Estimation of the sample covariance matrix from compressive measurements

Mls Pragna - 180050064
Rahul Puli - 180050080

1 Abstract

From compressive measurements of the data that is the low dimensional random projections, the sample covariance matrix is estimated. The paper offers an unbiased covariance matrix from the compressive measurements collected using a class of random projection matrices whose entries are independent and identically distributed random variables drawn from a zero mean distribution with finite first four moments. Experiments show that this method can accurately estimate the sample covariance matrix on a variety of real-world data sets.

2 Preliminaries

2.1 Unbiased estimator

An estimator of a parameter is said to be unbiased estimator if the expected value of the estimator is the parameter. If the estimator is S and the parameter being estimated is θ, S is a unbiased estimator if $E[S] = \theta$. This paper presents an unbiased estimator of the covariance matrix of data.

2.2 Moments of Probabilistic distribution

The moments of a function are quantitative measures that relate to the shape of the graph of the function.There are four moments for a probabilistic distribution. The first moment is expected value, second moment is variance, third standardized moment is skewness and the fourth standardized moment is kurtosis. Skewness is a measure of the asymmetry of a random variable's probability distribution around its mean. Kurtosis is the measure of the extremity of the deviations of a random variable's probability distribution. This paper assumes that random projection matrices are drawn i.i.d from a distribution with finite first four moments.
The kurtosis of a probability distribution function is,

$$\kappa = \frac{\mu_4}{\mu_2^2} - 3$$

where μ_2 and μ_4 are fourth central moments and $\mu_n = E[(x - \mu)^n]$.

3 The proposed unbiased covariance estimator

Let $X = [x_1, x_2, ...x_n] \epsilon R^{pxn}$ where each $x_i \epsilon R^n$ is a data sample. The covariance matrix of the data is

$$C_n = \frac{1}{n} \sum_{i=1}^{n} x_i {x_i}^T = \frac{1}{n} X X^T$$

The aim of the paper is to estimate the covariance matrix C_n from the low dimensional projections of the data $\{R_i^T x_i\}_{i=1}^{n} \epsilon R^m$ where $R_i \epsilon R^{mxp}$ with m¡p for all $i = 1, 2, \ldots, n$. The entries of R_i are independent and identically distributed random variables drawn from a zero mean distribution with finite first four moments.

3.1 Compressive covariance matrix estimation

First, the rescaled version of the covarinace matrix of projected data $\{R_i R_i^T x_i\}_{i=1}^{n} \epsilon R^p$ is considered.

$$\hat{C}_n = \frac{1}{(m^2+m)\mu_2^2} \frac{1}{n} \sum_{i=1}^{n} R_i R_i^T x_i x_i^T R_i R_i^T$$

where μ_2 is the second moment of the distribution.
The expectation of the covariance matrix $\hat{C}_n$ is,

$$E[\hat{C}_n] = C_n + \frac{\kappa}{m+1} diag(C_n) + \frac{1}{m+1} tr(C_n) I_{pxp}$$

The expected value of $\hat{C}_n$ has three terms, the first term is the covariance matrix that we want to estimate, the next two terms are bias terms. The second term can be can be interpreted as the estimator's bias towards the closest canonical basis vectors. The third term is not dependent on the shape of the distribution that is four moments of the distribution. It represents the energy of the compressed data scattered in different directions in R^p, as tr(C_n) represents the energy of the input data X.

Now we alter the covariance matrix C_n to $\hat{\sum}_n$ such that the $\hat{\sum}_n$ has bias zero i.e., it is a unbiased estimator of the sample covariance matrix C_n. To get such an estimator we first find the expectations of second bias and third bias terms.
The estimation of the first bias term is,

$$E[diag(\hat{C}_n)] = (1 + \frac{\kappa}{m+1}) diag(C_n) + n \frac{tr(C_n)}{m+1} I_{pxp}$$

The estimation of the second bias term is,

$$E[tr(\hat{C}_n)] = \frac{(m+1+\kappa+p)}{m+1} tr(C_n)$$

Now, the estimator $\hat{\sum}_n$ is defined as,

$$\hat{\sum}_n = \hat{C}_n - \alpha_1 diag(\hat{C}_n) - \alpha_2 tr(\hat{C}_n) I_{pxp}$$

where $\alpha_1 = \frac{\frac{\kappa}{m+1}}{1+\frac{\kappa}{m+1}}$ and $\alpha_2 = \frac{1}{(1+\frac{\kappa}{m+1})(m+1+\kappa+p)}$ The estimator $\hat{\sum}_n$ is an unbiased estimator,

$$\begin{aligned}
E[\hat{\sum_n}] &= E[\hat{C}_n - \alpha_1 diag(\hat{C}_n) - \alpha_2 tr(\hat{C}_n) I_{pxp}] \\
&= E[\hat{C}_n] - \alpha_1 E[diag(\hat{C}_n)] - \alpha_2 E[tr(\hat{C}_n) I_{pxp}] \\
&= E[\hat{C}_n] - \alpha_1 E[diag(\hat{C}_n)] - \alpha_2 E[tr(\hat{C}_n) I_{pxp}] \\
&= C_n + \frac{\kappa}{m+1} diag(C_n) + \frac{1}{m+1} tr(C_n) I_{pxp} \\
&\quad - \alpha_1((1+\frac{\kappa}{m+1}) diag(C_n) + n\frac{tr(C_n)}{m+1} I_{pxp}) \\
&\quad - \alpha_2(\frac{(m+1+\kappa+p)}{m+1} tr(C_n)) \\
&= C_n
\end{aligned}$$

Hence the estimator $\hat{\sum}_n$ is an unbiased estimator.

3.2 Compressive estimation using sparse random projections

In this paper, the random matrices are sparse random projections that is the entries of $\{R_i\}_{i=1}^n \epsilon R^{pxm}$ are distributed on -1,0,1 with probabilities $\frac{1}{2s}$, 1-$\frac{1}{s}$, $\frac{1}{2s}$. On average the number of non-zero entries in each column of $R_i \epsilon R^{pxm}$ are $\frac{p}{s}$. The bias term increases with the increase of s therefore the modification of covariance matrix $\hat{C}_n$ is more needed when the value of s is high.
The compressive factor is given by,

$$\gamma = \frac{m}{s} < 1$$

By choosing γ to be less than 1 we have two benefits. The first one is the computation cost for accessing each data sample is $O(\frac{mp}{s}) = O(\gamma p)$, $\gamma < 1$ is less when compared to acquiring the full data sample $O(p)$. The second is the computational cost to form the sample covariance matrix $\hat{C}_n$ is γ^2 times less that the computational cost required to form the actual covariance matrix C_n. This is because there are at most $O(\frac{mp}{s})$ non-zero entries in $R_i R_i^T x_i \epsilon R^p$. Thus the cost to form $R_i R_i^T x_i (R_i R_i^T x_i)^T$ is $O(\gamma^2 p^2)$ whereas the cost to form $x_i x_i^T$ is $O(p^2)$.

4 Experimental results

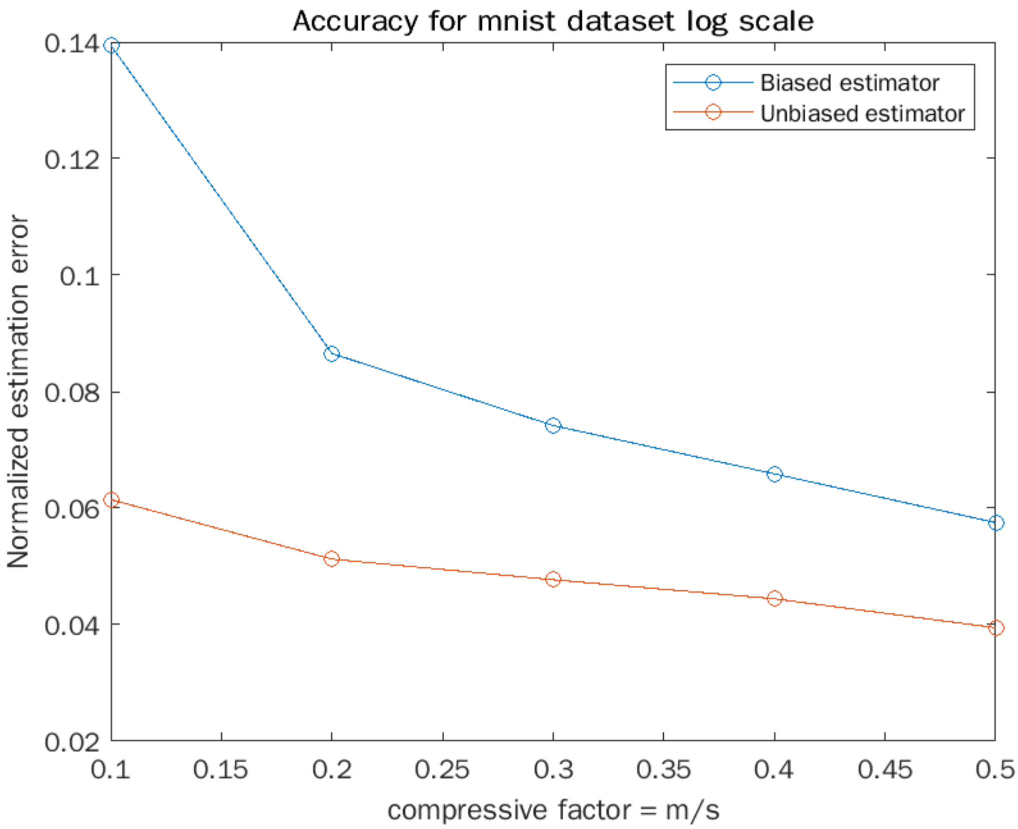
Normalized covariance estimation error is used for the estimation evaluation,

$$normalized\ estimation\ error = \frac{||\hat{\sum}_n - C_N||_2}{||C_n||_2}$$

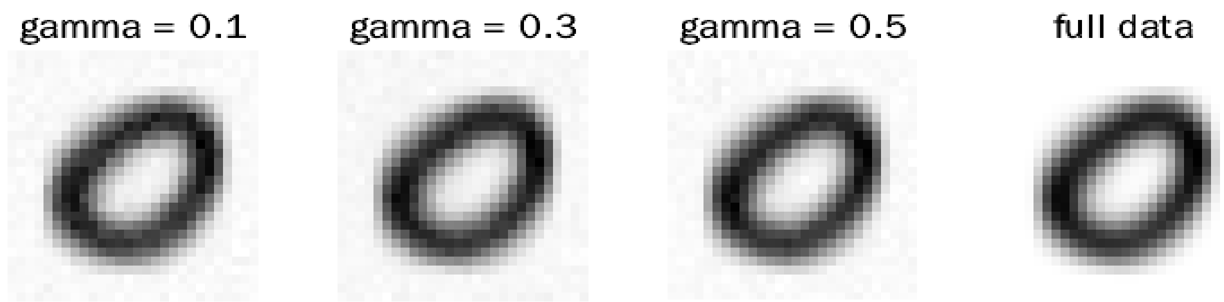
This evaluation metric measures the degree to which the predicted covariance matrix from compressive measurements is similar to the underlying covariance matrix. C_n.
In all the experiments, the parameter $\frac{m}{p} = 0.4$ is fixed and the accuracy is analyzed for the various values of compression factor $\gamma = \frac{m}{s}$. Here, the value of $\gamma < 1$ because our measurements are compressive.

4.1 MNIST data set

MNIST data set contains the images of the handwritten digits. Each image is a 28x28 size image and the vectorized image dimension is $p = 28 * 28 = 784$. The number of data samples taken are $n = 5923$. This below figure shows the normalized estimation error of the estimated biased and unbiased covariance matrices at $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$



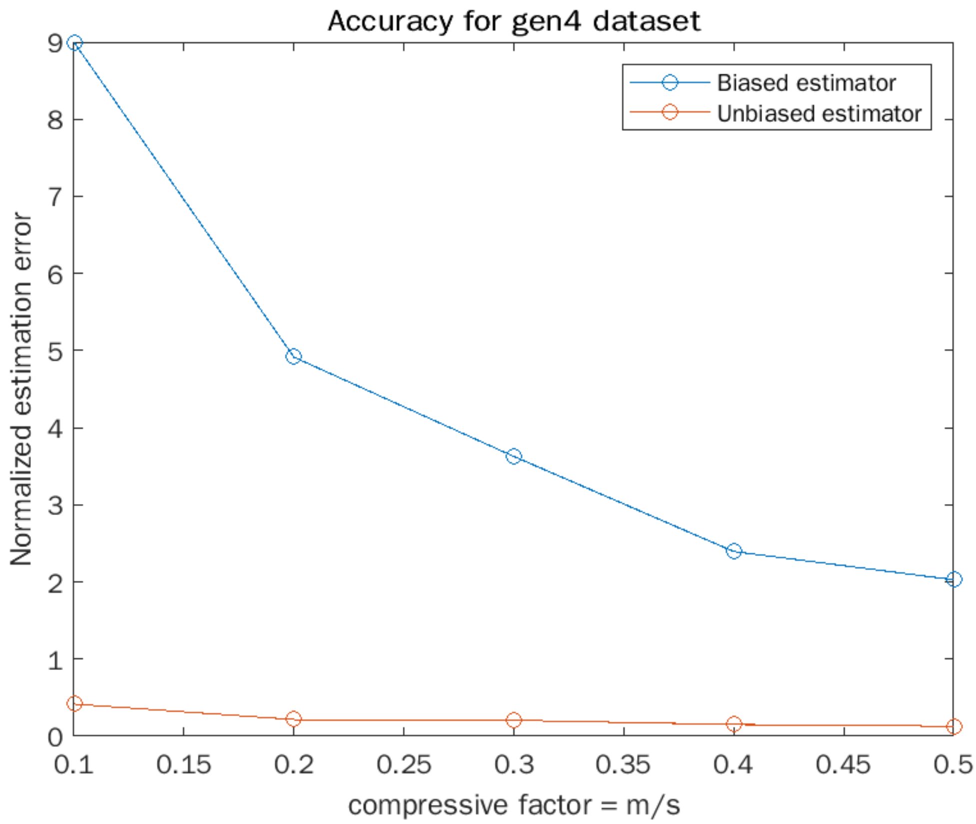
First eigenvector of the underlying covariance matrix and the first eigenvectors of the estimated unbiased covariance matrix for $\gamma = 0.1, 0.3, 0.5$ are shown in the below figure,



4.2 Gen4 data set

Gen4 data contains a matrix of size 1537 x 4298 that is the dimension of each x_i is $p = 1537$ and the number of data samples $n = 4298$.

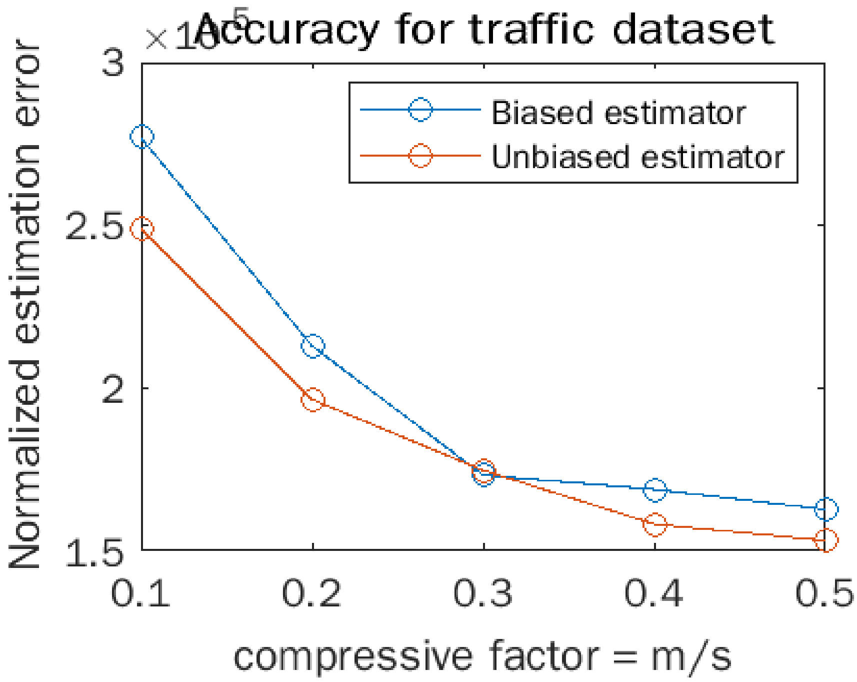
This below figure shows the normalized estimation error of the estimated biased and unbiased covariance matrices at $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$



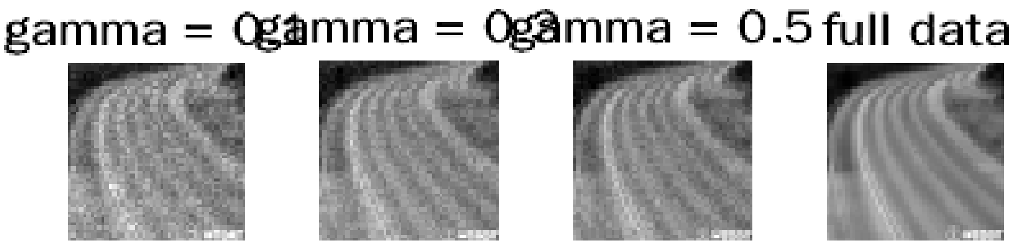
4.3 Traffic data set

Traffic data sets contains images which taken from a stationary camera that captures the video surveilliance of traffic. Each image is a 48x48 image and the vectorized image dimension is $p = 48 * 48 = 2304$. The number of data samples taken are n = 4621 frames.

This below figure shows the normalized estimation error of the estimated biased and unbiased covariance matrices at $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$



First eigenvector of the underlying covariance matrix and the first eigenvectors of the estimated unbiased covariance matrix for $\gamma = 0.1, 0.5$ are shown in the below figure,



4.4 Observations on graphs and images produced

The accuracy of the unbiased estimated covariance matrix is compared with accuracy of the biased estimated covariance matrix. In all three graphs, we can see that RMSE values for unbiased covariance matrix are less than the biased covariance matrix. Also, the accuracy increases(RMSE decreases) with the increase of compression factor that is with the decrease of s, this is because here s describes the sparsity of random variables R_i such that each column has $\frac{p}{s}$ non-zero entries. As, s decreases the bias term decreases, hence accuracy increases with the decrease of s.
As we can see that first eigenvector of the underlying covariance matrix C_n and the estimated unbiased covariance matrix are almost for $\gamma = 0.1, 0.3, 0.5$. Also as gamma increases, eigenvectors of estimated unbiased covariance matrix become more similar to the eigenvector of underlying covariance matrix, this is because the as gamma increases bias decreases.

5 Conclusions

- Unlike previous papers/studies, this paper doesn't make assumptions about the structure of the original covariance matrix such as being low-rank. Our theoretical result, on the other hand, is applicable to a wide range of random projections with entries drawn from a zero-mean distribution with finite first four moments.

- Without imposing distributional assumptions on the set of data samples, a fixed data setup is considered.
- To demonstrate the correctness of our suggested approach for varied compression factor values, we presented experimental findings on several real-world data sets like MNIST, Gen4 and traffic datasets.
- The proposed theoretical approach does not necessitate random projections that meet the Johnson-Lindenstrauss distributional characteristic.

6 References

- https://arxiv.org/pdf/1512.08887.pdf
- http://stat88.org/textbook/notebooks/Chapter_05/04_Unbiased_Estimators.html
- https://en.wikipedia.org/wiki/Moment_(mathematics)