



THE UNIVERSITY OF  
CHICAGO

DEPARTMENT OF STATISTICS

## MASTER'S THESIS PRESENTATION

---

SIMIAO JIAO

Department of Statistics  
The University of Chicago

LU decomposition and Toeplitz decomposition of a neural network

THURSDAY, April 20, 2023, at 4:30 PM

Jones 111, 5747 S. Ellis Avenue

### ABSTRACT

The LU decomposition of a matrix  $A$  is a well-known concept. However, it is less known that  $A$  also has a 'Toeplitz decomposition'  $A = T_1 T_2 \cdots T_r$ , where  $T_i$ 's are Toeplitz matrices. In this paper, we prove that any continuous function  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  can be approximated to arbitrary accuracy by a neural network that takes the form  $L_1 \sigma_1 U_1 \sigma_2 L_2 \sigma_3 U_2 \cdots L_r \sigma_{2r-1} U_r$ . The weight matrices alternate between lower and upper triangular matrices, and  $\sigma_i(x) \leq \sigma(x - b_i)$  for some bias vector  $b_i$ . Additionally, the activation function  $\sigma$  can be chosen to be essentially any uniformly continuous non-polynomial function. The same result also holds with Toeplitz and Hankel matrices, meaning that  $f$  can be approximated to arbitrary accuracy by  $T_1 \sigma_1 T_2 \sigma_2 \cdots \sigma_{r-1} T_r$ .

As our results apply to general neural networks, they can be regarded as LU and Toeplitz decompositions of a neural network. One of the practical implications of our results is that the number of weight parameters in a neural network can be significantly reduced without sacrificing its power of universal approximation. In fact, we show that imposing these structures on the weight matrices reduces the number of training parameters while having almost no noticeable effect on test accuracy.

Our paper also presents several experiments on real data sets that further demonstrate the effectiveness of these techniques. Furthermore, we provide a fixed-width universal approximation theorem for convolutional neural networks, which have only previously had arbitrary width versions.