ACMS Applied Math Seminar

Ruhui Jin
John Hopkins University

Wednesday, Nov. 11, 2026
154 Hurley Hall
3:30 PM – 4:30 PM

Data Selection for Reliable Learning: Probabilistic Experimental Design

Much effort in scientific machine learning is devoted to building more capable and efficient learning architectures. A complementary direction is to ask whether we can be more judicious about the data these methods rely on. This is particularly relevant to *inverse problems* - recovering the underlying mechanisms of physical and biological systems from indirect observations. However, practical data are rarely perfect: noise contamination, redundant information, and other limitations can largely jeopardize the quality of reconstruction. Rather than collecting arbitrary data, we turn to *optimal experimental design*, a statistical framework for guiding data acquisition to reduce uncertainty in inference. In PDE-based inverse problems, where should sensors be placed, when should snapshots be taken, which angles and frequencies should be chosen for incident waves? In this talk, we introduce the mathematical foundation of probabilistic experimental design to identify valuable measurements. We conclude by discussing how randomized algorithms can improve the computational efficiency of data selection, but at the expense of design optimality.