

Byron Yu

Carnegie Mellon University

## Kalman Filter

### A) Linear Dynamical Systems (LDS)

At timestep  $t=1, \dots, T$ , let:

$z_t \in \mathbb{R}^M$  be the state variable

(e.g. arm position and velocity)

$x_t \in \mathbb{R}^D$  be the observation

(e.g. spike count vector)

State model:

$$\begin{aligned}\underline{z}_t | \underline{z}_{t-1} &\sim N(A \underline{z}_{t-1}, Q) \\ \underline{z}_1 &\sim N(\underline{\Pi}, V)\end{aligned}\quad (1)$$

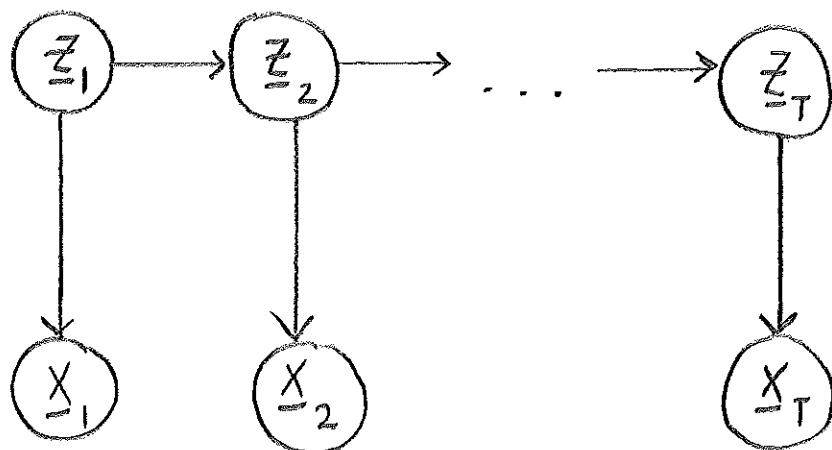
Observation model:

$$\underline{x}_t | \underline{z}_t \sim N(C \underline{z}_t, R) \quad (2)$$

Model parameters are  $\theta = \{A, Q, \underline{\Pi}, V, C, R\}$

What are the dimensions of each of these parameters?

Graphical model:



- State model describes how state evolves over time.
- Observation model describes how observation relates to the state.
- State model uses a Markov assumption:

$$\begin{aligned} P(\underline{z}_1, \dots, \underline{z}_T) &= P(z_1) P(z_2 | z_1) \dots P(z_T | z_1, \dots, z_{T-1}) \\ &= P(z_1) \prod_{t=2}^T P(z_t | z_{t-1}) \end{aligned} \quad \text{Markov assumption}$$

## B) Training phase

Goal: Estimate the model parameters

$\theta = \{A, Q, \Pi, V, C, R\}$  from the training data.

- If the values of the state variables  $\underline{z}_t$  are unknown during training, use EM algorithm (unsupervised learning).

Maximize  $P(\{\underline{x}\} | \theta)$  w.r.t.  $\theta$ .

- Here, we will consider the simpler case, where the  $\underline{z}_t$  are known during training (supervised learning).

Maximize  $P(\{x\}, \{\underline{z}\} | \theta)$  w.r.t.  $\theta$ .

For decoding arm trajectories from neural activity, the  $\underline{z}_t$  (arm states) are typically known during training.

$$P(\{x\}, \{\underline{z}\} | \theta) = P(\underline{z}_1) \prod_{t=2}^T P(\underline{z}_t | \underline{z}_{t-1}) \left( \prod_{t=1}^T P(x_t | \underline{z}_t) \right)$$

Let

$$\begin{aligned} \mathcal{L}(\theta) &= \log P(\{x\}, \{\underline{z}\} | \theta) \\ &= \log P(\underline{z}_1) + \sum_{t=2}^T \log P(\underline{z}_t | \underline{z}_{t-1}) + \sum_{t=1}^T \log P(x_t | \underline{z}_t) \\ &= -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log |V| - \frac{1}{2} (\underline{z}_1 - \underline{\mu})^T V^{-1} (\underline{z}_1 - \underline{\mu}) \\ &\quad + \sum_{t=2}^T \left( -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log |Q| - \frac{1}{2} (\underline{z}_t - A\underline{z}_{t-1})^T Q^{-1} (\cdot) \right) \\ &\quad + \sum_{t=1}^T \left( -\frac{D}{2} \log(2\pi) - \frac{1}{2} \log |R| - \frac{1}{2} (x_t - (\underline{z}_t)^T)^T R^{-1} (\cdot) \right) \end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta)}{\partial A} &= \frac{\partial}{\partial A} \left\{ \sum_{t=2}^T \left( -\underline{z}_t^\top A^\top Q^{-1} \underline{z}_t - \underline{z}_t^\top Q^{-1} A \underline{z}_{t-1} + \underline{z}_{t-1}^\top A^\top Q^{-1} A \underline{z}_{t-1} \right) \right\} \\
&= \frac{\partial}{\partial A} \left\{ -\text{Tr} \left( A^\top Q^{-1} \sum_{t=2}^T \underline{z}_t \underline{z}_{t-1}^\top \right) - \text{Tr} \left( A \left( \sum_{t=2}^T \underline{z}_{t-1} \underline{z}_t^\top \right) Q^{-1} \right) \right. \\
&\quad \left. + \text{Tr} \left( Q^{-1} A \left( \sum_{t=2}^T \underline{z}_{t-1} \underline{z}_{t-1}^\top \right) A^\top \right) \right\} \\
&= -Q^{-1} \left( \sum_{t=2}^T \underline{z}_t \underline{z}_{t-1}^\top \right) - Q^{-1} \left( \sum_{t=2}^T \underline{z}_{t-1} \underline{z}_{t-1}^\top \right) \\
&\quad + Q^{-1} A \left( \sum_{t=2}^T \underline{z}_{t-1} \underline{z}_{t-1}^\top \right) + Q^{-1} A \left( \sum_{t=2}^T \underline{z}_t \underline{z}_{t-1}^\top \right) \\
&= [0]
\end{aligned}$$

$$A = \left( \sum_{t=2}^T \underline{z}_t \underline{z}_{t-1}^\top \right) \left( \sum_{t=2}^T \underline{z}_{t-1} \underline{z}_{t-1}^\top \right)^{-1} \quad (3)$$

$$\begin{aligned}
\frac{\partial \mathcal{L}(\theta)}{\partial Q} &= \frac{\partial}{\partial Q} \left\{ -\frac{(T-1)}{2} \log |Q| - \frac{1}{2} \text{Tr} \left( Q^{-1} \sum_{n=2}^T (\underline{z}_n - A \underline{z}_{n-1}) (\underline{z}_n - A \underline{z}_{n-1})^\top \right) \right\} \\
&= -\frac{(T-1)}{2} Q^{-1} - \frac{1}{2} \left( -Q^{-1} \sum_{n=2}^T (\underline{z}_n - A \underline{z}_{n-1}) (\underline{z}_n - A \underline{z}_{n-1})^\top Q^{-1} \right) \\
&= [0]
\end{aligned}$$

$$Q = \frac{1}{T-1} \sum_{t=2}^T (\underline{z}_t - A \underline{z}_{t-1})(\underline{z}_t - A \underline{z}_{t-1})^T \quad (4)$$

use the  $A$  found in (3)

Note: The expressions for  $A$  and  $Q$  are entirely analogous to those in linear regression, where  $A$  is the "slope" and  $Q$  is the "minimum mean squared error".

(see Appendix ii))

Similarly,

$$\frac{\partial L(\theta)}{\partial C} = [0] \quad \text{yields}$$

$$C = \left( \sum_{t=1}^T X_t \underline{z}_t^T \right) \left( \sum_{t=1}^T \underline{z}_t \underline{z}_t^T \right)^{-1} \quad (5)$$

$$\frac{\partial L(\theta)}{\partial R} = [0] \quad \text{yields}$$

$$R = \frac{1}{T} \sum_{t=1}^T (X_t - C \underline{z}_t)(X_t - C \underline{z}_t)^T \quad (6)$$

use the  $C$  found in (5).

For notational simplicity, we consider only one sequence  $(\underline{z}_1, \dots, \underline{z}_T)$  here. In general,

there may be multiple sequences, each with a different number of time steps.

Let  $\{\underline{x}\}_n, \{\underline{z}\}_n$  represent the  $n$ th sequence ( $n=1, \dots, N$ )

Now, the goal is to maximize  $\prod_{n=1}^N P(\{\underline{x}\}_n, \{\underline{z}\}_n | \theta)$   
w.r.t.  $\theta$ .

The resulting expressions for (3) through (6) have the same form, but each summation sums over more elements.

$\bar{\Pi}$  and  $\bar{V}$  are the sample mean and covariance, respectively, of the  $N$  instances of  $\underline{z}_i$ .

C) Test phase: Decoding arm trajectories from neural activity

Goal: To compute  $P(\underline{z}_t | \underline{x}_1, \dots, \underline{x}_t)$   
for  $t=1, \dots, T$ .

We will use the shorthand  $\{\underline{x}\}_1^t$

The variables  $\underline{z}_1, \dots, \underline{z}_T, \underline{x}_1, \dots, \underline{x}_T$  are jointly Gaussian, so  $P(\underline{z}_t | \{\underline{x}\}_1^t)$  is Gaussian.

Thus, we need only find its mean and covariance.

We can compute  $P(\underline{z}_t | \{\underline{x}\}_1^t)$  recursively starting at  $t=1$ :

- One-step prediction

$$P(\underline{z}_t | \{\underline{x}\}_1^{t-1}) = \underbrace{\int P(\underline{z}_t | \underline{z}_{t-1})}_{\text{State model}} \underbrace{P(\underline{z}_{t-1} | \{\underline{x}\}_1^{t-1}) d\underline{z}_{t-1}}_{(7)}$$

- Measurement update

$$\underbrace{P(\underline{z}_t | \{\underline{x}\}_1^t)}_{\text{our goal}} = \frac{\underbrace{P(\underline{x}_t | \underline{z}_t)}_{\text{obs model}} P(\underline{z}_t | \{\underline{x}\}_1^{t-1})}{P(\underline{x}_t | \{\underline{x}\}_1^{t-1})} \quad (8)$$

Let

$$\mu_t^t = E[z_t | \{x\}_1^t]$$

$$\Sigma_t^t = \text{cov}(z_t | \{x\}_1^t)$$

We want to express (7) and (8) in terms of the model parameters.

Plugging the state and observation models into (7) and (8), then simplifying, is a method that will always work.

Here, we will recognize that all the distributions in (7) and (8) are Gaussian, and just solve for means and covariances.

- One-step prediction

$$z_t | \{x\}_1^{t-1} \sim N(\mu_t^{t-1}, \Sigma_t^{t-1})$$

Find  $\mu_t^{t-1}$  and  $\Sigma_t^{t-1}$ .

An equivalent way of writing (1) is

$$z_t = Az_{t-1} + v_t, \quad v_t \sim N(0, Q).$$

$$\mu_t^{t-1} = E[z_t | \{x\}_1^{t-1}]$$

$$= A E[z_{t-1} | \{x\}_1^{t-1}] + E[v_t | \{x\}_1^{t-1}]$$

$$\boxed{\mu_t^{t-1} = A \mu_{t-1}^{t-1}}$$

(9)

$$\begin{aligned}
 \Sigma_t^{t-1} &= \text{cov}(\underline{z}_t | \{\underline{x}\}_1^{t-1}) \\
 &= A \text{cov}(\underline{z}_{t-1} | \{\underline{x}\}_1^{t-1}) A^T + \text{cov}(\underline{v}_t | \{\underline{x}\}_1^{t-1}) \\
 \boxed{\Sigma_t^{t-1} = A \Sigma_{t-1}^{t-1} A^T + Q} \quad (10)
 \end{aligned}$$

• Measurement update

$$\underline{z}_t | \{\underline{x}\}_1^t \sim N(\underline{\mu}_t^t, \Sigma_t^t)$$

Find  $\underline{\mu}_t^t$  and  $\Sigma_t^t$ .

Recognize that (8) is Bayes rule for  $\underline{z}_t$  and  $\underline{x}_t$ ,  
with all terms conditioned on  $\{\underline{x}\}_1^{t-1}$ .

Thus, we will first find the joint probability of  $\underline{z}_t$   
and  $\underline{x}_t$  given  $\{\underline{x}\}_1^{t-1}$ , then apply the results  
of conditioning for jointly Gaussian random vars.

$$\left[ \begin{array}{c} \underline{x}_t \\ \underline{z}_t \end{array} \right] \mid \{\underline{x}\}_1^{t-1} \sim N \left( \left[ \begin{array}{c} C \underline{\mu}_t^{t-1} \\ \underline{\mu}_t^{t-1} \end{array} \right], \left[ \begin{array}{cc} C \Sigma_t^{t-1} C^T + R & C \Sigma_t^{t-1} \\ C \Sigma_t^{t-1} & \Sigma_t^{t-1} \end{array} \right] \right)$$

a)  
b)

(11)

An equivalent way of writing (2) is

$$\underline{x}_t = C \underline{z}_t + \underline{w}_t, \quad \underline{w}_t \sim N(\underline{0}, R).$$

$$\begin{aligned} a) E[\underline{x}_t | \{\underline{x}\}_{1,t-1}^{t-1}] &= C E[\underline{z}_t | \{\underline{x}\}_{1,t-1}^{t-1}] + E[\underline{w}_t | \{\underline{x}\}_{1,t-1}^{t-1}] \\ &= C \underline{\mu}_t^{t-1} \end{aligned}$$

$$\begin{aligned} \text{cov}(\underline{x}_t | \{\underline{x}\}_{1,t-1}^{t-1}) &= C \text{cov}(\underline{z}_t | \{\underline{x}\}_{1,t-1}^{t-1}) C^T + \text{cov}(\underline{w}_t | \{\underline{x}\}_{1,t-1}^{t-1}) \\ &= C \sum_t^{t-1} C^T + R \end{aligned}$$

$$\begin{aligned} b) E[\underline{x}_t \underline{z}_t^T | \{\underline{x}\}_{1,t-1}^{t-1}] - E[\underline{x}_t | \{\underline{x}\}_{1,t-1}^{t-1}] E[\underline{z}_t | \{\underline{x}\}_{1,t-1}^{t-1}]^T \\ &= E[C \underline{z}_t \underline{z}_t^T + \underline{w}_t \underline{z}_t^T | \{\underline{x}\}_{1,t-1}^{t-1}] - C \underline{\mu}_t^{t-1} \underline{\mu}_t^{t-1 T} \\ &= C \left( \sum_t^{t-1} + \underline{\mu}_t^{t-1} \underline{\mu}_t^{t-1 T} \right) + E[\underline{w}_t | \{\underline{x}\}_{1,t-1}^{t-1}] E[\underline{z}_t | \{\underline{x}\}_{1,t-1}^{t-1}]^T \\ &\quad - C \underline{\mu}_t^{t-1} \underline{\mu}_t^{t-1 T} \\ &= C \sum_t^{t-1} \end{aligned}$$

Applying the results of conditioning for jointly Gaussian random variables to (11),  
 (see Appendix iii))

$$\begin{aligned}\underline{\mu}_t^t &= E[\underline{z}_t | \underline{x}_t, \{\underline{x}\}_{t-1}^{t-1}] \\ &= \underline{\mu}_t^{t-1} + \underbrace{\sum_t^{t-1} C^T (C \sum_t^{t-1} C^T + R)^{-1} (\underline{x}_t - C \underline{\mu}_t^{t-1})}_{\text{call this } K_t, \text{ the "Kalman gain"}}$$

call this  $K_t$ , the  
 "Kalman gain"

Rewriting,

$$\boxed{\underline{\mu}_t^t = \underline{\mu}_t^{t-1} + K_t (\underline{x}_t - C \underline{\mu}_t^{t-1})} \quad (12)$$

$$\begin{aligned}\sum_t^t &= \text{cov}(\underline{z}_t | \underline{x}_t, \{\underline{x}\}_{t-1}^{t-1}) \\ &= \sum_t^{t-1} - \sum_t^{t-1} C^T (C \sum_t^{t-1} C^T + R)^{-1} C \sum_t^{t-1} \\ \boxed{\sum_t^t = \sum_t^{t-1} - K_t C \sum_t^{t-1}} &\quad (13)\end{aligned}$$

Taking the recursions defined by (9), (10), (12), (13),

we obtain  $\mu_t^t$  and  $\Sigma_t^t$  for  $t=1, \dots, T$ .

- $\mu_t^t$  is the arm state estimate at time  $t$
- $\Sigma_t^t$  is our uncertainty around that estimate at time  $t$ .

Initialize recursions with  $\mu_1^0 = \underline{\Pi}$

$$\Sigma_1^0 = V.$$

## Appendix

### i) Useful matrix properties

$$\frac{d}{dX} \text{Tr}(XA^T) = \frac{d}{dX} \text{Tr}(X^TA) = A$$

$$\frac{d}{dX} \text{Tr}(AXBX^TC) = A^TC^TXB^T + CAXB$$

$$\frac{d}{dX} \log |X| = X^{-T}$$

$$\frac{d}{dX} \text{Tr}(X^{-1}A) = -X^{-T}A^TX^{-T}$$

### ii) Linear regression

$$x_n = w z_n + \mu, \quad n=1, \dots, N$$

Minimizing mean squared error with respect to  $w$ ,

$$w^* = \frac{\sum_{n=1}^N (x_n - \mu) z_n}{\sum_{n=1}^N z_n^2}$$

Using this  $w^*$ , minimum mean squared error is

$$\frac{1}{N} \sum_{n=1}^N (x_n - w^* z_n - \mu)^2$$

## iii) Gaussian conditioning

If  $\underline{x} = \begin{bmatrix} x_a \\ x_b \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \right)$ ,

$P(x_a | x_b)$  is Gaussian with the following  
mean and covariance:

$$E[x_a | x_b] = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$$

$$\text{cov}(x_a | x_b) = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}$$