Ex. 1

1. • $u_1^T u_2 = u_1^T \left(a_2 - \frac{u_1^T a_2}{u_1^T u_1} \right) u_1 = u_1^T a_2 - \frac{u_1^T a_2}{u_1^T u_1} \cdot u_1^T u_1$

$$= u_1^T a_2 - u_1^T a_2 = 0.$$

• Let $v = \alpha_1 a_1 + \alpha_2 a_2$

As $a_1 = u_1$

$$a_2 = u_2 + \left(\frac{u_1^T a_2}{u_1^T u_1}\right) u_1$$

$$v = \alpha_1 u_1 + \alpha_2 \left(u_2 + \left(\frac{u_1^T a_2}{u_1^T u_1}\right) u_1 \right)$$

$$= \left(\alpha_1 + \alpha_2 \frac{u_1^T a_2}{u_1^T u_1}\right) u_1 + \alpha_2 u_2$$

$$= \tilde{\alpha}_1 u_1 + \tilde{\alpha}_2 u_2 .$$

2. • given $(a_1, \dots, a_k)$, assume that $(u_1, \dots, u_{n-1})$ $(n-1 < k)$ are orthogonal vectors.

Set $u_n = a_n - \frac{u_1^T a_n}{u_1^T u_1} u_1 - \dots - \frac{u_{n-1}^T a_n}{u_{n-1}^T u_{n-1}} u_{n-1}$ (1)

Then $u_i^T u_n = u_i^T a_n - \frac{u_1^T a_n}{u_1^T u_1} u_i^T u_1 - \dots - \frac{u_{n-1}^T a_n}{u_{n-1}^T u_{n-1}} u_i^T u_{n-1}$

for $i = 1 \dots n-1$.

By assumption : $u_i^T u_j = 0$ if $i \neq j$

Hence : $u_i^T u_n = u_i^T a_n - 0 \dots - \frac{u_i^T a_n}{u_i^T u_i} \cdot u_i^T u_i - 0 \dots$

$$= u_i^T a_n - u_i^T a_n = 0 .$$

Thus u_n is orthogonal to every vector in $(u_1, \dots, u_{n-1})$.
By induction, $(u_1, \dots, u_k)$ are orthogonal to each other.

(Ex 1)

- Assume that you can write any linear combination of $(a_1, \dots, a_{n-1})$ as a linear combination of $(u_1, \dots, u_{n-1})$.

If $v = \underbrace{\alpha_1 a_1 + \dots + \alpha_{n-1} a_{n-1}}_{\substack{\text{linear combination of} \\ (a_1, \dots, a_{n-1})}} + \underbrace{\alpha_n a_n}_{\text{new term!}}$

Use $a_n = u_n + \frac{u_1^T a_n}{u_1^T u_1} u_1 + \dots + \frac{u_{n-1}^T a_n}{u_{n-1}^T u_{n-1}} u_{n-1}$ from Equation (1) and the assumption that

$\alpha_1 a_1 + \dots + \alpha_{n-1} a_{n-1}$ can be written in terms of $(u_1 \dots, u_{n-1})$

as

$\tilde{\alpha}_1 u_1 + \dots + \tilde{\alpha}_{n-1} u_{n-1}$,

to write

$$v = \tilde{\alpha}_1 u_1 + \dots + \tilde{\alpha}_{n-1} u_{n-1} + \alpha_n u_n + \alpha_n \frac{u_1^T a_n}{u_1^T u_1} \cdot u_1 + \dots + \alpha_n \frac{u_{n-1}^T a_n}{u_{n-1}^T u_{n-1}} u_{n-1}$$

$$= \left(\tilde{\alpha}_1 + \alpha_n \frac{u_1^T a_n}{u_1^T u_1}\right) u_1 + \dots + \left(\tilde{\alpha}_{n-1} + \alpha_n \frac{u_{n-1}^T a_n}{u_{n-1}^T u_{n-1}}\right) u_{n-1} + \alpha_n u_n$$

Hence, if the assumption holds, we can also write any linear combination of $(a_1, \dots, a_n)$ as linear combination of $(u_1, \dots, u_n)$. From 1., we know that the assumption holds for $n=2$, thus (by induction) it will hold for $n=k$. (or any n).

Remark: From Eq. (1), we see that we must also assume that a_n cannot be written in terms of the other vectors $(u_1 \dots, u_{n-1})$, or the vectors $(a_1, \dots, a_{n-1})$.
Else $u_n = 0$.
Vectors $(a_1, \dots, a_n)$ which have this property are called linearly independent. The procedure of this exercise works thus only for linearly independent vectors $(a_1 \dots a_k)$.

Ex. 2

1. Let u_1 and $\vec{u}_2$ such that $A\vec{u}_1 = \lambda \vec{u}_1$

$$A\vec{u}_2 = \lambda \vec{u}_2$$

For $u = \alpha_1 \vec{u}_1 + \alpha_2 \vec{u}_2$

$$\begin{aligned} Au &= \alpha_1 A\vec{u}_1 + \alpha_2 A\vec{u}_2 \\ &= \alpha_1 \lambda \vec{u}_1 + \alpha_2 \lambda \vec{u}_2 \\ &= \lambda \left(\alpha_1 \vec{u}_1 + \alpha_2 \vec{u}_2 \right) \\ &= \lambda u . \end{aligned}$$

2. For $U = \left(\vec{u}_1 \; \dots \; \vec{u}_n \right)$

$$A \cdot U = \left(A\vec{u}_1 \; \dots \; A\vec{u}_n \right) = \left(\lambda_1 \vec{u}_1 \; \dots \; \lambda_n \vec{u}_n \right)$$

$$= U \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = U \Lambda$$

3. (a) From $AU = U\Lambda$, $A = U \Lambda U^{-1} = U \Lambda V^T$ (with $V^T = U^{-1}$)

(b) $\cdot$ $A_{ij} = u^{[i]} \Lambda v^{(j)}$ where $u^{[i]}$ is the i-th row of U

$v^{(j)}$ is the j-th column of V^T

$v^{[j]}$ is the j-th row of V

$$= u^{[i]} \Lambda v^{[j]T}$$

$$= u^{[i]} \begin{pmatrix} \lambda_1 V_{j1} \\ \vdots \\ \lambda_n V_{jn} \end{pmatrix}$$

$$= \sum_k \lambda_k V_{jk} U_{ik}$$

row vector $[0 \dots \overset{(i)}{1} \dots 0]$

column vector $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} j$

$\cdot$ On the other hand, $A_{ij} = \sum_n \lambda_n e^{[i]} \vec{u}_n \vec{v}_n^T e^{(j)}$

$$= \sum_n \lambda_n U_{in} V_{jn}$$

Hence $A = U \Lambda V^T = \sum_n \lambda_n \vec{u}_n \vec{v}_n^T .$

(c) From $A = U \Lambda U^{-1}$

$$A^{-1} = U \Lambda^{-1} U^{-1} = U \Lambda^{-1} V^{T}$$

(d) The same calculation as in (b), but for Λ^{-1}, shows that

$$A^{-1} = \sum_n \frac{1}{\lambda_n} \vec{U}_n \vec{V}_n^{T} .$$

Ex. 3

1. • Without using $Tr(AB) = Tr(BA)$.

From Ex. 3 $\quad A_{ij} = \sum_n \lambda_n U_{in} V_{jn}$ so that

$$A_{ii} = \sum_n \lambda_n U_{in} V_{in}$$

$$\sum_i A_{ii} = \sum_n \lambda_n \sum_i U_{in} V_{in} = \sum_n \lambda_n \sum_i (U^T)_{ni} (V^*)_{in}$$

$$= \sum_n \lambda_n \underbrace{\sum_i (U^T)_{ni} (U^{-T})_{in}}_{(U^T \cdot U^{-T})_{nn} = 1} \qquad (\text{using } V^T = U^{-1})$$

$$= \sum_n \lambda_n .$$

• With $Tr(AB) = Tr(BA)$:

$$Tr(A) = Tr(U \Lambda U^{-1}) = Tr(U^{-1} U \Lambda) = Tr(\Lambda) = \sum_n \lambda_n .$$

Why is $Tr(AB) = Tr(BA)$? :

$$Tr(AB) = \sum_i (AB)_{ii} = \sum_i \sum_n A_{in} B_{ni} = \sum_n \sum_i B_{ni} A_{in} = \sum_n (BA)_{nn}$$

$$= Tr(BA) .$$

$\det A \equiv |A|$ $\qquad$ properties of determinant

2. $$\det A = \det(U \Lambda U^{-1}) = |U|\,|\Lambda|\,|U^{-1}| = |U| \cdot 1/|U| \cdot |\Lambda|$$

$$= |\Lambda| = \prod_n \lambda_n . \qquad \text{because}$$

the determinant of $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ is the product of the diagonal elements.

Ex. 4

1. • $Au_1 = \lambda_1 u_1$
$Au_2 = \lambda_2 u_2$

• $u_2^T A u_1 = \lambda_1 u_2^T u_1$
$u_1^T A u_2 = \lambda_2 u_1^T u_2$ (1)

• Taking the transpose of (1) :
$u_2^T A^T u_1 = \lambda_2 u_2^T u_1$

As $A^T = A$, we have $u_2^T A u_1 = \lambda_2 u_2^T u_1$ and
$u_2^T A u_1 = \lambda_1 u_2^T u_1$

• Hence : $0 = u_2^T u_1 . (\lambda_2 - \lambda_1)$.
By assumption $\lambda_2 \neq \lambda_1$ so that $u_2^T u_1 = 0$.

2. • We have just seen that eigenvectors for eigenvalues that are distinct are orthogonal to each other. Normalization makes them orthonormal.

• For the case of equal eigenvalues, ~~the~~ we obtain a set of eigenvectors $(a_1 .. a_k)$ which are not orthogonal*. We know, however, that any linear combination of the $(a_1 ... a_k)$ is also an eigenvector. We can thus use Ex. 1 to make them orthogonal.

(* but linearly independent)

3. • Assume $v^T A v > 0 \quad \forall v \neq 0$

Take for v the eigenvector u_k,

Then $\quad u_k^T A u_k = u_k^T \lambda_k u_k = \lambda_k (u_k^T u_k) > 0$

As $(u_k^T u_k) = \|u_k\|^2 > 0$, we obtain $\lambda_k > 0$

• Assume $\lambda_i > 0$ and write v in terms of the orthogonal eigenvectors u_i as

$v = \alpha_1 u_1 + \dots \alpha_M u_M$

• $$\begin{aligned}
v^T A v &= (\alpha_1 u_1 + \dots + \alpha_M u_M)^T A (\alpha_1 u_1 + \dots + \alpha_M u_M) \\
&= (\alpha_1 u_1 + \dots + \alpha_M u_M)^T (\alpha_1 \lambda_1 u_1 + \dots + \alpha_M \lambda_M u_M) \\
&= \sum_{ij} \alpha_i u_i^T \alpha_j \lambda_j u_j \\
&= \sum_i \alpha_i \cdot \alpha_i \cdot \lambda_i \cdot u_i^T u_i \qquad \text{because } u_i^T u_j = 0 \text{ if } i \neq j \\
&= \sum_i \underbrace{(\alpha_i)^2}_{>0} \underbrace{\|u_i\|^2}_{>0} \underbrace{\lambda_i}_{>0} > 0 .
\end{aligned}$$

• Because $\lambda_i > 0$, we can use formula (10) of Ex. 3 to conclude that A^{-1} exists,

Note: For $(u_1 .. u_M)$ orthonormal eigenvectors,
$U^{-1} = U^T$ so that for any ~~sym~~ positive def. (symmetric)
matrix A: $\quad A^{-1} = \sum_i \frac{1}{\lambda_i} u_i \underline{\underline{u_i^T}}$

$$A = \sum_i \lambda_i u_i \underline{\underline{u_i^T}}$$

Ex.5

1. $L(\mu,\sigma) = \prod_{i=1}^{N} f(X_i;\mu,\sigma)$

$$= \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(X_i-\mu)^2}{\sigma^2} \right] \right)$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{N} \frac{(X_i-\mu)^2}{\sigma^2} \right]$$

2. $\sum_{i=1}^{N} (x_i-\mu)^2 = \sum_{i=1}^{N} \left\{ (X_i - \bar{X} + \bar{X} - \mu)^2 \right.$

$$= \sum_{i=1}^{N} \left\{ (X_i-\bar{X})^2 + 2(X_i-\bar{X})(\bar{X}-\mu) + (\bar{X}-\mu)^2 \right\}$$

$$= \sum_{i=1}^{N} (x_i-\bar{X})^2 + 2(\bar{X}-\mu) \underbrace{\sum_{i=1}^{N} (x_i-\bar{X})}_{0} + N(\bar{X}-\mu)^2$$

$$= \sum_{i=1}^{N} (x_i-\bar{X})^2 + N(\bar{X}-\mu)^2$$

$$= N S^2 + N(\bar{X}-\mu)^2$$

$$\ell(\mu,\sigma) = \log L(\mu,\sigma) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{N}{2} S^2 \cdot \frac{1}{\sigma^2} - \frac{N}{2}(\bar{X}-\mu)^2 \cdot \frac{1}{\sigma^2}$$

3 $\frac{\partial \ell}{\partial \sigma} = -\frac{N}{\sigma} + \frac{NS^2}{\sigma^3} + N\frac{(\bar{X}-\mu)^2}{\sigma^3} \stackrel{!}{=} 0$

$$\frac{\partial \ell}{\partial \mu} = +N(\bar{X}-\mu)\frac{1}{\sigma^2} \stackrel{!}{=} 0 \Rightarrow \underline{\underline{\mu = \bar{X}}}$$

And $\frac{\partial \ell}{\partial \sigma} = -\frac{N}{\sigma} + \frac{NS^2}{\sigma^3} \stackrel{!}{=} 0 \Rightarrow \underline{\underline{\sigma^2 = S^2}}$

The log is a monotonically increasing function. Hence the max arguments of $L(\mu,\sigma)$ and $\ell(\mu,\sigma)$ are the same.