



復旦大學  
FUDAN UNIVERSITY

# Applied Math Ph.D. Seminar

## Optimal Learning Rate Schedules under Functional Scaling Laws: Power Decay and Warmup-Stable-Decay

**Speaker:** Zilin Wang (Peking University)

**Time:** 2026-03-19, 16:10 to 17:00

**Location:** Rm 1801, Guanghua East Tower

**Advisor:** Lei Wu (Peking University)

**Abstract:** We study optimal learning rate schedules (LRSs) under the functional scaling law (FSL) framework, where the loss is written explicitly as a functional of the schedule. This formulation reveals a sharp phase transition governed by two exponents: a source exponent  $s > 0$  controlling signal learning and a capacity exponent  $\beta > 1$  controlling noise forgetting. In the easy-task regime, the optimal schedule follows a power decay to zero, with exponent determined by  $\beta$ . In the hard-task regime, the optimal schedule takes a warmup-stable-decay (WSD) form: it maintains the largest admissible learning rate for most of training and decays only near the end, with a vanishing decay fraction. We further analyze shape-fixed schedules, showing how the tail exponent governs both their optimality and their limitations through capacity saturation. This yields a principled evaluation of commonly used schedules such as cosine and linear decay. Finally, we apply the power-decay LRS to one-pass SGD for kernel regression and show that the last iterate attains the exact minimax-optimal rate.