CS 215 Assignment 4 Report

Diwan Anuj Jitendra - 170070005 Soumya Chatterjee - 170070010

November 10, 2018

Problem 2

Part (a)(Algorithm for generating samples)

A random vector X of dimension $D \times 1$ is a multivariate Gaussian iff it can be expressed as $\mu + AW$ where μ is a vector of dimension $D \times 1$, A is some vector of dimension $D \times N$ and W is a random vector composed of N independent standard normals. We also know that $C = AA^T$. Thus, given C, if we can find such an A, then we know that the r.v. $X = \mu + AW$ is a multivariate Gaussian with the given mean μ and covariance C.
Since C is symmetric and positive semidefinite, it has an eigen decomposition of the form:

$$C = Q\Lambda Q^T = Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q^T$$

where Q is an orthonormal eigenvector matrix and Λ is a diagonal eigenvalue matrix with non-negative diagonal entries(By SPSD property). Note that for a diagonal matrix, square root is defined as the matrix formed by taking the square root of each diagonal entry. Since diagonal entries are non-negative this is a real matrix.
Choosing $A = Q\Lambda^{\frac{1}{2}}$, we have:

$$AA^T = Q\Lambda^{\frac{1}{2}}(Q\Lambda^{\frac{1}{2}})^T = Q\Lambda^{\frac{1}{2}}(\Lambda^{\frac{1}{2}})^T Q^T = Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q^T = C$$

as desired. ($\Lambda^{\frac{1}{2}}$ is symmetric since it is diagonal) Thus the algorithm is as follows:

1. Perform eigen decomposition using the `eig` function and obtain Q and Λ. Compute $A = Q\Lambda^{\frac{1}{2}}$.
2. For each N, generate $2 \times N$ sample matrix using `randn`, left-multiply by A and add μ to each column to get N samples X_i from our desired Gaussian.
3. Compute the MLE estimates as:
$$\mu_{est} = \frac{1}{N}\sum_{i=1}^{N} X_i$$
and
$$C_{est} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_{est})(X_i - \mu_{est})^T$$
4. Compute the errors as given by the formulae.
5. For plotting the principal modes of variation, perform eigen decomposition of C_{est} and get Q_{est} and Λ_{est}. The columns in Q_{est} are the unit vectors pointing in the principal modes of variation and the corresponding diagonal values in Λ_{est} are the eigenvalues(variances). Thus using these, the required lines can be plotted.

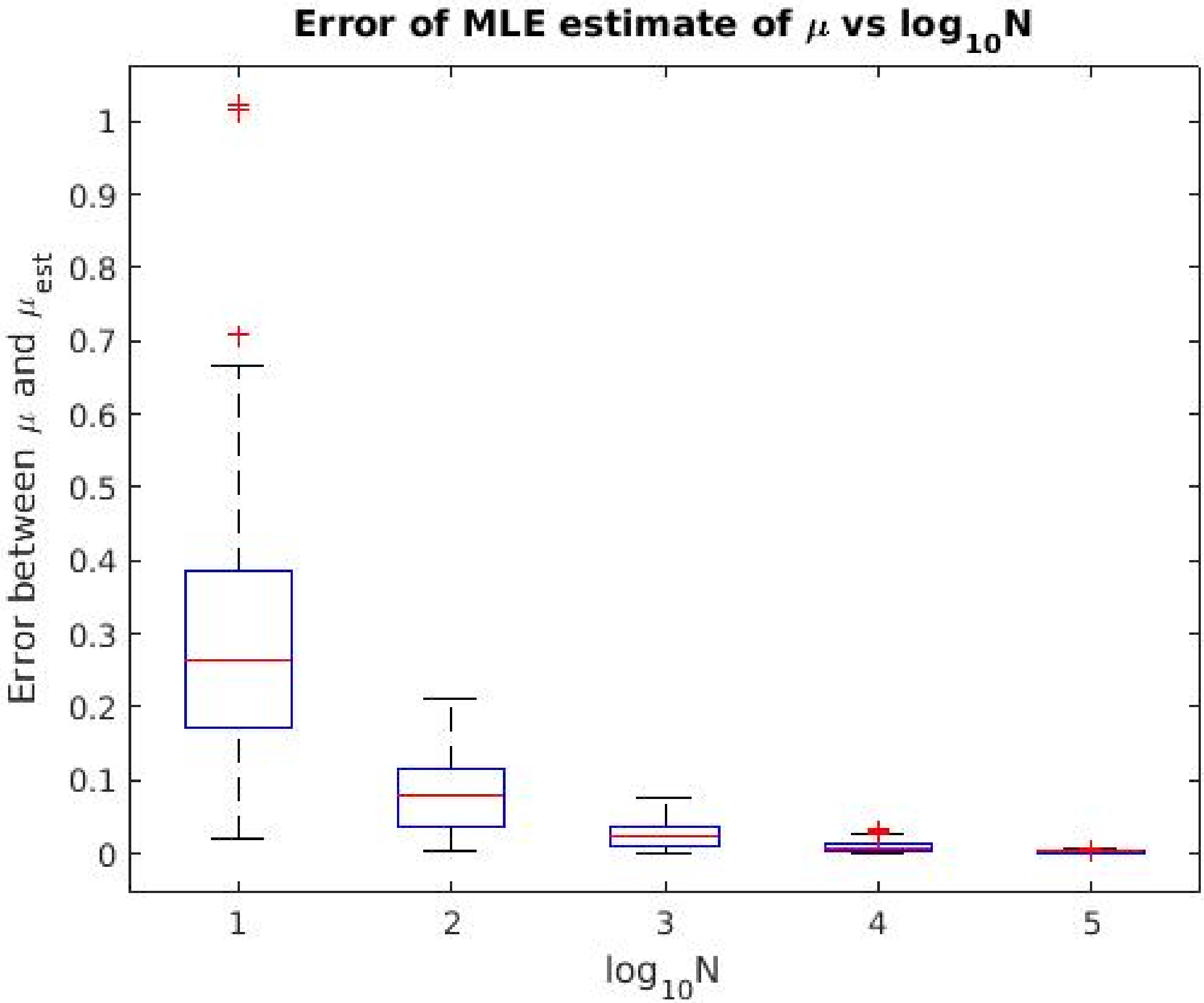


Figure 1: Error for mean

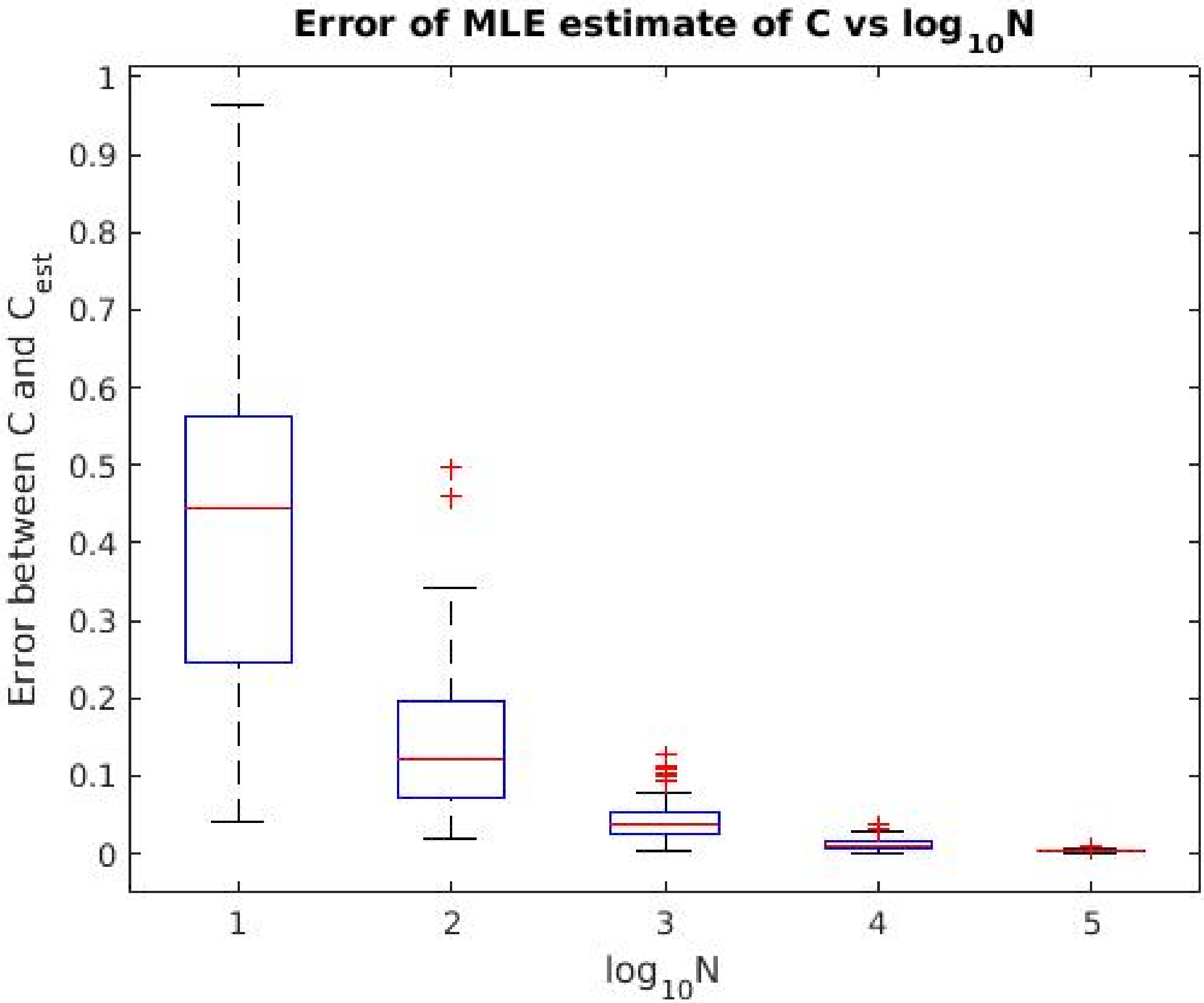
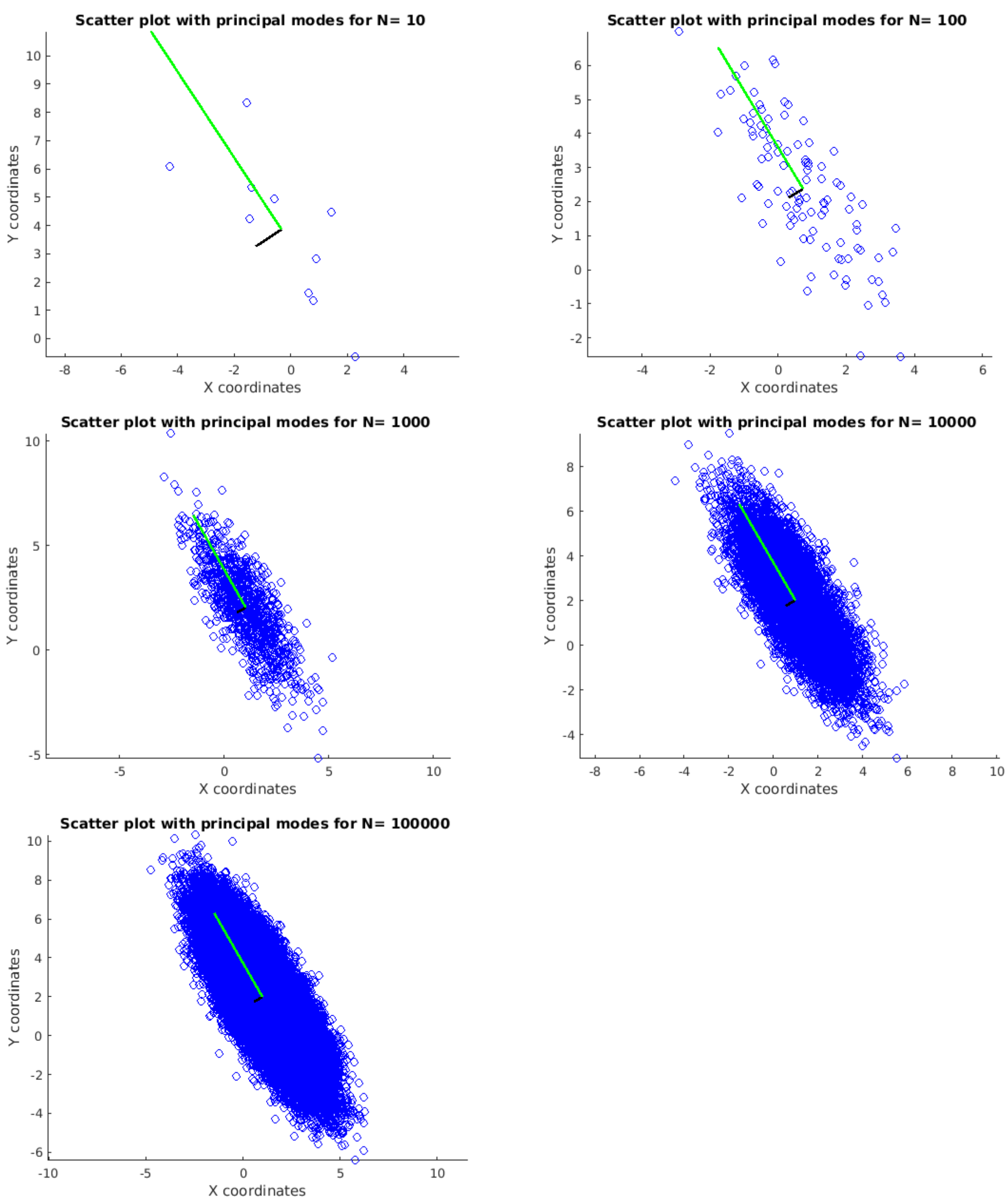


Figure 2: Error for covariance


Scatter plot with principal modes for N= 10
Scatter plot with principal modes for N= 100
Scatter plot with principal modes for N= 1000
Scatter plot with principal modes for N= 10000
Scatter plot with principal modes for N= 100000
X coordinates
Y coordinates