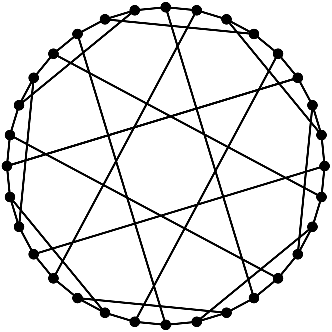
Sampling a Graph: Finding the important Vertices

Stefan Steinerberger

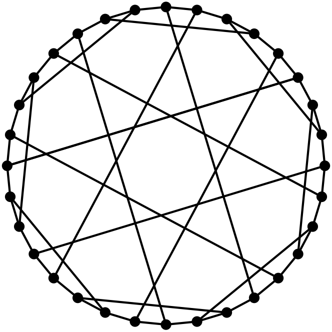
Trends in Mathematical Modelling, Simulation and Optimisation: Theory and Applications

UNIVERSITY *of* WASHINGTON

Suppose you are given a network like this.

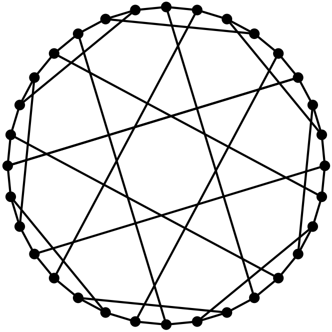


Suppose you are given a network like this.



There is a function $f : V \to \mathbb{R}$. We would like to understand the typical behavior of the function on this entire network by looking only in a few select nodes.

Suppose you are given a network like this.



There is a function $f : V \to \mathbb{R}$. We would like to understand the typical behavior of the function on this entire network by looking only in a few select nodes. **Which nodes do you check?**

My goal is to present a certain type of approach that

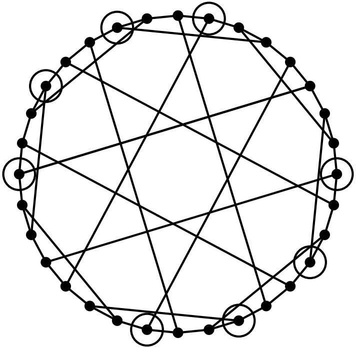
My goal is to present a certain type of approach that (1) has an underlying mathematical motivation and

My goal is to present a certain type of approach that (1) has an underlying mathematical motivation and (2) reduces (seemingly) to **mixed integer linear programing** problems.

My goal is to present a certain type of approach that (1) has an underlying mathematical motivation and (2) reduces (seemingly) to **mixed integer linear programing** problems. I also think the problem is **really important**.

My goal is to present a certain type of approach that (1) has an underlying mathematical motivation and (2) reduces (seemingly) to **mixed integer linear programing** problems. I also think the problem is **really important**. For the particular network discussed just now, a solution is

My goal is to present a certain type of approach that (1) has an underlying mathematical motivation and (2) reduces (seemingly) to **mixed integer linear programing** problems. I also think the problem is **really important**. For the particular network discussed just now, a solution is



Let's start in the continuous setting (which is strictly easier).

Let's start in the continuous setting (which is strictly easier). Suppose we want to approximate

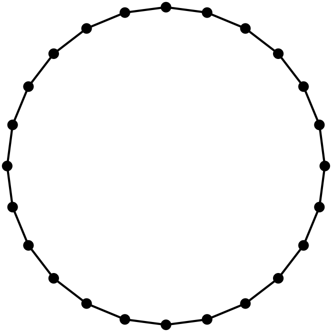
$$\frac{1}{|\mathbb{S}^d|} \int_{\mathbb{S}^d} f(x) dx \sim \frac{1}{N} \sum_{n=1}^{N} f(x_i).$$

Let's start in the continuous setting (which is strictly easier).
Suppose we want to approximate

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

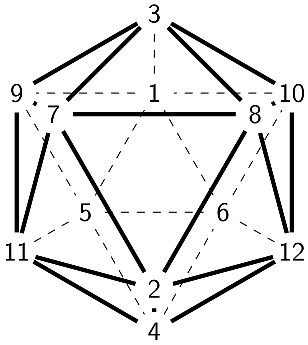
How should we select the points?

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: Platonic Solids



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

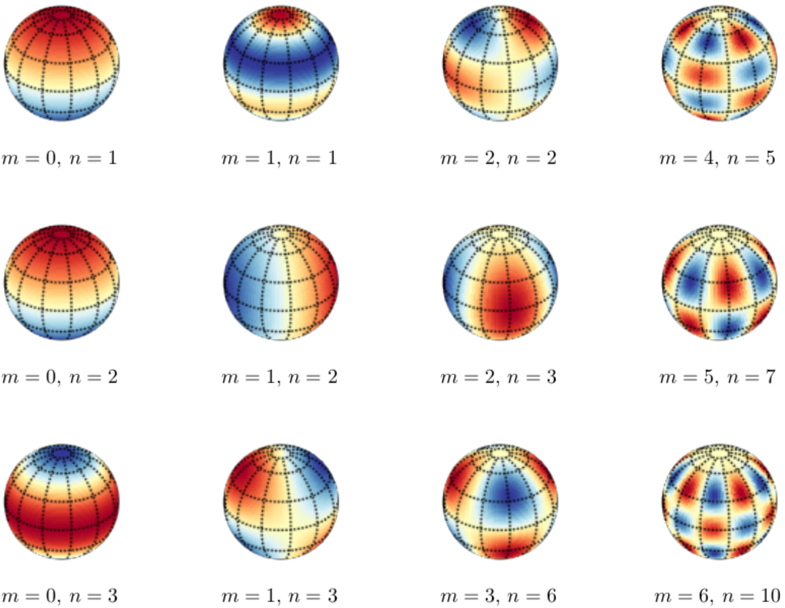
$$\frac{1}{|\mathbb{S}^2|}\int_{\mathbb{S}^2} f(x)dx \sim \frac{1}{N}\sum_{n=1}^{N} f(x_i).$$

How to select the points?

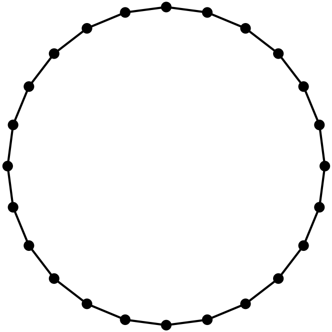
Idea (Sobolev 1962)

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.

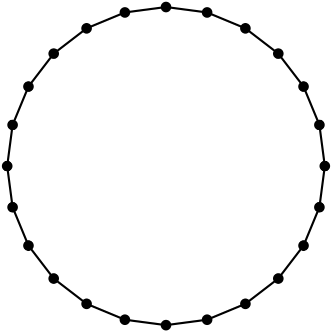
Spherical Harmonics



Checking our Minimal Requirements



Checking our Minimal Requirements

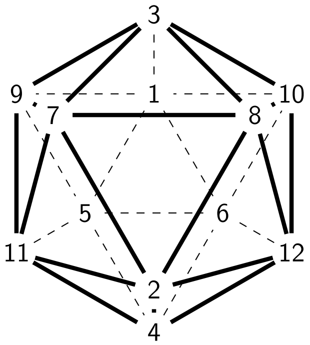


Polynomials in $\mathbb{R}^2$ look like $x^m y^n$. On $\mathbb{S}^1$, they start looking like

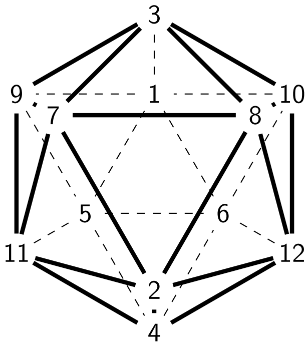
$$(\cos\theta)^m(\sin\theta)^n$$

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

Checking our Minimal Requirements



Checking 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).

Ahrens & Beylkin, Proc. Royal Soc. 2009

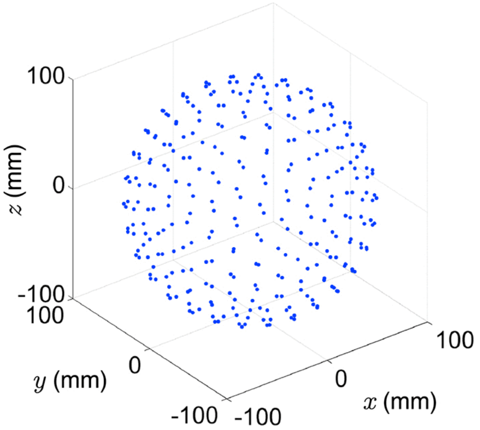


Figure: A set of 302 weighted points on $\mathbb{S}^2$ integrating all polynomials up to degree 29 exactly. Note that $30^2 = 900 \sim 3 \cdot 302$

Many people have studied these points

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

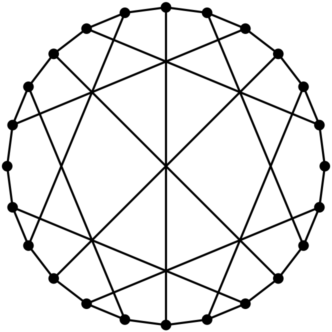
Now: move this entire philosophical framework over to Graphs.

Now: move this entire philosophical framework over to Graphs.

It is a priori not even clear that this works but it seems to work (for reasons not entirely clear at this point).

Some Graph Theory

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

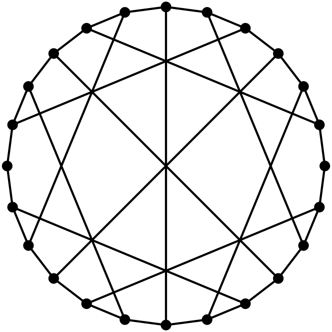


Functions are now simply maps $f : V \to \mathbb{R}$. The integral is merely a sum

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

Some Graph Theory

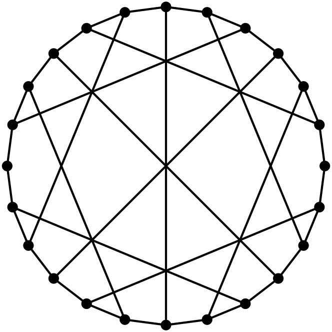
We will work with finite, simple, connected Graphs $G = (V, E)$.



Question. What is the analogue of a spherical harmonic polynomial?

Some Graph Theory

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



Question. What is the analogue of a spherical harmonic polynomial? Spherical harmonics are eigenfunctions of $\Delta_{\mathbb{S}^2}$.

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.

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. These eigenvectors are the *Graph-analogues* of the 'polynomials on the sphere'.

Some Graph Theory

Definition (Graphical Design)

I want to find subsets $W \subset V$ such that the average of ϕ_k is the same in W as in V for a large number of k.

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

Some Graph Theory

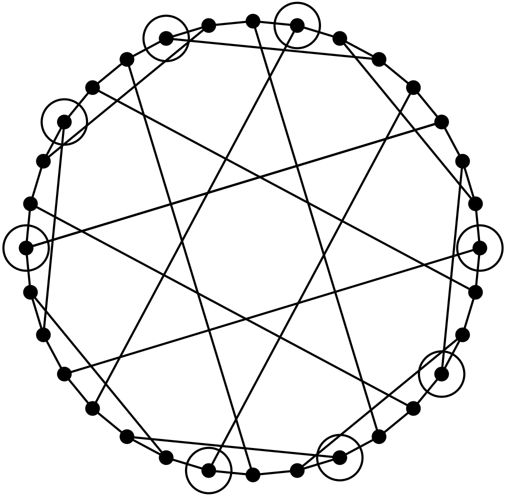
Definition (Graphical Design)

I want to find subsets $W \subset V$ such that the average of ϕ_k is the same in W as in V for a large number of k.

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

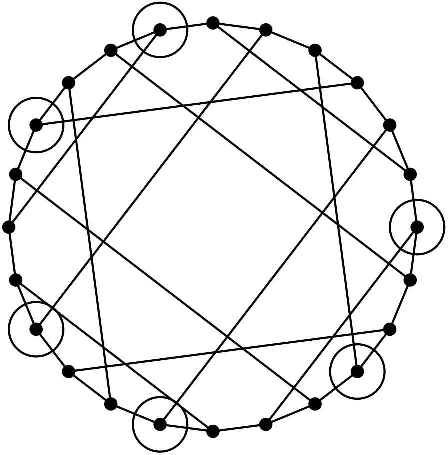
Why should they even exist at all? Originally I did not know! But let's have a look.

Graphical Design on Dyck Graph



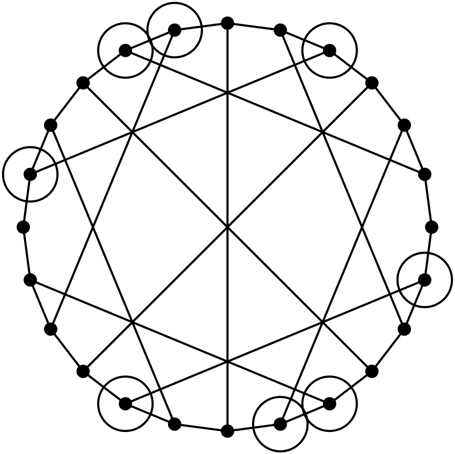
8 vertices integrate the first 16 of 32 eigenfunctions.

Graphical Design on Nauru Graph



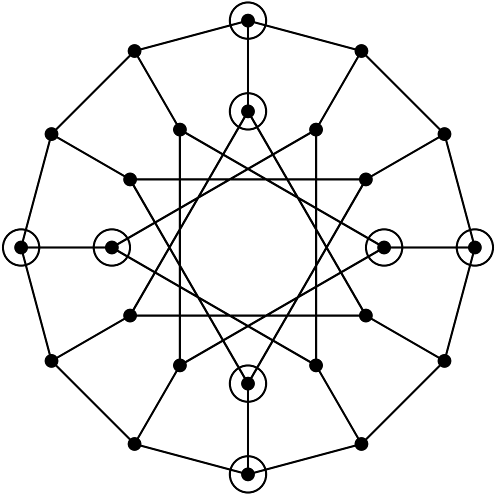
6 vertices integrate **19** out of 24 eigenfunctions exactly.

Graphical Design on McGee Graph



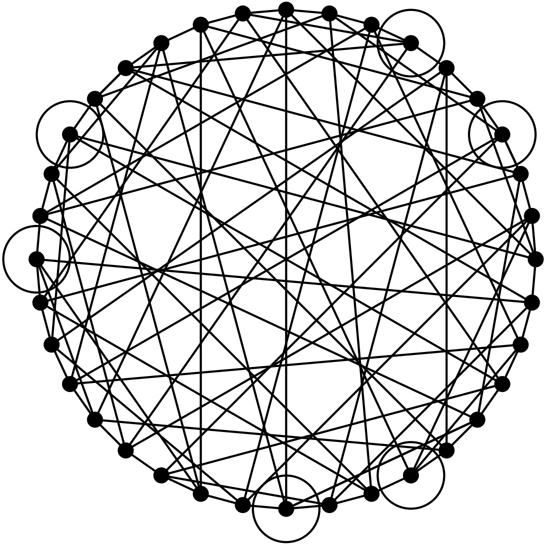
8 vertices integrate the first 21 of 24 eigenfunctions.

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.

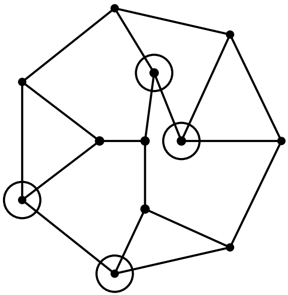


Figure: The Frucht Graph on 12 vertices: a subset W of 4 vertices integrates the first 11 eigenfunctions exactly.

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.

How do we find them?

Definition (Graphical Design)

I want to find subsets $W \subset V$ such that the average of ϕ_k is the same in W as in V.

How do we find them?

Definition (Graphical Design)

I want to find subsets $W \subset V$ such that the average of ϕ_k is the same in W as in V.

It's clear that the cases I have shown you are **algebraic miracles** and not robust. But certainly something similar works in general with a suitable relaxation.

Shahar Kovalsky found a nice way!

(Shahar Kovalsky, UNC Chapel Hill)

Shahar Kovalsky found a nice way!

Shahar Kovalsky proposes the ℓ^0-formulation of the Graphical Design Problem

$$\min_{\substack{a_v \in \{0,1\} \\ \|a\|_{\ell^1} = k}} \|Ua\|_{\ell^0}$$

where U is the matrix filled with entries of Laplacian eigenvectors.

Shahar Kovalsky found a nice way!

Shahar Kovalsky proposes the ℓ^0-formulation of the Graphical Design Problem

$$\min_{\substack{a_v \in \{0,1\} \\ \|a\|_{\ell^1} = k}} \|Ua\|_{\ell^0}$$

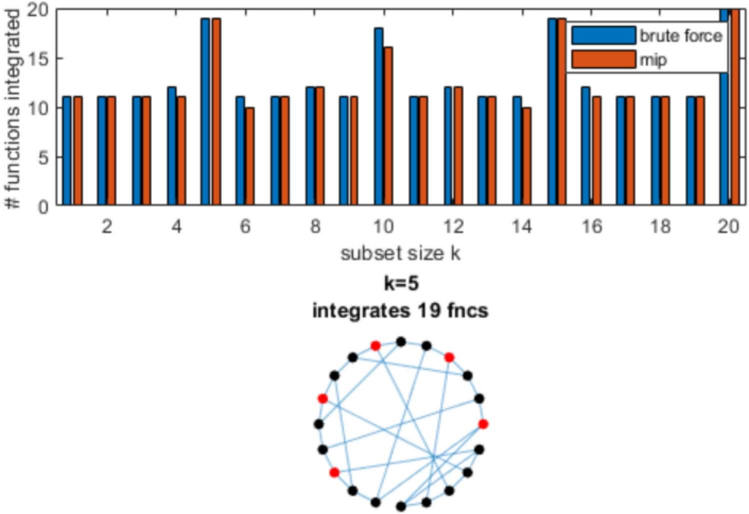
where U is the matrix filled with entries of Laplacian eigenvectors.

He discovered that the mixed integer linear programming formulation

$$\min_{\substack{a_v \in \{0,1\} \\ \|a\|_{\ell^1} = k}} \|Ua\|_{\ell^1}$$

seems to be exact in many cases. Standard solvers work!

Shahar Kovalsky found a nice way!



(Example by Shahar)

‘Hey George, I proved a pretty useless result.’

George Linderman

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

A basic problem in medicine(!): you are given a finite graph $G = (V, E)$ (think of a facebook graph, people and their friends).

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

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

A basic problem in medicine(!): 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.

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

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

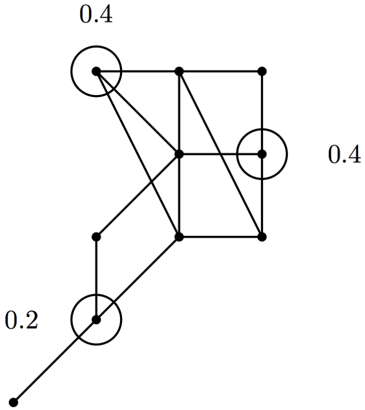
You want to understand the average value of f and are allowed to evaluate f in 3 vertices. Which 3 vertices do you choose?

A basic problem in medicine(!): 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.

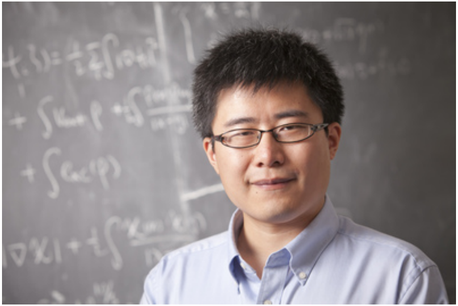


Linderman and S, Numerical Integration on Graphs, *Mathematics of Computation*, 2020. **This might be a really important problem.**

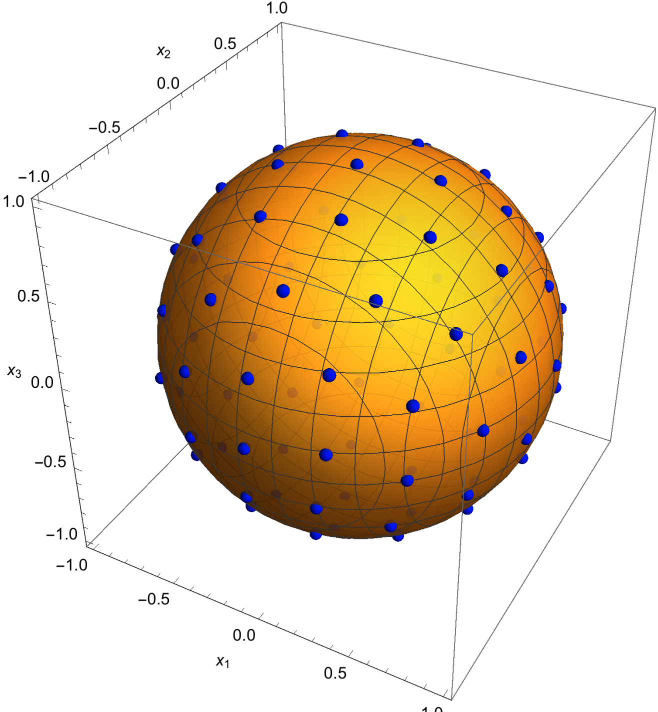
Another Fun Byproduct: ‘Actual’ Numerical Integration

The funny thing is that good ideas on graphs tend to also work in the continuous setting. A ‘continuous’ object is merely a particularly nice graph.

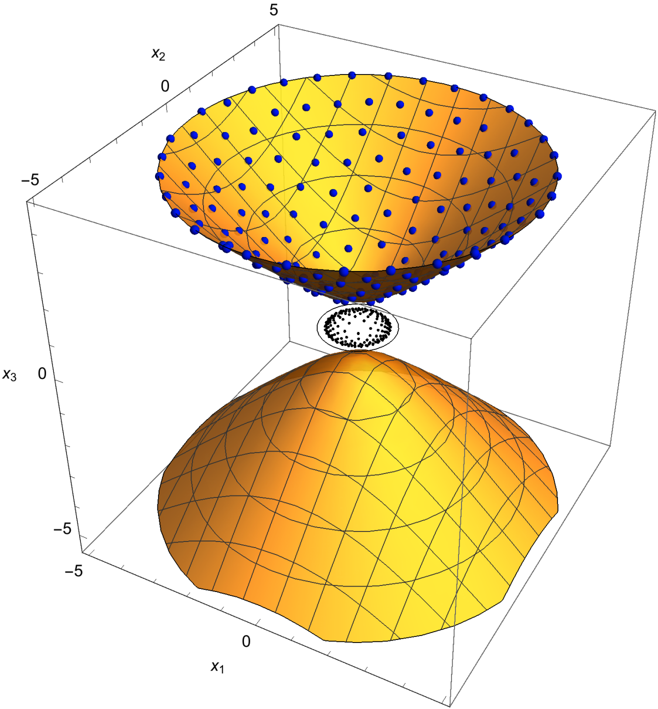
Matthias Sachs

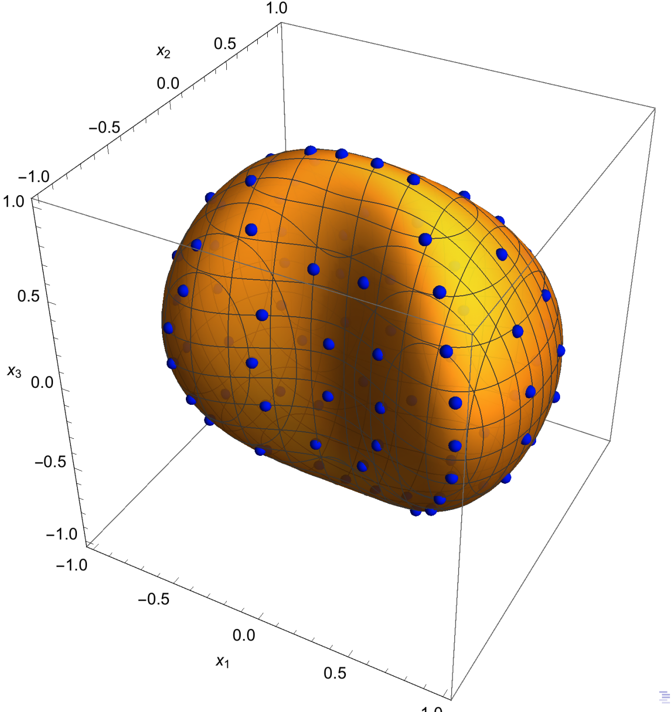


Jianfeng Lu



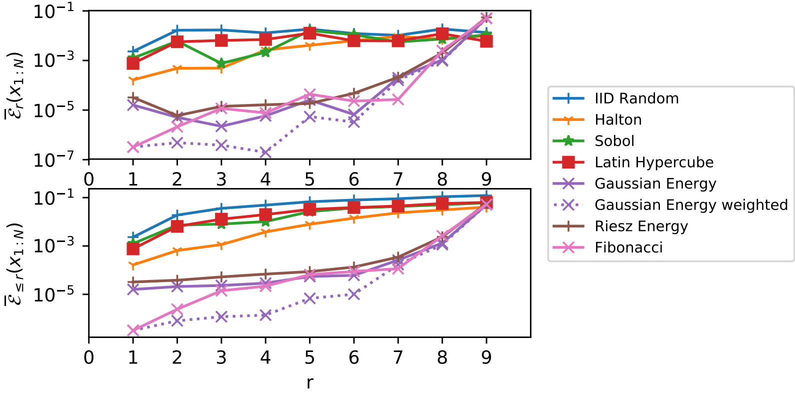
1.0
0.5
x_2
0.0
−0.5
−1.0
1.0
0.5
x_3
0.0
−0.5
−1.0
−1.0
−0.5
0.0
x_1
0.5


5
x_2
0
−5
x_3
0
−5
−5
0
x_1



x_2
1.0
0.5
0.0
−0.5
−1.0
1.0
0.5
x_3
0.0
−0.5
−1.0
−1.0
−0.5
0.0
x_1
0.5

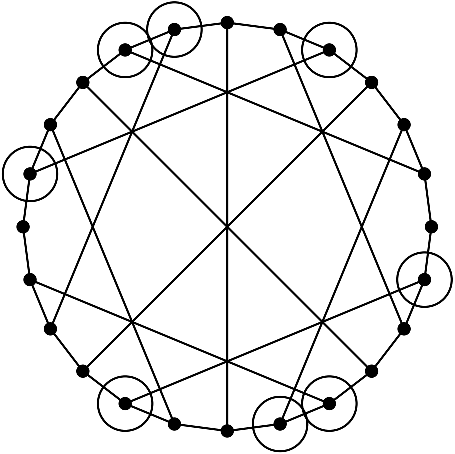
(Lu, Sachs, S, *Constructive Approximation*, '19).



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

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?

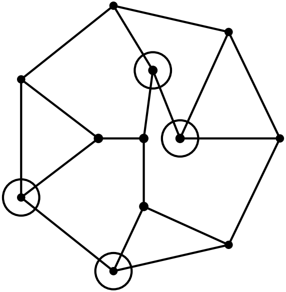


References

1. Spectral Limitations of Quadrature Rules and Generalized Spherical Designs, IMRN (2019)
2. Generalized Designs on Graphs: Sampling, Spectra, Symmetries, Journal of Graph Theory (2019)
3. J. Lu, M. Sachs and S, Quadrature Points via Heat Kernel Repulsion, Constructive Approximation (2020)
4. George Linderman and S, Numerical Integration on Graphs: where to sample and how to weigh, Mathematics of Computation (2020)

Also relevant:

1. Konstantin Golubev, Graphical Designs and Extremal Combinatorics (2020), **connects to the Erdős-Ko-Rado Theorem and the Deza-Frankl Theorem**
2. Catherine Babecki, Codes, Cubes, and Graphical Designs (2021) **connects to coding theory**



Thank you!