Lecture 19
Kalman Filter

Introduction

• We observe (measure) economic data, $\{z_t\}$, over time; but these measurements are noisy. There is an unobservable variable, y_t, that drives the observations. We call y_t the **state variable**.

• The **Kalman filter** (**KF**) uses the observed data to learn about the unobservable state variables, which describe the state of the model.

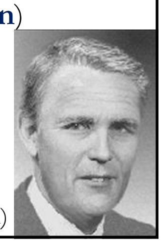
• KF models dynamically what we measure, z_t, and the state, y_t.

$y_t = g(y_{t-1}, u_t, w_t)$ (**state or transition equation**)

$z_t = f(\, y_t, x_t, v_t)$ (**measurement equation**)

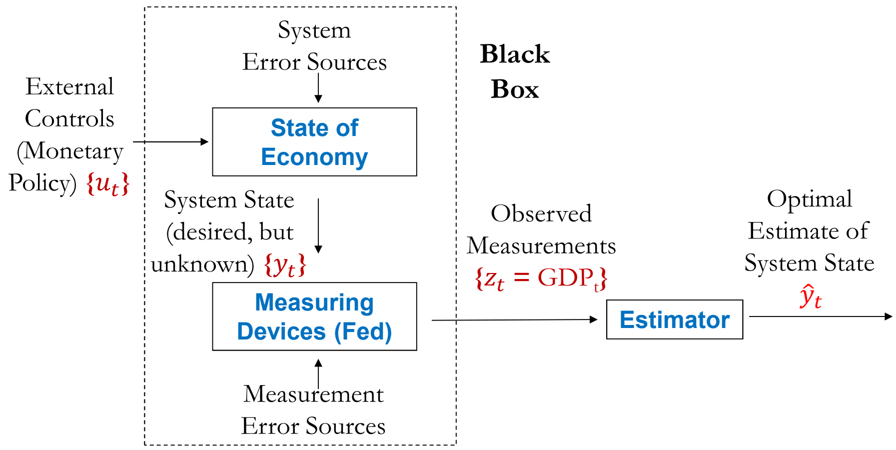
u_t, x_t: exogenous variables.

w_t, v_t: error terms.



Rudolf (Rudi) Emil Kalman (1930 – 2016, Hungary/USA)

Intuition: GDP and the State of the Economy



- The business cycle (the system state), y_t, cannot be measured directly
- We measure the system at time t: GDP (z_t).
- Need to estimate "optimally" the business cycle from GDP.

3

How does the KF work?

- It is a **recursive** data processing algorithm. As new information arrives, it updates predictions (it is a Bayesian algorithm).

- It generates **optimal** estimates of y_t, given measurements $\{z_t\}$.

- Optimal?
  - For linear system and white Gaussian errors, KF is "*best*" estimate based on all previous measurements (MSE sense).
  - For non-linear system optimality is '*qualified.*' (say, "*best linear*").

- Recursive?
  - No need to store all previous measurements and re-estimate system as new information arrives.

4

Notation

• For any vector s_t, we define the prediction of s_t at time t as:

$$s_{t|t-1} = \text{E}(s_t | I_{t-1})$$

That is, it is the **best guess** of s_t based on all the information available at time $t-1$, which we denote by $I_{t-1} = \{z_{t-1}, ..., z_1; u_{t-1}, ..., u_1; x_{t-1}, ..., x_1; ...\}$.

• As new information is released, we update our prediction:.

$$s_{t|t} = \text{E}[s_t | I_t]$$

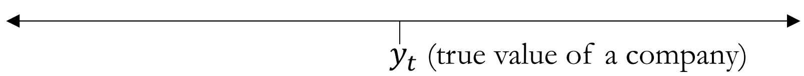
• The Kalman filter predicts $z_{t|t-1}$, $y_{t|t-1}$, and updates $y_{t|t}$.

At time t, we define a prediction error: $e_{t|t-1} = z_t - z_{t|t-1}$

The conditional variance of $e_{t|t-1}$: $F_{t|t-1} = \text{E}[e_{t|t-1}\, e_{t|t-1}']$

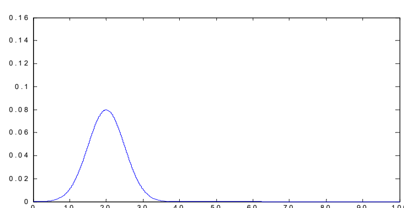
5

Intuitive Example: Prediction and Updating



• Distribution of true value, y_t, is unobservable

• Assume Gaussian distributed measurements

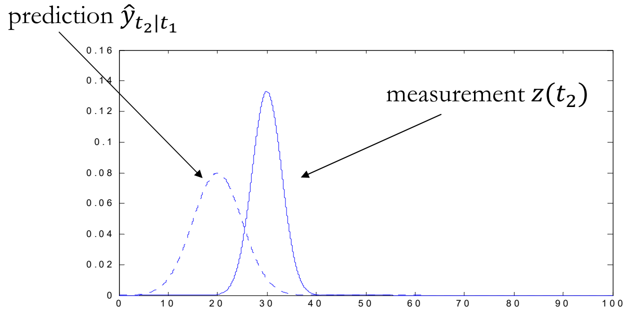


- Observed Measurement at t_1: Mean = z_1 & Variance = σ_{z_1}
- Optimal estimate of true value: $\hat{y}(t_1) = z_1$
- Variance of error in estimate: $\sigma_y^2(t_1) = \sigma_{z_1}^2$
- <u>Predicted</u> value at time t_2, using t_1 info: $\hat{y}_{t_2|t_1} = z_1$

6

Intuitive Example: Prediction and Updating

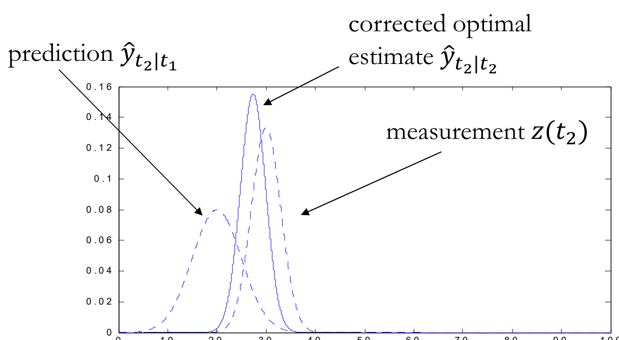
- We have the prediction $\hat{y}_{t_2|t_1}$. At t_2, we have a new measurement:



- New t_2 measurement:
  - Measurement at t_2: Mean = z_2 & Variance = σ_{z_2}
  - Update the prediction due to new measurement: $\hat{y}_{t_2|t_2}$
- Closer to more trusted measurement – linear interpolation?

7

Intuitive Example: Prediction and Updating



- Corrected mean is the new optimal estimate of true value

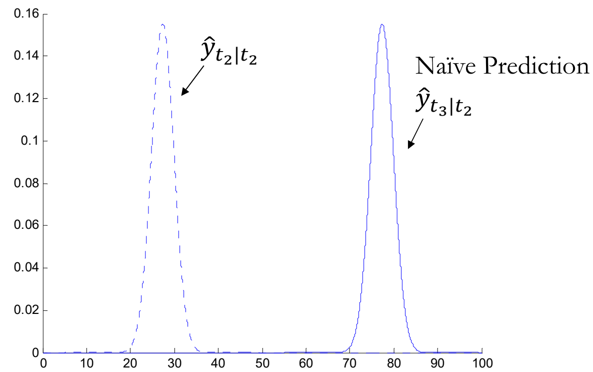
or Optimal (updated) estimate: $\hat{y}_{t|t} = \hat{y}_{t|t-1}$ + (Kalman Gain) * ($z_t - z_{t|t-1}$)

- New variance is smaller than either of the previous two variances

or Variance of estimate: Variance of prediction * (1 – Kalman Gain)

8

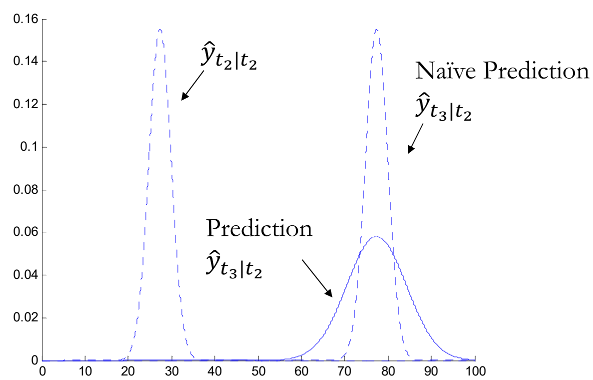
Intuitive Example: Prediction and Updating



- At time t_3, the true values changes at the rate $\frac{dy}{dt} = u$
- Naïve approach: Shift probability to the right to predict
- This would work if we knew the rate of change (perfect model). But, this is unrealistic.

9

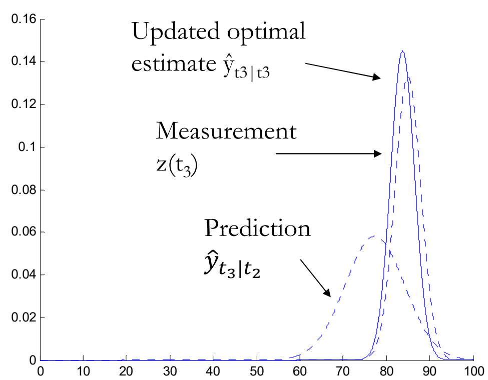
Intuitive Example: Prediction and Updating



- Then, we assume imperfect model by adding Gaussian noise, w.
- $\frac{dy}{dt} = u + w$
- Distribution for prediction moves and spreads out

10

Intuitive Example: Prediction and Updating



- Now we take a measurement at t_3
- Need to once again correct the prediction
- Same as before

11

Intuition: Prediction and Updating

• From the intuitive example, we see a simple process:

- **Initial conditions** ($\hat{y}_{t-1}$ & σ_{t-1})

- **Prediction** ($\hat{y}_{t|t-1}$, $\sigma_{t|t-1}$)
Use initial conditions and model (say, constant rate of change) to make prediction

- **Measurement** (z_t)
Take measurement, z_t and learn forecast error, $e_{t|t-1} = z_t - z_{t|t-1}$

- **Updating** ($\hat{y}_{t|t}$, $\sigma_{t|t}$)
Use measurement to update prediction by 'blending' prediction and residual – always a case of merging only two Gaussians

Optimal estimate with smaller variance

12

Terminology: Filtering and Smoothing

• **Prediction** is an *a priori* form of estimation. It attempts to provide information about what the quantity of interest will be at some time $t + \tau$ in the future by using data measured up to and including time $t - 1$ (usually, KF refers to one-step ahead prediction –i.e., $\tau = 1$).

• **Filtering** is an operation that involves the extraction of information about a quantity of interest at time t, by using data measured up to and including t.

• **Smoothing** is an *a posteriori* form of estimation. Data measured after the time of interest are used for the estimation. Specifically, the smoothed estimate at time t is obtained by using data measured over the interval $[0, T]$, where $t < T$.

13

Bayesian Optimal Filter

• A Bayesian optimal filter computes the distribution

$$\mathrm{P}(y_t \,|\, z_t, z_{t-1}, \ldots, z_1, y_t, \ldots, y_1) = \mathrm{P}(y_t \,|\, z_t)$$

• Given the following:

1. Prior distribution: $\mathrm{P}(y_0)$
2. State space model:

$$y_{t|t-1} \sim \mathrm{P}(y_t \,|\, y_{t-1})$$
$$z_{t|t-1} \sim \mathrm{P}(z_t \,|\, y_t)$$

3. Measurement sequence: $\{y_t\} = \{z_1, z_2, \ldots, z_t\}$

• Computation is based on recursion rule for incorporation of the new measurement y_k into the posterior:

$$\mathrm{P}(z_{t-1} \,|\, y_t, \ldots, y_1) \rightarrow \mathrm{P}(z_t \,|\, y_t, \ldots, y_1)$$

14

Bayesian Optimal Filter: Prediction Step

- Assume we know the posterior distribution of previous time step, $t-1$:

$$P(y_{t-1} | z_{t-1}, \dots, z_1)$$

- The joint pdf $P(y_t, y_{t-1} | \{z_{t-1}\})$ can be computed as (using the Markov property):

$$P(y_t, y_{t-1} | \{z_{t-1}\}) = P(y_t | y_{t-1}, \{z_{t-1}\}) * P(y_{t-1} | \{z_{t-1}\})$$
$$= P(y_t | y_{t-1}) * P(y_{t-1} | \{z_{t-1}\})$$

- Integrating over y_{t-1} gives the **Chapman-Kolmogorov equation**:

$$P(y_t | \{z_{t-1}\}) = \int P(y_t | y_{t-1}) * P(y_{t-1} | \{z_{t-1}\})\, dy_t$$

- This is the *prediction step* of the optimal filter.

15

Bayesian Optimal Filter: Update Step

- Now we have:
  1. Prior distribution from the Chapman-Kolmogorov equation

     $P(y_t | \{z_{t-1}\})$

  2. Measurement likelihood:

     $P(z_t | y_t)$

- The posterior distribution (= Likelihood x Prior):

$$P(y_t | \{z_{t-1}\}) \propto P(z_t | y_t) * P(y_t | \{z_{t-1}\})$$

(ignoring *normalizing constant,*

$$P(z_t | \{z_{t-1}\}) = \int P(z_t | y_t) * P(y_t | \{z_{t-1}\})\, dy_t).$$

- This is the **update step** of the optimal filter.

16

State Space Form

- Model to be estimated:

$$y_t = \boldsymbol{A}\, y_{t-1} + \mathbf{B}\, u_t + w_t$$

w_t: *state noise* ~ WN(0, **Q**)
u_t: exogenous variable.
$\boldsymbol{A}$: *state transition matrix*
B: coefficient matrix for u_t.

$$z_t = \boldsymbol{H}\, y_t + v_t$$

v_t: *measurement noise* ~ WN(0, **R**)
$\boldsymbol{H}$: *measurement matrix*

Initial conditions: y_0, usually a RV.

We call both equations **state space form**. Many economic models can be written in this form.

Note: The model is linear, with constant coefficient matrices, $\boldsymbol{A}$, **B**, and $\boldsymbol{H}$. It can be generalized –see Harvey(1989).

17

State Space Form: Examples

Example: We can write a VAR(1) in state space form:

$$\boldsymbol{Y}_t = \boldsymbol{A}\boldsymbol{Y}_{t-1} + \boldsymbol{w}_t$$

$$\boldsymbol{Y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{21} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix}$$ *Transition equation*

$$\boldsymbol{z}_t = \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \boldsymbol{v}_t$$ *Measurement equation*

18

State Space Form: Examples

Example: VAR(p) written in state-space form.

$$\boldsymbol{Y}_t = \boldsymbol{\Phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{Y}_{t-2} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{Y}_{t-p} + \boldsymbol{w}_t$$

$$\eta_t \equiv \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix} \qquad F \equiv \begin{bmatrix} \Phi_1 \; \Phi_2 \ldots\ldots \Phi_{p-1} \; \Phi_p \\ I_n \quad 0 \ldots\ldots\ldots\ldots 0 \\ 0 \quad I_n \ldots\ldots\ldots\ldots 0 \\ \vdots \\ 0 \quad 0 \ldots\ldots\ldots I_n \quad 0 \end{bmatrix} \qquad w_t = \begin{bmatrix} a_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then, let $\boldsymbol{A} = \boldsymbol{F}$, we write the *transition equation* as:

$$\eta_t = \boldsymbol{A}\, \eta_{t-1} + w_t$$

and letting $\boldsymbol{H} = \boldsymbol{I}$ the *measurement equation* as:

$$z_t = \boldsymbol{H}\, \eta_t$$

19

State Space Form: Examples

Example: In a linear model, we allow for time-varying coefficients.

$$z_t = \alpha_t + \beta_t u_t + v_t \qquad \sim N(0, V_t)$$
$$\alpha_t = \alpha_{t-1} + w_{\alpha,t} \qquad \sim N(0, W_{\alpha,t})$$
$$\beta_t = \beta_{t-1} + w_{\beta,t} \qquad \sim N(0, W_{\beta,t})$$

Define $y_t = (\alpha_t, \beta_t)'$ & $\boldsymbol{H}_t = [1, u_t]$, then, measurement equation is:

$$z_t = \boldsymbol{H}_t\, y_t + v_t$$

and let $\boldsymbol{A} = \mathbf{I}$, then, transition equation is:

$$y_t = \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_{\alpha,t} \\ w_{\beta,t} \end{bmatrix}$$

Example: Stochastic volatility.

$z_t = \boldsymbol{H}\, h_t + \mathbf{C}\, x_t + v_t \qquad v_t \sim \text{WN}(0, \mathbf{R}) \quad$ (*Measurement equation*)

$h_t = \boldsymbol{A}\, h_{t-1} + \mathbf{B}\, u_t + w_t \qquad w_t \sim \text{WN}(0, \mathbf{Q}) \quad$ (*Transition equation*)

20

State Space Form: Variations

• When $\{w_t\}$ is WN and independent of y_0, $\{y_t\}$ is **Markov**.
- Linearity is not essential for this property.
- $\{z_t\}$ is usually not Markov. z_t is conditionally independent given y_t:

$$P(z_t \,|\, z_t, z_{t-1}, ..., z_1, y_t, ..., y_1) = P(\, z_t | y_t)$$

• If, in addition, $\{w_t, v_t, y_0\}$ are jointly normal, the model is called **Gauss-Markov Model.**

• Usually, we assume that only the 1st and 2nd order statistics of $\{w_t\}$, $\{v_t\}$ are known. Under the following assumptions, we have the **Standard second-order model**:
- $\{w_t\}$ moments: $E[w_t] = 0$; $Cov[w_k\, w_l] = Q_k\, \delta_{k,l}$
- $\{v_t\}$ moments: $E[v_t] = 0$; $Cov[v_k\, v_l] = R_k\, \delta_{k,l}$
- $Cov[w_t, v_t] = 0$
- y_0 independent of $\{w_t\}$ & $\{v_t\}$; with $E[y_0] = y_{0|0}$; & $Cov(y_0) = P_{0|0}$.

21

Mean and Variance of State Vector y_t

• y_t is a random variable following: $y_t = \boldsymbol{A}y_{t-1} + \mathbf{B}u_t + w_t$
- may be unobservable and, thus, we have no data for y_t.
- it is normally distributed; a sum of normal variables, $w_t \sim N(0, \mathbf{Q})$:

$$P(y_t | I_{t-1}) = N(E[y_t | I_{t-1}], \mathrm{Var}[y_t | I_{t-1}])$$

• Conditional Mean: $\quad E[y_t | I_{t-1}] = y_{t|t-1} = \boldsymbol{A}y_{t-1|t-1} + \mathbf{B}u_{t|t-1}$

In an AR(1) model: $\quad E[y_t | I_{t-1}] = \mu + \phi\, y_{t-1|t-1}$.

• Conditional Variance: $\mathrm{Var}[y_t | I_{t-1}] = \boldsymbol{P}_{t|t-1} = \boldsymbol{A}\, \boldsymbol{P}_{t-1|t-1} \boldsymbol{A}^T + \mathbf{Q}$

<u>Note</u>: There are 2 source of noise:

1) w_t

2) Difference between y_{t-1} & $y_{t|t-1}$ may not be zero.

• $Cov(y_{t-1}, w_t) = 0$.

22

Mean and Variance of z_t & Joint (y_t, z_t)

• z_t is a random variable following: $z_t = \boldsymbol{H}\, y_t + v_t$.
- it is normally distributed; a sum of normal variables, $v_t \sim$ N(0, **R**):

$$\mathrm{P}(z_t|I_{t-1}) = \mathrm{N}(\mathrm{E}[z_t|I_{t-1}], \mathrm{Var}[z_t|I_{t-1}])$$

• Conditional Mean: $\mathrm{E}[z_t|I_{t-1}] = z_{t|t-1} = \boldsymbol{H}\, y_{t|t-1}$

• Conditional Variance: $\mathrm{Var}[z_t|I_{t-1}] = \boldsymbol{H}\, \boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}} + \mathbf{R}$

Note: $\mathrm{Cov}(y_{t-1}, v_t)$=0 (since $\mathrm{E}[w_t, v_t]$=0).

• Covariance between z_t & y_t: $\mathrm{Cov}[z_t, y_t|I_{t-1}] = \boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}$

• Joint pdf of $\mathrm{P}(z_t, y_t|I_{t-1})$:

$$\begin{pmatrix} y_t \mid I_{t-1} \\ z_t \mid I_{t-1} \end{pmatrix} \sim N\left(\begin{bmatrix} Ay_{t-1|t-1} + Bu_{t|t-1} \\ Hy_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}H' \\ P_{t|t-1}H' & HP_{t-1|t-1}H' + R \end{bmatrix} \right)$$

Kalman Filter

• Given $y_{0|0}$ & $\boldsymbol{P}_{0|0}$ (initialization), the Kalman Filter solves the following equations for $t = 1, \dots, T$.

Prediction: $y_{t|t-1}$ is estimate based on measurements at previous t:

$$y_{t|t-1} = \boldsymbol{A}\, y_{t-1|t-1} + \mathbf{B}\, u_t$$
$$\boldsymbol{P}_{t|t-1} = \boldsymbol{A}\, \boldsymbol{P}_{t-1}\boldsymbol{A}^{\mathrm{T}} + \mathbf{Q}$$

Update: y_t has additional information – the measurement at time t:

$$y_{t|t} = y_{t|t-1} + \boldsymbol{K}_t\, (z_t - \boldsymbol{H}\, y_{t|t-1})$$
$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t\, \mathrm{H}\, \boldsymbol{P}_{t|t-1} = (\boldsymbol{I} - \boldsymbol{K}_t\, \boldsymbol{H})\, \boldsymbol{P}_{t|t-1}$$
$$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\, \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\, \boldsymbol{P}_{t|t-1}\, \mathrm{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$ ("**Kalman gain**")

• The forecast error is: $e_{t|t-1} = z_t - z_{t|t-1} = z_t - \boldsymbol{H}\, y_{t|t-1}$

• The variance of $e_{t|t-1}$: $\boldsymbol{F}_{t|t-1} = \boldsymbol{H}\, \boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}} + \mathbf{R}$

Kalman Filter

- Recall: $e_{t|t-1} = z_t - z_{t|t-1} = z_t - \boldsymbol{H}\, y_{t|t-1}$ (forecast error)
  $\boldsymbol{F}_{t|t-1} = \boldsymbol{H}\, \boldsymbol{P}_{t|t-1} \boldsymbol{H}^{\mathrm{T}} + \mathbf{R}$ (variance of $e_{t|t-1}$)

- The initial values, $y_{0|0}$ & $\boldsymbol{P}_{0|0}$, are set to unconditional mean and variance, and reflect prior beliefs about the distribution of y_t.

- The update of the state variable, $y_{t|t}$, and its variance, $\boldsymbol{P}_{t|t}$, are linear combinations of previous guess and forecast error:

$$y_{t|t} = y_{t|t-1} + \boldsymbol{K}_t\, (z_t - \boldsymbol{H}\, y_{t|t-1}) = y_{t|t-1} + \boldsymbol{K}_t e_{t|t-1}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t\, \boldsymbol{H}\, \boldsymbol{P}_{t|t-1} \qquad \text{(conditional variance)}$$

$$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\, \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\, \boldsymbol{P}_{t|t-1}\, \boldsymbol{H}^{\mathrm{T}} + \mathbf{R})^{-1} = \mathrm{Cov}[z_t, y_t | I_{t-1}]\, (\boldsymbol{F}_{t|t-1})^{-1}$$

Since we observe z_t, the uncertainty (measured by $\boldsymbol{P}_{t|t}$) declines.

Note: The bigger $\boldsymbol{F}_{t|t-1}$, the smaller $\boldsymbol{K}_t$ & less weight put to updating.

Kalman Gain

- $\boldsymbol{K}_t$ depends on the relationship between (z_t & y_t) and $\boldsymbol{F}_{t|t-1}$:

$$\boldsymbol{K}_t = \mathrm{Cov}[z_t, y_t | I_{t-1}]\, (\boldsymbol{F}_{t|t-1})^{-1}$$

- The stronger $\mathrm{Cov}[z_t, y_t | I_{t-1}]$, the more relevant $\boldsymbol{K}_t$ in the update.

- If the relationship is weaker, we do not put much weight as it is likely not driven by y_t.

- If we are sure about measurements, **R** decreases to zero and, thus, $\boldsymbol{F}_{t|t-1}$ decreases. Then, $\boldsymbol{K}_t$ increases and we weight residuals more heavily in the update than prediction.

- If the model is time-invariant (**Q** and **R** are, in fact, constant) the Kalman gain quickly converges to a constant: $\boldsymbol{K}_t \rightarrow \mathbf{K}$. In this case, the filter becomes stationary.

Estimation: MLE

• Once we have $e_{t|t-1}$ & $\boldsymbol{F}_{t|t-1}$, under normality, we can do MLE:

$$\ell = -\frac{Tn}{2}\ln(2\pi) + \frac{T}{2}\sum_{t=p+1}^{T}\ln\left|F_t^{-1}\right| - \frac{1}{2}e'_t\, F_t^{-1}e_t + \ln(f(z_p, z_{p-1}, ..z_1))$$

Note: Under normality, the KF is optimal in an MSE sense.

• Algorithm (after initialization & ignoring first observations)

```
for (i in 1:T) {
    zhat = t(H) %*% state10
    zvar = (t(H )%*% P10 %*% H + R
    zvarinv = solve(zvar)
    eps = t(z[i,]) - zhat
    f0 = f0 - log(det(zvar)) - t(eps) %*% zvarinv %*% eps
    state11 = state10 + P10 %*% H %*% zvarinv %*% eps
    P11 = P10 - P10 %*% H %*% zvarinv %*% t(H) %*% P10
    state10 = A %*% state11
    P10 = A %*% P11 %*% t(A) + Q
  }
f0 = -(T*n/2) * log(2*pi)  + f0/2
```

27

Derivation of Update

• We wrote the joint of $(z_t, y_t)|I_{t-1}$. But, we could have also written the joint of $(e_t, y_t)|I_{t-1}$:

$$\begin{pmatrix} y_t \mid I_{t-1} \\ e_t \mid I_{t-1} \end{pmatrix} \sim N\left(\begin{bmatrix} y_{t|t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}H' \\ HP_{t|t-1} & F_{t|t-1} \end{bmatrix}\right)$$

• Recall a property of the multivariate normal distribution:
If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are jointly normally distributed, the conditional distribution of $\boldsymbol{x}_1|\boldsymbol{x}_2$ is also normal with mean $\mu_{1|2}$ & variance $\Sigma_{1|2}$:

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Then, $$y_{t|t} = y_{t|t-1} + \boldsymbol{P}_{t|t-1}\,\mathbf{H}^{\mathrm{T}}\,((\boldsymbol{F}_{t|t-1})^{-1}\,e_{t|t-1}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\,\mathbf{H}^{\mathrm{T}}(\boldsymbol{F}_{t|t-1})^{-1}\,\mathbf{H}\,\boldsymbol{P}_{t|t-1}$$

28

Feedback

Prediction (Time Update)

(1) Project the state ahead

$y_{t|t-1} = \boldsymbol{A}\, y_{t-1|t-1} + \mathbf{B}\, u_{t|t-1}$

(2) Project the error Cov ahead

$\boldsymbol{P}_{t|t-1} = \boldsymbol{A}\, \boldsymbol{P}_{t-1} \boldsymbol{A}^{\mathrm{T}} + \mathbf{Q}$

Correction (Measurement Update)

(1) Compute the Kalman Gain

$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\, \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\, \boldsymbol{P}_{t|t-1}\, \boldsymbol{H}^{\mathrm{T}} + \mathbf{R})^{-1}$

(2) Update $y_{t|t-1}$ with measurement z_t

$y_{t|t} = y_{t|t-1} + \boldsymbol{K}_t\, (z_t - \boldsymbol{H}\, y_{t|t-1})$

(3) Update Error Covariance

$\boldsymbol{P}_{t|t} = (\boldsymbol{I} - \boldsymbol{K}_t\, \boldsymbol{H})\, \boldsymbol{P}_{t|t-1}$

29

Kalman Filter: Remarks

• KF works by minimizing E$[(y_t - y_{t|t-1})'\,(y_t - y_{t|t-1})]$. Under this metric, the expected value is the optimal estimator.

• Conditions for optimality (MSE) –Anderson and Moore (1979):
- The DGP (linear system & its state space model) is exactly known.
- Noise vector (w_t, v_t) is white noise.
- The noise covariances are known.

• KF is an MSE estimator among all linear estimators, but in the case of a Gaussian model it is the MSE estimator among *all* estimators.

• In practice, difficult to meet the three conditions. A lot of tuning by ad-hoc methods, to get KF that work "sufficiently well."

30

Kalman Filter: Practical Considerations

• Estimating **Q** and **R** usually involves NW-style SE, based on autocovariances. Estimating **Q** is, in general, complicated. Tuning **Q** and **R** (sometimes called **system identification**) is usual to improve performance.

• The inversion of $\boldsymbol{F}_{t|t-1}$ can be difficult. Usually, the problem comes from having **Q** singular. In practice, approximations (pseudo-inversion) are used.

• When knowledge/confidence about y_0 is low, a diffuse prior for y_0 would set $\boldsymbol{P}_{0|0}$ high.

• The noise vector (w_t, v_t) should be WN. Autocorrelograms and LB tests can be used to check this.

Kalman Filter: Problems and Variations

• When the model –i.e., $g(.)$ and $f(.)$– is non-linear, the **extended Kalman filter** (EKF) works by linearizing the model (similar to NLLS, **A** & **H** are Jacobian matrices of partial derivatives.).

Problem: The distributions of the RVs are no longer normal after their respective nonlinear transformations. EKF is an ad-hoc method.

• When the model is highly non-linear, EKF will not work well. The **unscented Kalman filter** (UKF), which uses MC method to calculate the updates, works better.

• The KF struggles under non-normality and when the dimensions of the state vector increase. **Particle filters** (sequential Monte Carlo), which is another Bayesian filter, are very popular in these cases.

Kalman Filter: Smoothing

- We can use all the data to re-estimate our prediction. That is, $y_{t|T}$.

- Inputs: Initial distribution y_0 and data z_t, z_{t-1}, ..., z_1

- Algorithm: Forward-backward pass (**Rauch-Tung-Striebel algorithm**)

- **Forward pass**:
  Kalman filter: Compute $y_{t+1|t}$ and $y_{t+1|t+1}$ for $0 \leq t < T$

- **Backward pass**:
  Compute $y_{t|T}$ for $0 \leq t < T$
  Reverse process in our intuitive example.

33

Kalman Filter: Smoothing – Backward Pass

- Compute $y_{t|T}$ given $y_{t+1|T} \sim \mathrm{N}(y_{t+1|T}, \boldsymbol{P}_{t+1|T})$

- Reverse movement from filter: $X_{t|t} \rightarrow X_{t+1|t}$.

- Same as incorporating measurement in filter
  1. Compute joint pdf of $(y_{t|t}, y_{t+1|t})$
  2. Compute conditional distribution $(y_{t|t} | \, y_{t+1|t} = y_{t+1})$

- But: y_{t+1} is not "known", we only know its distribution:
  3. "Uncondition" on y_{t+1} to compute $y_{t|T}$ using laws of total expectation and variance

34

Kalman Filter: Smoothing – Backward Pass

• **Step 1**. Compute joint pdf of $(y_{t|t}, y_{t+1|t})$

$$\begin{pmatrix} y_{t|t} \\ y_{t+1|t} \end{pmatrix} \sim N\left(\begin{pmatrix} y_{t|t} \\ y_{t+1|t} \end{pmatrix}, \begin{pmatrix} \text{Var}(y_{t|t}) & \text{Cov}(y_{t|t}, y_{t+1|t}) \\ \text{Cov}(y_{t+1|t}, y_{t|t}) & \text{Var}(y_{t+1|t}) \end{pmatrix}\right)$$

$$\sim N\left(\begin{pmatrix} y_{t|t} \\ y_{t+1|t} \end{pmatrix}, \begin{pmatrix} P_{t|t} & P_{t|t}A^T \\ AP_{t|t} & P_{t+1|t} \end{pmatrix}\right)$$

• **Step 2**. Compute conditional pdf of $(y_{t|t} |\, y_{t+1|t} = y_{t+1})$

$$(y_{t|t} \mid y_{t+1|t} = y_{t+1}) = N\left(y_{t|t} + P_{t|t}A^T P_{t+1|t}{}^{-1}(y_{t+1} - y_{t+1|t}), P_{t|t} - P_{t|t}A^T P_{t+1|t}{}^{-1} AP_{t|t}\right)$$

where we used the conditional result: $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}{}^{-1}(x_2 - \mu_2)$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{22}$$

35

Kalman Filter: Smoothing – Backward Pass

• **Step 3**. "Uncondition" on y_{t+1} to compute $y_{t|T}$. We do not know its value, but only its distribution: $y_{t+1} \sim N()$.

• Uncondition on y_{t+1} to compute $y_{t|T}$ using the Law of total expectation and the Law of total variance:

Law of total expectation:

$$E[X] = E_Z[E[X|Y = Z]]$$
$$E[y_{t|T}] = E_{y_{t+1|T}}[E[y_{t|t}|y_{t+1|t} = y_{t+1|T}]]$$

Law of total variance:

$$\text{Var}(X) = E_Z[\text{Var}(X|Y = Z)] + \text{Var}_Z(E[X|Y = Z])$$
$$\text{Var}(y_{t|T}) = E_{y_{t+1|T}}[\text{Var}(y_{t|t} \mid y_{t+1|t} = y_{t+1|T})] + \text{Var}_{y_{t+1|T}}\left(E[y_{t|t} \mid y_{t+1|t} = y_{t+1|T}]\right)$$

36

Kalman Filter: Smoothing – Backward Pass

- **Step 3 (continuation).** From Step 2 we know:

$$E(y_{t|t} \mid y_{t+1|t} = y_{t+1|T}) = y_{t|t} + L_t\left(y_{t+1|T} - y_{t+1|t}\right)$$

$$\text{Var}(y_{t|t} \mid y_{t+1|t} = y_{t+1|T}) = P_{t|t} - L_t P_{t+1|t} L_t^T$$

Then,

$$\begin{aligned} E(y_{t|T}) &= E_{y_{t+1|T}}\left(E(y_{t|t} \mid y_{t+1|t} = y_{t+1|T})\right) \\ &= y_{t|t} + L_t\left(y_{t+1|T} - y_{t+1|t}\right) \end{aligned}$$

$$\begin{aligned} \text{Var}(y_{t|T}) &= E_{y_{t+1|T}}\left(\text{Var}(y_{t|t} \mid y_{t+1|t} = y_{t+1|T})\right) + \\ &\quad \text{Var}_{y_{t+1|T}}\left(E(y_{t|t} \mid y_{t+1|t} = y_{t+1|T})\right) \\ &= P_{t|t} - L_t P_{t+1|t} L_t^T + L_t P_{t+1|T} L_t^T \\ &= P_{t|t} + L_t (P_{t+1|T} - P_{t+1|t}) L_t^T \end{aligned}$$

Kalman Filter: Smoothing – Backward Pass

- Summary:

$$\begin{aligned} L_t &= P_{t|t} A^T P_{t+1|t}^{-1} \\ y_{t|T} &= y_{t|t} + L_t\left(y_{t+1|T} - y_{t+1|t}\right) \\ P_{t|T} &= P_{t|t} + L_t (P_{t+1|T} - P_{t+1|t}) L_t^T \end{aligned}$$

- Algorithm (after initialization):

```
for (it in 1:T-1) {
    P0T = Reshape(Psmo[T-it+1,],rx,rx)
    state0T = statesmo[T-it+1,]
    Pfilt = Reshape(P[T-it,],rx,rx)              # this is P t given t
    Pnext = A %*% Pfilt %*% t(A) + Q             # this is P t+1 given t
    Lt = Pfilt %*% t(A) %*%inv(Pnext)
    P0T = Pfilt + Lt %*% (P0T - Pnext) %*%t(Lt)
    state0T = t(state[T-it,]) + Lt %*% (state0T - A %*% t(state[T-it,]))
    statesmo[T-it,] = t(state0T)
    Psmo[T-it,] = t(vec(P0T))
}
```

Kalman Filter: Smoothing – Algorithm

- for ($t = 0$; $t < T$; $++\ t$) (**Kalman filter**)

$$
\begin{aligned}
y_{t+1|t} &= A y_{t|t} \\
P_{t+1|t} &= A P_{t|t} A^T + Q \\
K_{t+1} &= P_{t+1|t} H^T \left(H P_{t+1|t} H^T + R \right)^{-1} \\
y_{t+1|t+1} &= y_{t+1|t} + K_{t+1} \left(z_{t+1} - H y_{t+1|t} \right) \\
P_{t+1|t+1} &= P_{t+1|t} - K_{t+1} H P_{t+1|t}
\end{aligned}
$$

- for ($t = T - 1$; $t \geq 0$; -- t) (**Backward pass**)

$$
\begin{aligned}
L_t &= P_{t|t} A^T P_{t+1|t}^{-1} \\
y_{t|T} &= y_{t|t} + L_t \left(y_{t+1|T} - y_{t+1|t} \right) \\
P_{t|T} &= P_{t|t} + L_t (P_{t+1|T} - P_{t+1|t}) L_t^T
\end{aligned}
$$

Kalman Filter: Smoothing - Remarks

- Kalman smoother is a post-processing method.
- Use $y_{t|T}$'s as optimal estimate of state at time t, and use $\boldsymbol{P}_{t|T}$ as a measure of uncertainty.
- The smoothing recursion consists of the backward recursion that uses the filtered values of y and $\boldsymbol{P}$.