This week we built upon our discussion of probability and Gaussian distributions from last week to use Gaussian distributions for probabilistic estimations. In particular we covered the concept of minimum mean square error (MMSE). The important
takeway from this is the fact that the average performance will increase given conditional information, even though our estimate is the same.

We learned how to perform MMSE estimation using the Gaussian distribution and learned about close form solutions of the estimate and error based on observed random variables. Conditional independence was also discussed in the context of graphical models and the corresponding covariance matrix entries.

We also went over how to use the Schur complement. The Schur complement is really useful since it allows us to update the inverse of a matrix with new data without recomputing our existing inverse. This means for real time critical applications we can easily incorporate new data to obtain a better model on the fly.

Akash Patel
MFML HMWK 8

2. Let G be the set of zero-mean Gaussian random variables. Since if X and Y are zero-mean Gaussian $aX+bY$ is also zero-mean Gaussian for all $a,b \in \mathbb{R}$, it should be clear that G is a linear vector space. It is easy to check that the correlation $E[XY]=\langle X,Y\rangle$ is a valid inner product on G that induces the norm $\|X\|^2 = \text{var}(X)$.

a) Translate the Pythagorean Theorem and the Cauchy-Schwartz inequality into statements about correlation and variance of pairs of Gaussian random variables. Discuss.

Pythagorean: $\langle X,Y\rangle = 0 \Rightarrow \|X+Y\|^2 = \|X\|^2 + \|Y\|^2$

$$E[XY] = 0 \Rightarrow \text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$$

In general the $\text{var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$

$$= \text{var}(X) + \text{Var}(Y) + 2(E[XY] - E[X]E[Y])$$

$E[XY]=0 \Rightarrow$ $$= \text{var}(X) + \text{Var}(Y) + 2(-E[X]E[Y])$$

$E[XY]$ also implies

$E[X]E[Y]=0$ for a Gaussian

$$= \text{var}(X) + \text{var}(Y) \quad \checkmark$$

For the Gaussian distribution the Pythagorean theorem implies independence if X and Y are uncorrelated.

Cauchy-Schwartz: $|\langle X,Y\rangle| \le \|X\|\,\|Y\|$

$$\begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

$$|E[XY]| \le \sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}$$

$$|E[XY]| \le \sqrt{E[XX^T]E[YY^T]}$$

$$E[XY]^2 \le E[XX^T]E[YY^T]$$

$$\sigma_{XY}^2 \le \sigma_X^2 \cdot \sigma_Y^2$$

The C-S version of Gaussian distributions states that the covariance between X and Y squared can atmost be the variance of X squared times the Variance of Y squared. This makes sense as the covariance between two random variables cannot be higher than the variances of each.

b) Let $X_1, \ldots, X_N \in G$ with correlations captured in entries of the matrix $R_{ij} = E[X_i X_j]$. Let Y be another arbitrary point in G. Using what you know about finding the closest point (in induced norm) to a subspace, describe how to find the best linear predictor of Y from the $\{X_n\}$. That is, describe how to solve the optimization problem

$$\underset{\omega_1, \ldots, \omega_N}{\text{minimize}} \ \text{var}\left(Y - \sum_{n=1}^{N} w_n X_n\right)$$

using the covariance matrix R and the correlations $E[YX_n]$. Note that we mean "best" above in the mean-square sense, as

$$\text{var}\left(Y - \sum_{n=1}^{N} w_n X_n\right) = E\left[\left(Y - \sum_{n=1}^{N} w_n X_n\right)^2\right]$$

$$= E\left[\left(Y - \sum_{n=1}^{N} w_n X_n\right)\left(Y - \sum_{n=1}^{N} w_n X_n\right)\right] = E\left[YY^T - 2Y\sum_{n=1}^{N} w_n X_n + \left(\sum_{n=1}^{N} w_n X_n\right)^2\right]$$

$$= E[YY^T] - E[2Y w^T X] + E[w^T X X^T w]$$

$$= E[YY^T] - 2w^T E[YX] + w^T E[XX^T] w$$

$$= E[YY^T] - 2\sum_{n=1}^{N} w_n E[YX_n] + w^T R w$$

$$0 = \nabla_w \text{var}\left(Y - \sum_{n=1}^{N} w_n X_n\right) = \nabla_w \left(E[YY^T] - 2\sum_{n=1}^{N} w_n E[YX_n] + w^T R w\right)$$

$$0 = -E[YX] + Rw$$

$$Rw = E[YX]$$

$$\boxed{\hat{w} = R^{-1} E[YX]}$$

c) Let $\hat{\omega}_1, \ldots, \hat{\omega}_N$ be the solution to the optimization program above. Give the simplest expression possible for $E[Y(\hat{\omega}_1 X_1 + \cdots + \hat{\omega}_N X_N)]$.

$$E[Y \hat{\omega}^T X] = \hat{\omega}^T E[YX] = (R^{-1} E[YX])^T E[YX] = E[YX]^T R^{-T} E[YX]$$

$$= \boxed{E[YX]^T R^{-1} E[YX]}$$

symmetric

3. Let X_1 and X_2 be Gaussian random variables with

$E[X_1]=E[X_2]=0 \quad E[X_1^2]=E[X_2^2]=2 \quad E[X_1X_2]=1.$

Let $Y=X_1+X_2$. Suppose we observe $Y=1.5$. Find the conditional densities $f_{X_1}(x|Y=1.5)$ and $f(x_2|Y=1.5)$.

$$\begin{bmatrix} R_Y & R_{YX} \\ R_{XY} & R_X \end{bmatrix}$$

$$\begin{bmatrix} Y \\ X_1 \end{bmatrix} \sim \text{Normal}\left(0, \begin{bmatrix} E[YY] & E[YX_1] \\ E[X_1Y] & E[X_1X_1] \end{bmatrix}\right)$$

$E[X_1X_1]=2$

$E[YX_1]=E[(X_1+X_2)X_1]=E[X_1^2+X_2X_1]=E[X_1^2]+E[X_1X_2]=2+1=3$

similarly $E[X_1Y]=3$

$E[YY]=E[(X_1+X_2)(X_1+X_2)]=E[X_1^2]+E[X_2^2]+2E[X_1X_2]=2+2+2=6$

$$\begin{bmatrix} X_o \\ X_h \end{bmatrix} \sim \text{Normal}\left(0, \begin{bmatrix} R_o & R_{oh} \\ R_{oh}^T & R_h \end{bmatrix}\right)$$
$X_h|X_o=x_o \sim \text{Normal}(R_{oh}^T R_o^{-1} x_o, R_h - R_{oh}^T R_o^{-1} R_{oh})$ } From pg 23 of notes.

$$f_{X_1}(x|Y=1.5) = X_1|Y=\underset{y_o}{1.5} \sim \text{Normal}(R_{YX}^T R_Y^{-1} y_o, R_X - R_{YX}^T R_Y^{-1} R_{YX})$$

$R_{YX}^T R_Y^{-1} y_o = 3 \cdot \frac{1}{6} \cdot \frac{3}{2} = \frac{3}{4}$

$R_X - R_{YX}^T R_Y^{-1} R_{YX} = 2 - 3 \cdot \frac{1}{6} \cdot 3 = 2 - \frac{3}{2} = \frac{1}{2}$

$$\boxed{f_{X_1}(x|Y=1.5) \sim \text{Normal}\left(\tfrac{3}{4}, \tfrac{1}{2}\right)}$$

Due to the symmetry of the problem $f_{X_1}(x|Y=1.5) = f_{X_2}(x|Y=1.5) = \text{Normal}\left(\frac{3}{4}, \frac{1}{2}\right)$

4. Let X be a Gaussian random vector taking values in $\mathbb{R}^N$, let E be a Gaussian random vector taking values in $\mathbb{R}^M$, and let A be a $M \times N$ matrix. We have

$$X \sim \text{Normal}(0, R_x), \quad E \sim \text{Normal}(0, R_e), \quad X, E \text{ independent},$$

We will make observation of the random vector.

$$Y = AX + E$$

a) From your work above, it is clear that Y is a Gaussian random vector in $\mathbb{R}^M$ and that $E[Y] = 0$. Find the covariance matrix for the Gaussian random vector $\begin{bmatrix} X \\ Y \end{bmatrix}$ that takes values in R^{N+M}.

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \text{Normal}\left(0, \begin{bmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{bmatrix}\right) \qquad R_x = E[XX^T] = R_x$$

$$R_{xy} = E[XY^T] = E[X(AX+E)^T] = E[X(E^T + X^T A^T)]$$
$$= E[XE^T] + E[XX^T A^T] = E[X]E[E^T] + E[XX^T]A^T = 0 + R_x A^T = R_x A^T$$
$$R_{xy}^T = (R_x A^T)^T = A R_x^T = A R_x$$
$$R_y = E[YY^T] = E[(AX+E)(AX+E)^T] = E[AXX^T A^T + 2AXE^T + EE^T]$$
$$= AE[XX^T]A^T + 2AE[XE^T] + E[EE^T]$$
$$= AR_x A^T + R_e$$

$$\boxed{\begin{bmatrix} R_x & R_x A^T \\ AR_x & AR_x A^T + R_e \end{bmatrix}}$$

b) Suppose we observe $Y = y$. What is the minimum mean-square error estimate of X given $Y = y$?

$$\begin{bmatrix} Y \\ X \end{bmatrix} \sim \text{Normal}\left(0, \begin{bmatrix} AR_x A^T + R_e & AR_x \\ R_x A^T & R_x \end{bmatrix}\right)$$

$$\hat{x} = R_{yx}^T R_y^{-1} y = \boxed{R_x A^T (AR_x A^T + R_e)^{-1} y}$$ ← pg. 23 of notes

c) Suppose $R_x = \sigma_x^2 I$ and $R_e = \sigma_e^2 I$. In this case, your MMSE estimator should look familiar, and you should see immediately that $\hat{x}_{MMSE}$ is in the row space of A. What are the $\hat{\alpha}_n$ in the expression below?

$\hat{x}_{MMSE} = \sum_{n=1}^{N} \alpha_n v_n$ where the v_n are the right singular vectors of A

$$\hat{x}_{MMSE} = R_x A^T (A R_x A^T + R_e)^{-1} y$$
$$= \sigma_x^2 I A^T (A \sigma_x^2 I A^T + \sigma_e^2 I)^{-1} y$$
$$= \sigma_x^2 A^T (\sigma_x^2 A A^T + \sigma_e^2 I)^{-1} y$$
$$= A^T \left(A A^T + \frac{\sigma_e^2}{\sigma_x^2} I\right)^{-1} y$$
$$= A^T (A A^T + \delta I)^{-1} y$$
$$= V \Sigma U^T (U \Sigma V^T V \Sigma U^T + \delta I)^{-1} y$$
$$= V \Sigma U^T (U \Sigma^2 U^T + U \delta U^T)^{-1} y$$
$$= V \Sigma U^T (U (\Sigma^2 + \delta I)^{-1} U^T) y$$
$$= V \Sigma (\Sigma^2 + \delta I)^{-1} U^T y$$
$$= \sum_{n=1}^{N} v_n \underbrace{\left(\frac{\sigma_n}{\sigma_n^2 + \delta}\right) u_n^T y}_{\alpha_n} = \sum_{n=1}^{N} \alpha_n v_n$$

Let $\frac{\sigma_e^2}{\sigma_x^2} = \delta$ for convenience

SVD: $A = U \Sigma V^T$

$A^T = V \Sigma U^T$

Let σ_n be the singular values in $\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \sigma_3 & \\ & & & \ddots \end{bmatrix}$

$$\boxed{\alpha_n = \left(\frac{\sigma_n}{\sigma_n^2 + \frac{\sigma_e^2}{\sigma_x^2}}\right) u_n^T y}$$

d) Take R_x and R_e as in part c), and assume that A has full column rank. What is MSE $E[\|\hat{x}_{MMSE} - X\|_2^2]$ of the MMSE estimate $\hat{x}_{MMSE}$?

$$E[\|\hat{x}_{MMSE} - X\|_2^2 \mid Y = y] = \text{trace}(R_x - R_{yx}^T R_y^{-1} R_{yx})$$
$$= \text{trace}(R_x - R_x A^T (A R_x A^T + R_e)^{-1} A R_x)$$
$$= \text{trace}(\sigma_x^2 I - \sigma_x^2 I A^T (A \sigma_x^2 I A^T + \sigma_e^2 I)^{-1} A \sigma_x^2 I)$$
$$= \text{trace}(\sigma_x^2 I - A^T (A A^T + \delta I)^{-1} A \sigma_x^2)$$
$$= \sigma_x^2 (\text{trace}(I - A^T (A A^T + \delta I)^{-1} A))$$

Let $\delta = \frac{\sigma_e^2}{\sigma_x^2}$

$A = U\Sigma V^T \quad A^T = V\Sigma U^T$

$$= \sigma_x^2 \left(\text{trace}\left(I - V\Sigma U^T (U\Sigma V^T V\Sigma V^T + U\delta U^T)^{-1} U\Sigma V^T\right)\right)$$

$$= \sigma_x^2 \left(\text{trace}\left(I - V\Sigma U^T (U\Sigma^2 U + U\delta U^T)^{-1} U\Sigma V^T\right)\right)$$

$$= \sigma_x^2 \left(\text{trace}\left(I - V\Sigma U^T (U(\Sigma^2 + \delta I)^{-1} U^T) U\Sigma V^T\right)\right)$$

$$= \sigma_x^2 \left(\text{trace}\left(I - V\Sigma(\Sigma^2 + \delta I)^{-1} \Sigma V^T\right)\right)$$

$$= \sigma_x^2 \left(\text{trace}(I) - \text{trace}\left(V\Sigma(\Sigma^2 + \delta I)^{-1} \Sigma V^T\right)\right)$$

$$= \sigma_x^2 \left(\text{trace}(I) - \text{trace}\left(\Sigma V^T V \Sigma (\Sigma^2 + \delta I)^{-1}\right)\right) \qquad \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_N \end{bmatrix}$$

$$= \sigma_x^2 \left(\text{trace}(I) - \text{trace}\left(\Sigma^2 (\Sigma^2 + \delta I)^{-1}\right)\right)$$

$$= \sigma_x^2 N \quad - \sigma_x^2 \sum_{n=1}^{N} \frac{\sigma_n^2}{\sigma_n^2 + \frac{\sigma_\varepsilon^2}{\sigma_x^2}}$$

$$= \boxed{\sum_{n=1}^{N} \sigma_x^2 \left(1 - \frac{\sigma_n^2}{\sigma_n^2 + \frac{\sigma_\varepsilon^2}{\sigma_x^2}}\right)}$$

5. Let A be an $M \times N$ matrix. Suppose we have computed $P = (AA^T + \delta I)^{-1}$ and used it to form the ridge estimate $\hat{x} = A^T P y$ for some observation vector $y \in \mathbb{R}^M$. We now add a row to A, and an entry to y, forming

$$A' = \begin{bmatrix} A \\ a_{M+1}^T \end{bmatrix}, \quad y' = \begin{bmatrix} y \\ y_{M+1} \end{bmatrix}.$$

Using the Schur complement, describe how to form the updated ridge estimate $\hat{x}' = A'^T (A'A'^T + \delta I)^{-1} y'$ using only a few matrix-vector multiplications, dot products, and vector additions (and no additional matrix inversions).

$$P' = (A'A'^T + \delta I_{M+1})^{-1} = \left(\begin{bmatrix} A \\ a_{M+1}^T \end{bmatrix} \begin{bmatrix} A^T & a_{M+1} \end{bmatrix} + \delta I_{M+1} \right)^{-1}$$

$$= \left(\begin{bmatrix} AA^T + \delta I & A a_{M+1} \\ a_{M+1}^T A^T & a_{M+1}^T a_{M+1} + \delta \end{bmatrix} \right)^{-1}$$

$$P'^{-1} = \begin{bmatrix} AA^T + \delta I & A a_{M+1} + \delta \\ a_{M+1}^T A^T + \delta & a_{M+1}^T a_{M+1} + \delta \end{bmatrix}$$

Let $M = P'^{-1}$ then $M^{-1} = P'$

$M_{11} = AA^T + \delta I = P^{-1}$

$M_{12} = A a_{M+1}$

$M_{21} = a_{M+1}^T A^T$

$M_{22} = a_{M+1}^T a_{M+1} + \delta$

$M_{11}^{-1} = P$

Schur Complement:

$$P' = M^{-1} = \begin{bmatrix} M_{11}^{-1} + M_{11}^{-1} M_{12} S^{-1} M_{21} M_{11}^{-1} & -M_{11}^{-1} M_{12} S^{-1} \\ -S^{-1} M_{21} M_{11}^{-1} & S^{-1} \end{bmatrix}$$

$$S = M_{22} - M_{21} M_{11}^{-1} M_{12}$$

S is scalar in our case:

$$S = a_{M+1}^T a_{M+1} + \delta - a_{M+1}^T A^T P A a_{M+1}$$

$$\underline{S = \delta + a_{M+1}^T (I - A^T P A) a_{M+1}}$$

$$\frac{1}{a+b} = \frac{1}{b} + \frac{1}{a}$$

$$P'_{11} = P + P a_{M+1}^T A^T S^{-1} A a_{M+1} P$$

$$P'_{12} = -P A a_{M+1} S^{-1}$$

$$P'_{21} = -S^{-1} a_{M+1}^T A^T P$$

$$P'_{22} = S^{-1}$$

$$\hat{x}' = A'^T P' y'$$
$$= \begin{bmatrix} A^T & a_{m+1} \end{bmatrix} \begin{bmatrix} P'_{11} & P'_{12} \\ P'_{21} & P'_{22} \end{bmatrix} \begin{bmatrix} y \\ y_{m+1} \end{bmatrix}$$
$$= \begin{bmatrix} A^T & a_{m+1} \end{bmatrix} \begin{bmatrix} P'_{11} y + P'_{12} y_{m+1} \\ P'_{21} y + P'_{22} y_{m+1} \end{bmatrix}$$
$$= A^T (P'_{11} y + P'_{12} y_{m+1}) + a_{m+1} (P'_{21} y + P'_{22} y_{m+1})$$

$$\hat{x} \quad \boxed{\begin{aligned} &= A^T\left[\left(P + P a^T_{m+1} A^T S^{-1} A_{a_{m+1}} P\right) y - P A_{a_{m+1}} S^{-1} y_{m+1}\right] \\ &\quad + a_{m+1}\left[-S^{-1} a^T_{m+1} A^T P_y + S^{-1} y_{m+1}\right] \end{aligned}}$$

In addition to abusing the Schur complement to remove additional matrix inversions many of the above martrix-matrix, matrix-vector computations can be precomputed.

Let $A^T P = C_1$

$A^T P_y = C_1 y = C_2$

$a^T_{m+1} A^T = C_3$

$A a_{m+1} = C_3^T = C_4$

$A^T P A = C_1 A = C_5$

$S = \delta + a^T_{m+1}(I - A^T P A) a_{m+1} = \delta + a^T_{m+1}(I - C_5) a_{m+1} = C_6 \quad S^{-1} = \frac{1}{C_6} = C_7$

Then

$$\boxed{\hat{x} = C_2 + C_7 \left(C_1 C_3 C_4 P_y - P C_4 y_{m+1} - a_{m+1}\left(a^T_{m+1} C_2 + y_{m+1}\right)\right)}$$

6. In this problem, $X \in \mathbb{R}^D$ is a Gaussian random vector with $E[X]=0$ and $E[XX^T]=R$.

(a) Suppose that R is block diagonal in that it can be written

$$R=\begin{bmatrix} R_a & 0 \\ 0 & R_b \end{bmatrix},$$

where is R_a is a $D_a \times D_a$ symmetric positive definite matrix, R_b is a $D_b \times D_b$ sym+def matrix, $D_a+D_b=D$, and the zero matrices above are the appropriate sizes so that the dimensions work out. Argue that R^{-1} is also block diagonal and the joint pdf can be factored as,

$$f_{X_1,\dots,X_D}(x_1,\dots,x_D)=f_{X_1,\dots,X_{D_a}}(x_1,\dots,x_{D_a})\cdot f_{X_{D_a+1},\dots,X_D}(x_{D_a+1},\dots,x_D)$$

meaning that the random vectors

$$X_A=\begin{bmatrix} X_1 \\ \vdots \\ X_{D_a} \end{bmatrix} \quad X_B=\begin{bmatrix} X_{D_a+1} \\ \vdots \\ X_D \end{bmatrix}$$

are independent.

R is invertible if R_a and R_b are invertible.

$$R^{-1}=\begin{bmatrix} R_a^{-1} & 0 \\ 0 & R_b^{-1} \end{bmatrix} \qquad RR^{-1}=\begin{bmatrix} R_a & 0 \\ 0 & R_b \end{bmatrix}\begin{bmatrix} R_a^{-1} & 0 \\ 0 & R_b^{-1} \end{bmatrix}=\begin{bmatrix} R_aR_a^{-1} & 0 \\ 0 & R_bR_b^{-1} \end{bmatrix}$$

$$=\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

$$f_X(x) \propto \exp\left(-\tfrac{1}{2}x^TR^{-1}x\right)$$

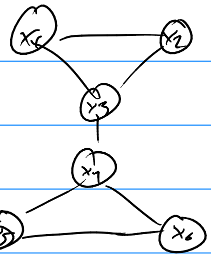
$$\propto \exp\left(-\frac{1}{2}\begin{bmatrix} x_A \\ x_b \end{bmatrix}^T\begin{bmatrix} R_a^{-1} & 0 \\ 0 & R_b^{-1} \end{bmatrix}\begin{bmatrix} x_a \\ x_b \end{bmatrix}\right)$$

$$\propto \exp\left(-\tfrac{1}{2}x_a^TR_a^{-1}x_a - \tfrac{1}{2}x_b^TR_b^{-1}x_b\right)$$

$$\propto \exp\left(-\tfrac{1}{2}x_a^TR_a^{-1}x_a\right)\exp\left(-\tfrac{1}{2}x_b^TR_b^{-1}x_b\right)$$

Thus $f_X(x)=f_{X_a}(x_a)\cdot f_{X_b}(x_b) \Rightarrow$ X_a, X_b are independent.

b) Suppose that X has inverse covariance R^{-1} whose non-zero entries are described by the graph.



Argue as rigorously as you can that the random vector $(X_1, X_2)|X_3$ is independent of the random vector $(X_4, X_5, X_6)|X_3$.

Let us construct the corresponding matrix of the graph:

'X' refers to non-zero and '0' refers to a zero entry

$$R^{-1} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{bmatrix} X & X & X & 0 & 0 & 0 \\ X & X & X & 0 & 0 & 0 \\ X & X & X & X & 0 & 0 \\ 0 & 0 & X & X & X & X \\ 0 & 0 & 0 & X & X & X \\ 0 & 0 & 0 & X & X & X \end{bmatrix} \end{array}$$

Now if we are conditioning on X_3 then the corresponding X_3 row and column entries are not considered, thus our updated R^{-1} matrix would be block diagonal and look like the following:

$$R^{-1} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{bmatrix} X & X & \cancel{X} & 0 & 0 & 0 \\ X & X & \cancel{X} & 0 & 0 & 0 \\ \cancel{X} & \cancel{X} & \cancel{X} & \cancel{X} & \cancel{0} & \cancel{0} \\ 0 & 0 & \cancel{X} & X & X & X \\ 0 & 0 & \cancel{0} & X & X & X \\ 0 & 0 & \cancel{0} & X & X & X \end{bmatrix} \end{array}$$

Let R^{-1}_{12} be the first block and R^{-1}_{456} be the second block

$$\therefore R^{-1} = \begin{bmatrix} R^{-1}_{12} & 0 \\ 0 & R^{-1}_{456} \end{bmatrix}$$

With the current matrix form we can use the results from part a) and show that $f_{X_{1,2,4,5,6}}(X_1, X_2, X_4, X_5, X_6 | X_3) = f_{X_1,X_2}(X_1, X_2 | X_3) \cdot f_{X_4,X_5,X_6}(X_4, X_5, X_6 | X_3)$

Thus $(X_1, X_2 | X_3)$ is independent of $(X_4, X_5, X_6 | X_3)$. ✓