Bringing Approximation Theory to Data Science

Introduction. The field of approximation theory goes back as far as the 19th century with perhaps what is essentially the first textbook in approximation theory arising in the early 20th century from Émile Borel in 1905 [1]. Much of classical approximation theory deals with approximating functions with of polynomials, e.g., Chebyshev series, Fourier series, and Laurent series. With the dawn of the computer era in mid-20th century, this became especially important for evaluating functions such as $\sin(x)$ or e^x up to arbitrary (machine) precision. Thus approximation theory was foundational to the fields of numerical methods and numerical analysis.

Approximation theory has always fundamentally been tied to computing, and as computers changed, so did research in approximation theory. Splines are one of the most classical objects in approximation theory dating back to the 1940s when Isaac Schoenberg invented the spline [2]. Wavelet systems are another classical object which, one could argue, dates back to Alfréd Haar's thesis in 1910 where he introduced the Haar basis, the simplest wavelet system [3]. As computing evolved, splines and wavelets were found to be very useful for signal and image processing [4] as a way to bridge the gap between the digital (discrete) and analog (continuous) domains with the notions of analysis and synthesis operators. Neural networks are a more contemporary object in approximation theory which were of great interest in the 1980s and 1990s due to their interpolation and approximation properties [5, 6], with one of the most fundamental results being the universal approximation theorem [5]. Other work brought in tools from classical harmonic analysis (Littlewood–Paley theory) as well as computational/applied harmonic analysis (wavelet analysis) to understand how to effectively construct neural networks to approximate a given function [6].

With the (empirical) success of deep learning in the last few years, a theoretical understanding of how exactly these neural network models work is vital. To build this understanding, a *reemergence of neural network approximation theory* has occurred, which includes some of my own work (currently submitted to the *SIAM Journal on Mathematics of Data Science*) where I develop a general framework that explains the interpolation properties of overparameterized two-layer $\mathbb{R} \to \mathbb{R}$ feedforward neural networks through the lens of spline theory. A preprint is available on arXiv [7].

Research Plan. I would first like to generalize our preliminary result and work towards developing a general theory of the interpolation properties of feedforward neural networks (deep and multivariate architectures) based on splines. Such a theory also provides us a purely spline-theoretic corollary regarding *spline representations*. The most common spline representation is the B-spline basis. Another common spline representation follows from associating a spline to a linear and translation-invariant operator and then representing the spline as a linear combination of shifted Green's functions of this operator. A spline theory of general feedforward neural networks would lead to a new spline representation that is the neural network itself. The B-spline basis is preferred over the Green's function representation since B-spline basis functions are bounded and compactly supported by construction, while Green's functions are not. A neural network spline representation may have some "nice" properties over these two standard representations, e.g., a neural network representation would improve computing multivariate splines since classical methods become computationally intractable for dimensions more than, say, ten, and we already use neural networks in practice with very high-dimensional inputs.

Another research question I would like to explore is the notion of *self-similarity* in the context data science. Wavelets became popular in the image processing literature since natural images exhibit self-similarity. Understanding if *natural datasets* exhibit some notion of self-similarity is not currently well understood. If tools from wavelet analysis (or other affine systems, e.g., Gabor systems) can be used to exploit self-similarity in datasets, then new machine learning algorithms could be developed to to exploit this structure.

My prior knowledge in functional analysis, Fourier/harmonic analysis, and machine learning theory puts me in a unique position to attack these fundamental theoretical problems in data science. This work will be in collaboration with my advisor Prof. Robert Nowak. Being at the University of Wisconsin–Madison also puts me in a unique position since many approximation theory experts have been or currently are faculty. Former/emeritus faculty include (the late) Prof. Isaac Schoenberg (who invented the spline), Prof. Carl de Boor, and Prof. Grace Wahba. The current faculty include Prof. Robert Nowak and Prof. Amos Ron. I'm also in the unique position that the University of Wisconsin–Madison has the Institute for Foundations of Data Science (IFDS), directed by Prof. Stephen Wright. The IFDS is supported through the *NSF's Transdisciplinary Research in Principles of Data Science (TRIPODS) initiative* with the mission of advancing the mathematical foundations of data science. I am an active participant in the IFDS, giving me access to faculty across many departments who are experts in mathematical data science. With the availability of experts in both approximation theory and data science, the NSF GRF will allow me to collaborate with any of these professors without being tied to any specific grant or research topic.

Intellectual Merit. The proposed research will lead to theoretical understanding of important data science problems using tools from approximation theory. In particular, an application of this research will aid in understanding the "mysteries" or "magic" of deep learning from a spline-theoretic perspective. This research will also lead to using neural networks for spline representations, a purely spline-theoretic result. The proposed research will also address finding hidden structure in datasets using notions of self-similarity, leading to better machine learning algorithms that can be used in practice. With the current craze about "big data" and machine learning, the ability to exploit any kind of structure to improve performance of algorithms is critical.

Broader Impact. The broader impacts of the proposed research are expected to be significant as it will give us a theoretical understanding of the neural network models employed in practice and will thus impact applications ranging from image processing and robotics to smart cars and unmanned aerial vehicles. I plan to disseminate my work through publications as well as presentations understandable to a general audience. I am a part of the LUCID NSF Research Traineeship Program which promotes interdisciplinary research between data science/machine learning and cognitive science/psychology. I have already had experience presenting my (rather theoretical) work in [7] at our weekly HAMLET seminar to an audience composed of cognitive scientists and psychologists who are interested in understanding deep learning and its connections to human neural processing. I plan to continue such interdisciplinary presentations.

References.

[1] Émile Borel. *Leçons sur les fonctions de variables réelles et les développements en séries de polynomes*. Paris: Gauthier-Villars, 1905.

[2] Isaac J. Schoenberg. "Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions". In: *Quarterly Applied Math* 4 (1946), pp. 45–99.

[3] Alfréd Haar. "Zur theorie der orthogonalen funktionensysteme". In: *Mathematische Annalen* 69.3 (1910), pp. 331–371.

[4] Michael Unser. "Splines: A perfect fit for signal and image processing". In: *IEEE Signal processing magazine* 16 (1999), pp. 22–38.

[5] George Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314.

[6] Emmanuel J. Candès. "Harmonic analysis of neural networks". In: *Applied and Computational Harmonic Analysis* 6.2 (1999), pp. 197–218.

[7] Rahul Parhi and Robert D. Nowak. *Minimum "Norm" Neural Networks are Splines*. 2019. arXiv: `1910.02333 [stat.ML]`.