Using the power of iterative methods for the SVD in machine learning

Andreas Stathopoulos[1], J. Myers[1], L. Wu[2], E. Romero[1], F. Xu[1], Z. Liu[1]

[1] *Department of Computer Science, College of William & Mary, Williamsburg, VA, USA*
[2] *IBM Watson, USA*

Abstract

Machine learning has emerged as one of the primary clients for large scale singular value calculations. The matrices are typically very large so that dense SVD methods cannot be used. In computational sciences, iterative methods have been used to solve these type of problems, with methods such as the recently proposed Golub-Kahan-Davidson pushing the envelope in accuracy and convergence speed. In machine learning, combinations of randomized and iterative methods have risen drastically in popularity because of their extreme efficiency.

One important constraint in machine learning is the size and nature of the matrices. In some applications, matrices have no exploitable sparsity, which makes the matrix-vector multiplication expensive in terms of operations, and even more expensive in terms of memory accesses. In other applications, the matrix may only be available as streaming sets of columns or rows. However, the singular values often decay fast, giving rise to a near low rank structure.

Another important variable is how much of the singular space is needed and how accurately. Several applications require a small number (1-4) smallest or largest singular triplets to a good but not high accuracy. Others require the computation of a low rank approximation of the entire matrix. In these cases, low accuracy is sufficient, especially when the matrix itself is near low rank.

Given the size of the problems, one way to obtain a low rank approximation is to perform a randomized projection of the matrix down to a more manageable dimensionality where we can solve the SVD directly. The singular space approximations obtained by the randomized projection can be improved iteratively by applying a basic subspace iteration. These techniques have proven particularly efficient, requiring only 1-2 subspace iterations, and although the computed singular vectors are not accurate, the low rank approximation is. This has puzzled numerical analysts and only recently has research emerged that starts to explain the phenomenon.

In this talk, we examine the relative merits of these two classes of methods based on this recent research and some new observations. First, we observe that traditional block iterative methods are also randomized. Then we show that the difference between randomized and traditional iterative methods is only in their stopping criteria. Specifically, the requirement for a good low rank approximation is a weaker condition that does not depend on the singular gaps or even the angles between our approximations and the singular vectors. Therefore, a common framework is possible with flexible stopping criteria and variable block size for different problems.

References

[1] S. Goldenberg, A. Stathopoulos, Copper Mountain Conference on Iterative methods, 2018

[2] M. Gu, SIAM Journal on Scientific Computing 37,(3),2014

[3] I. Ipsen and P. Drineas, arXiv:1801.00670, 2018