SIAG/OPT Views and News

Contents

Are you receiving this by postal mail?
Do you prefer electronic delivery?

Email siagoptnews@lists.mcs.anl.gov!

Articles

Exascale Challenges for Optimization

Andreas Grothey
School of Mathematics
University of Edinburgh
Edinburgh, EH9 3FD, UK
A.Grothey@ed.ac.uk
http://www.maths.ed.ac.uk/~agr

Supercomputers have reached sustained petaflop performance a few years ago, and machines capable of exascale computing are predicted to arrive around 2020. These machines will provide massive concurrency on an unprecedented scale. While the exact architecture of such machines is still uncertain, they likely will feature on the order of 10^9 parallel threads in some way or other. Our current optimization algorithms are by and large ill-suited to take advantage of this degree of concurrency. Consequently high-performance computing has not made as big an impact on optimization as in other areas of applied mathematics. Here we review why this is the case and what must be done in order to change this state of affairs.

1 Introduction

State-of-the-art supercomputers are now reaching a performance of tens of petaflops, and it is predicted that machines capable of exaflop performance will be with us by about 2020. On the other hand there is a fear that few, if any, of current supercomputer applications across all areas are actually capable of exploiting such a computing resource efficiently. The major bottleneck lies in the software, rather than the hardware, although some hardware restrictions (such as lack of both fast communication and fast memory access) are adding to the difficulty.

As a response to the identified gap between hardware and software capabilities, various efforts such as the International Exascale Software Project (IESP [4]) and its European counterpart, the European Exascale Software Initiative (EESI [3]) were set up in 2009/2010 with a view to qualitatively assess the gap in software capabilities, identify promising routes to achieve exascale performance, and quantify the amount of resources (mainly in manpower for coding and research) needed to close the gap in time for the arrival of the first exascale machines. A more recent effort is the Exascale Mathematics Working Group charged with a similar remit in 2013 [2].

The author has been vice-chair of the working group on "numerical libraries, solvers and algorithms" as part of EESI

Performance Development
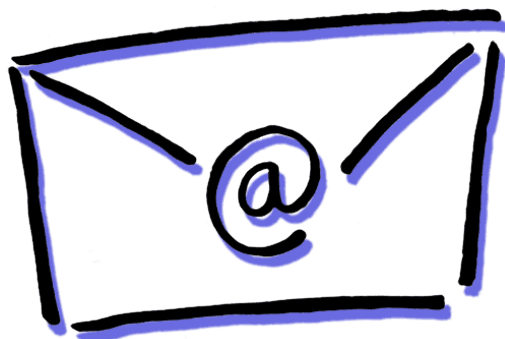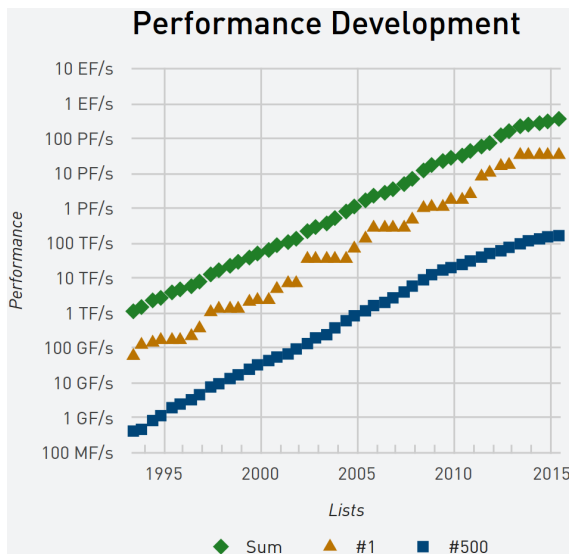
Figure 1: Performance of top500 list computers *(Image generated by* http://www.top500.org *on 11/11/15).*

and chair of the same working group in the follow-up EESI-2 over the past five years. This report presents the findings of the working group regarding the barriers to exaflop (and even petaflop) scalability for optimization algorithms and outlines possible future avenues to overcome these barriers.

2 Background and Setting

The reasons for exascale (and by implication massively parallel) computing are simple: Applications (that are deemed critical), such as seismic simulations, engineering (e.g., aircraft design), health, and military, always will require increasingly powerful computing resources. While Moore's law has reached its end in terms of pure processor speed, in terms of whole system performance we are still seeing an exponential increase (over the past 20 years a doubling of the R_{max} of the fastest system on the top500 list about every 15 months; see Figure 1). This increase is achieved almost solely by an increase in the number of concurrent threads. A machine with exascale performance ($R_{max} \approx 10^{18}$ flops) is expected to arrive around 2020 and is predicted to have around 10^9 parallel processes. Even optimistic estimates put the power consumption of such a machine at not less than 20 MW. That would translate into an annual electricity bill of $20 million. Clearly this machine is not a toy: it is designed to do real science in areas where exascale computing is likely to have the most impact.

EESI had been tasked to report on the current state of the art in applications of supercomputing and to identify gaps in the relevant areas that prevent achieving exascale performance by 2020. Internally the EESI remit has been split into several work packages of which the most relevant from an optimizer's point of view are *WP3: Application Grand Challenges* and *WP4: Enabling Technologies.* WP3 has identified applications for which (from an socio-economic viewpoint) achieving exascale performance is most critical. They

encompass a wide range of areas such as industrial and engineering applications, weather, climatology, earth sciences, fundamental sciences (chemistry, physics), life science, and health. WP4, on the other hand, is concerned with all levels of the software stack used by codes in these applications. It is subdivided into "hardware," "software eco-systems," "numerical libraries, solvers and algorithms," and "scientific software engineering." Obviously a strong interaction exists between these work packages: in particular, the applications should set the priorities for the enabling technologies.

The relevant aspect for an optimization audience is the work of the working group on "numerical libraries", which includes optimization in its remit. This working group has been made up of 15 experts, who cover all areas of numerical algorithms deemed critical—in particular, dense linear algebra, graph and hypergraph partitioning, sparse direct methods, iterative methods for sparse matrices, eigenvalue problems, model reduction, optimization, control of complex systems, and structured and unstructured grids.

It is notable, that few, if any, exascale-critical applications actually make use of optimization. The closest are probably inverse problems, where a simple parameter optimization (conjugate gradients or even gradient descent) is used over a complex, whole-machine simulation to optimize a (smallish) parameter set. Why this lack of use of optimization? The author can think of the following answers: First, the critical applications are to some extent self-selected. These are applications for which exascale computing is expected to make the most impact. Of course that means these applications are deemed important for socioeconomic reasons, but it also means that these are applications that have been shown to scale well on previous generations of supercomputers. Second, optimization algorithms generally don't scale well, at least not when compared with other algorithms in applied mathematics such as simulations or domain decomposition for PDEs. Our fundamental algorithms, such as the simplex method, are inherently serial; and even in problem classes that are prime candidates for parallelism, such as stochastic programming, a relatively tight interaction exists between different parts of the problem. Optimization simply doesn't scale enough for "cover-story" applications, and that's what we are talking about at the moment.

That is not, however, a reason to stop thinking about exascale. Optimization may not be at the forefront of algorithms that are heavily used in the first applications to run on exascale machines. But exascale machines will be here to stay. Today's flagship systems will be commodity hardware in 20–30 years. By that point, and likely much earlier, optimization algorithms will have to scale to exascale level.

3 Exascale Challenges

An exascale machine will pose some unique challenges in terms of programmability and scalability of algorithms. The massive task concurrency alone will pose a major challenge. While pure computing power is still increasing, memory availability and speed will lag behind. We are therefore looking at a significant decrease in available memory per node

and memory latency relative to calculation speed. The large power requirements make it important that whole-machine calculations be as efficient as possible. Even so, the power consumption target of 20 MW is extremely ambitious and will be achievable only through restrictions on the amount of memory accesses and data movements. Moreover, with increasing numbers of sockets on the system, hardware failures become more frequent (mean time to failure anywhere in the system significantly below 1 day), resulting in frequent need for checkpointing and restarting of some nodes, thus impacting load balancing.

3.1 Are we ready?

The immediate answer must be "no"! It is hard to see how any of the current optimization algorithms could scale to 10^9 parallel threads. On the other hand, in order to achieve application exascale performance, optimization is likely not going to be used on its own on the full machine. Often, good scalability of the underlying linear algebra will be inherited by the optimization routine, and algorithms that can use such scalable routines will be at an advantage.

More likely, an optimization algorithm will be a (small) part somewhere in a large hierarchy of algorithms with optimization used at the top level (optimizing a parameter set over a whole-machine simulation), at the bottom level (a relatively small optimization problem solved many times with slightly different parameters, probably concurrently, e.g., for ensemble runs), or somewhere in between. Therefore making algorithms fit for exascale performance means not only designing algorithms that scale to 10^9 threads but also reducing the flops/memory ratio and improving efficiency for many simultaneous runs (preferably without much communication). Keeping these points in mind, what is the current state of affairs regarding exascale readiness?

Linear programming is arguably the most important class of optimization problems. It is important both in its own right and as a subproblem in integer programming or nonlinear programming. Of the two competing algorithm classes, the simplex algorithm is notoriously difficult to parallelise, because of its inherently serial task dependency graph. Some progress can be made by exploiting parallelism in the linear algebra for special problem classes [7]. Interior point methods (IPMs) fare much better in this respect but are inefficient when used as subproblem solvers because of their ineffective warmstarts. Also, at least when using direct factorizations, IPMs with direct factorizations are memory hungry, which will become more of a problem. Dramatic breakthroughs are not likely for either of these limitations, although iterative solvers for IPM are starting to mature and will improve their memory requirements. To be able to solve a series of very large LP problems efficiently on a massively parallel hardware, we will likely need to look to new (hybrid?) algorithms.

Integer programming seems an obvious candidate for exascale computing. Often, realistic problems take too long to solve or need too much memory, so throwing vast computing resources at them is appealing. Will it work? Koch

et al. [5] point out that a (M)IP can be intractable for many reasons. Problems with inherently weak formulations may well result in search trees so large that their complete processing defeats all reasonable computing resources. Problems with large node problems and relatively small search trees are relying on the availability of warmstartable and parallelisable LP solvers (see above). Problems with huge search trees have more potential, but here are also problems with ramping (having not enough nodes early on), load balancing (transferring nodes to other processors dynamically is expensive), and the possible increase in the total number of nodes on the tree if processed in parallel because of delayed propagation of bounds. These issues need to be addressed.

Decomposition is an appealing strategy for certain classes of problems (notably, stochastic programming) and one that usually leads to scalable algorithms. However, the need to coordinate subproblems results in synchronization points that, in view of less-than-perfect load balancing (as is inevitable for more than a few hundred parallel tasks), hinders scalability.

In summary the main issues that prevent scalability of current algorithms to the required level for exascale are synchronization points, essentially linear task-dependency graphs and algorithms that are already memory-bound.

3.2 Strategies for exascale performance

So what can be done? A few general strategies are promising and in some cases are successfully used in other areas of applied mathematics:

Synchronization avoidance: Almost all our current optimization algorithms that do parallelise well do so by decomposition, and that almost invariably means synchronization when the master problem is solved. For a thousand concurrent tasks one may just hope that subproblems take an equal amount of time in order to achieve good load balancing. For machines with many more processors scalability is clearly bounded. Further, decomposition algorithms that require the solution of (M)ILP problems such as branch and price already have to deal with subproblem solves that can take vastly different amounts of time. In short, scalable decomposition algorithms clearly cannot impose barriers in the form of synchronization points. Some work has been done on asynchronous decomposition (e.g., Benders decomposition applied to stochastic linear programming [6]). However, much work remains to be done. One possible approach would be a totally asynchronous algorithm where each processor is responsible for one subproblem that it solves continuously, always taking the latest available data from other processors as input and always sending out the latest results asynchronously to the other processors to be used if and when they see fit. Such an algorithm may work with or without a coordinating master problem (which again would have to do its work without synchronising with other processors). Care would have to be taken to ensure convergence and stability of such an algorithm. A series of papers by Bertsekas [1] from the early 1990s gives a starting point for possible convergence analysis of such totally asynchronous algorithms.

Avoidance of memory accesses: Already on current machines memory fetches (especially uncached ones) can be orders of magnitude slower than computations. This situation is going to be even more pronounced on future machines. In addition, memory fetches are usually power hungry, and machines are likely to operate on some restricted power budget (energy-aware algorithms). Consequently, algorithms should aim for a much higher flops per byte ratio. Investigation of higher-order algorithms seems to be an obvious candidate.

Latency/communication hiding: Where memory accesses cannot be avoided, the next best strategy is to try to do useful work on data already available while waiting for the next set of data to arrive. Where data is required from other processors, nonblocking MPI communications can be used. Where data is required from RAM, however, much less support is available for coding such a strategy.

Blocking and hierarchical algorithms: In order to achieve a high flops per byte ratio and to avoid memory fetches, a good strategy is to partition the problem at hand into (data) blocks that can be worked on in situ. This is akin to the strategy that makes the use of BLAS level 3 routines so efficient. Extending this strategy will lead to hierarchical algorithms. We note, however, that use of this strategy usually requires (possibly expensive) restructuring or repartitioning of the problem data.

3.3 Beyond scalability: reliability, reproducibility, floating-point issues

Exascale challenges go beyond just achieving scalability of algorithms. In order to make maximal use of the scarce resources on such a system, some wider issues need to be addressed.

The tradeoff between solution accuracy and usage of scarce system resources (computation time, memory) is going to become more critical. Mixed-precision floating-point arithmetic will be relevant. Calculations in single precision are appealing since they reduce power- and cycle-hungry memory accesses. On the other hand, a whole-machine operation on 10^9 processors may well require quad precision to prevent the result from being flooded with floating-point errors. We will need to consider at every part of the calculations what precision is necessary. Too much will hurt performance; too little will hurt accuracy.

The use of asynchronous algorithms will introduce reproducibility into the equation. Reproducibility will be costly to achieve (if not impossible) for asynchronous algorithms. So, how important is it really that results are reproducible? A better paradigm might be that the optimization solver guarantees to always find a solution within the specified backward error bounds of the optimal solution (and does so reproducibly). What error bounds are acceptable for a given usage/application would need careful assessment.

On a machine with 10^9 processors some part inevitably will fail or even produce wrong results almost hourly. Obvious strategies to mitigate this situation are frequent node-by-node checkpointing and redundant calculations. Intuitively,

we feel that we can do better: many optimization algorithms (by being iterative) are robust enough to be able to cope with some errors or even partially missing results. Asynchronicity again would prevent delaying the whole application by downtime on a single compute node.

3.4 Training and software maintenance

Ultimately the aim (at least of efforts like IESP and EESI) are not (just) scalable algorithms but reliable and robust libraries that can and will be used by anyone who needs them. A mechanism is needed by which long-term funding for the development and maintenance of state-of-the-art scalable optimization libraries can be obtained. These libraries will be complex, so training opportunities for their users will also be needed. Clearly, in order for people to put in the effort to learn how to use such (likely complex) software packages, they must be confident that the packages will be actively supported for a reasonably long time.

Currently it is difficult (at least in Europe) to secure funding for the provision of robust, well-documented and user-friendly libraries.

4 The Future

Both IESP and EESI/EESI-2 have now finished. Follow-up projects to monitor the progress toward exascale readiness are under way (in the form of the workshop series on Big Data and Extreme Scale Computing, "BDEC") or in the application stages (in the form of the European eXtreme Data and Computing Initiative, "EXDCI"). The European Commissions Horizon 2020 program has included a call *"Towards Exascale High Performance Computing"* originating directly from the EESI recommendations. More funding calls are expected.

Currently Europe seems committed to the path to exascale readiness and many opportunities are available to be exploited.

REFERENCES

[1] D. P. Bertsekas and J. N. Tsitsiklis, Some aspects of parallel and distributed iterative algorithms – a survey, *Automatica*, 27 (1991), pp. 3–21.

[2] J. Dongarra et al., Applied Mathematics Research for Exascale Computing, (2014).

[3] European Exascale Software Initiative, http://www.eesi-project.eu.

[4] International Exascale Software Project, http://www.exascale.org.

[5] T. Koch, T. Ralphs, and Y. Shinano, Could we use a million cores to solve an integer program?, *Math. Method. Oper. Res.*, 76 (2012), pp. 67–93.

[6] J. Linderoth and S. J. Wright, Decomposition algorithms for stochastic programming on a computational grid, *Comput. Optim. Appl.*, 24 (2003), pp. 207–250.

[7] M. Lubin, J. A. J. Hall, C. G. Petra, and M. Anitescu, Parallel distributed-memory simplex for large-scale stochastic LP problems, *Comput. Optim. Appl.*, 55/3 (2013), pp. 571–596.

Mathematical Optimization for Packing Problems

Fernando Mário de Oliveira Filho

Instituto de Matemática e Estatística
Universidade de São Paulo
Rua do Matão 1010
05508-090 São Paulo/SP, Brazil
fmario@gmail.com
http://www.ime.usp.br/~fmario

Frank Vallentin

Mathematisches Institut
Universität zu Köln
Weyertal 86–90, 50931 Köln, Germany
frank.vallentin@uni-koeln.de
http://www.mi.uni-koeln.de/opt/
frank-vallentin

1 Introduction

During the past few years several new results on packing problems were obtained by using a blend of tools from semidefinite optimization, polynomial optimization, and harmonic analysis. Schrijver [19] used semidefinite optimization and the Terwilliger algebra to obtain new upper bounds for binary codes, Bachoc and Vallentin [2] used semidefinite optimization and spherical harmonics for spherical codes, and Cohn and Elkies [5] used linear optimization and Fourier analysis for sphere packings leading to the breakthrough result of Cohn and Kumar [6], who proved that the Leech lattice in dimension 24 gives the best lattice sphere packing in its dimension. De Laat, Oliveira, and Vallentin [13] generalized the approach of Cohn and Elkies to provide upper bounds for maximal densities of packings of spheres having different radii. The most recent extension was by Oliveira and Vallentin [17], providing new upper bounds for the density of packings of congruent copies of a given convex body.

Typical in all this work is the use of semidefinite optimization and harmonic analysis, which gives newcomers to the field—often overwhelmed with technical details—a hard time. Also typical is that the computational challenge grows dramatically if one goes from compact spaces, like binary Hamming space or the sphere, to noncompact spaces like the Euclidean space.

Our goal in this paper is to provide an introduction to this topic in an attempt to paint the big picture without losing essential detail. This paper is, however, not meant as a survey of results about geometric packing problems—that task would easily fill books! For a first orientation we refer the interested reader to the classical book by Conway and Sloane [7].

2 Some History

The *sphere packing problem* asks, How much of three-dimensional space can be filled with pairwise nonoverlapping translates of unit spheres? This was considered by Johannes Kepler (1571–1630) in his work *Strena seu de Nive Sexangula* (On the Six-Cornered Snowflake) from 1611, which was his New Year's gift to his friend and supporter Johann Matthäus Wacker von Wackenfels (1550–1619). He explains the formation of snowflakes into crystals having sixfold symmetry by drawing an analogy to dense sphere packings that possess the same kind of symmetry. The general acceptance of atomism was yet to come, so this explanation was a remarkable achievement. Kepler's work is the first scientific writing about crystal formation; in it he claims (essentially without any justification) that a specific periodic structure, the face-centered cubic lattice, describes the densest sphere packing having density $\pi/\sqrt{18} = 0.74\ldots$. This claim is now called Kepler's conjecture.

In 1998 Thomas Hales proved Kepler's conjecture. His proof makes heavy use of computers; and in 2009 he, together with his student Samuel P. Ferguson, was rewarded the Fulkerson Prize for his work.

The sphere packing problem, and more generally the problem of packing copies of a given body, was also considered by David Hilbert. He mentioned it as part of his 18th problem:

> 18. Building up of Space from Congruent Polyhedra
>
> ...I point out the following question, related to the preceding one, and important to number theory and perhaps sometimes useful to physics and chemistry: How can one arrange most densely in space an infinite number of equal solids of given form, e.g., spheres with given radii or regular tetrahedra with given edges (or in prescribed position), that is, how can one so fit them together that the ratio of the filled to the unfilled space may be as great as possible?

The problem of packing congruent copies of regular tetrahedra, mentioned by Hilbert, goes back to Aristotle's (384–322 BC) refutation of a theory of Plato (428–348 BC), presented in the *Timaeus*, that claimed that each of the four elements had a specific shape, namely, one of the Platonic solids, and that the properties of each element derived from its shape. So, for instance, earth, the most stable and plastic element, is cubic in shape, and fire, the most acute and most penetrating element, has the shape of a tetrahedron.

Aristotle presents several arguments against this theory in his treatise *De caelo*. In one of his arguments (see *De caelo*, Book III, Chapter VIII), he claims that it is irrational to assign geometrical shapes to the four elements, since not all of space can be thus filled. Indeed, says Aristotle, only the cube and the pyramid (i.e., the regular tetrahedron) can fill space. Thus, in order to refute Plato's theory, Aristotle's argument uses the idea of the impossibility of a vacuum, together with the fact that only two of the solids (corresponding to earth and fire) can fill the whole of space.

Aristotle's claim that one can tile space with tetrahedra was picked up by many of his commentators. Simplicius of Cilicia (c. 490–c. 560), one of the main commentators of

Aristotle in late antiquity, even stated that, as eight cubes are sufficient to fill the space around a given point, so are twelve regular tetrahedra (see page 42 in the translation by Mueller [20]).

In the Middle Ages, Aristotle's Arabic commentator, Averroës (1126–1198), restated the claim that twelve pyramids fill the space around a point, and he gives an argument for it. Three planes meet at the vertex of a cube, forming a so-called "solid angle" composed of three right angles. Eight cubes fill the space around a point in three-dimensional space, and these eight solid angles add up to a total of 8×3 right angles. Now, a solid angle of a tetrahedron is composed of three angles of $60°$ each, totaling two right angles. Since one needs 8×3 right angles to fill the space around a point (as can be seen from the cubes) and since $8 \times 3 = 12 \times 2$, it follows that twelve tetrahedra fill the space around a point.

Averroës' commentary introduced the problem to medieval schoolmen. Roger Bacon (c. 1214–1294) defended Averroës' position against the claim that not twelve, but twenty, tetrahedra are needed to fill the space around a point. Thomas Bradwardine (c. 1290–1349) disproved Averroës' claim with a simple argument: If indeed one could place twelve regular tetrahedra around a point in such a way that no empty space results, then in addition to the five Platonic solids, one would have another one, which is impossible. According to him, those who argue that twenty tetrahedra can be placed around a point have therefore a stronger position, since one can obtain twenty pyramids by joining the bases of a regular icosahedron to its center. Bradwardine observes that one still must check whether the pyramids so obtained are regular or not, but he leaves the question open (these pyramids are, as can be shown using the construction of the icosahedron given in the thirteenth book of Euclid's Elements, not regular).

The question was finally settled, it is believed, by Johannes Müller von Königsberg (1436–1476), known as Regiomontanus, who proved that it is impossible to tile space with regular tetrahedra. Of Regiomontanus's manuscript only the title, describing the contents of the work, has been preserved; but there is no doubt he had all the tools at his disposal to settle the problem. Francesco Maurolico (1494–1575) computed the angle between two faces of a regular tetrahedron. This angle, equal to $\arccos(1/3) \approx 70.52877°$, is greater than $60°$ and smaller than $72°$. Hence it follows that one cannot tile space with tetrahedra. Maurolico's work was recently rediscovered (see Addabbo [1]). For more on the fascinating history of the tetrahedra packing problem, including all the details presented here, see the historical survey by Struik [21] and the survey by Lagarias and Zong [14].

If one cannot tile space with regular tetrahedra, how much of space can be filled with them? Even today, the problem is far from being solved. In 2006, Conway and Torquato [8] found surprisingly dense packings of tetrahedra. This finding sparked renewed interest in the problem and a race for the best construction (see Lagarias and Zong [14] and Ziegler [23]). The current record is held by Chen, Engel, and Glotzer [4], who found in 2010 a packing with density

≈ 0.8563, a much larger fraction of space than that which can be covered by spheres. This prompted the quest for upper bounds; the current record rests with Gravel, Elser, and Kallus [10], who proved an upper bound of $1 - 2.6\ldots \cdot 10^{-25}$. They are themselves convinced that the bound can be greatly improved:

> In fact, we conjecture that the optimal packing density corresponds to a value of δ [the fraction of empty space] many orders of magnitude larger than the one presented here. We propose as a challenge the task of finding an upper bound with a significantly larger value of δ (e.g., $\delta > 0.01$) and the development of practical computational methods for establishing informative upper bounds.

3 Mathematical Modeling

How can one model mathematically the problem of packing spheres or regular tetrahedra in $\mathbb{R}^3$? Packing problems are optimization problems and can be seen as infinite analogues of a well-known problem in combinatorial optimization, namely, the problem of finding a maximum-weight independent set in a graph. To see this, let us first consider two kinds of packing problems.

Problem 1 (Translational body packings). *Given convex bodies $\mathcal{K}_1, \ldots, \mathcal{K}_N \subseteq \mathbb{R}^n$, how much of $\mathbb{R}^n$ can be filled with pairwise nonoverlapping translated copies of $\mathcal{K}_1, \ldots, \mathcal{K}_N$?*

The sphere packing problem is then obtained by taking $N = 1$ and letting $\mathcal{K}_1$ be the unit ball.

Problem 2 (Congruent body packings). *Given a convex body $\mathcal{K} \subseteq \mathbb{R}^n$, how much of $\mathbb{R}^n$ can be filled with pairwise nonoverlapping congruent (i.e., translated and rotated) copies of $\mathcal{K}$?*

Here letting $\mathcal{K}$ be the unit ball gives the sphere packing problem, and letting $\mathcal{K}$ be the regular tetrahedron gives the tetrahedra packing problem. In a sense, Problem 2 is a limiting case of Problem 1: Given a convex body $\mathcal{K}$ one tries to pack translative copies of infinitely many rotations $A\mathcal{K}$ of $\mathcal{K}$, where $A \in \mathrm{SO}(n)$ and $\mathrm{SO}(n)$ is the special orthogonal group of $\mathbb{R}^n$ (i.e., the group of all orthogonal $n \times n$ matrices with determinant 1).

We call a union of nonoverlapping (congruent or translated) copies of some bodies a *packing* of these bodies. In a packing bodies are allowed to touch on their boundaries but not to intersect in their interiors. The fraction of space covered by a packing is its *density*, so our goal is to find the maximum density of packings. Here we are using an informal definition of density; in Section 6 we will give a precise definition.

Let $G = (V, E)$ be a graph, finite or infinite. An *independent set* is a set $I \subseteq V$ that does not contain adjacent vertices. Packings of bodies can be seen as independent sets in some specially defined graphs called packing graphs. Given convex bodies $\mathcal{K}_1, \ldots, \mathcal{K}_N \subseteq \mathbb{R}^n$, the *translational packing graph* of $\mathcal{K}_1, \ldots, \mathcal{K}_N$ is the graph G whose vertex set is

$\{1, \ldots, N\} \times \mathbb{R}^n$. The vertices of G correspond to possible choices of bodies in the packing: vertex (i, x) corresponds to placing the body $x + \mathcal{K}_i$ in the packing. This interpretation defines the adjacency relation of G: vertices (i, x) and (j, y) are adjacent if the corresponding bodies overlap, that is, if

$$(x + \mathcal{K}_i)^\circ \cap (y + \mathcal{K}_j)^\circ \neq \emptyset,$$

where A° is the interior of set A. Hence, independent sets of G correspond to packings of translated copies of $\mathcal{K}_1, \ldots, \mathcal{K}_N$ and vice versa.

A similar idea can be used regarding packings of congruent copies of a given convex body $\mathcal{K}$. Given such a body, we consider its *congruent packing graph*, which is the graph G whose vertex set is $\mathrm{SO}(n) \times \mathbb{R}^n$. The elements of $\mathrm{SO}(n)$ correspond to the possible rotations of $\mathcal{K}$, so that a vertex (A, x) of G corresponds to placing the body $x + A\mathcal{K}$ in the packing. Again, this gives the adjacency relation of G: vertices (A, x) and (B, y) are adjacent if

$$(x + A\mathcal{K})^\circ \cap (y + B\mathcal{K})^\circ \neq \emptyset.$$

With this, independent sets of G correspond to packings of congruent copies of $\mathcal{K}$ and vice versa.

Packings therefore correspond to independent sets of the packing graph. If we measure the weight of an independent set by the density of the associated packing, then Problems 1 and 2 ask us to find maximum-weight independent sets in the corresponding packing graphs.

Does this modeling help? Finding a maximum cardinality independent set in a finite graph is a well-known NP-hard problem, figuring in Karp's list of 21 problems. Many techniques have been developed in combinatorial optimization to deal with hard problems: the basic approach is that one tries to develop efficient methods to find lower and upper bounds. In the case of the maximum-cardinality independent set problem, lower bounds are constructive and come from heuristics that try to find independent sets of large size. Analogously, for packing problems one has the adaptive shrinking cell scheme of Torquato and Jiao [22], which can successfully generate dense packings.

As for upper bounds, Lovász [16] introduced a graph parameter, the theta number, that provides an upper bound for the maximum cardinality of independent sets of a finite graph; Lovász's theta number can be computed efficiently by using semidefinite optimization. The most successful approaches to obtaining upper bounds for the maximum densities of packings all use extensions of the theta number. The theta number can be extended naturally to graphs having compact vertex sets, as we show in Section 4; still, this extension cannot be applied to the packing graphs we described above, because they have noncompact vertex sets. These graphs can be compactified, however, as we discuss in Section 6, and then the extension of the theta number can be applied.

4 The Lovász Theta Number and an Extension

The *independence number* of a graph $G = (V, E)$ (finite or infinite) is the graph parameter

$$\alpha(G) = \max\{ |I| : I \text{ is independent } \}.$$

Given a nonnegative weight function $w \colon V \to \mathbb{R}_+$, one may also define the *weighted independence number* of G as

$$\alpha_w(G) = \max\{ w(I) : I \text{ is independent } \},$$

where $w(I) = \sum_{x \in I} w(x)$. Weights will be useful in packing problems because, when we want to pack different kinds of bodies, such as spheres having different radii, the weight function allows us to distinguish between big and small bodies.

The theta number introduced by Lovász [16] provides an upper bound to the independence number of a graph. It was later strengthened and extended to the weighted case by Grötschel, Lovász, and Schrijver [11]. Many equivalent ways of defining their graph parameter exist; the one most convenient for us is the following. Given a finite graph $G = (V, E)$ and a weight function $w \colon V \to \mathbb{R}_+$, we define

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

where $w^{1/2} \in \mathbb{R}^V$ is such that $w^{1/2}(x) = w(x)^{1/2}$.

Theorem 1. *Let $G = (V, E)$ be a finite graph and $w \colon V \to \mathbb{R}_+$ be a weight function. Then $\alpha_w(G) \leq \vartheta_w'(G)$.*

Proof. Let $I \subseteq V$ be an independent set such that $w(I) > 0$ (if there is no such independent set, then $\alpha_w(G) = 0$, and the theorem follows trivially) and let M and K be a feasible solution of (1).

Consider the sum

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

This sum is at least

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

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

The same sum is also at most

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

because $K(x, x) \leq M$ and $K(x, y) \leq 0$ for distinct $x, y \in I$. Combining both inequalities proves the theorem. $\qquad\square$

Notice that the formulation we use for $\vartheta'_w(G)$ is a *dual formulation*, since any feasible solution gives an upper bound for the independence number.

Thus, $\vartheta'_w(G)$ provides an upper bound for $\alpha_w(G)$ when G is finite. When more generally V is a separable and compact measure space satisfying a mild technical condition, graph parameter ϑ'_w can be extended in a natural way to provide an upper bound for the weighted independence number.

This extension relies on a basic notion from functional analysis, that of kernel. Let V be a separable and compact topological space and μ be a finite Borel measure over V. A kernel is a complex-valued function $K \in L^2(V \times V)$.

A kernel K can be seen as a generalization of a matrix. Like a matrix, a kernel defines an operator on $L^2(V)$ by

$$(Kf)(x) = \int_V K(x,y) f(y) \, d\mu(y).$$

Kernel K is *Hermitian* if $K(x,y) = \overline{K(y,x)}$ for all x, $y \in V$. Hermitian kernels are the analogues of Hermitian matrices and an analogue of the spectral decomposition theorem, known as the Hilbert-Schmidt theorem, holds, as we describe now.

A function $f \in L^2(V)$, $f \neq 0$, is an *eigenfunction* of K if $Kf = \lambda f$ for some number λ, which is the *associated eigenvalue* of f. We say λ is an eigenvalue of K if it is the associated eigenvalue of some eigenfunction of K. The Hilbert-Schmidt theorem states that for a Hermitian kernel K, there is a complete orthonormal system $\varphi_1, \varphi_2, \dots$ of $L^2(V)$ consisting of eigenfunctions of K such that

$$K(x,y) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(x) \overline{\varphi_i(y)}$$

with L^2 convergence, where the real number λ_i is the associated eigenvalue of φ_i. Then the λ_i with their multiplicities are all the eigenvalues of K.

A Hermitian kernel K is *positive* if all its eigenvalues are nonnegative; this is the analogue of a positive semidefinite matrix. An equivalent definition is as follows: K is positive if for every $\rho \in L^2(V)$ we have

$$\int_V \int_V K(x,y) \rho(x) \overline{\rho(y)} \, d\mu(x) d\mu(y) \geq 0.$$

Using kernels, one may extend the definition of ϑ'_w to graphs defined over separable and compact measure spaces, simply by replacing the matrices in (1) by continuous kernels. In other words we define

$$
\begin{aligned}
\vartheta'_w(G) = \quad & \inf M \\
\text{s.t.} \quad & K(x,x) \leq M \quad \forall x \in V, \\
& K(x,y) \leq 0 \quad \forall \{x,y\} \notin E \text{ with } x \neq y, \quad (2)\\
& K \text{ is continuous \& Hermitian}, \\
& K - W \text{ is positive},
\end{aligned}
$$

where $K \colon V \times V \to \mathbb{R}$ and $W \in L^2(V \times V)$ is the kernel such that $W(x,y) = w(x)^{1/2} w(y)^{1/2}$.

One then has the following theorem.

Theorem 2. *Let $G = (V, E)$ be a graph where V is a separable and compact measure space in which any open set has nonzero measure. Let $w \colon V \to \mathbb{R}_+$ be a continuous weight function. Then $\alpha_w(G) \leq \vartheta'_w(G)$.*

Proof. Since V is compact and since we assume that every open subset of V has nonzero measure, we may use the following observation of Bochner [3]: a continuous kernel K is positive if and only if for any choice of N and points $x_1, \dots, x_N \in V$ we have that the matrix $\big(K(x_i, x_j)\big)_{i,j=1}^N$ is positive semidefinite.

Using this characterization of continuous and positive kernels, we may mimic the proof of Theorem 1 and obtain the desired result. This is why we require K to be continuous in the definition of ϑ'_w and also why we require w to be a continuous function: because we want to apply Bochner's characterization to $K - W$.   □

As was the case with Theorem 1, any feasible solution of (2) gives an upper bound for the weighted independence number. This is useful in the infinite setting because then it is often harder to obtain optimal solutions. Notice that it might also be that $\alpha_w(G) = \infty$. In this case, the theorem still holds, since (2) will be infeasible, and therefore $\vartheta'_w(G) = \infty$.

5 Exploiting Symmetry with Harmonic Analysis

If G is a finite graph, then computing $\vartheta'_w(G)$ is solving a semidefinite program whose value can be found with the help of a computer; that is, it can be approximated up to arbitrary precision in polynomial time. This is a theoretical assertion however; in practice, for moderately big graphs (say, with thousands of vertices), if one cannot exploit any special structure of the graph, then computing the theta number is often impossible with today's methods and computers.

If the graph G is infinite, we are dealing with an infinite-dimensional semidefinite program. If one then desires to use computational optimization methods, at some point the transition from infinite to finite has to be made. One way to make this transition is to use finer and finer grids to discretize the infinite graph and solve the corresponding finite semidefinite programs, obtaining bounds for the infinite problem. For coarse grids, however, this approach performs poorly, and for fine grids it soon becomes computationally infeasible. Moreover, with this approach one loses the entire geometrical structure of the packing graphs.

The alternative is to use harmonic analysis. Instead of computing ϑ'_w in the "time domain," we could formulate the optimization problem in the "Fourier domain." This has a twofold advantage. First, the Fourier domain can be discretized essentially by truncation and in doing so we do not lose too much, since we expect that most of the information in a well-structured problem (like a packing problem) is to be concentrated in the beginning of the spectrum. Second, the translation group $\mathbb{R}^n$ acts on the translational packing graph, and the group of Euclidean motions $SO(n) \ltimes \mathbb{R}^n$ acts on the congruent packing graph; using harmonic analysis we

can exploit the symmetry of this situation. On the downside, an explicit understanding of the harmonic analysis of these two groups is needed, which in the case of the motion group can be cumbersome.

To be more concrete, we demonstrate the basic strategy using the cyclic group $\mathbb{Z}_n$. This group is finite, so that discretization is unnecessary, and Abelian, so that harmonic analysis becomes simple. Nevertheless, this simple example already carries many essential features, and it ought to be kept in mind by the reader when the more complicated cases are treated later.

Let $\Sigma \subseteq \mathbb{Z}_n$ with $0 \notin \Sigma$ be closed under taking negatives, i.e., $\Sigma = -\Sigma$. Then we define the *Cayley graph*

$$\text{Cayley}(\mathbb{Z}_n, \Sigma) = (\mathbb{Z}_n, \{\, \{x, y\} : x - y \in \Sigma \,\}),$$

which is an undirected graph whose vertices are the elements of $\mathbb{Z}_n$ and where Σ defines the neighborhood of the neutral element 0; this neighborhood is then transported to every vertex by group translations. Since $\Sigma = -\Sigma$, the definition is consistent, and since $0 \notin \Sigma$, the Cayley graph does not have loops. For example, the n-cycle can be represented as a Cayley graph:

$$C_n = \text{Cayley}(\mathbb{Z}_n, \Sigma) \quad \text{with} \quad \Sigma = \{1, -1\}.$$

The goal in this section is to show that the computation of the theta number $\vartheta'_e(\text{Cayley}(\mathbb{Z}_n, \Sigma))$ with unit weights $e = (1, \ldots, 1)$ reduces from a semidefinite program to a linear program if one works in the Fourier domain.

For this we need the *characters* of $\mathbb{Z}_n$, which are group homomorphisms $\chi \colon \mathbb{Z}_n \to \mathbb{T}$, where $\mathbb{T}$ is the unit circle in the complex plane. Thus, every character χ satisfies

$$\chi(x + y) = \chi(x)\chi(y) \qquad \forall x, y \in \mathbb{Z}_n.$$

The characters themselves form a group with the operation of pointwise multiplication $(\chi\psi)(x) = \chi(x)\psi(x)$; this is the *dual group* $\widehat{\mathbb{Z}}_n$ of $\mathbb{Z}_n$. The *trivial character* e of $\mathbb{Z}_n$ defined by $e(x) = 1$ for all $x \in \mathbb{Z}_n$ is the unit element. Moreover, if χ is a character, then its inverse is its complex conjugate $\overline{\chi}$ such that $\overline{\chi}(x) = \overline{\chi(x)}$ for all $x \in \mathbb{Z}_n$. We often view characters as vectors in the vector space $\mathbb{C}^{\mathbb{Z}_n}$.

Lemma 1. *Let χ and ψ be characters of $\mathbb{Z}_n$. Then the following orthogonality relation holds:*

$$\chi^*\psi = \sum_{x \in \mathbb{Z}_n} \overline{\chi(x)}\psi(x) = \begin{cases} |\mathbb{Z}_n| & \text{if } \chi = \psi, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. If $\chi = \psi$, then

$$\chi^*\chi = \sum_{x \in \mathbb{Z}_n} \overline{\chi}(x)\chi(x) = \sum_{x \in \mathbb{Z}_n} 1 = |\mathbb{Z}_n|$$

holds. If $\chi \neq \psi$, then there is $y \in \mathbb{Z}_n$, so that $(\overline{\chi}\psi)(y) \neq 1$. Furthermore, we have

$$\begin{aligned} (\overline{\chi}\psi)(y)\chi^*\psi &= (\overline{\chi}\psi)(y) \sum_{x \in \mathbb{Z}_n} \overline{\chi}(x)\psi(x) \\ &= \sum_{x \in \mathbb{Z}_n} \overline{\chi}(x + y)\psi(x + y) \\ &= \sum_{x \in \mathbb{Z}_n} \overline{\chi}(x)\psi(x) = \chi^*\psi, \end{aligned}$$

so $\chi^*\psi$ has to be zero. $\qquad\square$

As a corollary we can explicitly give all characters of $\mathbb{Z}_n$ and see that they form an orthogonal basis of $\mathbb{C}^{\mathbb{Z}_n}$. It follows that the dual group $\widehat{\mathbb{Z}}_n$ is isomorphic to $\mathbb{Z}_n$.

Corollary 3. *Every element $u \in \mathbb{Z}_n$ defines a character of $\mathbb{Z}_n$ by $\chi_u(x) = e^{2\pi i u x/n}$. The map $u \mapsto \chi_u$ is a group isomorphism between $\mathbb{Z}_n$ and its dual group $\widehat{\mathbb{Z}}_n$.*

Proof. One immediately verifies that the map $u \mapsto \chi_u$ is well defined, that it is an injective group homomorphism, and that χ_u is a character of $\mathbb{Z}_n$. By the orthogonality relation we see that the number of different characters of $\mathbb{Z}_n$ is at most the dimension of the space $\mathbb{C}^{\mathbb{Z}_n}$. Hence $|\mathbb{Z}_n|$ equals $|\widehat{\mathbb{Z}}_n|$, and the map is a bijection. $\qquad\square$

Given a function $f \colon \mathbb{Z}_n \to \mathbb{C}$, the function $\hat{f} \colon \widehat{\mathbb{Z}}_n \to \mathbb{C}$ such that

$$\hat{f}(\chi) = \frac{1}{|\mathbb{Z}_n|} \sum_{x \in \mathbb{Z}_n} f(x)\chi^{-1}(x)$$

is the *discrete Fourier transform* of f; the coefficients $\hat{f}(\chi)$ are called the *Fourier coefficients* of f. We then have the *Fourier inversion formula*:

$$f(x) = \sum_{\chi \in \widehat{\mathbb{Z}}_n} \hat{f}(\chi)\chi(x).$$

We say that $f \colon \mathbb{Z}_n \to \mathbb{C}$ is of *positive type* if $f(x) = \overline{f(-x)}$ for all $x \in \mathbb{Z}_n$ and for all $\rho \colon \mathbb{Z}_n \to \mathbb{C}$ we have

$$\sum_{x, y \in \mathbb{Z}_n} f(x - y)\rho(x)\overline{\rho(y)} \geq 0.$$

Thus, f is of positive type if and only if the matrix $K(x, y) = f(x - y)$ is positive semidefinite. With this we have the following characterization for the theta number of $\text{Cayley}(\mathbb{Z}_n, \Sigma)$.

Theorem 4. *We have that*

$$\begin{aligned} \vartheta'_e(\text{Cayley}(\mathbb{Z}_n, \Sigma)) = {} & \\ \min \quad & f(0) \\ \text{s.t.} \quad & f(x) \leq 0 \quad \text{for all } x \notin \Sigma \cup \{0\}, \\ & \textstyle\sum_{x \in \mathbb{Z}_n} f(x) \geq |\mathbb{Z}_n|, \\ & f \colon \mathbb{Z}_n \to \mathbb{R} \text{ is of positive type.} \end{aligned} \tag{3}$$

Alternatively, expressing f in the Fourier domain, we obtain

$$\begin{aligned} \vartheta'_e(\text{Cayley}(\mathbb{Z}_n, \Sigma)) = {} & \\ \min \quad & \textstyle\sum_{\chi \in \widehat{\mathbb{Z}}_n} \hat{f}(\chi) \\ \text{s.t.} \quad & \textstyle\sum_{\chi \in \widehat{\mathbb{Z}}_n} \hat{f}(\chi)\chi(x) \leq 0 \quad \text{for all } x \notin \Sigma \cup \{0\}, \\ & \hat{f}(e) \geq 1, \\ & \hat{f}(\chi) \geq 0 \text{ and } \hat{f}(\chi) = \hat{f}(\chi^{-1}) \text{ for all } \chi \in \widehat{\mathbb{Z}}_n. \end{aligned} \tag{4}$$

Proof. Functions $f \colon \mathbb{Z}_n \to \mathbb{C}$ correspond to $\mathbb{Z}_n$-invariant matrices $K \colon \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{C}$, which are matrices such that $K(x + z, y + z) = K(x, y)$ for all $x, y, z \in \mathbb{Z}_n$.

In solving problem (1) for computing ϑ'_e we may restrict ourselves to $\mathbb{Z}_n$-invariant matrices. This fact can be seen by a symmetrization argument: If (M, K) is an optimal solution of (1), then so is $(M, \overline{K})$ with

$$\overline{K}(x, y) = \frac{1}{|\mathbb{Z}_n|} \sum_{z \in \mathbb{Z}_n} K(x + z, y + z),$$

which is $\mathbb{Z}_n$-invariant.

Thus, we can translate problem (1) into (3). The objective function and the constraint on nonedges translate easily. The positive-semidefiniteness constraint requires a bit more work.

First, observe that to require K to be real and symmetric is to require f to be real and such that $f(x) = f(-x)$ for all $x \in \mathbb{Z}_n$. We claim that each character χ of $\mathbb{Z}_n$ gives an eigenvector of K with eigenvalue $|\mathbb{Z}_n|\hat{f}(\chi)$. Indeed, using the inversion formula, we have

$$\begin{aligned}
(K\chi)(x) &= \sum_{y \in \mathbb{Z}_n} K(x, y)\chi(y) = \sum_{y \in \mathbb{Z}_n} f(x - y)\chi(y) \\
&= \sum_{y \in \mathbb{Z}_n} \sum_{\psi \in \widehat{\mathbb{Z}}_n} \hat{f}(\psi)\psi(x - y)\chi(y) \\
&= \sum_{\psi \in \widehat{\mathbb{Z}}_n} \hat{f}(\psi) \sum_{y \in \mathbb{Z}_n} \psi(y)\chi(x - y) \\
&= \sum_{\psi \in \widehat{\mathbb{Z}}_n} \hat{f}(\psi)\chi(x) \sum_{y \in \mathbb{Z}_n} \psi(y)\overline{\chi(y)} \\
&= |\mathbb{Z}_n|\hat{f}(\chi)\chi(x),
\end{aligned}$$

as claimed.

This immediately implies that K is positive semidefinite—or, equivalently, f is of positive type—if and only if $\hat{f}(\chi) \geq 0$ for all characters χ. Now, since $\hat{f}(e) = |\mathbb{Z}_n|^{-1}\sum_{x \in \mathbb{Z}_n} f(x)$ and since e is an eigenvalue of K, then $K - ee^\mathsf{T}$ is positive semidefinite if and only if $\sum_{x \in \mathbb{Z}_n} f(x) \geq |\mathbb{Z}_n|$ and f is of positive type.

We thus see that (1) can be translated into (3). Using the inversion formula and noting that f is real-valued if and only if $\hat{f}(\chi) = \hat{f}(\chi^{-1})$ for all χ, one immediately obtains (4). □

Cayley graphs on the cyclic group are not particularly exciting. Everything in this section, however, can be straightforwardly applied to any finite Abelian group. If, for instance, one considers the group $\mathbb{Z}_2^n$, then binary codes can be modeled as independent sets of Cayley graphs, and the analogue of Theorem 4 gives Delsarte's linear programming bound [9].

6 Upper Bounds for Congruent and Translational Body Packings

The packing graphs described above have noncompact vertex sets, but we said they can be compactified so that the theta number can be applied. Let us now see how that can be done.

First we need a definition of packing density. Given a packing $\mathcal{P}$, we say that its density is Δ if for every $p \in \mathbb{R}^n$

we have

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

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

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

We say that a packing $\mathcal{P}$ is *periodic* if there is a lattice[1] $L \subseteq \mathbb{R}^n$ that leaves $\mathcal{P}$ invariant, that is, $\mathcal{P} = x + \mathcal{P}$ for every $x \in L$. Lattice L is the *periodicity lattice* of $\mathcal{P}$. In other words, $\mathcal{P}$ consists of some bodies placed inside the fundamental cell of L, and this arrangement repeats itself at each copy of the fundamental cell translated by vectors of the lattice.

Periodic packings always have a density. Moreover, given any packing $\mathcal{P}$, one may define a sequence of periodic packings whose fundamental cells have volumes approaching infinity and whose densities converge to the upper density of $\mathcal{P}$. Therefore, when computing bounds for the maximum density of packings, we may restrict ourselves to periodic packings.

This is the key observation that allows us to compactify the packing graphs. Let $\mathcal{K}_1, \ldots, \mathcal{K}_N \subseteq \mathbb{R}^n$ be some given convex bodies. We have defined the translational packing graph of $\mathcal{K}_1, \ldots, \mathcal{K}_N$. Given a lattice $L \subseteq \mathbb{R}^n$, we may define a periodic version of the packing graph. This is the graph G_L, whose vertex set is $V = \{1, \ldots, N\} \times (\mathbb{R}^n/L)$. Now, vertex (i, x) of G_L corresponds not only to one body, but to many: specifically, to all the bodies $x + v + \mathcal{K}_i$, for $v \in L$. Vertices (i, x) and (j, y) are then adjacent if for some $v \in L$ we have

$$(x + v + \mathcal{K}_i)^\circ \cap (y + \mathcal{K}_j)^\circ \neq \emptyset.$$

Then an independent set of G_L corresponds to a periodic packing of translations of $\mathcal{K}_1, \ldots, \mathcal{K}_N$ with periodicity lattice L, and vice versa.

Graph G_L has a compact vertex set and each one of its independent sets is finite. If we consider the weight function $w \colon V \to \mathbb{R}_+$ such that $w(i, x) = \text{vol } \mathcal{K}_i$, then the maximum density of a periodic packing with periodicity lattice L is given by

$$\frac{\alpha_w(G_L)}{\text{vol}(\mathbb{R}^n/L)}.$$

Thus, one strategy to find an upper bound for the maximum density of a packing is to find an upper bound for $\alpha_w(G_L)$ for every L.

Notice that V is actually a separable and compact measure space that satisfies the hypothesis of Theorem 2. Therefore, $\vartheta'_w(G_L)$ provides an upper bound for $\alpha_w(G_L)$. Let us see how one may obtain a feasible solution of (2) for every graph G_L.

Let $f \colon \mathbb{R}^n \to \mathbb{C}$ be a rapidly decreasing function. This is an infinitely differentiable function with the following property: any derivative, multiplied by any polynomial, is a bounded function.

---

[1]A lattice is a discrete subgroup of $(\mathbb{R}^n, +)$.

The *Fourier transform* of f computed at $u \in \mathbb{R}^n$ is

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

where $u \cdot x = u_1 x_1 + \cdots + u_n x_n$. Since f is rapidly decreasing, the inversion formula holds, giving

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

Consider now a matrix-valued function $f \colon \mathbb{R}^n \to \mathbb{C}^{N \times N}$, where $f(x) = \big(f_{ij}(x)\big)_{i,j=1}^N$ and each function f_{ij} is rapidly decreasing. For $u \in \mathbb{R}^n$ we write

$$\hat{f}(u) = \big(\hat{f}_{ij}(u)\big)_{i,j=1}^N.$$

Hence, the Fourier transform of f is also a matrix-valued function.

We say f is of *positive type* if $f(x) = f(-x)^*$ for every $x \in \mathbb{R}^n$ and for every L^∞ function $\rho \colon \mathbb{R}^n \to \mathbb{C}^N$ we have

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

One may prove that f is of positive type if and only if $\hat{f}(u)$ is positive semidefinite for every $u \in \mathbb{R}^n$.

We now have the following theorem.

Theorem 5. *Let $\mathcal{K}_1, \ldots, \mathcal{K}_N \subseteq \mathbb{R}^n$ be convex bodies. Suppose $f \colon \mathbb{R}^n \to \mathbb{R}^{N \times N}$ is such that each f_{ij} is rapidly decreasing and that it satisfies the following conditions:*
(i) $f_{ij}(x) \leq 0$ whenever $(x + \mathcal{K}_i)^\circ \cap \mathcal{K}_j^\circ = \emptyset$;
(ii) $\hat{f}(0) - \big((\operatorname{vol}\mathcal{K}_i)^{1/2} (\operatorname{vol}\mathcal{K}_j)^{1/2}\big)_{i,j=1}^N$ is positive semidefinite; and
(iii) f is of positive type.
Then the maximum density of a packing of translated copies of $\mathcal{K}_1, \ldots, \mathcal{K}_N$ is at most $\max\{ f_{ii}(0) : i = 1, \ldots, N \}$.

Proof. Let $w \colon V \to \mathbb{R}_+$ be the weight function such that $w(i,x) = \operatorname{vol}\mathcal{K}_i$ for all $(i,x) \in V$. The proof of the theorem consists in deriving from f, for *every* lattice $L \subseteq \mathbb{R}^n$, a kernel $K_L \in L^2(V \times V)$, where $V = \{1, \ldots, N\} \times (\mathbb{R}^n/L)$, and a number M_L that together give a feasible solution of (2), thus obtaining an upper bound for $\alpha_w(G_L)$.

For a given lattice L, we let

$$K_L((i,x),(j,y)) = \operatorname{vol}(\mathbb{R}^n/L) \sum_{v \in L} f_{ij}(x - y + v).$$

The above sum is well defined since each f_{ij} is rapidly decreasing. Moreover, this implies that K_L is continuous.

Given two distinct, nonadjacent vertices (i,x) and (j,y) of G_L, we have that for all $v \in L$,

$$(x+v+\mathcal{K}_i)^\circ \cap (y+\mathcal{K}_j)^\circ = \emptyset \iff (x-y+v+\mathcal{K}_i)^\circ \cap \mathcal{K}_j^\circ = \emptyset.$$

This means that $f_{ij}(x - y + v) \leq 0$ for all $v \in L$. But then $K_L((i,x),(j,y)) \leq 0$, as we wanted.

Now we show that $K_L - W$ is a positive kernel, where $W((i,x),(j,y)) = (\operatorname{vol}\mathcal{K}_i)^{1/2}(\operatorname{vol}\mathcal{K}_j)^{1/2}$. This is implied by conditions (ii) and (iii) of the theorem and can be proven directly by combining the definition of a positive kernel with that of a function of positive type. We take another road, however, and exhibit a complete list of eigenfunctions and eigenvalues of $K_L - W$.

Let $L^* = \{ u \in \mathbb{R}^n : u \cdot v \in \mathbb{Z} \text{ for all } v \in L \}$ be the *dual lattice* of L, and consider the matrix $W' \in \mathbb{R}^{N \times N}$ with $W'_{ij} = (\operatorname{vol}\mathcal{K}_i)^{1/2}(\operatorname{vol}\mathcal{K}_j)^{1/2}$. Since f is of positive type, for each $u \in \mathbb{R}^n$ we have that $\hat{f}(u)$ is positive semidefinite. Moreover, from condition (ii) we have that $\hat{f}(0) - W'$ is positive semidefinite. Hence, the matrices $\hat{f}(u) - \delta_u W'$, where δ_u equals 1 if $u = 0$ and 0 otherwise, are positive semidefinite.

For $u \in L^*$, let $a_{1,u}, \ldots, a_{N,u}$ be an orthonormal basis of $\mathbb{R}^N$ consisting of eigenvectors of $\hat{f}(u) - \delta_u W'$, with associated eigenvalues $\lambda_{1,u}, \ldots, \lambda_{N,u}$, which are all nonnegative.

Also for $u \in L^*$, let $\chi_u(x) = e^{2\pi i u \cdot x}$. Then $(\operatorname{vol}(\mathbb{R}^n/L))^{1/2}\chi_u$, $u \in L^*$, forms a complete orthonormal system of $L^2(\mathbb{R}^n/L)$, and so

$$(\operatorname{vol}(\mathbb{R}^n/L))^{1/2} a_{k,u} \otimes \chi_u$$

for $k = 1, \ldots, N$ and $u \in L^*$ forms a complete orthonormal system of $L^2(V)$. We claim that each such function is an eigenfunction of $K_L - W$.

Indeed, let $(i,x) \in V$ be given. Notice that

$$\begin{aligned}
&[W(a_{k,u} \otimes \chi_u)]((i,x)) \\
&= \int_V W((i,x),(j,y))(a_{k,u} \otimes \chi_u)(j,y) \, d(j,y) \\
&= \sum_{j=1}^N W'_{ij}(a_{k,u})_j \int_{\mathbb{R}^n/L} e^{2\pi i u \cdot y} \, dy \\
&= \sum_{j=1}^N W'_{ij}(a_{k,u})_j \operatorname{vol}(\mathbb{R}^n/L)\delta_u \\
&= \operatorname{vol}(\mathbb{R}^n/L)(W' a_{k,u})_i \delta_u.
\end{aligned}$$

Similarly we have

$$\begin{aligned}
&[K_L(a_{k,u} \otimes \chi_u)](i,x) \\
&= \int_V K_L((i,x),(j,y))(a_{k,u} \otimes \chi_u)(j,y) \, d(j,y) \\
&= \operatorname{vol}(\mathbb{R}^n/L) \sum_{j=1}^N \int_{\mathbb{R}^n/L} \sum_{v \in L} f_{ij}(x - y + v)(a_{k,u})_j e^{2\pi i u \cdot y} \, dy \\
&= \operatorname{vol}(\mathbb{R}^n/L) \sum_{j=1}^N (a_{k,u})_j \int_{\mathbb{R}^n} f_{ij}(x-y) e^{2\pi i u \cdot y} \, dy \\
&= \operatorname{vol}(\mathbb{R}^n/L) \sum_{j=1}^N (a_{k,u})_j \int_{\mathbb{R}^n} f_{ij}(y) e^{2\pi i u \cdot (x-y)} \, dy \\
&= \operatorname{vol}(\mathbb{R}^n/L) \sum_{j=1}^N \hat{f}_{ij}(u)(a_{k,u})_j e^{2\pi i u \cdot x} \\
&= \operatorname{vol}(\mathbb{R}^n/L)(\hat{f}(u) a_{k,u})_i e^{2\pi i u \cdot x}.
\end{aligned}$$

Putting everything together, we have

$$\begin{aligned}
&[(K_L - W)(a_{k,u} \otimes \chi_u)](i,x) \\
&= \operatorname{vol}(\mathbb{R}^n/L)(\hat{f}(u) a_{k,u} - \delta_u W' a_{k,u})_i e^{2\pi i u \cdot x} \\
&= \operatorname{vol}(\mathbb{R}^n/L)\lambda_{k,u}(a_{k,u})_i e^{2\pi i u \cdot x} \\
&= \operatorname{vol}(\mathbb{R}^n/L)\lambda_{k,u}(a_{k,u} \otimes \chi_u)(i,x).
\end{aligned}$$

We see that all the functions $a_{k,u} \otimes \chi_u$ are eigenfunctions of $K_L - W$ with nonnegative associated eigenvalues, and it follows that $K_L - W$ is a positive kernel.

We now need to provide the bound M_L on the diagonal elements of K_L. To do so, we assume that the minimum vector of L is large enough that $(v + \mathcal{K}_i)^\circ \cap \mathcal{K}_i^\circ = \emptyset$ for all nonzero $v \in L$; this involves no loss of generality, since we care only about lattices with large fundamental cells, and one can scale L appropriately. But this means that $f_{ii}(v) \leq 0$ for all i and nonzero $v \in L$. Then from the definition of K_L we have that

$$K_L((i,x),(i,x)) \leq \mathrm{vol}(\mathbb{R}^n/L) f_{ii}(0)$$

for all $(i,x) \in V$, and we can take

$$M_L = \mathrm{vol}(\mathbb{R}^n/L) \max\{\, f_{ii}(0) : i = 1, \ldots, N \,\}.$$

We now have that the maximum density of a periodic packing with periodicity lattice L is

$$\frac{\alpha_w(G_L)}{\mathrm{vol}(\mathbb{R}^n/L)} \leq \frac{\vartheta'_w(G_L)}{\mathrm{vol}(\mathbb{R}^n/L)} \leq \max\{\, f_{ii}(0) : i = 0, \ldots, N \,\},$$

proving the theorem. $\square$

Theorem 5 was stated in the time domain; but using the inversion formula and the fact that a matrix-valued function is of positive type if and only if its Fourier transform is everywhere positive-semidefinite, we can restate it in the Fourier domain. We will use this alternative version in the next section, when we discuss a computational approach for finding functions f satisfying the conditions required in the theorem.

When $N = 1$, Theorem 5 is a direct analogue of Theorem 4. Indeed, then the translational packing graph is actually a Cayley graph with $\mathbb{R}^n$ as its vertex set. Although noncompact, $\mathbb{R}^n$ is an Abelian group, and the functions $\chi_u(x) = e^{2\pi i u \cdot x}$, for $u \in \mathbb{R}^n$, give its characters. The Fourier transform for $\mathbb{R}^n$ is the direct analogue of the discrete Fourier transform for $\mathbb{Z}_n$.

Moreover, except for the compactification step and other technical issues stemming from analysis, the proof of Theorem 5 follows exactly the same pattern of the proof of Theorem 4. Notice in particular how the characters give eigenvectors of the translation-invariant kernel K defined by f.

A theorem similar to Theorem 5 can be proven for packings of congruent copies of a given convex body $\mathcal{K} \subseteq \mathbb{R}^n$. Recall that the congruent packing graph has as vertex set $V = \mathrm{SO}(n) \times \mathbb{R}^n$. Set V is actually a group with identity $(I,0)$, where I is the identity matrix, under the operation

$$(A,x)(B,y) = (AB, x + Ay).$$

This group is denoted by $\mathrm{M}(n)$ and called the *Euclidean motion group*.

We will now work with complex-valued functions over $\mathrm{M}(n)$. There is also a definition of what it means for such a function to be rapidly decreasing, although it is more technical than the definition for functions over $\mathbb{R}^n$.

A function $f \in L^1(\mathrm{M}(n))$ is of *positive type* if $f(A,x) = \overline{f((A,x)^{-1})}$ for all $(A,x) \in \mathrm{M}(n)$ and for all $\rho \in L^\infty(\mathrm{M}(n))$ we have that

$$\int_{\mathrm{M}(n)} \int_{\mathrm{M}(n)} f((B,y)^{-1}(A,x)) \rho(A,x) \overline{\rho(B,y)} \, d(A,x) d(B,y)$$

is nonnegative. Here, we take as the measure the product of the Haar measure for $\mathrm{SO}(n)$, normalized so that $\mathrm{SO}(n)$ has total measure 1, with the Lebesgue measure for $\mathbb{R}^n$.

Theorem 6. *Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a convex body. Suppose $f \colon \mathrm{M}(n) \to \mathbb{R}$ is rapidly decreasing and that it satisfies the following conditions:*
(i) $f(A,x) \leq 0$ whenever $(x + A\mathcal{K})^\circ \cap \mathcal{K}^\circ = \emptyset$;
(ii) $\int_{\mathrm{M}(n)} f(A,x) \, d(A,x) \geq \mathrm{vol}\, \mathcal{K}$; and
(iii) f is of positive type.
Then the maximum density of a packing of congruent copies of $\mathcal{K}$ is at most $f(I,0)$.

The proof of Theorem 6 is slightly more technical than the proof of Theorem 5, but otherwise it follows the same pattern.

Notice that the congruent packing graph is a Cayley graph whose vertex set is the Euclidean motion group. Thus, Theorem 6 is also an analogue of Theorem 4. It is, however, more distant from Theorem 4 than Theorem 5 is, since $\mathbb{R}^n$ is Abelian but $\mathrm{M}(n)$ is not. Therefore, when one does harmonic analysis over $\mathrm{M}(n)$, using the characters is not enough: one also needs to consider higher-dimensional irreducible representations, most of them are even infinite-dimensional.

Although Theorem 6 can be restated in the Fourier domain just as Theorem 4 could, it now becomes harder to carry out this procedure explicitly—already for $n = 2$ or 3, the formulas involved are significantly more complicated than the ones for $\mathbb{R}^n$. Using the formulas effectively in a computational approach remains the main obstacle in applying Theorem 6.

7 A Computational Approach

Theorem 5 and Theorem 6 might be mathematically pleasing per se, but the real challenge is to determine explicit functions giving good bounds. So far this has been done for only a few cases. When $N = 1$, Theorem 5 becomes a theorem of Cohn and Elkies [5]. The Cohn-Elkies bound provides the basic framework for proving the best known upper bounds for the maximum density of sphere packings in dimensions $4, \ldots, 36$. It is also conjectured to provide tight bounds in dimensions 8 and 24, and strong numerical evidence supports this conjecture. De Laat, Oliveira, and Vallentin [13] have proposed a strengthening of the Cohn-Elkies bound and computed better upper bounds for the maximum density of sphere packings in dimensions 4, 5, 6, 7, and 9.

Here we want to give an idea of how to set up a semidefinite program for finding good functions. Let B_n be the unit ball in $\mathbb{R}^n$. To find bounds for the density of a sphere-packing, we want to find a function $f \colon \mathbb{R}^n \to \mathbb{R}$ with $f(0)$ as small as possible such that

(i) $f(x) \leq 0$ whenever $(x + B_n)^\circ \cap B_n^\circ = \emptyset$;

(ii) $\hat{f}(0) - \text{vol}\, B_n \geq 0$; and

(iii) f is of positive type, which means that $\hat{f}(u)$ is nonnegative for all $u \in \mathbb{R}^n$.

Without loss of generality we can assume that the function f is even and radial; in other words, $f(x)$ depends only on the norm of x, so it is essentially an even univariate function. Another good feature is that the Fourier transform of a radial function is radial again. Functions whose Fourier transform have the form

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

where p is an even and univariate polynomial, are dense in the space of rapidly decreasing even and radial functions. Then by the Fourier inversion formula we can compute f explicitly, monomial by monomial, through

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

where $L_k^{n/2-1}$ is the Laguerre polynomial of degree k with parameter $n/2 - 1$. These are orthogonal polynomials on the half-open interval $[0, \infty)$ with respect to the measure $x^{n/2-1}e^{-x}\, dx$.

We specify function f via the polynomial p. To do so, we fix $d > 0$ and work with polynomials of degree up to $2d$, that is, with polynomials of the form

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

Working with finite d is our way of discretizing the Fourier domain, a necessary step as we observed in Section 5.

Now constraints on f become constraints on p, which can be modeled as sum-of-squares constraints (see, e.g., the expository papers of Lasserre and Parrilo in *SIAG/OPT Views and News* 15-2 (2004)). Thus, we can set up a semidefinite programming problem to find a function f satisfying the required constraints:

$$
\begin{aligned}
\min \quad & \sum_{k=0}^{d} a_k k!\, \pi^{-k} L_k^{n/2-1}(0) \\
\text{s.t.} \quad & p(t) = \sum_{k=0}^{d} a_{2k} t^{2k}, \\
& 0 = \sum_{k=0}^{d} a_k k!\, \pi^{-k} L_k^{n/2-1}(\pi w^2) \\
& \qquad + v_d^{\mathsf{T}}(w)Rv_d(w) + (w^2 - 2^2)v_{d-2}^{\mathsf{T}}(w)Sv_{d-2}(w), \\
& p(0) - \text{vol}\, B_n \geq 0, \\
& p(t) = v_d^{\mathsf{T}}(t)Qv_d(t), \\
& Q,\, R,\, S \text{ are positive semidefinite matrices,}
\end{aligned}
$$

where $v_d(z) = (1, z, \ldots, z^d)$ is the vector of all monomials up to degree d.

From a numerical perspective this formulation is a catastrophe—a fact well known to specialists in the field—since the monomial basis is used. Even though the resulting semidefinite program is small, say when we use $d = 10$, one cannot get a solution from standard numerical solvers. On the other hand this program can be implemented in many other, equivalent ways by using different choices of polynomial bases. Here we have two choices: one for the vectors

v_d and one for testing the polynomial identities. With some experimentation we found that the basis

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

where μ_k is the absolute value of the coefficient of $L_k^{n/2-1}(2\pi t)$ with largest absolute value, performs well.

We believe that the problem of finding a good basis deserves further investigation. Currently almost nothing (to the best of our knowledge only the papers by Löfberg and Parrilo [15] and Roh and Vandenberghe [18] address this issue) is known about it, although it is a crucial factor for solving polynomial optimization problems in practice.

Another use of Theorem 5 is to provide bounds for binary sphere packings. These are packings of balls of two different sizes; i.e., we have $N = 2$ and $\mathcal{K}_1$, $\mathcal{K}_2$ are balls. Binary sphere packings occur naturally in applications such as materials science and chemistry. De Laat, Oliveira, and Vallentin [13] used Theorem 5 to compute upper bounds for the maximum densities of binary sphere packings in dimensions $2, \ldots, 5$.

Recently, Oliveira and Vallentin [17] used Theorem 6 to compute upper bounds for the densities of pentagon packings. Here a new challenge arises: The Fourier transform is no longer matrix-valued but takes infinite-dimensional Hilbert-Schmidt kernels as values. Oliveira and Vallentin determined a first upper bound (0.98 compared with the best known lower bound of 0.92), and the numerical result obtained gives hope that the theorem will also be useful in the case of tetrahedra packings to meet the challenge of Gravel, Elser, and Kallus.

8 Conclusion

It is natural to consider optimization methods when dealing with geometric packing problems, and we have described here how well-known methods from combinatorial optimization—namely, the Lovász theta number and its variants—can be extended to provide upper bounds for the packing density. Such extensions provide a uniform framework to deal with geometric packing problems.

For finite graphs, only in specific cases does the Lovász theta number provide tight bounds. The same happens for geometric packing graphs: only in a few cases are the bounds coming from extensions of the theta number tight; in most cases, such bounds are but a first step in solving the problem.

The link made with combinatorial optimization techniques not only allows us to provide a unified framework and to have access to well-known optimization tools; it also points out to ways in which such bounds can be strengthened. The obvious approach is to extend ideas such as the Lasserre hierarchy to geometric packing problems. Such higher-order bounds can incorporate more sophisticated constraints like those coming from the local interaction of more than two vertices; in other words, we then deal with k-point correlation functions and not only with 2-point correlation functions.

Schrijver [19] considered 3-point correlation functions for binary codes, and Bachoc and Vallentin [2] used 3-point correlation functions for packings of spherical caps on the unit

sphere. De Laat and Vallentin [12] recently showed that this approach has the (theoretical) potential to solve all geometric packing problems. However, the price to pay is that the size of the optimization problems involved grows rapidly.

The success of such techniques will depend on several factors, including (i) how to analyze the optimization problem without using a computer, for instance to find asymptotic results; (ii) how to automatize the use of harmonic analysis; and (iii) how to solve semidefinite programs involving sums-of-squares constraints in an efficient and numerically stable manner.

REFERENCES

[1] C. Addabbo, Il libellus de impletione loci di Francesco Maurolico, to appear in *Quaderni del Dipartimento di Matematica, Sezione Didattica e Storia*, 2013, 59pp.

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

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

[4] E.R. Chen, M. Engel, and S.C. Glotzer, Dense crystalline dimer packings of regular tetrahedra, *Discrete & Computational Geometry* 44 (2010) 253–280.

[5] H. Cohn and N.D. Elkies, New upper bounds on sphere packings I, *Ann. of Math.* 157 (2003) 689–714.

[6] H. Cohn and A. Kumar, Optimality and uniqueness of the Leech lattice among lattices, *Ann. of Math.* 170 (2009) 1003–1050.

[7] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York, 1988.

[8] J.H. Conway and S. Torquato, Packing, tiling, and covering with tetrahedra, *Proc. Natl. Acad. Sci. USA* 103 (2006) 10612–10617.

[9] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Research Reports Supplements 1973 No. 10, Philips Research Laboratories, Eindhoven, 1973.

[10] S. Gravel, V. Elser, and Y. Kallus, Upper bound on the packing density of regular tetrahedra and octahedra, *Discrete & Computational Geometry* 46 (2011) 799–818.

[11] M. Grötschel, L. Lovász, and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* 1 (1981) 169–197.

[12] D. de Laat and F. Vallentin, A semidefinite programming hierarchy for packing problems in discrete geometry, *Mathematical Programming* 151 (2015) 529–553.

[13] D. de Laat, F.M. de Oliveira Filho, and F. Vallentin, Upper bounds for packings of spheres of several radii, *Forum of Mathematics, Sigma* 2 (2014), 42pp.

[14] J.C. Lagarias and C. Zong, Mysteries in packing regular tetrahedra, *Notices of the AMS* 59 (2012) 1540–1549.

[15] J. Löfberg and P.A. Parrilo, From coefficient to samples: A new approach to SOS optimization, in: *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004, pp. 3154–3159.

[16] L. Lovász, On the Shannon capacity of a graph, *IEEE Transactions on Information Theory* IT-25 (1979) 1–7.

[17] F.M. de Oliveira Filho and F. Vallentin, Computing upper bounds for the packing density of congruent copies of convex bodies I, arXiv:1308.4893, 2013, 28pp.

[18] T. Roh and L. Vandenberghe, Discrete transforms, semidefinite programming, and sum-of-squares representations of nonnegative polynomials, *SIAM J. Optim.* 16 (2006) 939–964.

[19] A. Schrijver, New code upper bounds from the Terwilliger algebra and semidefinite programming, *IEEE Transactions on Information Theory* 51 (2005) 2859–2866.

[20] Simplicius, *On Aristotle On the Heavens 3.7–4.6* (translated by I. Mueller), Gerald Duckworth & Co. Ltd., London, 2009.

[21] D.J. Struik, Het probleem "de impletione loci", *Nieuw Archief voor Wiskunde II* 15 (1926) 121–137.

[22] S. Torquato and Y. Jiao, Dense packings of the Platonic and Archimedean solids, *Nature* 460 (2009) 876–879.

[23] G.M. Ziegler, Three Mathematics Competitions, in: *An Invitation to Mathematics: From Competitions to Research* (D. Schleicher and M. Lackmann, eds.), Springer-Verlag, Berlin, 2011, pp. 195–206.

In Memoriam

Che-Lin Su (1974–2015)

Che-Lin Su on top of the Titlis (Switzerland) in March 2013 *(Photo courtesy of Ben Skrainka).*

Che-Lin Su, a highly respected associate professor at the University of Chicago Booth School of Business, passed away on July 31, 2015, after a short illness. Che-Lin was a young man of great promise and achievement. He earned an international reputation as a solid researcher, a fine teacher, and a contributor to several scholarly communities such as economics, marketing, optimization, operations management, and operations research. In his relatively short professional lifetime, Che-Lin built a collection of collaborators, admirers, and close friends that spanned the globe.

A native of Taiwan born in 1974, Che-Lin attended the National Taiwan University, graduating in 1996 with a B.S. degree in agricultural engineering. In the autumn of 1998, he entered the M.S. program of the newly merged Department of Engineering-Economic Systems and Operations Research (EES-OR) at Stanford. The degree was conferred in January 2001. After just one quarter of master's-level course work, however, Che-Lin was strongly encouraged to add doctoral-level courses to his program of studies and to apply to the department's separate Ph.D. program. After having demonstrated his ability to achieve honor grades in the EES-OR Department's doctoral-level optimization courses, Che-Lin gained admission to its Ph.D. program. As it happened, the EES-OR Department was to undergo another merger, this one with the Department of Industrial Engineering and Engineering Management (IE-EM). Formed in January 2000, the new department chose the name Management Science & Engineering (MS&E). It is from this department that Che-Lin earned his Ph.D. in 2005.

The title of Che-Lin's doctoral dissertation was *Equilibrium Problems with Equilibrium Constraints: Stationarity, Algorithms and Applications* [3]. The thesis was supervised by myself and approved by the reading committee that included Hung-po Chao (Electric Power Research Institute), Kenneth Judd (Hoover Institution), and Michael Saunders (MS&E). Equilibrium problems with equilibrium constraints (abbreviated as EPECs) are a kind of extension of mathematical programs with equilibrium constraints (called MPECs for short). The latter is the title of a pioneering monograph by Zhi-Quan Luo, Jong-Shi Pang, and Daniel Ralph published in 1996. In August 2004, while he was still a graduate student, Che-Lin met Jong-Shi Pang, organizer and host of the (first) International Conference on Continuous Optimization. On that occasion Jong-Shi encouraged Che-Lin to make contact with Kenneth Judd. These two connections would grow much closer in the years to come.

Ken played a major role in Che-Lin's early professional development and indeed in his later life, right up to its last day. Che-Lin had a research assistantship with Ken during the summer of 2005. In August of that year they attended the International Conference on Complementarity Problems held at Stanford. Each delivered one of two papers they had co-authored that summer. The one presented by Che-Lin was on MPEC approaches to moral-hazard problems. This summer's experience would prove to be the beginning of an enduring research collaboration.

Che-Lin's writings, speaking appearances, and teaching reveal his commitment to a challenging mission, one that occupied his mind from his days as a doctoral student to the end of his life: bringing state-of-the-art modeling and computational methods to problems in economics, management science, and operations. After completing his doctoral studies at Stanford, Che-Lin became a postdoc at Northwestern University's Kellogg School of Management, a position he held from 2005 to 2008. There he was attached to the Center for Mathematical Studies in Economics and Management Science. From 2006 to 2008, he was also a postdoctoral research fellow at the National Bureau of Economic Research. His postdoc advisors were Ken Judd and Karl Schmedders. While he was at the Kellogg School of Management, he collaborated with Ken on what became Che-Lin's most frequently cited publication: "Constrained Optimization Approaches to Estimation of Structural Models" [4]. This paper appeared in the prestigious journal *Econometrica* (2012). In this article, the authors demonstrate the effectiveness and efficiency of the MPEC formulation and solution methodology on single-agent dynamic discrete-choice models. The results of Monte Carlo experiments comparing `AMPL` and `MATLAB` implementations with the so-called nested fixed-point (NFXP) algorithm (proposed by Rust in 1987) show the superiority of the constrained optimization strategy and MPEC formulation. Another important and widely cited publication of Che-Lin is one called "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation" [1]. This paper, which was coauthored with Jean-Pierre Dubé

and Jeremy T. Fox, appeared in the same issue of *Econometrica* as the aforementioned Judd-Su paper.

In July 2008 Che-Lin joined the faculty at the University of Chicago Booth School of Business as an assistant professor of operations management. His teaching and research duties in Chicago were augmented by external lecturing and, in November 2009, research at the Cowles Foundation for Research in Economics at Yale University. There he held a visiting faculty position in the program area devoted to structural microeconomics, macroeconomics, and econometrics.

Che-Lin was promoted to associate professor of operations management at the University of Chicago Booth School of Business in 2012, where he continued teaching both MBA- and Ph.D.-level courses. The latter, obviously much closer to his research interests, included *Numerical Methods in Economics* and *Empirical Research in Operations Management*. In 2011 and 2013 he was invited to teach a course called *Computational Economics* at Harvard, and in 2012 he taught a half-semester course called *Computational Methods: Economic Dynamics* at Yale.

In every year between 2007 and 2015, and in many different places, Che-Lin gave a "mini-course" titled *Mathematical Programming Methods for Structural Estimation: Dynamic Programming, Demand Systems, and Games with Multiple Equilibria.* The principal venues for these lectures were the Chicago-Argonne Initiative for Computational Economics and the Zurich Initiative on Computational Economics held at the University of Zurich. Che-Lin was a featured speaker at these events and served on their program and selection committees. Other places where this mini-course was delivered include the University of Southern California Marshall School of Business, the Department of Industrial Economics and Technology Management at the Norwegian University of Science and Technology, the Cowles Foundation at Yale University, the Department of Economics at Johns Hopkins University, the Department of Econometrics and Operations Research of Tilburg University, and the University of Rochester Simon School of Business.

A noteworthy item on Che-Lin's publication list is a paper he coauthored with Jong-Shi Pang and Yu-Ching Lee. Recently accepted by the journal *Operations Research*, this paper is titled "A Constructive Approach to Estimating Pure Characteristics Demand Models with Pricing" [2]. Here again, a novel optimization approach to a structural estimation problem is expected to promote its fruitful application in econometric and marketing problems, previously hindered by computational challenges. An extension of the model with a game-theoretic flavor allows for producers to be competitive Nash-Bertrand players in price setting. Altogether, Che-Lin's mini-courses, teaching, invited lectures, and research publications demonstrate his advocacy of state-of-the-art numerical optimization in economics and operations management.

As many of his friends have said, Che-Lin loved good food and good company. He enjoyed professional baseball and attended Giants games in San Francisco and Cubs games in

Chicago. He also loved classical music and was a subscriber to Chicago Symphony Orchestra concerts.

On August 10, 2015, Che-Lin was to have spoken in a session that he organized at the recent ICIAM meeting in Beijing, but this came too late. Instead, he became the subject of a moving memorial session in the slot made open by his untimely death. The session included fond remembrances contributed by his colleagues and friends and was attended by many.

Che-Lin Su is survived by his wife, Bella Yang; his parents; and his brother and sister. Che-Lin's many colleagues, friends, and admirers join his family in sorrow.

Richard W. Cottle, Professor Emeritus
Department of MS&E, Stanford University, Stanford, CA, USA, rwc@stanford.edu

REFERENCES

[1] J.-P. Dubé, J. T. Fox, and C.-L. Su. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267, 2012.

[2] J.-S. Pang, C.-L. Su, and Y.-C. Lee. A constructive approach to estimating pure characteristics demand models with pricing. *Operations Research*, 2015.

[3] C.-L. Su. *Equilibrium problems with equilibrium constraints: Stationarity, algorithms, and applications.* Stanford University, 2005.

[4] C.-L. Su and K. L. Judd. Constrained optimization approaches to estimation of structural models. *Econometrica*, 80(5):2213–2230, 2012.

Bulletin

Email items to siagoptnews@lists.mcs.anl.gov *for consideration in the bulletin of forthcoming issues.*

1 Event Announcements

1.1 SIAM Workshop on Combinatorial Scientific Computing 2016 (CSC16)

The SIAM Workshop on Combinatorial Scientific Computing 2016 (CSC16) will take place in Albuquerque, NM, October 10–12, 2016.

CSC16 follows five earlier successful CSC workshops held roughly biennially since 2004. A new feature of CSC16 is that there will be a peer-reviewed proceedings associated with the workshop. The proceedings will be published by SIAM at http://epubs.siam.org/series/pr and will be made permanently available with DOI links.

Authors are invited to submit manuscripts that present original unpublished research in all areas of combinatorial scientific computing. Topics of interest include, but are not limited to,

- Combinatorial (discrete) models in scientific computing
- Graph and hypergraph algorithms in scientific computing
- Sparse matrix computations
- Combinatorial problems in data science and network science
- Combinatorial problems in Algorithmic Differentiation
- Combinatorial problems in optimization
- Applications of combinatorial scientific computing

Types of Submissions:

1. **Full Papers.** Papers should not exceed 10 pages using double-column format and 11-pt font size. The SIAM macro to be used for paper preparation will be provided. Accepted papers will be included in the proceedings, which will be published electronically by SIAM in its proceedings platform. Accepted papers will be presented at the workshop as talks.

2. **Extended Abstracts.** Abstracts should not exceed 2 pages using double-column format and 11-pt font size. Accepted abstracts will be presented as talks at the workshop, but will not be published in the proceedings.

Submission will be handled through EasyChair.

Important Dates:

Submission deadline (for both papers and abstracts): May 2, 2016, 11:59pm PDT.

Notification of Acceptance: July 5, 2016.

More details are available on the workshop website http://www.siam.org/meetings/csc16.

1.2 ICCOPT 2016 in Tokyo

The Fifth International Conference on Continuous Optimization (ICCOPT 2016) will take place in Tokyo, Japan, August 6–11, 2016. The summer school will be held from August 6th to 7th at the National Olympics Memorial Youth Center, Yoyogi, Tokyo, and the technical program August 8–11 at the National Graduate Institute for Policy Studies, Roppongi, Tokyo. ICCOPT is a flagship conference of the Mathematical Optimization Society, organized every three years. ICCOPT 2016 is cosponsored by the Operations Research Society of Japan.

Important Dates:

March 15 Summer School Accommodation Deadline (Student)

April 15 Abstract Submission for Parallel Session Deadline

April 15 Summer School Accommodation Deadline (General)

May 16 Abstract Submission for Poster Session Deadline

May 31 Early Registration Deadline

**August 6 & 7** Summer School

Plenary Speakers: Francis Bach, INRIA, France; Florian Jarre, Heinrich Heine Universität Düsseldorf, Germany; Jong-Shi Pang, University of Southern California, USA; Shuzhong Zhang, University of Minnesota, USA

Semi-Plenary Speakers: Yu-hong Dai, Chinese Academy of Sciences, China; Erick Delage, HEC Montréal, Canada; Mirjam Dür, Universität Trier, Germany; Katsuki Fujisawa, Kyushu University, Japan; Elad Hazan, Princeton University, USA; Jonathan Kelner, MIT, USA; Caroline Uhler, Institute of Science and Technology, Austria; Rachel Ward, University of Texas at Austin, USA

More details are available on the conference website at http://www.iccopt2016.tokyo.

2 Book Announcements

2.1 The Shapes of Things: *A Practical Guide to Differential Geometry and the Shape Derivative*
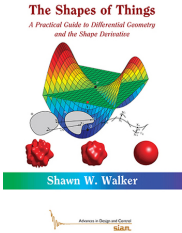
By Shawn W. Walker
Publisher: SIAM
Series: Advances in Design and Control, Vol. 28
*ISBN: 978-1-611973-95-2, x + 152 pages*
Published: June 2015
http://bookstore.siam.org/dc28/

ABOUT THE BOOK: Many things around us have properties that depend on their shape–for example, the drag characteristics of a rigid body in a flow. This self-contained overview of differential geometry explains how to differentiate a function (in the calculus sense) with respect to a "shape variable." This approach, which is useful for understanding mathematical models containing geometric partial differential equations, allows readers to obtain formulas for geometric quantities (such as curvature) that are clearer than those usually offered in differential geometry texts.

Readers will learn how to compute sensitivities with respect to geometry by developing basic calculus tools on surfaces and combining them with the calculus of variations. Several applications that utilize shape derivatives and many illustrations that help build intuition are included.

2.2 Variational Methods for the Numerical Solution of Nonlinear Elliptic Problems
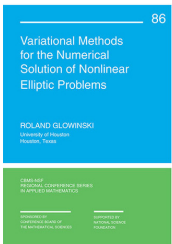
By Roland Glowinski
Publisher: SIAM
*Series: CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 86*
ISBN: 978-1-611973-77-8, xvi + 429 pages
Published: November 2015
http://bookstore.siam.org/cb86/

ABOUT THE BOOK: The author addresses computational methods that have proven efficient for the solution of a large variety of nonlinear elliptic problems. These methods can be applied to many problems in science and engineering, but this book focuses on their application to problems in continuum mechanics and physics. The book differs from others on the topic by presenting examples of the power and versatility of operator-splitting methods; providing a detailed introduction to alternating direction methods of multipliers and their applicability to the solution of nonlinear (possibly nonsmooth) problems from science and engineering; and showing that nonlinear least-squares methods, combined with operator-splitting and conjugate gradient algorithms, provide efficient tools for the solution of highly nonlinear problems.

2.3 Electrical Transmission System Cascades and Vulnerability: An Operations Research Viewpoint
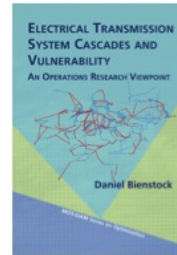
By Daniel Bienstock
Publisher: SIAM
Series: MOS-SIAM Series on Optimization, Vol. 22
*ISBN: 978-1-611974-15-7, viii + 296 pages*
Published: December 2015
http://bookstore.siam.org/mo22/

ABOUT THE BOOK: The power grid can be considered one of twentieth-century engineering's greatest achievements, and as grids and populations grow, robustness is a factor that planners must take into account. Power grid robustness is a complex problem for two reasons: the underlying physics is mathematically complex, and modeling is complicated by lack of accurate data.

This book sheds light on this complex problem by introducing the engineering details of power grid operations from the basic to the detailed; describing how to use optimization and stochastic modeling, with special focus on the modeling of cascading failures and robustness; providing numerical examples that show "how things work"; and detailing the application of a number of optimization theories to power grids.

2.4 AIMD Dynamics and Distributed Resource Allocation
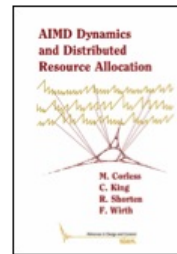
By M. Corless, C. King, R. Shorten, and F. Wirth
Publisher: SIAM
Series: Advances in Design and Control, Vol. 29
ISBN: 978-1-611974-21-8
Planned Publication: January 2016
http://bookstore.siam.org/dc29/

ABOUT THE BOOK: This is the first comprehensive book on the AIMD algorithm, the most widely used method for allocating a limited resource among competing agents without centralized control. The authors offer a new approach that is based on positive switched linear systems. It is used to develop most of the main results found in the book, and fundamental results on stochastic switched nonnegative and consensus systems are derived to obtain these results.

The original and best known application of the algorithm is in the context of congestion control and resource allocation on the Internet, and readers will find details of several variants of the algorithm in order of increasing complexity, including deterministic, random, linear, and nonlinear versions. In each case, stability and convergence results are derived based on unifying principles. Basic and fundamental properties of the algorithm are described, examples are used to illustrate the richness of the resulting dynamical systems, and applications are provided to show how the algorithm can be used in the context of smart cities, intelligent transportation systems, and the smart grid.

3 Other Announcements

2016–17 SAMSI Research Program on Optimization
Optimization is an exceptional field where the cooperation between mathematics and statistics can produce truly groundbreaking advances. The year-long SAMSI optimization (OPT) program aims to capitalize on this synergy and produce advances in:

1. Optimization for large-scale statistical analyses;
2. Statistical approaches for numerical solution of large-scale optimization problems; and
3. Applications of optimization.

The OPT program will start with two events:

- Summer School August 8–12, 2016
- Opening Workshop August 29–September 2, 2016

There are various ways to participate in the OPT program: Membership in remote research working groups, visiting researcher fellowships, and postdoctoral fellowships.

Remote Research Working Groups: The formation of the working groups is finalized at the Opening Workshop. From then on, they meet weekly throughout 2016–17, and on an electronic basis. The goal is to produce research on a well-defined, specific topic.

Research topics include: Optimization under uncertainty; convex and semidefinite optimization; robust optimization sparse regression, and stochastic gradient descent; E/M and M/M algorithms; mixed integer, linear and nonlinear optimization; PDE-constrained optimization with uncertainties; statistical inverse problems; computation of high-dimensional covariance functions; Bayesian optimization & computational decision analysis.

Application areas include: Machine learning; image and signal processing, and compressed sensing; energy and power grids; and finance.

Visiting Researcher Fellowships are available for periods ranging from a few weeks to the whole year 2016–17. Visitors are provided with office space and facilities at SAMSI, and various types of financial support.

Postdoctoral Fellowships begin in August 2016 and typically last two years. Appointments are made jointly between SAMSI and one of its partner universities, where teaching is a possibility. Extremely competitive salaries, travel stipends, and health insurance will be offered. Application deadline is 15 December 2015.

Questions should be sent to opt@samsi.info.

About SAMSI: The Statistical and Applied Mathematical Sciences Institute was established in 2002 as a partnership of Duke University, North Carolina State University, the University of North Carolina at Chapel Hill, and the National Institute of Statistical Sciences. SAMSI's mission is to forge a synthesis of the statistical sciences and the applied mathematical sciences with disciplinary science to confront the hardest and most important data- and model-driven scientific challenges. SAMSI is part of the Mathematical Sciences Institutes program of the Division of Mathematical Sciences at the National Science Foundation.

Comments from the Editors

Optimization Around the World

Congratulations on enjoying the 32nd issue of the SIAM Activity Group on Optimization's newsletter. We've managed to cover the globe, learning about optimization opportunities in Europe in the run up to the first exascale supercomputer, packing problems from Brazil (and Germany), preparations for ICCOPT in Tokyo, and an upcoming year focused on optimization at SAMSI in the US.

Still Printing (Some Copies)

Many thanks for your feedback to last issue's question regarding the printing of physical copies of *Views and News*. The strong consensus was that there was sufficient value in printing some copies.

Since electronic copies are distributed to the SIAM-OPT email list and archived online, we will do our part to aggressively reduce the number of physical copies. In addition to opting out of receiving a physical copy, you can also request that your physical copy be instead mailed to your organization (e.g., for distribution in a student lounge). Please do so via e-mail (siagoptnews@lists.mcs.anl.gov), which was also the recommended form of communication in the 1992 debut issue of *Views and News*! Some electronic highlights from that issue: intbib.bib has been renamed but is still available from netlib; most listed email addresses are still operational (sadly, the .bitnet address is not).

Other Electronic Resources

In light of the remarkable endurance of digital communication and archives, this issue we solicit your input on other ways that the activity group could benefit from electronic resources such as the activity group's wiki (did you know there was one?).

Are optimization-related job postings sufficiently represented? Would an archive (available not just to members of the activity group) of such job postings be valuable? What activities are underrepresented by other societies?

We welcome your feedback, (e-)mailed directly to us or to siagoptnews@lists.mcs.anl.gov. Suggestions for new issues, comments, and papers are always welcome! Best wishes for the new year, see you in 2016,

Stefan Wild, Editor
Mathematics and Computer Science Division, Argonne National Laboratory, USA, wild@anl.gov, http://www.mcs.anl.gov/~wild
Jennifer Erway, Editor
Department of Mathematics, Wake Forest University, USA, erwayjb@wfu.edu, http://www.wfu.edu/~erwayjb