Homework for "Algorithms For Big Data Analysis"

Zaiwen Wen
Beijing International Center for Mathematical Research
Peking University

February 27, 2020

1 Submission Requirement

1. Prepare a report including
   - detailed answers to each question
   - numerical results and their iterpretation
2. The programming language can be either matlab, Python or c/c++.
3. Pack all of your codes named as "svd-ID-name.zip" and upload the file to
   `https://file.admin.cluster-bicmr.com/u/d/045d80868f524d0bab11/`
   作业提交需要统一打包成压缩文件，命名格式为：svd-学号-姓名，文件类型随意。文件名中不要出现空格，最好不要出现中文。
4. 请勿大量将代码粘在报告中，涉及到实际结果需要打表或者作图，不要截图或者直接从命令行拷贝结果。
5. 提交word 的同学需要提供word 原文件并将其转换成pdf 文件。
6. If you get significant help from others on one routine, write down the source of references at the beginning of this routine.

2 Randomized Singular Value Decomposition Algorithms

Given a matrix $A \in \mathbb{R}^{m\times n}$, compute p-largest singular values and their corresponding left and right singular vectors.

1. Write down and implement one of the algorithms in (extra credit for choosing both algorithms)
   - LinearTimeSVD Algorithm on page 166 of "Petros Drineas, Ravi Kannan, and Michael W. Mahoney, Fast Monte Carlo Algorithms for Matrices II: Computing a Low-Rank Approximation to a Matrix, SIAM J. Comput., 36(1), 158183"
   - Prototype for Randomized SVD on page 227 of "N. Halko, P. G. Martinsson, and J. A. Tropp, Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions,SIAM Rev., 53(2), 217288. "

2. Compute $r \in \{5, 10, 15, 20\}$ largest singular values and their corresponding singular vectors on a random matrix A generated as follows:

```
m = 2048;
n = 512;
p = 20;
A = randn(m,p)*randn(p,n);
```

3. Try one of the practical data sets in section 7 of

   - N. Halko, P. G. Martinsson, and J. A. Tropp, Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions,SIAM Rev., 53(2), 217288.

   The datasets can be found at
   `https://github.com/WenjianYu/rSVD-single-pass`
   DO NOT copy the codes online directly!

4. Extra-credit: Accelerate the speed for solving the following matrix completion problem using the randomized SVD techniques:

$$\min_{X \in R^{m \times n}} \quad \frac{1}{2} \sum_{(i,j) \in \Omega} (X_{ij} - M_{ij})^2 + \mu \|X\|_*.$$