

# Kernel Truncated Randomized Ridge Regression: Optimal Rates and Low Noise Acceleration

Kwang-Sung Jun (The University of Arizona)

Ashok Cutkosky (Google Research)

Francesco Orabona (Boston University)

# Setup

- We consider the problem of nonparametric regression in Reproducing Kernel Hilbert Space (RKHS).
- We follow the standard parameterization of the problem complexity parameterization  $(b, \beta)$  where
  - $b$  is the eigenvalue decay rate of the integral operator and
  - $\beta$  is a complexity measure of the optimal predictor (related to its norm).

# Contributions

1. We achieve the optimal rate in certain problem regime on  $(b, \beta)$  (previously called a “hard regime”), resolving a long-standing open problem.
2. We also show an even faster convergence is possible when the Bayes error is 0.
3. Furthermore, when Bayes error is 0, the best regularization is 0, which connects to recent interest on the generalization ability of the interpolator.

# Key ingredients for the proof

1. Online-to-batch conversion:

Our algorithm is essentially an online learning algorithm at its heart, but we turn it into a batch algorithm with randomization.

2. “The identity” for Kernel Ridge Regression (KRR)\*:

A known, but rather obscure result that the *online cumulative prediction error* of KRR, adjusted by some weights, is exactly equal to the minimum of the *batch regularized training error* objective.

\*Zhdanov, Fedor, and Yuri Kalnishkan. "An identity for kernel ridge regression." In International Conference on Algorithmic Learning Theory, pp. 405-419, 2010.