

NUMERICAL ANALYSIS AND SCIENTIFIC COMPUTING
SEMINAR

Structure-conforming Operator Learning via Transformers

Shuhao Cao
University of Missouri-Kansas City

Abstract: GPT, Stable Diffusion, AlphaFold 2, etc., all these state-of-the-art deep learning models use a neural architecture called "Transformer". Since the emergence of "Attention Is All You Need" paper by Google, Transformer is now the ubiquitous architecture in deep learning. At Transformer's heart and soul is the "attention mechanism". In this talk, we shall dissect the "attention mechanism" through the lens of traditional numerical methods, such as Galerkin methods, and hierarchical matrix decomposition. We will report some numerical results on designing attention-based neural networks according to the structure of a problem in traditional scientific computing, such as inverse problems for Neumann-to-Dirichlet operator (EIT) or multiscale elliptic problems. Progresses within different communities will be briefed to answer some open problems on the mathematical properties of the attention mechanism in Transformers.

Thursday, March 21, 2024, 10:00 am
Mathematics and Science Center: MSC W201

MATHEMATICS
EMORY UNIVERSITY