Juxtaposition of Analytical and Empirical Performances of Random Sampling Methods in Randomized Algorithms

Dongping Qi, Tianyi Shi, Lily Wang

October 26, 2017

Abstract

In existing literatures, randomized algorithms have proven to be powerful tools to deal with various matrix problems. While random matrices generated by Gaussian distribution are most investigated and widely utilized, there are other random sampling methods that gives distinct advantages in terms of time complexity, robustness, and preservation of certain structures under different randomized algorithmic settings. This project explores three random sampling methods – approximation by Gaussian random matrices, *subsampled random Fourier transform* (SRFT), and approximation by *sparse sketching matrices*- under the settings of three randomized algorithms – *singular-value decomposition* (SVD), *least squares problems* (LS), and *single-pass algorithm* related problems. We will specifically review features such as speed, accuracy, special characteristics, etc. of the three random sampling methods in existing literatures, as well as analyze their empirical performances through conducting experiments on each combination of random sampling method and randomized algorithm.

Key words. randomized algorithms, Gaussian test matrix, SRFT, sparse sketching, SVD, LSR, single-pass algorithm

1 Introduction

Constructing a low-rank approximation for a general matrix assumes a key role in various fields such as data analysis and scientific computing. Recent researches demonstrates that randomized algorithms can be of great use in solving this problem[3]. Several randomized algorithms have already been developed and examined in dealing with this problem, and the main idea consists of generating a random matrix Q and using matrix operation to construct an approximation $A^* \approx A$.

Among all such randomized algorithms, the algorithm which takes on a Gaussian test matrix as our Q is most used. Gaussian random matrices (matrices all of whose entries are generated by the normal distribution $\mathcal{N}(0,1)$) are the top choice because they are well studied analytically, from which abundant analytical properties were derived. However, the random matrix Q is not restricted to the Gaussian class. There are several other ways to generate the test matrix and these methods may show more desirable characteristics than Gaussian test matrices do in specific cases. Among all these methods, two typical ones are subsampled random Fourier transform (SRFT) and sparse sketching method. The former incorporates a discrete Fourier transform multiplied by two random matrices[6], whereas the latter adopts a subspace embedding technique.

Besides matrix approximation, these randomized algorithms are also applicable to many other problems. For example, we can use single-pass algorithm to solve eigenvalue decomposition problems. Single-pass algorithm is preferable when the entries of our target matrix A are not easily accessible, as the algorithm traverses the entries of A only once [3] and avoids repeated matrix multiplications.

2 Objectives

In this project, we intend to make sense of theories of and test the performances of three specific randomized methods: the prototype algorithm using Gaussian test matrix, SRFT method and sparse sketching method. All these methods are applied to three different problems: SVD, least squares regression and single-pass problems. We will evaluate their effectiveness through three basic properties: the accuracy of matrix approximation, the efficiency of implementation and the robustness of each algorithm. Several numerical examples will be provided for the purpose of comparison and a few applications will also be included.

2.1 Theories

We will start by introducing the random sampling methods we use – Gaussian random matrices, SRFT and sparse sketching, and randomized algorithms we aim to implement – singular value decomposition, least squares problem and single-pass algorithm in solving problems such as eigenvalue decomposition. In their 2009 review, Halko, Martinsson, and Tropp discussed solving SVD and single-pass EVD with Gaussian random matrices[3]. We will follow this paper for proof of successful functioning and error bounds for these two complete algorithms. Another paper by Meng, Saunders, and Mahoney analyzes LSRN, an iterative solver utilizing Gaussian random matrix to solve least squares problems[4]. We will then look closely at a 2008 work by Woolfe, Liberty, Rokhlin and Tygert, where a discussion of solving the SVD using SRFT is included[6]. We will especially examine the speed improvement of SRFT method compared to SVD using Gaussian random matrices. Tygert's 2009 paper on solving least squares problem with SRFT enables us to replicate the entire algorithm, proof of accuracy and computational cost of the problem[5]. Clarkson and Woodruff's recent paper takes a close look at analysis and implementations of SVD and least squares regression with sparse sketching matrices[1]. Finally, Clarkson and Woodruff's paper in 2009 on numerical linear algebra in streaming model offers us insights on the single-pass algorithm with sparse sketching[2]. For the final two papers, we will focus on the advantages of using sparse sketching on sparse matrices.

2.2 Implementations

After some scrutiny of the theoretical performances of said models, we will use Matlab to implement each of the nine problems and test our codes with designed target matrices. For the singular value decomposition, we will test our three random sampling methods on dense matrices with fast decaying spectrum, dense matrices with slow decaying spectrum and sparse matrices. Moreover, we will compare the speed among these sampling methods and examine whether SRFT performs faster as theory suggests. Accuracy will also be compared to see if exactness is compensated for speed. Similarly, we will test least squares problems and single-pass algorithms with different structures of matrices and make comparison of accuracy, speed and robustness between the random sampling methods. Our ultimate goal is to decide the optimal solver when we are given a problem and a target matrix with a known structure.

2.3 Applications

We will end our project with applications where these algorithms are tentatively suitable. Indeed, SVD and LS are essential in signal processing, noise reduction, etc. We will also give instances, such as difficulty to access original data, where single-pass algorithms are suitable for problem solving.

3 Qualifications

This project explores the topic of randomized algorithms and random sampling methods, which are closely related to the material all three applicants are taking class on. The project consists of two major sections: a literature review on the theoretical analysis of the three random sampling methods in different randomized algorithms, and empirical experiments on the models we discuss.

For the first section, certain level of acquaintance with linear algebra, probability theory, algorithm analysis and functional analysis is required. All three applicants are familiar with linear algebra, algorithm and probability to at least an undergraduate level. Specifically, Lily Wang is taking probability theory at a graduate level; Tianyi Shi is auditing algorithm at a graduate level; and Dongping Qi and Lily Wang is taking real analysis at a graduate level where functional analysis is covered.

For the second section, the experiments will be realized through implementation in Matlab. We will utilize several built-in functions and operators, such as *svd*, linear system solver *backslash* ($\backslash$), etc. All three applicants have some experience with programming in Matlab language. In this project, Dongping Qi and Tianyi Shi will take on most of the implementation work as both have completed projects in Matlab.

References

[1] Kenneth L Clarkson and David P Woodruff. "Low rank approximation and regression in input sparsity time". In: *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM. 2013, pp. 81–90.

[2] Kenneth L Clarkson and David P Woodruff. "Numerical linear algebra in the streaming model". In: *Proceedings of the forty-first annual ACM symposium on Theory of computing*. ACM. 2009, pp. 205–214.

[3] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". In: *SIAM review* 53.2 (2011), pp. 217–288.

[4] Xiangrui Meng, Michael A Saunders, and Michael W Mahoney. "LSRN: A parallel iterative solver for strongly over-or underdetermined systems". In: *SIAM Journal on Scientific Computing* 36.2 (2014), pp. C95–C118.

[5] Mark Tygert. "A fast algorithm for computing minimal-norm solutions to underdetermined systems of linear equations". In: *arXiv preprint arXiv:0905.4745* (2009).

[6] Franco Woolfe et al. "A fast randomized algorithm for the approximation of matrices". In: *Applied and Computational Harmonic Analysis* 25.3 (2008), pp. 335–366.