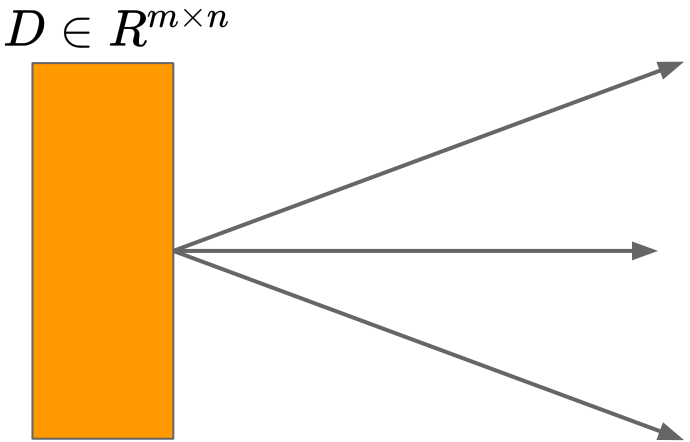
Truncating the SVD via Lanczos

Marco Bornstein
AMSC 763 Final Project

Project Goal

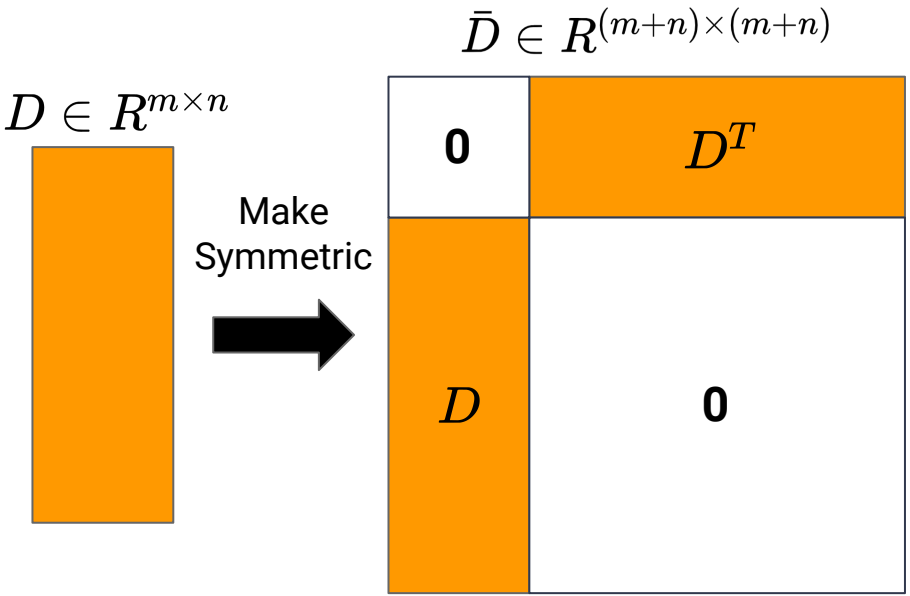


Can we visualize this large Data matrix?

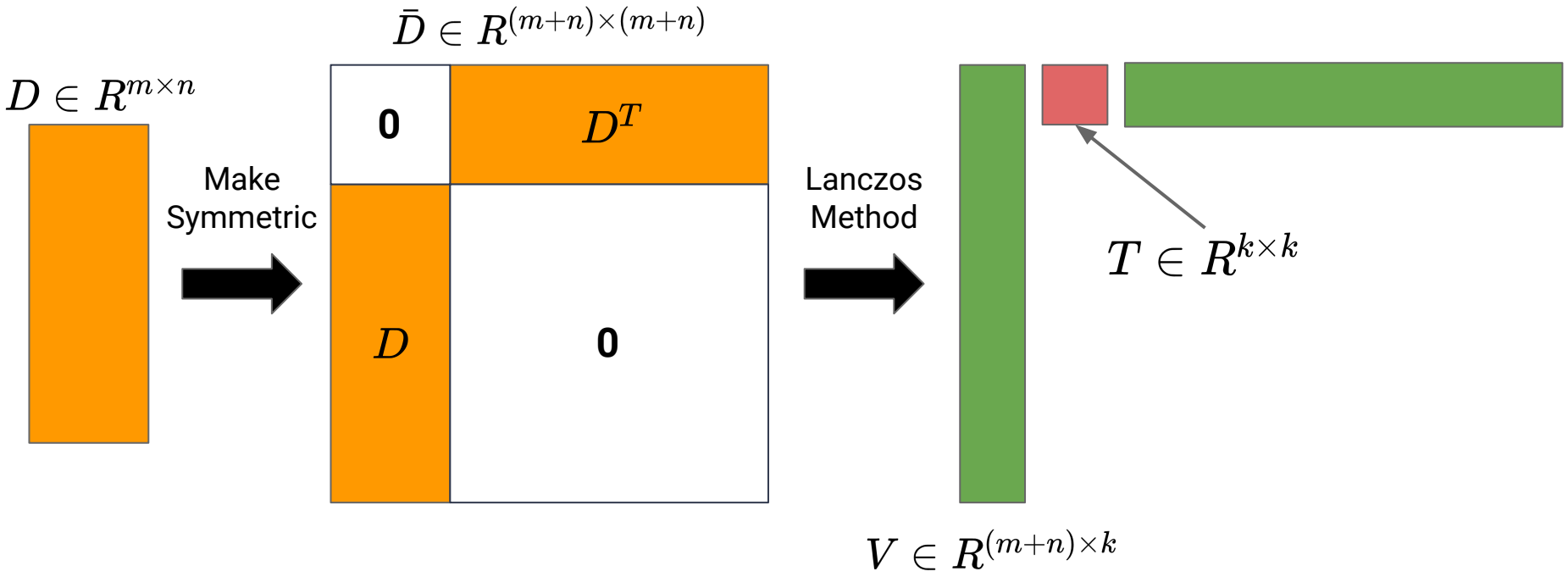
Can this be done in a more efficient manner than computing the full SVD?

Can this be implemented and sped up in parallel?

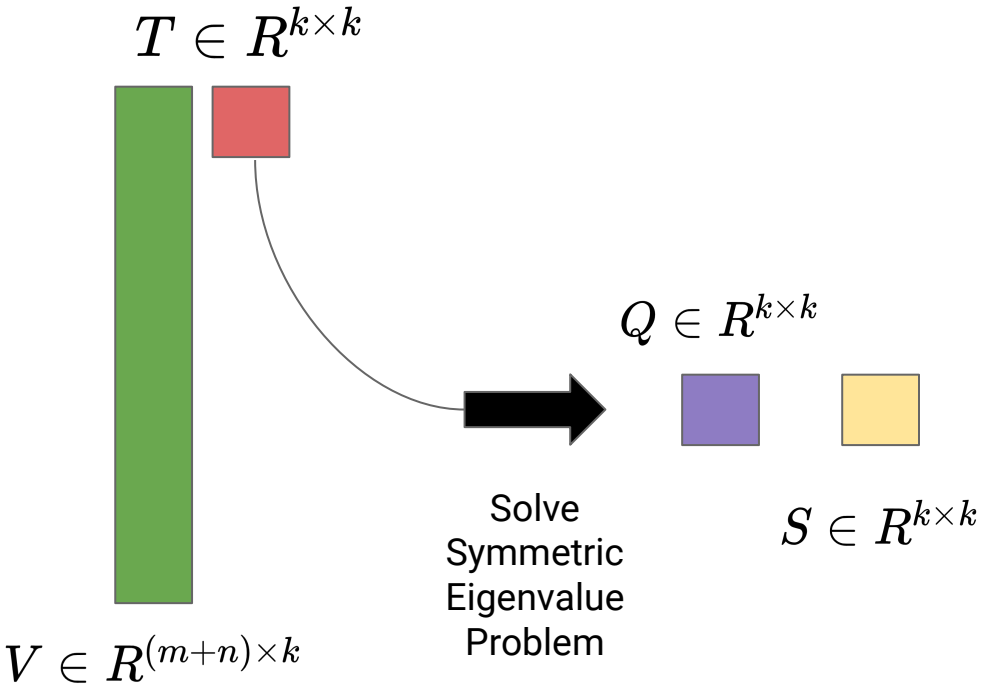
Project Overview: Symmetry



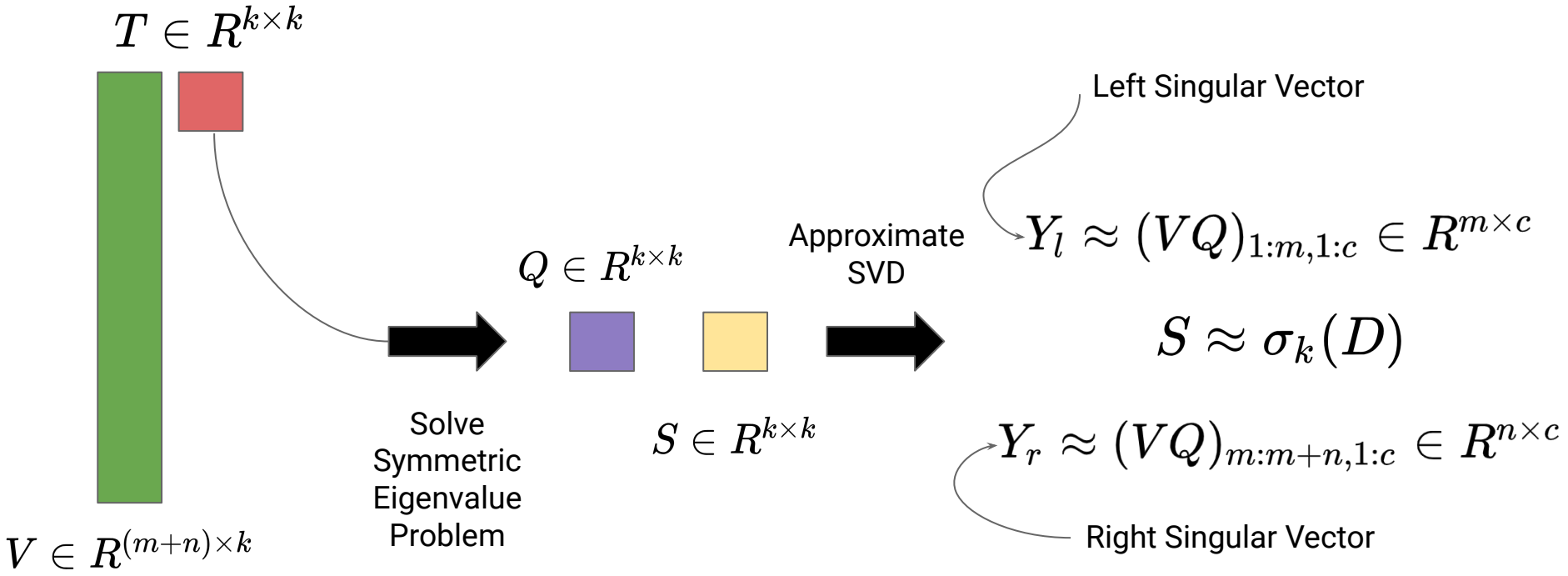
Project Overview: Applying Lanczos



Project Overview: Approximating & Truncating

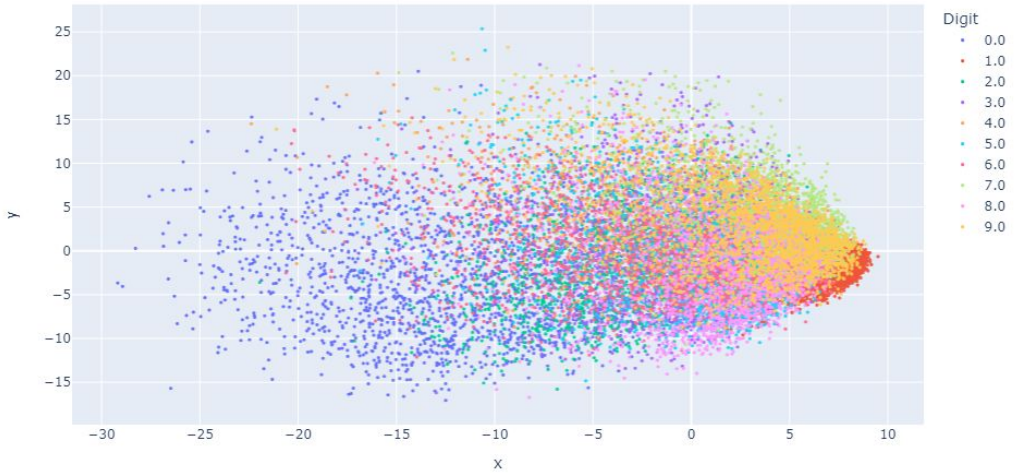


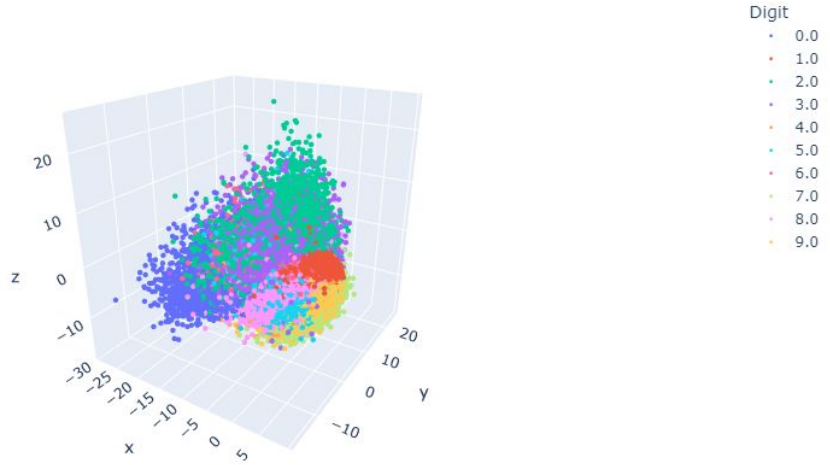
Project Overview: Approximating & Truncating



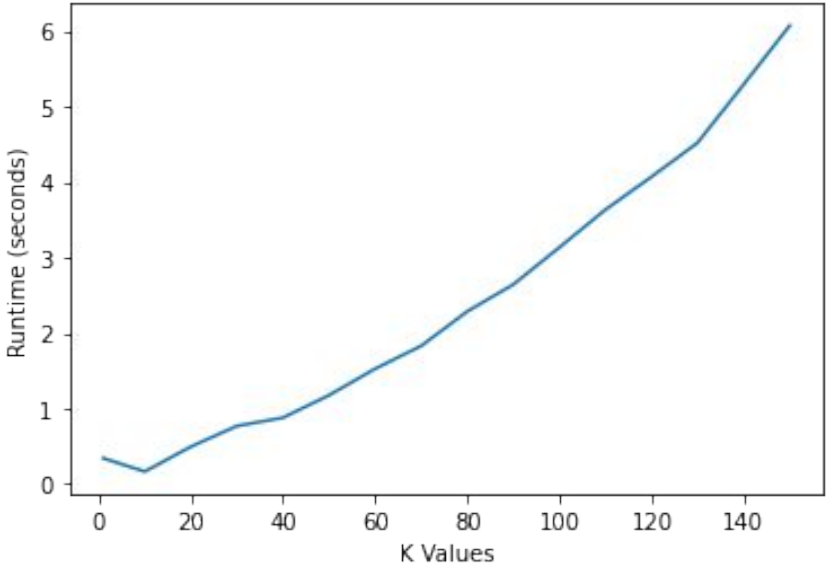
Results: Visualization & Performance

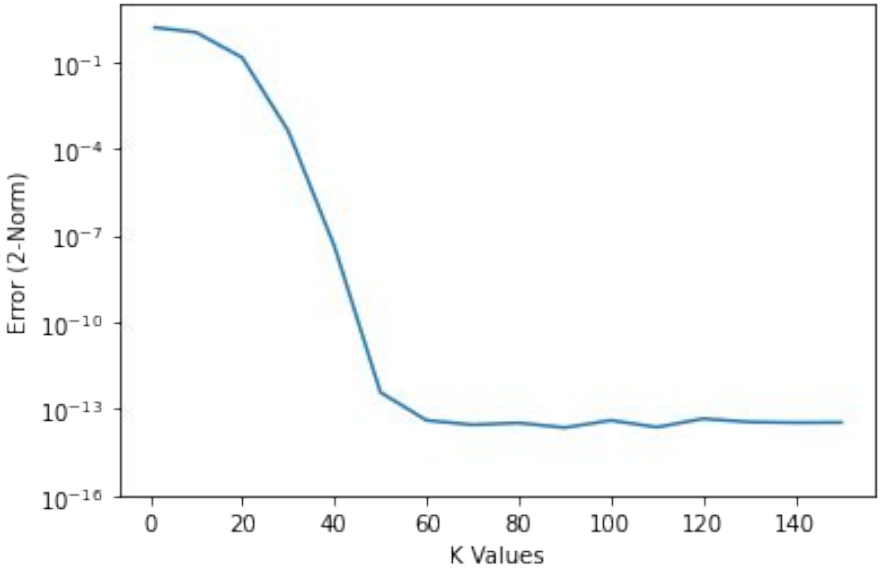
Performed PCA in 2D and 3D on the MNIST Dataset. Below are the results for 20,000 samples:





Results: Visualization & Performance





Results: Visualization & Performance

```
Accuracy and Runtime for 30000 samples and k = 150
Error of Lanczos Serial SVD vs True SVD:
5.795996113956645e-14
Error of Lanczos Parallel SVD vs True SVD:
5.99406546077072e-14
Serial Runtime (Minimum):
4.111012935638428
Serial Runtime (Average):
4.77892279624939
All Serial Runtimes:
[4.35722423 4.65536213 5.99209189 4.11101294]
Parallel Runtime (Minimum):
1.7424683570861816
Parallel Runtime (Average):
1.786321759223938
All Parallel Runtimes:
[1.90062666 1.74680662 1.74246836 1.7553854 ]
```

```
Accuracy and Runtime for 50000 samples and k = 150
Serial Runtime (Minimum):
6.6949944496154785
Serial Runtime (Average):
7.470055162906647
All Serial Runtimes:
[7.88747907 7.4348321  6.69499445 7.86291504]
Parallel Runtime (Minimum):
1.8563508987426758
Parallel Runtime (Average):
1.913610577583313
All Parallel Runtimes:
[2.06331515 1.86419678 1.87057948 1.8563509 ]
```