Homework 1

Solutions

1. $(6 + 7 + 7 = 20$ points)

   (a) Let $f_A(a)$ represent the PDF of the time Alice spends to complete her problem set. Let $f_B(b)$ represent the PDF of the time Bob spends to complete his problem set. Since Alice and Bob work independently on their problem sets, we know the joint PDF is

$$\begin{aligned} f_{A,B}(a,b) &= f_A(a)f_B(b) \\ &= \left(\frac{1}{4}e^{-\frac{a}{4}}\right)\left(\frac{1}{5}e^{-\frac{b}{6}}\right) \\ &= \frac{1}{24}e^{-\frac{6a+4b}{24}} \end{aligned}$$

   Thus, the probability that Alice finishes her homework before Bob, i.e., the probability of the event $\{A < B\}$, is given by

$$\begin{aligned} P(A < B) &= \int_0^\infty \int_0^b f_{A,B}(a,b)dadb \\ &= \frac{1}{24}\int_0^\infty \int_0^b e^{-\frac{6a+4b}{24}} dadb \\ &= \frac{3}{5}. \end{aligned}$$

   (b) The probability that Alice finishes the problem set before Bob given that Alice requires more than 4 hours is given by

$$\begin{aligned} P(A < B|A > 4) &= \frac{P(4 < A < B)}{P(A > 4)} \\ &= \frac{\int_4^\infty \int_4^b f_{A,B}(a,b)dadb}{\int_4^\infty f_A(a)da} \\ &= \frac{3}{5}e^{-\frac{2}{3}}. \end{aligned}$$

   (c) The desired probability is $P(|A - B| > 1)$, which can be calculated via

$$P(|A - B| > 1) = P(A - B > 1 \cup A - B < -1) = P(A > B + 1) + P(A < B - 1).$$

   And, we find

$$\begin{aligned} P(A > B + 1) &= \int_0^\infty \int_{b+1}^\infty f_{A,B}(a,b)dadb \\ &= \frac{2}{5}e^{-\frac{1}{4}}, \\ P(A < B - 1) &= \int_1^\infty \int_0^{b-1} f_{A,B}(a,b)dadb \\ &= \frac{3}{5}e^{-\frac{1}{6}}. \end{aligned}$$

   Thus,

$$P(|A - B| > 1) = \frac{2}{5}e^{-\frac{1}{4}} + \frac{3}{5}e^{-\frac{1}{6}}.$$

2. $(10 + 10 = 20$ points)

(a) It is clear that Y is also geometric, therefore having a PMF

$$P_Y(y) = (1-p)^{y-1}p.$$

And, for the random variable $X+Y$, $X+Y=k$ means the kth flip comes up a head and there has exactly one head in the first $k-1$ flips. Thus, we have the PMF of $X+Y$

$$P_{X+Y}(k) = P(X+Y=k) = \binom{k-1}{1}(1-p)^{k-2}p^2 = (k-1)(1-p)^{k-2}p^2.$$

(b)

$$P(X=k|X+Y=n) = \frac{P(X=k, X+Y=n)}{P(X+Y=n)} = \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)},$$

where the last equality uses the fact that X and Y are independent. It follows, from the result of part (a), that

$$P(X=k|X+Y=n) = \frac{1}{n-1}.$$

With this, we can find $E[X|X+Y=n]$ by the definition of the conditional expectation as

$$E[X|X+Y=n] = \sum_{x=1}^{n-1} xP(X=k|X+Y=n) = \frac{n}{2}.$$

Alternatively, we can solve the problem from another perspective. We know that $E[X|X+Y=n] = E[Y|X+Y=n]$ since X and Y are identically distributed. Also, $E[X+Y|X+Y=n] = n$. It follows that $E[X|X+Y=n] = n/2$, regardless of the true distributions of X and Y. This is also an intuitively correct result.

You can see the other two similar examples in our textbook Example 4.2-4 and Example 4.2-5.

3. $(5+10+10=25$ points) Prove the followings.

(a) Assume discrete random variable case.

$$\begin{aligned}
E[X] &= \sum_x x p_X(x) \\
&= \sum_x x \underbrace{\left(\sum_y p_{X,Y}(x,y)\right)}_{=p_X(x)} \\
&= \sum_x x \left(\sum_y p_{X|Y}(x|y)p_Y(y)\right) \\
&= \sum_y p_Y(y) \cdot \underbrace{\left(\sum_x x p_{X|Y}(x|y)\right)}_{=E[X|Y]}.
\end{aligned}$$

The proof for continuous random variable case follows a similar way.

(b) The result of this problem is intuitively obvious. The intuition is, given that $Y=y$, there is nothing random about $h(Y)|_{Y=y}$. Thus, it serves as a deterministic (non-random) value and can be pulled out of the expectation.

But we still need to justify it mathematically. We consider the discrete random variables case here. Continuous random variables case can also be shown in a similar fashion, and is left for you to work out by yourself.

Let X and Y be two discrete random variables. From the rule of expected value, the conditional expectation is

$$\begin{aligned}
E\Big[g(X)\cdot h(Y)\Big|Y=y\Big] &= \sum_{x,y'} g(x)h(y')p_{X,Y|Y}(x,y'|y) \\
&= \sum_{x,y'} g(x)h(y')\frac{P[X=x, Y=y', Y=y]}{P[Y=y]},
\end{aligned}$$

where
$$P[X=x,Y=y',Y=y]=\begin{cases} P[X=x,Y=y] & \text{if } y'=y,\\ 0 & \text{otherwise.}\end{cases}$$
Thus, the conditional expectation is evaluated only when $y'=y$, which gives

$$\begin{aligned}
E\Big[g(X)\cdot h(Y)\Big|Y=y\Big] &= \sum_{x,y'} g(x)h(y')\frac{P[X=x,Y=y',Y=y]}{P[Y=y]}\\
&= \sum_{x} g(x)h(y)\frac{P[X=x,Y=y]}{P[Y=y]}\\
&\quad \text{(We see only the event } \{y'=y\} \text{ yields nonzero } P[X=x,Y=y',Y=y])\\
&= h(y)\sum_{x} g(x)\frac{P[X=x,Y=y]}{P[Y=y]}\\
&= h(y)\left(\sum_{x} g(x)p_{X|Y}(x|y)\right)\\
&= h(y)E[g(X)|Y=y].
\end{aligned}$$

(c) For any function $k(\cdot)$, we have

$$\begin{aligned}
E\big[\big(X-E[X|Y]\big)\cdot k(Y)\big] &= E\big[Xk(Y)-k(Y)E[X|Y]\big]\\
&= E\Big[k(Y)X-E\big[k(Y)X|Y\big]\Big] \quad \text{(from the result of part (b))}\\
&= E\big[k(Y)X\big]-E\big[k(Y)X|Y\big]\big]\\
&= E\big[k(Y)X\big]-E\big[E\big[k(Y)X|Y\big]\big]\\
&= E\big[k(Y)X\big]-E\big[k(Y)X\big] \quad \text{(from the result of part (a))}\\
&= 0.
\end{aligned}$$

4. $(5+10+10=25$ points)

(a) Consider an eigenvalue λ of $\mathbf{A}^H\mathbf{A}$ associated with the eigenvector $\mathbf{v}$. From the definition of eigenvector/eigenvalue, we have
$$\mathbf{A}^H\mathbf{A}\cdot\mathbf{v}=\lambda\cdot\mathbf{v}.$$
Multiplying the above equation by $\mathbf{A}$ on the left yields
$$\mathbf{A}\mathbf{A}^H\mathbf{A}\cdot\mathbf{v}=\lambda\mathbf{A}\cdot\mathbf{v},$$
which shows that λ is also an eigenvalue of $\mathbf{A}\mathbf{A}^H$ associated with the eigenvector $\mathbf{A}\cdot\mathbf{v}$.

(b) The rank of a matrix can be regarded as the dimension of its range space. For $\mathbf{A}\in\mathcal{R}^{m\times n}$, the the dimension theorem tells us that
$$\text{rank}(\mathbf{A})+\dim(\text{N}(\mathbf{A}))=\dim(\mathcal{R}^n),$$
where $\text{N}(\mathbf{A})$ is the null space of the matrix $\mathbf{A}$. Similarly, for $\mathbf{A}^H\mathbf{A}\in\mathcal{R}^{n\times n}$, we know
$$\text{rank}(\mathbf{A}^H\mathbf{A})+\dim(\text{N}(\mathbf{A}^H\mathbf{A}))=\dim(\mathcal{R}^n).$$
It is not difficult to show that $\text{N}(\mathbf{A})=\text{N}(\mathbf{A}^H\mathbf{A})$. Thus, we have
$$\text{rank}(\mathbf{A})=\text{rank}(\mathbf{A}^H\mathbf{A}).$$
(For any $\mathbf{y}\in\text{N}(\mathbf{A})$, we know $\mathbf{A}\mathbf{y}=\mathbf{0}$, implying that $\mathbf{A}^H\mathbf{A}\mathbf{y}=\mathbf{0}$. Thus, $\mathbf{y}\in\text{N}(\mathbf{A}^H\mathbf{A})$. Thus, $\text{N}(\mathbf{A})\subseteq\text{N}(\mathbf{A}^H\mathbf{A})$. On the other hand, we can also show $\text{N}(\mathbf{A}^H\mathbf{A})\subseteq\text{N}(\mathbf{A})$, thereby completing the proof that $\text{N}(\mathbf{A})=\text{N}(\mathbf{A}^H\mathbf{A})$.)

(c) To prove two vector spaces S and T are identical in a rigorous manner, we need to show that S is a subset of T "and" T is also a subset of S. (S $\subseteq$ T and T $\subseteq$ S)
First, we show that $\text{R}(\mathbf{U}_1)=\text{R}(\mathbf{A})$.
I. $\underline{(\text{R}(\mathbf{U}_1)\subseteq\text{R}(\mathbf{A}))}$

Let $\mathbf{y}$ be a vector in the range of $\mathbf{U}_1$. Then, there exists a vector $\mathbf{x} \in \mathcal{R}^r$ such that $\mathbf{y} = \mathbf{U}_1\mathbf{x}$. And, since $\mathbf{U}_1 = \mathbf{A}\mathbf{V}_1\mathbf{\Sigma}_r^{-1}$, it follows that

$$\mathbf{y} = \mathbf{A}\mathbf{V}_1\mathbf{\Sigma}_r^{-1}\mathbf{x},$$

which clearly is in $\mathrm{R}(\mathbf{A})$.
II. $\underline{(\mathrm{R}(\mathbf{A}) \subseteq \mathrm{R}(\mathbf{U}_1))}$
Now let $\mathbf{y} \in \mathrm{R}(\mathbf{A})$. Then, there exists a vector $\mathbf{z} \in \mathcal{R}^n$ such that $\mathbf{y} = \mathbf{A}\mathbf{z}$. Further, $\mathbf{z}$ can be represented by a linear combination of any basis of $\mathcal{R}^n$. By choosing the column vectors of $\mathbf{V}$ as the basis for $\mathcal{R}^n$, we have

$$\mathbf{z} = \mathbf{V}\mathbf{p} = \mathbf{V}_1\mathbf{p}_1 + \mathbf{V}_2\mathbf{p}_2,$$

for a unique $\mathbf{p}$, where $\mathbf{p} = [\mathbf{p}_1^T\mathbf{p}_2^T]^T$. It follows that

$$\begin{aligned}
\mathbf{y} &= \mathbf{A}\mathbf{z} = \mathbf{A}(\mathbf{V}_1\mathbf{p}_1 + \mathbf{V}_2\mathbf{p}_2) \\
&= \mathbf{A}\mathbf{V}_1\mathbf{p}_1 \quad (\text{since } \ \mathbf{A}\mathbf{V}_2 = \mathbf{0}) \\
&= \mathbf{U}_1\mathbf{\Sigma}_r\mathbf{p}_1,
\end{aligned}$$

which is clearly an element in $\mathrm{R}(\mathbf{U}_1)$.
Thus, we conclude that $\mathrm{R}(\mathbf{U}_1) = \mathrm{R}(\mathbf{A})$.

Finally, it is relatively easier to show that $\mathrm{R}(\mathbf{V}_2) = \mathrm{N}(\mathbf{A})$. With $\mathbf{A}\mathbf{V}_2 = \mathbf{0}$, we know the column vectors of $\mathbf{V}_2$ are in $\mathrm{N}(\mathbf{A})$. From the dimension theorem

$$\mathrm{rank}(\mathbf{A}) + \dim(\mathrm{N}(\mathbf{A})) = \dim(\mathcal{R}^{\mathbf{n}}),$$

we know $\dim(\mathrm{N}(\mathbf{A})) = n - r$, which is exactly the number of columns of $\mathbf{V}_2$. This means the linearly independent columns of $\mathbf{V}_2$ form a basis of $\mathrm{N}(\mathbf{A})$. We can conclude that $\mathrm{R}(\mathbf{V}_2) = \mathrm{N}(\mathbf{A})$.

5. (5+5=10 points)

   (a) Singular value decomposition (SVD) tells us that we can decompose $\mathbf{H}$ into $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H$ with unitary matrices $\mathbf{U}$, $\mathbf{V}$ and diagonal matrix $\mathbf{D}$. With this,

$$\begin{aligned}
\tilde{\mathbf{y}} &= \mathbf{U}^H\mathbf{H}\mathbf{V}\tilde{\mathbf{x}} + \mathbf{U}^H\mathbf{n} \\
&= \mathbf{U}^H\mathbf{U}\mathbf{D}\mathbf{V}^H\mathbf{V}\tilde{\mathbf{x}} + \mathbf{U}^H\mathbf{n} \\
&= \mathbf{D}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}.
\end{aligned}$$

   Hence, we design $\mathbf{U}$ and $\mathbf{V}$ as the two unitary matrices in SVD of $\mathbf{H}$. And the diagonal entries of $\mathbf{D}$ will be the singular values of $\mathbf{H}$.
   Therefore, we can find $\mathbf{U}$ to be the matrix of eigenvectors of $\mathbf{H}\mathbf{H}^H$ and $\mathbf{V}$ to be the matrix of eigenvectors of $\mathbf{H}^H\mathbf{H}$. It's easy to find

$$\mathbf{H}\mathbf{H}^H = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}, \quad \text{and} \quad \mathbf{H}^H\mathbf{H} = \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix}.$$

   With some algebraic efforts, we have

$$\mathbf{U} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{V} = \frac{1}{2}\begin{bmatrix} -\sqrt{3} & 1 \\ -1 & -\sqrt{3} \end{bmatrix}, \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} \sqrt{6} & 0 \\ 0 & \sqrt{2} \end{bmatrix}.$$

   (b) First, this problem will not be graded. You automatically get the 5 pts.
   With the answer in (a), we get

$$\begin{aligned}
\tilde{y}_1 &= \sqrt{6} * \tilde{x}_1 + \tilde{n}_1 \\
\tilde{y}_2 &= \sqrt{2} * \tilde{x}_2 + \tilde{n}_2.
\end{aligned}$$

   We should give more power to the one which has a better channel, *i.e.*, larger channel gain(since the $\tilde{n}_1$ and $\tilde{n}_2$ have same attribution.) In this case, giving $\tilde{x}_1$ more power is better. (The answer may be $\tilde{x}_2$ if you do (a) differently and result in $\mathbf{D} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{6} \end{bmatrix}$.)