



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

MASTER'S THESIS PRESENTATION

YUCONG LIU

Department of Statistics
The University of Chicago

LU decomposition and Toeplitz decomposition of a neural network

THURSDAY, April 20, 2023, at 1:00 PM

Jones 304, 5747 S. Ellis Avenue

ABSTRACT

It is well-known that any matrix A has an LU decomposition. Less well-known is the fact that it has a 'Toeplitz decomposition' $A = T_1 T_2 \cdots T_r$ where T_i 's are Toeplitz matrices. We will prove that any continuous function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ has an approximation to arbitrary accuracy by a neural network that takes the form $L_1 \sigma_1 U_1 \sigma_2 L_2 \sigma_3 U_2 \cdots L_r \sigma_{2r-1} U_r$, i.e., where the weight matrices alternate between lower and upper triangular matrices, $\sigma_i(x) \coloneqq \sigma(x - b_i)$ for some bias vector b_i , and the activation σ may be chosen to be essentially any uniformly continuous nonpolynomial function. The same result also holds with Toeplitz matrices, i.e., $f \approx T_1 \sigma_1 T_2 \sigma_2 \cdots \sigma_{r-1} T_r$ to arbitrary accuracy, and likewise for Hankel matrices. A consequence of our Toeplitz result is a fixed-width universal approximation theorem for convolutional neural networks, which so far have only arbitrary width versions. Since our results apply in particular to the case when f is a general neural network, we may regard them as LU and Toeplitz decompositions of a neural network. The practical implication of our results is that one may vastly reduce the number of weight parameters in a neural network without sacrificing its power of universal approximation. We will present several experiments on real data sets to show that imposing such structures on the weight matrices sharply reduces the number of training parameters with almost no noticeable effect on test accuracy.