

1 LSD model specification

The latents have first order Markov structure with linear dynamics and i.i.d. gaussian noise:

$$\mathbf{z}_i = \mathbf{A}\mathbf{z}_{i-1} + \mathbf{w}_i \quad (1)$$

where $\mathbf{z}_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$ and $\mathbf{w}_i \sim \mathcal{N}(0, \mathbf{Q})$ (to avoid degeneracy, we usually take $\mathbf{Q} = \mathbf{I}$).

Observations are obtained by another linear projection with i.i.d. gaussian noise:

$$\mathbf{x}_i = \mathbf{C}\mathbf{z}_i + \mathbf{v}_i \quad (2)$$

with $\mathbf{v}_i \sim \mathcal{N}(0, \mathbf{R})$.

2 Kalman filter

The goal is to compute the posterior distribution of the latent state \mathbf{z}_i given past observations $\{\mathbf{x}_1, \dots, \mathbf{x}_i\}$. The proof proceeds by induction: we start by inferring the posterior distribution of latent variable \mathbf{z}_1 condition on observation \mathbf{x}_1 (marginalizing out the unobserved initial state \mathbf{z}_0). The resulting posterior is also gaussian so it can be treated as prior for inferring the next latent state and so on.

To compute the posterior $P(\mathbf{z}_1 | \mathbf{x}_1)$ we start from the joint distribution $P(\mathbf{z}_1, \mathbf{x}_1)$ and then use the general formula for conditioning in multivariate Gaussians:

$$\mu_{z|x} = \mu_z + \Sigma_{zx} \Sigma_{xx}^{-1} (\mathbf{x} - \mu_x) \quad (3)$$

$$\Sigma_{z|x} = \Sigma_{zz} - \Sigma_{zx} \Sigma_{xx}^{-1} \Sigma_{xz} \quad (4)$$

To compute the means, we start with $\mathbf{z}_1 = \mathbf{A}\mathbf{z}_0 + \mathbf{w}_1$ which is a sum of 2 independent multivariate gaussian distributions with means $\mathbf{A}\mu_0$ and 0, and covariances $\mathbf{A}\Sigma_0\mathbf{A}^t$ and \mathbf{Q} , respectively. Putting the two together, yields the prior $\mathbf{z}_1 \sim \mathcal{N}(\mu_{1|0}, \Sigma_{1|0})$, with parameters: ¹

$$\mu_z = \mu_{1|0} = \mathbf{A}\mu_0 \quad (5)$$

$$\Sigma_{zz} = \Sigma_{1|0} = \mathbf{A}\Sigma_0\mathbf{A}^t + \mathbf{Q} \quad (6)$$

Since \mathbf{x}_1 is generated using the same kind of process, we can get the corresponding moments exactly in the same way:

$$\mu_x = \mathbf{C}\mu_{1|0} \quad (7)$$

$$\Sigma_{xx} = \mathbf{C}\Sigma_{1|0}\mathbf{C}^t + \mathbf{R} \quad (8)$$

Lastly, we need to compute the covariance:²

$$\Sigma_{zx} = \text{cov}[\mathbf{z}_1, \mathbf{x}_1] = \text{cov}[\mathbf{z}_1, \mathbf{C}\mathbf{z}_1 + \mathbf{v}_1] = \Sigma_{1|0}\mathbf{C}^t \quad (9)$$

where we have used the fact that the observation noise is independent (so the second term cancels out).

Plugging the different elements in Eq. 3 we obtain the posterior mean for \mathbf{z}_1 :

$$\mu_{1|1} = \mu_{1|0} + \mathbf{K}(\mathbf{x}_1 - \mathbf{C}\mu_{1|0}) \quad (10)$$

where we use matrix $\mathbf{K} = \Sigma_{zx} \Sigma_{xx}^{-1}$ as shorthand for the Kalman gain.

¹Notational convention for the double indexing of $\mu_{i|j}$ and $\Sigma_{i|j}$ means that we are computing the moments of latent \mathbf{z}_i , conditioned on observations up to j , $\{\mathbf{x}_1, \dots, \mathbf{x}_j\}$.

²Remember that linear combinations distribute when computing covariances.

The corresponding posterior covariance matrix is (using Eq. 4):

$$\Sigma_{1|1} = \Sigma_{1|0} - \mathbf{K}\mathbf{C}\Sigma_{1|0}. \quad (11)$$

In general, for $i > 1$ we first compute the prior for \mathbf{z}_{i+1} as:

$$\mu_{i|i-1} = \mathbf{A}\mu_{i-1|i-1} \quad (12)$$

$$\Sigma_{i|i-1} = \mathbf{A}\Sigma_{i-1|i-1}\mathbf{A}^t + \mathbf{Q} \quad (13)$$

Then we incorporate the evidence, to obtain the posterior with parameters:

$$\mu_{i|i} = \mu_{i|i-1} + \mathbf{K}_i (x_i - \mathbf{C}\mu_{i|i-1}) \quad (14)$$

$$\Sigma_{i|i} = \Sigma_{i|i-1} - \mathbf{K}_i\mathbf{C}\Sigma_{i|i-1} \quad (15)$$

where the Kalman gain is $\mathbf{K}_i = \Sigma_{i|i-1}\mathbf{C}^t (\mathbf{C}\Sigma_{i|i-1}\mathbf{C}^t + R)^{-1}$.

3 Kalman smoothing

The goal of smoothing is to compute the posterior distribution of the latent state \mathbf{z}_i but given the full sequence of observations $\{\mathbf{x}_1, \dots, \mathbf{x}_t\}$. There are very different ways of deriving the smoother; here we use the approach of Ainsley and Kohn (1982), as presented in Chapter 6 of Shumway and Stoffer.

The starting point of this version of the derivation is focusing on the dependency between \mathbf{z}_i and \mathbf{z}_{i+1} , and explicitly separating out different components of the generative model such that the parts are (conditionally) independent. Specifically, we start by defining some shorthand variables $\mathbf{x}_{1:i} = \{\mathbf{x}_1, \dots, \mathbf{x}_i\}$ and $\eta_{i+1} = \{\mathbf{w}_{i+2}, \dots, \mathbf{w}_t, \mathbf{v}_{i+1}, \dots, \mathbf{v}_t\}$.

We also introduce an additional auxiliary random variable m_i , defined as follows:

$$m_i = \mathbb{E}[\mathbf{z}_i | \mathbf{x}_{1:i}, \mathbf{z}_{i+1} - \mu_{i+1|i}, \eta_{i+1}] \quad (16)$$

This estimates the uncertainty about \mathbf{z}_i after taking into account observations up to \mathbf{x}_i , the prediction error at step $i + 1$ (which depends implicitly on \mathbf{w}_i) and all the subsequent noise terms, which are independent of \mathbf{z}_i . Note also that the variables $\mathbf{x}_{1:i}$, $\mathbf{z}_{i+1} - \mu_{i+1|i}$, η_i are mutually independent.

To compute m_i we use again the trick of computing joint statistics for \mathbf{z}_i and $(\mathbf{z}_{i+1} - \mu_{i+1|i})$, conditioned on the rest and then recovering m_i via the multivariate normal conditioning formula, as done for the Kalman filter above.

First, the moments of \mathbf{z}_i conditioned on $\mathbf{x}_{1:i}$ are $\mu_{i|i}$ and $\Sigma_{i|i}$, according to our original definition.

Second, the corresponding moments for random variable $(\mathbf{z}_{i+1} - \mu_{i+1|i})$ are:³

$$\mathbb{E}[\mathbf{z}_{i+1} - \mu_{i+1|i} | \mathbf{x}_{1:i}] = \mathbb{E}[\mathbf{z}_{i+1} | \mathbf{x}_{1:i}] - \mu_{i+1|i} = 0 \quad (17)$$

$$\text{Var}[\mathbf{z}_{i+1} - \mu_{i+1|i} | \mathbf{x}_{1:i}] = \Sigma_{i+1|i} \quad (18)$$

Lastly the covariance term is:

$$\text{cov}[\mathbf{z}_i, \mathbf{z}_{i+1} - \mu_{i+1|i} | \mathbf{x}_{1:i}] = \text{cov}[\mathbf{z}_i, \mathbf{A}\mathbf{z}_{i+1} + w_t - \mu_{i+1|i} | \mathbf{x}_{1:i}] = \Sigma_{i|i}\mathbf{A}^t \quad (19)$$

since the rest of the terms cancel out.

Putting everything together (Eq. 3) we get the final expression for m_i :

$$m_i = \mu_{i|i} + \mathbf{F}_i (\mathbf{z}_{i+1} - \mu_{i+1|i}) \quad (20)$$

where $\mathbf{F}_i = \Sigma_{i|i}\mathbf{A}^t\Sigma_{i+1|i}^{-1}$.

³After conditioning, $\mathbf{z}_{i+1|i}$ is just a constant.

Lastly, we use m_i to compute the mean of \mathbf{z}_i conditioned on the full sequence of observations $\mathbf{x}_{1:t}$:

$$\mu_{i|t} = \mathbb{E}[m_i | \mathbf{x}_{1:t}] \quad (21)$$

$$= \mathbb{E}[\mu_{i|i} + \mathbf{F}_i (\mathbf{z}_{i+1} - \mu_{i+1|i}) | \mathbf{x}_{1:t}] \quad (22)$$

$$= \mu_{i|i} + \mathbf{F}_i (\mathbb{E}[\mathbf{z}_{i+1} | \mathbf{x}_{1:t}] - \mu_{i+1|i}) \quad (23)$$

$$= \mu_{i|i} + \mathbf{F}_i (\mu_{i+1|t} - \mu_{i+1|i}) . \quad (24)$$

For the covariance update rules, we start from the expression derived from the mean, $\mu_{i|t} = \mu_{i|i} + \mathbf{F}_i (\mu_{i+1|t} - \mu_{i+1|i})$, subtracting \mathbf{z}_i on both sides and rewriting slightly we get:

$$\mathbf{z}_i - \mu_{i|t} + \mathbf{F}_i \mu_{i+1|t} = \mathbf{z}_i - \mu_{i|i} + \mathbf{F}_i \mathbf{A} \mu_{i|i} \quad (25)$$

where we have used $\mu_{i+1|i} = \mathbf{A} \mu_{i|i}$. We then multiply each side with its transpose and take expectations (for \mathbf{z}_i). With some appropriate ordering of terms this yields:

$$\mathbb{E}[(\mathbf{z}_i - \mu_{i|t})(\mathbf{z}_i - \mu_{i|t})^t] + \mathbf{F}_i \mathbb{E}[\mu_{i+1|t} \mu_{i+1|t}^t] \mathbf{F}_i^t = \mathbb{E}[(\mathbf{z}_i - \mu_{i|i})(\mathbf{z}_i - \mu_{i|i})^t] + \mathbf{F}_i \mathbf{A} \mathbb{E}[(\mu_{i|i})(\mu_{i|i})^t] \mathbf{A}^t \mathbf{F}_i^t \quad (26)$$

$$\Sigma_{i|t} + \mathbf{F}_i \mathbb{E}[\mu_{i+1|t} \mu_{i+1|t}^t] \mathbf{F}_i^t = \Sigma_{i|i} + \mathbf{F}_i \mathbf{A} \mathbb{E}[\mu_{i|i} \mu_{i|i}^t] \mathbf{A}^t \mathbf{F}_i^t \quad (27)$$

where we used the fact that the cross-products cancel out. Finally, the remaining expectations can be rewritten as:⁴:

$$\mathbb{E}[\mu_{i+1|t} \mu_{i+1|t}^t] = \mathbb{E}[\mathbf{z}_{i+1}, \mathbf{z}_{i+1}^t] - \Sigma_{i+1|t} \quad (28)$$

$$= \mathbf{A} \mathbb{E}[\mathbf{z}_i, \mathbf{z}_i^t] \mathbf{A}^t + \mathbf{Q} - \Sigma_{i+1|t} \quad (29)$$

$$\mathbb{E}[\mu_{i|i} \mu_{i|i}^t] = \mathbb{E}[\mathbf{z}_i, \mathbf{z}_i^t] - \Sigma_{i|i} . \quad (30)$$

where we have used the one step linear dependence between \mathbf{z}_i and \mathbf{z}_{i+1} .

Including these expressions, and rewriting $\mathbf{A} \Sigma_{i|i} \mathbf{A}^t + \mathbf{Q}$ back as $\Sigma_{i+1|i}$, the posterior covariance for smoothing simplifies to the final expression:

$$\Sigma_{i|t} = \Sigma_{i|i} + \mathbf{F}_i (\Sigma_{i+1|t} - \Sigma_{i+1|i}) \mathbf{F}_i^t . \quad (31)$$

Overall, filtering (forward pass) includes information flowing rightwards from index 1 to t with the current state \mathbf{z}_i being updates as a function of past state \mathbf{z}_{i-1} and current observation \mathbf{x}_i . In contrast, during smoothing (backward pass) the current state \mathbf{z}_i is updated based on posterior for the next state \mathbf{z}_{i+1} which already incorporates information from the full sequence.

⁴Here we use the fact that the covariance is 2nd moment - squared mean.