NSF CSSI PI Meeting, Seattle, WA, Feb. 13-14, 2020

Award #: 1835821

CSSI Element: A high performance suite of SVD related solvers for machine learning

PI: Andreas Stathopoulos , Co-PIs: Zhenming Liu
Institutions: College of William & Mary

1.SVD methods in distance-based learning

➢ Introduction

Singular value decomposition (SVD) has been used for a wide range of applications:

- Recommender system
- Pattern (text/image) recognition
- Signal processing (adaptive beamforming modeling)

➢ Current stage

- ML researchers use SVD solvers as black boxes :
  - unaware of the properties for each SVD solvers
  - choose inappropriate solvers for tackling their own problems
- ML researchers develop tailor-made randomized algorithms (often sketches based) for solving SVD
  - not mature SVD solvers an
  - unlikely to beat carefully chosen SVD solvers

➢ Motivation

- Interplay between SVD solvers and ML algorithm
- Clarify the best practice guidelines
- Propose new algorithms
- Improve the current existing SVD related algorithms.

➢ Our contribution : we examine 3 categories

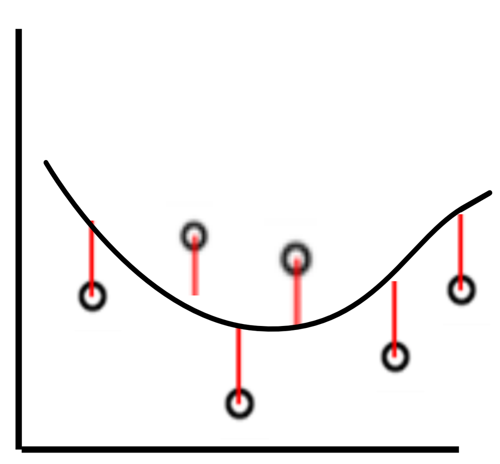
1. Kernel ridge regression and speedup methods

Our goal:

1. Identify the termination conditions with ML performance?

Assume $[\lambda_1, \lambda_2 \ldots \lambda_n]$ is the eigenvalues

- ❖ Use a hard threshold: keep the eigenvectors where $\lambda_i > \eta$ (η is around noise level)
- ❖ Keep the eigenvectors where gap $(\lambda_{i+1} - \lambda_{i+1}) > \sigma$
- ❖ Use the tail sum: $\sum_{i=1}^{k} \lambda_i > \Delta$
- ❖ By rank, use the first k eigenvectors



$$\beta = (\lambda I + K)^{-1} y$$

2. Speed up the kernel ridge regression for big and dense data ?

- More flexible
- with $O(N^3)$ training
- Hard for big data

2. Matrix completion

- Increased attention in recommender system
- Filling in the missing entries of a partially observed matrix

Movies

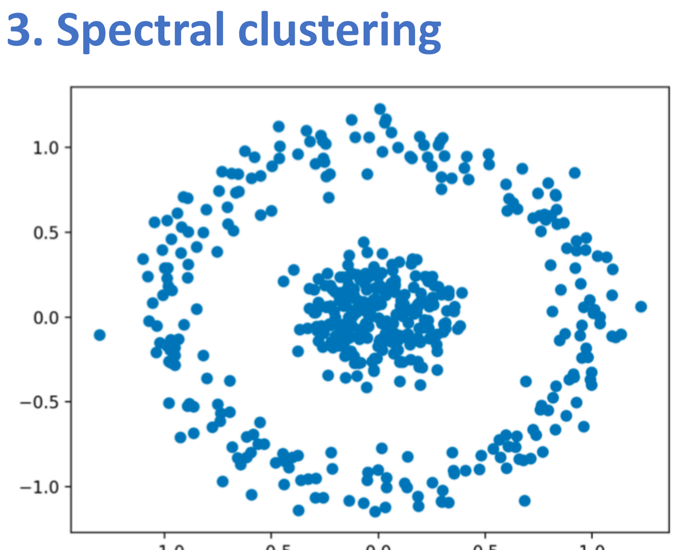
1	?	3	?
?	4	?	1
3	?	5	2

Users

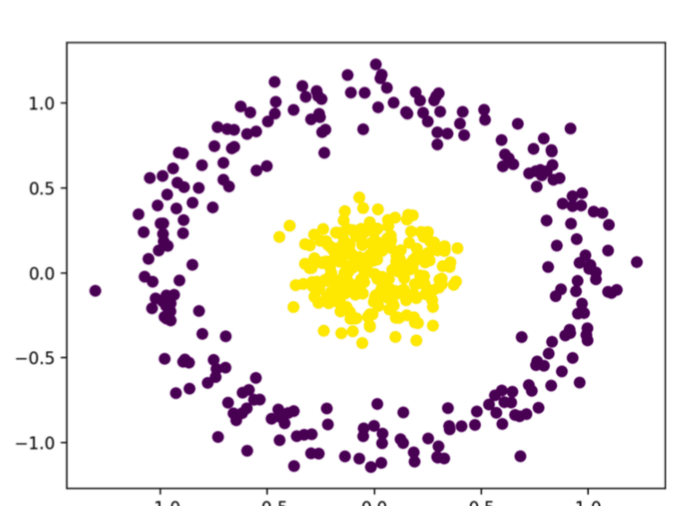
Our goal:

1. How to explore the limitations of these common terminal conditions ?
2. How to design dynamic terminal condition/ or improve the existing works with varying decay rates?

3. Spectral clustering



Kmeans is not appropriate



Spectral clustering results

Our goal:

1. Compare the properties of different graph Laplacians (unnormalized and normalized) and categorize all these algorithms.
2. Summarize the usage of SVD in different categorize

2. Low-Rank Stopping Criteria for Block Parallel SVD

➢ Introduction

Low rank approximation problem:

$$\begin{aligned} &\text{minimize} \quad \|Z\|_* \\ &\text{subject to} \quad \sum_{(i,j)\in\Omega} (A_{ij} - Z_{ij})^2 < \delta, \end{aligned}$$

where A is the $m \times n$ true data matrix, Z is an approximation, and Ω is set of the observed entries

➢ Motivation

- Most SVD algorithms only provide a residual stopping criterion - - iterative method is to limit
  - the number of total iterations,
  - matrix vector multiplications,
  - its execution time
- Lack guarantee on the accuracy

➢ Our contribution

- ➢ New stopping criteria for block parallel SVD algorithms
- ➢ Provide heuristics for dynamically changing both block and restart sizes when necessary
- ➢ Show their performance in both synthetic and real-world applications

Algorithm 2.1

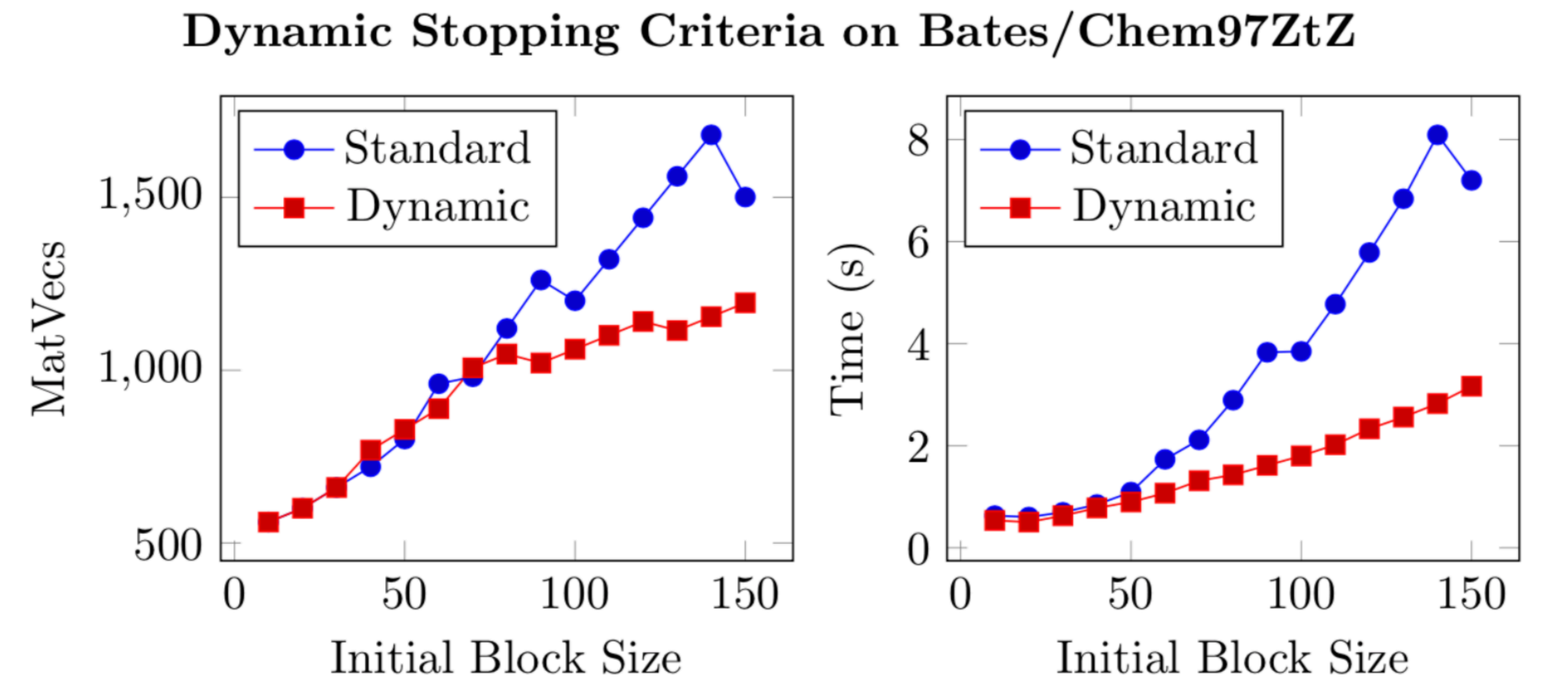
Current Basis Size $= c$, Max Basis $= m$, Restart Size $= s$, Block Size $= b$.
if $i > 2$ **then**
  if $1.5\tilde{k} > c$ **then**
    $m = \max(1.3\tilde{k}, \tilde{k} + 4b)$
    $s = \tilde{k} + b$
    if $4b > \tilde{k}$ **then**
      $b = \lfloor \tilde{k}/4 \rfloor$
    end if
  end if
end if

- Dynamic control is done with the heuristics given in Algorithm 2.1
- Choose to target the largest values that are unconverged to a residual tolerance of 1e-6
- Require all values above θ to converge to a residual tolerance of 1e-6

➢ Results

❖Comparing time and matrix vector multiplications

Dynamic Stopping Criteria on Bates/Chem97ZtZ



❖ Comparing various SVD tolerances

Hard Impute with Varying SVD Tolerance