Mathematics Colloquium

Mathematics of AI safety: The power of convexity in a highly nonconvex world

Brendon Anderson
Cal Poly, San Luis Obispo

Friday, November 1, 2024
11:10 am – 12 pm
Building 180, Room 113

Abstract

Despite the many notable achievements of artificial intelligence (AI), it has been shown to exhibit catastrophic failures when subjected to adversarial manipulation. In order to safely deploy AI in mission-critical applications, such as autonomous driving and healthcare systems, robust performance in the presence of adversaries must be guaranteed. This talk begins by reviewing the optimization-based framework for certifying the robustness of machine learning (ML) models. We then present recent advancements in the theory and computation of such robustness certificates. We will emphasize how the presented methods all intimately utilize the mathematics of convex analysis and convex optimization theory in order to rigorously prove robustness guarantees for highly nonconvex ML models.

About the speaker: Brendon Anderson is an Assistant Professor in the Mechanical Engineering Department at California Polytechnic State University. He earned his Ph.D. in Mechanical Engineering and his M.A. in Mathematics at the University of California, Berkeley, where he was advised by Professor Somayeh Sojoudi. Prior to that, he earned his B.S. in Mechanical Engineering at UCLA, where he conducted research in the Applied Math Lab under the supervision of Professor Andrea Bertozzi. His research spans interdisciplinary areas, including machine learning, optimization, control theory, and game theory, with an emphasis on proving mathematical guarantees of reliability and robustness.