Midterm

1. xn = αxn−1 + wn

a) joint pdf

Notice that $x_n = \alpha x_{n-1} + w_n, w_n \sim N(0,1)$
Therefore, x_i only depends on x_{i-1} and independ on $x_{i-2}, ..., x_0$.
$x_i|(x_{i-1}\alpha) \sim N(\alpha x_{i-1}, 1)$, where $i = 2, 3, ..., N$
$x_1|\alpha \sim N(\alpha x_0, 1)$

$$f(x_i|x_{i-1}\alpha) = \frac{e^{-\frac{1}{2}(x_i - \alpha x_{i-1})^2}}{\sqrt{2\pi}}$$

$$f(x_1|\alpha) = \frac{e^{-\frac{1}{2}(x_1 - \alpha x_0)^2}}{\sqrt{2\pi}}$$

Therefore, $f(x_1, x_2, ..., x_N|\alpha) = f(x_1|\alpha)f(x_2|x_1\alpha)...f(x_N|x_{N-1}\alpha)$

$$f(x_1, ..., x_N|\alpha) = \prod_{i=1}^{N} \frac{e^{-\frac{1}{2}(x_i - \alpha x_{i-1})^2}}{\sqrt{2\pi}} = \frac{e^{-\frac{1}{2}\sum_{i=1}^{N}(x_i - \alpha x_{i-1})^2}}{\sqrt{(2\pi)^N}}$$

b) $\hat{\alpha}_{MLE}$

$$\hat{\alpha}_{MLE} = \arg\max_{\alpha} \prod_{i=1}^{N} \frac{e^{-\frac{1}{2}(x_i - \alpha x_{i-1})^2}}{\sqrt{2\pi}} = \arg\max_{\alpha} \sum_{i=1}^{N}(ln(e^{-\frac{1}{2}(x_i - \alpha x_{i-1})^2}) - ln(\sqrt{2\pi}))$$

Therefore, $\hat{\alpha}_{MLE} = \arg\max_{\alpha} \sum_{i=1}^{N} ln(e^{-\frac{1}{2}(x_i - \alpha x_{i-1})^2}) = \arg\max_{\alpha} \sum_{i=1}^{N}(-\frac{1}{2}(x_i - \alpha x_{i-1})^2) = \arg\min_{\alpha} \sum_{i=1}^{N}(x_i - \alpha x_{i-1})^2$

Let $\nabla_\alpha \sum_{i=1}^{N}(x_i - \alpha x_{i-1})^2 = 0$
$\sum_{i=1}^{N}(x_{i-1}^2\alpha - x_{i-1}x_i) = 0$
$\hat{\alpha}_{MLE} = \frac{\sum_{i=1}^{N} x_i x_{i-1}}{\sum_{i=1}^{N} x_{i-1}^2}$

2 H0: xn = -sαn + wn, H1: xn = sαn + wn

I am not sure whether all x_n are from the same hypothesis or each x_n is independent and can from either H_0 or H_1.
Therefore, I write 2 different answers.

I. All x_n are from the same hypothesis

a) the optimum decision mechanism

Notice that the key point is to minimize the probability of making an error.
Let $C_{ij} = \begin{cases} 1 & ,i \neq j \\ 0 & ,i = j \end{cases}$, where $i, j = 0, 1$.

Notice that $w \sim N(0, \Sigma)$, where $\Sigma_{ij} = \begin{cases} \sigma^2 & ,i \neq j \\ 0 & ,i = j \end{cases}$, where $i, j = 0, 1, ..., N$

Therefore:
$H_0 : x|s\sigma \sim N(-s\alpha, \Sigma)$
$H_1 : x|s\sigma \sim N(s\alpha, \Sigma)$

$$f_0(x|s\sigma) = \frac{e^{-\frac{1}{2}(x+s\alpha)^T\Sigma^{-1}(x+s\alpha)}}{\sqrt{(2\pi)^N|\Sigma|}} = \frac{e^{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^N(x_i+s\alpha_i)^2}}{\sqrt{(2\pi\sigma^2)^N}}$$

$$f_1(x|s\sigma) = \frac{e^{-\frac{1}{2}(x-s\alpha)^T\Sigma^{-1}(x-s\alpha)}}{\sqrt{(2\pi)^N|\Sigma|}} = \frac{e^{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^N(x_i-s\alpha_i)^2}}{\sqrt{(2\pi\sigma^2)^N}}$$

Consider $C(\delta_0, \delta_1) = P(D_0H_1) + P(D_1H_0) = \int(\delta_0 f_1(x|s\sigma)P(H_1) + \delta_1 f_0(x|s\sigma)P(H_0))dx$

Therefore, in order to minimize $C(\delta_0, \delta_1)$:
If $f_1(x|s\sigma)P(H_1) > f_0(x|s\sigma)P(H_0)$, let $\delta_0 = 0, \delta_1 = 1$, i.e. we decide in favor of H_1.
If $f_1(x|s\sigma)P(H_1) = f_0(x|s\sigma)P(H_0)$, δ_0, δ_1 can be any number ($\delta_0 + \delta_1$ = 1, $0 \leq \delta_0, \delta_1 \leq 1$), i.e. we can decide in favor of either H_0 or H_1.
If $f_1(x|s\sigma)P(H_1) < f_0(x|s\sigma)P(H_0)$, let $\delta_0 = 1, \delta_1 = 0$, i.e. we decide in favor of H_0.

Note that to compare $f_1(x|s\sigma)P(H_1)$ and $f_0(x|s\sigma)P(H_0)$ is to compare $\frac{f_1(x|s\sigma)}{f_0(x|s\sigma)}$ and $\frac{P(H_0)}{P(H_1)}$.

Let $T = \frac{f_1(x|s\sigma)}{f_0(x|s\sigma)} = \frac{e^{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^N(x_i-s\alpha_i)^2}}{e^{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^N(x_i+s\alpha_i)^2}}$

Notice $P(H_0) = P(H_1) = 0.5$:
$\frac{P(H_0)}{P(H_1)} = 1$.

Therefore:
If $T > 1$, we decide in favor of H_1.
If $T = 1$, we can decide in favor of either H_0 or H_1.
If $T < 1$, we decide in favor of H_0.

Note that to compare T and 1 is to compare $\sum_{i=1}^{N}(x_i - s\alpha_i)^2$ and $\sum_{i=1}^{N}(x_i + s\alpha_i)^2$: (Because $lnT = \frac{\sum_{i=1}^{N}(x_i+s\alpha_i)^2 - \sum_{i=1}^{N}(x_i-s\alpha_i)^2}{2\sigma^2}$, $ln1 = 0$. Note that there is a factor $-\frac{1}{2}$.)

If $\sum_{i=1}^{N}(x_i - s\alpha_i)^2 < \sum_{i=1}^{N}(x_i + s\alpha_i)^2$, we decide in favor of H_1.
If $\sum_{i=1}^{N}(x_i - s\alpha_i)^2 = \sum_{i=1}^{N}(x_i + s\alpha_i)^2$, we can decide in favor of either H_0 or H_1.
If $\sum_{i=1}^{N}(x_i - s\alpha_i)^2 > \sum_{i=1}^{N}(x_i + s\alpha_i)^2$, we decide in favor of H_0.

b) equivalent mechanism

$$\sum_{i=1}^{N}(x_i + s\alpha_i)^2 = \sum_{i=1}^{N}(x_i^2 + 2x_i s\alpha_i + s^2\alpha_i^2)$$
$$\sum_{i=1}^{N}(x_i - s\alpha_i)^2 = \sum_{i=1}^{N}(x_i^2 - 2x_i s\alpha_i + s^2\alpha_i^2)$$

Therefore, to compare $\sum_{i=1}^{N}(x_i - s\alpha_i)^2$ and $\sum_{i=1}^{N}(x_i + s\alpha_i)^2$ is to compare $m_1 = -\sum_{i=1}^{N} x_i s\alpha_i$ and $m_0 = \sum_{i=1}^{N} x_i s\alpha_i$.

Notice that $s > 0$
Therefore, let $n_0 = \frac{m_0}{s} = \sum_{i=1}^{N} x_i\alpha_i, n_1 = \frac{m_1}{s} = -\sum_{i=1}^{N} x_i\alpha_i$

If $n_1 < n_0$, i.e. $\sum_{i=1}^{N} x_i\alpha_i > 0$, we decide in favor of H_1.
If $n_1 = n_0$, i.e. $\sum_{i=1}^{N} x_i\alpha_i = 0$, we can decide in favor of either H_0 or H_1.
If $n_1 > n_0$, i.e. $\sum_{i=1}^{N} x_i\alpha_i < 0$, we decide in favor of H_0.

c) optimality properties

1. This mechanism provided the least expectation of times of making an error. Namely, for a given realization of x_n, it may not be the optimal one. However, if we keep it working for a number of realizations, it will be optimal on average. Because we minimized $C(\delta_0, \delta_1) = E(D_1 H_0 + D_0 H_1)$

2. Also, it does not depend on unknown variables.

3. For each realization, this mechanism does not depand on any decisions of previous realizations.

4. Furthermore, it can be computed in linear time and space, which is also optimal because at least we have to store α with $O(n)$ space and process x with $O(n)$time.

II. Each x_n is independent and can from either H_0 or H_1

In this case, we have to decide each x_n belonging to H_0 or H_1.

a) the optimum decision mechanism

Notice that the key point is to minimize the probability of making an error.
Let $C_{ij} = \begin{cases} 1 & ,i \neq j \\ 0 & ,i = j \end{cases}$, where $i, j = 0, 1$.

Notice that $w_i \sim N(0, \sigma^2)$, where $i = 0, 1, ..., N$

Therefore:
$H_0 : x_i | s\sigma \sim N(-s\alpha, \sigma^2)$
$H_1 : x_i | s\sigma \sim N(s\alpha, \sigma^2)$

$$f_0(x_i|s\sigma) = \frac{e^{-\frac{1}{2}\frac{1}{\sigma^2}(x_i+s\alpha_i)^2}}{\sqrt{(2\pi\sigma^2)^N}}$$

$$f_1(x_i|s\sigma) = \frac{e^{-\frac{1}{2}\frac{1}{\sigma^2}(x_i-s\alpha_i)^2}}{\sqrt{(2\pi\sigma^2)^N}}$$

Consider $C(\delta_0, \delta_1) = P(D_0 H_1) + P(D_1 H_0) = \int(\delta_0 f_1(x_i|s\sigma)P(H_1) + \delta_1 f_0(x_i|s\sigma)P(H_0))dx$

Therefore, in order to minimize $C(\delta_0, \delta_1)$, for x_i:
If $f_1(x|s\sigma)P(H_1) > f_0(x|s\sigma)P(H_0)$, let $\delta_0 = 0, \delta_1 = 1$, i.e. we decide in favor of H_1.

If $f_1(x|s\sigma)P(H_1) = f_0(x|s\sigma)P(H_0)$, δ_0, δ_1 can be any number ($\delta_0 + \delta_1$ = 1, $0 \leq \delta_0, \delta_1 \leq 1$), i.e. we can decide in favor of either H_0 or H_1.
If $f_1(x|s\sigma)P(H_1) < f_0(x|s\sigma)P(H_0)$, let $\delta_0 = 1, \delta_1 = 0$, i.e. we decide in favor of H_0.

Note that to compare $f_1(x|s\sigma)P(H_1)$ and $f_0(x|s\sigma)P(H_0)$ is to compare $\frac{f_1(x|s\sigma)}{f_0(x|s\sigma)}$ and $\frac{P(H_0)}{P(H_1)}$.

Let $T = \frac{f_1(x|s\sigma)}{f_0(x|s\sigma)} = \frac{e^{-\frac{1}{2}\frac{1}{\sigma^2}(x_i - s\alpha_i)^2}}{e^{-\frac{1}{2}\frac{1}{\sigma^2}(x_i + s\alpha_i)^2}}$

Notice $P(H_0) = P(H_1) = 0.5$:
$\frac{P(H_0)}{P(H_1)} = 1$.

Therefore, for x_i:
If $T > 1$, we decide in favor of H_1.
If $T = 1$, we can decide in favor of either H_0 or H_1.
If $T < 1$, we decide in favor of H_0.

Note that to compare T and 1 is to compare $(x_i - s\alpha_i)^2$ and $(x_i + s\alpha_i)^2$: (Because $lnT = \frac{(x_i + s\alpha_i)^2 - (x_i - s\alpha_i)^2}{2\sigma^2}, ln1 = 0$. Note that there is a factor $-\frac{1}{2}$.)

If $(x_i - s\alpha_i)^2 < (x_i + s\alpha_i)^2$, we decide in favor of H_1.
If $(x_i - s\alpha_i)^2 = (x_i + s\alpha_i)^2$, we can decide in favor of either H_0 or H_1.
If $(x_i - s\alpha_i)^2 > (x_i + s\alpha_i)^2$, we decide in favor of H_0.

b) equivalent mechanism

$(x_i + s\alpha_i)^2 = (x_i^2 + 2x_i s\alpha_i + s^2\alpha_i^2)$
$(x_i - s\alpha_i)^2 = (x_i^2 - 2x_i s\alpha_i + s^2\alpha_i^2)$

Therefore, to compare $(x_i - s\alpha_i)^2$ and $(x_i + s\alpha_i)^2$ is to compare $m_1 = -x_i s\alpha_i$ and $m_0 = x_i s\alpha_i$.

Notice that $s > 0$
Therefore, let $n_0 = \frac{m_0}{s} = x_i\alpha_i, n_1 = \frac{m_1}{s} = -x_i\alpha_i$

If $n_1 < n_0$, i.e. $x_i\alpha_i > 0$, we decide in favor of H_1.
If $n_1 = n_0$, i.e. $x_i\alpha_i = 0$, we can decide in favor of either H_0 or H_1.
If $n_1 > n_0$, i.e. $x_i\alpha_i < 0$, we decide in favor of H_0.

c) optimality properties

1. This mechanism provided the least expectation of times of making an error. Namely, for a given realization of x_n, it may not be the optimal one. However, if we keep it working for a number of realizations, it will be optimal on average. Because we minimized $C(\delta_0, \delta_1) = E(D_1 H_0 + D_0 H_1)$

2. Also, it does not depend on unknown variables.

3. For each $x_i, i = 1, 2, ..., N$, this mechanism does not depand on any decisions of $x_j, j \neq i$, because we only care about $x_i \alpha_i$.

4. Furthermore, it can be computed in linear time and space, which is also optimal because at least we have to store α with $O(n)$ space and process x with $O(n)$time.

3 H0 : X ~ f0(X), H1 : X ~ f1(X), H2 : X ~ f2(X)

Notice that the key point is to minimize the probability of making an error.

Let $C_{ij} = \begin{cases} 1 & , i \neq j \\ 0 & , i = j \end{cases}$, where $i, j = 0, 1, 2$.

$C(\delta_0, \delta_1, \delta_2) = \int (\delta_0 (f_1(X) P(H_1) + f_2(X) P(H_2)) + \delta_1 (f_0(X) P(H_0) + f_2(X) P(H_2)) + \delta_2 (f_0(X) P(H_0) + f_1(X) P(H_1))) dX$

Notice that $P(H_1) = P(H_2) = P(H_3)$,
Therefore, $\arg \min_{\delta_0, \delta_1, \delta_2} C(\delta_0, \delta_1, \delta_2) = \arg \min_{\delta_0, \delta_1, \delta_2} \int (\delta_0 (f_1(X) + f_2(X)) + \delta_1 (f_0(X) + f_2(X)) + \delta_2 (f_0(X) + f_1(X))) dX$

Therefore, let $m_0 = f_1(X) + f_2(X), m_1 = f_0(X) + f_2(X), m_2 = f_0(X) + f_1(X)$
The optimum decision mechanism can be $\delta_i = 1, \delta_j = 0$, where $i = \arg \min_i m_i, j \neq i$

Notice that $f_0(X) > 0$,
$n_0 = \frac{m_0}{f_0(X)} = L_1 + L_2$
$n_1 = \frac{m_1}{f_0(X)} = 1 + L_2$
$n_2 = \frac{m_2}{f_0(X)} = L_1 + 1$
Therefore, the optimum decision mechanism can be $\delta_i = 1, \delta_j = 0$, where $i = \arg \min_i n_i, j \neq i$

(Note: say $m_0 < m_1, m_0 < m_2$:
$f_1(X) + f_2(X) < f_0(X) + f_2(X), f_1(X) + f_2(X) < f_0(X) + f_1(X)$
$f_1(X) < f_0(X), f_2(X) < f_0(X)$

Actually, $\arg\min_i m_i = \arg\max_i f_i(X)$
But I prefer to use "$\arg\min_i m_i$" because it is much eaiser to notice where L_1, L_2 come from.)

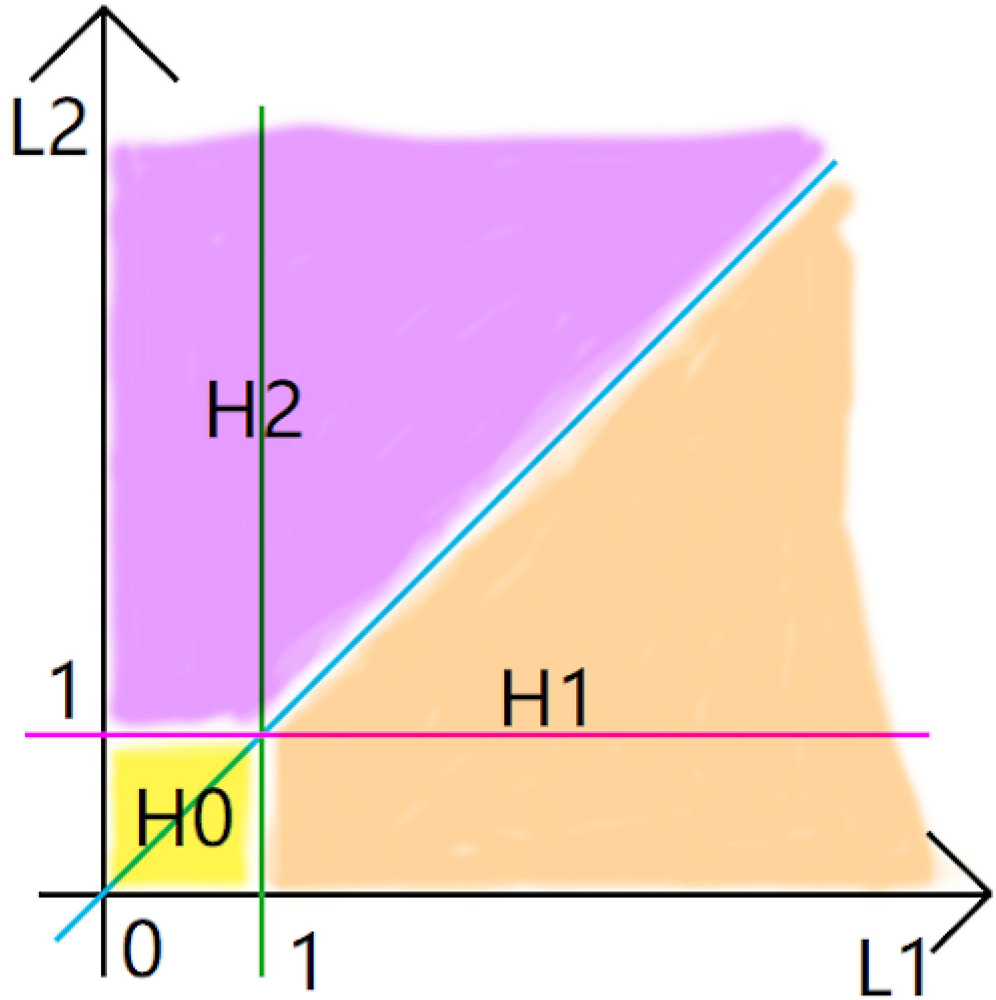
a) regions

Let $n_0 \leq n_1$:
$L_1 \leq 1$

Let $n_0 \leq n_2$:
$L_2 \leq 1$

Let $n_1 \leq n_2$:
$L_2 \leq L_1$

Notice $L_1, L_2 > 0$,
Therefore, the regions for which we decide in favor of each of the three scenarios are shown as below:



b) boundaries

There are 3 boundaries:

1. $L_2 = 1, 0 < L_1 \leq 1$: We can decide in favor of either H_0 or H_2
2. $L_1 = 1, 0 < L_2 \leq 1$: We can decide in favor of either H_0 or H_1
3. $L_1 = L_2 \geq 1$: We can decide in favor of either H_1 or H_2

Notice that when $L_1 = 1, L_2 = 1$, we can decide in favor of anyone of H_1, H_2, H_3.

4 x^ = az + bw

a) orthogonality principle

Notice $L = \{y : y = az + bw, a, b \in R\}$
$\hat{x}_* = \arg\min_{\hat{x} \in L} ||x - \hat{x}||$ satisfies $< x - \hat{x}, y >= 0, \forall y \in L$.

Therefore:
$< x - \hat{x}_*, z >=< x - (az + bw), z >= E((x - (az + bw))z) = 0$
$< x - \hat{x}_*, w >=< x - (az + bw), w >= E((x - (az + bw))w) = 0$

$E(xz) = E(z^2)a + E(wz)b$
$E(xw) = E(wz)a + E(w^2)b$

Therefore:
$a_* = \frac{E(xz)E(w^2) - E(xw)E(wz)}{E(z^2)E(w^2) - (E(wz))^2}$
$b_* = \frac{E(xw)E(z^2) - E(xz)E(wz)}{E(z^2)E(w^2) - (E(wz))^2}$

b) distance and x^*

$||x - \hat{x}_*|| = \sqrt{E((x - \hat{x}_*)^2)} = \sqrt{E(x^2 - 2x\hat{x}_* + \hat{x}_*^2)} = \sqrt{E(x^2) - E(2x\hat{x}_* - \hat{x}_*^2)}$

Notice that $E(2x\hat{x}_* - \hat{x}_*^2) = E(2(x - \hat{x}_*)\hat{x}_* + \hat{x}_*^2) = 2E((x - \hat{x}_*)\hat{x}_*) + E(\hat{x}_*^2)$
Notice that $\hat{x}_* \in L$:
$E((x - \hat{x}_*)\hat{x}_*) = 0$
Therefore, $||x - \hat{x}_*|| = \sqrt{E(x^2) - E(\hat{x}_*^2)} =$

$\sqrt{E(x^2) - (a_*^2 E(z^2) + b_*^2 E(w^2) + 2a_* b_* E(zw))}$, where $a_* = \frac{E(xz)E(w^2) - E(xw)E(wz)}{E(z^2)E(w^2) - (E(wz))^2}$, $b_* = \frac{E(xw)E(z^2) - E(xz)E(wz)}{E(z^2)E(w^2) - (E(wz))^2}$

Notice $a_*^2 E(z^2) + b_*^2 E(w^2) + 2a_* b_* E(zw) = \frac{(E(z^2)E(w^2) - E(wz)^2)((E(xz))^2 E(w^2) + (E(xw))^2 E(z^2) - 2E(xw)E(xz)E(wz))}{E(z^2)E(w^2) - E(wz)^2} = (E(xz))^2 E(w^2) + (E(xw))^2 E(z^2) - 2E(xw)E(xz)E(wz)$

Therefore, $||x - \hat{x}_*|| = \sqrt{E(x^2) - ((E(xz))^2 E(w^2) + (E(xw))^2 E(z^2) - 2E(xw)E(xz)E(wz))}$

$\hat{x}_* = a_* z + b_* w = \frac{E(xz)E(w^2) - E(xw)E(wz)}{E(z^2)E(w^2) - (E(wz))^2} z + \frac{E(xw)E(z^2) - E(xz)E(wz)}{E(z^2)E(w^2) - (E(wz))^2} w$

c) the physical meaning

Notice that $||x - \hat{x}_*|| = \sqrt{E(x^2) - E(\hat{x}_*^2)} = \sqrt{||x||^2 - ||\hat{x}_*||^2}$
Therefore, $\hat{x}_*$ is the projection of x onto the 2D plane of $span\{w, z\}$.
(If w and z are linear dependent, it becomes 1D line of $span\{w, z\}$.)