

# Lecture 13

## State-Space Models I

Readings: SS17 Chapter 6.1-6.2; BD Chapter 9.1-9.3

*MATH 8090 Time Series Analysis*  
Week 13

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

Whitney Huang  
Clemson University

Background

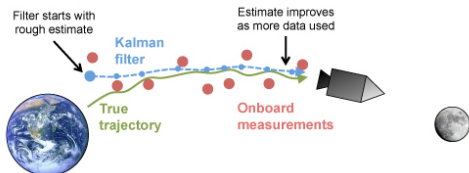
Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

- 1 Background
- 2 Multivariate Gaussian and Regression Lemmas
- 3 Forecasting, Filtering, and Smoothing

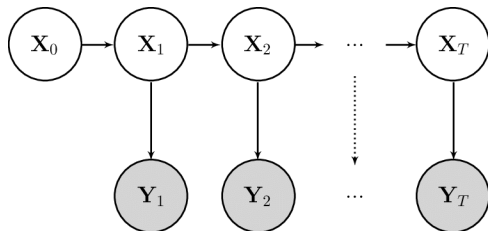
## Historical Background

- The original model emerged in the context of space tracking [Kalman, 1960, Kalman and Bucy, 1961]
- The “state equation” defines the motion equations for the position of a spacecraft with location  $x_t$



- The data  $y_t$  reflect information that can be observed from a tracking device, such as velocity and azimuth

The main goal was to retrieve the underlying state  $\{x_t\}$  based on observed data  $\{y_t\}$



## Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

State:  $\mathbf{X}_t = \mathbf{M}_t \mathbf{X}_{t-1} + \mathbf{V}_t$ ,  $\mathbf{V}_t \stackrel{i.i.d.}{\sim} \text{WN}(\mathbf{0}, \mathbf{Q}_t)$ ,  $t = 1, 2, \dots$

Observation:  $\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \mathbf{W}_t$ ,  $\mathbf{W}_t \stackrel{i.i.d.}{\sim} \text{WN}(\mathbf{0}, \mathbf{R}_t)$ ,  $t = 1, 2, \dots$

- $\mathbf{X}_t \in \mathbb{R}^p$  and  $\mathbf{Y}_t \in \mathbb{R}^q$  are the **state vector** and the **observation vector** at time  $t$
- $\mathbf{M}_t$  is the  $p \times p$  **transition matrix**, and  $\mathbf{H}_t$  is the  $q \times p$  **observation matrix**
- $\mathbf{V}_t$  and  $\mathbf{W}_t$  are the state and observation noises

State equation:

$$\mathbf{X}_t = \mathbf{M}_t \mathbf{X}_{t-1} + \mathbf{V}_t, \quad t = 1, 2, \dots$$

Observation equation:

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \mathbf{W}_t, \quad t = 1, 2, \dots$$

- $E(\mathbf{W}_s \mathbf{V}_t^T) = 0$  for all  $s$  and  $t$ , that is, **every observation noise is uncorrelated with every state-transition noise**

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

State equation:

$$\mathbf{X}_t = \mathbf{M}_t \mathbf{X}_{t-1} + \mathbf{V}_t, \quad t = 1, 2, \dots$$

Observation equation:

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \mathbf{W}_t, \quad t = 1, 2, \dots$$

- $E(\mathbf{W}_s \mathbf{V}_t^T) = 0$  for all  $s$  and  $t$ , that is, every observation noise is uncorrelated with every state-transition noise
- Assuming  $E(\mathbf{X}_0) = \boldsymbol{\mu}_0$ ,  $E(\mathbf{X}_0 \mathbf{W}_t^T) = 0$  and  $E(\mathbf{X}_0 \mathbf{V}_t^T) = 0$  for all  $t$ , that is, initial state vector are uncorrelated with both observation and state transition noises

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)



- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The **Kalman recursions** for state-space models provide elegant solution for **forecasting**, **filtering**, and **smoothing**

To estimate  $X_t$  with  $\mathbf{Y}_{1:s} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_s\}$ :

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The Kalman recursions for state-space models provide elegant solution for forecasting, filtering, and smoothing

To estimate  $X_t$  with  $\mathbf{Y}_{1:s} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_s\}$ :

- When  $s < t \Rightarrow$  forecasting

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The **Kalman recursions** for state-space models provide elegant solution for **forecasting**, **filtering**, and **smoothing**

To estimate  $X_t$  with  $\mathbf{Y}_{1:s} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_s\}$ :

- When  $s < t \Rightarrow$  **forecasting**
- When  $s = t \Rightarrow$  **filtering**

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The **Kalman recursions** for state-space models provide elegant solution for **forecasting**, **filtering**, and **smoothing**

To estimate  $X_t$  with  $\mathbf{Y}_{1:s} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_s\}$ :

- When  $s < t \Rightarrow$  **forecasting**
- When  $s = t \Rightarrow$  **filtering**
- When  $s > t \Rightarrow$  **smoothing**

- State-space models, defined through two seemingly simple equations, constitute a rich class of processes that have proven effective as models for time series
  - (S)ARIMA(X)
  - Hidden Markov Models (HMMs)
  - Vector Autoregression (VAR)
- The **Kalman recursions** for state-space models provide elegant solution for **forecasting**, **filtering**, and **smoothing**

To estimate  $X_t$  with  $\mathbf{Y}_{1:s} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_s\}$ :

- When  $s < t \Rightarrow$  **forecasting**
- When  $s = t \Rightarrow$  **filtering**
- When  $s > t \Rightarrow$  **smoothing**
- State-space models and Kalman recursions can be readily adapted to handle time series with **missing values**

## AR(1) Process as a State-Space Model: I

- State-transition equation

$$\mathbf{X}_t = M_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  and  $|\phi| < 1$



## AR(1) Process as a State-Space Model: I

- State-transition equation

$$\mathbf{X}_t = M_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  and  $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

# AR(1) Process as a State-Space Model: I

- State-transition equation

$$\mathbf{X}_t = M_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  and  $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting
  - $\mathbf{X}_t = Y_t; M_t = \phi$

# AR(1) Process as a State-Space Model: I

- State-transition equation

$$\mathbf{X}_t = M_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  and  $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting
  - $\mathbf{X}_t = Y_t$ ;  $M_t = \phi$
  - $\mathbf{V}_t = Z_t$  along with  $Q_t \stackrel{\text{def}}{=} \text{E}(\mathbf{V}_t \mathbf{V}_t^T) = \text{E}(Z_t^2) = \sigma_Z^2$

# AR(1) Process as a State-Space Model: I

- State-transition equation

$$\mathbf{X}_t = M_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  and  $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting
  - $\mathbf{X}_t = Y_t$ ;  $M_t = \phi$
  - $\mathbf{V}_t = Z_t$  along with  $Q_t \stackrel{\text{def}}{=} \text{E}(\mathbf{V}_t \mathbf{V}_t^T) = \text{E}(Z_t^2) = \sigma_Z^2$

## AR(1) Process as a State-Space Model: I

- State-transition equation

$$\mathbf{X}_t = M_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

is reminiscent of a causal AR(1) model:

$$Y_t = \phi Y_{t-1} + Z_t,$$

with  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  and  $|\phi| < 1$

- AR(1) can be expressed in state-space formulation by setting

- $\mathbf{X}_t = Y_t; M_t = \phi$
- $\mathbf{V}_t = Z_t$  along with  $Q_t \stackrel{\text{def}}{=} \text{E}(\mathbf{V}_t \mathbf{V}_t^T) = \text{E}(Z_t^2) = \sigma_Z^2$

and by using a **degenerate form of the observation equation**:  $\mathbf{Y}_t = H_t \mathbf{X}_t + \mathbf{W}_t$  in which  $H_t = 1$  and  $\mathbf{W}_t = 0$  so that  $\mathbf{Y}_t = X_t$

Need to define the initial state  $X_0$  in order to complete the model:

- A natural choice is

$$X_0 = \sum_{j=1}^{\infty} \phi^j Z_{1-j}, \quad \text{for which } \text{Var}(X_0) = \frac{\sigma^2}{1 - \phi^2}$$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

Need to define the initial state  $X_0$  in order to complete the model:

- A natural choice is

$$X_0 = \sum_{j=1}^{\infty} \phi^j Z_{1-j}, \quad \text{for which } \text{Var}(X_0) = \frac{\sigma^2}{1 - \phi^2}$$

- With this choice, the required conditions, namely,  $E(X_0 \mathbf{W}_t^T) = 0$  and  $E(X_0 \mathbf{V}_t^T) = 0$  hold

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

Need to define the initial state  $X_0$  in order to complete the model:

- A natural choice is

$$X_0 = \sum_{j=1}^{\infty} \phi^j Z_{1-j}, \quad \text{for which } \text{Var}(X_0) = \frac{\sigma^2}{1 - \phi^2}$$

- With this choice, the required conditions, namely,  $E(X_0 \mathbf{W}_t^T) = 0$  and  $E(X_0 \mathbf{V}_t^T) = 0$  hold
- Could also set  $X_0 = Z_0 \frac{\sigma}{\sqrt{1-\phi^2}}$  to get a AR(1) process, but using  $X_0 = Z_0$  would lead to a valid state-space model that is **not** a true AR(1) model



AR(1) process with  $0 < \phi < 1$  is known as “red noise”, red noise is related to a 1st order stochastic differential equation, rendering it a model for various geophysical processes:

- Typically only observe red noise process of interest in presence of observational noise (often taken to be white noise)

AR(1) process with  $0 < \phi < 1$  is known as “red noise”, red noise is related to a 1st order stochastic differential equation, rendering it a model for various geophysical processes:

- Typically only observe red noise process of interest in presence of observational noise (often taken to be white noise)
- Can modify this setup by changing observational noise from  $\mathbf{W}_t = 0$  to  $\mathbf{W}_t = W_t \sim \text{WN}(0, \sigma_W^2)$ , where  $W_t$  is uncorrelated with  $Z_t$ 's

Background

Multivariate Gaussian  
and Regression  
LemmasForecasting, Filtering,  
and Smoothing

AR(1) process with  $0 < \phi < 1$  is known as “red noise”, red noise is related to a 1st order stochastic differential equation, rendering it a model for various geophysical processes:

- Typically only observe red noise process of interest in presence of observational noise (often taken to be white noise)
- Can modify this setup by changing observational noise from  $W_t = 0$  to  $W_t = W_t \sim \text{WN}(0, \sigma_W^2)$ , where  $W_t$  is uncorrelated with  $Z_t$ 's
- The observation and state-transition equations become

$$Y_t = X_t + W_t \text{ and } X_t = \phi X_{t-1} + Z_t$$

## ARMA(1,1) Process as a State-Space Model: I

Recall ARMA(1,1) process  $Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}$

- Expressing ARMA(1,1) as  $\phi(B)Y_t = \theta(B)Z_t$ , note that one can create  $Y_t$  by taking causal AR(1) process  $X_t = \phi^{-1}(B)Z_t$  and subjecting it to a  $\theta(B)$  filter to obtain output  $Y_t = \theta(B)X_t = \theta(B)\phi^{-1}(B)Z_t$

### Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## ARMA(1,1) Process as a State-Space Model: I

Recall ARMA(1,1) process  $Y_t - \phi Y_{t-1} = Z_t + \theta Z_{t-1}$

- Expressing ARMA(1,1) as  $\phi(B)Y_t = \theta(B)Z_t$ , note that one can create  $Y_t$  by taking causal AR(1) process  $X_t = \phi^{-1}(B)Z_t$  and subjecting it to a  $\theta(B)$  filter to obtain output  $Y_t = \theta(B)X_t = \theta(B)\phi^{-1}(B)Z_t$
- Can express filtering of AR(1) process by

$$Y_t = [1 \quad \theta] \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix},$$

which matches up with observation equation

$$\mathbf{Y}_t = H_t \mathbf{X}_t + \mathbf{W}_t$$

$$\text{if } \mathbf{Y}_t = Y_t, H_t = [1 \quad \theta], \mathbf{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} \text{ and } \mathbf{W}_t = 0$$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## ARMA(1,1) Process as a State-Space Model: II

- Given  $\mathbf{X}_t = [X_t \ X_{t-1}]^T$ , can express  $X_t = \phi X_{t-1} + Z_t$  in the 1st row of matrix equation

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} Z_t \\ 0 \end{bmatrix},$$

which matches up with state-transition equation

$$\mathbf{X}_t = \mathbf{M}_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

if  $\mathbf{M}_t = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix}$  and  $\mathbf{V}_t = \begin{bmatrix} Z_t \\ 0 \end{bmatrix}$  with

$$\mathbf{Q}_t \stackrel{\text{def}}{=} \mathbf{E}(\mathbf{V}_t \mathbf{V}_t^T) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## ARMA(1,1) Process as a State-Space Model: II

- Given  $\mathbf{X}_t = \begin{bmatrix} X_t & X_{t-1} \end{bmatrix}^T$ , can express  $X_t = \phi X_{t-1} + Z_t$  in the 1st row of matrix equation

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} Z_t \\ 0 \end{bmatrix},$$

which matches up with state-transition equation

$$\mathbf{X}_t = \mathbf{M}_t \mathbf{X}_{t-1} + \mathbf{V}_t$$

if  $\mathbf{M}_t = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix}$  and  $\mathbf{V}_t = \begin{bmatrix} Z_t \\ 0 \end{bmatrix}$  with

$$\mathbf{Q}_t \stackrel{\text{def}}{=} \mathbf{E}(\mathbf{V}_t \mathbf{V}_t^T) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

- to complete the model, let

$$\mathbf{X}_0 = \begin{bmatrix} X_0 \\ X_{-1} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{\infty} \phi^j Z_{1-j} \\ \sum_{j=1}^{\infty} \phi^j Z_{-j} \end{bmatrix},$$

noting that  $X_0$  and  $\mathbf{V}_t$  for  $t \geq 1$  are uncorrelated, as required

Since

$$\mathbf{E}(\mathbf{X}_0 \mathbf{X}_0^T) = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix} = \frac{\sigma^2}{1 - \phi^2} \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix},$$

can alternatively stipulate

$$\mathbf{X}_0 = \begin{bmatrix} 1 & \frac{\phi}{\sqrt{1-\phi^2}} \\ 0 & \frac{\phi}{\sqrt{1-\phi^2}} \end{bmatrix} \begin{bmatrix} Z_0 \\ Z_{-1} \end{bmatrix},$$

yielding

$$\begin{aligned} \mathbf{E}(\mathbf{X}_0 \mathbf{X}_0^T) &= \begin{bmatrix} 1 & \frac{\phi}{\sqrt{1-\phi^2}} \\ 0 & \frac{\phi}{\sqrt{1-\phi^2}} \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{\phi}{\sqrt{1-\phi^2}} & \frac{1}{\sqrt{1-\phi^2}} \end{bmatrix} \\ &= \frac{\sigma^2}{1 - \phi^2} \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix} \end{aligned}$$

as required

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing



- State equation:

$$\mathbf{X}_t = \mathbf{M}_t \mathbf{X}_{t-1} + \mathbf{V}_t,$$

where  $\mathbf{V}_t \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$  with  $\mathbf{X}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$

- Observation equation:

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \mathbf{W}_t,$$

where  $\mathbf{W}_t \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$

- Additional assumptions:  $\mathbf{X}_0$ ,  $\{\mathbf{V}_t\}$ , and  $\{\mathbf{W}_t\}$  are uncorrelated

**Goal:** To estimate the underlying unobserved signal  $X_t$ , given the data  $y_{1:s} = \{y_1, y_2, \dots, y_s\}$ :

- When  $s < t$ , the problem is called **forecasting** or **prediction**
- When  $s = t$ , the problem is called **filtering**
- When  $s > t$ , the problem is called **smoothing**

In addition to these estimates, we would also want to measure their precision. The solution to these problems is accomplished via the **Kalman filter** and **Kalman smoother**

## The Kalman Filter: General Results

Assume the filtering distribution at time  $t - 1$  is

$$[\mathbf{X}_{t-1} | \mathbf{y}_{1:t-1}] \sim \mathcal{N}(\boldsymbol{\mu}_{t-1}^a, \Sigma_{t-1}^a)$$

- **Forecast Step:** Gives the forecast distribution at time  $t$ :

$$[\mathbf{X}_t | \mathbf{y}_{1:t-1}] \sim \mathcal{N}(\boldsymbol{\mu}_t^f, \Sigma_t^f),$$

where  $\boldsymbol{\mu}_t^f = M_t \boldsymbol{\mu}_{t-1}^a$ , and  $\Sigma_t^f = M_t \Sigma_{t-1}^a M_t^T + Q_t$ .

## The Kalman Filter: General Results

Assume the filtering distribution at time  $t - 1$  is

$$[\mathbf{X}_{t-1} | \mathbf{y}_{1:t-1}] \sim N(\boldsymbol{\mu}_{t-1}^a, \Sigma_{t-1}^a)$$

- **Forecast Step:** Gives the forecast distribution at time  $t$ :

$$[\mathbf{X}_t | \mathbf{y}_{1:t-1}] \sim N(\boldsymbol{\mu}_t^f, \Sigma_t^f),$$

where  $\boldsymbol{\mu}_t^f = M_t \boldsymbol{\mu}_{t-1}^a$ , and  $\Sigma_t^f = M_t \Sigma_{t-1}^a M_t^T + Q_t$ .

- **Update Step:** updates the forecast distribution using new data  $\mathbf{y}_t$

$$[\mathbf{X}_t | \mathbf{y}_{1:t}] \sim N(\boldsymbol{\mu}_t^a, \Sigma_t^a),$$

where  $\boldsymbol{\mu}_t^a = \boldsymbol{\mu}_t^f + K_t (\mathbf{y}_t - H_t \boldsymbol{\mu}_t^f)$ , and  $\Sigma_t^a = (I - K_t H_t^T) \Sigma_t^f$ ,  
and

$$K_t = \Sigma_t^f H_t^T (H_t \Sigma_t^f H_t^T + R_t)^{-1}$$

is the **Kalman gain matrix**

Let's begin with a particularly simple example of a state space model: the **local level model**. We will develop the basic state space techniques for this model.

- **Observation equation:**

$$Y_t = X_t + W_t, \quad \{W_t\} \stackrel{iid}{\sim} N(0, \sigma_W^2)$$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

Let's begin with a particularly simple example of a state space model: the **local level model**. We will develop the basic state space techniques for this model.

- **Observation equation:**

$$Y_t = X_t + W_t, \quad \{W_t\} \stackrel{iid}{\sim} N(0, \sigma_W^2)$$

- **State equation:**

$$X_t = X_{t-1} + V_t, \quad \{V_t\} \stackrel{iid}{\sim} N(0, \sigma_V^2)$$

Let's begin with a particularly simple example of a state space model: the **local level model**. We will develop the basic state space techniques for this model.

- **Observation equation:**

$$Y_t = X_t + W_t, \quad \{W_t\} \stackrel{iid}{\sim} N(0, \sigma_W^2)$$

- **State equation:**

$$X_t = X_{t-1} + V_t, \quad \{V_t\} \stackrel{iid}{\sim} N(0, \sigma_V^2)$$

- Assume  $E(X_0) = \mu_0$  and  $\text{Var}(X_0) = \sigma_0^2$  and  $X_0$  is uncorrected with  $W_t$ 's and  $V_t$ 's

- Since  $X_t = X_{t-1} + V_t$ , state variable  $X_t$  is a **random walk** starting from  $\mu_0$  (intended to model a slowly varying trend)



- Since  $X_t = X_{t-1} + V_t$ , state variable  $X_t$  is a **random walk** starting from  $\mu_0$  (intended to model a slowly varying trend)
- Since  $V_t$  and  $X_t$  are uncorrelated,

$$E(X_{t+1}|X_t) = E(X_t + V_t|X_t) = X_t + E(V_t) = X_t;$$

i.e., if state variable is at a certain 'level' at time  $t$ , we can expect no change in its level at time  $t + 1$

- Since  $X_t = X_{t-1} + V_t$ , state variable  $X_t$  is a **random walk** starting from  $\mu_0$  (intended to model a slowly varying trend)
- Since  $V_t$  and  $X_t$  are uncorrelated,

$$E(X_{t+1}|X_t) = E(X_t + V_t|X_t) = X_t + E(V_t) = X_t;$$

i.e., if state variable is at a certain ‘level’ at time  $t$ , we can expect no change in its level at time  $t + 1$

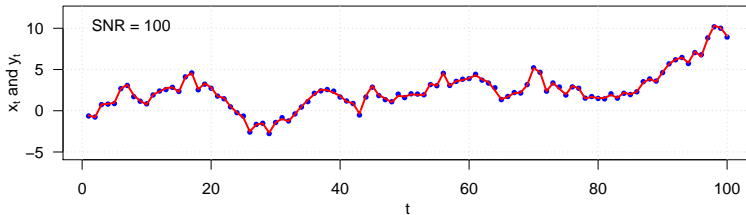
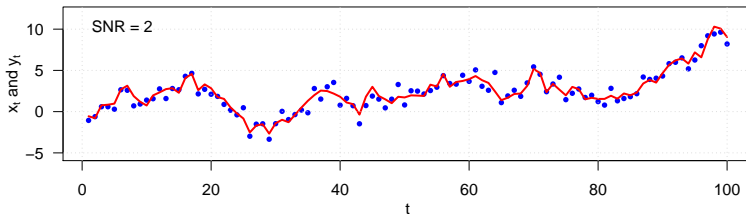
- When  $\sigma_W^2 > 0$ , trend is corrupted by noise, so ability to pick out trend depends upon “**signal to noise**” ratio (SNR)  $\frac{\sigma_V^2}{\sigma_W^2}$

# Local Level Model: Examples of Different SNR

## Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing



# Four Problems in State-Space Models

Given observations  $\{Y_i\}_{i=1}^t$  of a local level process,

- 1 **Filtering**: what is best predictor of state  $X_t$ ?
- 2 **Forecasting**: what is best predictor of state  $X_{t+1}$ ?
- 3 **Smoothing**: what is best predictor of state  $X_s$  for  $s < t$ ?
- 4 **Estimation**: what are best estimates of model parameters  $\sigma_W^2, \sigma_V^2, \mu_0, \sigma_0^2$ ?

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## Four Problems in State-Space Models

Given observations  $\{Y_i\}_{i=1}^t$  of a local level process,

- 1 **Filtering**: what is best predictor of state  $X_t$ ?
- 2 **Forecasting**: what is best predictor of state  $X_{t+1}$ ?
- 3 **Smoothing**: what is best predictor of state  $X_s$  for  $s < t$ ?
- 4 **Estimation**: what are best estimates of model parameters  $\sigma_W^2, \sigma_V^2, \mu_0, \sigma_0^2$ ?

First, we will focus on filtering and forecasting problems, with 'best' defined as the **minimum mean square error** (MSE).

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## Four Problems in State-Space Models

## Background

Multivariate Gaussian  
and Regression  
LemmasForecasting, Filtering,  
and Smoothing

Given observations  $\{Y_i\}_{i=1}^t$  of a local level process,

- 1 **Filtering**: what is best predictor of state  $X_t$ ?
- 2 **Forecasting**: what is best predictor of state  $X_{t+1}$ ?
- 3 **Smoothing**: what is best predictor of state  $X_s$  for  $s < t$ ?
- 4 **Estimation**: what are best estimates of model parameters  $\sigma_W^2, \sigma_V^2, \mu_0, \sigma_0^2$ ?

First, we will focus on filtering and forecasting problems, with 'best' defined as the **minimum mean square error** (MSE).

To facilitate discussion, let's assume that  $X_0$ ,  $V_t$ 's, and  $W_t$  are normals, implying that  $Y_t$  and the remaining  $X_t$ 's share this property.

- Suppose random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  are jointly normal with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ , to be denoted by

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

- Suppose random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  are jointly normal with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ , to be denoted by

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim N(\boldsymbol{\mu}, \Sigma)$$

- Can partition both  $\boldsymbol{\mu}$  and  $\Sigma$ :

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}\right),$$

where  $\boldsymbol{\mu}_X$  ( $\boldsymbol{\mu}_Y$ ) and  $\Sigma_{XX}$  ( $\Sigma_{YY}$ ) are mean and covariance matrix for  $\mathbf{X}$  ( $\mathbf{Y}$ );  $\Sigma_{XY}$  is the cross-covariance matrix between  $\mathbf{X}$  and  $\mathbf{Y}$



- Conditional distribution of  $\mathbf{X}$  given  $\mathbf{Y} = \mathbf{y}$  is multivariate normal with mean vector

$$\boldsymbol{\mu}_{\mathbf{X}|\mathbf{y}} = \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\Sigma}_{\mathbf{XY}} \boldsymbol{\Sigma}_{\mathbf{YY}}^{-1} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}})$$

and covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{X}|\mathbf{y}} = \boldsymbol{\Sigma}_{\mathbf{XX}} - \boldsymbol{\Sigma}_{\mathbf{XY}} \boldsymbol{\Sigma}_{\mathbf{YY}}^{-1} \boldsymbol{\Sigma}_{\mathbf{XY}}^T$$

- Conditional distribution of  $X$  given  $Y = y$  is multivariate normal with mean vector

$$\mu_{X|y} = \mu_X + \Sigma_{XY} \Sigma_{YY}^{-1} (y - \mu_Y)$$

and covariance matrix

$$\Sigma_{X|y} = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T$$

- Best (under MSE) predictor of  $X$  given  $Y$  is

$$E(X|Y) = \mu_{X|Y} = \mu_X + \Sigma_{XY} \Sigma_{YY}^{-1} (Y - \mu_Y)$$

## Regression Lemma III

- Recall that, if random vector  $U$  has covariance matrix  $\Sigma_U$ , then covariance matrix for  $AU$  is  $A\Sigma_U A^T$

$\Rightarrow$  covariance matrix of  $c + A(U - \mu_U)$  is also  $A\Sigma_U A^T$

- Recall that, if random vector  $U$  has covariance matrix  $\Sigma_U$ , then covariance matrix for  $AU$  is  $A\Sigma_U A^T$

$\Rightarrow$  covariance matrix of  $c + A(U - \mu_U)$  is also  $A\Sigma_U A^T$

- Covariance matrix for

$$E(X|Y) = \mu_{X|Y} = \mu_X + \Sigma_{XY} \Sigma_{YY}^{-1} (Y - \mu_Y)$$

is thus

$$\Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YY} \Sigma_{YY}^{-1} \Sigma_{XY}^T = \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T$$

**Note:** it is not the same as  $\Sigma_{X|y} = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T$

## Regression Lemma IV

Consider prediction error  $U$  associated with best linear predictor of  $X$ :

$$U = X - E(X|Y)$$

- Since  $E[E(X|Y)] = \mu_X \Rightarrow E(U) = 0$

Consider prediction error  $U$  associated with best linear predictor of  $X$ :

$$U = X - E(X|Y)$$

- Since  $E[E(X|Y)] = \mu_X \Rightarrow E(U) = 0$
- Covariance matrix for  $U$  is given by

$$\begin{aligned} E(UU^T) &= E\left([X - E(X|Y)][X - E(X|Y)]^T\right) \\ &= E(XX^T) + E[E(X|Y)E(X|Y)^T] \\ &\quad - E[XE(X|Y)^T] - E[E(X|Y)X^T] \\ &= \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}^T, \end{aligned}$$

which is equal to  $\Sigma_{X|y}$ , the conditional covariance matrix

Specialize now to case where  $X$  has just one element, say,  $X$

- Corollary: conditional distribution of  $X$  given  $Y = y$  is normal with mean

$$\mu_X + \Sigma_{XY}^T \Sigma_{YY}^{-1} (y - \mu_Y)$$

and conditional variance

$$\Sigma_{X|y} = \sigma_X^2 - \Sigma_{XY}^T \Sigma_{YY}^{-1} \Sigma_{XY},$$

where  $\sigma_X^2 = \text{Var}(X)$  and  $\Sigma_{XY}$  is a column vector containing covariance between  $X$  and  $Y$

Specialize now to case where  $X$  has just one element, say,  $X$

- Corollary: conditional distribution of  $X$  given  $Y = y$  is normal with mean

$$\mu_X + \Sigma_{XY}^T \Sigma_{YY}^{-1} (y - \mu_Y)$$

and conditional variance

$$\Sigma_{X|y} = \sigma_X^2 - \Sigma_{XY}^T \Sigma_{YY}^{-1} \Sigma_{XY},$$

where  $\sigma_X^2 = \text{Var}(X)$  and  $\Sigma_{XY}$  is a column vector containing covariance between  $X$  and  $Y$

- Since conditional variance is same as MSE for  $X$ , will refer to  $\Sigma_{X|y}$  as MSE



Suppose  $\{X_t\}$  is zero mean stationary process with ACF  $\gamma(h)$

- Set  $X$  to  $X_{n+1}$  and put  $X_1, \dots, X_n$  into  $\mathbf{Y}$
- Corollary says best linear predictor  $\hat{X}_{n+1}$  of  $X_{n+1}$  given  $X_1, \dots, X_n$  is

$$\hat{X}_{n+1} = \Sigma_{X\mathbf{Y}}^T \Sigma_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{Y} = \gamma_n^T \Gamma_n^{-1} \mathbf{Y} \stackrel{\text{def}}{=} \phi_n^T \mathbf{Y},$$

where

- 1  $\gamma_n = [\gamma(1), \gamma(2), \dots, \gamma(n)]^T = \Sigma_{X\mathbf{Y}}$
- 2  $(i, j)$ th entry of matrix  $\Gamma_n = \Sigma_{\mathbf{Y}\mathbf{Y}}$  is  $\gamma(i - j)$
- 3  $\phi_n^T \stackrel{\text{def}}{=} \gamma_n^T \Gamma_n^{-1}$  and hence  $\phi_n = \Gamma_n^{-1} \gamma_n$

Recall that MSE for  $\hat{X}_{n+1}$  is

$$\begin{aligned}v_n &= \text{Var}(X_{n+1}) - \phi_n^T \gamma_n \\&= \sigma_X^2 - \gamma_n^T \Gamma_n^{-1} \gamma_n \\&= \sigma_X^2 - \Sigma_{XY}^T \Sigma_{YY}^{-1} \Sigma_{XY} \\&= \Sigma_{X|Y}\end{aligned}$$

This is a special case of regression corollary

Let's begin with a particularly simple example of a state space model: the **local level model**

- Local level model:

$$Y_t = X_t + W_t, \quad \{W_t\} \sim N(0, \sigma_W^2)$$
$$X_t = X_{t-1} + V_t, \quad \{V_t\} \sim N(0, \sigma_V^2)$$

and  $X_0$  is a R.V. that

- is uncorrelated with  $W_t$ 's and  $V_t$ 's
- has  $E(X_0) = \mu_0$  and  $\text{Var}(X_0) = \sigma_0^2$
- Filtering problem is to predict unknown state  $X_t$  based on data up to time  $t$ , i.e.,  $\mathbf{Y}_{1:t} = (y_1, \dots, y_t)^T$

Background

Multivariate Gaussian  
and Regression  
LemmasForecasting, Filtering,  
and Smoothing

## Filtering for Local Level Model: II

Best linear predictor of  $X_t$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_t^a \stackrel{\text{def}}{=} E(X_t | \mathbf{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = E(X_t)$ ,  $\boldsymbol{\mu}_{1:t}$  is a vector containing, for  $j = 1, \dots, t$ ,

$$\mu_j \stackrel{\text{def}}{=} E(X_j) = E(X_j + W_j) = E(Y_j)$$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## Filtering for Local Level Model: II

Best linear predictor of  $X_t$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_t^a \stackrel{\text{def}}{=} E(X_t | \mathbf{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = E(X_t)$ ,  $\boldsymbol{\mu}_{1:t}$  is a vector containing, for  $j = 1, \dots, t$ ,

$$\mu_j \stackrel{\text{def}}{=} E(X_j) = E(X_j + W_j) = E(Y_j)$$

- Vector  $\Sigma_{t,t}$  contains covarinces between  $X_t$  and  $\mathbf{Y}_{1:t}$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## Filtering for Local Level Model: II

Best linear predictor of  $X_t$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_t^a \stackrel{\text{def}}{=} E(X_t | \mathbf{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = E(X_t)$ ,  $\boldsymbol{\mu}_{1:t}$  is a vector containing, for  $j = 1, \dots, t$ ,

$$\mu_j \stackrel{\text{def}}{=} E(X_j) = E(X_j + W_j) = E(Y_j)$$

- Vector  $\Sigma_{t,t}$  contains covarinces between  $X_t$  and  $\mathbf{Y}_{1:t}$
- $(i, j)$ th element of matrix  $\Sigma_{Y,t}$  is covariance between  $Y_i$  and  $Y_j$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## Filtering for Local Level Model: II

Best linear predictor of  $X_t$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_t^a \stackrel{\text{def}}{=} E(X_t | \mathbf{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = E(X_t)$ ,  $\boldsymbol{\mu}_{1:t}$  is a vector containing, for  $j = 1, \dots, t$ ,

$$\mu_j \stackrel{\text{def}}{=} E(X_j) = E(X_j + W_j) = E(Y_j)$$

- Vector  $\Sigma_{t,t}$  contains covarinces between  $X_t$  and  $\mathbf{Y}_{1:t}$
- $(i, j)$ th element of matrix  $\Sigma_{Y,t}$  is covariance between  $Y_i$  and  $Y_j$
- Note:  $E(\mu_t^a) = E[E(X_t | \mathbf{Y}_{1:t})] = E(X_t) = \mu_t$

## Filtering for Local Level Model: II

Best linear predictor of  $X_t$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_t^a \stackrel{\text{def}}{=} E(X_t | \mathbf{Y}_{1:t}) = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where

- $\mu_t = E(X_t)$ ,  $\boldsymbol{\mu}_{1:t}$  is a vector containing, for  $j = 1, \dots, t$ ,

$$\mu_j \stackrel{\text{def}}{=} E(X_j) = E(X_j + W_j) = E(Y_j)$$

- Vector  $\Sigma_{t,t}$  contains covariances between  $X_t$  and  $\mathbf{Y}_{1:t}$
- $(i, j)$ th element of matrix  $\Sigma_{Y,t}$  is covariance between  $Y_i$  and  $Y_j$
- Note:  $E(\mu_t^a) = E[E(X_t | \mathbf{Y}_{1:t})] = E(X_t) = \mu_t$
- With  $\sigma_t^2 \stackrel{\text{def}}{=} \text{Var}(X_t)$ , MSE for predictor is

$$E[(X_t - \mu_t^a)^2] = \sigma_t^2 - \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} \Sigma_{t,t} \stackrel{\text{def}}{=} \Sigma_t^a$$



**Forecasting:** estimate  $X_{t+1}$  given  $\mathbf{Y}_{1:t}$

- Best linear predictor of  $X_{t+1}$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_{t+1}^f \stackrel{\text{def}}{=} E(X_{t+1} | \mathbf{Y}_{1:t}) = \mu_{t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where vector  $\Sigma_{t+1,t}$  has covaraince between  $X_{t+1}$  and  $\mathbf{Y}_{1:t}$

**Forecasting:** estimate  $X_{t+1}$  given  $\mathbf{Y}_{1:t}$

- Best linear predictor of  $X_{t+1}$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_{t+1}^f \stackrel{\text{def}}{=} \mathbb{E}(X_{t+1} | \mathbf{Y}_{1:t}) = \mu_{t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where vector  $\Sigma_{t+1,t}$  has covaraince between  $X_{t+1}$  and  $\mathbf{Y}_{1:t}$

- Note:  $\mathbb{E}(\mu_{t+1}^f) = \mathbb{E}[\mathbb{E}(X_{t+1} | \mathbf{Y}_{1:t})] = \mathbb{E}(X_{t+1}) = \mu_{t+1}$

**Forecasting:** estimate  $X_{t+1}$  given  $\mathbf{Y}_{1:t}$

- Best linear predictor of  $X_{t+1}$  given  $\mathbf{Y}_{1:t}$  is

$$\mu_{t+1}^f \stackrel{\text{def}}{=} \mathbb{E}(X_{t+1} | \mathbf{Y}_{1:t}) = \mu_{t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t}),$$

where vector  $\Sigma_{t+1,t}$  has covaraince between  $X_{t+1}$  and  $\mathbf{Y}_{1:t}$

- Note:  $\mathbb{E}(\mu_{t+1}^f) = \mathbb{E}[\mathbb{E}(X_{t+1} | \mathbf{Y}_{1:t})] = \mathbb{E}(X_{t+1}) = \mu_{t+1}$
- MSE for predictor is

$$\mathbb{E}[(X_{t+1} - \mu_{t+1}^f)^2] = \sigma_{t+1}^2 - \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} \Sigma_{t+1,t} \stackrel{\text{def}}{=} \Sigma_{t+1}^f$$

## Forecasting for Local Level Model: II

- Let's also consider best linear predictor of  $Y_{t+1}$  given  $\mathbf{Y}_{1:t}$  :

$$Y_{t+1}^t \stackrel{\text{def}}{=} E(Y_{t+1} | \mathbf{Y}_{1:t}) = \mu_{Y,t+1} + \tilde{\Sigma}_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}),$$

where the vector  $\tilde{\Sigma}_{t+1,t}$  has covarainces between  $Y_{t+1}$  and  $\mathbf{Y}_{1:t}$

## Forecasting for Local Level Model: II

- Let's also consider best linear predictor of  $Y_{t+1}$  given  $\mathbf{Y}_{1:t}$  :

$$Y_{t+1}^t \stackrel{\text{def}}{=} E(Y_{t+1} | \mathbf{Y}_{1:t}) = \mu_{Y,t+1} + \tilde{\Sigma}_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}),$$

where the vector  $\tilde{\Sigma}_{t+1,t}$  has covarainces between  $Y_{t+1}$  and  $\mathbf{Y}_{1:t}$

- However, note that, for  $j = 1, \dots, t$

$$\text{Cov}(Y_{t+1}, Y_j) = \text{Cov}(X_{t+1} + W_{t+1}, Y_j) = \text{Cov}(X_{t+1}, Y_j)$$

## Forecasting for Local Level Model: II

- Let's also consider best linear predictor of  $Y_{t+1}$  given  $\mathbf{Y}_{1:t}$ :

$$Y_{t+1}^t \stackrel{\text{def}}{=} E(Y_{t+1} | \mathbf{Y}_{1:t}) = \mu_{Y,t+1} + \tilde{\Sigma}_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}),$$

where the vector  $\tilde{\Sigma}_{t+1,t}$  has covarainces between  $Y_{t+1}$  and  $\mathbf{Y}_{1:t}$

- However, note that, for  $j = 1, \dots, t$

$$\text{Cov}(Y_{t+1}, Y_j) = \text{Cov}(X_{t+1} + W_{t+1}, Y_j) = \text{Cov}(X_{t+1}, Y_j)$$

- Thus  $\tilde{\Sigma}_{t+1,t} = \Sigma_{t+1,t}$ , yielding

$$Y_{t+1}^t = \mu_{Y,t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{y}_{1:t} - \boldsymbol{\mu}_{Y,1:t}) = \mu_{t+1}^f$$

$\Rightarrow$  difference between  $Y_{t+1}$  and  $X_{t+1}$  is  $W_{t+1}$ , therefore they have the same estimator, but their MSEs differ:

$$E \left[ (Y_{t+1} - Y_{t+1}^f)^2 \right] = \Sigma_{t+1}^f + \sigma_W^2$$

- To implement filtering, i.e., compute  $\mu_t^a$ , need to determine:

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

- To implement filtering, i.e., compute  $\mu_t^a$ , need to determine:

1  $\mu_j = E(X_j), j = 1, \dots, t$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing



- To implement filtering, i.e., compute  $\mu_t^a$ , need to determine:

1  $\mu_j = E(X_j), j = 1, \dots, t$

2 Elements of  $\Sigma_{t,t}$ , i.e., covariance between  $X_t$  and  $Y_{1:t}$

- To implement filtering, i.e., compute  $\mu_t^a$ , need to determine:
  - 1  $\mu_j = E(X_j)$ ,  $j = 1, \dots, t$
  - 2 Elements of  $\Sigma_{t,t}$ , i.e., covaraince between  $X_t$  and  $Y_{1:t}$
  - 3 Elements of  $\Sigma_{Y,t}$ , i.e., covariances between  $Y_j$  and  $Y_k$ ,  $1 \leq j \leq k \leq t$

- To implement filtering, i.e., compute  $\mu_t^a$ , need to determine:
  - 1  $\mu_j = E(X_j), j = 1, \dots, t$
  - 2 Elements of  $\Sigma_{t,t}$ , i.e., covaraince between  $X_t$  and  $Y_{1:t}$
  - 3 Elements of  $\Sigma_{Y,t}$ , i.e., covariances between  $Y_j$  and  $Y_k$ ,  
 $1 \leq j \leq k \leq t$
- To compute  $\Sigma_t^a$ , i.e., MSE for  $\mu_t^a$ , need  $\sigma_t^2 = \text{Var}(X_t)$  in addition to 2 and 3 above

- To implement filtering, i.e., compute  $\mu_t^a$ , need to determine:
  - 1  $\mu_j = E(X_j)$ ,  $j = 1, \dots, t$
  - 2 Elements of  $\Sigma_{t,t}$ , i.e., covariance between  $X_t$  and  $Y_{1:t}$
  - 3 Elements of  $\Sigma_{Y,t}$ , i.e., covariances between  $Y_j$  and  $Y_k$ ,  $1 \leq j \leq k \leq t$
- To compute  $\Sigma_t^a$ , i.e., MSE for  $\mu_t^a$ , need  $\sigma_t^2 = \text{Var}(X_t)$  in addition to 2 and 3 above
- Since  $X_t = X_{t-1} + V_t$  and  $Y_t = X_t + W_t$ , telescoping yields  $X_j = X_0 + \sum_{l=1}^j V_l$  and  $Y_j = X_0 + \sum_{l=1}^j V_l + W_j$ ,  $j = 1, \dots, t$

Using

$$X_j = X_0 + \sum_{l=1}^j V_l \text{ and } Y_j = X_0 + \sum_{l=1}^j W_l, \quad j = 1, \dots, t,$$

get  $\mu_j = E[X_j] = E[X_0] = \mu_0$  and (assuming  $j \leq k \leq t$ )

$$\begin{aligned} \text{Cov}(X_t, Y_j) &= \text{Cov}\left(X_0 + \sum_{l=1}^t V_l, X_0 + \sum_{l=1}^j V_l + W_j\right) \\ &= \sigma_0^2 + j\sigma_V^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}(Y_j, Y_k) &= \text{Cov}\left(X_0 + \sum_{l=1}^j V_l + W_j, X_0 + \sum_{l=1}^k V_l + W_k\right) \\ &= \sigma_0^2 + j\sigma_V^2 + \delta_{jk}\sigma_W^2, \end{aligned}$$

where  $\delta_{jk} = 1$  if  $j = k$  and  $\delta_{jk} = 0$  if  $j \neq k$

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

Using

$$X_t = X_0 + \sum_{l=1}^t V_l,$$

get

$$\sigma_t^2 = \text{Var}(X_t) = \sigma_0^2 + t\sigma_V^2$$

- Now we have all the pieces needed to form  $\mu_t^a$  and its MSE  $\Sigma_t^a$
- **Note:** similar argument leads to pieces needed to form forecast  $\mu_{t+1}^f$  and its MSE  $\Sigma_{t+1}^f$

While straightforward conceptually, forming

$$\mu_t^a = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

and

$$\mu_{t+1}^f = \mu_{t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

via these equations requires inversion of matrix  $\Sigma_{Y,t}$  whose dimension  $t \times t$  becomes problematic as  $t$  gets large 😞

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

While straightforward conceptually, forming

$$\mu_t^a = \mu_t + \Sigma_{t,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

and

$$\mu_{t+1}^f = \mu_{t+1} + \Sigma_{t+1,t}^T \Sigma_{Y,t}^{-1} (\mathbf{Y}_{1:t} - \boldsymbol{\mu}_{1:t})$$

via these equations requires inversion of matrix  $\Sigma_{Y,t}$  whose dimension  $t \times t$  becomes problematic as  $t$  gets large 😞

⇒ The celebrated **Kalman recursions** give a recipe that avoids explicit matrix inversion

- **Idea:** at time  $t - 1$ , we have 4 quantities of interest: fitted value  $\mu_{t-1}^a$ , and forecast  $\mu_t^f$  and their associated MSEs  $\Sigma_{t-1}^a$  and  $\Sigma_t^f$
- **Note:**  $\mu_{t-1}^a = \mu_t^f$  for local level model (but not others)



- At time  $t$ , new observation  $Y_t$  becomes available

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## Kalman Recursions for Filtering/Forecasting: II

- At time  $t$ , new observation  $Y_t$  becomes available
- Kalman recursion takes  $\mu_t^f$ ,  $\Sigma_t^f$  and  $Y_t$  and yields

Background

Multivariate Gaussian  
and Regression  
Lemmas

Forecasting, Filtering,  
and Smoothing

## Kalman Recursions for Filtering/Forecasting: II

- At time  $t$ , new observation  $Y_t$  becomes available
- Kalman recursion takes  $\mu_t^f$ ,  $\Sigma_t^f$  and  $Y_t$  and yields
  - fitted values  $\mu_t^a$  and forecast  $\mu_{t+1}^f$

- At time  $t$ , new observation  $Y_t$  becomes available
- Kalman recursion takes  $\mu_t^f$ ,  $\Sigma_t^f$  and  $Y_t$  and yields
  - fitted values  $\mu_t^a$  and forecast  $\mu_{t+1}^f$
  - associated MSEs  $\Sigma_t^a$  and  $\Sigma_{t+1}^f$

- At time  $t$ , new observation  $Y_t$  becomes available
- Kalman recursion takes  $\mu_t^f$ ,  $\Sigma_t^f$  and  $Y_t$  and yields
  - fitted values  $\mu_t^a$  and forecast  $\mu_{t+1}^f$
  - associated MSEs  $\Sigma_t^a$  and  $\Sigma_{t+1}^f$
- There are six steps in the Kalman recursion:

- At time  $t$ , new observation  $Y_t$  becomes available
- Kalman recursion takes  $\mu_t^f$ ,  $\Sigma_t^f$  and  $Y_t$  and yields
  - fitted values  $\mu_t^a$  and forecast  $\mu_{t+1}^f$
  - associated MSEs  $\Sigma_t^a$  and  $\Sigma_{t+1}^f$
- There are six steps in the Kalman recursion:
  - 1 steps 1 and 2 are preparatory

- At time  $t$ , new observation  $Y_t$  becomes available
- Kalman recursion takes  $\mu_t^f$ ,  $\Sigma_t^f$  and  $Y_t$  and yields
  - fitted values  $\mu_t^a$  and forecast  $\mu_{t+1}^f$
  - associated MSEs  $\Sigma_t^a$  and  $\Sigma_{t+1}^f$
- There are six steps in the Kalman recursion:
  - 1 steps 1 and 2 are preparatory
  - 2 steps 3 and 4 yield  $\mu_t^a$  and  $\Sigma_t^a$  (filtering)

- At time  $t$ , new observation  $Y_t$  becomes available
- Kalman recursion takes  $\mu_t^f$ ,  $\Sigma_t^f$  and  $Y_t$  and yields
  - fitted values  $\mu_t^a$  and forecast  $\mu_{t+1}^f$
  - associated MSEs  $\Sigma_t^a$  and  $\Sigma_{t+1}^f$
- There are six steps in the Kalman recursion:
  - steps 1 and 2 are preparatory
  - steps 3 and 4 yield  $\mu_t^a$  and  $\Sigma_t^a$  (filtering)
  - steps 5 and 6 yield  $\mu_{t+1}^f$  and  $\Sigma_{t+1}^f$  (forecasting)



1. Compute innovation:

$$U_t = Y_t - Y_t^{t-1} = Y_t - \mu_t^f$$

2. Compute MSE for  $Y_t^{t-1}$ :

$$\Sigma_t^f + \sigma_W^2 \stackrel{\text{def}}{=} F_t$$

3. Compute new filtered value:

$$\mu_t^a = \mu_t^f + K_t U_t,$$

where  $K_t \stackrel{\text{def}}{=} \Sigma_t^f / F_t$  is the so-called **Kalman gain**

4. Compute MSE for new filtered value:

$$\Sigma_t^a = \Sigma_t^f (1 - K_t)$$

5. Compute new forecast:

$$\mu_{t+1}^f = \mu_t^f + K_t U_t = \mu_t^a$$

6. Compute MSE for new forecast:

$$\Sigma_{t+1}^f = \Sigma_t(1 - K_t) + \sigma_V^2 = \Sigma_t^a + \sigma_V^2$$

Recursions are carried out for  $t = 0, \dots, n$  with inputs  $E[X_0] = \mu_0$ ,  $\text{Var}(X_0) = \sigma_0^2$  and  $Y_t'$ s

## Kalman Recursions for Filtering/Forecasting: V

To prove validity of steps 3 and 4, need to show that  $\mu_t^f + K_t U_t$  is equal to  $\mu_t^a$ , and  $\Sigma_t^f (1 - K_t)$  is equal to  $\Sigma_t^a$

## Kalman Recursions for Filtering/Forecasting: V

To prove validity of steps 3 and 4, need to show that  $\mu_t^f + K_t U_t$  is equal to  $\mu_t^a$ , and  $\Sigma_t^f (1 - K_t)$  is equal to  $\Sigma_t^a$

- **Key fact:**  $X_t$  conditioned on both  $U_t = Y_t - Y_t^{t-1}$  and  $\mathbf{Y}_{1:t-1}$  is the same as  $X_t$  conditioned on  $\mathbf{Y}_{1:t-1}$  because

$$\begin{aligned}\text{Cov}(X_t, U_t | \mathbf{Y}_{1:t-1}) &= \text{Cov}(X_t, Y_t - Y_t^{t-1} | \mathbf{Y}_{1:t-1}) \\ &= \text{Cov}(X_t, X_t + W_t | \mathbf{Y}_{1:t-1}) = \text{Var}(X_t | \mathbf{Y}_{1:t-1}) \\ &= \Sigma_t^f\end{aligned}$$

We have

$$\mu_t^a = \mu_t^f + \frac{\Sigma_t^f}{F_t} U_t, \text{ and } \Sigma_t^a = \Sigma_t^f - \frac{(\Sigma_t^f)^2}{F_t}$$

since  $K_t \stackrel{\text{def}}{=} \frac{\Sigma_t^f}{F_t}$ , we get required

$$\mu_t^a = \mu_t^f + K_t U_t \text{ and } \Sigma_t^a = \Sigma_t^f (1 - K_t)$$