

# 1 Introduction:

$$\xi_{t+1} = F\xi_t + v_{t+1} \quad (1.1)$$

$$y_t = A + H\xi_t + w_t \quad (1.2)$$

where

$$\begin{aligned} \xi_t &= \begin{bmatrix} \zeta_t \\ \zeta_{t-1} \end{bmatrix} & F &= \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \\ v_t &= \begin{bmatrix} u_t \\ 0 \end{bmatrix} & y_t &= \begin{bmatrix} y_t^\infty \\ y_t^n \end{bmatrix} \\ A &= \begin{bmatrix} 0 \\ a \end{bmatrix} & H &= \begin{bmatrix} 1 & 0 \\ h & 0 \end{bmatrix} \\ w_t &= \begin{bmatrix} w_t^\infty \\ w_t^n \end{bmatrix} \end{aligned}$$

Equation (1.1) governs state evolution, which allows more sophisticated structures by changing its specification. Equation (1.2) governs how various measures interact with  $\xi_t$ . This setup admits only one additional measure  $y_t^n$ , but we can potentially add more measures with different  $n$ .

We further assume white noise on error terms:

$$E(v_t v_\tau') = \begin{cases} Q & t = \tau \\ 0 & otherwise \end{cases}$$

$$E(w_t w_\tau') = \begin{cases} R & t = \tau \\ 0 & otherwise \end{cases}$$

In addition, we assume that  $\xi_t$  is uncorrelated with any realizations of  $v_t$  or  $w_t$ .

## 2 Solve the Kalman filter

Here I use EM algorithm to solve the Kalman filter<sup>1</sup>. The challenge is that we will have partially missing data. For the remainder of this section I will lay out the algorithm to solve our problem.

---

<sup>1</sup>Alternatively, one can use MCMC, which is widely used in Bayesian statistics. In addition, UKF and EnKF will be explored in the future to account for non-linearity.

## 2.1 EM Algorithm Premier:

Denote  $Y$  as observed measurements,  $\Xi$  as hidden states, and  $\theta$  as parameters to be estimated. We want to optimize:

$$\begin{aligned} L(\theta) &= \log[P(Y|\theta)] \\ &= \log \left[ \frac{P(Y, \xi|\theta)}{P(\xi|Y, \theta)} \right] \\ &= \log[P(Y, \xi|\theta)] - \log[P(\xi|Y, \theta)] \end{aligned} \quad (2.1)$$

Take expectation of equation (2.1) wrt some distribution of  $\xi$  with pdf  $f(\xi)$  and get

$$\begin{aligned} L(\theta) &= \int f(\xi) \log[P(Y, \xi|\theta)] d\xi \\ &\quad - \int f(\xi) \log[P(\xi|Y, \theta)] d\xi \end{aligned}$$

To optimize  $L(\theta)$  we can iterate through the E steps and M steps to achieve a local maximum. By Jensen's inequality, we have the second term in equation (2.1) maximized when  $f(\xi) = P(\xi|Y, \theta)$  (E-step). If we define:

$$Q = \int \log[P(Y, \xi|\theta)] f(\xi|Y, \theta) d\xi$$

then  $Q = L(\theta)$ . For a given  $\hat{\theta}$  and  $P(\xi|Y, \hat{\theta})$ , we find  $\theta$  to optimize the first term (M-step). Use the new  $\theta$  as  $\hat{\theta}$  for the next iteration, and we will reach a local maximum. It is important to note that for a given  $\hat{\theta}$ ,  $f(\xi|Y, \hat{\theta})$  is a given quantity and does not change wrt.  $\theta$ .

## 2.2 E-step:

The E-steps involves forward updating and backward smoothing. First, I make the following definitions:

$$\begin{aligned} \hat{\xi}_{t+1|t} &\equiv \hat{E}(\xi_{t+1}|\mathcal{Y}_t) \\ \hat{\xi}_{t|t} &\equiv \hat{E}(\xi_t|\mathcal{Y}_t) \\ P_{t+1|t} &\equiv E[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})'] \end{aligned}$$

where  $\mathcal{Y}_t \equiv (y_t, y_{t-1}, \dots, y_1)$ .

We start off by providing initial estimates on  $\hat{\xi}_{1|0}$  and  $P_{1|0}$ . By Kalman Update rule<sup>2</sup>, we have:

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t-1} + FK_t(y_t - A - H\hat{\xi}_{t|t-1}) \quad (2.2)$$

$$P_{t|t} = P_{t|t-1} - K_tHP_{t|t-1} \quad (2.3)$$

$$P_{t+1|t} = FP_{t|t}F' + Q \quad (2.4)$$

$$K_t \equiv P_{t|t-1}H'(HP_{t|t-1}H' + R)^{-1} \quad (2.5)$$

Note that  $K_t$  in equation (2.5) is the Kalman Gain Matrix. It incorporates  $y_t$  and updates  $\hat{\xi}_{t|t}$ . By using equation (2.2) to (2.4) recursively, we are able to calculate  $\{\hat{\xi}_{t+1|t}\}$  and  $\{P_{t+1|t}\}$  for all  $t$ . Since we have data for  $t \in (1, 2, \dots, T)$ , we may use Kalman Smoothing to obtain better fit. The updating formula are the following:

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} + J_t(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}) \quad (2.6)$$

$$P_{t|T} = P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_t' \quad (2.7)$$

$$J_t \equiv P_{t|t}F'P_{t+1|t}^{-1} \quad (2.8)$$

We use equation (2.6) and (2.7) recursively backwards to get  $\{P_{t|T}\}$  and  $\{\hat{\xi}_{t|T}\}$ .

In theory, equation (2.3) should always yield positive semi-definite matrix. In practice, it is possible that we have non-PSD matrix due to numerical rounding errors. I use Joseph Form to guarantee we have a PSD matrix. Rewrite equation (2.3) as:

$$\begin{aligned} \hat{\xi}_{t|t} &= P_{t|t-1} - K_tHP_{t|t-1} \\ &= (I - K_tH)P_{t|t-1}(I - K_tH)' + K_tRK_t' \end{aligned} \quad (2.9)$$

Equation (2.9) guarantees the result is PSD

## 2.3 Sequential Processing and Missing Measurements:

In practice, we use sequential updating and smoothing. One obvious reason is improvement in numerical performance. Instead of performing the full matrix inverse, which can be prohibitively expensive if the either observation or state vector is large. But most importantly, by using sequential updating, I am able to update Kalman

---

<sup>2</sup>See Hamilton's Time Series Analysis for details.

Filter using only partial information, which is not handled by existing Kalman filter packages.

To perform Kalman updating sequentially, consider again equation (1.2):

$$y_t = A + H\xi_t + w_t$$

where  $E(w_t w_t') = R$ . We use LDL decomposition method to obtain  $L$  such that  $R = LDL'$ . Rewrite equation (1.2) as:

$$\bar{y}_t = L^{-1}A + L^{-1}H\xi_t + L^{-1}w_t$$

where  $\bar{w}_t = L^{-1}w_t$ . Note that  $\bar{w}_t$  has diagonal variance matrix  $D$ . For the remainder of this section, we assume that we work with transformed measures (thus omit the "-" notation).

Assume at time  $t$  we have  $k$  measurements. We initialize the measurement update process with  $\hat{\xi}_{t|t}^0 = \hat{\xi}_{t|t-1}$ , and  $P_{t|t}^0 = P_{t|t-1}$ . Consider the measurement update for the  $i$ th measurement,  $y_t^i$ :

$$\begin{aligned}\hat{\xi}_t^i &= \hat{\xi}_t^{i-1} + K_t^i(y_t^i - A - H\hat{\xi}_{t|t}^{i-1}) \\ P_{t|t}^i &= (I - K_t^i H_i)P_{t|t}^{i-1}(I - K_t^i H_i)' + K_t^i R_{i,i}(K_t^i)' \\ K_t^i &= \frac{P_{t|t}^{i-1} H_i'}{H_i P_{t|t}^{i-1} H_i' + R_{i,i}}\end{aligned}$$

If we don't have complete set of measurements at time  $t$ , we can simply skip the updating from any missing measurement.

## 2.4 M-step:

In this section, I detail steps to solve MLE iteratively. Note that to goal is to find parameters  $\theta \equiv A, H, F, R, Q$  to minimize:

$$L(\theta, y) = \sum_t^{t=T} \log[\mathbb{P}(y_t|\theta)] = \sum_t^{t=T} \log\{E[\mathbb{P}(y_t, \xi_t|\theta)]_{\xi_t}\} \quad (2.10)$$

By Markov Properties, we have:

$$\begin{aligned}
\sum_t^{t=T} E\{\log[\mathbb{P}(y_t, \xi_t)]\}_{\xi_t} &= \sum_t^{t=T} \log\{E[\mathbb{P}(\xi_t|\xi_{t-1}, \theta)]_{\xi_t}\} + \sum_t^{t=T} \log\{E[\mathbb{P}(y_t|\xi_t, \theta)]_{\xi_t}\} \\
&= \sum_t^{t=T} \log\left[\int \mathbb{P}(y_t|\xi_t, \theta)d\xi_t\right] + \sum_t^{t=T} \log\left[\int \mathbb{P}(\xi_t|\xi_{t-1}, \theta)d\xi_t\right]
\end{aligned}$$

If we assume Gaussian Process on the error terms, we have:

$$\begin{aligned}
\log[\mathbb{P}(\xi_t|\xi_{t-1}, \theta)] &= \frac{1}{2}\log|Q^{-1}| - \frac{1}{2}(\xi_t - F\xi_{t-1})'Q^{-1}(\xi_t - F\xi_{t-1}) + C \\
&= \frac{1}{2}\log|Q^{-1}| - \frac{1}{2}\{Trace[(\xi_t - F\xi_{t-1})'Q^{-1}(\xi_t - F\xi_{t-1})]\} + C \\
&= \frac{1}{2}\log|Q^{-1}| - \frac{1}{2}\{Trace[Q^{-1}(\xi_t - F\xi_{t-1})(\xi_t - F\xi_{t-1})']\} + C \quad (2.11)
\end{aligned}$$

The second equality follows from  $x = Trace(x)$  if  $x$  is a scalar, and the third equality follows from  $Trace(AB) = Trace(BA)$ . Similarly, for  $\mathbb{P}(y_t|\xi_t, \theta)$ , we have:

$$\begin{aligned}
\log[\mathbb{P}(y_t|\xi_t, \theta)] &= \frac{1}{2}\log|R^{-1}| - \frac{1}{2}(y_t - A - H\xi_t)'R^{-1}(y_t - A - H\xi_t) + C \\
&= \frac{1}{2}\log|R^{-1}| - \frac{1}{2}\{Trace[R^{-1}(y_t - A - H\xi_t)(y_t - A - H\xi_t)']\} + C \quad (2.12)
\end{aligned}$$

Note that we have incomplete observations, so we use cholesky decomposition.  
(Replace current formulation)

Before we take derivatives, I list here some useful formula:

$$\begin{aligned}
\frac{\partial B'AB}{\partial B} &= B'(A' + A) \\
\frac{\partial Trace(AB)}{\partial A} &= \frac{\partial Trace(BA)}{\partial A} = \frac{\partial Trace(B'A')}{\partial A} = B' \\
\frac{\partial \log(A)}{\partial A} &= A^{-1} \\
\frac{\partial \det(A)}{\partial A} &= \det(A)(A^{-1})'
\end{aligned}$$

Now we find  $\hat{\theta}$  by taking derivations of equation (2.10). First let's assume we observe  $\{\xi\}_t$ , we have:

$$\begin{aligned}
\log[\mathbb{P}(\xi, y|\theta)] = & \frac{T-1}{2} \log|Q^{-1}| + \frac{T}{2} \log|R^{-1}| \\
& - \frac{1}{2} \text{Trace} \left[ Q^{-1} \sum_1^{T-1} (\xi_t - F\xi_{t-1})' (\xi_t - F\xi_{t-1}) \right] \\
& - \frac{1}{2} \text{Trace} \left[ R^{-1} \sum_1^T (y_t - A - H\xi_t)(y_t - A - H\xi_t)' \right] \quad (2.13)
\end{aligned}$$

Take FOC on (2.13), we have:

$$\begin{aligned}
\frac{\partial \log[\mathbb{P}(\xi, y|\theta)]}{\partial F} &= Q^{-1} \left( \sum_1^T \xi_{t+1} \xi_t' - F \xi_{t+1} \xi_{t+1}' \right) \\
\frac{\partial \log[\mathbb{P}(\xi, y|\theta)]}{\partial H} &= R^{-1} \left( \sum_1^T y_t \xi_t' - A \xi_t - H \xi_t \xi_t' \right) \\
\frac{\partial \log[\mathbb{P}(\xi, y|\theta)]}{\partial A} &= R^{-1} \left( \sum_1^T y_t - A - H \xi_t \right) \\
\frac{\partial \log[\mathbb{P}(\xi, y|\theta)]}{\partial Q^{-1}} &= \frac{T-1}{2} Q - \frac{1}{2} \left( \sum_1^{T-1} (\xi_t - F\xi_{t-1})' (\xi_t - F\xi_{t-1}) \right) \\
\frac{\partial \log[\mathbb{P}(\xi, y|\theta)]}{\partial R^{-1}} &= \frac{T}{2} R - \frac{1}{2} \left( \sum_1^T (y_t - A - H\xi_t)(y_t - A - H\xi_t)' \right)
\end{aligned}$$

Set FOC to 0 and take conditional expectations of  $\xi$  with pdf  $f(\xi_t|y_1^T)$ , we have the update rules for parameters of interest:

$$F = \left[ \sum_1^T E(\xi_{t+1} \xi_t | y_1^T)' \right] \left[ \sum_1^T E(\xi_{t+1} \xi_{t+1} | y_1^T)' \right]^{-1} \quad (2.14)$$

$$H = \left[ \sum_1^T y_t E(\xi_t | y_1^T)' - \frac{1}{T} \left( \sum_1^T y_t \right) \left( \sum_1^T E(\xi_t | y_1^T) \right) \right] \left[ \sum_1^T E(\xi_t \xi_t | y_1^T)' \right]^{-1} \quad (2.15)$$

$$A = \quad (2.16)$$

$$Q = \quad (2.17)$$

$$R = \quad (2.18)$$

### 3 Implementation:

The Implementation is simple (we can use ) implement use `scipy.linalg.dtrtri` to solve inverse of triangular matrix. use cholesky decomposition to make sure inverse is still PSD potential next step is to use efficient smoothing. and make a general kalman filter. Note we might need to redo structure. As the F need to have eigenvalue less than 1. Need to check whether diff is necessary to get good estimates. This section should be completed last, as the model is still subject to change. Start with general case, and have a section about linear gaussian implementation. revise M-step formula, as it does not take into account of changing vec of updates, use cholesky decomposition to get sequential update