AN INTRODUCTION TO GRAPHICAL DESIGNS

CATHERINE BABECKI

A graphical design is a generalization of a quadrature rule to the domain of graphs. Informally, a quadrature rule is a set of points that represent a continuous domain well in the context of numerical integration. That is, if we would like to average a suitably chosen function over the entire domain, it should suffice to know what values the function takes on the quadrature points. An example of a quadrature rule on a sphere is a *spherical t-design*, which is a subset of points $\{x_1, \dots, x_N\} \subset \mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ and weights $\alpha_i \in \mathbb{R}$ chosen so that

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

whenever f is a polynomial of degree $\leq t$. For instance, the vertices of an icosahedron when embedded into a sphere form a spherical 5-design (Figure 1). The average of any function which is well approximated by a polynomial of degree ≤ 5 will be well approximated by these quadrature points.

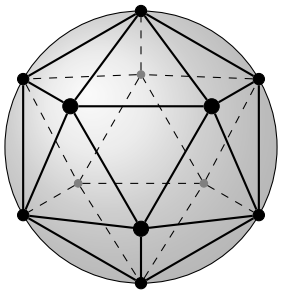


FIGURE 1. A spherical 5-design.

For a graph $G = (V, E)$, a graphical design, first defined in [Ste20], is a subset $W \subset V$ which approximates G in a similar sense; to know the average of a suitable function over the whole graph, we need only look at the function values on the graphical design. We take the class of suitable functions on a graph to be the low frequency eigenvectors of the graph Laplacian $L = AD^{-1} - I$, where A is the adjacency matrix, D is the diagonal degree matrix, and I is the identity. This mimics the construction of spherical t-designs, since low degree polynomials are the low frequency eigenfunctions of the Laplace-Beltrami operator Δ on the sphere. In many senses, L is for graphs what Δ is for smooth manifolds (see [Sin06; BIK13; HAL07] for some references).

There are many matrices which are called graph Laplacians (see [Chu97]). In the smooth case, a Taylor expansion shows that for a function $f : \mathbb{S}^{d-1} \to \mathbb{R}$, $\Delta f(x)$ is essentially the average of f in a neighborhood of x. We take $L = AD^{-1} - I$

Date: November 5, 2021.

since it analogously captures averaging over the neighborhood of a vertex; given $f : V \to \mathbb{R}$,

$$(Lf)(v) = \sum_{u:uv \in E} \left(\frac{f(u)}{\deg(u)} - \frac{f(v)}{\deg(v)} \right).$$

We say that an eigenvector φ_j of L and its eigenvalue λ_j have low frequency if j is early in the ordering

$$1 = |\lambda_1 + 1| \geq |\lambda_2 + 1| \geq \ldots \geq |\lambda_n + 1| \geq 0.$$

That is, λ_1 has the lowest frequency and λ_n has the highest frequency. The eigenvalues of L are contained in $[-2, 0]$ – see Figure 2 for an illustration of how frequency is arranged in this interval. Low frequency eigenfunctions of Δ are smooth objects. With this ordering of L, low frequency eigenvectors respect graph structure, which is not always smooth. Let's see how this works with an example, the cycle on 50 vertices, which we denote C_{50}.

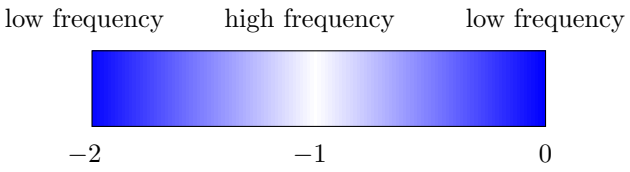
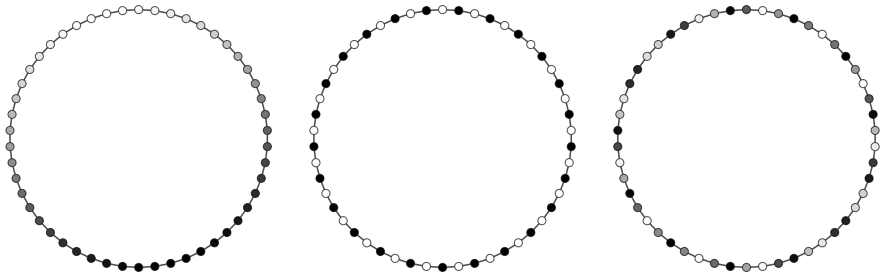


FIGURE 2. Ordering the eigenvalues of L by frequency

A graph is regular if and only if $L\mathbb{1} = 0$, where $\mathbb{1}$ denotes the all-ones vector. The constant vector is the smoothest possible function. In Figure 3A, we see that φ_3, a low frequency eigenvector, creates a smooth gradient around the cycle. An eigenvector φ with eigenvalue near 0 is smooth in the sense of continuous functions – the average of φ in a neighborhood of v is close to $\varphi(v)$. The eigenvector φ_2 (Figure 3B) is also low frequency. While it is highly non-smooth in the traditional sense, it does respect the structure of C_{50} by exactly separating the two halves of its bipartition. An eigenvector with eigenvalue near -2 is 'anti-smooth' – the average of φ in a neighborhood of v is close to $-\varphi(v)$. These functions are highly oscillating, but in a structured way. Thus low frequency eigenvectors like $\varphi_1 = \mathbb{1}, \varphi_2$, and φ_3 relate to the structure of C_{50}.



(A) φ_3. $\lambda_3 \approx -0.0079$ (B) φ_2. $\lambda_2 = -2$ (C) φ_{50}. $\lambda_{50} \approx -1.063$

FIGURE 3. Three eigenvectors of C_{50}

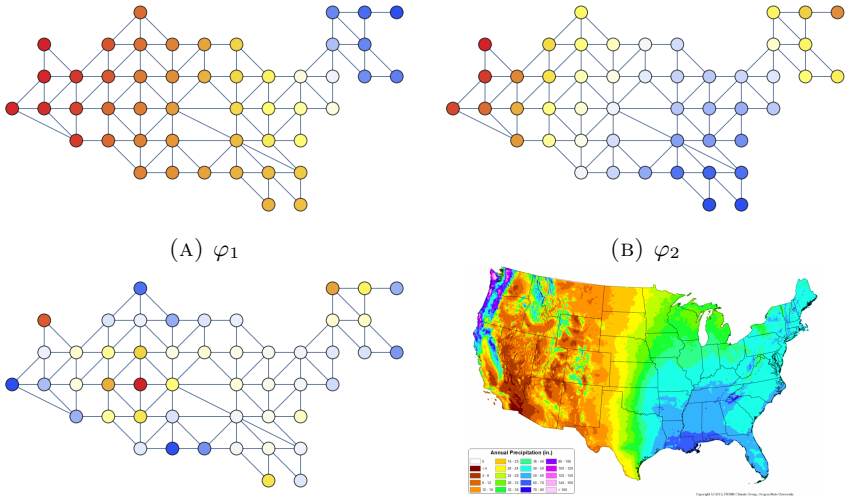
High frequency eigenvectors are not useful in this sense. In classical smooth domains, functions which decay under Δ are less significant objects, as their contributions die out over time. An eigenvector with eigenvalue near -1 is such that

$|(L\varphi)(v)| < |\varphi(v)|$; that is, these are decaying functions under L. Indeed, by inspection the high frequency eigenvector φ_{50} (Figure 3C) does not contain any clear information about C_{50}. In this sense, the frequency ordering of L prioritizes eigenvectors which best respect the structure of the graph, leaving the 'noisiest' eigenvectors in the higher frequencies.

We can see another illustration of this principle in the graph of the contiguous United States (Fig. 4). Vertices represent states, and edges connect states which share a border. The first two eigenvectors by frequency are smooth with respect to the graph geometry, but the eleventh eigenvector is not so smooth. The eigenvectors of L form a basis for for the set of all functions $f : V \to \mathbb{R}$. If we expand a function which relates to a graph's structure, the coefficients c_i of the expansion

$$f = c_1\varphi_1 + \ldots + c_{|V|}\varphi_{|V|}$$

should decay as i increases. This notion was made more precise in [LS20]. For the contiguous United States graph, the graph structure is the geography of the country. We expect that a function which is largely determined by geography, such as average annual precipitation by state, will be approximated well by a truncated expansion in terms of the low frequency eigenvectors.



(A) φ_1 (B) φ_2 (C) φ_{11} (D) Average Annual Precipitation in the Contiguous US, 1981-2010[1]

FIGURE 4. The contiguous United States graph.

Thus we make the following definitions. Definition 1.1 first appeared in [Ste20], Definitions 1.2 and 2 first appeared in [Bab21], and are variations of the problem that was initially posed.

Definition 1. Consider a graph $G = (V, E)$.

(1) A subset $W \subset V$ *integrates an eigenvector* φ of L if

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

[1]©2015, PRISM Climate Group, Oregon State University. http://prism.oregonstate.edu/normals/

(2) A subset $W \subset V$ *integrates an eigenspace* Λ of L if W integrates every vector in Λ, or equivalently, if W integrates *any* basis of Λ.

We have restricted our attention to equal weights. It is a more general problem to say that for a graph $G = (V, E)$ and weights $a_v \in \mathbb{R}$, $W \subset V$ integrates φ if

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

To mimic the qualities of a spherical t-design on a graph, we seek subsets which exactly integrate the first eigenspaces by frequency. Throughout this paper, let $G = (V, E)$ be a graph whose Laplacian has m distinct eigenspaces ordered from low to high frequency as $\Lambda_1 \leq ... \leq \Lambda_m$.

Definition 2. A *k-graphical design* is $W \subset V$ such that W integrates $\Lambda_1, \ldots, \Lambda_k$.

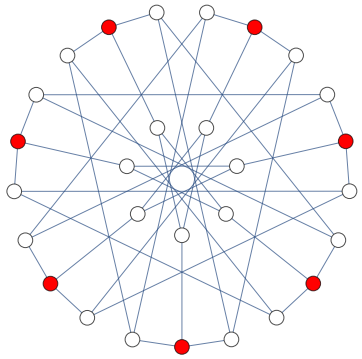


FIGURE 5. A 13-graphical design on the J7 Flower Snark

There is not yet any consensus on what makes a k-graphical design 'good'. So, we define the following three variations on the graphical design problem. An optimal design seeks to optimize the trade-off between number of vertices and number of eigenspaces integrated. A maximal design integrates as many eigenspaces as is possible for a given graph, using as few vertices as possible. An extremal design integrates all but the last eigenspace by frequency. In precise terms, these are defined as follows. Definition 3 was first introduced in [Bab21].

Definition 3 (Optimal Designs and Efficacy)**.** Let $W \subset V$ be a k-graphical design but not a $(k+1)$-graphical design. We define the *efficacy* of W as

$$\text{efficacy}(W) = \frac{|W|}{\sum_{i=1}^{k} \dim(\Lambda_i)}.$$

An *optimal design* is a subset $W^* \subset V$ such that

$$\text{efficacy}(W^*) = \min \{\text{efficacy}(W) : W \subset V\}.$$

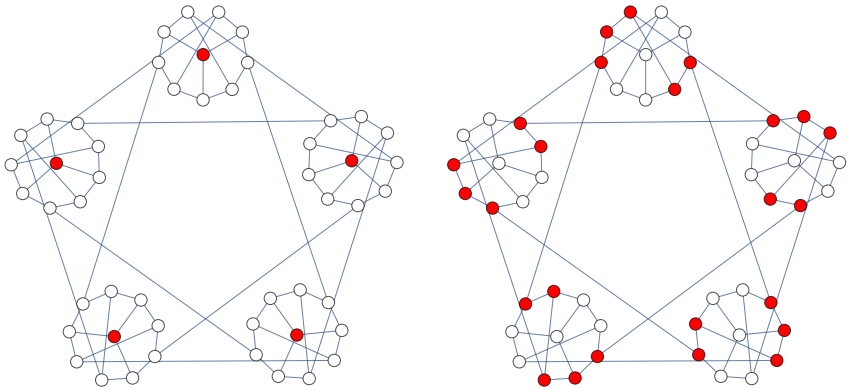
Definition 4 (Maximal Designs)**.** Let $K \in [m-1]$ be maximal such that G has a K-design. A *maximal design* is a minimum cardinality K-design.

The 13-graphical design in Figure 5 is both optimal and maximal. We will see an example shortly where the optimal and maximal designs are distinct; it can be that the optimal design W^* integrates few eigenspaces, but $|W^*|$ is very small.

Definition 5 (Extremal Designs)**.** An *extremal design* is an $(m-1)$-design.

This is a variation of the definition of extremal from [Gol20], which did not consider the frequency order and only asked for a subset which integrated all but one eigenspace. Using our language, a minimum cardinality extremal design must be a maximal design. Not every graph has extremal designs, however.

Example 6. Consider the Szekeres Snark (Figure 6), a 3-regular graph on 50 vertices with 11 eigenspaces. An optimal design W^* consists of 5 vertices and is a 3-graphical design, with efficacy$(W^*) = 5/9$. A maximal design $W_{\max}$ consists of 25 vertices and is an 8-graphical design, with efficacy$(W_{\max}) = 25/33$. There are no extremal designs on this graph.



(A) An optimal design (B) A maximal design

FIGURE 6. The Szekeres Snark.

The graphical design problem is closely tied to signal processing on graphs and the graph sampling problem. Signal processing aims to optimize, modify, or analyze data that represent some physical events or measurements. Some common goals of signal processing are to speed up data transmission, to store data more efficiently, to improve the subjective quality of the data (say, a grainy image), and to emphasize or detect notable components of the data. Traditional signal processing techniques consider data from the time, space, or frequency domains with either continuous or discrete measurements. There are many challenges when trying to translate the techniques of continuous and discrete signal processing to the domain of graphs, see [Shu+13] for an overview of this field. In particular, there is no natural notion of downsampling. To downsample a discrete signal, one takes every j-th data point. For a graph signal $f : V \to \mathbb{R}$ on $G = (V, E)$, there is no way to define 'every j-th vertex,' since any labeling of the vertices is arbitrary. Graphical designs provide a meaningful way to downsample a graph, as a design is constructed to respect functions which respect the structure of G.

We summarize some of the literature adjacent to graphical designs. As previously noted, graphical designs were first introduced in [Ste20]. This paper also shows, loosely speaking, that if W is a good graphical design, then either $|W|$ is large, or the j-neighborhoods of W grow exponentially. [LS20] finds an upper bound on the integration error for any quadrature rule on a graph, which multiplicatively separates the function f and a quantity which can be interpreted as the quality of the quadrature scheme. [Gol20] introduces (unordered) extremal designs, which are then connected to the Hoffman and Cheeger bounds. [Bab21] connects graphical

designs on graphs of the cube to linear error correcting codes and distinguishes graphical designs from several related concepts: the extremal designs of [Gol20], maximum stable sets in distance graphs, and t-designs in association schemes. See [Val19] for an overview of the regime which connects maximum stable sets in distance graphs to coding theory, and Chapter 21 of [MS77] for an introduction to association schemes. [MS77] is also the standard text for the theory of error correcting codes. For more on spectral graph theory, [Chu97] is the classical reference; we also refer to [Spi19]. For a short introduction to spherical harmonics, see [Moh]. Spherical t-designs were introduced in [DGS77], see also [BRV13; BRV15]. For quadrature rules on general smooth manifolds, see, for instance, [GG18; Ste19]. The graph sampling problem has been investigated primarily from an engineering perspective, such as [AGA16; TBL16; Tan+20]. We also refer to the work of Pesenson (e.g. [Pes00; Pes08; Pes19]).

References

[AGA16] A. Anis, A. Gadde, and Antonio A. Ortega. "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies". In: *IEEE transactions on signal processing a publication of the IEEE Signal Processing Society.* 64.14 (2016), pp. 3775–3789. ISSN: 1053-587X.

[Bab21] C. Babecki. "Cubes, codes, and graphical designs". In: *J Fourier Anal Appl* 27 (2021), p. 81. URL: https://doi.org/10.1007/s00041-021-09852-z.

[BIK13] D. Burago, S. Ivanov, and Y. Kurylev. "A graph discretization of the Laplace-Beltrami operator". In: *Journal of Spectral Theory* 4 (Jan. 2013). DOI: 10.4171/JST/83.

[BRV13] A. Bondarenko, D. Radchenko, and M. Viazovska. "Optimal asymptotic bounds for spherical designs". In: *Annals of mathematics.* 178.2 (2013), pp. 443–452. ISSN: 0003-486X.

[BRV15] A. Bondarenko, D. Radchenko, and M. Viazovska. "Well-separated spherical designs". In: *Constructive approximation.* 41.1 (2015), pp. 93–112. ISSN: 0176-4276.

[Chu97] F.R.K. Chung. *Spectral graph theory.* Providence, R.I. : Published for the Conference Board of the mathematical sciences by the American Mathematical Society, 1997. ISBN: 9780821803158.

[DGS77] P. Delsarte, J. M. Goethals, and J. J. Seidel. "Spherical codes and designs". In: *Geometriae Dedicata* 6.3 (1977), pp. 363–388. DOI: 10.1007/bf03187604. URL: https://doi-org.offcampus.lib.washington.edu/10.1007/bf03187604.

[GG18] B. Gariboldi and G. Gigante. *Optimal asymptotic bounds for designs on manifolds.* 2018. arXiv: 1811.12676 [math.AP].

[Gol20] K. Golubev. "Graphical designs and extremal combinatorics". In: *Linear Algebra and its Applications* 604 (2020).

[HAL07] M. Hein, J.-Y. Audibert, and U. von Luxburg. "Graph Laplacians and Their Convergence on Random Neighborhood Graphs". In: *Journal of Machine Learning Reseach* 8 (2007).

[LS20] G. C. Linderman and S. Steinerberger. "Numerical integration on graphs: Where to sample and how to weigh". In: *Mathematics of computation.* 89.324 (2020), pp. 1933–1952. ISSN: 0025-5718.

[Moh] M. J. Mohlenkamp. *A User's Guide to Spherical Harmonics.* URL: http://www.ohiouniversityfaculty.com/mohlenka/research/uguide.pdf. (version: 10.18.2016).

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error Correcting Codes*. North Holland, 1977.

[Pes00] I. Z. Pesenson. "A sampling theorem on homogeneous manifolds". In: *Transactions of the American Mathematical Society*. 352.9 (2000), pp. 4257–4269. ISSN: 0002-9947.

[Pes08] I. Z. Pesenson. "Sampling in Paley-Wiener spaces on combinatorial graphs". In: *Transactions of the American Mathematical Society*. 360.10 (2008), pp. 5603–5627. ISSN: 0002-9947.

[Pes19] I. Z. Pesenson. *Sampling by averages and average splines on Dirichlet spaces and on combinatorial graphs*. 2019. arXiv: 1901.08726.

[Shu+13] D. Shuman et al. "The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and other Irregular Domains". In: *IEEE Signal Processing Magazine* 30.3 (2013).

[Sin06] A. Singer. "From Graph to Manifold Laplacian: The Convergence Rate". In: *Applied and Computational HarmonicAnalysis* 21 (2006).

[Spi19] D. Spielman. *Spectral and Algebraic Graph Theory*. 2019. URL: http://cs-www.cs.yale.edu/homes/spielman/sagt/.

[Ste19] S. Steinerberger. "Spectral Limitations of Quadrature Rules and Generalized Spherical Designs". In: *IMRN* (2019). URL: https://arxiv.org/abs/1708.08736. (accepted).

[Ste20] S. Steinerberger. "Generalized designs on graphs: Sampling, spectra, symmetries". In: *Journal of graph theory*. 93.2 (2020), pp. 253–267. ISSN: 0364-9024.

[Tan+20] Y. Tanaka et al. "Sampling signals on graphs: from theory to applications". In: *IEEE Signal Processing Magazine* 37.6 (Nov. 2020), pp. 14–30. ISSN: 1558-0792. DOI: 10.1109/msp.2020.3016908. URL: http://dx.doi.org/10.1109/MSP.2020.3016908.

[TBL16] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo. "Signals on graphs: uncertainty principle and sampling". In: *IEEE transactions on signal processing a publication of the IEEE Signal Processing Society*. 64.18 (2016), pp. 4845–4860. ISSN: 1053-587X.

[Val19] F. Vallentin. "Semidefinite programming bounds for error-correcting codes". In: *CoRR* abs/1902.01253 (2019). arXiv: 1902.01253. URL: http://arxiv.org/abs/1902.01253.