
RAHUL PADMANABHAN, Concordia University

Deep Learning Approximation of Matrix Functions: From Feedforward Neural Networks to Transformers

Deep Neural Networks (DNNs) have been at the forefront of Artificial Intelligence (AI) over the last decade. Transformers, a type of DNN, have revolutionized Natural Language Processing (NLP) through models like ChatGPT, Llama and more recently, Deepseek. While transformers are used mostly in NLP tasks, their potential for advanced numerical computations remains largely unexplored. This presents opportunities in areas like surrogate modeling and raises fundamental questions about AI's mathematical capabilities.

We investigate the use of transformers for approximating matrix functions, which are mappings that extend scalar functions to matrices. These functions are ubiquitous in scientific applications, from continuous-time Markov chains (matrix exponential) to stability analysis of dynamical systems (matrix sign function). Our work makes two main contributions. First, we prove theoretical bounds on the depth and width requirements for ReLU DNNs to approximate the matrix exponential. Second, we use transformers with encoded matrix data to approximate general matrix functions and compare their performance to feedforward DNNs. Through extensive numerical experiments, we demonstrate that the choice of matrix encoding scheme significantly impacts transformer performance. Our results show strong accuracy in approximating the matrix sign function, suggesting transformers' potential for advanced mathematical computations.