sampling complexity is essential. Our results are constructive and the sampling complexity is much larger than the number of samples used in practice. The other possibility is that the learning algorithm that is mostly used is not the algorithm that can find a good approximation. To get further insight into this question we continue this work with the analysis of different expressivity results with respect to their computability. We then aim to consider also a more restrictive family of functions to reduce the sampling complexity. Finally, the weights in the construction are very large and we aim to reduce them as well with more sophisticated architecture choices.

In the long run, we want to get a good understanding of the ground truth, especially in image classification to understand which problems are computable and sensible. In this line, a measure that allows us to investigate both the stability and accuracy of the network approximation is of high interest.

References

[1] K. Hornik, M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators*, Neural networks, **2** (1989), 359–366.

[2] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, *Understanding deep learning (still) requires rethinking generalization*, Communications of the ACM 64 **3** (2021) 107–115.

[3] H. Bölcskei, P. Grohs, G. Kutyniok, and P. Petersen, *Optimal approximation with sparsely connected deep neural networks*, SIAM Journal on Mathematics of Data Science, 1 **1** (2019) 8–45.

Norm bounds for a scattering transform on graphs

Bernhard G. Bodmann

(joint work with Íris Emilsdóttir)

Introduction

In the present work, we examine signals on graphs whose structure is motivated by time series analysis. In typical models for time series data, the values that occur closely together in time show a stronger dependence than observations that are spaced further apart [9]. However, this property may need to be modified to explain recurring patterns, such as daily or weekly periodicities in traffic intensities. We assume that the underlying periodicities are known and encoded in a graph structure, where neighboring vertices are immediate successors in time or related by a shift in time corresponding to a period of the observed process. Based on the graph structure, one may devise a type of scattering transform in the spirit of Mallat's method to generate feature vectors with convolutional networks in a non-adaptive way [1, 7]. This has been done by Zou and Lehrman [11] based on graph wavelets [4], see also [2]. Here, we pursue a parallel strategy that is based on heat kernels.

Preliminaries

An oriented graph (V, E) is described by a vertex set V and an edge set E, for which E contains *ordered* pairs of vertices. When considering a directed graph, we also speak of an edge without orientation when passing from $(i, j) \in E$ to $\{i, j\}$. Two edges are adjacent if they have a vertex in common. A graph is connected if any two vertices in V appear in a sequence of vertices such that each pair of consecutive elements in this sequence forms an edge. A directed graph is weakly connected if any two vertices appear in a sequence of adjacent edges without orientation.

The Hilbert space $\ell^2(V)$ is the space of all real-valued functions $f : V \to \mathbb{R}$, equipped with the canonical inner product that associates $\langle f, g\rangle = \sum_{j\in V} f(j)g(j)$ with $f, g \in \ell^2(V)$. The standard graph Laplacian Δ is the self-adjoint operator corresponding to the quadratic form defined by $Q(f) = \sum_{i,j\in E} |f(i) - f(j)|^2$ for $f \in \ell^2(V)$. Given a directed graph (V, E) and a function $a : E \to \mathbb{R}$, we let Δ_a be the operator on $\ell^2(V)$ corresponding to $Q_a(f) = \sum_{(i,j)\in E} |e^{a(i,j)} f(i) - f(j)|^2$. In this context, we call the function a a connection. Finally, for two functions w, a on the edge set of a directed graph with w assuming only strictly positive values, we let $\Delta_{w,a}$ be defined via $Q_{w,a}(f) = \sum_{(i,j)\in E} w(i, j)|e^{a(i,j)} f(i) - f(j)|^2 = -\langle \Delta_{w,a} f, f\rangle$. We then say that the edges are weighted by w. If there is $\phi : V \to \mathbb{R}$ such that for each $(i, j) \in E$, $a(i, j) = \phi(j) - \phi(i)$, then we say that a is a gradient function.

Main results

With the help of the Laplacian $\Delta_{w,a}$, we define a cascade of transforms.

Definition. Let (V, E) be a directed graph with weights $w : E \to \mathbb{R}^+$, connection $a : E \to \mathbb{R}$ and Laplacian $\Delta_{w,a}$, $\langle \Delta_{w,a} f, f\rangle = -\sum_{\langle i,j\rangle} w_{i,j} |e^{a(i,j)} f(i) - f(j)|^2$. For $\epsilon > 0$, $f : V \to \mathbb{R}$, let $S_0(f) = f$, we inductively set for $m \in \mathbb{N}$

$$S_m(f) = (I - (I - \epsilon\Delta_{w,a})e^{\epsilon\Delta_{w,a}})^{1/2} |S_{m-1}(f)|$$

and

$$T_m(f) = (I - \epsilon\Delta_{w,a})^{1/2} e^{\frac{\epsilon}{2}\Delta_{w,a}} |S_{m-1}(f)| \,.$$

Here, for any function g on V, $|g| = \max\{g, -g\}$.

It is a direct consequence of this definition that the norm of a signal is preserved under this transform.

Proposition. Let (V, E), w, a, and $\Delta_{w,a}$, S_m and T_m be as above, then for $N \in \mathbb{N}$,

$$\|f\|^2 = \|S_N(f)\|^2 + \sum_{m=1}^{N} \|T_m(f)\|^2$$

Next, we observe that if a is a gradient function, then the functions in the kernel of the Laplacian saturate the norm of T_1.

Proposition. Let (V, E) be a weakly connected, directed graph, w, a, and $\Delta_{w,a}$, S_1 and T_1 as above, and there is $\phi : V \to \mathbb{R}$ such that $a(i,j) = \phi(j) - \phi(i)$, then $\|T_1(f)\| = \|f\|$ if and only if there is $c \in \mathbb{R}$ and for each $i \in V$, $f(i) = ce^{\phi(i)}$.

More generally, we study the behavior of the norm when applying S_1 or T_1, with a similar motivation as in [10]. Here, we relate the norm to the Rayleigh quotient of the Laplacian. Because of the Parseval-type identity $\|f\|^2 = \|S_1(f)\|^2 + \|T_1(f)\|^2$, it is enough to investigate T_1.

Theorem. Let (V, E) be a weakly connected, directed graph with weights w, connection a, and Laplacian $\Delta_{w,a}$ such that λ_1 is the first non-zero eigenvalue of $-\Delta_{w,a}$, and $\lambda_{\max}$ the maximal eigenvalue of $-\Delta_{w,a}$. If $\epsilon > 0$, $0 \le \rho \le \lambda_1$, and

$$-\langle \Delta_{w,a} f, f\rangle = \rho \|f\|^2$$

then T_1 as defined above satisfies

$$\|T_1(f)\|^2 \ge \big[(1 - \frac{\rho}{\lambda_1}) + (1 + \epsilon\lambda_{\max}) e^{-\epsilon\lambda_{\max}} \frac{\rho}{\lambda_{\max}}\big] \|f\|^2$$

and

$$\|T_1(f)\|^2 \le \big[(1 - \frac{\rho}{\lambda_{\max}}) + (1 + \epsilon\lambda_1) e^{-\epsilon\lambda_1} \frac{\rho}{\lambda_1}\big] \|f\|^2 .$$

Corollary. If the assumptions of the preceding theorem hold, $f \in \ell^2(V) \setminus \{0\}$, and $\epsilon > 0$ is sufficiently small so that $(1 + \epsilon\lambda_1)e^{-\epsilon\lambda_1}\lambda_{\max} > \lambda_1$ then

$$\frac{\|T_1(f)\|^2/\|f\|^2 - 1}{(1 + \epsilon\lambda_1)e^{-\epsilon\lambda_1}/\lambda_1 - 1/\lambda_{\max}} \le \rho .$$

Estimating the Rayleigh quotient is useful when it is used as a statistic to infer whether an observed graph signal is consistent with a stochastic model for it. This will be pursued in an application to traffic counts in forthcoming work.

References

[1] J. Bruna and S. Mallat, *Invariant scattering convolution networks*, IEEE Trans. Pattern Anal. Mach. Intell. **35** (8) (2013), 1872-1886.

[2] R. R. Coifman and M. Maggioni, *Diffusion wavelets*, Applied and Computational Harmonic Analysis **21**(1) (2006), 53–94.

[3] F. Gama, A. Ribeiro, and J. Bruna, *Diffusion Scattering Transforms on Graphs*, International Conference on Learning Representations, 2019, Paper 110.

[4] D. K. Hammond, P. Vandergheynst, and R. Gribonval, *Wavelets on graphs via spectral graph theory*, Applied and Computational Harmonic Analysis **30**(2) (2011), 129–150.

[5] A. Sandryhaila and J. M. F. Moura, *Discrete signal processing on graphs*, IEEE Trans. Signal Proc. **61** (2013), 1644–1656.

[6] A. Sandryhaila and J. M. F. Moura, *Discrete signal processing on graphs: Frequency analysis*, IEEE Trans. Signal Proc. **62** (2014), 3042–3054.

[7] Stéphane Mallat, *Group invariant scattering*, Communications on Pure and Applied Mathematics **65** (10) (2012), 1331–1398.

[8] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, *Graph signal processing: Overview, challenges, and applications*, Proceedings of the IEEE **106** (2018), 808–828.

[9] M. B. Priestley, *Spectral Analysis and Time Series*, vols. 1 & 2, Academic Press, London, 1981.

[10] T. Wiatowski, P. Grohs, H. Bölcskei, *Energy propagation in deep convolutional neural networks*, IEEE Trans. Inform. Theory **64** (7) (2017), 4819–4842.
[11] D. Zou and G. Lerman, *Graph convolutional neural networks via scattering*, Applied and Computational Harmonic Analysis, **49**(3) (2020), 1046-1074.

Gaussian lattice sums

Markus Faulhuber
(joint work with Laurent Bétermin, Stefan Steinerberger)

1. Definition and Main Result

In [2], we characterize optimizers for a variational problem with applications in various fields. Let Λ be a lattice in $\mathbb{R}^2$ and consider the function

$$E_\Lambda(z;\alpha) = \sum_{\lambda\in\Lambda} e^{-\pi\alpha|\lambda+z|^2} \qquad z\in\mathbb{R}^2,\ \alpha>0. \tag{1}$$

The function $E_\Lambda(z;\alpha)$ is simply the sum of (scaled) Gaussians centered at points given by a (shifted) lattice: it may thus be understood as the two-dimensional analogue of a Jacobi theta function. Given the fundamental nature of this object, the function $E_\Lambda(z;\alpha)$ naturally arises in many different areas of mathematics. In [2], we are concerned with minimizing and maximizing the function $E_\Lambda(z;\alpha)$. The canonical candidate for solving the variational problem is the hexagonal lattice;

$$\Lambda_2 = \sqrt{\tfrac{2}{\sqrt{3}}}\begin{pmatrix}1 & \frac{1}{2}\\ 0 & \frac{\sqrt{3}}{2}\end{pmatrix}\mathbb{Z}^2.$$

Theorem (Montgomery, 1988)**.** *Among all lattices $\Lambda\subset\mathbb{R}^2$ with fixed density,*

$$\max_{z\in\mathbb{R}^2} E_\Lambda(z;\alpha) \qquad \textit{is minimized}$$

if and only if Λ is the hexagonal lattice Λ_2.

Main Result (Bétermin, Faulhuber, Steinerberger, 2021)**.** *Among all lattices $\Lambda\subset\mathbb{R}^2$ with fixed density,*

$$\min_{z\in\mathbb{R}^2} E_\Lambda(z;\alpha) \qquad \textit{is maximized}$$

if and only if Λ is the hexagonal lattice Λ_2.

One nice aspect of Montgomery's result is that the maximum is assumed in a lattice point; in contrast, we have relatively little control over the point z in which the minimum is assumed (see Figure 1) which makes the proof significantly harder. One important consequence of our Main Result is that the hexagonal lattice maximizes the minimum while *simultaneously* minimizing the maximum of E_Λ (the latter being due to Montgomery). We expect this to be a very rare property if (1) is generalized to higher dimensions. This reaffirms the special role that the hexagonal lattice Λ_2 plays for variational problems in $\mathbb{R}^2$. The Main Result has many consequences, one of which is, in combination with Montgomery's result, that the conjecture of Strohmer and Beaver [9] is finally affirmed.