THE UNIVERSITY  OF HONG KONG

*Institute of Mathematical Research*
*Department of Mathematics*

# "MATH for Machine Learning" Seminar

# On the Proof of Global Convergence of Gradient Descent for Deep ReLU Networks with Linear Widths

**Mr. Tianrui He**
HKU

## Abstract

Why can gradient descent converge to a global minimum when training a deep neural network, despite the nonconvexity of its loss function? This talk presents a proof of global convergence for deep ReLU networks trained with the square loss, based on the work of Quynh Nguyen (2021). We begin by introducing the basic structure of ReLU networks and the gradient descent algorithm. We then explain the key observation connecting the output-layer gradient, the training loss, and the smallest singular value of the last hidden layer's feature matrix. Building on this observation, we present the main convergence theorem and its proof strategy: controlling weight updates, preserving the nondegeneracy of the feature matrix, and establishing linear convergence of the training loss. The analysis avoids explicitly tracking changes in ReLU activation patterns or the entire neural tangent kernel. Finally, we discuss the roles of initialization and network width, clarifying the conditions under which a single wide hidden layer with width proportional to the number of training samples is sufficient.

| | |
|---|---|
| Date: | November 19, 2026 (Thursday) |
| Time: | 3:00 pm – 4:00 pm |
| Venue: | Room 210, Run Run Shaw Building<br>HKU |

*All are welcome*