Designs: a fun Mix of Analysis, Combinatorics, Graph/Number/Spectral Theory

Stefan Steinerberger

University of Washington, Jan 2020

Yale University

The (Very Vague) Guiding Question

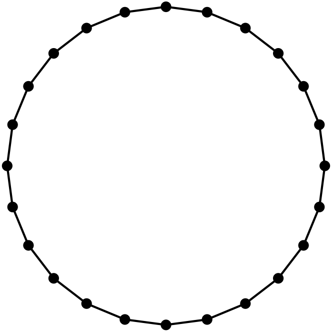
What is the most even distribution of N points on the sphere $\mathbb{S}^d$?

Suppose we want to approximate

$$\frac{1}{|\mathbb{S}^d|}\int_{\mathbb{S}^d} f(x)dx \sim \frac{1}{N}\sum_{n=1}^{N} f(x_i).$$

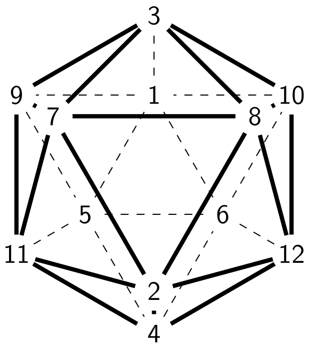
How should we select the points? What if the domain is not $\mathbb{S}^d$?

Minimal Requirements



If we can pick 24 points on $\mathbb{S}^1$, this is how we should do it (admittedly: up to rotation but only up to rotation).

Minimal Requirements



If we can pick 12 points on $\mathbb{S}^2$, this is how we should do it (up to rotation but only up to rotation).

Sobolev-Lebedev Quadrature

Sergei Sobolev (1908–1989)

(Vyacheslav Lebedev, 1930–2010)

Sobolev (1962) gets right to the point

10. Cubature Formulas on the Sphere Invariant under Finite Groups of Rotations*

S. L. Sobolev

A cubature formula on the surface of the sphere

$$(l, f) = \int_S f(\vartheta, \varphi)\, dS - \sum_{k=1}^{N} c_k f(x^{(k)}) \cong 0 \tag{1}$$

is called *invariant* under transformations of a certain group G of sphere rotations if

$$\Big(l, f\big(\vartheta_1(\vartheta, \varphi), \varphi_1(\vartheta, \varphi)\big)\Big) = \Big(l, f(\vartheta, \varphi)\Big), \tag{2}$$

where

$$\vartheta_1(\vartheta, \varphi), \quad \varphi_1(\vartheta, \varphi) \tag{3}$$

is a substitution in G.

Integrate as many low-degree polynomials as possible exactly.

Spherical Designs

Suppose we want to approximate

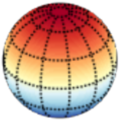
$$\frac{1}{|\mathbb{S}^2|}\int_{\mathbb{S}^2} f(x)dx \sim \frac{1}{N}\sum_{n=1}^{N} f(x_i).$$

How to select the points?

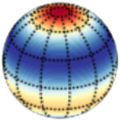
Idea (Sobolev 1962)

Pick the points in such a way that as many spherical harmonics (these are polynomials in $\mathbb{R}^3$ restricted to $\mathbb{S}^2$) as possible are integrated exactly.

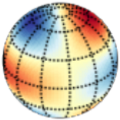
Spherical Harmonics



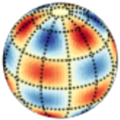
$m = 0,\ n = 1$



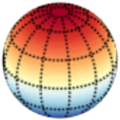
$m = 1,\ n = 1$



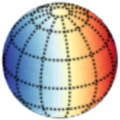
$m = 2,\ n = 2$



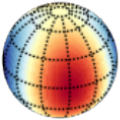
$m = 4,\ n = 5$



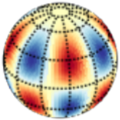
$m = 0,\ n = 2$



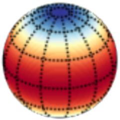
$m = 1,\ n = 2$



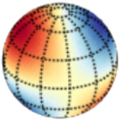
$m = 2,\ n = 3$



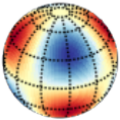
$m = 5,\ n = 7$



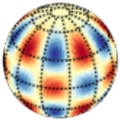
$m = 0,\ n = 3$



$m = 1,\ n = 3$

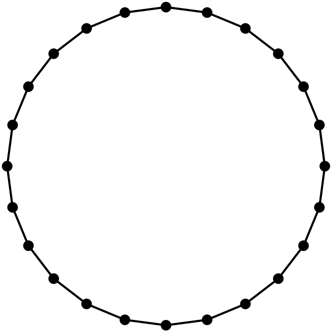


$m = 3,\ n = 6$



$m = 6,\ n = 10$

Returning to our Minimal Requirements



Polynomials in $\mathbb{R}^2$ look like $x^m y^n$. On $\mathbb{S}^1$, they start looking like

$$(\cos\theta)^m(\sin\theta)^n$$

and trigonometric identities simplify this to classical Fourier series $\sin\theta, \cos\theta, \sin 2\theta, \cos 2\theta, \ldots.$

Returning to our Minimal Requirements

Basic Fact

n equispaced points integrate

$$\sin k\theta, \cos k\theta \qquad \text{for all } 1 \leq k \leq n$$

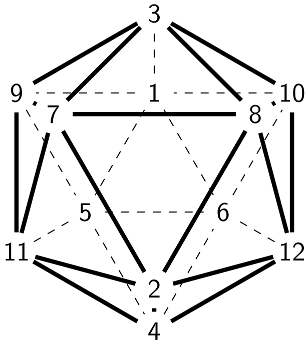
exactly.

Proof.

$$\begin{aligned}\sum_{\ell=0}^{n-1} \cos\left(2\pi\frac{\ell}{n}\right) + i\sin\left(2\pi\frac{\ell}{n}\right) &= \sum_{\ell=0}^{n-1} e^{2\pi i \frac{\ell}{n}} \\ &= \text{sum of roots of unity} \\ &= 0.\end{aligned}$$

$\square$

Returning to our Minimal Requirements



The Dodecahedron has a great degree of symmetry. It integrates all polynomials on $\mathbb{S}^2$ up to degree 5 exactly ($\dim(V) = 36$). (Not quite as basic: this is optimal).

What can I hope for?

I want

$$\frac{1}{|\mathbb{S}^2|}\int_{\mathbb{S}^2} f(x)dx \sim \frac{1}{N}\sum_{n=1}^{N} f(x_i)$$

to be true for many polynomials.

What can we hope for? $2N$ parameters *should* solve $2N$ equations. If we add weights $a_i \in \mathbb{R}$

$$\frac{1}{|\mathbb{S}^2|}\int_{\mathbb{S}^2} f(x)dx \sim \frac{1}{N}\sum_{n=1}^{N} a_i f(x_i)$$

we *should* be able to do $3N$ equations. (*Open problem.*)

Ahrens & Beylkin, Proc. Royal Soc. 2009

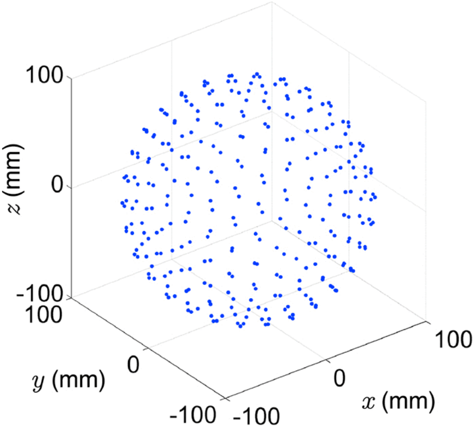


Figure: A set of 302 weighted points on $\mathbb{S}^2$ integrating all polynomials up to degree 29 exactly. Note that $30^2 = 900 \sim 3 \cdot 302$

The No-Miracles Theorem

Definition (Delsarte-Goethals-Seidels, 1977)

A set of points $\{x_1, \ldots, x_n\} \subset \mathbb{S}^d$ is called a spherical $t-$design if

$$\frac{1}{|\mathbb{S}^d|} \int_{\mathbb{S}^d} f(x) dx = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

holds for all polynomials up to degree t. Note that

$$\text{dim(polynomials degree less than } t \text{)} \sim t^d$$

Theorem (No Miracles, Delsarte-Goethals-Seidels, 1977)

If a spherical design integrates all polynomial of degree $\leq t$, then

$$N \gtrsim t^d$$

This field is hard to summarize

An, Andreev, Astola, Bachoc, Bannai, Bojnak, Bondarenko, Boumova, Boyvalenkov, Brauchart, Breger, Cameron, Chen, Coulangeon, Dai, Damelin, Damerell, Danev, De La Harpe, Delsarte, Dhillon, Ding, Dragnev, Dunkl, Ehler, Etayo, Fliege, Frommer, Grabner, Godsil, Goethals, Gorbachev, Gräf, Gross, Haemers, Hamkins, Hardin, Heath, Hirao, Hoggar, Hong, Koike, Krahmer, Korevaar, Kueng, Kuperberg, Kulina, Lang, Lazzarini, Leopardi, Levenshtein, Lovett, Lyubich, Maier, Marzo, Meyers, Mhaskar, Mimura, Munemasa, Nakata, Nebe, Neumaier, Neutsch, Nikov, Nikova, Okuda, Ortego-Cerda, Pache, Peled, Potts, Rabau, Radchenko, Reznick, Roy, Saff, Sawa, Scott, Seidel, Seymour, Shinohara, Sidelnikov, Sloan, Sloane, Suda, Suprijanto, Sustik, Tagami, Tanaka, Tichy, Tiep, Tropp, Vallentin, Venkov, Viazovska, Womersley, Xu, Yudin, Zaslavsky, Zeger, Zhou

Theorem (Delsarte-Goethals-Seidels, 1977)

$$N \gtrsim t^d$$

Theorem (Bondarenko-Radchenko-Viazovska, Annals 2013)

$$N \lesssim t^d \qquad \textit{is always possible.}$$

Summary: Linear Algebra is approximately correct.

$$N \sim \mathrm{dim}(\text{polynomials degree less than } t).$$

Now for something completely different: 'Toroidal Designs'

Same question would be interesting for the two-dimensional Torus

$$\frac{1}{4\pi^2}\int_{\mathbb{T}^2} f(x)dx \sim \frac{1}{N}\sum_{i=1}^{N} f(x_i).$$

Let's replace polynomials by trigonometric functions $e^{i\langle k,x\rangle}$ where $k \in \mathbb{Z}^2$. There is a natural ordering by $\|k\|$.

Question

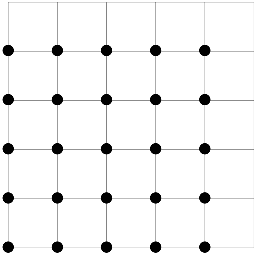
Is the Linear Algebra Heuristic still correct?

'Toroidal Designs'

Same question would be interesting for the two-dimensional Torus

$$\frac{1}{4\pi^2} \int_{\mathbb{T}^2} f(x) dx \sim \frac{1}{N} \sum_{i=1}^{N} f(x_i).$$

Taking a regular grid of points shows that N points can integrate the first $\sim N$ exponentials. Proof is actually exactly the same as before: the sum of roots of unity sum to 0, twice.



'Toroidal Designs': No Miracles

There is a result in Combinatorics/Fourier Analysis/Number Theory that implies No Miracles. It was originally proven for a very different reason (and is not very well known).

Lemma (Hugh Montgomery, 1980s)

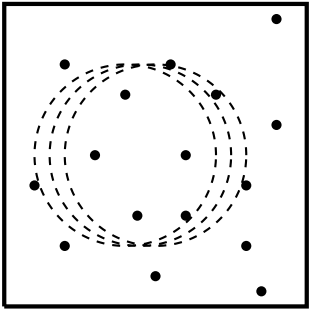
N points can integrate no more than $\sim N$ exponentials exactly.

I will first show you the beautiful result that he used it for.

Theorem (Montgomery)

Let $\{x_1, \ldots, x_N\} \subset \mathbb{T}^2$. There exists a disk $D \subset \mathbb{T}^2$ with radius $1/4$ or $1/2$ such that the number of elements in the disk substantially deviates from its expectation

$$\left| \# \left\{ 1 \leq i \leq N : x_i \in D \right\} - N|D| \right| \gtrsim N^{1/4}.$$



(J. Beck, 1980s): sharp up to possibly $\log N$ (*Open Problem.*)

The crucial ingredient: Montgomery's Lemma

Lemma (Montgomery, 1984)

For any $\{x_1, \ldots, x_N\} \subset \mathbb{T}^2$ *and* $X \geq 0$

$$\sum_{\substack{k \in \mathbb{Z}^2 \\ \|k\| \leq X}} \left| \sum_{n=1}^{N} e^{2\pi i \langle k, x_n \rangle} \right|^2 \geq N X^2.$$

It says that N points are not orthogonal to all low-frequency trigonometric polynomials.

Lemma (Montgomery, 1984)

For any $\{x_1, \dots, x_N\} \subset \mathbb{T}^2$ *and* $X \geq 0$

$$\sum_{\substack{k \in \mathbb{Z}^2 \\ \|k\| \leq X}} \left| \sum_{n=1}^{N} e^{2\pi i \langle k, x_n \rangle} \right|^2 \geq N X^2.$$

Theorem (S, Journal of Number Theory, 2017)

$$\sum_{\|k\| \leq X} \left| \sum_{n=1}^{N} e^{2\pi i \langle k, x_n \rangle} \right|^2 \gtrsim \sum_{i,j=1}^{N} \frac{X^2}{1 + X^4 \|x_i - x_j\|^4} \gtrsim N X^2.$$

- refinement of Montgomery's Lemma: as soon as there is any form of clustering, we get an automatic improvement.
- these ideas also refine Montgomery's orginal circle result

Erwin Schrödinger Institute, Fall 2017

(Dmitriy Bilyk, U Minnesota)

His talk (on joint work with Feng Dai, U Alberta) used a Montgomery-type inequality on the sphere $\mathbb{S}^d$ (same statement, spherical harmonics instead of exponentials).

Erwin Schrödinger Institute, Fall 2017

The other memorable impression from that conference.

Laplacian Eigenfunctions

What unites

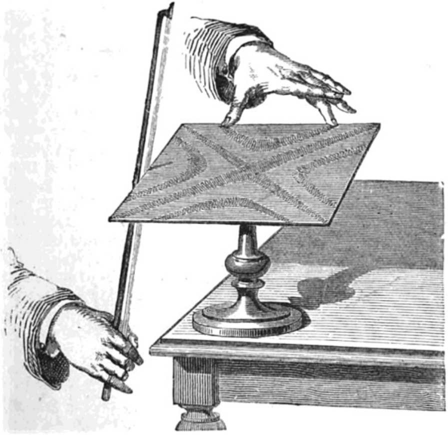
- spherical harmonics (polynomials) on the sphere $\mathbb{S}^d$ and
- exponential functions $\exp(i \langle k, x \rangle)$ on $\mathbb{T}^d$?

They are both eigenfunctions of the Laplacian on the manifold

$$-\Delta_g \phi_k = \lambda_k \phi_k.$$

If you feel you don't have a good handle on what that means, *welcome to the club*: People have been trying to understand them for 200 years.

Ernst Florens Friedrich Chladni (1756 - 1827)



Montgomery's Lemma on general Manifolds

Let us use ϕ_k to denote the L^2-normalized eigenfunctions of $-\Delta_g$ (ordered w.r.t, increasing eigenvalue).

Lemma (Montgomery, 1980s)

For any $\{x_1, \ldots, x_N\} \subset \mathbb{T}^d$

$$\sum_{k \le X} \left(\frac{1}{N} \sum_{n=1}^{N} \phi_k(x_n) \right)^2 \gtrsim_d NX.$$

Lemma (Dmitriy Bilyk and Feng Dai, 2016)

For any $\{x_1, \ldots, x_N\} \subset \mathbb{S}^d$

$$\sum_{k \le X} \left(\frac{1}{N} \sum_{n=1}^{N} \phi_k(x_n) \right)^2 \gtrsim_d NX.$$

Montgomery's Lemma on general Manifolds

Conjecture. On a general compact manifold M

$$\sum_{k \le X} \left(\frac{1}{N} \sum_{n=1}^{N} \phi_k(x_n) \right)^2 \gtrsim_d NX.$$

Theorem (Bilyk, Dai, S., Mathematische Annalen, 2018)

For any $\{x_1, \dots, x_N\} \subset M$ *(smooth, compact manifold without boundary) and* **positive weights** a_n *(suitably normalized)*

$$\sum_{k \le X} \left(\frac{1}{N} \sum_{n=1}^{N} a_n \phi_k(x_n) \right)^2 \gtrsim_M \frac{NX}{(\log X)^{\frac{d}{2}}}.$$

(Generalizes Montgomery on $\mathbb{T}^d, \mathbb{S}^d$. Removing log is *open problem*.)

Universal Limitations on Quadrature

Theorem (General No Miracle Theorem, S., IMRN 2019)

There exists a $k \leq c_d N + o(N)$ *such that*

$$\frac{1}{N} \sum_{n=1}^{N} \phi_k(x_n) \neq 0.$$

Moreover, the constant is independent of the manifold

$$c_d \leq \frac{(d/2+1)^{d/2+1}}{\Gamma(d/2+1)}.$$

Conjecture.

$$c_d = d + 1.$$

The rest of the talk

Until now we have seen

- spherical designs, the existence and non-existence theory
- what such a theory could look like on a manifold
- and some applications: exponential sum estimates in Number Theory, structure statements for Laplacian eigenfunctions

Vague Philosophy

A good proof gives you more than just a theorem.

The rest of the talk

Vague Philosophy

A good proof gives you more than just a theorem.

Outline for the rest of the talk: applications to

- data science (George Linderman, Yale Medical)
- actual numerical integration (M. Sachs and J. Lu, Duke)
- extremal combinatorics (Konstantin Golubev, ETH)

'Hey George, I proved a pretty useless result.'

(since May: Dr.) George Linderman (M.D./Ph.D. program Yale)

The Lancet Cover from Last Year

Prevalence, awareness, treatment, and control of hypertension in China: data from 1·7 million adults in a population-based screening study (China PEACE Million Persons Project)

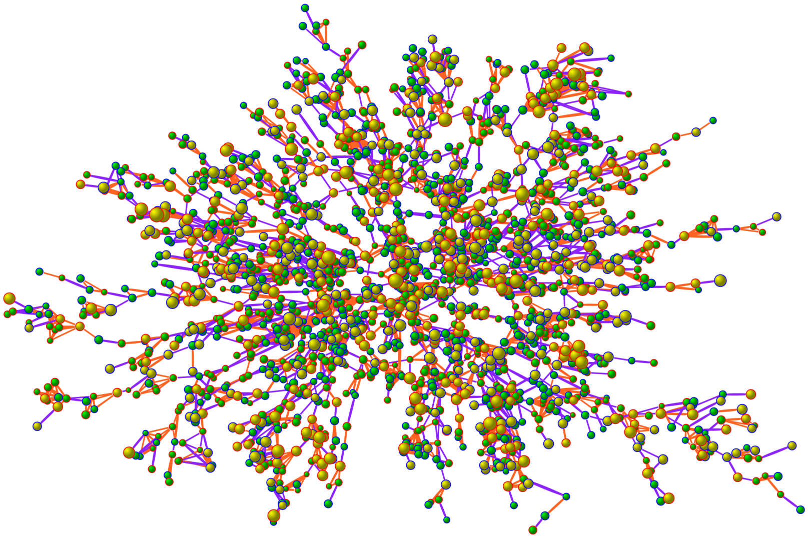
Jiapeng Lu, Yuan Lu*, Xiaochen Wang, Xinyue Li, George C Linderman, Chaoqun Wu, Xiuyuan Cheng, Lin Mu, Haibo Zhang, Jiamin Liu, Meng Su, Hongyu Zhao, Erica S Spatz, John A Spertus, Frederick A Masoudi, Harlan M Krumholz†, Lixin Jiang†*

Summary

Background Hypertension is common in China and its prevalence is rising, yet it remains inadequately controlled. Few studies have the capacity to characterise the epidemiology and management of hypertension across many heterogeneous subgroups. We did a study of the prevalence, awareness, treatment, and control of hypertension in China and assessed their variations across many subpopulations.

Lancet 2017; 390: 2549–58

Published Online
October 25, 2017
http://dx.doi.org/10.1016/
S0140-6736(17)32478-9



(Christakis and Fowler, The Spread of Obesity in a Large Social Network Over 32 Years, New England Journal of Medicine 2007.)

Sampling on Graphs

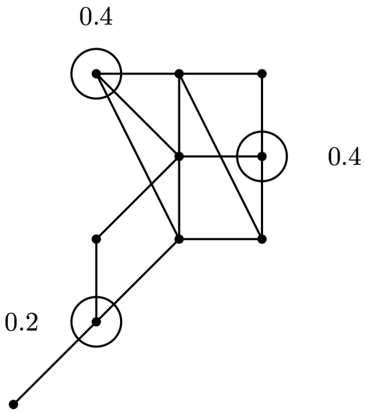
A basic problem: you are given a finite graph $G = (V, E)$ (think of a facebook graph, people and their friends). You have an unknown function

$$f : V \to \mathbb{R} \qquad \text{(say, blood pressure).}$$

You want to understand the average value of f and are allowed to evaluate f in 3 vertices. Which 3 vertices do you choose?

This was the actual question George had to deal with.

Sampling on Graphs



Linderman and S, Numerical Integration on Graphs, *Mathematics of Computation*, 2020

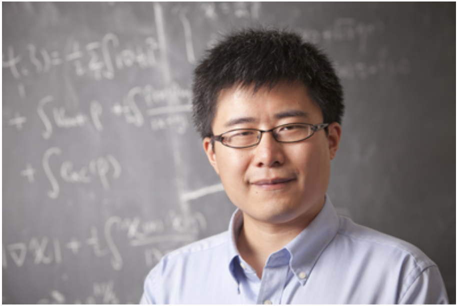
Open Problem: everything. Might be very important.

Another Fun Byproduct: 'Actual' Numerical Integration

In all these proofs, there is one term that sort of consistently popped up:

$$\text{minimize} \quad \sum_{i,j=1}^{N} e^{-N^{2/d}\|x_i - x_j\|^2}.$$

Matthias Sachs

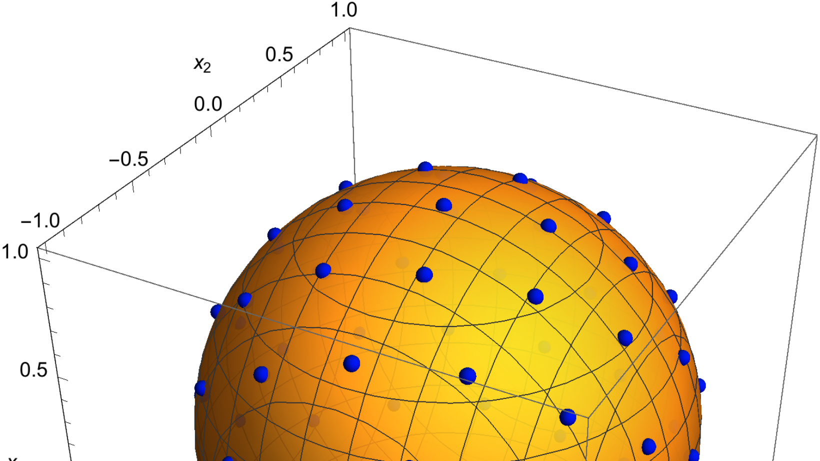


Jianfeng Lu

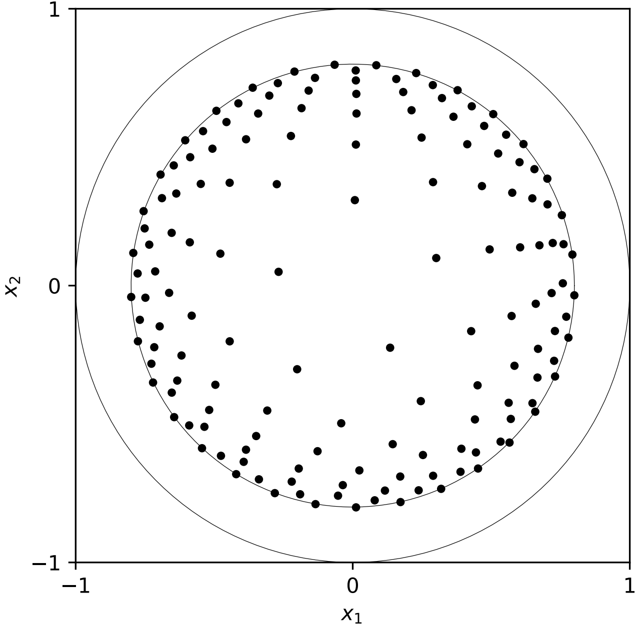
Returning to Euclidean Space

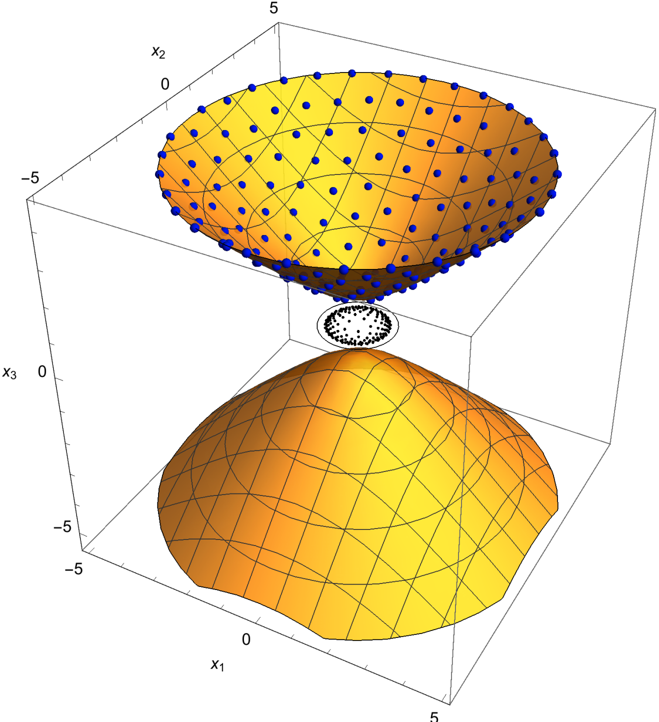
Going back to the continuous setting: we want to put points $\{x_1, \ldots, x_n\}$ such that

$$\sum_{i,j=1}^{n} \exp(-n^{2/d} \|x_i - x_j\|^2) \to \min .$$

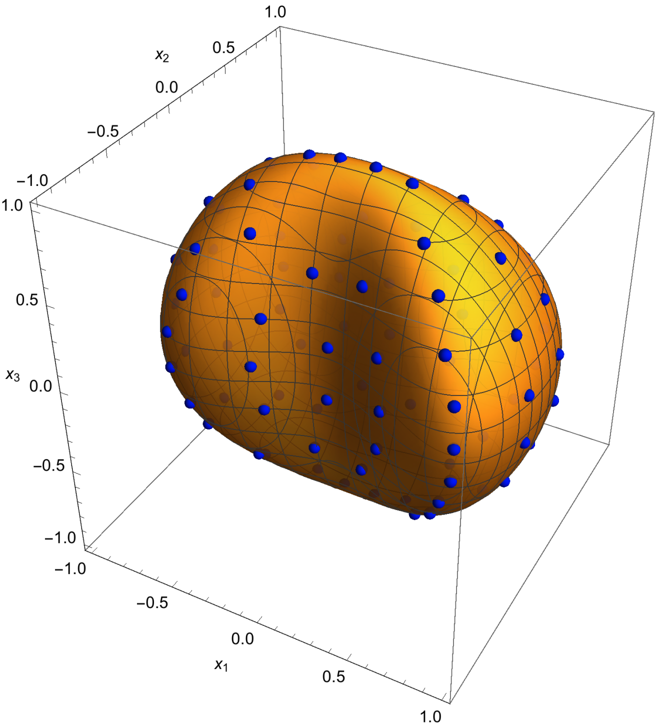


On the truncated Poincare disk





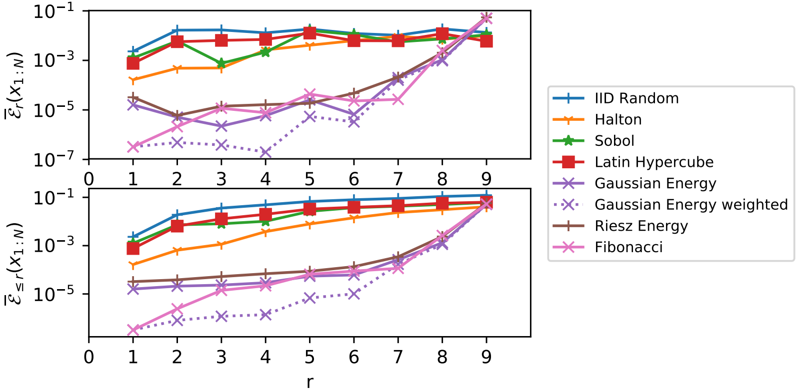
5
x_2
0
−5
x_3
0
−5
−5
0
x_1
5



x_1
x_2
x_3
-1.0
-0.5
0.0
0.5
1.0

Returning to Euclidean Space

Works on $\mathbb{T}^d$ (Lu, Sachs, S, *Constructive Approximation*, '19).



Integration Errors: purple is Graph Design heuristic. It works!

Let's move this philosophy somewhere else!

Some Vague Philosophy I

Sobolev Lebedev in Euclidean Space give Platonic Bodies.

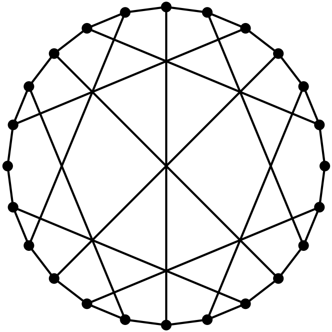
Some Vague Philosophy II

If we define the same thing on other abstract objects, we can find the 'Euclidean bodies' in that object.

What's the analogue of a platonic body on a Graph?

Some Graph Theory

We will work with finite, simple, connected Graphs $G = (V, E)$.

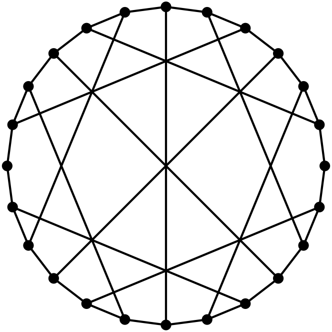


Functions are now simply maps $f : V \to \mathbb{R}$. The integral is merely a sum

$$\int_G f := \frac{1}{|V|} \sum_{v \in V} f(v).$$

Some Graph Theory

We will work with finite, simple, connected Graphs $G = (V, E)$.



Question. What is the analogue of a polynomial, a trigonometric polynomial or a Laplacian eigenfunctions?

Some Graph Theory

Definition (Graphic Laplacian)

If $f : V \to \mathbb{R}$, then the Graph Laplacian $(Lf) : V \to \mathbb{R}$ is given by

$$(Lf)(u) = \sum_{v \sim_E u} \left(\frac{f(v)}{\deg(v)} - \frac{f(u)}{\deg(u)} \right).$$

where the sum runs over all vertices v adjacent to u.

This is merely a linear operator, a $|V| \times |V|$ matrix. It has eigenvalues and eigenvectors.

Some Graph Theory

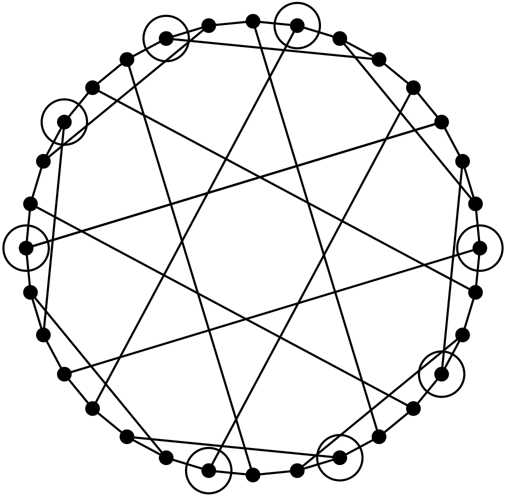
Definition (Graphical Design)

A graphical design on a Graph $G = (V, E)$ is a subset $W \subset V$ such that, for as many eigenfunctions (eigenvectors) ϕ_k of the Graph Laplacian (the matrix) as possible,

$$\frac{1}{|W|} \sum_{w \in W} \phi_k(w) = \frac{1}{|V|} \sum_{v \in V} \phi_k(v).$$

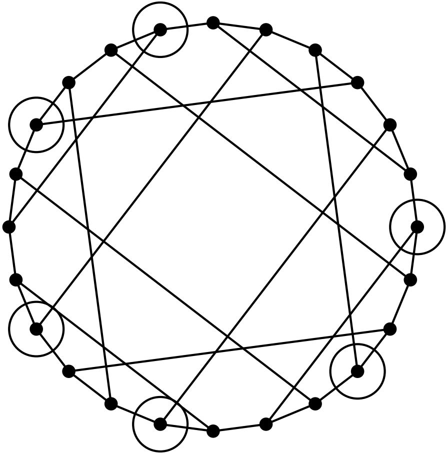
Why should they even exist at all? I do not know! But let's have a look.

Graphical Design on Dyck Graph



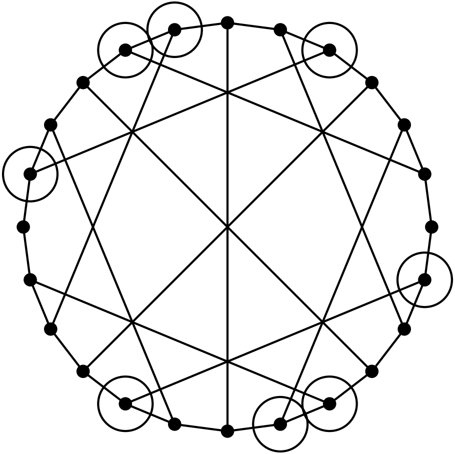
8 vertices integrate the first 16 of 32 eigenfunctions.

Graphical Design on Nauru Graph



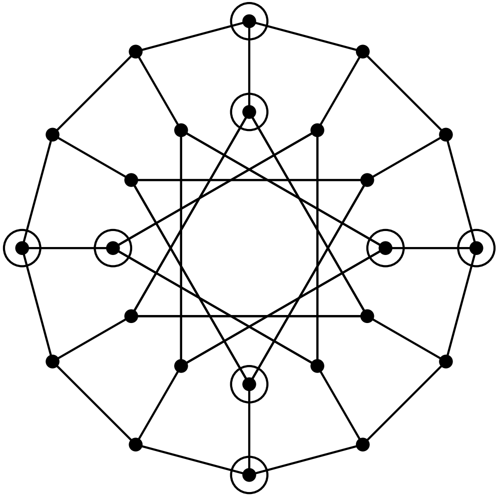
6 vertices integrate **19** out of 24 eigenfunctions exactly.

Graphical Design on McGee Graph



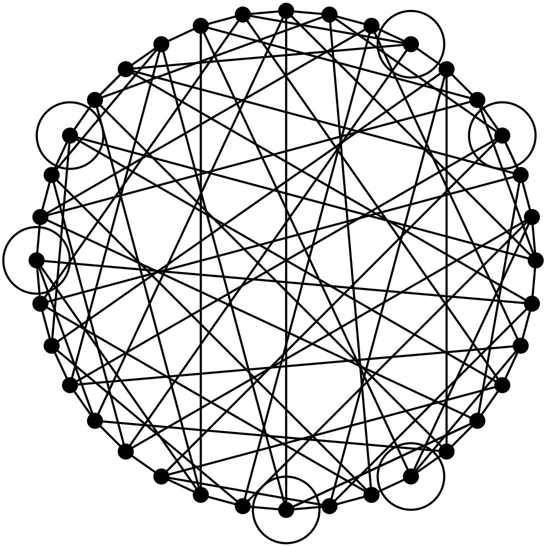
8 vertices integrate the first 21 of 24 eigenfunctions.

Graphical Design on Generalized Petersen Graph



8 vertices integrate the first 22 of 24 eigenfunctions.

Graphical Design on Sylvester Graph



6 vertices integrate the first 26 of 36 eigenfunctions.

Main Result

Graphical Designs are amazing (S, J. Graph Theory, 2019)

A Graphical Design W is either

1. not particularly good
2. has W large (for example $W = V$)
3. or has exponential growth of neighborhoods.

I was pretty sure that it's an important definition. However, I couldn't even show that interesting examples exist.

ETH 2019: “Can you send me your slides?”

Konstantin Golubev

Sep 2019: visit in New Haven

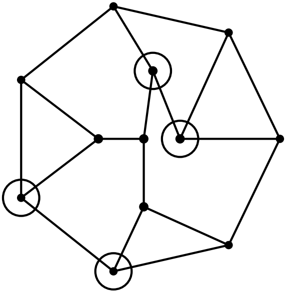


Figure: The Frucht Graph on 12 vertices: a subset W of 4 vertices integrates the first 11 eigenfunctions exactly.

Definition (Extremal Graphical Design)

We call a graphical design extremal if it integrates all but 1 eigenvector exactly. (This is best possible.)

Konstantin's Great Insight

The independence number $\alpha(G)$ of a Graph is the largest subset of vertices that has no edges between them. There is a classical 1970 bound of Hoffman in terms of the spectrum of the Graph Laplacian matrix

$$\alpha(G) \leq \frac{\lambda_n}{\lambda_n - 1}.$$

This is sometimes sharp.

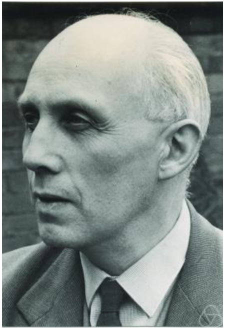
Theorem (Golubev, Oct. 2019)

Assume the Hoffman bound is sharp. Then the independence set is an extremal graphical design.

Erdős-Ko-Rado Theorem


Paul Erdős


Ke Zhao


Richard Rado

"Actually our first joint paper was done with Chao Ko, and was essentially finished in 1938, Curiously enough it was published only in 1961. One of the reasons for the delay was that at that time there was relatively little interest in combinatorics." (Erdős)

Erdős-Ko-Rado Theorem

Erdős-Ko-Rado Theorem

Let A be a collection of $k-$element subsets of $\{1, 2, \ldots, n\}$ such that any two elements in A have a non-empty intersection. Then

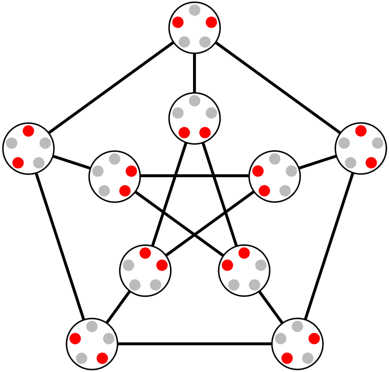
$$|A| \leq \binom{n-1}{k-1}.$$

This is sharp: take all subsets that contain the element 1.

Erdős-Ko-Rado Theorem

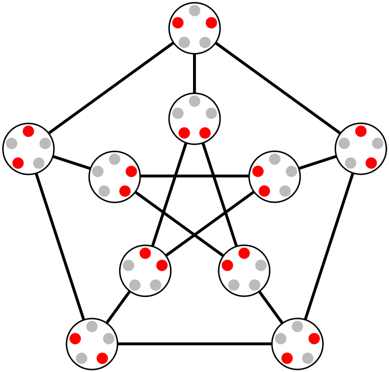
Definition (Kneser Graph)

Let V be the set of all $k-$element subsets of $\{1, \ldots, n\}$. Connect two vertices if the subsets are disjoint.



Erdős-Ko-Rado determines the independence number of the graph.

Erdős-Ko-Rado Theorem



Theorem (Golubev, Oct. 2019)

Extremal Configurations for Erdos-Ko-Rado in the Kneser Graph are an extremal graphical design.

The Story does not end here...

The same is true for the Deza-Frankl Theorem (a similar theorem about fixed points of permutations on $\{1, \dots, n\}$).

Extremal Combinatorics $\rightarrow$ Graphical Designs BUT...

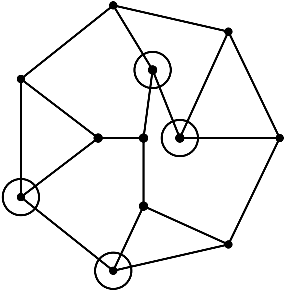
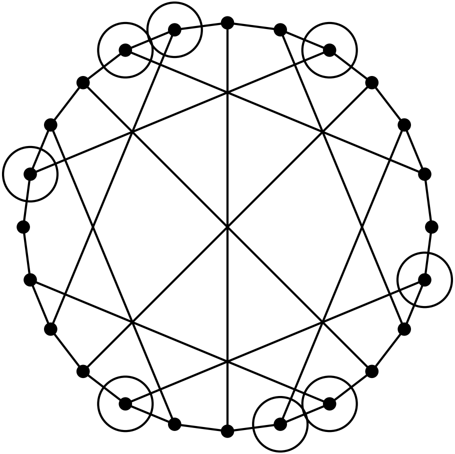
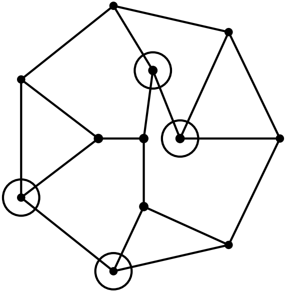


Figure: This is an extremal graphical design **but** not an independence set.

Many Questions Remain!

When do these magical Graphical Designs exist? What is required? What about weights? How do we find them? (Mixed Integer Programming?) How are they connected to classical Graph theory?





THANK YOU!