Linear time series

TSIA202

Definition 1.3.2 (Strict stationarity). *Set $T = \mathbb{Z}$ or $T = \mathbb{N}$. A random process $(X_t)_{t\in T}$ is strictly stationary if X and $\mathrm{S}\circ X$ have the same law, i.e. $\mathbb{P}^{\mathrm{S}\circ X} = \mathbb{P}^X$.*

Example 1.3.1 (I.i.d process). *Let $(Z_t)_{t\in T}$ be a sequence of* independent and identically distributed *(i.i.d) with values in $\mathbb{R}^d$. Then $(Z_t)_{t\in T}$ is a strictly stationary process , since, for all finite set $I = \{t_1, < t_2 < \cdots < t_n\}$ and all Borel set $A_1, \ldots, A_n$ of $\mathbb{R}^d$, we have*

$$\mathbb{P}(Z_{t_1} \in A_1, \cdots, Z_{t_n} \in A_n) = \prod_{j=1}^{n} \mathbb{P}(Z_0 \in A_j) ,$$

which does not depend on $t_1, \ldots, t_n$. Observe that, from Example 1.2.1, for all probability ν on $\mathbb{R}^d$, we can define a random process $(Z_t)_{t\in T}$ which is i.i.d. with marginal distribution *ν, that is, such that $Z_t \sim \nu$ for all $t \in T$.*

Example 1.3.2 (Moving transformation of an i.i.d. process). *Let Z be an i.i.d. process (see Example 1.3.1). Let k be an integer and g a measurable function from $\mathbb{R}^k$ to $\mathbb{R}$. One can check that the process $(X_t)_{t\in\mathbb{Z}}$ defined by*

$$X_t = g(Z_t, Z_{t-1}, \cdots, Z_{t-k+1})$$

also is a stationary random process in the strict sense. On the other hand, the obtained process is not i.i.d. in general since for $k \geq 1$, $X_t, X_{t+1}, \ldots, X_{t+k-1}$ are identically distributed but are in general dependent variables as they all depend on the same random variables Z_t. Nevertheless such a process is said to be k-dependent *because $(X_s)_{s\leq t}$ and $(X_s)_{s>t+k}$ are independent for all t.*

Definition 2.1.1 (L^2 Processes). *The process $\mathbf{X} = (X_t)_{t\in T}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{C}^d$ is an L^2 process if $\mathbf{X}_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ for all $t \in T$.*

The *mean function* defined on T by $\boldsymbol{\mu}(t) = \mathbb{E}[\mathbf{X}_t]$ takes its values in $\mathbb{C}^d$ and the *covariance function* is defined on $T \times T$ by

$$\Gamma(s,t) = \mathrm{Cov}(\mathbf{X}_s, \mathbf{X}_t) = \mathbb{E}\left[(\mathbf{X}_s - \boldsymbol{\mu}(s))(\mathbf{X}_t - \boldsymbol{\mu}(t))^H\right] ,$$

Proposition 2.1.1. *Let Γ be the covariance function of a L^2 process $\mathbf{X} = (\mathbf{X}_t)_{t\in T}$ with values in $\mathbb{C}^d$. The following properties hold.*

(i) Hermitian symmetry: for all $s, t \in T$,

$$\Gamma(s,t) = \Gamma(t,s)^H \tag{2.1}$$

(ii) Nonnegativity: for all $n \geq 1$, $t_1, \ldots, t_n \in T$ and $a_1, \cdots, a_n \in \mathbb{C}^d$,

$$\sum_{1\leq k,m\leq n} a_k^H \Gamma(t_k, t_m) a_m \geq 0 \tag{2.2}$$

Conversely, if Γ satisfy these two properties, there exists an L^2 process $\mathbf{X} = (\mathbf{X}_t)_{t\in T}$ with values in $\mathbb{C}^d$ with covariance function Γ.

Definition 2.2.1 (Weakly stationary processes). *Let $\boldsymbol{\mu} \in \mathbb{C}^d$ and $\Gamma : \mathbb{Z} \to \mathbb{C}^{d\times d}$. A process $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ with values in $\mathbb{C}^d$ is said* weakly stationary *with mean $\boldsymbol{\mu}$ and autocovariance function Γ if all the following assertions hold:*

(i) $\mathbf{X}$ is an L^2 process, i.e. $\mathbb{E}\left[|\mathbf{X}_t|^2\right] < +\infty$,

(ii) for all $t \in \mathbb{Z}$, $\mathbb{E}[\mathbf{X}_t] = \boldsymbol{\mu}$,

(iii) for all $(s,t) \in \mathbb{Z}\times\mathbb{Z}$, $\mathrm{Cov}(\mathbf{X}_s, \mathbf{X}_t) = \Gamma(s-t)$.

Proposition 2.2.1. *The autocovariance function $\gamma : \mathbb{Z} \to \mathbb{C}$ of a complex valued weakly stationary process satisfies the following properties.*

(i) Hermitian symmetry : for all $s \in \mathbb{Z}$,

$$\gamma(-s) = \overline{\gamma(s)}$$

(ii) Nonnegative definiteness : for all integer $n \geq 1$ and $a_1, \cdots, a_n \in \mathbb{C}$,

$$\sum_{s=1}^{n} \sum_{t=1}^{n} \overline{a_s} \gamma(s-t) a_t \geq 0$$

The autocovariance matrix Γ_n of n consecutive samples $X_1, \dots, X_n$ of the time series has a particular structure, namely it is constant on its diagonals, $(\Gamma_n)_{ij} = \gamma(i-j)$,

$$\Gamma_n^+ = \mathrm{Cov}([X_1 \ \dots \ X_n]^T)$$
$$= \begin{bmatrix} \gamma(0) & \gamma(-1) & \cdots & \gamma(1-n) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(2-n) \\ \vdots & & & \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix} \tag{2.3}$$

Definition 2.2.3 (White noise). *A weak white noise is a centered weakly stationary process whose autocovariance function satisfies $\gamma(0) = \sigma^2 > 0$ and $\gamma(s) = 0$ for all $s \neq 0$. We will denote $(X_t) \sim \mathrm{WN}(0, \sigma^2)$. When a weak white noise is an i.i.d. process, it is called a* strong white noise. *We will denote $(X_t) \sim \mathrm{IID}(0, \sigma^2)$.*

Of course a strong white noise is a weak white noise. However the converse is in general not true. The two definitions only coincide for Gaussian processes because in this case the independence is equivalent to being uncorrelated.

Example 2.2.1 (MA(1) process). *Define, for all $t \in \mathbb{Z}$,*

$$X_t = Z_t + \theta Z_{t-1} \, , \tag{2.4}$$

where $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$ and $\theta \in \mathbb{R}$. Then $\mathbb{E}[X_t] = 0$ and the autocovariance function reads

$$\gamma(s) = \begin{cases} \sigma^2(1+\theta^2) & \text{if } s = 0, \\ \sigma^2\theta & \text{if } s = \pm 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Such a weakly stationary process is called a Moving Average of order 1 MA(1).

Example 2.2.2 (Harmonic process). *Let $(A_k)_{1 \leq k \leq N}$ be N real valued L^2 random variables. Denote $\sigma_k^2 = \mathbb{E}\left[A_k^2\right]$. Let $(\Phi_k)_{1 \leq k \leq N}$ be N i.i.d. random variables with a uniform distribution on $[-\pi, \pi]$, and independent of $(A_k)_{1 \leq k \leq N}$. Define*

$$X_t = \sum_{k=1}^{N} A_k \cos(\lambda_k t + \Phi_k) \, , \tag{2.6}$$

where $(\lambda_k)_{1 \leq k \leq N} \in [-\pi, \pi]$ are N frequencies. The process (X_t) is called an harmonic process. *It satisfies $\mathbb{E}[X_t] = 0$ and, for all $s, t \in \mathbb{Z}$,*

$$\mathbb{E}[X_s X_t] = \frac{1}{2} \sum_{k=1}^{N} \sigma_k^2 \cos(\lambda_k (s-t)) \, .$$

It is thus a weakly stationary process.

Example 2.2.3 (Random walk). *Let (S_t) be a random process defined on $t \in \mathbb{N}$ by $S_t = X_0 + X_1 + \cdots + X_t$, where (X_t) is a strong white noise. Such a process is called a* random walk. *We have $\mathbb{E}[S_t] = 0$, $\mathbb{E}\left[S_t^2\right] = t\sigma^2$ and for all $s \leq t \in \mathbb{N}$,*

$$\mathbb{E}[S_s S_t] = \mathbb{E}[(S_s + X_{s+1} + \cdots + X_t) S_s] = s \, \sigma^2$$

The process (S_t) is not weakly stationary.

Definition 2.3.1. *The* empirical mean *(or* sample mean*) and the* empirical autocovariance function *of the sample $X_{1:n}$ are respectively defined as*

$$\widehat{\mu}_n = \frac{1}{n}\sum_{t=1}^{n} X_t \tag{2.8}$$

$$\widehat{\gamma}_n(h) = \begin{cases} n^{-1}\sum_{t=1}^{n-h}(X_{t+h}-\widehat{\mu}_n)\overline{(X_t-\widehat{\mu}_n)} & \text{if } 0 \le h \le n-1\,, \\ n^{-1}\sum_{t=1-h}^{n}(X_{t+h}-\widehat{\mu}_n)\overline{(X_t-\widehat{\mu}_n)} & \text{if } 0 \le -h \le n-1\,, \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

Theorem 2.4.1 (Herglotz). *A sequence $(\gamma(h))_{h\in\mathbb{Z}}$ is a nonnegative definite hermitian sequence in the sense of Proposition 2.2.1 if and only if there exists a finite nonnegative measure ν on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ such that :*

$$\gamma(h) = \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}h\lambda}\nu(\mathrm{d}\lambda),\ \forall h \in \mathbb{Z}\,. \tag{2.12}$$

Moreover this relation defines ν uniquely.

Corollary 2.4.2 (The ℓ^1 case). *Let $(\gamma(h))_{h\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$. Then it is a nonnegative definite hermitian sequence in the sense of Proposition 2.2.1 if and only if*

$$f(\lambda) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \gamma(h)\mathrm{e}^{-\mathrm{i}h\lambda} \ge 0\,,$$

for all $\lambda \in \mathbb{T}$. Moreover, in the case where this condition holds, f is the spectral density function associated to γ.

Example 2.4.1 (MA(1), Continued from Example 2.2.4). *Consider Example 2.2.4. Then $(\chi(h))$ is in $\ell^1(\mathbb{Z})$ and*

$$f(\lambda) = \frac{1}{2\pi}\sum_{h} \chi(h)\mathrm{e}^{-\mathrm{i}h\lambda} = \frac{1}{2\pi}(1+2\rho\cos(\lambda))\,.$$

Thus we obtain that χ is nonnegative definite if and only if $|\rho| \le 1/2$. An exemple of such a spectral density function is displayed in Figure 2.3.

Example 2.4.2 (Spectral density function of a white noise). *Recall the definition of a white noise, Definition 2.2.3. We easily get that the white noise* $\mathrm{IID}(0,\sigma^2)$ *admits a spectral density function given by*

$$f(\lambda) = \frac{\sigma^2}{2\pi}\,,$$

that is, a constant spectral density function. Hence the name "white noise", referring to white color that corresponds to a constant frequency spectrum.

Example 2.4.3 (Spectral measure of an harmonic process, continued from Example 2.2.2). *The autocovariance function of X is given by (see Example 2.2.2)*

$$\gamma(h) = \frac{1}{2}\sum_{k=1}^{N} \sigma_k^2\cos(\lambda_k h)\,, \tag{2.14}$$

where $\sigma_k^2 = \mathbb{E}\left[A_k^2\right]$. Observing that

$$\cos(\lambda_k h) = \frac{1}{2}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}h\lambda}(\delta_{\lambda_k}(\mathrm{d}\lambda) + \delta_{-\lambda_k}(\mathrm{d}\lambda))$$

where $\delta_{x_0}(\mathrm{d}\lambda)$ denote the Dirac mass at point x_0, the spectral measure of X reads

$$\nu(\mathrm{d}\lambda) = \frac{1}{4}\sum_{k=1}^{N}\sigma_k^2\delta_{\lambda_k}(\mathrm{d}\lambda) + \frac{1}{4}\sum_{k=1}^{N}\sigma_k^2\delta_{-\lambda_k}(\mathrm{d}\lambda)\,.$$

We get a sum of Dirac masses with weights σ_k^2 and located at the frequencies of the harmonic functions.

Definition 2.4.1 (Linearly predictable processes). *A weakly stationary process X is called* linearly predictable *if there exists $n \ge 1$ such that for all $t \ge n$, $X_t \in \mathrm{Span}\,(X_1, \ldots, X_n)$ (in the L^2 sense).*

Proposition 2.4.3. *Let γ be the autocovariance function of a weakly stationary process X. If $\gamma(0) \neq 0$ and $\gamma(t) \to 0$ as $t \to \infty$ then X is not linearly predictable.*

Definition 2.5.1 (Periodogram)**.** *The periodogram of the sample $X_{1:n}$ is the function valued in $\mathbb{C}$ and defined on $\mathbb{T}$ by*

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} (X_t - \widehat{\mu}_n) \mathrm{e}^{-\mathrm{i}t\lambda} \right|^2 . \tag{2.15}$$

$$\begin{aligned} \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}\lambda h} I_n(\lambda) \mathrm{d}\lambda &= \frac{1}{n} \sum_{s=1}^{n} \sum_{t=1}^{n} (X_s - \widehat{\mu}_n) \overline{(X_t - \widehat{\mu}_n)} \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\lambda(h-s+t)} \\ &= \widehat{\gamma}_n(h) , \end{aligned}$$

3.1 Linear filtering using absolutely summable coefficients

Let $\psi = (\psi_t)_{t\in\mathbb{Z}}$ be an absolutely summable sequence of $\mathbb{C}^{\mathbb{Z}}$, we will write $\psi \in \ell^1(\mathbb{Z})$, or simply $\psi \in \ell^1$.

In this section we consider the linear filter defined by

$$\mathrm{F}_\psi : x = (x_t)_{t\in\mathbb{Z}} \mapsto y = \psi \star x , \tag{3.1}$$

where $\star$ denotes the convolution product on sequences, that is, for all $t \in \mathbb{Z}$,

$$y_t = \sum_{k\in\mathbb{Z}} \psi_k x_{t-k} . \tag{3.2}$$

We introduce some usual terminology about such linear filters.

Definition 3.1.1. *We have the following definitions.*

(i) If ψ is finitely supported, F_ψ is called a finite impulse response (FIR) *filter.*

(ii) If $\psi_t = 0$ for all $t < 0$, F_ψ is said to be causal.

(iii) If $\psi_t = 0$ for all $t \geq 0$, F_ψ is said to be anticausal.

Of course (3.2) is not always well defined. In fact, F_ψ is well defined only on

$$\ell_\psi = \left\{ (x_t)_{t\in\mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}} \ : \ \text{for all } t \in \mathbb{Z}, \sum_{k\in\mathbb{Z}} |\psi_k \, x_{t-k}| < \infty \right\} .$$

Moreover we observe that a FIR filter can be written as

$$\mathrm{F}_\psi = \sum_{k\in\mathbb{Z}} \psi_k \, \mathrm{B}^k , \tag{3.3}$$

where B is the Backshift operator of Definition 1.3.1. This sum is well defined for a finitely supported ψ since it is a finite sum of linear operators.

An immediate consequence of this result is that F_ψ applies to any weakly stationary process and its output is also weakly stationary.

Theorem 3.1.2. *Let $\psi \in \ell^1$ and $X = (X_t)_{t\in\mathbb{Z}}$ be a weakly stationary process with mean μ, autocovariance function γ and spectral measure ν. Then $\mathrm{F}_\psi(X)$ is well defined and is a weakly stationary process with mean*

$$\mu' = \mu \sum_{t\in\mathbb{Z}} \psi_t , \tag{3.7}$$

autocovariance function given for all $h \in \mathbb{Z}$ by

$$\gamma'(h) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \psi_j \overline{\psi}_k \, \gamma_X(h + k - j) , \tag{3.8}$$

and spectral measure ν' defined as the measure with density $|\psi^*(\lambda)|^2$ with respect to ν, where

$$\psi^*(\lambda) = \sum_{t\in\mathbb{Z}} \psi_t \mathrm{e}^{-\mathrm{i}t\lambda} . \tag{3.9}$$

Definition 3.3.1 (MA(q) processes). *A random process $X = (X_t)_{t\in\mathbb{Z}}$ is called a* moving average *process of order q (MA(q)) with coefficients $\theta_1, \dots, \theta_q$ if it satisfies the MA(q) equation*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} , \tag{3.17}$$

where $Z \sim \mathrm{WN}(0, \sigma^2)$.

In other word $X = \mathrm{F}_\alpha(Z)$, where F_α is a FIR filter with coefficients

$$\alpha_t = \begin{cases} 1 & \text{if } t = 0, \\ \theta_k & \text{if } t = 1, \dots, q, \\ 0 & \text{otherwise.} \end{cases} \tag{3.18}$$

Equivalently, we can write

$$X = [\Theta(\mathrm{B})](Z) ,$$

where B is the Backshift operator and Θ is the polynomial defined by $\Theta(z) = 1 + \sum_{k=1}^{p} \theta_k z^k$.

Hence it is a linear process with short memory, and by Corollary 3.1.3, it is a centered weakly stationary process with autocovariance function given by

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} \overline{\theta_k}\theta_{k+h}, & \text{if } 0 \le h \le q , \\ \sigma^2 \sum_{k=0}^{q+h} \overline{\theta}_k\theta_{k-h}, & \text{if } -q \le h \le 0 , \\ 0, & \text{otherwise ,} \end{cases} \tag{3.19}$$

and with spectral density function given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 + \sum_{k=1}^{q} \theta_k \mathrm{e}^{-\mathrm{i}k\lambda} \right|^2 .$$

Definition 3.2.2 (All-pass filters). *Let $\psi \in \ell^1$. The linear filter F_ψ is called an all-pass filter if there exists $c > 0$ such that, for all z on the unit circle Γ_1,*

$$\left| \sum_{k\in\mathbb{Z}} \psi_k z^k \right| = c .$$

An interesting obvious property of these filters is the following.

Lemma 3.2.5. *Let $\psi \in \ell^1$ such that F_ψ is an all-pass filter. Then if Z is a weak white noise, so is $\mathrm{F}_\psi(Z)$.*

Definition 3.3.2 (AR(p) processes). *A random process $X = (X_t)_{t\in\mathbb{Z}}$ is called an* autoregressive *process of order p (AR(p)) with coefficients $\phi_1, \dots, \phi_p$ if it satisfies the AR(p) equation*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t \,, \tag{3.20}$$

where $Z \sim \mathrm{WN}(0, \sigma^2)$.

Theorem 3.3.1 (Existence and uniqueness of a weakly stationary solution of the AR(p) equation). *Let $Z \sim \mathrm{WN}(0, \sigma^2)$ with $\sigma^2 > 0$ and $\phi_1, \dots, \phi_p \in \mathbb{C}$. Define the polynomial*

$$\Phi(z) = 1 - \sum_{k=1}^{p} \phi_k z^k \,.$$

Then the AR(p) equation (3.20) has a unique weakly stationary solution X if and only if Φ does not vanish on the unit circle $\mathbb{U}$. Moreover, in this case, we have $X = \mathrm{F}_\psi(Z)$, where $\psi \in \ell^1$ is uniquely defined by

$$\sum_{t\in\mathbb{Z}} \psi_t z^t = \frac{1}{\Phi(z)} \quad \text{on} \quad z \in \mathbb{U} \,.$$

Let us just mention that it easily follows from our result on the inversion of FIR filters (see Corollary 3.2.4) by observing that, as for MA processes, the AR(p) equation can be interpreted as a FIR filter equation, namely, $Z = \mathrm{F}_\beta(X)$, where F_β is a FIR filter with coefficients

$$\beta_t = \begin{cases} 1 & \text{if } t = 0, \\ -\phi_t & \text{if } t = 1, \dots, p, \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

Or, equivalently, $Z = [\Phi(\mathrm{B})](X)$.

Definition 3.3.3 (ARMA(p, q) processes). *A random process $X = (X_t)_{t\in\mathbb{Z}}$ is called an* autoregressive moving average *process of order (p, q) (ARMA(p, q)) with AR coefficients $\phi_1, \dots, \phi_p$ and MA coefficients $\theta_1, \dots, \theta_q$ if it satisfies the ARMA(p, q) equation*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \,, \tag{3.25}$$

where $Z \sim \mathrm{WN}(0, \sigma^2)$.

Before stating this result, let us recall how the ARMA equation can be rewritten using linear filer operators. The ARMA(p, q) equation can be written as

$$\Phi(\mathrm{B})(X) = \Theta(\mathrm{B})(Z) \,, \tag{3.26}$$

where B is the Backshift operator and Φ and Θ are the polynomials defined by

$$\Phi(z) = 1 - \sum_{k=1}^{p} \phi_k z^k \quad \text{and} \quad \Theta(z) = 1 + \sum_{k=1}^{p} \theta_k z^k \,. \tag{3.27}$$

Theorem 3.3.2 (Existence and uniqueness of a weakly stationary solution of the ARMA(p, q) equation). *Let $Z \sim \mathrm{WN}(0, \sigma^2)$ with $\sigma^2 > 0$ and $\phi_1, \dots, \phi_p$, $\theta_1, \dots, \theta_q \in \mathbb{C}$. Assume that the polynomials Φ and Θ defined by (3.27) have no common roots. Then the ARMA(p, q) equation (3.20) has a unique weakly stationary solution X if and only if Φ does not vanish on the unit circle $\mathbb{U}$. Moreover, in this case, we have $X = \mathrm{F}_\psi(Z)$, where $\psi \in \ell^1$ is uniquely defined by*

$$\sum_{t\in\mathbb{Z}} \psi_t z^t = \frac{\Theta}{\Phi}(z) \quad \text{on} \quad z \in \mathbb{U} \,. \tag{3.28}$$

As a consequence, X admits a spectral density function given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\Theta}{\Phi}(\mathrm{e}^{-\mathrm{i}\lambda}) \right|^2 \,. \tag{3.29}$$

Definition 3.4.1 (Representations of ARMA(p,q) processes). *If the ARMA equation (3.25) has a weakly stationary solution $X = \mathrm{F}_\psi(Z)$, it is said to provide*

(i) a causal representation *of X if F_ψ is a causal filter,*

(ii) an invertible representation *of X if $\mathrm{F}_\psi(Z)$ is an invertible representation and its inverse filter is causal,*

(iii) a canonical representation *of X if $\mathrm{F}_\psi(Z)$ is a causal and invertible representation.*

Theorem 3.4.1. *Under the assumptions and notation of Theorem 3.3.2, the ARMA equation (3.25) provides*

(i) a causal representation *of X if and only if Φ does not vanish on the unit closed disk Δ_1,*

(ii) an invertible representation *of X if and only if Θ does not vanish on the unit closed disk Δ_1,*

(iii) a canonical representation *of X if and only if neither Φ nor Θ does vanish on the unit closed disk Δ_1.*

Theorem 3.4.2. *Let X be the weakly stationary solution of the ARMA equation (3.25), where Φ and Θ defined by (3.27) have no common roots and no roots on the unit circles. Then there exists AR coefficients $\tilde\phi_1, \dots, \tilde\phi_p$ and MA coefficients $\tilde\theta_1, \dots, \tilde\theta_q$ and $\tilde Z \sim \mathrm{WN}(0, \sigma^2)$ such that X satisfies the ARMA(p,q) equation*

$$X_t = \tilde\phi_1 X_{t-1} + \dots + \tilde\phi_p X_{t-p} + \tilde Z_t + \tilde\theta_1 \tilde Z_{t-1} + \dots + \tilde\theta_q \tilde Z_{t-q} \,, \tag{3.30}$$

and the corresponding polynomials $\tilde\Phi$ and $\tilde\Theta$ do not vanish on the unit closed disk Δ_1. In particular, (3.30) is a canonical representation of X. Moreover, if the original AR and MA coefficients ϕ_k's and θ_k's are real, so are the canonical ones $\tilde\phi_k$'s and $\tilde\theta_k$'s.

Algorithm 1: Computation of the filter coefficients and the autocovariance function from a causal ARMA representation.

Data: AR and MA coefficients $\phi_1, \dots, \phi_r$, $\theta_1, \dots, \theta_r$, and variance σ^2 of the white noise.

Result: Causal filter coefficients $(\psi_k)_{k\geq 0}$ and autocovariance function γ.

Step 1 Initialization: set $\psi_0 = 1$.

for $k = 1, 2, \dots, r$ **do**

Compute

$$\psi_k = \theta_k + \sum_{j=1}^{k} \psi_{k-j}\phi_j \,. \tag{3.32}$$

end

for $k = r+1, r+2, \dots$ **do**

Compute

$$\psi_k = \sum_{j=1}^{r} \psi_{k-j}\phi_j \,. \tag{3.33}$$

end

Step 2 **for** $\tau = 0, 1, 2, \dots$ **do**

Compute

$$\gamma(\tau) = \sigma^2 \sum_{k=0}^{\infty} \overline{\psi_k}\psi_{k+\tau} \,. \tag{3.34}$$

end

and **for** $\tau = -1, -2, \dots$ **do**

Set

$$\gamma(\tau) = \overline{\gamma(-\tau)} \,.$$

end

Theorem 3.5.1. *Let X be the weakly stationary solution of the ARMA(p,q) equation (3.31), which is assumed to be a causal representation, that is, for all $z \in \mathbb{C}$ such that $|z| \leq 1$,*

$$1 - \sum_{k=1}^{p} \phi_k z^k \neq 0 \,.$$

Define $r = \max(p,q)$ and set $\theta_j = 0$ for $q < j \leq r$ or $\phi_j = 0$ for $p < j \leq r$. Then Algorithm 1 applies.

Algorithm 2: Computation of the autocovariance function from a causal ARMA representation.

Data: AR and MA coefficients $\phi_1, \ldots, \phi_r$, $\theta_1, \ldots, \theta_r$, and variance σ^2 of the white noise, a lag m.

Result: Causal filter coefficients ψ_k for $k = 0, \ldots, r$ and autocovariance function $\gamma(\tau)$ for $\tau = -m, \ldots, m$.

Step 1 Initialization: set $\psi_0 = 1$.

for $k = 1, 2, \ldots, r$ **do**

Compute ψ_k by applying (3.32).

end

Step 2 Using that $\gamma(-j) = \overline{\gamma(j)}$ for all j and setting $\theta_0 = 1$, solve the linear system

$$\gamma(\tau) - \phi_1\gamma(\tau - 1) - \cdots - \phi_r\gamma(\tau - r) = \sigma^2 \sum_{\tau \le j \le r} \theta_j \overline{\psi}_{j-\tau}, \quad 0 \le \tau \le r , \tag{3.35}$$

in $\gamma(\tau)$, $\tau = 0, 1, 2, \ldots, r$.

Step 3 Then apply the following induction.

for $\tau = r + 1, r + 2, \ldots, m$ **do**

Compute

$$\gamma(\tau) = \phi_1\gamma(\tau - 1) + \cdots + \phi_r\gamma(\tau - r) . \tag{3.36}$$

end

for $\tau = -1, -2, \ldots, -m$ **do**

Set

$$\gamma(\tau) = \overline{\gamma(-\tau)} .$$

end

Theorem 3.5.2. *Under the same assumptions as Theorem 3.5.1, Algorithm 1 applies.*

Let us define the *linear past* of a process $X = (X_t)_{t\in\mathbb{Z}}$ up to time t by

$$\mathcal{H}_t^X = \overline{\mathrm{Span}}\,(X_s, s \le t) .$$

It is related to the already mentioned space $\mathcal{H}_\infty^X$ as follows

$$\mathcal{H}_\infty^X = \overline{\bigcup_{t\in\mathbb{Z}} \mathcal{H}_t^X} .$$

Definition 4.1.1 (Innovation process)**.** *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a centered weakly stationary process. We call* innovation process *the process $\epsilon = (\epsilon_t)_{t\in\mathbb{Z}}$ defined by*

$$\epsilon_t = X_t - \mathrm{proj}\left(X_t \middle| \mathcal{H}_{t-1}^X \right) . \tag{4.1}$$

By the orthogonal principle of projections in L^2, each ϵ_t is characterized by the fact that $X_t - \epsilon_t \in \mathcal{H}_{t-1}^X$ (which implies $\epsilon_t \in \mathcal{H}_t^X$) and $\epsilon_t \perp \mathcal{H}_{t-1}^X$. As a consequence $(\epsilon_t)_{t\in\mathbb{Z}}$ is a centered orthogonal sequence. We shall see below that it is in fact a white noise, that is, the variance of the innovation

$$\sigma^2 = \|\epsilon_t\|^2 = \mathbb{E}\left[|\epsilon_t|^2\right] \text{ does not depend on } t. \tag{4.2}$$

Example 4.1.1 (Innovation process of a white noise). *The innovation process of a white noise $X \sim \mathrm{WN}(0, \sigma^2)$ is $\epsilon = X$.*

Example 4.1.2 (Innovation process of a MA(1), continued from Example 2.2.1). *Consider the process X defined in Example 2.2.1. Observe that $Z_t \perp \mathcal{H}_{t-1}^X$. Thus, if $\theta Z_{t-1} \in \mathcal{H}_{t-1}^X$, we immediately get that $\epsilon_t = Z_t$. The questions are thus: is Z_{t-1} in $\mathcal{H}_{t-1}^X$? and, if not, what can be done to compute ϵ_t ?*

Because the projection in (4.1) is done on an infinite dimension space, it is interesting to compute it as a limit of finite dimensional projections. To this end, define, for $p \geq 0$, the finite dimensional space

$$\mathcal{H}_{t,p}^X = \mathrm{Span}\left(X_s, t-p < s \leq t\right) ,$$

and observe that $(\mathcal{H}_{t,p}^X)_p$ is an increasing sequence of linear space whose union has closure $\mathcal{H}_t^X$. In this case we have, for any L^2 variable Y,

$$\lim_{p\to\infty} \mathrm{proj}\left(Y \middle| \mathcal{H}_{t,p}^X\right) = \mathrm{proj}\left(Y \middle| \mathcal{H}_t^X\right) , \tag{4.3}$$

where the limit holds in the L^2 sense.

Definition 4.1.2 (Prediction coefficients). *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a centered weakly stationary process. We call the* predictor of order p *the random variable* $\mathrm{proj}\left(X_t \middle| \mathcal{H}_{t-1,p}^X\right)$ *and the* partial innovation process of order p *the process $\epsilon_p^+ = \left(\epsilon_{t,p}^+\right)_{t\in\mathbb{Z}}$ defined by*

$$\epsilon_{t,p}^+ = X_t - \mathrm{proj}\left(X_t \middle| \mathcal{H}_{t-1,p}^X\right) .$$

The prediction coefficients *are any coefficients $\phi_p^+ = \left(\phi_{k,p}^+\right)_{k=1,\dots,p}$ which satisfy, for all $t \in \mathbb{Z}$,*

$$\mathrm{proj}\left(X_t \middle| \mathcal{H}_{t-1,p}^X\right) = \sum_{k=1}^{p} \phi_{k,p}^+ X_{t-k} . \tag{4.4}$$

Observe that, by the orthogonality principle, (4.4) is equivalent to

$$\Gamma_p^+ \phi_p^+ = \gamma_p^+ , \tag{4.5}$$

where $\gamma_p^+ = [\gamma(1), \gamma(2), \cdots, \gamma(p)]^T$ and

$$\begin{aligned}
\Gamma_p^+ &= \mathrm{Cov}\left([X_{t-1} \; \dots \; X_{t-p}]^T\right)^T \\
&= \begin{bmatrix}
\gamma(0) & \gamma(-1) & \cdots & & \gamma(-p+1) \\
\gamma(1) & \gamma(0) & \gamma(-1) & & \vdots \\
\vdots & \ddots & \ddots & \ddots & \\
\vdots & & & & \gamma(-1) \\
\gamma(p-1) & \gamma(p-2) & \cdots & \gamma(1) & \gamma(0)
\end{bmatrix} ,
\end{aligned}$$

Observing that Equation (4.5) does not depend on t and that the orthogonal projection is always well defined, such coefficients $\left(\phi^+_{k,p}\right)_{k=1,\dots,p}$ always exist. However they are uniquely defined if and only if Γ^+_p is invertible.

Let us now compute the variance of the order-p prediction error $\epsilon^+_{t,p}$, denoted as

$$\sigma_p^2 = \|X_t - \mathrm{proj}\left(X_t|\,\mathcal{H}_{t-1,p}\right)\|^2 = \mathbb{E}\left[|X_t - \mathrm{proj}\left(X_t|\,\mathcal{H}_{t-1,p}\right)|^2\right] . \tag{4.6}$$

By (4.4) and the usual orthogonality condition of the projection, we have

$$\begin{aligned}\sigma_p^2 &= \langle X_t, X_t - \mathrm{proj}\left(X_t|\,\mathcal{H}_{t-1,p}\right)\rangle \\ &= \gamma(0) - \sum_{k=1}^{p} \overline{\phi^+_{k,p}}\gamma(k) \\ &= \gamma(0) - (\phi^+_p)^H \gamma^+_p . \end{aligned} \tag{4.7}$$

Equations (4.5) and (4.7) are called *Yule-Walker equations*. An important consequence of these equations is that σ_p^2 does not depend on t, and since (4.3) implies

$$\sigma^2 = \lim_{p\to\infty} \sigma_p^2 ,$$

we obtain that, as claimed above, the variance of the innovation defined in (4.2) is also independent of t. So we can state the following result.

Corollary 4.1.1. *The innovation process of a centered weakly stationary process X is a (centered) weak white noise. Its variance is called the* innovation variance *of the process X.*

The innovation variance is not necessarily positive, that is, the innovation process can be zero a.s., as shown by the following example.

Example 4.1.3 (Innovations of the harmonic process (continued from Example 2.2.2)). *Consider the harmonic process $X_t = A\cos(\lambda_0 t + \Phi)$ where A is a centered random variable with finite variance σ_A^2 and Φ is a random variable, independent of A, with uniform distribution on $(0, 2\pi)$. Then X is a centered weakly stationary process with autocovariance function $\gamma(\tau) = (\sigma_A^2/2)\cos(\lambda_0\tau)$. The prediction coefficients or order 2 are given by*

$$\begin{bmatrix}\phi^+_{1,2}\\ \phi^+_{2,2}\end{bmatrix} = \begin{bmatrix}1 & \cos(\lambda_0)\\ \cos(\lambda_0) & 1\end{bmatrix}^{-1}\begin{bmatrix}\cos(\lambda_0)\\ \cos(2\lambda_0)\end{bmatrix} = \begin{bmatrix}2\cos(\lambda_0)\\ -1\end{bmatrix}$$

We then obtain that $\sigma_2^2 = \|X_t - \mathrm{proj}\left(X_t|\,\mathcal{H}^X_{t-1,2}\right)\|^2 = 0$ and thus

$$X_t = \mathrm{proj}\left(X_t|\,\mathcal{H}^X_{t-1,2}\right) = 2\cos(\lambda_0)X_{t-1} - X_{t-2} \in \mathcal{H}^X_{t-1}$$

Hence in this case the innovation process is zero: one can exactly predict the value of X_t from its past.

Theorem 4.2.1. *Let X be the weakly stationary solution to a canonical ARMA(p,q) equation of the form*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} ,$$

where $Z \sim \mathrm{WN}(0,\sigma^2)$. Then Z is the innovation process of X.

Proof. By definition of the canonical representation, there exists $\psi, \tilde{\psi} \in \ell^1$ such that $\psi_k = \tilde{\psi}_k = 0$ for all $k < 0$, $X = \mathrm{F}_\psi(Z)$ and $Z = \mathrm{F}_{\tilde{\psi}}(X)$. We deduce that, for all $t \in \mathbb{Z}$, $\mathcal{H}^Z_t = \mathcal{H}^X_t$. Consequently, for all $t \in \mathbb{Z}$,

$$\hat{X}_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \in \mathcal{H}^X_{t-1} ,$$

and

$$X_t - \hat{X}_t = Z_t \in \mathcal{H}^Z_t \perp \mathcal{H}^Z_{t-1} = \mathcal{H}^X_{t-1} .$$

Hence, by the orthogonality principle of projection, we obtain that

$$\mathrm{proj}\left(X_t|\,\mathcal{H}^X_{t-1}\right) = \hat{X}_t .$$

Theorem 4.2.2. *Let X be a centered weakly stationary process with autocovariance function γ. Then X is an MA(q) process if and only if $\gamma(h) = 0$ for all $|h| > q$.*

Theorem 4.2.3. *Let X be a weakly stationary AR(p) process with causal representation*

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t \;,$$

where $Z \sim \mathrm{WN}(0, \sigma^2)$. Then, for all $m \geq p$, the prediction coefficients are given by

$$\phi_p^+ = [\phi_1, \ldots, \phi_p, \underbrace{0, \ldots, 0}_{m-p}]^T \;,$$

that is, for all $t \in \mathbb{Z}$,

$$\operatorname{proj}\left(X_t \middle| \mathcal{H}_{t-1,m}^X\right) = \sum_{k=1}^{p} \phi_k X_{t-k} \;. \tag{4.8}$$

In particular the prediction error of order m is Z_t and has variance σ^2 and thus is constant for all $m \geq p$.

Definition 4.2.1 (Partial autocorrelation function)**.** *Let X be a weakly stationary process. The* partial autocorrelation function *of X is the function defined by*

$$\kappa(p) = \phi_{p,p}^+, \quad p = 1, 2, \ldots$$

where $\phi_p^+ = \left(\phi_{k,p}^+\right)_{k=1,\ldots,p}$ denote the prediction coefficients of X, that is, for all $t \in \mathbb{Z}$,

$$\operatorname{proj}\left(X_t \middle| \mathcal{H}_{t-1,p}^X\right) = \sum_{k=1}^{p} \phi_{k,p}^+ X_{t-k} \;,$$

with the convention that $\kappa(p) = 0$ if this equation does not defines uniquely ϕ_p^+, that is, if Γ_p^+ is not invertible.

We see from Theorem 4.2.3 that if X is an AR process, then its partial autocorrelation function vanishes for all $m > p$. It is in fact a characterization of AR processes, as shown by the following result.

Theorem 4.2.4. *Let X be a centered weakly stationary process with partial autocorrelation function κ. Then X is an AR(p) process if and only if $\kappa(m) = 0$ for all $m > p$.*