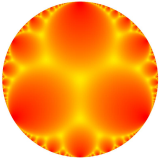
Sphere packing

Henry Cohn

IAP Math Lecture Series
January 16, 2015



The sphere packing problem

How densely can we pack identical spheres into space?

Not allowed to overlap (but can be tangent).

Density = fraction of space filled by the spheres.

Why should we care?

The densest packing is pretty obvious.
It's not difficult to stack cannonballs or oranges.

It's *profoundly difficult* to prove (Hales 1998, 2014).
But why should anyone but mathematicians care?

One answer is that it's a toy model for:

- Granular materials.
- Packing more complicated shapes into containers.

Sphere packing is a first step towards these more complex problems.

Varying the dimension

What if we didn't work in three-dimensional space?

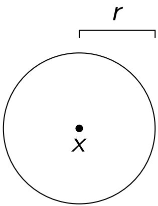
The two-dimensional analogue is packing circles in the plane.

Still tricky to prove, but not nearly as difficult (Thue 1892).

What about one dimension? What's a one-dimensional sphere?

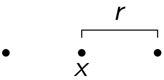
Spheres in different dimensions

Sphere centered at x with radius r means the points at distance r from x.

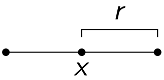


Ordinary sphere in three dimensions, circle in two dimensions.

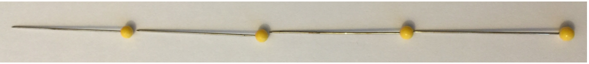
Just two points in one dimension:



The inside of a one-dimensional sphere is an interval.



One-dimensional sphere packing is boring:



(density $= 1$)

Two-dimensional sphere packing is prettier and more interesting:

(density ≈ 0.91)

Three dimensions strains human ability to prove:

(density ≈ 0.74)

What about *four dimensions*? Is that just crazy?

Some history

Thomas Harriot (1560–1621)

Mathematical assistant to Sir Walter Raleigh.

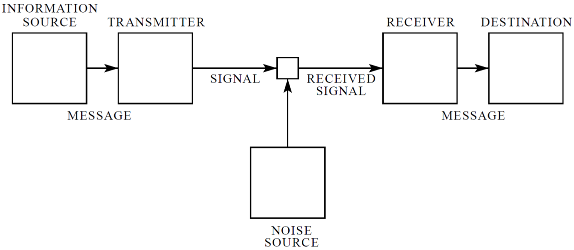
A Brief and True Report of the New Found Land of Virginia (1588)

First to study the sphere packing problem.

Claude Shannon (1916–2001)

Developed information theory.

A Mathematical Theory of Communication (1948)



Practical importance of sphere packing in higher dimensions!
We'll return to this later.

Distances and volumes in $\mathbb{R}^n$

The distance between (a, b, c, d) and (w, x, y, z) is

$$\sqrt{(a-w)^2+(b-x)^2+(c-y)^2+(d-z)^2}.$$

Just like two or three dimensions, but with an extra coordinate.
(n-dimensional Pythagorean theorem)

4d volume of right-angled $a \times b \times c \times d$ box $=$

product $abcd$ of lengths in each dimension.

Just like area of a rectangle or volume of a 3d box.

Higher dimensions work analogously. Just use all the coordinates.

Why should we measure things this way? We don't have to.

We could measure distances and volumes however we like, to get *different geometries* that are useful for different purposes.

But we're going to focus on Euclidean geometry today.

Applications

Anything you can measure using n numbers is a point in n dimensions.

Your height, weight, and age form a point in $\mathbb{R}^3$.

Twenty measurements in an experiment yield a point in $\mathbb{R}^{20}$.

One pixel in an image is described by a point in $\mathbb{R}^3$ (red, green, and blue components). A one-megapixel image is a point in $\mathbb{R}^{3000000}$.

Some climate models have ten billion variables, so their states are points in $\mathbb{R}^{10000000000}$.

High dimensions are everywhere! Low dimensions are anomalous. All data can be described by numbers, so any large collection of data is high dimensional.

Big data

Classical statistics originated in an information-poor environment.

Goal: extract as much information as possible from limited data.
Put up with a high ratio of computation and thought to data.
Only valuable data is collected, to answer specific questions.

Nowadays massive data sets are common. We often collect vast quantities of data without knowing exactly what we are looking for.

How can we find a needle in a high-dimensional haystack?

Curse of dimensionality

Volume scales like the nth power in n dimensions:
rescaling by a factor of r multiplies volume by r^n.

Exponential growth as the dimension varies!

Volume of $2 \times 2 \times \cdots \times 2$ cube in $\mathbb{R}^n$ is 2^n.

$$2^{10} = 1024$$

$$2^{100} = 1267650600228229401496703205376$$

$$\begin{aligned}2^{1000} = \ & 10715086071862673209484250490600018105614048117055336074437503883703510511249361224931983788156958581275946729175531468251871452856923140435984577574698574803934567774824230985421074605062371141877954182153046474983581941267398767559165543946077062914571196477686542167660429831652624386837205668069376\end{aligned}$$

Why is exponential volume growth a curse?

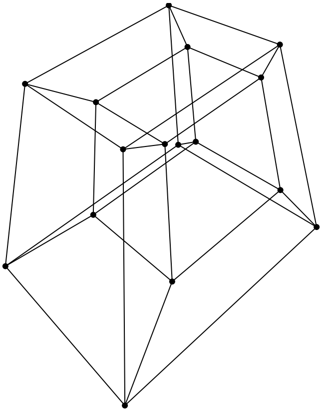
Exponentials grow unfathomably quickly, so there's a ridiculous amount of space in high dimensions.

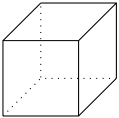
Naively searching a high-dimensional space is incredibly slow. Nobody can search a hundred-dimensional space by brute force, let alone a billion-dimensional space.

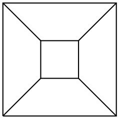
To get anywhere in practice, we need better algorithms. But that's another story for another time. . .

How can we visualize four dimensions?

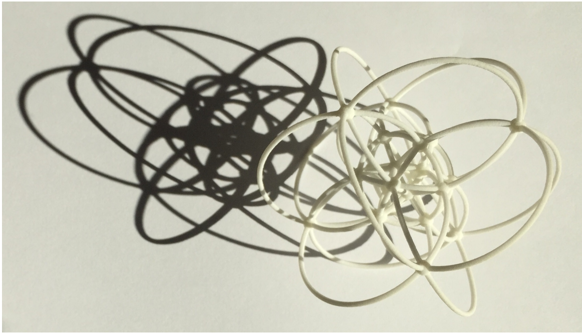
Perspective picture of a hypercube.





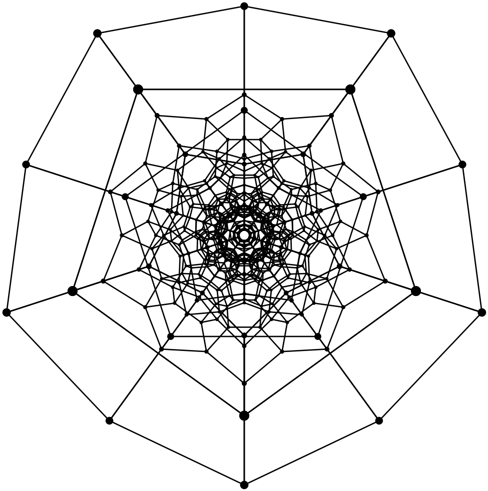


Shadow cast in a lower dimension ("projection").



Fundamentally the same as a perspective picture.

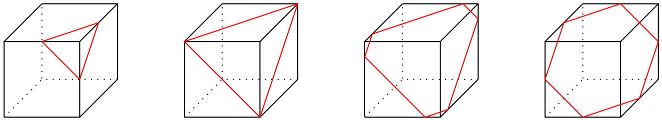
Projections can get complicated.



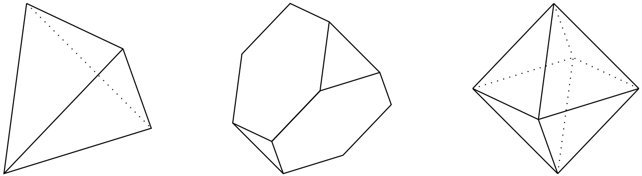
Cross sections

Cubes have square cross sections.
Hypercubes have cubic cross sections.

Cubes have other cross sections too:



So do hypercubes:

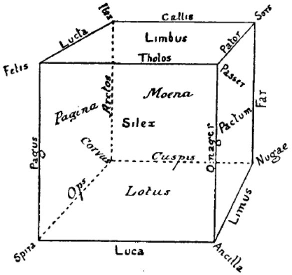


Hinton cubes

Charles Howard Hinton (1853–1907)

A New Era of Thought, Swan Sonnenschein & Co., London, 1888

MODEL 7. PLUVIUM.



COLOURS: PLUVIUM, DARK-STONE.

Points: Spira, Silver. Ancilla, Turquoise. Nugæ, Fawn. Corvus, Gold. Felis, Quaker-green. Passer, Peacock-blue. Sors, Dull-purple. Ilex, Light-blue.

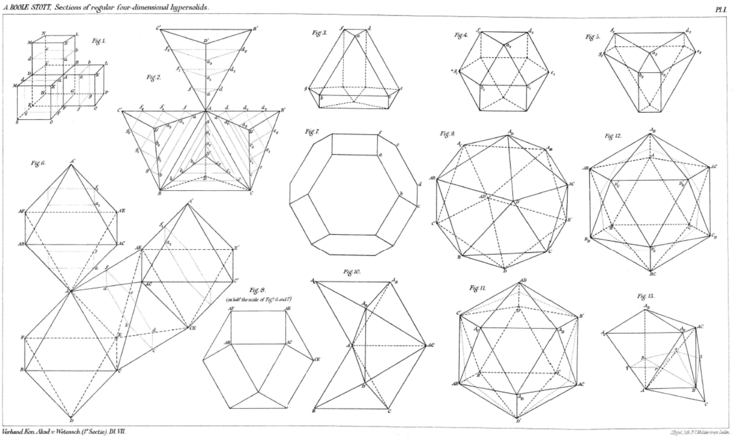
Lines: Luca, Leaf-green. Limus, Smoke. Cuspis, Orange. Ops, Stone. Pagus, Dull-blue. Onager, Dark-pink. Far, French-grey. Arctos, Brown. Tholos, Brown-green. Pator, Green-blue. Callis, Reddish. Lucta, Rich-red.

Surfaces: Silex, Burnt-Sienna. Pactum, Yellow-green. Moena, Dark-blue. Pagina, Yellow. Limbus, Ochre. Lotus, Azure.

(Hinton introduced the term "tesseract" and invented the automatic pitching machine.)

Alicia Boole Stott (1860–1940)

On certain series of sections of the regular four-dimensional hypersolids, Verhandelingen der Koninklijke Akademie van Wetenschappen te Amsterdam **7** (1900), no. 3, 1–21.



Information theory

Shannon's application of high-dimensional sphere packings.

Represent signals by points s in $\mathbb{R}^n$.

E.g., radio with coordinates = amplitudes at different frequencies.
In most applications, n will be large. Often hundreds or thousands.

Send stream of signals over this channel.

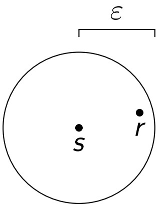
Noise in the communication channel

The key difficulty in communication is noise.

Send signal s in $\mathbb{R}^n$; receive r on other end.
Noise means generally $r \neq s$.

The channel has some *noise level* ε, and r is almost always within distance ε of s.

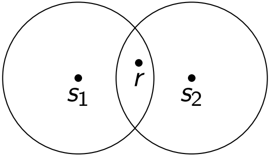
Imagine an *error sphere* of radius ε about each sent signal, showing how it could be received.



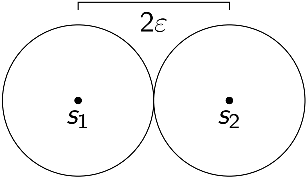
How can we communicate without error?

Agree ahead of time on a restricted vocabulary of signals.

If s_1 and s_2 get too close, received signals could get confused:



Did r come from s_1 or s_2? Therefore, keep all signals in S at least 2ε apart, so the error spheres don't overlap:



This is sphere packing!

The error spheres should form a sphere packing. This is called an *error-correcting code*.

For rapid communication, want as large a vocabulary as possible. I.e., to use space efficiently, want to maximize the packing density.

Rapid, error-free communication requires a dense sphere packing. Real-world channels correspond to high dimensions.

Of course some channels require more elaborate noise models, but sphere packing is the most fundamental case.

What is known?

Each dimension seems to behave differently.

Good constructions are known for low dimensions.

No idea what the best high-dimensional packings look like (they may even be disordered).

Upper/lower density bounds in general.

Bounds are very far apart:
  For $n = 36$, differ by a *multiplicative factor of 58*.
  This factor grows exponentially as $n \to \infty$.

Packing in high dimensions

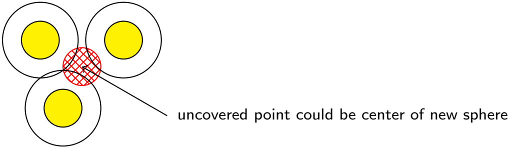
On a scale from one to infinity, a million is small, but we know almost nothing about sphere packing in a million dimensions.

Simple lower bound: can achieve density at least 2^{-n} in $\mathbb{R}^n$.

How to get density at least 2^{-n}

Consider any packing in $\mathbb{R}^n$ with spheres of radius r, such that no further spheres can be added without overlap.

No point in $\mathbb{R}^n$ can be $2r$ units away from all sphere centers. I.e., radius $2r$ spheres cover space completely.



Doubling the radius multiplies the volume by 2^n.

Thus, the radius r packing has density at least 2^{-n} (since the radius $2r$ packing covers all of space). Q.E.D.

This is very nearly all we know!

Nonconstructive proof

That proof showed that the best packing in $\mathbb{R}^n$ must have density at least 2^{-n}.

But it didn't describe where to put the spheres in an actual packing. It's a "nonconstructive proof."

In fact, nobody has any idea how to build such a good packing. Our constructions are all much worse.

For example, using a rectangular grid gives density

$$\frac{\pi^{n/2}}{(n/2)!2^n}$$

(this is not obvious), and the factorial ruins everything.

Packing in high dimensions (large n)

We just showed:
density at least 2^{-n}.

Hermann Minkowski (1905):
at least $2 \cdot 2^{-n}$.

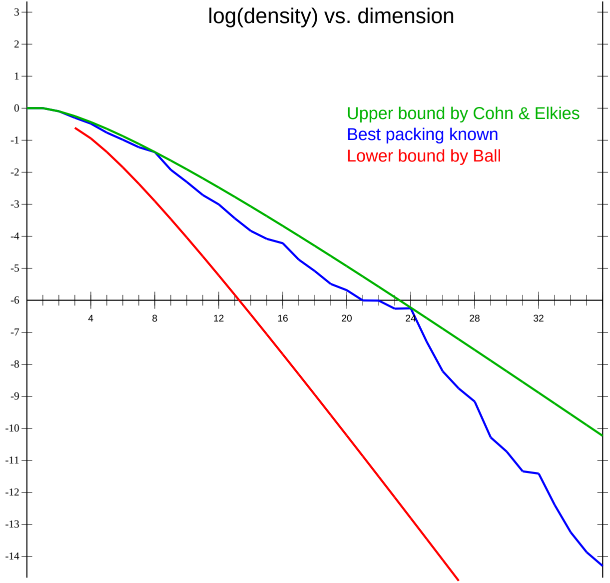
...[many further research papers]...

Keith Ball (1992):
at least $2n \cdot 2^{-n}$.

Stephanie Vance (2011):
at least $\frac{6}{e} n \cdot 2^{-n}$ when n is divisible by 4.

Akshay Venkatesh (2012):
at least $\frac{e^{-\gamma}}{2} n \log\log n \cdot 2^{-n}$ for certain sparse sequence of dimensions.
First superlinear improvement on 2^{-n}!

For comparison, the best upper bound known is $2^{-0.599n}$, due to Grigorii Kabatiansky and Vladimir Levenshtein (1978).


log(density) vs. dimension
Upper bound by Cohn & Elkies
Best packing known
Lower bound by Ball
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
-11
-12
-13
-14
4
8
12
16
20
24
28
32

The most remarkable packings

Certain dimensions have amazing packings.

$\mathbb{R}^8$: E_8 root lattice
$\mathbb{R}^{24}$: Leech lattice [named after John Leech (1926–1992)]

Extremely symmetrical and dense packings of spheres.
Must be optimal, but this has not been proved.

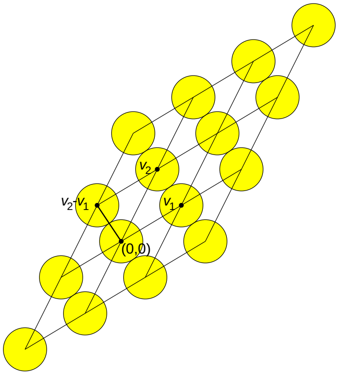
Connected with many areas in mathematics, physics
(such as string theory, modular forms, finite simple groups).

What makes these cases work out so well?

How could we pack spheres?

Lattice: integer span of n linearly independent vectors. I.e., for basis $v_1, \ldots, v_n$, center spheres at

$$\{a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \mid a_1, \ldots, a_n \in \mathbb{Z}\}.$$



Packing radius = half shortest non-zero vector length.

Periodic packing

Union of finitely many translates of a lattice.

Spheres are not restricted to just the corners of a fundamental domain.

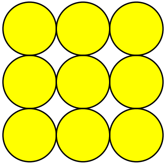
No reason to believe densest packing must be periodic, but periodic packings come arbitrarily close to the maximum density.

By contrast, lattices probably do not.

Quadratically many parameters don't give enough flexibility to fill all the gaps in an exponential amount of space.

The best sphere packings currently known are **not** always lattice packings, but many good packings are.

Simplest lattice: $\mathbb{Z}^n$, lousy packing.



Better: for $n \geq 3$, "checkerboard" packing

$$D_n = \{(x_1, \ldots, x_n) \in \mathbb{Z}^n \mid x_1 + \cdots + x_n \text{ is even}\}.$$

D_3, D_4, D_5 are best known packings in those dimensions, and provably best lattice packings.

What about $n \geq 6$?

Holes in D_n grow greater and greater.

A hole is a local maximum for distance from nearest lattice point.

What about $n \geq 6$?

Holes in D_n grow greater and greater.

A hole is a local maximum for distance from nearest lattice point.

Two classes of holes in D_n (for $n \geq 3$):

- $(1, 0, \ldots, 0)$ at distance 1 from lattice.
- $(1/2, 1/2, \ldots, 1/2)$ at distance

$$\sqrt{\left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^2} = \sqrt{\frac{n}{4}}.$$

How big are the spheres in a D_n packing?

Packing radius for D_n: nearest points are

$$(0, 0, \ldots, 0) \qquad \text{and} \qquad (1, 1, 0, \ldots, 0)$$

at distance $\sqrt{2}$, so use spheres of radius $\sqrt{2}/2$.

Wonderful properties of dimension 8

When $n = 8$, radius $\sqrt{n/4}$ of deep hole equals distance $\sqrt{2}$ between lattice points.

We can slip another copy of D_8 into the holes! This doubles the packing density.

Result called E_8 lattice.

E_6, E_7 are certain cross sections of E_8.

These are densest lattice packings for $n = 6, 7, 8$. Almost certainly densest packings.

Leech lattice ($n = 24$) is similar in spirit, but more complicated.

Why is the sphere packing problem difficult?

Many local maxima for density.

Lots of space.

Complicated geometrical configurations.

No good way to rule out implausible configurations rigorously.

High dimensions are weird.

Visualization, take two

There are many ways to try to visualize four dimensions.
But they become incomprehensible as the dimension increases.

Is there any way we can conceive of a million dimensions?
What could it look like?

Let's start by thinking about cubes, because they are easy to pack.

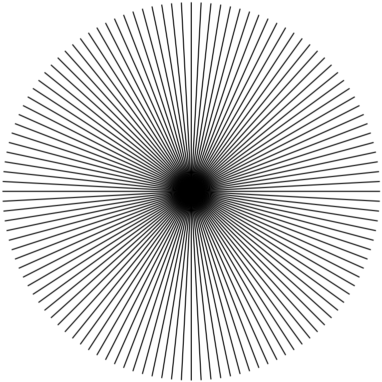
Even the cubes go crazy

Let's look at the cube $\{(x_1, \ldots, x_n) \mid |x_i| \leq 1 \text{ for all } i\}$.

2^n vertices $(\pm 1, \ldots, \pm 1)$.

Distance from center to vertex is $\sqrt{1^2 + \cdots + 1^2} = \sqrt{n}$.

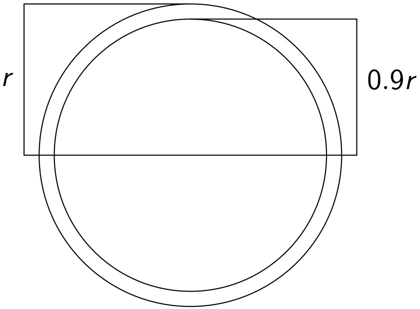
Imagine $n = 10^6$. How can this be?



Concentration of volume

Almost all volume in high-dimensional solids is concentrated near the boundary.

Why? Imagine a sphere. Fraction of volume less than 90% of way to boundary is 0.9^n. This tends rapidly to zero as $n \to \infty$.
For example, $0.9^{1000} \approx 1.75 \cdot 10^{-46}$.



We can replace 0.9 with 0.99, 0.999, or whatever we want.

[Aside for those who know calculus:
As $n \to \infty$ with c fixed, fraction of volume less than $1 - c/n$ of the way to the boundary is $(1 - c/n)^n \to e^{-c}$.]

All low-dimensional intuition must be revised.

Boundaries, where all the interaction takes place, become increasingly important in high dimensions.

Relevant for packing and for analysis of algorithms.

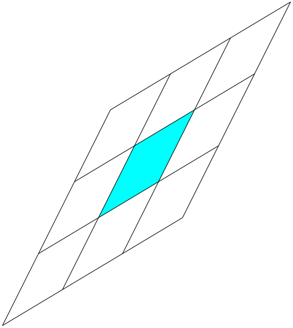
Also the second law of thermodynamics (increase of entropy)!

Hard computational problems

Hard to tell how dense a lattice packing is!

$$\text{density} = (\text{sphere vol.}) \cdot (\#\ \text{spheres/unit vol.})$$

2nd factor $= 1/(\text{vol. of fundamental domain})$



Absolute value of the determinant of the basis.

Hard part is figuring out which radius spheres will fit. How long is the shortest nonzero vector in the lattice?

NP-complete (Ajtai, 1997).

In other words, many search problems can be reduced to it. No proof is known that it cannot be solved efficiently, but that is almost certainly the case.

There are good algorithms that give pretty short vectors, at least in low dimensions. These vectors are short enough for some applications.

Practical applications

Cryptosystems based on the difficulty of finding short lattice vectors, or ones near a given point.

Many direct applications in computational mathematics, and numerical exploration.

Goldreich-Goldwasser-Halevi cryptosystem (1997)

Public key: high dimensional lattice

Private key: secret nearly orthogonal basis, that makes it easy to find the nearest lattice point to any given point

Encode messages as lattice points.

To encrypt, add a small random perturbation. Can decrypt using the private key, but not without it.

This system has weaknesses (Nguyen 1999), but there are other, stronger lattice-based systems.

Recognizing algebraic numbers:

The number

$$\alpha = -7.82646099323767402929927644895$$

is a 30-digit approximation to a root of a 5th-degree equation. Which equation?

Of course, infinitely many answers, but we want the simplest.

Trivial example: recognizing that

$$0.1345345345345345345345345345345$$

is an approximation to $1/10 + 345/9990$.

Let $C = 10^{20}$ (based on precision of α), and look at lattice generated by

$$\begin{aligned} v_0 &= (1,0,0,0,0,0,C), \\ v_1 &= (0,1,0,0,0,0,C\alpha), \\ v_2 &= (0,0,1,0,0,0,C\alpha^2), \\ v_3 &= (0,0,0,1,0,0,C\alpha^3), \\ v_4 &= (0,0,0,0,1,0,C\alpha^4), \\ v_5 &= (0,0,0,0,0,1,C\alpha^5). \end{aligned}$$

Lattice vector ($a_0, \ldots, a_5 \in \mathbb{Z}$):

$$a_0 v_0 + \cdots + a_5 v_5 = \left(a_0, a_1, a_2, a_3, a_4, a_5, C \left(\sum_{i=0}^{5} a_i \alpha^i \right) \right)$$

For this to be small, we want small a_i's and really small $\sum_i a_i \alpha^i$, since C is huge.

We can find a short vector using a computer:

$$(71,\ -5,\ 12,\ -19,\ 13,\ 2,\ 0.000004135\ldots).$$

That tells us that

$$71 - 5\alpha + 12\alpha^2 - 19\alpha^3 + 13\alpha^4 + 2\alpha^5 \approx 0.$$

(More precisely, it is about $0.000004135/C \approx 4 \cdot 10^{-26}$.)
In fact, this is the equation I used to generate α.

How do we get good sphere packings?

There are several potential UROP problems:

Computer searches in low dimensions.

Nonconstructive proofs in high dimensions.

Computer searches

Yoav Kallus, Veit Elser, and Simon Gravel, *Method for dense packing discovery*, Phys. Rev. E **82**, 056707 (2010)

Étienne Marcotte and Salvatore Torquato, *Efficient linear programming algorithm to generate the densest lattice sphere packings*, Phys. Rev. E **87**, 063303 (2013)

Yoav Kallus, *Statistical mechanics of the lattice sphere packing problem*, Phys. Rev. E **87**, 063307 (2013)

These papers recover the densest lattices known up to 20 dimensions.

Can we push the calculations further, to unknown territory?

What about periodic packings?

The Siegel mean value theorem

What does it mean to average over all lattices?

It's not quite obvious, but there is a canonical probability measure on lattices with fixed determinant (i.e., fundamental cell volume). Key concept: $SL_n(\mathbb{R})$-invariance.

Recall that $SL_n(\mathbb{R})$ is the group of all $n \times n$ real matrices with determinant 1. They act on $\mathbb{R}^n$ by left multiplication and preserve volume.

What does the average pair correlation function look like? It measures the average number of neighbors at each distance.

Siegel mean value theorem: it is exactly the same as for a Poisson distribution (uniformly scattered points).

More precisely: given nice $f\colon \mathbb{R}^n \to \mathbb{R}$, the average of

$$\sum_{x \in \Lambda \setminus \{0\}} f(x)$$

over all lattices Λ of determinant 1 equals

$$\int_{\mathbb{R}^n} f(x)\, dx.$$

Why is this true? There is enough symmetry to rule out any other possible answer.

Specifically, by linearity the answer must be $\int f\, d\mu$ for some measure μ on $\mathbb{R}^n \setminus \{0\}$ that is invariant under $SL_n(\mathbb{R})$. This is only one such measure, up to scaling, and some simple consistency checks determine the constant of proportionality.

Meta principle: averaging over all possible structures is the same as having no structure at all.

This is certainly not always true. It generally depends on having a big enough symmetry group.

Breaks down a little for higher correlation functions, but get Poisson statistics modulo obvious restrictions (e.g., three points in a row for three-point correlations).

Why do we care? Sphere packing bounds.

Siegel mean value theorem $\Rightarrow$ density lower bound of $2 \cdot 2^{-n}$.

Let B be a ball of volume 2 centered at the origin.

For a random lattice of determinant 1, the expected number of nonzero lattice points in B is $\mathrm{vol}(B) = 2$.

These lattice points come in pairs, and some lattices have a lot of them. Since the average is 2, some lattices must have none.

No lattice points in B means lattice gives a packing of $B/2$.

Get packing with one copy of $B/2$ per unit volume, density

$$\frac{\mathrm{vol}(B)}{2^n} = 2 \cdot 2^{-n}.$$

Vance's idea

The extra factor of 2 came because lattice vectors occur in pairs of the same length.

What if we impose additional symmetry?

Intuition: the average number of neighbors remains the same, but now they occur in bigger clumps, so the chances of no nearby neighbors go up.

Vance (2011): quaternion algebras

Venkatesh (2012): cyclotomic fields

Is this the best one can do? Only certain symmetry groups work here: we need a big centralizer to get enough invariance for the Siegel mean value theorem proof, and only division algebras will do. Cyclotomic fields are the best division algebras for this purpose.

Other sorts of groups will distort the pair correlation function away from Poisson statistics. But is that good or bad?

Is there anything similar one can do besides imposing symmetry?

The results so far cannot be the end of the story.

In summary

High dimensions are deeply weird. Even four dimensions can be puzzling, and a million dimensions are outlandish.

We really don't know what the best sphere packing in a million dimensions looks like. Our estimate of its density might be off by an exponential factor. We don't even know whether it is ordered (like a crystal) or disordered (like a lump of dirt).

But these problems really matter. Every time you use a cell phone, you can thank Shannon and his relationship between information theory and high-dimensional sphere packing.

For more information

Papers are available from:

`http://research.microsoft.com/~cohn`

Specifically, see *Order and disorder in energy minimization*:

`http://arXiv.org/abs/1003.3053`