Lecture 14

State-Space Models I

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

MATH 8090 Time Series Analysis
Week 14

Background

Forecasting, Filtering, and Smoothing

Multivariate Gaussian and Regression Lemmas

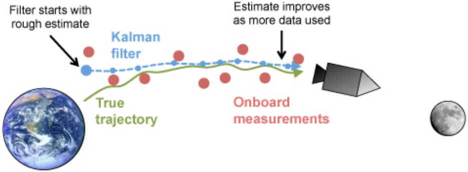
Whitney Huang
Clemson University

Agenda

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

Historical Background

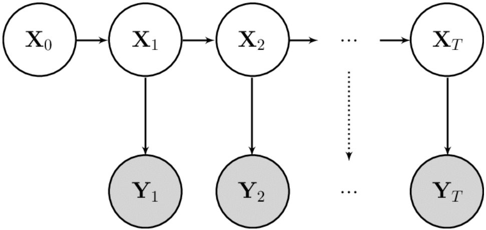
- The original model emerged in the context of space tracking [Kalman, 1960, Kalman and Bucy, 1961]
- The "state equation" defines the motion equations for the position of a spacecraft with location x_t



- 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 underling state $\{x_t\}$ based on observed data $\{y_t\}$

State-Space Model



$$\text{State: } \boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t, \quad \boldsymbol{V}_t \overset{i.i.d.}{\sim} \text{WN}(\mathbf{0}, Q_t), \quad t = 1, 2, \cdots$$

$$\text{Observation: } \boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t, \quad \boldsymbol{W}_t \overset{i.i.d.}{\sim} \text{WN}(\mathbf{0}, R_t), \quad t = 1, 2, \cdots$$

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

Additional Assumptions of State-Space Models

State equation:

$$\boldsymbol{X}_t = M_t \boldsymbol{X}_{t-1} + \boldsymbol{V}_t, \quad t = 1, 2\cdots$$

Observation equation:

$$\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t, \quad t = 1, 2, \cdots$$

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

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

Applications of State-Space Models

- 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 $\boldsymbol{X}_t$ with $\boldsymbol{Y}_{1:s} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_s\}$:
  - 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

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

is reminiscent of a causal AR(1) model:

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

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

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

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

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

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

- A natural choice is

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

- With this choice, the required conditions, namely, $\mathrm{E}(X_0 \boldsymbol{W}_t^T) = 0$ and $\mathrm{E}(X_0 \boldsymbol{V}_t^T) = 0$ 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 as a State-Space Model: III

Background

Forecasting, Filtering, and Smoothing

Multivariate Gaussian and Regression Lemmas

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 $\boldsymbol{W}_t = 0$ to $\boldsymbol{W}_t = W_t \sim \mathrm{WN}(0, \sigma_W^2)$, 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 $Y_t = \theta(B)X_t = \theta(B)\phi^{-1}(B)Z_t$

- Can express filtering of AR(1) process by

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

  which matches up with observation equation

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

  if $\boldsymbol{Y}_t = Y_t$, $H_t = \begin{bmatrix} 1 & \theta \end{bmatrix}$, $\boldsymbol{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix}$ and $\boldsymbol{W}_t = 0$

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

- Given $\boldsymbol{X}_t = \begin{bmatrix} X_t & X_{t-1} \end{bmatrix}^T$, 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

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

  if $M_t = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix}$ and $\boldsymbol{V}_t = \begin{bmatrix} Z_t \\ 0 \end{bmatrix}$ with $Q_t \overset{\text{def}}{=} \mathrm{E}(\boldsymbol{V}_t \boldsymbol{V}_t^T) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$

- to complete the model, let

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

  noting that $\boldsymbol{X}_0$ and $\boldsymbol{V}_t$ for $t \geq 1$ are uncorrelated, as required

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

- Since

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

can alternatively stipulate

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

yielding

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

as required

The Linear Gaussian State-Space Model

- State equation:

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

where $\boldsymbol{V}_t \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, Q_t)$ with $\boldsymbol{X}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0, \Sigma_0)$

- Observation equation:

$$\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{W}_t,$$

where $\boldsymbol{W}_t \overset{iid}{\sim} \mathrm{N}(\mathbf{0}, R_t)$

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

CLEMSON UNIVERSITY

Background
Forecasting, Filtering, and Smoothing
Multivariate Gaussian and Regression Lemmas

Forecasting, Filtering, and Smoothing

Goal: To estimate the underlying unobserved signal $\boldsymbol{X}_t$, given the data $\boldsymbol{y}_{1:s} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_s\}$:

- 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

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

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

$$[\boldsymbol{X}_{t}|\boldsymbol{y}_{1:t-1}] \sim \mathrm{N}\left(\boldsymbol{\mu}_{t}^{f}, \Sigma_{t}^{f}\right),$$

  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 $\boldsymbol{y}_t$

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

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

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

  is the Kalman gain matrix

Local Level Model: Part I

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**:

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

- **State equation**:

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

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

Background
Forecasting, Filtering, and Smoothing
Multivariate Gaussian and Regression Lemmas

Local Level Model: Part II

Background
Forecasting, Filtering, and Smoothing
Multivariate Gaussian and Regression Lemmas

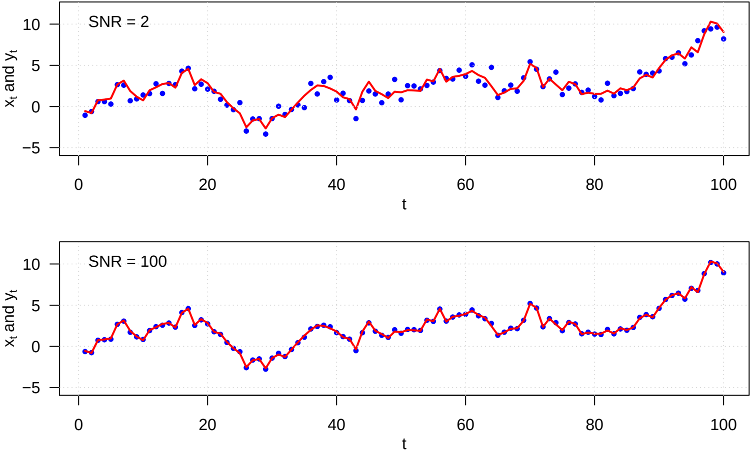
- Since $X_t = X_{t-1} + V_t$, state variable X_t is a random walk starting from μ_0 (intended to model a slowly varying trend)
- Since V_t and X_t are uncorrelated,

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

  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



Four Problems in State-Space Models

Background
Forecasting, Filtering, and Smoothing
Multivariate Gaussian and Regression Lemmas

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.

Regression Lemma I

Background
Forecasting, Filtering, and Smoothing
Multivariate Gaussian and Regression Lemmas

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

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

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

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

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

Regression Lemma II

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

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

and covariance matrix

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

- Best (under MSE) predictor of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y}} = \boldsymbol{\mu}_{\boldsymbol{X}} + \Sigma_{\boldsymbol{XY}} \Sigma_{\boldsymbol{YY}}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

Regression Lemma III

- Recall that, if random vector $\boldsymbol{U}$ has covariance matrix $\Sigma_{\boldsymbol{U}}$, then covariance matrix for $A\boldsymbol{U}$ is $A\Sigma_{\boldsymbol{U}}A^T$

  $\Rightarrow$ covariance matrix of $c + A(\boldsymbol{U} - \mu_{\boldsymbol{U}})$ is also $A\Sigma_{\boldsymbol{U}}A^T$

- Covariance matrix for

$$\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y}} = \boldsymbol{\mu}_{\boldsymbol{X}} + \Sigma_{\boldsymbol{X}\boldsymbol{X}}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

  is thus

$$\Sigma_{\boldsymbol{X}\boldsymbol{Y}}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}\Sigma_{\boldsymbol{X}\boldsymbol{Y}}^{T} = \Sigma_{\boldsymbol{X}\boldsymbol{Y}}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}\Sigma_{\boldsymbol{X}\boldsymbol{Y}}^{T}$$

Note: it is not the same as $\Sigma_{\boldsymbol{X}|\boldsymbol{y}} = \Sigma_{\boldsymbol{X}\boldsymbol{X}} - \Sigma_{\boldsymbol{X}\boldsymbol{Y}}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}\Sigma_{\boldsymbol{X}\boldsymbol{Y}}^{T}$

Regression Lemma IV

Consider prediction error $\boldsymbol{U}$ associated with best linear predictor of $\boldsymbol{X}$:

$$\boldsymbol{U} = \boldsymbol{X} - \mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})$$

- Since $\mathrm{E}[\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})] = \boldsymbol{\mu}_{\boldsymbol{X}} \Rightarrow \mathrm{E}(\boldsymbol{U}) = \boldsymbol{0}$
- Covariance matrix for $\boldsymbol{U}$ is given by

$$\begin{aligned}\mathrm{E}(\boldsymbol{U}\boldsymbol{U}^T) =& \mathrm{E}\left([\boldsymbol{X} - \mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})][\boldsymbol{X} - \mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})]^T\right)\\ =& \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^T) + \mathrm{E}[\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})^T]\\ &- \mathrm{E}[\boldsymbol{X}\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})^T] - \mathrm{E}[\mathrm{E}(\boldsymbol{X}|\boldsymbol{Y})\boldsymbol{X}^T]\\ =& \Sigma_{\boldsymbol{X}\boldsymbol{X}} - \Sigma_{\boldsymbol{X}\boldsymbol{Y}}\Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1}\Sigma_{\boldsymbol{X}\boldsymbol{Y}}^T,\end{aligned}$$

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

Regression Corollary

Specialize now to case where $\boldsymbol{X}$ has just one element, say, X

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

$$\mu_X + \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{Y}})$$

  and conditional variance

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

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

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

Aside – Revisiting Time Series Prediction: I

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

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

$$\hat{X}_{n+1} = \Sigma^T_{X\boldsymbol{Y}} \Sigma^{-1}_{\boldsymbol{Y}\boldsymbol{Y}} \boldsymbol{Y} = \gamma_n^T \Gamma_n^{-1} \boldsymbol{Y} \stackrel{\text{def}}{=} \boldsymbol{\phi}_n^T \boldsymbol{Y},$$

where

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

Aside – Revisiting Time Series Prediction: II

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

$$
\begin{aligned}
v_n &= \mathrm{Var}(X_{n+1}) - \boldsymbol{\phi}_n^T \boldsymbol{\gamma}_n \\
&= \sigma_X^2 - \boldsymbol{\gamma}_n^T \Gamma_n^{-1} \boldsymbol{\gamma}_n \\
&= \sigma_X^2 - \Sigma_{X\boldsymbol{Y}}^T \Sigma_{\boldsymbol{Y}\boldsymbol{Y}}^{-1} \Sigma_{X\boldsymbol{Y}} \\
&= \Sigma_{X|\boldsymbol{y}}
\end{aligned}
$$

This is a special case of regression corollary