Moment methods in extremal geometry

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op

vrijdag 29 januari 2016 om 12.30 uur

door

David de Laat

Master of Science,
Universiteit Groningen,
geboren te Delft, Nederland

This dissertation has been approved by the promotores:

Prof. dr. F. Vallentin and Prof. dr. ir. K.I. Aardal

Composition of the doctoral committee:

Rector Magnificus,	Chairman
Prof. dr. ir. K.I. Aardal,	Delft University of Technology
Prof. dr. F. Vallentin,	University of Cologne

Independent members:

Prof. dr. H. Cohn,	Microsoft Research and MIT
Prof. dr. E. de Klerk,	Tilburg University and Delft University of Technology
Prof. dr. J.M.A.M. van Neerven,	Delft University of Technology
Prof. dr. A. Schürmann,	University of Rostock
Dr. D.C. Gijswijt,	Delft University of Technology

The research described in this dissertation was financed by Vidi grant 639.032.917
of the Netherlands Organisation for Scientific Research (NWO).

ISBN 978-94-6186-582-3

Cover design by Sinds1961 Grafisch Ontwerp
Printed and bound by Printservice Ede

Contents

CHAPTER 1

Introduction

What is the ground state energy of a system of interacting particles? How do we pack objects together as densely as possible? These are questions of *extremal geometry*. Applications range from the study of error correcting codes, approximation theory, and computational complexity to the modeling of materials in chemistry and physics. In these problems the search space consists of infinitely many configurations among which there can be many suboptimal local optima. This makes it notoriously difficult to certify the optimality of a construction, and for all but the simplest of these problems we do not expect there will ever be purely human generated proofs. We work on methods which allow us to use computers to search for small proofs in the form of optimality certificates. These certificates are given by dual objects which we call obstructions. For the two examples above an obstruction gives an energy lower bound or a density upper bound, and when such a bound is sharp the obstruction provides an optimality certificate. On the one hand we show our methods can find arbitrarily good obstructions in principle. On the other hand we compute new obstructions for concrete geometric problems, where the symmetry of the problems is often of decisive importance.

We give an infinite dimensional generalization of *moment methods* from polynomial optimization. By using infinite dimensional optimization we deal with the infinite set of possible locations of each particle or object, and by using moments we deal with the suboptimal local optima. The theory of moments has a rich history, but here we only describe its use in optimization via the Lasserre hierarchy [**63**]: An example of a moment is a value $y_\alpha = \int x^\alpha \, d\mu(x)$, where μ is a probability measure on a compact set $K \subseteq \mathbb{R}^n$. Here, for $\alpha \in \mathbb{N}_0^n = \{0, 1, 2, \ldots\}^n$, we use the notation $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and $|\alpha| = \sum_{i=1}^n \alpha_i$. The transformation mapping μ to its sequence of moments preserves positivity: $\{y_\alpha\}$ is of positive type; that is, the finite principal submatrices of the infinite matrix $(y_{\alpha+\beta})_{\alpha,\beta\in\mathbb{N}_0^n}$ are positive semidefinite. Consider the problem of finding the minimal value of a polynomial $p = \sum_\alpha p_\alpha x^\alpha \in \mathbb{R}[x_1, \ldots, x_n]$ over the set K, which is equivalent to minimizing $\int p \, d\mu$ over all probability measures supported on K. Upper bounds on the minimum can be obtained by evaluating p at points $x \in K$. For lower bounds we assume K is basic closed semialgebraic: $K = \{x \in \mathbb{R}^n : g(x) \geq 0 \text{ for } g \in \mathcal{G}\}$, where $\mathcal{G}$ is a finite subset of $\mathbb{R}[x_1, \ldots, x_n]$. We select an integer $t \geq \lceil \deg(p)/2 \rceil$ and minimize $\sum_\alpha p_\alpha y_\alpha$ over all sequences $\{y_\alpha\}_{|\alpha|\leq 2t}$ with $y_0 = 1$, where the matrix $(y_{\alpha+\beta})_{|\alpha|,|\beta|\leq t}$ is positive semidefinite, and where some additional moment conditions involving the set $\mathcal{G}$ are satisfied (see Section 2.6). For each t we obtain a relaxed problem whose optimal value is computable through semidefinite programming (see below) and

which lower bounds the minimum. These bounds improve as t gets bigger, and for many interesting classes of problems the bound is sharp for some finite value of t.

Before we discuss our adaptation of the above approach to problems in extremal geometry, we first show how packing problems can be modeled as independent set problems in infinite graphs. Consider the problem of finding a sphere packing of maximal density, where a sphere packing is a set of translates of unit balls in Euclidean space such that the pairwise interiors do not intersect. In 3 dimensions this is the Kepler conjecture which was solved by Hales in 1998 through a computer assisted proof [**44**]. This proof does not use dual certificates as discussed above, and since its large size made it difficult to verify its correctness, a formal, fully computer verified version was finished in 2015 [**45**]. The spherical cap packing (or spherical code) problems are compact analogues of the sphere packing problem. Here we ask for the optimal density of a packing of equally sized spherical caps on a unit sphere. If we take Euclidean space as vertex set and connect two vertices whenever their distance is strictly less than two, or take the unit sphere as vertex set and connect two vertices whenever their inner product is strictly larger than some value corresponding to the cap size, then the independent sets in these graphs (the subsets that do not contain adjacent vertices) correspond precisely to valid packings. For the spherical cap packing problem the independence number (the size of a largest independent set) is finite and proportional to the optimal density.

The independent set problem for finite graphs is one of the main NP-hard problems in combinatorial optimization [**55**]. To find upper bounds we can use the graph parameter known as the Lovász ϑ-number, which was introduced in the celebrated paper [**71**]. This number upper bounds the independence number of a finite graph and is efficiently computable through semidefinite programming. In semidefinite programming we optimize a linear functional over an affine section of the cone of positive semidefinite matrices. Semidefinite programs form a powerful generalization of linear programs but can still be solved efficiently; they form the main computational tool in this thesis. Some important bounds in extremal geometry can be interpreted as analogues of the ϑ-number for infinite graphs. For the spherical code problem there is the Delsarte–Goethals–Seidel linear programming upper bound [**27**], which we can view this as a symmetry reduced (see below) version of a generalization of the ϑ-number to the infinite spherical code graph [**8**]. Similarly, we can view the Cohn–Elkies [**21**] linear programming bound for the sphere packing problem as a symmetry reduced analogue of the ϑ-number for the infinite sphere packing graph; see Section 4.1.1.

We extend the above approach to compute new bounds for packings of spherical caps and spheres of multiple sizes; see Chapter 4. Although we use a semidefinite programming solver – and hence floating point arithmetic – to find these bounds, we obtain proofs through a rounding procedure, where we round the solutions to matrices containing rational or algebraic numbers which satisfy all the constraints of the semidefinite programs. For instance, the binary sphere packing with the structure of sodium chloride has density approximately 79.3%, and we prove an upper bound of approximately 81.3%. We give an example of a binary spherical cap packing where our bound is sharp, which leads to a simple optimality proof. We

also give the best known bounds for the classical sphere packing problem (where all spheres are congruent) in dimensions 4 to 7 and 9 by giving a slight improvement on the Cohn–Elkies bound. This leads to the following question which we will discuss now: can we obtain arbitrarily good bounds?

A relaxation of an optimization problem $\inf_{x\in X} f(x)$ is another, typically easier, problem $\inf_{y\in Y} g(y)$ together with a map $R\colon X \hookrightarrow Y$ such that $g \circ R \leq f$. The moment bounds defined above are examples of relaxations and so is the Lovász ϑ-number. We call the above bounds for packing problems 2-point bounds because these are relaxations where we replace optimization over all geometric configurations by optimization over computationally tractable information on the pair distribution of configurations. To obtain better bounds we can consider relaxations which use information on triples, and Schrijver [**89**] found an approach to compute 3-point bounds for binary codes. This approach was put in a representation theoretic framework and extended to the spherical code problem and by Bachoc and Vallentin [**9**]. An extension to energy minimization was given by Cohn and Woo [**23**]. These techniques led to many new optimality proofs and for many problems these bounds still give the best available results. An extension to k-point bounds is considered in [**76**], but when applied to the sphere $S^{n-1} \subseteq \mathbb{R}^n$ it cannot go beyond n-point bounds. For a new, but related, approach to obtaining relaxations for these problems we continue our discussion of moment methods.

The independent set problem for a finite graph $G = ([n], E)$ can be stated as a polynomial optimization problem where we maximize the objective function $\sum_{i=1}^n x_i$ over all $x \in \mathbb{R}^n$ which satisfy the constraints $x_i(1 - x_i) = 0$ for $i \in [n]$ to enforce 0/1 valued variables, and $1 - x_i - x_j \geq 0$ for $\{i, j\} \in E$ to enforce the edge conditions. By applying the moment techniques discussed above we obtain a hierarchy of optimization problems whose optimal values give increasingly good upper bounds on the independence number $\alpha(G)$. In [**65**] Laurent showed this hierarchy converges to the independence number in $t = \alpha(G)$. In Chapter 6 we generalize this approach to infinite graphs. For this we define topological packing graphs as an abstraction for the infinite graphs coming from packing problems. We use functional analytic tools to give a definition of the moments of measures defined on sets of geometric configurations. Now, instead of a moment sequence, the moments form a measure defined on the set of independent sets. We obtain relaxations by optimizing over measures defined on the independent sets up to cardinality $2t$. This gives a sequence of infinite dimensional maximization problems whose optimal values give increasingly good upper bounds on the optimal density. We prove this sequence converges to the optimal packing density. We also show the first step of this hierarchy is equivalent to a generalization of the ϑ-number to topological packing graphs, which shows the first step equals well-known bounds for packing problems.

To go from relaxations to obstructions we use what is arguably the most beautiful topic in optimization: duality. Given a maximization problem, which we call the primal, there exist dual minimization problems whose optimal values upper bound the primal's optimal value. The obstructions mentioned in the first paragraph of

this introduction are given by feasible solutions to the duals of the relaxations discussed above. In our primal optimization problems we optimize over measures, which naturally means the dual variables are continuous functions. In general, the primal and dual optimal values are not equal but there can be a strictly positive duality gap. Using a closed cone condition and convex geometric arguments we prove that for each step in our hierarchy there is no duality gap. Together with the convergence result mentioned in the previous paragraph, this shows we can obtain arbitrarily good bounds on the optimal density by finding good feasible solutions to these dual programs.

When an optimization problem admits symmetry, then the relaxations and their duals typically inherit this symmetry. The symmetry is expressed by a group action on the space of variables for which the constraints and objective are invariant. If such an optimization problem is convex (and if the group is compact), then we can restrict to invariant variables, which can simplify the problem significantly. The 2 and 3-point bounds for the spherical code problem are good examples where this symmetry can be used. Here the variables are continuous, positive definite kernels $K\colon S^2 \times S^2 \to \mathbb{R}$, which for 2-point bounds can be assumed to be invariant under the orthogonal group $O(3)$, and for 3-point bounds under the stabilizer subgroup with respect to a point $e \in S^2$. In the dual problems of our hierarchy, the variables are continuous, positive definite kernels $K\colon I_t \times I_t \to \mathbb{R}$, where I_t is the set of independent sets in the packing graph that have size at most t. These kernels can be assumed to be invariant under the symmetry group of the graph.

To exploit the symmetry we use harmonic analysis. The main idea is to reduce to a finite dimensional variable space by optimizing over truncated Fourier series of the kernel K. Since the Fourier coefficients of a positive definite kernel are positive semidefinite, this results in approximating optimization problems where we optimize over positive semidefinite matrices. We can view this as a block diagonalization, and the bigger the group action the smaller the blocks. In the case of 2-point bounds for the spherical code problem, these blocks are of size 1×1, and the problem reduces to an infinite dimensional linear program. We consider theoretical issues, such a the existence of a Fourier basis for the kernels and convergence of these approximations, as well as more practical issues such as how to explicitly construct the Fourier basis for the spaces I_t by using tensor representations. We show that for the case where the vertex set is a sphere, the programs in our dual hierarchy can be approximated in this way by a sequence of semidefinite programs with polynomial constraints.

We expect that the class of semidefinite programs with polynomial constraints will become increasingly important. Here, by a polynomial constraint we mean the requirement that a polynomial, whose coefficients depend linearly on the entries of the positive semidefinite matrix variable(s), is positive on a basic closed semialgebraic set. This includes the problem of finding the minimum of a polynomial as discussed above, but instead of considering the moments we now take the dual sum of squares viewpoint. A sum of squares polynomial is nonnegative, and in real algebraic geometry we study when and how a polynomial that is nonnegative (or strictly positive) on a set can be represented using sum of squares. This is useful from a computational perspective because the cone of sum of squares polynomials

of fixed degree is isomorphic to a cone of positive semidefinite matrices. Using these techniques a semidefinite program with polynomial constraints can be approximated by a sequence of semidefinite programs. In applying this there are three important points to consider: numerical conditioning, symmetry, and sparsity. In Chapter 4 we consider the first point, where we show the correct choice of bases is essential to be able to solve the resulting semidefinite programs. In Section 7.8 we show how symmetry in the polynomial constraints can be exploited to get block diagonalized sum of squares characterizations, and we show how much we gain from this when applied to our hierarchy. It is an open question whether sparse sum of squares characterization as for instance discussed in [**58**] yield significant computational savings for the type of problems considered in this thesis.

We use our generalized moment techniques to construct a converging hierarchy for approximating the ground state energy of a system of N interacting particles. An important example is the Thomson problem, where we minimize the pairwise sum of $\|x_i - x_j\|_2^{-1}$ over all sets $\{x_1, \dots, x_N\}$ of N distinct elements in the unit sphere $S^2 \subseteq \mathbb{R}^3$. We show the N-th step E_N in this hierarchy is guaranteed to give the optimal energy E. It could be, however, that for many problems the bound E_t is sharp for much smaller t. After symmetry reduction, the dual of the first step E_1 essentially reduces to Yudin's bound [**100**], which is an adaptation of the Delsarte–Goethals–Seidel bound mentioned above for energy minimization. This means E_1 is sharp for the Thomson problem with $N = 2, 3, 4, 6, 12$. It would be very interesting if this pattern continues; that is, if the second step E_2 is sharp for several new values of N. As a first step into investigating this – and to show that it is possible to compute the second step of the hierarchy – we compute E_2 numerically for $N = 5$, where the computational results suggest this bound is sharp. This is the first time a 4-point bound has been computed for a continuous problem.

The 5 particle case is especially interesting as this is one of the simplest mathematical models of a phase transition. By this we mean there is a discontinuous jump from one globally optimal solution to another as the pair potential changes only slightly. We compute the bound for the Riesz s-energy potentials for $s = 2, 4$, where the numerical results again suggest the bound is sharp. It would be very interesting if E_2 is universally sharp for 5 particles, by which we mean it is sharp for a large class of pair potentials and hence also throughout the phase transition.

***

This thesis consists of seven chapters including this introductory chapter. Chapters 2 and 3 mainly contain background material (the former on optimization and the latter on harmonic analysis) and chapters 4 to 7 are based on papers and contain their own introductions.

CHAPTER 2

A ten page introduction to conic programming

This background chapter gives an introduction to conic programming. We do not give proofs, but focus on important (for this thesis) tools and concepts.

2.1. Optimization and computational hardness

Optimization is about maximizing or minimizing a function over a set. The set is typically described more implicitly than just an enumeration of its elements, and the structure in this description is essential in developing good optimization techniques. The set is known as the *feasible set* and its elements the *feasible solutions*. The function is called the *objective function* and its range the *objective values*. We write a minimization problem as $p = \inf_{x \in S} f(x)$, and we often use p to refer to the optimization problem as a whole instead of just its optimal value. We are not only interested in finding the optimal value, but also in finding optimal solutions, and if this is too difficult (or if they do not exist) we seek close to optimal feasible solutions. An important topic is finding certificates asserting the solution's optimality or quality of approximation. In fact, by solving an optimization problem we often mean finding an optimal solution together with a certificate.

Linear programming is foundational in conic optimization. Consider the problem of finding a vector x satisfying a linear system $Ax = b$. We can find such an x by Gaussian elimination, but when we also require the entries of x to be nonnegative, then we need different algorithms. In a linear program we optimize a linear functional over all nonnegative vectors satisfying a given linear system. It is, however, the positivity condition, and not the fact that we are optimizing a functional, that moves a linear problem into the field of optimization: Using complementary slackness (see Section 2.3) we can add variables and constraints to a linear program so that all its feasible solutions are optimal. Alternatively, we can constrain a minimization problem's objective value to be at most some number b, and then bisect on b to solve the optimization problem by solving a number of feasibility problems.

When we discuss the hardness (in some computational model) of solving or approximating a class of optimization problems, we need to define an explicit encoding of the feasible sets and objective functions. In this way it is clear what constitutes the input data for the algorithms. This is important because the efficiency of an algorithm is determined by the dependence of the running time on the input size. Geometrically, linear programming is the optimization of a linear functional over a polyhedron, and although a polyhedron can be described in different ways, when we discuss computational hardness we assume a facial description. This means the polyhedron is given by all vectors x satisfying some linear inequality $Ax \geq b$. We

can, however, use any description that is easy to transform into and derive from this one, such as the description from the previous paragraph. Linear programs can be solved efficiently in practice by simplex methods, although it is not known whether there exists a simplex method that runs in polynomial time. The ellipsoid method can solve a rational linear program in polynomial time (in the bit model) but appears to be too slow in practice. In Section 2.4 we discuss interior point methods which are fast in practice and that can be made to run in polynomial time.

If a linear program's input is rational, then its optimal value is a rational number whose bit size is bounded by a fixed polynomial in the input bit size [**87**]. For semidefinite programming, which is a powerful generalization of linear programming, there exist rational instances whose optimal values are algebraic numbers of high degree [**77**], and it is not known whether a polynomial time algorithm for semidefinite programing exists. However, if the feasible set of a semidefinite program contains a ball of radius r and is contained in a ball of radius R, then for each $\varepsilon > 0$ we can find an ε-optimal solution in polynomial time (where ε, r, and R are part of the input of the algorithm); see also Section 2.4.

We distinguish between convex and nonconvex optimization problems, where a convex optimization problem has a convex feasible set and convex (concave) objective function in case it is a minimization (maximization) problem. Convex problems have the advantage that local optima are globally optimal, but this does not mean they are necessarily easy to solve.

2.2. Lifts and relaxations

When optimization problems are difficult, we can try to use their description to derive easier optimization problems which give information about the original problems. Lifts provide one such technique. A *lift* of an optimization problem is another optimization problem with a surjective map P from its feasible set onto the original problem's feasible set, and whose objective function is given by composing the original objective function with P. This technique originated from the observation that there exist polytopes which are projections of higher dimensional polytopes with drastically simpler facial structure. Lifts contain all information of the original problems; they have the same optimal value and we can project their optimal solutions to optimal solutions of the original problem.

Typically we do not lift a single problem, but we systematically lift an entire class of problems. When the worst case instances in this class are inherently difficult to solve – for instance, the class is NP-hard and P $\neq$ NP – then it appears to be difficult for lifts to recognize the easy problems; that is, all of them will be hard to solve. More successful in this respect are relaxations. A *relaxation* of a problem $\inf_{x\in A} f(x)$ is another problem $\inf_{x\in B} g(x)$ together with an injective map $R\colon A \hookrightarrow B$ such that $g\circ R \leq f$. Relaxations are often obtained by relaxing the constraint set, in which case R is the identity. For example, by removing the integrality constraints in an integer linear program we obtain the linear programming relaxation. Also common are Lagrangian relaxations which we discuss in Section 2.3.

A lift of a relaxation is a relaxation, and we will encounter instances which are naturally interpreted in this way. When R is surjective and $g\circ R = f$, the relaxation

is a lift by taking $P = R^{-1}$. Even when R is not surjective, it can happen that it maps optimal solutions to optimal solutions having the same objective value, and in this case we say the relaxation is *sharp*. Any optimization problem $\inf_{x \in S} f(x)$ admits a sharp convex relaxation $\inf_{x \in C} g(x)$ by taking C to be the convex hull of the basis elements δ_x of the vector space $\mathbb{R}^S$ of finitely supported functions and g to be the linear functional satisfying $g(\delta_x) = f(x)$ for $x \in S$.

2.3. Conic programming and duality

In a *conic program* we optimize a linear functional over the intersection of a closed convex cone with an affine space. A convex cone K is a nonempty subset of a real vector space E such that $ax + by \in K$ for all $a, b \geq 0$ and $x, y \in K$. We define the affine space by the set of solutions to the equation $Ax = b$, where A is a linear operator from E to another real vector space F, and b is an element from F. The objective function is a linear functional $c\colon E \to \mathbb{R}$. A conic program is an optimization problem in the form

$$p = \inf \{c(x) : x \in K,\ Ax = b\}.$$

Any convex optimization problem $\inf_{x \in S} f(x)$ can be written as a conic program: First write it as a minimization problem with linear objective $(x, b) \mapsto b$ and convex feasible set $C = \{(x, b) \in S \times \mathbb{R} : f(x) \leq b\}$, then write it as a conic program over the cone $\{(x, t) : t \geq 0,\ x \in tC\}$. The power of conic programming, however, lies in the fact that we only need a few classes of convex cones to express a wide variety of optimization problems. The type of optimization problem is encoded by the cone, and the problem data is given by the affine space and objective function. Linear programs are conic programs over a nonnegative orthant cone $\mathbb{R}^n_{\geq 0}$, and semidefinite programs use a cone of positive semidefinite matrices.

Positivity — as modeled by the cone constraints in a conic program — is fundamental in convex optimization. A second fundamental concept is duality. We first discuss Lagrangian duality, which is based on removing constraints and penalizing violations of those constraints in the objective. Consider a problem of the form

$$q = \inf \Big\{ f(x) : x \in S,\ g_i(x) = 0 \text{ for } i \in [l],\ h_j(x) \geq 0 \text{ for } j \in [m] \Big\},$$

where $[l] = \{1, \ldots, l\}$. We call this the primal problem. For simplicity we assume all functions to be real-valued and continuously differentiable, and we assume S to be an open subset of $\mathbb{R}^n$. We define the *Lagrangian* by

$$L\colon S \times \mathbb{R}^l \times \mathbb{R}^m_{\leq 0} \to \mathbb{R},\ (x, u, v) \mapsto f(x) + \sum_{i=1}^{l} u_i g_i(x) + \sum_{j=1}^{m} v_j h_j(x).$$

When $m = 0$, the constrained stationary points of f correspond precisely to the stationary points of L. The geometric explanation is that $\nabla_u L = 0$ forces x to be feasible, and $\nabla_x L = 0$ forces the direction of steepest descent of f at x to be a normal vector of the feasible set. The entries of the vector u in a stationary point (x, u) of L are called *Lagrange multipliers*. In the general case where $m > 0$ the situation is more subtle. The constrained stationary points of L are known as *Karush-Kuhn-Tucker points*. For each such point (x, u, v), the vector x is a

constrained stationary point of f. In general not all constrained stationary points of f can be obtained in this way, but there are sufficient conditions known as global constraint qualifications under which this is true. The most well-known is Slater's condition which requires the problem to be convex and to admit a strictly (all inequalities are strictly satisfied) feasible point. When the functions $f, g_1, \ldots, g_m$ are convex, the set S is convex, and the functions $h_1, \ldots, h_m$ are linear, then the problem is convex. In convex problems the global constrained minima are precisely the constrained stationary points.

We define the *Lagrangian dual function*

$$R\colon \mathbb{R}^l \times \mathbb{R}^m_{\leq 0} \to \mathbb{R},\ R(u,v) = \inf_{x\in S} L(x,u,v),$$

so that for each u and each $v \leq 0$, the problem $R(u,v)$ is a relaxation of q. The *Lagrangian dual problem* is given by maximizing this function over its domain:

$$q^* = \sup_{(u,v)\in\mathbb{R}^l\times\mathbb{R}^m_{\leq 0}} R(u,v).$$

The primal problem can be written as

$$\inf_{x\in S} \sup_{(u,v)\in\mathbb{R}^l\times\mathbb{R}^m_{\leq 0}} L(x,u,v),$$

so that we simply interchange sup and inf to go from the primal to the dual problem. A global constraint qualification such as Slater's condition guarantees the optima of the primal and dual are the same.

To apply Lagrangian duality to general conic programs we extend the above discussion to conic constraints. In q^* the objective function is an optimization problem itself, and the reduction to a more explicit form requires problem specific information. An advantage of conic programming is that all nonlinearities are contained in the cone constraint, and an explicit description of the dual cone is all we need for an explicit description of the dual program. The dual program is a conic program over the dual cone, and the situation is symmetric in the sense that we recover the original problem by taking the dual again.

Let E^* and F^* be the algebraic duals of E and F; that is, the vector spaces of real-valued linear functionals on E and F. Then $c \in E^*$. We have two nondegenerate bilinear pairings $E \times E^* \to \mathbb{R}$ and $F \times F^* \to \mathbb{R}$, each denoted and defined by $\langle x, y\rangle = y(x)$. The dual cone K^* is defined by $\{y \in E^* : \langle x, y\rangle \geq 0 \text{ for } x \in K\}$. The adjoint operator $A^*\colon F^* \to E^*$ is defined by $A^*f = f \circ A$ for all $f \in F^*$. The Lagrangian of the conic program p is naturally given by

$$L\colon K^* \times E \to \mathbb{R},\ (y,x) \mapsto c(x) - \langle x, y\rangle,$$

so that the Lagrangian dual program becomes

$$p^* = \sup\{\langle b, y\rangle : y \in F^*,\ c - A^*y \in K^*\}.$$

To reconstruct the primal from the dual we write the dual as a conic program in standard form, take the dual, and write this in standard form. The symmetry here becomes more apparent when we write both programs in a more geometric form. For e an element such that $Ae = b$ and $P = \ker(A)$, the primal and dual become

$$\inf\{\langle x, c\rangle : x \in (e + P) \cap K\} \quad \text{and} \quad \sup\{\langle y, e\rangle : y \in P^\perp \cap (K^* + c)\},$$

so that both programs optimize a linear functional over the intersection of a (translated) cone with an (affine) linear subspace.

When the vector spaces E and F are infinite dimensional, their algebraic duals are so large that the algebraic dual conic programs have too many variables and constraints to be useful. Instead we endow E and F with topologies and restrict E^* and F^* to contain only *continuous* linear functionals. We require these topologies to agree with the data by requiring c and A to be continuous, so that c is in E^* and the adjoint A^* maps E^* into F^*. We also require these topologies to be Hausdorff and locally convex so that there are — by the Hahn–Banach theorem — enough continuous linear functionals to separate points. This insures nondegeneracy of the bilinear forms, so that (E, E^*) and (F, F^*) are *dual pairs*. We form the dual cone and the dual conic program in the same way as before, and if we equip E and F with very strong topologies, such as the topologies of algebraically open sets, then we get the same duals as in the algebraic case. To keep the situation symmetric we equip E^* and F^* with weak* topologies; that is, we give them the weakest topologies for which all linear functionals $x \mapsto \langle x, y\rangle$ are continuous. Using nondegeneracy of the pairings we see that $(E^*)^*$ and $(F^*)^*$ are isomorphic to E and F, and by identifying them we obtain $(A^*)^* = A$, $(K^*)^* = K$, and $(p^*)^* = p$.

Suppose x is feasible for p and y is feasible for p^*. We always have $p \geq p^*$, which we call *weak duality* and which follows from $\langle x, c\rangle \geq \langle x, A^*y\rangle = \langle Ax, y\rangle = \langle b, y\rangle$. We also have *complementary slackness*, which says $\langle x, c - A^*y\rangle = 0$ if and only if both x and y are optimal and have the same objective value. There can be a strictly positive *duality gap* $p - p^*$, and we say *strong duality* holds when this gap is 0. Like for the constraint qualifications in Lagrangian duality, we have sufficient conditions for strong duality. To Slater's condition corresponds the following interior point condition: If the interior of K admits a primal feasible point and the primal problem is bounded, then $p = p^*$, and the supremum in the dual is attained.

In infinite dimensional spaces there are many interesting cones whose interiors are empty, which means we cannot use an interior point condition. We have the following alternative closed cone condition: If the cone $\{(Ax, \langle x, c\rangle) : x \in K\}$ is closed in $F \times \mathbb{R}$ and there is a primal feasible solution, then $p = p^*$, and if in addition the primal is bounded, then the infimum in the primal is attained [**10**]. Choosing stronger topologies on E and F makes it easier for strong duality to hold: K will have more interior points and $F \times \mathbb{R}$ more closed sets. But the duality gap cannot always be closed by choosing a stronger topology; even finite dimensional problems such as semidefinite programs can have a strictly positive duality gap. Notice that the interior point condition benefits from E having a stronger topology, while the closed cone condition benefits from F having a stronger topology (and indirectly by E having a stronger topology to keep A continuous). The crucial ingredient in the proofs of these conditions is the Hahn–Banach separation theorem. This theorem says that if we have a point and a closed convex set, then either the point lies in the set or it can be strictly separated from it by two parallel hyperplanes in between the set and the point. This resembles the situation that given strong duality, a number λ either is an upper bound on the optimal objective of a minimization problem, or

there is a dual feasible solution whose objective is in between λ and the optimal objective.

2.4. Semidefinite programming and interior point methods

The *positive semidefinite cone* $\mathcal{S}^n_{\succeq 0}$ consists of the positive semidefinite matrices of size $n \times n$. A *positive semidefinite matrix* is a real symmetric matrix whose eigenvalues are nonnegative, or equivalently, a matrix that can be written as RR^T where R is a real rectangular matrix. Such an R can be found efficiently by performing a Cholesky factorization, which moreover gives R in lower triangular form which is useful for solving a system of the form $RR^\mathsf{T}x = b$. The positive semidefinite cones are convex and for $n \geq 2$ they are not polyhedral; the extreme rays are spanned by the rank one matrices xx^T, where $x \in \mathbb{R}^n$. The positive semidefinite cones are self dual, where the dual pairings, denoted by $\langle \cdot, \cdot \rangle$, are defined by taking the trace of the matrix product. Here we view $\mathcal{S}^n_{\succeq 0}$ as a subset of the $n(n+1)/2$ dimensional vector space $\mathcal{S}_n$ of $n \times n$ real symmetric matrices. The interior of $\mathcal{S}^n_{\succeq 0}$ is the cone $S^n_{\succ 0}$ of positive definite matrices, which are real symmetric matrices with strictly positive eigenvalues. The cones $\mathcal{S}^n_{\succeq 0}$ and $\mathcal{S}^n_{\succ 0}$ induce partial orders, denoted $\succeq$ and $\succ$, on the vector space $\mathcal{S}^n$. The *Schur complement condition* says that if A, B, and C are matrices with A invertible, then $\left(\begin{smallmatrix} A & B \\ B^\mathsf{T} & C \end{smallmatrix}\right) \succeq 0$ if and only if $A \succ 0$ and $C - B^\mathsf{T}A^{-1}B \succeq 0$.

A semidefinite program is a conic program over a cone of positive semidefinite matrices. We can write such a program as

$$p = \inf \Big\{ \langle X, C \rangle : X \in \mathcal{S}^n_{\succeq 0},\ \langle X, A_i \rangle = b_i \text{ for } i \in [m] \Big\},$$

where $C, A_1, \ldots, A_m \in \mathcal{S}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$. The dual program is given by

$$p^* = \sup \Big\{ \langle b, y \rangle : y \in \mathbb{R}^m,\ C - \sum_{i=1}^m y_i A_i \in \mathcal{S}^n_{\succeq 0} \Big\}.$$

Checking whether a matrix X is positive semidefinite is easy, and when a matrix is not positive semidefinite, we can easily certify this with a positive semidefinite matrix C for which $\langle X, C \rangle < 0$. So, under the condition that the feasible set is contained in a ball and contains a ball, the ellipsoid method can efficiently solve a semidefinite program. This works as follows: First we only consider the feasibiliy problem since binary search allows us to solve the optimization problem by solving a sequence of such problems [**43**]. We start with the large ball. If its center is not feasible, we can separate it from the feasible set by a halfspace. Then we select a smaller ellipsoid containing the intersection of the current ellipsoid with the halfspace and iterate this process. This yields a polynomial time algorithm.

For an approach that is also fast in practice we use interior point methods, in which we reduce a problem to a sequence of stationary point finding problems which we solve using *Newton's method*. This is an iterative method to find roots of (multivariate) vector functions and stationary points of (multivariate) scalar functions. Given a continuously differentiable function $g \colon \mathbb{R}^n \to \mathbb{R}^n$ and a point close enough to a root r, Newton's method generates a sequence of points rapidly converging to r by applying successive Newton steps. A *Newton step* moves a point to the

root of the linear approximation of g at that point. To find a stationary point of a twice continuously differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$ we apply the above method to the gradient ∇f. In this case a Newton step maps a point x to the stationary point of the second order Taylor approximation of f at x; that is, it maps x to $x - (Hf(x))^{-1}\nabla f(x)$, where Hf is the Hessian. If the domain of f is an affine space in $\mathbb{R}^n$, then we use Lagrange multipliers to optimize the Taylor approximation subject to linear equality constraints. In our applications the linear systems to be solved to determine the Newton steps will have a positive definite matrix so that we can use a Cholesky factorization. Although a Cholesky factorization can be computed efficiently, this is a relatively expensive step in interior point methods, so we typically only perform a single Newton step when we invoke Newton's method.

The function $\beta\colon S^n_{\succ 0} \to \mathbb{R}$ defined by $\beta(X) = -\log(\det(X))$ is strongly convex and grows to infinity as X nears the boundary of the cone. This is an example of a *barrier functional*, which lies at the heart of any interior point method. We use this to define the primal and dual central paths $\{X_\eta\}_{\eta\geq 0}$ and $\{(y_\eta, Z_\eta)\}_{\eta\geq 0}$, where X_η and (y_η, Z_η) are the unique optimal solutions to the *barrier problems*

$$p_\eta = \min\left\{\langle X, C\rangle + \eta\beta(X) : X \in S^n_{\succ 0},\ \langle X, A_i\rangle = b_i \text{ for } i \in [m]\right\}$$

and

$$p^*_\eta = \max\left\{\langle b, y\rangle - \eta\beta(Z) : y \in \mathbb{R}^m,\ Z \in S^n_{\succ 0},\ Z = C - \sum_{i=1}^m y_i A_i\right\}.$$

To guarantee the existence and uniqueness of optimal solutions we assume strict feasibility of p and p^* and linear independence of the matrices A_i. The central paths converge to optimal solutions of p and p^* as η tends to 0.

In the (short-step) *primal barrier method* we first solve an auxiliary problem to find a primal feasible solution X close to the primal central path; that is, close to X_η for some η. Then we iteratively decrease η and apply a constrained Newton step to X for the function $\langle X, C\rangle + \eta\beta(X)$ and the constraints $\langle X, A_i\rangle = b_i$ for $i \in [m]$. If we decrease η slowly enough, this results in a sequence of matrices which lie close to the central path and for which it is guaranteed that they are positive definite. As $\eta \to 0$ they converge towards the optimal solution $\lim_{\eta\downarrow 0} X_\eta$, and by choosing the right parameters this algorithm finds, for each $\varepsilon > 0$, an ε-optimal solution in polynomial time.

In *primal-dual methods* we maintain both primal and dual iterates which are allowed to violate the affine constraints. To find new iterates we use both primal and dual information, and this results in excellent performance in practice. The main observation is that the Lagrangian

$$L_\eta\colon \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n_{\succ 0} \to \mathbb{R},\ (X, y, Z) \mapsto \langle b, y\rangle - \eta\beta(Z) + \langle C - \sum_{i=1}^m y_i A_i - Z, X\rangle$$

of p^*_η has (X_η, y_η, Z_η) as unique stationary point. The stationarity condition

$$0 = \nabla_Z L_\eta(X_\eta, y_\eta, Z_\eta) = -\eta Z_\eta^{-1} + X_\eta$$

can be written as $X_\eta Z_\eta = \eta I$ so that $\eta = \langle X_\eta, Z_\eta\rangle/n$. Since $\langle X_\eta, Z_\eta\rangle$ is the duality gap, this tells us how fast the primal and dual central paths converge to optimality

as $\eta \downarrow 0$. Moreover, this formula allows us to compute an η value for iterates which do not lie on the central paths or are not even feasible.

The basic idea in primal-dual methods is to start with arbitrary positive definite matrices X and Z and corresponding vector y. Then we iteratively set η to the current η value of X and Z, multiply this by a factor between 0 and 1, and perform a Newton step for the function L_η to get new iterates X and Z. This Newton step is not necessarily positive definite, so instead of jumping to the Newton iterate we move into the direction of this iterate by for instance performing a line search.

In the above method we take an optimizing Newton step for L_η, which is the same as taking a root finding Newton step for ∇L_η. In practice, we often use variations that are obtained by first rewriting the equation $\nabla_Z L_\eta(X, y, Z) = 0$ as, for instance, $ZX - \eta I = 0$. In this variation we have to symmetrize the Z matrix after each Newton step because the product ZX of two symmetric matrices is not necessarily symmetric, so we have to apply Newton's root finding method to maps whose domain and codomain is $\mathcal{S}^n \times \mathbb{R}^m \times \mathbb{R}^{n\times n}$ instead of $\mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$. This reformulation of the nonlinear gradient condition is used in the CSDP solver, which uses a predictor-corrector variant of the above algorithm [**16**].

These interior point methods can be generalized to methods for *symmetric cones*, which have been classified as being products of Lorentz cones, real, complex, and quaternionic positive semidefinite cones, and one exceptional cone. Semidefinite programming is the main case in the sense that a conic program over a product of cones from these families can easily be transformed into a semidefinite program: A conic program over a product of positive semidefinite cones is a semidefinite program by taking direct sums of the data matrices with zero blocks at appropriate places. This also shows linear programming is a special case of semidefinite programming. A second order cone program transforms into a semidefinite program using a Schur complement. The complex plane embeds into the algebra of real antisymmetric 2×2 matrices by mapping $x + iy$ to the matrix $\left(\begin{smallmatrix} x & y \\ -y & x \end{smallmatrix}\right)$. To transform a complex semidefinite program into a semidefinite program we simply replace each entry in the data matrices by such a block. For the quaternionic case we do the same using an embedding of the quaternions in the algebra of real antisymmetric 4×4 matrices. Of course, the complexity of solving a resulting semidefinite program can be higher than the original problem, and for especially linear and second order cone programming we use specialized solvers. Moreover, semidefinite programming solvers typically work with products of semidefinite cones; that is, they exploit the block structure in semidefinite programs.

2.5. Symmetry in semidefinite programming

A problem $p = \inf_{x\in S} f(x)$ can contain symmetry if the underlying data has symmetry or if the modeling method introduces symmetry. Exploiting this symmetry can reduce the problem size significantly and can remove problematic degeneracies. Given a group Γ with an action on S, we say p is Γ-invariant if f is Γ-invariant. If S is a closed convex set in a locally convex topological vector space, f is a continuous linear functional, and Γ is a compact group with a continuous

action on S, then we can use the symmetry to derive a simpler optimization problem. For this we let μ be the normalized Haar measure of Γ and notice that for each $x \in S$ the group average $\bar{x} = \int \gamma x \, d\mu(\gamma)$, defined through a weak vector valued integral, also lies in S, is invariant under the action of Γ, and satisfies $f(\bar{x}) = f(x)$. We obtain a simpler optimization problem $p^\Gamma = \inf_{x \in S^\Gamma} f(x)$, where S^Γ is the set of Γ-invariant vectors in S. Convexity is essential here: A nonconvex symmetric optimization problem does not necessarily admit symmetric optimal solutions.

Given a unitary representation ρ of a finite group Γ on $\mathbb{C}^n$; that is, a group homomorphism $\rho\colon \Gamma \to U(\mathbb{C}^n)$, we get an action of Γ on the space of Hermitian $n \times n$-matrices by $\gamma X = \rho(\gamma)^* X \rho(\gamma)$. This action is eigenvalue preserving, so it preserves positive semidefiniteness, and a complex semidefinite program p is invariant whenever its objective and affine space are invariant. We obtain p^Γ by restricting to the cone of Γ-invariant, complex, positive semidefinite matrices.

There are several related ways to simplify the program p^Γ. The matrix $*$-algebra $(\mathbb{C}^{n\times n})^\Gamma$ is $*$-isomorphic to a direct sum $\oplus_{i=1}^d \mathbb{C}^{m_i \times m_i}$ [**7**], and since $*$-isomorphisms between unital $*$-algebras preserve eigenvalues, this provides a block diagonalization of p^Γ as a conic program over a product of smaller complex positive semidefinite cones. Another viewpoint, where we use the representation more explicitly, is that invariant matrices X commute with ρ: for each $\gamma \in \Gamma$ we have $\rho(\gamma)^* X = X\rho(\gamma)$. Schur's lemma [**33**] provides a coordinate transform $T\colon \mathbb{C}^n \to \mathbb{C}^n$ such that T^*XT has identical block structure for all $X \in (\mathbb{C}^{n\times n})^\Gamma$. This is a block diagonal structure with d diagonal blocks where the ith block is again block diagonal and consists of identical blocks of size m_i. Applying this transformation and removing redundant blocks yields the same block diagonalization as above. Here d is the number of inequivalent irreducible subrepresentations of ρ and m_i is the number of equivalent copies of the ith of these representations. A third approach applies when ρ maps into the set of permutation matrices. Then we view an invariant matrix as an invariant kernel $[n] \times [n] \to \mathbb{C}$ and apply Bochner's theorem to obtain a diagonalization with the kernel's Fourier coefficients as blocks; see Chapter 3.

2.6. Moment hierarchies in polynomial optimization

When constructing relaxations we need to find a balance between their complexity and the quality of the bounds they gives. For an in general NP-hard optimization problem of the form

$$p = \inf_{x \in S} f(x), \quad S = \{x \in \mathbb{R}^n : g(x) \geq 0 \text{ for } g \in \mathcal{G}\},$$

where $\{f\} \cup \mathcal{G}$ is a finite set of polynomials, we use moment techniques to define a hierarchy of semidefinite programs which give increasingly good bounds. The program p admits the sharp relaxation $\inf_{\mu \in \mathcal{P}(S)} \mu(f)$, where $\mathcal{P}(S)$ is the set of probability measures on S. Let $y_\alpha = \int x^\alpha \, d\mu(x)$, where $\alpha \in \mathbb{N}_0^n$ and $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The moment sequence $\{y_\alpha\}_{\alpha \in \mathbb{N}_0^n}$ satisfies $y_0 = 1$ and is of *positive type*. This means the infinite *moment matrix* $M(y)$, defined by $M(y)_{\alpha,\beta} = y_{\alpha+\beta}$, is positive semidefinite (all its finite principal submatrices are positive semidefinite). Moreover, the *localizing matrices* $M(y * g)$, where $y * g$ is the convolution $(y * g)_\alpha = \sum_\gamma y_{\alpha+\gamma} g_\gamma$, are positive semidefinite.

We obtain a relaxation by optimizing over truncated moment sequences y that satisfy only finitely many of these constraints. Let $M_t(y)$ be the submatrix of $M(y)$ whose entries are indexed by (α, β) with $|\alpha|, |\beta| \leq t$. Let $M_t^g(y)$ be the partial matrix whose entries are indexed by (α, β) with $|\alpha|, |\beta| \leq t$ where the (α, β) entry is given by $(y * g)_{\alpha+\beta}$ if $|\alpha + \beta| \leq 2t - \deg(g)$ and remains unspecified otherwise. By $M_t^g(y) \succeq 0$ we mean that $M_t^g(y)$ can be completed to a positive semidefinite matrix.

For $t \geq \lceil \deg(f)/2 \rceil$, we have the semidefinite programming relaxation

$$L_t = \inf \Big\{ \sum_\alpha f_\alpha y_\alpha : y \in \mathbb{R}^{\{\alpha : |\alpha| \leq 2t\}},\ y_0 = 1,\ M_t(y) \succeq 0,\ M_t^g(y) \succeq 0 \text{ for } g \in \mathcal{G} \Big\}.$$

This is a (strengthened) variation on the Lasserre hierarchy [**63**]. This gives a nondecreasing sequence of lower bounds on p and under mild conditions on $\mathcal{G}$ these bounds converge to p.

In the case where we enforce the variables to be binary by using the constraints $x_i^2 - x_i \geq 0$ and $x_i - x_i^2 \geq 0$ for $i \in [n]$, we can simplify the hierarchy. For each feasible y the localizing matrix corresponding to a constraint $x_i^2 - x_i \geq 0$ is both positive and negative definite, and hence equal to zero. It follows that $y_\alpha = y_{\bar{\alpha}}$ for each $\alpha \in \mathbb{N}_0^n$, where $\bar{\alpha}$ is obtained from α by replacing all nonzero entries by ones. By restricting the vectors to be of this form and removing the polynomials $x_i^2 - x_i$ and $x_i - x_i^2$ from $\mathcal{G}$ we simplify the hierarchy. We may assume all polynomials to be square free and we index their entries by subsets of $[n]$ instead of 0/1 vectors. The moment matrix of a real vector y indexed by elements from $[n]_{2t} = \{S \subseteq [n] : |S| \leq 2t\}$ is now defined as $M(y)_{J,J'} = y_{J \cup J'}$ for $J, J' \in [n]_t$, and we modify the truncated/localizing matrices in the same way. The hierarchy becomes

$$L_t = \inf \Big\{ \sum_{S \in [n]_{2t}} f_S\, y_S : y \in \mathbb{R}^{[n]_{2t}},\ y_\emptyset = 1,\ M_t(y) \succeq 0,\ M_t^g(y) \succeq 0 \text{ for } g \in \mathcal{G} \Big\}.$$

In [**64**] it is shown that the relaxation is sharp for $t = n$.

The maximum independent set problem, which asks for a largest set of pairwise nonadjacent vertices in a finite graph $G = (V, E)$, can be written as a polynomial optimization problem with a binary variable x_v for each vertex $v \in V$ and a constraint $x_u + x_v \leq 1$ for each edge $\{u, v\} \in E$. In [**65**] it is shown that for $t \geq 2$ the t-th step of the (maximization version of the) Lasserre hierarchy reduces to

$$\vartheta_t(G) = \max \Big\{ \sum_{x \in V} y_{\{x\}} : y \in \mathbb{R}^{[n]_{2t}},\ y_\emptyset = 1,\ M_t(y) \succeq 0,\ y_S = 0 \text{ for } S \text{ dependent} \Big\}.$$

Our strengtened version reduces to this hierarchy for all $t \geq 1$. This hierarchy converges to the indepence number $\alpha(G)$ in $\alpha(G)$ steps. The map $P \colon \mathbb{R}^{V_{2t}} \to \mathbb{R}^V$ defined by $P(y)_v = y_{\{v\}}$ identifies $\vartheta_t(G)$ as a lift (see Section 2.2) of the relaxation $\max\{\sum_{x \in V} x_v : x \in P(F_t)\}$ where F_t is the feasible set of $\vartheta_t(G)$. The first step is equivalent to the Lovász ϑ-number [**88**, Theorem 67.10] which a well-known relaxation in combinatorial optimization. When the edge set is invariant under a group action on the vertices, this is a good example where the symmetrization procedure from the previous section applies.

CHAPTER 3

Invariant positive definite kernels

In this chapter we consider cones of invariant positive definite kernels. In particular, we show how to construct simultaneous block diagonalizations of such kernels. This is mostly a background chapter where the main topics are the Peter–Weyl and Bochner theorems from harmonic analysis. New contributions are the generalization of some results about positive type functions to kernels, and results for kernels that are invariant under group actions with infinitely many orbits. Apart from some Hilbert space theory and results about unitary representations we give full proofs.

3.1. Introduction: From matrices to kernels

We can view a matrix in $\mathbb{C}^{n\times n}$ as a map $[n]\times[n]\to\mathbb{C}$, or we can view it as a linear operator on $\mathbb{C}^n$. In the first interpretation we generalize the set $[n]$ to a compact Hausdorff space X and generalize from matrices to continuous functions $X\times X\to\mathbb{C}$. We call such functions *(continuous) kernels* on X. In the second interpretation we generalize the space $\mathbb{C}^n$ to the Hilbert space $L^2_{\mathbb{C}}(X,\mu)$, where μ is a strictly positive Radon measure on X, and consider *Hilbert–Schmidt integral operators*. These are operators of the form

$$T_K\colon L^2_{\mathbb{C}}(X,\mu)\to L^2_{\mathbb{C}}(X,\mu),\ T_Kf(x)=\int K(x,y)f(y)\,d\mu(y),$$

where $K\in L^2_{\mathbb{C}}(X\times X,\mu\otimes\mu)$ is called a *Hilbert–Schmidt kernel*. The subscript $\mathbb{C}$ here indicates the functions are complex-valued.

A continuous kernel K is said to be *positive definite* if the matrix $(K(x_i,x_j))_{i,j=1}^n$ is positive semidefinite for all $n\in\mathbb{N}$ and $x\in X^n$. If we view a kernel $K\colon X\times X\to\mathbb{C}$ as an infinite matrix whose rows and columns are indexed by X, then the above condition requires all finite principal submatrices to be positive semidefinite. A Hilbert–Schmidt kernel K is said to be *positive definite* if T_K is a positive operator; that is, $\langle T_Kf,f\rangle\geq 0$ for all $f\in L^2_{\mathbb{C}}(X,\mu)$, where $\langle\cdot,\cdot\rangle$ denotes the inner product of the Hilbert space $L^2_{\mathbb{C}}(X,\mu)$. In other words, a Hilbert–Schmidt kernel K is positive definite if

$$\iint K(x,y)f(x)\overline{f(y)}\,d\mu(x)d\mu(y)\geq 0\quad\text{for all}\quad f\in L^2_{\mathbb{C}}(X,\mu).$$

A continuous kernel is positive definite if and only if it is positive definite as a Hilbert–Schmidt kernel; see Lemma 3.4.2. Positive definite kernels are Hermitian (for Hilbert–Schmidt kernels this follows from the polarization identity), and the sets $\mathcal{C}_{\mathbb{C}}(X\times X)_{\succeq 0}$ and $L^2_{\mathbb{C}}(X\times X,\mu\otimes\mu)_{\succeq 0}$ of positive definite kernels form cones in the real vector spaces $\mathcal{C}_{\mathbb{C}}(X\times X)_{\mathrm{her}}$ and $L^2_{\mathbb{C}}(X\times X,\mu\otimes\mu)_{\mathrm{her}}$ of Hermitian kernels.

Let Γ be a compact topological group acting continuously on X and assume μ to be Γ-invariant; that is, $\mu(\gamma E) = \mu(E)$ for all $\gamma \in \Gamma$ and all measurable subsets E of X. A continuous kernel K is Γ-*invariant* if $K(\gamma x, \gamma y) = K(x, y)$ for all $\gamma \in \Gamma$ and $x, y \in X$. A Hilbert–Schmidt kernel is Γ-invariant if the operator T_K commutes with $L(\gamma)$ for all $\gamma \in \Gamma$, where $L(\gamma)$ is the unitary operator on $L^2_{\mathbb{C}}(X, \mu)$ defined by $L(\gamma)f(x) = f(\gamma^{-1}x)$. By strict positivity of μ this is equivalent to requiring $K(\gamma x, \gamma y) = K(x, y)$ for all $\gamma \in \Gamma$ and μ-almost all $x, y \in X$, which shows that a continuous kernel is Γ-invariant if and only if it is Γ-invariant as a Hilbert–Schmidt kernel. The spaces of Γ-invariant Hermitian kernels are complete; that is, $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\mathrm{her}}$ is a Banach space with the supremum norm, and $L^2_{\mathbb{C}}(X \times X, \mu \otimes \mu)^{\Gamma}_{\mathrm{her}}$ is a Hilbert space.

The goal of this chapter is to understand the structure of the cone $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$ of continuous, Γ-invariant, positive definite kernels. In particular, we want to find a simultaneous block diagonalization of the elements in this cone. In Section 3.2 we characterize the extreme rays of $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$ and show how this suggests a block form. In Section 3.3 we use the Peter–Weyl theorem to show that X always admits a symmetry adapted system, and in Section 3.4 we use Bochner's theorem to give a sequence of inner approximating cones consisting of block diagonalized kernels.

3.2. A characterization of the extreme rays

In this section we characterize the extreme rays of the cone $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$ of continuous, Γ-invariant, positive definite kernels on X. For the results in this section we only require Γ and X to be locally compact instead of compact.

For the case where X equals the group Γ, we can identify $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$ with the cone of *positive definite functions* on Γ. These are continuous functions $f \colon \Gamma \to \mathbb{C}$ for which the matrix $(f(\gamma_j^{-1}\gamma_i))_{i,j=1}^n$ is positive semidefinite for all $n \in \mathbb{N}$ and $\gamma_1, \ldots, \gamma_n \in \Gamma$. Positive definite functions are well studied objects in harmonic analysis, and the results in this section generalize some results about these functions as described in Folland's book [**33**] to the case of kernels.

An *extreme ray* of a cone K is a set $\mathbb{R}_{\geq 0}x$, with $x \in K$, such that for all $x_1, x_2 \in K$ we have $x_1, x_2 \in \mathbb{R}_{\geq 0}x$ whenever $x = x_1 + x_2$. A vector x for which $\mathbb{R}_{\geq 0}x$ is an extreme ray is called an *extreme direction*. Since $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$ lies in an infinite dimensional space, it is not immediately clear that it admits any extreme rays. For instance, the cone $\mathcal{C}(X)_{\geq 0}$ of nonnegative, continuous functions on X does not have extreme rays unless X has isolated points. We will see, however, that $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$ has sufficiently many extreme rays to approximate any kernel in this cone by convex combinations of extreme directions.

To characterize the extreme rays we use representation theory. A *unitary representation* of Γ is a strongly continuous group homomorphism π from Γ to the group $U(\mathcal{H})$ of unitary operators on a nonzero Hilbert space $\mathcal{H}$. *Strong continuity* here means that we require π to be continuous given that $U(\mathcal{H})$ is endowed with the strong operator topology. In other words, we require the map $\gamma \mapsto \pi(\gamma)u$ to be continuous for each $u \in \mathcal{H}$. On $U(\mathcal{H})$ the weak and strong operator topologies coincide, so we can equivalently require the map $\gamma \mapsto \langle \pi(\gamma)u, v \rangle$ to be continuous for all $u, v \in \mathcal{H}$ [**33**].

Given two topological spaces X and Y with continuous actions of the group Γ, we denote by $\mathrm{Hom}_\Gamma(X,Y)$ the set of continuous Γ-equivariant maps from X to Y. Here a map $\varphi\colon X \to Y$ is said to be Γ-*equivariant* if $\varphi(\gamma x) = \gamma\varphi(x)$ for all $\gamma \in \Gamma$ and $x \in X$. For our applications Y typically is a Hilbert space where the action comes from a unitary representation on this space.

We take the following theorem from [**23**].

THEOREM 3.2.1. *For each kernel $K \in \mathcal{C}_\mathbb{C}(X \times X)^\Gamma_{\succeq 0}$, there exists a unitary representation $\pi_K\colon \Gamma \to U(\mathcal{H}_K)$ and a map $\varphi_K \in \mathrm{Hom}_\Gamma(X, \mathcal{H}_K)$ such that*

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

PROOF. Let $\mathbb{C}^X$ be the vector space of formal complex 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 linearly in the first and antilinearly in the second component. The completion of $\mathbb{C}^X/N$ is a Hilbert space which we denote by $\mathcal{H}_K$, and the action of Γ on X extends to the homomorphism $\pi_K\colon \Gamma \to U(\mathcal{H}_K)$, where $\pi_K(\gamma)$ is inner product preserving because K is Γ-invariant.

Since $\langle \pi_K(\gamma)x + N, y + N\rangle = K(\gamma x, y)$, it follows from both K and the action of Γ on X being continuous, that the map $\gamma \to \langle \pi_K(\gamma)x + N, y + N\rangle$ is continuous. So π_K is a unitary representation.

We define the Γ-equivariant map $\varphi_K\colon X \to \mathcal{H}_K$ by $\varphi_K(x) = x + N$. This map is continuous because

$$\|\varphi_K(y) - \varphi_K(x)\|^2 \le K(x,x) + K(y,y) - K(x,y) - K(y,x). \qquad \square$$

The image of the map φ_K constructed in the above theorem has dense span in $\mathcal{H}_K$. In the following lemma we show that under this condition π_K and φ_K are essentially unique.

LEMMA 3.2.2. *Assume that for $i = 1, 2$, $\pi_i\colon \Gamma \to \mathcal{H}_i$ is a unitary representation and $\varphi_i \in \mathrm{Hom}_\Gamma(\Gamma, \mathcal{H}_i)$ is a function whose image has dense span in $\mathcal{H}_i$. If*

$$\langle \varphi_1(x), \varphi_1(y)\rangle = \langle \varphi_2(x), \varphi_2(y)\rangle \quad \textit{for all} \quad x, y \in X,$$

then there exists a unitary operator $T \in \mathrm{Hom}_\Gamma(\mathcal{H}_1, \mathcal{H}_2)$ such that $\varphi_2 = T \circ \varphi_1$.

PROOF. Let $x, y \in X$. If $\varphi_1(x) = \varphi_1(y)$, then

$$\begin{aligned}\|\varphi_2(x) - \varphi_2(y)\|^2 &= \langle \varphi_2(x) - \varphi_2(y), \varphi_2(x) - \varphi_2(y)\rangle \\ &= \langle \varphi_1(x) - \varphi_1(y), \varphi_1(x) - \varphi_1(y)\rangle = \|\varphi_1(x) - \varphi_1(y)\|^2 = 0,\end{aligned}$$

so $\varphi_2(x) = \varphi_2(y)$. This shows the map $T\colon \{\varphi_1(x) : x \in X\} \to \mathcal{H}_2$ defined by $T(\varphi_1(x)) = \varphi_2(x)$ is well-defined. Since the image of φ_1 has dense span in $\mathcal{H}_1$, we can extend T to an operator $\mathcal{H}_1 \to \mathcal{H}_2$. Since the span of the image of φ_2 is dense, the operator T is surjective, and since

$$\|T\varphi_1(x)\|^2 = \|\varphi_2(x)\|^2 = \langle \varphi_2(x), \varphi_2(x)\rangle = \langle \varphi_1(x), \varphi_1(x)\rangle = \|\varphi_1(x)\|^2,$$

it is an isometry, so T is a unitary operator. It is also Γ-equivariant:

$$T\pi_1(\gamma)\varphi_1(x) = T\varphi_1(\gamma^{-1}x) = \varphi_2(\gamma^{-1}x) = \pi_2(\gamma)\varphi_2(x). \qquad \square$$

In the following theorem we characterize the extreme rays of the cone of continuous, Γ-invariant, positive definite kernels. When specialized to matrices; that is, when $X = [n]$ and Γ is the trivial group, it says that the extreme directions of the positive semidefinite cone $\mathcal{S}^n_{\succeq 0}$ are of the form xx^* for $x \in \mathbb{C}^n$. For the more general case where X is a locally compact topological space and Γ the trivial group it says that the extreme rays are given by $\{\mathbb{R}_{\geq 0} f \otimes \bar{f} : f \in \mathcal{C}_{\mathbb{C}}(X)\}$.

For the statement and proof of this result we first need more representation theory. A subspace $\mathcal{M}$ of the Hilbert space $\mathcal{H}$ of a unitary representation $\pi\colon \Gamma \to U(\mathcal{H})$ is said to be Γ-*invariant* if it is closed and if $\pi(\gamma)u \in \mathcal{M}$ for all $\gamma \in \Gamma$ and $u \in \mathcal{M}$. A unitary representation π is irreducible if the trivial representation and the representation π itself are the only invariant subspaces. Two representations $\pi_1\colon \Gamma \to U(\mathcal{H}_1)$ and $\pi_2\colon \Gamma \to U(\mathcal{H}_2)$ are said to be *equivalent* if $\mathrm{Hom}_\Gamma(\mathcal{H}_1, \mathcal{H}_2)$ contains a unitary operator. The first result we need is the observation that any reducible unitary representation is the direct sum of two nontrivial unitary representations. This follows from the fact that when $\mathcal{M}$ is an invariant subspace, then also its orthogonal complement $\mathcal{M}^\perp$ is an invariant subspace. The second result we need is Schur's lemma, which says that the space $\mathrm{Hom}_\Gamma(\mathcal{H}, \mathcal{H})$ consists of scalar multiples of the identity operator if and only if $\mathcal{H}$ is irreducible, and the space $\mathrm{Hom}_\Gamma(\mathcal{H}_1, \mathcal{H}_2)$ is one dimensional if and only if $\mathcal{H}_1$ and $\mathcal{H}_2$ are both irreducible.

THEOREM 3.2.3. *A kernel $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$ is an extreme direction if and only if π_K is irreducible.*

PROOF. If π_K is reducible, then $\mathcal{H}_K$ admits a nontrivial orthogonal decomposition $\mathcal{M}_1 \oplus \mathcal{M}_2$ into π_K-invariant subspaces. Let $\varphi_i = P_i \circ \varphi_K$, where $P_i\colon \mathcal{H}_K \to \mathcal{M}_i$ is the projection operator onto $\mathcal{M}_i$, and where φ_K is the function defined in Theorem 3.2.1. Let $K_i(x, y) = \langle \varphi_i(x), \varphi_i(y)\rangle$, so that $K = K_1 + K_2$. The kernels K_1 and K_2 do not lie on the same ray: The image of φ_i has dense span in $\mathcal{M}_i$, so if $K_1 = |c|^2 K_2$ for some $c \in \mathbb{C}$, then by Lemma 3.2.2 there exists a unitary, Γ-equivariant operator $T\colon \mathcal{M}_1 \to \mathcal{M}_2$ such that $\varphi_2 = cT \circ \varphi_1$. But this means $\varphi_K = \varphi_1 + \varphi_2 = \varphi_1 + cT\varphi_1$, which contradicts with the image of φ_K having dense span in $\mathcal{H}_K$. Hence K is not an extreme direction.

Now assume π_K is irreducible and $K = K_1 + K_2$ for $K_1, K_2 \in \mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$. We have $K_1(x, x) = K(x, x) - K_2(x, x) \leq K(x, x)$ for all $x \in X$, so

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

So we can use K_1 to define a bounded Hermitian form on $\mathcal{H}_K$, which defines a bounded self-adjoint operator T in $\mathcal{H}_K$ for which $K_1(x, y) = \langle T\varphi_K(x), \varphi_K(y)\rangle$ for all $x, y \in X$. This operator T is Γ-equivariant: For all $x, y \in X$ we have

$$\begin{aligned}\langle T\pi_K(\gamma)\varphi_K(x), \varphi_K(y)\rangle &= \langle T\varphi_K(\gamma^{-1}x), \varphi_K(y)\rangle = K_1(\gamma^{-1}x, y) = K_1(x, \gamma y)\\ &= \langle T\varphi_K(x), \varphi_K(\gamma y)\rangle = \langle T\varphi_K(x), \pi_K(\gamma^{-1})\varphi_K(y)\rangle\\ &= \langle \pi_K(\gamma)T\varphi_K(x), \varphi_K(y)\rangle.\end{aligned}$$

Since π_K is irreducible, Schur's lemma implies $T = cI$ for some $c \in \mathbb{C}$. Thus $K_1(x, y) = c\langle \varphi_K(x), \varphi_K(y)\rangle = cK(x, y)$ for all $x, y \in X$, and hence $K_1 = cK$ and $K_2 = (1 - c)K$ which means K spans an extreme ray. □

Denote by $\hat{\Gamma}$ a complete set of non-equivalent irreducible unitary representations. The above theorems suggest that in a block diagonalization of a kernel we should have a block for each $\pi \in \hat{\Gamma}$. Let E_π be the cone of all kernels corresponding to the irreducible representation π of Γ; that is,

$$E_\pi = \operatorname{cone}\Big\{ K_\varphi : \varphi \in \operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi) \Big\}, \quad \text{where} \quad K_\varphi(x,y) = \langle \varphi(x), \varphi(y) \rangle.$$

If $m = \dim(\operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)) < \infty$, then E_π is isomorphic to the cone of complex positive semidefinite $m \times m$ matrices: Let $\varphi_1, \dots, \varphi_m$ be a basis of $\operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)$, then $\varphi = a_1\varphi_1 + \cdots + a_m\varphi_m$ for some $a_1, \dots, a_m \in \mathbb{R}$, so

$$K_\varphi = \sum_{i,j=1}^{m} a_i \overline{a_j} K_{\varphi_i, \varphi_j} \quad \text{where} \quad K_{\varphi,\psi}(x,y) = \langle \varphi(x), \psi(y) \rangle.$$

We are especially interested in situations where $\operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)$ is infinite dimensional, and where there are infinitely many non-equivalent irreducible representations. This means we have to consider convergence, which we do in the remainder of this chapter.

3.3. Symmetry adapted systems

From now on we assume X to be a compact, metrizable topological space with a continuous action of a compact group Γ. The space $\mathcal{C}_\mathbb{C}(X)$ is separable, so there exists a linearly independent sequence $\{e_i\}$ whose span is uniformly dense in $\mathcal{C}_\mathbb{C}(X)$. Given a Radon measure μ on X, by Gram–Schmidt orthogonalization we may assume $\{e_i\}$ to be orthonormal in $L^2_\mathbb{C}(X, \mu)$. We call such a sequence a *complete orthonormal system* of X. The goal of this section is to show the existence of a complete orthonormal system that is in harmony with the group action. We later use this to construct a Fourier basis for invariant kernels.

We start by fixing a strictly positive, Γ-invariant, Radon probability measure μ on X, which by the following lemma always exists.

LEMMA 3.3.1. *The space X admits a Radon probability measure that is strictly positive and Γ-invariant.*

PROOF. Let $\{x_i\}$ be a dense sequence in the separable space X and $\{a_i\}$ a sequence of strictly positive numbers that sums to one. Define a Borel probability measure μ_0 by setting

$$\mu_0(U) = \sum_{i : x_i \in U} a_i$$

for each open subset U of X. This measure is strictly positive by construction. We define a Γ-invariant Borel probability measure μ by setting $\mu(U) = \int_\Gamma \mu_0(\gamma U)\, d\gamma$ for $U \subseteq X$ open, where integration is over the normalized Haar measure of Γ. The measure μ is a strictly positive probability measure since the total measure and strict positivity are preserved by invariant integration. The measure μ is finite, and since X is a separable metric space and μ a Borel measure, μ is also inner regular and hence a Radon measure. □

The action of Γ on X induces the unitary representation

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

Similarly, we have the representation $L\colon \Gamma \to \mathcal{L}(\mathcal{C}(X))$, where $\mathcal{L}(\mathcal{C}_{\mathbb{C}}(X))$ is the space of bounded operators on $\mathcal{C}_{\mathbb{C}}(X)$, so that each finite dimensional subrepresentation of $\mathcal{C}_{\mathbb{C}}(X)$ is a unitary subrepresentation of $L^2_{\mathbb{C}}(X,\mu)$.

Denote the dimension of the representation π by d_π. A complete orthonormal system of X is said to be a *symmetry adapted system* of X if there exist numbers $m_\pi \in \{0,1,\dots,\infty\}$ for which we can write the set as

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

where $H_{\pi,i} = \operatorname{span}\{e_{\pi,i,1},\dots,e_{\pi,i,d_\pi}\}$ is equivalent to π, and where there exist unitary operators $T_{\pi,i,i'} \in \operatorname{Hom}_\Gamma(H_{\pi,i}, H_{\pi,i'})$ with $e_{\pi,i',j} = T_{\pi,i,i'}e_{\pi,i,j}$ for all π, i, i', and j. It can be shown that the numbers m_π are the same for each symmetry adapted system of X and are given by the dimension of the space $\operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)$.

To prove that a symmetry adapted system always exists we use approximate identities, and to define these we use the integral operators T_K from Section 3.1. We say a sequence of kernels $\{I_n\}$ in $\mathcal{C}_{\mathbb{C}}(X\times X)$ is an *approximate identity* of X if $\|T_{I_n}f - f\|_\infty \to 0$ for each $f \in \mathcal{C}_{\mathbb{C}}(X)$.

LEMMA 3.3.2. *The space X admits an approximate identity $\{I_n\}$ where each I_n is real-valued, symmetric, and Γ-invariant.*

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

$$\mu(U^n_i \setminus C^n_i) \le \mu(C^n_i)/n,$$

which is possible by inner regularity of μ, and remove C^n_i from the sets U^n_j for $j \neq i$. We then have $C^n_i \cap U^n_{i'} = \emptyset$ for all n and all distinct i and i'.

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

$$K_n(x,y) = \sum_i \frac{p^n_i(x)p^n_i(y)}{\mu(C^n_i)}.$$

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

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

Then for each $x \in X$,

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

with

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

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 real-valued, symmetric, and Γ-invariant for each n, and the sequence $\{I_n\}$ is an approximate identity: For $f \in \mathcal{C}_\mathbb{C}(X)$ and $\bar{f}(x) = \int_\Gamma f(\gamma x)\, d\gamma$ we have

$$\|T_{I_n} f - f\|_\infty = \sup_{x \in X} \left| \int_X \int_\Gamma (K_n(\gamma x, \gamma y) f(y) - f(x))\, d\gamma\, d\mu(y) \right|$$
$$= \sup_{x \in X} \left| \int_X \int_\Gamma (K_n(x, y) f(\gamma^{-1} y) - f(\gamma^{-1} x))\, d\gamma\, d\mu(y) \right|$$
$$= \|T_{K_n} \bar{f} - \bar{f}\|_\infty \to 0. \qquad \square$$

Now that we have established the existence of invariant, strictly positive measures and invariant approximate identities, we can prove $\mathcal{C}_\mathbb{C}(X)$ has enough finite-dimensional invariant subspaces to span a dense subspace. This result is an important part of the Peter–Weyl theorem, and the proof we give here is a direct adaptation of the original proof, which can for instance be found in [**33**] or [**101**] for the left regular representation, to the setting of a compact group acting on another topological space. Below the sum of a set of subspaces is defined as the set of all finite sums from elements of those spaces.

LEMMA 3.3.3. *The space $\mathcal{C}_\mathbb{C}(X)$ is equal to the closure of the sum of its finite dimensional Γ-invariant subspaces.*

PROOF. Let $f \in \mathcal{C}_\mathbb{C}(X)$ and $\varepsilon > 0$. By Lemma 3.3.2 there exists a continuous, Hermitian, Γ-invariant kernel $K \in \mathcal{C}_\mathbb{C}(X \times X)$ such that $\|T_K f - f\|_\infty \leq \varepsilon$. We will show that $T_K f$ is the uniform limit of linear combinations of functions from finite dimensional, Γ-invariant subspaces.

Using Fubini's theorem we have $\langle T_K f, g \rangle = \langle g, T_K f \rangle$ for all $f, g \in L^2(X, \mu)$, so T_K is self-adjoint. Let d be a metric on X that agrees with the topology of X. Since X is compact, the kernel K is uniformly continuous, and this implies that for each $\kappa > 0$ there is a $\delta > 0$ such that $|K(x_1, y) - K(x_2, y)| \leq \kappa$ for all $y \in X$ and

$x_1, x_2 \in X$ satisfying $d(x_1, x_2) \leq \delta$. This means that for each f in the closed unit ball B of $L^2_{\mathbb{C}}(X, \mu)$, we have

$$|T_K f(x_1) - T_K f(x_2)| \leq \kappa \|f\|_1 \leq \kappa \|f\|_2 \leq \kappa,$$

which shows equicontinuity of the set $T_K B$. The set $T_K B$ is also pointwise bounded: For each $f \in B$ we have

$$\|T_K f\|_\infty \leq \|K\|_\infty \|f\|_1 \leq \|K\|_\infty \|f\|_2 \leq \|K\|_\infty.$$

So, by the Arzelà–Ascoli theorem $T_K B$ is relatively compact in $\mathcal{C}_{\mathbb{C}}(X)$ in the uniform topology, and hence compact in $L^2_{\mathbb{C}}(X, \mu)$. This shows T_K is a compact operator.

The spectral theorem for compact self-adjoint operators on separable Hilbert spaces tells us that $L^2(X, \mu)$ is the Hilbert space direct sum of the spaces

$$E_\lambda = \{f \in L^2_{\mathbb{C}}(X, \mu) : T_K f = \lambda f\}.$$

Moreover, for each $\lambda > 0$ the space E_λ is finite dimensional, and the values of λ for which E_λ is nonzero can only accumulate at 0. Each of these eigenspaces is invariant: If $T_K f = \lambda f$, then

$$\begin{aligned}(T_K L(\gamma) f)(x) &= \int K(x, y) L(\gamma) f(y)\, d\mu(y) = \int K(x, y) f(\gamma^{-1} y)\, d\mu(y) \\ &= \int K(\gamma^{-1} x, y) f(y)\, d\mu(y) = T_K f(\gamma^{-1} x) \\ &= L(\gamma) T_K f(x) = \lambda L(\gamma) f(x),\end{aligned}$$

for all $x \in X$.

Let f_t be the projection of f on $\oplus_{\lambda \geq t} E_\lambda$. Then $f_t \to f$ in $L^2(X, \mu)$, so

$$\|T_K f_t - T_K f\|_\infty \leq \|K\|_\infty \|f_t - f\|_1 \leq \|K\|_\infty \|f_t - f\|_2 \to 0.$$

Each f_t is of the form $f_t = \sum_{\lambda \geq t} f_{t,\lambda}$, with $f_{t,\lambda} \in E_\lambda$. We have $T_K f_{t,\lambda} \in E_\lambda$, which means that $T_K f_t = \sum_{\lambda \geq t} T_K f_{t,\lambda}$ is a sum of functions from finite dimensional Γ-invariant subspaces. □

We will need the following variation on the Schur orthogonality relations. We take the proof from [**98**].

LEMMA 3.3.4. *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.*

PROOF. Define the operator $A \colon \mathcal{M} \to \mathcal{M}'$ by

$$Ae_i = \sum_j \langle e_i, e'_j \rangle e'_j,$$

and extending by linearity. This operator is Γ-invariant:

$$\begin{aligned} A\pi(\gamma)e_i &= \sum_j \langle \pi(\gamma)e_i, e_j\rangle e_j = \sum_j \langle e_i, \pi(\gamma)e_j\rangle e_j \\ &= \sum_j \langle e_i, \sum_k \pi(\gamma^{-1})_{k,j} e_k\rangle e_j = \sum_k \langle e_i, e_k\rangle \sum_j \overline{\pi(\gamma^{-1})_{k,j}} e_j \\ &= \sum_k \langle e_i, e_k\rangle \sum_j \pi(\gamma)_{j,k} e_j = \pi(\gamma)Ae_i. \end{aligned}$$

So, if $\mathcal{M}$ and $\mathcal{M}'$ are not equivalent, then by Schur's lemma A must be the zero operator. Since the e'_j are linearly independent this implies $\langle e_i, e'_j\rangle = 0$ for all i and j, which proves the first part of the lemma.

For the second part we note that the operator $T^{-1}A$ is Γ-invariant, so by Schur's lemma $T^{-1}A = cI$ for some $c \in \mathbb{C}$. This means

$$\sum_j \langle e_i, e'_j\rangle e'_j = Ae_i = cTe_i = ce'_i$$

for all i, which implies $\langle e_i, e'_j\rangle = c\delta_{i,j}$ for all i and j. □

In the following proof we use the concept of linearly independent subspaces. We say that a set S of nonzero subspaces is *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 the union of any set of bases of each of the subspaces to be linearly independent. We will use the following property: If S is a linearly independent set of subspaces and A is a nonzero subspace such that the intersection of A with the sum of any finite set of spaces from S is the zero space, then $S \cup \{A\}$ is a linearly independent set of subspaces.

THEOREM 3.3.5. *The space X admits a symmetry adapted system.*

PROOF. Let C be the set containing all linearly independent sets of nonzero, finite dimensional, Γ-invariant subspaces of $\mathcal{C}(X)$. If X is nonempty, then $\mathcal{C}(X)$ is nonempty, so by Lemma 3.3.3 the set C is nonempty.

We 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 nonzero, 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 $\bar{P}$ be the closure of P in $\mathcal{C}(X)$. If $\bar{P}$ is not equal to $\mathcal{C}(X)$, then by Lemma 3.3.3 there exists a finite dimensional Γ-invariant subspace V of $\mathcal{C}(X)$ containing a vector u that does not lie in P. The cyclic subspace $W = \operatorname{span}\{L(\gamma)u : \gamma \in \Gamma\}$ is finite dimensional, Γ-invariant, and has trivial intersection with P: For all $\gamma \in \Gamma$ we have $L(\gamma)u \notin L(\gamma)P = P$. So $M \cup \{W\}$ is linearly independent, and hence contained in C. This contradicts maximality of M, thus $\bar{P}$ must be equal to $\mathcal{C}(X)$.

The finite dimensional representations in M are unitary subrepresentations of $L^2(X, \mu)$ and decompose into irreducible subrepresentations of $\mathcal{C}(X)$. So we may

assume M to be a linearly independent set of finite dimensional, Γ-irreducible subspaces of $\mathcal{C}(X)$.

Denote by $m_\pi \in \{0, 1, \ldots, \infty\}$ the number of representations in M that are equivalent to π. Let $\{\tilde{e}_{\pi,i,j} : \pi \in \hat{\Gamma},\, i \in [m_\pi],\, j \in [d_\pi]\}$ be a complete symmetry adapted system of $\mathcal{C}(X)$ by selecting appropriate orthonormal bases $\tilde{e}_{\pi,i,1}, \ldots, \tilde{e}_{\pi,i,d_\pi}$ of the representations in M. Now give this system any ordering where $\tilde{e}_{\pi,i,j}$ occurs before $\tilde{e}_{\pi,i',j'}$ whenever $i < i'$, and apply the Gram–Schmidt process to obtain a new sequence $\{e_{\pi,i,j}\}$. This process preserves linear independence and also preservers the span of the sequence, so the new sequence $\{e_{\pi,i,j}\}$ is a complete orthonormal system in $\mathcal{C}(X)$. By Lemma 3.3.4 we have

$$e_{\pi,i,j} = \tilde{e}_{\pi,i,j} - \sum_{k=1}^{i-1} \langle \tilde{e}_{\pi,i,j}, \tilde{e}_{\pi,k,j} \rangle \tilde{e}_{\pi,k,j} = \tilde{e}_{\pi,i,j} - \sum_{k=1}^{i-1} c_{\pi,i,k} \tilde{e}_{\pi,k,j},$$

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

3.4. Block diagonalized kernels

We use the symmetry adapted system from the previous section to construct simultaneous block diagonalizations of invariant kernels, and we discuss convergence of these block diagonalizations.

First we consider the case where Γ is the trivial group. Then any complete orthonormal system $\{e_i\}$ on X is symmetry adapted. The sequence $\{e_i \otimes \overline{e_{i'}}\}_{i,i'}$ forms a complete orthonormal system in $\mathcal{C}_{\mathbb{C}}(X \times X)$ and hence also in $L^2(X \times X, \mu \otimes \mu)$. We define the possibly infinite matrix $Z(x, y)$ by

$$Z(x, y) = E(x)E(y)^*,$$

where $E(x)$ is the vector with $E(x)_i = e_i(x)$. For a kernel $K \in \mathcal{C}_{\mathbb{C}}(X \times X)$, we then have the Fourier series

$$K(x, y) = \sum_{i,i'} \hat{K}_{i,i'} Z(x, y)_{i,i'},$$

with convergence in L^2, where the Fourier coefficient matrix is given by

$$\hat{K} = \iint K(x, y) Z(x, y)^* \, d\mu(x) d\mu(y),$$

where the integral of the matrix is entrywise.

When Γ is a nontrivial group with nontrivial action on X, then we can block diagonalize the above construction, and under certain conditions we can use the completeness of the symmetry adapted system in $\mathcal{C}(X)$ to prove uniform convergence. Then instead of one Fourier matrix we have a Fourier matrix for each irreducible representation of Γ. We define

$$\hat{K}(\pi) = \iint K(x, y) Z_\pi(x, y)^* \, d\mu(x) d\mu(y),$$

where

$$Z_\pi(x, y) = E_\pi(x) E_\pi(y)^* \quad \text{and} \quad E_\pi(x)_{i,j} = e_{\pi,i,j}(x).$$

The matrix-valued kernels $Z_\pi(x,y)$, which we call *zonal matrices*, are Γ-invariant and uniquely defined up to unitary basis transformations. They are also positive definite, by which we mean that $(Z_\pi(x_a,x_b)_{i_a,i_b})_{a,b=1}^n$ is positive semidefinite for each $n\in\mathbb{N}$, $i_1,\dots,i_n\in[m_\pi]$, and $x_1,\dots,x_n\in X$. With a similar proof as in Theorem 3.2.1, it is shown in [**23**] that there exist functions $\varphi_1,\dots,\varphi_{m_\pi}\in\mathrm{Hom}_\Gamma(X,\mathcal{H}_\pi)$ for which

$$Z_\pi(x,y)_{i,i'}=\langle\varphi_i(x),\varphi_{i'}(y)\rangle.$$

In the following proposition we show that the matrix entries of the zonal matrices have dense span in the Hilbert space of invariant Hilbert–Schmidt kernels, which proves Fourier inversion and Parseval's theorem for kernels. This gives L^2 converging block diagonalizations of invariant kernels.

PROPOSITION 3.4.1. *The matrix entries of the zonal matrices Z_π, for $\pi\in\hat\Gamma$, form a complete orhonormal system in the Hilbert space $L^2_{\mathbb{C}}(X\times X,\mu\otimes\mu)^\Gamma$. Given $K,G\in L^2_{\mathbb{C}}(X\times X,\mu\otimes\mu)^\Gamma$, we have*

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

with convergence in L^2, and

$$\langle K,G\rangle=\sum_{\pi\in\hat\Gamma}\langle\hat K(\pi),\hat G(\pi)\rangle.$$

PROOF. Let $K\in L^2_{\mathbb{C}}(X\times X,\mu\otimes\mu)$. The sequence $\{e_{\pi,i,j}\otimes\overline{e_{\pi',i',j'}}\}$ forms a complete orthonormal system in $L^2(X\times X,\mu\otimes\mu)$, so

$$K(x,y)=\sum_{\pi,\pi'\in\hat\Gamma}\sum_{i,i'=1}^{m_\pi}\sum_{j,j'=1}^{d_\pi}\langle e_{\pi,i,j},e_{\pi',i',j'}\rangle_K\,e_{\pi,i,j}(x)\overline{e_{\pi',i',j'}(y)},$$

with convergence in L^2, where $\langle\cdot,\cdot\rangle_K$ is the Γ-invariant, sesquilinear form

$$\langle f,g\rangle_K=\iint K(x,y)f(x)\overline{g(y)}\,d\mu(x)d\mu(y).$$

By Lemma 3.3.4 we have $\langle e_{\pi,i,j},e_{\pi',i',j'}\rangle_K=0$ whenever $\pi\neq\pi'$ or $j\neq j'$, and $\langle e_{\pi,i,j},e_{\pi,i',j}\rangle_K$ does not depend on j. So we have

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

with convergence in L^2. This means the matrix entries of the zonal matrices form a complete orthonormal system in the Hilbert space $L^2(X\times X,\mu\otimes\mu)^\Gamma$. The relation $\langle K,G\rangle=\sum_{\pi\in\hat\Gamma}\langle\hat K(\pi),\hat G(\pi)\rangle$ follows from Parseval's identity for Hilbert spaces. □

We show that a continuous kernel is positive definite if and only if it is positive definite as a Hilbert–Schmidt kernel; our proof is an adaptation of [**33**, Proposition 3.35] from functions to kernels. We then use this to show that a kernel is positive definite if and only if its Fourier coefficient matrices are positive semidefinite.

LEMMA 3.4.2. *We have*

$$L^2_{\mathbb{C}}(X \times X, \mu \otimes \mu)^{\Gamma}_{\succeq 0} \cap \mathcal{C}_{\mathbb{C}}(X \times X) = \mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}.$$

PROOF. Let $K \in L^2_{\mathbb{C}}(X \times X, \mu \otimes \mu)^{\Gamma}_{\succeq 0} \cap \mathcal{C}_{\mathbb{C}}(X \times X)$ and let $m \in \mathbb{N}$, $c \in \mathbb{C}^m$, and $x \in X^m$. By Lemma 3.3.2 there exists an approximate identity $\{I_m\}$ of X where each I_m is real-valued, symmetric, and Γ-invariant. Define

$$g_n(x) = \sum_{i=1}^{m} c_i I_n(x_i, x).$$

Then

$$\begin{aligned} 0 &\leq \lim_{n\to\infty} \iint K(x,y) g_n(x) \overline{g_n(y)}\, d\mu(x) d\mu(y) \\ &= \sum_{i,j=1}^{m} c_i \overline{c_j} \lim_{n\to\infty} \iint K(x,y) I_n(x_i, x) I_n(x_j, y)\, d\mu(x) d\mu(y) \\ &= \sum_{i,j=1}^{m} c_i \overline{c_j} K(x_i, x_j). \end{aligned}$$

Hence, $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$.

Now we assume $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$. Since K is continuous and X compact, there exists, for each $\varepsilon > 0$, a partition $X = E_1 \cup \ldots \cup E_N$ such that the difference between the supremum and infimum of K on $E_i \times E_j$ is at most ε. Let $g \in \mathcal{C}_{\mathbb{C}}(X)$, $x_i \in E_i$, and $c_i = \int_{E_i} g\, d\mu$. Then

$$\Big| \sum_{i,j=1}^{N} c_i \overline{c_j} K(x_i, x_j) - \iint K(x,y) g(x) \overline{g(y)}\, d\mu(x) d\mu(y) \Big| \leq \varepsilon |\mu(g)|^2,$$

which shows $K \in L^2_{\mathbb{C}}(X \times X, \mu \otimes \mu)^{\Gamma}_{\succeq 0}$. □

In the following proof we take the idea of using Parseval's theorem from [**25**], where the same theorem is proved for positive definite functions.

PROPOSITION 3.4.3. *Let $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}$. Then K is positive definite if and only if $\hat{K}(\pi)$ is positive semidefinite for each $\pi \in \hat{\Gamma}$.*

PROOF. For all $n \in \mathbb{N}$, $i_1, \ldots, i_n \in [m_\pi]$, and $c_1, \ldots, c_n \in \mathbb{C}$ we have

$$\begin{aligned} \sum_{a,b=1}^{n} c_a \overline{c_b} \hat{K}(\pi)_{i_a, i_b} &= \iint K(x,y) \sum_{a,b=1}^{n} c_a \overline{c_b} Z_\pi(x,y)_{i_a, i_b}\, d\mu(x) d\mu(y) \\ &= \iint K(x,y) \sum_{a,b=1}^{n} c_a \overline{c_b} Z_\pi(x,y)_{i_a, i_b}\, d\mu(x) d\mu(y) \\ &= \sum_{j=1}^{d_\pi} \iint K(x,y) \Big(\sum_{a=1}^{n} c_i e_{\pi, i_a, j}(x) \Big) \overline{\Big(\sum_{a=1}^{n} c_i e_{\pi, i_a, j}(y) \Big)}\, d\mu(x) d\mu(y), \end{aligned}$$

and by Lemma 3.4.2 each of the terms in the right hand side is nonnegative, so $\hat{K}(\pi)$ is positive semidefinite.

For the other direction we let $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}$ be a kernel whose Fourier coefficient matrices are all positive semidefinite. For each $g \in \mathcal{C}_{\mathbb{C}}(X)$, the kernel $g \otimes \bar{g}$ is positive definite which means all its Fourier coefficients are positive semidefinite. Hence by Parseval's theorem as proved in Proposition 3.4.1 we have

$$\iint K(x,y)g(x)\overline{g(y)}\,d\mu(x)d\mu(y) = \langle K, \bar{g} \otimes g\rangle = \sum_{\pi \in \hat{\Gamma}} \langle \hat{K}(\pi), \widehat{\bar{g} \otimes g}(\pi)\rangle \geq 0,$$

so by Lemma 3.4.2 the kernel K is positive definite. □

The Fourier series of a kernel converges in L^2, but in general does not converge uniformly. However, if the action of Γ on X has finitely many orbits and the kernel is positive definite, then this series does converge uniformly. For a transitive action, this result together with the first part of Proposition 3.4.1 and Theorem 3.4.3 is known as Bochner's theorem [**15**]. Our proof here is a generalization of the argument in [**25**] from positive type functions to kernels.

THEOREM 3.4.4. *If the action of Γ on X has finitely many orbits, then for each $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$, the series $\sum_{\pi \in \hat{\Gamma}} \langle \hat{K}(\pi), Z_\pi(x,y)\rangle$ converges absolutely-uniformly.*

PROOF. For each $x \in X$, the map $\Gamma \to X, \gamma \mapsto \gamma x$ is continuous, and since Γ is compact and X Hausdorff, this map is closed. This implies the orbits of the action of Γ on X are closed. Since there are only finitely many orbits, the union of all but one orbit is also closed, and hence each orbit is also open. This means we can construct a symmetry adapted system of X by combining symmetry adapted systems of the orbits, where each function on an orbit is extended to a function on X by setting it to zero outside the orbit. The numbers m_π corresponding to a symmetry adapted system of an orbit of X are upper bounded by d_π, and this shows that the numbers m_π corresponding to the resulting symmetry adapted system of X are finite.

Since K is positive definite, Theorem 3.4.3 says the Fourier coefficients $\hat{K}(\pi)$ are positive semidefinite matrices. This implies the kernel $(x,y) \mapsto \langle \hat{K}(\pi), Z_\pi(x,y)\rangle$ is positive definite, so

$$|\langle \hat{K}(\pi), Z_\pi(x,y)\rangle| \leq \sqrt{\langle \hat{K}(\pi), Z_\pi(x,x)\rangle}\sqrt{\langle \hat{K}(\pi), Z_\pi(y,y)\rangle} \quad \text{for all} \quad x, y \in X.$$

Denote the orbit containing the point $x \in X$ by O_x. Then,

$$\int_X Z_\pi(z,z)\,d\mu(z) = d_\pi I \quad \text{and} \quad \int_{O_x} Z_\pi(z,z)\,d\mu(z) \preceq d_\pi I \quad \text{for each} \quad x \in X,$$

where by $A \preceq B$ we mean that $B - A$ is a positive semidefinite matrix. Since Z_π is Γ-invariant and $Z_\pi(z,z)$ is positive semidefinite for every $z \in X$, we have $Z_\pi(x,x) \preceq d_\pi/\mu(O_x)I$. Since $\hat{K}(\pi)$ is also positive semidefinite, we have

$$\langle \hat{K}(\pi), Z_\pi(x,x)\rangle \leq d_\pi/\mu(O_x)\mathrm{Tr}(\hat{K}(\pi)) \quad \text{for all} \quad x \in X.$$

With $c = \max_{x \in X} 1/\mu(O_x)$, we then have

$$|\langle \hat{K}(\pi), Z_\pi(x,y)\rangle| \leq c\, d_\pi \mathrm{Tr}(\hat{K}(\pi)) \quad \text{for all} \quad x, y \in X.$$

In the remainder of the proof we show $\sum_{\pi \in \hat{\Gamma}} \mathrm{Tr}(\hat{K}(\pi)) < \infty$, which completes the proof by the Weierstrass M-test we are done when we show

Let F be a finite subset of $\hat{\Gamma}$. The series

$$\sum_{\pi \in \hat{\Gamma} \setminus F} \langle \hat{K}(\pi), Z_\pi(x,y) \rangle$$

converges to $K - G$ in $L^2_{\mathbb{C}}(X \times X, \mu \otimes \mu)$, where

$$G(x,y) = \sum_{\pi \in F} \langle \hat{K}(\pi), Z_\pi(x,y) \rangle.$$

This means that each Fourier coefficient of $K - G$ either is the zero matrix or is given by $\hat{K}(\pi)$ for some $\pi \in \hat{\Gamma}$. So, each Fourier coefficient is positive semidefinite, thus $K - G$ is a positive definite kernel.

Hence,

$$\begin{aligned}\sum_{\pi \in F} d_\pi \mathrm{Tr}(\hat{K}(\pi)) &= \sum_{\pi \in F} \left\langle \hat{K}(\pi), \int Z_\pi(z,z)\, d\mu(z) \right\rangle = \int G(z,z)\, d\mu(z) \\ &= \int K(z,z)\, d\mu(z) - \int (K-G)(z,z)\, d\mu(z) \leq \int K(z,z)\, d\mu(z),\end{aligned}$$

and since $\int K(z,z)\, d\mu(z)$ does not depend on F, this shows

$$\sum_{\pi \in \hat{\Gamma}} d_\pi \mathrm{Tr}(\hat{K}(\pi)) < \infty. \qquad \square$$

In optimization problems where the variables are positive definite kernels that are invariant under a group action with finitely many orbits, Theorem 3.4.4 can be useful in showing that the optimal values of the approximations obtained by optimizing over truncated Fourier series converge to the optimal value of the original problem. However, if there are infinitely many orbits, then in general the Fourier series of a positive definite kernel does not converge uniformly. Instead we construct a hierarchy

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

of Γ-invariant cones where each C_d is isomorphic to a finite product of complex positive semidefinite cones. We show that $\cup_{d=0}^\infty C_d$ is dense in $\mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$ in the uniform topology, and this can be used in showing that optimization problems obtained by approximating $\mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$ by C_d become arbitrarily good as $d \to \infty$. A similar theorem is proved in [**6**] but there it is required that the group acting on the space is contained in a bigger group that has a transitive action. By using the symmetry adapted basis from the previous section we have no such requirement.

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 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 complex positive semidefinite matrices of size $|R_{\pi,d}|$. Let C_d be the Minkowski sum $\sum_{\pi \in \hat{\Gamma}} C_{\pi,d}$. Then we have

$$C_0 \subseteq C_1 \subseteq \ldots \subseteq \mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}.$$

THEOREM 3.4.5. *The cone $\bigcup_{d=0}^\infty C_d$ is uniformly dense in $\mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$.*

PROOF. By Lemma 3.3.2 there exists a Hermitian, Γ-invariant approximate identity $\{I_n\}$ of X, and by Theorem 3.3.5 there exists a symmetry adapted system $\{e_{\pi,i,j}\}$ of X. Let

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

Then $S_0 \subseteq S_1 \subseteq \ldots \subseteq \mathcal{C}_{\mathbb{C}}(X)$ and $\cup_{d=0}^{\infty} S_d$ is uniformly dense in $\mathcal{C}_{\mathbb{C}}(X)$, so for each n there exists a sequence $\{I_{n,d}\}_d$ of real-valued kernels converging uniformly to I_n with $I_{n,d} \in S_d \otimes S_d$. We may assume the kernels $I_{n,d}$ to be Γ-invariant: If $\|I_{n,d} - I_n\|_\infty \to 0$ as $d \to \infty$, then also $\|\bar{I}_{n,d} - I_n\|_\infty \to 0$, where $\bar{I}_{n,d}(x,y) = \int I_{n,d}(\gamma x, \gamma y)\, d\gamma$ is defined by integrating over the normalized Haar measure of Γ. Moreover, since S_d is Γ-invariant, we have $\bar{I}_{n,d} \in S_d \otimes S_d$.

Let $J_n = I_{n,d}$ for some d for which $\|I_{n,d} - I_n\|_\infty \leq 1/n$. Then $\{J_n\}$ is a Hermititan, Γ-invariant approximate identity of X which satisfies $J_n \in S_n \otimes S_n$ for all n. For each n we define $K_n \in \mathcal{C}_{\mathbb{C}}(Y \times Y)$, where $Y = X \times X$, by

$$K_n((x,y),(x',y')) = J_n(x,x')\overline{J_n(y,y')} \quad \text{for} \quad (x,x'),(y,y') \in Y.$$

Let $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$. We will show that $\{T_{K_n} K\}$ converges uniformly to K and that $T_{K_n} K$ is contained in C_n. For $f, g \in \mathcal{C}_{\mathbb{C}}(X)$ we have

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

and since the span of kernels of the form $f \otimes g$ is uniformly dense in $\mathcal{C}_{\mathbb{C}}(X \times X)$, $T_{K_n} K$ converges uniformly to K.

There exist $k \in \mathbb{N}$ and $h_1, \ldots, h_k, h'_1, \ldots, h'_k \in S_n$ for which

$$J_n(x,x') = \sum_{i=1}^{k} h_k(x) h'_k(x'),$$

so by defining

$$c_{i,j} = \iint K(x',y') h'_i(x') \overline{h'_j(y')}\, d\mu(x') d\mu(y')$$

we have

$$T_{K_n} K(x,y) = \sum_{i,j=1}^{k} c_{i,j} h_i \otimes \overline{h_j}(x,y) \in S_n \otimes S_n.$$

By Γ-invariance of K, μ, and J_n, the kernels $T_{K_n} K$ are Γ-invariant. By Lemma 3.4.2 they are also also positive definite: For $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{C}$, and $x_1, \ldots, x_n \in X$,

$$\sum_{i,j=1}^{n} c_i \overline{c_j}\, T_{K_n} K(x_i, x_j) = \iint f(x') \overline{f(y')} K(x',y')\, d\mu(x') d\mu(y') \geq 0,$$

where $f(x) = \sum_{i=1}^{n} c_i J_m(x_i, x)$. □

Until now we have only considered complex-valued kernels because complex representation theory is more natural and simpler than real representation theory. In optimization, however, we are usually concerned with real-valued kernels and we prefer to work with real semidefinite cones instead of complex semidefinite cones.

Here we modify the sequence $\{C_d\}$ defined above to a sequence of inner approximating cones of the cone $\mathcal{C}(X \times X)^{\Gamma}_{\succeq 0}$ of real-valued, Γ-invariant, positive definite kernels, where each cone is isomorphic to a finite direct product of real (instead of complex) positive semidefinite cones. If the zonal matrices Z_π are real-valued, then the fourier coefficients $\hat{K}(\pi)$ of a real-valued kernel K are real-valued. If this is the case, we can simply define $C_{\pi,d}$ as the cone containing kernels of the form $(x, y) \mapsto \langle A, Z_{\pi,d}(x, y)^{\mathsf{T}} \rangle$, where A ranges over the real positive semidefinite matrices, and $\{C_d\}$ becomes the desired inner approximating sequence of $\mathcal{C}(X \times X)^{\Gamma}_{\succeq 0}$. If all irreducible representations π of Γ are of real type, that is, if each π is equivalent to a representation $\Gamma \to O(d_\pi) \subseteq U(d_\pi)$, then we can always construct the symmetry adapted system in such a way that the zonal matrices are real-valued. In this thesis we only consider groups where all representations are of real type, and all symmetry adapted systems are constructed so that the zonal matrices are real-valued. If $\hat{\Gamma}$ also contains representations of complex or quaternionic type (these are the two remaining possibilities), then we should construct a *real* symmetry adapted system, where π ranges over the real irreducible representations. See [**37**] or [**91**] where this is discussed for finite groups.

CHAPTER 4

Upper bounds for packings of spheres of several radii

This chapter is based on the publication *"D. de Laat, F.M. de Oliveira Filho, F. Vallentin, Upper bounds for packings of spheres of several radii, Forum Math. Sigma 2 (2014), e23 (42 pages)."*

Abstract. We give theorems that can be used to upper bound the densities of packings of different spherical caps in the unit sphere and of translates of different convex bodies in Euclidean space. These theorems extend the linear programming bounds for packings of spherical caps and of convex bodies through the use of semidefinite programming. We perform explicit computations, obtaining new bounds for packings of spherical caps of two different sizes and for binary sphere packings. We also slightly improve bounds for the classical problem of packing identical spheres.

4.1. Introduction

How densely can one pack given objects into a given container? Problems of this sort, generally called *packing problems*, are fundamental problems in geometric optimization.

An important example having a rich history is the sphere packing problem. Here one tries to place equal-sized spheres with pairwise disjoint interiors into n-dimensional Euclidean space while maximizing the fraction of covered space. In two dimensions the best packing is given by placing open disks centered at the points of the hexagonal lattice. In three dimensions, the statement that the best sphere packing has density $\pi/\sqrt{18} = 0.7404\ldots$ was known as Kepler's conjecture; it was proved by Hales [**44**] in 1998 by means of a computer-assisted proof.

Currently, one of the best methods for obtaining upper bounds for the density of sphere packings is due to Cohn and Elkies [**21**]. In 2003 they used linear programming to obtain the best known upper bounds for the densities of sphere packings in dimensions $4, \ldots, 36$. They almost closed the gap between lower and upper bounds in dimensions 8 and 24. Their method is the noncompact version of the linear programming method of Delsarte, Goethals, and Seidel [**27**] for upper-bounding the densities of packings of spherical caps on the unit sphere.

From a physical point of view, packings of spheres of different sizes are relevant as they can be used to model chemical mixtures which consist of multiple atoms or, more generally, to model the structure of composite material. For more about

technological applications of these kind of systems of polydisperse, totally impenetrable spheres we refer to Torquato [**96**, Chapter 6]. In recent work, Hopkins, Jiao, Stillinger, and Torquato [**52**, **53**] presented lower bounds for the densities of packings of spheres of two different sizes, also called *binary sphere packings*.

In coding theory, packings of spheres of different sizes are important in the design of error-correcting codes that can be used for unequal error protection. Masnick and Wolf [**73**] were the first who considered codes with this property.

In this paper we extend the linear programming method of Cohn and Elkies to obtain new upper bounds for the densities of multiple-size sphere packings. We also extend the linear programming method of Delsarte, Goethals, and Seidel to obtain new upper bounds for the densities of multiple-size spherical cap packings.

We perform explicit calculations for binary packings in both cases using semidefinite, instead of linear, programming. In particular we complement the constructive lower bounds of Hopkins, Jiao, Stillinger, and Torquato by non-constructive upper bounds. Insights gained from our computational approach are then used to improve known upper bounds for the densities of monodisperse sphere packings in dimensions 4, ... 9, except 8. The bounds we present improve on the best-known bounds due to Cohn and Elkies [**21**].

4.1.1. Methods and theorems. We model the packing problems using tools from combinatorial optimization. All possible positions of the objects that we can use for the packing are vertices of a graph and we draw edges between two vertices whenever the two corresponding objects cannot be simultaneously present in the packing because they overlap in their interiors. Now every independent set in this conflict graph gives a valid packing and vice versa. To determine the density of the packing we use vertex weights since we want to distinguish between "small" and "big" objects. For finite graphs it is known that the weighted independence number can be upper bounded by the weighted theta number. Our theorems for packings of spherical caps and spheres are infinite-dimensional analogues of this result.

Let $G = (V, E)$ be a finite graph. A set $I \subseteq V$ is *independent* if no two vertices in I are adjacent. Given a weight function $w\colon V \to \mathbb{R}_{\geq 0}$, the *weighted independence number* of G is the maximum weight of an independent set, i.e.,

$$\alpha_w(G) = \max\Big\{\sum_{x\in I} w(x) : I \subseteq V \text{ is independent}\Big\}.$$

Finding $\alpha_w(G)$ is an NP-hard problem.

Grötschel, Lovász, and Schrijver [**43**] defined a graph parameter that gives an upper bound for α_w and which can be computed efficiently by semidefinite optimization. It can be presented in many different, yet equivalent ways, but the one convenient for us is

$$\begin{aligned}
\vartheta'_w(G) = \min \quad & M & \\
& K - (w^{1/2})(w^{1/2})^{\mathsf{T}} & \text{is positive semidefinite,} \\
& K(x,x) \leq M & \text{for all } x \in V, \\
& K(x,y) \leq 0 & \text{for all } \{x,y\} \notin E \text{ where } x \neq y, \\
& M \in \mathbb{R},\ K \in \mathbb{R}^{V\times V} & \text{is symmetric.}
\end{aligned}$$

Here we give a proof of the fact that $\vartheta'_w(G)$ upper bounds $\alpha_w(G)$. In a sense, after discarding the analytical arguments in the proofs of Theorems 4.1.2 and 4.1.3, we are left with this simple proof.

THEOREM 4.1.1. *For any finite graph $G = (V, E)$ with weight function $w\colon V \to \mathbb{R}_{\geq 0}$ we have $\alpha_w(G) \leq \vartheta'_w(G)$.*

PROOF. Let $I \subseteq V$ be an independent set of nonzero weight and let $K \in \mathbb{R}^{V \times V}$, $M \in \mathbb{R}$ be a feasible solution of $\vartheta'_w(G)$. Consider the sum

$$\sum_{x,y \in I} w(x)^{1/2} w(y)^{1/2} K(x, y).$$

This sum is at least

$$\sum_{x,y \in I} w(x)^{1/2} w(y)^{1/2} w(x)^{1/2} w(y)^{1/2} = \left(\sum_{x \in I} w(x) \right)^2$$

because $K - (w^{1/2})(w^{1/2})^{\mathsf{T}}$ is positive semidefinite.

The sum is also at most

$$\sum_{x \in I} w(x) K(x, x) \leq M \sum_{x \in I} w(x)$$

because $K(x, x) \leq M$ and because $K(x, y) \leq 0$ whenever $x \neq y$ as I forms an independent set. Now combining both inequalities proves the theorem. □

Multiple-size spherical cap packings. We first consider packings of spherical caps of several radii on the unit sphere $S^{n-1} = \{\, x \in \mathbb{R}^n : x \cdot x = 1 \,\}$. The spherical cap with angle $\alpha \in [0, \pi]$ and center $x \in S^{n-1}$ is given by

$$C(x, \alpha) = \{\, y \in S^{n-1} : x \cdot y \geq \cos \alpha \,\}.$$

Its normalized volume equals

$$w(\alpha) = \frac{\omega_{n-1}(S^{n-2})}{\omega_n(S^{n-1})} \int_{\cos \alpha}^{1} (1 - u^2)^{(n-3)/2} \, du,$$

where $\omega_n(S^{n-1}) = (2\pi^{n/2})/\Gamma(n/2)$ is the surface area of the unit sphere. Two spherical caps $C(x_1, \alpha_1)$ and $C(x_2, \alpha_2)$ intersect in their topological interiors if and only if the inner product of x_1 and x_2 lies in the interval $(\cos(\alpha_1 + \alpha_2), 1]$. Conversely we have

$$C(x_1, \alpha_1)^\circ \cap C(x_2, \alpha_2)^\circ = \emptyset \quad \iff \quad x_1 \cdot x_2 \leq \cos(\alpha_1 + \alpha_2).$$

A *packing* of spherical caps with angles $\alpha_1, \ldots, \alpha_N$ is a union of any number of spherical caps with these angles and pairwise-disjoint interiors. The density of the packing is the sum of the normalized volumes of the constituting spherical caps.

The optimal packing density is given by the weighted independence number of the *spherical cap packing graph*. This is the graph with vertex set $S^{n-1} \times \{1, \ldots, N\}$, where a vertex (x, i) has weight $w(\alpha_i)$, and where two distinct vertices (x, i) and (y, j) are adjacent if $\cos(\alpha_i + \alpha_j) < x \cdot y$.

In Section 4.2 we will extend the weighted theta prime number to the spherical cap packing graph. There we will also derive Theorem 4.1.2 below, which gives

upper bounds for the densities of packings of spherical caps. We will show that the sharpest bound given by this theorem is in fact equal to the theta prime number.

In what follows we denote by P_k^n the Jacobi polynomial $P_k^{((n-3)/2,(n-3)/2)}$ of degree k, normalized so that $P_k^n(1) = 1$. Jacobi polynomials are orthogonal polynomials defined on the interval $[-1, 1]$ with respect to the measure $(1-u^2)^{(n-3)/2}du$. See for instance Andrews, Askey, and Roy [**2**] for more information.

THEOREM 4.1.2. *Let $\alpha_1, \ldots, \alpha_N \in (0, \pi]$ be angles and for $i, j = 1, \ldots, N$ and $k \geq 0$ let $f_{ij,k}$ be real numbers such that $f_{ij,k} = f_{ji,k}$ and $\sum_{k=0}^{\infty} |f_{ij,k}| < \infty$ for all i, j. Write*

$$f_{ij}(u) = \sum_{k=0}^{\infty} f_{ij,k} P_k^n(u). \tag{1}$$

Suppose the functions f_{ij} satisfy the following conditions:

(i) $\left(f_{ij,0} - w(\alpha_i)^{1/2} w(\alpha_j)^{1/2}\right)_{i,j=1}^N$ *is positive semidefinite;*

(ii) $\left(f_{ij,k}\right)_{i,j=1}^N$ *is positive semidefinite for $k \geq 1$;*

(iii) $f_{ij}(u) \leq 0$ *whenever $-1 \leq u \leq \cos(\alpha_i + \alpha_j)$.*

Then the density of every packing of spherical caps with angles $\alpha_1, \ldots, \alpha_N$ on the unit sphere S^{n-1} is at most $\max\{ f_{ii}(1) : i = 1, \ldots, N \}$.

When $N = 1$, Theorem 4.1.2 reduces to the linear programming bound for spherical cap packings of Delsarte, Goethals, and Seidel [**27**]. In Section 4.4 we use semidefinite programming instead of linear programming to perform explicit computations for $N = 2$.

Translational packings of bodies and multiple-size sphere packings. We now deal with packings of spheres with several radii in $\mathbb{R}^n$. Theorem 4.1.3 presented below can be used to find upper bounds for the densities of such packings. In fact, it is more general and can be applied to packings of translates of different convex bodies.

Let $\mathcal{K}_1, \ldots, \mathcal{K}_N$ be convex bodies in $\mathbb{R}^n$. A translational *packing* of $\mathcal{K}_1, \ldots, \mathcal{K}_N$ is a union of translations of these bodies in which any two copies have disjoint interiors. The *density* of a packing is the fraction of space covered by it. There are different ways to formalize this definition, and questions appear as to whether every packing has a density and so on. We postpone further discussion on this matter until Section 4.3 where we give a proof of Theorem 4.1.3.

Our theorem can be seen as an analogue of the weighted theta prime number ϑ'_w for the infinite graph G whose vertex set is $\mathbb{R}^n \times \{1, \ldots, N\}$ and in which vertices (x, i) and (y, j) are adjacent if $x + \mathcal{K}_i$ and $y + \mathcal{K}_j$ have disjoint interiors. The weight function we consider assigns weight $\operatorname{vol} \mathcal{K}_i$ to vertex $(x, i) \in \mathbb{R}^n \times \{1, \ldots, N\}$. We will say more about this interpretation in Section 4.3.

For the statement of the theorem we need some basic facts from harmonic analysis. Let $f \colon \mathbb{R}^n \to \mathbb{C}$ be an L^1 function. For $u \in \mathbb{R}^n$, the Fourier transform of f at u is

$$\hat{f}(u) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i u \cdot x} \, dx.$$

We say that function f is a *Schwartz function* (also called a *rapidly-decreasing function*) if it is infinitely differentiable, and if any derivative of f, multiplied by any power of the variables $x_1, \ldots, x_n$, is a bounded function. The Fourier transform of a Schwartz function is a Schwartz function, too. A Schwartz function can be recovered from its Fourier transform by means of the *inversion formula*:

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(u) e^{2\pi i u \cdot x}\, du$$

for all $x \in \mathbb{R}^n$.

THEOREM 4.1.3. *Let $\mathcal{K}_1, \ldots, \mathcal{K}_N$ be convex bodies in $\mathbb{R}^n$ and let $f\colon \mathbb{R}^n \to \mathbb{R}^{N\times N}$ be a matrix-valued function whose every component f_{ij} is a Schwartz function. Suppose f satisfies the following conditions:*

(i) *the matrix $\big(\hat{f}_{ij}(0) - (\operatorname{vol}\mathcal{K}_i)^{1/2}(\operatorname{vol}\mathcal{K}_j)^{1/2}\big)_{i,j=1}^{N}$ is positive semidefinite;*

(ii) *the matrix of Fourier transforms $\big(\hat{f}_{ij}(u)\big)_{i,j=1}^{N}$ is positive semidefinite for every $u \in \mathbb{R}^n \setminus \{0\}$;*

(iii) *$f_{ij}(x) \le 0$ whenever $\mathcal{K}_i^\circ \cap (x + \mathcal{K}_j^\circ) = \emptyset$.*

Then the density of any packing of translates of $\mathcal{K}_1, \ldots, \mathcal{K}_N$ in the Euclidean space $\mathbb{R}^n$ is at most $\max\{\, f_{ii}(0) : i = 1, \ldots, N \,\}$.

We give a proof of this theorem in Section 4.3. When $N = 1$ and when the convex body $\mathcal{K}_1$ is centrally symmetric (an assumption that is in fact not needed) then this theorem reduces to the linear programming method of Cohn and Elkies [**21**].

We apply this theorem to obtain upper bounds for the densities of binary sphere packings, as we discuss in Section 4.1.3.

4.1.2. Computational results for binary spherical cap packings. We applied Theorem 4.1.2 to compute upper bounds for the densities of binary spherical cap packings. The results we obtained are summarized in the plots of Figure 1.

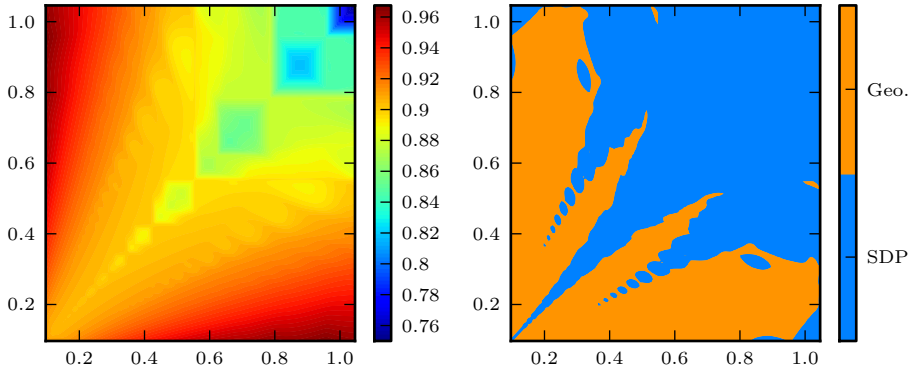
For $n = 3$, Florian [**30**, **31**] provides a geometric upper bound for the density of a spherical cap packing. He shows that the density of a packing on S^2 of spherical caps with angles $\alpha_1, \ldots, \alpha_N \in (0, \pi/3]$ is at most

$$\max_{1\le i\le j\le k\le N} D(\alpha_i, \alpha_j, \alpha_k),$$

where $D(\alpha_i, \alpha_j, \alpha_k)$ is defined as follows: Let $\mathcal{T}$ be a spherical triangle in S^2 such that if we center the spherical caps with angles α_i, α_j, and α_k at the vertices of $\mathcal{T}$, then the caps intersect pairwise at their boundaries. The number $D(\alpha_i, \alpha_j, \alpha_k)$ is then defined as the fraction of the area of $\mathcal{T}$ covered by the caps.

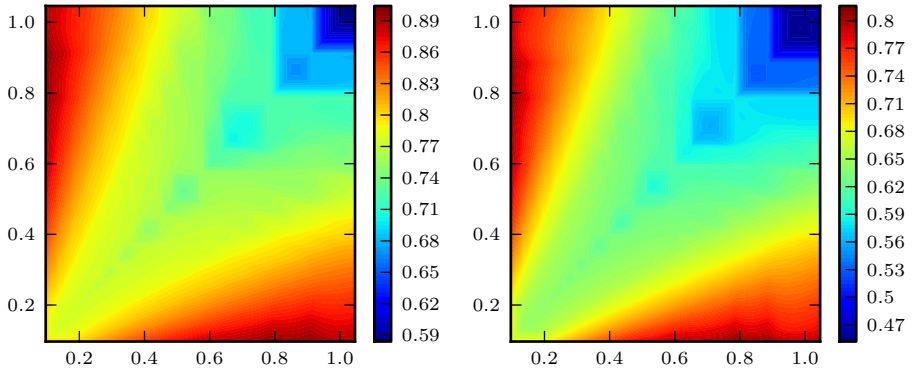
In Figure 1 (B) we see that for $N = 2$ it depends on the angles whether the geometric or the semidefinite programming bound is sharper. In particular we see that near the diagonal the semidefinite programming bound is at least as good as the geometric bound; see also Figure 1 (A).

We can construct natural multiple-size spherical cap packings by taking the incircles of the faces of spherical Archimedean tilings. A sequence of binary packings is for instance obtained by taking the incircles of the prism tilings. These are the Archimedean tilings with vertex figure $(4, 4, m)$ for $m \ge 3$ (although strictly



(A) $n = 3$

(B) SDP bound/geometric bound for $n = 3$



(C) $n = 4$

(D) $n = 5$

FIGURE 1. Upper bounds on the packing density for $N = 2$. The horizontal and vertical axes carry the spherical cap angle; the colors indicate the density, or in the case of plot (B) whether the SDP bound or the geometric bound is sharper.

speaking for $m = 4$ this is a spherical Platonic tiling). The question then is whether the packing associated with the m-prism has maximal density among all packings with the same cap angles π/m and $\pi/2 - \pi/m$, that is, whether the packing is *maximal*. The packing for $m = 3$ is not maximal while the one for $m = 4$ trivially is, since here there is only one cap size, and adding a 9th cap yields a density greater than 1.

Heppes and Kertész [**48**] showed that the configurations for $m \geq 6$ are maximal, and the remaining case $m = 5$ was later shown to maximal by Florian and Heppes [**32**]. Florian [**30**] showed that the geometric bound given above is in fact

sharp for the cases where $m \geq 6$, and for the $m = 5$ case it is not sharp but still good enough to prove maximality (notice that given a finite number of cap angles, the set of obtainable densities is finite).

Now we illustrate that Theorem 4.1.2 gives a sharp bound for the density of the packing associated to the 5-prism, thus giving a simple proof of its maximality. The theorem also provides a sharp bound for $m = 4$ but whether it can provide sharp bounds for the cases $m \geq 6$ we do not know at the moment. The numerical results are not decisive.

We shall exhibit functions

$$f_{ij}(u) = \sum_{k=0}^{4} f_{ij,k} P_k^n(u)$$

that satisfy the conditions of Theorem 4.1.2 with $f_{11}(1) = f_{22}(1) = 5w(\alpha_1)+2w(\alpha_2)$ where

$$\alpha_1 = \frac{\pi}{5},\ \alpha_2 = \frac{3\pi}{10},\ w(\alpha_1) = \frac{1}{2}\left(1 - \cos\frac{\pi}{5}\right),\ w(\alpha_2) = \frac{1}{2}\left(1 - \cos\frac{3\pi}{10}\right).$$

For a sharp solution, the inequalities in the proof of Theorem 4.1.2 must be equalities, so the $f_{ij,k}$ have to satisfy the following linear conditions:

$$0 = f_{11}\left(\cos\frac{2\pi}{5}\right) = f_{11}\left(\cos\frac{4\pi}{5}\right) = f'_{11}\left(\cos\frac{4\pi}{5}\right) = f_{12}(0) = f_{22}(-1);$$

the product

$$\begin{pmatrix} f_{11,0} & f_{12,0} \\ f_{12,0} & f_{22,0} \end{pmatrix} \begin{pmatrix} 25w(\alpha_1) & 10\sqrt{w(\alpha_1)w(\alpha_2)} \\ 10\sqrt{w(\alpha_1)w(\alpha_2)} & 4w(\alpha_2) \end{pmatrix}$$

equals

$$\begin{pmatrix} 25w(\alpha_1)^2 + 10w(\alpha_1)w(\alpha_2) & \sqrt{w(\alpha_1)w(\alpha_2)}(10w(\alpha_1) + 4w(\alpha_2)) \\ \sqrt{w(\alpha_1)w(\alpha_2)}(25w(\alpha_1) + 10w(\alpha_2)) & 10w(\alpha_1)w(\alpha_2) + 4w(\alpha_2)^2 \end{pmatrix};$$

for $k = 1, \ldots, 4$ the product of the two matrices $\begin{pmatrix} f_{11,k} & f_{12,k} \\ f_{12,k} & f_{22,k} \end{pmatrix}$ and

$$\begin{pmatrix} w(\alpha_1)(5P_k(1) + 10P_k(\cos\frac{2\pi}{5}) + 10P_k(\cos\frac{2\pi}{4})) & \sqrt{w(\alpha_1)w(\alpha_2)}10P_k(0) \\ \sqrt{w(\alpha_1)w(\alpha_2)}10P_k(0) & w(\alpha_2)(2P_k(1) + 2P_k(-1)) \end{pmatrix}$$

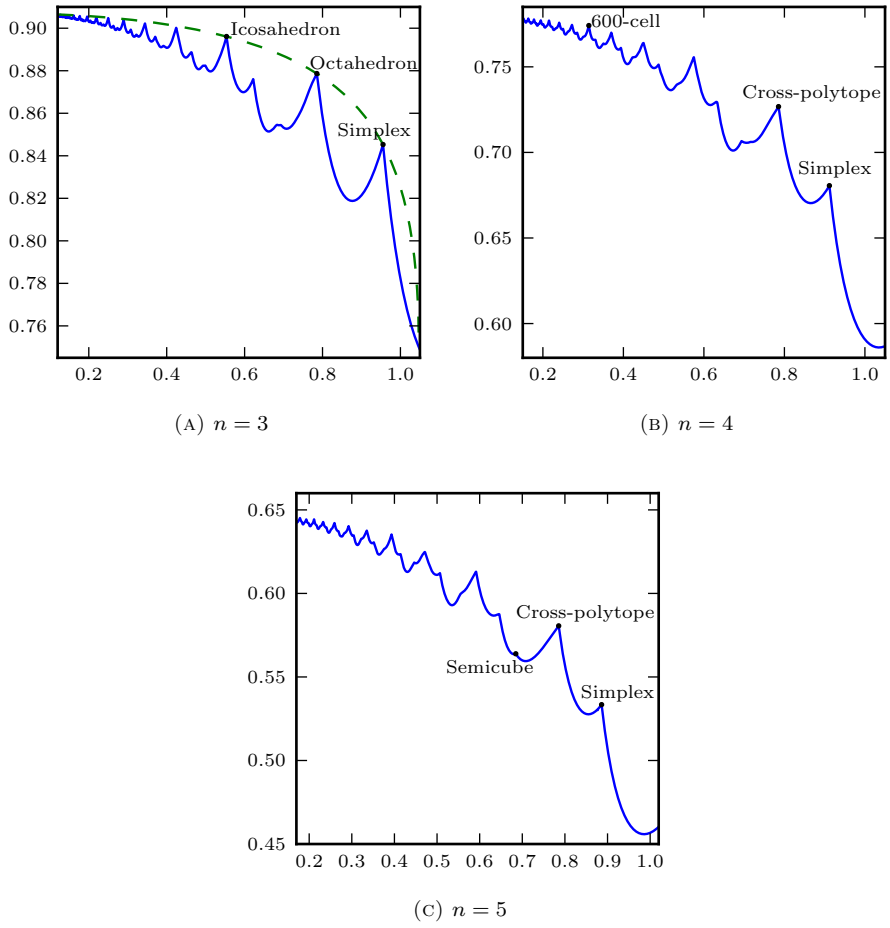
equals zero. This linear system together with the additional assumptions

$$0 = f_{11}(-1) = f_{12}\left(-\frac{95}{100}\right) = f'_{12}\left(-\frac{95}{100}\right)$$

has a one-dimensional space of solutions from which it is easy to select one that fulfills all requirements of Theorem 4.1.2.

For the remaining 13 Archimedean solids in dimension $n = 3$ we are only able to show maximality of the packing associated to the truncated octahedron, the Archimedean solid with vertex figure $(6, 6, 5)$. Its density is $0.9056\ldots$, the geometric bound shows that the density is at most $0.9088\ldots$, and using the semidefinite program we get $0.9079\ldots$ as an upper bound. The first packing with caps of angles $\arcsin(1/3)$ and $\arcsin(1/\sqrt{3})$ that would be denser is obtained by taking 19 of the

smaller caps and 4 of the bigger caps, and has density 0.9103... The upper bounds show however that it is not possible to obtain this dense a packing, thus showing that the truncated octahedron packing is maximal.



(A) $n = 3$

(B) $n = 4$

(C) $n = 5$

FIGURE 2. Upper bounds on the packing density for $N = 1$, the horizontal axis represents the spherical cap angle and the vertical axis the packing density.

We also used our programs to plot the upper bounds for $N = 1$, the classical linear programming bound of Delsarte, Goethals, and Seidel [**27**], for dimensions $n = 3$, 4, and 5 in Figure 2. To the best of our knowledge these kinds of plots were not made before and they seem to reveal interesting properties of the bound. For better orientation we show in the plots the packings where the linear programming bound is sharp (cf. Levenshtein [**68**]; Cohn and Kumar [**22**] proved the much

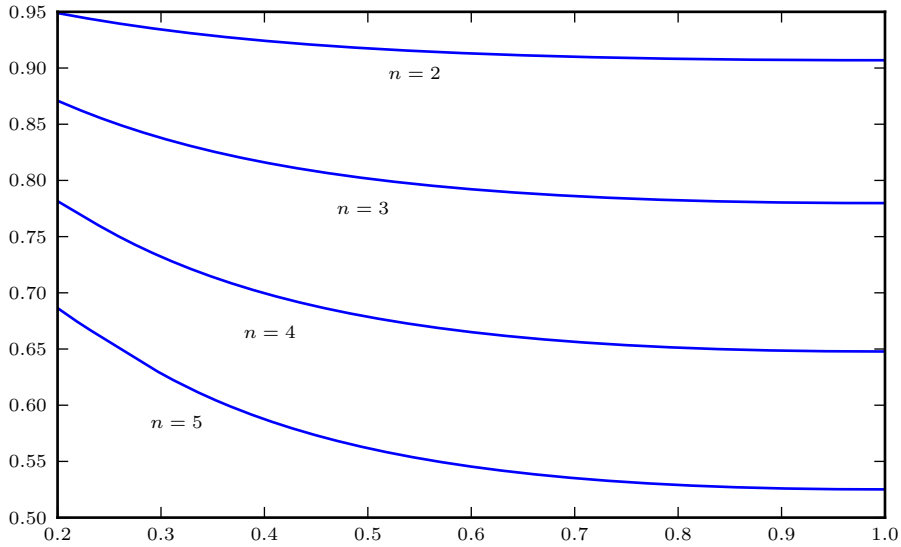


FIGURE 3. The horizontal axis carries the ratio between the radii of the small and the large spheres. The vertical axis carries our upper bound. Our bounds for dimensions 2, ..., 5 are shown together.

stronger statement that these packings provide point configurations that are universally optimal). The dotted line in the plot for $n = 3$ is the geometric bound, and since we know that both the geometric (cf. Florian [**30**]) and the semidefinite programming bounds are sharp for the given configurations, we know that at these peaks the bounds meet.

An interesting feature of the upper bound seems to be that it has some periodic behavior. Indeed, the numerical results suggest that for $n = 3$, the two bounds in fact meet infinitely often as the angle decreases, and that between any two of these meeting points the semidefinite programming bound has a similar shape. Although in higher dimensions we do not have a geometric bound, the semidefinite programming bound seems to admit the same kind of periodic behavior.

4.1.3. Computational results for binary sphere packings. We applied Theorem 4.1.3 to compute upper bounds for the densities of binary sphere packings. The results we obtained are summarized in the plot of Figure 3, where we show bounds computed for dimensions 2, ..., 5. A detailed account of our approach is given in Section 4.5. We now quickly discuss the bounds presented in Figure 3.

Dimension 2. Only in dimension 2 have binary sphere (i.e., circle) packings been studied in depth. We refer to the introduction in the paper of Heppes [**47**] which surveys the known results about binary circle packings in the plane.

Currently one of the best-known upper bounds for the maximum density of a binary circle packing is due to Florian [**29**]. Florian's bound states that a packing

of circles in which the ratio between the radii of the smallest and largest circles is r has density at most

$$\frac{\pi r^2 + 2(1-r^2)\arcsin(r/(1+r))}{2r\sqrt{2r+1}},$$

and that this bound is achieved exactly for $r = 1$ (i.e., for classical circle packings) and for $r = 0$ in the limit.

The question arises of which bound is better, our bound or Florian's bound. From our experiments, it seems that our bound is worse than Florian's bound, at least for $r < 1$. For instance, for $r = 1/2$ we obtain the upper bound $0.9174426\ldots$, whereas Florian's bound is $0.9158118\ldots$ Whether this really means that the bound of Theorem 4.1.3 is worse than Florian's bound, or just that the computational approach of Section 4.5 is too restrictive to attain his bound, we do not know.

It is interesting to note that for $r = 1$, that is, for packings of circles of one size, our bound clearly coincides with the one of Cohn and Elkies [**21**]. This bound seems to be equal to $\pi/\sqrt{12}$, but no proof of this is known.

Dimension 3. Much less is known in dimension 3. In fact we do not know about other attempts to find upper bounds for the densities of binary sphere packings in dimensions 3 and higher.

Let us compare our upper bound with the lower bound by Hopkins, Jiao, Stillinger, and Torquato [**52**]. The record holder for $r \geq 0.2$ in terms of highest density occurs for $r = 0.224744\ldots$ and its density is $0.824539\ldots$ Our computations show that there cannot be a packing with this r having density more than $0.8617125\ldots$, so this leaves a margin of 5%.

Another interesting case is $r = \sqrt{2} - 1 = 0.414\ldots$ Here the best-known lower bound of $0.793\ldots$ comes from the crystall structure of sodium chloride NaCl. The large spheres are centered at a face centered cubic lattice and the small spheres are centered at a translated copy of the face centered cubic lattice so that they form a jammed packing. Our upper bound for $r = \sqrt{2} - 1$ is $0.813\ldots$, less than 3% away from the lower bound. Therefore, we believe that proving optimality of the NaCl packing might be doable.

Dimension 4 and beyond. In higher dimensions even less is known about binary sphere packings. We observed from Figure 3 that it seems that the upper bound is decreasing: as the radius of the small sphere increases from 0.2 to 1, the bound seems to decrease. This suggests that the bound given by Theorem 4.1.3 is decreasing in this sense, but we do not know a proof of this.

We also do not know the limit behavior of our bound when r approaches 0. Due to numerical instabilities we could not perform numerical calculations in this regime of r.

4.1.4. Improving the Cohn-Elkies bounds. We now present a theorem that can be used to find better upper bounds for the densities of monodisperse sphere packings than those provided by Cohn and Elkies [**21**]; our theorem is a strengthening of theirs.

Fix $\varepsilon > 0$. Given a packing of spheres of radius $1/2$, we consider its *ε-tangency graph*, a graph whose vertices are the spheres in the packing, and in which two

vertices are adjacent if the distance between the centers of the respective spheres lies in the interval $[1, 1+\varepsilon)$.

Let $M(\varepsilon)$ be the least upper bound on the average degree of the ε-tangency graph of any sphere packing. Our theorem is the following:

THEOREM 4.1.4. *Take $0 = \varepsilon_0 < \varepsilon_1 < \cdots < \varepsilon_m$ and let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a Schwartz function such that*

(i) $\hat{f}(0) \geq \operatorname{vol} B$, *where B is the ball of radius $1/2$;*
(ii) $\hat{f}(u) \geq 0$ *for all $u \in \mathbb{R}^n \setminus \{0\}$;*
(iii) $f(x) \leq 0$ *whenever $\|x\| \geq 1 + \varepsilon_m$;*
(iv) $f(x) \leq \eta_k$ *whenever $\|x\| \in [1+\varepsilon_{k-1}, 1+\varepsilon_k)$ with $\eta_k \geq 0$, for $k \in [m]$.*

Then the density of a sphere packing is at most the optimal value of the following linear programming problem in variables $A_1, \ldots, A_m$:

$$\begin{array}{lll} \max & f(0) + \eta_1 A_1 + \cdots + \eta_m A_m & \\ & A_1 + \cdots + A_k \leq U(\varepsilon_k) & \text{for } k \in [m], \\ & A_k \geq 0 & \text{for } k \in [m], \end{array} \tag{2}$$

where $U(\varepsilon_k) \geq M(\varepsilon_k)$ for $k = 1, \ldots, m$.

In Section 4.6 we give a proof of Theorem 4.1.4 and show how to compute upper bounds for $M(\varepsilon)$ using the semidefinite programming bounds of Bachoc and Vallentin [**9**] for the sizes of spherical codes. There we also show how to use semidefinite programming and the same ideas we employ in the computations for binary sphere packings (cf. Section 4.5) to compute better upper bounds for the densities of sphere packings.

Dimension	Lower bound	Cohn-Elkies bound	New upper bound
4	0.12500	0.13126	0.130587
5	0.08839	0.09975	0.099408
6	0.07217	0.08084	0.080618
7	0.06250	0.06933	0.069193
9	0.04419	0.05900	0.058951

TABLE 1. For each dimension we show the best lower bound known, the bound by Cohn and Elkies [**21**], and the upper bound coming from Theorem 4.1.4.

In Table 1 we show the upper bounds obtained through our application of Theorem 4.1.4. To better compare our bounds with those of Cohn and Elkies, on Table 1 we show bounds for the center density of a packing, the *center density* of a packing of unit spheres being equal to $\Delta / \operatorname{vol} B$, where Δ is the density of the packing, and B is a unit ball.

We omit dimension 8 because for this dimension it is already believed that the Cohn-Elkies bound is itself optimal, and therefore as is to be expected we did not

manage to obtain any improvement over their bound. We also note that the bounds by Cohn and Elkies are the best known upper bounds in all other dimensions shown.

In dimension 3 the Cohn-Elkies bound is 0.18616 whereas the optimal sphere packing has center density 0.17678. We can improve the Cohn-Elkies bound to 0.184559 which is also better than the upper bound 0.1847 due to Rogers [**83**].

4.2. Multiple-size spherical cap packings

In this section we prove Theorem 4.1.2 and discuss its relation to an extension of the weighted theta prime number for the spherical cap packing graph.

4.2.1. Proof of Theorem 4.1.2. Let $x_1, \ldots, x_m \in S^{n-1}$ and $r\colon \{1, \ldots, m\} \to \{1, \ldots, N\}$ be such that

$$\bigcup_{i=1}^{m} C(x_i, \alpha_{r(i)})$$

is a packing of spherical caps on S^{n-1}.

Consider the sum

$$\sum_{i,j=1}^{m} w(\alpha_{r(i)})^{1/2} w(\alpha_{r(j)})^{1/2} f_{r(i)r(j)}(x_i \cdot x_j). \tag{3}$$

By expanding $f_{r(i)r(j)}(x_i \cdot x_j)$ according to (1) this sum is equal to

$$\sum_{k=0}^{\infty} \sum_{i,j=1}^{m} w(\alpha_{r(i)})^{1/2} w(\alpha_{r(j)})^{1/2} f_{r(i)r(j),k} P_k^n(x_i \cdot x_j).$$

By the addition formula (cf. e.g. Section 9.6 of Andrews, Askey, and Roy [**2**]) for the Jacobi polynomials P_k^n the matrix $\left(P_k^n(x_i \cdot x_j)\right)_{i,j=1}^m$ is positive semidefinite. From condition (ii) of the theorem, we also know that the matrix $\left(f_{r(i)r(j),k}\right)_{i,j=1}^m$ is positive semidefinite for $k \geq 1$. So the inner sum above is nonnegative for $k \geq 1$. If we then consider only the summand for $k = 0$ we see that (3) is at least

$$\sum_{i,j=1}^{m} w(\alpha_{r(i)})^{1/2} w(\alpha_{r(j)})^{1/2} f_{r(i)r(j),0} P_0^n(x_i \cdot x_j) \geq \left(\sum_{i=1}^{m} w(\alpha_i) \right)^2, \tag{4}$$

where the inequality follows from condition (i) of the theorem.

Now, notice that whenever $i \neq j$, the caps $C(x_i, \alpha_{r(i)})$ and $C(x_j, \alpha_{r(j)})$ have disjoint interiors. Condition (iii) then implies that $f_{r(i)r(j)}(x_i \cdot x_j) \leq 0$. So we see that (3) is at most

$$\sum_{i=1}^{m} w(\alpha_i) f_{r(i)r(i)}(1) \leq \max\{\, f_{ii}(1) : i = 1, \ldots, N \,\} \sum_{i=1}^{m} w(\alpha_i). \tag{5}$$

So (3) is at least (4) and at most (5), yielding

$$\sum_{i=1}^{m} w(\alpha_i) \leq \max\{\, f_{ii}(1) : i = 1, \ldots, N \,\}. \qquad \square$$

4.2.2. Theorem 4.1.2 and the Lovász theta number. We now briefly discuss a generalization of ϑ'_w to infinite graphs and its relation to the bound of Theorem 4.1.2. Similar ideas were developed by Bachoc, Nebe, Oliveira, and Vallentin [**8**].

Let $G = (V, E)$ be a graph, where V is a compact space, and let $w\colon V \to \mathbb{R}_{\geq 0}$ be a continuous weight function. An element in the space $\mathcal{C}(V \times V)$ of real-valued continuous functions over $V \times V$ is called a *kernel*. A kernel K is *symmetric* if $K(x, y) = K(y, x)$ for all $x, y \in V$. It is *positive* if it is symmetric and if for any $m \in \mathbb{N}$ and for any $x_1, \ldots, x_m \in V$, the matrix $\big(K(x_i, x_j)\big)_{i,j=1}^{m}$ is positive semidefinite. The *weighted theta prime number* of G is defined as

$$
\begin{aligned}
\vartheta'_w(G) = \inf \quad & M \\
& K - w^{1/2} \otimes (w^{1/2})^* && \text{is a positive kernel,} \\
& K(x, x) \leq M && \text{for all } x \in V, \\
& K(x, y) \leq 0 && \text{for all } \{x, y\} \notin E \text{ where } x \neq y, \\
& M \in \mathbb{R},\ K \in \mathcal{C}(V \times V) && \text{is symmetric.}
\end{aligned} \tag{6}
$$

One may show, mimicking the proof of Theorem 4.1.1, that $\vartheta'_w(G) \geq \alpha_w(G)$.

Let $G = (V, E)$ be the spherical cap packing graph as defined in Section 4.1.1. We will use the symmetry of this graph to show that (6) gives the sharpest bound obtainable by Theorem 4.1.2.

The orthogonal group $\mathrm{O}(n)$ acts on S^{n-1}, and this defines the action of $\mathrm{O}(n)$ on the vertex set $V = S^{n-1} \times \{1, \ldots, N\}$ by $A(x, i) = (Ax, i)$ for $A \in \mathrm{O}(n)$. The group average of a kernel $K \in \mathcal{C}(V \times V)$ is given by

$$\overline{K}((x, i), (y, j)) = \int_{\mathrm{O}(n)} K(A(x, i), A(y, j))\, d\mu(A),$$

where μ is the Haar measure on $\mathrm{O}(n)$ normalized so that $\mu(\mathrm{O}(n)) = 1$. If (K, M) is feasible for (6), then $(\overline{K}, M)$ is feasible too. This follows since for each $A \in \mathrm{O}(n)$, a point (x, i) has the same weight as $A(x, i)$, and two points (x, i) and (y, j) are adjacent if and only if $A(x, i)$ and $A(y, j)$ are adjacent. Since (K, M) and $(\overline{K}, M)$ have the same objective value M, and since $\overline{K}$ is invariant under the action of $\mathrm{O}(n)$, we may restrict to $\mathrm{O}(n)$-invariant kernels (i.e., kernels K such that $K(Au, Av) = K(u, v)$ for all $A \in \mathrm{O}(n)$ and $u, v \in V$) in finding the infimum of (6).

Schoenberg [**85**] showed that a symmetric kernel $K \in \mathcal{C}(S^{n-1} \times S^{n-1})$ is positive and $\mathrm{O}(n)$-invariant if and only if it lies in the cone spanned by the kernels $(x, y) \mapsto P_k^n(x \cdot y)$. We will use the following generalization for kernels over $V \times V$.

THEOREM 4.2.1. *A symmetric kernel $K \in \mathcal{C}(V \times V)$, with $V = S^{n-1} \times \{1, \ldots, N\}$, is positive and $\mathrm{O}(n)$-invariant if and only if*

$$K((x, i), (y, j)) = f_{ij}(x \cdot y) \tag{7}$$

with

$$f_{ij}(u) = \sum_{k=0}^{\infty} f_{ij,k} P_k^n(u),$$

where $\big(f_{ij,k}\big)_{i,j=1}^{N}$ is positive semidefinite for all $k \geq 0$ and $\sum_{k=0}^{\infty} |f_{ij,k}| < \infty$ for all $i, j = 1, \ldots, N$, implying in particular that we have uniform convergence above.

Before we prove the theorem we apply it to simplify problem (6). If K is an $\mathrm{O}(n)$-invariant feasible solution of (6), then $K - w^{1/2} \otimes (w^{1/2})^*$ is a positive $\mathrm{O}(n)$-invariant kernel, and hence can be written in the form (7). Using in addition that $P_0^n = 1$, problem (6) reduces to

$$\begin{aligned} \vartheta'_w(G) = \inf \quad & M \\ & f_{ii}(0) + w(\alpha_i) \le M && \text{for all } 1 \le i \le N, \\ & f_{ij}(u) + (w(\alpha_i)w(\alpha_j))^{1/2} \le 0 && \text{when } -1 \le u \le \cos(\alpha_i + \alpha_j), \\ & M \in \mathbb{R} \text{ and } \big(f_{ij,k}\big)_{i,j=1}^N \text{ positive semidefinite for all } k \ge 0. \end{aligned}$$

By substituting $f_{ij,0} - (w(\alpha_i)w(\alpha_j))^{1/2}$ for $f_{ij,0}$ we see that the solution to this problem indeed equals the sharpest bound given by Theorem 4.1.2.

Proof of Theorem 4.2.1. If we endow the space $\mathcal{C}(S^{n-1})$ of real-valued continuous function on the unit sphere S^{n-1} with the usual L^2 inner product, then for f, $g \in \mathcal{C}(V)$,

$$\langle f, g \rangle = \sum_{i=1}^N \int_{S^{n-1}} f(x,i)g(x,i)\, d\omega(x)$$

gives an inner product on $\mathcal{C}(V)$. The space $\mathcal{C}(S^{n-1})$ decomposes orthogonally as

$$\mathcal{C}(S^{n-1}) = \bigoplus_{k=0}^\infty H_k,$$

where H_k is the space of homogeneous harmonic polynomials of degree k restricted to S^{n-1}. With

$$H_{k,i} = \{\, f \in \mathcal{C}(V) : \text{there is a } g \in H_k \text{ such that } f(\cdot, j) = \delta_{ij} g(\cdot) \,\},$$

it follows that $\mathcal{C}(V)$ decomposes orthogonally as

$$\mathcal{C}(V) = \bigoplus_{k=0}^\infty \bigoplus_{i=1}^N H_{k,i}.$$

Given the action of $\mathrm{O}(n)$ on V, we have the natural unitary representation on $\mathcal{C}(V)$ given by $(Af)(x,i) = f(A^{-1}x, i)$ for $A \in \mathrm{O}(n)$ and $f \in \mathcal{C}(V)$. It follows that each space $H_{k,i}$ is $\mathrm{O}(n)$-irreducible and that two spaces $H_{k,i}$ and $H_{k',i'}$ are $\mathrm{O}(n)$-equivalent if and only if $k = k'$. Let

$$\{\, e_{k,i,l} : k \ge 0,\ 1 \le i \le N, \text{ and } 1 \le l \le h_k \,\}$$

be a complete orthonormal system of $\mathcal{C}(V)$ such that $e_{k,i,1}, \ldots, e_{k,i,h_k}$ is a basis of $H_{k,i}$. By Bochner's characterization [**15**], a kernel $K \in \mathcal{C}(V \times V)$ is positive and $\mathrm{O}(n)$-invariant if and only if

$$K((x,i),(y,j)) = \sum_{k=0}^\infty \sum_{i',j'=1}^N f_{ij,k} \sum_{l=1}^{h_k} e_{k,i',l}(x,i) e_{k,j',l}(y,j), \tag{8}$$

where each $\big(f_{ij,k}\big)_{i,j=1}^N$ is positive semidefinite and $\sum_{k=0}^\infty |f_{ij,k}| < \infty$ for all i, j.

By the addition formula (cf. Chapter 9.6 of Andrews, Askey, and Roy [**2**]) we have

$$\sum_{l=1}^{h_k} e_{k,l}(x)e_{k,l}(y) = \frac{h_k}{\omega_n(S^{n-1})} P_k^n(x \cdot y)$$

for any orthonormal basis $e_{k,1}, \ldots, e_{k,h_k}$ of H_k. It follows that

$$\sum_{l=1}^{h_k} e_{k,i',l}(x,i)e_{k,j',l}(y,j) = \delta_{ii'}\delta_{jj'} \frac{h_k}{\omega_n(S^{n-1})} P_k^n(x \cdot y),$$

and substituting this into (8) completes the proof. □

Bochner's characterization for the kernel K, which we used above, usually assumes that the spaces under consideration are homogeneous, so that the decompositions into isotypic irreducible spaces are guaranteed to be finite. This finiteness is then used to conclude uniform convergence. Since the action of $\mathrm{O}(n)$ on V is not transitive, we do not immediately have this guarantee. We can still use the characterization, however, since irreducible subspaces of $\mathcal{C}(V)$ have finite multiplicity.

4.3. Translational packings and multiple-size sphere packings

Before giving a proof of Theorem 4.1.3 we quickly present some technical considerations regarding density. Here we follow closely Appendix A of Cohn and Elkies [**21**].

Let $\mathcal{K}_1, \ldots, \mathcal{K}_N$ be convex bodies and $\mathcal{P}$ be a packing of translated copies of $\mathcal{K}_1, \ldots, \mathcal{K}_N$, that is, $\mathcal{P}$ is a union of translated copies of the bodies, any two copies having disjoint interiors. We say that the *density* of $\mathcal{P}$ is Δ if for all $p \in \mathbb{R}^n$ we have

$$\Delta = \lim_{r\to\infty} \frac{\mathrm{vol}(B(p,r) \cap \mathcal{P})}{\mathrm{vol}\, B(p,r)},$$

where $B(p,r)$ is the ball of radius r centered at p. Not every packing has a density, but every packing has an *upper density* given by

$$\limsup_{r\to\infty} \sup_{p\in\mathbb{R}^n} \frac{\mathrm{vol}(B(p,r) \cap \mathcal{P})}{\mathrm{vol}\, B(p,r)}.$$

We say that a packing $\mathcal{P}$ is *periodic* if there is a lattice $L \subseteq \mathbb{R}^n$ that leaves $\mathcal{P}$ invariant, that is, which is such that $\mathcal{P} = x + \mathcal{P}$ for all $x \in L$. In other words, a periodic packing consists of some translated copies of the bodies $\mathcal{K}_1, \ldots, \mathcal{K}_N$ arranged inside the fundamental parallelotope of L, and this arrangement repeats itself at each copy of the fundamental parallelotope translated by vectors of the lattice.

It is easy to see that a periodic packing has a density. This is particularly interesting for us, since in computing upper bounds for the maximum possible density of a packing we may restrict ourselves to periodic packings, as it is known (and also easy to see) that the supremum of the upper densities of packings can be approximated arbitrary well by periodic packings (cf. Appendix A in Cohn and Elkies [**21**]).

To provide a proof of the theorem we need another fact from harmonic analysis, the Poisson summation formula. Let $f\colon \mathbb{R}^n \to \mathbb{C}$ be a Schwartz function and $L \subseteq \mathbb{R}^n$ be a lattice. The *Poisson summation formula* states that, for every $x \in \mathbb{R}^n$,

$$\sum_{v\in L} f(x+v) = \frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{u\in L^*} \hat{f}(u) e^{2\pi i u\cdot x},$$

where $L^* = \{\, u \in \mathbb{R}^n : u\cdot x \in \mathbb{Z} \text{ for all } x \in L \,\}$ is the *dual lattice* of L and where $\operatorname{vol}(\mathbb{R}^n/L)$ is the volume of a fundamental domain of the lattice L.

Proof of Theorem 4.1.3. As observed above, we may restrict ourselves to periodic packings. Let $L \subseteq \mathbb{R}^n$ be a lattice and $x_1, \dots, x_m \in \mathbb{R}^n$ and $r\colon \{1,\dots,m\} \to \{1,\dots,N\}$ be such that

$$\mathcal{P} = \bigcup_{v\in L} \bigcup_{i=1}^{m} v + x_i + \mathcal{K}_{r(i)}$$

is a packing. This means that, whenever $i \neq j$ or $v \neq 0$, bodies $x_i + \mathcal{K}_{r(i)}$ and $v + x_j + \mathcal{K}_{r(j)}$ have disjoint interiors. This packing is periodic and therefore has a well-defined density, which equals

$$\frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{i=1}^{m} \operatorname{vol} \mathcal{K}_{r(i)}.$$

Consider the sum

$$\sum_{v\in L} \sum_{i,j=1}^{m} (\operatorname{vol} \mathcal{K}_{r(i)})^{1/2} (\operatorname{vol} \mathcal{K}_{r(j)})^{1/2} f_{r(i)r(j)}(v + x_j - x_i). \tag{9}$$

Applying the Poisson summation formula we may express (9) in terms of Fourier transform of f, obtaining

$$\frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{u\in L^*} \sum_{i,j=1}^{m} (\operatorname{vol} K_{r(i)})^{1/2} (\operatorname{vol} K_{r(j)})^{1/2} \hat{f}_{r(i)r(j)}(u) e^{2\pi i u\cdot(x_j - x_i)},$$

where L^* is the dual lattice of L.

Since f satisfies condition (ii) of the theorem, matrix $\big(\hat{f}_{r(i)r(j)}(u)\big)_{i,j=1}^{m}$ is positive semidefinite for every $u \in \mathbb{R}^n$. So the inner sum above is always nonnegative. If we then consider only the summand for $u = 0$, we see that (9) is at least

$$\begin{aligned} &\frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{i,j=1}^{m} (\operatorname{vol} \mathcal{K}_{r(i)})^{1/2} (\operatorname{vol} \mathcal{K}_{r(j)})^{1/2} \hat{f}_{r(i)r(j)}(0) \\ &\qquad \geq \frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{i,j=1}^{m} \operatorname{vol} \mathcal{K}_{r(i)} \operatorname{vol} \mathcal{K}_{r(j)} \qquad (10) \\ &\qquad = \frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \left(\sum_{i=1}^{m} \operatorname{vol} \mathcal{K}_{r(i)} \right)^2, \end{aligned}$$

where the inequality comes from condition (i) of the theorem.

Now, notice that whenever $v \neq 0$ or $i \neq j$ one has $f_{r(i)r(j)}(v + x_j - x_i) \leq 0$. Indeed, since $\mathcal{P}$ is a packing, if $v \neq 0$ or $i \neq j$ then the bodies $x_i + \mathcal{K}_{r(i)}$

and $v+x_j+\mathcal{K}_{r(j)}$ have disjoint interiors. But then also $\mathcal{K}_{r(i)}$ and $v+x_j-x_i+\mathcal{K}_{r(j)}$ have disjoint interiors, and then from (iii) we see that $f_{r(i)r(j)}(v+x_j-x_i)\le 0$.

From this observation we see immediately that (9) is at most

$$\sum_{i=1}^{m} \operatorname{vol}\mathcal{K}_{r(i)} f_{r(i)r(i)}(0) \le \max\{\, f_{ii}(0) : i=1,\,\ldots,\,N\,\}\sum_{i=1}^{m}\operatorname{vol}\mathcal{K}_{r(i)}. \tag{11}$$

So (9) is at least (10) and at most (11). Putting it all together we get that

$$\frac{1}{\operatorname{vol}(\mathbb{R}^n/L)}\sum_{i=1}^{m}\operatorname{vol}\mathcal{K}_{r(i)} \le \max\{\, f_{ii}(0) : i=1,\,\ldots,\,N\,\},$$

proving the theorem. □

We mentioned in the beginning of the section that Theorem 4.1.3 is an analogue of the weighted theta prime number for a certain infinite graph. The connection will become more clear after we present a slightly more general version of Theorem 4.1.3.

An L^∞ function $f\colon \mathbb{R}^n \to \mathbb{C}^{N\times N}$ is said to be of *positive type* if $f(x)=f(-x)^*$ for all $x\in\mathbb{R}^n$ and for all L^1 functions $\rho\colon \mathbb{R}^n\to\mathbb{C}^N$ we have

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \rho(y)^* f(x-y)\rho(x)\,dxdy \ge 0.$$

When $N=1$ we have the classical theory of functions of positive type (see e.g. the book by Folland [**33**] for background). Many useful properties of such functions can be extended to the matrix-valued case (that is, to the $N>1$ case) via a simple observation: a function $f\colon\mathbb{R}^n\to\mathbb{C}^{N\times N}$ is of positive type if and only if for all $p\in\mathbb{C}^N$ the function $g_p\colon\mathbb{R}^n\to\mathbb{C}$ such that

$$g_p(x)=p^*f(x)p$$

is of positive type.

From this observation two useful classical characterizations of functions of positive type can be extended to the matrix-valued case. The first one is useful when dealing with continuous functions of positive type. It states that a continuous and bounded function $f\colon\mathbb{R}^n\to\mathbb{C}^{N\times N}$ is of positive type if and only if for every choice $x_1,\,\ldots,\,x_m$ of finitely many points in $\mathbb{R}^n$, the block matrix $\big(f(x_i-x_j)\big)_{i,j=1}^m$ is positive semidefinite.

The second characterization is given in terms of the Fourier transform. It states that an L^1 function $f\colon\mathbb{R}^n\to\mathbb{C}^{N\times N}$ is of positive type if and only if the matrix $\big(\hat{f}_{ij}(u)\big)_{i,j=1}^N$ is positive semidefinite for all $u\in\mathbb{R}^n$. So in the statement of Theorem 4.1.3, for instance, one could replace condition (i) by the equivalent condition that f be a function of positive type.

When $N=1$, the previous two characterizations of functions of positive type date back to Bochner [**14**].

With this we may give an alternative and more general version of Theorem 4.1.3.

THEOREM 4.3.1. *Let $\mathcal{K}_1,\,\ldots,\,\mathcal{K}_N$ be convex bodies in $\mathbb{R}^n$ and let $f\colon\mathbb{R}^n\to\mathbb{R}^{N\times N}$ be a continuous and L^1 function. Suppose f satisfies the following conditions:*

(i) *the matrix $\big(\hat{f}_{ij}(0)-(\operatorname{vol}\mathcal{K}_i)^{1/2}(\operatorname{vol}\mathcal{K}_j)^{1/2}\big)_{i,j=1}^N$ is positive semidefinite;*

(ii) *f is of positive type;*
(iii) *$f_{ij}(x) \leq 0$ whenever $\mathcal{K}_i^\circ \cap (x + \mathcal{K}_j^\circ) = \emptyset$.*

Then the density of every packing of translates of $\mathcal{K}_1, \ldots, \mathcal{K}_N$ in the Euclidean space $\mathbb{R}^n$ is at most $\max\{ f_{ii}(0) : i = 1, \ldots, N \}$.

Let $V = \mathbb{R}^n \times \{1, \ldots, N\}$. Notice that the kernel $K \colon V \times V \to \mathbb{R}$ such that

$$K((x,i),(y,j)) = f_{ij}(x-y),$$

implicitly defined by the function f, plays the same role as the matrix K from the definition of the theta prime number (cf. Section 4.2.2). For instance, this is a positive kernel, since f is of positive type and hence for any L^1 function $\rho \colon V \to \mathbb{R}$ we have that

$$\int_V \int_V K((x,i),(y,j))\rho(x,i)\rho(y,j)\, d(x,i)d(y,j) \geq 0.$$

Theorem 4.3.1 can then be seen as an analogue of the weighted theta prime number for the packing graph with vertex set V that we consider.

When one reads through the proof of Theorem 4.1.3, the one step that fails when f is L^1 instead of Schwartz is the use of the Poisson summation formula. Indeed, sum (9) is not anymore well-defined in such a situation. The summation formula also holds, however, under somewhat different conditions that are just what we need to make the proof go through. The proof of the following lemma makes use of the well-known interpretation of the Poisson summation formula as a trace formula, which for instance is explained by Terras [**94**, Chapter 1.3].

LEMMA 4.3.2. *Let $f \colon \mathbb{R}^n \to \mathbb{C}^{N \times N}$ be a continuous function of bounded support and positive type. Then for every lattice $L \subseteq \mathbb{R}^n$, every $x \in \mathbb{R}^n$, and all $i, j = 1, \ldots, N$ we have*

$$\sum_{v \in L} f_{ij}(x+v) = \frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{u \in L^*} \hat{f}_{ij}(u) e^{2\pi i u \cdot x}.$$

PROOF. Since each function f_{ij} is continuous and of bounded support, the functions $g_{ij} \colon \mathbb{R}^n/L \to \mathbb{C}$ such that

$$g_{ij}(x) = \sum_{v \in L} f_{ij}(x+v)$$

are continuous. Indeed, the sum above is well-defined, being in fact a finite sum (since f_{ij} has bounded support), and therefore g_{ij} can be seen locally as a sum of finitely many continuous functions.

Let us now compute the Fourier transform of g_{ij}. For $u \in L^*$ we have that

$$\begin{aligned}\hat{g}_{ij}(u) &= \int_{\mathbb{R}^n/L} g_{ij}(x) e^{-2\pi i u\cdot x}\,dx \\ &= \int_{\mathbb{R}^n/L} \sum_{v\in L} f_{ij}(x+v) e^{-2\pi i u\cdot x}\,dx \\ &= \int_{\mathbb{R}^n} f_{ij}(x) e^{-2\pi i u\cdot x}\,dx \\ &= \hat{f}_{ij}(u).\end{aligned}$$

So we know that

$$g_{ij}(x) = \frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{u\in L^*} \hat{f}_{ij}(u) e^{2\pi i u\cdot x} \tag{12}$$

in the sense of L^2 convergence. Our goal is to prove that pointwise convergence also holds above.

To this end we consider for $i = 1, \ldots, N$ the kernel $K_i \colon (\mathbb{R}^n/L) \times (\mathbb{R}^n/L) \to \mathbb{C}$ such that

$$K_i(x,y) = \sum_{v\in L} f_{ii}(v + x - y).$$

Since each function f_{ii} is of bounded support and continuous, each kernel K_i is continuous. Since for each i we have that $f_{ii}(x) = \overline{f_{ii}(-x)}$ for all $x \in \mathbb{R}^n$ (since f is of positive type), each kernel K_i is self-adjoint. Notice that the functions $x \mapsto (\operatorname{vol}(\mathbb{R}^n/L))^{-1/2} e^{2\pi i u\cdot x}$, for $u \in L^*$, form a complete orthonormal system of $L^2(\mathbb{R}^n/L)$. Each such function is also an eigenfunction of K_i, with eigenvalue $\hat{f}_{ii}(u)$. Indeed, we have

$$\begin{aligned}&\int_{\mathbb{R}^n/L} K_i(x,y)(\operatorname{vol}(\mathbb{R}^n/L))^{-1/2} e^{2\pi i u\cdot y}\,dy \\ &\quad= (\operatorname{vol}(\mathbb{R}^n/L))^{-1/2} \int_{\mathbb{R}^n/L} \sum_{v\in L} f_{ii}(v+x-y) e^{2\pi i u\cdot y}\,dy \\ &\quad= (\operatorname{vol}(\mathbb{R}^n/L))^{-1/2} \int_{\mathbb{R}^n} f_{ii}(x-y) e^{2\pi i u\cdot y}\,dy \\ &\quad= (\operatorname{vol}(\mathbb{R}^n/L))^{-1/2} \int_{\mathbb{R}^n} f_{ii}(y) e^{2\pi i u\cdot(x-y)}\,dy \\ &\quad= \hat{f}_{ii}(u)(\operatorname{vol}(\mathbb{R}^n/L))^{-1/2} e^{2\pi i u\cdot x}.\end{aligned}$$

Since f is of positive type, the matrices of Fourier transforms $\big(\hat{f}_{ij}(u)\big)_{i,j=1}^N$, for $u \in \mathbb{R}^n$, are all positive semidefinite. In particular this implies that the Fourier transforms of f_{ii}, for $i = 1, \ldots, N$, are nonnegative. So we see that each K_i is a continuous and positive kernel. Mercer's theorem (see for instance Courant and Hilbert [**24**]) then implies that K_i is trace-class, its trace being the sum of all its eigenvalues. So for each $i = 1, \ldots, N$, the series

$$\sum_{u\in L^*} \hat{f}_{ii}(u) \tag{13}$$

converges, and since each summand is nonnegative, it converges absolutely.

Suppose now that i, $j = 1, \ldots, N$ are such that $i \neq j$. Since the matrices of Fourier transforms are nonnegative, for all $u \in \mathbb{R}^n$ we have that the matrix

$$\begin{pmatrix} \hat{f}_{ii}(u) & \hat{f}_{ij}(u) \\ \overline{\hat{f}_{ij}(u)} & \hat{f}_{jj}(u) \end{pmatrix}$$

is positive semidefinite, and this in turn implies that $|\hat{f}_{ij}(u)|^2 \leq \hat{f}_{ii}(u)\hat{f}_{jj}(u)$ for all $u \in \mathbb{R}^n$. Using then the convergence of the series (13) and the Cauchy-Schwarz inequality, one gets

$$\sum_{u \in L^*} |\hat{f}_{ij}(u)| \leq \sum_{u \in L^*} (\hat{f}_{ii}(u)\hat{f}_{jj}(u))^{1/2} \leq \Big(\sum_{u \in L^*} \hat{f}_{ii}(u) \Big)^{1/2} \Big(\sum_{u \in L^*} \hat{f}_{jj}(u) \Big)^{1/2},$$

and we see that in fact for all i, $j = 1, \ldots, N$ the series

$$\sum_{u \in L^*} \hat{f}_{ij}(u)$$

converges absolutely.

This convergence result shows that the sum in (12) converges absolutely and uniformly for all $x \in \mathbb{R}^n/L$. This means that the function defined by this sum is a continuous function, and since g_{ij} is also a continuous function, and in (12) we have convergence in the L^2 sense, we must also then have pointwise convergence, as we aimed to establish. □

With this we may give a proof of Theorem 4.3.1:

Proof of Theorem 4.3.1. Using Lemma 4.3.2, we may repeat the proof of Theorem 4.1.3 given before, proving the theorem for continuous functions of bounded support. To extend the proof also to continuous L^1 functions we use the following trick.

Let $f\colon \mathbb{R}^n \to \mathbb{R}^{N \times N}$ be a continuous and L^1 function satisfying the hypothesis of the theorem. For each $T > 0$ consider the function $g^T\colon \mathbb{R}^n \to \mathbb{R}^{N \times N}$ defined such that

$$g^T(x) = \frac{\operatorname{vol}(B(0,T) \cap B(x,T))}{\operatorname{vol} B(0,T)} f(x),$$

where $B(p,T)$ is the ball of radius T centered at p.

It is easy to see that g^T is a continuous function of bounded support. It is also clear that it satisfies condition (iii) from the statement of the theorem. We now show that g^T is a function of positive type, that is, it satisfies condition (ii).

For this pick any points $x_1, \ldots, x_m \in \mathbb{R}^n$. Let $\chi_i\colon \mathbb{R}^n \to \{0,1\}$ be the characteristic function of $B(x_i,T)$ and denote by $\langle \phi, \psi \rangle$ the standard inner product

between functions ϕ and ψ in the Hilbert space $L^2(\mathbb{R}^n)$. Then

$$\begin{aligned} g^T(x_i - x_j) &= \frac{\operatorname{vol}(B(0,T) \cap B(x_i - x_j, T))}{\operatorname{vol} B(0,T)} f(x_i - x_j) \\ &= \frac{\operatorname{vol}(B(x_i,T) \cap B(x_j, T))}{\operatorname{vol} B(0,T)} f(x_i - x_j) \\ &= \frac{\langle \chi_i, \chi_j \rangle}{\operatorname{vol} B(0,T)} f(x_i - x_j). \end{aligned}$$

This shows that the matrix $\big(g^T(x_i - x_j)\big)_{i,j=1}^m$ is positive semidefinite, being the Hadamard product, i.e. entrywise product, of two positive semidefinite matrices. We therefore have that g^T is of positive type.

Now, g^T is a continuous function of positive type and bounded support, satisfying condition (iii). It is very possible, however, that g^T does not satisfy condition (i), and so the conclusion of the theorem may not apply to g^T. Let us now fix this problem.

Notice that g_{ij}^T converges pointwise to f_{ij} as $T \to \infty$. Moreover, for all $T > 0$ we have $|g_{ij}^T(x)| \leq |f_{ij}(x)|$. It then follows from Lebesgue's dominated convergence theorem that $\hat{g}_{ij}^T(0) \to \hat{f}_{ij}(0)$ as $T \to \infty$. This means that there exists a number $T_0 > 0$ such that for each $T \geq T_0$ we may pick a number $\alpha(T) \geq 1$ so that the function $h^T \colon \mathbb{R}^n \to \mathbb{C}^{N \times N}$ such that

$$\begin{aligned} h_{ii}^T(x) &= \alpha(T) g_{ii}^T(x) \quad \text{for } i = 1, \ldots, N, \\ h_{ij}^T(x) &= g_{ij}^T(x) \qquad\quad \text{for } i, j = 1, \ldots, N \text{ with } i \neq j \end{aligned}$$

for all $x \in \mathbb{R}^n$ satisfies condition (i). We may moreover pick the numbers $\alpha(T)$ in such a way that $\lim_{T \to \infty} \alpha(T) = 1$.

It is also easy to see that each function h^T is of positive type and bounded support and satisfies condition (iii). Hence the conclusion of the theorem applies for each h^T, and so for every $T \geq T_0$ we see that

$$M_T = \max\{\, h_{ii}^T(0) : i = 1, \ldots, N \,\}$$

is an upper bound for the density of any packing of translated copies of $\mathcal{K}_1, \ldots, \mathcal{K}_N$. But then, since $g_{ii}^T(0) = f_{ii}(0)$ for all $T \geq 0$, and since $\lim_{T \to \infty} \alpha(T) = 1$, we see that

$$\max\{\, f_{ii}(0) : i = 1, \ldots, N \,\} = \lim_{T \to \infty} M_T,$$

finishing the proof. □

4.4. Computations for binary spherical cap packings

In this and the next section we describe how we obtained the numerical results of Sections 4.1.2 and 4.1.3. Our approach is computational: to apply Theorems 4.1.2 and 4.1.3 we use techniques from semidefinite programming and polynomial optimization.

We start by briefly discussing the case of binary spherical cap packings. Next we will discuss the more computationally challenging case of binary sphere packings.

It is a classical result of Lukács (see e.g. Theorem 1.21.1 in Szegö [**93**]) that a real univariate polynomial p of degree $2d$ is nonnegative on the interval $[a, b]$ if and only

if there are real polynomials q and r such that $p(x) = (q(x))^2 + (x-a)(b-x)(r(x))^2$. This characterization is useful when we combine it with the elementary but powerful observation (discovered independently by several authors, cf. Laurent [**67**]) that a real univariate polynomial p of degree $2d$ is a sum of squares of polynomials if and only if $p(x) = v(x)^{\mathsf{T}} Q v(x)$ for some positive semidefinite matrix Q, where $v(x) = (1, x, \ldots, x^d)$ is a vector whose components are the monomial basis.

Let $\alpha_1, \ldots, \alpha_N \in (0, \pi]$ be angles and d be an integer. Write $v_0(x) = (1, x, \ldots, x^d)$ and $v_1(x) = (1, x, \ldots, x^{d-1})$. Using this characterization together with Theorem 4.1.2, we see that the optimal value of the following optimization problem gives an upper bound for the density of a packing of spherical caps with angles $\alpha_1, \ldots, \alpha_N$.

Problem A. For $k = 0, \ldots, 2d$, find positive semidefinite matrices $\left(f_{ij,k}\right)_{i,j=1}^{N}$, and for $i, j = 1, \ldots, N$, find $(d+1) \times (d+1)$ positive semidefinite matrices Q_{ij} and $d \times d$ positive semidefinite matrices R_{ij} that minimize

$$\max\left\{ \sum_{k=0}^{2d} f_{ii,k} : i = 1, \ldots, N \right\}$$

and are such that

$$\left(f_{ij,0} - w(\alpha_i)^{1/2} w(\alpha_j)^{1/2}\right)_{i,j=1}^{N}$$

is positive semidefinite and the polynomial identities

$$\sum_{k=0}^{2d} f_{ij,k} P_k^n(u) + \langle Q_{ij}, v_0(u) v_0(u)^{\mathsf{T}} \rangle + \langle R_{ij}, (u+1)(\cos(\alpha_i + \alpha_j) - u) v_1(u) v_1(u)^{\mathsf{T}} \rangle = 0 \tag{14}$$

are satisfied for $i, j = 1, \ldots, N$. ◁

Above, $\langle A, B \rangle$ denotes the trace inner product between matrices A and B. Problem A is a semidefinite programming problem, as the polynomial identities (14) can each be expressed as $2d+1$ linear constraints on the entries of the matrices involved. Indeed, to check that a polynomial is identically zero, it suffices to check that the coefficient of each monomial $1, x, \ldots, x^{2d}$ is zero, and for each such monomial we get a linear constraint.

In the above, we work with the standard monomial basis $1, x, \ldots, x^{2d}$, but we could use any other basis of the space of polynomials of degree at most $2d$, both to define the vectors v_0 and v_1 and to check the polynomial identity (14). Such a change of basis does not change the problem from a formal point of view, but can drastically improve the performance of the solvers used. In our computations for binary spherical cap packings it was enough to use the standard monomial basis. We will see in the next section, when we present our computations for the Euclidean space, that a different choice of basis is essential.

We reported in Section 4.1.2 on our calculations for $N = 1$, and 2 and $n = 3$, 4, and 5. The bounds, for the angles under consideration, do not seem to improve beyond $d = 25$, so we use this value for d in all computations. To obtain these bounds we used the solver SDPA-QD, which works with quadruple precision floating

point numbers, from the SDPA family [**36**]. To generate the input for the solver we wrote a SAGE [**84**] program using SDPSL [**79**] (our programs are distributed as part of SDPSL).

4.5. Computations for binary sphere packings

In this section we discuss our computational approach to find upper bounds for the density of binary sphere packings using Theorem 4.1.3. This is a more difficult application of semidefinite programming and polynomial optimization techniques than the one described in Section 4.4.

It is often the case in applications of sum of squares techniques that, if one formulates the problems carelessly, high numerical instability invalidates the final results, or even numerical results cannot easily be obtained. This raises questions of how to improve the formulations used and the precision of the computations, so that we may provide *rigorous* bounds. We also address these questions and, since the techniques we use and develop might be of interest to the reader who wants to perform computations in polynomial optimization, we include some details.

4.5.1. Theorem 4.1.3 for multiple-size sphere packings. In the case of sphere packings, Theorem 4.1.3 can be simplified. The key observation here is that, when all the bodies $\mathcal{K}_i$ are spheres, then condition (iii) depends only on the norm of the vector x. More specifically, if each $\mathcal{K}_i$ is a sphere of radius r_i, then $\mathcal{K}_i^\circ \cap (x + \mathcal{K}_j^\circ) = \emptyset$ if and only if $\|x\| \geq r_i + r_j$.

So in Theorem 4.1.3 one can choose to restrict oneself to radial functions. A function $f\colon \mathbb{R}^n \to \mathbb{C}$ is *radial* if the value of $f(x)$ depends only on the norm of x. If $f\colon \mathbb{R}^n \to \mathbb{C}$ is radial, for $t \geq 0$ we denote by $f(t)$ the common value of f for vectors of norm t.

The Fourier transform $\hat{f}(u)$ of a radial function f also depends only on the norm of u; in other words, the Fourier transform of a radial function is also radial. By restricting ourselves to radial functions, we obtain the following version of Theorem 4.1.3.

THEOREM 4.5.1. *Let $r_1, \ldots, r_N > 0$ and let $f\colon \mathbb{R}^n \to \mathbb{R}^{N\times N}$ be a matrix-valued function whose every component f_{ij} is a radial Schwartz function. Suppose f satisfies the following conditions:*

(i) *the matrix $\left(\hat{f}_{ij}(0) - (\operatorname{vol} B(r_i))^{1/2}(\operatorname{vol} B(r_j))^{1/2}\right)_{i,j=1}^N$ is positive semidefinite, where $B(r)$ is the ball of radius r centered at the origin;*

(ii) *the matrix of Fourier transforms $\left(\hat{f}_{ij}(t)\right)_{i,j=1}^N$ is positive semidefinite for every $t > 0$;*

(iii) *$f_{ij}(w) \leq 0$ if $w \geq r_i + r_j$, for $i, j = 1, \ldots, N$.*

Then the density of any packing of spheres of radii $r_1, \ldots, r_N$ in the Euclidean space $\mathbb{R}^n$ is at most $\max\{ f_{ii}(0) : i = 1, \ldots, N \}$.

One might ask whether the restriction to radial functions worsens the bound of Theorem 4.1.3. For spheres, this is not the case. Indeed, suppose each body $\mathcal{K}_i$ is a sphere. If $f\colon \mathbb{R}^n \to \mathbb{R}^{N\times N}$ is a function satisfying the conditions of the theorem,

then its radialized version, the function

$$\overline{f}(x) = \int_{S^{n-1}} f(\|x\|\xi)\, d\omega_n(\xi),$$

also satisfies the conditions of the theorem, and it provides the same upper bound. This shows in particular that, for the case of multiple-size sphere packings, Theorem 4.5.1 is equivalent to Theorem 4.1.3.

4.5.2. A semidefinite programming formulation. To simplify notation and because it is the case of our main interest we now take $N = 2$. Everything in the following also goes through for arbitrary N with obvious modifications.

To find a function f satisfying the conditions of Theorem 4.5.1 we specify f via its Fourier transform. Let $d \geq 0$ be an odd integer and consider the even function $\varphi\colon \mathbb{R}_{\geq 0} \to \mathbb{R}^{2\times 2}$ such that

$$\varphi_{ij}(t) = \sum_{k=0}^{d} a_{ij,k} t^{2k},$$

where each $a_{ij,k}$ is a real number and $a_{ij,k} = a_{ji,k}$ for all k. We set the Fourier transform of f to be

$$\hat{f}_{ij}(u) = \varphi_{ij}(\|u\|)e^{-\pi\|u\|^2}.$$

Notice that each $\hat{f}_{ij}$ is a Schwartz function, so its Fourier inverse is also Schwartz.

The reason why we choose this form for the Fourier transform of f is that it makes it simple to compute f from its Fourier transform by using the following result.

LEMMA 4.5.2. *We have that*

$$\int_{\mathbb{R}^n} \|u\|^{2k} e^{-\pi\|u\|^2} e^{2\pi i u\cdot x}\, du = k!\, \pi^{-k} e^{-\pi\|x\|^2} L_k^{n/2-1}(\pi\|x\|^2), \tag{15}$$

where $L_k^{n/2-1}$ is the Laguerre polynomial of degree k with parameter $n/2-1$.

For background on Laguerre polynomials, we refer the reader to the book by Andrews, Askey, and Roy [**2**].

PROOF. With $f(u) = \|u\|^{2k} e^{-\pi\|u\|^2}$, the left hand side of (15) is equal to $\hat{f}(-x)$. By [**2**, Theorem 9.10.3] we have

$$\hat{f}(-x) = 2\pi\|x\|^{1-n/2} \int_0^\infty s^{2k} e^{-\pi s^2} J_{n/2-1}(2\pi s\|x\|) s^{n/2}\, ds,$$

where $J_{n/2-1}$ is the Bessel function of the first kind with parameter $n/2-1$. Using [**2**, Corollary 4.11.8] we see that this is equal to

$$\pi^{-k} \frac{\Gamma(k+n/2)}{\Gamma(n/2)} e^{-\pi\|x\|^2}\, {}_1F_1\Big(\begin{matrix}-k\\ n/2\end{matrix}; \pi\|x\|^2\Big), \tag{16}$$

where ${}_1F_1$ is a hypergeometric series.

By [**2**, (6.2.2)] we have

$${}_1F_1\Big(\begin{matrix}-k\\ n/2\end{matrix}; \pi\|x\|^2\Big) = \frac{k!}{(n/2)_k} L_k^{n/2-1}(\pi\|x\|^2),$$

where $(n/2)_k = (n/2)(1+n/2)\cdots(k-1+n/2)$.

By substituting this in (16), and using the property that $\Gamma(x+1) = x\Gamma(x)$ for all $x \neq 0, -1, -2, \ldots$, we obtain the right hand side of (15) as desired. □

So we have

$$f_{ij}(x) = \int_{\mathbb{R}^n} \varphi_{ij}(\|u\|) e^{-\pi\|u\|^2} e^{2\pi i u\cdot x}\, du = \sum_{k=0}^{d} a_{ij,k}\, k!\, \pi^{-k} e^{-\pi\|x\|^2} L_k^{n/2-1}(\pi\|x\|^2).$$

Notice that it becomes clear that f_{ij} is indeed real-valued, as required by the theorem.

Consider the polynomial

$$p(t) = \sum_{k=0}^{d} a_k t^{2k}.$$

According to Lemma 4.5.2, if $g(x)$ is the Fourier inverse of $\hat{g}(u) = p(\|u\|)e^{-\pi\|u\|^2}$, then $g(\|x\|) = q(\|x\|)e^{-\pi\|x\|^2}$, where

$$q(w) = \sum_{k=0}^{d} a_k\, k!\, \pi^{-k} L_k^{n/2-1}(\pi w^2)$$

is a univariate polynomial. We denote the polynomial q above by $\mathcal{F}^{-1}[p]$. Notice that $\mathcal{F}^{-1}[p]$ is obtained from p via a linear transformation, i.e., its coefficients are linear combinations of the coefficients of p. With this notation we have

$$f_{ij}(x) = \mathcal{F}^{-1}[\varphi_{ij}](\|x\|)e^{-\pi\|x\|^2}.$$

Let

$$\sigma(t, y_1, y_2) = \sum_{i,j=1}^{2} \sum_{k=0}^{d} a_{ij,k} t^{2k} y_i y_j. \tag{17}$$

If this polynomial is a sum of squares, then it is nonnegative everywhere, and hence the matrices $\big(\varphi_{ij}(t)\big)_{i,j=1}^{2}$ are positive semidefinite for all $t \geq 0$. This implies that f satisfies condition (ii) of Theorem 4.5.1. (The converse is also true, that if the matrices $\big(\varphi_{ij}(t)\big)_{i,j=1}^{2}$ are positive semidefinite for all $t \geq 0$, then σ is a sum of squares; For a proof see Choi, Lam, Reznick [**20**]. This fact is related to the Kalman-Yakubovich-Popov lemma in systems and control; see the discussion in Aylward, Itani, and Parrilo [**5**].)

Moreover, we may recover φ, and hence $\hat{f}$, from σ. Indeed we have

$$\begin{aligned} \varphi_{11}(t) &= \sigma(t, 1, 0), \\ \varphi_{22}(t) &= \sigma(t, 0, 1), \quad \text{and} \\ \varphi_{12}(t) &= (1/2)(\sigma(t, 1, 1) - \sigma(t, 1, 0) - \sigma(t, 0, 1)). \end{aligned} \tag{18}$$

So we can express condition (i) of Theorem 4.5.1 in terms of σ. We may also express condition (iii) in terms of σ, since it can be translated as

$$\mathcal{F}^{-1}[\varphi_{ij}](w) \leq 0 \quad \text{for all } w \geq r_i + r_j \text{ and } i,\, j = 1,\, 2 \text{ with } i \leq j. \tag{19}$$

If we find a polynomial σ of the form (17) that is a sum of squares, is such that

$$\left(\varphi_{ij}(0) - (\operatorname{vol} B(r_i))^{1/2}(\operatorname{vol} B(r_j))^{1/2}\right)_{i,j=1}^{2} \tag{20}$$

is positive semidefinite, and satisfies (19), then the density of a packing of spheres of radii r_1 and r_2 is upper bounded by

$$\max\{\mathcal{F}^{-1}[\varphi_{11}](0), \mathcal{F}^{-1}[\varphi_{22}](0)\}.$$

We may encode conditions (19) in terms of sums of squares polynomials (cf. Section 4.4), and therefore we may encode the problem of finding a σ as above as a semidefinite programming problem, as we show now.

Let P_0, P_1, ... be a sequence of univariate polynomials where polynomial P_k has degree k. Consider the vector of polynomials v, which has entries indexed by $\{0, \dots, \lfloor d/2 \rfloor\}$ given by

$$v(t)_k = P_k(t^2)$$

for $k = 0, \dots, \lfloor d/2 \rfloor$. We also write $V(t) = v(t)v(t)^{\mathsf{T}}$.

Consider also the vector of polynomials m with entries indexed by $\{1, 2\} \times \{0, \dots, \lfloor d/2 \rfloor\}$ given by

$$m(t, y_1, y_2)_{i,k} = P_k(t^2)y_i$$

for i, $j = 1, 2$ and $k = 0, \dots, \lfloor d/2 \rfloor$.

Since σ is an even polynomial, it is a sum of squares if and only if there are positive semidefinite matrices S_0, $S_1 \in \mathbb{R}^{(d+1)\times(d+1)}$ such that

$$\sigma(t, y_1, y_2) = \langle S_0, m(t, y_1, y_2)m(t, y_1, y_2)^{\mathsf{T}}\rangle + \langle S_1, t^2 m(t, y_1, y_2)m(t, y_1, y_2)^{\mathsf{T}}\rangle.$$

From the matrices S_0 and S_1 we may then recover φ_{ij} and also $\mathcal{F}^{-1}[\varphi_{ij}]$. A more convenient way for expressing φ_{ij} in terms of S_0 and S_1 is as follows. Consider the matrices

$$Y_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Y_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad Y_{12} = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}.$$

Then

$$\varphi_{ij}(t) = \langle S_0, V(t) \otimes Y_{ij}\rangle + \langle S_1, t^2 V(t) \otimes Y_{ij}\rangle$$

and

$$\mathcal{F}^{-1}[\varphi_{ij}](w) = \langle S_0, \mathcal{F}^{-1}[V(t)](w) \otimes Y_{ij}\rangle + \langle S_1, \mathcal{F}^{-1}[t^2 V(t)](w) \otimes Y_{ij}\rangle,$$

where $\mathcal{F}^{-1}$, when applied to a matrix, is applied to each entry individually.

With this, we may consider the following semidefinite programming problem for finding a polynomial σ satisfying the conditions we need.

Problem B. Find $(d+1) \times (d+1)$ real positive semidefinite matrices S_0 and S_1, and $(\lfloor d/2 \rfloor + 1) \times (\lfloor d/2 \rfloor + 1)$ real positive semidefinite matrices Q_{11}, Q_{22}, and Q_{12} that minimize

$$\begin{aligned}\max\{&\langle S_0, \mathcal{F}^{-1}[V(t)](0) \otimes Y_{11}\rangle + \langle S_1, \mathcal{F}^{-1}[t^2 V(t)](0) \otimes Y_{11}\rangle, \\ &\langle S_0, \mathcal{F}^{-1}[V(t)](0) \otimes Y_{22}\rangle + \langle S_1, \mathcal{F}^{-1}[t^2 V(t)](0) \otimes Y_{22}\rangle\}\end{aligned}$$

and are such that

$$\left(\langle S_0, V(0) \otimes Y_{ij}\rangle - (\operatorname{vol} B(r_i))^{1/2}(\operatorname{vol}(B(r_j))^{1/2}\right)_{i,j=1}^{2} \tag{21}$$

is positive definite and the polynomial identities

$$\begin{aligned}(22)\qquad &\langle S_0, \mathcal{F}^{-1}[V(t)](w) \otimes Y_{ij}\rangle + \langle S_1, \mathcal{F}^{-1}[t^2 V(t)](w) \otimes Y_{ij}\rangle \\ &\qquad + \langle Q_{ij}, (w^2 - (r_i + r_j)^2) V(w)\rangle = 0\end{aligned}$$

are satisfied for i, $j = 1$, 2 and $i \leq j$. ◁

Any solution to this problem gives us a polynomial σ of the shape (17) which is a sum of squares and satisfies conditions (19) and (20), and so the optimal value is an upper bound for the density of any packing of spheres of radius r_1 and r_2. There might be, however, polynomials σ satisfying these conditions that cannot be obtained as feasible solutions to Problem B, since condition (22) is potentially more restrictive than condition (19) (compare Problem B above with Lukács' result mentioned in Section 4.4). In our practical computations this restriction was not problematic and we found very good functions.

Observe also that Problem B is really a semidefinite programming problem. Indeed, the polynomial identities in (22) can each be represented as $d + 1$ linear constraints in the entries of the matrices S_i and Q_{ij}. This is the case because testing whether a polynomial is identically zero is the same as testing whether each monomial has a zero coefficient and so, since all our polynomials are even and of degree $2d$, we need only check if the coefficients of the monomials x^{2k} are zero for $k = 0, \ldots, d$.

4.5.3. Numerical results. When solving Problem B, we need to choose a sequence P_0, P_1, ... of polynomials. A choice which works well in practice is

$$P_k(t) = \mu_k^{-1} L_k^{n/2-1}(2\pi t),$$

where μ_k is the absolute value of the coefficient of $L_k^{n/2-1}(2\pi t)$ with largest absolute value. We observed in practice that the standard monomial basis performs poorly.

To represent the polynomial identities in (22) as linear constraints we may check that each monomial x^{2k} of the resulting polynomial has coefficient zero. We may use, however, any basis of the space of even polynomials of degree at most $2d$ to represent such identities. Given such a basis, we expand each polynomial in it and check that the expansion has only zero coefficients. The basis we use to represent the identities is $P_0(t^2)$, $P_1(t^2)$, ..., $P_d(t^2)$, which we observed to work much better than t^0, t^2, ..., t^{2d}. Notice that no extra variables are necessary if we use a different basis to represent the identities. We need only keep, for each polynomial in the matrices $\mathcal{F}^{-1}[V(t)](w)$, $\mathcal{F}^{-1}[t^2 V(t)](w)$, $w^2 V(w)$, and $V(w)$, its expansion in the basis we want to use.

The plot in Figure 3 was generated by solving Problem B with $d = 31$ using the solver SDPA-GMP from the SDPA family [**36**]. To generate the solver input we wrote a SAGE [**84**] program using SDPSL [**79**] working with floating-point arithmetic and precision of 256 bits; see the examples in the source distribution of SDPSL for the source code. For each dimension 2, ..., 5 we solved Problem B with $r_1 = r/1000$ and $r_2 = 1$ for $r = 200, 201, \ldots, 1000$; the reason we start with $r = 200$ is that for smaller values of r the solver runs into numerical stability problems. We also note that the solver has failed to solve some of the problems, and these points have been ignored when generating the plot. The number of problems

that could not be solved was small though: for $n = 2$ all problems could be solved, for $n = 3$ there were 6 failures, for $n = 4$ we had 18 failures, and finally for $n = 5$ the solver failed for 137 problems.

With our methods we can achieve higher values for d, but we noticed that the bound does not improve much after $d = 31$. For instance, in dimension 2 for $r_1 = 1/2$ and $r_2 = 1$, we obtain the bound $0.9174466\ldots$ for $d = 31$ and the bound $0.9174426\ldots$ for $d = 51$.

* * *

In the previous account of how the plot in Figure 3 was generated, we swept under the rug all precision issues. We generate the data for the solver using floating-point arithmetic, and the solver also uses floating-point arithmetic. We cannot therefore be sure that the optimal value found by the solver gives a valid bound at all.

If we knew *a priori* that Problem B is strictly feasible (that is, that it admits a solution in which the matrices S_i and Q_{ij} are positive definite), and if we had some control over the dual solutions, then we could use semidefinite programming duality to argue that the bounds we compute are rigorous; see for instance Gijswijt [**39**, Chapter 7.2] for an application of this approach in coding theory. The matter is however that we do not know that Problem B is strictly feasible, neither do we have knowledge about the dual solutions. In fact, most of our approach to provide rigorous bounds consists in finding a strictly feasible solution.

A naive idea to turn the bound returned by the solver into a rigorous bound would be to simply project a solution returned by the solver onto the subspace given by the constraints in (22). If the original solution is of good quality, then this would yield a feasible solution.

There are two problems with this approach, though. The first problem is that the matrices returned by the solver will have eigenvalues too close to zero, and therefore after the projection they might not be positive semidefinite anymore. We discuss how to handle this issue below.

The second problem is that to obtain a rigorous bound one would need to perform the projection using symbolic computations and rational arithmetic, and the computational cost is just too big. For instance, we failed to do so even for $d = 7$.

Our approach avoids projecting the solution using symbolic computations. Here is an outline of our method.

(1) Obtain a solution to the problem with objective value close the optimal value returned by the solver, but in which every matrix S_i and Q_{ij} is positive definite by a good margin and the maximum violation of the constraints is very small.
(2) Approximate matrices S_i and Q_{ij} by rational positive semidefinite matrices $\bar{S}_i$ and $\bar{Q}_{ij}$ having minimum eigenvalues at least λ_i and μ_{ij}, respectively.
(3) Compute a bound on how much constraints (22) are violated by $\bar{S}_i$ and $\bar{Q}_{ij}$ using rational arithmetic. If the maximum violation of the constraints is small compared to the bounds λ_i and μ_{ij} on the minimum eigenvalues, then we may be sure that the solution can be changed into a feasible solution without changing its objective value too much.

We now explain how each step above can be accomplished.

First, most likely the matrices S_i, Q_{ij} returned by the solver will have eigenvalues very close to zero, or even slightly negative due to the numerical method which might allow infeasible steps.

To obtain a solution with positive definite matrices we may use the following trick (cf. Löfberg [**70**]). We solve Problem B to find its optimal value, say z^*. Then we solve a feasibility version of Problem B in which the objective function is absent, but we add a constraint to ensure that

$$\begin{aligned}\max\{&\langle S_0, \mathcal{F}^{-1}[V(t)](0) \otimes Y_{11}\rangle + \langle S_1, \mathcal{F}^{-1}[t^2 V(t)](0) \otimes Y_{11}\rangle,\\ &\langle S_0, \mathcal{F}^{-1}[V(t)](0) \otimes Y_{22}\rangle + \langle S_1, \mathcal{F}^{-1}[t^2 V(t)](0) \otimes Y_{22}\rangle\} \le z^* + \eta,\end{aligned}$$

where $\eta > 0$ should be small enough so that we do not jeopardize the objective value of the solution, but not too small so that a good strictly feasible solution exists. (We take $\eta = 10^{-5}$, which works well for the purpose of making a plot.) The trick here is that most semidefinite programming solvers, when solving a feasibility problem, will return a strictly feasible solution — the analytical center —, if one can be found.

This partially addresses step (1), because though the solution we find will be strictly feasible, it might violate the constraints too much. To quickly obtain a solution that violates the constraints only slightly, we may project our original solution onto the subspace given by constraints (22) using floating-point arithmetic of high enough precision. If the solution returned by the solver had good precision to begin with, then the projected solution will still be strictly feasible.

As an example, for our problems with $d = 31$, SDPA-GMP returns solutions that violate the constraints by at most 10^{-30}. By doing a projection using floating-point arithmetic with 256 bits of precision in SAGE, we can bring the violation down to about 10^{-70} without affecting much the eigenvalues of the matrices.

So we have addressed step (1). For step (2) we observe that simply converting the floating-point matrices S_i, Q_{ij} to rational matrices would work, but then we would be in trouble to estimate the minimum eigenvalues of the resulting rational matrices in a rigorous way. Another idea of how to make the conversion is as follows.

Say we want to approximate floating-point matrix A by a rational matrix $\bar{A}$. We start by computing numerically an approximation to the least eigenvalue of A. Say $\tilde{\lambda}$ is this approximation. We then use binary search in the interval $[\tilde{\lambda}/2, \tilde{\lambda}]$ to find the largest λ so that the matrix $A - \lambda I$ has a Cholesky decomposition; this we do using floating-point arithmetic of high enough precision. If we have this largest λ, then

$$A = LL^{\mathsf{T}} + \lambda I$$

where L is the Cholesky factor of $A - \lambda I$. Then we approximate L by a rational matrix $\bar{L}$ and we set

$$\bar{A} = \bar{L}\bar{L}^{\mathsf{T}} + \lambda I,$$

obtaining thus a rational approximation of A and a bound on its minimum eigenvalue.

Our idea for step (3) is to compare the maximum violation of constraints (22) with the minimum eigenvalues of the matrices. To formalize this idea, suppose that

constraints (22) are slightly violated by $\bar{S}_i$, $\bar{Q}_{ij}$. So for instance we have

$$\begin{aligned}\langle \bar{S}_0, \mathcal{F}^{-1}[V(t)](w) \otimes Y_{11}\rangle + \langle \bar{S}_1, \mathcal{F}^{-1}[t^2 V(t)](w) \otimes Y_{11}\rangle \\ + \langle \bar{Q}_{11}, (w^2 - (2r_1)^2)V(w)\rangle = p,\end{aligned} \tag{23}$$

where p is an even polynomial of degree at most $2d$. Notice that we may compute an upper bound on the absolute values of the coefficients of p using rational arithmetic.

To fix this constraint we may distribute the coefficients of p in the matrices $\bar{S}_0$ and $\bar{Q}_{11}$ (a very similar idea was presented by Löfberg [**70**]). To make things precise, for $k = 1, \ldots, d$ write

$$\begin{aligned} i(k) &= \min\{\lfloor d/2 \rfloor, k-1\}, \\ j(k) &= k - 1 - i(k). \end{aligned}$$

Pairs $(i(k), j(k))$ correspond to entries of the matrix $V(w)$. Notice that the polynomial $(w^2 - (2r_1)^2)V(w)_{i(k)j(k)}$ has degree $2k$.

So the polynomials

$$\begin{aligned} R_0 &= \mathcal{F}^{-1}[V(t)_{00}](w), \\ R_1 &= (w^2 - (2r_1)^2)V(w)_{i(1)j(1)}, \\ &\vdots \\ R_d &= (w^2 - (2r_1)^2)V(w)_{i(d)j(d)} \end{aligned}$$

form a basis of the space of even polynomials of degree at most $2d$. We may then express our polynomial p in this basis as

$$p = \alpha_0 R_0 + \cdots + \alpha_d R_d.$$

Now, we subtract α_0 from $(\bar{S}_0)_{(1,0),(1,0)}$ and α_k from $(\bar{Q}_{11})_{i(k)j(k)}$, for $k = 1, \ldots, d$. The resulting matrices satisfy constraint (23), and as long as the α_k are small enough, they should remain positive semidefinite. More precisely, it suffices to require that $d\,\|(\alpha_1, \ldots, \alpha_d)\|_\infty \leq \mu_{11}$ and $|\alpha_0| \leq \lambda_0$.

There are two issues to note in our approach. The first one is that it has to be applied again twice to fix the other two constraints in (22). The applications do not conflict with each other: in each one we change a different matrix $\bar{Q}_{ij}$ and different entries of $\bar{S}_0$. We have to be careful though that we consider the changes to $\bar{S}_0$ at once in order to check that it remains positive semidefinite.

The second issue is how to compute the coefficients α_k. Computing them explicitly using symbolic computation is infeasible. One way to do it then is to consider the basis change matrix between the bases x^{2k}, for $k = 0, \ldots, d$, and $R_0, \ldots, R_d$, which we denote by U. Then we know that

$$\|(\alpha_0, \ldots, \alpha_d)\|_\infty \leq \|U^{-1}\|_\infty \|p\|_\infty,$$

where $\|p\|_\infty$ is the ∞-norm of the vector of coefficients of p in the basis x^{2k}.

So if we have an upper bound for $\|U^{-1}\|_\infty$ we are done. To quickly find such an upper bound, we use an algorithm of Higham [**49**] (cf. also Higham [**50**]) which works for triangular matrices, like U. This bound proved to be good enough for our purposes.

4.6. Improving sphere packing bounds

We now prove Theorem 4.1.4 and show how to use it in order to compute the bounds presented in Table 1.

PROOF OF THEOREM 4.1.4. Let $x_1, \ldots, x_N \in \mathbb{R}^n$ and $L \subseteq \mathbb{R}^n$ be a lattice such that

$$\bigcup_{v \in L} \bigcup_{i=1}^{N} v + x_i + B$$

is a sphere packing, where B is the ball of radius $1/2$ centered at the origin. We may assume that, if $i \neq j$ and $v \neq 0$, then the distance between the centers of $v + x_i + B$ and $x_j + B$ is greater than $1 + \varepsilon_m$. Indeed, we could discard all x_i that lie at distance less than $1 + \varepsilon_m$ from the boundary of the fundamental parallelotope of L. If the fundamental parallelotope is big enough (and if it is not, we may consider a dilated version of L instead), this will only slightly alter the density of the packing, and the resulting packing will have the desired property.

Consider the sum

$$\sum_{i,j=1}^{N} \sum_{v \in L} f(v + x_i - x_j). \tag{24}$$

Using the Poisson summation formula, we may rewrite it as

$$\frac{1}{\operatorname{vol}(\mathbb{R}^n/L)} \sum_{i,j=1}^{N} \sum_{u \in L^*} \hat{f}(u) e^{2\pi i u \cdot (x_i - x_j)}.$$

By discarding all summands in the inner sum above except the one for $u = 0$, we see that (24) is at least

$$\frac{N^2 \operatorname{vol} B}{\operatorname{vol}(\mathbb{R}^n/L)}.$$

For $k = 1, \ldots, m$, write $F_k = \{\, (i,j) : \|x_i - x_j\| \in [1 + \varepsilon_{k-1}, 1 + \varepsilon_k) \,\}$. Then we see that (24) is at most

$$N f(0) + \eta_1 |F_1| + \cdots + \eta_m |F_m|.$$

So we see that

$$\frac{N \operatorname{vol} B}{\operatorname{vol}(\mathbb{R}^n/L)} \leq f(0) + \eta_1 \frac{|F_1|}{N} + \cdots + \eta_m \frac{|F_m|}{N}.$$

Notice that the left-hand side above is exactly the density of our packing. Now, from the definition of $M(\varepsilon)$, it is clear that for $k = 1, \ldots, m$ we have

$$\frac{|F_1|}{N} + \cdots + \frac{|F_k|}{N} \leq M(\varepsilon_k),$$

and the theorem follows. □

To find good functions f satisfying the conditions required by Theorem 4.1.4 we use the approach from Section 4.5. We fix an odd positive integer d and specify f via its Fourier transform, writing

$$\varphi(t) = \sum_{k=0}^{d} a_k t^{2k}$$

and setting

$$\hat{f}(u) = \varphi(\|u\|)e^{-\pi\|u\|^2}.$$

Using Lemma 4.5.2 we then have that

$$f(x) = \mathcal{F}^{-1}[\varphi](\|x\|)e^{-\pi\|x\|^2},$$

where

$$\mathcal{F}^{-1}[\varphi](w) = \sum_{k=0}^{d} a_k k!\, \pi^{-k} L_k^{n/2-1}(\pi w^2)$$

is a polynomial obtained as a linear transformation of φ.

Constraint (ii), requiring that $\hat{f}(u) \geq 0$ for all $u \in \mathbb{R}^n$, can be equivalently expressed as requiring that the polynomial φ should be a sum of squares.

Recalling the result of Lukács mentioned in Section 4.4, one may also express constraint (iii) in terms of sums of squares: one simply has to require that there exist polynomials $p_0(w)$ and $q_0(w)$ such that

$$\mathcal{F}^{-1}[\varphi](w) = -(p_0(w))^2 - (w^2 - (1+\varepsilon_m)^2)(q_0(w))^2.$$

In a similar way, one may express constraints (iv). For instance, for a given k, we require that there should exist polynomials $p_k(w)$ and $q_k(w)$ such that

$$\mathcal{F}^{-1}[\varphi](w)e^{-\pi(1+\varepsilon_{k-1})^2} - \eta_1 = -(p_k(w))^2 - (w - (1+\varepsilon_{k-1}))((1+\varepsilon_k) - w)(q_k(w))^2,$$

and this implies (iv).

So we may represent the constraints on f in terms of sums of squares, and therefore also in terms of semidefinite programming, as we did in Sections 4.4 and 4.5. There is only the issue that now we want to find a function f that satisfies constraints (i)–(iv) of the theorem and that minimizes the maximum in (2). This does not look like a linear objective function, but since by linear programming duality this maximum is equal to

$$\begin{array}{lll} \min & f(0) + y_1 U(\varepsilon_1) + \cdots + y_m U(\varepsilon_m) & \\ & y_i + \cdots + y_m \geq \eta_i & \text{for } i = 1, \ldots, m, \\ & y_k \geq 0 & \text{for } k = 1, \ldots, m, \end{array}$$

we may transform our original problem into a single minimization semidefinite programming problem, the optimal value of which provides an upper bound for the densities of sphere packings.

It is still a question how to compute upper bounds for $M(\varepsilon)$. For this we use upper bounds on the sizes of spherical codes. A *spherical code* with *minimum angular distance* $0 < \theta \leq \pi$ is a set $C \subseteq S^{n-1}$ such that the angle between any two distinct points in C is at least θ. In other words, a spherical code with minimum angular distance θ gives a packing of spherical caps with angle $\theta/2$. We denote

by $A(n, \theta)$ the maximum cardinality of any spherical code in S^{n-1} with minimum angular distance θ.

For $\varepsilon \leq (\sqrt{5}-1)/2$ we have

$$M(\varepsilon) \leq A(n, \arccos t(\varepsilon)), \quad \text{where} \quad t(\varepsilon) = 1 - \frac{1}{2(1+\varepsilon)^2}.$$

This follows from the proof of [**22**, Lemma 4.1] which we replicate here for the convenience of the reader: Suppose x, $y \in \mathbb{R}^n$ are such that $\|x\|$, $\|y\| \in [1, 1+\varepsilon]$ and $\|x-y\| \geq 1$. Then by the law of cosines

$$\cos \angle(x, y) = \frac{\|x\|^2 + \|y\|^2 - \|x-y\|^2}{2\|x\|\,\|y\|} \leq \frac{\|x\|^2 + \|y\|^2 - 1}{2\|x\|\,\|y\|}.$$

The right hand side is convex as a function of $\|x\|$ or $\|y\|$ individually, so it is upper bounded by the maximal value at the four vertices of the square $[1, 1+\varepsilon]^2$. Since $\varepsilon \leq (\sqrt{5}-1)/2$, the maximum occurs for $\|x\| = \|y\| = 1+\varepsilon$, which gives the desired bound.

For the bounds of Table 1 we took $d = 31$. To compute upper bounds for $A(n, \theta)$ we used the semidefinite programming bound of Bachoc and Vallentin [**9**]. The bounds we used for computing Table 1 are given in Table 2.

Finally, we mention that all numerical issues discussed in Section 4.5 also happen with the approach we sketched in this section. In particular, the choices of bases are important for the stability of the semidefinite programming problems involved. We use the same bases as described in Section 4.5 though, so we skip a detailed discussion here. Notice moreover that our bounds are rigorous, having been checked with the same approach described in Section 4.5.

Dimension	$(\varepsilon, U(\varepsilon))$ pairs
4	(0.008097, 24), (0.017446, 25), (0.025978, 26), (0.036951, 27)
5	(0.003013, 45), (0.008097, 46), (0.013259, 47), (0.017446, 48)
6	(0.002006, 79), (0.004024, 80), (0.006054, 81), (0.008097, 82)
7	(0.001001, 136), (0.002006, 137), (0.003013, 138), (0.004024, 139), (0.005037, 140)
9	(0.003013, 373), (0.029233, 457), (0.030325, 459), (0.031421, 464), (0.032520, 468), (0.033622, 473)

TABLE 2. For each dimension considered in Table 1 we show here the sequence $\varepsilon_1 < \cdots < \varepsilon_m$ and the upper bounds $U(\varepsilon_k)$ used in our application of Theorem 4.1.4.

We refrained from performing similar calculations for higher dimensions because of two reasons. Firstly, we expect that the improvements are only minor. Secondly, the computations of the upper bounds for $M(\varepsilon)$ in higher dimensions require substantially more time as one needs to solve the semidefinite programs with a high accuracy solver, see Mittelmann and Vallentin [**74**].

CHAPTER 5

Optimal polydisperse packing densities using objects with large size ratio

This chapter is based on the publication *"D. de Laat, Optimal polydisperse packing densities using objects with large size ratio, in preparation"*.

Abstract. Let Δ denote the optimal packing density of Euclidean space by unit balls. We show the optimal packing density using two sizes of balls approaches $\Delta+(1-\Delta)\Delta$ as the ratio of the radii tends to infinity. More generally, if B is a body and $\mathcal{B}$ a finite set of bodies, then we show the optimal density $\Delta_{\{rB\}\cup\mathcal{B}}$ of packings using congruent copies of the bodies $\{rB\}\cup\mathcal{B}$ converges to $\Delta_{\mathcal{B}}+(1-\Delta_{\mathcal{B}})\Delta_{\{B\}}$ as r tends to zero.

5.1. Introduction

There has been extensive research into the determination of optimal monodisperse packing densities. A well-known example is Hales's proof of the Kepler conjecture on the optimal sphere packing density in $\mathbb{R}^3$ [**44**]. More recently, packings with polydispersity have been investigated: New lower and upper bounds for the density of packings of spheres using several sizes have been given in respectively [**53**] and [**60**]. In applications, sphere packings can be used to model many-particle systems, and here it is important to also consider polydispersity as this can "dramatically affect the microstructure and the effective properties of the materials" [**97**]. In this note we discuss the case of wide dispersity; that is, the case where the size ratio of the larger to the smaller objects grows large. One would expect boundary behavior to become negligible as the ratio of the radii tends to infinity, which intuitively means the density converges to $\Delta+(1-\Delta)\Delta$, see for instance [**96**]. To the best of our knowledge a proof of this has not yet been published. Here we provide such a proof which used standard techniques albeit it is not trivial.

We prove the following theorem:

THEOREM 5.1.1. *Suppose B is a body and $\mathcal{B}$ is a finite set of bodies. Then,*

$$\lim_{r\downarrow 0}\Delta_{\{rB\}\cup\mathcal{B}}=\Delta_{\mathcal{B}}+(1-\Delta_{\mathcal{B}})\Delta_{\{B\}}.$$

Here $\Delta_{\mathcal{B}}$ denotes the optimal packing density using the bodies from the set $\mathcal{B}$. In $\mathbb{R}^2$ this means that by taking B to be the unit disk and $\mathcal{B}=\{B\}$, the theorem says the optimal packing density using two sizes of disks converges to $0.9913\ldots$ as the ratio of the radii goes to infinity. In $\mathbb{R}^3$ this means the optimal packing density using two sizes of balls converges to $0.9326\ldots$ as the ratio of the radii goes to infinity.

5.2. Packings and density

We define a *body* to be a bounded subset of $\mathbb{R}^n$ that has nonempty interior and whose boundary has Lebesgue measure zero. Such a set B is Jordan measurable [**13**], which means it admits inner and outer approximations by simple sets whose measures are arbitrarily close to the Lebesgue measure of B. Moreover, since the interior of a body is nonempty it contains a nontrivial ball and hence has strictly positive Lebesgue measure.

A *packing* using a set of bodies $\mathcal{B}$ is a set of congruent copies of the elements in $\mathcal{B}$ such that the interiors of the copies are pairwise disjoint. In other words, a packing is of the form

$$P = \Big\{ R_i B_i + t_i : i \in \mathbb{N}, R_i \in O(n),\, t_i \in \mathbb{R}^n,\, B_i \in \mathcal{B} \Big\},$$

where $(R_i B_i^\circ + t_i) \cap (R_j B_j^\circ + t_j) = \emptyset$ for all $i \neq j$, where B_i° denotes the interior of B_i, and where $O(n)$ is the orthogonal group. Let $\Sigma_{\mathcal{B}}$ be the set of packings that use bodies from $\mathcal{B}$ and $\Lambda_{\mathcal{B}}$ the set of packings $P \in \Sigma_{\mathcal{B}}$ that have *rational box periodicity*; that is, for which there exists a $p \in \mathbb{Q}$ such that $|P| + pe_i = |P|$ for all $i \in [n]$, where $|P| = \bigcup P$ denotes the *carrier* of P, and where e_i is the ith unit vector.

The *density* and *upper density* (provided these exist) of a subset S of $\mathbb{R}^n$ are defined as

$$\rho(S) = \lim_{r\to\infty} \frac{\lambda(S \cap rC)}{r^n} \quad \text{and} \quad \overline{\rho}(S) = \limsup_{r\to\infty} \frac{\lambda(S \cap rC)}{r^n}.$$

Here λ is the Lebesgue measure on $\mathbb{R}^n$, and C is the unit cube centered about the origin. The upper density $\bar{\rho}(|P|)$ is defined for every $P \in \Sigma_{\mathcal{B}}$, because for each $r > 0$, the set $|P| \cap rC$ is Lebesgue measurable with measure at most r^n. The density $\rho(|P|)$ is defined for all $P \in \Lambda_{\mathcal{B}}$: Let $p \in \mathbb{Q}$ be a period of P, then $D = \lambda(|P| \cap kpC)/(kp)^n$ does not depend on $k \in \mathbb{N}$, and for r inbetween kp and $(k+1)p$ we have

$$\frac{\lambda(|P| \cap kpC)}{((k+1)p)^n} \leq \frac{\lambda(|P| \cap rC)}{r^n} \leq \frac{\lambda(|P| \cap kpC)}{((k+1)p)^n} + \frac{((k+1)p)^n - (kp)^n}{((k+1)p)^n},$$

where the rightmost term as well as $|\frac{\lambda(|P| \cap kpC)}{((k+1)p)^n} - D|$ converge to 0 as $k \to \infty$.

We define the *optimal packing density* for packings that use bodies from $\mathcal{B}$ by

$$\Delta_{\mathcal{B}} = \sup_{P \in \Sigma_{\mathcal{B}}} \overline{\rho}(|P|) = \sup_{P \in \Lambda_{\mathcal{B}}} \rho(|P|).$$

The second equality follows because for each $P \in \Sigma_{\mathcal{B}}$, we can construct a periodic packing with rational box periodicity whose density is arbitrarily close to $\overline{\rho}(|P|)$ by taking the subpacking contained in a sufficiently large cube and tiling space with this part of the packing. One could wonder whether the optimal density depends on the fact that C is a cube, but it follows from [**42**] that the optimal density $\Delta_{\mathcal{B}}$ is also equal to $\sup_{P \in \Lambda_{\mathcal{B}}} \lim_{r\to\infty} \lambda(|P| \cap (rB + t))/\lambda(rB)$, where t is any point in $\mathbb{R}^n$ and where B is any compact set that is the closure of its interior and contains the origin in its interior.

5.3. Packings of wide polydispersity

We first show that a packing, and hence the interstitial space of a packing, can be approximated uniformly by grid cubes. Given $S \subseteq \mathbb{R}^n$ and $k \in \mathbb{Z}$, define the packings

$$G_k(S) = \Big\{C_{k,t} : t \in \mathbb{Z}^n,\, C_{k,t} \subseteq S\Big\} \quad \text{and} \quad G^k(S) = \Big\{C_{k,t} : t \in \mathbb{Z}^n,\, C_{k,t} \cap S \neq \emptyset\Big\},$$

where $C_{k,t}$ is the cube $[\frac{t_1}{2^k}, \frac{t_1+1}{2^k}] \times \cdots \times [\frac{t_n}{2^k}, \frac{t_n+1}{2^k}]$ having side length $\frac{1}{2^k}$. Given a set P of subsets of $\mathbb{R}^n$, let $P^c = \mathbb{R}^n \setminus |P|$.

Lemma 5.3.1. *Let $\mathcal{B}$ be a finite set of bodies. Then*

$$\rho(|G_k(|P|)|) \uparrow \rho(|P|) \quad \textit{and} \quad \rho(|G^k(|P|)|) \downarrow \rho(|P|)$$

and hence

$$\rho(|G_k(P^c)|) \uparrow 1 - \rho(|P|) \quad \textit{and} \quad \rho(|G^k(P^c)|) \downarrow 1 - \rho(|P|)$$

as $k \to \infty$ for $P \in \Lambda_{\mathcal{B}}$ uniformly.

Proof. Let $\varepsilon > 0$ and $P \in \Lambda_{\mathcal{B}}$. Since P has rational box periodicity, the packings $G_k(|P|)$ and $G^k(|P|)$ have rational box periodicity, which means the densities $\rho(|G_k(|P|)|)$ and $\rho(|G^k(|P|)|)$ are defined. For each $k \in \mathbb{Z}$ we have

$$|G_k(|P|)| \subseteq |P| \subseteq |G^k(|P|)|,$$

hence

$$\rho(|G_k(|P|)|) \leq \rho(|P|) \leq \rho(|G^k(|P|)|).$$

We have

$$\rho(|G_k(|P|)|) = \lim_{r\to\infty} \frac{\lambda(|G_k(|P|)| \cap rC)}{r^n} \geq \lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P : B\subseteq rC} \lambda(|G_k(B)|)$$

and

$$\rho(|G^k(|P|)|) = \lim_{r\to\infty} \frac{\lambda(|G^k(|P|)| \cap rC)}{r^n} \leq \lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P : B\cap rC\neq\emptyset} \lambda(|G^k(B)|).$$

Every $B \in \mathcal{B}$ is Jordan measurable, so there exists a number $K = K(B, \varepsilon)$ such that

$$\lambda(|G_k(B)|) \geq \lambda(B) - \varepsilon \quad \text{and} \quad \lambda(|G^k(B)|) \leq \lambda(B) + \varepsilon$$

for all $k \geq K$. Since $\mathcal{B}$ is a finite set, this implies

$$\rho(|G_k(|P|)|) \geq \lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P : B\subseteq rC} (\lambda(B) - \varepsilon)$$

and

$$\rho(|G^k(|P|)|) \leq \lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P : B\cap rC\neq\emptyset} (\lambda(B) + \varepsilon)$$

for all $k \geq \max_{B\in\mathcal{B}} K(B,\varepsilon)$. Since each body B in the finite set $\mathcal{B}$ is bounded, there exists a number $r_0 = r_0(\mathcal{B}) \geq 0$ such that

$$\lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P: B\subseteq rC} \lambda(B) \geq \lim_{r\to\infty} \frac{\lambda(|P| \cap (r-r_0)C)}{r^n} = \lim_{r\to\infty} \frac{\lambda(|P| \cap rC)}{(r+r_0)^n} = \rho(|P|)$$

and

$$\lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P: B\cap rC\neq\emptyset} \lambda(B) \leq \lim_{r\to\infty} \frac{\lambda(|P| \cap (r+r_0)C)}{r^n} = \lim_{r\to\infty} \frac{\lambda(|P| \cap rC)}{(r-r_0)^n} = \rho(|P|).$$

Moreover, each body in the finite set $\mathcal{B}$ has nonempty interior, so there exists a constant $c = c(\mathcal{B})$ such that the number of congruent copies of elements from $\mathcal{B}$ that fit in a cube of radius $r + r_0$ is at most cr^n. Hence,

$$\lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P: B\subseteq rC} \varepsilon \leq c\varepsilon \quad \text{and} \quad \lim_{r\to\infty} \frac{1}{r^n} \sum_{B\in P: B\cap rC\neq\emptyset} \varepsilon \leq c\varepsilon.$$

Hence, for all $k \geq \max_{B\in\mathcal{B}} K(B,\varepsilon)$ we have

$$\rho(|G_k(|P|)|) \geq \rho(|P|) - c\varepsilon \quad \text{and} \quad \rho(|G^k(|P|)|) \leq \rho(|P|) + c\varepsilon,$$

which implies

$$\rho(|G_k(|P|)|) \uparrow \rho(|P|) \quad \text{and} \quad \rho(|G^k(|P|)|) \downarrow \rho(|P|),$$

as $k \to \infty$ for $P \in \Lambda_{\mathcal{B}}$ uniformly. This then implies

$$\rho(|G_k(P^c)|) = 1 - \rho(|G^k(|P|)|) \uparrow 1 - \rho(|P|)$$

and

$$\rho(|G^k(P^c)|) = 1 - \rho(|G_k(|P|)|) \downarrow 1 - \rho(|P|)$$

as $k \to \infty$ for $P \in \Lambda_{\mathcal{B}}$ uniformly. □

Let $\mathcal{B}$ and $\mathcal{B}'$ be sets of bodies. Given $P \in \Lambda_{\mathcal{B}}$, define

$$\Lambda_{\mathcal{B}'}(P) = \big\{Q \in \Lambda_{\mathcal{B}\cup\mathcal{B}'} : Q = P \cup R,\, R \in \Lambda_{\mathcal{B}'}\big\}.$$

The optimal density of such packings is given by $\Delta_{\mathcal{B}'}(P) = \sup_{Q\in\Lambda_{\mathcal{B}'}(P)} \rho(|Q|)$. In the following lemma we give the optimal density when a part of the packing is already fixed.

LEMMA 5.3.2. *Suppose B is a body and $\mathcal{B}$ is a finite set of bodies. Then*

$$\lim_{r\downarrow 0} \Delta_{\{rB\}}(P) = \rho(|P|) + (1 - \rho(|P|))\Delta_{\{B\}} \quad \textit{for} \quad P \in \Lambda_{\mathcal{B}} \quad \textit{uniformly.}$$

PROOF. Let $\varepsilon > 0$ and $P \in \Lambda_{\mathcal{B}}$. By Lemma 5.3.1 there exists an integer $K = K(\mathcal{B},\varepsilon)$ such that

$$\rho(|G_k(P^c)|) \geq 1 - \rho(|P|) - \varepsilon \quad \text{for all} \quad k \geq K.$$

By the definition of density there exists a scalar $R = R(k, B, \varepsilon) > 0$ such that for every $0 < r \leq R$ we can pack each cube in $G_k(P^c)$ with congruent copies of rB

with density at least $\Delta_{\{B\}} - \varepsilon$. Taking the union of these packings together with $\{P\}$ gives a packing from $\Delta_{\{rB\}}(P)$ which has density at least

$$\begin{aligned}\rho(|P|) + \rho(|G_k(P^c)|)(\Delta_{\{B\}} - \varepsilon) &\geq \rho(|P|) + (1 - \rho(|P|) - \varepsilon)(\Delta_{\{B\}} - \varepsilon) \\ &\geq \rho(|P|) + (1 - \rho(|P|))\Delta_{\{B\}} - 2\varepsilon.\end{aligned}$$

This implies

$$\Delta_{\{rB\}}(P) \geq \rho(|P|) + (1 - \rho(|P|))\Delta_{\{B\}} - 2\varepsilon \quad \text{for all} \quad 0 < r \leq R.$$

By Lemma 5.3.1 there exists an integer $K' = K'(\mathcal{B}, \varepsilon)$ such that

$$\rho(|G^k(P^c)|) \leq 1 - \rho(|P|) + \varepsilon \quad \text{for all} \quad k \geq K.$$

Again by the definition of density there exists a scalar $R' = R'(k, B, \varepsilon)$ such that for each $0 < r \leq R'$ and each $Q \in \Lambda_{\{rB\}}(P)$, the intersection of $Q \setminus P$ with a cube from $G^k(P^c)$ has density at most $\Delta_{\{B\}} + \varepsilon$ in that cube. So,

$$\begin{aligned}\rho(|Q|) = \rho(|P|) + \rho(|Q \setminus P|) &\leq \rho(|P|) + \rho(|G^k(P^c)|)(\Delta_{\{B\}} + \varepsilon) \\ &\leq \rho(|P|) + (1 - \rho(|P|) + \varepsilon)(\Delta_{\{B\}} + \varepsilon) \\ &\leq \rho(|P|) + (1 - \rho(|P|))\Delta_{\{B\}} + 2\varepsilon + \varepsilon^2,\end{aligned}$$

hence

$$\Delta_{\{rB\}}(P) \leq \rho(|P|) + (1 - \rho(|P|))\Delta_{\{B\}} + 2\varepsilon + \varepsilon^2 \quad \text{for all} \quad 0 < r \leq R'. \qquad \square$$

We prove the main theorem by using the uniform convergence in the above lemma:

Proof of Theorem 5.1.1. We have $\Lambda_{\{rB\}\cup\mathcal{B}} = \bigcup_{Q\in\Lambda_{\mathcal{B}}} \Lambda_{\{rB\}}(Q)$, so

$$\Delta_{\{rB\}\cup\mathcal{B}} = \sup_{P\in\Lambda_{\{rB\}\cup\mathcal{B}}} \rho(|P|) = \sup_{Q\in\Lambda_{\mathcal{B}}} \sup_{P\in\Lambda_{\{rB\}}(Q)} \rho(|P|) = \sup_{Q\in\Lambda_{\mathcal{B}}} \Delta_{\{rB\}}(Q),$$

and

$$\lim_{r\downarrow 0} \Delta_{\{rB\}\cup\mathcal{B}} = \lim_{r\downarrow 0} \sup_{Q\in\Lambda_{\mathcal{B}}} \Delta_{\{rB\}}(Q).$$

By Lemma 5.3.2, we have

$$\lim_{r\downarrow 0} \Delta_{\{rB\}}(Q) = \rho(|Q|) + (1 - \rho(|Q|))\Delta_{\{B\}},$$

and since convergence is uniform for $Q \in \Lambda_{\mathcal{B}}$, we can interchange limit and supremum and obtain

$$\begin{aligned}\lim_{r\downarrow 0} \sup_{Q\in\Lambda_{\mathcal{B}}} \Delta_{\{rB\}}(Q) &= \sup_{Q\in\Lambda_{\mathcal{B}}} \lim_{r\downarrow 0} \Delta_{\{rB\}}(Q) = \sup_{Q\in\Lambda_{\mathcal{B}}} (\rho(|Q|) + (1 - \rho(|Q|))\Delta_{\{B\}}) \\ &= \Delta_{\{B\}} + (1 - \Delta_{\{B\}}) \sup_{Q\in\Lambda_{\mathcal{B}}} \rho(|Q|) = \Delta_{\{B\}} + (1 - \Delta_{\{B\}})\Delta_{\mathcal{B}},\end{aligned}$$

and since $\Delta_{\{B\}} + (1 - \Delta_{\{B\}})\Delta_{\mathcal{B}} = \Delta_{\mathcal{B}} + (1 - \Delta_{\mathcal{B}})\Delta_{\{B\}}$ this completes the proof. $\square$

As a special case of the above theorem we obtain the result in the abstract of this note: We have $\lim_{r\downarrow 0} \Delta_{\{B,rB\}} = \Delta + (1 - \Delta)\Delta$, where B is the closed unit ball

and where $\Delta = \Delta_{\{B\}}$. By recursively applying the above theorem and rewriting the resulting expression we see that if $B_1, \ldots, B_k$ are bodies, then

$$\begin{aligned}\lim_{r_k \downarrow 0} \cdots \lim_{r_1 \downarrow 0} \Delta_{\{r_1 B_1, \ldots, r_k B_k\}} &= \Delta_{\{B_1\}} + (1 - \Delta_{\{B_1\}}) \lim_{r_k \downarrow 0} \cdots \lim_{r_2 \downarrow 0} \Delta_{\{r_2 B_2, \ldots, r_k B_k\}} \\ &= 1 - (1 - \Delta_{\{B_1\}}) \cdots (1 - \Delta_{\{B_k\}}).\end{aligned}$$

Moreover, if B is a body and $\mathcal{B} = \{rB : r > 0\}$, then for each k we have

$$\Delta_{\mathcal{B}} \geq \lim_{r_k \downarrow 0} \cdots \lim_{r_1 \downarrow 0} \Delta_{\{r_1 B, \ldots, r_k B\}} = 1 - (1 - \Delta_{\{B\}})^k,$$

so we get the intuitive result $\Delta_{\mathcal{B}} = 1$.

CHAPTER 6

A semidefinite programming hierarchy for packing problems in discrete geometry

This chapter is based on the publication *"D. de Laat, F. Vallentin, A semidefinite programming hierarchy for packing problems in discrete geometry, Math. Program., Ser. B 151 (2015), 529-553."*

Abstract. Packing problems in discrete geometry can be modeled as finding independent sets in infinite graphs where one is interested in independent sets which are as large as possible. For finite graphs one popular way to compute upper bounds for the maximal size of an independent set is to use Lasserre's semidefinite programming hierarchy. We generalize this approach to infinite graphs. For this we introduce topological packing graphs as an abstraction for infinite graphs coming from packing problems in discrete geometry. We show that our hierarchy converges to the independence number.

6.1. Packing problems in discrete geometry

Many, often notoriously difficult, problems in discrete geometry can be modeled as packing problems in graphs where the vertex set is an uncountable set having additional geometric structure.

The most famous example is the sphere packing problem in 3-dimensional space, the Kepler problem, which was solved by Hales [**44**] in 1998. Here the vertex set is $\mathbb{R}^3$ and two points are adjacent whenever their Euclidean distance is in the open interval $(0, 2)$.

An *independent set* of an undirected graph $G = (V, E)$ is a subset of the vertex set which does not span an edge. In the sphere packing case, an independent set corresponds to centers of unit balls which do not intersect in their interior. Now one is trying to find an independent set which covers as much space as possible. What "much" means depends on the situation. When the vertex set V, the *container*, is compact and when we pack identical shapes we can simply count and we use the *independence number*

$$\alpha(G) = \sup\{|I| : I \subseteq V,\ I \text{ is independent}\}.$$

If the objects are of different size we provide them with a weight $w(x)$ and we use the *weighted independence number*

$$\alpha_w(G) = \sup\Big\{ \sum_{x \in I} w(x) : I \subseteq V,\ I \text{ is independent} \Big\}.$$

In the non-compact sphere packing case one needs to use a density version of the independence number since maximal independent sets have infinite cardinality: The *(upper) point density* of an independent set $I \subset \mathbb{R}^3$ is

$$\delta(I) = \limsup_{R\to\infty} \frac{|I \cap [-R, R]^3|}{\mathrm{vol}([-R, R]^3)},$$

where $[-R, R]^3$ is the cube centered at the origin with side length $2R$. This measures the number of centers of unit balls per unit volume. To determine the geometric density of the corresponding sphere packing we multiply $\delta(I)$ by the volume of the unit ball.

More examples include:

— *Error correcting q-ary codes*: $V = \mathbb{F}_q^n$, where $\{x, y\} \in E$ if their Hamming distance lies in the open interval $(0, d)$. If $q = 2$ we speak about *binary codes* and if we restrict to all code words having the same Hamming norm we speak about *constant weight codes*.

— *Spherical codes:* $V = S^{n-1}$, where $\{x, y\} \in E$ if their inner product lies in the open interval $(\cos(\theta), 1)$.

— *Codes in real projective space:* $V = \mathbb{R}\mathrm{P}^{n-1}$, where $\{x, y\} \in E$ if their distance lies in the open interval $(0, d)$.

— *Sphere packings:* $V = \mathbb{R}^n$, where $\{x, y\} \in E$ if their Euclidean distance lies in the open interval $(0, 2)$.

— *Binary sphere packings:* $V = \mathbb{R}^n \times \{1, 2\}$ where $\{(x, i), (y, j)\} \in E$ if the Euclidean distance between x and y lies in the open interval $(0, r_i + r_j)$ and $w(x, i) = r_i^n \,\mathrm{vol}\, B_n$, where B_n is the unit ball.

— *Binary spherical cap packings:* $V = S^{n-1} \times \{1, 2\}$ where $\{(x, i), (y, j)\} \in E$ if the inner product of x and y lies in the open interval $(\cos(\theta_i + \theta_j), 1)$ and $w(x, i)$ is the volume of the spherical cap $\{z \in S^{n-1} : x \cdot z \geq \cos(\theta_i)\}$.

— *Packings of congruent copies of a convex body:* $V = \mathbb{R}^n \rtimes \mathrm{SO}(n)$ where (x, A) and (y, B) are adjacent if $x + A\mathcal{K}^\circ \cap y + B\mathcal{K}^\circ \neq \emptyset$, where $\mathcal{K}^\circ$ is the interior of the convex body $\mathcal{K}$.

Currently, these problems have been solved in only a few special cases. One might expect that they will never be solved in full generality, for all parameters. Finding good lower bounds by constructions and good upper bounds by obstructions are both challenging tasks. Over the last years the best known results were achieved with computer assistance: Algorithms like the adaptive shrinking cell scheme of Torquato and Jiao [**95**] generate dense packings and give very good lower bounds. The combination of semidefinite programming and harmonic analysis often gives the best known upper bounds for these packing problems. This method originated from work of Hoffman [**51**], Delsarte [**26**], and Lovász [**71**].

6.2. Lasserre's hierarchy for finite graphs

Computing the independence number of a finite graph is an NP-hard problem as shown by Karp [**55**]. Approximating optimal solutions of NP-hard problems in combinatorial optimization with the help of linear and semidefinite optimization is a very wide and active area of research. The most popular semidefinite programming hierarchies for NP-hard combinatorial optimization problems are the Lovász-Schrijver hierarchy [**72**] (the N^+-operator) and the hierarchy of Lasserre [**64**]. Laurent [**65**] showed that Lasserre's hierarchy is stronger than the Lovász-Schrijver hierarchy.

We now give a formulation of Lasserre's hierarchy for computing the independence number of a finite graph $G = (V, E)$. Here we follow Laurent [**65**]. The *t-th step of Lasserre's hierarchy* is:

$$\mathrm{las}_t(G) = \max\Big\{ \sum_{x\in V} y_{\{x\}} : y \in \mathbb{R}^{I_{2t}}_{\geq 0},\ y_\emptyset = 1,\ M_t(y) \text{ is positive semidefinite}\Big\},$$

where I_t is the set of all independent sets with at most t elements and where $M_t(y) \in \mathbb{R}^{I_t \times I_t}$ is the *moment matrix* defined by the vector y: Its (J, J')-entry equals

$$(M_t(y))_{J,J'} = \begin{cases} y_{J\cup J'} & \text{if } J \cup J' \in I_{2t}, \\ 0 & \text{otherwise.} \end{cases}$$

The first step in Lasserre's hierarchy coincides with the ϑ'-number, the strengthened version of Lovász ϑ-number [**71**] which is due to Schrijver [**86**]; for a proof see for instance the book by Schrijver [**88**, Theorem 67.11]. Furthermore the hierarchy converges to $\alpha(G)$ after at most $\alpha(G)$ steps:

$$\vartheta'(G) = \mathrm{las}_1(G) \geq \mathrm{las}_2(G) \geq \ldots \geq \mathrm{las}_{\alpha(G)}(G) = \alpha(G).$$

Lasserre [**64**] showed this convergence in the general setting of hierarchies for 0/1 polynomial optimization problems by using Putinar's Positivstellensatz [**81**]. Laurent [**65**] gave an elementary proof, which we discuss in Section 6.8.

Many variations are possible to set up a semidefinite programming hierarchy: For instance one can consider only "interesting" principal submatrices to simplify the computation and one can also add more constraints coming from problem specific arguments. In fact, in the definition of $\mathrm{las}_t(G)$ we used the nonnegativity constraints $y_S \geq 0$ for $S \in I_{2t}$. Even without them, the convergence result holds, and the first step in the hierarchy coincides with the Lovász ϑ-number.

A rough classification for all these variations can be given in terms of *n-point bounds*. This refers to all variations which make use of variables y_S with $|S| \leq n$. An n-point bound is capable of using obstructions coming from the local interaction of configurations having at most n points. For instance the Lovász ϑ-number is a 2-point bound and the t-th step in Lasserre's hierarchy is a $2t$-point bound. The relation between n-point bounds and Lasserre's hierarchy was first made explicit by Laurent [**66**] in the case of bounds for binary codes.

6.3. Topological packing graphs

The aim of this paper is to define and analyze a semidefinite programming hierarchy which upper bounds the independence number for infinite graphs arising

from packing problems in discrete geometry. For this we consider graphs where vertices which are close are adjacent, and where vertices which are adjacent will stay adjacent after small enough perturbations. These two conditions will be essential at many places in this paper. We formalize them by the following definition.

DEFINITION 6.3.1. *A graph whose vertex set is a Hausdorff topological space is called a* topological packing graph *if each finite clique is contained in an open clique. An* open clique *is an open subset of the vertex set where every two vertices are adjacent.*

It clearly suffices to verify the condition for cliques of size one and two.

Of course, every graph is a packing graph when we endow the vertex set with the discrete topology. However, weaker topologies give stronger conditions on the edge sets. For instance, when the vertex set of a topological packing graph is compact, then the independence number is finite because every single vertex is a clique.

A *distance graph* $G = (V, E)$ is a graph where (V, d) is a metric space, and where there exists $D \subseteq (0, \infty)$ such that x and y are adjacent precisely when $d(x, y) \in D$. If D is open and contains the interval $(0, \delta)$ for some $\delta > 0$, then G is a topological packing graph. That D contains an interval starting from 0 implies that vertices which are close are adjacent, and that D is open implies that adjacent vertices will stay adjacent after small enough perturbations. The binary spherical cap packing graph as defined in Section 6.1 is a compact topological packing graph with the usual topology on the vertex set $S^{n-1} \times \{1, 2\}$. And although there exists a metric compatible with this topology which gives the graph as a distance graph[1], it is easier and more natural to work directly with the topological packing graph structure.

Notice that in Definition 6.3.1 requiring all cliques to be contained in an open clique — which by Zorn's lemma is equivalent to all maximal cliques being open — would give a strictly stronger condition.[2]

6.4. Generalization of Lasserre's hierarchy

Now we introduce our generalization of Lasserre's hierarchy for compact topological packing graphs.

Before we go into the technical details we like to comment on the choice of spaces in our generalization: In Lasserre's hierarchy for finite graphs the optimization variable y lies in the cone[3] $\mathbb{R}^{I_{2t}}_{\geq 0}$. One might try to use the same cone when I_{2t} is uncountable. But then there are too many variables and it is impossible to express the objective function. At the other extreme one might try to restrict this cone to finitely (or countably) supported vectors. But then we do not know how to develop a duality theory like the one in Section 6.7. A duality theory is important for concrete computations: Minimization problems can be used to derive upper bounds

[1]Assume $\theta_1 < \theta_2$ and let ε be some number strictly between $(1 - \theta_1/\theta_2)/2$ and 1. Let $D = (0, 1)$, and let $d((x, i), (y, j))$ be given by $\varepsilon\delta_{i \neq j} + (1 - \varepsilon\delta_{i \neq j}) \arccos(x \cdot y)\,(\theta_1 + \theta_2)^{-1}$ when $x \cdot y < \cos(\theta_i + \theta_j)$ and 1 otherwise.

[2]Consider the graph with vertex set $[0, 1] \times \mathbb{Z}$ where (x, i) and (y, j) are adjacent if $i = j$ or when x and y are both strictly smaller than $|i - j|^{-1}$ (for $i \neq j$). Here each finite clique is contained in an open clique, but the countable clique $\{0\} \times \mathbb{Z}$ is not.

[3]In this paper cones are always assumed to be convex.

rigorously. We use a cone of Borel measures where we have "one degree of freedom" for every open set.

In Section 6.6 we use the topology of V to equip the set I_t, consisting of the independent sets which have at most t elements, with a Hausdorff topology. There we also use the topological packing graph condition to show that I_t is compact.

Let $\mathcal{C}(I_{2t})$ be the set of continuous real-valued functions on I_{2t}. By the Riesz representation theorem (see e.g. [**11**, Chapter 2.2]) the topological dual of $\mathcal{C}(I_{2t})$, where the topology is defined by the supremum norm, can be identified with the space $\mathcal{M}(I_{2t})$ of signed Radon measures. A *signed Radon measure* is the difference of two Radon measures, where a *Radon measure* ν is a locally finite measure on the Borel algebra satisfying *inner regularity*: $\nu(B) = \sup\{\nu(C) : C \subseteq B, C \text{ compact}\}$ for each Borel set B. Nonnegative functions in $\mathcal{C}(I_{2t})$ form the cone $\mathcal{C}(I_{2t})_{\geq 0}$. Its conic dual $(\mathcal{C}(I_{2t})_{\geq 0})^*$ is the cone of *positive Radon measures*

$$\mathcal{M}(I_{2t})_{\geq 0} = \{\lambda \in \mathcal{M}(I_{2t}) : \lambda(f) \geq 0 \text{ for all } f \in \mathcal{C}(I_{2t})_{\geq 0}\}.$$

Denote by $\mathcal{C}(I_t \times I_t)_{\text{sym}}$ the space of *symmetric kernels*, which are the continuous functions $K \colon I_t \times I_t \to \mathbb{R}$ such that

$$K(J, J') = K(J', J) \text{ for all } J, J' \in I_t.$$

We say that a symmetric kernel K is *positive definite* if

$$(K(J_i, J_j))_{i,j=1}^m \text{ is positive semidefinite for all } m \in \mathbb{N} \text{ and } J_1, \ldots, J_m \in I_t.$$

The positive definite kernels form the cone $\mathcal{C}(I_t \times I_t)_{\succeq 0}$. The dual of $\mathcal{C}(I_t \times I_t)_{\text{sym}}$ can be identified with the space of symmetric signed Radon measures $\mathcal{M}(I_t \times I_t)_{\text{sym}}$. Here a signed Radon measure $\mu \in \mathcal{M}(I_t \times I_t)$ is *symmetric* if

$$\mu(E \times E') = \mu(E' \times E) \text{ for all Borel sets } E \text{ and } E'.$$

We say that a measure $\mu \in \mathcal{M}(I_t \times I_t)_{\text{sym}}$ is *positive definite* if it lies in the dual cone $\mathcal{M}(I_t \times I_t)_{\succeq 0} = (\mathcal{C}(I_t \times I_t)_{\succeq 0})^*$.

Now we are ready to define our generalization:

— The optimization variable is $\lambda \in \mathcal{M}(I_{2t})_{\geq 0}$.

— The objective function evaluates λ at $I_{=1}$, where in general,

$$I_{=t} = \{S \in I_t : |S| = t\},$$

and so when $t = 1$ we simply deal with all vertices, as singleton sets. This is similar to the objective function $\sum_{x \in V} y_{\{x\}}$ in Lasserre's hierarchy for finite graphs.

— The normalization condition reads $\lambda(\{\emptyset\}) = 1$.

— For generalizing the moment matrix condition "$M_t(y)$ is positive semidefinite" we use a dual approach. Let T_t be the operator such that for all vectors y and all matrices Y we have $\langle M_t(y), Y\rangle_1 = \langle y, T_t Y\rangle_2$, where $\langle \cdot, \cdot \rangle_1$ is the trace inner product of matrices and $\langle \cdot, \cdot \rangle_2$ is standard vector inner

product. Instead of directly generalizing the operator M_t, we will dualize the following generalization of the operator T_t:

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

We have $\|A_t K\|_\infty \leq 2^{2t} \|K\|_\infty$, so A_t is bounded and hence continuous. Thus there exists the adjoint $A_t^* \colon \mathcal{M}(I_{2t}) \to \mathcal{M}(I_t \times I_t)_{\mathrm{sym}}$ and the moment matrix condition reads $A_t^* \lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0}$.

Definition 6.4.1. *The t-th step of the* generalized hierarchy *is*

$$\mathrm{las}_t(G) = \sup \Big\{ \lambda(I_{=1}) : \lambda \in \mathcal{M}(I_{2t})_{\geq 0},\ \lambda(\{\emptyset\}) = 1,\ A_t^* \lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0} \Big\}.$$

Clearly, we have a nonincreasing chain

$$\mathrm{las}_1(G) \geq \mathrm{las}_2(G) \geq \ldots \geq \mathrm{las}_{\alpha(G)}(G) = \mathrm{las}_{\alpha(G)+1}(G) = \ldots \tag{25}$$

which stabilizes after $\alpha(G)$ steps, and specializes to the original hierarchy if G is a finite graph. Each step gives an upper bound for $\alpha(G)$ because for every independent set S the measure

$$\lambda = \sum_{R \in I_{2t} : R \subseteq S} \delta_R, \quad \text{where } \delta_R \text{ is the delta measure at } R,$$

is a feasible solution for $\mathrm{las}_t(G)$ with objective value $|S|$. To see this we note that $\lambda(\{\emptyset\}) = 1$, and for any $K \in \mathcal{C}(I_t \times I_t)_{\succeq 0}$ we have

$$\begin{aligned} \langle K, A_t^* \lambda \rangle = \langle A_t K, \lambda \rangle &= \sum_{R \in I_{2t} : R \subseteq S} \ \sum_{J,J' \in I_t : J \cup J' = R} K(J, J') \\ &= \sum_{J,J' \in I_t : J, J' \subseteq S} K(J, J') \geq 0. \end{aligned}$$

In Section 6.7 we consider the dual program of $\mathrm{las}_t(G)$, which is

$$\begin{aligned} \mathrm{las}_t(G)^* = \inf \Big\{ K(\emptyset, \emptyset) : K \in \mathcal{C}(I_t \times I_t)_{\succeq 0}, \\ A_t K(S) \leq -1_{I_{=1}}(S) \text{ for } S \in I_{2t} \setminus \{\emptyset\} \Big\}, \end{aligned}$$

and we show that strong duality holds in every step:

Theorem 6.4.2. *Let G be a compact topological packing graph. For every $t \in \mathbb{N}$ we have $\mathrm{las}_t(G) = \mathrm{las}_t(G)^*$, and if $\mathrm{las}_t(G)$ is finite*[4]*, then the optimum in $\mathrm{las}_t(G)$ is attained.*

In Section 6.8 we show that the chain (25) converges to the independence number:

Theorem 6.4.3. *Let G be a compact topological packing graph. Then,*

$$\mathrm{las}_{\alpha(G)}(G) = \alpha(G).$$

[4]We show this in Remark 6.9.4.

A variation of $\text{las}_t(G)$ can be used to upper bound the weighted independence number of a weighted compact topological packing graph G with a continuous weight function $w\colon V \to \mathbb{R}_{\geq 0}$. We extend w, with the obvious abuse of notation, to a function $w\colon I_{2t} \to \mathbb{R}_{\geq 0}$ where only singleton sets have positive weight. It turns out, by Lemma 6.6.2, that also the extension is continuous. Then we replace the objective function $\lambda(I_{=1})$ by $\lambda(w)$.

6.5. Explicit computations in the literature

Explicit computations of n-point bounds have been done in a variety of situations. The following table provides a guide to the literature:

Packing problem	**2-point bound**	**3-point bound**	**4-point bound**
Binary codes	Delsarte [**26**]	Schrijver [**89**]	Gijswijt, Mittelmann, Schrijver [**40**]
q-ary codes	Delsarte [**26**]	Gijswijt, Schrijver, Tanaka [**41**]	
Constant weight codes	Delsarte [**26**]	Schrijver [**89**], Regts [**82**]	
Spherical codes	Delsarte, Goethals, Seidel [**27**]	Bachoc, Vallentin [**9**]	
Codes in $\mathbb{R}\text{P}^{n-1}$	Kabatiansky, Levenshtein [**54**]	Cohn, Woo [**23**]	
Sphere packings	Cohn, Elkies [**21**]		
Binary sphere and spherical cap packings	de Laat, Oliveira, Vallentin [**60**]		
Congruent copies of a convex body	Oliveira, Vallentin [**80**]		

For the first three packing problems in this table one can use Lasserre's hierarchy for finite graphs. For the last five packing problems in this table our generalization can be used, where in the last three cases one has to perform a compactification of the vertex set first.

We elaborate on the connection between these n-point bounds and our hierarchy in Section 6.9. The convergence of the hierarchy, shows that this approach is in theory capable of solving any given packing problem in discrete geometry. One attractive feature of the hierarchy is that already its first steps give strong upper bounds as one can see from the papers cited in the table above.

6.6. Topology on sets of independent sets

Let $G = (V, E)$ be a topological packing graph. In this section we introduce a topology on I_t, the set of independent sets having cardinality at most t.

We equip the direct product V^t with the product topology and the image of V^t under the map

$$q\colon (v_1, \ldots, v_t) \mapsto \{v_1, \ldots, v_t\}$$

with the quotient topology. When we add the empty set to the image we obtain the collection $\mathrm{sub}_t(V)$ of all subsets of V of cardinality at most t, which obtains its topology from the disjoint union topology. Compactness of $\mathrm{sub}_t(V)$ follows immediately from compactness of V. Handel [**46**, Proposition 2.7] shows that it is Hausdorff.

Given $U_1, \dots, U_r \subseteq V$, define

$$(U_1, \dots, U_r)_t = \{S \in \mathrm{sub}_t(V) : S \subseteq U_1 \cup \dots \cup U_r,\ S \cap U_i \neq \emptyset \text{ for } 1 \leq i \leq r\}.$$

Handel [**46**] observes

$$q^{-1}((U_1, \dots, U_r)_t) = \bigcup_{\substack{\tau:\{1,\dots,t\}\to\{1,\dots,r\} \\ \tau \text{ surjective}}} U_{\tau(1)} \times \dots \times U_{\tau(t)}.$$

This shows that if the sets U_i are open, then $(U_1, \dots, U_r)_t$ is open. In fact, if $\mathcal{B}$ is a base for V, then

$$\mathcal{B}_t = \{(U_1, \dots, U_r)_t : 1 \leq r \leq t,\ U_1, \dots, U_r \in \mathcal{B}\}$$

is a base for $\mathrm{sub}_t(V)$. Moreover, if $\{u_1, \dots, u_r\}$ is an element in an open set U in $\mathrm{sub}_t(V)$, then there are open neighborhoods U_i of u_i such that the open neighborhood $(U_1, \dots, U_r)_t$ of $\{u_1, \dots, u_r\}$ is a contained in U.

We now endow I_t with a topology as a subset of $\mathrm{sub}_t(V)$. Clearly, $I_{=1}$ is homeomorphic to V. It is also immediate that I_t is Hausdorff. Furthermore, it is compact:

LEMMA 6.6.1. *Let $G = (V, E)$ be a compact topological packing graph. Then I_t is compact for every $t \in \mathbb{N}$.*

PROOF. We will show that I_t is closed, respectively that its complement $D_t = \mathrm{sub}_t(V) \setminus I_t$ is open in the compact space $\mathrm{sub}_t(V)$. Let $\{x_1, \dots, x_r\} \in D_t$ be arbitrary. Without loss of generality we may assume that x_1 and x_2 are adjacent. By the topological packing graph condition there exists an open clique $U \subseteq V$ containing both x_1 and x_2. Since V is a Hausdorff space there exist disjoint open sets U_1 and U_2 such that $x_1 \in U_1 \subseteq U$ and $x_2 \in U_2 \subseteq U$. Each set in $(U_1, U_2, V, \dots, V)_t$ contains at least one edge, so $(U_1, U_2, V, \dots, V)_t \subseteq D_t$. The set $(U_1, U_2, V, \dots, V)_t$ is an open neighborhood of $\{x_1, \dots, x_r\}$. Hence, D_t is open. □

If the topology on V comes from a metric, then the topology on $\mathrm{sub}_t(V)$ is given by the Hausdorff distance, see for example Borsuk and Ulam [**17**]. This indicates that subsets of nonequal cardinality can be close in the topology on $\mathrm{sub}_t(V)$. However, in the following lemma, we use the topological packing graph condition to show that independent sets of different cardinality are in different connected components of I_t.

LEMMA 6.6.2. *Let $G = (V, E)$ be a topological packing graph. The map $I_t \to \mathbb{N}$, $S \mapsto |S|$ is continuous for every $t \in \mathbb{N}$. In particular, $I_{=t}$ is both open and closed.*

PROOF. Let $\{S_\alpha\}$ be a net in I_t converging to $\{x_1, \dots, x_r\} \in I_t$, where we assume the x_i to be pairwise different. By the topological packing graph condition, there exist pairwise disjoint open cliques U_i such that $x_i \in U_i$. The set $(U_1, \dots, U_r)_t$

is open and contains $\{x_1, \dots, x_r\}$. Hence, we eventually have $S_\alpha \in (U_1, \dots, U_r)_t$. Then $|S_\alpha| \geq r$ since the U_i are pairwise disjoint and $|S_\alpha| \leq r$ since the U_i are cliques. □

6.7. Duality theory of the generalized hierarchy

6.7.1. A primal-dual pair. In this section we derive the dual program of the t-th step in our hierarchy $\mathrm{las}_t(G)$.

We want to have a symmetric situation between primal and dual. We consider the dual pairs $(\mathcal{C}(I_{2t}), \mathcal{M}(I_{2t}))$ and $(\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}, \mathcal{M}(I_t \times I_t)_{\mathrm{sym}})$ together with the corresponding nondegenerate bilinear forms

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

We endow the spaces with the weakest topologies compatible with the pairing: the weak topology on the function spaces and the weak* topology on the measure spaces. From now on we will always use these topologies unless explicitly stated otherwise. Because the cones defined in Section 6.4 are closed, it follows from the bipolar theorem that

$$(\mathcal{M}(I_{2t})_{\geq 0})^* = \mathcal{C}(I_{2t})_{\geq 0} \quad \text{and} \quad (\mathcal{M}(I_t \times I_t)_{\succeq 0})^* = \mathcal{C}(I_t \times I_t)_{\succeq 0}.$$

Hence, the situation is completely symmetric.

Recall that the operator

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

is continuous in the norm topologies, so it follows that it is continuous in the weak topologies. In the next subsection we use that its adjoint A_t^* is injective:

LEMMA 6.7.1. *Let $G = (V, E)$ be a compact topological packing graph. Then the operator A_t is surjective for every $t \in \mathbb{N}$.*

PROOF. Let g be a function in $\mathcal{C}(I_{2t})$. The continuity of

$$u \colon I_t \times I_t \to \mathrm{sub}_{2t}(V),\ (J, J') \mapsto J \cup J'$$

follows from [**46**]. Hence

$$h \colon u^{-1}(I_{2t}) \to \mathbb{R},\ (J, J') \mapsto \frac{g(J \cup J')}{A_t \mathcal{J}(J \cup J')}$$

is continuous where $\mathcal{J}$ is the kernel which evaluates to 1 everywhere.

The set I_{2t} is closed in $\mathrm{sub}_{2t}(V)$, so the preimage $u^{-1}(I_{2t})$ is closed in $I_t \times I_t$. Since $I_t \times I_t$ is a compact Hausdorff space there exists, by Tietze's extension theorem, a function $H \in \mathcal{C}(I_t \times I_t)$ such that $H(J, J') = h(J, J')$ for all $J, J' \in I_t$. For each $S \in I_{2t}$ we then have

$$\begin{aligned} A_t H(S) &= \sum_{J, J' \in I_t : J \cup J' = S} H(J, J') = \sum_{J, J' \in I_t : J \cup J' = S} h(J, J') \\ &= \frac{1}{A_t \mathcal{J}(S)} \sum_{J, J' \in I_t : J \cup J' = S} g(J \cup J') = g(S). \end{aligned}$$

□

Using the theory of duality in conic optimization problems, see for instance Barvinok [**10**], we derive the dual hierarchy:

$$\operatorname{las}_t(G)^* = \inf\Big\{K(\emptyset,\emptyset) : K \in \mathcal{C}(I_t \times I_t)_{\succeq 0}, \\ A_t K(S) \le -1_{I_{=1}}(S) \text{ for } S \in I_{2t} \setminus \{\emptyset\}\Big\},$$

where one should note that by Lemma 6.6.2 the characteristic function $1_{I_{=1}}$ is continuous. It follows from weak duality that $\operatorname{las}_t(G) \le \operatorname{las}_t(G)^*$, and hence $\operatorname{las}_t(G)^*$ upper bounds the independence number. In the following lemma we give a simple direct proof.

LEMMA 6.7.2. *Let $G = (V, E)$ be a compact topological packing graph. Then*

$$\alpha(G) \le \operatorname{las}_t(G)^*$$

holds for all $t \in \mathbb{N}$.

PROOF. Suppose K is feasible and L is an independent set. Then

$$0 \le \sum_{J,J' \in \operatorname{sub}_t(L)} K(J,J') = \sum_{S \in \operatorname{sub}_{2t}(L)} A_t K(S) \\ = K(\emptyset,\emptyset) + \sum_{x \in L} A_t K(\{x\}) + \sum_{S \in \operatorname{sub}_{2t}(L) \setminus \operatorname{sub}_1(L)} A_t K(S) \le K(\emptyset,\emptyset) - |L|. \qquad \square$$

The hierarchy $\operatorname{las}_t(G)^*$ stabilizes after $\alpha(G)$ steps, because the variables and constraints are the same for each $t \ge \alpha(G)$. By Lemma 6.6.2 the set I_t is both open and closed in I_{t+1}, which means we can extend a feasible kernel K of $\operatorname{las}_t(G)^*$ by zeros to obtain a feasible solution to $\operatorname{las}_{t+1}(G)^*$ with the same objective value. This shows that the hierarchy is nonincreasing; that is, $\operatorname{las}_{t+1}(G)^* \le \operatorname{las}_t(G)^*$ for all t. These results also follow from strong duality as discussed next.

6.7.2. Strong duality. In this section we prove Theorem 6.4.2: We have strong duality between the problems $\operatorname{las}_t(G)$ and $\operatorname{las}_t(G)^*$. We will show the finiteness of $\operatorname{las}_t(G)^*$ in Remark 6.9.4.

For proving Theorem 6.4.2 we make use of a closed cone condition, which for example is explained in Barvinok [**10**, Chapter IV.7]. For this we have to show that $\operatorname{las}_t(G)$ has a feasible solution, which we already know from Section 6.4, and that the cone

$$K = \big\{(A_t^*\xi - \mu, \xi(\{\emptyset\}), \xi(I_{=1})) : \mu \in \mathcal{M}(I_t \times I_t)_{\succeq 0},\ \xi \in \mathcal{M}(I_{2t})_{\ge 0}\big\}$$

is closed in $\mathcal{M}(I_t \times I_t)_{\mathrm{sym}} \times \mathbb{R} \times \mathbb{R}$. The above cone is the Minkowski difference of

$$K_1 = \big\{(A_t^*\xi, \xi(\{\emptyset\}), \xi(I_{=1})) : \xi \in \mathcal{M}(I_{2t})_{\ge 0}\big\}$$

and

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

By a theorem of Klee [**56**] and Dieudonné [**28**] the Minkowski difference $K_1 - K_2$ is closed when the three conditions

(A) $K_1 \cap K_2 = \{0\}$,
(B) K_1 and K_2 are closed,
(C) K_1 is locally compact.

are satisfied. The fact that K_2 is closed follows immediately since $\mathcal{M}(I_t \times I_t)_{\succeq 0}$ is closed. We now verify the other conditions:

LEMMA 6.7.3. $K_1 \cap K_2 = \{0\}$.

PROOF. We will show that $\xi \in \mathcal{M}(I_{2t})_{\geq 0}$ with $\xi(\{\emptyset\}) = 0$ is the zero measure if $A_t^*\xi \in \mathcal{M}(I_t \times I_t)_{\succeq 0}$.

Let $f \in \mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$ be given by

$$f(J, J') = \begin{cases} 1 & \text{if } J = J' = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Then $A_t^*\xi(\{(\emptyset, \emptyset)\}) = \langle f, A_t^*\xi \rangle = \langle A_t f, \xi \rangle = \xi(\{\emptyset\}) = 0$.

For $n \in \mathbb{Z}$ define $g_n \in \mathcal{C}(I_t)$ by

$$g_n(S) = \begin{cases} |n| & \text{if } S = \emptyset, \\ 1/n & \text{otherwise.} \end{cases}$$

Since $g_n \otimes g_n \in \mathcal{C}(I_t \times I_t)_{\succeq 0}$ and $A_t^*\xi \in \mathcal{M}(I_t \times I_t)_{\succeq 0}$ we have $A_t^*\xi(g_n \otimes g_n) \geq 0$. We have that $A_t^*\xi(g_n \otimes g_n)$ equates to

$$n^2 A_t^*\xi\big(\{(\emptyset, \emptyset)\}\big) \; + \; \frac{1}{n^2} A_t^*\xi\big(I_t \setminus \{\emptyset\} \times I_t \setminus \{\emptyset\}\big) \; + \; 2\,\mathrm{sign}(n) A_t^*\xi\big(\{\emptyset\} \times I_t \setminus \{\emptyset\}\big).$$

The first term is zero, so the sum of the last two terms is nonnegative for each n. By letting n tend to plus and minus infinity we see that $A_t^*\xi(\{\emptyset\} \times I_t \setminus \{\emptyset\}) = 0$.

Define $h \in \mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$ by

$$h(J, J') = \begin{cases} 1 & \text{if } J = \emptyset \text{ and } J' = \emptyset, \\ 1/2 & \text{if } J = \emptyset \text{ or } J' = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Since ξ is a positive measure we have $\|\xi\| = \xi(I_{2t})$, but

$$\xi(I_{2t}) = \langle A_t h, \xi \rangle = \langle h, A_t^*\xi \rangle = A_t^*\xi(\{(\emptyset, \emptyset)\}) + A_t^*\xi(\{\emptyset\} \times I_t \setminus \{\emptyset\}) = 0,$$

so $\xi = 0$. □

REMARK 6.7.4. The set I_{2t} is a subset of the power set 2^V. A power set is a *monoid* with the associative binary operation $\cup$ and unit element $\emptyset$. Monoids have sufficient structure for defining functions of *positive type*, which in this case are functions $f\colon 2^V \to \mathbb{R}$ for which the matrices $(f(J_i \cup J_j))_{i,j=1}^m$ are positive semidefinite for all $m \in \mathbb{N}$ and $J_1, \ldots, J_m \in 2^V$. This monoid is *commutative* (i.e., $J \cup J' = J' \cup J$ for all $J, J' \in 2^V$) and *idempotent* (i.e., $J \cup J = J$ for all $J \in 2^V$), so the matrix

$$\begin{pmatrix} f(\emptyset) & f(J) \\ f(J) & f(J) \end{pmatrix} \quad \text{is positive semidefinite,}$$

and so $0 \leq f(J) \leq f(\emptyset)$ for all $J \in 2^V$ [**11**, p. 119]. In particular, a function of positive type which vanishes at the unit element is identically zero. This resembles the situation in the proof of Lemma 6.7.3. To see this we show that one can view $\lambda \in \mathcal{M}(I_{2t})$ with $A_t^*\lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0}$ as a "measure of positive type". For this we notice that a function $f\colon 2^V \to \mathbb{R}$ is of positive type if and only if

$\sum_{S\in 2^V} f(S)\sum_{J\cup J'=S} g(J)g(J') \geq 0$ for all finitely supported functions $g\colon 2^V \to \mathbb{R}$. Going from the monoid 2^V to the "truncated monoid" I_{2t}, and from functions to measures, we have the natural definition that a measure $\lambda \in \mathcal{M}(I_{2t})$ is of positive type if $\int A_t(g\otimes g)(S)\, d\lambda(S) \geq 0$ for all $g \in \mathcal{C}(I_{2t})$, which is the case if and only if $A_t^*\lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0}$. Moreover, if we define a convolution and an involution on $\mathcal{C}(I_{2t})$ by $f * g = A_t(f\otimes g)$ and $f^* = f$, respectively, then a measure λ is of positive type if and only if $\lambda(f^* * f) \geq 0$ for all $f \in \mathcal{C}(I_{2t})$. This agrees with the definition of measures of positive type as given for instance in [**33**, Chapter 6.3] for locally compact groups, where a different algebra is used.

Before we consider condition (C) we need some background: A cone is *locally compact* if it is locally compact as a topological space, that is, each point in the cone is contained in a compact neighborhood relative to the cone. A cone is locally compact if the origin has a compact neighborhood relative to the cone: For each point x in the cone and each neighborhood U of the origin there is an $r > 0$ such that $x \in rU$. A *convex base* B of a cone K is a convex subset of the cone such that every nonzero $x \in K$ can be written in a unique way as a positive multiple of an element in B. A cone is *pointed* if it does not contain a line. Now we can state a theorem of Klee and Dieudonné [**56**, (2.4)]: A nonempty pointed cone in a locally convex vector space is closed and locally compact if and only if it admits a compact convex base.

LEMMA 6.7.5. *K_1 is closed and locally compact.*

PROOF. Set

$$B = \{\xi \in \mathcal{M}(I_{2t})_{\geq 0} : \langle 1_{I_{2t}}, \xi\rangle = 1\}.$$

The map

$$\mathcal{M}(I_{2t}) \to \mathbb{R},\ \xi \mapsto \langle 1_{I_{2t}}, \xi\rangle$$

is continuous, so the preimage of $\{1\}$ under this is closed. Hence, B is closed in the space of probability measures on I_{2t}, which is compact by the Banach-Alaoglu theorem. So, B is compact as well.

By Lemma 6.7.1 A_t^* is injective, so the map $\xi \mapsto (A_t^*\xi, \xi(\{\emptyset\}), \xi(I_{=1}))$ is injective and the image of B under this map is a compact convex base for K_1. Hence, by Klee, Dieudonné, the cone K_1 is closed and locally compact. □

REMARK 6.7.6. In this remark we show that for infinite graphs the cone K_2 is not locally compact, and hence it is important that only one of the two cones is required to be locally compact in condition (C). If V is an infinite set, then so is I_t, which means that $\mathcal{M}(I_t)$ is an infinite dimensional (Hausdorff) topological vector space which is therefore not locally compact. The Banach-Alaoglu theorem says that the closed ball of radius r centered about the origin in $\mathcal{M}(I_t)$ is compact. This means that it cannot be a neighborhood of the origin. Thus, for each $r > 0$ there exists a net $\{\lambda_\beta\} \subseteq \mathcal{M}(I_t)$ converging to the origin, such that $\|\lambda_\beta\| = r$ for all β.

Let $f \in \mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$ and $\varepsilon > 0$. The set

$$\operatorname{span}\{c\, g\otimes g : c \in \mathbb{R},\ g \in \mathcal{C}(I_t)\}$$

is a point separating and nowhere vanishing subalgebra of $\mathcal{C}(I_t \times I_t)_{\mathrm{sym}}$, so it follows from the Stone-Weierstrass theorem that it is dense in the uniform topology. This

means that there exists a function $\tilde{f} = \sum_{i=1}^m c_i g_i \otimes g_i$ such that $\|\tilde{f} - f\|_\infty \leq \varepsilon/r^2$. Then,

$$\begin{aligned}|\lambda_\beta \otimes \lambda_\beta(f)| &\leq |\lambda_\beta \otimes \lambda_\beta(f) - \lambda_\beta \otimes \lambda_\beta(\tilde{f})| + |\lambda_\beta \otimes \lambda_\beta(\tilde{f})| \\ &\leq \|\lambda_\beta \otimes \lambda_\beta\| \|f - \tilde{f}\|_\infty + \sum_{i=1}^m c_i \lambda_\beta(g_i)^2 \to \varepsilon.\end{aligned}$$

So, the net $\{\lambda_\beta \otimes \lambda_\beta\}$ in $\mathcal{M}(I_t \times I_t)_{\succeq 0}$, which satisfies $\|\lambda_\beta \otimes \lambda_\beta\| = r^2$ for each β, converges to the origin. Therefore, none of the closed balls centered about the origin is a neighborhood of the origin in $\mathcal{M}(I_t \times I_t)_{\succeq 0}$. Since compact sets are bounded, this means that the origin does not have a compact neighborhood in $\mathcal{M}(I_t \times I_t)_{\succeq 0}$, so this cone is not locally compact and neither is K_2.

6.8. Convergence to the independence number

In this section we prove Theorem 6.4.3: The chain (25) converges to the independence number $\alpha(G)$.

Our proof can be seen as an infinite-dimensional version of Laurent's proof of the convergence of the hierarchy for finite graphs $G = (V, E)$. In [**65**] she makes use of the fact that the cone of positive semidefinite moment matrices where rows and columns are indexed by the power set 2^V is a simplicial polyhedral cone; an observation due to Lindström [**69**] and Wilf [**99**]. More specifically,

$$\{M \in \mathbb{R}^{2^V \times 2^V} : M \succeq 0, M \text{ is a moment matrix}\} = \operatorname{cone}\{\chi_S \chi_S^\mathsf{T} : S \subseteq V\}, \tag{26}$$

where a moment matrix M is a matrix where the entry $M_{J,J'}$ only depends on the union $J \cup J'$ and where the vector $\chi_S \in \mathbb{R}^{2^V}$ is defined componentwise by

$$\chi_S(R) = \begin{cases} 1 & \text{if } R \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

The proof of (26) uses the inclusion-exclusion principle. In our proof the following form of the inclusion-exclusion principle will be crucial: Given finite sets A and C,

$$\begin{aligned}\sum_{B : A \subseteq B \subseteq C} (-1)^{|B|} &= (-1)^{|A|} \sum_{B \subseteq C \setminus A} (-1)^{|B|} \\ &= (-1)^{|A|} \sum_{i=0}^{|C \setminus A|} \binom{|C \setminus A|}{i} 1^{|C \setminus A| - i} (-1)^i \\ &= (-1)^{|A|} (1-1)^{|C \setminus A|} = \begin{cases} (-1)^{|A|} & \text{if } A = C, \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

In our proof we are also faced with two analytical difficuties because we consider infinite graphs: 1. The cone $\{A_t^* \lambda : \lambda \in \mathcal{M}(I_{2t})\} \cap \mathcal{M}(I_t \times I_t)_{\succeq 0}$ is not finitely generated. 2. Also the power set 2^V is too large.

The second problem we solve by considering the set $I = I_{\alpha(G)}$ instead of 2^V. In fact, already when we defined the hierarchy we used measures on independent sets instead of measures on all subsets of the vertices.

The first problem we solve by using weak vector valued integrals (as discussed in for instance [**33**, Appendix 3]) instead of finite conic combinations: Let $\tau \in \mathcal{M}(I)$ and $\nu_S \in \mathcal{M}(I)$ so that $S \mapsto \nu_S$ is a continuous map from I to $\mathcal{M}(I)$ with $\sup_{S\in I} \|\nu_S\| < \infty$. Then $f \mapsto \int \nu_S(f)\, d\tau(S)$ is a bounded linear map on $\mathcal{C}(I)$, and hence defines a unique signed Radon measure ν on I which we denote by $\nu = \int \nu_S\, d\tau(S)$. The point measures

$$\delta_S \quad \text{and} \quad \chi_R = \sum_{Q\subseteq R} \delta_Q$$

which we will use in the next proposition satisfy the above conditions, so we can use them as integrants in vector valued integrals.

Now the proof of Theorem 6.4.3 will follow immediately from the following proposition.

PROPOSITION 6.8.1. *Let G be a compact topological packing graph and suppose λ is feasible for $\mathrm{las}_{\alpha(G)}(G)$. Then there exists a unique probability measure*

$$\sigma \in \mathcal{P}(I) = \{\lambda \in \mathcal{M}(I)_{\geq 0} : \|\lambda\| = 1\}$$

such that

$$\lambda = \int \chi_R\, d\sigma(R).$$

PROOF. *Existence:* We have

$$\lambda = \int \delta_S\, d\lambda(S) = \int \sum_{R\subseteq S} (-1)^{|S\setminus R|} \chi_R\, d\lambda(S),$$

because by the inclusion-exclusion principle

$$\sum_{R\subseteq S} (-1)^{|S\setminus R|}\chi_R = \sum_{R\subseteq S} (-1)^{|S\setminus R|} \sum_{Q\subseteq R} \delta_Q = \sum_{Q\subseteq S} \delta_Q \sum_{R:Q\subseteq R\subseteq S} (-1)^{|S\setminus R|} = \delta_S.$$

The image of $f \in \mathcal{C}(I)$ under the linear map

$$\mathcal{C}(I) \to \mathbb{R},\ f \mapsto \int \sum_{R\subseteq S} (-1)^{|S\setminus R|} f(R)\, d\lambda(S)$$

has norm at most $2^{\alpha(G)}\|\lambda\|\|f\|_\infty$, so the above linear functional is bounded and hence defines a signed Radon measure σ on I. Then

$$\int \chi_R(f)\, d\sigma(R) = \int \sum_{R\subseteq S} (-1)^{|S\setminus R|} \chi_R(f)\, d\lambda(S) = \lambda(f),$$

for each $f \in \mathcal{C}(I)$, so $\lambda = \int \chi_R\, d\sigma(R)$.

Uniqueness: If $\sigma' \in \mathcal{M}(I_{2t})$ is another measure such that $\lambda = \int \chi_R\, d\sigma'(R)$, then $\int \chi_R\, d(\sigma - \sigma')(R) = 0$. Evaluating the above measure at a Borel set $L \subseteq I_{=t}$ with $t = \alpha(G)$ gives

$$0 = \int \chi_R(L)\, d(\sigma - \sigma')(R) = (\sigma - \sigma')(L),$$

so $\sigma|_{I_{=t}} = \sigma'|_{I_{=t}}$. Repeating this argument for $t = \alpha(G) - 1, \ldots, 1, 0$ shows $\sigma = \sigma'$, which shows that σ is unique.

Positivity: Let $g \in \mathcal{C}(I)_{\geq 0}$ be arbitrary and define $f \in \mathcal{C}(I)$ by

$$f(Q) = \sum_{P \subseteq Q} (-1)^{|Q \setminus P|} \sqrt{g(P)},$$

so that

$$\begin{aligned} \sum_{Q \subseteq R} f(Q) &= \sum_{Q \subseteq R} \sum_{P \subseteq Q} (-1)^{|Q \setminus P|} \sqrt{g(P)} \\ &= \sum_{P \subseteq R} (-1)^{|P|} \sqrt{g(P)} \sum_{Q : P \subseteq Q \subseteq R} (-1)^{|Q|} = \sqrt{g(R)}. \end{aligned}$$

We have

$$0 \leq \langle f \otimes f, A^*_{\alpha(G)} \lambda \rangle = \langle A_{\alpha(G)} f \otimes f, \lambda \rangle,$$

and since $\lambda = \int \chi_R \, d\sigma(R)$, the right hand side above is equal to

$$\int \sum_{Q \subseteq R} A_{\alpha(G)}(f \otimes f)(Q) \, d\sigma(R).$$

Since we are in the final step of the hierarchy, we have that $A_{\alpha(G)}(f \otimes f)(Q)$ can be written as $\sum_{J \cup J' = Q} f(J) f(J')$, so the above equals

$$\int \sum_{Q \subseteq R} \sum_{J \cup J' = Q} f(J) f(J') \, d\sigma(R) = \int \left(\sum_{Q \subseteq R} f(Q) \right)^2 d\sigma(R) = \int g(R) \, d\sigma(R),$$

which shows that σ is a positive measure.

Normalization: σ is a probability measure, because

$$1 = \lambda(\{\emptyset\}) = \int \chi_S(\{\emptyset\}) \, d\sigma(S) = \|\sigma\|. \qquad \square$$

Proposition 6.8.2. *Let G be a compact topological packing graph. Then the extreme points of the feasible region of $\mathrm{las}_{\alpha(G)}(G)$ are precisely the measures χ_R with $R \in I$.*

Proof. If $\sigma \in \mathcal{P}(I)$ and $\lambda = \int \chi_R \, d\sigma(R)$, then

$$\lambda(\{\emptyset\}) = \int \chi_R(\{\emptyset\}) \, d\sigma(R) = 1,$$

and for each $K \in \mathcal{C}(I \times I)_{\succeq 0}$ we have

$$\langle K, A^*_{\alpha(G)} \lambda \rangle = \int \chi_R(A_{\alpha(G)} K) \, d\sigma(R) = \int \sum_{J, J' \subseteq R} K(J, J') \, d\sigma(R) \geq 0,$$

so λ is feasible for $\mathrm{las}_{\alpha(G)}(G)$. So we have the surjective linear map

$$L \colon \mathcal{P}(I) \to \mathcal{F}, \ \sigma \mapsto \int \chi_R \, d\sigma(R),$$

where $\mathcal{F}$ denotes the feasible set of $\mathrm{las}_{\alpha(G)}(G)$. By Proposition 6.8.1 the map L is also injective. This means that

$$\mathrm{ex}(\mathcal{F}) = \mathrm{ex}(L(\mathcal{P}(I))) = L(\mathrm{ex}(\mathcal{P}(I)))$$

and since $\mathrm{ex}(\mathcal{P}(I)) = \{\delta_S : S \in I\}$ (see for instance Barvinok [**10**, Proposition 8.4]), the right hand side above is equal to $L(\{\delta_S : S \in I\}) = \{\chi_R : R \in I\}$. □

PROOF OF THEOREM 6.4.3 . Let λ be feasible for $\mathrm{las}_{\alpha(G)}(G)$. By Proposition 6.8.1 there exists a probability measure σ on I such that $\lambda = \int \chi_S \, d\sigma(S)$. Substituting this integral for λ in the definition of $\mathrm{las}_{\alpha(G)}(G)$ gives

$$\mathrm{las}_{\alpha(G)}(G) \leq \max\Big\{ \int \underbrace{\chi_R(I_{=1})}_{|R|} \, d\sigma(R) : \sigma \in \mathcal{P}(I) \Big\} = \alpha(G),$$

and since we already know that $\mathrm{las}_{\alpha(G)}(G) \geq \alpha(G)$, this completes the proof. □

6.9. Two and three-point bounds

6.9.1. Two-point bounds. The Lovász ϑ-number is a two-point bound originally defined for finite graphs. Bachoc, Nebe, Oliveira, and Vallentin [**8**] generalized this to the spherical code graph, and they showed that it is equivalent to the linear programming bound of Delsarte, Goethals, and Seidel [**27**]. The following generalization of the ϑ'-number for compact topological packing graphs G is natural:

$$\begin{aligned} \vartheta'(G)^* = \inf \Big\{ a : a \in \mathbb{R},\ & F \in \mathcal{C}(V \times V)_{\succeq 0}, \\ & F(x,x) \leq a - 1 \text{ for } x \in V, \\ & F(x,y) \leq -1 \text{ for } \{x,y\} \in I_{=2} \Big\}. \end{aligned}$$

LEMMA 6.9.1. *Let G be a compact topological packing graph. Then $\vartheta'(G)^*$ has a feasible solution.*

For finite graphs one can show $\vartheta'(G)^*$ admits a feasible solution by selecting a matrix F with $F(x,y) = -1$ for $\{x,y\} \in I_{=2}$ and the diagonal of F large enough so as to make it diagonally dominant and hence positive semidefinite. For infinite graphs it is not clear how to adapt this argument, so we use a different approach.

PROOF LEMMA 6.9.1 . By the topological packing graph condition there is for each $x \in V$ an open clique C_x containing x. Since V is a compact Hausdorff space, it is a normal space, so there exists an open neighborhood U_x of x such that its closure does not intersect $V \setminus C_x$. By compactness there exists an $S \subseteq V$ such that $\{U_x : x \in S\}$ is a finite open cover of V. By Urysohn's lemma there is a function $f_x \in \mathcal{C}(V)$ such that

$$f_x(y) \begin{cases} = |S| & \text{if } y \in U_x, \\ \in [-1, |S|] & \text{if } y \in C_x \setminus U_x, \\ = -1 & \text{if } y \in V \setminus C_x. \end{cases}$$

Define

$$F \in \mathcal{C}(V \times V)_{\succeq 0} \text{ by } F = \sum_{x \in S} f_x \otimes f_x, \text{ and } a = |S|^3 + 1.$$

Then,

$$F(y,y) = \sum_{x \in S} f_x(y)^2 \leq |S|^3 = a - 1 \text{ for all } y \in V.$$

Moreover, if $\{y, y'\} \in I_{=2}$, then at most one of y and y' lies in C_x for every given $x \in S$. So, $f_x(y)f_x(y') = -|S|$ if either y or y' lies in U_x and $f_x(y)f_x(y') \leq 1$ if neither y nor y' lies in U_x. Hence, $F(y, y') \leq -1$ for all $\{y, y'\} \in I_{=2}$, and it follows that (a, F) is feasible for $\vartheta'(G)^*$. □

Now we show that the first step of our hierarchy equals the ϑ'-number for compact topological packing graphs, as it is known for finite graphs.

THEOREM 6.9.2. *Let G be a compact topological packing graph. Then*

$$\mathrm{las}_1(G)^* = \vartheta'(G)^*.$$

We prove this theorem by Lemma 6.9.3 and Lemma 6.9.6. We first show the easy inequality.

LEMMA 6.9.3. $\mathrm{las}_1(G)^* \leq \vartheta'(G)^*$.

PROOF. Assume (a, F) is feasible for $\vartheta'(G)^*$ and define $K \in \mathcal{C}(I_1 \times I_1)_{\mathrm{sym}}$ by

$$\begin{aligned} &K(\emptyset, \emptyset) = a, \\ &K(\emptyset, \{x\}) = K(\{x\}, \emptyset) = -1 \text{ for } x \in V, \\ &K(\{x\}, \{y\}) = (F(x, y) + 1)/a \text{ for } x, y \in V. \end{aligned}$$

To show that K is positive definite we show that the matrix $(K(J_i, J_j))_{i,j=1}^m$ is positive semidefinite for all $m \in \mathbb{N}$ and $J_1, \ldots, J_m \in I_1$ pairwise different. If none of the J_i's is empty, then it follows directly. Otherwise we may assume that there are $x_2, \ldots, x_m \in V$ such that $J_1 = \emptyset$ and $J_i = \{x_i\}$ for $i = 2, \ldots, m$. We have

$$\Big(K(J_i, J_j) - K(J_i, J_1)K(J_1, J_1)^{-1}K(J_1, J_j)\Big)_{i,j=2}^m = a^{-1}\big(F(x_i, x_j)\big)_{i,j=2}^m,$$

so by the Schur complement $\big(K(J_i, J_j)\big)_{i,j=1}^m$ is positive semidefinite.

For $x \in V$ we have

$$A_1K(\{x\}) = K(\{x\}, \{x\}) + K(\{x\}, \emptyset) + K(\emptyset, \{x\}) = (F(x, x) + 1)/a - 2 \leq -1,$$

and for $\{x, y\} \in I_{=2}$ we have

$$\begin{aligned} A_1K(\{x, y\}) &= K(\{x\}, \{y\}) + K(\{y\}, \{x\}) \\ &= (F(x, y) + 1)/a + (F(y, x) + 1)/a \leq 0. \end{aligned}$$

So K is feasible for $\mathrm{las}_t(G)^*$ and since $K(\emptyset, \emptyset) = a$ we have $\mathrm{las}_t(G)^* \leq \vartheta'(G)^*$. □

REMARK 6.9.4. From this lemma we can see that for each $t \in \mathbb{N}$ the optimization problem $\mathrm{las}_t(G)^*$ has a feasible solution and so by strong duality the maximum in $\mathrm{las}_t(G)$ is attained: By Lemma 6.9.1, $\vartheta'(G)^*$ has a feasible solution, hence by the lemma above $\mathrm{las}_1(G)^*$ also has one. Then this can be extended trivially to a feasible solution for every $\mathrm{las}_t(G)^*$.

To prove the other inequality we will use the following generalization of the Schur complement.

LEMMA 6.9.5. *Let X be a compact Hausdorff space and let $x_1, \dots, x_n \in X$ be elements such that the singletons $\{x_i\}$ are open. Suppose $\mu \in \mathcal{M}(X \times X)_{sym}$ is such that the matrix $A = (\mu(\{(x_i, x_j)\}))_{i,j=1}^n$ is positive definite. Denote by $\mathcal{F} \subseteq \mathcal{C}(X)$ the set of functions which are zero on $\{x_1, \dots, x_n\}$ and for $g \in \mathcal{F}$ define the vector $v_g \in \mathbb{R}^n$ by $(v_g)_i = \mu(1_{\{x_i\}} \otimes g)$. Then μ is positive definite if and only if*

$$\mu(g \otimes g) - v_g^{\mathsf{T}} A^{-1} v_g \geq 0 \quad \text{for all} \quad g \in \mathcal{F}.$$

PROOF. Mercer's theorem says that a kernel $K \in \mathcal{C}(X \times X)_{\mathrm{sym}}$ is positive definite if and only if there exist sequences $(f_i)_i$ and $(\lambda_i)_i$ in $\mathcal{C}(X)$ and $\mathbb{R}_{\geq 0}$ such that $K(x, y) = \sum_{i=1}^{\infty} \lambda_i f_i \otimes f_i(x, y)$, where convergence is uniform and absolute. It follows that $\mu \in \mathcal{M}(X \times X)_{\succeq 0}$ if and only if $\mu(f \otimes f) \geq 0$ for all $f \in \mathcal{C}(X)$. Now we use the technique as described in for instance the book by Boyd and Vandenberghe [**19**, Appendix A.5.5] and note that the measure μ is positive definite if and only if the function $p \colon \mathbb{R}^n \times \mathcal{F} \to \mathbb{R}$ given by

$$\begin{aligned} p(r, g) &= \mu((r_1 1_{\{x_1\}} + \cdots + r_n 1_{\{x_n\}} + g) \otimes (r_1 1_{\{x_1\}} + \cdots + r_n 1_{\{x_n\}} + g)) \\ &= \mu(g \otimes g) + r^{\mathsf{T}} A r + 2 r^{\mathsf{T}} v_g \end{aligned}$$

is nonnegative on its domain. We have $\nabla_r p(r, g) = 2Ar + 2v_g$, so for fixed g, the minimum of p is attained for $r = -A^{-1} v_g$. Hence p is nonnegative on its domain if and only if $\mu(g \otimes g) - v_g^{\mathsf{T}} A^{-1} v_g \geq 0$ for all $g \in \mathcal{F}$. □

LEMMA 6.9.6. $\mathrm{las}_1(G)^* \geq \vartheta'(G)^*$.

PROOF. We will use the duals of $\vartheta'(G)^*$ and $\mathrm{las}_1(G)^*$. We derive the dual $\vartheta'(G)$ of $\vartheta'(G)^*$ similarly to Section 6.7.1. We have

$$\begin{aligned} \vartheta'(G) = \sup \Big\{ \eta(I_2 \setminus \{\emptyset\}) : \eta \in \mathcal{M}(I_2 \setminus \{\emptyset\})_{\geq 0}, \\ \eta(I_{=1}) = 1,\ T^* \eta \in \mathcal{M}(I_{=1} \times I_{=1})_{\succeq 0} \Big\}, \end{aligned}$$

where $T \colon \mathcal{C}(I_{=1} \times I_{=1}) \to \mathcal{C}(I_2 \setminus \{\emptyset\})$ is the operator defined by

$$TF(S) = \begin{cases} F(\{x\}, \{x\}) & \text{if } S = \{x\}, \\ \frac{1}{2}(F(\{x\}, \{y\}) + F(\{y\}, \{x\})) & \text{if } S = \{x, y\}. \end{cases}$$

Now we prove strong duality: $\vartheta'(G) = \vartheta'(G)^*$ and the optimum in $\vartheta'(G)$ is attained. Following the approach from Section 6.7.2 we first observe that every probability measure on $I_{=1}$ is feasible for $\vartheta'(G)$. To complete the proof we show that

$$\begin{aligned} K = \{(T^* \eta - \nu, \eta(I_2 \setminus \{\emptyset\})) : \nu \in \mathcal{M}(I_{=1} \times I_{=1})_{\succeq 0}, \\ \eta \in \mathcal{M}(I_2 \setminus \{\emptyset\})_{\geq 0},\ \eta(I_{=1}) = 0\} \end{aligned}$$

is closed in $\mathcal{M}(I_{=1} \times I_{=1})_{\mathrm{sym}} \times \mathbb{R}$. We decompose K as the Minkowski difference of

$$K_1 = \{(T^* \eta, \eta(I_2 \setminus \{\emptyset\})) : \eta \in \mathcal{M}(I_2 \setminus \{\emptyset\})_{\geq 0},\ \eta(I_{=1}) = 0\}$$

and

$$K_2 = \{(\nu, 0) : \nu \in \mathcal{M}(I_{=1} \times I_{=1})_{\succeq 0}\}.$$

It is immediate that $K_1 \cap K_2 = \{0\}$ and again using the approach from Section 6.7.2 we see that K_1 and K_2 are closed and that K_1 is locally compact.

Now we show the inequality $\vartheta'(G) \leq \mathrm{las}_1(G)$. Let η be an optimal solution for $\vartheta'(G)$ and define $\lambda \in \mathcal{M}(I_2)$ by $\lambda(\{\emptyset\}) = 1$ and

$$\lambda(L) = \begin{cases} \vartheta'(G)\eta(L) & \text{if } L \text{ is a Borel set in } I_{=1}, \\ \frac{1}{2}\vartheta'(G)\eta(L) & \text{if } L \text{ is a Borel set in } I_{=2}. \end{cases}$$

Then

$$\lambda(I_{=1}) = \vartheta'(G)\eta(I_{=1}) = \vartheta'(G).$$

To complete the proof we have to show $A_1^*\lambda \in \mathcal{M}(I_1 \times I_1)_{\succeq 0}$. We apply our generalized Schur complement: Let $g \in \mathcal{C}(I_1)$ be a function with $g(\emptyset) = 0$. We have

$$A_1^*\lambda(g \otimes g) = \vartheta'(G)T^*\eta(g \otimes g).$$

The symmetric bilinear form $(h, g) \mapsto T^*\eta(h \otimes g)$ is positive semidefinite because $T^*\eta \in \mathcal{M}(I_{=1} \times I_{=1})_{\succeq 0}$, so we can apply the Cauchy-Schwarz inequality and optimality of η to obtain

$$\vartheta'(G)T^*\eta(g \otimes g) \geq \frac{\vartheta'(G)}{T^*\eta(1_{I_{=1}} \otimes 1_{I_{=1}})}(T^*\eta(1_{I_{=1}} \otimes g))^2 = (T^*\eta(1_{I_{=1}} \otimes g))^2.$$

In the remainder of this proof we show

$$T^*\eta(1_{I_{=1}} \otimes g) = \vartheta'(G)\eta(g).$$

Since

$$\vartheta'(G)\eta(g) = \lambda(g) = A_1^*\lambda(1_\emptyset \otimes g)$$

the proof is then complete by using the generalized Schur complement, Lemma 6.9.5.

Inspired by Schrijver [**88**, Theorem 67.10] we use Lagrange multipliers. First observe that

$$T(1_{I_{=1}} \otimes 1_{I_{=1}}) = 1_{I_2 \setminus \{\emptyset\}} \quad \text{and} \quad T^*\eta(1_{I_{=1}} \otimes 1_{I_{=1}}) = \eta(I_2 \setminus \{\emptyset\}).$$

For $u \in \mathbb{R}^2$ define $g_u \in \mathcal{C}(I_{=1})$ by $g_u = u_1 g + u_2(1_{I_{=1}} - g)$. For each $u \in \mathbb{R}^2$ with $\eta(g_u^2) = 1$, the measure $\tilde{\eta}$ defined by $d\tilde{\eta}(S) = T(g_u \otimes g_u)(S)d\eta(S)$ is feasible for $\vartheta'(G)$. So, if we consider the problem of maximizing $T^*\eta(g_u \otimes g_u)$ over all $u \in \mathbb{R}^2$ for which $\eta(g_u^2) = 1$, then optimality of η implies that an optimal solution is attained for $u = 1$.

It follows that there exists a Lagrange multiplier $c \in \mathbb{R}$ such that

$$\frac{\partial}{\partial u_i}\bigg|_{u=(1,1)} T^*\eta(g_u \otimes g_u) = c\frac{\partial}{\partial u_i}\bigg|_{u=(1,1)} \eta(g_u^2) \quad \text{for} \quad i = 1, 2.$$

Since

$$T^*\eta(g_u \otimes g_u) = u^\mathsf{T} \begin{pmatrix} T^*\eta(g \otimes g) & T^*\eta(g \otimes (1_{I_{=1}} - g)) \\ T^*\eta(g \otimes (1_{I_{=1}} - g)) & T^*\eta((1_{I_{=1}} - g) \otimes (1_{I_{=1}} - g)) \end{pmatrix} u$$

and

$$\eta(g_u^2) = u^\mathsf{T} \begin{pmatrix} \eta(g^2) & \eta(g(1_{I_{=1}} - g)) \\ \eta(g(1_{I_{=1}} - g)) & \eta((1_{I_{=1}} - g)^2) \end{pmatrix} u$$

we have

$$T^*\eta(g \otimes 1_{I_{=1}}) = c\eta(g) \quad \text{and} \quad T^*\eta((1_{I_{=1}} - g) \otimes 1_{I_{=1}}) = c\eta(1_{I_{=1}} - g).$$

By summing the last two equations we see that $c = \vartheta'(G)$, hence we have the desired equality $T^*\eta(g \otimes 1_{I_{=1}}) = \vartheta'(G)\eta(g)$. □

6.9.2. Three-point bounds. In this section we modify the $2t$-point bound $\mathrm{las}_t(G)$ to obtain a $2t+1$-point bound for sufficiently symmetric graphs G.

Let $G = (V, E)$ be a compact topological packing graph. We are interested in two groups related to G. The group of graph automorphisms of G and the group of homeomorphisms of the topological space V. When we endow the latter group with the compact-open topology, it is a topological group with a continuous action on V; see Arens [**4**]. In the special case when G is a distance graph, as defined in Section 6.3, the former group is contained in the latter. We say that G is *homogeneous* if there exists a compact subgroup of the group of homeomorphisms which consists only of graph automorphisms and is such that the action of Γ on V is transitive.

Fix a point $e \in V$. By G^e we denote the induced subgraph of G with vertex set

$$V^e = \{x \in V : x \neq e \text{ and } \{e, x\} \notin E\}.$$

It follows that G^e is also a compact topological packing graph. We have $\alpha(G) \geq 1 + \alpha(G^e)$, and if G is homogeneous, then $\alpha(G) = 1 + \alpha(G^e)$: If S is an independent set of G, then there exists a graph automorphism γ with $e \in \gamma S$, and $(\gamma S) \setminus \{e\} \subseteq V^e$ is an independent set for $\alpha(G^e)$. So, for computing an upper bound on the independence number of G we can also compute $1 + \mathrm{las}_t(G^e)$. This yields a bound which is at least as good as $\mathrm{las}_t(G)$:

LEMMA 6.9.7. *Suppose G is a compact topological packing graph. Then*

$$1 + \mathrm{las}_t(G^e) \leq \mathrm{las}_t(G).$$

PROOF. We denote the sets of independent sets of G^e by I_t^e and $I_{=t}^e$. Suppose λ^e is feasible for $\mathrm{las}_t(G^e)$. Let $\lambda = \delta_e + \lambda^e$. We have $\lambda \geq 0$ and $\lambda(\{\emptyset\}) = 1$. Moreover, since $A_t^*\lambda = \delta_e \otimes \delta_e + A_t^*\lambda^e$ and $A_t^*\lambda^e \in \mathcal{M}(I_t^e \times I_t^e)_{\succeq 0} \subseteq \mathcal{M}(I_t \times I_t)_{\succeq 0}$, we have $A_t^*\lambda \in \mathcal{M}(I_t \times I_t)_{\succeq 0}$. So λ is feasible for $\mathrm{las}_t(G)$. We have $1 + \lambda^e(I_{=1}^e) = \lambda(I_{=1})$, which completes the proof. □

CHAPTER 7

Moment methods in energy minimization: New bounds for Riesz minimal energy problems

This chapter is based on the publication *"D. de Laat, Moment methods in energy minimization: New bounds for Riesz minimal energy problems, In preparation".*

Abstract. We use moment techniques to construct a converging sequence of optimization problems to find the ground state energy of interacting particle systems. We approximate the problems in this sequence by block diagonalized semidefinite programs. For this we develop harmonic analysis techniques 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 the Thomson problem and Riesz s-energy problems with $s = 2, 4$. Here the numerical results suggest these bounds are sharp for the five particle case. This is the first time a 4-point bound has been computed for a continuous problem.

7.1. 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 [**34**], but for $N > 3$ we seem to require more involved techniques. In 1992, Yudin [**100**] introduced a beautiful method, based on earlier work for spherical codes by Delsarte, Goethals, and Seidel [**27**], which in addition to the $N \leq 3$ cases can be used to prove optimality for 4, 6, and 12 particles (see [**1**] for the 12 particle case). 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 optimization problem whose optimal value lower bounds the minimal energy. This means the feasible solutions of the dual of this relaxation, which is a maximization problem in the form of an infinite dimensional linear program, provide energy lower bounds. For $N = 2, 3, 4, 6, 12$ the bound is sharp, and the optimal dual solutions become optimality certificates. In 2006, Cohn and Kumar [**22**] used this

method in their proof of universal optimality of the above configurations (as well as many other configurations in higher dimensional spheres), 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 we consider conditions on pairs of particles, and hence this is called a 2-*point bound*. In 2012, Cohn and Woo [**23**] derived 3-point bounds for energy minimization based on earlier work by Schrijver [**41**] for binary codes and Bachoc and Vallentin [**9**] for spherical codes. They used this to prove universal optimality of the vertices of the rhombic dodecahedron in $\mathbb{RP}^2$. In [**76**] 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 the Thomson problem.

In Section 7.2 we construct a hierarchy $E_1, E_2, \dots$ of increasingly strong relaxations of the energy minimization problem. 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 [**61**], which generalize techniques from the Lasserre hierarchy [**64**] 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 7.4 we prove convergence to the optimal energy in at most N steps. In addition to the moment conditions we derive in Section 7.3 a set of linear constraints, where on the one hand we have enough constraints to ensure convergence of the hierarchy and on the other hand have a small enough set of constraints to still allow for a satisfying duality theory which is necessary for doing conrete computations.

The problems E_t are infinite dimensional optimization problems where the optimization variables are measures. This naturally means that the optimization variables in the dual problems E_t^* are continuous functions, which in our case are continuous kernels. In this chapter 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 by doing this we demonstrate the computational applicability of the theoretical moment techniques developed in [**61**]. Here we are interested in the second step because after symmetry reduction (see below) the 2-point bound E_1 is essentially the same as Yudin's bound.

In Section 7.5 we define a class of infinite dimensional optimization problems which occur naturally when forming moment relaxations for problems with infinitely many binary variables. The relaxations E_t fit into this framework, as well as the relaxations for packing problem in discrete geometry as derived in [**61**]. To find good energy lower bounds we need to find good feasible solutions of the dual optimization problems E_t^*. For this we work out a duality and symmetry reduction theory for this more general class of problems. The symmetry reduced dual programs, denoted by $\bar{E}_t^*$, are conic programs over the cone $\mathcal{C}(X \times X)_{\succeq 0}^{\Gamma}$ of continuous, positive definite, Γ-invariant kernels on a certain compact metric space X (see below) equipped with a continuous action of a compact group Γ. These are continuous functions $K\colon X \times X \to \mathbb{R}$ for which the matrix $(K(x_i, x_j))_{i,j=1}^n$ is positive semidefinite for all $n \in$

$\mathbb{N}$ and $x_1, \ldots, x_n \in X$, and for which $K(\gamma x, \gamma y) = K(x, y)$ for all $\gamma \in \Gamma$ and $x, y \in X$. We obtain an approximation to this optimization problem by replacing the cone $\mathcal{C}(X \times X)^{\Gamma}_{\succeq 0}$ by an inner approximating cone. Given a sequence of inner approximating cones whose union is uniformly dense, we give a sufficient condition, which is satisfied in the case of energy minimization, under which the optimal values of the corresponding sequence of approximating optimization problems converge to the optimal value of the original problem.

To find good dual solutions 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 dual programs E_t^* are quite different. In the 2 and 3-point bounds for the Thomson problem the dual variables are continuous kernels $K \colon S^2 \times S^2 \to \mathbb{R}$, which for 2-point bounds can be assumed to be invariant under the orthogonal group $O(3)$, and for 3-point bounds under the stabilizer subgroup of a point $e \in S^2$. In the dual programs $\bar{E}_t^*$, the variables are kernels $K \in \mathcal{C}(X \times X)^{\Gamma}_{\succeq 0}$, where X is the set I_t of independent sets of size at most t in a certain Γ-invariant graph G whose vertex set is the container of the problem. For the Thomson problem the vertex set of G is the sphere S^2 and $\Gamma = O(3)$. In Chapter 3 we give a nonconstructive proof that for each X there exists a sequence of finite dimensional, block diagonalized, inner approximating kernels whose union is uniformly dense in this cone. By using harmonic analysis and symmetric tensor powers, we show in Section 7.6 how to construct such a sequence explicitly for the case where $X = I_t$ given that we have some knowledge about the harmonic analysis of the group action of Γ on the vertex set V of the graph. In the case where $V = S^2$ we explicitly derive this information and hence give an explicit construction of the inner approximating cones. Together with the results from the previous paragraph, this gives a concrete sequence of optimization problems, each having finite dimensional variable space, whose objective values converge to E_t.

Using a result from invariant theory, we show (see Section 7.7) how we can write these problems, which now have finite dimensional variable space but still infinitely many constraints, as semidefinite programs with finitely many polynomial constraints. A *semidefinite program* is an optimization problem where we optimize a linear functional over the intersection of an affine space with a cone of positive semidefinite matrices. Semidefinite programming forms a powerful generalization of linear programming and, as for linear programming, we have efficient algorithms for solving them. A *polynomial constraint* here is the requirement that a polynomial, whose coefficients depend on the entries of the matrix variable(s), is nonnegative on a basic closed semialgebraic set. We can model these polynomial constraints as semidefinite constraints using sum-of-squares characterizations from real algebraic geometry. Together with the aforementioned results this yields a sequence of increasingly large semidefinite programs whose optimal values converge to the ground state energy.

To block diagonalize the inner approximating cones of the cone of invariant positive definite kernels we use the symmetry of the sphere and the pair potential. Energy minimization problems, however, admit even more symmetry: the particles

are interchangeable. This means the polynomial constraints in the problems discussed above admit some additional symmetries. In Section 7.8 we give a simple extension of Putinar's theorem which allows us to exploit such symmetries. This result in even futher block diagonalized semidefinite programs. These techniques also apply to the 3-point bounds mentioned above, where this symmetry has not been used before.

Although the N-th relaxation E_N is guaranteed to give the optimal energy E, it is possible that E_t is already 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 [**23**] to be sharp for $N = 8$, and since E_2 is a 4-point bound one would expect it to be at least as good and hence also 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 – we compute E_2 numerically for the Thomson problem with $N = 5$. The first 7 digits of the solution obtained by the semidefinite programming solver agree with the energy of the optimal configuration, which is a good indication that the bound is sharp. This is the first time a 4-point bound has been computed for a continuous problem.

The 5 particle case on S^2 is particularly interesting because it provides one of the simplest mathematical models of a phase transition. 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 summing $\|x_i - x_j\|_2^{-s}$ over all $1 \leq i < j \leq N$. The configuration consisting of the vertices of the triangular bipyramid is believed to be optimal for $0 < s \leq 15.04\ldots$, and the vertices of the square pyramid (where the latitude of the base depends on the specific value of s) is believed to be optimal for the remaining positive values of s. For no value of s an optimality certificate is known, although optimality has been proved for $s = 1$ and $s = 2$ by essentially enumerating all possibilities [**90**]. In addition to the $s = 1$ case we compute E_2 for $s = 2$ and $s = 4$ where the numerical results suggest the bound is sharp. This leads to the following conjecture:

CONJECTURE 7.1.1. *Given* 5 *particles on* S^2*, the bound* E_2 *is sharp for the Riesz s-energy potential for* $s = 1, 2, 4$.

We have not been able to reliably compute the bound for larger values of s, but it would be very interesting if the bound stays sharp throughout the phase transition. Inspired by Conjecture 14 in [**23**] about the optimality of the 3-point bound for $N = 8$ we pose the following question:

QUESTION 7.1.2. Is the bound E_2 universally sharp for 5 particles on S^2?

Here by *universally sharp* we mean the bound is sharp for all completely monotonic pair potentials in the squared chordal distance.

7.2. A hierarchy of relaxations for energy minimization

In this section we derive a sequence of relaxations for the energy minimization problem. 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 $h\colon (0, \operatorname{diam}(V)] \to \mathbb{R}$, where $h(s) \to \infty$ as $s \downarrow 0$. We denote the number of particles in the system by N. The *optimal energy* or *ground state energy* is given by the minimum of

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

over all sets $\{x_1, \dots, x_N\}$ of N distinct points from V. For the Thomson problem we have $V = S^2$ with metric $d(x, y) = \|x - y\|_2$ and pair potential the Coulomb energy $h(s) = 1/s$.

To compactify this problem, which will be important when we discuss duality, we introduce a graph which allows us to discard some configurations which are clearly nonoptimal. We 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. Let G be the graph with vertex set V, where distinct vertices x and y are adjacent whenever $h(d(x, y)) \ge B$. Let I_t be the set of independent sets which have cardinality at most t, where an *independent set* is a subset of the vertices for which no two vertices are adjacent. We endow $I_t \setminus \{\emptyset\}$ with a topology as a subset of the quotient space V^t/q, where q maps a tuple $(x_1, \dots, x_t)$ to the set $\{x_1, \dots, x_t\}$. 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. The set $I_{=t}$ obtains a topology as a subset of I_t. The graph G is an example of a compact topological packing graph as defined in Chapter 6. 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 B it follows that $I_{=N}$ is nonempty.

The ground state energy can be computed as

$$E = \min\Big\{\chi_S(f) : S \in I_{=N}\Big\}.$$

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

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

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 Lemma 6.6.2. Since we minimize the continuous function $S \mapsto \chi_S(f)$ over a compact set the minimum is attained.

To obtain energy lower bounds we construct a hierarchy $E_1, E_2, \dots$ of relaxations of the above problem. These are minimization problems where for every feasible solution of E we can construct a feasible solution of E_t having the same objective value. The optimization problems E_t have the important feature that we can explicitly give the dual optimization problems, and we can prove the duality gap to be zero. In Section 7.4 we show the N-th step E_N in this hierarchy gives the optimal energy, and that the extreme points of the feasible set of E_N are precisely the measures χ_S with $S \in I_{=N}$. In other words, we show E_N is a *sharp* relaxation on E.

Denote by $\mathcal{M}(I_{2t})$ the space of signed Radon measures on I_{2t}. Given $S \in I_{=N}$, define $\lambda_S \in \mathcal{M}(I_{2t})$ by restricting χ_S to I_{2t} if $2t \leq N$, or by extending χ_S to I_{2t} by zeros if $2t \geq N$. In the t-th step E_t of the hierarchy we will optimize over measures λ on I_{2t} which we require to satisfy some of the 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 tools from Chapter 6 to define what we mean by this. Define the operator

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

Here $\mathcal{C}(I_t \times I_t)_{\text{sym}}$ is the space of symmetric, continuous functions $I_t \times I_t \to \mathbb{R}$, which we call symmetric *kernels*. A symmetric kernel K is *positive definite* if the matrix $(K(J_i, J_j))_{i,j=1}^n$ is positive semidefinite for all $n \in \mathbb{N}$ and $J_1, \dots, J_n \in I_t$. 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 $\mathcal{C}(I_t \times I_t)_{\text{sym}}$ and $\mathcal{C}(I_{2t})$ can be identified with the space $\mathcal{M}(I_t \times I_t)_{\text{sym}}$ of symmetric Radon measures and the space $\mathcal{M}(I_{2t})$ of Radon measures. Here a measure $\mu \in \mathcal{M}(I_t \times I_t)$ is said to be symmetric if

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

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

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

We have the dual operator

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

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

Definition 7.2.1. *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 Remark 6.7.4 for an explanation why we use the term positive type here. The measure λ_S is of positive type: For each $K \in \mathcal{C}(I_t \times I_t)_{\succeq 0}$, we have

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

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

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

In Section 7.5.1 we prove strong duality, which implies the minimum here is attained. The measure λ_S is feasible for E_t by construction, so $E_t \leq E$ for all t. In a similar way we have $E_t \leq E_{t+1}$ for all t.

7.3. Connection to the Lasserre hierarchy

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

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

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

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

These are, of course, special cases of polynomial optimization problems, because we can enforce the constraints $x \in \{0,1\}^n$ by adding the polynomial constraints $x_i(1-x_i) \geq 0$ and $-x_i(1-x_i) \geq 0$ 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

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

where $\kappa(x) = \sum_{i=1}^n x_i - N$, and where f is the pair potential as defined in the previous section. To obtain a sequence of relaxations we apply a variation of the Lasserre hierarchy to this problem. From this we then derive a possibly weaker sequence of relaxations that admits a useful generalization to energy minimization problems with an infinite container. This then leads to the hierarchy $\{E_t\}$ as defined in Section 7.2, and this is how we originally came up with these relaxations.

Denote by $[n]_t$ the set of all subsets of $[n]$ of cardinality at most t and by $[n]_{=t}$ the subsets of cardinality t. In a binary polynomial optimization problem we may assume all polynomials to be square free, and for such a polynomial we write

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

Given an integer $t \in \mathbb{N}$ and a vector $y \in \mathbb{R}^{[n]_{2t}}$, the t-th *moment matrix* $M_t(y) \in \mathbb{R}^{[n]_t \times [n]_t}$ is defined by $M_t(y)_{J,J'} = y_{J \cup J'}$. The t-th *localizing matrix* with respect to a polynomial $g \in \mathbb{R}[x_1, \ldots, x_n]$ is 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

$$\sum_{R \in [n]_{\deg(g)}} y_{J \cup J' \cup R}\, g_R$$

whenever $|J \cup J'| \leq 2t - \deg(g)$. 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 (which 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:

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

These relaxations where introduced by Lasserre in [**64**]. The only modifications we make here is that we restrict y to be nonnegative, and originally the localizing matrices $M_t^{g_i}(y)$ are defined to be full matrices indexed by $[n]_{t - \lceil \deg(g_i)/2 \rceil}$, but here we take them to be partial matrices indexed by $[n]_t$. Typically, this does not make the semidefinite programs much more difficult to solve, but in some cases, such as the case of energy minimization as discussed here, it can lead to much stronger bounds.

In the binary polynomial optimization problem formulation for the energy minimization problem 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 $-M_t^\kappa(y) = M_t^{-\kappa}(y) \succeq 0$, which 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

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

So, for energy minimization we get the relaxations

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

The linear constraints in these problems become problematic when we want to generalize V from the finite set $[n]$ to an infinite set. The reason being that in the infinite dimensional generalization we want to use measures λ instead of vectors y

(because this allows for a satisfying duality theory; see Section 7.5.1), and these then become uncountably many "thin" constraints on λ. By thin we mean the constraints are of the form $\lambda(E) = b$, where E is a measurable set with empty interior. This means these constraints have no grip on measures which are zero on sets with empty interior.

In the following proposition we show these constraints imply $2t + 1$ natural constraints on the vectors y. In particular, this proposition 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 new constraints, then we obtain the problem E_t for the finite container $V = [n]$.

PROPOSITION 7.3.1. *Let $t \in \mathbb{N}$ and $y \in \mathbb{R}^{[n]_{2t}}$. If*

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

then

$$\sum_{S\in[n]_{=i}} y_S = \binom{N}{i} \quad \textit{for all} \quad 0 \le i \le 2t.$$

PROOF. For $i = 0$ we have

$$\sum_{S\in[n]_{=i}} y_S = y_\emptyset = 1 = \binom{N}{i}.$$

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

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

So, the proof follows by induction. □

In the following proposition we show that the constraints

$$\sum_{S\in[n]_{=i}} y_S = \binom{N}{i} \quad \text{for} \quad 0 \le i \le 2t$$

imply a large number of the linear constraints in the formulation of L_t. In particular, together with the previous proposition it shows that for $t = N$, the feasible set does not change when we replace all $\binom{n}{0} + \ldots + \binom{n}{2t-1}$ linear constraints in L_t with these $2t + 1$ constraints.

PROPOSITION 7.3.2. *Let $t \in \mathbb{N}$ and $y \in \mathbb{R}^{[n]_{2t}}$. If $M_t(y) \succeq 0$ and*

$$\sum_{S \in [n]_{=j}} y_S = \binom{N}{j} \quad \text{for all} \quad 0 \leq j \leq t,$$

then

$$y_\emptyset = 1 \quad \text{and} \quad N y_S = \sum_{i=1}^{n} y_{S \cup \{i\}} \quad \text{for all} \quad S \in [n]_t.$$

PROOF. We have

$$y_\emptyset = \sum_{S \in [n]_{=0}} y_S = \binom{N}{0} = 1.$$

The identity $y_\emptyset = 1$ together with $M_t(y) \succeq 0$ implies $y_S \geq 0$ and $1 y_S - y_S^2 \geq 0$ for all $S \in [n]_t$. That is, $y_S \in [0,1]$ for all $S \in [n]_t$. For all $J, J' \in [n]_t$ we have $y_J y_{J'} - y_{J \cup J'} \geq 0$, and hence $y_{J \cup J'} \leq y_J y_{J'} \leq y_J$.

Then, $M_t(y) \succeq 0$ implies $y_R \geq y_S$ whenever $R \subseteq S \in [n]_t$. Hence,

$$v_S := (N - |S|) y_S - \sum_{x \in [n] \setminus S} y_{S \cup \{x\}} \geq 0 \quad \text{for all} \quad S \in [n]_t.$$

Then,

$$\begin{aligned}
\frac{1}{i+1} \sum_{S \in [n]_{=i}} v_S &= \frac{N-i}{i+1} \sum_{S \in [n]_{=i}} y_S - \frac{1}{i+1} \sum_{S \in [n]_{=i}} \sum_{x \in [n] \setminus S} y_{S \cup \{x\}} \\
&= \frac{N-i}{i+1} \sum_{S \in [n]_{=i}} y_S - \frac{1}{i+1}(1+i) \sum_{S \in [n]_{=i+1}} y_S \\
&= \frac{N-i}{i+1} \binom{N}{i} - \binom{N}{i+1} = \binom{N}{i+1} - \binom{N}{i+1} = 0,
\end{aligned}$$

so $v_S = 0$ for all $S \in [n]_t$, hence

$$N y_S = \sum_{i=1}^{n} y_{S \cup \{i\}} \quad \text{for all} \quad S \in [n]_t. \qquad \square$$

7.4. Convergence to the ground state energy

In this section we show 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 Proposition 6.8.1.

PROPOSITION 7.4.1. *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 7.4.2. *The N-th step E_N gives the optimal energy E.*

PROOF. Let $\lambda \in \mathcal{M}(I_{2N})$ be feasible for E_N. Then λ is supported on I_N and is of positive type. By Proposition 7.4.1 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

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

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

Using the ideas of the above proof together with the proof of Proposition 6.8.2 it follows that the extreme points of the feasible set of E_N are precisely the measures χ_S with $S \in I_{=N}$.

7.5. Optimization with infinitely many binary variables

We discuss the duality theory and symmetry reduction for a more general type of optimization problems which arise when we form 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*; see Section 6.3. This means the vertex set V is a compact Hausdorff space where every clique is contained in an open clique, where an *open clique* is an open subset of V where every two vertices are adjacent.

DEFINITION 7.5.1. *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 $H = H^{\inf}_{G,t}(h; g_1, \ldots, g_m; b_1, \ldots, b_m)$ by*

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

For energy minimization we have

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

where G is the graph defined in Section 7.2, and f is the pair potential. For packing problems, the t-th step of the hierarchy from Chapter 6 is given by $H^{\max}_{G,t}(1_{I_{=1}}; 1_{\{\emptyset\}}, 1)$, where G is the packing graph defining the packing problem.

7.5.1. 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 the reader to Section 2.3 or [**10**] for an introduction to conic programming. If we endow both $\mathcal{M}(I_t \times I_t)_{\text{sym}}$ and $\mathcal{M}(I_{2t})$ with the weak* topologies, then the topological dual spaces can be identified with $\mathcal{C}(I_t \times I_t)_{\text{sym}}$ and $\mathcal{C}(I_{2t})$. The tuples $(\mathcal{C}(I_{2t}), \mathcal{M}(I_{2t})$ and $(\mathcal{C}(I_t \times I_t)_{\text{sym}}, \mathcal{M}(I_t \times I_t)_{\text{sym}})$ are dual pairs, and the dual pairings $\langle f, \lambda \rangle = \lambda(f) = \int f \, d\lambda$ and $\langle K, \mu \rangle = \mu(K) = \int K \, d\mu$ 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 Section 6.7.2, gives a sufficient condition for strong duality.

THEOREM 7.5.2. *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 strong duality holds we use a closed cone condition; see Section 2.3. 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 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 [**56**] and Dieudonné [**28**], 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 6.7.5 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, which follows since it is a dual cone. □

In Lemma 6.7.3 we show 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 7.5.2 applies whenever 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 7.2), 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^*).

7.5.2. 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 $H^{\inf}_{G,t}(f; g_1, \ldots, g_m; b_1, \ldots, b_m)$ is Γ-*invariant* if

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

Using this definition the relaxations E_t for energy minimization 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 obtain the optimal symmetry reduction 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).

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

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

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 $K \in \mathcal{C}(I_t \times I_t)_{\succeq 0}$, 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 $\mu \in \mathcal{M}(I_t \times I_t)_{\text{sym}}$ 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 $\mathcal{M}(I_t \times I_t)^\Gamma_{\text{sym}}$ can be identified with the duals of $\mathcal{C}(I_{2t})^\Gamma$ and $\mathcal{C}(I_t \times I_t)^\Gamma_{\text{sym}}$, and as in the nonsymmetrize situation these form dual pairs.

Let $\mathcal{C}(I_{2t})^\Gamma_{\succeq 0}$ be the cone defined by intersection $\mathcal{C}(I_{2t})_{\succeq 0}$ with $\mathcal{C}(I_{2t})^\Gamma$, and define the cones $\mathcal{C}(I_t \times I_t)^\Gamma_{\succeq 0}$, $\mathcal{M}(I_{2t})^\Gamma_{\succeq 0}$, and $\mathcal{M}(I_t \times I_t)^\Gamma_{\succeq 0}$ in a similar manner.

The operator A_t maps Γ-invariant kernels to Γ-invariant functions, so we can view A_t as an operator $\mathcal{C}(I_t \times I_t)^\Gamma \to \mathcal{C}(I_{2t})^\Gamma$ which means we can view $(A^\Gamma_t)^*$ as an operator $\mathcal{M}(I_{2t})^\Gamma \to \mathcal{M}(I_t \times I_t)^\Gamma_{\text{sym}}$.

We can now define the symmetrization of a Γ-invariant primal problem H by

$$\bar{H} = \min\Big\{\lambda(f) : \lambda \in \mathcal{M}(I_{2t})^\Gamma_{\succeq 0},\ (A^*_t)^\Gamma \lambda \in \mathcal{M}(I_t \times I_t)^\Gamma_{\succeq 0},\ \lambda(g_i) = b_i \text{ for } i \in [m]\Big\},$$

and the symmetrization of the dual program H^* by

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

If follows that the optimal values of H and $\bar{H}$ coincide, and the optimal values of H^* and $\bar{H}^*$ coincide. This implies $\bar{H} = \bar{H}^*$ (which alternatively could be shown by proving strong duality as is done in the previous section).

In the following section we show how to construct a nested sequence $\{C_d\}$ of inner approximating cones of $\mathcal{C}(I_t \times I_t)^\Gamma_{\succeq 0}$ such that the union $\cup_{d=0}^\infty C_d$ is uniformly dense in $\mathcal{C}(I_t \times I_t)^\Gamma_{\succeq 0}$. The cones C_d are constructed so that optimization over the cone C_d is easier than optimization over the cones $C_{d'}$, with $d' > d$, and the cone $\mathcal{C}(I_t \times I_t)^\Gamma_{\succeq 0}$. Moreover, we will use the Γ-invariance to block diagonalize the

kernels in C_d so that optimization over C_d becomes feasible for small d for concrete, sufficiently symmetric, problems. Let H_d^* be the optimization problem $\bar{H}^*$ with the cone of invariant, positive definite kernels replaced by its inner approximation C_d. Here we give a sufficient condition for the programs H_d^* to approximate the program $\bar{H}^*$. To apply the following proposition for energy minimization we take the symmetrized version of E_t^* and select $c = 0$ and $y = -e$, where e is the all 1 vector.

PROPOSITION 7.5.3. *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 $\bar{H}_d^* \to \bar{H}^*$ 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 $\bar{H}^*$ 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. Let $\delta > 0$ be such that

$$\left| \frac{\delta}{1 + \delta c} \sum_{i=1}^m (y_i - c a_i) b_i \right| \leq \varepsilon.$$

We have $f - \sum_{i=1}^m a_i g_i - A_t K \geq 0$, so

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

which implies

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

Since $\cup_{d=0}^\infty C_d$ is uniformly dense in $\mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma$, and since A_t is a bounded operator, there exists an integer d and a kernel $L \in C_d$ such that

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

This means that

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

so $((a + \delta y)/(1 + \delta c), L)$ is feasible for $\bar{H}_d^*$, and by the choice of δ the objective value of this feasible solution lies withing distance ε of the objective of (a, K). □

7.6. Inner approximating cones via harmonic analysis

In this section we show how to construct a sequence of inner approximating cones whose union is uniformly dense in $\mathcal{C}(I_t \times I_t)_{\succeq 0}^\Gamma$. As shown in the previous section we can use this to construct a sequence of approximatings to $\bar{H}^*$. In certain cases, including the case of energy minimization, the optimal values of these

approximations converge to $\bar{H}^*$. The inner approximating cones are finite dimensional, which means each of the approximating optimization problems has finite dimensional variable space. Moreover, we use the symmetry to give a simultaneous block diagonalization of the kernels in these inner approximating cones, which is crucial for performing computations.

A good way to approximate kernels is to express them in terms of their Fourier coefficients and set all but finitely many of these coefficients to zero. To perform the Fourier transform we need the zonal matrices, and to construct the zonal matrices we need a symmetry adapted system. In Section 7.6.1 we explain these concepts for the general cone $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$ of Hermitian, Γ-invariant, positive definite kernels on a compact metrizable space X with a continuous action of the compact group Γ, and we show how this yields a sequence of tractable inner approximating cones. The relevant proofs are in Chapter 3, where we also give the extreme rays of the above cone, prove that the union of the inner approximating cones is uniformly dense, and give a nonconstructive proof that a symmetry adapted system, and hence the desired sequence of inner approximating cones, always exists.

The main goal of this section is to construct the inner approximating cones explicitly. In Section 7.6.2 we do this by embedding the space I_t into a simpler space X_t and constructing inner approximations of $\mathcal{C}_{\mathbb{C}}(X_t \times X_t)^{\Gamma}_{\succeq 0}$. We then show this yields such a sequence of $\mathcal{C}_{\mathbb{C}}(I_t \times I_t)^{\Gamma}_{\succeq 0}$ by restricting the kernels to $I_t \times I_t$. We show how symmetric tensor powers can be used to construct a symmetry adapted system for the space X_t. For this we need a symmetry adapted system for the vertex set V and we need to know how tensor products and symmetric tensor powers of the irreducible representations spanned by this system decompose into irreducibles. In Section 7.6.3 we show how to do this for the case where $t = 2$, $V = S^2$ and $\Gamma = O(3)$. Together this yields an explicit sequence of inner approximating cones of $\mathcal{C}(I_2 \times I_2)^{\Gamma}_{\succeq 0}$ for the case where $V = S^2$ and $\Gamma = O(3)$. We will use this to perform concrete computations.

7.6.1. Symmetry adapted systems and zonal matrices. Let X be a compact metrizable space and Γ a compact group with a continuous action on X. We start by defining a complete orthonormal system of X. Let μ be a Radon probability measure on X which is strictly positive and Γ-invariant; that is, $\mu(U) > 0$ and $\mu(\gamma U) = \mu(U)$ for every open set $U \subseteq X$ and every $\gamma \in \Gamma$. Such a measure always exists as is shown in Lemma 3.3.1. A *continuous orthonormal system* of X is a set of continuous, complex-valued functions on X which are orthonormal with respect to the $L^2_{\mathbb{C}}(X, \mu)$ 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 $\mathcal{C}_{\mathbb{C}}(X)$ 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 $T \colon \mathcal{H}_1 \to \mathcal{H}_2$ which 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 such a representation π by d_π. A particularly important example of a unitary representation is given by

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

A complete continuous 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

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

where $\mathcal{H}_{\pi,i} = \operatorname{span}\{e_{\pi,i,1}, \ldots, e_{\pi,i,d_\pi}\}$ is 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 3.3.5 we show such a system always exists. The number m_π is given by the dimension of the space $\operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)$ of Γ-equivariant, continuous functions from X to the Hilbert space $\mathcal{H}_\pi$ of the representation π, and hence does not depend on the choice of symmetry adapted system.

The spaces $H_{\pi,i}$ defined above are finite dimensional, irreducible subrepresentations of L spanned by continuous functions and which are pairwise orthogonal and whose (algebraic) sum is uniformly dense in $\mathcal{C}(X; \mathbb{C})$. When we are given a set of spaces satisfying these properties, then we can construct a complete orthonormal symmetry adapted system by simply selecting appropriate bases of each of these spaces.

Let $\mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$ be the cone of complex-valued, Γ-invariant, positive definite kernels, where a complex-valued kernel K is said to be *positive definite* if $\sum_{i,j=1}^n c_i \overline{c_j} K(x_i, x_j) \geq 0$ for all $n \in \mathbb{N}$, $x \in X^n$, and $c \in \mathbb{C}^n$. An *extreme direction* of a cone is an element x in the cone for which $\mathbb{R}_{\geq 0} x$ is an *extreme ray*, which means that for all $x_1, x_2 \in K$ with $x = x_1 + x_2$ we have $x_1, x_2 \in \mathbb{R}_{\geq 0} x$. In Theorem 3.2.3 we show a kernel $K \in \mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$ is an extreme direction if and only if there exists $\pi \in \hat{\Gamma}$ and $\varphi \in \operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)$ such that

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

If $m_\pi < \infty$, then the conic hull of all extreme directions corresponding to π is of the form

$$\Big\{ \sum_{i,j=1}^{m_\pi} A_{i,j} \langle \varphi_i(\cdot), \varphi_j(\cdot) \rangle : A \in \mathbb{C}^{m_\pi \times m_\pi},\ A \succeq 0 \Big\},$$

where $\varphi_1, \ldots, \varphi_{m_\pi}$ is a basis of $\operatorname{Hom}_\Gamma(X, \mathcal{H}_\pi)$. That is, this conic hull is isomorphic to a complex positive semidefinite cone. This suggests how to block diagonalize the cone $\mathcal{C}_{\mathbb{C}}(X \times X)^\Gamma_{\succeq 0}$, but since $\hat{\Gamma}$ and m_π are infinite for the cases we are interested in, we need to consider convergence.

Given a symmetry adapted system $\{e_{\pi,i,j}\}$ of X, for each $\pi \in \hat{\Gamma}$ we define the *zonal matrix*

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

where $E_\pi(x)$ is the matrix defined as $E_\pi(x)_{i,j} = e_{\pi,i,j}(x)$ for $x \in X$, $i \in [m_\pi]$, and $j \in [d_\pi]$. The *Fourier coefficients* of a kernel $K \in \mathcal{C}_{\mathbb{C}}(X \times X)$ are then defined by

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

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 converges in L^2 (see Proposition 3.4.1). By Proposition 3.4.3 the kernel K is positive definite if and only if $\hat{K}(\pi)$ is positive semidefinite for all $\pi \in \hat{\Gamma}$. 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 a part of Bochner's theorem, see Theorem 3.4.4 for an alternative proof). We are interested in the situation of infinitely many orbits where the above series generally 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 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 complex positive semidefinite matrices. Let C_d be the Minkowski sum $\sum_{\pi \in \hat{\Gamma}} C_{\pi,d}$. Then we have

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

In Theorem 3.4.5 we show $\bigcup_{d=0}^{\infty} C_d$ is uniformly dense in $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$.

7.6.2. Harmonic analysis on subset spaces. Here we show how to construct a sequence $\{C_d\}$ of inner approximating cones of $\mathcal{C}_{\mathbb{C}}(I_t \times I_t)^{\Gamma}_{\succeq 0}$ as discussed in the previous section. We give a construction in two steps: First we construct such a sequence for the cone $\mathcal{C}_{\mathbb{C}}(X_t \times X_t)^{\Gamma}_{\succeq 0}$, where X_t is a larger – but simpler – space containing I_t as an embedding. Then we restrict the kernels in the approximating cones to the smaller space $I_t \times I_t$.

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. The group Γ has a continuous action 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 which sends $\{x_1, \ldots, x_i\} \in I_{=i}$ to $\{(x_{\sigma(1)}, \ldots, x_{\sigma(i)}) : \sigma \in S_i\}$. Notice we could have

embedded 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 use X_t.

We first show that each kernel in $\mathcal{C}_{\mathbb{C}}(I_t \times I_t)^{\Gamma}_{\succeq 0}$ is the restriction to $I_t \times I_t$ of a kernel from $\mathcal{C}_{\mathbb{C}}(X_t \times X_t)^{\Gamma}_{\succeq 0}$. It then follows that a sequence of inner approximating cones whose union is uniformly dense in the latter cone yields such a sequence for the former cone.

LEMMA 7.6.1. *Let X 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. Every kernel in $\mathcal{C}_{\mathbb{C}}(Y \times Y)^{\Gamma}_{\succeq 0}$ is the restriction to $Y \times Y$ of a kernel in $\mathcal{C}_{\mathbb{C}}(X \times X)^{\Gamma}_{\succeq 0}$.*

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

Given a point $x \in X$, define

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

where d is the metric on X. For each i we define the lower semicontinuous function c_i on X by

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

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 the Tietze extension theorem there exist functions $f_i \in \mathcal{C}(X; \mathbb{C})$ with $f_i|_Y = e_i$. Let

$$Q_i = \{x \in X : |f_i(x)|^2 \geq c_i(x) + 1/2^i\}.$$

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

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

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,

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

We have

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

We can use the axiom of choice to select an element $y_x \in Y_x$ for each $x \in X$. It follows that

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

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

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

We use symmetric tensor powers to give an explicit construction of a symmetry adapted system of X_t. Here it is convenient to define the measure ν on X_t in terms of a strictly positive Radon probability measure μ on V by

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

First some background on symmetric tensor powers: The (algebraic) *tensor product* $\mathcal{U}\otimes\mathcal{V}$ of two complex vector spaces $\mathcal{U}$ and $\mathcal{V}$ is the unique (up to isomorphisms) complex vector space for which there exists a bilinear map $\phi\colon \mathcal{U}\times\mathcal{V}\to\mathcal{U}\otimes\mathcal{V}$ such that each bilinear map $h\colon \mathcal{U}\times\mathcal{V}\to\mathcal{W}$ into another complex vector space $\mathcal{W}$ admits a unique linearization through ϕ; that is, there is a unique linear map $\tilde{h}\colon \mathcal{U}\otimes\mathcal{V}\to\mathcal{W}$ with $h = \tilde{h}\circ\phi$. The map ϕ is unique, and we use the notation $a\otimes b = \phi(a,b)$. In general, ϕ is not surjective, but if $\{u_i\}$ and $\{v_j\}$ are bases of $\mathcal{U}$ and $\mathcal{V}$, then $\{u_i\otimes v_j\}$ is a basis of $\mathcal{U}\otimes\mathcal{V}$ (the elements in the image of ϕ are called rank-1 tensors). The tensor product is associative and we use the notation $\mathcal{V}^{\otimes n}$ for the nth tensor power; that is, $\mathcal{V}^{\otimes n} = \mathcal{V}\otimes\cdots\otimes\mathcal{V}$ (n times) of $\mathcal{V}$. Given vector spaces $\mathcal{V}_1,\dots,\mathcal{V}_n$, vectors $v_i\in\mathcal{V}_i$, and an element σ in the symmetric group S_n, we use the notation $(\otimes_{i=1}^n v_i)^\sigma = \otimes_{i=1}^n v_{\sigma(i)}$, and extend this linearly to $\otimes_{i=1}^n\mathcal{V}_i$. We use this to define the nth *symmetric tensor power* of $\mathcal{V}$ by

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

It follows from the polarization identity that $\mathcal{V}^{\odot n} = \operatorname{span}\{v^{\otimes n} : v\in\mathcal{V}\}$, and this shows $w^\sigma = w$ for all $w\in\mathcal{V}^{\odot n}$ and $\sigma\in S_n$.

If $\mathcal{U}$ and $\mathcal{V}$ have inner products, then we equip the tensor product $\mathcal{U}\otimes\mathcal{V}$ 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. This defines a topology on the tensor product, and the tensor product of two Hilbert spaces is defined to be the completion of the vector space tensor product in this topology. This extends to finite products, and we define

$$\mathcal{H}^{\odot n} = \Big\{\sum_{\sigma\in S_n} w^\sigma : w\in\mathcal{H}^{\otimes n}\Big\} = \operatorname{cl}(\operatorname{span}\{v^{\otimes n} : v\in\mathcal{H}\}).$$

The *(inner) tensor product representation* $\pi_1 \otimes \pi_2 \colon \Gamma \to U(\mathcal{H}_1 \otimes \mathcal{H}_2)$ 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

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

This can also be extended to finite products and finite (symmetric) powers.

Let $L_i \colon L^2_{\mathbb{C}}(V,\mu)^{\odot i} \to L^2_{\mathbb{C}}(X_t,\nu)$ be the operator defined by

$$L_i(f)|_(\{(x_{\sigma(1)},\dots,x_{\sigma(i)}) : \sigma \in S_i\}) = f(x_1,\dots,x_i).$$

These are isometric, Γ-equivariant operators with pairwise orthogonal images, and the sum of the images under T_i of the ith symmetric algebraic tensor powers of $\mathcal{C}_{\mathbb{C}}(V)$ is uniformly dense in $\mathcal{C}_{\mathbb{C}}(X_t)$.

We assume we have a symmetry adapted system of V. Such a system defines a sequence $\{\mathcal{H}_k\}_{k=1}^m$ (where $1 \le m \le \infty$) of irreducible and pairwise orthogonal subrepresentations of $L^2_{\mathbb{C}}(V,\mu)$ whose sum is uniformly dense in $\mathcal{C}_{\mathbb{C}}(V)$. The spaces

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

are orthogonal, Γ-invariant subspaces of $L^2_{\mathbb{C}}(V,\mu)^{\otimes i}$ (in the tensor product here we exclude the factors $\mathcal{H}_k^{\odot \tau_k}$ where $\tau_k = 0$, so that this becomes a finite tensor product).

For each $0 \le i \le t$, we will define a unitary, Γ-equivariant operator

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

such that the algebraic sum of $T_i(\otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k})$ over all $\tau \in D_i$ is uniformly dense in the ith algebraic tensor product of $\mathcal{C}_{\mathbb{C}}(V)$.

The space $\otimes_{k=1}^m \mathcal{H}_k^{\odot \tau_k}$ decomposes into 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 R_τ is a finite subset of $\hat{\Gamma}$, and where $\mathcal{H}_\pi$ is the Hilbert space of the irreducible representation $\pi \in \hat{\Gamma}$. In the following section we construct these operators M_τ explicitly for the case where $t = 2$, $V = S^2$, μ is the surface measure on S^2, and $\Gamma = O(3)$.

Let $\{e_{\pi,1},\dots,e_{\pi,d_\pi}\}$ be an orthonormal basis of $\mathcal{H}_\pi$. Then,

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

is a symmetry adapted system of X_t.

In the remainder of this section we give the definition of T_i and show it is a Γ-equivariant and unitary operator. This generalizes a result from [**3**] to infinite direct sums, and we additionally consider unitarity and equivariance. Given $\tau \in D_i$, we let A_τ be the subgroup of all $\sigma \in S_i$ for which the set

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

is invariant under the permutation $\sigma\colon [i] \to [i]$ 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

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

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 if $\tau \neq \tau'$ or $\sigma \neq \sigma'$. So we can define T_i by sending an element $w \in \otimes_{k=1}^m \mathcal{H}_k^{\odot\tau_k}$, where $\tau \in D_i$, to $1/|B_\tau| \sum_{[\sigma]\in B_\tau} w^\sigma$ and extending by linearity.

LEMMA 7.6.2. *The operators T_i are unitary and Γ-equivariant, and the algebraic sum of $T_i(\otimes_{k=1}^m \mathcal{H}_k^{\odot\tau_k})$ over all $\tau \in D_i$ is uniformly dense in the ith algebraic symmetric tensor power of $\mathcal{C}_{\mathbb{C}}(V)$.*

PROOF. By definition T_i is linear. It is also an isometry: Given $w \in \otimes_{k=1}^m \mathcal{H}_k^{\odot\tau_k}$, we have

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

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 the ith algebraic tensor power of $\mathcal{C}_{\mathbb{C}}(V)$. 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, the image of algebraic direct sum of the spaces $\otimes_{k=1}^m \mathcal{H}_k^{\odot\tau_k}$ over all $\tau \in D_i$ under T_i is uniformly dense in ith algebraic tensor product of $\mathcal{C}_{\mathbb{C}}(V)$ and hence is dense in $L^2_{\mathbb{C}}(V)^{\odot i}$. Hence, T_i is an isometry whose image is dense in $L^2_{\mathbb{C}}(V)^{\odot i}$ and therefore is a unitary operator.

Since the spaces $\mathcal{H}_k$ are π-invariant, the spaces $\otimes_{k=1}^m \mathcal{H}_k^{\odot\tau_k}$, for $\tau \in D_i$, are $\pi^{\otimes i}$-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. □

7.6.3. Explicit computations for the sphere. In this section we construct a symmetry adapted system of X_2, where

$$X_2 = \bigcup_{i=0}^{2} V^i/S_i, \quad V = S^2, \quad \text{and} \quad \Gamma = O(3),$$

and where X_2 has the measure ν as defined in the previous section. We then use this to construct the zonal matrices and inner approximating cones of $\mathcal{C}_{\mathbb{C}}(X_2 \times X_2)^{\Gamma}_{\succeq 0}$. For some material we only state the results and provide references. For less familiar material in this context, such as our use of $O(3)$ instead of $SO(3)$, the focus on cartesian coordinates instead of spherical coordinates, and the focus on symmetric tensor powers, we give more details. The formulas and constants are given explicitly, so that these can be used to construct a software implementation to compute the zonal matrices.

Let $\mathcal{H}_\ell \subseteq \mathcal{C}_{\mathbb{C}}(S^2)$ 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 [**38**]. The space $\mathcal{H}_\ell$ has dimension $2\ell + 1$. Moreover, these spaces are orthogonal and form irreducible subrepresentations of the unitary representation

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

where ω is the invariant measure on the sphere with normalization $\omega(S^2) = 4\pi$. 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 $\mathcal{C}_{\mathbb{C}}(S^2)$, and $L^2_{\mathbb{C}}(S^2, \omega)$ decomposes as the Hilbert space direct sum of the spaces $\mathcal{H}_\ell$. Moreover, up to equivalence these are all irreducible representations of $SO(3)$.

The *Laplace spherical harmonics* Y_ℓ^m provide an explicit set of orthonormal bases of the spaces $\mathcal{H}_\ell = \operatorname{span}\{Y_\ell^m : m = -\ell, \dots, \ell\}$. 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

$$c_\ell^m = (-1)^m \sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}$$

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

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

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 through the Clebsch–Gordan coefficients [**38**], which can be expressed as

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

by setting

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

and extending by linearity. Since the Clebsch–Gordan coefficients are real numbers, it follows that

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

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

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

and by using the symmetry relation

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

of the Clebsch–Gordan coefficients, 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, so

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

We have

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

so

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

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

We have shown how $L^2_{\mathbb{C}}(S^2,\omega)$ 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 $\mathbb{Z}_2 \times SO(3)$. Consequently, for each irreducible representation $\mathcal{H}_\ell$ of $SO(3)$, we define two nonequivalent irreducible representations $\pi_\ell^p \colon O(3) \to U(\mathcal{H}_\ell^p)$, with $p = \pm 1$, where $\mathcal{H}_\ell^{+1}$ and $\mathcal{H}_\ell^{-1}$ are both isomorphic to $\mathcal{H}_\ell$ as Hilbert spaces, and where $\pi_\ell^p|_{SO(3)}$ is equivalent to $\mathcal{H}_\ell$, but where $\pi_\ell^p(-I)f = pf$. It follows that π_ℓ^p is a subrepresentation of the unitary representation

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

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

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

which implies

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

We can view the operators Φ_{ℓ_1,ℓ_2} and Φ_ℓ defined above as $O(3)$-equivariant, unitary operators $\mathcal{H}_{\ell_1}^{p_1} \otimes \mathcal{H}_{\ell_2}^{p_2} \to \mathcal{H}_{|\ell_1-\ell_2|}^{p_1p_2} \oplus \cdots \oplus \mathcal{H}_{\ell_1+\ell_2}^{p_1p_2}$ and $(\mathcal{H}_{\ell'}^{p})^{\odot 2} \to \mathcal{H}_0^{+1} \oplus \mathcal{H}_2^{+1} \oplus \cdots \oplus \mathcal{H}_{2\ell'}^{+1}$.

We use these operators to give an explicit definition of the operators M_τ from the previous section. Let e_ℓ denote the vector in $\mathbb{N}_0^\infty$ where $(e_\ell)_{\ell'} = \delta_{\ell,\ell'}$. 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 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 $R_{(\ell,p)} = R_{(\ell,p)}^0 \cup R_{(\ell,p)}^1 \cup R_{(\ell,p)}^2$,

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

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

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 $\cup_{d=0}^\infty R_{(\ell,p),d} = R_{(\ell,p)}$, and for each d only finitely many sets $R_{(\ell,p),d}$ are nonempty. The natural way to do that here is to define

$$R_{(\ell,p),d} = \big\{\tau \in R_{(\ell,p)} : \sum_{\ell'} \tau_{\ell'} \leq d\big\};$$

that is, if we view the basis elements as multivariate polynomials, we restrict the degree to be at most d.

Although the symmetry adapted system constructed in this section does not consist of real-valued functions, it can easily be verified that the resulting zonal matrices do have real-valued entries. Following the discussion at the end of Section 3.4 we see that we then also get a sequence of inner approximating cones of the cone $\mathcal{C}(X_2 \times X_2)_{\succeq 0}^\Gamma$ consisting of real-valued kernels. The proof of Lemma 7.6.1 also works for real-valued kernels, so this yields the desired sequence of inner approximating cones of $\mathcal{C}(I_2 \times I_2)_{\succeq 0}^\Gamma$.

7.7. Reduction to semidefinite programs with polynomial constraints

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

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

where $g_1, \dots, g_m \in \mathbb{R}[x_1, \dots, x_n]$, and the polynomials defining the semialgebraic sets are assumed to be part of the description of the problem. In the next section we show how these programs can be approximated by semidefinite programs and how symmetries in these polynomial constraints can be exploited to (further) block diagonalize the semidefinite programming formulations.

Following Section 7.2 and Section 7.5 the t-th step in the dual hierarchy for the Thomson problem reads

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

where I_t is the set of independent sets of cardinality at most t in the graph on the unit sphere S^2 defined in Section 7.2, and where

$$f(S) = \begin{cases} \|x - y\|_2^{-1} & \text{if } S = \{x, y\} \text{ with } x \neq y, \\ 0 & \text{otherwise.} \end{cases}$$

In the symmetrized version of this problem, as derived in Section 7.5.2, we restrict the kernels in the cone $\mathcal{C}(I_t \times I_t)_{\succeq 0}$ to be $O(3)$-invariant, and in Section 7.6.3 we construct the sequence

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

of inner approximations to the latter cone. By replacing $\mathcal{C}(I_t \times I_t)_{\succeq 0}^{O(3)}$ with C_d we obtain the sequence of approximations:

$$\tilde{Q}_{t,d} = \sup\Big\{ \sum_{i=0}^{2t} \binom{N}{i} a_i : a \in \mathbb{R}^{\{0,\dots,2t\}},\ A \in \bigoplus_{\pi \in \hat{\Gamma}} S_{\succeq 0}^{R_{\pi,d}},\\ r_{a,A}^i(S) \leq f(S) \text{ for } S \in I_{=i} \text{ and } i = 0, \dots, 2t\Big\},$$

where

$$r_{a,A}^i(S) = a_i + A_t\big(\sum_{\pi \in \hat{\Gamma}} \big\langle A_\pi, Z_{\pi,d}(\cdot,\cdot) \big\rangle \big)(S).$$

For each t and d, the problem $Q_{t,d}$ has finite dimensional variable space, and we have $\lim_{d\to\infty} Q_{t,d} = E_t^* = E_t$ and $\lim_{d\to\infty} Q_{N,d} = E$.

Let $\mathbb{R}[x_1, \ldots, x_i]$ be the ring of real polynomials in $3i$ variables, where each x_j is a vector of 3 variables. In Section 7.6.3 we construct the matrix entries of Z_π using the spherical harmonics, and this defines polynomials $p^i_{a,A} \in \mathbb{R}[x_1, \ldots, x_i]$ with

$$r^i_{a,A}(\{x_1, \ldots, x_i\}) = p^i_{a,A}(x_1, \ldots, x_i) \quad \text{for all} \quad \{x_1, \ldots, x_i\} \in I_{=i}.$$

The coefficients of $p^i_{a,A}$ depend linearly on the vector a and the matrices A_π. These polynomials $p^i_{a,A} \in \mathbb{R}[x_1, \ldots, x_i]$ are $O(3)$-invariant:

$$p^i_{a,A}(x_1, \ldots, x_i) = p(\gamma x_1, \ldots, \gamma x_i) \quad \text{for all} \quad x_1, \ldots, x_i \in S^2 \quad \text{and} \quad \gamma \in O(3).$$

It follows from the first fundamental theorem of invariant theory for the orthogonal group [**59**, Theorem 10.2] that $p^i_{a,A}$ can be written as a polynomial in the $\binom{i+1}{2}$ inner products. Restricted to sphere we have the identities $x_j \cdot x_j = 1$, for $j \in [i]$, so there is a polynomial $q^i_{a,A} \in \mathbb{R}[u_1, \ldots, u_{\binom{i}{2}}]$ such that

$$p^i_{a,A}(x_1, \ldots, x_i) = q^i_{a,A}(x_1 \cdot x_2, x_1 \cdot x_3, \ldots, x_i \cdot x_i) \quad \text{for all} \quad x_1, \ldots, x_i \in S^2.$$

The polynomial $q^i_{a,A}$ is not unique in general. The use of this theorem is why, as mentioned in the previous section, we need $O(3)$-invariance, instead of just $SO(3)$-invariance. For otherwise the polynomials $q^i_{a,A}$ would also depend on the determinants of the 3×3 matrices whose columns are given by vectors from $\{x_1, \ldots, x_i\}$, which would mean we have too many variables.

The degenerate polynomials $q^0_{a,A}$ and $q^1_{a,A}$ have 0 variables; they are linear combinations of the entries of the vector a and the matrices A_π. The constraints $r^i_{a,A}|_{I_{=i}} \leq 0$ for $i = 0, 1$ in $Q_{t,d}$ therefore reduce to the two linear constraints $q^0_{a,A} \leq 0$ and $q^1_{a,A} \leq 0$.

For distinct $x, y \in S^2$ we have

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

So, by using the substitution $w = \sqrt{2 - 2u}$ we can write the constraint $r^2_{a,A}|_{I_{=2t}} \leq f|_{I_{=2t}}$ in $Q_{t,d}$ as the polynomial nonnegativity constraint

$$1 - w - w\, q^2_{a,A}(1 - w^2/2) \geq 0$$

on the interval $[B^{-1}, 2]$.

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

$$I_{=i} = \Big\{\{x_1, \ldots, x_i\} \subseteq S^2 : x_j \cdot x_{j'} \leq 1 - \frac{1}{2}B^{-2} \text{ for } 1 \leq j < j' \leq i\Big\}.$$

Let

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

A constraint $r^i_{a,A}|_{I=i} \le 0$ for $i \in \{3,\dots,2t\}$ can be written as $q^i_{a,A}|_{\mathcal{I}=i} \le 0$. This means we can write the problem $Q_{t,d}$ as

$$\begin{aligned}\tilde{Q}_{t,d} = \sup\Big\{ \sum_{i=0}^{2t} \binom{N}{i} a_i : a \in \mathbb{R}^{\{0,\dots,2t\}},\ A \in \bigoplus_{\pi\in\tilde{\Gamma}} S^{R_{\pi,d}}_{\succeq 0},\\ q^0_{a,A} \le 0,\ q^1_{a,A} \le 0,\\ 1 - w - w\, q^2_{a,A}(1 - w^2/2) \ge 0 \text{ for } w \in [B^{-1}, 2],\\ q^i_{a,A}|_{\mathcal{I}=i} \le 0 \text{ for } i = 3,\dots,2t\Big\}.\end{aligned}$$

Given a feasible solution (a, K) of $\tilde{Q}_{t,d}$, we can modify a slightly so that all polynomial inequalities are satisfied strictly. So, the program $Q_{t,d}$, defined to be the same as $\tilde{Q}_{t,d}$ except that we replace all inequalities by strict inequalities, has the same optimal value.

To describe $\mathcal{I}_{=i}$ as a semialgebraic set we first observe that by using the Gram decomposition of a positive semidefinite matrix, it can be written as

$$\mathcal{I}_{=i} = \Big\{u \in \mathbb{R}^{\binom{i}{2}} : u_j \le 1 - 1/(2B^2) \text{ for } j \in [\tbinom{i}{2}],\ \mathcal{E}(u) \succeq 0,\ \operatorname{rank}(\mathcal{E}(u)) \le 3\Big\},$$

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

$$\begin{aligned}\mathcal{I}_{=i} = \Big\{u \in \mathbb{R}^{\binom{i}{2}} : u_j \le 1 - 1/(2B^2) \text{ for } j \in [\tbinom{i}{2}],\\ g(u) \ge 0 \text{ for } g \in G_{i,j} \text{ with } 2 \le j \le 3,\\ g(u) = 0 \text{ for } g \in G_{i,j} \text{ with } 4 \le j \le i\Big\},\end{aligned}$$

where $G_{i,j}$ is the set of principal minors (the determinants of principal submatrices) of $\mathcal{E}(u)$ of order j.

From the symmetry of the sphere and pair potential it follows that the polynomials $p^i_{a,A}$ are $O(3)$-invariant. The Thomson problem, however, admits even more symmetry: the particles are interchangeable. The polynomials $p^i_{a,A}$ are invariant under the symmetric group; that is, $p^i_{a,A}(x_1,\dots,x_i) = p^i_{a,A}(x_{\sigma(1)},\dots,x_{\sigma(i)})$ for all $x_1,\dots,x_i \in S^2$ and $\sigma \in S_i$. This extra symmetry translates into symmetry in the polynomials $q^i_{a,A}$: The relation between the x and u variables gives a group homomorphism τ_i from the symmetric group on i elements to the symmetric group on $\binom{i}{2}$ elements. For example, τ_2 is the trivial map $S_2 \to S_1$, and τ_3 sends the generators $(1,2)$ and $(1,3)$ of S_3 to the generators $(2,3)$ and $(1,2)$ of S_3. The map τ_4 sends the generators $(1,2)$, $(2,3)$, and $(3,4)$ of S_4 to the elements $(3,5)(2,4), (1,2)(5,6)$, and $(4,5)(2,3)$ of S_6. We can view $\tau_i(S_4)$ as a subgroup of the orthogonal group $O(\binom{i}{2})$, where we represent the group elements as permutation matrices. The polynomials $q^i_{a,A}$ are invariant under the representation $L_i\colon \tau_i(S_i) \to O(\mathbb{R}[u])$, $L_i(\gamma)p(u) = p(\gamma^{-1}u)$. Moreover, the sets $G_{i,j}$, for $0 \le j \le i$, are invariant under this representation L_i.

7.8. Invariant polynomials in the quadratic module

In this section we first explain how Putinar's theorem can be used to approximate a semidefinite program with polynomial constraints by a sequence of block diagonal semidefinite programs. We then show how symmetry of the polynomial constraints can be used to further block diagonalize these semidefinite programming formulations. By using the problems $Q_{2,d}$ from the previous section as example we show this can lead to significant computational savings.

The *quadratic module* generated by a set of polynomials $g_1, \ldots, g_m$ from the ring $\mathbb{R}[x] = \mathbb{R}[x_1, \ldots, x_n]$ reads

$$M(g_1, \ldots, g_m) = \Big\{ \sum_{i=0}^{m} g_i s_i : s_0, \ldots, s_m \in \mathbb{R}[x] \text{ are sum of squares polynomials} \Big\},$$

where for convenience g_0 always denotes the constant one polynomial. 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 [**81**], which says that under the condition that $\{g_1, \ldots, g_m\}$ has the Archimedean property, every strictly positive polynomial on $S(g_1, \ldots, g_m)$ is contained in $M(g_1, \ldots, g_m)$. A set of polynomials $\{g_1, \ldots, g_m\}$ has the *Archimedean property* if its quadratic module contains a polynomial p whose semialgebraic set $S(p)$ is compact; this is an algebraic certificate of the compactness of $S(g_1, \ldots, g_m)$.

For $e \in \mathbb{N}_0$, we define the *truncated quadratic module* $M_e(g_1, \ldots, g_m)$ in the same way as we defined $M(g_1, \ldots, g_m)$, except that we require s_i to have degree at most $2h_i$, where $h_i = \lfloor (e - \deg(g_i))/2 \rfloor$. This is different from the set we obtain when we restrict the degrees of the polynomials in $M(g_1, \ldots, g_m)$ to be at most $2h_i$, since terms can cancel each other out. Putinar's theorem shows that each polynomial p that is strictly positive on $S(g_1, \ldots, g_m)$ is contained in $M_e(g_1, \ldots, g_m)$ for every large enough e. In [**78**] an upper bound on the smallest e for which this is true is given in terms of the polynomials $g_1, \ldots, g_m$, the degree of p, and how close p is to having a zero on $S(g_1, \ldots, g_m)$.

Let $v_i(x)$ be a vector whose entries form a basis of the polynomials of degree at most h_i. The cone of sum of squares polynomials of degree at most $2h_i$ consists of polynomials of the form

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

where Q_i ranges over the positive semidefinite matrices of size $\binom{n+h_i}{n}$. Here a Cholesky factorization $Q_i = R_i^{\mathsf{T}} R_i$ can be used to prove $v_i(x)^{\mathsf{T}} Q_i v_i(x)$ is a sum of squares. This implies

$$M_e(g_1, \ldots, g_m) \simeq S_{\succeq 0}^{\binom{n+h_0}{n}} \times \cdots \times S_{\succeq 0}^{\binom{n+h_m}{n}}.$$

Given a semidefinite program with polynomial constraints, we say we model a polynomial constraint $p|_{S(g_1,\ldots,g_m)} > 0$ by a degree d sum of squares characterization if we introduce the additional positive semidefinite matrix variables $Q_0, \ldots, Q_m$ and

replace the above constraint by a set of linear constraints which enforce the identity

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

To obtain a set of linear constraints for the above identity we can express the left and right hand sides in terms of the same basis and equate the coefficients with respect to this basis. As discussed in Chapter 4, the choice of basis for the entries of v_i and the basis choice for the linear constaints can have great impact on the numerical conditioning of the resulting semidefinite program. The semidefinite programs we obtain in this way become arbitrarily good as we take characterizations of higher degrees, and if the semidefinite program with polynomial constraints has an optimal solution, then the optimum is obtained for finite degree sum of squares characterizations.

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 $O(n)$. This induces the orthogonal representation $L\colon \Gamma \to O(\mathbb{R}[x])$, $L(\gamma)p(x) = p(\gamma^{-1}x)$, where the inner product on $\mathbb{R}[x]$ is given by $\langle p, q\rangle = \sum_\alpha p_\alpha q_\alpha$. A polynomial p is Γ-invariant if $L(\gamma)p = p$ for all $\gamma \in \Gamma$, and a set of polynomials $\{g_1, \ldots, g_m\}$ is Γ-invariant if $\{L(\gamma)g_1, \ldots, L(\gamma)g_m\} = \{g_1, \ldots, g_m\}$ for all $\gamma \in \Gamma$. We denote by Γ_{g_i} the stabilizer subgroup of Γ with respect to g_i. In the next proposition we show that if the polynomials p and the set $\{g_1, \ldots, g_m\}$ are invariant, then the sum of squares polynomials are invariant under the corresponding stabilizer subgroups.

PROPOSITION 7.8.1. *Every Γ-invariant polynomial $p \in M_d(g_1, \ldots, g_m)$ can be written as $p = \sum_{i=0}^m g_i s_i$, where s_i is a Γ_{g_i}-invariant sum of squares polynomial with $\deg(s_i) \le h_i$.*

PROOF. Let $\Delta_{i,j} = \{\gamma \in \Gamma : L(\gamma)g_j = g_i\}$, so that $\Delta_{i,i} = \Gamma_{g_i}$. We have

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

So, if we define

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

then $p(x) = \sum_{i=0}^m g_i(x)\bar{s}_i(x)$. The functions $s_0, \ldots, s_m$ are sums of squares polynomials because the cone of sum of squares polynomials is $O(n)$-invariant. Moreover, for $\eta \in \Delta_{i,i}$ we have

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

so $\Delta_{i,i}$-invariance of s_i follows from the fact that $\eta\Delta_{i,j} = \Delta_{i,j}$ for all $\eta \in \Delta_{i,i}$. □

As shown in [**37**], the representation of a polynomial as a sum of squares can be block diagonalized when the polynomial is Γ-invariant. In our application to $Q_{2,d}$, the representation L only uses permutation matrices, so we give a derivation for this slightly simpler case. Assume $v(x)$ consists of monomials. We represent $L(\gamma)$ as a real $\binom{n+h}{h} \times \binom{n+h}{h}$ matrix with respect to the basis spanned by the entries of $v(x)$, so that $L(\gamma)$ is a permutation matrix and $v(\gamma^{-1}x) = L(\gamma)v(x)$ for each $\gamma \in \Gamma$. Then,

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

where

$$\bar{Q} = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma} L(\gamma)^{\mathsf{T}} Q L(\gamma).$$

The matrix $\bar{Q}$ commutes with $L(\gamma)$ for all $\gamma \in \Gamma$, so we can view it as a Γ-invariant, positive definite kernel $[\binom{n+h}{h}] \times [\binom{n+h}{h}] \to \mathbb{R}$, and we can use the techniques from Chapter 3 to block diagonalize it. The block sizes are given by the numbers m_π, where m_π denotes the number of times the real irreducible representation π occurs in a complete orthonormal symmetry adapted system of $\mathbb{R}[x]_h$. These numbers can be computed through the use of a Hilbert–Poincaré series. The subspaces $\mathbb{R}[x]_{=k}$ of $\mathbb{R}[x]_d$ consisting of homogeneous polynomial of degree k are Γ-invariant, and we have $\mathbb{R}[x]_d = \oplus_{k=0}^d \mathbb{R}[x]_{=k}$. We denote the number of times π occurs in a complete orthonormal symmetry adapted system of $\mathbb{R}[x]_{=k}$ by m_π^k, so that $m_\pi = \sum_{k=0}^d m_\pi^k$. The Molien series is the formal power series

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

where the second equality is a theorem of Molien [**75**]. We use a Magma [**18**] implementation to compute the block sizes in the block diagonalization of Q.

We apply this to the problems $Q_{2,d}$. We model the polynomial constraints $q_{a,A}^i|_{\mathcal{I}_{=i}} < 0$, for $i = 3, 4$, using the semialgebraic description of $\mathcal{I}_{=4}$ as given in the previous section. In the description of $\mathcal{I}_{=4}$ we do not just have polynomial inequalities but also a polynomial equality: we restrict to u for which $\det(\mathcal{E}(u)) = 0$. We could model this constraint by just having the constraints $\det(\mathcal{E}(u)) \geq 0$ and $-\det(\mathcal{E}(u)) \geq 0$, but this is inefficient. Instead we have free variables z_α and the characterization

$$\begin{aligned} q_{a,A}^4(u) + v_0(u)^{\mathsf{T}} Q_0 v_0(u) + \sum_{j=2}^{3} \sum_{g\in G_{4,j}} g(u) v_g(u)^{\mathsf{T}} Q_g v_g(u) \\ + \det(\mathcal{E}(u)) \sum_{\alpha\in\mathbb{N}_0^6 : \|\alpha\|_1 \leq e-4} z_\alpha u_1^{\alpha_1}\cdots u_6^{\alpha_6} = 0. \end{aligned} \tag{27}$$

e	$\binom{6+\lfloor e/2 \rfloor}{6}$	g_0	$G_{4,2}$	$G_{4,3}$
0	1	1	0	0
1	1	1	0	0
2	7	2	1	0
3	7	2	1	1
4	28	5	4	1
5	28	5	4	3
6	84	12	13	3
7	84	12	13	9
8	210	29	33	9
9	210	29	33	27
10	462	63	75	27
11	462	63	75	69
12	924	124	153	69
13	924	124	153	153
14	1716	228	291	153
15	1716	228	291	306
16	3003	395	519	306
17	3003	395	519	570
18	5005	654	882	570
19	5005	654	882	999
20	8008	1040	1435	999

TABLE 1. Largest block sizes in semidefinite programming formulations of $Q_{2,d}$ with and without exploiting symmetry.

In practice, the occurence of large blocks in a semidefinite program is problematic, and given that the numerical conditioning does not get much worse, we prefer to have several smaller blocks instead. The sizes of the matrices A_π in $Q_{2,d}$ grow slowly in d, so the largest block in the semidefinite programming approximations of $Q_{2,d}$ will occur in modeling the polynomial constraint $q^4_{a,A}|_{\mathcal{I}_{=4}} < 0$ by a sum squares characterization. Here $q^4_{a,A}$ is a polynomial of degree d in 6 variables. If we use sum of squares characterizations of degree e (in practice we usually take $d \leq e \leq 2d$), then without the use of symmetry, the largest block will have size $\binom{6+\lfloor e/2 \rfloor}{6}$. The stabilizer subgroup of $\tau_4(S_4)$ with respect to the constant 1 polynomial is isomorphic to S_4, and with respect to a polynomial $g \in G_{4,j}$, it is isomorphic to the Klein-Four group for $j = 2$ and to S_3 for $j = 3$. In Table 1 we show in the second column the size of Q_0 in (27) when not using symmetry reduction. In the third column we show the size of the largest block size in the block diagonalization of Q_0. In the fourth and fifth columns we show the size of the larges block size in the block diagonalization of Q_g for $g \in G_{4,j}$. Here we see that (at least in this range) the largest matrix in the semidefinite program is approximately a factor 6 smaller when using this symmetry reduction.

7.9. Computations

Given a system of N particles, we know the N-th step E_N is guaranteed to give the optimal energy E. It could be the case, however, that for many problems the bound E_t is sharp for much smaller values of t. After symmetry reduction the dual of the first step E_1^* essentially reduces to Yudin's bound, and hence is sharp for the Thomson problem with $N = 2, 3, 4, 6, 12$. The goal here is to show it is possible to

compute the second step E_2, which is a 4-point bound, and to show it is sharp (up to solver precision) for $N = 5$.

To compute E_2 for the Thomson problem with $N = 5$ we write a program which generates the semidefinite program $Q_{2,6}$ from Section 7.7, where we model the polynomial inequalities using the techniques described in Section 7.8. We compute the optimal value of this semidefinite program with a semidefinite programming solver and check that the optimal objective is equal, up to solver precision, to the energy of the vertices of the triangular bipyramid.

Our code is written in Julia [**12**], which is a high-level dynamic language which allows for quick experimentation (which we have done extensively for this project) with different algorithms and data structures, and it has a high quality just-in-time compiler which makes code execution fast.

First we generate the symmetry adapted system and zonal matrices described in Section 7.6.3. For this we constructed a simple library for sparse multivariate polynomials, which includes generators for the Laplace spherical harmonics and code for generating the Clebsch–Gordan coefficients. To generate high quality input for the solver we perform the computations in high precision arithmetic using the MPFR library [**35**].

As described in Section 7.7 we write the generated polynomial entries of the zonal matrices in terms of the inner products. For this we need to solve a large number of instances of the following problem: Let $p \in \mathbb{R}[x_1, \dots, x_i]$, where each x_i is a vector of 3 variables, be $O(3)$-invariant. We want to find a polynomial $q \in \mathbb{R}[u_1, \dots, u_s]$, with $s = \binom{i+1}{2}$, such that

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

As mentioned before, such a q is guaranteed to exist, and hence p must have even degree $2d$. If $m \in \mathbb{R}[u_1, \dots, u_s]$ is a monomial, then the polynomial $m(x_1 \cdot x_2, \dots, x_{i-1} \cdot x_i)$ 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 monomials in $3i$ variables up to degree $2d$, and the columns of A and rows of x by the monomials in s variables up to degree d. The size of A quickly grows large: for $i = 4$ and $d = 6$ it has about 2.7 million rows. The matrix is sparse, however, where the maximum number of nonzero's in a row is 3^d, and although this is exponential, 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 typically is not of full column rank (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 which can work with sparse data structures is not readily available, we implement a simple pivoting 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 use Magma [**18**] code to generate the block diagonalized sum of squares characterizations described in Section 7.8. This code uses Magma functions to compute the stabilizer subgroups and irreducible representations automatically. Then we implement the algorithm described in [**91**] to generate the symmetry adapted systems. We transfer these into our Julia code to construct the block diagonalized sum of squares formulations.

We develop a simple semidefinite programming specification library in Julia which we use in combination with the above code to generate the semidefinite programs. This program outputs files in the SDPA-sparse format which can function as input file for many semidefinite programming solvers. We solve the generated semidefinite programs using CSDP [**16**], which implements a machine precision primal-dual interior point method to solve the semidefinite programs.

In addition to computing the Coulomb energy for the Thomson problem, we use a different variable substitution in Section 7.7 to compute the Riesz-s energy for $s > 1$. Due to this variable substitution we, the degree that we need in the sums of squares characterization for modeling the $i = 2$ polynomial constraint, depend on s. The configuration consisting of vertices of the triangular bipyramid has Riesz-s energy

$$\frac{6}{2^{s/2}} + \frac{3}{3^{s/2}} + \frac{1}{4^{s/2}}.$$

We compute our bound for $s = 1, 2, 4$. For $s = 1, 2$ we use $e = 12$ for the $i = 2$ sum of squares characterization and for $s = 4$ we use $e = 16$. In all cases we use $e = 8$ for the $i = 3$ and $i = 4$ sum of squares characterizations. For each of these values of s the solver solves the problem to optimality, and at least the first 7 digits of the solver output agree with the corresponding energy of the vertices of the triangular bipyramid. For $s \in \{3, 5, 6, 7\}$, the solver solves the problems with reduced accuracy, but the bound seems sharp. For higher values of s we cannot draw a conclusion from the solver output.

A next step would be to solve these problems with a high precision solver. However, current high precision solvers such as SDPA-QD and SDPA-GMP [**92**] do not currently perform well if the primal or dual semidefinite program is not strictly feasible. In our problems we have free variables (the a_i variables and the variables in the sum of squares modeling of the $i = 4$ polynomial constraints), and we model each of these as the difference of two nonnegative variables, which implies the primal problems are unbounded and the dual problems are not strictly feasible. It would be very useful to find an approach such as in [**57**] to model these free variables in a way that preserves strict feasibility and that also preserves the block diagonal structure.

In [**23**] a 3-point bound is used to obtain rigorous proofs. Here the floating point output of the semidefinite programming solver is turned into an optimality proof by rounding the matrices to symbolic matrices in such a way they become feasible and sharp. To do the same for our bound we need to have a symbolic (instead of floating point) formulation of the semidefinite program. This means we need to compute the zonal matrices exactly. One approach for this would be to compute the polynomials $p^i_{a,A}$ symbolically (using number fields for the coefficients), and use Gröbner bases to compute the polynomials $q^i_{a,A}$ exactly. Alternatively, it would be

very interesting to derive explicit closed form formulas for the zonal matrices. By formulating these problems as semidefinite programs where both the primal and dual are strictly feasible, it may be possible to solve these with a high precision solver and round the solutions to symbolic solutions which are optimal and feasible as can be checked using the exact formulation of the problem.

Bibliography

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

[2] G. E. Andrews, R. Askey, and R. Roy, *Special Functions*, Cambridge University Press, 1999.

[3] 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.

[4] R. Arens, *Topologies for homeomorphism groups*, Amer. J. Math. **68** (1946), 593–610.

[5] E. Aylward, S. Itani, and P. A. Parrilo, *Explicit SOS decompositions of univariate polynomial matrices and the Kalman-Yakubovich-Popov Lemma*, Proceedings of the 46th IEEE Conference on Decision and Control (2007).

[6] 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`.

[7] C. Bachoc, D. C. Gijswijt, A. Schrijver, and F. Vallentin, *Invariant semidefinite programs*, Handbook on Semidefinite, Conic and Polynomial Optimization (M. F. Anjos and J. B. Lasserre, eds.), Springer, 2002, available at `http://arxiv.org/abs/1007.2905`.

[8] C. Bachoc, G. Nebe, F. M. de Oliveira Filho, and F. Vallentin, *Lower bounds for measurable chromatic numbers*, Geom. Funct. Anal. **19** (2009), no. 3, 645–661.

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

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

[11] C. Berg, J. P. R. Christensen, and P. Ressel, *Harmonic analysis on semigroups: theory of positive definite and related functions*, Springer-Verlag, 1984.

[12] 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`.

[13] V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007.

[14] S. Bochner, *Vorlesungen über Fouriersche Integrale*, Akademische Verlagsgesellschaft, Leipzig (1932).

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

[16] 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.

[17] K. Borsuk and S. Ulam, *On symmetric products of topological spaces*, Bull. Amer. Math. Soc. **37** (1931), no. 12, 875–882.

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

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[20] M. D. Choi, T. Y. Lam, and B. Reznick, *Real zeros of positive semidefinite forms. I*, Math. Z. **171** (1980), no. 1, 1–26.

[21] H. Cohn and N. Elkies, *New upper bounds on sphere packings I*, Ann. of Math. (2) **157** (2003), no. 2, 689–714, DOI 10.4007/annals.2003.157.689.

[22] 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.

[23] 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.
[24] R. Courant and D. Hilbert, *Methods of mathematical physics*, Interscience Publishers (1953).
[25] P. E. B. DeCorte, *The eigenvalue method for extremal problems on infinite vertex-transitive graphs*, PhD Thesis, Delft University of Technology, 2015.
[26] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Res. Rep. Suppl. **10** (1973), vi+97.
[27] P. Delsarte, J. M. Goethals, and J. J. Seidel, *Spherical codes and designs*, Geometriae Dedicata **6** (1977), no. 3, 363–388.
[28] J. Dieudonné, *Sur la séparation des ensembles convexes*, Math. Ann. **163** (1966), 1–3 (French).
[29] A. Florian, *Ausfüllung der Ebene durch Kreise*, Rend. Circ. Mat. Palermo **9** (1960), 300–312.
[30] A. Florian, *Packing of incongruent circles on the sphere*, Monatsh. Math. **133** (2001), 111–129.
[31] A. Florian, *Remarks on my paper: packing of incongruent circles on the sphere*, Monatsh. Math. **152** (2007), 39–43.
[32] A. Florian and A. Heppes, *Packing Circles of Two Different Sizes on the Sphere II*, Period. Math. Hungar. **39** (1999), 125–127.
[33] G. B. Folland, *A course in abstract harmonic analysis*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1995.
[34] L. Föppl, *Stabile Anordnungen von Elektronen im Atom*, J. Reine Angew. Math. **141** (1912), 251–302, DOI 10.1515/crll.1912.141.251 (German).
[35] 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.
[36] 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.
[37] 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.
[38] 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.
[39] D. C. Gijswijt, *Matrix Algebras and Semidefinite Programming Techniques for Codes*, PhD Thesis, University of Amsterdam, 2005.
[40] D. C. Gijswijt, H. D. Mittelmann, and A. Schrijver, *Semidefinite code bounds based on quadruple distances*, IEEE Trans. Inform. Theory **58** (2012), no. 5, 2697–2705, DOI 10.1109/TIT.2012.2184845.
[41] D. Gijswijt, A. Schrijver, and H. Tanaka, *New upper bounds for nonbinary codes based on the Terwilliger algebra and semidefinite programming*, J. Combin. Theory Ser. A **113** (2006), no. 8, 1719–1731, DOI 10.1016/j.jcta.2006.03.010.
[42] H. Groemer, *Existenzsätze für Lagerungen im Euklidischen Raum*, Math. Zeitschr. **81** (1963), 260–278.
[43] M. Grötschel, L. Lovász, and A. Schrijver, *The ellipsoid method and its consequences in combinatorial optimization*, Combinatorica **1** (1981), no. 2, 169–197, DOI 10.1007/BF02579273.
[44] T. C. Hales, *A proof of the Kepler conjecture*, Ann. of Math. (2) **162** (2005), no. 3, 1065–1185, DOI 10.4007/annals.2005.162.1065.
[45] T. Hales, M. Adams, G. Bauer, Dat Tat Dang, J. Harrison, T. Le Hoang, C. Kaliszyk, V. Magron, S. McLaughlin, T. T. Nguyen, T. Q. Nguyen, T. Nipkow, S. Obua, J. Pleso, J. Rute, A. Solovyev, An Hoai Thi Ta, Trung Nam Tran, Diep Thi Trieu, J. Urban, Ky Khac Vu, and R. Zumkeller, *A formal proof of the Kepler conjecture* (2015), available at `http://arxiv.org/abs/1501.02155`.

[46] D. Handel, *Some homotopy properties of spaces of finite subsets of topological spaces*, Houston J. Math. **26** (2000), no. 4, 747–764.
[47] A. Heppes, *Some densest two-size disc packings in the plane*, Discrete Comput. Geom. **30** (2003), no. 2, 241–262, DOI 10.1007/s00454-003-0007-6. U.S.-Hungarian Workshops on Discrete Geometry and Convexity (Budapest, 1999/Auburn, AL, 2000).
[48] A. Heppes and G. Kertész, *Packing circles of two different sizes on the sphere*, Intuitive geometry (Budapest, 1995), Bolyai Soc. Math. Stud., vol. 6, János Bolyai Math. Soc., Budapest, 1997, pp. 357–365.
[49] N. J. Higham, *Upper bounds for the condition number of a triangular matrix*: *Numerical Analysis Report*, 86th ed., University of Manchester, Manchester, 1983.
[50] N. J. Higham, *A survey of condition number estimation for triangular matrices*, SIAM Rev. **29** (1987), no. 4, 575–596, DOI 10.1137/1029112.
[51] A. J. Hoffman, *On eigenvalues and colorings of graphs*, Graph Theory and its Applications (Proc. Advanced Sem., Math. Research Center, Univ. of Wisconsin, Madison, Wis., 1969), Academic Press, New York, 1970, pp. 79–91.
[52] A. B. Hopkins, Y. Jiao, F. H. Stillinger, and S. Torquato, *Phase diagram and structural diversity of the densest binary sphere packings*, Phys. Rev. Lett. **107** (2011).
[53] A. B. Hopkins, F. H. Stillinger, and S. Torquato, *Densest binary sphere packings*, Phys. Rev. E **85** (2012).
[54] G. A. Kabatiansky and V. I. Levenshtein, *On bounds for packings on a sphere and in space*, Probl. Peredachi Inf. **14** (1978), 3–25.
[55] R. M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972), Plenum, New York, 1972, pp. 85–103.
[56] V. L. Klee Jr., *Separation properties of convex cones*, Proc. Amer. Math. Soc. **6** (1955), 313–318.
[57] K Kobayashi, K Nakata, and M Kojima, *A covnersion of an SDP having free variables into the standard form SDP*, Comput. Optim. Appl. **36** (2007), 289–307.
[58] M. Kojima, S. Kim, and H. Waki, *Sparsity in sums of squares of polynomials*, Math. Program. **103** (2005), no. 1, Ser. A, 45–62, DOI 10.1007/s10107-004-0554-3.
[59] H. Kraft and C. Procesi, *Classical invariant theory: a primer*, 1996.
[60] 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.
[61] 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.
[62] E. DeCorte, D. de Laat, and F. Vallentin, *Fourier analysis on finite groups and the Lovász ϑ-number of Cayley graphs*, Exp. Math. **23** (2014), no. 2, 146–152, DOI 10.1080/10586458.2014.882170.
[63] 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.
[64] 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.
[65] M. Laurent, *A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28** (2003), no. 3, 470–496, DOI 10.1287/moor.28.3.470.16391.
[66] M. Laurent, *Strengthened semidefinite programming bounds for codes*, Math. Program. **109** (2007), no. 2-3, Ser. B, 239–261, DOI 10.1007/s10107-006-0030-3.
[67] M. Laurent, *Sums of squares, moment matrices and optimization over polynomials*, Emerging applications of algebraic geometry, IMA Vol. Math. Appl., vol. 149, Springer, New York, 2009, pp. 157–270.
[68] V. I. Levenshtein, *Universal bounds for codes and designs*, Handbook of coding theory, Vol. I, II, North-Holland, Amsterdam, 1998, pp. 499–648.
[69] B. Lindström, *Determinants on semilattices*, Proc. Amer. Math. Soc. **20** (1969), 207–208.

[70] J. Löfberg, *Pre- and post-processing sum-of-squares programs in practice*, IEEE Trans. Automat. Control **54** (2009), no. 5, 1007–1011, DOI 10.1109/TAC.2009.2017144.
[71] L. Lovász, *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory **25** (1979), no. 1, 1–7, DOI 10.1109/TIT.1979.1055985.
[72] L. Lovász and A. Schrijver, *Cones of matrices and set-functions and* 0-1 *optimization*, SIAM J. Optim. **1** (1991), no. 2, 166–190, DOI 10.1137/0801013.
[73] B. Masnick and J. Wolf, *On linear unequal error protection codes*, IEEE Transactions on Information Theory **13** (1967), 600–607.
[74] H. D. Mittelmann and F. Vallentin, *em High accuracy semidefinite programming bounds for kissing numbers*, Experiment. Math. **19** (2010), 174–178.
[75] T. Molien, *Uber die Invarianten der linearen Substitutionsgruppe*, Sitzungsber. Knig. Preuss. Akad. Wiss. (1897), 1152–1156.
[76] 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.
[77] J. Nie, K. Ranestad, and B. Sturmfels, *The algebraic degree of semidefinite programming*, Math. Program. **122** (2010), no. 2, Ser. A, 379–405, DOI 10.1007/s10107-008-0253-6.
[78] 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.
[79] F. Oliveira, *SDPSL: A semidefinite programming specification library*, `http://www.ime.usp.br/fmario/sdpsl.html`.
[80] F. M. de Oliveira Filho and F. Vallentin, *Computing upper bounds for packing densities of congruent copies of a convex body I* (2013), available at `http://arxiv.org/abs/1308.4893`.
[81] 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.
[82] G. Regts, *Upper bounds for ternary constant weight codes from semidefinite programming and representation theory*, Master thesis, University of Amsterdam, 2009.
[83] C. A. Rogers, *The packing of equal spheres*, Proc. London Math. Soc. **8** (1958), 609–620.
[84] The Sage Development Team, *Sage Mathematics Software (Version 4.8)*, 2012.
[85] I. J. Schoenberg, *Positive definite functions on spheres*, Duke Math. J. **9** (1942), 96–108.
[86] A. Schrijver, *A comparison of the Delsarte and Lovász bounds*, IEEE Trans. Inform. Theory **25** (1979), no. 4, 425–429, DOI 10.1109/TIT.1979.1056072.
[87] A. Schrijver, *Theory of linear and integer programming*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, Ltd., Chichester, 1986. A Wiley-Interscience Publication.
[88] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*, Springer-Verlag, 2003.
[89] A. Schrijver, *New code upper bounds from the Terwilliger algebra and semidefinite programming*, IEEE Trans. Inform. Theory **51** (2005), no. 8, 2859–2866, DOI 10.1109/TIT.2005.851748.
[90] 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.
[91] 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.
[92] 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.
[93] G. Szegő, *Orthogonal polynomials*, 4th ed., American Mathematical Society, Providence, R.I., 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.
[94] A. Terras, *Harmonic analysis on symmetric spaces and applications. I*, Springer-Verlag, New York, 1985.
[95] S. Torquato and Y. Jiao, *Dense packings of the Platonic and Archimedean solids*, Nature **460** (2009), 876–879.

[96] S. Torquato, *Random Heterogeneous Materials, Microstructure and macroscopic properties*, Springer-Verlag, New York, 2002.

[97] O. U. Uche, F. H. Stillinger, and S. Torquato, *Concerning maximal packing arrangements of binary disk mixtures*, Physica A **342** (2004), 428–446.

[98] F. Vallentin, *Lecture notes: Semidefinite programs and harmonic analysis* (2008), 31.

[99] H. S. Wilf, *Hadamard determinants, Möbius functions, and the chromatic number of a graph*, Bull. Amer. Math. Soc. **74** (1968), 960–964.

[100] 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.

[101] A. Zaman, *Peter-Weyl Theorem for Compact Groups*, 2011, `http://www.math.toronto.edu/asif/Researchfiles/peter-weyl.pdf`.

Summary

In this thesis we develop techniques for solving problems in extremal geometry. We give an infinite dimensional generalization of moment techniques from polynomial optimization. We use this to construct semidefinite programming hierarchies for approximating optimal packing densities and ground state energies of particle systems. For this we define topological packing graphs as an abstraction for the graphs arising from geometric packing problems, and we prove results concerning convergence and strong duality. We use harmonic analysis to perform symmetry reduction and reduce to a finite dimensional variable space in the optimization problems. For this we explicitly work out the harmonic analysis for kernels on spaces consisting of subsets of another space. We show how sums of squares characterizations from real algebraic geometry can be used to reduce the infinitely many constraints to finitely many semidefinite constraints, where we focus in particular on numerical conditioning and symmetry reduction. We perform explicit computations for concrete problems: We give new bounds for binary spherical cap packings, binary sphere packings, and classical sphere packing problems. This can be used, for instance, to give a simple optimality proof of a binary spherical cap packing. We compute the second step of our hierarchy where the numerical results suggest the bound is sharp for the 5-particle case of the Thomson and related problems. This is the first time a 4-point bound has been computed for a continuous problem.

Samenvatting

In dit proefschrift ontwikkelen we technieken voor het oplossen van problemen in de extremale meetkunde. We geven een oneindigdimensionale generalisatie van de momentwerkwijze uit de polynomiale optimalisatie. Deze gebruiken we om hiërarchieën van semidefiniete programma's te construeren, waarmee we optimale dichtheden van stapelingen en grondtoestanden van deeltjessystemen kunnen berekenen. Hiervoor definiëren we topologische stapelingsgrafen als een abstractie voor de grafen die behoren tot meetkundige stapelingsproblemen, en we bewijzen resultaten aangaande convergentie en sterke dualiteit. We gebruiken harmonische analyse voor symmetriereductie en voor het eindigdimensionaal maken van het toegelaten gebied van de optimalisatieproblemen. Hiervoor werken we de harmonische analyse voor kernen op ruimtes bestaande uit deelverzamelingen van andere ruimtes expliciet uit. We laten zien hoe som-van-kwadraten-technieken uit de reële algebraïsche meetkunde gebruikt kunnen worden om van oneindig veel restricties naar eindig veel semidefiniete restricties te gaan, waarbij we in het bijzonder ingaan op numerieke conditionering en symmetriereductie. We voeren expliciete berekeningen voor concrete problemen uit: We geven nieuwe grenzen voor binaire bolkapstapelingen, binaire bolstapelingen en het klassieke bolstapelingsprobleem. Dit gebruiken we bijvoorbeeld voor het geven van een eenvoudig optimaliteitsbewijs voor een binaire bolkapstapeling. We berekenen de tweede stap in onze hiërarchie, waar de numerieke resultaten suggereren dat dit een scherpe afschatting geeft voor het vijfdeeltjesgeval van het Thomsonprobleem en van gerelateerde problemen. Dit is de eerste keer dat een vierpuntgrens is berekend voor een continu probleem.

Acknowledgments

First of all I would like to thank my advisor, Frank Vallentin, for giving me such an interesting project to work on, for his guidance, and for the many helpful and inspiring conversations we had. I also wish to thank Karen Aardal who has been very supportive in her role as promotor.

I would like to thank my co-authors Evan DeCorte, Fernando Oliveira, and Frank for their collaboration. With the latter two I worked on projects which form an important part of this thesis. It has been great fun working with you!

Thanks to the collegeas in Delft from "the 4th floor" (the Analysis and Optimization groups) for the pleasant working atmosphere and the many interesting discussions and seminars. In particular, I would like to thank Jan Rozendaal, Nikita Moriakov, and Evan for their enthusiasm in our harmonic analysis reading group.

Also thanks to the collegeas in Cologne from the "Optimierung, Geometrie und diskrete Mathematik" group for always being very welcoming on my visits there.

Thanks to the many great people I got to know at conferences and workshops, for sharing their research and for taking an interest in mine.

I would like to thank Henry Cohn for being a great host on my visit to Microsoft Research, and I would like to thank him and the other committee members Dion Gijswijt, Etienne de Klerk, Jan van Neerven, and Achill Schürmann, for taking an interest in my thesis.

Thanks to my family and friends for their support. And a special thanks to Willemieke for making me happy!

Curriculum Vitae

David de Laat was born on 8 February 1987 in Delft, the Netherlands. He completed his secondary education in 2004 at Roelof van Echten College in Hoogeveen. After two years of studying Informatics at Windesheim in Zwolle he started his Mathematics studies at the University of Groningen. In 2009 he obtained his BSc degree (cum laude) and in 2011 he obtained his MSc degree (cum laude) with specialization Algebra and Geometry. In the same year he started his PhD research at Delft University of Technology in the Optimization group.

Publication list

D. de Laat, *Moment methods in energy minimization: New bounds for Riesz minimal energy problems*, in preparation.

D. de Laat, *Optimal polydisperse packing densities using objects with large size ratio*, in preparation.

D. de Laat, F. Vallentin, *A semidefinite programming hierarchy for packing problems in discrete geometry*, Math. Program., Ser. B 151 (2015), 529-553. [**61**]

D. de Laat, F.M. de Oliveira Filho, F. Vallentin, *Upper bounds for packings of spheres of several radii*, Forum Math. Sigma 2 (2014), e23 (42 pages). [**60**]

P.E.B. DeCorte, D. de Laat, F. Vallentin, *Fourier analysis on finite groups and the Lovász ϑ-number of Cayley graphs*, Experiment. Math. 23 (2014), 146-152. [**62**]