Linear Dynamical Systems (Kalman Filters)

(a) Filtering and Smoothing in LDS
(b) EM in LDS
Linear Dynamical Systems (LDS)

\[ x_t = W y_t + e_t \]

Transition model

\[ y_1 = \mu_0 + u \]
\[ y_t = A y_{t-1} + v_t \]

Parameters: \( \theta = \{W, A, \mu_0, \Sigma, \Gamma, P_0\} \)
Linear Dynamical Systems (LDS)

\[
p(y_1) = N(y_1 | \mu_0, P_0)
\]

Transition Probability:  \[
p(y_t | y_{t-1}) = N(y_t | Ay_{t-1}, \Gamma)
\]

Emission:  \[
p(x_t | y_t) = N(x_t | Wy_t, \Sigma)
\]
## HMM vs LDS

<table>
<thead>
<tr>
<th>HMM</th>
<th>LDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markov Chain with discrete latent variables</td>
<td>Markov Chain with continuous latent variables</td>
</tr>
<tr>
<td>$p(y_1)$</td>
<td>$p(y_1) = N(y_1</td>
</tr>
<tr>
<td>$\pi$</td>
<td></td>
</tr>
<tr>
<td>$Kx1$</td>
<td></td>
</tr>
<tr>
<td>$p(y_t</td>
<td>y_{t-1})$</td>
</tr>
<tr>
<td>$A$</td>
<td></td>
</tr>
<tr>
<td>$KxK$</td>
<td></td>
</tr>
<tr>
<td>$p(x_t</td>
<td>y_t)$</td>
</tr>
<tr>
<td>$B$</td>
<td></td>
</tr>
<tr>
<td>$LxK$</td>
<td></td>
</tr>
<tr>
<td>or</td>
<td></td>
</tr>
<tr>
<td>$K$ distributions</td>
<td></td>
</tr>
</tbody>
</table>
LDS

\[ p(y_t | x_1, x_2, ..., x_t) \]

\[ p(y_{t-\tau} | x_1, x_2, ..., x_t) \]

\[ p(y_t | x_1, x_2, ..., x_T) \]

\[ p(y_{t+\delta} | x_1, x_2, ..., x_t) \]

\[ p(x_{t+\delta} | x_1, x_2, ..., x_t) \]
Filtering

Filtering: \( \hat{a}(y_t) = p(y_t | x_1, x_2, ..., x_t) \)

The filtered probability is a Gaussian: \( \hat{a}(y_t) = N(y_t | \mu_t, V_t) \)

Hence we need to recursively compute: \( \mu_t, V_t \)

HMM

\[ c_t \hat{a}(z_t) = p(x_t | z_t) \sum_{z_{t-1}} \hat{a}(z_{t-1}) p(z_t | z_{t-1}) \]

\[ c_t \hat{a}(y_t) = p(x_t | y_t) \int \hat{a}(y_{t-1}) p(y_t | y_{t-1}) dy_{t-1} \]
Filtering

\[
\int_{y_{t-1}} \hat{\alpha}(y_{t-1}) \rho(y_t|y_{t-1}) \, dy_{t-1} \\
= \int N(y_t|Ay_{t-1}, \Gamma)N(y_{t-1}|\mu_{t-1}, V_{t-1}) \, dy_{t-1} \\
= N(y_t|A\mu_{t-1}, P_{t-1})
\]

Using the technique “completing the square

\[
P_{t-1} = AV_{t-1}A^T + \Gamma
\]
Filtering

\[ c_t \hat{a}(y_t) = N(x_t | W y_t, \Sigma)N(y_t | Ay_{t-1}, P_{t-1}) \]

\[ c_t N(y_t | \mu_t, V_t) = N(x_t | W y_t, \Sigma)N(y_t | Ay_{t-1}, P_{t-1}) \]

Which gives the updates:

\[ \mu_t = A \mu_{t-1} + K_t (x_t - WA \mu_{t-1}) \]

\[ V_t = (I - K_t W) P_{t-1} \]

Kalman Gain

\[ K_t = P_{t-1} W^T (WP_{t-1} W^T + \Sigma)^{-1} \]

and:

\[ c_t = N(x_t | WA \mu_{t-1}, WP_{t-1} W^T + \Sigma) \]
Filtering

Start of the recursion

\[
c_1 \hat{a}(y_1) = p(y_1)p(x_1|y_1)
\]

\[
c_1 \mathcal{N}(y_1|\mu_1, V_1) = \mathcal{N}(y_1|\mu_0, P_0)\mathcal{N}(x_t|Wy_t, \Sigma)
\]

which gives

\[
\mu_1 = \mu_0 + K_1(x_1 - Wy_0)
\]

\[
V_1 = (I - K_1W)P_0
\]

\[
K_1 = P_0W^T(WP_0W^T + \Sigma)^{-1}
\]

\[
c_1 = \mathcal{N}(x_1|Wy_0, WP_0W^T + \Sigma)
\]
Smoothing

Smoothing: \( \gamma(y_t) = p(y_t | x_1, x_2, ..., x_T) \)

\[ \gamma(y_t) = \hat{a}(y_t) \hat{\beta}(y_t) = N(y_t | \mu_t, \nu_t) \]

We have computed \( \hat{a}(y_t) \) now let us compute \( \hat{\beta}(y_t) \) (backward step)

\[
HMM \ c_{t+1} \hat{\beta}(z_t) = \sum_{z_{t+1}} \hat{\beta}(z_{t+1}) p(x_{t+1} | z_{t+1}) p(z_{t+1} | z_t)
\]

\[
c_{t+1} \hat{\beta}(y_t) = \int \hat{\beta}(y_{t+1}) p(x_{t+1} | y_{t+1}) p(y_{t+1} | y_t) dy_{t+1}
\]
Smoothing

\[ \gamma(y_t) = N(y_t | \mu_t, V_t) \]

After similar manipulations as in \( \hat{a}(y_t) \) we get the updates

\[ \widehat{\mu}_t = \mu_t + J_t (\mu_{t+1} - A \mu_t) \]

\[ \widehat{V}_t = V_t + J_t (V_{t+1} - P_t) J_t^T \]

\[ J_t = V_t A^T (P_t)^{-1} \]

It's necessary to complete the forward step so that we have \( P_t \) and \( \mu_t \) computed.
Smoothing

\[ p(y_{t-1}, y_t | x_1, \ldots, x_T) = \]

\[ \xi(y_{t-1}, y_t) = (c_t)^{-1} \hat{\alpha} (y_{t-1})p(x_t | y_t)p(y_t | y_{t-1})\hat{\beta} (y_t) \]

Which is again a Gaussian

\[ \xi(y_{t-1}, z_t) = N([y_{t-1}^T \ y_t] | \hat{\mu}_t, R_t) \]

\[ \hat{\mu}_t = \begin{bmatrix} \hat{\mu}_t^{-1} \\ \hat{\mu}_t \end{bmatrix} \]

\[ R_t = \begin{bmatrix} \tilde{V}_{t-1}^T & J_{t-1} \tilde{V}_t \\ (J_{t-1} \tilde{V}_t)^T & \tilde{V}_t \end{bmatrix} \]
E: Step

\[
E[y_t] = \int_{y_t} N(y_t | \hat{\mu}_t, \hat{V}_t) \, dy_t = \hat{\mu}_n
\]

\[
E[y_t y_t^T] = \hat{V}_t + \hat{\mu}_t \hat{\mu}_t^T
\]

\[
E[y_t y_{t-1}^T] = \hat{V}_t J_{t-1}^T + \hat{\mu}_t \hat{\mu}_{t-1}^T
\]
EM

Assume the sequence $x_1, x_2, \ldots, x_T$

The complete likelihood is given by:

$$p(X, Y | \theta) = p(x_1, x_2, \ldots, x_T, y_1, y_2, \ldots, y_T | \theta)$$

$$= \prod_{t=1}^{T} p(x_t | y_t) p(y_1) \prod_{t=2}^{T} p(y_t | y_{t-1})$$

$$\Rightarrow \ln p(X, Y | \theta) = \ln p(y_1 | \mu_0, P_0) + \sum_{t=2}^{T} \ln p(y_t | y_{t-1}, A, \Gamma)$$

$$+ \sum_{t=1}^{T} \ln p(x_t | y_t, W, \Sigma)$$
EM

Now we take the expectation with regards to \( Y|X \)

\[
\Rightarrow E[\ln p(X, Y|\theta)]
\]

\[
E[\ln p(X, Y|\theta)] = E[\ln p(y_1|\mu_0, P_0)] + E[\sum_{t=2}^{T} \ln p(y_t|y_{t-1}, A, \Gamma)] \\
+ E[\sum_{t=1}^{T} \ln p(x_t|y_t, W, \Sigma)]
\]
To find $\mu_0, P_0$ we need the first term only

$$E[\ln p(y_1 | \mu_0, P_0)]$$

$$= -\frac{1}{2} \ln |P_0| - E \left[ \frac{1}{2} (y_1 - \mu_0)^T P_0^{-1} (y_1 - \mu_0) \right]$$

Taking the derivative of the above and making equal to zero we get

$$\mu_0^{new} = E[y_1]$$

$$P_0^{new} = E[y_1 y_1^T] - E[y_1]E[y_1^T]$$
EM

To find $A$, $\Gamma$ we need the second term only

$$
E \left[ \sum_{t=2}^{N} \ln p(y_t | y_{t-1}, A, \Gamma) \right] = -\frac{N - 1}{2} \ln |\Gamma|
$$

$$
-E \left[ \frac{1}{2} \sum_{t=2}^{N} (y_t - Ay_{t-1})^T \Gamma^{-1} (y_t - Ay_{t-1}) \right]
$$

$$
A^{new} = \left( \sum_{t=2}^{T} E [y_t y_{t-1}^T] \right) \left( \sum_{t=2}^{T} E [y_{t-1} y_{t-1}^T] \right)^{-1}
$$

$$
\Gamma^{new} = \frac{1}{N - 1} \sum_{t=2}^{T} \{ E[y_t y_t^T] - A^{new} E[y_{t-1} y_t^T] 
$$

$$
- E[y_t y_{t-1}^T] (A^{new})^T E[y_{t-1} y_{t-1}^T] (A^{new})^T \}$$
EM

To find $C, W$ we need the third term only

$$E \left[ \sum_{t=1}^{T} \ln p(x_t | y_t, W, \Sigma) \right]$$

$$= - \frac{N}{2} \ln |\Sigma| - E \left[ \frac{1}{2} \sum_{t=1}^{T} (x_t - Wy_t)^T \Sigma^{-1} (x_t - Wy_t) \right]$$
Taking the derivative and forcing to zero we get

\[ \mathbf{W}^{\text{new}} = \left( \sum_{t=1}^{T} \mathbf{x}_t \mathbb{E}[\mathbf{y}_t^T] \right) \left( \sum_{t=1}^{T} \mathbb{E}[\mathbf{y}_t \mathbf{y}_t^T] \right)^{-1} \]

\[ \Sigma^{\text{new}} = \frac{1}{T} \sum_{t=1}^{T} \{ \mathbf{x}_t \mathbf{x}_t^T - \mathbf{W}^{\text{new}} \mathbb{E}[\mathbf{y}_t] \mathbf{x}_t^T \}
- \mathbf{x}_t \mathbb{E}[\mathbf{y}_t^T] \mathbf{W}^{\text{new}} + (\mathbf{W}^{\text{new}})^T \mathbb{E}[\mathbf{y}_t \mathbf{y}_t^T] \mathbf{W}^{\text{new}} \} \]