Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4267 Linear statistical models
Recommended exercises 1 – solutions

Problem 1 Simple matrix calculations

R code follows below.

c) $$\begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} = 9x^2 - 4xy^2 + 6y^2 = 9\Big(x^2 - \frac{4}{9}xy + \frac{4}{81}y^2\Big) - \frac{4}{9}y^2 + 6y^2 = 9\Big(x - \frac{2}{9}y\Big)^2 + \frac{50}{9}y^2,$$

which is nonnegative (sum of two squares) and zero only for $x = y = 0$. This shows that A is positive definite.

d) Find eigenvalues by solving the equation $\det(\lambda I - A)$ for λ. They are 5 and 10. Then find eigenvectors by finding nonzero solutions $\boldsymbol{v}$ of $(\lambda I - A)\boldsymbol{v}$ when $\lambda = 5$ and when $\lambda = 10$. You may find $(1 \;\; 2)^{\mathrm{T}}$ and $(2 \;\; -1)^{\mathrm{T}}$ for 5 and 10, respectively.

e) The eigenvectors found above are orthogonal since they belong to distinct eigenvalues. Consider the matrix P having columns that are normalized versions of the eigenvectors, and the diagonal matrix Λ having eigenvalues on the diagonal (in the same order as the eigenvectors),

$$P = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}.$$

Then we know from linear algebra that P is orthogonal ($PP^{\mathrm{T}} = P^{\mathrm{T}}P = I$) and that $A = P\Lambda P^{\mathrm{T}}$.

f) To find A^{-1}, you may perform elementary row operations on the block matrix $(A \;\; I)$ so you get a matrix of the form $(I \;\; B)$. Then $B = A^{-1}$. Alternatively, you can use the formula $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)^{-1} = \frac{1}{ad-bc}\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$ for 2×2-matrices.

g) That $\boldsymbol{v} \neq \boldsymbol{0}$ is an eigenvector of A corresponding to an eigenvalue λ, means that $A\boldsymbol{v} = \lambda\boldsymbol{v}$. This is equivalent to $A^{-1}A\boldsymbol{v} = \lambda A^{-1}\boldsymbol{v}$, that is, $A^{-1}\boldsymbol{v} = \frac{1}{\lambda}\boldsymbol{v}$, which means that $\boldsymbol{v}$ is an eigenvector of A^{-1} corresponding to the eigenvalue $1/\lambda$ of A^{-1}. So $\boldsymbol{v}$ is an eigenvector of A corresponding to λ if and only if $\boldsymbol{v}$ is an eigenvector of A^{-1} corresponding to $1/\lambda$.

In our case, the eigenvectors of A^{-1} are the same as for A, but corresponding to eigenvalues $1/5$ and $1/10$.

h) A can be a covariance matrix because it is symmetric and positive definite. We already know that covariance matrices are symmetric and positive semidefinite. We will see later in the course that for all positive definite symmetric matrices, there exists a multivariate normal distribution having that covariance matrix.

i) Remember that the correlation coefficient of two random variables X and Y is $\text{Cov}(X,Y)/\sqrt{(\text{Var}\,X)(\text{Var}\,Y)}$.

j) Remember that $E(A\boldsymbol{X}) = AE\boldsymbol{X}$ and that $\text{Cov}(A\boldsymbol{X}) = A(\text{Cov}\,\boldsymbol{X})A^{\text{T}}$. For the last case, remember that the ij entry of $\text{Cov}\,Y$ is in general the covariance of the ith and jth component of $\boldsymbol{Y}$.

R code:

```
## a) construct A

A <- matrix(c(9,-2,-2, 6), ncol = 2)
A

## b) symmetric?

t(A)
t(A) == A
# yes t(A)=A

## c) positive definite

# t(x) %*% A %*% x > 0 for all x
# just showing how this is calculated
x <- matrix(c(1, 2), ncol = 1)
t(x) %*% A %*% x

## d)

ev <- eigen(A)
names(ev)
ev$values
# yes, positive eigenvalues

# normalized eigenvectors?
```

```
ev$vectors

# first eigenvector, length
sum(ev$vectors[, 1] ^ 2)
# or
t(ev$vectors[, 1]) %*% ev$vectors[, 1]
# second
t(ev$vectors[, 2]) %*% ev$vectors[, 2]

## e) orthogonal diagonalization

P <- ev$vectors
lambda <- diag(ev$values)

P %*% lambda %*% t(P)

## f) inverse

Ainv <- solve(A)
Ainv
# or using the diagonalization (also see g)
lambdainv <- diag(1 / ev$values)
P %*% lambdainv %*% t(P)

## g)

eigen(Ainv)$values
diag(lambdainv) # the same, but different order (by coincidence)
eigen(Ainv)$vectors
eigen(A)$vectors # the same, but different order and sign

## h) since A is symmetric positive definite, it can be a covariance matrix

## i) correlation matrix

varvec <- diag(A) # variances
invsdmat <- diag(1 / sqrt(varvec))

# divide ij entry by sd of ith and by sd of jth component of vector
corrmat <- invsdmat %*% A %*% invsdmat
corrmat
```

```
# builtin
cov2cor(A)

## j)

# X has mean mu and covariance matrix A
mu <- matrix(c(3, 1), ncol = 1)
B <- matrix(c(1, 1, 1, 2), ncol = 2)
d <- matrix(c(1, 2), nrow = 1)

# E and Cov for s=BX
# mean of s
B %*% mu
# cov(s) is B A B^T
B %*% A %*% t(B)

# E and Cov for d X
# mean is
d %*% mu
# cov(t) is
d %*% A %*% t(d)

# E and Cov for v rbind X and 3X
# mean of 3X is 3mu
# cov of 3X is 9 covX
# mean
rbind(mu, 3 * mu) # concatenation
# cov v is a matrix with four blocks
# block1 is cov of X
block1 <- A
# block 2 is cov of X and 3X=3 covX
block2 <- 3 * A
# block 3 is block 2 transposed
block3 <- t(block2)
# block4 is Cov(3X)=9A
block4 <- 9 * A

covv <- cbind(rbind(block1, block2), rbind(block3, block4))
covv
```

Problem 2 Mean and covariance of linear combinations

Let

$$A = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \boldsymbol{X}.$$

We use the formulas $E(A\boldsymbol{X}) = AE\boldsymbol{X}$ and $\operatorname{Cov}(A\boldsymbol{X}) = A(\operatorname{Cov}\boldsymbol{X})A^{\mathrm{T}}$:

$$E(A\boldsymbol{X}) = AE\boldsymbol{X} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

$$\operatorname{Cov}(A\boldsymbol{X}) = A(\operatorname{Cov}\boldsymbol{X})A^{\mathrm{T}} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} I \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Note that A is *idempotent*, that is, $A^2 = A$.

Problem 3 Covariance formula

We use the formulas $E(\boldsymbol{X}+\boldsymbol{Y}) = E\boldsymbol{X} + E\boldsymbol{Y}$, $E(A\boldsymbol{X}B) = A(E\boldsymbol{X})B$ and $E(\boldsymbol{X}^{\mathrm{T}}) = (E\boldsymbol{X})^{\mathrm{T}}$. Remember that $E\boldsymbol{V}$ and $E\boldsymbol{W}$ are just constant vectors.

$$\begin{aligned}
\operatorname{Cov}(\boldsymbol{V},\boldsymbol{W}) &= E\Big((\boldsymbol{V} - E\boldsymbol{V})(\boldsymbol{W} - E\boldsymbol{W})^{\mathrm{T}}\Big) \\
&= E\Big(\boldsymbol{V}\boldsymbol{W}^{\mathrm{T}} - (E\boldsymbol{V})\boldsymbol{W}^{\mathrm{T}} - \boldsymbol{V}(E\boldsymbol{W})^{\mathrm{T}} + (E\boldsymbol{V})(E\boldsymbol{W})^{\mathrm{T}}\Big) \\
&= E(\boldsymbol{V}\boldsymbol{W}^{\mathrm{T}}) - (E\boldsymbol{V})E(\boldsymbol{W}^{\mathrm{T}}) - E(\boldsymbol{V})(E\boldsymbol{W})^{\mathrm{T}} + (E\boldsymbol{V})(E\boldsymbol{W})^{\mathrm{T}} \\
&= E(\boldsymbol{V}\boldsymbol{W}^{\mathrm{T}}) - (E\boldsymbol{V})E(\boldsymbol{W})^{\mathrm{T}}.
\end{aligned}$$

Problem 4 The square root matrix and the Mahalanobis transform

a) First, Σ is symmetric, since its ij entry, $\operatorname{Cov}(X_i, X_j)$, is equal to its ji entry, $\operatorname{Cov}(X_j, X_i)$, where X_i denotes the ith component of $\boldsymbol{X}$. Next, let $\boldsymbol{z}$ be a vector of the same length as $\boldsymbol{X}$. Then (by the formula $\operatorname{Cov}(A\boldsymbol{X}) = A(\operatorname{Cov}\boldsymbol{X})A^{\mathrm{T}}$), $\boldsymbol{z}^{\mathrm{T}}\Sigma\boldsymbol{z} = \operatorname{Cov}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{X}) = \operatorname{Var}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{X}) \geq 0$ (the covariance matrix $\operatorname{Cov}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{X})$ is 1×1 and contains only $\operatorname{Var}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{X})$, and variances are always nonnegative).

b) Since Σ is positive definite, $\boldsymbol{z}^{\mathrm{T}}\Sigma\boldsymbol{z} > 0$ for all vectors $\boldsymbol{z} \neq \boldsymbol{0}$. Choose $\boldsymbol{z} = P\boldsymbol{e}_j$, where $\boldsymbol{e}_j$ is the vector having 1 as its jth component and all other components 0. Then $\boldsymbol{z} \neq \boldsymbol{0}$, since P is nonsingular, and $0 < \boldsymbol{z}^{\mathrm{T}}\Sigma\boldsymbol{z} = \boldsymbol{e}_j^{\mathrm{T}}P^{\mathrm{T}}\Sigma P\boldsymbol{e}_j = \boldsymbol{e}_j^{\mathrm{T}}\Lambda\boldsymbol{e}_j = \lambda_j$, the jth diagonal entry of Λ. Since the eigenvalues of Σ are the diagonal entries of Λ, we have shown that all eigenvalues are positive.

(Note that the opposite is also true: If all eigenvalues of a symmetric matrix are positive, then the matrix is positive definite: $\boldsymbol{z}^{\mathrm{T}}\Sigma\boldsymbol{z} = \boldsymbol{z}^{\mathrm{T}}P\Lambda P^{\mathrm{T}}\boldsymbol{z} = \sum_{j=1}^{p}\lambda_j b_j^2$, where $P^{\mathrm{T}}\boldsymbol{z} = (b_1 \ \cdots \ b_p)^{\mathrm{T}}$. The sum at the right is zero only if all $b_j = 0$, that is, $P^{\mathrm{T}}\boldsymbol{z} = \boldsymbol{0}$, which, since P is nonsingular, implies $\boldsymbol{z} = \boldsymbol{0}$.)

Σ is invertible, since none of the eigenvectors are zero (they are all positive).

Assume that $\boldsymbol{p}$ is an eigenvector of Σ^{-1}, corresponding to the eigenvalue λ, that is, $\Sigma^{-1}\boldsymbol{p} = \lambda\boldsymbol{p}$. Then $\boldsymbol{p} = \Sigma\Sigma^{-1}\boldsymbol{p} = \lambda\Sigma\boldsymbol{p}$, $\ \Sigma\boldsymbol{p} = \frac{1}{\lambda}\Sigma$, showing that $\boldsymbol{p}$ is an eigenvector of Σ corresponding to the eigenvalue $1/\lambda$. Alternatively, $\Sigma = P\Lambda P^{\mathrm{T}}$ is equivalent to $\Sigma^{-1} = P\Lambda^{-1}P^{\mathrm{T}}$. In both diagonalizations, eigenvectors are the columns of P, and the corresponding eigenvalues the diagonal of Λ or Λ^{-1}. So the eigenvalues of Σ^{-1} are reciprocals of those of Σ and vice versa. An eigenvector of one of the matrices corresponding to λ is an eigenvector of the other corresponding to $1/\lambda$.

c) (Note that $\Lambda^{-1/2}$ is a diagonal matrix having the reciprocals of the diagonal entries of $\Lambda^{1/2}$ on the diagonal.)

$$(\Sigma^{1/2})^{\mathrm{T}} = (P\Lambda^{1/2}P^{\mathrm{T}})^{\mathrm{T}} = (P^{\mathrm{T}})^{\mathrm{T}}(\Lambda^{1/2})^{\mathrm{T}}P^{\mathrm{T}} = P\Lambda^{1/2}P^{\mathrm{T}} = \Sigma^{1/2},$$

showing that $\Sigma^{1/2}$ is symmetric. Below we show that $\Sigma^{-1/2}$ is the inverse of $\Sigma^{1/2}$, and in general, the inverse of a symmetric matrix is symmetric. So also $\Sigma^{-1/2}$ is symmetric.

Remember that $P^{\mathrm{T}}P = I$, since P is orthogonal.

$$\begin{aligned}
\Sigma^{1/2}\Sigma^{1/2} &= P\Lambda^{1/2}P^{\mathrm{T}}P\Lambda^{1/2}P^{\mathrm{T}} = P\Lambda^{1/2}\Lambda^{1/2}P^{\mathrm{T}} = P\Lambda P^{\mathrm{T}} = \Sigma,\\
\Sigma^{-1/2}\Sigma^{-1/2} &= P\Lambda^{-1/2}P^{\mathrm{T}}P\Lambda^{-1/2}P^{\mathrm{T}} = P\Lambda^{-1/2}\Lambda^{-1/2}P^{\mathrm{T}} = P\Lambda^{-1}P^{\mathrm{T}} = \Sigma^{-1}, \text{ and}\\
\Sigma^{1/2}\Sigma^{-1/2} &= P\Lambda^{1/2}P^{\mathrm{T}}P\Lambda^{-1/2}P^{\mathrm{T}} = P\Lambda^{1/2}\Lambda^{-1/2}P^{\mathrm{T}} = PP^{\mathrm{T}} = I.
\end{aligned}$$

d) We use the formulas $E(A\boldsymbol{X} + \boldsymbol{b}) = AE\boldsymbol{X} + \boldsymbol{b}$ and $\mathrm{Cov}(A\boldsymbol{X} + \boldsymbol{b}) = A(\mathrm{Cov}\,\boldsymbol{X})A^{\mathrm{T}}$:

$$\begin{aligned}
E\boldsymbol{Y} &= E\Big(\Sigma^{-1/2}(\boldsymbol{X} - \mu)\Big) = \Sigma^{-1/2}E(\boldsymbol{X} - \mu) = \Sigma^{-1/2}(\boldsymbol{E}X - \mu) = \Sigma^{-1/2}(\boldsymbol{\mu} - \boldsymbol{\mu}) = \boldsymbol{0},\\
\mathrm{Cov}\,\boldsymbol{Y} &= \mathrm{Cov}\Big(\Sigma^{-1/2}(\boldsymbol{X} - \mu)\Big) = \Sigma^{-1/2}\,\mathrm{Cov}(\boldsymbol{X} - \mu)(\Sigma^{-1/2})^{\mathrm{T}}\\
&= \Sigma^{-1/2}\,\mathrm{Cov}(\boldsymbol{X})(\Sigma^{-1/2})^{\mathrm{T}} = \Sigma^{-1/2}\Sigma\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{1/2}\Sigma^{1/2}\Sigma^{-1/2} = I
\end{aligned}$$

($(\Sigma^{-1/2})^{\mathrm{T}} = \Sigma^{-1/2}$ since the inverse, $\Sigma^{-1/2}$, of the symmetric matrix $\Sigma^{1/2}$ is symmetric.)