Homework 2

Your Name

February 27, 2025

We will now start reflecting on the coding questions of Homework 2. The code base that can be used to replicate the result can be found in the following link: `https://github.com/TagoreZhao/STAT260/tree/main/HW2`

1 Problem 6, 7, and 8

The code needed for generating these three matrices is given below:

```python
import numpy as np

def generate_covariance_matrix(d):
    indices = np.arange(d)
    Sigma = 2 * 0.5 ** np.abs(indices[:, None] - indices[None, :])
    return Sigma

def generate_gaussian_A(n, d, seed=1234):
    rng = np.random.default_rng(seed)
    Sigma = generate_covariance_matrix(d)
    mean = np.ones(d)
    A = rng.multivariate_normal(mean, Sigma, size=n)
    return A

def generate_t_distribution_A(n, d, df, seed=1234):
    rng = np.random.default_rng(seed)
    Sigma = generate_covariance_matrix(d)
    mean = np.ones(d)
    z = rng.multivariate_normal(mean, Sigma, size=n)
    chi2_samples = rng.chisquare(df, size=(n, 1))
    A = z / np.sqrt(chi2_samples / df)
    return A
```

Listing 1: Python code for generating matrices

Since numpy does not provide built in functions for generating t-distributed random variables, we have to generate the random variables ourselves. The Gaussian random variables are generated using the `multivariate_normal` function, while the t-distributed random variables are generated using the formula $A = Z/\sqrt{\chi^2/df}$, where Z is the Gaussian random variable, χ^2 is the chi-squared random variable, and df is the degrees of freedom.

Problem 9

We will first plot the norm based probability distribution for all three matrices that we generated using seed 1234.

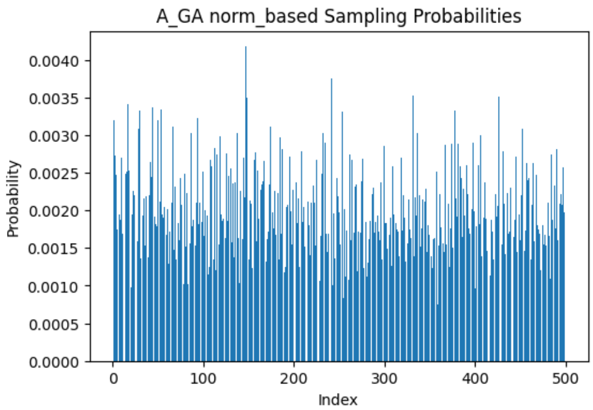


Figure 1: GA Norm based probability distribution

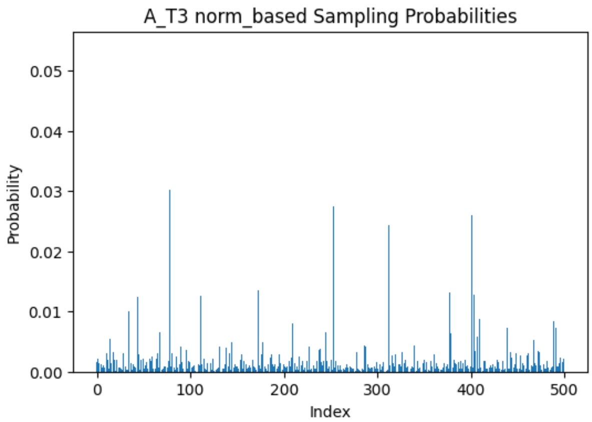


Figure 2: T3 Norm based probability distribution

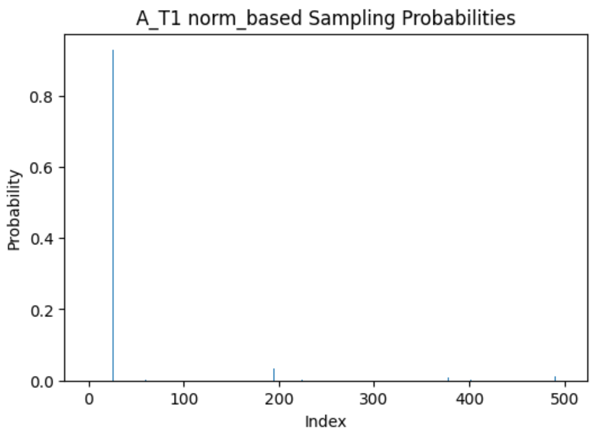


Figure 3: T1 Norm based probability distribution

We will now plot the Frobenius and spectral error for the approximations of the three matrix multiplications.

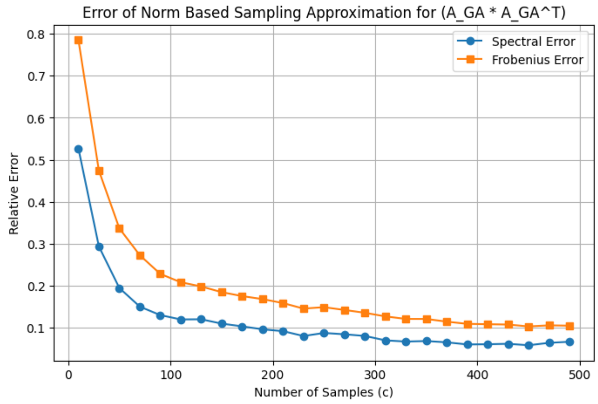


Figure 4: Error of Norm Based Sampling Approximation for $(A_{GA}^{\top} A_{GA})$

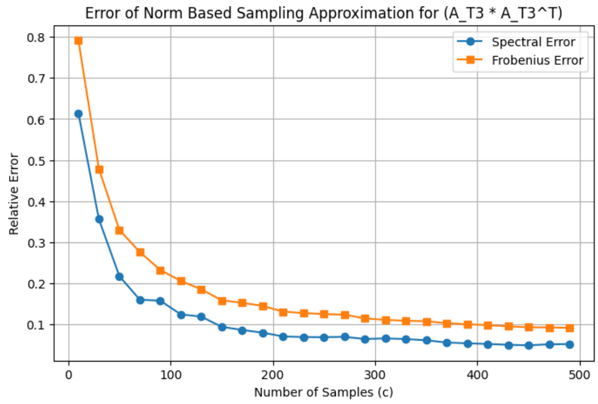


Figure 5: Error of Norm Based Sampling Approximation for $(A_{T3}^{\top} A_{T3})$

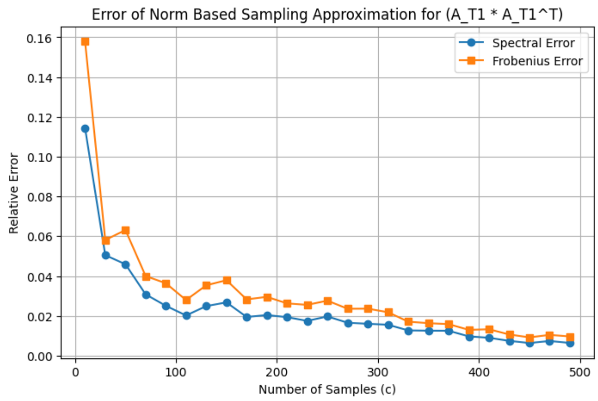


Figure 6: Error of Norm Based Sampling Approximation for $(A_{T1}^{\top} A_{T1})$

Problem 10

We will now plot the Frobenius and spectral error for the approximations of the three left singular matrices multiplication.

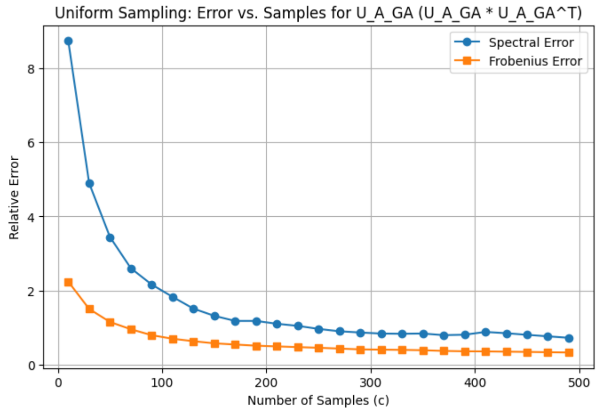


Figure 7: Error of Uniform Based Sampling Approximation for $(U_{GA}^{\top} U_{GA})$

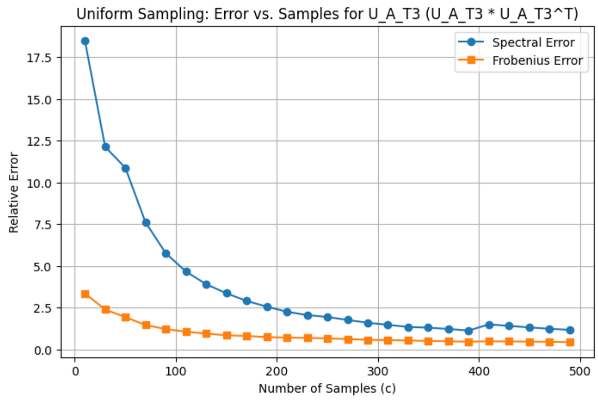


Figure 8: Error of Uniform Based Sampling Approximation for $(U_{T3}^{\top} U_{T3})$

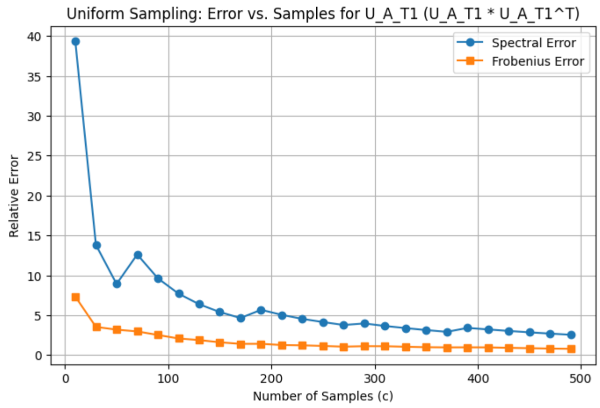


Figure 9: Error of Uniform Based Sampling Approximation for $(U_{T1}^{\top} U_{T1})$

Problem 11

We will now plot the Frobenius and spectral error for the approximations of $A^\top A$ using leverage score sampling.

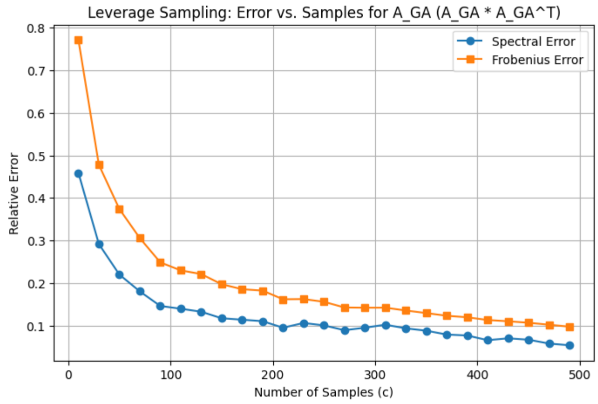


Figure 10: Error of Leverage Based Sampling Approximation for $(A_{GA}^\top A_{GA})$

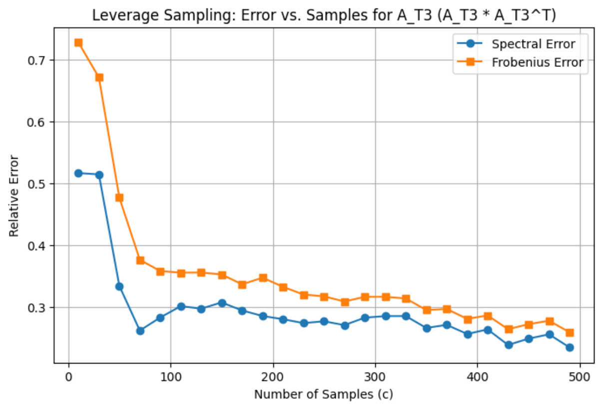


Figure 11: Error of Leverage Based Sampling Approximation for $(A_{T3}^\top A_{T3})$

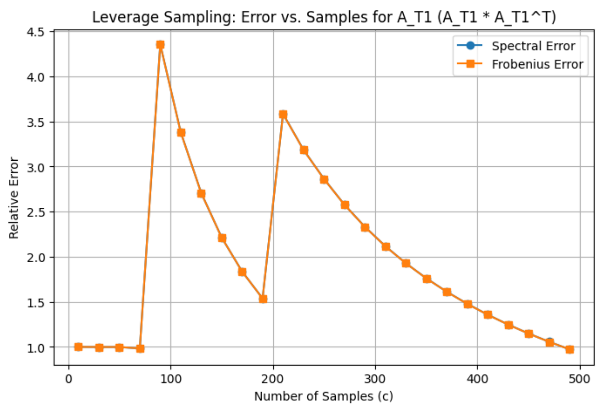


Figure 12: Error of Leverage Based Sampling Approximation for $(A_{T1}^\top A_{T1})$

The results looks similar for A_{GA} and A_{T3}, but the error for A_{T1} is significantly higher than the other two matrices. This means that the leverage score sampling is not as effective for A_{T1} as it is for the other two matrices.

Problem 12

We will now plot the Frobenius and spectral error for the approximations of $AA^\top$ using gaussian projection and $\{\pm 1\}$ projection.

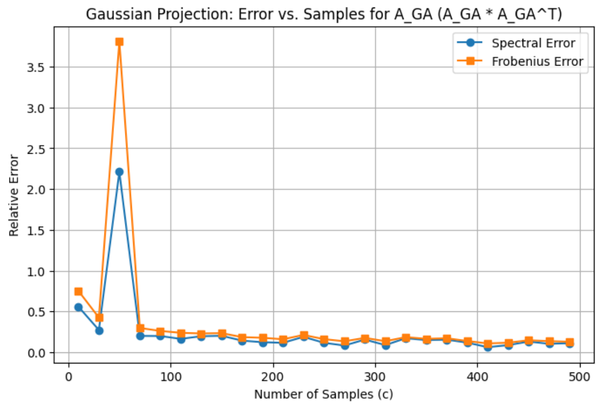


Figure 13: Error of Gaussian Projection Approximation for $(A_{GA}A_{GA}^\top)$

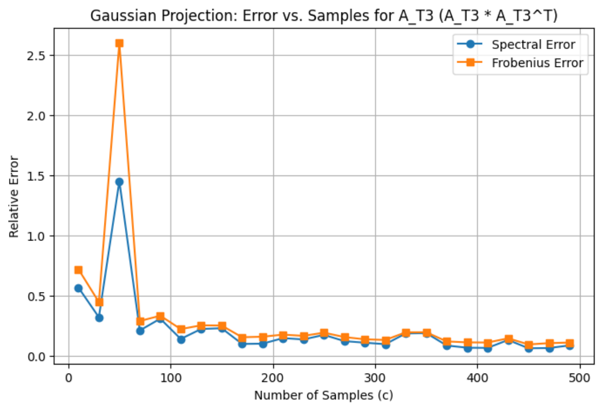


Figure 14: Error of Gaussian Projection Approximation for $(A_{T3}A_{T3}^\top)$

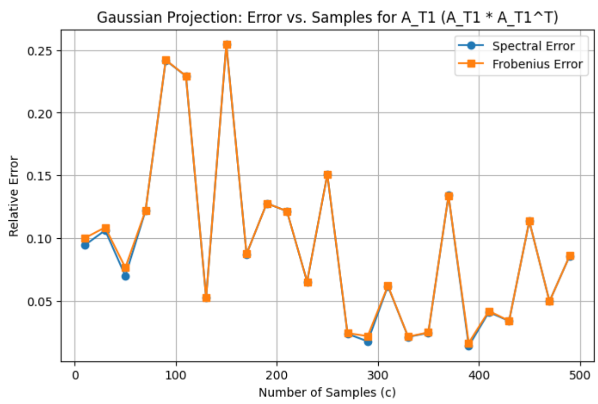


Figure 15: Error of Gaussian Projection Approximation for $(A_{T1}A_{T1}^\top)$

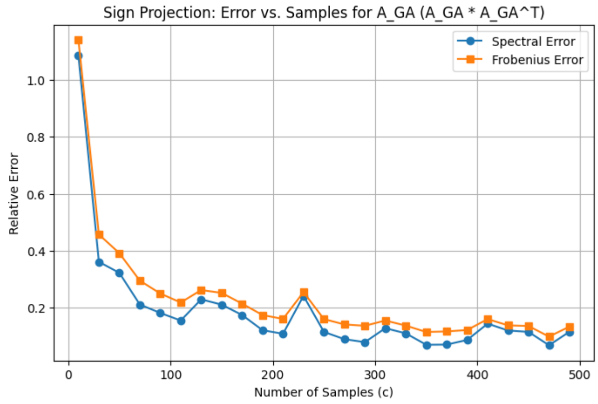


Figure 16: Error of $\{\pm 1\}$ Projection Approximation for $(A_{GA}A_{GA}^\top)$

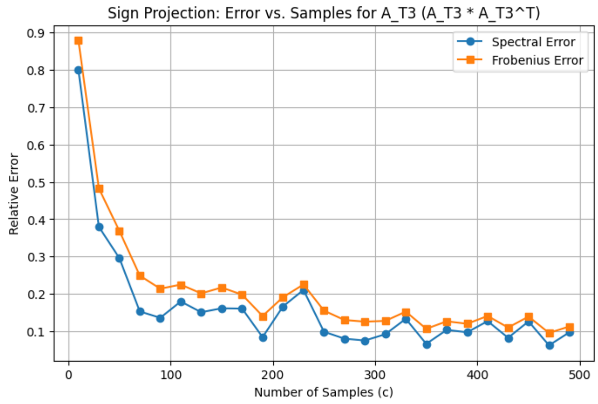


Figure 17: Error of $\{\pm 1\}$ Projection Approximation for $(A_{T3}A_{T3}^\top)$

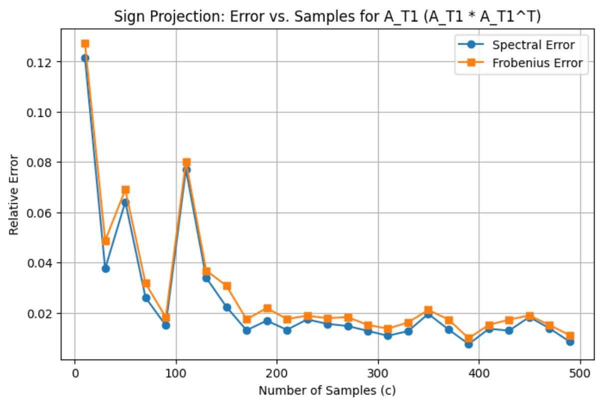


Figure 18: Error of $\{\pm 1\}$ Projection Approximation for $(A_{T1}A_{T1}^\top)$

Comparing the results of the Gaussian projection and $\{\pm 1\}$ projection: The gaussian projection produces more stable results and has slightly smaller error for projecting A_{GA} and A_{T3} when there is reasonbale amount of dimensions. The guassian projection performs extremelly well for A_{T1}, while the $\{\pm 1\}$ projection has a higher error for all three matrices. The only extreme case is that sign projection seems to outperform gaussian projection for A_{T1} when the number of dimensions is small.

Comparing the results of projection and sampling: The projection method has much lower error compare to uniform sampling and overall slight lower error when comparing to norm based sampling and leverage score sampling. The projection method is more stable and has lower error for all three matrices. However, there are times where the projection method shows outlier errors, such as the gaussian projection for A_{GA} when the number of dimensions is small.

Problem 13

We will now plot 3D plot of the error of the sparse approximations of $AA^\top$ using gaussian projection and $\{\pm 1\}$ projection. The two axes are the sparsity and the number of dimensions, and the third axis is the error.

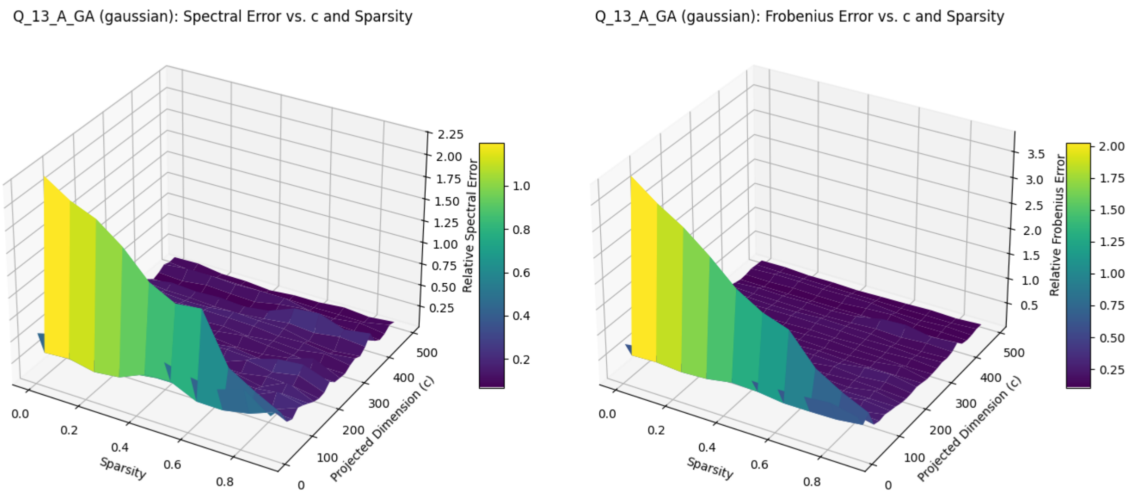


Figure 19: Error of Gaussian Projection Approximation for $(A_{GA}A_{GA}^\top)$

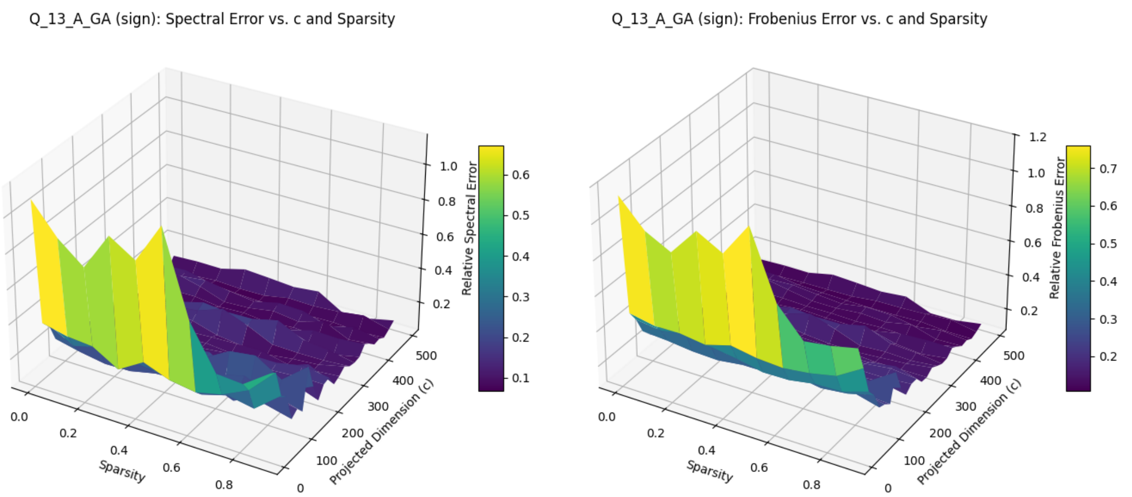


Figure 20: Error of $\{\pm 1\}$ Projection Approximation for $(A_{GA}A_{GA}^\top)$

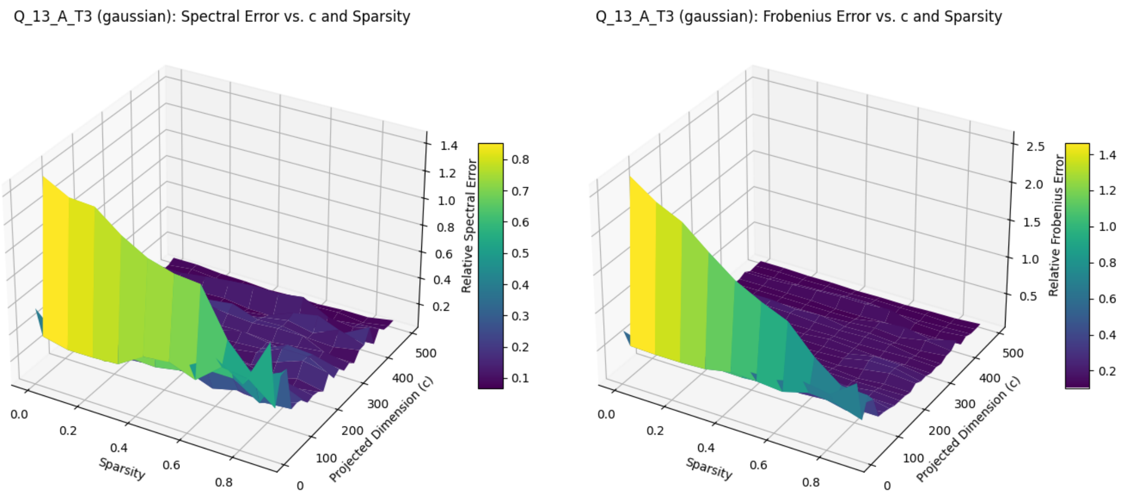


Figure 21: Error of Gaussian Projection Approximation for $(A_{T3}A_{T3}^{\top})$

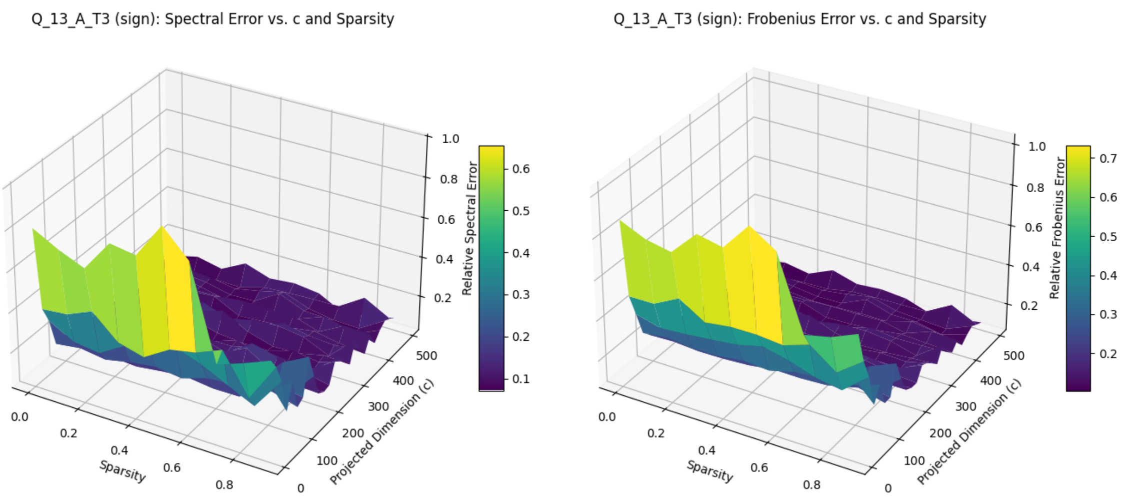


Figure 22: Error of $\{\pm 1\}$ Projection Approximation for $(A_{T3}A_{T3}^{\top})$

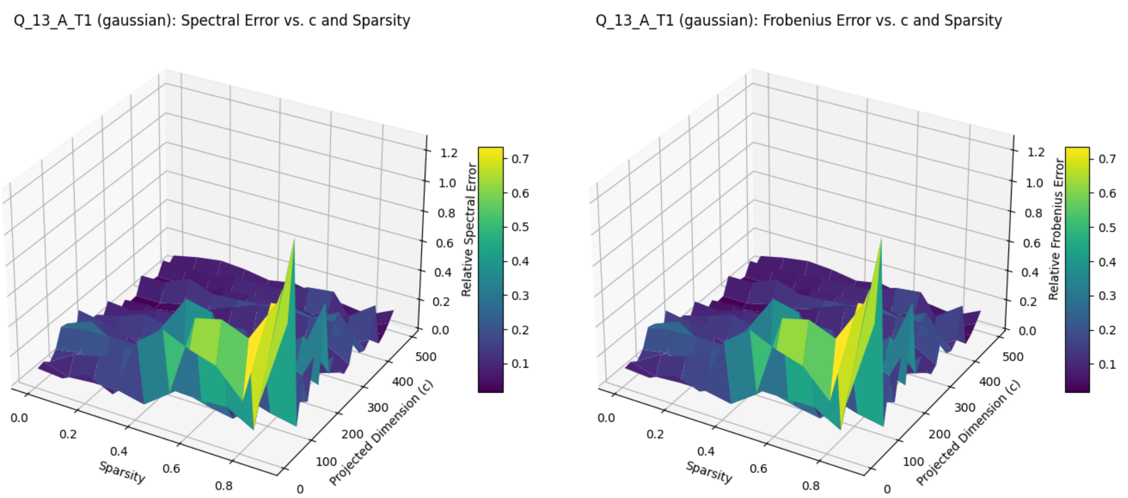


Figure 23: Error of Gaussian Projection Approximation for $(A_{T1}A_{T1}^{\top})$

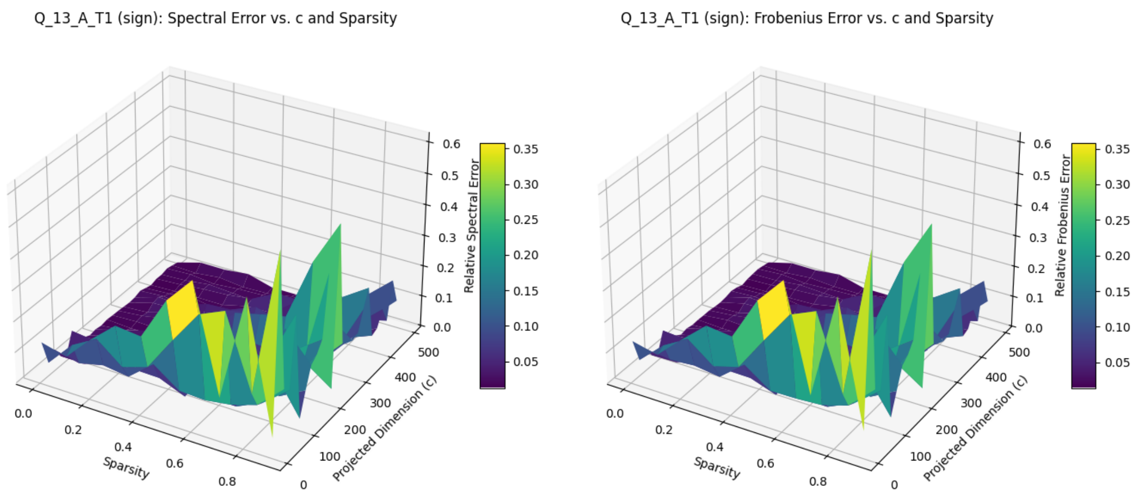
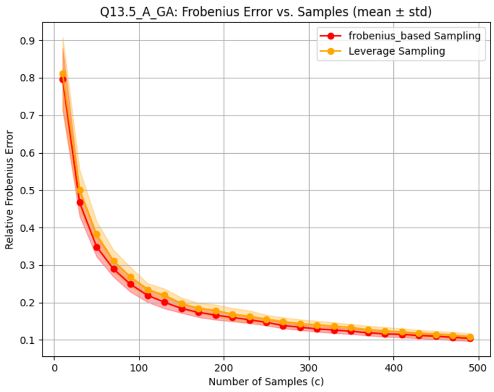


Figure 24: Error of $\{\pm 1\}$ Projection Approximation for $(A_{T1}A_{T1}^{\top})$

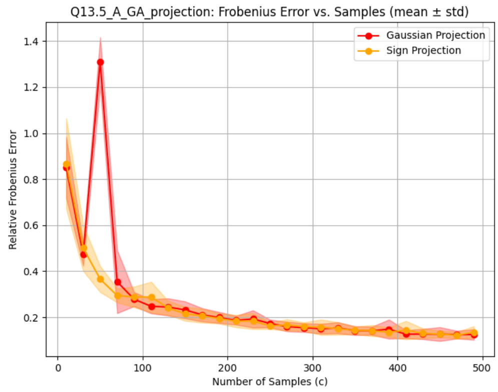
Both Gaussian projection and $\{\pm 1\}$ projection have similar error patterns for all three matrices. For A_{GA} and A_{T3}, the error is relatively stable and low when the number of dimensions is large. In addition, both projection seems to perform well as we increase sparsity. However, for A_{T1}, the error is significantly higher than the other two matrices, and the error is not as stable as the other two matrices. For both projection, the errors are higher when the sparsity is high, and the errors are more unstable when the sparsity is high.

Problem 13.5 Variability Analysis

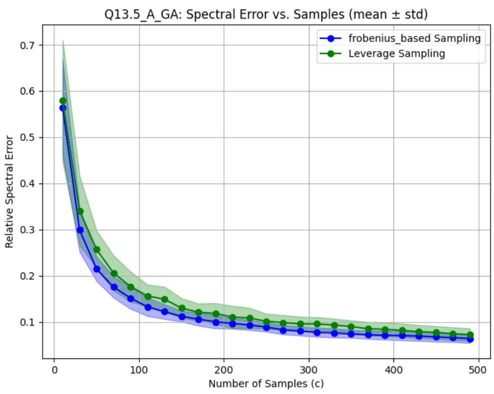
A_{GA}



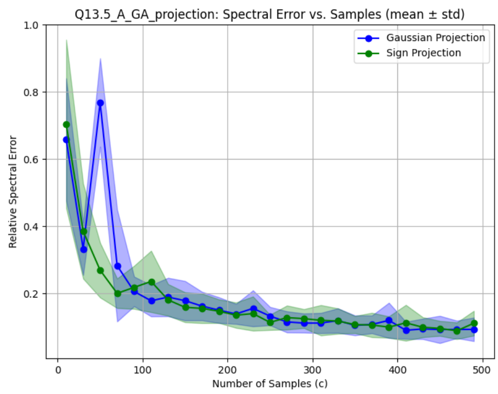
(a) Sampling: Frobenius Error



(b) Projection: Frobenius Error



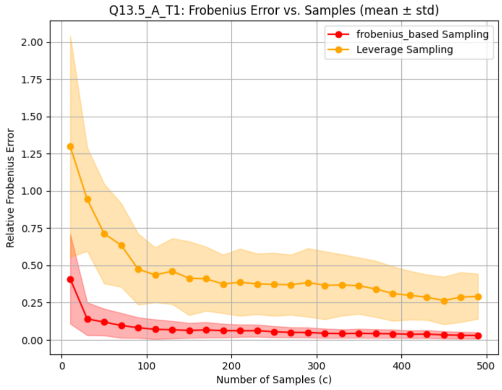
(c) Sampling: Spectral Error



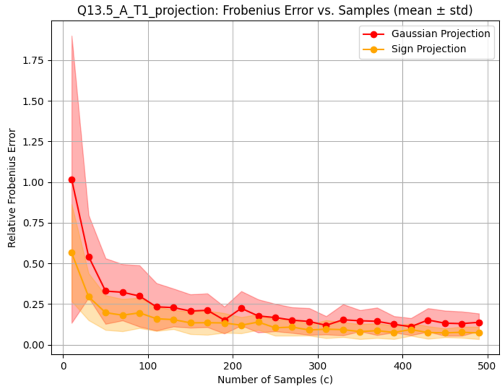
(d) Projection: Spectral Error

Figure 25: A_{GA}: Comparison of Random Sampling and Random Projection

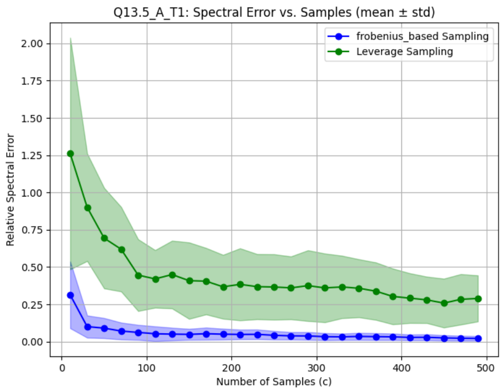
A_{T1}



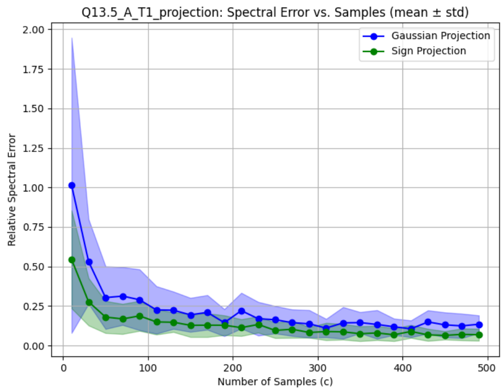
(a) Sampling: Frobenius Error



(b) Projection: Frobenius Error



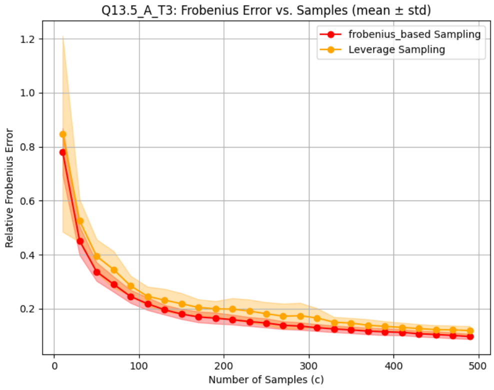
(c) Sampling: Spectral Error



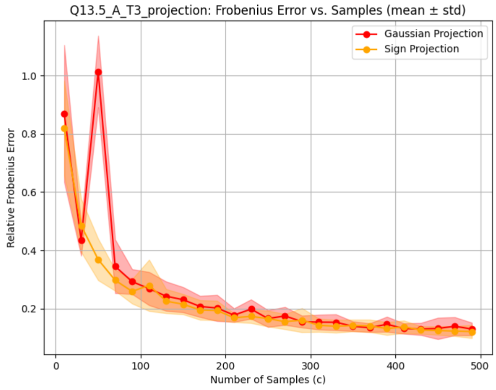
(d) Projection: Spectral Error

Figure 26: A_{T1}: Comparison of Random Sampling and Random Projection

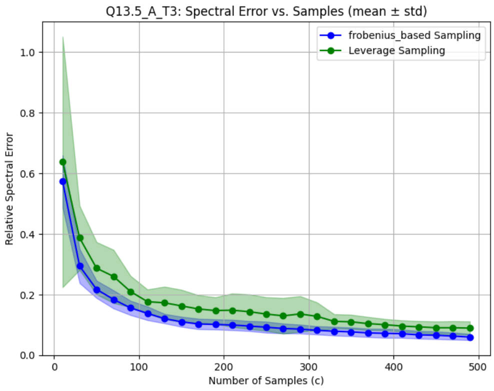
A_{T3}



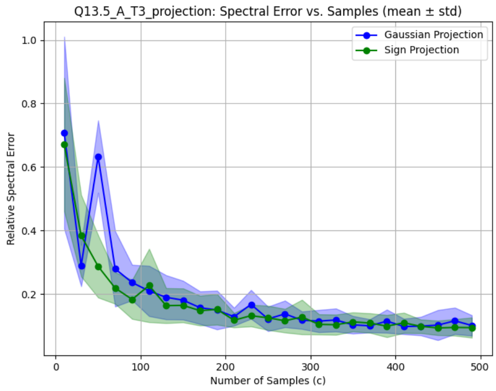
(a) Sampling: Frobenius Error



(b) Projection: Frobenius Error



(c) Sampling: Spectral Error



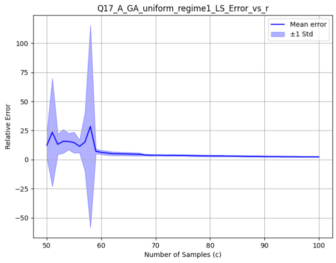
(d) Projection: Spectral Error

Figure 27: A_{T3}: Comparison of Random Sampling and Random Projection

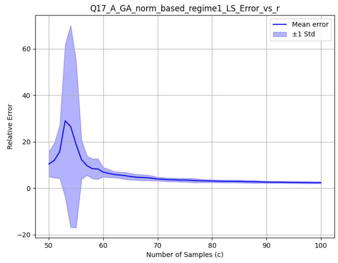
Problem 17

For each matrix, the following figures show the LS error versus the number of samples r using three sampling methods. The top row corresponds to Regime 1 (r = d ... 2d) and the bottom row to Regime 2 (r = 2d, 3d, . . .).

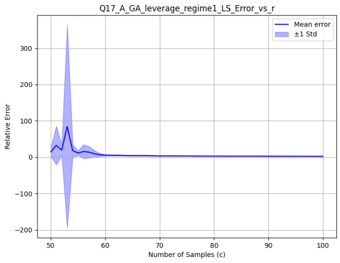
A_{GA}



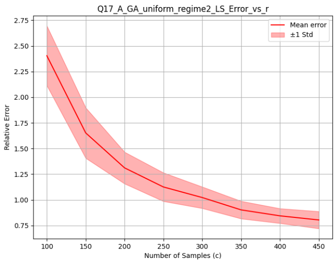
(a) Uniform, Regime 1



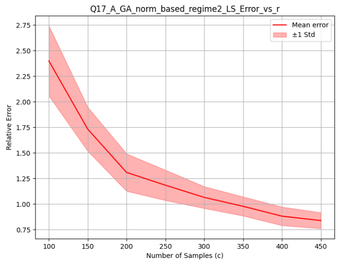
(b) Norm-based, Regime 1



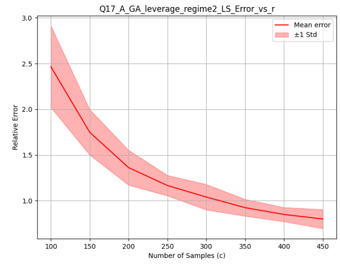
(c) Leverage, Regime 1



(d) Uniform, Regime 2



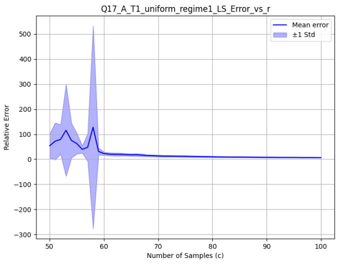
(e) Norm-based, Regime 2



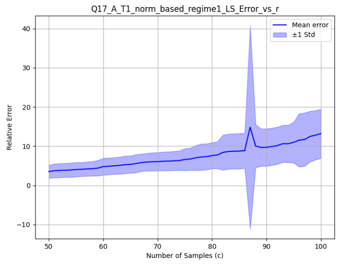
(f) Leverage, Regime 2

Figure 28: A_{GA}: LS Error vs. r for Different Sampling Methods and Regimes

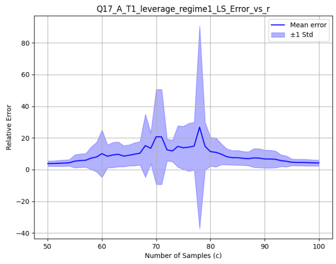
A_{T1}



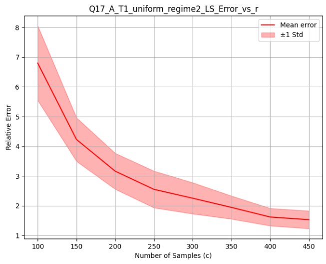
(a) Uniform, Regime 1



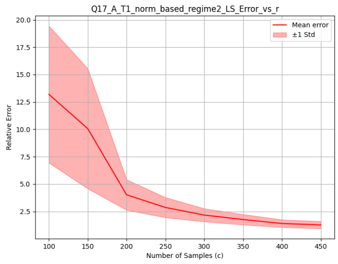
(b) Norm-based, Regime 1



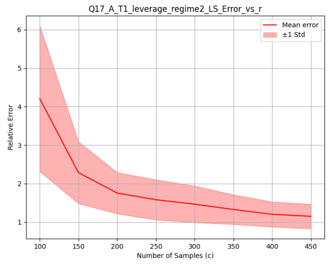
(c) Leverage, Regime 1



(d) Uniform, Regime 2



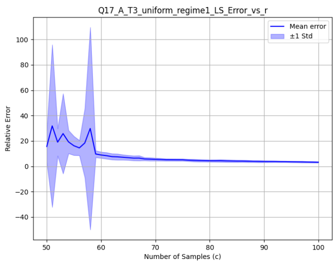
(e) Norm-based, Regime 2



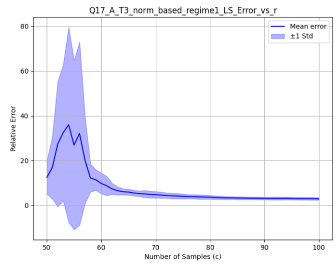
(f) Leverage, Regime 2

Figure 29: A_{T1}: LS Error vs. r for Different Sampling Methods and Regimes

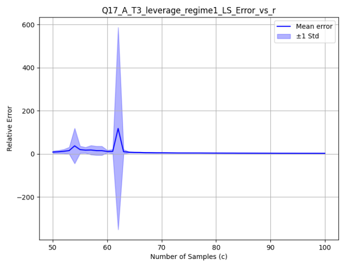
A_{T3}



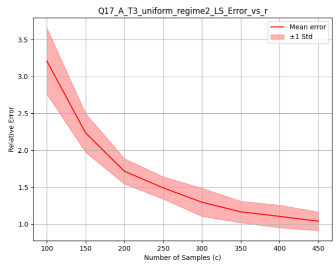
(a) Uniform, Regime 1



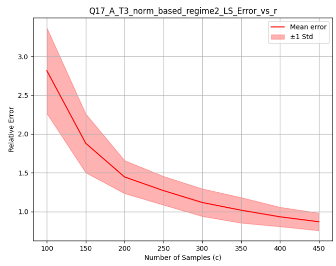
(b) Norm-based, Regime 1



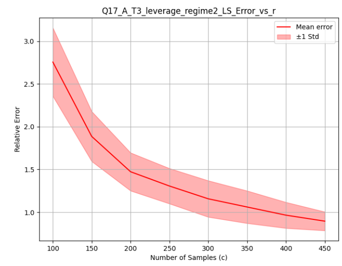
(c) Leverage, Regime 1



(d) Uniform, Regime 2



(e) Norm-based, Regime 2



(f) Leverage, Regime 2

Figure 30: A_{T3}: LS Error vs. r for Different Sampling Methods and Regimes