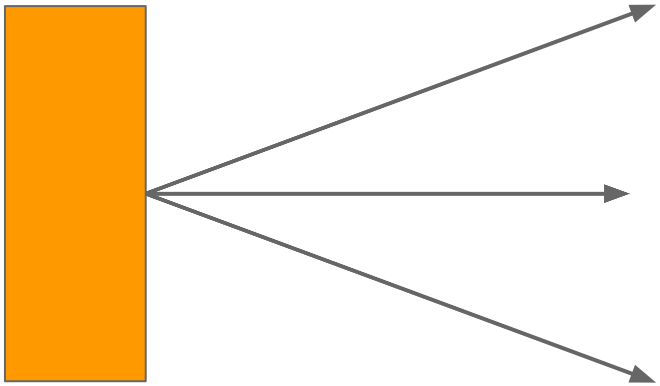
Truncating the SVD via Lanczos

Marco Bornstein
AMSC 763 Final Project

Project Goal

$D \in R^{m \times n}$

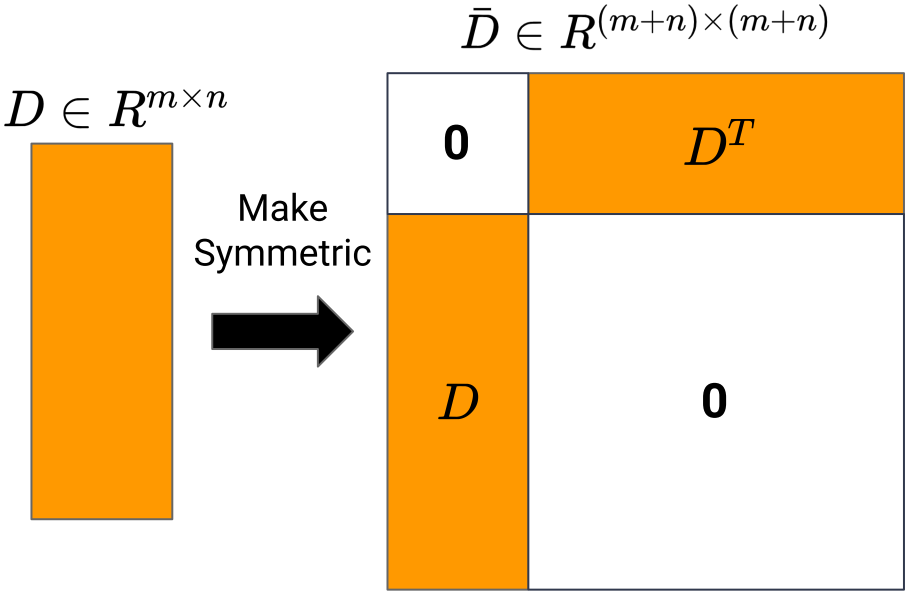


Can we visualize this large Data matrix?

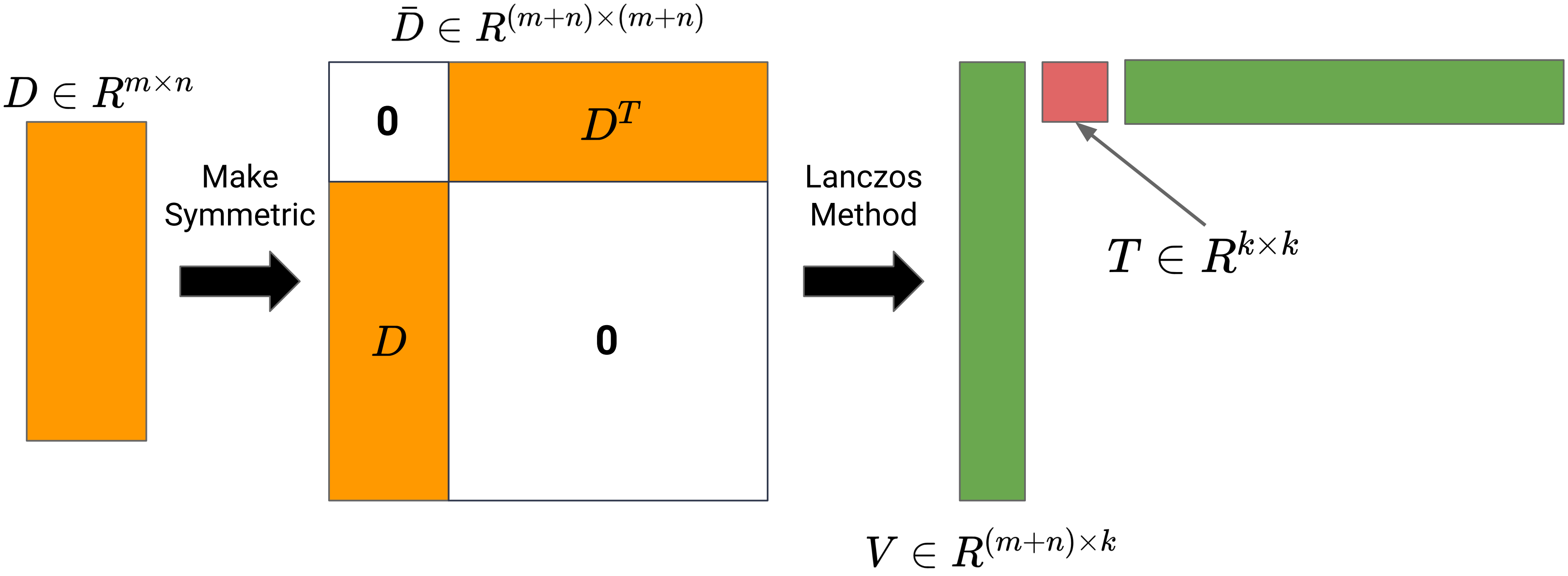
Can this be done in a more efficient manner than computing the full SVD?

Can this be implemented and sped up in parallel?

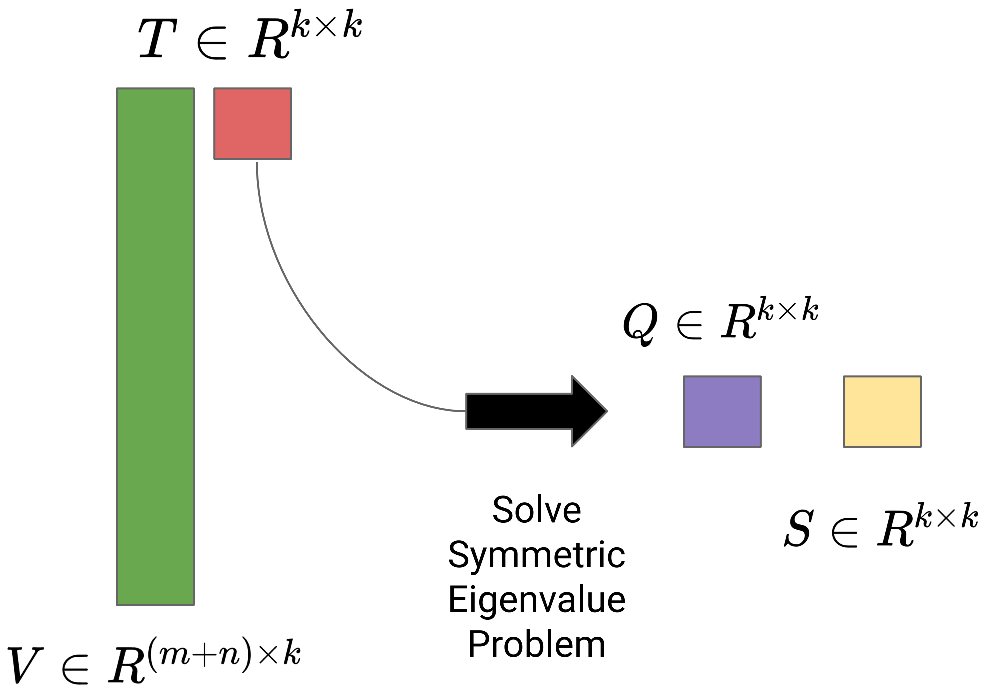
Project Overview: Symmetry



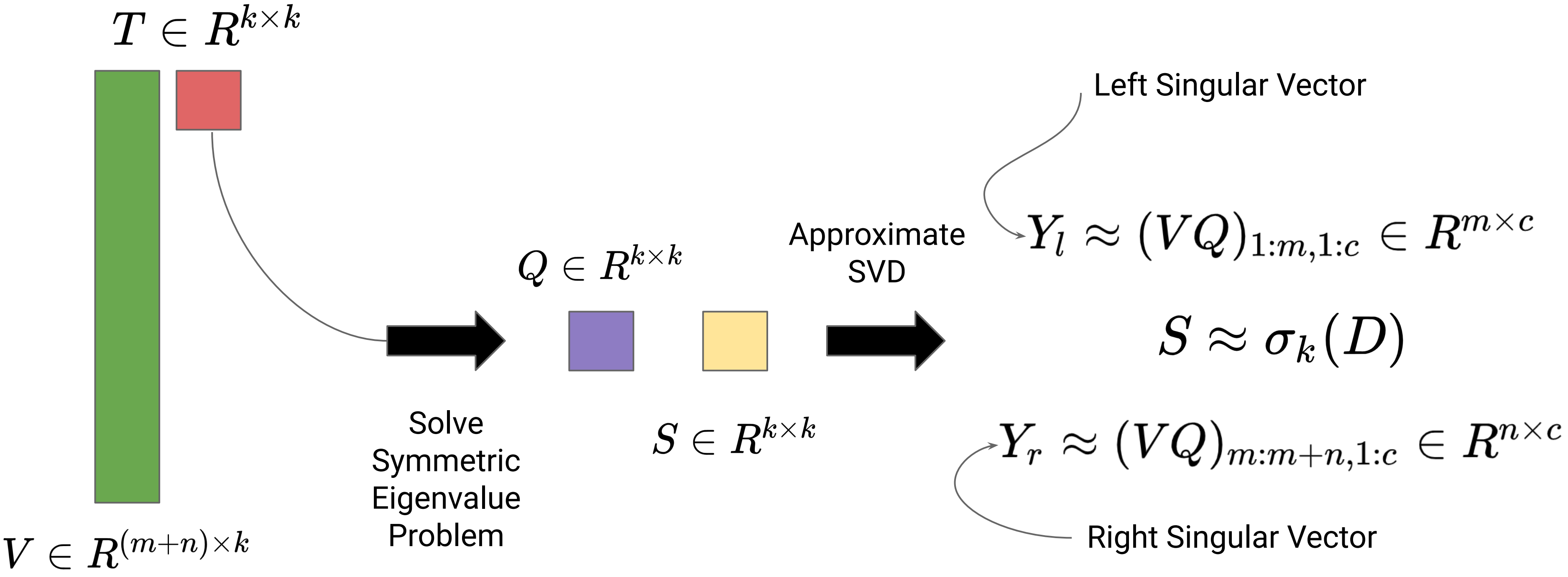
Project Overview: Applying Lanczos



Project Overview: Approximating & Truncating

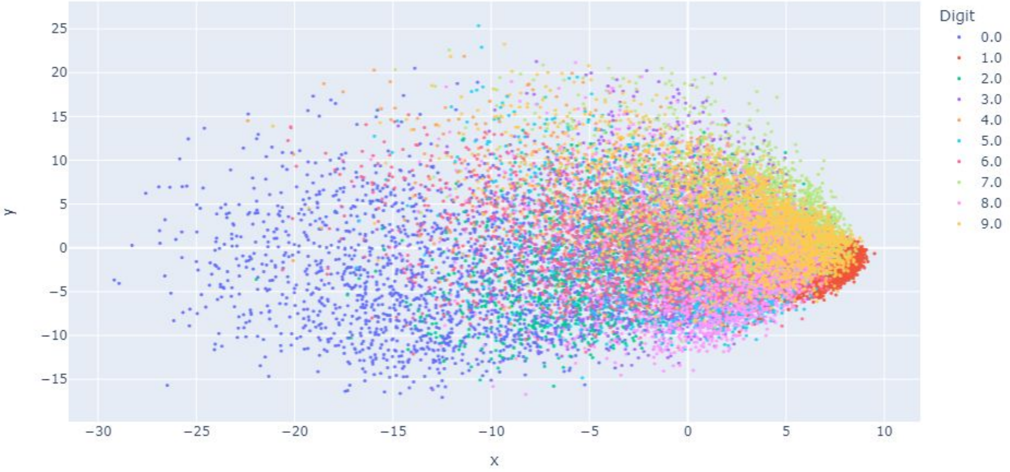


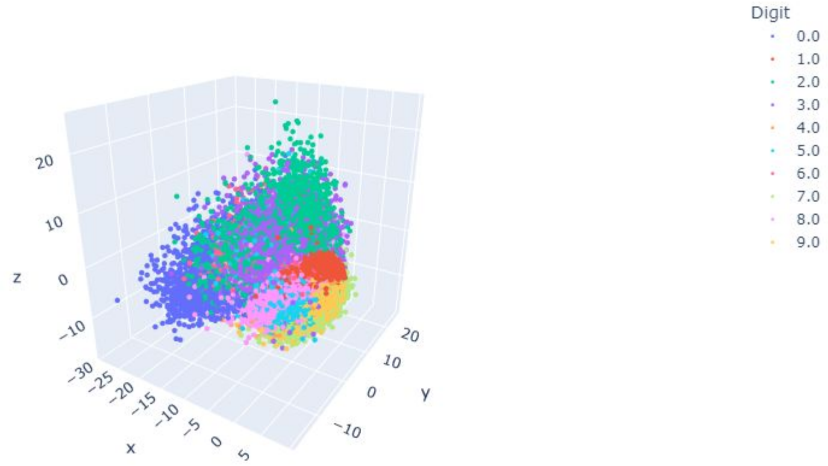
Project Overview: Approximating & Truncating



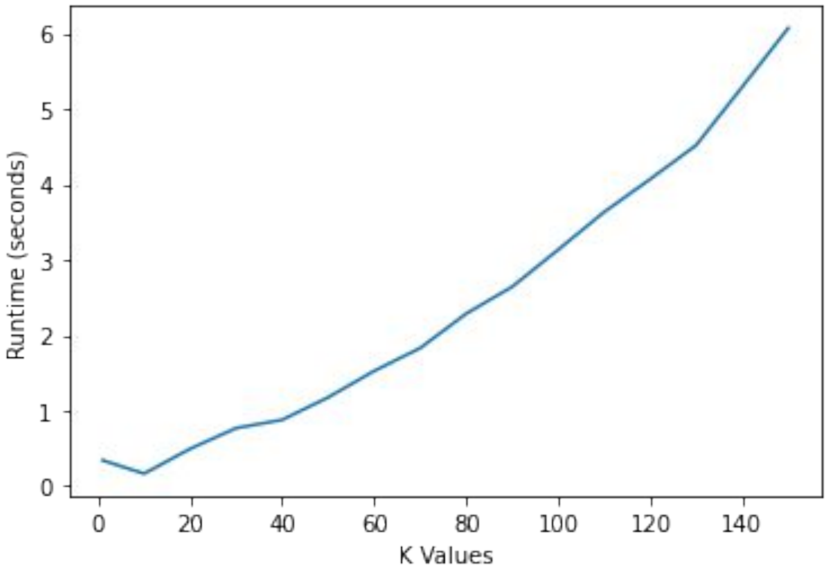
Results: Visualization & Performance

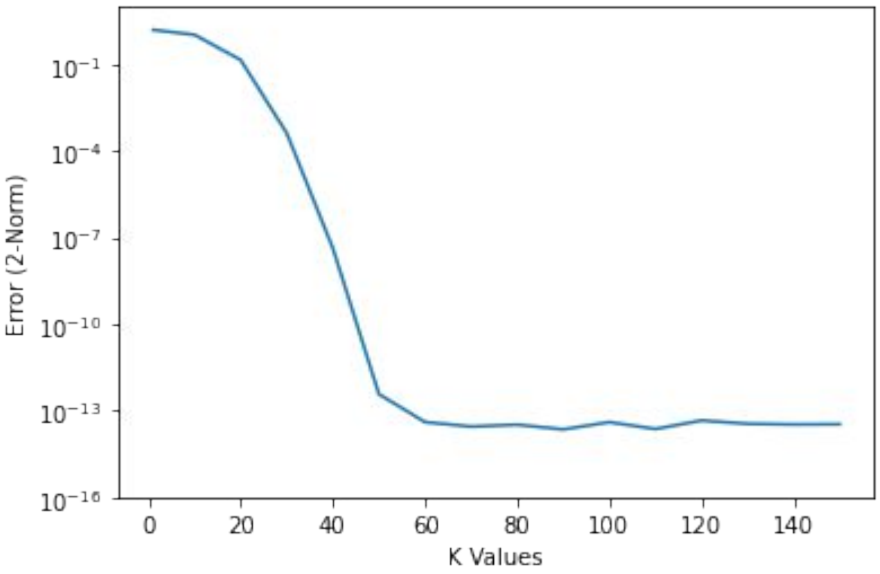
Performed PCA in 2D and 3D on the MNIST Dataset. Below are the results for 20,000 samples:





Results: Visualization & Performance





Results: Visualization & Performance

```
Accuracy and Runtime for 30000 samples and k = 150
Error of Lanczos Serial SVD vs True SVD:
5.833667808510153e-14
Error of Lanczos Parallel SVD vs True SVD:
5.965192431464469e-14
Serial Runtime (Minimum):
20.648642539978027
Serial Runtime (Average):
22.4166455864906З
All Serial Runtimes:
[22.42034674 23.46281552 20.64864254 23.13477755]
Parallel Runtime (Minimum):
1.750953197479248
Parallel Runtime (Average):
1.7855660915374756
All Parallel Runtimes:
[1.8630693  1.7509532  1.76441503 1.76382685]
```

```
==================================
==================================
==================================
Accuracy and Runtime for 50000 samples and k = 150
Serial Runtime (Minimum):
53.493218183517456
Serial Runtime (Average):
60.13505846261978
All Serial Runtimes:
[53.49321818 54.38296747 73.10990381 59.55414438]
Parallel Runtime (Minimum):
1.9418103694915771
Parallel Runtime (Average):
2.0611024498939514
All Parallel Runtimes:
[2.36202955 1.9616251  1.97894478 1.94181037]
==================================
==================================
==================================
```