

# Improved Nystrom Kernel Matrix Approximation for Large-Scale Learning: Practical and Theoretical Aspects

Farhad Pourkamali

Department of Computer Science, UMass Lowell

**Abstract:** The Nystrom method is a popular technique for generating fixed-rank approximations of large kernel matrices that arise in many machine learning problems. The approximation quality of the Nystrom method depends crucially on the number of selected landmark points and the selection procedure. In practice, to ensure high quality approximations, the number of landmark points is chosen to be greater than the target rank. However, the standard Nystrom method uses a sub-optimal procedure for rank reduction. This talk highlights the drawbacks of standard Nystrom in terms of poor performance and lack of theoretical guarantees. To address these issues, we present an efficient method for generating improved fixed-rank Nystrom approximations. Furthermore, we present a randomized algorithm for generating landmark points that is scalable to large high-dimensional data sets and streaming scenarios. The proposed method performs K-means clustering on low-dimensional random projections of a data set and thus leads to significant savings. Extensive numerical experiments on classification and regression tasks demonstrate the superior performance and efficiency of the proposed method compared with state-of-the-art methods.