HW1 — CSE8803: IUQ

Ashish Dhiman — ashish.dhiman9@gatech.edu

May 7, 2023

Contents

1 Question 1: Joint Normal

We are given two random vectors $X = (X, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$ with joint distribution:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} m_X \\ m_Y \end{bmatrix}, \begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix}\right)$$

1.1 part a: Marginal Distribution

Let us define Z as a RV vector as $[X \quad Y]^T$, with

$$m_X = [m_X \quad m_Y]^T \quad n+m, 1 \text{ dimensional vector}$$

$$C_Z = \begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix} \quad n+m, n+m \text{ dimensional vector}$$

$$\implies Z \sim \mathcal{N}(m_Z, C_Z)$$

$$\text{We can write X as} : X = S_X Z = \begin{bmatrix} 1 & 0 & 0\dots0 \\ 0 & 1 & 0\dots0 \\ 0 & 0 & 1\dots0 \\ \dots & & \\ \dots & & \\ 0 & 0 & \dots0 \end{bmatrix} * \begin{bmatrix} X \\ \dots \\ X_n \\ Y_1 \\ \dots \\ Y_m \end{bmatrix} \tag{1}$$

Thus S_X is $(n, n+m)$matrix with $s_{ij} = 1$ if ith element in Z is X_j

Now we know that linear transformation of a Normal RV is Normal,i.e.

$$W \sim \mathcal{N}(0, \sigma_w^2) \implies aW + c \sim \mathcal{N}(a * 0 + c, a^2\sigma_w^2)$$

Hence it follows:

$$X \sim \mathcal{N}(S_X m_Z, S_X * C_Z * S_X^T)$$

$$S_X * m_Z = \begin{bmatrix} 1 & 0 & 0\dots0 \\ 0 & 1 & 0\dots0 \\ 0 & 0 & 1\dots0 \\ \dots & & \\ \dots & & \\ 0 & 0 & \dots0 \end{bmatrix}_{n,n+m} * \begin{bmatrix} m_X \\ \dots \\ m_{X_n} \\ m_{Y_1} \\ \dots \\ m_{Y_m} \end{bmatrix} = \begin{bmatrix} m_X \\ \dots \\ m_{X_n} \\ 0 \\ \dots \\ 0 \end{bmatrix} = m_X$$

$$\implies S_X * C_Z * S_X^T = \begin{bmatrix} 1 & 0 & 0\dots0 \\ 0 & 1 & 0\dots0 \\ 0 & 0 & 1\dots0 \\ \dots & & \\ \dots & & \\ 0 & 0 & \dots0 \end{bmatrix}_{n,n+m} * \begin{bmatrix} C_{X\,n,n} & C_{XY\,n,m} \\ C_{YX\,m,n} & C_{Y\,m,m} \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0\dots0 \\ 0 & 1 & 0\dots0 \\ 0 & 0 & 1\dots0 \\ \dots & & \\ \dots & & \\ 0 & 0 & \dots0 \end{bmatrix}^T_{n+m,n}$$

$$= \begin{bmatrix} C_{X\,n,n} & C_{XY\,n,m} \end{bmatrix}_{n,n+m} * \begin{bmatrix} 1 & 0 & 0\dots0 \\ 0 & 1 & 0\dots0 \\ 0 & 0 & 1\dots0 \\ \dots & & \\ \dots & & \\ 0 & 0 & \dots0 \end{bmatrix}^T_{n+m,n} = C_X \tag{2}$$

Hence we have proved that marginal distribution of X

$$\implies X \sim \mathcal{N}(S_X m_Z, S_X * C_Z * S_X^T) = \mathcal{N}(m_X, C_X)$$

We can use the same approach to prove for Y, with $Y = S_Y Z$

1.2 part b: Conditional Distribution

We define conditional distribution

$$p(X|Y=y) = \frac{p(x,y)}{p(y)}$$

$$\implies p(X|Y) = \frac{\mathcal{N}(m_Z, C_Z)}{\mathcal{N}(m_Y, C_Y)}$$

with Z defined in part a above.

$$\implies p(X|Y) = \frac{1/\sqrt{(2\pi)^{n+m}|C_Z|} \cdot \exp\left[-\frac{1}{2}(z-m_Z)^{\mathrm{T}} C_Z^{-1}(z-m_Z)\right]}{1/\sqrt{(2\pi)^m |C_Y|} \cdot \exp\left[-\frac{1}{2}(y-m_Y)^{\mathrm{T}} C_Y^{-1}(Y-m_Y)\right]}$$

$$= \frac{1}{\sqrt{(2\pi)^n}} \cdot \sqrt{\frac{|C_Y|}{|C_Z|}} \cdot \exp\left[-\frac{1}{2}(z-m_Z)^{\mathrm{T}} C_Z^{-1}(x-m_Z) + \frac{1}{2}(y-m_Y)^{\mathrm{T}} C_Y^{-1}(Y-m_Y)\right] \tag{3}$$

Let's look at the determinant term first:

$$\begin{aligned} det(C_Z) = \begin{vmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{vmatrix} &= |C_Y| \cdot |C_X - C_{XY} C_Y^{-1} C_{YX}| \\ \implies \frac{|C_Y|}{|C_Z|} &= \frac{|C_Y|}{|C_Y| \cdot |C_X - C_{XY} C_Y^{-1} C_{YX}|} \\ &= \frac{1}{|C_X - C_{XY} C_Y^{-1} C_{YX}|} \end{aligned} \tag{4}$$

Simplification of the exponent terms:

$$= \left[-\frac{1}{2}\left(\begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} m_X \\ m_Y \end{bmatrix}\right)^{\mathrm{T}} \begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix}^{-1} \left(\begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} m_X \\ m_Y \end{bmatrix}\right) \right.$$
$$\left. + \frac{1}{2}(Y - m_Y)^{\mathrm{T}} {}_Y^{-1} (Y - m_Y)\right]$$

Now the inverse of a matrix is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$

Thus

$$\begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix}^{-1} = \begin{bmatrix} (C_X - C_{XY}C_Y^{-1}C_{YX})^{-1} & -(C_X - C_{XY}C_Y^{-1}C_{YX})^{-1}C_{XY}C_Y^{-1} \\ -C_Y^{-1}C_{YX}(C_X - C_{XY}C_Y^{-1}C_{YX})^{-1} & C_Y^{-1} + C_Y^{-1}C_{YX}(C_X - C_{XY}C_Y^{-1}C_{YX})^{-1}C_{XY}C_Y^{-1} \end{bmatrix} \tag{5}$$

Plugging the inverse in and simplifying we are left with:

$$= -\frac{1}{2} * \left[X - \left(m_X + C_{XY}C_Y^{-1}(y - m_Y)\right)\right]^{\mathrm{T}} (C_X - C_{XY}C_Y^{-1}C_{YX})^{-1} \left[X - \left(m_X + C_{XY}C_Y^{-1}(y - m_Y)\right)\right] \tag{6}$$

Hence we have

$$
\begin{aligned}
p(X|Y) = -\frac{1}{\sqrt{(2\pi)^n}}* \\
(\frac{1}{|C_X - C_{XY}C_Y^{-1}C_{YX}|})^{1/2}* \\
\exp(\frac{1}{2} * \left[X - \left(m_X + C_{XY}C_Y^{-1}(y - m_Y)\right)\right]^{\mathrm{T}} * \\
(C_X - C_{XY}C_Y^{-1}C_{YX})^{-1}\left[X - \left(m_X + C_{XY}C_Y^{-1}(y - m_Y)\right)\right]
\end{aligned} \tag{7}
$$

Now Let $m_{X|Y} = m_X + C_{XY}C_Y^{-1}(y - m_Y)$ and $C_{X|Y} = C_X - C_{XY}C_Y^{-1}C_{YX}$ Hence,

$$
\implies p(X|Y) = \frac{1}{\sqrt{(2\pi)^n det(C_{X|Y})}} \cdot \exp(\frac{1}{2} * (X - m_{X|Y})^T * C_{X|Y} * (X - m_{X|Y}))
$$

Thus conditional distribution of X is

$$
X \sim \mathcal{N}(m_{X|Y}, C_{X|Y})
$$

1.3 part c: MLE estimate

We have i.i.d. random samples $(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})$

We need to find

$$
\theta_{mle} = \arg\max_{\theta} Lik(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})|\theta)
$$

with $\theta = [m_X, m_Y, C_X, C_Y]$

Let us again define Z as a RV vector= $[X \quad Y]^T$

Hence $z^i \sim \mathcal{N}(m_Z, C_Z)$, where:

$$
z^i \in \mathcal{R}^{n+m}, m_Z \in \mathcal{R}^{n+m}, C_Z \in \mathcal{R}^{n+m}
$$

We can now find m_Z, C_Z, such that we maximise the likelihood, i.e.

$$
m_{Zmle}, C_{Zmle} = \arg\max_{m_Z, C_Z} = Lik(z_1, z_2 ... z_n | m_Z, C_Z)
$$

We can simplify:

$$
\begin{aligned}
Lik = p(z^1, z^2 ... z^N | m_Z, C_Z) \\
= p(z^1|m_Z, C_Z) * p(z^2|m_Z, C_Z) ... p(z^N|m_Z, C_Z) = \prod_{i=1}^{N} p(z^i|m_Z, C_Z) \\
= \prod_{i=1}^{N} \frac{1}{2\pi^{(n+m)/2} det(C_Z)^{1/2}} * \exp(\frac{-1}{2}(z^i - m_Z)^T C_Z^{-1}(z^i - m_Z)) \quad \text{(From pdf)} \\
= \frac{1}{2\pi^{N(n+m)/2} det(C_Z)^{N/2}} * \exp\frac{-1}{2}\sum_{i=1}^{N}((z^i - m_Z)^T C_Z^{-1}(z^i - m_Z)) \\
\text{Taking log and dropping constant terms, we get} \\
\log Lik \propto -\frac{N}{2} log(det(C_Z)) \frac{-1}{2}\sum_{i=1}^{N}((z^i - m_Z)^T C_Z^{-1}(z^i - m_Z)) \\
\propto -N \log(det(C_Z)) - \sum_{i=1}^{N}((z^i - m_Z)^T C_Z^{-1}(z^i - m_Z))
\end{aligned} \tag{8}
$$

Now max -x is same as min x, and the problem above is Convex, since we know that C_Z is positive definite, therefore C_Z^{-1} is also positive definite.

Therefore we can apply first order optimality condition and put gradient wrt to $m_Z, C_Z = 0$

$$\begin{aligned}\nabla_{m_Z} \log Lik = 0 + \sum_{i=1}^{N} 2C_Z^{-1}(z^i - m_Z) = 0 \\ \sum_{i=1}^{n} m_Z = \sum_{i=1}^{n}(z^i) \\ \text{Thus we have:} \\ \implies m_{Z\,mle} = \frac{1}{N}\sum_{i=1}^{N}(z^i) \\ \implies [m_{X\,mle} \quad m_{Y\,mle}]^T = \frac{1}{N}\sum_{i=1}^{N}([x^i \quad y^i]^T)\end{aligned} \tag{9}$$

Similarly taking gradient wrt C_Z^{-1}

$$\begin{aligned}\nabla_{C_Z^{-1}} \log Lik = \nabla_{C_Z^{-1}}(-N\log(det(C_Z))) + \sum_{i=1}^{N}((z^i - m_Z)^T(z^i - m_Z)) = 0 \\ \text{Since } det(A) = 1/det(A^{-1}) \\ \nabla_{C_Z^{-1}}(N\log(det(C_Z))) = N(C_Z^{-1})^{-1T} \quad \text{From } \nabla_A log(det(A) = (A^{-1})^T \\ \implies \nabla_{C_Z^{-1}} \log Lik = -NC_Z^T + (\sum_{i=1}^{N}((z^i - m_Z)(z^i - m_Z)^T))^T = 0 \\ \implies C_{Z\,mle} = \frac{1}{N}\sum_{i=1}^{N}(z^i - \hat{m_Z})(z^i - \hat{m_Z})^T \\ \text{where :} C_Z = \begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix} \quad z^i = [x^i \quad y^i]^T\end{aligned} \tag{10}$$

Thus we ca take n, n sub matrix from left top corner of $C_{Z\,mle}$ to get estimate of $C_{X\,mle}$. Similarly $C_{Y\,mle}$ will be sub matrix of m, m from right bottom corner.

2 Question 2: Gaussian RV

2.1 part a: drawing samples

- $d = 1$, $\alpha = 1, 2$, $\kappa = 1, 4, 16$

  A Gauss Random Field with one dimension is same as a Gaussian process. Given below are some samples generated for the above parameters:

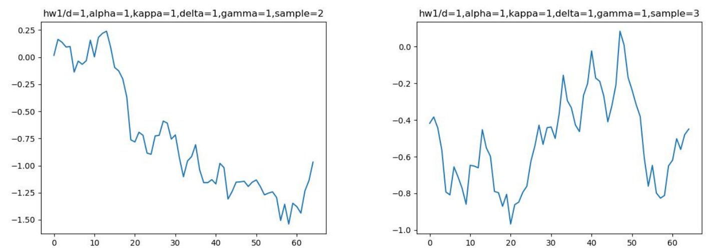

Figure 1: Two samples for d=1,alpha=1,kappa=1

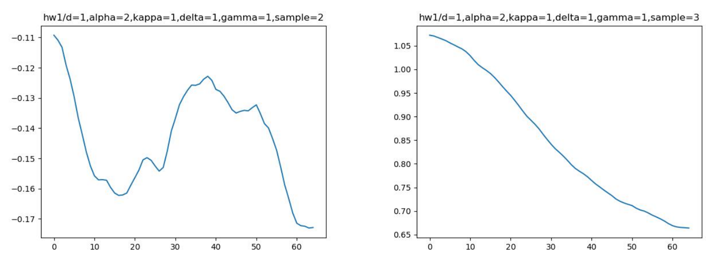

Figure 2: Two samples for d=1,alpha=2,kappa=1

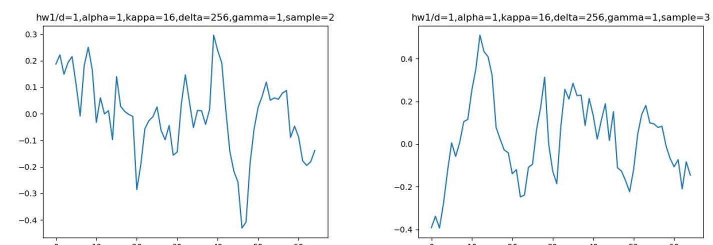

Figure 3: Two samples for d=1,alpha=1,kappa=16

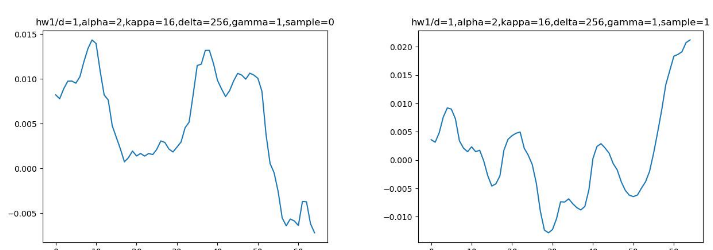

Figure 4: Two samples for d=1,alpha=2,kappa=16

From above plots it is apparent that higher the α, greater is the smoothness of the process. While increasing κ decreases corelation.

- $d = 2$, $\alpha = 1, 2$, $\kappa = 1, 4, 16$
  Plots of samples in 2D case are given below:

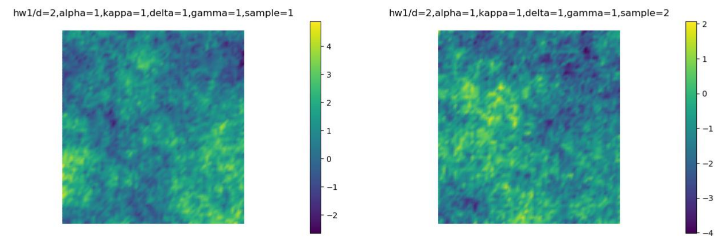

Figure 5: Two samples for d=2,alpha=1,kappa=1

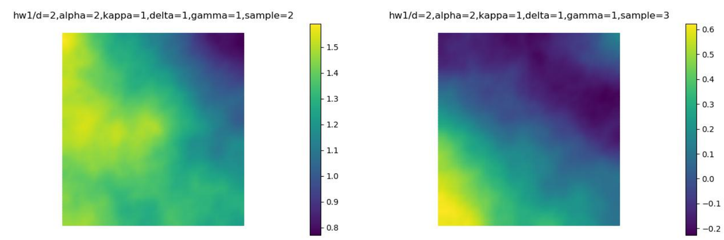

Figure 6: Two samples for d=2,alpha=2,kappa=1

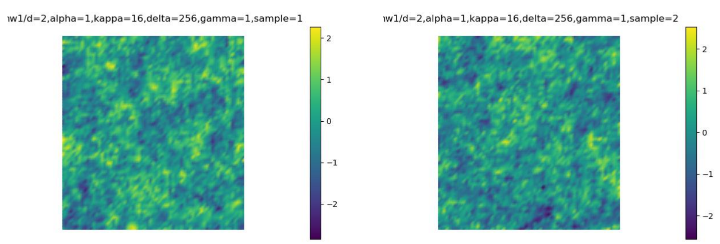

Figure 7: Two samples for d=2,alpha=1,kappa=16

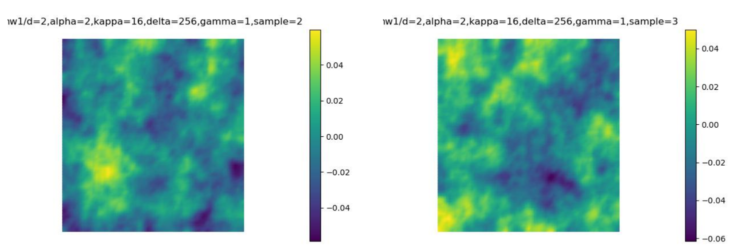

Figure 8: Two samples for d=2,alpha=1,kappa=16

Here to similar to 1D case, increasing alpha makes the mesh smoother, while increasing kappa makes values less corelated. Infact the effect of kappa is more pronounced in the 2D case relative to 1D case.

$d = 3$, $\alpha = 1, 2$, $\kappa = 1, 4, 16$

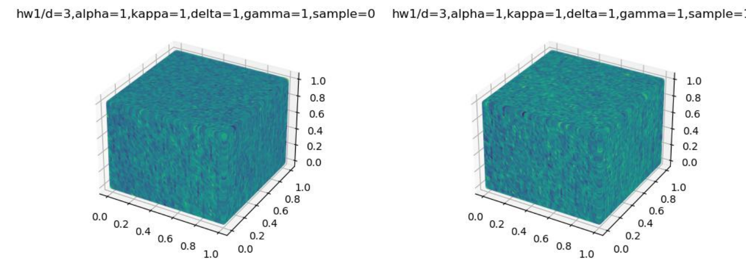

Figure 9: Two samples for d=3,alpha=1,kappa=1

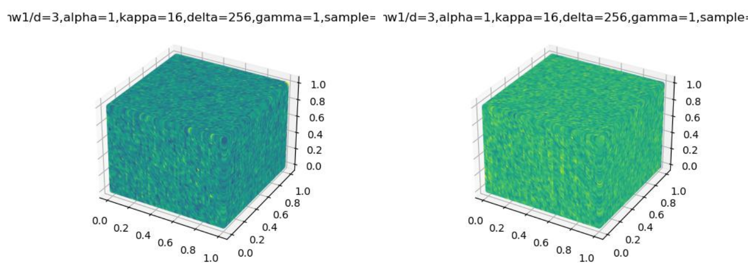

Figure 10: Two samples for d=3,alpha=2,kappa=16

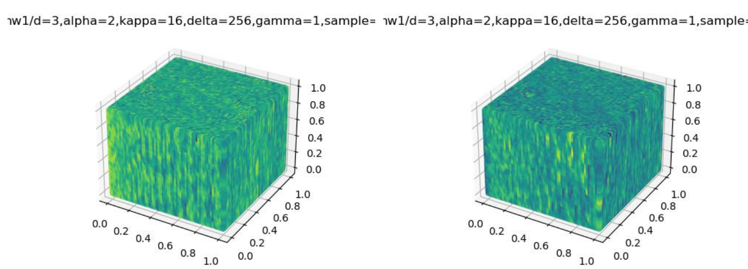

Figure 11: Two samples for d=3,alpha=2,kappa=16

Rest of the samples for all the enumerations can be found in the datafiles submitted on Canvas :)

2.2 part b: KL expansion

We are asked to use KL expansion to sample from Gaussian Random field with KL expansion

Case 1: $d = 1, \alpha = 1, \kappa = 1$

Plot of X_r and Truncation error:

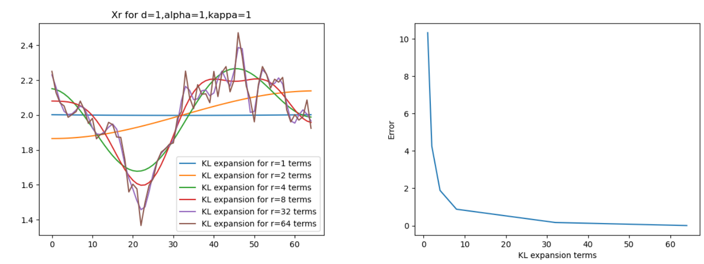

Figure 12: Xr for different r terms and corresponding Truncation error

Thus we see that as we increase r, or terms in KL Explansion, the X_r starts to look more and more like samples from part (a). This point is further validated by the truncation error plot, which decreases as we increase 'r'. In other words, as we include more terms in the KL expansion, X_r starts to converge. This is driven by the diminishing scale of eigen values, since in KL expansion the additive terms are scaled by root of eigen values. So greater the number of eigen values we pick higher is the accuracy.

Plot of Eigen values and eigen vectors

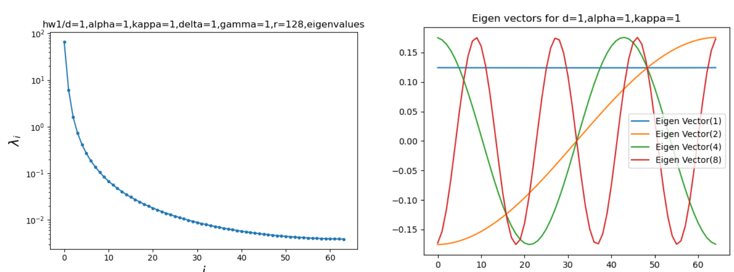

Figure 13: Plot of eigen values (left) and eigen vectors (right)

The truncation error here is calculated from the remaining eigen values, i.e r+1 to 128(benchmark).

As discussed above, since the truncation error is proportional to the magnitude of eigen values, we can see that including uptill 10th term reduces the truncation error a lot. After the 10th index the decay of eigen values slows down and this is also represented in the Truncation error graph, which decays much slowly after the 10th index.

Case 2: $d = 2, \alpha = 2, \kappa = 4$
Plot of X_r:

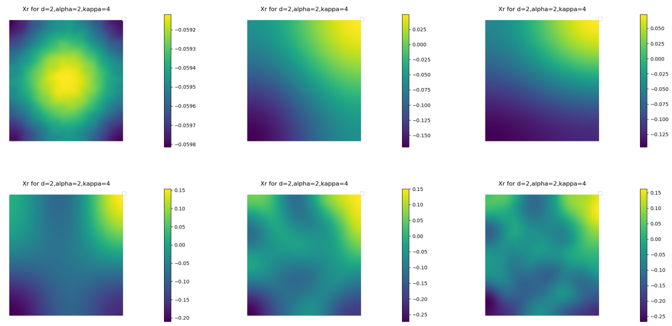

Figure 14: Plot of eigen values (left) and eigen vectors (right)

Similar to 1d case we see that as we increase r, or terms in KL Explansion, the X_r starts to look more and more like samples from part (a). This point is further validated by the truncation error plot, which decreases as we increase 'r'.

Eigenvalues, Eigen functions and truncation error

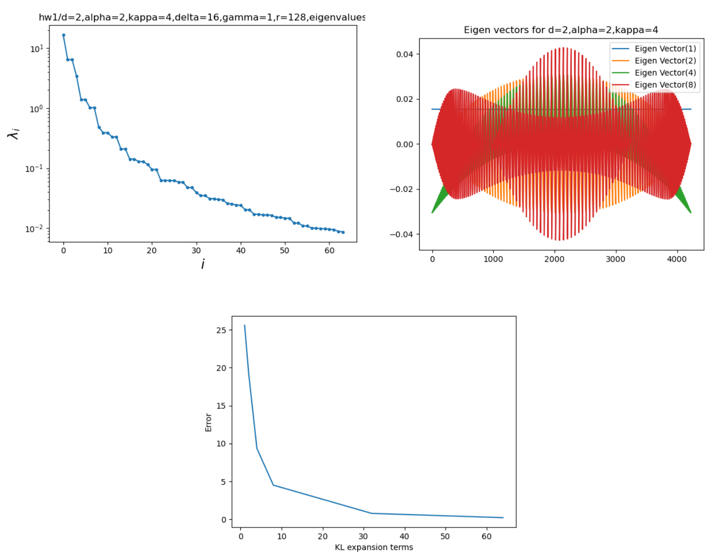

Figure 15: Plot of Eigen vectors, eigen values and truncation error

From plot of eigen vectors we see that higher the index, greater is the oscillation frequency of the vector.

And since the truncation error is proportional to the magnitude of eigen values, we can see that beyond the 30th index, the decay of eigen values slows down. This is also represented in the Truncation error graph, which decays much slowly after the 30th index. This is in contrast to 1d case, where error plateaued at around 10.

2.3 part c: Estimated Covariance

For this part we draw 100 samples of Gaussian field from FEM method, with $d = 2, \alpha = 2, \kappa = 4$.

Because we discretize to 64 points in x and y domain, we have a total of 65*65 points. From the 100 realisations of these 4225 RV we estimate the covariance of the GRF.

The plot of true vs estimated eigen values is given below: *Eigenvalues, Eigen functions and truncation error*

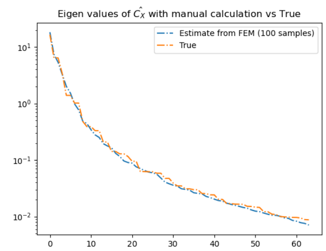

Figure 16: Plot of true vs estimated eigen values of covariance matrix

We can see that the true and estimated eigen values very closely track each other. However, if we are to pick say only 10 samples, eigen values beyond 10 drop to zero.

Eigenvalues, Eigen functions and truncation error

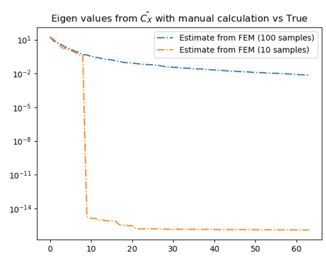

Figure 17: Eigen with less samples

3 References:

1. https://en.wikipedia.org/wiki/Multivariate_normal_distribution
2. https://stats.stackexchange.com/questions/73225/joint-distribution-of-two-multivariate-
3. https://statproofbook.github.io/P/mvn-marg.html
4. https://statproofbook.github.io/P/mvn-ltt
5. https://stats.stackexchange.com/questions/7263/difference-between-the-terms-joint-dist
~:text=A%20joint%20normal%20distribution%20is,to%20be%20called%
20out%20separately.
6. https://online.stat.psu.edu/stat505/lesson/6/6.1
7. https://stats.stackexchange.com/questions/30588/deriving-the-conditional-distributions-