Systems Dynamics

Course ID: 267MI – Fall 2019

Thomas Parisini
Gianfranco Fenu

University of Trieste
Department of Engineering and Architecture

267MI –Fall 2019

Lecture 10
Solution of the Prediction Problem

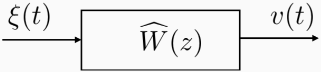
Lecture 10: Table of Contents

10. Solution of the Prediction Problem

Solution of the Prediction Problem

Solution of the Prediction Problem

Consider the process $v(t)$ with rational complex spectrum:



where $\xi(\cdot) \sim WN(0, \lambda^2)$ and $\widehat{W}(z) = \dfrac{N(z)}{D(z)}$ is the **spectral canonical factor**, that is:

- $N(z)$ and $D(z)$ are monic, co-prime and of the same degree
- All roots of $N(z)$ (zeros of $\widehat{W}(z)$) have $|\cdot| \leq 1$
- All roots of $D(z)$ (poles of $\widehat{W}(z)$) have $|\cdot| < 1$

Solution of the Prediction Problem (cont.)

- Then:

$$v(t) = \sum_{i=0}^{\infty} \hat{w}(i)\,\xi(t-i) = \hat{w}(0)\xi(t) + \hat{w}(1)\xi(t-1) + \cdots$$

where $\hat{w}(0)$, $\hat{w}(1)$, $\ldots$ are the samples of the impulse response of the system with transfer function $\widehat{W}(z)$:

$$\hat{w}(k) = \mathcal{Z}^{-1}\left[\widehat{W}(z)\right]$$

- Let us introduce the **additional assumption**:

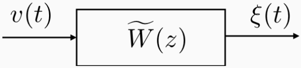
All roots of $N(z)$ (zeros of $\widehat{W}(z)$) have $|\cdot| < 1$

Hence: the spectral factorisation theorem also holds for

$$\widetilde{W}(z) = \frac{1}{\widehat{W}(z)}$$

Solution of the Prediction Problem (cont.)

- Then, we are able to consider



  that is, feeding the system having transfer function $\widetilde{W}(z)$ with the process $v(t)$, at the output we obtain exactly the white process $\xi(t)$
- $\widetilde{W}(z)$ is called **whitening filter** and

$$\xi(t) = \sum_{i=0}^{\infty} \tilde{w}(i)\, v(t-i) = \tilde{w}(0)v(t) + \tilde{w}(1)v(t-1) + \cdots$$

- Thus, the whitening filter is kind of a **inverse filter** with respect to the canonical representation of the original process $v(t)$

Solution of the Prediction Problem (cont.)

- Let us now consider

$$v(t) = \sum_{i=0}^{\infty} \hat{w}(i)\,\xi(t-i) = \hat{w}(0)\xi(t) + \hat{w}(1)\xi(t-1) + \cdots$$

  Clearly $v(t) \in \mathcal{H}_t[\xi]$ where we recall that $\mathcal{H}_t[\xi]$ is the space of all infinite linear combinations of $\xi(t), \xi(t-1), \ldots$.
  Analogously:

$$v(t-1) = \sum_{i=0}^{\infty} \hat{w}(i)\,\xi(t-1-i) = \hat{w}(0)\xi(t-1) + \hat{w}(1)\xi(t-2) + \cdots$$

$$v(t-2) = \sum_{i=0}^{\infty} \hat{w}(i)\,\xi(t-2-i) = \hat{w}(0)\xi(t-2) + \hat{w}(1)\xi(t-3) + \cdots$$

  $\ldots$

- Hence linear combinations of $v(t), v(t-1), \ldots$ can be expressed as linear combinations of $\xi(t), \xi(t-1), \ldots$ which implies:

$$\mathcal{H}_t[v] \subseteq \mathcal{H}_t[\xi]$$

Solution of the Prediction Problem (cont.)

- In the same way, one gets:

$$\xi(t) = \sum_{i=0}^{\infty} \tilde{w}(i)\, v(t-i) = \tilde{w}(0)v(t) + \tilde{w}(1)v(t-1) + \cdots$$

  Clearly $\xi(t) \in \mathcal{H}_t[v]$ and

$$\xi(t-1) = \sum_{i=0}^{\infty} \tilde{w}(i)\, v(t-1-i) = \tilde{w}(0)v(t-1) + \tilde{w}(1)v(t-2) + \cdots$$

$$\xi(t-2) = \sum_{i=0}^{\infty} \tilde{w}(i)\, v(t-2-i) = \tilde{w}(0)v(t-2) + \tilde{w}(1)v(t-3) + \cdots$$

$$\ldots$$

- Hence linear combinations of $\xi(t), \xi(t-1), \ldots$ can be expressed as linear combinations of $v(t), v(t-1), \ldots$ which implies:

$$\mathcal{H}_t[\xi] \subseteq \mathcal{H}_t[v]$$

- Thus, we finally conclude that:

$$\mathcal{H}_t[\xi] = \mathcal{H}_t[v]$$

Solution of the Prediction Problem (cont.)

The Prediction Problem

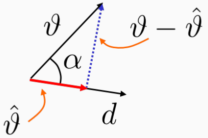
- Given the rational spectrum stationary process $v(t)$, we want to estimate $v(t+r),\ r \geq 1$ as a function of the past observations $v(t),\, v(t-1),\, \ldots$
- The observations $v(t),\, v(t-1),\, \ldots$ clearly make up an **a-priori knowledge** with respect to the quantity to be estimated $v(t+r)$
- Therefore, it is quite natural to cast the prediction problem in the framework of **Bayes estimation**:

$$\hat{v}(t+r\,|\,t) = E\left[v(t+r)\,|\,v(t), v(t-1), \ldots\right]$$

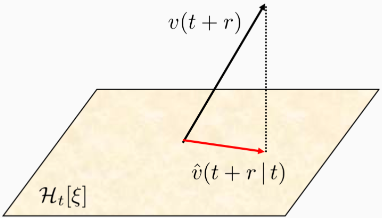
Solution of the Prediction Problem (cont.)

- Recall the geometric interpretation of the Bayes estimation:

$$\hat{\vartheta} = \frac{\lambda_{\vartheta d}}{\lambda_{dd}}\, d = \|\vartheta\|\, \cos\alpha\, \frac{d}{\|d\|}$$



Hence, in the case of the prediction problem $\hat{v}(t+r\,|\,t)$ is the projection of $v(t+r)$ (interpreted as a geometric vector) on the subspace (hyper-plane) $\mathcal{H}_t[\xi]\;(=\mathcal{H}_t[v])$



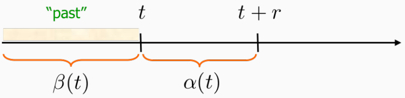
Solution of the Prediction Problem (cont.)

- Let us now determine $\hat{v}(t+r\,|\,t)$:

$$
\begin{aligned}
v(t+r) &= \sum_{i=0}^{\infty} \hat{w}(i)\,\xi(t+r-i) \\
&= \underbrace{\hat{w}(0)\xi(t+r) + \hat{w}(1)\xi(t+r-1) + \cdots + \hat{w}(r-1)\xi(t+1)}_{\alpha(t)} \\
&\quad + \underbrace{\hat{w}(r)\xi(t) + \hat{w}(r+1)\xi(t-1) + \cdots}_{\beta(t)} \\
&= \alpha(t) + \beta(t)
\end{aligned}
$$

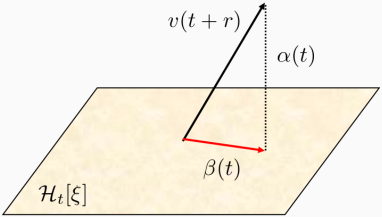
where:

- $\alpha(t)$: lin. comb. of white process samples in $[t+1, t+r] \cap \mathbb{Z}$
- $\beta(t)$: lin. comb. of white process samples in $(-\infty, t] \cap \mathbb{Z}$



Solution of the Prediction Problem (cont.)

- But: $\xi(t)$ is white $\Longrightarrow$ $\alpha(t)$ and $\beta(t)$ are uncorrelated
- Hence, vectors associated with $\alpha(t)$ and $\beta(t)$ are orthogonal



- Thus: the **optimal prediction** coincides with $\beta(t)$:

$$\hat{v}(t+r\,|\,t) = \hat{w}(r)\xi(t) + \hat{w}(r+1)\xi(t-1) + \cdots$$

Solution of the Prediction Problem (cont.)

- Instead, the **prediction error** coincides with $\alpha(t)$ which is orthogonal to $\mathcal{H}_t[\xi]\,(=\mathcal{H}_t[v])$:

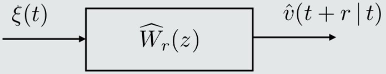
$$\begin{aligned}\varepsilon(t) &= v(t+r) - \hat{v}(t+r\,|\,t) \\ &= \hat{w}(0)\xi(t+r) + \hat{w}(1)\xi(t+r-1) + \cdots + \hat{w}(r-1)\xi(t+1)\end{aligned}$$

- Therefore, by defining

$$\widehat{W}_r(z) = \hat{w}(r) + \hat{w}(r+1)z^{-1} + \cdots$$

Optimal Predictor

$\widehat{W}_r(z)$ is the transfer function of the r-th steps ahead optimal predictor from the white process samples $\xi(t)$



Solution of the Prediction Problem

Determination of the Predictor

Determination of the Predictor

The computation of $\widehat{W}_r(z)$ is very simple: just carry out the **long-division** between the numerator and denominator of $\widehat{W}(z)$:

$$
\begin{aligned}
\widehat{W}(z) &= \hat{w}(0) + \hat{w}(1)z^{-1} + \cdots + \hat{w}(r-1)z^{-r+1} \\
&\quad + \hat{w}(r)z^{-r} + \hat{w}(r+1)z^{-r-1} + \cdots \\
&= \hat{w}(0) + \hat{w}(1)z^{-1} + \cdots + \hat{w}(r-1)z^{-r+1} \\
&\quad + z^{-r}\left[\hat{w}(r) + \hat{w}(r+1)z^{-1} + \cdots\right] \\
&= \hat{w}(0) + \hat{w}(1)z^{-1} + \cdots + \hat{w}(r-1)z^{-r+1} + z^{-r}\,\widehat{W}_r(z)
\end{aligned}
$$

Determination of the Optimal Predictor

$\widehat{W}_r(z)$ is obtained as a result of the r-times repeated division: the **remainder**, multiplied by z^r is the $\widehat{W}_r(z)$ we were looking for:

$$
\widehat{W}(z) = \frac{N(z)}{D(z)} \implies \frac{N(z)}{D(z)} = E(z) + z^{-r}\widehat{W}_r(z)
$$

Determination of the Predictor: Basic Example

Consider:

$$
\begin{aligned}
v(t) + \frac{5}{6}v(t-1) + \frac{1}{6}v(t-2) &= \xi(t) + \frac{1}{9}\xi(t-1) \\
\implies (1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2})v(t) &= (1 + \frac{1}{9}z^{-1})\xi(t) \\
\implies v(t) &= \frac{z(z+\frac{1}{9})}{z^2 + \frac{5}{6}z + \frac{1}{6}}\xi(t)
\end{aligned}
$$

The assumptions of the spectral factorization theorem are satisfied because the poles are $-\frac{1}{2}, -\frac{1}{3}$ and the zeros are $0, -\frac{1}{9}$ and hence they lie strictly inside the unit-circle.

Determination of the Predictor: Basic Example (cont.)

One-step ahead predictor:

$$
\begin{array}{l|l}
1 + \frac{1}{9}z^{-1} & 1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2} \\ \cline{2-2}
1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2} & 1 \\ \cline{1-1}
-\frac{13}{18}z^{-1} - \frac{1}{6}z^{-2} &
\end{array}
$$

$$\widehat{W}(z) = 1 + z^{-1}\,\frac{-\frac{13}{18} - \frac{1}{6}z^{-1}}{1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}} \implies \widehat{W}_1(z) = \frac{-\frac{13}{18} - \frac{1}{6}z^{-1}}{1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}}$$

Hence:

$$\hat{v}(t+1|t) = -\frac{5}{6}\hat{v}(t|t-1) - \frac{1}{6}\hat{v}(t-1|t-2) - \frac{13}{18}\xi(t) - \frac{1}{6}\xi(t-1)$$

Determination of the Predictor: Basic Example (cont.)

Two-steps ahead predictor:

$$
\begin{array}{l|l}
1+\frac{1}{9}z^{-1} & 1+\frac{5}{6}z^{-1}+\frac{1}{6}z^{-2} \\
\underline{1+\frac{5}{6}z^{-1}+\frac{1}{6}z^{-2}} & \overline{1-\frac{13}{18}z^{-1}} \\
-\frac{13}{18}z^{-1}-\frac{1}{6}z^{-2} & \\
\underline{-\frac{13}{18}z^{-1}-\frac{65}{108}z^{-2}-\frac{13}{108}z^{-3}} & \\
\frac{47}{108}z^{-2}+\frac{13}{108}z^{-3} &
\end{array}
$$

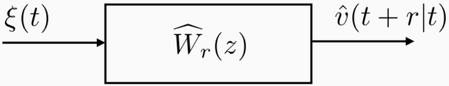
$$\widehat{W}(z) = 1 - \frac{13}{18}z^{-1} + z^{-2}\,\frac{\frac{47}{108}+\frac{13}{108}z^{-1}}{1+\frac{5}{6}z^{-1}+\frac{1}{6}z^{-2}} \implies \widehat{W}_2(z) = \frac{\frac{47}{108}+\frac{13}{108}z^{-1}}{1+\frac{5}{6}z^{-1}+\frac{1}{6}z^{-2}}$$

Hence:

$$\hat{v}(t+2|t) = -\frac{5}{6}\hat{v}(t+1|t-1) - \frac{1}{6}\hat{v}(t|t-2) + \frac{47}{108}\xi(t) + \frac{13}{108}\xi(t-1)$$

Determination of the Predictor from Observed Data

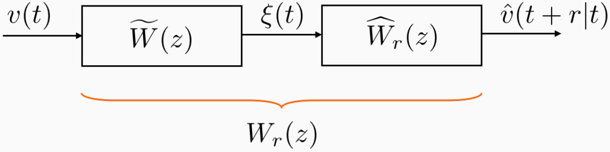
- Starting from the spectral canonical factor $\widehat{W}(z) = \dfrac{C(z)}{A(z)}$ we have obtained the transfer function $\widehat{W}_r(z) = \dfrac{C_r(z)}{A(z)}$ of the r-th steps ahead optimal predictor from the samples of $\xi(t)$:



- However, the process $\xi(t)$ is just a **mathematical abstraction** but certainly it is not a measurable entity. Instead, the goal is to determine a predictor yielding the prediction $\hat{v}(t+r|t)$ using the **measurable** past observations $v(t), v(t-1), v(t-2), \ldots$.

Determination of the Predictor from Observed Data

- Recall that $\mathcal{H}_t[\xi] = \mathcal{H}_t[v]$. Hence, it is sufficient to suitably use the **whitening filter**:



$$\widetilde{W}(z) = \frac{1}{\widehat{W}(z)} = \frac{A(z)}{C(z)} \implies W_r(z) = \frac{A(z)}{C(z)} \frac{C_r(z)}{A(z)} = \frac{C_r(z)}{C(z)}$$

Remark. The additional assumption for which the zeroes of $C(z)$ should lie strictly inside the unit-circle is unavoidable to **guarantee the stability of the predictor**.

Determination of the Predictor: Basic Example (cont.)

Continuing the previous example, the one-step ahead predictor from the observed data is:

$$W_1(z) = \frac{1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}}{1 + \frac{1}{9}z^{-1}} \cdot \frac{-\frac{13}{18} - \frac{1}{6}z^{-1}}{1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}} = \frac{-\frac{13}{18} - \frac{1}{6}z^{-1}}{1 + \frac{1}{9}z^{-1}}$$

and hence

$$\hat{v}(t+1|t) = -\frac{1}{9}\hat{v}(t|t-1) - \frac{13}{18}v(t) - \frac{1}{6}v(t-1)$$

Analogously, the two-steps ahead predictor from the observed data is:

$$W_2(z) = \frac{1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}}{1 + \frac{1}{9}z^{-1}} \cdot \frac{\frac{47}{108} + \frac{13}{108}z^{-1}}{1 + \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}} = \frac{\frac{47}{108} + \frac{13}{108}z^{-1}}{1 + \frac{1}{9}z^{-1}}$$

and hence

$$\hat{v}(t+2|t) = -\frac{1}{9}\hat{v}(t+1|t-1) + \frac{47}{108}v(t) + \frac{13}{108}v(t-1)$$

Solution of the Prediction Problem

Prediction Errors

Prediction Errors

One has:

$$\begin{aligned}\varepsilon(t) &= v(t+r) - \hat{v}(t+r\,|\,t)\\ &= \hat{w}(0)\xi(t+r) + \hat{w}(1)\xi(t+r-1) + \cdots + \hat{w}(r-1)\xi(t+1)\end{aligned}$$

Notice that $\varepsilon(t)$ is a $MA(r)$ process. Therefore:

- $$\begin{aligned}E[\varepsilon(t)] = \hat{w}(0)E[\xi(t+r)] + \hat{w}(1)E[\xi(t+r-1)]\\ + \cdots + \hat{w}(r-1)E[\xi(t+1)] = 0\end{aligned}$$

- $$\text{var}\left[\varepsilon(t)\right] = \left[\hat{w}(0)^2 + \hat{w}(1)^2 + \cdots + \hat{w}(r-1)^2\right]\ \lambda^2$$

Remark. The variance of the prediction error **increases as** r **increases** and asymptotically converges to the variance of the process $v(t)$ (the variance is finite thanks to the stability assumption).

Solution of the Prediction Problem

A Key Example

A notable/Key Example

We want to solve the prediction problems for a generic $AR(1)$ process.

$$v(t) = av(t-1) + \xi(t)\,, \quad \xi(\cdot) \sim WN(0, \lambda^2)\,, \quad |a| < 1$$

Hence:

$$\left(1 - az^{-1}\right)\, v(t) = \xi(t) \implies v(t) = \frac{1}{1 - az^{-1}}\xi(t) = \frac{1}{A(z)}\xi(t)$$

Since $|a| < 1$, it follows that $\widehat{W}(z) = \dfrac{1}{A(z)}$ is a canonical factor.

A notable/Key Example (cont.)

Then:

$$
\begin{array}{l|l}
z & z-a \\
\underline{z-a} & \overline{1 + az^{-1} + a^2 z^{-2} + \cdots} \\
\quad a & \\
\quad \underline{a - a^2 z^{-1}} & \\
\qquad a^2 z^{-1} & \\
\qquad \underline{a^2 z^{-1} - a^3 z^{-2}} & \\
\qquad\qquad a^3 z^{-2} & \\
\qquad\qquad \ddots &
\end{array}
$$

$$\widehat{W}(z) = \frac{1}{A(z)} = \frac{z}{z-a} = 1 + az^{-1} + a^2 z^{-2} + \cdots + z^{-r} \frac{a^r z}{z-a}$$

Hence:

$$\widehat{W}_r(z) = \frac{a^r z}{z-a} = \frac{a^r}{1 - az^{-1}} \Longrightarrow \hat{v}(t+r\,|t) = a\hat{v}(t+r-1\,|t-1) + a^r \xi(t)$$

$$W_r(z) = \frac{C_r(z)}{C(z)} = \frac{a^r z}{z} = a^r \Longrightarrow \hat{v}(t+r\,|t) = a^r v(t)$$

A notable/Key Example (cont.)

- The outcome for which $\hat{v}(t+r\,|t) = a^r\, v(t)$ is not surprising: we have the process $v(t) = av(t-1) + \xi(t)$ and hence it is reasonable that the one-step ahead prediction of $v(t+1)$ is $av(t)$ as, at time t, a white noise is added to $v(t)$.
- Notice that $\hat{v}(t+r\,|t) = a^r\, v(t) \longrightarrow 0$ for $r \to \infty$. This is consistent with $E[v(t)] = 0$ and then, for $r \to \infty$, the prediction has to coincide with the expected value of the process
- Prediction error variance:

$$
\begin{aligned}
\varepsilon(t) &= v(t+r) - \hat{v}(t+r\,|\,t) \\
&= a^r v(t) + \xi(t+r) + a\xi(t+r-1) + \cdots + a^{r-1}\xi(t+1) - a^r v(t) \\
&= \xi(t+r) + a\xi(t+r-1) + \cdots + a^{r-1}\xi(t+1)
\end{aligned}
$$

A notable/Key Example (cont.)

- Therefore, the prediction error is a $MA(r-1)$ process for which

$$\text{var}\left[\varepsilon(t)\right] = \left[1 + a^2 + a^4 + \cdots + a^{2(r-1)}\right] \lambda^2$$

and hence the variance of the prediction error grows with respect to r.

- Moreover:

$$\lim_{r\to\infty} \text{var}\left[\varepsilon(t)\right] = \frac{\lambda^2}{1-a^2} = \text{var}\left[v(t)\right]$$

because

$$\text{var}\left[v(t)\right] = E\,\left[\left[v(t)^2\right]\right] = E\left\{\left[\sum_{i=0}^{\infty} \hat{w}(i)\,\xi(t-i)\right]^2\right\}$$

$$= \sum_{i=0}^{\infty} \hat{w}(i)^2\, E\left[\xi(t-i)^2\right] = \lambda^2 \sum_{i=0}^{\infty} \hat{w}(i)^2 = \lambda^2\,\frac{1}{1-a^2}$$

Solution of the Prediction Problem

One-step Ahead Prediction for ARMA Processes

One-step Ahead Prediction for ARMA Processes

- Consider the process $ARMA(n_a, n_c)$, $\xi(\cdot) \sim WN(0, \lambda^2)$:

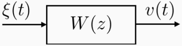
$$
\begin{aligned}
v(t) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) \\
+\xi(t) + c_1 \xi(t-1) + c_2 \xi(t-2) + \cdots + c_n \xi(t-n)
\end{aligned}
$$

Hence:

$$A(z)v(t) = C(z)\xi(t)$$

with

$$
\begin{aligned}
A(z) &= 1 - a_1 z^{-1} - \cdots - a_{n_a} z^{-n_a} \\
C(z) &= 1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}
\end{aligned}
$$



$$W(z) = \frac{C(z)}{A(z)} = \frac{1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}}{1 - a_1 z^{-1} - \cdots - a_{n_a} z^{-n_a}}$$

Setting $n = \max(n_a, n_c)$: $W(z) = \dfrac{z^n + c_1 z^{n-1} + \cdots + c_{n_c} z^{n-n_c}}{z^n - a_1 z^{n-1} - \cdots - a_{n_a} z^{n-n_a}}$

One-step Ahead Prediction for ARMA Processes (cont.)

- Assume that the zeros and poles of $W(z)$ are different from each other and that they all lie strictly inside the unit circle
- Since we are determining the one-step ahead predictor, we get:

$$
\begin{array}{l|l}
C(z) & A(z) \\ \cline{2-2}
A(z) & 1 \\ \cline{1-1}
C(z) - A(z) &
\end{array}
$$

Thus:

$$\frac{C(z)}{A(z)} = 1 + \frac{C(z) - A(z)}{A(z)} = 1 + z^{-1} \, \frac{z\left[C(z) - A(z)\right]}{A(z)}$$

and hence

$$\widehat{W}_1(z) = \frac{z\left[C(z) - A(z)\right]}{A(z)}$$

$$W_1(z) = \frac{z\left[C(z) - A(z)\right]}{C(z)}$$

One-step Ahead Prediction for ARMA Processes (cont.)

- Since $A(z)$ and $C(z)$ are monic, in $C(z) - A(z)$ the constant term is missing:

$$C(z) - A(z) = \left(1 + c_1 z^{-1} + \cdots + c_n z^{-n}\right) - \left(1 - a_1 z^{-1} - \cdots - a_n z^{-n}\right)$$
$$= (c_1 + a_1) z^{-1} + \cdots + (c_n + a_n) z^{-n}$$

Hence:

$$\begin{aligned} C(z)\,\hat{v}(t+1\,|t) &= [C(z) - A(z)]\, z\, v(t) \\ &= [C(z) - A(z)]\, v(t+1) \\ &= \left[(c_1 + a_1) z^{-1} + \cdots + (c_n + a_n) z^{-n}\right]\, v(t+1) \\ &= (c_1 + a_1) v(t) + (c_2 + a_2) v(t-1) + \cdots + (c_n + a_n) v(t-n+1) \end{aligned}$$

and then:

$$\begin{aligned} \hat{v}(t+1\,|t) = &-c_1 \hat{v}(t\,|t-1) - c_2 \hat{v}(t-1\,|t-2) \cdots - c_n \hat{v}(t-n+1\,|t-n) \\ &+ (c_1 + a_1) v(t) + (c_2 + a_2) v(t-1) + \cdots + (c_n + a_n) v(t-n+1) \end{aligned}$$

Remark. Stability of the predictor guaranteed because zeros of $C(z)$ are assumed to lie inside the unit circle

One-step Ahead Prediction for ARMA Processes (cont.)

Alternative Procedure

$$A(z)v(t) = C(z)\xi(t)$$

- Add and subtract to the right-hand side the term $C(z)v(t)$:

$$\begin{aligned} A(z)v(t) &= C(z)\xi(t) + C(z)v(t) - C(z)v(t) \\ \implies\ C(z)v(t) &= [C(z) - A(z)]\, v(t) + C(z)\, \xi(t) \\ \implies\ v(t) &= \frac{[C(z) - A(z)]}{C(z)}\, v(t) + \xi(t) \quad (\star) \end{aligned}$$

- But $\dfrac{[C(z) - A(z)]}{C(z)} = \# z^{-1} + \# z^{-2} + \cdots$ and hence $v(t)$ in $(\star)$ is a function of $v(t-1), v(t-2), \ldots$
- Moreover $\xi(t)$ is uncorrelated with the past of $v(t)$. Then:

$$\hat{v}(t \,|\, t-1) = \frac{[C(z) - A(z)]}{C(z)}\, v(t)$$

where $\xi(t)$ has been dropped since it is uncorrelated with the first term and it is unpredictable from the past

Solution of the Prediction Problem

Prediction in Presence of External Inputs

Prediction in Presence of External Inputs

- First, consider the simple case

$$v(t) = av(t-1) + u + \xi(t)\,, \quad |a| < 1\,, \quad \xi(\cdot) \sim WN(0, \lambda^2)$$

where u is **constant**, **known**, and **deterministic**.

- Clearly:

$$E[v(t)] = aE[v(t-1)] + u + E[\xi(t)]$$
$$\implies (1-a)\,E[v(t)] = u \implies E[v(t)] = \frac{u}{1-a}$$

- Set $\bar{v} = \dfrac{u}{1-a}$ and $\tilde{v}(t) = v(t) - \bar{v}$. Then:

$$\begin{aligned}\tilde{v}(t) = v(t) - \bar{v} &= av(t-1) + u + \xi(t) - \bar{v} \\ \implies \tilde{v}(t) &= av(t-1) - a\bar{v} + u + \xi(t) + (a-1)\bar{v} \\ &= a\tilde{v}(t-1) + u + \xi(t) + (a-1)\bar{v} \\ &= a\tilde{v}(t-1) + \xi(t)\end{aligned}$$

Prediction in Presence of External Inputs (cont.)

- Let us write the process in terms of "variations":

$$\tilde{v}(t) = a\tilde{v}(t-1) + \xi(t)$$

This process is $AR(1)$ and hence:

$$\hat{\tilde{v}}(t \,|\, t-1) = a\tilde{v}(t-1)$$

But $v(t) = \tilde{v}(t) + \bar{v}$ and thus:

$$\begin{aligned} \hat{v}(t \,|\, t-1) = \hat{\tilde{v}}(t \,|\, t-1) + \bar{v} &= a\tilde{v}(t-1) + \bar{v} \\ &= a[v(t-1) - \bar{v}] + \bar{v} \\ &= av(t-1) + u \end{aligned}$$

To sum-up:

the one-step ahead predictor can be obtained by adding the known external input to the predictor obtained without considering the external input

Prediction in Presence of External Inputs (cont.)

- Let us generalize (without proof) to the case of ARMAX models:

$$A(z)v(t) = B(z)u(t) + C(z)\xi(t)$$

with:

$$\begin{aligned} A(z) &= 1 - a_1 z^{-1} - \cdots - a_n z^{-n} \\ B(z) &= b_1 z^{-1} + \cdots + b_n z^{-n} \\ C(z) &= 1 + c_1 z^{-1} + \cdots + c_n z^{-n} \end{aligned}$$

The one-step ahead predictor can be obtained by adding the known (deterministic or not) external term $B(z)u(t)$ to the predictor obtained without considering the external input:

$$\begin{aligned} \hat{v}(t+1\,|t) = & -c_1\hat{v}(t\,|t-1) - c_2\hat{v}(t-1\,|t-2) \cdots - c_n\hat{v}(t-n+1\,|t-n) \\ & +(c_1+a_1)v(t) + (c_2+a_2)v(t-1) + \cdots + (c_n+a_n)v(t-n+1) \\ & +b_1u(t) + b_2u(t-1) + \cdots + b_nu(t-n+1) \end{aligned}$$

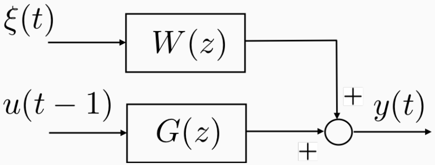
Models and Predictors

Models and Predictors

- Consider the general model

$$\mathcal{M}(\vartheta): \quad y(t) = G(z)\,u(t-1) + W(z)\,\xi(t)$$

where ϑ denotes a **vector of parameters characterizing the model** in which the one-step delay between input and output is explicitly enhanced (a widely used convention)



Models and Predictors (cont.)

- Let us determine the optimal predictor:

$$
\begin{aligned}
& y(t) = G(z)\,u(t-1) + W(z)\,\xi(t) \\
& \implies \frac{1}{W(z)}\,y(t) = \frac{G(z)}{W(z)}\,u(t-1) + \xi(t) \\
& \implies y(t) + \frac{1}{W(z)}\,y(t) = y(t) + \frac{G(z)}{W(z)}\,u(t-1) + \xi(t) \\
& \implies y(t) = \left[1 - \frac{1}{W(z)}\right]\,y(t) + \frac{G(z)}{W(z)}\,u(t-1) + \xi(t)
\end{aligned}
$$

- But $W(z)$ is monic and hence $1 - \frac{1}{W(z)} = \#z^{-1} + \#z^{-2} + \cdots$.
  Therefore, $\left[1 - \frac{1}{W(z)}\right]\,y(t)$ depends on $y(t-1), y(t-2), \ldots$.
- Moreover, $\frac{G(z)}{W(z)}\,u(t-1)$ depends on $u(t-1), u(t-2), \ldots$

Models and Predictors (cont.)

- Therefore, since $\xi(t)$ is white, the class of optimal predictors $\widehat{\mathcal{M}}(\vartheta)$ associated with the class of models $\mathcal{M}(\vartheta)$ is:

$$\widehat{\mathcal{M}}(\vartheta): \quad \hat{y}(t\,|\,t-1) = \left[1 - \frac{1}{W(z)}\right] y(t) + \frac{G(z)}{W(z)}\, u(t-1)$$

  where the optimality stems from the fact that the prediction error

$$\hat{\varepsilon}(t) = y(t) - \hat{y}(t\,|\,t-1) = \xi(t)$$

  is white (zero expected value and variance equal to the variance of $\xi(t)$).

- Let us now consider another predictor $\widetilde{\mathcal{M}}(\vartheta)$ with a white prediction error $\tilde{\varepsilon}(t)$ with zero expected value. Assume that $\widetilde{\mathcal{M}}(\vartheta)$ is “better” than $\widehat{\mathcal{M}}(\vartheta)$, that is

$$\mathrm{var}\left[\tilde{\varepsilon}(t)\right] < \mathrm{var}\left[\hat{\varepsilon}(t)\right]$$

Models and Predictors (cont.)

- But:

$$\begin{aligned}\tilde{\varepsilon}(t) = y(t) - \tilde{y}(t\,|\,t-1) &= y(t) - \hat{y}(t\,|\,t-1) + \hat{y}(t\,|\,t-1) - \tilde{y}(t\,|\,t-1) \\ &= \xi(t) + \hat{y}(t\,|\,t-1) - \tilde{y}(t\,|\,t-1)\end{aligned}$$

On the other hand, $\widehat{\mathcal{M}}(\vartheta)$ and $\widetilde{\mathcal{M}}(\vartheta)$ are predictors and hence:

  - $\hat{y}(t\,|\,t-1)$ depends on $y(t-1), y(t-2), \ldots$
  - $\tilde{y}(t\,|\,t-1)$ depends on $y(t-1), y(t-2), \ldots$

Therefore $\hat{y}(t\,|\,t-1) - \tilde{y}(t\,|\,t-1)$ is uncorrelated with $\xi(t)$ and hence

$$\begin{aligned}\mathrm{var}[\tilde{\varepsilon}(t)] &= \mathrm{var}[\xi(t) + \hat{y}(t\,|\,t-1) - \tilde{y}(t\,|\,t-1)] \\ &= \mathrm{var}[\xi(t)] + \mathrm{var}[\hat{y}(t\,|\,t-1) - \tilde{y}(t\,|\,t-1)] \\ &\geq \mathrm{var}[\xi(t)] = \mathrm{var}[\hat{\varepsilon}(t)]\end{aligned}$$

which **contradicts** the assumption $\mathrm{var}\left[\tilde{\varepsilon}(t)\right] < \mathrm{var}\left[\hat{\varepsilon}(t)\right]$ hence proving that $\widehat{\mathcal{M}}(\vartheta)$ is optimal.

Models and Predictors (cont.)

Summing up:

The model and its associated predictor

$$\mathcal{M}(\vartheta): \quad y(t) = G(z)\,u(t-1) + W(z)\,\xi(t)$$

$$\Longrightarrow \quad \widehat{\mathcal{M}}(\vartheta): \quad \hat{y}(t\,|\,t-1) = \left[1 - \frac{1}{W(z)}\right] y(t) + \frac{G(z)}{W(z)}\,u(t-1)$$

$\widehat{\mathcal{M}}(\vartheta)$ is called **model in prediction form**.

Models and Predictors

Predictors for ARX Models

Predictors for ARX Models

$$\mathcal{M}(\vartheta): \quad A(z)\,y(t) = B(z)\,u(t-1) + \xi(t)$$

$$\Longrightarrow \quad G(z) = \frac{B(z)}{A(z)} \qquad W(z) = \frac{1}{A(z)} \qquad \vartheta = \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Then:

$$\begin{aligned} \hat{y}(t\,|\,t-1) &= \left[1 - \frac{1}{W(z)}\right] y(t) + \frac{G(z)}{W(z)}\,u(t-1) \\ &= [1 - A(z)]\; y(t) + B(z)\,u(t-1) \\ &= a_1\,y(t-1) + a_2\,y(t-2) + \cdots + a_n\,y(t-n) \\ &\quad + b_1\,u(t-1) + b_2\,u(t-2) + \cdots + b_n\,u(t-n) \end{aligned}$$

Observe that $\hat{y}(t\,|\,t-1)$ does not depend on its past values, that is, **the predictor is not dynamic and hence it is always stable**

Models and Predictors

Predictors for ARMAX Models

Predictors for ARMAX Models

$$\mathcal{M}(\vartheta): \quad A(z)\, y(t) = B(z)\, u(t-1) + C(z)\, \xi(t)$$

$$\Longrightarrow \quad G(z) = \frac{B(z)}{A(z)} \qquad W(z) = \frac{C(z)}{A(z)} \qquad \vartheta = \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \\ c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Then:

$$\begin{aligned} \hat{y}(t\,|\,t-1) &= \left[1 - \frac{1}{W(z)}\right] y(t) + \frac{G(z)}{W(z)}\, u(t-1) \\ &= \left[1 - \frac{A(z)}{C(z)}\right] y(t) + \frac{B(z)}{C(z)}\, u(t-1) \\ &= \left[\frac{C(z) - A(z)}{C(z)}\right] y(t) + \frac{B(z)}{C(z)}\, u(t-1) \end{aligned}$$

Predictors for ARMAX Models (cont.)

Hence:

$$\begin{aligned}\hat{y}(t\,|t-1) = -c_1\hat{y}(t-1\,|t-2) - c_2\hat{y}(t-2\,|t-3)\cdots - c_n\hat{y}(t-n\,|t-n-1)\\ +(c_1+a_1)y(t-1) + (c_2+a_2)y(t-2) + \cdots + (c_n+a_n)y(t-n)\\ +b_1u(t-1) + b_2u(t-2) + \cdots + b_nu(t-n)\end{aligned}$$

Observe that $\hat{y}(t\,|t-1)$ now depends on its past values, that is, **the predictor is dynamic**.

Therefore, its stability depends on the position in the complex plane of the zeroes of $C(z)$

Models and Predictors

Predictors for MA Models

Predictors for MA Models

$$\mathcal{M}(\vartheta): \quad y(t) = C(z)\,\xi(t)$$

$$\implies \quad G(z) = 0 \qquad W(z) = C(z) \qquad \vartheta = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Then:

$$\begin{aligned}
\hat{y}(t\,|\,t-1) &= \left[1 - \frac{1}{W(z)}\right] y(t) + \frac{G(z)}{W(z)}\,u(t-1) \\
&= \left[1 - \frac{1}{C(z)}\right] y(t) = \left[\frac{C(z)-1}{C(z)}\right] y(t) \\
&= -c_1\hat{y}(t-1\,|t-2) - c_2\hat{y}(t-2\,|t-3) \cdots - c_n\hat{y}(t-n\,|t-n-1) \\
&\quad +c_1 y(t-1) + c_2 y(t-2) + \cdots + c_n y(t-n)
\end{aligned}$$

Analogously to the ARMAX case, observe that $\hat{y}(t\,|\,t-1)$ depends on its past values, that is, **the predictor is dynamic**.

Therefore, its stability depends on the position in the complex plane of the zeroes of $C(z)$.

Models and Predictors

Predictors for ARXAR Models

Predictors for ARXAR Models

$$\mathcal{M}(\vartheta): \quad A(z)\,y(t) = B(z)\,u(t-1) + \frac{1}{D(z)}\xi(t)$$

$$\Longrightarrow \quad G(z) = \frac{B(z)}{A(z)} \qquad W(z) = \frac{1}{A(z)D(z)} \qquad \vartheta = \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \\ d_1 \\ \vdots \\ d_n \end{bmatrix}$$

Then:

$$\begin{aligned} \hat{y}(t\,|\,t-1) &= \left[1 - \frac{1}{W(z)}\right] y(t) + \frac{G(z)}{W(z)}\,u(t-1) \\ &= [1 - A(z)D(z)]\; y(t) + B(z)D(z)\,u(t-1) \end{aligned}$$

Analogously to the ARX case, **the predictor is not dynamic and hence it is always stable**

Models and Predictors

Concluding Remarks

Models and Predictors: Remarks

- All models in prediction form $\widehat{\mathcal{M}}(\vartheta)$ **depend linearly** on $y(t)$ and $u(t)$
- In general, the stability of the model in prediction form $\widehat{\mathcal{M}}(\vartheta)$ has **nothing to do** with the stability of the associated model $\mathcal{M}(\vartheta)$: for all considered models, the stability depends on the zeroes of $A(z)$ (**poles of the model**) whereas, for the models in prediction form $\widehat{\mathcal{M}}(\vartheta)$, the stability depends on the zeroes of $C(z)$ (**poles of the model in prediction form**)
- Consider the ARX model in prediction form:

$$\begin{aligned}\hat{y}(t\,|\,t-1) = a_1\, y(t-1) + a_2\, y(t-1) + \cdots + a_n\, y(t-n) \\ + b_1\, u(t-1) + b_2\, u(t-1) + \cdots + b_n\, u(t-n)\end{aligned}$$

  Hence, $\hat{y}(t\,|\,t-1)$ **depends linearly on the parameters** a_i, b_i. This property is typically exploited in the identification algorithms

Models and Predictors: Remarks (cont.)

- Consider the ARXAR model in prediction form:

$$\hat{y}(t \,|\, t-1) = [1 - A(z)D(z)]\; y(t) + B(z)D(z)\, u(t-1)$$

Hence:

  - For a given $D(z)$, $\hat{y}(t \,|\, t-1)$ **depends linearly on the parameters** a_i, b_i
  - For given $A(z), B(z)$, $\hat{y}(t \,|\, t-1)$ **depends linearly on the parameters** d_i

This property is typically exploited in the identification algorithms

Models and Predictors: Remarks (cont.)

- On the other hand, consider a first-order ARMAX model in prediction form:

$$\hat{y}(t\,|\,t-1) = \left[\frac{C(z)-A(z)}{C(z)}\right] y(t) + \frac{B(z)}{C(z)}\, u(t-1)$$

But:

$$\left[\frac{C(z)-A(z)}{C(z)}\right] = \frac{(a+c)z^{-1}}{1+cz^{-1}} = (a+c)z^{-1} - c(a+c)z^{-2} + \cdots$$

$$\frac{B(z)}{C(z)} = \frac{b}{1+cz^{-1}} = b - cbz^{-1} + \cdots$$

Hence, $\hat{y}(t\,|\,t-1)$ depends in a **nonlinear** on the parameters a_i, b_i, c_i.

This nonlinear dependence will make the identification algorithms much more complicated

267MI –Fall 2019

Lecture 10
Solution of the Prediction Problem

END