{\mathsf{T}} R_i$). This implies

$$M_\delta(g_1, \ldots, g_m) \simeq \mathcal{S}_{\succeq 0}^{\binom{n+h_0}{n}} \times \cdots \times \mathcal{S}_{\succeq 0}^{\binom{n+h_m}{n}}.$$

In a semidefinite program with semialgebraic constraints, we can now approximate a constraint of the form

$$p(x_1, \ldots, x_n) \geq 0 \quad \text{for} \quad (x_1, \ldots, x_n) \in S(g_1, \ldots, g_m),$$

by a degree δ sum of squares characterization. By this we mean that we introduce additional positive semidefinite matrix variables $Q_0, \ldots, Q_m$ and replace the constraint $p|_{S(g_1,\ldots,g_m)} \geq 0$ by a set of linear constraints that enforces the identity

$$p(x) = \sum_{i=0}^m g_i(x) v_{h_i}(x)^{\mathsf{T}} Q_i v_{h_i}(x).$$

To obtain a set of linear constraints that enforces the above identity, we can express the left and right hand sides in terms of the same polynomial basis and equate the coefficients with respect to this basis. Here the basis choice for the entries of v_i and the basis choice for the linear constraints can have great impact on the numerical conditioning of the resulting semidefinite program (see, for instance, [29]), but in the computations in this paper we only use the standard basis because we only use polynomials of low degree.

The approximations given by the semidefinite programs obtained in this way become arbitrarily good as we take sum-of-squares characterizations of higher degrees. Moreover, if a semidefinite program with semialgebraic constraints has an optimal solution where all the inequalities are strict, then the optimum is obtained for a finite degree sum-of-squares characterization.

We will show that if p is invariant under the action of a group, then we can further block diagonalize the matrices Q_i. Let Γ be a finite subgroup of $U(\mathbb{C}^n)$. This induces the unitary representation

$$L\colon \Gamma \to U(\mathbb{C}[x_1,\dots,x_n]),\ L(\gamma)p(x) = p(\gamma^{-1}x), \tag{7}$$

where $\mathbb{C}[x_1,\dots,x_n]$ has the inner product $\langle p, q\rangle = \sum_\alpha p_\alpha \overline{q_\alpha}$. A polynomial p is said to be Γ-invariant if $L(\gamma)p = p$ for all $\gamma \in \Gamma$, and a set of polynomials $\{g_1,\dots,g_m\}$ is said to be Γ-invariant if

$$\{L(\gamma)g_1,\dots,L(\gamma)g_m\} = \{g_1,\dots,g_m\} \quad \text{for all} \quad \gamma \in \Gamma.$$

Let Γ_i be the stabilizer subgroup of Γ with respect to g_i; that is,

$$\Gamma_i = \{\gamma \in \Gamma : L(\gamma)g_i = g_i\}.$$

Since $g_0 = 1$ we have $\Gamma_0 = \Gamma$. In the next proposition we show that if the polynomials p and the set $\{g_1,\dots,g_m\}$ are invariant under the group action, then the sum of squares polynomials can be taken to be invariant under the corresponding stabilizer subgroups.

Proposition 8.1. *If $p \in M_\delta(g_1,\dots,g_m)$ is Γ-invariant, then there are Γ_i-invariant sum of squares polynomials $s_i \in \mathbb{R}[x_1,\dots,x_n]_{h_i}$ such that*

$$p = \sum_{i=0}^m g_i s_i.$$

Proof. Since p is Γ-invariant we have

$$p(x) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} L(\gamma)p(x) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} p(\gamma^{-1}x),$$

and since p lies in $M_\delta(g_1,\dots,g_m)$, we have

$$\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} p(\gamma^{-1}x) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\sum_{i=0}^m g_i(\gamma^{-1}x)s_i(\gamma^{-1}x).$$

Let $\Gamma_{i,j} = \{\gamma \in \Gamma : L(\gamma)g_i = g_j\}$, so that, for each $0 \le i \le m$, we have $\Gamma_{i,i} = \Gamma_i$ and Γ is the disjoint union of $\Gamma_{i,0},\dots,\Gamma_{i,m}$. Then,

$$\begin{aligned}\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\sum_{i=0}^m g_i(\gamma^{-1}x)s_i(\gamma^{-1}x) &= \frac{1}{|\Gamma|}\sum_{j=0}^m\sum_{i=0}^m\sum_{\gamma\in\Gamma_{i,j}} g_i(\gamma^{-1}x)s_i(\gamma^{-1}x)\\ &= \frac{1}{|\Gamma|}\sum_{j=0}^m g_j(x)\sum_{i=0}^m\sum_{\gamma\in\Gamma_{i,j}} s_i(\gamma^{-1}x)\end{aligned}$$

So, if we define

$$\bar{s}_j(x) = \frac{1}{|\Gamma|}\sum_{i=0}^m \sum_{\gamma\in\Gamma_{i,j}} s_i(\gamma^{-1}x),$$

then $p(x) = \sum_{j=0}^m g_j(x)\bar{s}_j(x)$. Since the cone of sum of squares polynomials is $\mathrm{GL}(\mathbb{R}[x])$-invariant, we see that the functions $\bar{s}_0,\ldots,\bar{s}_m$ are sums of squares polynomials. Moreover, for $\eta\in\Gamma_j$, we have

$$L(\eta)\bar{s}_j(x) = \frac{1}{|\Gamma|}\sum_{i=0}^m \sum_{\gamma\in\Gamma_{i,j}} s_i(\gamma^{-1}\eta^{-1}x) = \frac{1}{|\Gamma|}\sum_{i=0}^m \sum_{\gamma\in\eta\Gamma_{i,j}} s_i(\gamma^{-1}x),$$

so Γ_i-invariance of $\bar{s}_i$ follows from the identity

$$\Gamma_{j,k}\Gamma_{i,j} = \Gamma_{i,k} \quad \text{for all} \quad 0\le i,j,k\le m. \qquad \square$$

In [23] it is shown how the matrix used to represent an invariant sum of squares polynomial can be block diagonalized, and how this can be used to simplify semidefinite programs involving such polynomials. We combine this with Putinar's theorem and the above proposition to block diagonalize the representation of a positive invariant polynomial on an invariant semialgebraic set. To describe how this block diagonalization works we assume the group Γ in (7) consists of permutation matrices, because in this special case we can use the setting from Section 6, and in our application to $E_{t,d}^*$ all relevant groups are of this form.

We can view the matrix Q in a sum of squares representation

$$s(x) = v_h(x)^{\mathsf{T}} Q v_h(x)$$

as a positive definite kernel $[x_1,\ldots,x_n]_h \times [x_1,\ldots,x_n]_h \to \mathbb{R}$, where $[x_1,\ldots,x_n]_h$ is the set of monomials of degree at most h. The group Γ has an obvious action on $[x_1,\ldots,x_n]_h$, and if s is Γ-invariant, then we may assume Q to be a Γ-invariant kernel: Represent $L(\gamma)$ in the monomial basis, so that $L(\gamma)v(x) = v(\gamma^{-1}x)$, and

$$\begin{aligned} s(x) &= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} s(\gamma^{-1}x) = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} v(\gamma^{-1}x)^{\mathsf{T}} Q v(\gamma^{-1}x) \\ &= \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} v(x)^{\mathsf{T}} L(\gamma)^{\mathsf{T}} Q L(\gamma) v(x) = v(x)^{\mathsf{T}}\left(\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} L(\gamma)^{\mathsf{T}} Q L(\gamma)\right) v(x), \end{aligned}$$

which means we may replace Q by its symmetrization $1/|\Gamma|\sum_{\gamma\in\Gamma} L(\gamma)^{\mathsf{T}} Q L(\gamma)$.

By viewing Q as a Γ-invariant kernel $[x_1,\ldots,x_n]_h \times [x_1,\ldots,x_n]_h \to \mathbb{R}$, we get

$$Q(u,v) = \sum_{\pi\in\widehat{\Gamma}} \langle G_\pi, Z_\pi(u,v)\rangle, \quad \text{for} \quad u,v\in[x_1,\ldots,x_n]_h,$$

where the G_π are Hermitian positive semidefinite matrices. Here the Z_π are the zonal matrices as defined in Section 6.1, where the topological space X is now the finite set $[x_1,\ldots,x_n]_h$. We have,

$$\begin{aligned} s(x) &= \sum_{u,v\in[x_1,\ldots,x_n]_h} Q(u,v)uv \\ &= \sum_{\pi\in\widehat{\Gamma}} \langle \widehat{Q}(\pi), \sum_{u,v\in[x_1,\ldots,x_n]_h} Z_\pi(u,v)uv\rangle = \sum_{\pi\in\widehat{\Gamma}} \langle \widehat{Q}(\pi), Z_\pi(x)\rangle, \end{aligned}$$

where we define the modified zonal matrices

$$Z_\pi(x) = \sum_{u,v\in[x_1,\ldots,x_n]_h} Z_\pi(u,v)uv.$$

In general we have to use Hermitian positive semidefinite blocks G_π, but in our computations all groups only have real irreducible representations, so, as explained at the end of Section 6.1 we can use real blocks. Since the groups here are finite, we can compute the symmetry adapted system algorithmically. Before we explain how this is done we note that we can compute the size m_π of G_π without having to compute the actual block diagonalization.

For this we first notice that m_π now denotes the number of representations in an orthogonal decomposition of $\mathbb{R}[x]_h$ into irreducible unitary representations that are unitarily equivalent to π. First observe that $m_\pi = m_\pi(0)+\cdots+m_\pi(h)$, where $m_\pi(k)$ denotes the denotes the number of representations in an orthogonal decomposition of $\mathbb{R}[x]_{=k}$ into irreducible unitary representations that are unitarily equivalent to π. Here $\mathbb{R}[x]_{=k}$ is the space of homogeneous polynomials of degree k. We can compute the numbers $m_\pi(k)$ by a theorem of Molien [37], which gives the equality of the formal power series

$$\sum_{k=0}^{\infty} m_\pi(k) t^k = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \frac{\operatorname{trace}(\pi(\gamma))}{\det(I - tL(\gamma))}.$$

To compute the actual block diagonalization we use the projection algorithm as described in [46], which generates the symmetry adapted systems used to construct the zonal matrices. This algorithm works as follows: First define the operators

$$p^\pi_{j,j'} = \frac{d_\pi}{|\Gamma|} \sum_{\gamma \in \Gamma} \pi(\gamma^{-1})_{j,j'} L(\gamma).$$

Then, for each $\pi \in \hat{\Gamma}$ and $i \in [m_\pi]$, we let $\{e_{\pi,i,1}\}_{i=1}^{m_\pi}$ be a a basis of $\operatorname{Im}(p^\pi_{1,1})$, and we set $e_{\pi,i,j} = p^\pi_{j,1} e_{\pi,i,1}$ for all π, i, and $j > 1$. In [46] it is shown that this yields a (not necessarily orhtonormal) symmetry adapted system. Moreover, if we choose the bases $\{e_{\pi,i,1}\}_{i=1}^{m_\pi}$ of $\operatorname{Im}(p^\pi_{j,j'})$ to be orthonormal, then it is not difficult to show the resulting symmetry adapted system is also orthonormal.

We apply the above techniques to the problems $E^*_{2,d}$ from Section 7 for Riesz s-energy problems. First we show how to model the semialgebraic constraints using sum of squares characterizations without exploiting the additional symmetry. For the constraints $q_3|_{P_3} \leq 0$ and $q_4|_{P_4} \leq 0$ we use sum of squares characterizations of degree δ, and we denote the resulting semidefinite program by $E^*_{2,d,\delta}$.

If s is odd, we model the semialgebraic constraint

$$w^s\, q_2(1 - w^2/2) \leq 1 \quad \text{for} \quad w \in [\sqrt{2 - 2U}, 2]$$

by introducing positive semidefinite matrices $Q_{2,1}$ and $Q_{2,2}$ and adding a set of linear constraints to enforce the identity

$$w^s\, q_2(1-w^2/2) = 1+(w-\sqrt{2 - 2U})\, v_{h_1}(w)^{\mathsf{T}} Q_{2,1} v_{h_1}(w)+(2-w) v_{h_1}(w)^{\mathsf{T}} Q_{2,2} v_{h_1}(w).$$

As explained above, we know that for sufficiently large h_1, this sum-of-squares constraint approximates the above semialgebraic constraint arbitrarily well. However, for this special case, where we have an odd degree polynomial that is nonnegative on a compact interval, a result of Lukács (see, for instance, [40, 41]) says that for $h_1 = (2d+s-1)/2$ the above semialgebraic constraint and sum-of-squares constraint are identical.

If s is even, we model the semialgebraic constraint

$$(2 - 2u)^{s/2} q_2(u) \leq 1 \quad \text{for} \quad u \in [-1, U]$$

by introducing positive semidefinite matrices $Q_{2,1}$ and $Q_{2,2}$ and adding a set of linear constraints to enforce the identity

$$(2 - 2u)^{s/2} q_2(u) = 1 + v_{h_2}(u)^{\mathsf{T}} Q_{2,1} v_{h_2}(u) + (u - 1)(U - u) v_{h_3}(u)^{\mathsf{T}} Q_{2,2} v_{h_3}(u),$$

if $s/2+d$ is even, and

$$(2-2u)^{s/2}q_2(u) = 1 + (u-1)v_{h_4}(u)^{\mathsf{T}}Q_{2,1}v_{h_4}(u) + (U-u)v_{h_4}(u)^{\mathsf{T}}Q_{2,2}v_{h_4}(u),$$

if $s/2+d$ is odd. By the same result of Lukács mentioned above, the above semialgebraic constraint and sum of squares constraint are identical if we take $h_2 = (s/2+d)/2$, $h_3 = (s/2+d)/2-1$, and $h_4 = (s/2+d-1)/2$.

We model the semialgebraic constraint $q_3|_{P_3} \leq 0$ by introducing positive semidefinite matrix variables Q_3, $Q_{3,2,g}$ for $g \in G_{3,2}$, and $Q_{3,3,g}$ for $g \in G_{3,3}$, and adding a set of linear constraints to enforce the identity

$$q_3(u) + v_{h_1'}(u)^{\mathsf{T}}Q_3 v_{h_1'}(u) + \sum_{k\in\{2,3\}}\sum_{g\in G_{3,k}} g(u)v_{h_k'}(u)^{\mathsf{T}}Q_{3,g}v_{h_k'}(u) = 0, \tag{8}$$

where $h_1' = \lfloor \delta/2 \rfloor$, $h_2' = \lfloor (\delta-2)/2 \rfloor$, $h_3' = \lfloor (\delta-3)/2 \rfloor$.

In the semialgebraic description of P_4 in (5) we do not just have polynomial inequalities constraints but also the polynomial equality constraint $\det(\mathcal{E}(u)) = 0$. We could replace this equality constraint by two inequality constraints, but this would be computationally inefficient. Instead, we introduce positive semidefinite matrix variables Q_4, $Q_{4,2,g}$ for $g \in G_{4,2}$, $Q_{4,3,g}$ for $g \in G_{4,3}$, and $q_{\alpha,\pm 1} \in \mathbb{R}_{\geq 0}$ for $\alpha \in \mathbb{N}_0$ with $\|\alpha\|_1 = \sum_i \alpha_i \leq \delta - 6$, and use the sum of squares characterization

$$\begin{aligned} q_4(u) + v_{h_1'}(u)^{\mathsf{T}}Q_4 v_{h_1'}(u) + \sum_{k\in\{2,3\}}\sum_{g\in G_{4,k}} g(u)v_{h_k'}(u)^{\mathsf{T}}Q_{4,g}v_{h_k'}(u) \\ + \det(\mathcal{E}(u)) \sum_{\alpha\in\mathbb{N}_0^6:\|\alpha\|_1\leq\delta-6} (q_{\alpha,1}-q_{\alpha,-1})u_1^{\alpha_1}\cdots u_6^{\alpha_6} = 0. \end{aligned} \tag{9}$$

	Without symmetry reduction			With symmetry reduction		
δ	Q_4	$Q_{4,2,g}$	$Q_{4,3,g}$	Q_4	$Q_{4,2,g}$	$Q_{4,3,g}$
0	1	0	0	1	0	0
1	1	0	0	1	0	0
2	7	1	0	2	1	0
3	7	1	1	2	1	1
4	28	7	1	5	4	1
5	28	7	7	5	4	3
6	84	28	7	12	13	3
7	84	28	28	12	13	9
8	210	84	28	29	33	9
9	210	84	84	29	33	27
10	462	210	84	63	75	27
11	462	210	210	63	75	69
12	924	462	210	124	153	69
13	924	462	462	124	153	153
14	1716	924	462	228	291	153
15	1716	924	924	228	291	306
16	3003	1716	924	395	519	306
17	3003	1716	1716	395	519	570
18	5005	3003	1716	654	882	570
19	5005	3003	3003	654	882	999
20	8008	5005	3003	1040	1435	999

TABLE 1. Block sizes in the sum-of-squares modeling of the $i=4$ constraints in $E_{2,d,\delta}^*$ with and without symmetry reduction.

Now we show by how much we can reduce the largest block size in the semidefinite program $E_{2,d,\delta}^*$ by exploiting the symmetry in the semialgebraic constraints. The matrix

$$Q_4 \in S_{\succeq 0}^{\binom{6+\lfloor\delta/2\rfloor}{6}}$$

from (9) typically forms the largest block in $E_{2,d,\delta}^*$. This block is larger than any other matrix used in the sum-of-squares modeling, and unless d is much larger than

δ, it is larger than any of the F_π blocks. As explained at the end of Section 7, the polynomial q_4 and the set of polynomials in the semialgebraic description of P_4 are invariant under the group $\Gamma = \phi_4(S_4)$. The stabilizer subgroup Γ_1 of Γ with respect to the constant 1 polynomial is isomorphic to S_4. The stabilizer subgroup Γ_g of Γ with respect to a polynomial $g \in G_{4,k}$ is isomorphic to the Klein-Four group for $k = 2$ and to S_3 for $k = 3$. In Table 1 we first show the size of Q_4, $Q_{4,2,g}$, and $Q_{4,3,g}$ for different values of δ for the case where we are not using symmetry. We use a Magma [10] implementation of the Molien series mentioned above to compute the blocksizes that we get when we do exploit the symmetry. In Table 1 we then show the largest of these block sizes when block diagonalizing the matrices Q_4, $Q_{4,2,g}$, and $Q_{4,3,g}$. In our computation we will use $\delta = 6$ and $\delta = 8$, where we see this gives a 6 fold reduction in the largest block size in the semidefinite program.

9. Computations and discussion

Our goal here is to show how one can compute the 4-point bound E_2^* for Riesz s-energy problems on S^2, and to observe that these bounds are numerically (with high precision) sharp for $N = 5$ and $s = 1, 2, \ldots, 7$. To do this we develop a program that can generate the semidefinite programs $E_{2,d,\delta}^*$ from Section 8. We then solve these semidefinite programs for $d = 6$ and $\delta = 6, 8$ with a semidefinite programming solver and check that the optimal objective values given by the solver (consisting of 28 decimal digits) coincide with the first 28 decimals of the Riesz s-energy

$$\frac{6}{2^{s/2}} + \frac{3}{3^{s/2}} + \frac{1}{4^{s/2}}, \tag{10}$$

of the triangular bipiramid.

We implement the program in the Julia language [7], which is a high level language that allows for quick experimentation with different algorithms and data structures (which we did extensively for this project), and has a modern type system and JIT compiler that allows for fast execution of the code. We first generate the symmetry adapted system and the zonal matrices as described in Section 6.3. For this we develop a simple Julia library for sparse multivariate polynomials, which includes generators for the Laplace spherical harmonics in cartesian coordinates and a generator for the Clebsch–Gordan coefficients. To generate high precision solver input we perform all computations in high precision arithmetic using the MPFR library [20].

To compute the polynomials $q_0, \ldots, q_{2t}$ from Section 7, we need to write the polynomials from (2) as polynomials in the inner products. For this we need to solve a large number of instances of the following problem: Suppose $p \in \mathbb{R}[x_1, \ldots, x_i]_{2d}$, where each x_k is a vector of 3 variables, is $O(3)$-invariant. We want to find a polynomial $q \in \mathbb{R}[u_1, \ldots, u_r]$, with $r = \binom{i+1}{2}$, such that

$$p(x_1, \ldots, x_i) = q(x_1 \cdot x_1, x_1 \cdot x_2, \ldots, x_i \cdot x_i).$$

As mentioned in Section 7, by a nonconstructive theorem from invariant theory such a polynomial q is guaranteed to exist. If $m \in \mathbb{R}[u_1, \ldots, u_r]$ is a monomial, then the polynomial

$$m(x_1 \cdot x_1, x_1 \cdot x_2, \ldots, x_i \cdot x_i)$$

is homogeneous of degree $2\deg(m)$. This means we may assume $\deg(q) \leq d$. We construct a linear system $Ax = b$, where the rows of A and b are indexed by the monomials in $3i$ variables of degree at most $2d$, and the columns of A and rows of x by the monomials in s variables up to degree d. The size of A increases rapidly: for $i = 4$ and $d = 6$ it has about 2.7 million rows. The matrix is sparse, however, where the maximum number of nonzeros in a row is 3^d, and although this is exponential in d, for $d = 6$ this is just 729. We therefore store A in a sparse data structure. For $i = 4$, the system $Ax = b$ has more rows than columns. So we use a least squares approach and solve $A^{\mathsf{T}}Ax = A^{\mathsf{T}}b$ instead. The matrix A is in general not of full column rank (in general, q is not unique), which means $A^{\mathsf{T}}A$ is singular, so instead we solve the system $(A^{\mathsf{T}}A + \varepsilon I)x = A^{\mathsf{T}}b$, where $\varepsilon > 0$ is small. Because a high precision solver that can work with sparse data structures is not readily available, we implement a simple pivoting, sparse, high precision, Cholesky factorization algorithm. We use this to compute the Cholesky factorization $A^{\mathsf{T}}A + \varepsilon I = PR^{\mathsf{T}}RP^{\mathsf{T}}$, where P is a permutation matrix, and retrieve x using backwards substitution. Finally, we use the equation relating p and q to verify the correctness of the computed polynomial up to a large number of digits. We then use equation (6) to symmetrize the polynomial q.

We develop a GAP [10] script to generate the symmetry adapted systems used in Section 8 for the symmetrized sum-of-squares characterizations. For this we need the orthogonal (real unitary) irreducible representations of the relevant stabilizer subgroups of the symmetric groups S_3 and S_4. Here, the only groups with nonobvious irreducible representations are the symmetric groups themselves, and we use Young's orthogonal form (see, for instance, [6]) for these representations.

We develop a semidefinite programming specification library in Julia that allows for modeling polynomial equality constraints involving multivariate polynomials. We use this together with the above to generate the semidefinite programs and output these in the SDPA-sparse format [36].

Just as we did for the variables $q_{\alpha,\pm 1}$ from equation (9), we model the free variables $a_0, \ldots, a_5$ from (4) as the difference of two 1×1 positive semidefinite matrices. This, however, means that the resulting semidefinite programs are unbounded, which implies the dual programs are not strictly feasible. That is, the dual programs do not admit feasible solutions where all blocks are positive definite. For many semidefinite programming solvers this is a problem, and, in particular, the high precision solvers SDPA-QD and SDPA-GMP cannot be used in this situation. We therefore introduce a new parameter M and add the constraints that the variables used to model the free variables are at most M. In this way the dual problems become strictly feasible and can be solved with high precision solvers. Notice that for any value of M it is guaranteed we get a lower bound on the energy, and if M is large enough (we use $M = 1000$) this does not change the bound.

We model the polynomial equality constraints (8) and (9) by a linear constraint for each monomial. When we use the additional symmetry from Section 8, then this results in linearly dependent constraints. For some solvers such as the machine precision solver CSDP [9] this is not a problem, but solvers from the SDPA family do not work well in this case. Therefore we first remove identical constraints and then use a QR factorization of the constraint matrix of the semidefinite program to remove any remaining linearly dependent constraints.

The solver SDPA-QD works with quad double precision, which means solving a semidefinite program with this solver yields a solution with approximately 28 decimal digits of precision. In all sharp instances that we compute we verify that at least the first 28 decimal digits given by the solver agree with the first 28 decimals of the energy of a configuration. Notice that to get more digits we do not have have to increase the parameters d and δ – we can use the same semidefinite program – but we simply have to increase the precision parameter in the Julia code that generates the semidefinite program and increase the precision parameter in the solver (where we switch to SDPA-GMP for variable precision instead of SDPA-QD). Here, however, there is no reason to use higher precision. Notice that this very different from the sphere packing problem, which was recently solved in dimensions 8 and 24 using 2-point bounds [12, 49], where one needs to increase the number of terms in the inverse Fourier transform (which would correspond to increasing d in our bound $E^*_{t,d,\delta}$) for the bound to get closer to the exact optimal value.

As is to be expected, the computation time increases strongly with δ. Computing the bound $E^*_{2,6,6}$ with SDPA-QD takes approximately 10 minutes (on a standard desktop computer) if we do use the additional symmetry from Section 8, and takes approximately 80 minutes if we do not use this additional symmetry. Computing the bound $E^*_{2,6,8}$ takes approximately 7 hours with additional symmetry and 150 hours without additional symmetry. Here, the value of s itself has virtually no impact on the computation time, however, as s increases we do need to increase the parameter δ to get a sharp bound. We observe that for $s = 1, \ldots, 5$, the bound $E^*_{2,6,6}$ is numerically sharp, and for $s = 6, 7$ the bound $E^*_{2,6,8}$ is numerically sharp; that is, the 28 decimal digits given by the solver agree with (10).

It would be of interest to use the (high precision) floating point output of the solver to construct optimality certificates for the triangular bipiramid. In [13], 3-point semidefinite programming bounds were used in this way to prove optimality of the rhombic dodecahedron. Here the floating point solver output, which provides a near feasible and near optimal solution, was rounded to a solution (consisting of algebraic numbers) that lies on the optimal face, and a very simple computer program is then used to verify (in exact arithmetic) that the solution is indeed feasible for the semidefinite program. There seem to be no principal objections to using the same approach for the bounds $E^*_{2,d,\delta}$. The only difference is that in our case not only the solver uses floating point arithmetic, but also the solver input consists of high precision floating point numbers. The reason for this is that we use numerical linear algebra to compute the polynomials $q_0, \ldots, q_4$ in terms of the inner products (see above). One approach to generate the semidefinite programs exactly (using algebraic numbers) would be to use Gröbner bases instead of numerical linear algebra for computing these polynomials.

To find new sharp instances, it would also be of interest to compute E_2 for energy minimization problems on higher dimensional spheres, or other compact spaces. As observed in [13], of particular interest is the case of 24 particles on S^3, as here the 24-cell seems to be optimal for some potential functions, but for other potentials the optimal configurations seem to be more exotic. It would be remarkable if E_2 would be universally sharp for 24 particles on S^3. It would also be interesting to use the techniques developed in this paper to compute 4-point bounds for packing problems such as spherical code problems on S^{n-1}. Of particular interest would be the spherical code problem $A(4, \arccos(1/3))$, where a construction of 14 points exists, and where the 2 and 3-point bounds give the upper bounds 16 and 15 [4].

Appendix A. Invariant positive definite kernels

In this appendix we prove theorems concerning the "simulateneous block diagonalization" of invariant positive definite kernels. These results are used in Section 6.1, where we also give more background information and introduce some of the notation used in this appendix.

The first theorem characterizes the extreme rays of the cone of invariant positive definite kernels. As a special case, this shows

$$\partial_r\big(\mathcal{C}(X \times X; \mathbb{C})_{\succeq 0}\big) = \Big\{ f \otimes \bar{f} : f \in \mathcal{C}(X; \mathbb{C}) \Big\},$$

where we use the notation ∂_r for the extreme rays of a cone. This theorem and its proof are a generalization to kernels of a result in harmonic analysis about functions of positive type as given in [19]. In this appendix X is a compact metric space with a continuous action of a compact group Γ (in the following theorem, however, we may assume X and Γ to be locally compact).

Theorem A.1. *We have*

$$\partial_r\big(\mathcal{C}(X \times X; \mathbb{C})_{\succeq 0}^{\Gamma}\big) = \Big\{K_\varphi : \pi \in \hat{\Gamma},\, \varphi \in \mathrm{Hom}_\Gamma(X, \mathcal{H}_\pi)\Big\},$$

where $\mathcal{H}_\pi$ is the Hilbert space of the representation $\pi \in \hat{\Gamma}$, where $\mathrm{Hom}_\Gamma(X, \mathcal{H}_\pi)$ is the space of Γ-equivariant maps $X \to \mathcal{H}_\pi$, and where K_φ is defined by

$$K_\varphi(x, y) = \big\langle \varphi(x), \varphi(y) \big\rangle \quad \textit{for all} \quad x, y \in X.$$

Proof. Let K be a nonzero kernel in $\mathcal{C}(X \times X; \mathbb{C})_{\succeq 0}^{\Gamma}$. As shown in [13], we can use a Gelfand-Naimark-Segal type construction to build a unitary representation $\pi \colon \Gamma \to U(\mathcal{H}_\pi)$ and a nonzero function $\varphi \in \mathrm{Hom}_\Gamma(X, \mathcal{H}_\pi)$, so that

$$K(x, y) = \big\langle \varphi(x), \varphi(y) \big\rangle \quad \text{for all} \quad x, y \in X.$$

Indeed, let $\mathbb{C}^X$ be the complex vector space of formal linear combinations of elements in X, and define the subspace $N = \mathrm{span}\{x \in X : K(x, x) = 0\}$. Define an inner product on the quotient space $\mathbb{C}^X/N$ by setting $\langle x + N, y + N\rangle = K(x, y)$ for all $x, y \in X$ and extending by (anti)linearity. The action of Γ on X extends to the homomorphism $\pi \colon \Gamma \to U(\mathcal{H}_\pi)$, where $\mathcal{H}_\pi$ is the Hilbert space obtained by completing $\mathbb{C}^X/N$ in the metric defined by the inner product $\langle \cdot, \cdot\rangle$. Here $\pi(\gamma)$ is an isometry because K is Γ-invariant, and because $\pi(\gamma)$ is invertible, it is a unitary operator. Since $\langle \pi(\gamma)x + N, y + N\rangle = K(\gamma x, y)$, it follows from K being continuous and the action of Γ on X being continuous, that the map $\gamma \to \langle \pi(\gamma)x + N, y + N\rangle$ is continuous. So π is a unitary representation. We define the Γ-equivariant map $\varphi \colon X \to \mathcal{H}$ by $\varphi(x) = x + N$. This map is continuous, because

$$\|\varphi(y) - \varphi(x)\|^2 \leq K(x, x) + K(y, y) - K(x, y) - K(y, x),$$

and, moreover, φ is injective and has dense span.

Now assume K spans an extreme ray. If π is reducible, then $\mathcal{H}_\pi$ admits a nontrivial orthogonal decomposition $\mathcal{M}_1 \oplus \mathcal{M}_2$ into Γ-invariant subspaces. Let $P_i \colon \mathcal{H} \to \mathcal{M}_i$ be the projector onto $\mathcal{M}_i$, and set $\varphi_i = P_i \circ \varphi$. Let

$$K_i(x, y) = \langle \varphi_i(x), \varphi_i(y)\rangle \quad \text{for} \quad x, y \in X,$$

so that $K = K_1 + K_2$. Now we show the kernels K_1 and K_2 do not lie on the same ray: If $K_2 = |c|^2 K_1$ for some nonzero $c \in \mathbb{C}$, then we can define a Γ-equivariant unitary operator $T \colon \mathcal{M}_2 \to \mathcal{M}_1$ by setting $T(\varphi_2(x)) = c\,\varphi_1(x)$ for all $x \in X$. But this implies $\varphi = \varphi_1 + \varphi_2 = c^{-1}T \circ \varphi_2 + \varphi_2$, and this contradicts with φ being injective and having dense span in $\mathcal{H}_\pi$. Therefore, π must be irreducible.

Now assume K is a nonzero kernel in $\mathcal{C}(X \times X; \mathbb{C})_{\succeq 0}^{\Gamma}$, such that

$$K(x, y) = \big\langle \varphi(x), \varphi(y) \big\rangle, \quad \text{for all} \quad x, y \in X,$$

for some irreducible unitary representation $\pi \colon \Gamma \to U(\mathcal{H}_\pi)$ and $\varphi \in \mathrm{Hom}_\Gamma(X, \mathcal{H}_\pi)$. Let K_1 and K_2 be two kernels in $\mathcal{C}(X \times X; \mathbb{C})_{\succeq 0}^{\Gamma}$ with $K = K_1 + K_2$. We have $K_1(x, x) = K(x, x) - K_2(x, x) \leq K(x, x)$ for all $x \in X$, so

$$|K_1(x, y)| \leq K_1(x, x)^{1/2} K_1(y, y)^{1/2} \leq K(x, x)^{1/2} K(y, y)^{1/2} \quad \text{for all} \quad x, y \in X.$$

This means we can define the bounded Hermitian form $\langle \cdot, \cdot\rangle_1$ on $\mathbb{C}^X/N$ by setting

$$\langle \varphi(x) + N, \varphi(y) + N\rangle_1 = K_1(x, y)$$

and extending by (anti)linearity. The form $\langle \cdot, \cdot\rangle_1$ is continuous since it is bounded, so we can extend it to the Hilbert space $\mathcal{H}_\pi$. By the Riesz representation theorem for Hilbert spaces there is a bounded self-adjoint operator T on $\mathcal{H}_\pi$ such that

$$\langle \varphi(x) + N, \varphi(y) + N\rangle_1 = \langle T(\varphi(x) + N), \varphi(y) + N\rangle \quad \text{for all} \quad x, y \in X.$$

This operator is Γ-equivariant: For all $x, y \in X$ and $\gamma \in \Gamma$ we have

$$\begin{aligned}\langle T\pi(\gamma)(\varphi(x)+N), \varphi(y)+N\rangle &= \langle \varphi(\gamma^{-1}x)+N, \varphi(y)+N\rangle_1\\ &= K_1(\gamma^{-1}x, y) = K_1(x, \gamma y)\\ &= \langle \varphi(x), \varphi(\gamma y)\rangle_1 = \langle T\varphi(x), \pi(\gamma^{-1})\varphi(y)\rangle\\ &= \langle \pi(\gamma)T\varphi(x), \varphi(y)\rangle.\end{aligned}$$

Since π is irreducible, Schur's lemma states there is a $c \in \mathbb{C}$ such that $T = cI$. But this means that

$$K_1(x, y) = \langle \varphi(x)+N, \varphi(y)+N\rangle_1 = \langle T(\varphi(x)+N), \varphi(y)+N\rangle = cK(x, y),$$

for all $x, y \in X$, and hence $K_1 = cK$ and $K_2 = (1-c)K$, which shows that K spans an extreme ray. □

Next, we prove the existence of a symmetry adapted system. For this we first need a few lemmas.

Lemma A.2. *The space X admits a strictly positive, Γ-invariant, Radon probability measure.*

Proof. Let $\{x_i\}$ be a dense sequence in X and let $\{a_i\}$ be a sequence of strictly positive numbers that sum to one. Define a Borel probability measure μ by setting $\mu(U) = \sum_{i : x_i \in U} a_i$ for all Borel sets U. The desired measure is obtained by averaging μ over the Haar measure of Γ. □

We say that a sequence $\{I_n\}$ of kernels in $\mathcal{C}(X \times X; \mathbb{C})$ is an *approximate identity* of X if $\|T_{I_n} f - f\|_\infty \to 0$ as $n \to \infty$ for each $f \in \mathcal{C}(X; \mathbb{C})$, where

$$T_K \colon \mathcal{C}(X; \mathbb{C}) \to \mathcal{C}(X; \mathbb{C}),\ T_K f(x) = \int K(x, y) f(y)\, d\mu(y),$$

and where μ is some fixed strictly positive Γ-invariant Radon probability measure.

Lemma A.3. *The space X admits an approximate identity $\{I_n\}$, where each kernel I_n may be assumed to be real-valued, symmetric, and Γ-invariant.*

Proof. Let d be a compatible metric on X. Let $\{U_i^1\}, \{U_i^2\}, \ldots$ be a sequence of finite open covers of X such that for all i and n the diameter of U_i^n is at most $1/n$. For each i and n inductively select a compact set $C_i^n \subseteq U_i^n$ such that

$$\mu(U_i^n \setminus C_i^n) \le \mu(C_i^n)/n,$$

(this is possible by inner regularity of μ), and remove C_i^n from the sets U_j^n for $j \neq i$. We then have $C_i^n \cap U_{i'}^n = \emptyset$ for all n and all distinct i and i'.

Let $\{p_i^n\}_i$ be a partition of unity subordinate to the cover $\{U_i^n\}_i$, so that the restriction of p_i^n to C_i^n is identically 1, and define the kernel $K_n \in \mathcal{C}(X \times X)$ by the finite sum

$$K_n(x, y) = \sum_i \frac{p_i^n(x) p_i^n(y)}{\mu(C_i^n)}.$$

Let $f \in \mathcal{C}(X; \mathbb{C})$ and $\varepsilon > 0$. For large enough n we have

$$\mu(U_i^n \setminus C_i^n) \le \frac{\mu(C_i^n)}{2\|f\|_\infty}\varepsilon \quad \text{and} \quad \sup_{x, y \in C_i^n} |f(x) - f(y)| \le \frac{1}{2}\varepsilon \quad \text{for all} \quad i.$$

Then for each $x \in X$,

$$|T_{K_n} f(x) - f(x)| = \left| \sum_i \int_{U_i^n} \frac{p_i^n(x) p_i^n(y)}{\mu(C_i)} f(y)\, d\mu(y) - f(x) \right| \le A + B,$$

where

$$A = \left| \sum_i \int_{C_i^n} \frac{p_i^n(x) p_i^n(y)}{\mu(C_i^n)} f(y)\, d\mu(y) - f(x) \right|$$
$$= \left| \sum_i \frac{p_i^n(x)}{\mu(C_i^n)} \int_{C_i^n} |f(y) - f(x)|\, d\mu(y) \right| \leq \sum_i p_i^n(x) \frac{\varepsilon}{2} = \frac{\varepsilon}{2}$$

and

$$B = \left| \sum_i \int_{U_i^n \setminus C_i^n} \frac{p_i^n(x) p_i^n(y)}{\mu(C_i^n)} f(y)\, d\mu(y) \right|$$
$$= \sum_i \frac{p_i^n(x)}{\mu(C_i^n)} \int_{U_i^n \setminus C_i^n} p_i^n(y) |f(y)|\, d\mu(y) = \sum_i p_i^n(x) \frac{\mu(U_i^n \setminus C_i^n)}{\mu(C_i^n)} \|f\|_\infty \leq \frac{\varepsilon}{2}.$$

So, for each $\varepsilon > 0$ we have $\|T_{K_n} f - f\|_\infty \leq \varepsilon$ for sufficiently large n, which means that the sequence $\{K_n\}$ is an approximate identity.

Let

$$I_n(x, y) = \int_\Gamma K_n(\gamma x, \gamma y)\, d\gamma,$$

where we integrate against the normalized Haar measure of Γ. Then $\{I_n\}$ is an approximate identity, and each I_n is real-valued, symmetric, and Γ-invariant. □

We need the following part of the Peter–Weyl theorem. A proof for the case where a compact group acts on itself can be found in for instance [19], and a generalization of this to the setting of a compact group acting on a compact metric space can be found in [28].

Lemma A.4. *The space $\mathcal{C}(X; \mathbb{C})$ is equal to the closure of the sum of its finite dimensional Γ-invariant subspaces.*

We also need the following variation on the Schur orthogonality relations, for which a proof can be found in [50].

Lemma A.5. *Let $\pi \colon \Gamma \to U(\mathcal{H})$ be a unitary representation, and let $\langle \cdot, \cdot \rangle$ be a Γ-invariant sesquilinear from on $\mathcal{H}$. Let $\mathcal{M}$ and $\mathcal{M}'$ be finite-dimensional, irreducible subrepresentations with orthonormal bases $\{e_i\}$ and $\{e_j'\}$.*

1. *If $\mathcal{M}$ and $\mathcal{M}'$ are not equivalent, then $\langle e_i, e_j' \rangle = 0$ for all i and j.*
2. *If there exists a Γ-equivariant bijection $T \colon \mathcal{M} \to \mathcal{M}'$ such that $Te_i = e_i'$ for all i, then there is a $c \in \mathbb{C}$ such that $\langle e_i, e_j' \rangle = c\delta_{i,j}$ for all i and j.*

Theorem A.6. *Let X be a compact metric space with a continuous action of a compact group Γ. The space X admits a symmetry adapted system.*

Proof. Let C be the set of all linearly independent sets of nontrivial, finite dimensional, Γ-invariant subspaces of $\mathcal{C}(X; \mathbb{C})$. Here, a set S of subspaces is said to be *linearly independent* if for any $n \in \mathbb{N}$ and distinct $A, B_1, \ldots, B_n \in S$, the intersection of A with the sum $B_1 + \cdots + B_n$ is the zero space. This is equivalent to requiring that the union of any set of bases of the subspaces in S is linearly independent. If X is nonempty, then $\mathcal{C}(X; \mathbb{C})$ is nonempty, so by Lemma A.4 the set C is nonempty.

Define a partial order on C by set inclusion. Given a chain T in C, the union of the sets in T is also in C: Given $n \in \mathbb{N}$ and distinct $A, B_1, \ldots, B_n \in \bigcup T$, there must be some set in T containing the sets $A, B_1, \ldots, B_n$, hence these sets are nontrivial, finite dimensional, Γ-invariant, and $A \cap (B_1 + \cdots + B_n) = \{0\}$, which means that $\bigcup T \in C$. Therefore, any chain in C has an upper bound, and by Zorn's lemma C contains a maximal element M.

Let P be the sum of all sets in M and let $\overline{P}$ be the closure of P in the uniform topology of $\mathcal{C}(X;\mathbb{C})$. If $\overline{P}$ is not equal to $\mathcal{C}(X;\mathbb{C})$, then by Lemma A.4 there exists a finite dimensional, Γ-invariant subspace V of $\mathcal{C}(X;\mathbb{C})$ containing a vector u that does not lie in P. The cyclic subspace $W = \operatorname{span}\{L(\gamma)u : \gamma \in \Gamma\}$ has trivial intersection with P because $L(\gamma)u \notin L(\gamma)P = P$ for all $\gamma \in \Gamma$. This means that $M \cup \{W\}$ is a linearly independent set of subspaces. The space W is Γ-invariant, and moreover, W is finite dimensional since it is a subspace of the finite dimensional space V. So $M \cup \{W\}$ is contained in C. This contradicts maximality of M, so $\overline{P}$ must be equal to $\mathcal{C}(X;\mathbb{C})$.

Since the representation in M are finite dimensional, by Maschke's theorem for compact groups they decompose into irreducible subrepresentations of $\mathcal{C}(X;\mathbb{C})$, and we may assume M to be a linearly independent set of Γ-irreducible subspaces of $\mathcal{C}(X;\mathbb{C})$ whose sum is uniformly dense.

Denote by $m_\pi \in \{0, 1, \ldots, \infty\}$ the number of representations in M that are equivalent to π. Select appropriate orthonormal bases $f_{\pi,i,1}, \ldots, f_{\pi,i,d_\pi}$ of the representations in M, so that the span of

$$\big\{f_{\pi,i,j} : \pi \in \hat{\Gamma},\, i \in [m_\pi],\, j \in [d_\pi]\big\}$$

is uniformly dense in $\mathcal{C}(X;\mathbb{C})$, and so that there are Γ-equivariant unitary operators $T_{\pi,i,i'} \colon \mathcal{H}_{\pi,i} \to \mathcal{H}_{\pi,i'}$ with $f_{\pi,i',j} = T_{\pi,i,i'} f_{\pi,i,j}$ for all π, i, i', and j. Now give this system any ordering where $f_{\pi,i,j}$ occurs before $f_{\pi,i',j'}$ whenever $i < i'$, and apply the Gram–Schmidt process to obtain a complete orthonormal system $\{e_{\pi,i,j}\}$. By Lemma A.5 we have

$$e_{\pi,i,j} = f_{\pi,i,j} - \sum_{i'=1}^{i-1} \langle e_{\pi,i,j}, e_{\pi,i',j}\rangle e_{\pi,i',j} = f_{\pi,i,j} - \sum_{i'=1}^{i-1} c_{\pi,i,i'} e_{\pi,i',j},$$

where $c_{\pi,i,i'} = \langle e_{\pi,i,j}, e_{\pi,i',j}\rangle$ does not depend on j. It follows that the system $\{e_{\pi,i,j}\}$ is symmetry adapted, which completes the proof. □

We now use the previous theorem to prove that the union of the sequence of inner approximations constructed in Section 6.1 is uniformly dense. For this we need one more lemma, for which a proof can be found in [28], which is a generalization of a proof from [15].

Lemma A.7. *A Γ-invariant kernel $K \in \mathcal{C}(X \times X)$ is positive definite if and only if $\hat{K}(\pi)$ is positive semidefinite for every $\pi \in \hat{\Gamma}$.*

Now we can show the sequence of inner approximations converges. A similar results is shown in [3], but there it is required that Γ is contained in a bigger group that has a transitive action. Using the existence of a symmetry adapted system as proved above we can avoid this requirement.

Theorem A.8. *The cone $\bigcup_{d=0}^\infty C_d$ is uniformly dense in $\mathcal{C}(X \times X;\mathbb{C})_{\succeq 0}^\Gamma$.*

Proof. Lemma A.3 shows there exists an approximate identity $\{I_n\}$ of X, where each I_n is real-valued, symmetric, and Γ-invariant. By Theorem A.6 there exists a symmetry adapted system $\{e_{\pi,i,j}\}$ of X. We use this to define

$$S_d = \operatorname{span}\Big\{e_{\pi,i,j} : \pi \in \hat{\Gamma},\, i \in R_{\pi,d},\, j \in [d_\pi]\Big\}.$$

We have the inclusions

$$S_0 \subseteq S_1 \subseteq \ldots \subseteq \mathcal{C}(X;\mathbb{C}),$$

and $\bigcup_{d=0}^\infty S_d$ is uniformly dense in $\mathcal{C}(X;\mathbb{C})$. This means that for each n, there exists a sequence $\{I_{n,d}\}_d$ of real-valued kernels, with $I_{n,d} \in S_d \times S_d$ for all d and $I_{n,d} \to I_n$ uniformly as $d \to \infty$. We may assume the kernels $I_{n,d}$ to be symmetric and Γ-invariant: The sequence $\{I_{n,d}\}_d$ converges uniformly to I_n, and since I_n

is symmetric and Γ-invariant, it follows that $\bar{I}_{n,d}$ also converges to I_n uniformly, where $\bar{I}_{n,d}$ is the symmetric Γ-invariant kernel defined by integrating against the Haar measure of Γ:

$$\bar{I}_{n,d}(x,y) = \frac{1}{2}\int \Big(I_{n,d}(\gamma x, \gamma y) + I_{n,d}(\gamma y, \gamma x)\Big)\, d\gamma.$$

Since S_d is Γ-invariant, we have $\bar{I}_{n,d} \in S_d \otimes S_d$, so we can replace $I_{n,d}$ by $\bar{I}_{n,d}$. For each n, let d_n be an integer such that $\|I_{n,d_n} - I_n\|_\infty \leq 1/n$. It follows that $\{I_{n,d_n}\}_n$ is an approximate identity of X.

For each n we define the kernel $\Pi_n \in \mathcal{C}(X^2 \times X^2)$ by

$$\Pi_n((x,y),(x',y')) = I_{n,d_n}(x,x')\overline{I_{n,d_n}(y,y')} \quad \text{for} \quad x,x',y,y' \in X.$$

In the remainder of the proof we show $\{\Pi_n\}$ is an approximate identity of X^2, and we show that the range of T_{Π_n} is contained in C_{d_n}, and hence in $\bigcup_{d=0}^\infty C_d$.

For $f, g \in \mathcal{C}(X;\mathbb{C})$, we have

$$\begin{aligned}\|T_{K_n}(f\otimes g) - f\otimes g\|_\infty &= \|T_{I_{n,d}} f \otimes T_{I_{n,d}} g - f \otimes g\|_\infty \\ &\leq \|T_{I_{n,d}} f\|_\infty \|T_{I_{n,d}} g - g\|_\infty + \|T_{J_n} f - f\|_\infty \|g\|_\infty \to 0.\end{aligned}$$

The span of kernels of the form $f \otimes g$ is uniformly dense in $\mathcal{C}(X \times X;\mathbb{C})$, so, given a kernel $K \in \mathcal{C}(X\times X;\mathbb{C})_{\succeq 0}^\Gamma$, the sequence $T_{\Pi_n}K$ converges uniformly to K. This shows $\{\Pi_n\}$ is an approximate identity.

The kernel $T_{\Pi_n}K$ lies in $S_{d_n} \otimes S_{d_n}$ because I_{n,d_n} lies in $S_{d_n} \otimes S_{d_n}$. Thus,

$$T_{\Pi_n}K = \sum_{\pi,\pi' \in \hat{\Gamma}} \sum_{i \in R_{\pi,d_n}} \sum_{i' \in R_{\pi',d_n}} \sum_{j=1}^{d_\pi} \sum_{j'=1}^{d_{\pi'}} \Big\langle e_{\pi,i,j}, e_{\pi',i',j'} \Big\rangle_{T_{\Pi_n}K} e_{\pi,i,j} \otimes \overline{e_{\pi',i',j'}},$$

where

$$\Big\langle e_{\pi,i,j}, e_{\pi',i',j'} \Big\rangle_{T_{\Pi_d}K} = \iint T_{\Pi_d}K(x,y) e_{\pi,i,j}(x) \overline{e_{\pi',i',j'}(y)}\, d\mu(x) d\mu(y)$$

is a sesquilinear form that is Γ-invariant, because $T_{\Pi_d}K$ is Γ-invariant. Lemma A.5 shows $\langle e_{\pi,i,j}, e_{\pi',i',j'}\rangle_K = 0$ when $\pi \neq \pi'$ or $j \neq j'$, and $\langle e_{\pi,i,j}, e_{\pi,i',j}\rangle_K$ does not depend on j. This shows

$$T_{\Pi_n}K(x,y) = \sum_{\pi \in \widehat{\Gamma}} \Big\langle \widehat{T_{\Pi_n}K}(\pi), Z_{\pi,d_n}(x,y) \Big\rangle.$$

The kernel $T_{\Pi_n}K$ is positive definite, so by Lemma A.7, the matrices $\widehat{T_{\Pi_n}K}(\pi)$ are positive semidefinite. So, the kernel

$$(x,y) \mapsto \langle \widehat{T_{\Pi_n}K}(\pi), Z_{\pi,d_n}(x,y)\rangle$$

lies in C_{π,d_n}, and $T_{\Pi_n}K \in \sum_{\pi\in\hat{\Gamma}} C_{\pi,d_n} = C_{d_n}$. □

Acknoledgements

I would like to thank Henry Cohn, Dion Gijswijt, and Frank Vallentin for helpful discussions.

References

[1] N. N. Andreev, *An extremal property of the icosahedron*, East J. Approx. **2** (1996), no. 4, 459–462.

[2] J. M. Ansemil and K. Floret, *The symmetric tensor product of a direct sum of locally convex spaces*, Studia Math. **129** (1998), no. 3, 285–295.

[3] C. Bachoc, *Lecture notes: Semidefinite programming, harmonic analysis and coding theory* (2010), 48 pp., available at `https://hal.inria.fr/file/index/docid/515969/filename/CIMPA.pdf`.

[4] C. Bachoc and F. Vallentin, *New upper bounds for kissing numbers from semidefinite programming*, J. Amer. Math. Soc. **21** (2008), no. 3, 909–924.

[5] A. Barvinok, *A Course in Convexity*, Grad. Stud. Math., vol. 54, Amer. Math. Soc., 2002.

[6] E. van Beveren, *Some notes on group theory* (2002), available at `http://cft.fis.uc.pt/eef/evbgroups.pdf`.

[7] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, *Julia: A fresh approach to numerical computing.* (2014), available at `http://arxiv.org/abs/1411.1607`.

[8] S. Bochner, *Hilbert distances and positive definite functions*, Ann. of Math. (2) **42** (1941), 647–656.

[9] B. Borchers, *CSDP, a C library for semidefinite programming*, Optim. Methods Softw. **11/12** (1999), no. 1-4, 613–623, DOI 10.1080/10556789908805765. Interior point methods.

[10] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), 235–265.

[11] H. Cohn and A. Kumar, *Universally optimal distribution of points on spheres*, J. Amer. Math. Soc. **20** (2007), no. 1, 99–148, DOI 10.1090/S0894-0347-06-00546-7.

[12] H. Cohn, A. Kumar, S. D. Miller, D. Radchenko, and M. S. Viazovska, *The sphere packing problem in dimension* 24 (2016), 12 pp., available at `http://arxiv.org/abs/1603.06518`.

[13] H. Cohn and J. Woo, *Three-point bounds for energy minimization*, J. Amer. Math. Soc. **25** (2012), no. 4, 929–958, DOI 10.1090/S0894-0347-2012-00737-1.

[14] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl. **30** (2008), no. 3, 1254–1279, DOI 10.1137/060661569.

[15] P. E. B. DeCorte, *The eigenvalue method for extremal problems on infinite vertex-transitive graphs*, PhD Thesis, Delft University of Technology, 2015.

[16] P. Delsarte, J. M. Goethals, and J. J. Seidel, *Spherical codes and designs*, Geometriae Dedicata **6** (1977), no. 3, 363–388.

[17] J. Dieudonné, *Sur la séparation des ensembles convexes*, Math. Ann. **163** (1966), 1–3 (French).

[18] L. Föppl, *Stabile Anordnungen von Elektronen im Atom*, J. Reine Angew. Math. **141** (1912), 251–302, DOI 10.1515/crll.1912.141.251 (German).

[19] G. B. Folland, *A course in abstract harmonic analysis*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1995.

[20] L. Fousse, G. Hanrot, V. Lefèvre, P. Pélissier, and P. Zimmermann, *MPFR: A Multiple-precision Binary Floating-point Library with Correct Rounding*, ACM Trans. Math. Softw. **33** (2007), no. 2, DOI 10.1145/1236463.1236468.

[21] K. Fujisawa, M. Fukuda, K. Kobayashi, M. Kojima, K. Nakata, M. Nakata, and M. Yamashita, *SDPA (SemiDefinite Programming Algorithm) Users Manual - Version 7.0.5*, Technical Report B-448, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, 2008.

[22] *GAP – Groups, Algorithms, and Programming, Version 4.8.5*, The GAP Group, 2016.

[23] K. Gatermann and P. A. Parrilo, *Symmetry groups, semidefinite programs, and sums of squares*, J. Pure Appl. Algebra **192** (2004), no. 1-3, 95–128, DOI 10.1016/j.jpaa.2003.12.011.

[24] I. M. Gel′fand, R. A. Minlos, and Z. Ja. Šapiro, *Representations of the rotation group and of the Lorentz group, and their applications (translation)*, Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1958.

[25] V. L. Klee Jr., *Separation properties of convex cones*, Proc. Amer. Math. Soc. **6** (1955), 313–318.

[26] H. Kraft and C. Procesi, *Classical invariant theory: a primer*, 1996.

[27] A. B. J. Kuijlaars, E. B. Saff, and X. Sun, *On separation of minimal Riesz energy points on spheres in Euclidean spaces*, J. Comput. Appl. Math. **199** (2007), no. 1, 172–180, DOI 10.1016/j.cam.2005.04.074.

[28] D. de Laat, *Moment methods in extremal geometry*, PhD Thesis, Delft University of Technology, 2016.

[29] D. de Laat, F. M. de Oliveira Filho, and F. Vallentin, *Upper bounds for packings of spheres of several radii*, Forum Math. Sigma **2** (2014), e23, 42, DOI 10.1017/fms.2014.24.

[30] D. de Laat and F. Vallentin, *A semidefinite programming hierarchy for packing problems in discrete geometry*, Math. Program. **151** (2015), no. 2, Ser. B, 529–553, DOI 10.1007/s10107-014-0843-4.

[31] J. B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM J. Optim. **11** (2000/01), no. 3, 796–817, DOI 10.1137/S1052623400366802.

[32] J. B. Lasserre, *An explicit equivalent positive semidefinite program for nonlinear* 0-1 *programs*, SIAM J. Optim. **12** (2002), no. 3, 756–769, DOI 10.1137/S1052623400380079.

[33] V. I. Levenshtein, *Designs as maximum codes in polynomial metric spaces*, Acta Appl. Math. **29** (1992), 1–82.

[34] H. Whitney, *Congruent Graphs and the Connectivity of Graphs*, Amer. J. Math. **54** (1932), no. 1, 150–168, DOI 10.2307/2371086.
[35] T. W. Melnik, O. Knop, and W. R. Smith, *Extremal arrangements of points and unit charges on a sphere: equilibrium configurations revised*, Can. J. Chem. **55** (1977), 1745–1761.
[36] H. D. Mittelman, *The SDP Problem*, available at `http://plato.asu.edu/ftp/sdpa_format.txt`.
[37] T. Molien, *Uber die Invarianten der linearen Substitutionsgruppe*, Sitzungsber. Knig. Preuss. Akad. Wiss. (1897), 1152–1156.
[38] O. R. Musin, *Multivariate positive definite functions on spheres*, Discrete geometry and algebraic combinatorics, Contemp. Math., vol. 625, Amer. Math. Soc., Providence, RI, 2007, pp. 177–190, DOI 10.1090/conm/625/12498.
[39] J. Nie and M. Schweighofer, *On the complexity of Putinar's Positivstellensatz*, J. Complexity **23** (2007), no. 1, 135–150, DOI 10.1016/j.jco.2006.07.002.
[40] G. Pólya and G. Szegő, *Problems and theorems in analysis. II*, Classics in Mathematics, Springer-Verlag, Berlin, 1998. Theory of functions, zeros, polynomials, determinants, number theory, geometry; Translated from the German by C. E. Billigheimer; Reprint of the 1976 English translation.
[41] V. Powers and B. Reznick, *Polynomials that are positive on an interval*, Trans. Amer. Math. Soc. **352** (2000), no. 10, 4677–4692, DOI 10.1090/S0002-9947-00-02595-2.
[42] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana Univ. Math. J. **42** (1993), no. 3, 969–984, DOI 10.1512/iumj.1993.42.42045.
[43] A. Schrijver, *New code upper bounds from the Terwilliger algebra and semidefinite programming*, IEEE Trans. Inf. Th. **51** (2005), 2859–2866.
[44] R. E. Schwartz, *The five-electron case of Thomson's problem*, Exp. Math. **22** (2013), no. 2, 157–186, DOI 10.1080/10586458.2013.766570.
[45] R. E. Schwartz, *The triangular bi-pyramid minimizes a range of power law potentials*, 57 pages pp., available at `arXiv:1512.04628`.
[46] J.-P. Serre, *Linear representations of finite groups*, Springer-Verlag, New York-Heidelberg, 1977. Translated from the second French edition by Leonard L. Scott; Graduate Texts in Mathematics, Vol. 42.
[47] K. Fujisawa, M. Fukuda, K. Kobayashi, M. Kojima, K. Nakata, M. Nakata, and M. Yamashita, *SDPA (SemiDefinite Programming Algorithm) Users Manual Version 7.0.5*, Technical Report B-448, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, 2008.
[48] A. Tumanov, *Minimal biquadratic energy of 5 particles on 2-sphere*, Indiana Univ. Math Journal **62** (2013), 1717–1731.
[49] M. S. Viazovska, *The sphere packing problem in dimension* 8 (2016), 22 pp., available at `http://arxiv.org/abs/1603.04246`.
[50] F. Vallentin, *Lecture notes: Semidefinite programs and harmonic analysis* (2008), 31.
[51] V. A. Yudin, *Minimum potential energy of a point system of charges*, Diskret. Mat. **4** (1992), no. 2, 115–121, DOI 10.1515/dma.1993.3.1.75 (Russian, with Russian summary); English transl., Discrete Math. Appl. **3** (1993), no. 1, 75–81.