SPHERE PACKING LECTURE

ALEX BLUMENTHAL

Abstract. What is the most space-efficient way to stack spheres in three-dimensional space? Although the answer is obvious, a rigorous proof has eluded mathematicians for centuries, having only recently been found by Hales in 1998, who used an immense computerized proof-by-exhaustion. The more general problem of packing (n-1)-spheres into n-dimensional Euclidean (or other) space is still only poorly understood, in spite of many the striking connections linking this problem to other areas of mathematics. In this talk, we shall explore these connections, discuss some of the methods used to prove sphere packing density bounds, and review some of the results that are known about sphere packings in high dimensions.

Contents

1. Statement of the Problem

A sphere packing $\Omega \subset \mathbb{R}^n$ is the union of a set of solid n-balls of a fixed (arbitrary) radius, where we require that the balls only 'kiss' or intersect at the boundary. It is clear that a sphere packing in $\mathbb{R}^n$ is specified by the locations of the centers of the spheres, i.e. a discrete subset $\mathcal{P} \subset \mathbb{R}^n$ such that for any $x, y \in \mathcal{P}$, we have $|x-y| \geq 2r$ for some fixed positive $r > 0$, the largest radius such that spheres of radius r placed at the points of $\mathcal{P}$ intersect only on their boundaries.

To define the "density" for an arbitrary sphere packing, it is natural to first consider packings that exhibit some sort of periodicity condition.

Definition 1. A *lattice* in $\mathbb{R}^n$ is the infinite discrete subset of $\mathbb{R}^n$ generated by integral linear combinations of some linearly-independent set $\{v_1, \cdots, v_n\} \subset \mathbb{R}^n$.

We call a packing $\mathcal{P}$ a lattice packing if $\mathcal{P} = \Lambda$ is a lattice. More important is the general notion of a *periodic packing*, which is any packing obtained from finitely many translations of a lattice Λ.

Suppose a periodic packing Ω is obtained from N translations of a lattice Λ. Then, we define the density $\Delta(\mathcal{P})$ to be

$$\Delta(\mathcal{P}) = \frac{N\omega_n r^n}{|\det \Lambda|}$$

where we identify Λ with the $n \times n$ matrix whose columns are the basis vectors $\{v_1, \cdots, v_n\}$, ω_n is the volume of a unit S^{n-1} in $\mathbb{R}^n$ and $r > 0$ is the radius of the packing. Often it is more convenient to work with the center density $\delta(\mathcal{P})$ defined from

$$\delta(\mathcal{P}) = \Delta(\mathcal{P})/\omega_n = \frac{N}{|\det \Lambda|}$$

We can generalize this to a definition of density for a packing without any periodic structure: let $p \in \mathbb{R}^n$ and denote by $B(r, p) \subset \mathbb{R}^3$ the ball of radius $r > 0$ about p. Then for an arbitrary packing $\Omega, \mathcal{P}$ we define

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

It is known that when the limit exists, it exists and is equal for all $p \in \mathbb{R}^n$. Otherwise, the density defined above is referred to as the *upper density*.

It is not clear a priori that we lose generality by considering only lattice sphere packings. In fact, the crystallographer Barlow showed how to construct an uncountable family of sphere packings in $\mathbb{R}^3$ whose density is that of the cannonball packing, the densest (and unique) lattice packing known, due to Thue (Zong).

On the other hand, it is known (Groemer) that there exist optimally (upper) dense sphere packings for which the limit defining Δ exists *uniformly* for all $p \in \mathbb{R}^n$. This seems to justify using the upper density to characterize the global density of a packing.

We define the density bound Δ to be

$$\Delta = \sup_{\text{Packings } \mathcal{P}} \Delta(\mathcal{P})$$

where the supremum is over all sphere packings $\mathcal{P}$.

One reduction we can use: for the purposes of bounding Δ, it is sufficient to consider periodic packings, since (Cohn-Elkies, 2003) Δ is obtained as the limit of a sequence of densities of periodic packings.

2. History of the Sphere Packing Problem

The following is a brief timeline of the significant developments in the sphere packing problem.

- 1611 - Kepler conjectures that the most space-efficient way of packing spheres into $\mathbb{R}^3$ is the cannonball, Kepler or face-centered cubic packing, formed by repeating the tetrahedral cell throughout $\mathbb{R}^3$.
- 1773 - By studying extremal quadratic forms, Lagrange was able to deduce that the hexagonal packing is optimal among lattice packings of S^2.
- 1831 - Gauss uses similar methods to deduce that the cannonball packing is the densest among lattice packings of spheres.

- 1900 - Hilbert makes the Kepler Conjecture the third part of his 18th problem.
- 1953 - Toth suggests a method of proving the Kepler conjecture by checking a finite number of cases, i.e. proof-by-exhaustion.
- 1940 - 1960 - In his groundbreaking work on formal communication theory, Shannon discovers the connection between designing codes to optimize error-free information transfer rates in a Gaussian white noise band-limited channel and finding the densest sphere packings.
- 1993 - Hsiang announces he has proven the Kepler conjecture using the program of Toth and a computer; later the proof is discovered to contain errors.
- 1997 - Hales fulfills the program of Toth and announces his proof of the Kepler conjecture. A panel of mathematicians (including Toth himself) later asserts they are "99% certain" that the proof is correct, as it is impossible to check all 3-GB of computations required by hand.

3. Applications of the Problem to Computer Science

There are two applications of the sphere packing problem to this field.

3.1. Design of Codes for Band-limited Gaussian white noise Channels.

Let $T > 0$ be a fixed length of time corresponding to the length of a signal transmission.

Definition 2. A *signal* is a continuous map $f : [0, T] \to \mathbb{R}$ such that the frequencies of the signal in f do not surpass a fixed 'limit' W. Precisely, the Fourier transform $\hat{f}$ has compact support contained in $[-W, W] \subset \mathbb{R}$. A *code* is a finite collection of signals $\{f_1, \cdots, f_M\}$.

We shall adopt the convention of writing $f : \mathbb{R} \to \mathbb{R}$ by extending f by the zero function to the rest of $\mathbb{R}$.

The channel is the physical conduit through which signals are sent from one computer (the source) to another (the destination). Digital data from the Source is encoded using some finite set of signals, the *code*. This can be thought of as a is a symbolic alphabet for the two computers to communicate with via the channel. A simple example of such an alphabet is $\{0, 1\}$, the binary code.

The Shannon-Nyquist Sampling Theorem asserts that a bandlimited signal f is uniquely characterized by a finite set of its samples

$$S(f) = \{f(0), f(\frac{1}{2W}), \cdots, f(\frac{n-1}{2W})\}$$

where $n = 2TW$. It is natural to identify $S(f)$ as a vector in $\mathbb{R}^n$.

Indeed, we have the ability to reconstruct f entirely from $S(f)$ by the Cardinal Series:

$$f(t) = \sum_{k \in \mathbb{Z}} f(\frac{k}{2W}) \frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)}$$

The function in the summand is called 'sinc' by the electrical engineers, and satisfies the orthogonality relation

$$\int_{t \in \mathbb{R}} \frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)} \frac{\sin 2\pi W(t - l/2W)}{2\pi W(t - l/2W)} dt = \frac{\delta_{kl}}{2W}$$

We see that a code for a band-limited channel amounts to a finite subset of $\mathbb{R}^n$. Really, $\mathbb{R}^n$ is a configuration space for the data received at the end of the channel by the decoder, which samples the signal, plots it in this configuration space and determines which letter best 'fits' the signal being uttered in the channel.

We say that the channel has Gaussian white noise when we model the noise (unintentional signals arising from the environment) in the channel modeled by n random variables (i.e. a random vector) in Gaussian distribution with mean zero and variance σ^2, one each indicating the components of the vector Y corresponding to the noise's signal.

Suppose the source sends a signal f through the channel corresponding to the point $F \in \mathbb{R}^n$, and in doing so incurs the Gaussian white noise signal represented by the point $Y \in \mathbb{R}^n$. Then the received signal is $F + Y$, and an error may occur in the data transmission if and only if the resulting message $F + Y$ happens to be closer to another signal $F' \neq F$ in the code.

Naive intuition would suggest separating the $\mathbb{R}^n$ positions of the signals as much as possible. However, consider the average power

$$P = \frac{1}{T} \int_0^T |f(t)|^2 dt$$

exerted in delivering the signal. We see that by the orthogonality relation,

$$F \cdot F = |F|^2 = \sum_{k \in \mathbb{Z}} \left(f(\frac{k}{2W}) \right)^2 = 2W \int_0^T |f(t)|^2 dt = 2WTP = nP$$

Thus, the more we separate the signals F, the more power is exerted and the less efficient the code! An 'optimal' choice of channel code must somehow balance these two competing effects.

Recall that each of the components Y_j of Y is a Gaussian random variable, and so $\mathbb{P}(-2\sigma \leq Y_j \leq 2\sigma)$ is close to one, so that we conclude

$$\mathbb{P}(Y \cdot Y = |Y|^2 \leq 4n\sigma^2)$$

is very close to one. So, to make sure that two signals $F, G \in \mathbb{R}^n$ are not mistaken for one-another with high probability, we see that sufficient to separate them by a distance

$$d = 4\sqrt{n\sigma^2}$$

since we are then guaranteed almost surely that $|Y| < d$. Geometrically, this means that we want to consider designing codes as a kind of sphere-packing problem: we're insulating the points F, G with balls of radius $r = d/2$.

Suddenly, the problem of how to design a code for a band-limited channel is equivalent to the problem of packing spheres of radius $r = d/2 = 2\sqrt{n\sigma^2}$ such that all the centers lie within $\sqrt{nP}$ of the origin.

Already, there are modems on the market with codes designed from the lattice $E_8 \subset \mathbb{R}^8$, which is conjectured (and most certainly *is*, although there is no formal proof) to be the densest packing in $\mathbb{R}^8$.

3.2. **Design of Optimal Quantizers.** When apparatus obtains some kind of empirical or real-world datum which is then converted into a digital datum, the *quantizer* is the device that selects the digital representative of the analogue input datum.

Suppose that $x \in \mathbb{R}^n$ is an analogue datum to be converted to a digital one. The quantizer is an algorithm which selects some point p on a fixed discrete lattice $\mathcal{P} \subset \mathbb{R}^n$ that most accurately reflects the actual value of x, i.e. x has been quantized to $p \in \mathcal{P}$. All commercial quantizers are implemented with $n = 1$, but it is known that higher-dimensional quantizers are theoretically more accurate (on average).

A natural question is to ask for the statistics of a particular quantizer algorithm as a function of $\mathcal{P}$. For instance, what is the expectation of the second moment (the variance) of the error incurred by quantizing?

We need the notion of a *Voronoi Cell* in $\mathbb{R}^n$: Given some lattice or discrete subset $\mathcal{P}$, the Voronoi Cell $V(p)$ at a point $p \in \mathcal{P}$ is the set

$$V(p) = \{x \in \mathbb{R}^n \mid |x - q| \geq |x - p| \; \forall q \in \Lambda\}$$

It is known that for any $\mathcal{P}$,

$$\mathbb{R}^n - \bigcup_{p \in \mathcal{P}} V(p)$$

is a set of zero measure.

One useful statistic is the normalized mean squared error per symbol,

$$G(\Lambda) = \frac{1}{n} |\det \Lambda|^{-\frac{n+2}{n}} \int_{V(0)} x \cdot x \, dx$$

which quantifies how 'accurate' a generic quantization is. Interestingly,

$$G(E_8) = \frac{929}{12960} = 0.0716821 \ldots$$

is a far better quantizer according to this measure than is $\mathbb{Z}^3$, the cube packing, for which $G(\mathbb{Z}^3) = \frac{1}{12} = 0.08333 \ldots$.

The problem of designing an efficient quantizer for dimensions $n \geq 1$ has led to the result that for low dimensions, the best (most efficient at minimizing the possible error of the digitalization) quantizing lattices are the duals of the densest sphere packing lattices, where we recall that for a lattice $\Lambda \subset \mathbb{R}^n$, its dual is defined to be

$$\Lambda^* = \{x \mid \langle x, y \rangle \in \mathbb{Z} \; \forall y \in \Lambda\} \subset \mathbb{R}^n$$

where $\langle , \rangle$ is the Euclidean inner product.

4. Generating Upper Bounds on the Optimal Density

It is clear that a lower bound on the optimal density of sphere packings in $\mathbb{R}^n$, or any space for that matter, can be found by constructively proving the density of a concrete packing $\Delta(\mathcal{P})$. The situation is trickier with an upper bound on Δ; in this section we'll discuss one method for generating upper bounds on the density of sphere packings.

4.1. **The Cohn-Elkies Upper Bound Theorem.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a 'radial' function, i.e. f is constant on balls of constant radius centered at the origin. We write $f(r)$ for the common value of f on the ball of radius r.

Then, either by averaging over the rotation group $O(n)$ or by using the radial Laplacian operator for $\mathbb{R}^n$, it is possible to derive a so-called 'radial' fourier transform,

$$\hat{f}(t) = 2\pi |t|^{-\alpha} \int_0^\infty f(r) J_\alpha(2\pi r |t|) r^{n/2} dr$$

where $\alpha = n/2 - 1$ and J_α denotes the Bessel function of order α.

In addition, we shall assume that f and $\hat{f}$ are of decay sufficient for the Poisson Summation Formula to hold, i.e. for all $v \in \mathbb{R}^n$,

$$\sum_{x\in\Lambda} f(x+v) = \frac{1}{|\Lambda|}\sum_{t\in\Lambda^*} e^{-2\pi i\langle v,t\rangle}\hat{f}(t)$$

where Λ is any lattice, as defined earlier, and Λ^* represents the dual lattice. We shall now sketch a proof of the following:

Theorem 4.1. *Suppose that a function $f : \mathbb{R}^n \to \mathbb{R}$ is radial, is not identically zero, and satisfies*

- $f(r) \leq 0$ *when* $r \geq 1$
- $\hat{f}(t) \geq 0$ *for all* $t \in \mathbb{R}$

Then the center density δ of any *n-dimensional sphere packing satisfies*

$$\delta \leq \frac{f(0)}{2^n \hat{f}(0)}$$

Proof. It suffices to consider only periodic packings $\mathcal{P}$ with lattice Λ and translation vectors $\{v_1, \cdots, v_N\}$. Then recall that the center density is

$$\delta(\mathcal{P}) = \frac{N}{2^n|\Lambda|}$$

Without loss, select the scale of Λ so that spheres have radius $R = 1/2$. From Poisson Summation, we have

$$\sum_{1\leq j,k\leq N}\sum_{x\in\Lambda} f(x+v_j-v_k) = \frac{1}{|\Lambda|}\sum_{t\in\Lambda^*}\hat{f}(t)\sum_{1\leq j,k\leq N} e^{-2\pi i\langle v_j - v_k, t\rangle} = \frac{1}{|\Lambda|}\sum_{t\in\Lambda^*}\hat{f}(t)\left|\sum_{1\leq j\leq N} e^{2\pi i\langle v_j,t\rangle}\right|^2$$

Each term in the sum on the RHS over Λ^* is positive, and so the whole is bounded from below by the $t = 0$ term $\hat{f}(0)N^2/|\Lambda|$.

For the LHS, we see that $|x+v_j-v_k| < 1$ iff $x = 0$ and $j = k$, so that from the hypotheses these are the only positive terms arising in the left hand sum. Therefore we obtain the upper bound $Nf(0)$. Collecting these bounds and recalling the definition of $\Delta(\mathcal{P})$ and $\delta(\mathcal{P})$, we have obtained the conclusion of the theorem. □

4.2. **Results for Packings in $\mathbb{R}^n$.** Cohn-Elkies were able to use the upper-bound theorem to re-derive the second-best known bounds for sphere packing densities at high dimensions, due to Levenshtein:

$$\Delta \leq \frac{j_{n/2}^n}{(n/2)!^2 4^n}$$

where j_t is the first positive root of the Bessel function J_t.

Using linear programming methods, they were able to compute the best-known upper bounds on the densities of packings in 8 and 24 dimensions. These are known to be sharp, since in these dimensions there are already two beautiful lattices, E_8 and the Leech lattice, which are all but certainly the optimal packing lattices in their respective dimensions, and which nearly match the Cohn-Elkies upper bound.

5. Generalizing the Problem to Hyperbolic Space

Let X be a metric space, equipped with a metric $d : X \times X \to \mathbb{R}_{\geq 0}$. It is a natural generalization of the sphere packing problem in $\mathbb{R}^n$ to consider the same question of "spheres" defined to be the loci

$$\{y \in X \mid d(x, y) = R\}$$

for some $x \in X$, the "center", and a fixed $R > 0$.

5.1. The Radial Hyperbolic Fourier Transform and Upper Bounds for the Optimal Density.

Let $X = \mathbb{H}^n$ be n-dimensional hyperbolic space. The radial laplacian in this space, an operator acting on functions of the hyperbolic radius r (hereafter called the 'spatial parameter'), is given by

$$Lu = \frac{\partial^2 u}{\partial r^2} + (n-1)\coth r \frac{\partial u}{\partial r}$$

where $u = u(r)$ is used to denote the common value of a radial function $u : \mathbb{H}^n \to \mathbb{C}$ on hyperbolic balls of radius $r \in \mathbb{R}^+$. We have the eigenvalue problem

$$Lu = -\big((\frac{n-1}{2})^2 + t^2\big)u$$

The parameter t is called the 'spectral parameter.' We can treat this as a self-adjoint Sturm-Liouville ODE in the spatial parameter, and with the initial conditions $u(0) = 1$ and $u'(0) = 0$, we have specified a unique solution which we shall denote $u = x_n(r, t)$.

The functions $x_n(r, t)$ are the even eigenfunctions of the laplacian, and define the radial hyperbolic fourier transform. It is here that we stop considering even n, since there are no closed forms for the eigenfunctions x_{2k}, but there are for x_{2k+1}. The first few are

$$x_1(r, t) = \cos rt$$
$$x_3(r, t) = \frac{\sin rt}{t \sinh r}$$

and subsequent eigenfunctions can be determined from the simple raising operator relation

$$x_{n+2}(r, t) = \frac{-n}{(t^2 + (\frac{n-1}{2})^2)\sinh r} \frac{\partial}{\partial r} x_n(r, t)$$

Recall that the case $n = 1$ corresponds to the cosine Fourier transform on $\mathbb{R}_+$, so for brevity I'll only display the $\mathbb{H}^3$ transform and its inversion:

$$\hat{f}_3(t) = \mathcal{F}_3[f](t) = \int\limits_{\mathbb{R}_{\geq 0}} f(r)x_3(r, t)\sinh^2 r \, dr$$
$$f(r) = \mathcal{F}_3^{-1}[\hat{f}_3](r) = \frac{1}{\pi}\int\limits_{\mathbb{R}} \hat{f}_3(t)x_3(r, t)t^2 dt$$

In a derivation completely analogous to that of Cohn-Elkies, Kerzhner was able to prove the extension of their theorem to the hyperbolic situation.

Theorem 5.1. *Let $f : \mathbb{H}^n \to \mathbb{C}$ be radial, and fix $R > 0$. Suppose that*

- $\hat{f}(t) \geq 0$ *for all* t
- $f(r) \leq 0$ *for all* $r \geq R$

Then the density of all periodic packings of $\mathbb{H}^n$ of radius R obeys the bound

$$\Delta \leq \frac{\int_0^{R/2} \sinh^{n-1} r \, dr \cdot f(0)}{\hat{f}(\frac{n-1}{2}i)}$$

Notice that unlike in the Euclidean case, the 'scale' R of the spheres is very much important.

5.2. **Results in $\mathbb{H}^2$ and $\mathbb{H}^3$.** We were able to derive a few results concerning sphere packings for small R in low dimensions.

For $n = 3$ and small radius, Kerzhner was able to prove that for asymptotically small radius R, the optimal density of sphere packings comes arbitrarily close to the packing density of the Kepler packing, verifying that 'locally' (small radius) one recovers the Euclidean packing structure.

For $R = 1$, $n = 3$, we were able to prove bounds using a few different functions. One of our functions was basically a generalization of an example used in the paper of Cohn-Elkies:

$$f_G(r) = (1 - \frac{r^2}{R^2})\left(\frac{\Gamma^2(x)}{\Gamma(x - yr)\Gamma(x + yr)}\right)^2 \frac{r}{\sinh r} e^{-\alpha r^2}$$

To determine that this function was positive-definite (has positive $\mathbb{H}^3$- transform) we made use of the analogue of the 'convolution theorem' in $\mathbb{R}^1$:

$$\mathcal{F}_3[fg](t) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{t-u}{t} \hat{f}_3(t-u)\hat{g}_1(u)du \tag{5.1}$$

This was used to show that the possible negative values of the transform $\hat{f}_G$ occur inside a finite interval; a short numerical computation completed the proof. The bound we obtained was close to $\Delta \leq 0.8369$.

Among our other ideas at this scale were two functions: one of them was an extension of another idea used in Cohn-Elkies paper, and involved a calculus of variations argument to optimize the bound obtained over a class of functions satisfying a certain constraint, but the bound obtained was strictly weaker than this one. We also used another function with the heat kernel in $\mathbb{H}^3$,

$$\frac{r}{\sinh r} e^{-\alpha r^2}$$

multiplied against a many-termed polynomial. The bounds obtained however were not quite as rigorous, since the problem of determining the positivity of a real polynomial is in general intractable.

We were dismayed to learn that it is possible to prove a Sobolev-type inequality on the weight space with weight function $w = w(t) = t^2$ which implies

Lemma 5.1. *There exists a finite $R > 0$ such that for any function of the form $f(r) = (R^2 - r^2)g^2(r)$ which is positive-definite in $\mathbb{H}^3$, the bound obtained on the sphere packing density is trivial.*

5.3. **Bounds for $\mathbb{H}^n$.** Using the function

$$f_X(r) = \frac{x_{n+2}^2(r, b_R)}{\sinh^2 R - \sinh^2 r}$$

where b_R is the first positive real root of the nonlinear equation $x_{n+2}(R, b_R) = 0$, we were able to compute the asymptotic ($n \to \infty$) bound

$$\Delta \leq \frac{2P_n(b_R) \int\limits_0^{R/2} \sinh^{n-1} r \, dr}{(n-2)!!^2 b_R}$$

where $P_n(t) := t(t^2+1)(t^2+4)(\cdots)(t^2+(\frac{n-1}{2})^2)$ and $n!!$ denotes the rising factorial (e.g. $5!! = 5 \cdot 3 \cdot 1$). This bound is valid in any odd dimension, and gives nontrivial bounds only for very small radii R. However, notice that the bound relies on an estimate of the roots of the eigenfunction to the laplacian, so that these bounds could be seen as an analogue of the Levenshtein bounds in $\mathbb{H}^n$. To make this a concrete bound, we employ the estimate

$$b_R \leq \frac{\pi(n+1)}{2R}$$

which is obtained from elementary estimates on the locations of the roots and a simple observation regarding the raising operator mentioned above.