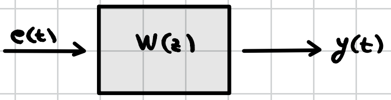
04/03 - Prediction

It is possible to represent a SP in several way, but these representation are not unique. For a given SSP $y(t)$ there may exist infinitely many $W(z)$ and $e(t)$ that reproduce $y(t)$ as output



Case 1: Consider a simple process $y(t) = W(z)e(t)$ with $e(t) \sim WN(0,\lambda^2)$. We can write, for $\alpha \in \mathbb{R}$, that

$$y(t) = \frac{\alpha}{\alpha} W(z) e(t)$$

Define $\tilde{W}(z) = 1/\alpha \, W(z)$ and $\tilde{e}(t) = \alpha e(t)$ that $\tilde{e}(t) \sim WN(0, \alpha^2 \lambda^2)$. So I can write the process as

$$y(t) = \tilde{W}(z)\tilde{e}(t)$$

Case 2: Consider a simple process $y(t) = W(z)e(t)$ with $e(t) \sim WN(0,\lambda^2)$. We can write

$$y(t) = \frac{z^n}{z^n} W(z) e(t)$$

Define $\tilde{W}(z) = z^n W(z)$ and $\tilde{e}(t) = z^{-1} e(t) = e(t-1)$ so that $\tilde{e}(t) \sim WN(0,\lambda^2)$. So I can write the process as

$$y(t) = \tilde{W}(z)\tilde{e}(t)$$

Case 3: Consider a simple process $y(t) = W(z)e(t)$ with $e(t) \sim WN(0,\lambda^2)$. We can write, for $p \in \mathbb{C}$, $|p| < 1$,

$$y(t) = W(z) \frac{z-p}{z-p} e(t)$$

Define $\tilde{W}(t) = W(z)\, {}^{z-p}/_{z-p}$ (obtaining an higher degree polynom - different from the original but same dynamic) and $\tilde{e}(t) = e(t)$. So I can write the process as $y(t) = \tilde{W}(z)\tilde{e}(t)$

Case 4: Consider a simple process $y(t) = W(z)e(t)$ with $e(t) \sim WN(0,\lambda^2)$.

We can write $W(z)$ as $W(z) = W_1(z)(z-9)$ so $y(t) = W_1(z)\left(z-\frac{1}{9}\right)\frac{z-9}{z-1/9} e(t)$ (high pass filter). Define $\tilde{W}(t) = W_1(z)\left(z-\frac{1}{9}\right)$ and $\tilde{e}(t) = {}^{z-9}/_{z-1/9} \cdot e(t)$ (which is still a white noise). So I can write the process as $y(t) = \tilde{W}(t)\tilde{e}(t)$

Proof: $\tilde{e}(t) = \frac{z-9}{z-1/9} e(t)$ is a white noise

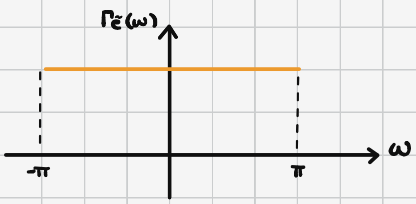
We know that $\tilde{e}(t)$ is computed as output of $G(z)$, so the spectrum of the process $\tilde{e}(t)$ will be defined as

$$\Gamma_{\tilde{e}}(\omega) = \Gamma_e(\omega)\,|G(e^{j\omega})|^2 = \Gamma_e(\omega)\left|e^{j\omega}-9/e^{j\omega}-1/9\right|^2$$

$$= \frac{(e^{j\omega}-9)(e^{-j\omega}-9)}{(e^{j\omega}-1/9)(e^{-j\omega}-1/9)}\lambda^2$$

$$= \frac{1+9^2-9(e^{j\omega}+e^{-j\omega})}{1+1/9^2-1/9(e^{j\omega}+e^{-j\omega})}\lambda^2 = \frac{1+9^2-9(2\cos\omega)}{1+1/9^2-1/9(2\cos\omega)}\lambda^2$$

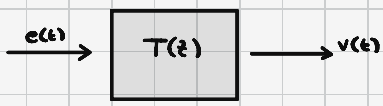
$$= \frac{9^2\cancel{(1/9^2+1-1/9(2\cos\omega))}}{\cancel{1+1/9^2-1/9(2\cos\omega)}}\lambda^2 = 9^2\lambda^2 \quad \forall\omega$$



So the shape is again flat, as happened for the white noise. So yes, $\tilde{e}(t)$ is a white noise with scaled variance

$$T = \frac{z-9}{z-1/9}\cdot\frac{1}{9}$$

is called ALL-PASS FILTER (all frequencies are buffered).



$$\Gamma_y(\omega) = \underbrace{|T(e^{j\omega})|^2}_{1}\Gamma_e(\omega) = \Gamma_e(\omega)$$

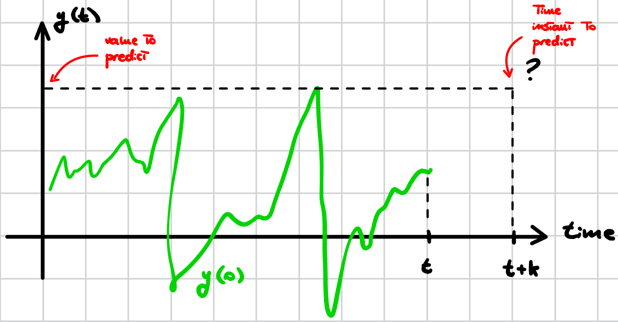
$y(t)$ and $e(t)$ are equivalent (not equal): $e(t) \sim WN(0,\lambda^2)$ $\quad y(t) \sim WN(0,\lambda^2)$

In the following we will need to define any SSP in a unique way. The following result allows us to select the so called CANONICAL REPRESENTATION.

THEOREM: Let $y(t)$ be a SSP with $\Gamma_y(\omega)$ rational. There exists a unique pair $(W(z), e(t))$ if and only if, given $W(z) = C(z)/A(z)$

1) $C(z)$ and $A(z)$ are monic $\left(C(z) = 1 + c_1 z^{-1} + \dots \quad A(z) = 1 + a_1 z^{-1} + \dots \right)$
2) $C(z)$ and $A(z)$ have null relative degree $(\nu = m - n$ ~ this means that ν = #poles - #zeroes$)$
3) $C(z)$ and $A(z)$ are coprime (no common factor)
4) $C(z)$ and $A(z)$ have roots inside the unit circle

Given a SSP $y(t)$ in canonical form and a specific realization $y(0), \dots, y(t)$, how can we predict the future value of the process at a generic time $t+k$?



$$y(t) = \frac{C(z)}{A(z)} e(t) \qquad e(t) \sim WN(\mu, \lambda^2) \qquad C(z), A(z), \mu, \lambda^2$$

We denote $\hat{y}(t+k|t)$ the PREDICTOR at time $t+k$, given t.
(or $\hat{y}(t|t-1)$ the predictor at time t given a past sample at $t-k$)

Example of predictors:

1) $y(t) = W(z)\, e(t) = \dfrac{C(z)}{A(z)} e(t) \Rightarrow m_y = \dfrac{C(1)}{A(1)} m_e$

2) $\hat{y}(t+k|t) = y(t)$ or $\hat{y}(t+k|t) = \frac{1}{3}\left(y(t) + y(t-1) + y(t-2)\right)$ or $\hat{y}(t+k|t) = \frac{1}{3}\left(2y(t) + \frac{1}{2}y(t-1) + \frac{1}{2}y(t-2)\right)$

Under some assumption, they are all good "naive" predictors, but:
1) uses only information about the model ($c_1, \dots, c_n, a_1, \dots, a_m$ are all model parameters)
2) uses only the data

We would like $\hat{y}(t+k|t)$ to be "optimal" in some sense (so we need an optimality criteria) and we would like to exploit all the information we have, so both model and data values

Model $\quad \xrightarrow{e(t)} \boxed{W(z)} \rightarrow y(t)$

past data $\quad \dots, y(t-100), \dots, y(t)$

$\rightarrow \hat{y}(t+k|t)$ predictor

How can we measure the quality of a predictor?

MEAN SQUARE prediction error: $\qquad J(\hat{y}) = E\left[\left(y(t+k) - \hat{y}(t+k|t)\right)^2\right] = E\left[\varepsilon(t+k|t)^2\right]$

which predictor $\hat{y}(t+k|t)$ minimizes $J(\hat{y})$? The predictor must be a function of both model information and data

$$\hat{y}(t+k|t) = f(\text{model}, \text{data}) \qquad \text{HARD TO SOLVE (in the general setting)}$$

Since we work with linear models, we will focus on linear predictors

Linear predictors:

$$\hat{y}(t+k|t) = \alpha_0 y(t) + \alpha_1 y(t-1) + \dots \overset{\tau \to -\infty}{=} \sum_{0}^{100} \alpha_i\, y(t-i) \qquad \alpha_i \in \mathbb{R} \text{ such that } \sum \alpha_i < \infty$$

where
- $\alpha_0, \alpha_1, \dots$ are parameters about the model
- $y(t), y(t-1)$ refers to the available data

$$\xrightarrow{y(t)} \boxed{F_\alpha(z)} \longrightarrow \hat{y}(t+k|t)$$

Using the backward-shift operator, the predictor will be written as

$$\hat{y}(t+k|t) = (\alpha_0 + \alpha_1 z^{-1} + \dots)\, y(t) = F_\alpha(z)\, y(t)$$

The goal now is to find $\alpha_0, \alpha_1, \dots$ of the linear predictor $\hat{y}(t+k|t, \alpha) = \sum \alpha_i\, y(t-i)$ so that the mean square prediction error $J(\alpha) = E\left[\left(y(t+k) - \hat{y}(t+k|t, \alpha)\right)^2\right]$ is minimized.

$$\hat{\alpha} = \arg\min_{\alpha} E\left[\left(y(t+k|t) - \hat{y}(t+k|t, \alpha)\right)^2\right]$$

Solution of the design problem: Consider a simple process $y(t) = W(z)e(t)$ with $e(t) \sim WN(0,\lambda^2)$. We can write the ARMA $y(t)$ into an $MA(\infty)$

$$y(t) = w_0 e(t) + w_1 e(t-1) + w_2 e(t-2) + \dots = \sum_0^{+\infty} w_i\, e(t-i)$$

but at time $t-1$, the process will simply be the time translation of $y(t)$

$$y(t-1) = w_0 e(t-1) + w_1 e(t-2) + \dots = \sum_0^{\infty} w_i\, e(t-i-1)$$

At the same time, the predictor can be written as

$$\hat{y}(t+k|t) = \alpha_0 \sum_0^{+\infty} w_i\, e(t-i) + \alpha_1 \sum_0^{+\infty} w_i\, e(t-i-1) + \dots$$

So $\hat{y}$ is written as past values of $e(t)$ (past values of white noise)

$$\hat{y}(t+k|t) = \beta_0 e(t) + \beta_1 e(t-1) + \dots = \sum_0^{+\infty} \beta_i\, e(t-i)$$

Reformulate the optimization problem with respect to α in terms of a new optimization problem with respect to β.

$$V(\beta) = E[(y(t+k) - \hat{y}(t+k|t,\beta))^2]$$

and $\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, V(\beta)$. In this way, it will be easier to find the solution. We can write the prediction at time $t+k$ as

$$y(t+k) = \sum_0^{+\infty} w_i\, e(t-i+k) = \underbrace{\boxed{\sum_{j=0}^{k-1} w_j\, e(t+k-j)}}_{\text{combination of future sample of } e(t)} + \underbrace{\boxed{\sum_{i=0}^{+\infty} w_{i+k}\, e(t-i)}}_{\text{past / present value of } e(t)}$$

The optimization problem $V(\beta)$ in this way become

$$
\begin{aligned}
V(\beta) &= E[(y(t+k) - \hat{y}(t+k|t))^2] \\
&= E\Big[\Big(\sum_{j=0}^{k-1} w_j\, e(t+k-j) + \sum_{i=0}^{+\infty} w_{k+i}\, e(t-i) - \sum_{i=0}^{+\infty} \beta_i\, e(t-i)\Big)^2\Big] \\
&= E\Big[\underbrace{\cancel{\Big(\sum_{j=0}^{k-1} w_j\, e(t+k-j)\Big)^2}}_{\text{not function of } \beta \Rightarrow \text{ offset}} + \Big(\sum_{i=0}^{+\infty} (w_{k+i} - \beta_i)\, e(t-i)\Big)^2 + \underbrace{\cancel{2\Big(\sum_{j=0}^{k-1} w_j e(t+k-j)\Big)\Big(\sum_{i=0}^{+\infty} (w_{k+i} - \beta_i)\, e(t-i)\Big)}}_{0}\Big]
\end{aligned}
$$

all $e(t)$'s are uncorrelated at different time instant (never coincide in time instant ⇒ always 0)

The optimal predictor is given by $\quad E\Big[\sum_0^{+\infty} ((w_{k+i} - \beta_i)\, e(t-i))^2\Big] \geq 0 \iff \beta_i = w_{k+i} \quad \forall i = 0, \dots$

optimal linear predictor: From the noise ⇒ we need $e(t)$ and infinite amount of coefficient (not computable)

$$\hat{y}(t+k|t) = \sum_0^{+\infty} w_{k+i}\, e(t-i)$$